MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ AMAR TELIDJI -LAGHOUAT

FACULTÉ DES SCIENCES EXACTES ET DES SCIENCES DE LA NATURE ET DE LA VIE

Département de Mathématiques et Informatique

N° d'ordre : ……..

Mémoire

Présenté en vue de l'obtention du diplôme de magister en informatique

Option :

Informatique Répartie et Mobile (IRM)

Thème :

CLASSIFICATION MULTICRITÈRE DE DONNÉES ÉVOLUTIVES : APPLICATION DIAGNOSTIC MÉDICAL

Présenté Par :

ABDELHAMID LEBAL

Devant le jury composé de :

YAGOUBI MOHAMED BACHIR	**PROFESSEUR**	**PRÉSIDENT**	**UNIVERSITÉ DE LAGHOUAT**
MOUSSAOUI ABDELOUAHAB	**PROFESSEUR**	**RAPPORTEUR**	**UNIVERSITÉ DE SÉTIF 1**
CHERROUN HADDA	**MCA**	**EXAMINATRICE**	**UNIVERSITÉ DE LAGHOUAT**
OUINTEN YOUCEF	**PROFESSEUR**	**EXAMINATEUR**	**UNIVERSITÉ DE LAGHOUAT**

Année universitaire 2015-2016

Dédicaces :

Á mes parents.
Á mes frères et sœurs.
Á Ayoub et Aya.

Remerciements :

Tout d'abord, je tiens à remercie chaleureusement, Mr : Abdelouahab MOUSSAOUI, Professeur à l'université Ferhat ABBAS de Sétif-1 pour avoir accepté d'être le rapporteur de ce mémoire. Merci beaucoup pour tous vos conseils, votre soutien, vos encouragements, votre confiance et aussi pour votre sympathie. J'ai vraiment eu beaucoup de plaisir à travailler avec vous.

Je remercie le président du jury Monsieur : Mohamed Bachir YAGOUBI, Professeur à l'université de Laghouat qui nous a fait l'honneur de présider le jury.

Je remercie également Madame : Hadda CHERROUN, Maître de conférences à l'université de Laghouat pour m'avoir fait l'honneur d'évaluer ce travail.

Je remercie aussi Monsieur : Youcef OUINTEN, Professeur à l'université de Laghouat d'avoir accepté d'assurer la tâche d'examinateur et d'avoir consacré leur temps à l'examen de ce mémoire.

ملخـــص:

التصنيف هو العملية المستخدمة لتوزيع البيانات على مجموعات متجانسة تسمى فئات، حيث أنه في الفئة الواحدة تكون البيانات في أقصى تشابه وأقصى اختلاف ممكن مع بيانات الفئات الأخرى. التصنيف مجال نشط في العديد من المجالات الأخرى مثل الإحصاء، التعرف على الأنماط و التعلم الآلي وأيضا في استخراج المعارف من معالجة عدد كبير من البيانات.

وقد تم تطوير العديد من التقنيات لتصنيف البيانات في نظام ثابت، ومع ذلك في الكثير من تطبيقات الحياة سواء كانت طبيعية أو اصطناعية هي غير ثابتة، أي أنها متغيرة مع الزمن.

يعتبر التصنيف الديناميكي ذلك المجال المعقد عند معالجة حجم كبير من البيانات الغير الثابتة. ولتحسين أداء التصنيف الديناميكي، ظهرت فكرة مثيرة للاهتمام في سنوات التسعينيات من القرن الماضي هي منهج التعاون بين العديد من المصنفات ولهذا وجهنا حلنا للتصنيف الديناميكي للبيانات الغير الثابتة إلى منهج التعاون بين المصنفات المختلفة الموجودة في مجال التصنيف المتعدد المعايير وأيضا في مجال التعلم الآلي بالاعتماد على نظرية الاعتقاد للباحثين Dempster-Shafer.

في هذه الدراسة ، نقدم أسلوبا جديدا للتصنيف الديناميكي للبيانات الغامضة والغير الثابتة، هذه الطريقة متكيفة مع البيانات والفئات المتغيرة مع الزمن. هذه الطريقة متكونة من ست مراحل: الحصول على البيانات، معالجة البيانات إذا لزم الأمر،استخراج خصائص البيانات، مرحلة التصنيف تقوم بها عدة مصنفات، دمج نتائج المصنفات باستخدام نظرية الاعتقاد، كشف تغير البيانات والفئات وأخيرا تحديث نموذج التصنيف لتجنب إعادة العملية من البداية.

تم تطبيق الطريقة المقترحة على تصنيف البيانات البيوطبية وهي الرسم الكهربائي للدماغ التي تعتبر غير ثابتة وغير خطية.

الكلمات المفتاحية: ***التصنيف الديناميكي، التصنيف المتعدد المعايير، البيانات الغير ثابتة، الفئات الغير ثابتة، التعاون، التشخيص الطبي، إشارات الرسم الكهربائي للدماغ.***

Résumé :

La classification est le processus qui permet d'affecter les données en groupes homogènes appelés classes. Tels que les objets dans chaque classe ont une similarité maximale et une dissemblance maximale avec les objets d'autres classes. La classification est le domaine de recherche actif dans plusieurs autres sujets de recherche tels que, la reconnaissance des formes, l'apprentissage automatique, les statistiques. La classification est également étudiée dans l'extraction de connaissances où un très grand nombre de données sera traitées.

De nombreuses techniques ont été développées pour la classification des données dans un système statique. Cependant, dans de nombreux processus naturels ou artificiels de la vie sont non-stationnaires, c'est-à-dire variant dans le temps. La classification est considérée comme science difficile lorsque nous traitons un large volume d'informations variables dans le temps.

Pour améliorer la performance de la classification dynamique, une idée intéressante est apparue dans les années 90 consiste à combiner les différents classificateurs. Nous sommes orientés notre solution de la classification dynamique des données évolutives à une architecture de coopération et/ou de coordination de différents classifieurs flous situés dans le domaine d'Aide MultiCritère à la Décision (MCDA) et dans l'apprentissage automatique, cette combinaison est réalisée par la théorie de Dempster-Shafer.

Dans cette étude, nous présentons une nouvelle approche pour la classification dynamique floue de données non-stationnaires basée sur une architecture de fusion de plusieurs classificateurs. Cette approche est adaptative aux données et classes évolutives. L'approche développée est faite en six étapes : acquisition de données, le prétraitement des données si nécessaire, la classification, la fusion des résultats de classifieurs à l'aide de théorie des fonctions de croyance, détection des changements des données et les classes, enfin l'incrémentation du modèle de classification.

L'approche proposée est appliquée à la classification des données biomédicales à l'instar des signaux EEG, réputés être dynamiques et non-linéaires.

Mots-clés : ***la classification dynamique, classification multicritère, données évolutives, classes évolutives, coopération, diagnostic médical, EEG.***

Abstract:

Clustering is the process that classifiers data in homogeneous groups called clusters such as the objects, in each cluster, have a maximum similarity and maximum dissimilarity with the objects of other clusters. Clustering is the subject of active research in several fields such as statistics, pattern recognition, and machine learning. Clustering is also studied in data mining filed where large data sets which many attribute of different type considered.

Numerous techniques have been developed for classification data in static system. However, in many real life applications, non-stationary (i.e. time-varying) data are generally common and class models have to evolve in time. The dynamic classification is seen as difficult science when we process large volumes of information varying over time.

To improve the performance of the dynamic data classification, an interesting idea appeared in 90s consists in combining classifiers for benefits their performances. We are oriented our solution of dynamic classification in evolutionary data to a collaborative architecture based on fuzzy clustering in Machine Learning (ML) classifiers and fuzzy classifiers in Multi-Criteria Decision Aid (MCDA) by Dempster-Shafer theory, where several classifiers corporate with them.

In this study, we present a new approach for fuzzy dynamic classification on non stationary data, based on classifiers cooperation architecture. This approach is adaptive to evolutionary data and classes. The developed approach is made in six stages: data acquisition, data preprocessing, features extraction, modeling data for to construct a matrix of membership, merging the membership degrees using the theory of belief function (Dempster-Shafer theory), detection of changes of data and classes, finally, incremental of model of classification step.

The approach proposed is applied to the classification of EEG signals. These signals characterized by non-linear dynamics.

Key-words: ***Dynamic classification, multi-criteria classification, evolutionary data, evolutionary classes, medical diagnosis, EEG, cooperation.***

TABLE DES MATIÈRES

LISTE DES FIGURES

Liste des Tableaux

INTRODUCTION GÉNÉRALE :

i. Contexte et motivation :

L'extraction de connaissances à partir de données provenant de sources hétérogènes nécessite des techniques plus intelligentes classées dans une nouvelle branche de la recherche scientifique nommée Apprentissage Automatique. Le principe de l'apprentissage automatique est de concevoir des systèmes autonomes capables d'extraire des connaissances utiles à partir de données disponibles servant à la prise de décision. Cependant, la fiabilité des sources de données, la nature des connaissances elles-mêmes ainsi que le choix des techniques à utiliser pour extraire ces connaissances sont les principaux problèmes liés à l'hétérogénéité des sources de données. Très peu de systèmes d'apprentissage sont dotés de cette faculté d'adaptation à l'évolution et la multi-modalité de données. La Classification Dynamique est un nouvel axe de l'apprentissage automatique qui s'intéresse justement à ce type de problèmes. La majorité des algorithmes de classification classique ne prennent pas en considération le caractère évolutif des classes et la multi- modalité des données, ce qui conduit à de mauvais résultats de classifications comme dans beaucoup de domaine (biologie, diagnostic médicale, biotechnologie, reconnaissance de visages, …etc.) où les connaissances évolues dans le temps et où la véracité des connaissances varie en fonction des capteurs qui peuvent être à la fois complémentaires et même parfois contradictoires (exemple : classification de la végétation qui varie en fonction de la saison).

De plus, en aide à la décision multicritère (MCDA ou MultiCriteria Decision Aiding), on suppose que les différentes alternatives se présentant au décideur peuvent être décrites sur un certain nombre de propriétés ou attributs. Si l'on est capable d'établir une échelle de préférence sur un attribut, on parle alors de critère. Les valeurs des alternatives sur ces critères représentent la prise en compte de points de vue diversifiés, en général non réductibles à un seul critère. Ces critères sont souvent conflictuels. La difficulté est d'arriver

à obtenir une comparaison relative des alternatives, afin de guider le choix du décideur vers la solution qui lui paraîtra optimale. En effet, la notion d'optimisation "dans l'absolu" est vide de sens en décision multicritère, car il n'existe généralement pas d'alternative optimisant tous les critères simultanément. Il est donc nécessaire de prendre en compte de l'information supplémentaire, en particulier l'importance relative de chaque critère et les compensations possibles entre les différents critères. La collaboration des classificateurs est une des meilleures solutions pour la classification dynamique des données, où le processus d'apprentissage ainsi que la classification sont partagés entre plusieurs classificateurs qui coopèrent d'une façon intelligente et complètement autonome.

ii. Contribution :

La contribution de ce mémoire comporte les points suivants :

- Proposer de nouvelles méthodes pour la classification dynamique floue de données, tout en exploitant les différentes techniques et algorithmes de classification dynamique utiles pour l'extraction de connaissances à partir de données évolutifs (non-stationnaires) et multimodales. L'aspect multicritères est mis en avant à travers l'application de la méthodologie au diagnostic médical.
- Montrer l'efficacité des techniques d'apprentissage automatique pour le diagnostic médical.
- Proposer une architecture de fusion et/ou de coopération de classifieurs pour surmonter le problème de la multi-modalité, de l'évolutivité des données et de la dynamicité des classes à l'instar des données médicales.

iii. Organisation :

Outre l'introduction générale et la conclusion générale, Ce mémoire est organisé en sept chapitres répartis de la façon suivante :

Chapitre 01 :

Le premier chapitre est dédié à un état de l'art sur le croisement des systèmes d'aide multicritère à la décision (MultiCriteria Decision Aiding, en anglais : MCDA) et le domaine d'apprentissage automatique (Machine Learning, en anglais : ML). Nous

dressons d'abord quelques formulations des systèmes d'aide multicritères à la décision. Nous décrivons, en suite, les méthodes d'agrégation multicritères les plus connues et les plus utilisées. Puis, nous détaillons les méthodes automatiques pour construire les poids des critères.

Chapitre 02 :

Ce chapitre traite les phénomènes engendrés à cause des données évolutives, nous avons illustré le processus général de suivi un système évolutive ainsi, on a montré quelques définitions et caractéristiques des données évolutives. Nous montrons dans ce chapitre les principales problématiques de la classification dynamique. Telle que la modélisation adaptative, apprentissage incrémental, classification supervisé.

Chapitre 03 :

Ce chapitre est consacré à l'identification de quelques algorithmes de classification existent dans la littérature pour construire un classifieur dynamique. Il existe deux grands types de méthodes de classification, il y a les méthodes de classification supervisées et les méthodes de classification non supervisées, ces derniers sont aussi classés selon deux catégories : les méthodes floues et les méthodes non-floues : cependant, en combinant deux types de méthodes de classification multicritères et de classification en apprentissage automatique.

Chapitre 04 :

Dans ce chapitre, nous présentons les fondements théoriques de la fusion de données. En premier temps, nous décrivons les différents formalismes permettant de représenter les connaissances incertaines et/ou imprécises. Nous présentons les trois théories de fusion de données suivantes : la théorie des probabilités, la théorie des possibilités et la théorie des croyances. Ces méthodes sont actuellement considérées comme les plus adaptées à la représentation et au traitement des informations imparfaites. Le traitement des données imparfaites dans ces trois méthodes a été détaillé suivant les quatre étape : la modélisation, l'estimation, la combinaison et la décision.

Chapitre 05 :

Nous présentons dans ce chapitre les fondements terminologies et les connaissances de base sur l'outil de diagnostic médical l'ÉlectroEncéphaloGraphie (EEG), le cerveau humain, les signaux EEG, la maladie d'épilepsie et les différentes techniques de traitement du signal à savoir, l'analyse de Fourier et la transformée en ondelettes.

Chapitre 06 :

Ce chapitre est réservé à présenter en détail notre contribution, nous présenterons l'architecture du système de classification dynamique, où les données peuvent varier dans le temps. Description des algorithmes de classification utilisés et leurs paramètres. Un algorithme générique de fusion proposé, ces composants essentiels et les différentes étapes de fusion dans le cadre de théorie des croyances. Nous avons proposé une architecture de classification coopérative à base des méthodes de classification floues en apprentissage automatique et en aide multicritère à la décision. L'approche développée est composée en six phases : acquisition des données, prétraitement des données, extraction des caractéristiques des données, modélisation des données pour construire une matrice des degrés d'appartenance, fusion les matrices d'appartenance en utilisant la théorie des fonctions des croyances (théorie de dempster-Shafer), la détection des évolutions des données et des classes et la phase l'incrémentation du modèle de classification.

Chapitre 07:

Ce chapitre est réservé à la validation expérimentale. Tout d'abord, l'approche proposée est appliquée à la classification des signaux EEG. Ces signaux sont caractérisés par des dynamiques non linéaires.

Chapitre 01 : Aide à la Décision MultiCritère et Apprentissage Automatique :

I.1. Introduction :

Au cours de ces dernières décennies années est apparu un ensemble de disciplines fortement interdépendantes, portant sur le traitement de l'information, la théorie de la décision et l'apprentissage automatique.

L'apprentissage automatique (*machine learning* en anglais) est scientifique qui fait référence au développement, à l'analyse et à l'implémentation de méthodes qui permettent à une machine d'évoluer grâce à un processus d'apprentissage et ainsi de remplir des tâches qu'il est difficile ou impossible de remplir par des moyens algorithmiques plus classiques.

En plus, l'aide multicritère à la décision ou (MCDA) MultiCriteria Decision Aiding (en anglais), MCDA est un domaine en plein essor dont les développements scientifiques, se situent à la fois dans le champ des sciences fondamentales et appliquées telles que les mathématiques, l'informatique, la recherche opérationnelle, ….etc. Et aussi dans les sciences humaines et de gestion telles que la sociologie, psychologie, management,…etc.

Le champ de l'aide multicritère à la décision propose des manières de modéliser formellement les préférences d'un décideur de façon à lui apporter des éclaircissements. Le champ s'intéresse aux problèmes impliquant une décision et faisant intervenir plusieurs critères, points de vue utilisés pour évaluer les actions (ou alternatives) disponibles.

En aide multicritère à la décision, on distingue généralement trois types de problématiques : les problématiques du choix (α), du rangement (γ) et du Tri (β). La première consiste à sélectionner au sein d'un ensemble d'objets, un sous ensemble considéré comme optimale pour le problème posé. La problématique de tri consiste à formuler le problème en terme d'affectation les objets à des classes prédéfinies. La problématique de rangement consiste à ranger les objets du meilleur à la moins bonne.

Dans ce chapitre nous allons définir quelques notions et concepts relatifs à l'aide à la décision, et les structures de préférences, ainsi que nous aurons montré en détail le fonctionnement de quelques méthodes concernant l'aide multicritère à la décision et nous voulons attirer l'attention de MCDA sur les progrès récents dans l'apprentissage automatique en matière de **modélisation humaine de comportement de préférence**.

I.2. Décision, Aide à la décision, Acteurs:

I.2.1. Décision :

Prendre une décision c'est trancher entre plusieurs possibilités et choisir celle qui sera effectivement mise en œuvre. **Bernard Roy** a donne une définition d'aide à la décision.

I.2.2. Aide à la décision :

Définition: L'aide à la décision est l'activité de celui qui, prenant appui sur des modèles clairement explicités mais non nécessairement clairement formalisés, aide à obtenir des éléments de réponse aux questions que se pose un intervenant dans un processus de décision, éléments concourants à éclairer la décision et normalement à prescrire, ou simplement à favoriser, un comportement de nature à accroître la cohérence entre l'évolution d'un processus d'une part, les objectifs et le système de valeurs au service desquels cet intervenant se trouve placé d'autre part. Donc l'objectif d'aide à la décision va être de guider et d'éclairer une entité appelée *décideur* tout au long de son processus de décision. **[Roy93]**

Au cours de processus d'aide à la décision, on distingue principalement deux intervenants : *l'homme d'étude* et le *décideur*. Roy et al en donnent les définitions suivantes **[ROY93]**:

I.2.3. L'homme d'étude et décideur

Définitions:

Décideur : est l'entité intervenant dans le processus de décision que les modèles mis en œuvre cherchent à éclairer, et aussi de justifier ses décisions. Contrairement à certaines idées reçues le décideur n'est pas nécessairement un individu en particulier. Ce peut être dans certains cas un individu mais ce peut être aussi un groupe d'individus pas nécessairement bien identifié (ex. le gouvernement, les personnes qui le composent peuvent évoluer rapidement sans pour autant empêcher que des décisions soient prises en continu).

Le rôle de décideur :

La responsabilité principale de décideur est d'évaluer les différentes alternatives possibles afin de proposer ou de mettre en œuvre une solution (ou des solutions).

L'homme d'étude ou l'analyste : celui qui conduit l'activité d'aide à la décision.

I.3. Modélisation des préférences :

I.3.1. Le concept d'Action:

Toutes discipline scientifique est amenée à manipuler est à raisonner sur des objets plus au moins bien formalisés. Par exemple, dans le domaine des statistiques, les objets manipulés vont par exemple être des points décrits par des vecteurs [HER00].

I.3.1.1. Définition d'une action :

Dans [VIN89], l'auteur a défini l'ensemble des actions comme suit : «est l'ensemble des objets, décisions, candidats, … que l'on va explorer dans le processus de décision ».

De plus, il convient de différencier les **actions globales** et les **actions fragmentaires**

Action globale: action dont la mise à exécution est exclusive de toute autre action introduite dans le modèle c-à-dire, le choix d'une action unique par exemple : achat d'une voiture, tracé d'autoroute.

Action fragmentaire : dans le cas contraire de précédente, où un groupe d'actions peut-être choisi.

En aide multicritère à la décision, les actions manipulées vont être les actions potentielles.

La construction des actions est la première et parfois la plus difficile des tâches de processus de décision, une action est appelé aussi *Alternative*, solution ou décision. Les caractéristiques des actions sont contingentes à un état d'avancement du processus de décision, c'est-à-dire qu'elle peut être considérée pour elle-même, isolément, sans perdre de sa portée décisionnelle. Une action peut être réelle (déjà réalisée), **réaliste** (envisageable raisonnablement), **fictive** (associée à un projet idéalisé) ou **irréaliste** (associée à un projet irréalisable et/ou satisfaisant des objectifs totalement incompatibles).

I.3.1.2. Action potentielle:

L'action potentielle est une action *réelle* ou *fictive* provisoirement jugé réaliste par un acteur au moins ou présumée telle par l'homme d'étude en vue de l'aide à la décision ;

l'ensemble des actions potentielles sur lequel l'aide à la décision prend appui au cours d'une phase d'étude. Cet ensemble est désigne par $A = (a_1, a_2, \dots, a_i, \dots, a_m)$ où a_i sont les actions potentielles. **[ROY93]**

Dans la suite de ce chapitre, on parle : action potentielle = action = alternative.

I.3.1.3. Action de référence :

Action servant, dans la problématique β de référence par rapport à laquelle les actions potentielles sont examinées. Les actions de référence servent de limites à des catégories auxquelles les actions potentielles sont effectuées. [ROY93]

Etant donnée la complexité des problèmes de décision, il n'est pas toujours possible de définir *a priori* l'ensemble *A*, il arrive même souvent que la définition de *A* se fasse progressivement au cours de la procédure d'aide à la décision, L'ensemble *A* peut être donc : [VIN89]

· **Stable** : s'il est défini *a priori* et n'est pas susceptible d'être changé au cours de la procédure.

· **Évolutif :** *s'il peut être modifié en cours de la procédure, soit à cause des résultats intermédiaires que cette procédure fait apparaître, soit parce que le problème de décision se pose dans un environnement changeant. (Les deux causes peuvent être simultanées).*

I.3.2 Le concept de critère :

I.3.2.1 Définition d'un critère:

Expression qualitative ou quantitative de point de vue objectifs, aptitudes ou contraintes relatives au contexte réel ; permettant de juger des personnes, des objets ou des évènements. Pour qu'une telle expression puisse devenir un critère, elle doit être utile pour le problème considéré et fiable. Un critère est doté d'une **structure de préférence** ; à chaque critère est associée une échelle, en valeurs ordinales ou cardinales [ROY93].

Vincke [VIN89] a donné une définition formelle du critère:

Définition: un critère est une fonction g, définie sur l'ensemble *A* des actions, qui prend ses valeurs dans un ensemble totalement ordonné, et qui représente les préférences du décideur selon un point de vue. Lorsque le problème repose sur la considération de plusieurs critères, nous notons $F = \{g_1, g_2, \dots, g_j, \dots, g_n\}$. Dans la suite de ce chapitre, nous parlerons l'évaluation d'une action *a* suivant le critère *j* est notée $g_j(a)$.

Donc, nous appellerons critère une fonction réelles g qui permet de déterminer le résultat de la comparaison de tous les paires d'actions de sorte que :

$$\begin{cases} g(a) = g(b) \Rightarrow aIb \\ g(a) > g(b) \Rightarrow aPb \end{cases}$$ Modèle de préférence traditionnel.

Afin de bien éclairer la différence entre un critère et un attribut, il nous faut les définir :

Attribut : caractéristique décrivant chaque objet « âge, diplôme, sexe, … ».

Critère : Exprime les préférences du décideur relativement à un point de vue, par exemple « prix d'une voiture, puissance, vitesse, … ». Cette notion intègre la structure de préférence du décideur sur ce critère « maximiser la puissance, maximiser la vitesse, minimiser le prix ».

I.3.2.2 Famille cohérente des critères :

On appelle famille cohérente de critères toute famille $F = \{g_1, g_2, \dots, g_j, \dots, g_n\}$ conforme aux trois conditions suivantes :

1- *Exhaustivité* : il ne faut pas oublier des critères.
2- *Cohérence* : cohérences entre les préférences locales de chaque critère et les préférences globales *i.e.* $\forall j \neq l;\ g_j(a) = g_j(b)\ et\ g_l(a) > g_l(b) \Rightarrow a\ domine\ b.$
3- *Non redondance* : il ne faut pas de critères qui se dupliquent.

I.3.3. Relation binaire et structures de préférence :

I.3.3.1. Relation binaire :

Les études d'aide à la décision se basent sur la comparaison des actions (alternatives). Cette comparaison fait apparaître en diverses situations de préférences qui peuvent être modélisées par des relations binaires.

I.3.3.1.1. Définition d'une relation binaire:

***Définition**: Une relation binaire R sur un ensemble A est un sous-ensemble du produit cartésien AxA. C'est un ensemble de couple (a, b) d'éléments de A.*

*Si le couple (a, b) appartient à l'ensemble R, on notera a R b, sinon on notera ¬a R b et on le lit **non** a **relation avec** b.*

I.3.3.1.2. Propriétés d'une relation binaire :

On a vue que le concept de relation binaire était crucial pour modéliser les préférences. Nous rappelons ici quelques structures remarquables de ces relations ayant un intérêt particulier en modélisation des préférences. [ROY93][BEL00]

- ***Réflexive*** *ssi* : $\forall a \in A, aRa$;
- ***Symétrique*** *ssi* :$\forall a. b \in A, aRb \Rightarrow bRa$; $(ex - aequo)$
- ***Antisymétrique*** *ssi* : $\forall a, b \in A, aRb\ et\ bRa \Rightarrow a = b$;
- ***Asymétrique*** *ssi* : $\forall a, b \in A, aRb \Rightarrow \neg bRa$;
- ***Connexe*** *ssi* **:** $\forall a, b \in A, a \neq b \Rightarrow aRb\ et/ou\ bRa$;
- ***Complète*** *ssi* **:** $\forall a, b \in A, aRb\ et/ou\ bRa\ tel\ que\ a \neq b$;
- ***Transitive*** *ssi* : $\forall a, b, c \in A, aRb\ et\ bRc\ \Rightarrow aRc$;
- ***Négativement transitive*** *ssi*: $\forall a, b, c \in A, \neg aRb\ et\ \neg bRc\ \Rightarrow\ \neg aRc$;
- ***pré-ordre*** : est une relation réflexive est transitive
- ***ordre :*** est un pré-ordre antisymétrique. (c'est-à-dire : dans un pré-ordre, des ex-aequo sont possibles, au contraire dans la relation d'ordre).
- ***Pré-ordre totale :*** pré-ordre dans lequel les éléments sont toujours comparables c'est-à-dire la relation de pré-ordre totale est une relation binaire complète (incomparabilité exclue).
- ***Pré-ordre partiel :*** pré-ordre dans lequel l'incomparabilité est permise.

I.3.3.1.3. Représentation graphique d'une relation binaire :

Une relation binaire R dans A peut être représentée par un graphe orienté (A, R), où A l'ensemble des sommets du graphe et R l'ensemble des arcs du graphe. [VIN89]

Dans la suite du ce chapitre on introduit des méthodes multicritère à la décision qui se basant sur la représentation par graphes d'une relation binaire.

I.3.3.2. Système de préférence :

I.3.3.2.1. Définition d'une relation de préférence :

Définition : Une relation de préférence est une relation binaire sur un ensemble A, que l'on note $\succsim$; $a \succsim b$ est lu "a est préféré à b"

I.3.3.2.2. Les types de la relation de préférence :

Roy [ROY93] a proposé quatre situations fondamentales de préférence. ***Indifférence, préférence stricte, préférence faible*** *et* ***l'incomparabilité.*** Ce sont les quatre relations que l'on retrouve dans la plupart des travaux sur la modélisation des préférences.

- ***Indifférence :*** cette relation traduit une situation dans laquelle il existe des raisons claires et satisfaisantes qui justifient une équivalence entre les deux actions :

  On notera **a∼b** ou **aIb.** Cette relation est généralement considérée comme étant réflexive et symétrique.

$$\mathbf{a} \sim \mathbf{b} \Leftrightarrow [\boldsymbol{a} \gtrsim \boldsymbol{b} \ \ \boldsymbol{et} \ \ \boldsymbol{b} \gtrsim \boldsymbol{a}]$$

- ***Préférence stricte:*** cette relation permet de traduire une situation dans laquelle il existe des raisons claires et suffisante qui justifie une préférence significative entre deux actions.

  On notera **a≻b** ou **aPb,** cette relation est considérée comme étant irréflexive et asymétrique.

  En plus, les relations I et P ne sont pas toujours transitives. [HER00]

$$\mathbf{a} \succ \mathbf{b} \Leftrightarrow [\boldsymbol{a} \gtrsim \boldsymbol{b} \ \ et \ \ \neg \boldsymbol{b} \gtrsim \boldsymbol{a}]$$

- ***Préférence faible :*** l'affirmation *a faiblement préféré à b* traduit le fait que l'homme d'étude juge que « ***b*** n'est pas préféré à ***a*** **»,** tout en hésitant entre « *a* est strictement préféré à *b* » et « *a* indifférent à *b* », les éléments dont il dispose entre ces deux dernières possibilités ne lui paraissant pas suffisamment probants, on notera **a≺b** ou **aQb.** Cette relation est considérée comme étant irréflexive et transitive.
- ***Incomparabilité :*** dans le domaine d'aide multicritère à la décision, il peut exister des situations où le décideur ne peut s'exprimer les trois relations précédente.

  On notera **aRb,** cette relation est considérée comme étant irréfléxive et transitive.

$$\boldsymbol{a}\mathbf{R}\boldsymbol{b} \Leftrightarrow [\neg \boldsymbol{a} \gtrsim \boldsymbol{b} \ \ et \ \ \neg \boldsymbol{b} \gtrsim \boldsymbol{a}]$$

I.3.3.2.3. La relation de sur-classement :

Par la combinaison des relations de préférence et la relation d'indéfférence, on peut former la relation de sur-classement. On notera **aSb,** qui signifie « *a* ***est moins aussi bonne que*** *b* ». C'est-à dire, **S** correspond les trois situations suivantes : *a* est préféré à *b*, a est indifférent à *b*, ou *a* est préféré faiblement à *b*. donc: $\boldsymbol{S} = \boldsymbol{P} \cup \boldsymbol{I} \cup \boldsymbol{Q}$. [VIN89]

On a alors :

- $aSb \Leftrightarrow (a \sim b) \vee (a \succ b) \vee (a \prec b)$
- $(aSb) \wedge (bSa) \Leftrightarrow a \sim b$
- $\neg(aSb) \wedge \neg(bSa) \Leftrightarrow aRb.$

On remarque que la relation *S* est réflexive i.e. aSa (*a* est toujours moins aussi que *a*)

I.3.3.2.4. La prise en compte d'un seuil d'indifférence et de préférence:

Le modèle de préférence traditionnel ne reflète pas la réalité par ce que les petits écarts positifs $\big(g(a) - g(b)\big)$ ne peuvent pas être considérés comme des préférences stricts.

Pour cela nous introduisons deux seuils suivants : p (seuil de préférence) et q (seuil d'indifférence) avec $p \geq q$. Tel que :

La notion de seuils conduit aux modèles suivants : [VIN89]

$$\begin{cases} aPb \Leftrightarrow g(a) > g(b) + q(g(b)) \\ aIb \Leftrightarrow g(a) = g(b) + q(g(b)) \end{cases}$$ Modèle à un seuil.

$$\begin{cases} aPb \Leftrightarrow g(a) > g(b) + p\big(g(b)\big). \\ aQb \Leftrightarrow g(b) + p\big(g(b)\big) \geq g(a) > g(b) + q\big(g(b)\big) \\ aIb \Leftrightarrow \begin{cases} g(b) + q\big(g(b)\big) \geq g(a), \\ g(a) + q\big(g(a)\big) \geq g(b). \end{cases} \end{cases}$$ Modèle à deux seuils.

On dit qu'un critère est un : [BEL00]

Vrai critère	si	$p = q = 0$. (modèle de préférence traditionnel).
Quasi-critère	si	$p = q \neq 0$. (modèle à un seuil d'indiffirence).
Pseudo-critère	si	$p \neq q \neq 0$. (modèle de préférence à deux seuils).
Pré-critère	si	$p \neq 0 \; et \; q = 0$. (modèle à un seuil de préférence).

I.4. L'analyse Multicritère (MCDA):

La résolution d'un problème de décision peut se faire par deux types d'analyse : *Monocritère* et *multicritère*. L'analyse monocritère est appliquée s'il y a un point de vue unique ou des points de vue multiples non conflictuels. Ce type de problème est simple dans le sens où il est mathématiquement bien posé.

L'analyse multicritère est un outil d'aide à la décision permettant d'effectuer en groupe un choix en fonction de critères préalablement définis : choisir un sujet, une solution, une action, un problème à traiter, lorsque le nombre de possibilités est important et de hiérarchiser les idées ou des solutions, l'intérêt des méthodes multicritères est de considérer un ensemble de

critères différents ***(exprimés en unités différentes « le problème de multi modalités des données »),*** sans nécessairement les transformer en critères similaires, ni en fonction unique. Il ne s'agit pas de rechercher un optimum, mais une solution compromis qui peut prendre diverses formes : choix, affectation ou classement.

I.4.1. Définition d'un problème multicritère:

On appelle le problème multicritère le couple :

$$\langle A\ ,F\rangle,$$

$$A=\{a_1,a_2,a_3,\dots,a_m\},$$

$$F=\{g_1,g_2,g_3,\dots,g_n\},$$

Où A : Ensemble des actions.

F : Est une famille cohérente des critères.

L'évaluation (ou la performance) de chaque action de *A* sur une famille de critères *F* est donnée par $g_j(a_i)$. le résultat de l'analyse des conséquences peuvent-être présentées dans un tableau de performances suivant : [NAF09]

Critères	g_1	g_2	……..	g_j	……..	g_n
poids	k_1	k_2	……..	k_j	……..	k_n
Seuils	p_1	p_2	……..	p_j	……..	p_n
	q_1	q_2	……..	q_j	……..	q_n
	v_1	v_2		v_j		v_n
Actions						
a_1	$g_1(a_1)$	$g_2(a_1)$	……..	$g_j(a_1)$	……..	$g_n(a_1)$
a_2	$g_1(a_2)$	$g_2(a_2)$	……..	$g_j(a_2)$	……..	$g_n(a_2)$
……..	……..	……..	……..	……..	……..	……..
a_i	$g_1(a_i)$	$g_2(a_i)$	……..	$g_j(a_i)$	……..	$g_n(a_i)$
……..	……..	……..	……..	……..	……..	……..
a_m	$g_1(a_m)$	$g_1(a_m)$	……..	$g_j(a_m)$	……..	$g_n(a_m)$

Tableau 1: Tableau des performances

I.4.2. Problématiques d'aide multicritère à la décision:

Lors de l'examen d'un *problème décisionnel discret*, trois problématiques sont habituellement distinguées [ROY93]: la problématique de *choix*, de *tri* et de *rangement*.

Le mot problématique doit-être compris comme désignant le type d'éclairage que la procédure d'aide à la décision vise à apporter au décideur.

- La problématique du choix (ou problématiqueα): qui consiste à vouloir identifier la meilleure solution ou sélectionner un nombre aussi faible que possible de meilleures alternatives.
- La problématique du classement (problématiqueγ) : qui consiste à vouloir Construire une hiérarchisation des alternatives à partir de la meilleure à la moins bonne.
- La problématique du Tri (ou problématiqueβ): orient vers une affectation de chaque alternative à une catégorie, jugée la plus appropriée, parmi une famille de catégories prédéfinies.

La figure suivante [DOU04] offre un aperçu graphique des types de problématique d'aide à la décision.

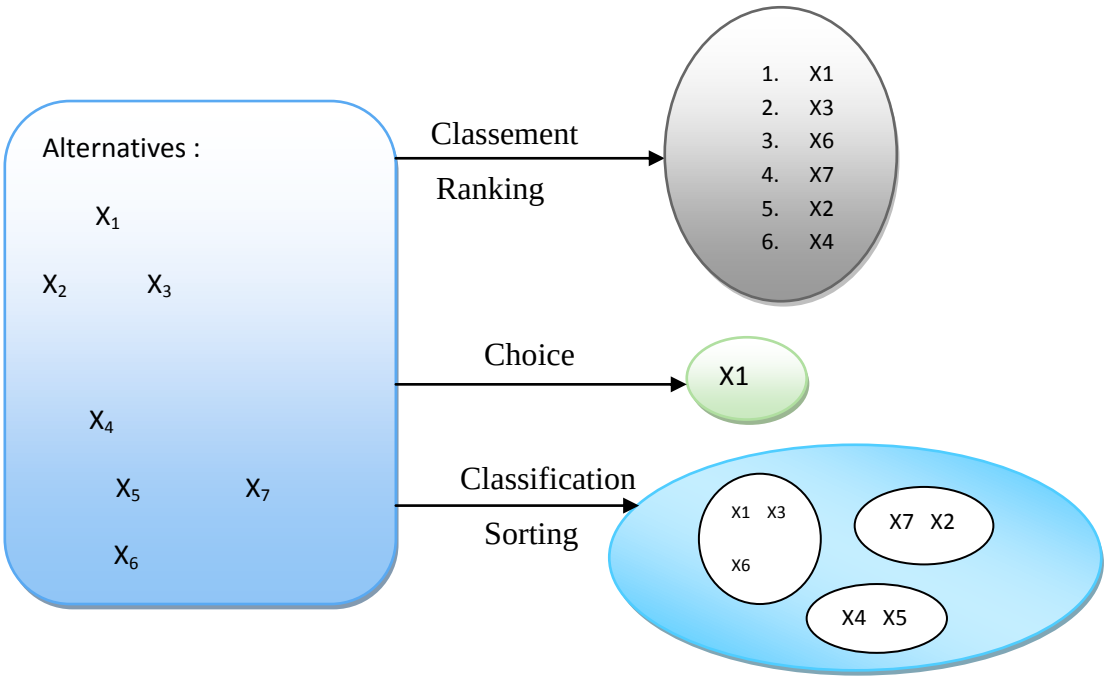


Figure 1: illustration de problématiques de MCDA

La problématique du tri peut être soit ordinale, dans le cas où les classes seraient complètement ordonnées soit nominale dans le cas où il serait difficile d'établir un ordre entre les catégories. [BEL00]

I.4.3. Processus d'aide multicritère à la décision :

L'objectif principal de MCDA est de fournir un ensemble des méthodes d'agrégation des critères qui permettent de développer des modèles d'aide à la décision, on prend en considération les préférences de décideur et son jugements.

La réalisation de cet objectif nécessite la mise en œuvre d'un processus complexe, *ce processus ne conduise pas forcément à une solution optimale, mais pour satisfaire les préférences de décideur* [ROL12].

En 1985, Roy a introduit un modèle qui couvre tous les aspects de MCDA.
Chaque processus d'aide multicritère à la décision doit passer par quatre étapes de traitement, qui seront représentés ci-après : [DOU13]

a) Définition du problème et l'objet de la décision (les actions potentielles) :

La définition du problème requiert une compréhension de la situation étudiée, du contexte et des auteurs impliqués. Dans la prise de décision, l'interaction avec les différents acteurs permet de comprendre le processus de décision, l'objet de la décision et la nature de décision à prendre, il s'agit donc de définir la nature de problème posé, soit de choix, d'affectation (classification) ou de rangement (classement).

La détermination de l'objet de la décision consiste à identifier l'ensemble des actions ou alternatives sur lesquelles va porter la décision.

b) Détermination des critères et l'analyse des conséquences:

La deuxième étape consiste à identifier tous les facteurs liés à la décision. En MCDA ces facteurs ont la forme de critères. Ensuite, il s'agit d'identifier et mesurer les conséquences des actions selon un critère donné.

c) Choix d'une méthode d'aide multicritère à la décision :

Plusieurs méthodes ont été développées. [BOU06], Le choix d'une méthode multicritère dépend de la nature du problème posé (sélection, classement ou classification). En suite détermination des différents seuils et poids des critères correspond la méthode choisit.

d) Analyse des résultats et décision finale.

I.5. Méthodes d'agrégation multicritère :

L'agrégation multicritère est une opération d'obtenir des informations sur la préférence globale entre les actions potentielles, à partir d'information sur les préférences partielles (préférences par critères).

Il existe un très grand nombre de méthodes d'aide multicritère à la décision, la plus d'entre elles sont rassemblées et expliquées par [BOU06], [VIN89], [JOS12], [ROG00], [ROY93], [MAY94], [DOU94], [ZOP04]….etc. on distingue généralement *deux* grands types de méthodes pour l'aide multicritère à la décision : on prenant la terminologie de [BOU06]. *L'approche à base sur la théorie d'utilité* et *l'approche de sur-classement*.

I.5.1. L'approche à base d'une fonction d'utilité :

La première c'est l'approche à base d'une fonction d'utilité, s'appelle aussi l'approche à relation unique de synthèse [ROL12]. Ce sont des méthodes développées par l'école américaine qui utilise le plus souvent une fonction d'utilité additive. Cette approche se base sur l'agrégation des critères de décision en un critère unique. La préférence $\gtrsim(a,b)$ est alors définie par agrégation des indices de préférence partielle. Formellement, on a :

$$\gtrsim (a,b) = f\left(\varphi_1\left(g_1(a), g_1(b)\right), \varphi_2\left(g_2(a), g_2(b)\right), \dots, \varphi_n\left(g_n(a), g_n(b)\right)\right)$$

Dans cette approche, chaque fonction φ_j est utilisée pour comparer les performances de deux alternatives sur le même critère. On distingue deux méthodes dans cette approche, selon la fonction d'utilité qu'on l'utilise *: Méthodes à base d'une fonction d'utilité additive*, *et les méthodes à base d'une fonction d'utilité non additive.*

I.5.1.1. Méthodes à base d'une fonction d'utilité additive:

La méthode la plus simple de cette catégorie est celle de la somme pondérée.

I.5.1.1.1. Somme pondérée :

Cette méthode n'est pas à proprement parler une théorie basée sur l'utilité, mais elle s'en approche. Le principe de cette méthode est simple où La note globale d'une action a notée A_a^* est la somme des toutes les valeurs $g_j(a)$ multipliée par le poids du critère (w).

$$A_a^* = \sum_{j=1}^{n} w_j . g_j(a).$$

L'avantage de cette méthode est qu'elle permet d'obtenir un résultat numérique, et un classement complet des actions.

I.5.1.1.2. Multi-attribute Utility Theory (MAUT):

La théorie de l'utilité multi-attribut est essentiellement d'inspiration anglo-saxons et largement utilisée aux Etat Unis, cette théorie a été l'une des premières méthodes de développement en MCDA, est une méthode développée vers la fin des années 60 par **howard raiffa** et **ralph keeney** [DOU13]. Cette approche est utilisée dans plusieurs problèmes d'aide à la décision comme des problèmes d'économie, de finance, d'aide au diagnostic médicale, …etc. cette théorie repose sur l'axiome fondamentale suivante : tout décideur essaye de maximiser la fonction d'agrégation de tous les points de vue à prendre en compte (Critères).i.e. déterminer d'une utilité totale $\mu(g_1, g_2, \dots, g_n)$ (utilité d'une alternative a) qui peut être tirée à partir des utilités élémentaires ou partielle $\mu_i(g_i)$ (utilité de chaque critère $g_i;\ i = 1..n$) déterminé comme suit:

- $\boldsymbol{\mu(a) > \mu(b) \Rightarrow a \succ b}$ (Alternative a est préféré à l'alternative b).
- $\boldsymbol{\mu(a) = \mu(b) \Rightarrow a \sim b}$ (Alternative a est indifférent de l'alternative b).

La difficulté principale de cette méthode réside dans la complexité d'estimer la fonction d'utilité. Plusieurs forme d'utilité existe dans la littérature telles que : additive, multiplicatif, Linéaire, …etc [VIN89]. La forme analytique la plus utilisée d'une fonction d'utilité est l'additive :

$$\mu(a) = \sum_{i=1}^{n} w_i . \mu_i(g_i(a))$$

Où : $\mu_1, \mu_2, \dots, \mu_n$ sont les fonctions d'utilités pour évaluation n critères. Où ces utilités servent à transformer les critères initiaux de manière à ce qu'ils s'expriment tous suivant la même échelle, par exemple [0,1]. [VIN89].

$w_1, w_2, \dots, w_n$ Sont des constantes représentant le jugement de décideur sur un critère. Ces constantes sont souvent considérées comme les poids des critères.

I.5.1.2. Méthodes à base d'une fonction d'utilité non-additive

Pour des raisons de faiblesse de la somme pondérée, il est apparu la notion de capacité et les intégrales floues dans le domaine d'aide multicritère à la décision.

Il existe plusieurs classes d'intégrales floues, parmi lesquelles les plus représentatives dans MCDA sont celles de *Choquet* [GRA96][GRA06] et *Sugeno* [DUB01] . Dans cette section nous étudions ces types d'intégrales en tant que fonctions d'agrégation.

I.5.1.2.1. La notion de capacité (Fuzzy Measure)

Mesure floue est un des concepts les plus importantes en mathématiques, particulièrement dans les *ensembles floues* [AMB09], elle est très adaptée à l'étude des situations dont les connaissances sont imparfaites (incertaines et imprécises).

Définition (capacité) : *soit* $N = \{1,2,3,\dots,n\}$ *un ensemble de critères, une capacité sur N (appelée aussi fuzzy measures ou mesure floue* [GRA08]*), est une fonction* $\mu: 2^N \to [0.1]$ $vérifiant\ les\ conditions\ suivantes$ *:*

(i) $\mu(\emptyset) = 0, \mu(N) = 1.$

(ii) $monotonie\ c'est - à - dire: \mu(A) \leq \mu(B), si\ A \subseteq B \subseteq N.$

Ainsi, une capacité sur *N* est dite *additive* si : $\mu(A \cup B) = \mu(A) + \mu(B)\ /\ A \cap B = \emptyset$

I.5.1.2.2. L'intégrale de Choquet

Si l'intégrale de Choquet est une notion qui est apparue en 1953 [GRA06], il faut attendre les années 90 pour voir les premières applications en aide multicritère à la décision (MCDA) [GRA96].

L'intégrale de *Choquet* généralise la somme pondérée si la capacité utilisée est additive.

I.5.1.2.2.1. Définition (intégrale de choquet) :

Définition*: Soit* $\boldsymbol{\mu}$ *une capacité sur* $N = \{x_1, x_2, \dots, x_n\}. et\ f: N \to R$ *une fonction représentant les scores d'un objet sur les* **n** *critères. L'intégrale de Choquet de* **f** *par rapport à* $\boldsymbol{\mu}$ *(score global de l'objet) est donné par :*

$$C_\mu(f) := \sum_{i=1}^{n} [f(x_i) - f(x_{i-1})]\mu(A_i).$$

Avec : $f(x_0) = 0.$

$A_i := \{i, i+1, .., n\}. est\ une\ permutation\ sur\ N\ telle\ que:\ f(x_1) \leq f(x_2) \leq \cdots \leq f(x_n).$

$\boldsymbol{\mu}: est\ une\ fonction\ d'ensemble\ croissante.$

$\boldsymbol{\mu}(\boldsymbol{A}): est\ l'importance\ du\ groupe\ de\ critères\ A.$

I.5.1.2.2.2. Explication de l'intégrale de Choquet :

Le concept de base de l'intégrale de Choquet peut être illustré par la figure ci-dessous [AYU09] :

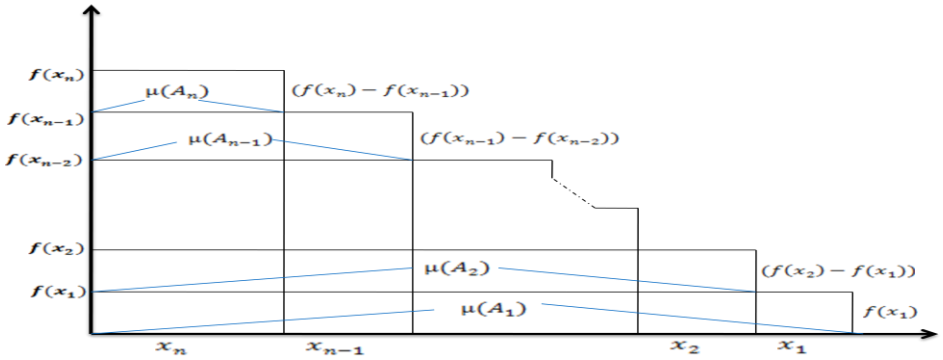


Figure 2 : Illustration du concept de l'intégrale de choquet.

Soit $N = \{x_1, x_2, \dots, x_n\}$ n critère, avec de scores suivants $\{f(x_1), f(x_2), \dots, f(x_n)\}$ telle que : $f(x_1) \leq f(x_2) \leq \cdots \leq f(x_n)$, et les mesures de ces critères sont : $\mu(A_1), \mu(A_2), \dots, \mu(A_n)$.

- L'aire de premier rectangle $= \big(f(x_1)\big)\mu(A_1)$ $avec$ $A_1 = \{x_1, x_2, \dots, x_n\}$.
- L'aire de deuxième rectangle $= \big(f(x_2) - f(x_1)\big)\mu(A_2)$ $avec$ $A_2 = \{x_2, \dots, x_n\}$.
- L'aire de $n^{ième}$ rectangle $= \big(f(x_n) - f(x_{n-1})\big)\mu(A_n)$ $avec$ $A_n = \{x_n\}$.

Dans le domaine d'aide multicritère à la décision, la somme de ces valeurs est considérée comme une degré de préférence de l'alternative correspondant.

I.5.1.2.3. Intégrale de Sugeno:

L'intégrale de Sugeno a été introduite par *Sugeno* en 1974 dans sa thèse de doctorat [SUG74], mais l'utilisation de cette notion comme un outil pour déterminer la relation de préférence entre les alternatives se faite au début des années 2000 [MAR00] [DUB01]. Comme l'integrale de Choquet, l'integrale de Sugeno se base sur la notion de capacité

Définition*: l'intégrale de Sugeno discret non additive est une fonction d'agrégation de* $[0,1]^n \rightarrow [0,1]$ *sur un ensemble de critère* $N = \{1,2, \dots ,3\}$ *avec des scores d'un alternative correspondant* $X = (x_1, x_2, x_3, .., x_n)$ *par rapport à une capacité* μ *est définie par :*

$$S_\mu(X) = \bigvee_{i=1}^{n} \left(x_i \bigwedge \mu\,(A_i)\right)$$

$$= \bigwedge_{i=1}^{n} \left(x_i \bigvee \mu(A_{i+1})\right)$$

$telle\ que: x_1 \leq x_2 \leq \cdots \leq x_n.$

Où :

- $\vee$: $le\ maximum\ et\ \wedge$: $le\ minimum$.
- 0 : le plus petit élément et 1 : le plus grand élément. Parce que l'intégrale de Sugeno joue un rôle important dans le cas où les données sont qualitatives [DUB01].
- $\mu: 2^N \rightarrow [0,1]$: $une\ mesure\ floue$.
- A_i: $Permutation\ sur\ N\ pour\ i = 1, \dots, n\ .A_i = \{i, i+1, \dots, n\} et\ A_{n+1} = \phi$.

Par exemple :

Si $x_3 \leq x_1 \leq x_2 : S_\mu(x_1, x_2, x_3) = \vee(x_3 \wedge \mu(x_1, x_2, x_3); x_1 \wedge (x_1, x_2); x_2 \wedge \mu(x_2))$

Selon la définition de l'intégrale de Sugeno, sa valeur est :

$$S_\mu(X) \in \{x_1, x_2, \dots, x_n\} \cup \mu(A) / \ A \subseteq N\ et\ X \in [0,1]^n.$$

Formellement, les intégrales de Sugeno et de Choquet sont des méthodes de classement, mais elles peuvent être utilisées comme méthodes de tri (sorting) en définissant un niveau de référence λ_t par catégorie C_t tel que: $a \in C_t \Leftrightarrow \mu(a) \geq \lambda_t$.[ROL12]

Pour comprendre bien l'application de deux intégrales de Choquet et de Sugeno dans le domaine MCDA, voire [RAK07][AYU07] qui présentent un exemple détaillé aux données médicales..

I.5.2. L'approche de sur-classement :

S'appelle aussi, l'approche à relation de synthèse [BOU06]. Ces méthodes sont développées par l'école Européenne, premièrement par Bernard ROY durant la fin des années soixante, où nous retrouvons les méthodes ELECTRE (Elimination Et Choix Traduisant la REalité).

Toutes les méthodes de sur-classement sont exécutées en deux étapes :

- Développement de la relation de sur-classement.
- Exploitation de la relation de sur-classement selon le problème posé (Choix, classement, classification).

Les relations de sur-classement permettent de construire le graphe de sur-classement. Ces méthodes permettent aussi d'établir des notions d'incomparabilité et d'indifférence, ce qui correspond souvent le mieux à la façon de penser du décideur.

De nombreuses modifications aux méthodes classiques (les méthodes à base d'utilité) ont été proposées. Ces modifications sont liées à l'introduction de plusieurs notions comme (seuil

de concordance, seuil de discordance, seuil de préférence, seuil d'indifférence, l'indice de crédibilité, indice de coupe,…etc

Les méthodes de sur-classement les mieux connues et les plus utilisées sont les méthodes ELECTRE [ROY 93], [VIN89], [MAY94] et les méthodes PROMETHEE [BRA 86].

I.5.2.1. Les Méthodes ELECTRE :

La mise au point des méthodes ELECTRE (Elimination Et Choix Traduisant la REalité) représente un travail conduit sur plus de 25 ans (1968 à 1994), toutes les méthodes ELECTRE appartiennent à l'approche de sur-classement de synthèse *acceptant l'incomparabilté.* En plus les méthodes ELECTRE se basant sur quelques concepts tels que, *l'indice de concordance, l'indice de discordance, l'indice de crédibilité, l'indice de coupe, pré-ordre total, pré-ordre partiel, l'indice de crédibilité,…etc.* dans la suite de cette section, nous présentons quelques méthodes de la famille ELECTRE les plus connues telles que (ELECTRE I, ELECTREII, ELECTREIII, ELECTRE Tri).

I.5.2.1.1.ELECTRE I :

Cette méthode a été construite par Bernard ROY en 1968 [MAY94], l'idée de cette algorithme est posé en terme de choix des alternatives qui sont préférées pour la plupart des critères, c'est-à-dire, elle relève de la problématique α (procédure de sélection).

I.5.2.1.1.1. Développement de la relation de sur-classement.

La construction de sous-ensemble préféré se fait en définissant une relation de sur-classement. Pour définir cette dernière, il faudrait de développer trois concepts : *matrice de concordance*, *matrice de discordance*, et *valeurs seuil*.

- ***Matrice de concordance :*** est une matrice de *mxm éléments*, chaque élément du cette matrice s'appel l'*indice de concordance* pour deux alternatives *a* et *b* est noté *C(a,b)*, calculé par suivant :

$$C(a,b) = \frac{\sum_{\forall j, g_j(a) \geq g_j(b)} K_j}{K}; \; avec \; K = \sum_{j=1}^{n} K_j.$$

Il est évident que l'indice de concordance pour deux alternatives compris entre 0 et 1.

- ***Matrice de discordance :***

L'indice de discordance entre deux alternatives *a* et *b*, noté *D(a,b)* est considérée comme une mesure d'insatisfaction de choisir *a* par rapport à *b*, les éléments du cette matrice est calculé selon la formule suivante :

$$D(a,b) = \begin{cases} 0 \quad Si \ \forall j, g_j(a) \geq g_j(b) \\ Sinon \\ \frac{1}{\delta} \max_j [g_j(b) - g_j(a)] \end{cases}$$

Avec $\boldsymbol{\delta}$: *est la différence maximale entre le même critère pour deux actions donnée*

La relation de sur-classement construite pour sélectionner les actions préférés, pour cela, en se basant sur les deux seuils, seuil de concordance ĉ (relativement grand) qui exprime le minimum de concordance ce requise pour que la proposition « *a* surclasse *b*» ne soit pas rejetée, et un seuil de discordance $\hat{d}$ (relativement petit), qui exprime le maximum de discordance tolérée pour l'hypothèse « *a* surclasse *b*» ne soit pas rejetée, tous les deux compris entre 0 et 1. On définit la relation de sur-classement *S* selon les deux conditions suivantes :

$$aSb \ \ ssi \begin{cases} C(a,b) \geq \hat{c} \\ D(a,b) \leq \hat{d} \end{cases}$$

I.5.2.1.1.2. Exploitation de la relation de sur-classement :

Pour simplifier le choix des alternatives (actions), La relation de sur-classement dans la méthode ELECRE I est représentée par un graphe orienté appelé un **graphe de sur-classement**, dont les sommets sont l'ensemble d'actions, et les arcs sont reliés entre deux actions *a* et *b* ; tel que *a* surclasse *b* (aSb).

Définition: noyau d'un graphe [VIN89] **:** Un noyau d'un graphe G = (A, R) est une partie N de l'ensemble des sommets A telle que :

- $\forall a_k \in A/N \Rightarrow \exists a_i \in N \ tel \ que \ a_i R a_k$.
- $\forall a_i, a_k \in N, \ \neg a_i \ R \ a_k$.

On recherche donc un sous-ensemble N d'actions tel que toute action qui n'est pas dans N est surclassé par au moins une action de N et les actions de N sont incomparables (cette deuxième condition permet de rendre N minimal).

Remarque :

- C'est très difficile pour déterminer un noyau dans un graphe s'il existe des circuits.
- un graphe sans circuit possède un noyau unique.

Exemple pratique : [VIN89]

A = {1,2,3,4,5,6,7} et R = {(2,1),(2,3),(4,3),(4,6),(4,5),(5,3),(5,4),(5,6)}.

Exemple : Soit G = (A, R), un graphe de relation binaire tel que :

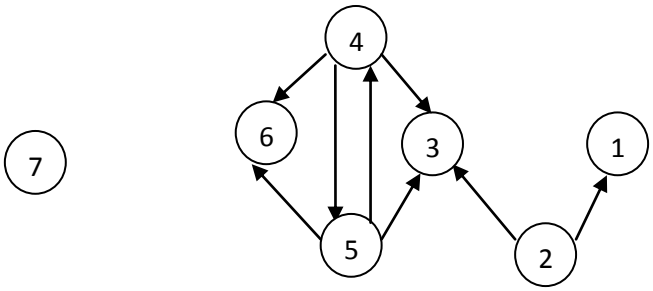


Figure 3 : Représentation graphique d'une relation binaire.

Il existe un circuit dans ce graphe, donc Les noyaux du graphe sont les sous ensembles {2, 4, 7} et {2, 5, 7}. La relation entre les sommets 4 et 5 est une relation transitive.

Après la construction d'un graphe de sur-classement, on choisit qui est équivalent à la recherche du *noyau du graphe G* représentant *S*.

I.5.2.1.2. ELECTRE III :

Avec l'évolution de la modélisation des préférences, des procédures qui prennent en compte des seuils d'indifférence et de préférence sont apparu. La méthode ELECTRE III est un bon exemple. ELECTRE III est une méthode d'analyse multicritère qui permet de résoudre des problèmes de sur-classement. La problématique concernée ici est celle de rangement.

Pour cette méthode, les seuils d'indifférence, de préférence et de veto sont la fonction de l'évaluation de l'action pour chaque critère. Pour une action *a* évaluée pour le critère *j*, dans ce cas le seuil d'indifférence est noté q_j , le seuil de préférence est noté p_j, et le seuil de véto est noté v_j.

Les seuils p_j, qj *et* v_j de chaque critère *j* satisfaisants la condition suivant :

$$v_j \geq p_j \geq q_j.$$

La méthode ELECTRE III s'appuie sur les étapes suivantes :

I.5.2.1.2.1. Construction de la relation de sur-classement :

- *Evaluation des indices de concordance :*

Dans ce cas, on considère le sens de préférence des critères, on distingue un sens de préférence croissant et décroissant. Dans le cas de préférence croissant, les indices de concordance sont obtenus comme suit :

$$c_j(a,b) = \begin{cases} 1 & si: g_j(b) - g_j(a) \leq q_j\left(g_j(a)\right) \\ 0 & si: g_j(b) - g_j(a) \geq p_j\left(g_j(a)\right) \\ sinon & \dfrac{p_j\left(g_j(a)\right) + g_j(b) - g_j(a)}{p_j\left(g_j(a)\right) - \quad q_j\left(g_j(a)\right)} \end{cases}$$

$c_j(a,b)$ d'un critère croissant est illustré par la figure ci-dessous :

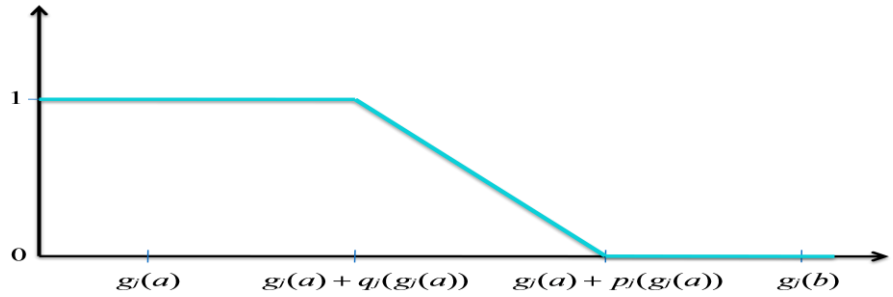


Figure 4: Détermination de l'indice de concordance (Préférence croissante).

Dans le cas de préférence décroissante $c_j(a,b)$ est illustré comme suit :

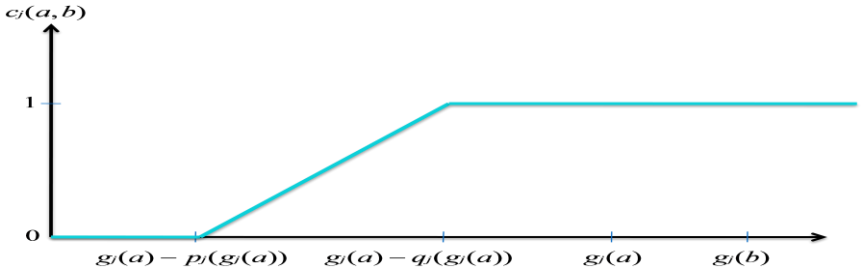


Figure 5: Détermination de l'indice de concordance (Préférence décroissante).

- *Calcule de l'indice de concordance global* $C(a,b)$*:*

$$C(a,b) = \sum_{j=1}^{n} k_j . c_j(a,b) \Big/ \sum_{j=1}^{n} k_j$$

- *Evaluation des indices de discordance :*

La discordance d'un critère g_j vie à prendre en compte le fait que ce critère est plus ou moins discordant avec l'affirmation *aSb*. L'indice de discordance atteint sa valeur maximale lorsque le critère g_j met son veto à la relation de sur-classement, il est minimale lorsque le critère g_j pas discordant avec cette relation, cette indice peut maintenant se présenter comme suit :

$$d_j(a,b) = \begin{cases} 0 & si: g_j(b) - g_j(a) \leq p_j\left(g_j(a)\right) \\ 1 & si: g_j(b) - g_j(a) \geq v_j\left(g_j(a)\right) \\ sinon & \dfrac{p_j\left(g_j(b)\right) - g_j(a) - p_j\left(g_j(a)\right)}{v_j\left(g_j(a)\right) - \quad q_j\left(g_j(a)\right)} \end{cases}$$

Comme illustré dans la figure suivante :

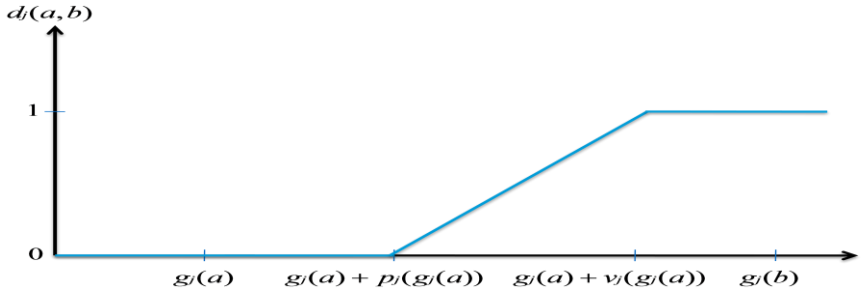


Figure 6: Détermination de l'indice de discordance (préférence croissante)

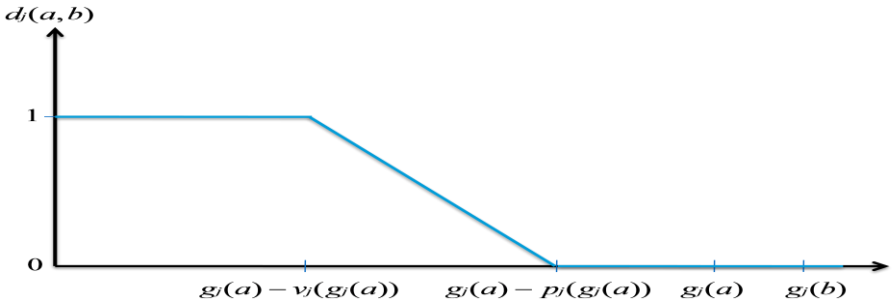


Figure 7: Détermination de l'indice de discordance (préférence décroissante)

Fig : détermination de l'indice de dis-concordance (préférence décroissante).

- *Calcul de l'indice de crédibilité et définissant la relation de sur-classement floue:*

Cette étape consiste à combiner les indices de concordances et de discordance pour chaque paire d'alternatives pour produire une mesure du degré de sur-classement i.e : une matrice de crédibilité qui évalue la force de l'affirmation « a est moins aussi bonne que b ». Le degré de crédibilité pour chaque paire d'alternatives est défini comme suit :

$$\sigma(a,b) = \begin{cases} d_j(a,b) \quad si \quad d_j(a,b) \leq C(a,b) \quad \forall j = 1..n. \\ sinon \qquad\qquad C(a,b) * \prod_{j \in \overline{F}} \dfrac{(1-d_j(a,b))}{1-c_j(a,b)} \end{cases}$$

Avec : $\overline{F} = \{g_j \in F : d_j(a,b) > C(a,b)\}$.

Dans [LI07], les auteurs ont développé une autre procédure pour construire et exploiter la relation de sur-classement dans la méthode ELECTRE III à partir de trois types de degré de crédibilité pour chaque action $a \in A$.

- Degré de concordance de crédibilité $\psi^+(a)$:

  Ce degré quantifie la manière dont une action a est globalement préférée **à** toutes les autres actions de A.

$$\psi^+(a) = \sum_{b \in A} \sigma(a,b).$$

- Degré de dis-concordance de crédibilité $\psi^{-}(a)$:

  Ce degré quantifie la manière dont une action *a* est globalement préférée **par** toutes les autres actions de *A*.

$$\psi^{-}(a) = \sum_{b \in A} \sigma(b, a).$$

- Degré net de crédibilité $\psi(a)$:

$$\psi(a) = \psi^{+}(a) - \psi^{-}(a)$$

I.5.2.1.2.2. Exploitation de la relation de sur-classement :

Dans [LI07], Le degré de net de crédibilité représente la valeur globale de chaque alternative où :

$$\psi(a) > \psi(b) \Leftrightarrow a \succ b$$
$$\psi(a) = \psi(b) \Leftrightarrow a \sim b$$

Pour obtenir un classement complet des actions, l'origine ELECTRE III utilise deux procédures de classement. La première est obtenue de manière descendante où les alternatives sont classées de meilleure au pire *(distillation descendante).* La seconde est ascendante où les alternatives sont classées de la plus mauvais à la meilleure alternative *(distillation descendante).* [ROG99].

I.5.2.1.3. ELECTRE Tri :

ELECTRE Tri est conçue pour attribuer un ensemble d'actions à des catégories homogènes. En ELECTRE Tri, les catégories sont ordonnées, supposons de pire C_1 au meilleur C_k. chaque catégorie doit-être caractérisée par un seuil inférieur et un seuil supérieur.

Par contre aux autres approches, les alternatives qui constituent l'objet de décision ne sont pas comparées entre elles, mais à des seuils traduisant la frontière entre les *k* classes prédéfinies noté : $C = \{C_1, C_2, \dots, C_h, \dots, C_k\}$, désigne l'ensemble des catégories et $B = \{b_0, b_1, \dots, b_h, \dots, b_k\}$, désigne l'ensemble des frontières des catégories.

L'attribution d'une action donnée à une certaine catégorie C_h basée sur la comparaison de deux seuils précédente, tel que b_h est la borne supérieure de la catégorie C_h et la borne inférieur de la catégorie C_{h+1}, $\forall\, h = 1..k$.

La figure ci-dessous illustre la problématique de Tri ou d'affectation : [MOU00], [ZOP04].

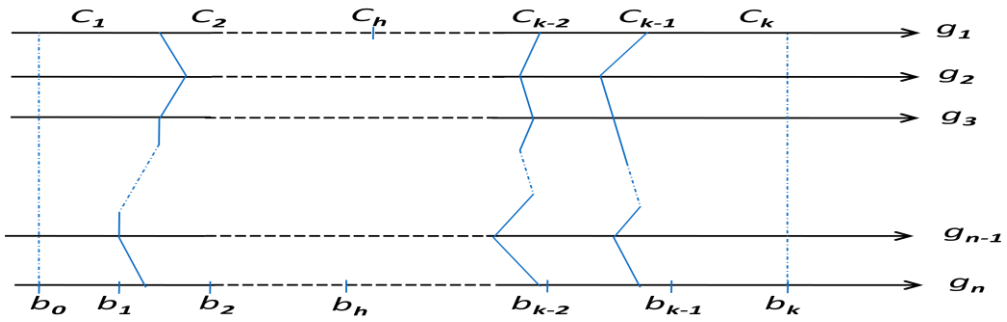


Figure 8: Illustration de la problématique de Tri.

Les premières étapes de la méthode ELECTRE Tri sont comme dans celle de ELECTRE III jusqu'à l'étape calcule de l'indice de crédibilité, Tel que le seuil de préférence $p_j(b_h)$, le seuil d'indifférence $q_j(b_h)$, et le seuil de veto $v_j(b_h)$, sont définis comme suit :

$$p_j(b_h) = max\big(g_j(a) - g_j(b_h)\big).$$
$$q_j(b_h) = min\big(g_j(a) - g_j(b_h)\big).$$
$$v_j(b_h) = min\big(g_j(b_h) - g_j(a)\big).$$

ELECCTRE Tri effectue les alternatives aux catégories selon deux étapes consécutives :

I.5.2.1.3.1. Construction de la relation de sur-classement :

ELECTRE Tri construit la relation de sur-classement entre chaque action a et l'ensemble des frontières des classes B, tel que aSb_h ou b_hSa à base sur l'indice de crédibilité $\sigma(a, b_h) \in [0,1]$ (respectivement, $\sigma(b_h, a)$) qui représente le degré de crédibilité de l'affirmation aSb_h (respectivement b_hSa), $\forall a \in A \; et \; \forall b \in B$. comme on le voit, le degré de crédibilité est défini dans la méthode ELECTRE III. En plus, la relation de sur-classement dans ELECTRE Tri définie à partir de *L'indice de coupe* $\lambda \in [0.5,1]$. tel que :

$$\sigma(a, b_h) \geq \lambda \; et \; \sigma(b_h, a) \geq \lambda \Rightarrow aSb_h \; et \; b_hSa \Rightarrow (\boldsymbol{a \; I \; b_h}); \; (a \; Indifférence \text{ à } b_h).$$
$$\sigma(a, b_h) \geq \lambda \; et \; \sigma(b_h, a) < \lambda \Rightarrow aSb_h \; et \; \neg b_hSa \Rightarrow (\boldsymbol{a \; P \; b_h}); (a \; Préféré \text{ à } b_h).$$
$$\sigma(a, b_h) < \lambda \; et \; \sigma(b_h, a) \geq \lambda \Rightarrow \neg aSb_h \; et \; b_hSa \Rightarrow (\boldsymbol{b_h \; P \; a}); (b_h \; Préféré \text{ à } a).$$
$$\sigma(a, b_h) < \lambda \; et \; \sigma(b_h, a) < \lambda \Rightarrow \neg aSb_h \; et \; \neg b_hSa \Rightarrow (\boldsymbol{b_h \; R \; a}); (a \; Incomparable \; avec \; b_h).$$

I.5.2.1.3.2. Exploitation de relation de sur-classement :

Le rôle de cette étape est déterminer la catégorie de chaque alternative à partir de le comparer avec les profiles b_h , Deux procédures d'affectation sont possibles :

Procédure pessimiste :

(i) Comparer successivement a à b_h tel que : $h=k, k-1,\ldots,0$.

(ii) b_h étant le premier profile de telle sorte que: aSb_h, donc, a est effectuée à la catégorie C_{h+1}.

Procédure optimiste :

(i) Comparer successivement a à b_h, tel que : $h=0,1,2,\ldots,k$.

(ii) b_h étant le premier profile de telle sorte que: b_hSa, a est effectuée à la catégorie C_h.

I.5.2.2 les Méthodes PROMETHEE :

Preference **R**anking **O**rganization **METH**od for **E**nrichment **E**valuation est une famille de méthodes d'agrégation multicritère qui appartiennent à la famille des méthodes de sur-classement, ces méthodes développées par Brans et al [BRA86] depuis 1986 en Belgique. Nous avons présenté le principe des méthodes PROMETHEE I (partial ranking) et PROMETHEE II (complet ranking).

I.5.2.2.1. Principe général :

Les deux méthodes ont même enchainement initial, mais leur but différent, PROMETHE I permet de dégager des relations partielles de classement alors que PROMETHEE II fournit un classement de toutes les actions.

La méthode PROMETHEE est exécutée en cinq étapes :

1ière étape : comparaison par paire : $d_j(a,b)$

Premièrement, une comparaison sera faite paire par paire entre toutes les actions pour chaque critère où : $d_j(a,b) = g_j(a) - g_j(b)$ avec $d_j(a,b)$ est la différence entre les évaluations de deux actions a, b pour le critère g_j. (Ces différences bien sûres dépendent des échelles de mesure utilisées).

2ième étape : Degré de préférence : $\pi_j(a,b)$

$$\pi_j(a,b) = P_j\left(d_j(a,b)\right)$$: En cas de maximisation d'un critère.

$$\pi_j(a,b) = P_j\left(-d_j(a,b)\right)$$: En cas de minimisation d'un critère.

Où : $P_j : \Re \rightarrow [0,1]$ est une fonction de préférence positif telle que :

$$P_j(0) = 0 \; et \; \pi_j(a,b) > 0 \Rightarrow \pi_j(b,a) = 0.$$

Six types de préférence sont proposés dans la définition originale PROMETHEE [JOS12]:

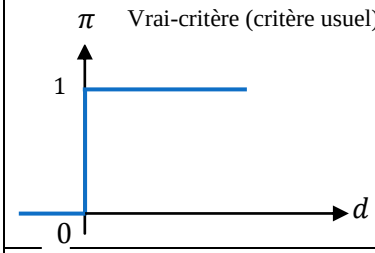
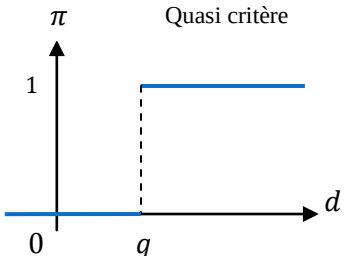
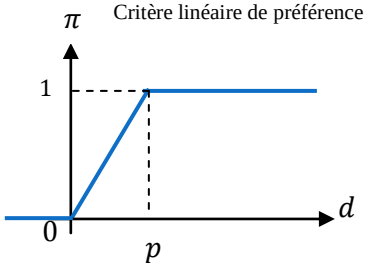
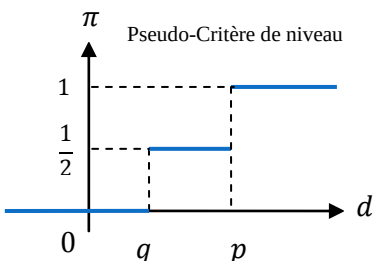
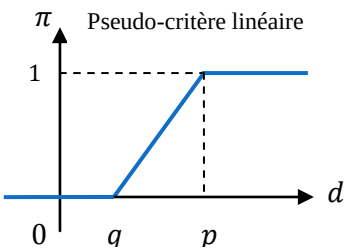
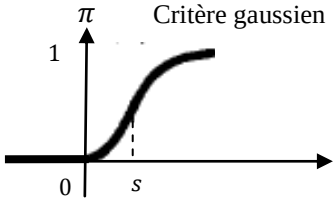
Critère généralisé	Définition	Paramètres à fixer
Vrai-critère (critère usuel) 	$\pi_j(a,b) = \begin{cases} 0, & si\ \ d \le 0 \\ 1, & si\ \ d > 0 \end{cases}$	-
Quasi critère 	$\pi_j(a,b) = \begin{cases} 0, & si\ \ d \le q \\ 1, & si\ \ d > q \end{cases}$	Q
Critère linéaire de préférence 	$\pi_j(a,b) = \begin{cases} 0, & si \quad d \le 0 \\ \frac{d}{p}, & si\ \ 0 < d < p \\ 1, & si \quad d \ge p \end{cases}$	P
Pseudo-Critère de niveau 	$\pi_j(a,b) = \begin{cases} 0, & si \quad d \le q \\ \frac{1}{2}, & si\ \ q < d < p \\ 1, & si \quad d \ge p \end{cases}$	p, q
Pseudo-critère linéaire 	$\pi_j(a,b) = \begin{cases} 0, & si \quad d \le 0 \\ \frac{d-q}{p-q}, & si\ \ 0 < d < p \\ 1, & si \quad d \ge p \end{cases}$	p, q
Critère gaussien 	$\pi_j(a,b) = \begin{cases} 0, & si\ d \le 0 \\ 1 - e^{-\frac{d^2}{2s^2}}, & si\ \ d > q \end{cases}$	S

Tableau 2: Les six types de critères dans la méthode PROMETHEE.

3ième étape : indice de préférence multicritère : $\pi(a,b)$

Cet indice est considéré comme le degré de préférence d'une action *a* par rapport à *b* sur tous les critères :

$$\pi(a,b) = \sum_{j=1}^{n} \pi_j(a,b).w_j$$

Où : w_j représente le poids du critère *j*.

4ième étape : flux de préférence multicritère :

Les valeurs de 3ième étape représentent le graphe de sur-classement, où les sommets sont l'ensemble des actions, l'arc *(a, b)* prend la valeur de $\pi(a,b)$. Nous définissons pour chaque sommet de *A* deux valeurs : $\varphi^+(a)\ et\ \varphi^-(a)$.

Avec :

Flux de sur-classement positif:

$$\varphi^+(a) = \frac{1}{n-1}\sum_{b\in A}\pi(a,b)$$

Flux de sur-classement négatif :

$$\varphi^-(a) = \frac{1}{n-1}\sum_{b\in A}\pi(b,a)$$

Où : $\varphi^+(a)$ désigne le flux positif (arc sortant) quantifie la manière dont une actions donnée *a* est globalement préférée **à** toutes les autres actions.

$\varphi^-(a)$ représente le flux négatif (arc entrant) quantifie la manière dont une action donnée est globalement préférée **par** toutes les autres actions.

5ième étape : détermination de la relation de sur-classement entre les actions :

Comme on le voit, les méthodes PROMETHEE permet de résoudre les problèmes multicritère de classement, deux types de classement sont définis dans la famille des méthodes PROMETHEE, *classement partiel* et *classement totale*.

I.5.2.2.2. PROMETHEE I :

La méthode PROMETHEE I consiste à classer les actions sur la base de relation de sur-classement, cette dernière est construite comme le suivant :

Nous définissons deux pré-ordres totale (P^+, I^+) et (P^-, I^-) comme suit :

$$\begin{cases} aP^+b & si \quad \varphi^+(a) > \varphi^+(b) \\ aP^-b & si \quad \varphi^-(a) < \varphi^-(b) \end{cases}$$

$$\begin{cases} aI^{+}b & si \quad \varphi^{+}(a) = \varphi^{+}(b) \\ aI^{-}b & si \quad \varphi^{-}(a) = \varphi^{-}(b) \end{cases}$$

En suite, on obtient le pré-ordre partiel *(P, I, R)*, en tenant compte leur intersection :

$$\begin{cases} aPb & si \begin{cases} aP^{+}b \ et \ aP^{-}b \\ aP^{+}b \ et \ aI^{+}b \\ aP^{+}b \ et \ aI^{-}b \end{cases} \\ aIb & si \ aP^{+}b \ et \ aI^{-}b \\ aRB & Sinon \end{cases}$$

C'est la relation partiel de la méthode PROMETHEE I, elle offre au décideur un graphe en certaines actions qui sont comparables, et d'autres actions ne sont pas comparables.

I.5.2.2.3. PROMETHEE II :

Dans cette méthode, on suppose le pré-ordre total (*P*, *I*) (la relation d'incomparabilité est exclue), nous définissons pour chaque action une valeur dites *flux net* $\varphi(a)$ comme suit :

$$\varphi(a) = \varphi^{+}(a) - \varphi^{-}(a)$$

La valeur de flux net est utilisée pour déterminer la relation de sur-classement entre deux actions de *A* comme suit :

$$aPb \Leftrightarrow \varphi(a) > \varphi(b)$$
$$aIb \Leftrightarrow \varphi(a) = \varphi(b)$$

I.6. Pondération et importance des critères :

I.6.1. Pondération des critères :

I.6.1.2. Principe général

Un des aspects significatifs dans les problèmes d'agrégation est la prise en considération *de l'importance des critères* considérés, laquelle puisque ces poids doivent être pris en considération durant la phase d'agrégation.

Les méthodes d'aide multicritère à la décision ne permet pas de générer de manière automatique les poids des critères, on est obligé d'intégré une autre approche pour pondérer les critères à l'intérieur de ces méthodes. [CHE12].

Détermination des poids des critères n'est pas une chose simple, elle nécessitée l'interaction de l'analyste avec un ou plusieurs décideurs. Il existe de nombreuses méthodes de pondération des critères. Deux approches existent à ce propos : [ROL12].

- Demander au décideur de fixer le poids de chaque critère directement.

- Retrouver les poids à partir d'un jeu de données existant : Cette approche se base sur la comparaison des critères deux à deux pour construire qu'on s'appelle une matrice de comparaison binaire *C*.

I.6.1.2. Matrice de comparaisons binaires :

La matrice de comparaison par paire *C* est un outil de prise de décision qui est utilisée pour évaluer les alternatives sur la base d'évaluation les poids des critères par paires, on effectue les valeurs numériques c_{ij} pour estimer le rapport $\frac{w_i}{w_j}$. [SAA77].

Avec : w_i *: le poids de critère i.*

w_j *: le poids de critère j.*

$$C = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{pmatrix} = \begin{pmatrix} \frac{w_1}{w_1} & \frac{w_1}{w_2} & \dots & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & \frac{w_2}{w_2} & \dots & \frac{w_2}{w_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \dots & \frac{w_n}{w_n} \end{pmatrix}$$

Figure 9: Matrice de comparaisons binaire.

I.6.1.3. Matrice de comparaisons binaires réciproque :

On dit qu'une matrice de comparaison est réciproque si et seulement si les opinions symétriques issues d'un même décideur inverse l'une de l'autre i.e : $c_{ij} = \frac{1}{c_{ji}}$. et les opinions diagonales égale à 1.

$$C_r = \begin{pmatrix} 1 & c_{12} & \dots & c_{1n} \\ \frac{1}{c_{12}} & 1 & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{c_{n1}} & \frac{1}{c_{n2}} & \dots & 1 \end{pmatrix}$$

Figure 10: Matrice de comparaisons binaires réciproque.

I.6.1.4. Les méthodes de pondération des critères :

Une méthode de pondération des critères se définit tout simplement comme l'algorithme qui, à partir d'un ensemble d'opinions exprimant des comparaisons binaires de critères, fournit un jeu de poids acceptable. Pour *n* critères à pondérer, il s'agit d'estimer des

importances relatives qui peuvent prendre la forme$c_{ij} = \frac{w_i}{w_j}, i, j = 1, \dots, n. w_i, w_j$, représentent respectivement les poids des critères g_i, g_j.

Une comparaison alors consiste à choisir un rapport d'importance c_{ij} sur chaque arrangement (g_i, g_j) parmi un ensemble de valeurs possibles dans l'intervalle$]0, +\infty[$. [LIM01]

Le schéma suivant illustre le principe des méthodes de pondération des critères selon une matrice de comparaison binaire : [LIM01]

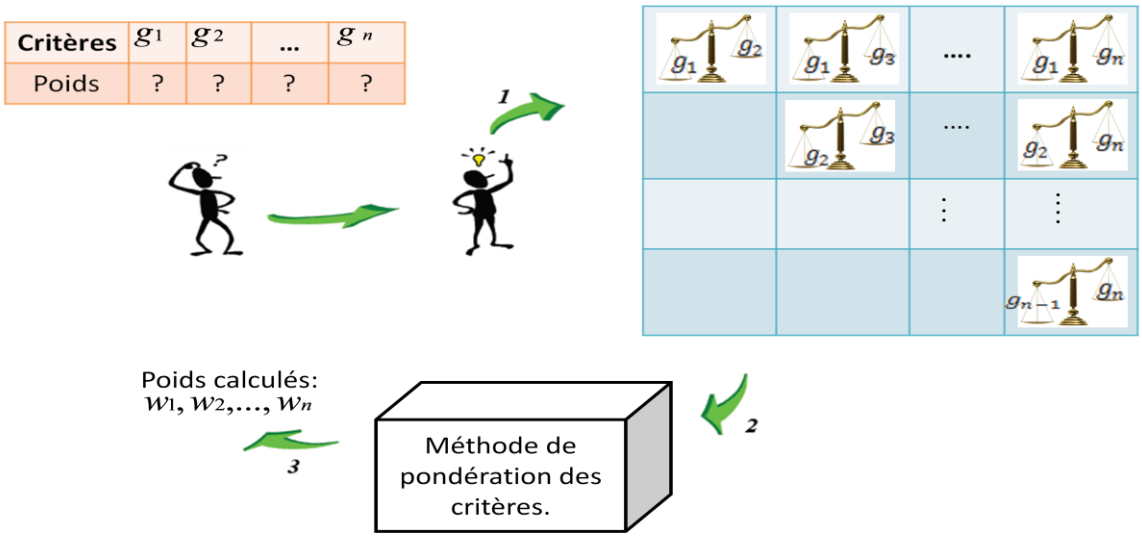


Figure 11: Principe de base d'une méthode de pondération des critères illustré sur une demi-matrice de comparaisons pour *n* critères à pondérer.

Nous avons présenté ci-après certaines méthodes de pondération des critères existent dans la littérature.

I.6.1.4.1. Approche basée sur l'analyse des valeurs propres (Eigen Values):

Principe de l'approche.

Dans [SAA77], l'auteur a introduit une méthode simple pour déterminer le poids de chaque critère basée sur la comparaison par paires des critères. Le principe de cette méthode est exécuté sur deux étapes :

En premier lieu, construire une matrice réciproque de comparaison *C(m,m)* (*m* :le nombre de critères): demander au décideur de comparer les critères par paires, nous assignons la valeur c_{ij} *et* c_{ji} à la matrice *C* comme suit :

Si le critère *k* est plus important que le critère *l*, alors on attribuer à C_{kl} un des numéros suivants $(1,2,3,4,5,6,7,8,9)$ selon le schéma suivant :

a) c_{ij} :le degré d'importance de critère i par rapport le critère$j$.

1: Égalité.

3: Faible.

5 : forte.

7 : démontré.

9 : Absolue.

2, 4, 6,8 : valeurs intermédiaires entre deux degrés d'importance précédentes.

b) $c_{ji} = \frac{1}{c_{ij}}$.

c) $c_{ii} = 1$

En deuxième lieu : calculer les valeurs propres (Algèbre Linéaire) $\theta_i : i = 1..m$ de cette matrice, et le vecteur propre de la plus grande valeur propre θ_{max} comme suit :

- Détermination des valeurs propres :

$$\mathbf{det}(\boldsymbol{C} - \boldsymbol{e\theta_{max}}) = \mathbf{0}.$$

$$avec: e = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

- Détermination de vecteur propre *V* (de plus grande valeur propre) :

$$\boldsymbol{CV} - \boldsymbol{\theta_{max}V} = \mathbf{0}.$$

Chaque élément *i* du vecteur *V* correspond le poids du critère *i*.

L'indice de cohérence d'une matrice basée sur l'analyse des valeurs propres :

Dans [SAA77], l'auteur a démontré qu'une matrice réciproque à valeurs positives admet une plus grande valeur propre θ_{max} , et que cette matrice est cohérente si et seulement si : $\theta_{max} = n$ $(n\ est\ le\ nombre\ des\ critères)$, mais avec les erreurs de jugement des décideurs, l'auteur a proposé un indicateur de cohérence (IC_{Saaty}) d'une matrice réciproque suivant :

$$IC_{Satty} = \frac{\theta_{max} - n}{1 - n}$$

I.6.1.4.2. La Moyenne géométrique sur les lignes (MGL) et la Moyenne Géométrique sur les Colonnes (MGC) :

Ou **R**ow **G**eometric **M**ean (RGM), respectivement **C**olumn **G**eometric **M**ean (CGM) en anglais. Ces méthodes exigent que la matrice de comparaison binaire doive être réciproque avec l'ordre. [LIM04].

Les deux formules suivantes permettent de calculer les poids des critères. La première, la moyenne géométrique sur les lignes et la deuxième, la moyenne géométrique sur les colonnes.

$$w_i = \frac{r_i}{\sum_{s=1}^{n} r_s}, \quad i = 1,2,..,n, \qquad with\ r_i = \left(\prod_{j=1}^{n} c_{ij}\right)^{1/n}$$

$$w_i = \frac{r_j}{\sum_{s=1}^{n} r_s}, \quad i = 1,2,..,n, \qquad with\ r_j = \left(\prod_{i=1}^{n} c_{ij}\right)^{-1/n}$$

I.6.1.4.3. La Moyenne Géométrique sur les Lignes et les Colonnes (MGLC).

Koczodaj et orolowiski [KOC99] ont proposé une généralisation des méthodes RGM et CGM à matrices non réciproques et non nécessairement cohérente. Ils offrent la plus proche matrice cohérente C^* à partir de minimisation de la formule de régression logarithmique des moindres carrés suivante : $\sum\left(\log\left(c_{ij}^*\right) - \log\left(c_{ij}\right)\right)^2$:

La matrice C^* est calculée à partir de moyennes géométriques des lignes et des colonnes de la matrice *C*. le résultat finale est :

$$c_{ij}^* = \sqrt{\frac{R_i^*}{R_j^*}\frac{G_j^*}{G_i^*}} \quad , \quad i,j = 1,2,\dots,n$$

$$avec \qquad R_i^* = \sqrt[n]{\left(\prod_{j=1}^{n} c_{ij}\right)} \quad et \quad G_i^* = \sqrt[n]{\left(\prod_{j=1}^{n} c_{ji}\right)}$$

Lorsqu'on applique n'importe quelle méthode précédente (EV, RGM, CGM) à la matrice cohérente C^*, elles donnent le vecteur de poids.

I.6.1.4.4. Approche basée sur la régression Logarithmique selon les moindres carrés : (R.L.M.C)

Pour la construction de la matrice *C*, tous les méthodes illustré ci-dessus : EV, RGM, CGM, RCGM. Il ***existe une seule*** opinion (un décideur).

La matrice de comparaison doit-être incomplète. En effet il est parfois difficile de comparer les critères entre eux. En plus, plusieurs opinions doivent pouvoir exprimées au niveau d'une comparaison (s'il existe plusieurs décideurs), la valeur de comparaison peut varier d'un décideur à l'autre.

Pour cela et d'autres raisons indiqués dans [LIM01], il est apparu d'autre méthodes de comparaison des critères qui prennent en considération ces raisons. Parmi ces méthodes, on peut présente l'algorithme R.L.M.C basée sur une régression logarithmique selon les moindres carrés.

La transformation logarithmique permet d'obtenir une relation linéaire entre une comparaison et la paire de poids qui lui est associé.

Soit $c_{ijk}, i \neq j = 1,2,\dots,n \; et \; k = 1,2,\dots,d$, l'opinion de $k^{ième}$ décideur pour comparer l'importance de critère i relativement à celle de critère j et d étant le nombre de décideurs.

Un décideur a le choix entre exprimer une seule opinion ou s'abstenir.

Pour avoir une notation homogène, posons :

$$\alpha_{ijk} = \begin{cases} \mathbf{1}, si\ le\ décideur\ exprime\ une\ opinion\ sur\ la\ comparaison\ c_{ij}. \\ \mathbf{0}, sinon. \end{cases}$$

Si $\alpha_{ijk} = 0$, la valeur de c_{ijk} est relativement fixée à une valeur positive ($c_{ijk} = c_0 > 0$).

Les auteurs de cette approche se basent sur la minimisation de la fonction de régression linéaire au sens des moindres carrées pour calculer les poids des critères. La transformation logarithmique permet d'obtenir une relation linéaire entre une comparaison et la paire de poids qui lui est associé.

La fonction à minimiser est donnée par la formule suivante :

$$f = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{d} \alpha_{ijk}\left(\log(c_{ijk}) - (\log(w_i) - \log(w_j))\right)^2.$$

Sous cette forme, les opinions symétriques sont supposées inverses de l'autre $\left(\alpha_{ijk}\, c_{ijk} = \alpha_{jik}\, \frac{1}{c_{jik}}; i,j = 1,2,\dots,n; k = 1,2,,\dots,d\right)$ et les opinions diagonales sont toutes supposées égales à 1.

La minimisation de la fonction des moindres carrées précédente conduit à la résolution des équations normale.

$$\theta_i \sum_{i\neq j}^{n} d_{ij} - \sum_{j\neq i}^{n-1} d_{ij}\,\theta_j = \sum_{j\neq 1}^{n}\sum_{k=1}^{d} \alpha_{ijk}\, b_{ijk}, \quad i = 1,2,\dots,n-1;$$

Avec :

$\theta_i = \log w_i, i = 1,2,\dots,n.$

$\theta_n = 0, (poids\ fixé\ arbtrairement\ \text{à}\ 1).$

$b_{ijk} = \log(c_{ijk}), i,j = 1,2,\dots,n; k = 1,2,\dots,d.$

$$d_{ij} = \sum_{k=1}^{d} \alpha_{ij\,k}^{2} = \sum_{k=1}^{d} \alpha_{ijk}\,, i,j = 1,2,\dots,n.$$

$(d_{ij} : nombre\ totale\ d^{'}\ opinionsexprimées\ sur\ la\ comparaison\ c_{ij})$.

Pour résoudre l'équation précédente nécessite de fixer arbitrairement un des n poids ($w_n = 1$).

La dernière étape consiste la normalisation pour que la somme des poids soit égale à 1, elle décrite par la formule suivante :

$$w_i = \frac{e^{\theta_i}}{\sum_{j=1}^{n} e^{\theta_j}}; i = 1,2,..,n$$

L'indice de cohérence de la matrice des comparaisons basée sur la régression Logarithmique selon les moindres carrés :

La cohérence de la matrice des comparaisons s'énonce comme suit : [LIM01]

$$\alpha_{ihk}\alpha_{hjk}\,c_{hjk} = \alpha_{ijk}\,c_{ijk}$$

I.6.2. Importances et interaction des critères selon le concept de capacité:

Dans un problème MCDA où l'on cherche à construire un modèle, il est nécessaire de prendre en compte de l'information supplémentaire, en particulier l'importance relative de chaque critère et le degré d'interaction entre les différents critères. La capacité permet de modéliser les importances et les interactions entre les critères en associant un poids à chaque **coalition de critères**. La valeur de la capacité seule $\mu(i)$ du critère i ne permet pas de rendre en compte l'importance d'un critère, mais également par tous les $\mu(A)$, tels que $i \subseteq A$.

I.6.2.1. La valeur de Shepley

La valeur de Shapley $\varphi(i)$ [GRA08] peut être calculée afin de donner une indication d'importance pour chaque critère (i) :

$$\varphi(i) = \sum_{A \subseteq N/i} \frac{(n - |A| - 1)!\,|A|!}{n!} [\mu(A \cup \{i\}) - \mu(A)].$$

Avec $|A|$: le cardinale de A.

$$\sum_{i=1}^{n} \varphi(i) = 1.$$

Cet indice devrait être une moyenne des quantités $\mu(A \cup \{i\}) - \mu(A)$, pour tout les groupes A possibles (y compris A=∅) [GRA06]. Si μ est additive alors $\varphi(i) = \mu(i)$.

Plus un critère est important, plus son indice de Shapley est grand.

I.6.2.2. Indice d'interaction:

Indice d'interaction entre deux critères *i* et *j* est la moyenne de la combinaison des critères *i* et *j*, en présence d'un groupe de critère *A*, pour tous les groupes *A* possibles (y compris $A=\emptyset$).

Indice d'interaction entre deux critères *i* et *j* est défini par la formule suivante [LON11].

$$I_{ij} = \sum_{A \subseteq N/\{i,j\}} \frac{(n-|A|-1)!\,|A|!}{(n-1)!} [\mu(A \cup \{i,j\}) - \mu(A \cup \{i\}) - \mu(A \cup \{j\}) + \mu\{A\}].$$

$\boldsymbol{I_{ij} > 0}$: Une interaction positive décrit la **complémentarité** entre les deux critères *i et j* i.e., pour qu'une alternative soit satisfaisante, elle doit satisfaisante sur les deux critères simultanément.

$\boldsymbol{I_{ij} < 0}$: *i et j* sont **redondants**, il suffit à une alternative d'être satisfaisante sur un des deux critères pour être satisfaisante.

$\boldsymbol{I_{ij} = 0}$: *i et j* sont **indépendants** i.e., les critères agissent indépendamment l'un de l'autre sur la qualité de l'alternative.

L'indice d'interaction a été étendu à plus de deux critères, pour chaque sous-ensemble $A \subset N$:

$$I(A) = \sum_{B \subset N \setminus A} \frac{(n-|B|-|A|)!\,|B|!}{(n-|A|+1)!} \sum_{C \subset A} (-1)^{|A \setminus C|} \mu(C \cup B).$$

I.6.2.3. Sugeno λ-measure :

Sugeno λ-measure est un cas particulier de mesure floue définie de manière itérative. Elle permet de calculer l'importance d'un ensemble des critères. Elle est définie comme suit :

Définition :

Soit : $X = \{x_1, x_2, .., x_n\}$ ensemble fini des critères (en MCDA),

Soit $\lambda \in [-1, +\infty[$: degré d'interaction entre les critères.

Sugeno λ-measure est une fonction μ: $2^X \rightarrow [0,1]$ à propriétés suivantes :

1. $\mu(X) = 1;\ \mu(\emptyset) = 0.$
2. $Si\ A, B \subseteq X\ i.e.\ A, B \in 2^X / A \cap B = \emptyset\ alors$:

$$\mu(\boldsymbol{A} \cup \boldsymbol{B}) = \mu(\boldsymbol{A}) + \mu(\boldsymbol{B}) + \boldsymbol{\lambda}\mu(\boldsymbol{A})\mu(\boldsymbol{B}).$$

   Cette formule permet de calculer l'importance de chaque groupe de critères.

3. En plus, λ satisfait la propriété suivante :

$$(\boldsymbol{\lambda}+\mathbf{1})=\prod_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}}(\mathbf{1}+\boldsymbol{\lambda}\mu(\boldsymbol{x_i})).$$

$$\mu(x_i), i = 1 \dots n: sont\ les\ poids\ des\ crit\grave{e}res\ x_i, \dots, x_n.$$

Un exemple illustratif pour calculer les importances des critères selon la méthode Sugeno λ-measure est présenté dans [AYU09] et [RAK07].

I.7. Conclusion :

Dans ce chapitre, nous avons défini quelques concepts et notions relatifs à l'aide à la décision et les structures de préférences existent dans ce domaine, ainsi que nous avons montré en détail le fonctionnement de quelques méthodes les plus connues concernant l'élicitation des préférences en aide multicritère à la décision.

Nous avons conclu que l'inférence des paramètres d'un modèle de préférence à partir des jugements d'un décideur reste un objectif commun, et un croisement entre les deux champs (Apprentissage automatique et Aide MultiCritère à la Décision.

Chapitre 02 : CLASSIFICATION DYNAMIQUE DE DONNÉES ÉVOLUTIVES :

II.1 Introduction :

DATA Mining que l'on peut traduire par «fouille de données », apparaît au milieu des années 1990 aux Etats-Unis comme une nouvelle discipline a pour objet d'extraction d'un savoir ou d'une connaissance à partir de grandes quantités de données. Le DATA Mining est apparue à l'interface de la statistique et des technologies de l'information : bases de données, Intelligence artificielle, Apprentissage automatique,…

Zekulin en donne la définition suivante: [FRE97]

Définition: « *Data mining is the process of extracting previously unkown, comperehensible, and actionable information from large data bases and using it to make crucial business decisions* ».

On distingue plusieurs familles de tâches réalisées en data mining, telles que la classification, la prédiction, la recherche d'association, le classement,… [TUF07]

Lorsque les connaissances évoluent dans le temps, il est important que le système d'apprentissage automatique soit capable de s'adapter en prenant en compte ces évolutions. Ce chapitre représente les fondements de la *classification dynamique* des données non-stationnaire. La classification dynamique est l'axe de l'apprentissage automatique qui s'intéresse aux problèmes de classification et modélisation adaptatives des données non-stationnaires. En classification dynamique les connaissances utiles à extraire (à apprendre) sans les classes susceptible d'évoluer dans le temps. Le temps est donc un aspect important à prendre en compte pour construire un classifieur plus performant. Cette évolution des classes peut être traduite par un déplacement, une scission, une rotation, suppression, apparition de nouvelles classes …etc.

Les objectifs des algorithmes de classification dynamique sont développés selon la disponibilité de données, le premier est *l'apprentissage supervisé* où le but de l'algorithme de classification est de construire en connaissant *a priori* les étiquettes d'appartenance des données aux classes. Le deuxième objectif est *L'apprentissage non supervisé* qui est l'inverse de premier où aucune information disponibles sur l'appartenance des données aux classes ainsi que le nombre des classes. Le troisième objectif est *l'apprentissage séquentiel*, dans ce cas la base d'apprentissage n'est pas disponible a *priori* ou incomplète, l'objectif de l'algorithme séquentiel est de pouvoir apprendre séquentiellement des connaissances supplémentaires au fur et à mesure que de nouvelles données sont acquises sans avoir à réaliser un apprentissage complet. [SAL12]. Le dernier objectifs est *l'apprentissage adaptatif*, dans ce cas il s'agit de la mise à jour du modèle de classification au fur et à mesure que les données se présentent.

II.2. Environnement statique et environnement dynamique :

Selon le type de système étudié, les données obtenues sur un système peuvent être statiques ou dynamiques lorsque leurs caractéristiques évoluent avec le temps. Une donnée statique est représentée par un point dans l'espace de représentation tandis qu'une donnée dynamique est représentée par une trajectoire multidimensionnelle. Dans ce dernier cas, l'espace de représentation possède une dimension supplémentaire qui est le temps.

De la même façon, les classes peuvent être statiques ou dynamiques. Les classes statiques sont représentées par des zones restreintes contenant des données similaires dans l'espace de représentation. Dans ce cas les paramètres de classifieur restent inchangés. Plusieurs méthodes existent pour traiter le cas des données statiques. On peut citer la méthode des K-Plus Proches voisins (K-PPV), les Séparateurs à Vaste Marge (SVM), le classifieur Bayésien, la méthode Fuzzy C-Means (FCM),…etc. ces méthodes sont détaillées dans le chapitre suivant. Cependant, plusieurs systèmes sont constamment en évolution entre les différents modes de fonctionnement, on parle alors de système évolutif [HAR10][AMA06] pour lesquels il est nécessaire d'utiliser des méthodes d'utiliser des méthodes de classification dynamiques.

II.3. Principe de suivi d'un système évolutif :

Dans [MER06], les auteurs ont défini un processus général pour suivre un système évolutive qui se base sur deux niveaux : Premièrement, il faut réaliser la modélisation de l'état du processus à chaque instant (figure 12) par l'extraction d'un vecteur représentatif de l'état de fonctionnement à partir de l'ensemble d'entrées et des sorties du système, un algorithme incrémental est réalisé pour estimer à chaque instant le modèle de linéaire du système suivi.

Après la modélisation de l'état du système, les connaissances apprises à partir les données, sont représentées par des classes à partir d'un algorithme de classification spécifique permettant de réaliser une modélisation en continue de modes de fonctionnement. [MER06]

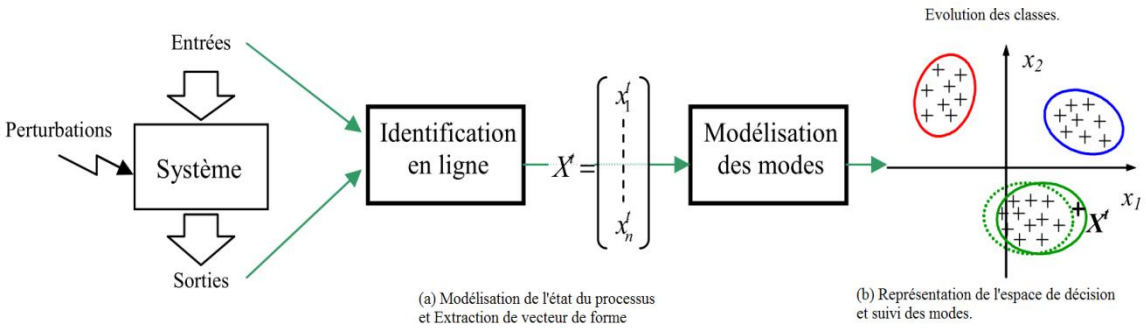


Figure 12. Modélisation de l'état de système et évolution de se mode de fonctionnement.

Dans la section suivante, on présente les phénomènes dynamiques qui se produisent dans un environnement non stationnaire.

II.4. Les données évolutives (non-stationnaires):

Les données stationnaires sont des données qui conservent leurs caractéristiques statistiques au cours du temps. Au contraire des données non-stationnaires qu'on peut les définissent comme suit : [AMA06]

Définition (Données non-stationnaires): les données non-stationnaires sont des données issues d'un processus dont le comportement est susceptible d'évoluer dans le temps. La non-stationnarité entraîne une évolution des caractéristiques des modèles de connaissances (classes) apprises de ces données. En classification, nous parlons de classes évolutives. Les données issues du monde réel sont non-stationnaires par ce que le monde est perpétuelle évolution.

Un des domaines représentatifs pour illustrer les données évolutifs est le domaine médical, où les données évoluent selon le cas du patient, par exemple dans [CAP06], l'auteur essaye d'analyser les crises d'épilepsie par la comparaison du signal EEG (**E**lectro**E**ncephaloGraphy) dans le cas de repos et dans une période critique. La figure 13 est un exemple de signal EEG dans une période précritique (340 à 350 secondes avant le départ d'une crise) et la période critique (de 350 à 360 secondes) illustre le signal EEG dans une période critique où la crise commence à la seconde 350 d'après les indications données par les médecins. On peut voir qu'elle se diffuse à un grand nombre de voie et que le signal semble plus ample. En plus, il y a des d'autres types de signaux dans cette figure, ce sont des artéfacts crées à cause de mouvement des yeux (artéfacts oculaires) ou à cause de fonctionnement de la machine (artéfacts hautes fréquences).

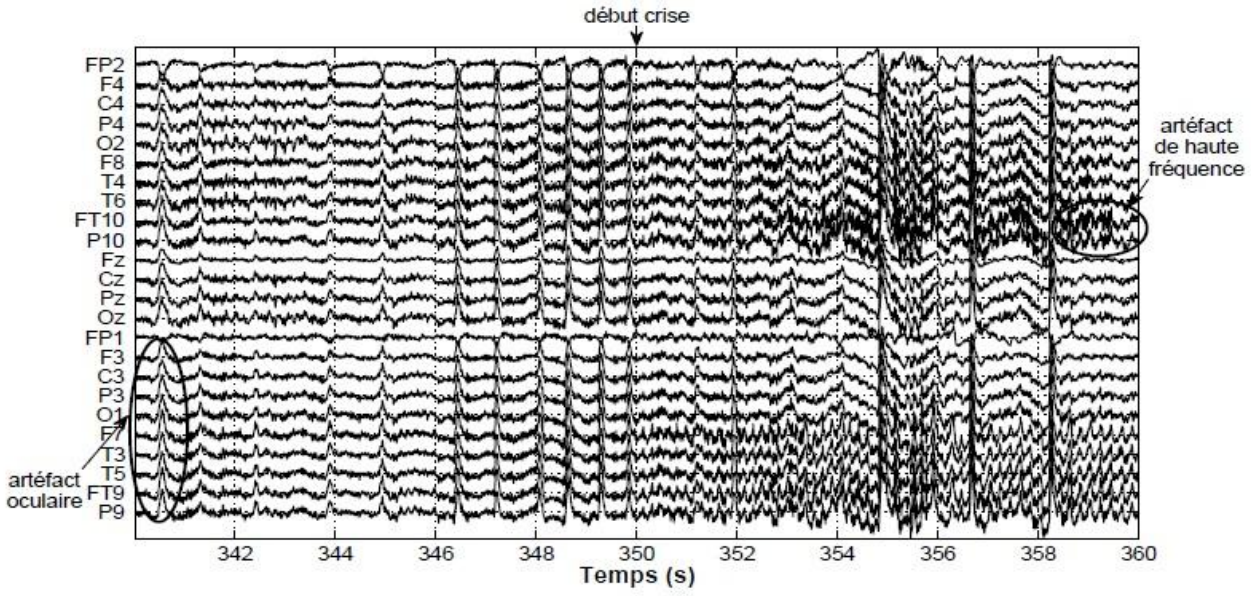


Figure 13: Exemple d'un signal EEG dans une période précritique suivi par un signal EEG dans une période critique.

II.5. Problématiques de la classification dynamique :

Dans [AMA06], l'auteur a distingué trois problématiques dans le domaine de classification dynamique, le premier c'est le problème de classification non supervisé, où aucune information sur les classes connue *a priori.* Le deuxième, concernant l'apprentissage incrémental (séquentiel) où les données enrichissent progressivement la base d'apprentissage. Le dernier lié à la modélisation adaptative qui concerne à la mise à jour du modèle de classification au fur et à mesure que les données se présentent.

II.5.1. Classification non supervisée :

Dans ce cas, aucune information sur les classes d'appartenance des données n'est disponible *a priori*, ainsi le nombre de classes ne sont connues *a priori*. La méthode de classification a pour objectif la partition des données et effectuer les données aux classes. L'attribution des données aux classes se fait généralement dans ce cas par un critère de ressemblance. La détermination de ce critère entre les données dépend de la nature du problème à traiter.

II.5.2. Apprentissage incrémental :

L'apprentissage incrémental correspond à un système capable d'extraire des connaissances supplémentaires à partir des nouvelles données sans avoir à réaliser un apprentissage complet. Dans ce cas la base d'apprentissage n'est pas complète *a priori* où les données enrichissent progressivement la base d'apprentissage [SAL12]. La classification consiste à incorporer chaque donnée dés qu'elle se présente, afin d'actualiser le modèle de connaissances (classes). Cette tâche doit se faire en utilisant les connaissances extraites des données déjà acquises et sans attendre que toutes les informations contenues dans les données futures sont disponibles. C'est l'enjeu des techniques de l'apprentissage incrémental.

II.5.3. Modélisation adaptative :

A cause de non stationnarité des données, il s'agit de la mise à jour du modèle de classification au fur et à mesure que les données se présentent. Cette tâche est réalisée par de règles d'adaptation tenant compte sur les changements engendrées sur la partition (fusion, scission, déplacement des classes,…) au cours de l'incorporation des nouvelles données dans les classes. Il est donc nécessaire de disposer de règle d'apprentissage continue et de suivi l'évolution. Cette tâche est plus importante dans le contexte de classification dynamique. [AMA06]

II.6. Phénomènes dynamiques en environnement non-stationnaire:

Cette section détaille les mécanismes généraux engendrés par l'incorporation des nouvelles données dans la partition et les phénomènes dynamiques causés par la non-stationnarité des données. Avant d'exposer ces mécanismes, nous allons tout d'abord donner

les définitions d'une classe et d'une partition dynamique dans le contexte de notre étude. [AMA06]

Définition (une classe):

Une classe est définie comme un ensemble (ou regroupement homogène) de données similaires. La similarité est évaluée d'un point de vue de la proximité géométrique des données. Lorsque les données d'une classe évoluent dans le temps, cette classe est dite évolutive. Son modèle doit donc être adapté.

Définition (Partition dynamique):

Une partition dynamique représente l'ensemble des regroupements homogènes des données non-stationnaires. Elle est associée à un modèle de connaissances (i.e. Classes) susceptible d'évoluer dans le temps en fonction de l'incorporation de nouvelles données.

Dans la suite de cette section, on présente des divers mécanismes se produisent à cause de la non stationnarité des données. Ces mécanismes concernent, *l'apparition de nouvelles classes, l'évolution de classe, la fusion, la scission et l'élimination de classes parasites ou obsolètes.*

II.6.1. Apparition de nouvelles classes:

La classification dynamique de données permet de résoudre le problème de l'incomplète de la base de connaissance. Lorsque des nouvelles données sont acquises dans une même région ou différentes régions non occupées par des classes existantes. Ces situations entrainent d'apparition une ou des nouvelles classes chaque classe contient des données similaires.

La création des classes dans ces situations sont déterminées par les deux points suivants :

- Premièrement, construire les données suffisamment proches pour former chaque classe, cela est défini généralement par un critère de similarité (mesure de distance).
- La seconde, consiste à calculer les modèles de nouvelles classes crées à partir de la mise œuvre de la procédure d'initialisation respectant la distribution des données.

La partition dynamique mettre à jour par l'insertion des nouvelles classes au cours de ce mécanisme de création.

II.6.1.1. Apparition d'une ou plusieurs classes dans un espace vide :

Lorsque l'espace de classification est vide c-à-dire, au départ de l'apprentissage où les premières données arrivent telles qu'aucune classe n'existe encore et si les données sont similaires, elles constituent une classe, sinon plusieurs classes sont construites. La figure suivante présente une situation de l'apparition d'une ou des nouvelles classes dans un espace vide.

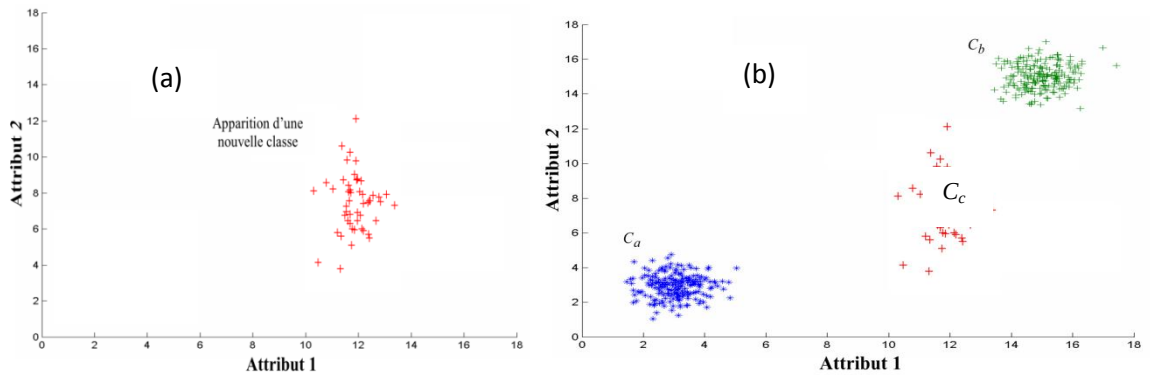


Figure 14 :(a)- apparition d'une classe dans un espace de classification vide. (b): Apparition de plusieurs classes dans un espace de classification vide.

Dans le cas où une classes ou plusieurs déjà crées dans l'espace de classification évoluent d'un état normal vers un autre état normal, soit d'une façon brutale ou d'une façon successive.

II.6.1.2. Changement brutale des données :

Une évolution est brutale si des nouvelles données similaires et suffisantes apparaissent soudainement à l'instant *t+1* n'est pas affectée à la même classe que ces données apparaissent à l'instant *t.* la figure représente le phénomène de changement brutale des données. La flèche identifie le sens de l'évolution. [LUR03]

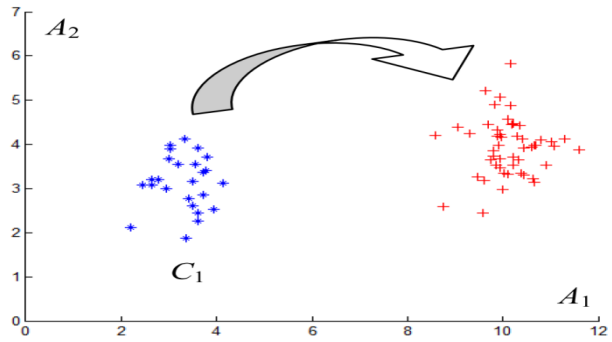


Figure 15:Apparition de classe après un changement brusque.

II.6.1.3. Changement progressive des données :

Par opposition à la situation précédente, un changement progressive se caractérise par l'évolution rapide de plusieurs données i.e. ils existent des transitions entre une classe de début de l'évolution et une autre classe de la fin de l'évolution. La figure représente le phénomène de changement progressif des données où le numéro de point indique la séquence de l'évolution.

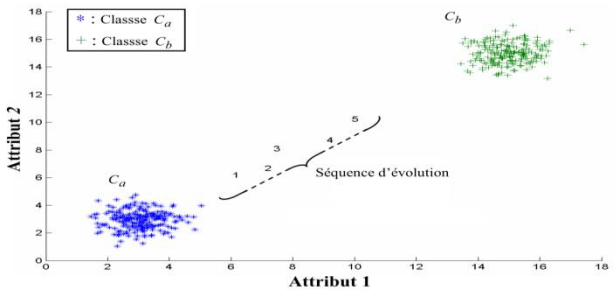


Figure 16:apparition de classe après une évolution rapide.

II.6.2. Evolution de la partition dynamique :

II.6.2.1. Modification locale de classes :

La mise à jour de modèle de classification après l'affectation de chaque nouvelle donnée conduira à une modification locale de la classe. Ces modifications se traduites par changement de structure géométrique de classe dans leur région de définition. Il existe plusieurs phénomènes de modification locale de la classe par exemple : *grossissement, Rotation,* …comme montre la figure 17, où le grossissement de la classe se produire tel que les nouveaux point indiqués en rouges affectés à la classe, et les points bleu représentent les points initiaux de l'ensemble d'apprentissage, les flèches indiquent le sens de grossissement. En plus le contour de la classe peut être diminué (*rétrécissement*).

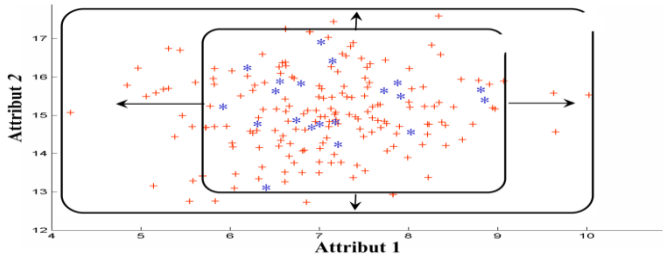


Figure 17:Grossissement de classe.

II.6.2.2. Déplacement de classes:

Le déplacement des classes est dû au non stationnarité des données. Il est réalisé par le changement successif des classes, où les contours des classes restent inchangeables durant le changement progressif. La figure 18 [BOUG07] représente le déplacement de la classe C_1 vers la classe C_2 puis vers la classe C_3, enfin vers la classe C_4.

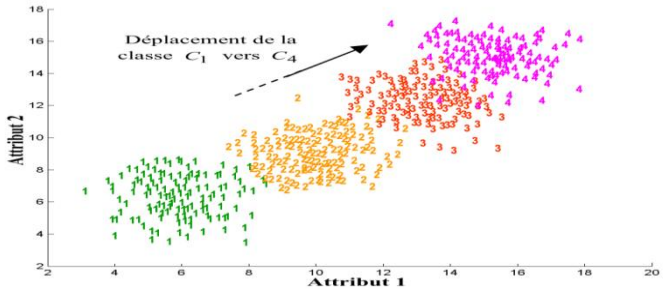


Figure 18:Déplacement de classe.

II.6.3. Fusion d'informations mutuelles :

Dans un environnement non stationnaire, des interactions entre deux ou plusieurs classes dans une même partition dynamique. Pour cela, plusieurs phénomènes peuvent se produire tels que, Fusion de deux classes, Scission d'une classe.

II.6.3.1. Fusion des classes :

Une fusion de deux ou plusieurs classes peut se produire lorsque ces classes partagent des informations mutuelles. En conséquence nous allons avoir une seule classe résultante. La détection d'informations mutuelles peut être effectuée par un *critère de ressemblance* [AMA06]. La détermination du modèle résultant de la fusion sans effectuer le réapprentissage sur toutes les données nécessite la mise en œuvre de *règles de fusion*, mais cette tâche est souvent compliquée [AMA06]. Le principe de fusion de classes est représenté dans la figure

19 [BOUG07], où une fusion de la classe C_1 et la classe C_2 en seule classe C_5 dû à la non stationnarité des données.

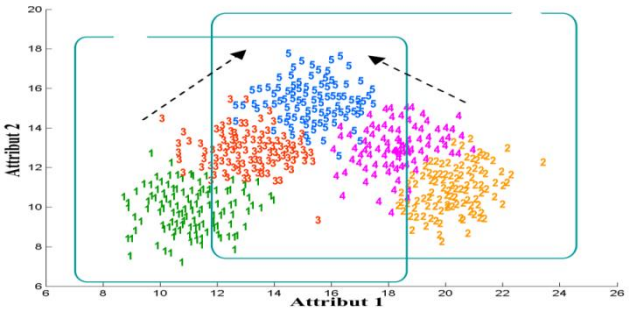


Figure 19:Fusion de classes.

5.3.2 Scission de classe:

Le phénomène inverse de la fusion des classes est la division des classes au cours de l'incorporation des nouvelles données dans l'environnement non stationnaire si nécessaire. La scission des classes est réalisée lorsque la distribution des données n'est plus homogène pour une classe. [HAR10]

Par exemple, les données caractérisant le mode de fonctionnement d'un processus, peuvent être initialement mélangées avec des données parasites causées par des perturbations extérieures. Lors de l'évaluation de ce mode de fonctionnement. La classe de départ peut se scinder en deux suite à une séparation de la classe normale de la classe parasite. [AMA06].

La figure 20 représente le phénomène de scission de classe en deux autre classes à cause de non stationnarité des données [BOUG07].

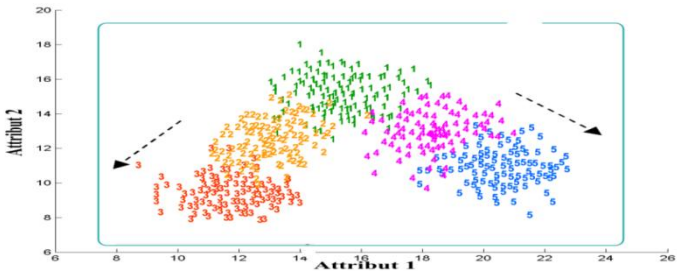


Figure 20:Scission de classe.

II.6.3.3. Suppression de connaissances obsolètes ou parasites :

Comme on le voit, les connaissances extraites à partir de données évolutives peuvent être inutiles à une instant donnée crée à cause du bruit et des perturbations extérieurs

(connaissances *parasites*). En plus, les connaissances ne soient pas valables à l'instant courant sont des connaissances *obsolètes*. Ces connaissances doivent être éliminées de l'espace de classification.

Un classifieur dynamique doit donc comporter une procédure d'élimination des classes non représentatives (classes parasites ou obsolètes), cela peut être défini par un *critère de cardinalité* de points ayant un niveau acceptable de similarité entre eux. Ainsi, la détection de la classe parasite dépend fortement du contexte de l'application. [AMA06].

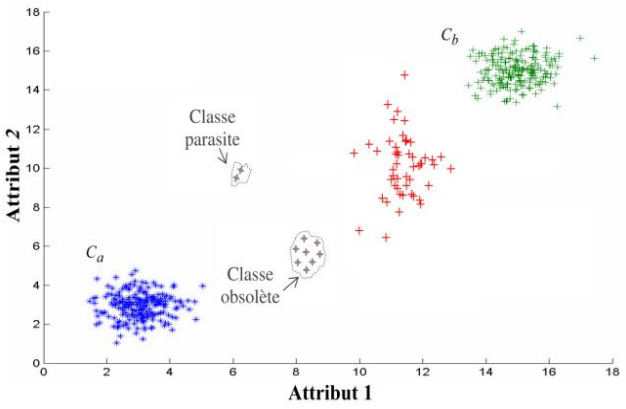


Figure 21 : Suppression de classes parasites et obsolètes.

II.7. Conclusion :

La classification dynamique est considérée comme étant un nouvel axe de recherche prometteur pour les procédures d'apprentissage automatique et d'extraction de connaissance. Ce type de classification touche plusieurs domaines surveillance vidéo « suivi de mouvements de la cible », audit de réseau informatique «détection des intrusions », diagnostic médicale « suivi l'état du patient », en l'industrie « suivi le mode de fonctionnement d'une machine », …etc.

Dans ce que passé, nous avons exposé les principes et les objectifs de classification dynamique de données évolutives, ainsi que les problématiques et les difficultés causées par la non stationnarité des données et les phénomènes engendrées dans un environnement dynamique. ***Le suivi de système évolutif nous permet d'améliorer le fonctionnement de ce système.*** Dans le suivant, nous allons présenter les différentes méthodes et algorithmes les plus abordés dans la littérature pour résoudre cette problématique de la classification dynamique.

Chapitre 03 : PANORAMA DES MÉTHODES DE CLASSIFICATION :

III.1. Introduction :

Une méthode de classification classique a pour objectif d'associer chaque élément de l'ensemble des *individus stationnaires* $X=\{x_1,x_2,....,x_n\}$, à l'une des classes statiques $C=\{C_1, C_2,..., C_k\}$, selon un critère de ressemblance. Un individu x_i est associé à la classe C_j si et seulement s'il est de la même nature des individus déjà regroupés dans la classe C_j. Ce qui explique une stationnarité permanente du modèle de classification au cours du temps. L'introduction de la propriété « *évolution ou non stationnarité* », où X devient un ensemble d'individus non-stationnaires, nous permet d'introduire la méthode de classification dynamique.

Développer un algorithme de classification dynamique est un travail assez complexe vu la difficulté de manipuler des données non-stationnaires, afin de les affectées dans des classes également non stationnaires au moyen d'une base d'apprentissage parfois inconnues ou incomplète, c'est-à-dire, l'étiquetage des données est inconnue *a priori.*

Peu d'algorithmes de classification dynamique sont développées jusqu'à présent, la plupart de ces algorithmes basés sur des méthodes de classification avec apprentissage supervisé (SVM, modèle de mélange gaussien, réseaux de neurones, …etc), ces méthodes sont détaillées dans le présent chapitre.

Dans le chapitre précédent, nous avons décrit les objectifs et les difficultés liées aux méthodes de la classification dynamique de données évolutives. La détermination d'une approche adéquate est notre objectif principal dans ce mémoire de magister. L'étude préalable de l'ensemble de techniques et d'algorithmes de classification les plus connus sont présentés dans ce chapitre.

On peut identifier ces méthodes en deux variantes : les méthodes de classification dites « *supervisées* » et méthodes de classification dites « *non-supervisées* », pour cela, ce chapitre est divisé en deux parties, en premier lieu, nous présentons l'ensemble des méthodes de

classification avec apprentissage supervisé et le deuxième partie présente des méthodes de classification avec apprentissage non supervisé.

Avant de présenter les méthodes de classification, nous décrivons quelques concepts importants dans le domaine de classification.

III.2. Définitions de base :

III.2.1. Degré de ressemblance (Indice de similarité):

Indice de similarité (ressemblance) sur *I* est une application : $Sim : IxI \rightarrow \Re^+$ ayant la propriété suivante :

$\forall x \in I, \forall y \in I, avec\ x \neq y$

$$Sim(x,y) = Sim(y,x) \tag{3.1}$$

$$Sim(x,x) = Sim(y,y) = Sim_{max} > Sim(x,y)$$

III.2.2. Degré de dissemblance (Indice de dissimilarité):

L'indice de dissimilarité (dissemblance) sur *I* est une application : $Diss : IxI \rightarrow \Re^+$ ayant la propriété suivante :

$$Diss(x,y) = Diss(y,x) \tag{3.2}$$
$$Diss(x,x) = 0.$$

Seuls les indices de similarité *Sim* seront considérés par la suite, car à un indice de dissimilarité *Diss* peut toujours être associé l'indice de similarité défini par :

$$Diss(x,y) = Sim_{max} - Sim(x,y) \tag{3.3}$$

III.2.3. Notion de distance :

Généralement, l'indice de similarité entre deux objets *x, y* représenté par la distance de *Minkowski d(x,y)*comme suivant :

$$\forall x, y \in I^p,\ d(x,y) = \left[\sum_{i=1}^{p} |x_i - y_i|^q\right]^{\frac{1}{q}} \tag{3.4}$$

La distance euclidienne et la distance de Manhattan sont deux cas particuliers de l'indice de Minkowsky :

$$Si, q = 2;\ \ d(x,y) = \sqrt{\sum_{i=1}^{p}(x_i - y_i)^2}\,, distance\ euclidienne. \tag{3.5}$$

$$Si, q = 1;\ \ d(x,y) = \sum_{i=1}^{p}|x_i - y_i|,\ \ distance\ de\ Manhattan \tag{3.6}$$

III.2.4 Notion de partition :

Les méthodes de classification couramment utilisées peuvent être classées en deux grandes familles suivant la structure de classification produite : *partition dure* (crisp-partition) ou *partition floue (fuzzy partition).*

III.2.4.1. Partition dure :

Considérant un ensemble fini des données : $X = \{x_1, x_2, \dots, x_m\}, et\ soit\ C = \{C_1, C_2, \dots, C_c\}$ une partition considérées de *c* classes c'est-à-dire des sous ensembles non vides de *X* et μ une fonction dite fonction d'appartenance qui associe à tout couple (X, C).

C est une partition dure si les conditions suivantes sont vérifiées :

1. $\forall\,(C_k, C_l) \in CxC, C_k \cap C_l = \emptyset.$
2. $C_k \in C, \bigcup_{k=1}^{c} C_k = X.$ (3.7)

La première condition signifie que chaque donnée de *X* est effectuée à une et une seule classe, la deuxième condition traduit que les données de toutes classes recouvrent l'ensemble. Dans ce cas la fonction d'appartenance de l'élément x_i à la classe C_k prend deux valeurs binaires (0 ou 1).

$$\mu_{ik} = \begin{cases} 1 & si\ \ x \in C_k \\ 0 & sinon \end{cases} \quad \forall i = 1..m\ \ et\ \ k = 1..c. \tag{3.8}$$

II.2.4.2. Partition floue :

Le concept de partition floue provient de la théorie des ensembles flous que nous avons présentés dans le premier chapitre. Dans ce cas les valeurs de la fonction d'appartenance μ incluses dans l'intervalle [0,1], une partition floue de *X* est définie telle que :

1. $\forall\, x_i \in X\ ;\ \sum_{k=1}^{c} \mu(x_i, C_k) = 1$

2. $\forall C_k \in C; \quad 0 < \sum_{i=1}^{m} \mu(x_i, C_k) < m$ (3.9)

La première condition traduit que chaque donnée de X peut appartenir à plusieurs classes avec différent degrés d'appartenance. La deuxième condition signifie toute classe contient au moins une donnée avec un degré d'appartenance non nul. [AMA06]

III.3. Méthodes de classification avec apprentissage supervisé:

Comme leur nom l'indique, ces méthodes sont basées sur l'apprentissage supervisé. L'apprentissage supervisé consiste à établir des règles de classification à partir de l'ensemble d'apprentissage connu *a priori*. Donc le principe de la classification supervisée consiste à regrouper ensemble des données étiquetées à leurs classes connues. Dans ce cas, le modèle de chaque classe est représenté par une fonction d'appartenance qui détermine la valeur d'appartenance d'une forme à une classe. La classification non supervisée cherche à former des clases dans un ensemble d'observations non étiquetées pour lesquelles aucune connaissance sur leur appartenance aux différentes classes. Plusieurs méthodes de classification supervisées existent dans la littérature, dans la section suivante, nous décrivons quelques méthodes les plus connues et les plus utilisées.

III.3.1 Classification Bayésienne :

Dans le cadre de l'apprentissage Bayésien, nous retrouvons plusieurs types de classificateurs : *classificateur optimal de Bayes, classificateur de Bayes naïf, classificateur de Gibbs,…*

Dans cette partie nous allons présenter le classificateur optimal de Bayes qui est la base d'autres méthode. [BEN09]

Classificateur optimal de Bayes :

Comme son nom l'indique, la classification bayésienne repose sur la théorie de Bayes. La classification bayésienne consiste à calculer pour chaque vecteur forme x_i la probabilité conditionnelle $P(C_k/x_i)$ pour chaque des classes $(C_1, C_2, \ldots, C_c)$ à l'aide de la règle de Bayes (l'équation 3.10):

$$P(C_k/x_i) = \frac{P(C_k).P(x_i/C_k)}{\sum_{l=1}^{c} P(C_l).P(x_i/C_l)} \quad (3.10)$$

Où :

$P(C_k/x_i)$: Est la probabilité d'appartenance de x_i à la classe C_k, cette probabilité est appelée *probabilité a posteriori.*

$P(C_k)$: La probabilité d'apparition de la classe C_k, probabilité *a priori.* Cette probabilité peut être calculée comme le rapport entre le nombre d'échantillons appartenant à la classe C_k et le nombre total d'échantillons.

$P(x_i/C_k)$: La densité de probabilité conditionnelle de x_i sachant que la classe C_k est vraie. Cette probabilité peut être effectuée soit :

- par estimation, soit

- par approche directe.

L'approche par *estimation* repose sur les lois de distribution de probabilités comme la loi normale (gaussienne) et la loi uniforme.

L'approche directe utilise des procédés qui convergent vers les distributions de chacune des classes sans faire aucune hypothèse préalable [OUK97]. Parmi ces procédés, on peut citer la méthode des K Plus Proche Voisins (K-PPV) et les méthodes à base de fonctions noyaux.

Si les probabilités *a priori* $P(C_k)$ et la densité de probabilité conditionnelle $P(x_i/C_k)$, la règle de décision de classer l'individu x_i à la classe C_k est donnée sous forme suivante :

$$x \in C_k \quad ssi \quad P(C_k/x_i) > P(C_l/x_i) \qquad \forall k \neq l\,.$$

III.3.2. Méthode des K-Plus Proche Voisins K-PPV :

La Méthode des K-Plus Proches Voisins *K-PPV* (K-Nearest Neighbors *K-NN* en anglais) est une méthode non paramétrique et supervisée largement utilisée en classification. Le principe de cette méthode est simple et se base sur le regroupement d'individus en fonction de leur voisinage. Donc le *K-PPV* repose sur l'établissement d'une règle de distance entre l'individu ***x*** à classifier et les individus de la base d'apprentissage qui contient leurs classes d'affectation. L'individu ***x*** est affecté à la classe la plus représentée parmi ses *k* plus proche voisin, où *k* est un nombre fixé *a priori*. (la distance *euclidienne*, la distance de et celle de *Mahalanobis* sont des exemples de métriques largement utilisées).

L'algorithme générique de classification d'un nouvel individu par la méthode des K-PPV est illustré comme suit : [COR03]

Algorithme de classification par K-Plus proche Voisins :

Entrée : x : le point à classifier, **K :** le nombre de voisins.

Ensemble d'individus et leurs classes **C**.

Pour chaque : individu **(y,C)** de l'ensemble d'apprentissage **Faire** :

- Calculer la distance **D(y,x)**.

Fin pour :

Dans les **K** points les plus proches de **x**.

- Compter le nombre d'occurrence de chaque classe

Attribuer **x** à la classe qui apparaît le plus souvent.

Pour illustrer cet algorithme, soit l'exemple de la figure 22 [BIS06] de deux classes (classe des points rouges et la classe des points bleus) avec *K=2*. Dans cet exemple les 2-Plus proches voisins d'élément ***x*** à classer sont les points rouges, donc l'élément ***x*** est affecté à la classe des points rouges.

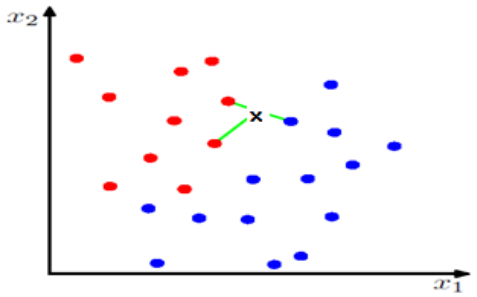


Figure 22:méthode des 2-Plus Proches Voisins (2-ppv).

III.3.3. Les séparateurs à vaste Marge (SVM) :

Les séparateurs à vaste marges ou SVM (Support Vector Machine, en anglais) ont été proposés dans les années 90 par *Vladimir Vapnik* [VAP95]. Les SVM ont été introduite initialement pour le problème de classification binaire par apprentissage supervisé i.e. les données traitées appartiennent à deux classes seulement, l'objectif principal des SVM est de rechercher le *meilleur hyperplan* séparant les données en deux classes et maximiser la distance entre ces deux classes (*notion de marge*). Pour cela *les SVM utilise seulement les points plus proches* c'est-à dire les points de la frontières entre les deux classes des données

parmi l'ensemble total d'apprentissage, ces points sont appelés les *vecteurs de support, figure 22* . Il est évident qu'il existe une infinité d'hyperplan capable de séparer les deux classes, mais la propriété remarquable des SVM est que cet hyperplan doit être optimal. *L'hyperplan optimal* est celui qui maximise la marge.

Dans le cas d'un nouvel individu à classifier est donnée par sa position par rapport à l'hyperplan optimal. Dans la figure 23 [COR03], l'élément *x* sera classifié dans la catégorie des « carrés ».

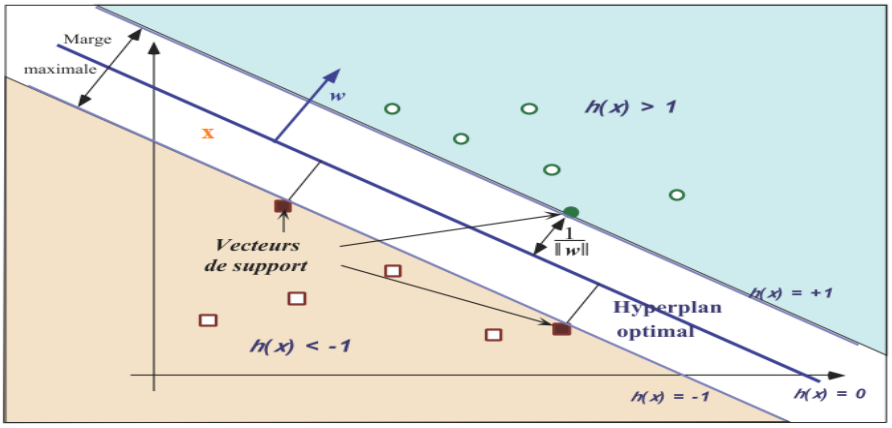


Figure 23: la géométrie d'un hyperplan.

On distingue deux modèles des SVM:

- Les cas linéairement séparable (figure 23)
- Les cas non linéairement séparable. (figure 24) [DJE12]

III.3.3.1. SVM linéaire :

Un classificateur est dit linéaire lorsqu'il est possible d'exprimer sa fonction de décision par une fonction linéaire.

Supposons donnée l'échantillon d'apprentissage : $S = \{(x_i, \mu_i)\}$ où, $i = 1..m$; $x_i \in \Re^d$ et $\mu_i = \{-1, +1\}$

On suppose qu'il existe une séparatrice linéaire permettant de distinguer les données positives (étiquetées +1) et des données négatives (étiquetées -1), les points *x* appartient à l'hyperplan optimal satisfont : $h(x) = w^t.x + b = 0$, où w est vecteur normal de l'hyperplan optimal. Donc la recherche d'un hyperplan optimal revient à chercher une fonction hypothèse : $h(x) = w^t.x + b$ telle que :

$$w^t.x_i + b \begin{cases} > 0 & \Rightarrow \mu_i = +1 \\ < 0 & \Rightarrow \mu_i = -1 \end{cases} \text{[COR03].} \quad (3.11)$$

III.3.3.2. SVM non linéaire (Fonctions Karnels) :

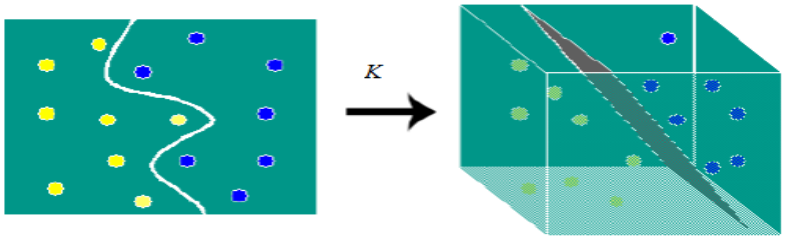


Figure 24: SVM non linéaire.

Dans le cas non linéairement séparable, l'idée des SVM consiste à transformer non linéaire des données vers un espace linéairement séparable où les données transformées seront linéairement séparable, comme on le voit sur la figure 24, l'espace initial (*l'espace de représentation*) est de 2 dimensions, alors que l'espace d'arrivé est de 3 dimensions (*l'espace de re-description*)[MOH06]. La transformation est réalisée via une fonction dite fonction noyau (*Kernel*).

Plusieurs fonctions noyau sont connues et il revient à l'utilisateur des SVM d'effectuer des tests pour déterminer celle qui convient le mieux pour son application. On peut citer les exemples de noyaux les plus connus et les plus utilisés suivants,

Fonction linéaire : $K(x, x') = x.x'$

Fonction polynomiale de degré p: $K(x, x') = (1 + x.x')^p$

Fonction à base radiale (RBF) : $K(x, x') = exp\left(\frac{\|x-x'\|^2}{2\sigma^2}\right)$

Fonction sigmoïde : $K(x, x') = \tanh(\alpha(x.x') - b)$

III.3.3.3. Approches Multi-classes pour les SVM :

D'abord les SVM sont développées pour le problème de classification binaire. en 1998, les SVM a été étendues aux problèmes multi-classes (nombre de classes c>2), pour cela deux approches existent :

III.3.3.3.1. Approche une contre Reste (1vsR) :

Le principe de cette méthode est de construire pour chaque classe un hyperplan qui discrimine toutes les classes, dans ce cas il y a ***c****(c : le nombre total de classes)* hyperplans possibles. Cette approche est illustrée dans la figure ci-dessous. [BOU11]

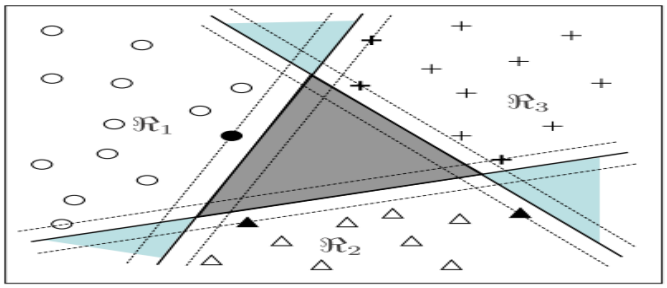


Figure 25: Séparateurs un contre reste (1vsR) de trois régions.

III.3.3.3.2. Approche une contre une (1vs1) :

Consiste à utiliser un classificateur pour chaque paire de classes, donc la méthode 1vs1 discrimine chaque classe de chaque autre classe. Dans ce cas il y a : ***c(c-1)/2*** hyperplans possibles. (figure 25).

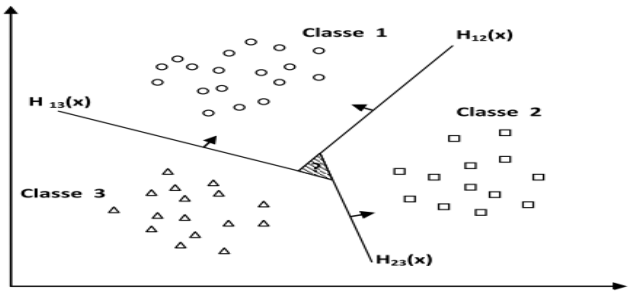


Figure 26: Séparateurs un contre un (1vs1) de trios régions.

III.3.4. Réseaux de neurones Artificiels:

III.3.4.1. Neurone formel:

Les réseaux de neurones sont des outils de l'intelligence artificielle, capables d'effectuer des opérations de classification des données. Les réseaux de neurones artificiels ou réseaux connexionnistes sont des réseaux dont l'architecture est inespérée de celle de réseaux de neurones biologiques. Un réseau de neurone est constitué d'un ensemble d'unités (ou neurones) ces unités sont reliées par des canaux de communication (les connexions, aussi appelées synapses d'après le terme biologique correspondant) qui transposent des données

numériques. Les unités peuvent agir uniquement sur leurs données locales et sur les entrées qu'elles reçoivent par leur connexion.

La première version de neurone formel est celle de McCulloch et x.pitts en 1943. Ils ont proposé le modèle de neurones formel qui se voit comme un opérateur effectuant une somme pondérée de ces entrées d'une fonction d'activation, le neurone formel est représenté par figure 27. [GUE09]

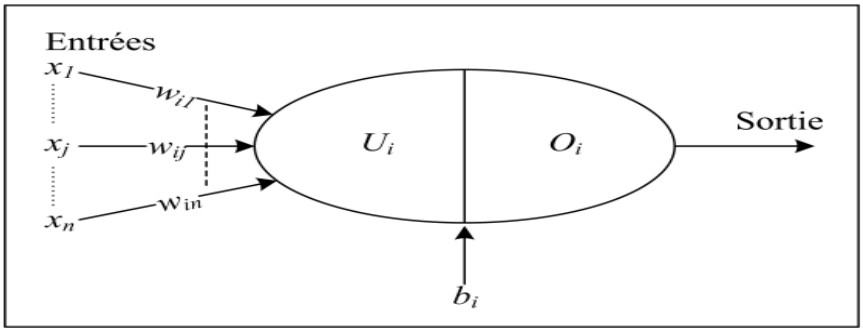


Figure 27: Modèle de base d'un neurone formel.

x_j : *représente l'entrée j connectée au neurone i.*

b_i : représente le biais (constante) : c'est le seuil interne du neurone *i*.

w_{ij} : *désigne le poids de la connexion reliant l'entréej au neurone.*

U_i : *représente la somme pondérée des entrées du neurone, elle est donnée par*:

$$U_i = \sum w_{ij} . x_j + b.$$

$O_i = f(U_i)$: *est la sortie du neurone et f sa fonction d'activation.*

III.3.4.2. Les fonctions d'activation :

Il existe dans la littérature plusieurs fonctions d'activations utilisées dans le cadre des réseaux de neurones. Le choix entre ces fonctions dépend de l'application considérée. On peut citer ces fonctions d'activation les plus connues :

Fonction Linéaire : $f(U_i) = U_i$.

Fonction sigmoïde : $f(U_i) = \frac{1}{1+e^{U_i}}$ *valeur dans* [0,1]

Fonction gaussienne $f(U_i) = \exp\left(\frac{-U_i}{2.\delta^2}\right)$ δ: $l'ecat\ type\ de\ la\ fonction\ gaussienne.$

Fonction d'Heaviside : $f(U_i) = \begin{cases} 1 & si\ U_i \geq 0 \\ 0 & sinon \end{cases}$

Fonction signe : $f(U_i) = \begin{cases} +1 & si\ U_i \geq 0 \\ -1 & sinon \end{cases}$

Fonction tangente hyperbolique (***tanh***) : $f(U_i) = \frac{e^{U_i} - e^{-U_i}}{e^{U_i} + e^{-U_i}}$ $valeur\ dans\ [-1, +1]$.

III.3.4.3. Types de réseaux de neurones :

L'ensemble des unités de réseau de neurones sont organisées sous forme de niveaux différents appelés *couches du réseau* où les neurones appartient à même couche possèdent les mêmes caractéristiques et utilisent le même type de fonction d'activation. Dans la partie suivante, nous présenterons les deux types de réseaux de neurones les plus populaire : le perceptron Multi-Couches (PMC) et le réseau à Fonction de Base Radial (RBF: Radial Basis Function).

III.3.4.3.1. Perceptron Multi-Couches (PMC) :

Les perceptrons multi-couches (MLP : Multi Layer Perceptron en anglais) sont les réseaux les plus populaires et les plus simples. Ce sont des réseaux à propagation directe sans cycle, avec au moins une couche cachée, les neurones sont généralement complètement connectés, où chaque neurone élémentaire est connecté à l'ensemble des neurones de la couche qui suit immédiatement celle à laquelle il appartient. Les neurones participant à la couche de sortie sont appelés « neurones de sorties ». La figure suivante schématise un MLP avec une couche cachée. [HAM08]

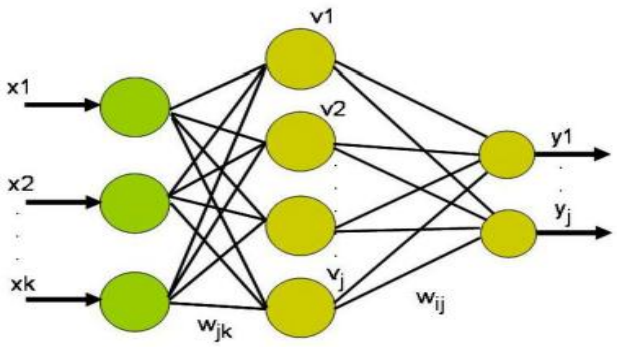


Figure 28: Perceptron Multi Couches (MLP).

III.3.4.3.2. Réseaux à fonctions de base radiale (RBF) :

En anglais, Radial Basis Function Network, sont des réseaux de type *feedforward* (réseaux à propagation avant) sont des réseaux de neurones supervisées.ils comportent trois couches, une couche d'entrée, une couche cachée qui contient les neurones à noyau RBF qui sont généralement gaussiennes, elle permet d'effectuer une transformation non linéaires des entrées, la couche de sortie établit une combinaison linéaire des sorties des neurones de la couche cachée. L'architecture générale d'un réseau à fonction de base radiale est donnée sur la figure suivante [MRA09].

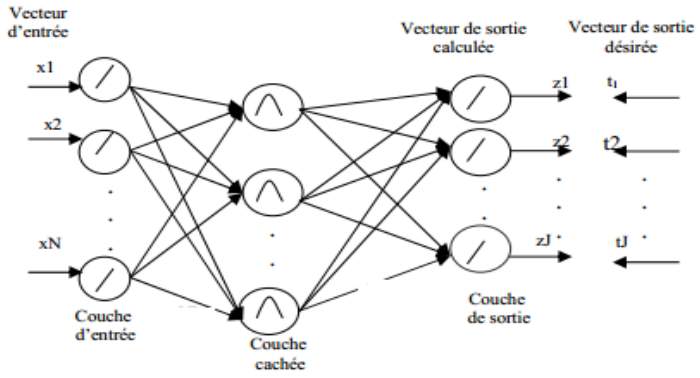


Figure 29: réseau de neurones RBF.

III.3.4.4. Apprentissage dans les réseaux de neurones:

Dans l'apprentissage supervisé, la valeur de sortie (la classe d'appartenance désirée) que le réseau de neurones doit associer au vecteur d'entrée ***x*** est connue. Cette valeur ou vecteur de sortie (appelé cible ou sortie désirée) est fournie par un superviseur (ou expert humain). L'apprentissage consiste à modifier les paramètres du réseau de neurones afin de minimiser l'erreur entre la sortie désirée et la sortie réelle du réseau de neurones. [HAM06]

Lors d'un apprentissage non supervisé, on ne fournit pas au réseau les sorties que l'on désire obtenir. On le laisse évoluer librement jusqu'à ce qu'il se stabilise.

III.4. Les méthodes de classification avec apprentissage non supervisé:

Dans ce type de méthodes, les étiquettes des classes ne sont pas connues *a priori*. Mais son nombre *c* est connu. Afin de définir l'homogénéité d'un groupe d'observations, il est nécessaire de mesurer le degré de *ressemblance* (similarité) ou le degré de *dissemblance* (dissimilarité) entre deux observations. [BOU07]

III.4.1. Les méthodes de classification non supervisées non floues :

III.4.1.1. Modèles de mélange pour la classification :

Un modèle de mélange est un modèle statistique exprimé selon une *densité mélange*. A travers ces modèles, la structure de chaque classe est décrite par une *composante* s'adaptant à la distribution des données où le mélange de ces composantes forme le modèle global qu'il s'appelle *le modèle de mélange*.

III.4.1.1.1. Fonction de densité mélange et le modèle de mélange:

Est une fonction de densité d'une variable aléatoire *x* paramétrée par Θ. Si on appelle $\Theta_1, \Theta_2, \dots, \Theta_c$ une famille de paramètres et $\pi_1, \pi_2, \dots, \pi_c$ et une famille de scalaires tels que : $\sum_{k=1}^{c} \pi_k = 1$. alors la fonction de densité de probabilité φ de *c* composantes s'exprime de la manière suivante:

$$\varphi(x)=\sum_{k=1}^{c} \pi_k f(x, \Theta_k). \tag{3.12}$$

Donc, le modèle de mélange est complètement décrit à partir des paramètres $\{\pi_k, \Theta_k\}_{k=1..c}$.

III.4.1.1.2. Modèle de mélange gaussien :

Soit un ensemble d'individus $\{x_1, x_2, \dots, x_m\}$ représentés sur un espace vectoriel $\Re^d$. Afin de décrire le modèle de mélange gaussien, nous définissons tout d'abord la fonction de densité de la distribution gaussienne à *d*-dimension qui prend la forme suivante :

$$f(x, \Theta) = \frac{1}{(2\pi)^{\frac{1}{d}} (det\, A_k)^{\frac{1}{2}} . e^{\frac{1}{2}(x-\mu_k) A_k^{-1} (x-\mu_k)^T}} \tag{3.13}$$

Où : μ_k est le vecteur moyenne du composant *k* de *d*-dimensions.

A_k: est la matrice de covariance du composant *k* de dimension *d*d* et $detA_k$ désigne le déterminant de la matrice A_k et A_k^{-1} est la matrice inverse de A_k.

Le mélange gaussien a été utilisée dans plusieurs domaines comme la reconnaissance de forme [PAA06], segmentation d'images [ALL07] [NAC09],…etc. Dans le cadre d'utilisation de modèle de mélange gaussien en classification des données, on considère les individus appartiennent chacun à un des *c* classes $(C_1, C_2, \dots, C_c)$, où *c* étant fixé *a priori* et

chaque classe C_k suive une loi normale (ou gaussienne) de *moyenne* μ_k et de *matrice de covariance* A_k i.e. le paramètre Θ_k de chaque composante est caractérisé par sa moyenne μ_k est sa matrice de covariance A_k et on note également $\Theta_k = \{\mu_k, A_k\}$. D'autre part on associé à chaque classe C_k une *probabilité a priori* π_k satisfait : $\pi_k \geq 0 \; et \; \sum_{k=1}^{c} \pi_k = 1$.

La densité de probabilité du modèle globale φ d'un individu x est définie comme suit [ALL07]:

$$\varphi(x|\Theta) = \sum_{k=1}^{c} \pi_k f(x, \Theta_k) = \sum_{k=1}^{c} \pi_k . \frac{1}{(2\pi)^{\frac{1}{d}} (detA)^{\frac{1}{2}}} . \exp\left(\frac{1}{2}(x-\mu_k) A_k^{-1} (x-\mu_k)^T\right) \quad (3.14)$$

III.4.1.1.2.1. Estimation de paramètres de modèle de mélange :

III.4.1.1.2.1.1. Algorithme EM (Expectation-Maximization):

Il existe plusieurs algorithmes pour estimer les paramètres du mélange $(\pi_k, \Theta_k)_{k=1..c}$, le plus utilisé d'entre eux est l'algorithme itératif *Expectation Maximisation (EM)* développé par *Dempster et al* en 1977 [DEM77] que nous allons détailler ci-après :

L'algorithme EM se base sur l'idée de maximisation de la *loi de vraisemblance* suivante :

$$V(X|\Theta) = \prod_{i=1}^{m} \varphi(x_i, \Theta) \quad (3.15)$$

Où, φ est un modèle de mélange. Alors, la loi de vraisemblance s'écrit :

$$V(X|\Theta) = \prod_{i=1}^{m} \sum_{k=1}^{c} \pi_k f(x, \boldsymbol{\Theta}_k) \quad (3.16)$$

Dans [CHI12], maximiser cette expression revient à maximiser le log vraisemblance:

$$\begin{cases} L(X|\Theta) = \sum_{i=1}^{m} log\left(\sum_{k=1}^{c} \pi_k f(x_i, \Theta_k)\right) \\ o\grave{u}: \Theta = (\pi_1, \dots, \pi_c, \Theta_1, \dots, \Theta_c), \; \pi_k \geq 0 \; et \; \sum_{k=1}^{c} \pi_k = 1 \end{cases} \quad (3.17)$$

Le but est de trouver $\widehat{\Theta}$ qui maximise cette vraisemblance où :

$\widehat{\Theta} = \text{argmax}_{\Theta} \; L(X, \boldsymbol{\Theta})$

L'algorithme EM est largement utilisé pour estimer les paramètres du MMG. C'est un algorithme itératif qui maximise la probabilité de vraisemblance générée par chaque gaussienne (classe). Il se déroule en deux étapes, *l'étape E* pour Estimation et l'étape M pour Maximisation. *L'étape E* calcule *la probabilité a posteriori* pour chaque donnée x_i qui dépend des paramètres courants $\Theta^{(t)}$ à partir du *théorème inverse de Bayes*. *L'étape M* qui permet de ré-estimer les paramètres $\Theta^{(t+1)}$ à partir des paramètres $\Theta^{(t)}$ pour maximiser la vraisemblance. L'algorithme (EM) est illustré comme suivant :

Algorithme Expectation Maximization (EM) :

Entrées : $\Theta^{(0)} = \left\{\pi_1^{(0)}, \dots, \pi_c^{(0)}, \Theta_1^{(0)}, \dots, \Theta_c^{(0)}\right\}$; *critère d'arrêt* $\boldsymbol{\varepsilon}$; *nombre de classe* $\boldsymbol{c}$.

Initialement de compteurs des itérations : $\mathbf{t} \leftarrow \mathbf{0}$.

Répéter :

L'étape E : calcule des probabilités a posteriori:

$$P\left(x_i|\Theta^{(t)}\right) = \frac{\pi_k f\left(x_i|\Theta_k^{(t)}\right)}{\sum_{l=1}^{c} \pi_l f\left(x_i|\Theta_l^{(t)}\right)} \quad k = 1..c, i = 1..m$$

L'étape M : ré-estimation des paramètres $\Theta^{(t+1)}$ à partir des paramètres $\Theta^{(t)}$.

$$\pi_{(k)}^{(t+1)} = \frac{1}{m}\sum_{i=1}^{m} P\left(x_i|\Theta^{(t)}\right)$$

Jusqu'à : $\left|L\left(X, \Theta^{(t+1)}\right) - L\left(X, \Theta^{(t)}\right)\right| \leq \varepsilon$

III.4.1.1.2.1.2. Algorithme EM pour le modèle de mélange gaussien:

Pour le modèle de mélange gaussien, les estimateurs du modèle sont calculés dans l'étape de Maximisation (M) de l'algorithme EM.

Ces estimateurs $\left(\mu_{(k)}^{(t+1)}, A_{(k)}^{(t+1)}\right), k = 1..c$, s'écrivent comme suit :

$$\mu_{(k)}^{(t+1)} = \frac{\sum_{i=1}^{m} \boldsymbol{x_i} P\left(x_i|\Theta^{(t)}\right)}{\sum_{i=1}^{m} P\left(x_i|\Theta^{(t)}\right)} \qquad (3.18)$$

$$A_{(k)}^{(t+1)} = \frac{\sum_{i=1}^{m} P\left(x_i | \Theta^{(t)}\right)\left[\left(x_i - \mu_{(k)}^{(t+1)}\right)\left(x_i - \mu_{(k)}^{(t+1)}\right)^t\right]}{\sum_{i=1}^{m} P\left(x_i | \Theta^{(t)}\right)} \qquad (3.19)$$

III.4.1.1.2.1.3. Algorithme CEM pour le modèle de mélange gaussien :

Plusieurs variantes de l'algorithme EM ont été proposées [BOU13][JOL03], on s'intéresse dans notre cas sur la méthode CEM (Classificatoire Expectation Maximisation). Cette méthode nous permet d'attribuer pour chaque itération chaque donnée x_i à sa classe. Cet algorithme a même structure que l'algorithme EM avec l'ajout d'une étape supplémentaire (étape C) entre les étapes E et M, dans ce cas chaque objet est affecté à une et une seule classe selon la règle suivante :

$$x_i \in C_k \quad si: \; P\left(x_i | \Theta_k\right) = \underset{\Theta}{\operatorname{argmax}}\left(P\left(x_i | \Theta_l\right)\right)$$

On remarque que la seule intervention de l'expert dans la classification par modèle de mélange gaussienne se situe à la fin du processus pour identifier les classes calculées avec les classes biologiques.

III.4.1.2. H-C-Means (Hard C-Means):

Est une des méthodes les plus connues pour la classification non supervisée crée par *Hatigant* et *Wong* en 1979 [HAR79], elle nécessite uniquement de fixer le nombre de classes, noté *c*. l'algorithme Hard C-means (ou k-means) consiste à choisir aléatoirement des centres de *c* classes. La classification des éléments s'effectue de manière itérative en alternant l'étape de classification et l'étape de mise à jour des centres jusqu'à stabilisation de la fonction objective. Avec un élément est attribué à une et une seule classe.

L'algorithme HCM basé sur la minimisation de la fonction objective suivante :

$$J(B, U, X) = \sum_{i=1}^{c} \sum_{j=1}^{m} \mu_{ij} . d^2\left(x_j, b_i\right) \qquad (3.20)$$

Où :

$B = (b_1, b_2, \ldots, b_c)$: $l'ensembledes\ centres\ de\ classes.$

$d^2(x_j, b_i)$: *la distance euclidienne entre la donnée* x_j *et le centre* b_i.

$U = [\mu_{ij}]$: *la matrice des degrés d'appartenancedu modèle de classification.*

Les solutions du problème s'écrivent :

$$\mu_{ij} = \begin{cases} 1 & ssi \quad d^2(x_j, b_i) < d^2(x_j, b_k) \quad \forall k \neq i \\ 0 & sinon \end{cases} \tag{3.21}$$

$$b_i = \frac{\sum_{j=1}^{m} \mu_{ij} . x_j}{\sum_{j=1}^{m} \mu_{ij}} \tag{3.22}$$

L'algorithme C-Moyennes est représenté comme suit :

Algorithme Hard C-Means (HCM)

Etape 01 :

Initialisation: fixer le nombre de classes c.

Initialiser la partition dure U_0 et fixer le seuil d'arrêt ε.

Etape 02 :

Répéter :

1) Calculer les centres de classes par l'équation 3.22.
2) Calculer les nouveaux degrés d'appartenance μ_{ij} et mettre à jour la partition U par l'équation 3.21.

Jusqu'à : $\left|J(B,U,X)^{(t+1)} - J(B,U,X)^{(t)}\right| \leq \varepsilon$

Dans cet algorithme les éléments sont classés de façon certaine, c'est-à-dire un élément est appartenant à une et une seule classe (*notion de partition dure*), cette assertion ne reflète pas la réalité physique de l'échantillon étudié comme l'existence d'un bruit et les données sont incomplètes,….

Dans le suivant, nous avons traités certaines qui nous permettent d'obtenir une *classification floue* qui prend en compte ces aspects imprécis et incertains.

III.4.2. Les méthodes de classification non supervisées floues :

Comme on le voit dans le premier chapitre, la théorie des ensembles flous est une théorie mathématique développée par Lotfi ZADEH en 1965 [AMB09], afin de représenter mathématiquement certaines informations imprécises, incomplètes et/ou incertaines par une fonction d'appartenance à valeur dans l'intervalle [0,1]. Contrairement au cas classique où la fonction d'appartenance prend deux valeurs : 0 ou 1.

Au contraire des méthodes de classification classique, Un algorithme de classification floue permet de partitionner l'espace des données en plusieurs classes, où chaque point de l'espace de caractéristiques peut appartenir à plusieurs classes avec différents degrés d'appartenance entre 0 et 1.

Ces algorithmes permettent de partager l'ensemble de *n* objets en *c* groupes, la similarité à l'intérieur d'un même groupe est élevée, mais faible entre les différentes classes. Pour ce faire, ces algorithmes sont exécutés en deux étapes, d'abord ils calculent les centres des groupes et ensuite ils assignent chaque objet au centre le plus proche.

La figure suivante illustre la notion des ensembles flous dans le domaine de classification [MIN06].

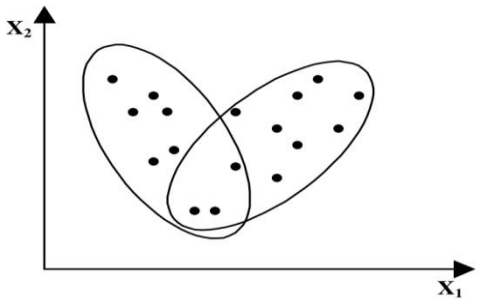


Figure 30: illustration de la classification floue.

Dans cette section nous allons présenter quelques méthodes de classification automatique qui utilisent la notion d'ensemble flou les plus connues dans la littérature. Parmi ces algorithmes, nous expliquons les algorithmes FCM (Fuzzy C-Means) et ses variantes, PCM (Possibiliste C-Means), FK-NN (Fuzzy K-Nearest-Neighbour).

III.4.2.1. Algorithme C-Moyennes floue (FCM)

Fuzzy c-means (FCM), l'une des méthodes les plus connues de classification floue dérivée de l'algorithme K-means, elle est développée par Dunn en 1974 et améliorée au plus

tard par Bezdek en 1981 [CHA11]. Cet algorithme nécessite de connaitre le nombre de classes au préalable et génère les classes par un processus itératif à bas sur la minimisation de la somme aux carrés des distances euclidiennes entre les objets et les centres des classes,

$$J(B,U,X) = \sum_{i=1}^{c}\sum_{j=1}^{N}\mu_{ij}^{\boldsymbol{m}} d_A^2(x_j, b_i). \quad (3.23)$$

Où :

$X = [x_1, x_2, \dots, x_j, \dots, x_n]$: L'ensemble des données, chaque donnée représentée par $\boldsymbol{N}$ **dimensions**.

$B = [b_1, b_2, \dots, b_i, \dots, b_c]$: Est le vecteur des centres de classes.

$U = [u_{ij}]$: Est la matrice de partition floue (de dimension *c*x*n*).

$d_A^2(x_j, b_i) = (x_j - b_i)^T . A . (x_j - b_i)$: Définit la mesure de distance entre l'observation x_j et le centre b_i au sens de la matrice définie positive A (Dimension N X N). (Une matrice définie positive est une matrice dans laquelle toutes les valeurs propres de cette matrice sont positives).

u_{ij} : représente le degré d'appartenance des données x_j à la classe *i.* pour avoir une bonne partition, on impose les éléments de *U* les contraintes suivantes :

$$\begin{aligned} &u_{ij} \in [0,1] \quad \forall i \in \{1..c\}, \ \forall j \in \{1..n\} \\ &0 < \sum_{j=1}^{n} u_{ij} < n \quad \forall i \in \{1..c\} \\ &\sum_{i=1}^{c} u_{ij} = 1 \quad \forall j \in \{1..n\} \end{aligned} \quad (3.24)$$

$\boldsymbol{m} \in [\mathbf{1}, +\infty[$**:** est un facteur qui désigne le degré de flou de la partition. Généralement m=2. [KHO98], [CHE09], [KEL85], [HAR10]. La minimisation de l'équation précédente nous donne :

- Matrice *U* des degrés d'appartenances sont représentés comme suit :

$$u_{ij} = \left(\sum_{k=1}^{c} \left(\frac{d^2(x_j, b_i)}{d^2(x_j, b_k)} \right)^{\frac{2}{(m-1)}} \right)^{-1} \quad (3.25)$$

- Les centres des classes sont calculés comme suit :

$$b_i = \frac{\sum_{j=1}^{n} u_{ij}^m x_j}{\sum_{j=1}^{n} u_{ij}^m} \quad (3.26)$$

L'algorithme FCM est illustré comme suit :

Algorithme FCM

Etape01. Initialisation: c, m, $\varepsilon > 0$: ε plus petite valeur (critère d'arrêt). Initialiser le vecteur de centres des classes b_i et la matrice de partition floue U. Définir la matrice A.

Etape02 :

Répéter :

1) Mettre à jour les centres b_i des classes par l'équation (3.26)

2) Mettre à jours la matrice de partition U par l'equation (3.25)

Jusqu'à : $|U^t - U^{t-1}| < \varepsilon$,

III.4.2.2. Algorithme FCM de Gustafson et Kessel (GK):

Dans, [KHO98], l'auteur a rassemblé et expliqué plusieurs algorithmes qui son considérés comme variantes de l'algorithme FCM. Dans cette section, nous allons expliquer L'extension de l'algorithme FCM la plus connu [OUA09] proposée par Gustafson et Kessel en 1979, dont le but de détecter les classes de forme différent plutôt de définir un modèle unique pour toutes les classes (hypersphérique ou hyperelliptique, comme dans l'algorithme FCM classique et l'algorithme PCM). Pour cela, *chaque classe C_i possède sa propre matrice de norme A_i*. Donc, la mesure de distance entre les centres des classes et l'objet est représentée comme suit :

$$d^2_{A_i}(x_j, b_i) = (x_j - b_i)^T . A_i . (x_j - b_i) \tag{3.27}$$

$$A_i = [\delta_i det(F_i)]^{\frac{1}{n}} F_i^{-1} \tag{3.28}$$

Où : F_i: Est la matrice de covariance floue de la i^{ème} classe donnée par :

$$F_i = \frac{\sum_{j=1}^{n} (u_{ij})^m (x_j - b_i)(x_j - b_i)^T}{\sum_{j=1}^{n} (u_{ij})^m} \tag{3.29}$$

δ_i: Exprime une contrainte sur le volume de la classe C_i permettant d'optimiser la forme de celle-ci. Donc, l'algorithme de Gustafson et Kessel est représenté comme suit :

Algorithme de Gustafson et Kessel

Etape01 : Initialisation : Choisir un nombre c de classes, l'exposant m, le critère d'arrêt ε , Initialiser les centres des classes b_i et la matrice de partition floue U. Fixer une valeur δ_i pour chaque classe C_i

Etape02 :

Répéter :

1) Calculer les centres des classes par l'équation (3.26)

2) Calculer les matrices A_i des classes par l'équation (3.28).

3) Mettre à jour la matrice de partition U par l'équation (3.25).

Jusqu'à : $|U^t - U^{t-1}| < \varepsilon$.

III.4.2.3. Algorithme C-Moyennes possibilistes (PCM) :

Pssibilistic C-Means est un algorithme développé par Krishnapuram et Keller en 1993 à partir des idées de l'algorithme FCM, selon laquelle la contrainte de la somme des degrés d'appartenance d'un élément j dans la matrice de partition floue U n'égale pas forcément 1. C'est-à-dire, lever la contrainte probabiliste et réaliser un algorithme possibiliste. Pour cela, cette approche a l'avantage de considérer l'appartenance d'un élément à une classe indépendamment de son appartenance aux autres classes. L'algorithme PCM basé sur la minimisation de la fonction suivante :

$$J(B,U,X) = \sum_{i=1}^{c}\sum_{j=1}^{N}\mu_{ij}^{\boldsymbol{m}}d^2(x_j,b_i) + \sum_{i=1}^{c}\eta_i\sum_{j=1}^{n}(1-u_{ij})^m \qquad (3.30)$$

Dans l'algorithme PCM, on impose les contraintes suivantes sur la matrice de partition floue U :

$$u_{ij} \in [0,1] \quad \forall\, i \in \{1..c\},\ \ \forall j \in \{1..n\}$$

$$0 < \sum_{j=1}^{n} u_{ij} < n \quad \forall i \in \{1..c\} \qquad (3.31)$$

$$\max_{i} u_{ij} > 0, \quad \forall j \in \{1..n\}$$

η_i: est l'indice qui pondère chaque classe C_i, c-à-dire, un réel positif déterminant la distance à laquelle le degré d'appartenance d'un vecteur j à la classe i est égale à 0.5[BAR00]. Cet indice est calculé comme suit :

$$\eta_i = \frac{\sum_{j=1}^{n}(u_{ij})^{\boldsymbol{m}}d^2(x_j,b_i)}{\sum_{j=1}^{n}(u_{ij})^{\boldsymbol{m}}} \qquad (3.32)$$

La minimisation de la fonction objective précédente nous donne l'expression des centres des classes b_i comme dans l'algorithme FCM. Par contre la relation de mise à jour de la partition floue U prend la forme suivante :

$$u_{ij} = \left(1 + \left(\frac{d^2(x_j, b_i)}{\eta_i}\right)^{\frac{1}{(m-1)}}\right)^{-1} \quad (3.33)$$

III.4.2.4. Algorithme des k-Plus Proche Voisin Floue (K-PPV):

Keller et al, [KEL85] ont proposé une extension d'un algorithme *k*-PPV classique avec l'introduction la notion de flou. Dans cette approche nécessite d'associer un degré d'appartenance u d'un élément x à classer. Où ce degré vérifié les conditions suivants :

$$u_{ij} \in [0,1] \quad \forall i \in \{1..c\}, \ \forall j \in \{1..n\}$$

$$\sum_{i=1}^{c} u_{ij} = 1 \quad \forall j \in \{1..n\}$$

Les auteurs ont proposé de calculer les degrés d'appartenance de chaque élément x et les k proches voisins comme suit:

$$u_i(x) = \frac{\sum_{j=1}^{\boldsymbol{k}} u_{ij} \left(\frac{1}{d(x, x_j)^{\frac{2}{(m-1)}}}\right)}{\sum_{j=1}^{\boldsymbol{k}} \left(\frac{1}{d(x, x_j)^{\frac{2}{(m-1)}}}\right)}$$

Où :

u_{ij} : est le degré d'appartenance du $j^{ième}$ plus proche voisin de x à la classe C_i.

$d(x, x_j)$: Représente la distance entre l'élément x à classifier et le $j^{\text{ième}}$ plus proche voisin.

Donc, l'élément x est associé à la classe C_i selon l'équation suivante : $argmax_{i=1}^{c} u_i(x)$.

Dans la table suivante, nous décrivons quelques méthodes de classification dynamique de données évolutives développées. Bien sûr, ces méthodes sont basées sur les algorithmes de classification supervisés et non-supervisés mentionnés dans ce chapitre.

Abréviation	Nom	Auteur	Année	Méthode de classification utilisée	Référence	Domaine d'application.	Types de classes évolutives étudiés
FMMC	Fuzzy Min Max Clustering	Simpson	1993	Réseau de neurones (MLP). Avec une règle de classification floue.	[SIM93]	IRIS	Ajout de nouvelles éléments (expansion de classes).
ESOM	Evolving Self Organizing Map	Deng	2003	Réseau de neurones, cartes organisatrices de Kohonen. Avec mesure de similarité (distance euclidienne)	[DEN03]	Pima indians diabets diagnosis.	Fusion, scission de classes et élimination des classes inutiles
AUDyC 1ière version	AUto –adaptive and Dynamical Clustering.1	Cristophe LURETTE	2003	Réseau de neurones La mesure de similarité (distance euclidienne, fonction noyau gaussienne)	[LUR03]	Supervision d'une presse hydraulique	Fusion, suppression,
SAKM	Self Adaptive Karel Machine	Amadou Boubacar HABIBOULAY	2006	Réseau de neurones SVM, méthodes à noyau.	[AMA06]	Suivi d'un processus thermique (thermorégulat--eur)	Fusion, élimination.
AUDyC 2ième version	AUto –adaptive and Dynamical Clustering.2	Amadou boubacar HABIBOULAY	2006	Réseau de neurones, modèle de mélange gaussien (MMG).	[AMA06]	Suivi d'un processus thermique (thermorégulat--eur)	Fusion, élimination, Scission
KPPVFDS	K-Plus Proche Voisins Flou Dynamique Supervisé	Laurent HARTERT	2010	K-plus proche voisins floue	[HAR10]	Système de commutation entre plusieurs modes.	fusion, scission et suppression des classes obsolètes.
KPPVFD Semi Supervisé	K-Plus Proche Voisins Flou Dynamique Semi Supervisé	Laurent HARTERT	2010	K-plus proche voisins floue	[HAR10]	-Machine de soudures. -Générateur de vapeurs des Réactions à Neutrons Rapides (RNR).	fusion, scission et **apparition de nouvelles classes** et suppression des classes obsolètes.
FPMD	Fuzzy Pattern Matching Dynamic	Laurent HARTERT	2010	Approche possibiliste.	[HAR10]	Système hydraulique à deux réservoirs.	fusion, scission et suppression des classes obsolètes.

Tableau 3: Table récapitulatif des méthodes de classification dynamique de données évolutives

III.5. Conclusion :

Nous décrivons dans ce chapitre quelques méthodes de classification supervisées et non supervisées (floues et non floues) les plus connues et les plus utilisées. A base de ces méthodes, nous essayons de développer une méthode de classification satisfaisante pour les données évolutives (non stationnaires) et multimodales, A partir de combiner et de coopérer deux ou plusieurs classifieurs avec l'ajout de certaines règles qui permet de suivi les classes évolutives (notion de partition dynamique).

Chapitre 04 : Fusion de données : Théories et techniques

IV.1. Introduction

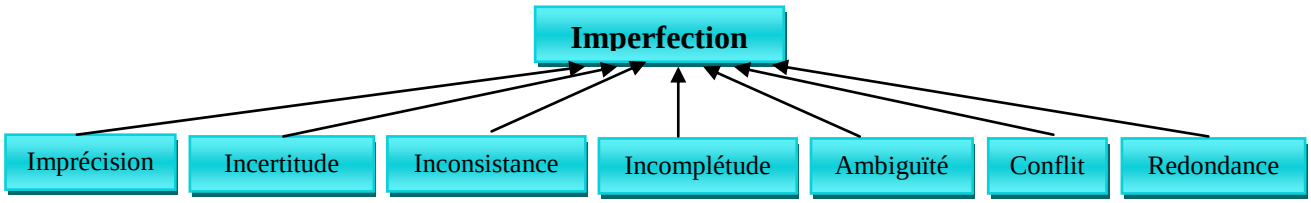
Dans de nombreuses applications nous sommes confrontés à un grand nombre de données imparfaites ce qui pose souvent problème pour les classifier. La fusion d'informations est apparue initialement afin de gérer des quantités très importantes de données multi-sources dans le domaine militaires. Depuis quelques années des méthodes de fusion ont été adaptées et développées pour les applications en traitement de signal et plus particulièrement pour la classification, dans le but de faire face à la problématique d'usage des données issues de divers sources ou capteurs externes avec des formes et types d'informations (attributs externes) différents.

Les bases théoriques de la fusion ont été établies par Zadeh, Shafer et Dempster [ZAD65][DEM67][SHA76], puis les modèles mathématiques de la fusion ont connus plus d'amélioration dans les années 1980, où ils étaient plus utilisée dans le domaine médical, le domaine militaire, la technologie spatiale et robotique, …

Dans ce chapitre, nous proposons un panorama des principales méthodes de fusion de données. Il s'agit des classiques méthodes probabilistes (en particulier de l'inférence Bayésienne), mais aussi des méthodes non probabilistes comme la théorie des ensembles flous, la théories des possibilités et la théorie des fonctions de croyance apparues plus récemment, mais qui connaissent un essor de plus en plus important. Pour chaque théorie, nous présentons les deux composantes essentielles de système de fusion : la représentation des connaissances et le raisonnement. Pour la première, nous montrons quelles sont les caractéristiques des données imparfaites que chaque théorie permet de modéliser. Pour la seconde, nous décrivons les modes de combinaison et les règles de décision. Les méthodes présentées ici sont particulièrement adaptées à la classification, mais elles peuvent être employées dans un cadre plus général d'aide à la décision.

IV.2. Caractéristiques générales des données à fusionner :

Dans le domaine des systèmes d'aides à la décision, les informations manipulées sont souvent imparfaites. Une des caractéristiques importantes de l'information en fusion est son imperfection. Celle-ci est toujours présente sinon la fusion ne serait pas nécessaire [LAM13]. Ces imperfections peuvent prendre diverses formes : *Imprécision, Incertitude, Incomplétude, Inconsistance et ambiguïté.* L'ensemble des informations imparfaites sont représentées dans le schéma suivant [CAV12].



- ***Imprécision :*** concerne le contenu de l'information et indique donc son défaut quantitatif de connaissance. Une information est dite imprécise si elle insuffisante pour permettre de comprendre une situation donnée. Une information imprécise prend la forme de disjonction de valeurs mutuellement exclusives. par exemple, l'âge de Abdelhamid entre 27 et 32 ans, ce qui se traduit par l'âge de Abdelhamid (noté X) appartient au sous-ensemble {27,28,29,30,31,32}, X=27 ou X=28 ou X=29 ou X=30 ou X=31 ou X=32.
- ***Incertitude :*** Relative à la vérité d'une information et caractérise son degré de conformité à la réalité. c'est-à-dire, une information est dite incertaine si l'on ne sait pas si elle vraie ou fausse. Une information certaine représente une connaissance complète de la réalité. alors qu'une information incertaine représente une connaissance partielle de la réalité par exemple : « il va pleuvoir demain ».

*L'imprécision concerne le contenu de l'information et porte sur un défaut quantitatif de connaissance. Tandis que l'incertitude est relative à la vérité d'une information, caractérisant sa conformité à la réalité. Une proposition peut être imprécise « cet homme est grand », incertaine « cette lettre arrivera demain » ou à la fois imprécise et incertaine « il pleuvra **beaucoup** demain».*

Dans notre problème de classification dynamique de données évolutive, et pour illustrer les concepts d'imprécision et d'incertitude dans le domaine médical, on prend

l'exemple suivant : « *La transition douce entre tissus sains et pathologique est une imprécision due au phénomène physiologique. En plus une information est dite incertaine si un agent ne sait pas si cette information est vraie ou fausse* ».

- ***Incomplétude :*** caractérise l'absence de l'information apportée par la source sur certains aspects du problème. Elle doit mesurer par la différence de la quantité d'information réellement fournie par la source et la quantité d'information que la source doit fournir. Par exemple : un radar ne permet pas de fournir une image des sous-marins immergés, l'information ne porte que sur la surface de l'eau.
- ***Inconsistance :*** l'inconsistance apparait quand plusieurs informations concernant la même situation sont en conflit c'est-à-dire, ces informations conduisant à des interprétations contradictoires et donc incompatibles. Par exemple : « Abdelhamid est célibataire » et « la femme de Abdelhamid s'appelle Maria » le conflit dans les données peut mener à une incohérence dans la conclusion. [ASS10].
- ***Ambiguïté :*** Exprime la capacité d'une information à conduire à deux interprétations. L'ambiguïté peut provenir des imperfections précédentes, mais pas nécessairement par exemple, un système de navigation qui n'arrive pas à identifier si un ensemble de pixels issue d'une image satellite représente un fleuve ou une route. Cette ambiguïté peut résulter d'une imprécision issue de la faible résolution des pixels de l'image à cause de bruit.
- ***Le conflit :*** caractérise plusieurs informations conduisant à des interprétations contradictoires et donc incompatibles. Les situations conflictuelles sont fréquentes dans les problèmes de fusion, et posent toujours des problèmes difficiles à résoudre.
- ***Redondance :*** est l'opposé du conflit, est la qualité de sources qui apportent plusieurs fois la même information. La redondance entre les sources est souvent observée, dans la mesure où les sources donnent des informations sur le même phénomène.

La fusion d'informations peut apporter des solutions qui permettent d'enlever certaine imperfections.

IV.3. Modélisation de l'information Imparfaite :

Dans ce qui suit, nous étudions les principales théories permettant de représenter les connaissances imparfaites. Ces théories sont : les probabilités, la théorie des possibilités, les ensembles flous et la théorie des fonctions de croyance. Ces théories permettent d'introduire le concept de fusion de données. Tous ces outils reposent sur une base mathématique ensembliste commune. [BEL08]

Notations :

Ω : représente le domaine de l'information considéré. Cela peut être l'ensemble des valeurs possibles d'un variable ou l'ensemble des réalisations possibles d'un évènement. Cet ensemble est appelé *« univers des possibles »* ou *« référentiel »*.

$\wp(\Omega)$: L'ensemble des parties de Ω ou encore 2^{Ω}.

A, B : sous ensembles de Ω, alors :

$A \cup B$: Représente la réalisation des évènements disjoints $A\ OU\ B$.

$A \cap B$: Représente la réalisation des évènements conjoints $A\ ET\ B$.

$\Omega / A = \bar{A}$: Désigne le contraire de A.

Ω : Désigne l'événement certain.

$\emptyset$: Désigne l'événement impossible.

IV.3.1. Théorie des probabilités:

IV.3.1.1. Notions fondamentales:

La théorie des probabilités est l'outil plus traditionnel qui permet de modéliser les phénomènes aléatoires. Un phénomène est dite aléatoire si l'on ne peut pas prédire avec *certitude* les résultats. La théorie probabiliste associée à la théorie Bayésienne pour la décision dans l'incertain où l'information y modélisée par une probabilité conditionnelle.

La probabilité P de l'occurrence d'un évènement $A \subseteq \Omega$ comme une fonction $P : \wp(\Omega) \rightarrow [0,1]$ qui associé à A un degré de probabilité compris dans l'intervalle [0,1] et qui satisfait les axiomes suivants :

$$(P(\phi)) \leq P(A) \leq (P(\Omega) = 1) \qquad (4.1)$$

$$si : A \cap B = \phi, P(A \cup B) = P(A) + P(B) \tag{4.2}$$

L'axiome (4.1) établit que Ω est un événement certain, contrairement à l'ensemble vide ϕ qui est évènement impossible. Le nombre non négatif *P(A)* quantifie donc dans quelle mesure l'événement *A* est probable.

L'axiome (4.2) signifie que si deux événements sont mutuellement exclusifs, alors la probabilité de l'occurrence d'au moins un d'entre les événements est la somme des leur probabilités individuelles.

On peut déduire de ces axiomes de base que la connaissance de la probabilité d'un évènement *A* détermine complètement celle de l'évènement contraire $\overline{A}$:

$$P(A) = 1 - P(\overline{A}) \tag{4.3}$$

Dans le cas où Ω est fini et discret, une mesure de probabilité *P* peut être définie à partir d'une distribution de probabilité $p : \Omega \to [0,1]$ sur les singletons de Ω :

$$P(A) = \sum_{x \in A} p(x) \tag{4.4}$$

A partir de l'axiome (4.4), on peut déduire l'axiome de normalisation dans la théorie des probabilités :

$$P(\Omega) = \sum_{x \in \Omega} p(x) = 1 \tag{4.5}$$

Dans le cas où Ω est continu, la fonction *p* devient une densité de probabilité *p :* $\Omega \to [0,1]$, telle que : $\int_{\Omega} p(x)dx = 1$ et $P(A) = \int_{A} p(x)dx$. (4.6)

IV.3.1.2. Cadre Bayésien :

Dans le cadre de probabilités conditionnelles, la théorie de Bayes permet de combiner plusieurs distribution de probabilités. Cette théorie est illustrée dans le deuxième chapitre concernant la classification bayésienne.

IV.3.2. Théorie des sous ensembles flous et théorie des possibilités:

La théorie des possibilités introduite en 1978 par le père de la logique floue *Zadeh Lotfi* [Zadeh78], puis développée en France par *Dubois et Prade* en 1988 [DUB88]. La théorie des possibilités étroitement liée à la théorie des sous-ensembles flous [ZAD65][AMB09]. Elle est présentée pour traiter les données à caractère *imprécis* « cet arbre est très haut » et/ou *incertain* « il va pleuvoir demain ». [ZOU08]

IV.3.2.1. Théorie des ensembles flous :

IV.3.2.1.1. Notion de sous-ensemble-flou :

Parmi les techniques non probabilistes est la théorie des ensembles flous ou la théorie de la logique floue [CAV12] fournit un très bon outil pour représenter explicitement des informations imprécises sous la forme de fonctions d'appartenance. La théorie des ensembles flous a été introduite par Zadeh en 1964 [ZAD64]. Comme on le voit dans le premier chapitre, l'appartenance d'un élément x à un ensemble A est binaire dans la théorie des ensembles classique, il peut lui appartenir ou non. La théorie floue permet de définir l'affectation d'un élément x de l'ensemble Ω à un ou plusieurs sous ensemble flous disjoint, avec des degrés d'appartenance différents entre 0 et 1 relatifs à chaque sous ensemble.

Pour illustrer la notion des sous ensembles flous, un exemple couramment utilisé [BOU10] est celui de la taille de l'individu (voir figure ci-dessous) pour laquelle on définit trois classes et leurs intervalles « petit=[40, 160[», « moyen=[160,180[» et « grand=[180,250] ». selon la logique classique, une personne x de 159.9 cm est qualifié de petite taille, contrairement à la logique floue, x sera considéré de petite taille avec un degré d'appartenance égal à 0.4 et de taille moyenne avec un degré d'appartenance égal à 0.8.

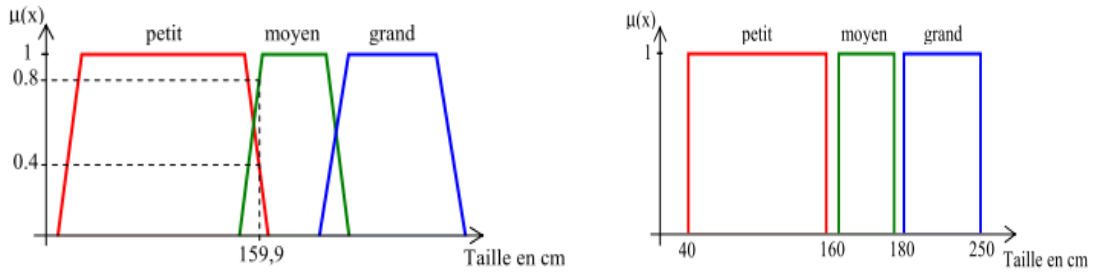


Figure 31:Valeurs de vérité pour les théories des ensembles flous et non-flous.

Formellement, la fonction $\mu_A(x)$ définie le degré d'appartenance de x à A dans l'intervalle [0,1] comme suit:

$$\mu_A(x): \Omega \rightarrow [0,1]$$

Où :

- $\mu_A(x)$: est le degré d'appartenance de *x* à *A*.
- $\boldsymbol{supp(A)}$: support de *A :* ensemble des éléments *x* ayant un degré d'appartenance non nul. Formellement :

$$supp(A) = \{x \in \Omega \ / f_A(A) \neq 0\}. \tag{4.7}$$

- $\boldsymbol{h(A)}$: hauteur de *A* : est le plus fort degré avec lequel de *x* appartient à A. Formellement :

$$h(A) = sup_{x \in A} f_A(x). \tag{4.8}$$

Si *h(A)=1,* alors *A* est dit normalisé.

- $\boldsymbol{noy(A)}$(noyau de *A*): est l'ensemble de tous les éléments qui lui appartiennent totalement (avec un degré 1). Formellement :

$$noy(A) = \{x \in \Omega / \mu_A(x) = 1\}. \tag{4.9}$$

- $\boldsymbol{\alpha - coupe}$ **:** pour toute valeur α de l'intervalle [0,1], on appel $\alpha - coupe$ d'un sous-ensemble flou *A* de 3, le sous –ensemble noté A_α des éléments de Ω pour lesquels la fonction d'appartenance est supérieure ou égale à α:

$$A_\alpha = \{x \in \Omega / \mu_\alpha(x) \geq \alpha\}. \tag{4.10}$$

Si nous choisissons $\alpha = 0$, alors A_α est l'univers de discours Ω. Si nous choisissons $\alpha = 1$, alors A_α est le noyau de *A, noy(A).* Sur la figure ci-dessous, nous illustrons les définitions précédentes par un exemple. [PAG11]

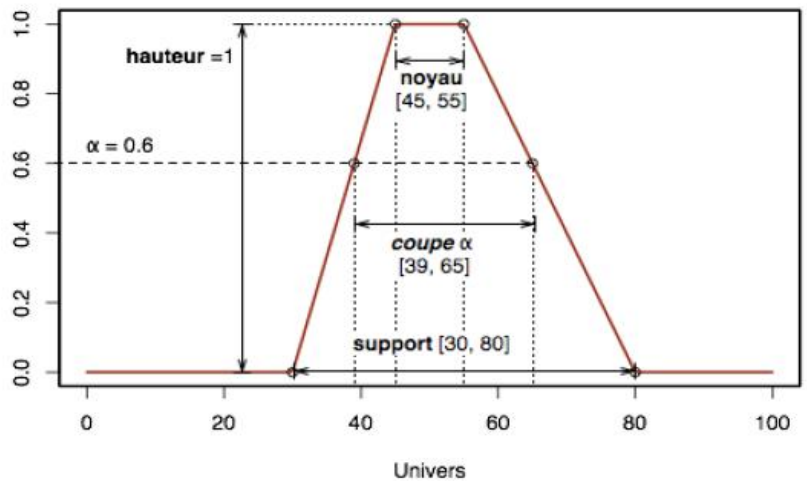


Figure 32:Attributs des ensembles flous.

IV.3.2.1.2- Propriétés des sous ensembles flous:

Nous avons dit précédemment que la théorie des sous-ensembles flous englobait la théorie des ensembles classiques. Nous allons donc définir les opérateurs de l'égalité, d'inclusion d'intersection, d'union et de complémentarité afin de conserver cette propriété.

Soient A et B deux sous-ensembles flous d'un ensemble Ω :

A et B sont dits *égaux*, propriété que l'on note $A=B$, si leurs fonctions d'appartenance prennent la même valeur en tout point de Ω. Formellement, $A = B$ est donné par :

$$\forall x \in \Omega, \quad \mu_A(x) = \mu_B(x). \tag{4.11}$$

A est dit *inclus* dans B, propriété que l'on note $A \subseteq B$, si tout élément x de Ω qui appartient à A appartient à B avec un degré moins aussi grand. Formellement, $A \subseteq B$ est donné par :

$$\forall x \in \Omega, \quad \mu_A(x) \leq \mu_B(x). \tag{4.12}$$

L'*intersection* (l'opérateur T-norme) de deux sous-ensembles flous est le sous-ensemble flou constitué des éléments de Ω affectés du *plus petit des degrés d'appartenance* avec lesquels ils appartiennent à A et B. formellement $A \cap B$ est donné par :

$$\forall x \in \Omega, \quad \mu_{A \cap B}(x) = min\left(\mu_A(x), \mu_B(x)\right) \tag{4.13}$$

L'*union* (l'opérateur T-conorme) de deux sous-ensembles flous est le sous-ensemble flou constitué des éléments de Ω affectés du *plus grand des degrés d'appartenance* avec lesquels ils appartiennent à *A* et *B*. formellement $A \cup B$ est donné par :

$$\forall x \in \Omega, \quad \mu_{A\cup B}(x) = max\big(\mu_A(x), \mu_B(x)\big) \tag{4.14}$$

Le *complément* du sous-ensemble flou *A*, que l'on note A^C *ou* $\overline{A}$, est le sous-ensemble flou de Ω constitué des éléments *x* lui appartenant d'autant plus qu'ils appartiennent peu à *A*. formellement, le complément de *A* est défini comme suit :

$$\forall x \in \Omega, \quad \mu_{A^c}(x) = 1 - \mu_A(x) \tag{4.15}$$

La théorie des possibilités est un moyen de dire dans quelle mesure la réalisation d'un événement est possible et dans quelle mesure on en est certain. Elle s'en éloigne par l'utilisation de deux évaluations duales, dites de *possibilités* et de *nécessité* [DUB06].

IV.3.2.2- Théorie des possibilités:

IV.3.2.2.1- Mesure de possibilité:

On peut définir la mesure de possibilité Π d'un événement *A* comme une fonction $\Pi : \wp(\Omega) \rightarrow [0,1]$ qui associe à *A* un coefficient (degré de possibilité) compris entre 0 et 1. De plus cette fonction doit vérifier les axiomes suivants :

$$\forall A \subseteq \Omega, \qquad (\Pi(\phi) = 0) \leq \Pi(A) \leq (\Pi(\Omega) = 1) \ ; \tag{4.16}$$

$$\forall A_i \in \wp(\Omega), \quad \Pi\big(\bigcup_i A_i\big) = \sup_i \Pi(A_i) \tag{4.17}$$

Dans le cas de deux sous ensemble, la condition (4.17) s'écrit :

$$\Pi(A_1 \cup A_2) = \max\big(\Pi(A_1), \Pi(A_2)\big). \tag{4.18}$$

Dans l'axiome (4.16) quantifie dans quelle mesure l'événement *A* est possible. Pour $\Pi(A) = 0$ c'est-à-dire que *A* est un événement impossible, pour $\Pi(A) = 1$, l'événement *A* est tout à fait possible. La mesure de possibilité est dite *normale* si $\Pi(\Omega) = 1$. [BOU07]

L'axiome (4.18) traduit le fait qu'une mesure de possibilité n'est pas additive mais a un caractère d'ordre, de plus cette mesure n'exige pas que les deux événements A_1 et A_2 soient disjoints. A partir des deux axiomes précédents, on peut déduire l'axiome des événements contraires suivant:

$$\forall A \subset \Omega, \qquad Max\left(\Pi(A), \Pi\left(\overline{A}\right)\right) = 1 \tag{4.19}$$

Une mesure de possibilité Π est totalement définie si toute partie de référence Ω se voit attribuer un coefficient dans [0,1]. Dans ce cas, si $|\Omega| = m$, il faut déterminer 2^m coefficients pour connaître complètement Π. Pour définir cette mesure, il suffit d'indiquer les coefficients attribués aux singletons de Ω, pour cela, on introduit une autre fonction appelée *distribution de possibilité* π, qui attribue à tout singleton de Ω un réel dans [0,1],et qui vérifie la condition de normalisation suivante :

$$\Pi(\{x\}) = \pi(x), \qquad \underset{x\in\Omega}{Sup}\,\pi(x) = 1,$$

$$\text{Avec : } \Pi(A) = \underset{x\in A}{Sup}\,\pi(x) \tag{4.20}$$

La distribution des possibilités $\pi(x)$ est composée des mesures de possibilités $\Pi(A)$ de tout les éléments $x \in A$ qui composent une hypothèse A. La mesure de possibilité $\Pi(A)$ est donc représentée par l'élément possédant la plus grande mesure de possibilité $\pi(x)$.

Contrairement à la théorie des probabilités, un événement A **ou** son contraire $\overline{A}$ sont tout à fait possible, ce qui signifie que, de deux événements contraires, l'un au moins est possible. Ceci dit, la possibilité de l'un n'implique pas l'impossibilité de l'autre. Ce qui peut conduire à une possibilité unitaire concernant ces deux événements contraires.

$$\Pi(A) = \Pi\left(\overline{A}\right) = 1. \tag{4.21}$$

Une mesure de possibilité ne présente pas *l'incertitude* d'un événement A, mais elle fournit uniquement une information sur l'occurrence de cet événement relatif à un ensemble de référence, plus ce degré est proche de 1, plus il est certain que l'événement sera réalisé. L'équation (4.21) décrit une situation d'ignorance totale, c'est-à-dire, elle exprime une indétermination complète sur la réalisation de l'événement A. pour lever cette ambiguïté, on

complète l'information sur *A* par un degré avec lequel la réalisation de *A* est certaine par une mesure appelée *mesure de nécessité*.

IV.3.2.2.2- Mesure de Nécessité :

La mesure de nécessité *N* défini le caractère prioritaire d'un événement, elle indique le degré avec lequel la réalisation d'un évènement *A* est *certaine* [ABB12]. Nous pouvons définir la mesure de nécessité *N* d'un évènement $A \subseteq \Omega$, comme une fonction $N : \wp(\Omega) \to [0,1]$ qui associe à *A* un degré de nécessité compris dans l'intervalle [0,1], tel que :

$$(N(\phi) = 0) \leq N(A) \leq (N(\Omega) = 1) \qquad (4.22)$$

Et

$$N\left(\bigcap_{i=1,2,\ldots} A_i\right) = Inf_{i=1,2,\ldots} N(A_i) \qquad (4.23)$$

Dans le cas de deux parties de Ω , (4.23) devient :

$$\forall A_1, A_2 \subseteq \Omega, \qquad N(A_1 \cap A_2) = Min(N(A_1), N(A_2)). \qquad (4.24)$$

De façon analogue à (4.19), on peut déduire à partir de (4.22) et (4.23), la relation suivante :

$$Min\left(N(A), N\left(\overline{A}\right)\right) = 0. \qquad (4.25)$$

On dit qu'un événement est nécessaire si son événement contraire est impossible $\Pi\left(\overline{A}\right) = 0$, donc, la mesure de nécessité *N* peut être obtenue par la mesure de possibilité Π correspondante :

$$\forall A \subseteq \Omega, \qquad N(A) = 1 - \Pi\left(\overline{A}\right). \qquad (4.26)$$

L'axiome (5.26) signifie que la maximisation de la nécessité de l'événement *A* minimise la possibilité de son événement contraire $\overline{A}$, et maximise la certitude de la réalisation de *A* [ABB12].

Comme la mesure de possibilité Π , La mesure de nécessité peut également s'exprimer en fonction de la distribution de possibilité π pour un événement *A*:

$$\forall A \subseteq \Omega, \qquad N(A) = Inf_{x \notin A}(1 - \pi(x)). \qquad (4.27)$$

La théorie des possibilités nous permet d'introduire la notion d'ignorance totale associée à un évènement et son contraire (5.21). Nous ne pouvons pas de distinguer ce cas d'ignorance totale, où tous les éléments ont une mesure de possibilité égale à 1. L'ignorance totale est représentée par la théorie des probabilités sous la forme d'équiprobabilités [BOU06].

Nous avons alors un compromis entre les mesures de possibilité et nécessité :

$$N(A) > 0 \Rightarrow \Pi(A) = 1 \tag{4.28}$$

$$\Pi(A) < 1 \Rightarrow N(A) = 0 \tag{4.29}$$

Le modèle de possibilité permet alors de modéliser l'imprécision et l'incertitude incluse dans une proposition A par un couple de mesures de possibilité et de nécessité $(\Pi(A), N(A))$. Cette théorie permet de traiter les incertitudes de nature non probabiliste. L'interprétation d'un degré de possibilité est très différente de celle de probabilité. Un fort de degré de possibilité n'implique pas un fort de degré de probabilité et qu'un faible degré de probabilité n'est pas synonyme d'un faible degré de possibilité. Seulement nous pouvons dire qu'un degré de possibilité nul implique une probabilité nulle. [CAV12]

IV.3.3. Théorie des Croyances :

La théorie des fonctions de croyances, appelée aussi la théorie de l'évidence [BEL08] a été développée par Shafer en 1976 [SHA76] grâce aux travaux de Dempster [DEM67] sur les probabilités supérieurs et inférieurs. Elle permet de manière analogue à la théorie des possibilités, de représenter à la fois l'imprécision et l'incertitude, à l'aide de fonction de masse *m*, de plausibilité *Pls* et de crédibilité *Cr*. Le formalisme mathématique repose tout d'abord sur la définition de masse accordée aux événements.

IV.3.3.1. Fonction de masse:

IV.3.3.1.1. Définition de la fonction de masse:

Dans cette théorie, la croyance sur une proposition est définie par la fonction de masse. Les fonctions de masse sont définies sur tous les sous ensembles de l'espace de discernement Ω et pas seulement sur les singletons comme les probabilités. Une fonction de masse *m* est définie de 2^{Ω} dans [0,1] par :

$m(\phi)=0$

Et une normalisation de la forme: (4.30)

$$\sum_{i=1}^{2^{\Omega}} m(A_i)=1.$$

Les éléments A de Ω tels que $m(A)>0$ sont appelés des *éléments focaux* de m et constituent le noyau de la structure de croyance [BEL08]. Lorsque les éléments focaux se réduisent aux singletons de Ω, les masses coïncident avec les probabilités. [LAM13]

IV.3.3.1.2. Interprétation particulières des fonctions de masse de croyance :

A la différence des probabilités, on voit que la théorie de Dempster-Shafer permet de répartir de la *masse* à des sous ensembles de Ω et non uniquement à des singletons. Cette possibilité fournit une modélisation très souple et très riche des connaissances imparfaites, en particulier de l'imprécision et de l'incertitude

La théorie des fonctions de croyance, contrairement à la théorie des probabilités, est donc bien adaptée pour représenter des connaissances à la fois imprécises et incertaines. De manière très naturelle et d'après [BOU07] [BEL08] [CAV12], soit un espace de discernement : $\Omega=\{A,B,C\}$. Nous pouvons définir :

- **Connaissance imprécise et incertaine :** correspond à l'affectation des fractions de la masse unité à plusieurs éléments focaux singletons et composés. Par exemple: $m(\{A\})=0,7$; $m(\{B\})=0,2$; $m(\{A,C\})=0,1$.
- **Connaissance imprécise et certaine :** correspond à l'allocation de toute la certitude à un élément focal non singleton. Par exemple: $m(\{A,C\})=1$.
- **Connaissance précise et certaine :** correspond à l'attribution de la totalité de la masse à un élément focal singleton de Ω. Par exemple : $m(\{A\})=1$.
- **Connaissance incertaine ou Bayésienne :** correspond à l'allocation de fraction de la masse unité à des éléments focaux singletons. Par exemple : $m(\{A\})=0,7$; $m(\{B\})=0,2$; $m(\{C\})=0,1$.

Un autre intérêt de modélisation par la théorie des fonctions de croyance est la prise en compte de *l'incomplétude* dans la connaissance ou s'il ya un conflit entre les sources, cette

théorie traite ce cas par la notion de monde ouvert est mal appréhendée par la théorie des probabilités et de possibilités. Dans ce cas $m(\phi)$ est alors mesure de conflit.

- **L'hypothèse du monde ouvert :** qui considère l'ensemble des hypothèses Ω ouvert à autre hypothèses (la notion d'incomplétude), c'est-à-dire qu'il peut exister certaines possibilités non recensées initialement. Dans ce cas, nous avons : $m(\phi) \neq 0$.
- **L'hypothèse du monde fermé :** suppose que le cadre de discernement recense toutes les réponses possibles de la question considérée, ce qui signifie que Ω désigne l'événement certain et ϕ l'événement impossible. Nous avons : $m(\phi) = 0$.
- **L'ignorance totale :** correspond à l'affectation de toute la certitude à l'ignorance Ω. Dans ce cas, nous avons : $m(\Omega) = 1$ et $m(A) = 0 \quad \forall A \neq \Omega$.

- **L'ignorance partiel le:** correspond à l'attribution de toute ou une fraction de certitude à une hypothèse composée (à l'exception de Ω). Elle représente les informations imprécises. , par exemple, $m(\{A, B\}) = 1$ ou $m(\{A, B\}) = 0.7$.

Donc, la fonction masse $m(A)$ représente une mesure subjective non additive des chances de réalisation de l'événement *A*, c'est-à-dire $m(A)$ quantifie dans quelle mesure l'événement *A* est crédible [RIC08]. Le grand problème réside dans cette théorie est l'estimation de masse, qui n'a pas de solution universelle. La difficulté est augmentée ici si l'on veut affecter des masses aux hypothèses composées. [BLO05]

IV.3.3.2. Fonctions associées :

A partir d'une distribution de masse qui représente une croyance stricte, il est possible d'obtenir des représentations équivalentes mais différentes sémantiquement. Les plus importantes sont les fonctions non additive de crédibilité notée *Cr* et de plausibilité notée *Pl* qui sont respectivement les limites inférieurs et supérieurs d'une masse *m*, ce qui défini comme suit :

IV.3.3.2.1. Fonction de crédibilité :

La fonction de crédibilité *Cr(A)* (Beleif, en anglais *Bel*(A)), représente le degré de *croyance minimum* correspondants à toutes les masses affectées à tous les éléments focaux *B* qui *impliquent A,* formellement la fonction de crédibilité est définie comme suit :

$$Cr : 2^{\Omega} \rightarrow [0,1]$$

Où : $Cr(\phi) = 0$

$$Cr(\Omega) = 1 \qquad (4.31)$$

$$Cr(A) = \sum_{B \subseteq A} m(B).$$

L'interprétation ensembliste de cette mesure correspond à la somme des poids de croyance de sous-ensembles de *A* et de lui-même, avec, $m(\phi)$ n'est pas pris en compte dans le calcul du degré de croyance (ϕ n'est pas impliquer *A*). [BEL08]

IV.3.3.2.1. Fonction de plausibilité :

En contrepartie de la fonction de crédibilité, la fonction de plausibilité *Pl(A)* représente la *croyance maximale* que l'on pourrait affecter à un sous-ensemble *A*. numériquement, la plausibilité d'un sous ensemble *A* prend la somme de toutes les masses attribuées aux éléments pour lesquels l'intersection avec *A* est non nulle.

$$Pl : 2^{\Omega} \rightarrow [0,1]$$

Où : $Pl(\phi) = 0$

$$Pl(\Omega) = 1$$

$$Pl(A) = \sum_{B \cap A \neq \phi} m(B). \qquad (4.32)$$

Donc, la sortie d'un processus de Dempster-Shafer est un ensemble d'intervalles de l'évidence [*Cr(A), Pl(A)*] qui encadrent la probabilité mal connue. [DES96, BOU07], où, $Pl(A) - Cr(A)$ mesure l'ignorance relative à *A*.

$$Cr(A) \leq P(A) \leq Pl(A) \qquad (4.33)$$

De la même façon, la plausibilité *Pl* d'un sous ensembles d'hypothèses *A* est définie par la crédibilité de l'évènement contraire $\overline{A}$ [DES96]:

$$Pl(A) = 1 - Cr(\overline{A}) \qquad (4.34)$$

L'équation (5.34) montre que plus nous augmentons la croyance dans l'hypothèse *A*, plus l'inverse de cette hypothèse deviens moins plausible.

Selon [LAM13], Shafer a été démontré dans [SHA76] que la connaissance d'une des trois fonctions *(m, Cr, Pl)* sur 2^{Ω} était suffisante pour en déduire les deux autres.

Dans la théorie de Dempster-Shafer, on modélise l'ignorance totale par l'affectation de toute la masse *m* à Ω (5.35), qui implique de l'affectation de toute la crédibilité *Cr* à Ω [BOU07]:

$$\forall A \subset \Omega : Cr(\Omega) = 1 \quad \text{et} \quad Cr(A) = 0. \tag{4.35}$$

Contrairement à l'ignorance totale, l'équi-répartition dans la théorie des croyances est représentée par la fonction de plausibilité, donc, la théorie des croyances est vérifiée l'équi-répartition si :

$\forall a \in \Omega, Pl(\{a\}) = cte$, par exemple, soit la distribution de plausibilité suivante :

$$\forall a \in \Omega, Pl(\{a\}) = \frac{1}{|\Omega|} \tag{4.36}$$

IV.4. Fusion de données:

Dans cette partie, nous montrons comment la fusion de données peut être envisagée dans les trois cadres théoriques introduits dans la section précédente. Le concept de fusion de données est facile à comprendre mais la difficulté réside de trouver une définition adéquate. Plusieurs définitions ont été proposées dans la littérature pour la fusion d'informations depuis 1987, nous reprenons ici quelques unes extraites de différents travaux:

IV.4.1 Définitions:

Dans [GRA06], « *la fusion de données est un procédé traitant de l'association, de la corrélation et de la combinaison des données et de l'information provenant de sources uniques ou multiples afin d'aboutir à des estimées affinées des positions et des identités, à une évaluation complète et en compte utile des situations et des menaces, et de leurs importances. Le procédé est caractérisé par une amélioration continu des estimées et des prévisions, et par*

l'évaluation du besoin en sources additionnelles ou de modification du procédé lui-même, afin d'atteindre des résultats améliorés ».

[BLO05], l'auteur a donné une définition de fusion de données comme suit : «*La fusion de données consiste à combiner des données (souvent imparfaites et hétérogènes) issues de plusieurs sources afin d'améliorer la prise de décision».*

Dans [ZOU07], le groupe européen SEE (Société d'Electricité et d'Eléctronoque), la branche française de l'Institute of Electric and Electronics Engineers (IEEE), et la branche européenne de International Society for Photogrammetry and Remote Sensing (ISPRS) ont proposé la définition suivante : « *la fusion de données consiste un cadre formel dans lequel s'expriment les moyens et techniques permettant l'alliance des données provenant de sources diverses».*

En français, le terme fusion est généralement adopté. En revanche, en anglais, plusieurs termes sont en concurrence dans la littérature, comme par exemple « merging » et « combining ». [ASS10].

IV.4.2. Processus de fusion de données:

Les étapes qui constituent le processus de fusion sont représentées ci-après [BOU12] [ABB12] :

- **Représentation homogène et recalage des informations pertinentes :** Dans cette étape initiale, nous commençons par la transformation de certaines informations initiales hétérogènes vers des informations homogènes dans un espace de représentation commun de fusion ;
- **Modélisation :** Cette seconde étape consiste à choisir un formalisme pour exprimer les informations à fusionner. La modélisation peut être guidée par les informations supplémentaires sur le contexte et le domaine d'application ;
- **Estimation :** Cette étape consiste à déterminer numériquement les valeurs des informations modélisées ;
- **Combinaison :** Cette étape consiste à choisir un opérateur de combinaisons qui s'approprie avec le formalisme de modélisation retenu ;

- **Décision :** C'est l'étape finale du processus de la fusion de données. Après la détermination d'un critère de décision selon le cadre théorique choisi et l'objectif à atteindre, nous allons sélectionner l'information *la plus vraisemblable* parmi toutes les hypothèses possibles.

IV.4.3. Classification des opérateurs de fusion :

Dans [BLO96], l'auteur a illustré trois grands types d'opérateurs de fusion de données. La notion de fusion s'applique au cas de *n* sources. Pour simplifier, nous considérons, dans la suite, la fusion d'informations issues de deux sources n_1, n_2. On cherche à agréger les informations fournies par n_1 et n_2 en exploitant au mieux l'ambiguïté et la complémentarité des données. Pour cela, on réalise l'agrégation par un opérateur binaire $F(.,.)$ auquel on impose une contrainte de *fermeture* c'est-à-dire, la valeur retournée doit être de même nature que les informations d'entrée, par exemple probabiliste.

La classification des opérateurs de fusion a été proposée dans [BLO96], le comportement de F est qualifié de :

- *Sévère* si : $F(n_1, n_2) \leq min(n_1, n_2)$;
- *Prudent* si : $min(n_1, n_2) \leq F(n_{1,} n_2) \leq max(n_1, n_2)$;
- *Indulgent* si : $F(n_1, n_2) \geq max(n_1, n_2)$;

En relation avec la notion de conjonctions (l'opérateur T-norme) et de disjonctions (l'opérateur T-conorme) illustrées dans [CAV12], Un opérateur *sévère* suppose que les deux capteurs sont fiables et exploite l'information commune aux deux mesures. Au contraire, un opérateur *indulgent* agit sur des sources *a priori* en conflit. Il augmente la certitude sur l'événement observé (et donc augmente l'imprécision…) et exprime la redondance entre les informations. Un opérateur *prudent* correspond à une position intermédiaire entre ces deux extrêmes [BAR00] [FRE04].

Cette distinction ne suffit pas à classer les opérateurs dont le comportement n'est pas toujours le même. Il existe différents types d'opérateurs [BLO96], suivant leur variabilité et leur dépendance au contexte et pas seulement comme disjonctifs ou conjonctifs. Dans [BAR00][BLO05], les auteurs ont rappelé brièvement la taxinomie qui nous permettra de caractériser le comportement des fusions présentées dans les paragraphes suivants.

IV.4.3.1. Opérateurs à comportement constant et indépendant du contexte (CCIC):

Cette première classe est appelée la classe des opérateurs à comportement constant et indépendant du contexte (CCIC), ce sont des opérateurs ayant le même comportement quelques soit les valeurs n_1 et n_2 à agréger. Le résultat de la fusion est indépendant du contexte de l'agrégation. L'opérateur F est donc exclusivement sévère, indulgent ou prudent. Parmi ces opérateurs, on compte par exemple les opérateurs *min* et *max*, les opérateurs *médians*, … etc.

IV.4.3.2. Opérateurs à comportement variable et indépendant du contexte (CVIC):

Cette seconde classe regroupe les opérateurs qui ne dépendent pas du contexte mais dont le résultat est fonction des valeurs n_1 et n_2 (de leur valeur absolue par exemple). Un exemple d'opérateurs de fusion à comportement variable et indépendant du contexte, on peut trouver l'opérateur de la somme symétrique dans la théorie des possibilités [BLO96]. Elle est de la forme : $\boldsymbol{\sigma = \frac{g(x,y)}{g(x,y)+g(1-x,1-y)}}$ avec σ définie de [0,1]×[0,1] dans [0,1] telles que, $\sigma(0,0) = 0$, σ est commutative, croissante, continue. Avec g est croissante, positive et continue telle que $g(0,0) = 0$.

IV.4.3.3. Opérateurs dépendant du Contexte (CDC):

Enfin, la dernière classe regroupe les opérateurs à Comportement Dépendant du Contexte (CDC). La valeur retournée par F ne dépend plus seulement de n_1 et n_2, mais aussi d'une connaissance *a priori* sur le système de capteurs ou le phénomène étudié. F peut être construit avec un comportement sévère par une fonction croissante de l'accord entre les deux capteurs (respectivement, indulgent-fonction décroissante).

IV.4.3.4- Quelques propriétés :

Un opérateur doit être si possible associatif et commutatif (l'opérateur étant alors indépendant de l'ordre de présentation des informations), continu (ce qui assure la robustesse de la combinaison pour des couples de mesures voisins) et strictement croissant par rapport au couple (n_1, n_2). [BAR00]

IV.5. Fusion en théorie des probabilités :

Dans cette théorie, il existe deux types d'opérateurs de fusion, la première basée sur la règle des probabilités conditionnelles de Bayes. La seconde s'appuie sur la combinaison des

probabilités par différents types d'opérateurs tels que, la *moyenne pondérée* avec un somme de poids égale à 1, *le minimum*, *le maximum*, *la médian* …etc, pour plus de détaille voir [KIT98]. l'objectif de fusion dans ce cas est de chercher la probabilité d'un objet observé appartient à une classe précise. La règle plus utilisée pour la combinaison en théorie des probabilités est la règle de Bayes. Nous détaillons dans la suite la première méthode de *fusion Bayésienne*.

IV.5.1. Combinaison Bayésienne:

Pour combiner l'ensemble des déclarations des capteurs, on peut utiliser la formule de Bayes. Soient les événements $C_1, C_2, \ldots . C_m$, (les événements sont les classes dans le problème de classification). Soient $n_1, n_2, \ldots, n_p$ les mesures arrivants des p capteurs $I_1, I_2, \ldots, I_p$ (p est le nombre de classifieurs utilisés). La règle de Bayes permet de calculer la probabilité d'obtenir l'événement C_i en ayant effectivement mesuré $n_1, n_2, \ldots, n_p$.

$$p(C_i|I_1, I_2, \ldots, I_p) = \frac{p(I_1|C_i).p(I_2|C_i,I_1)\ldots p(I_p|C_i,I_1,\ldots,I_{p-1}).p(C_i)}{p(I_1).p(I_2|I_1)\ldots p(I_p|I_1,\ldots,I_{p-1})} \quad (4.37)$$

En supposons que les sources sont indépendantes afin de simplifier les calculs et l'estimation des paramètres de l'équation de Bayes, la formule précédente devient alors [BLO99]:

L'indépendance conditionnelle signifie que les erreurs des mesures sont indépendantes. [GRA06]

$$p(C_i|I_1, I_2, \ldots, I_p) = \frac{p(C_i).\prod_{k=1}^{p} p(I_k|C_i)}{p(I_1, I_2, \ldots, I_p)} \quad (4.38)$$

Où :

$p(I_1, I_2, \ldots, I_p) = \sum_{i=1}^{m} p(C_i)\prod_{k=1}^{p} p(I_k|C_i)$: mesure la concordance entre les sources dont l'*a priori*.

En ayant estimé les probabilités conditionnelles comme illustré dans le troisième chapitre, section II.3.1. L'équation précédente décrit un opérateur de fusion entre les données

fournies par p capteurs ($p \geq 2$). Cet opérateur fait apparaître le type de combinaison des informations, sous forme d'un produit, donc une fusion conjonctif. En plus, cet opérateur est de type CCIC.

IV.5.2. Règle de décision :

La décision en théorie des probabilités est généralement la règle de maximum *a posteriori,* elle consiste à privilégier l'événement ayant la plus forte probabilité *a posteriori* à l'issue de l'étape de fusion. Dans notre problème de classification, le résultat est l'affectation d'un élément x à la classe la plus probable. La règle du maximum *a posteriori* définie comme suit :

$$x \in C_i \; si \;\; p\left(x \in C_i | n_1, n_2, \dots, n_p\right) = max\left(p\left(x \in C_i | n_1, n_2, \dots, n_p\right); 1 \leq k \leq m\right) \tag{4.39}$$

D'autres critères de décision Bayésienne existent dans la littérature, on peut citer, le maximum de vraisemblance, le maximum d'entropie [GRA06], la marginale maximale et l'espérance maximale [BLO04] [MAR05].

IV.6. Fusion en théorie de croyance :

IV.6.1. Combinaison :

Lors de la combinaison de sources distinctes d'informations dans la théorie des croyances, plusieurs opérateurs ont été proposés, nous trouvons principalement deux types d'opérateurs de combinaison, la combinaison *disjonctive* et la combinaison *conjonctive* qui ont été déclinés en un grand nombre d'opérateurs de combinaison dont des combinaisons *mixtes* [BEL08] :

IV.6.1.1. Combinaison conjonctive :

L'intérêt majeur de la théorie d'évidence en fusion de données repose sur la possibilité de construire une *fonction de masse unique par combinaison* de plusieurs fonctions de masse issues de plusieurs sources d'information distinctes. Ils existent des règles de combinaison qui permettent de fusionner p sources en combinant leurs masses. On rencontre l'opérateur de *Dempster-Shafer* ou l'opérateur de la *somme orthogonale* qui est le premier opérateur de combinaison défini dans le cadre de la théorie des croyances, il est développé en 1976

[SHA76], il permet de combiner deux ou plusieurs fonctions de masse en une seule, il est donné pour tout $A \in 2^{\Omega}$ par:

Dans le cas de deux sources, il s'écrit :

$$m(A) = (m_1 \oplus m_2)(A) = \sum_{B \cap C = A} m_1(B) m_2(C) \tag{4.40}$$

Dans le cas de plusieurs sources (p fonctions de masse), il s'écrit:

$$m(A) = (m_1 \oplus m_2 \oplus ... \oplus m_p)(A) = \frac{\sum_{B_1 \cap B_2 \cap .... \cap B_p = A} \prod_{k=1}^{p} m_k(B_k)}{L} \tag{4.41}$$

Avec :

$$L = 1 - \sum_{B_1 \cap B_2 \cap .... \cap B_p = \phi} \prod_{k=1}^{p} m_k(B_k) = \sum_{B_1 \cap B_2 \cap .... \cap B_p \neq \phi} \prod_{k=1}^{p} m_k(B_k) \tag{4.42}$$

Où :

$m(\phi) = 0$, m_k $(k = 1..p)$ est la fonction de masse définie pour la source k

On peut généraliser la règle de *Dempster-Shafer* à $p.n$ fonctions de masses :

$$. m(A) = \sum_{B_1 \cap B_2 \cap .... \cap B_p = A} \prod_{i=1}^{n} \prod_{k=1}^{p} m_{ki}(B_k) \tag{4.43}$$

La règle de combinaison de *Dempster-Shafer* vérifie les propriétés suivantes. Elle appartient à la classe CCIC des opérateurs de fusion [BLO96], elle est associative, commutative, possède un élément neutre $(m(\Omega) = 1)$.

L représente le degré de conflit entre les p sources. Et la quantité *1-L* représente la masse qui aurait été assignée à l'ensemble vide si les masses *n'étaient pas normalisées* ($m(\phi) = L - 1$). Il est très important de prendre en compte cette quantité pour évaluer la qualité de l'agrégation. En effet, si L est proche de 0, le conflit entre les sources est important. De plus, la règle de combinaison est discontinue pour des valeurs de L proches de 0 [BAR00].

Comme on le voit, l'opérateur de Dempster-Shafer est une forme normalisée, la forme non-normalisée développée plus tard par Smets en 1990 [SME90-a] pour aborder les *mondes ouverts*. Smets propose donc la combinaison :

$$\begin{cases} m(A) = \sum_{B_1 \cap B_2 \cap \ldots \cap B_p = A} \prod_{k=1}^{p} m_k(B_k) \\ m(\phi) = \sum_{B_1 \cap B_2 \cap \ldots \cap B_p = \phi} \prod_{k=1}^{p} m_k(B_k) \end{cases} \quad (4.44)$$

Il existe d'autres opérateurs de combinaison conjonctif ont été proposées, les principales sont de *Yager* et *Inagaki*. Voici une récapitulation des différentes règles de combinaison dans la théorie des croyances [DJI08][MAR05][SEN02].

IV.6.1.2. Combinaison disjonctive :

D'autres modes de combinaison en théorie des croyances, tels que des modes disjonctifs ou mixtes sont possibles, en remplaçant l'intersection dans la formule de combinaison de la somme orthogonale par une opération ensembliste, c'est-à-dire en fusion disjonctif en prenant la réunion. Alors la formule de combinaison disjonctive est décrite comme suit :

$$m(A) = (m_1 \oplus_{\cup} \ldots . \oplus_{\cup} m_p)(A) = \sum_{B_1 \cup B_2 \cup \ldots . \cup B_p = A} \prod_{k=1}^{p} m_k(B_k) \quad (4.45)$$

IV.6.1.3. Combinaison mixte :

Il existe aussi des combinaisons qui combinent deux approches conjonctive et disjonctive, pour conserver les avantages de combinaison conjonctive et disjonctive on trouve dans ce type l'opérateur adaptative de Dubois et Prade. Cette combinaison est donnée par tout $A \in 2^{\Omega}$, $A \neq \phi$, par :

$$m(A) = \sum_{B_1 \cap B_2 \cap \ldots . \cap B_p = A} \prod_{k=1}^{p} m_k(B_k) + \sum_{B_1 \cup B_2 \cup \ldots . \cup B_p = A et B_1 \cap B_2 \cap \ldots . \cap B_p = \phi} \prod_{k=1}^{p} m_k(B_k) \quad (4.46)$$

IV.6.2. Règles de décision :

A la fin de l'étape de combinaison dans la théorie des croyances, on obtient les masses relatives à chaque élément du cadre de discernement issue de la combinaison des différentes

sources. Il est alors aisé de calculer les fonctions de croyance (crédibilité, *Cr*) et de plausibilité (*Pl*) par les équations (4.31) et (4.32) respectivement. La dernière étape qui reste est celle de prise de décision.

Dans la théorie des fonctions de croyances, plusieurs règles de décision possibles pour le choit d'une hypothèse. Ce choix peut ce faire par la maximisation d'un critère. Parmi ces critères, nous citons, maximum de plausibilité, maximum de crédibilité, celle dont l'intervalle de confiance ou encore celle de probabilité de pignistique maximum [MAR05] :

- *Maximum de plausibilité :*

$$x \in A_i \quad si \quad Pl(A_i)(x) = max\{Pl(A_k)(x),\ 1 \leq k \leq m\}. \tag{4.47}$$

- Maximum de crédibilité :

$$x \in A_i \quad si \quad Cr(A_i)(x) = max\{Cr(A_k)(x),\ 1 \leq k \leq m\}. \tag{4.48}$$

- L'intervalle de confiance $[Cr, Pl]$:

En effet, un intervalle de confiance de [1, 1] certifie que cette hypothèse est réellement sûre, alors qu'un intervalle de confiance de [0,0] certifie que cette hypothèse est impossible. Un intervalle de [0.7, 0.9] indique que l'hypothèse est sûrement vraie, et un intervalle de [0.2, 0.4] décrit une hypothèse sûrement fausse. Plus l'intervalle de confiance est étroit, plus l'incertitude de l'hypothèse est faible. Plus il se rapproche des valeurs de [1,1], plus l'hypothèse est probable. [DRO98]

- Maximum de probabilité pignistique :

Le mot pignistique provient de *pignus* en latin signifiant pari (*bet* en anglais). Un compromis au critère du maximum de plausibilité et du maximum de crédibilité a été proposé par Philippe Smets en 1990 [SME90], où il introduit le maximum de probabilité pignistique. Il permet de transformer la distribution de croyance en une distribution de probabilité. Le maximum de probabilité pignistique, notée $\left(bet(A_i)\right)$ est définie Pour toute hypothèse A_i, la par la formule suivante [MAR05]:

$$bet(A_i) = \sum_{A \in 2^{\Omega}, A_i \in A} \frac{m(A)}{|A|\left(1-m(\emptyset)\right)} \tag{4.49}$$

Avec : $|A|$ représente la cardinalité de *A.* ainsi le critère de maximum de probabilité pignistique revient à décider A_i pour l'observation *x* si :

$$bet(A_i)(x) = \max_{1 \leq k \leq m} bet(A_k)(x) \quad (4.50)$$

Dans l'application de fusion de données représentée dans la figure 33 [GRA06], on voit la manière d'élaborer l'intervalle résultant de la combinaison des intervalles initiaux. Une décision est prise en fin de chaîne de traitement.

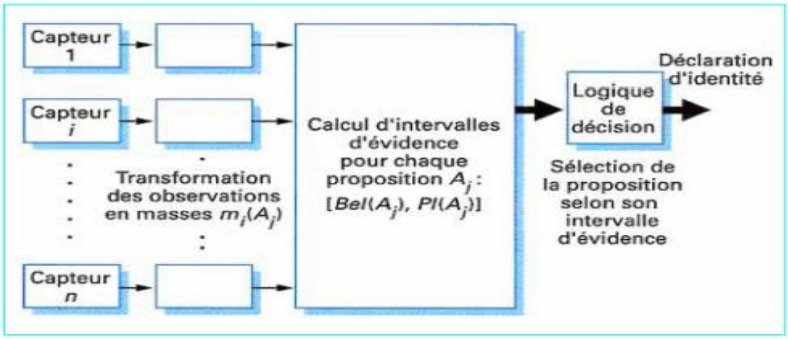


Figure 33:fusion utilisant la théorie de Dempster-Shafer

IV.7. Fusion en théorie des possibilités :

IV.7.1. Combinaison :

Un des avantages des ensembles flous et des possibilités est le nombre important d'opérateurs disponibles permettant la combinaison de fonctions d'appartenance ou des distributions de possibilités. L'idée générale derrière une approche possibiliste de la fusion d'informations est qu'il n'existe pas de mode unique de combinaison. Tout dépend de la situation étudiée et de la confiance accordée aux capteurs. De nombreux auteurs ont proposé une étude précise et détaillée des opérateurs de fusion dans le cadre possibiliste : [BLO96] [BLO04] [MAR05]. Le grand nombre d'opérateurs pousse à un classement de ceux-ci afin de permettre le choix de l'opérateur le plus adapté à l'application. Ce choix peut se faire selon plusieurs critères liés aux propriétés des opérateurs. Deux types de classification des opérateurs de fusion existent dans ce contexte sont présentés dans [BLO96]. Dans le premier temps, il est courant de considérer les opérateurs selon leur comportement sévère, indulgent ou prudent, c'est-à-dire les opérateurs conjonctifs, disjonctifs et de compromis. Dans le deuxième temps, les opérateurs sont classés en fonction de leur comportement selon les valeurs des informations à combiner (CCIC, CCVC, CDC). Nous en présentons les principaux de ces opérateurs :

IV.7.1.1. Les opérateurs possibiliste à comportement constant et indépendant du contexte :

Selon [BLO96], on distingue dans cette catégorie, les opérateurs conjonctifs, disjonctifs ou de compromis :

IV.7.1.1.1- Les opérateurs conjonctifs :

Les opérateurs conjonctifs combinent l'information à la façon d'un ET logique (conjonction). L'opérateur conjonctif possibiliste le plus connu et l'opérateur Triangulaire-norms.

T-norms (Triangular-norms) : cet opérateur généralise la notion d'intersection des ensembles. Pour tout T-norms *t*, nous avons :

$$\forall(x,y) \in [0,1]^2, t(x,y) \leq min(x,y) \qquad (4.51)$$

Dans l'opérateur T-norms, le résultat ne dépend que des valeurs à combiner (sans autre information) et le comportement est le même quelles que soient ces valeurs. Pour cela, cet opérateur appartient à la classe CCIC. On distingue plusieurs opérateurs de ce type [MAR05][BLO04], nous trouvons, T-norms de Zadeh, T-norms de Lukasiewicz, T-norms de Yager, T-norms de Hamacher, T-norms Archimédienne…etc.

IV.7.1.1.2- Opérateurs disjonctifs :

Les opérateurs disjonctifs combinent l'information à la façon du OU logique (disjonction). La valeur du résultat de la combinaison sera grande dès une des valeurs combinées le sera. Le comportement sera donc indulgent. Les principaux opérateurs disjonctifs sont les conorms Triguliares ou T-conorms.

Opérateurs T-conorms (Triangular-conorms) : cet opérateur généralise la notion de l'union des ensembles. Pour tout T-conorms *t*, nous avons :

$$\forall(x,y) \in [0,1]^2, t(x,y) \geq max(x,y) \qquad (4.52)$$

L'observation ne peut donc pas appartenir plus fortement à l'union des sous-ensembles qu'il n'appartient à chacun d'eux. Alors, l'opérateur T-conorms comme l'opérateur T-norms appartient à la classe des opérateurs à comportement constant et indépendant du contexte (CCIC). Comme on le voit dans les différents types d'opérateurs T-norms, même

chose avec les opérateurs T-conorms, où il y a l'opérateur T-conorms de Zadeh, de Lukasiewicz, de Yager, de Hamacher, T-Conorms Archimédienne…etc. [MAR05].

IV.7.1.1.3- Les opérateurs de compromis :

Les opérateurs de compromis ont un comportement moins sévère que les opérateurs conjonctifs, et moins indulgent que les opérateurs disjonctifs, c'est-à-dire, le résultat de la combinaison est toujours compris entre le *min* et le *max*, formellement, l'opérateur de compromis est une fonction t: [0,1]x[0,1] → [0,1] telle que :

$$\forall (x,y) \in [0,1]^2, min(x,y) \leq t(x,y) \leq max(x,y) \qquad (4.53)$$

Il existe plusieurs opérateurs de compromis possibiliste, on peut citer, Moyennes pondérées ordonnées (OWA), les intégrales flous (intégrale de Sugeno et l'intégrale de Choquet), les opérateurs de moyenne (moyenne harmonique, géométrique, arithmétique, …). Ces opérateurs sont détaillés dans le premier chapitre comme des opérateurs d'agrégation multicritères.

IV.7.1.2- Les opérateurs possibilistes à Comportement Variable et indépendant du contexte :

D'autres opérateurs de combinaison possibiliste ne peuvent pas être classés comme conjonctifs, disjonctifs ou compromis. Ces opérateurs sont classés comme des opérateurs à comportement variable et indépendant de contexte (CVIC) car, ces opérateurs change leurs résultats selon les valeurs à combiner [BLO96]. Nous avons illustré cette catégorie dans la section précédente et on prend comme un exemple les *sommes symétriques*. D'autres opérateurs sont classés dans cette catégorie, nous présentons l'opérateur hybride *T-norms et T-conorms* proposé par Zemmerman et Zysno [MAR05]. Cependant, ces opérateurs ont des propriétés limités (continuité, associativité, monotonie, neutralité). Cet opérateur est donné comme suit :

$$H(a_1,a_2,...,a_d) = \left(\prod_{k=1}^{d} a_k\right)^{1-\delta} \left(1-\prod_{k=1}^{d}(1-a_k)\right)^{\delta} \qquad (4.54)$$

Où : $\delta \in [0,1]$.

IV.7.1.3- Les opérateurs possibilistes à Comportement Dépendant du contexte :

Dubois et Prade ont introduit en 1992 des opérateurs adaptatifs en fonction du conflit entre les distributions de possibilité qui se comportent comme un minimum s'il y a peu de conflit entre les distributions et comme un maximum si le conflit est fort.

On prend l'exemple de [BLO96] sur les opérateurs possibiliste de type CDC, on note π_1 , π_2 deux distributions de possibilité définies sur un espace de discernement Ω et π' la distribution résultante, voici quelques opérateurs possibles de type CDC qui forme π' :

$$\pi'(s) = max\left[\frac{i[\pi_1(s),\pi_2(s)]}{h(\pi_1,\pi_2)}, 1 - h(\pi_1,\pi_2)\right] \quad (4.55)$$

$$\pi'(s) = min\left[1, \frac{i[\pi_1(s),\pi_2(s)]}{h(\pi_1,\pi_2)} + 1 - h(\pi_1,\pi_2)\right] \quad (4.56)$$

$$\pi'(s) = i[\pi_1(s),\pi_2(s)] + 1 - h(\pi_1,\pi_2) \quad (4.57)$$

Où :

$\mathbf{1} - \boldsymbol{h}(\boldsymbol{\pi_1},\boldsymbol{\pi_2})$ représente la mesure globale de conflit entre les deux distributions. Une mesure de conflit peut être donnée par:

$$h(\pi_1,\pi_2) = 1 - \max_{s\in\Omega} min\big(\pi_1(d),\pi_2(d)\big) \quad (4.58)$$

i est un opérateur conjonctif (T-norme).

IV.7.2- Règle de décision :

Une fois les informations issues des différentes sources combinées, la décision se faite à partir de maximum des degrés d'appartenance. Ainsi, nous choisissons l'hypothèse A_i pour l'élément *x* si :

$$\mu_i(x) = max\{\mu_k(x),\ k = 1 \ldots m\}. \quad (4.59)$$

Où μ_k désigne la fonction d'appartenance à la classe *k* (dans notre problème de classification) résultant de la combinaison.

La maximisation du degré d'appartenance revient à maximiser le degré de possibilité ou le degré de nécessité (choisir le critère le plus sévère).

IV.8. Conclusion :

Nous venons de présenter un panel non exhaustif des techniques de fusion de données. Cette présentation a néanmoins l'avantage de proposer un certain nombre de solutions particulièrement adaptées à la fusion ***multi-classifiuers*** dans des ***environnements dynamiques*** où il est nécessaire de prendre en compte une certitude et une précision fluctuante des classifieurs au cours du temps. Ces approches sont adaptées et adaptables à de nombreux problèmes en traitement d'images ,par exemple dans [HEG04][LEI92] pour la reconnaissance manuscrite, mais surtout dans le cas d'images médicales [BAR00][ZOU07][LAM13][LEL13] avec pour but de segmenter les images IRM ou de détecter les anomalies, la fusion d'images géographiques [OLT08][CHA95], les images radar pour la détection de mines [MIL03], le traitement de la parole [MAU04], Robotique [OUS01],…etc. Chacune des méthodes de fusion décrite a son propre modèle pour gérer les imperfections, le choix d'une méthode de fusion appropriée dans une situation décisionnelle donnée n'est pas trivial. Il n'y pas une méthode de fusion universelle offrant toujours les meilleurs résultats que les autres méthodes. En effet, le choix de la méthode appropriée dépend fortement de plusieurs facteurs reliés au contexte du travail.

Chapitre 05 : Système d'aide au diagnostic médical : L'ÉlectroEncéphaloGraphie (EEG)

V.1. Introduction:

Le cerveau humain est le centre de contrôle du corps, il est responsable de la perception, la cognition, l'attention, l'émotion, la mémoire, l'action, …etc. quand une personne pense, lire ou regarder la télévision, différentes parties du cerveau sont stimulés. Dans ce cas, des signaux électriques sont créés et avec des réactions chimiques laissent les parties du corps en communications. Ces signaux électriques peuvent être contrôlés grâce à des techniques scientifiques telles que, l'électroencéphalographie (EEG), électrocorticographie (ECoG), l'imagerie par résonance magnétique (IRM), l'imagerie par résonance magnétique fonctionnelle (IRMf), la tomographie par émission de position (TEP). Les techniques scientifiques mentionnées ci-dessus nous aider à acquérir une meilleure compréhension du cerveau tandis qu'une personne effectue un certain nombre de tâches. EEG est la technique la plus utilisée pour capturer les signaux du cerveau grâce à son excellente résolution temporelle, non invasif, la convivialité et les coûts de mise en place faibles. Un EEG peut monter dans quel état une personne se trouve, que ce soit dans sommeil, éveillé, anesthésié, par ce que les caractéristiques de chaque état électrique différents pour chacun d'autres états. EEG est de plus en plus important dans le diagnostic et le traitement des maladies et des anomalies mentales [SIU12]. *L'analyse, le suivi l'évolution et la classification* des signaux EEG sont essentielles à la fois pour soutenir le diagnostic des maladies du cerveau et de contribuer à une meilleure compréhension du processus cognitif. Le but principal de la classification des signaux EEG est de séparer les différents états de l'EEG et de décider si les gens sont en bonne santé ou d'estimer l'état mental. L'état d'un sujet lié à une tâche effectuée. Plus grandes quantités de données sont générées par EEG et l'inspection visuelle pour discriminer l'état du personne perd beaucoup de temps, et pas assez suffisante pour obtenir des informations fiables. Le développement de méthode de classification automatique, ainsi que une méthode pour suivi les signaux EEG est essentiel pour l'évaluation

et traitement des maladies neurologiques. Dans ce mémoire de magister, nous nous concentrons sur la classification, suivi l'évolution des signaux EEG dans le domaine le plus important, *l'épilepsie.*

Afin d'avoir une large compréhension du cerveau humain et l'outil de diagnostic (EEG), ce chapitre fournit essentiellement un aperçu et les connaissances de base sur le cerveau humain, l'EEG et l'épilepsie. Une brève terminologie est discutée à la section 5.2. La section 5.3 présente une idée sur l'outil de diagnostic EEG. La section 5.4 est consacré à déterminer quelques maladies neurologiques, en particulier l'épilepsie. La dernière section de ce chapitre discute les techniques de traitement du signal.

V.2. Les connaissances de base liées à des signaux EEG :

Cette section présente les terminologies et les informations connexes aux signaux EEG. Tant que les signaux EEG sont générés à partir du cerveau, cette section présente la structure anatomique et neurophysiologique du cerveau humain. Cette section se concentre particulièrement sur les fonctions des principales parties du cerveau et de la création des courants électrochimiques qui sont captés par des électrodes placés sur le cuir chevelu.

V.2.1. Cerveau humain :

Le cerveau est l'organe le plus complexe du corps humain, il régit notre comportement, nos actions, nos pensés, nos désire et nos instincts. Tout d'abords, dans cette section, nous expliquons les structures anatomiques du cerveau et leurs fonctions. En suite, nous nous concentrons sur comment, pourquoi et où le cerveau génère les activités électriques qui peuvent être capturés par EEG.

Le cerveau est un terme qui désigne le tissu que l'on trouve dans la crâne. Le cerveau possède deux moitiés relativement symétriques appelés des hémisphères, l'un à gauche et l'autre à droite. Chaque hémisphère est divisé en quatre lobes. *Le lobe pariétal, le lobe occipital, le lobe frontal, le lobe temporal.* La figure 34 [ROM10] représente les différents lobes du cerveau humain.

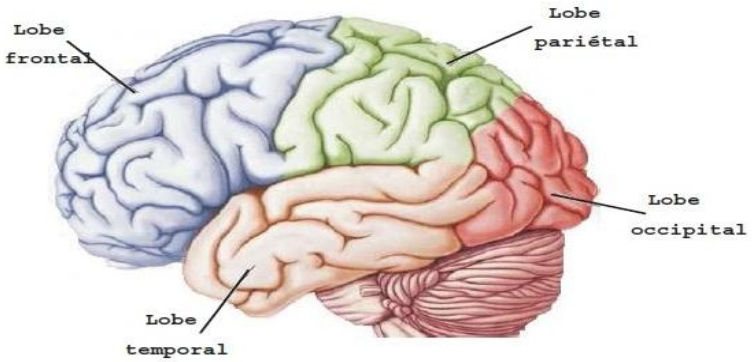


Figure 34: les différents lobes du cerveau humain.

Lobe frontal : est impliqué à la personnalité, les émotions, la résolution des problèmes, les actions motrices, le raisonnement, la planification.

Lobe pariétal : est responsable de la sensation (la douleur, le toucher), la reconnaissance, la compréhension sensorielle, la perception de stimuli, l'orientation et mouvement.

Lobe temporal : est responsable de la reconnaissance des stimuli auditifs, de la parole, de la perception de la mémoire.

Lobe frontal : est situé à l'avant de lobe occipital, il est impliqué dans le sens du toucher, dans la détection des mouvements dans l'environnement et la localisation des objets dans l'espace.

Lobe occipital: localisé à l'arrière du cortex, il est spécialisé de la vision.

V.2.2. Neurophysiologie:

En vue de comprendre les signaux EEG, cette section présente brièvement un aperçu de la neurophysiologie. L'ensemble de cette section est une synthèse provenant de [SIU12][HAD14][HOC12][MAC10][ROM10]. Le cerveau humain est constitué d'environ 100 milliards $\left(10^{11}\right)$ de cellules appelées *neurones.* Le neurone est la cellule de base du système nerveux. Un neurone est divisé en plusieurs régions ayant chacune une fonction propre, il composé de *l'axone*, s'étendant parfois sur une longue distance et servant à transmettre des signaux à d'autres cellules. *Les dendrites,* qui accroissent la surface disponible pour des connexions avec les axones d'autres neurones. *Les synapses,* entre leur axone et d'autres cellules, permettant ainsi une communication cellulaire directe. Un neurone est présenté à la figure 35. [HOC12].

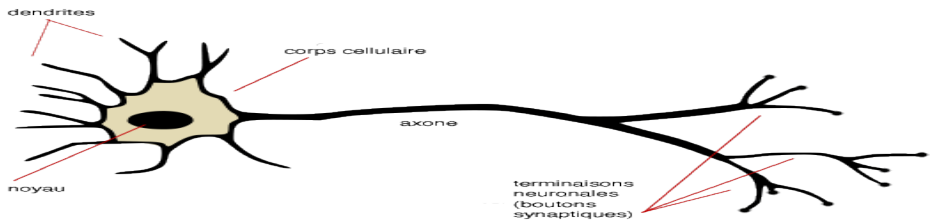


Figure 35: les composants du neurone.

V.2.3. L'activité électrochimique cérébrale :

Les neurones peuvent communiquer entre eux, cette communication est rendre possible grâce au potentiel d'action (PA) ou l'influx nerveux. Le PA est un événement court durant lequel les neurones sont chargés électriquement grâce aux ions (Sodium, Na^+), (Potassium, K^+), (Chlore, Cl^-) qui sont à l'extérieur et à l'intérieur des cellules. Le PA se présente sous la forme d'une onde où l'on peut distinguer 04 phases [LEH07] :

Lorsque le neurone est excité, il se dépolarise, c'est-à-dire on assiste à une modification qui va aboutir à une inversion de la polarité des membranes. Autrement dit la membrane de la cellule concernée devient positive à l'intérieur et négative à l'extérieur, c'est la phase de la *dépolarisation*. Cette phase est très rapide, le plus souvent inférieure à la milliseconde (msec), une amplitude maximale (pic ou spike), voisine de 110 µv (microvolts).

La phase de la *repolarisation* : c'est la phase de descente de la PA, cette phase est également très rapide (1 à 2 msec), le potentiel de membrane revenant alors vers son niveau initial.

Puis, à la fin de la phase de descente, le potentiel de membrane atteint une valeur plus négative que le niveau de son potentiel de repos (le niveau initial), c'est la phase *d'hyperpolarisation*.

La dernière phase c'est le retour à la valeur de potentiel initial, ce retour ce fait relativement plus lentement (quelques msec).

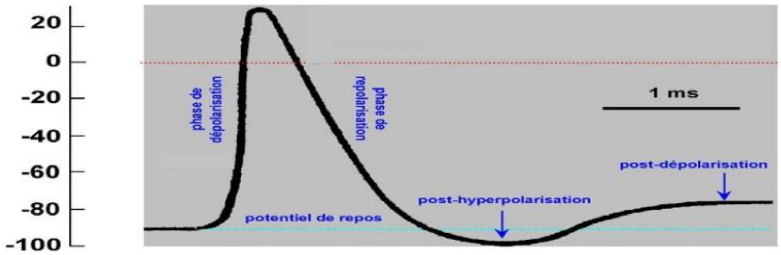


Figure 36: les étapes du potentiel d'action (PA).

Le potentiel d'action est donc une variation brève et rapide de la *charge électrique* des neurones. Il sert à véhiculer une information de neurone en neurone. L'activité électrique du cortex cérébral a une influence sur les enregistrements EEG. Dans la section suivante, nous présentons un aperçu sur l'outil de diagnostic EEG et ses signaux.

V.3. ÉlectroEncéphaloGraphie (EEG):

EEG est une mesure des potentiels qui reflètent l'activité électrique du cerveau avec grande précision, milliseconde par milliseconde. Cette mesure est réalisée à l'aide des électrodes (petites pièces métalliques) placés sur le cuir chevelu. Ces mesures sont représentées sous la forme d'une série de lignes d'ondes disposées les unes au dessus des autres appelées *élecroencéphalogramme*. Chacune d'elles correspond à l'activité électrique recueille au point de départ de deux électrodes disposées sur le scalp (ou cuir chevelu) et connectées au canal d'amplification.

L'EEG est largement utilisé par les physiologistes et les scientifiques pour étudier les fonctions cérébrales et diagnostiquer les troubles neurologiques. L'étude de l'activité électrique du cerveau à travers les enregistrements EEG est l'un des outils les plus importants pour les diagnostics des maladies neurologiques telles que, l'épilepsie, tumeur cérébral, un traumatisme crânien, trouble du sommeil, de la démence, …etc. il peut être recommandé pour le traitement des maladies des troubles du comportement (par exemple, l'autisme), trouble de l'attention, des problèmes d'apprentissage, retard de langage, …etc. Dans cette section, nous avons présenté l'historique de l'EEG, les méthodes de placement des électrodes, les rythmes du signal EEG.

V.3.1. Historique de l'EEG:

L'histoire de l'électroencéphalographie a commencé avec le physiologiste Britannique Richard Caton en 1875, qui met le premier en évidence la présence d'une activité au niveau du cerceau chez le singe et le lapin. En 1902, le neurologue Allemand Hans Berger confirme expérimentalement ces données, mais ne les applique chez l'homme qu'en 1924, il fut le premier à enregistrer le signal d'activité cérébrale en 1929, il fut aussi le premier à décrire les tracés en forme de vague (qu'il appelé « ondes alpha et ondes bêta » sous forme de variations permanentes de potentiels enregistrées avec les électrodes imposables sur la lacune crânienne ou à la surface du crâne intact.

En 1932, Disch obtenait, par calcul manuel, la transformée de Fourier d'un EEG numérisé. L'inscription à jet d'encre, introduite par Grass en 1935, permit de visualiser les activités électriques sur papier. Les bases de l'examen furent posées dès 1945 et sont toujours appliquées aujourd'hui. A partir de 1950 et avec l'avènement des micro-ordinateurs, l'enregistrement papier est remplacé par l'enregistrement numérique [TOU10].

V.3.2. Montage :

Pendant le test de l'électroencéphalographie, un certain nombre de petits disques métalliques appelés « électrodes » sont placées à des endroits différents sur la surface du cuir chevelu. Ensuite, chaque électrode est reliée à un amplificateur et un appareil d'enregistrement d'EEG. Enfin, les signaux électriques sont convertis en lignes d'ondes sur un écran d'ordinateur pour enregistrer les résultats. L'image 37 illustre l'emplacement de l'EEG sur le cuir chevelu.

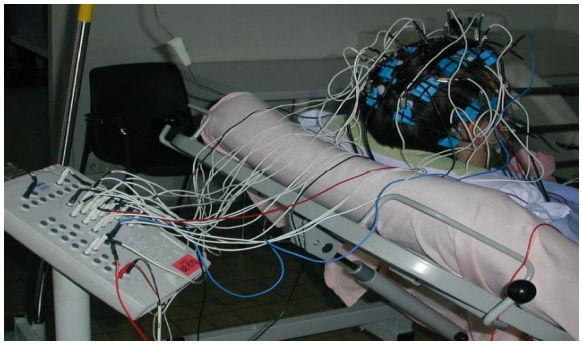


Figure 37: Montage de l'Electroencéphalographie.

Il existe deux types de l'EEG, selon l'endroit où les électrodes sont placées dans la tête, sur le cuir chevelu ou à la profondeur de tissu cérébral. Dans le premier, appelé EEG de surface, petites électrodes sont placées sur le cuir chevelu avec un bon contact mécanique et électrique. L'EEG de surface est examen indolore et non invasif (ne nécessite aucune opération chirurgicale). Pour le deuxième, appelé EEG intracrânienne, nécessite une opération chirurgicale pour mettre en place les électrodes qui sont appelées dans ce cas l'électrocorticographie (ECoG). Le chirurgien ouvrira la boite crânienne et placera les électrodes sur la surface du cerveau dans l'espace sous-dural, sur la première touche protectrice du cerveau.

Les EEG intracrâniennes donnent une lecture plus exacte que l'EEG de surface, mais c'est aussi une procédure complexe et invasive en comparaison à cette dernière.

L'amplitude d'un signal d'EEG va typiquement d'environ 1 à 100 µv à un adulte normal, et il est d'environ 10 à 20 µv lorsqu'il est mesuré avec des électrodes sous durales. Etant donné que l'architecture du cerveau est non uniforme, donc l'EEG peut varier en fonction de l'emplacement des électrodes d'enregistrement.

La question de comment placer les électrodes est important, par ce que les différents lobes du cortex cérébral sont responsables du traitement de différents types d'activités. La méthode standard pour la localisation des électrodes sur le cuir chevelu est le système international 10-20 [JAS58]. La ligne de départ de ce système est celle qui réunit le nasion et l'inion en passant par le vertex. Cette ligne est divisée en six parties, 10% de la longueur sont au dessus du nasion pour former le plan frontal et 10% au dessus de l'inion pour le plan occipital. Le reste st divisé en quatre parties égales représentant chacune 20% de la longueur totale. La figure 38 présente la position d'électrodes sur le cerveau selon le système international 10-20 [MAC10].

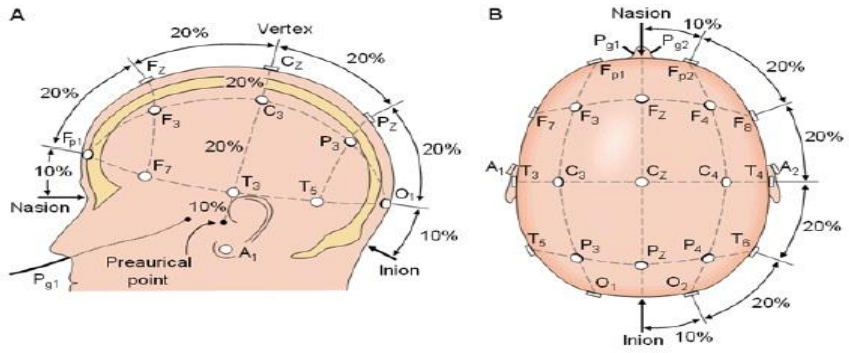


Figure 38: Le système international de montage d'électrodes 10-20

Chaque emplacement utilise une lettre pour identifier le lobe et un numéro pour identifier l'emplacement de l'hémisphère. Les lettres F, T, C, P et O représentent Frontal, Temporal, Central, Pariétal et Occipital, respectivement. Un z se réfère à une électrode placé sur la ligne médiane. Les chiffres pairs représentent les électrodes de l'hémisphère droit tandis que, les nombres impairs se réfèrent à ceux de l'hémisphère gauche. Un signal d'EEG représente une différence entre les tensions de deux électrodes.

Plusieurs montages sont effectués. Dans ce qui suit, nous citons les deux méthodes de montages les plus utilisées par les neurologues : le montage référentiel, le montage bipolaire. Il peut exister un troisième type de montage (moins utilisé), le montage laplacien.

V.3.2.1. Montage bipolaire :

Une paire d'électrodes représente un canal, chaque canal représente la différence entre deux électrodes adjacentes. L'ensemble de montage se compose d'une série de ces canaux. Par exemple, le canal (Fp1-F3) représente la différence de tension entre l'électrode Fp1 et F3. Le canal suivant dans le montage (F3-C3) représente la différence de tension entre C3 et F3, et ainsi de suite. Il ya principalement deux types de montage bipolaire : le montage bipolaire longitudinal qui explore d'avant en arrière et le montage bipolaire transverse de droite à gauche.

V.3.2.2. Montage référentiel :

Chaque canal représente la différence de potentiel entre une certaine électrode active et une électrode de référence. Il n'y a aucune position standard pour cette référence. Le problème du choix de l'électrode de référence est encore un problème d'actualité. Elle peut se située au niveau de l'oreille (A1 et A2 sur la figure 38), mais elle peut également placée en Cz ou être une moyenne calculée de toutes les électrodes sur des sites actifs.

V.3.2.3. Montage laplacien :

Le montage laplacien consiste à comparer chaque électrode à une moyenne de ses voisins directs. Chaque type de montage possède ses avantages et ses inconvénients. Certains appareillages EEG peuvent combiner ces différents montages.

V.3.3. Fréquences du signal EEG :

La fréquence est l'une des caractéristiques les plus importantes pour évaluer les anomalies dans l'EEG clinique et pour comprendre les comportements fonctionnels du cerveau humain. Un rythme cérébral désigne une oscillation électromagnétique résultant de l'activité électrique d'un grand nombre de neurones du cerveau, telle qu'on peut observer dans l'électroencéphalographie. Les caractéristiques des ondes cérébrales enregistrées à l'EEG dépendent de l'état de personne de conscience : d'éveil, sommeil léger, sommeil lent, coma,…, de l'état de vigilance : concentré, distrait,…, de l'état émotionnel : joyeux, triste, neutre, … et plus généralement de tout les états pouvant influencer à l'activité cérébrale : la stimulation, heure de la journée d'examen, …etc. il est possible d'observer dans l'EEG des modifications anormales résultants de certains états psychologiques et de certains pathologies neurologiques comme la dépression, tumeur cérébral, épilepsie,…etc. [SIU12]

Les activités cérébrales rythmiques sont classées selon leur fréquence en cinq groupes. Dans la section suivante, nous discutons ces fréquences dans des situations où la personne est dans un état de vigilance, de sommeil, soufrant d'un trouble du cerveau, …etc.

La figure39 [LOT10] illustre des exemples de ces fréquences de l'électroencéphalogramme.

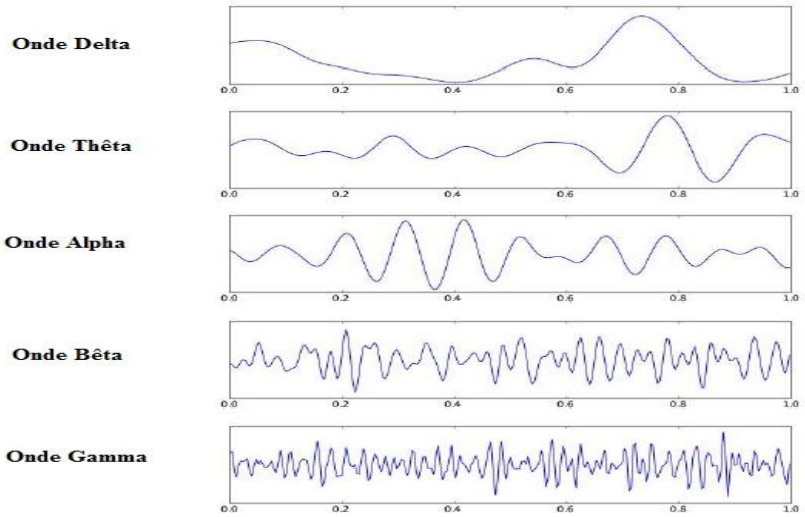


Figure 39: les différentes bandes de fréquences de l'électroencéphalogramme

V.3.3.1. Delta (δ) :

Les ondes Delta ont une basse fréquence de 01 jusqu'à 04Hz, avec une amplitude jusqu'à 200 µv. Il s'agit d'un rythme résulte des périodes de sommeil profond. L'activité Delta diminue au fur et à mesure du vieillissement. Certaines pathologies telles que démences, schizophrénie, épilepsie et tumeurs cérébrales sont liées à une activité *delta anormale*, de même que des pathologies du sommeil comme le somnambulisme et la somniloquie.

V.3.3.2. Thêta (θ) :

Les ondes thêta se trouvent dans la bande de fréquences de 4 à 8HZ, avec une amplitude habituellement plus grande que 5 µv. Ces ondes résultent de l'effort émotif, particulièrement l'anéantissement ou de la déception. Les ondes thêta se rencontrent également dans les états de veille, somnolence, sommeil léger, …etc.

V.3.2.3. Alpha (α) :

Ces fréquences comprises entre 8 à15HZ, avec amplitude entre 20 à 60µv. les ondes alpha se retrouvent particulièrement dans les régions postérieures de la tête (lobe occipital), et

sont émises lorsque le sujet a les yeux fermés. Elles caractérisent un état de conscience apaisé, activité mentale, stress, …etc. les ondes alpha disparaissent complètement pendant le sommeil.

V.3.2.4. Bêta (β) :

Les bandes bêta ont une fréquence de 15 à 30HZ. Le rythme bêta présent habituellement une amplitude faible 5 à 15µV. l'origine de ces fréquences se trouve dans le lobe frontal et pariétal. Les ondes bêta apparaissent en période d'éveil actif.

V.3.2.5. Gamma (γ) :

Les ondes gamma ont la fréquence de 30HZ à 120HZ. Ces ondes se retrouvent dans des états qualifiés de conscience active, par exemple la concentration sur un problème.

La table 5.1 [OUA12], récapitule les fréquences et l'intervalle de l'amplitude de chaque rythme de l'électroencéphalogramme.

Rythme EEG	Fréquence (Hz)	Amplitude (µv)
δ	0-4	[20-200]
θ	4-8	[5-100]
α	8-15	[20-60]
β	15-30	[5-15]
γ	30-120	[2-30]

Tableau 4:caractéristique des bandes de fréquence de l'EEG

La section suivante donne un aperçu sur certaines pathologies cérébrales les plus connues et les plus fréquentes telles que l'épilepsie, Al-Zheimer et la maladie du Parkinson. Mais, dans notre travail, nous intéressons sur l'épilepsie, les crises épileptiques et les signaux EEG épileptiques.

V.4. Les maladies neurologiques :

Les maladies neurologiques sont des maladies qui affectent les cellules nerveuses. Nous décrivons dans le suivant quelques maladies neurologiques les plus connues.

V.4.1. Maladie d'Alzheimer :

La maladie d'Alzheimer est une maladie dégénérative qui engendre un déclin progressif des facultés cognitives et de la mémoire. Peu à peu une destruction des cellules nerveuses se produit dans des régions du cerveau liées à la mémoire et au langage. Avec le temps, la personne atteinte à de plus en plus de difficulté à mémoriser les événements à reconnaitre les objets et les visages, à se rappeler la signification des mots et à exercer son jugement.

V.4.2. Maladie de Parkinson :

C'est une pathologie dégénérative dont le trait fondamental est la combinaison de la perte de neurones de la partie compacte de la substance noire de mésencéphale et la présence d'inclusions intracellulaires dans les neurones suivants. [TOU10]

V.4.3. L'épilepsie, les crises épileptiques et les signaux EEG épileptiques :

L'épilepsie est la conséquence d'une modification brutale de l'électrogènes du cerveau. C'est une crise qui résulte de la décharge électrique excessive d'un groupe plus ou moins vaste de neurones corticaux. Les neurones adoptent alors un fonctionnement anormal et incontrôlable. L'épilepsie est la deuxième cause d'hospitalisation en neurologie après les accidents vasculaire cérébraux. On estime aujourd'hui environ 50 million de la population mondiale de tout âge sont atteintes par une épilepsie [RAJ15], et en Algérie environ 400000 (1%) souffrent de crises épileptiques [APS14].

Jusqu'à présent, le principal test d'évaluation pré-chirurgicale est l'EEG. L'activité cérébrale du patient est enregistrée pendant des heures, des jours et parfois des semaines.

Au cours de la lecture d'un enregistrement EEG, le neurologue cherche à isoler le plus grand de points épileptiques pour pouvoir les analyser. En effet, plus le nombre de points épileptiques analysés est important, plus l'analyse sera robuste. Le neurologue doit donc inspecter le signal attentivement pour détecter ces événements. Ce travail est fastidieux et coûteux en temps, c'est dans ce contexte les scientifiques ont pensé au développement d'outil de détection automatique des points épileptiques.

Les crises épileptiques soit sont généralisées, soit partielles. Lorsque la décharge électrique excessive apparait dans une partie limitée du cortex, la crise est partielle. Alors que,

dans les crises généralisées, la décharge électrique excessive se produit dans les deux hémisphères cérébraux.

L'EEG du patient épileptique revêt trois états différents. L'état de crise est appelé *phase critique*, la phase critique est suivie par une phase *post-critique* correspond au retour du sujet vers un état normal qui peut durer de quelques minutes à plusieurs heures. La période qui sépare l'occurrence de deux crises est appelée la phase *inter-critique*, cette dernière est caractérisée par les signaux EEG transitoires suivants [HAD14][MAC10]:

- ✓ Pointe : est une onde très aiguë de décharge dont la durée varie entre 20 et 70 ms.
- ✓ Pointe lente : pic de décharge dont la durée varie entre 70 et 200 ms.
- ✓ Pointe onde : est une pointe suivie d'une onde lente de même polarité et de durée comprise entre 300 et 400 ms.
- ✓ Poly-pointe onde : plusieurs pointes qui se succèdent suivi d'une onde lente. Ces pointes onde peuvent aussi apparaître sous forme de décharges rythmiques de plusieurs secondes (5 à 30 sec).

Les crises épileptiques présentent des modèles de signaux EEG assez différents selon les types de crises en présence. On peut distinguer fréquemment un des éléments suivants [HAD14]:

- ✓ Synchronisation des signaux EEG qui se mettent à osciller avec des fortes amplitudes.
- ✓ Une accentuation des basses fréquences (thêta) aux alentours de 5Hz.
- ✓ Une accentuation de hautes fréquences aux alentours de 10 Hz.

V.5. Techniques numériques de traitement du signal :

Cette section a pour but de présenter un ensemble non exhaustif de méthodes de traitement du signal qui nous pouvons être interrogées dans l'élaboration du notre étude de classification des signaux EEG pour détecter les crises épileptiques. Dans ce contexte, il existe trois types d'analyses principalement utilisées.

- Analyse temporelle (d'amplitude) des signaux.
- Analyse spectrale par calcule fréquentielles.
- Analyse temps-fréquence.

V.5.1. Analyse temporelle :

L'analyse dans le domaine temporel et l'une des premières analyses qui ont été opérées sur le signal EEG. Notons d'abord, l'utilisation des lois statistiques (loi normal) [MUR13] pour mesurer aléatoire les variations des signaux, l'utilisation les caractéristiques de statistiques descriptive comme, minimum, maximum, moyenne, écart-type, médiane, mode, …[SIU12], les caractéristiques de corrélation avec les signaux de référence [CAP06]. Les caractéristiques de forme telles que, coefficient d'aplatissement, coefficient d'asymétrie [ROM10].

V.5.1. Analyse fréquentielle spectrale (Transformée de Fourier):

Les signaux EEG sont représentés dans le domaine temporel, cette représentation du signal est de type temps-amplitude. Elle n'est pas toujours la meilleure pour toutes les applications de traitement du signal. Dans beaucoup de cas, l'information la plus pertinente est cachée dans la composante fréquentielle du signal.

En 1822, Fourier a montré qu'un signal périodique pouvait être décomposé en une somme de sinusoïdes de fonctions exponentielles périodiques. Donc, la transformée de Fourier est une opération qui transforme un signal de domaine temporel vers le domaine fréquentiel. Sa définition mathématique pour un signal x est le suivant :

$$X(f) = \int_{-\infty}^{+\infty} x(t).e^{j2\pi ft}\,dt. \tag{5.1}$$

On définit de même la transformée de Fourier inverse :

$$x(t) = \int_{-\infty}^{+\infty} X(f).e^{-j2\pi ft}\,dt. \tag{5.2}$$

Plusieurs variantes d'analyses de Fourier existent dans la littérature, la plus utilisée est la Transformée de Fourier Discrète (TFD) qui consiste à transformer un bloc de N échantillons régulièrement répartis dans le temps provenant d'un signal, et qui donne N points régulièrement espacés dans le domaine des fréquences. Où, chacun des points fréquentiels de rang *k* élaborés par TFD, voici la relation :

$$x(k) = \sum_{n=0}^{N-1} X(n).e^{j\frac{2\pi kn}{N}} \quad . \tag{5.3}$$

En 1956, les deux mathématiciens américains Cooley et Tukey ont proposé une autre variante de transformée de Fourier discrète, c'est l'algorithme de Transformée de Fourier Rapide (TFR) ou, en anglais Fast Fourier Transformer (FFT). Cet algorithme permet de réduire le temps de calcule de TFD dont le nombre d'échantillons N décomposable selon des puissances 2, 4, ...etc. FFT est tout simplement un moyen rapide pour calculer la TFD qui réduit le nombre d'opérations nécessaire (N points) de $2N^2$ à 2NlogN.

Cet algorithme est couramment utilisé en traitement numérique du signal EEG pour transformer les données discrètes du domaine temporel dans le domaine fréquentiel, on peut citer les travaux [MUR13] pour la reconnaissance des émotions humains et dans [THI15] pour diagnostiquer l'épilepsie.

Comme s'est indiqué au paravent, la Transformée de Fourier Discrète permet d'obtenir ce spectre seulement. La transformée de Fourier n'est pas appropriée si le signal a une fréquence variable dans le temps (cas d'un signal non-stationnaire). Dans ce cadre, on peut dire que presque tous les signaux biologiques sont non-stationnaires, en particulier l'EEG. C'est-à-dire, l'analyse de Fourier permet de connaître les différentes fréquences excitées dans un signal, mais ne permet pas de savoir à quels instants ces fréquences ont été émises.

L'analyse de Fourier ne permet pas l'étude de signaux dont la fréquence varie dans le temps, elle n'est pas l'outil adéquat pour les signaux non-stationnaires, à une exception, elle peut être utilisée pour les signaux non-stationnaires si on ne s'intéresse qu'aux composantes spectrales qui existent dans le signal et non-instants où elles apparaissent. De tels signaux nécessitent la mise en place d'une analyse temps fréquence qui permettra une localisation des périodicités dans le temps. [TOU10]

Quand nous avons besoin de localiser dans le temps des composantes spectrales, on peut utiliser une parmi les méthodes suivantes [DUM01]:

- La transformée de Fourier fenêtrée (Short time Fourier Transform (STFT)) ;
- La transformée en ondelettes (Wavelet Transform (WT))

V.5.3. Analyse temps-fréquence :

V.5.3.1. Transformée de Fourier sur fenêtre glissante :

La transformée de Fourier ne peut seulement s'appliquer à l'analyse de signaux stationnaires. Si le signal à analyser est non stationnaire, l'idée de STFT est de partager ce signal en fonctions supposées stationnaires. Pour chaque fraction temporelle, une transformée de Fourier (FFT) est appliquée, le signal est décomposée au moyen d'une fenêtre « *g* » ou l'indice représente le positionnement temporel de cette fenêtre et donc le positionnement du spectre correspondant. La formule utilisée pour calculer la STFT est [DUM01]:

$$X(\tau, f) = \int_{-\infty}^{+\infty} x(t).g(t-\tau).e^{-j2\pi f(t-\tau)}. \tag{5.4}$$

Où, τ représente le paramètre de la fenêtre *g*. $X(\tau, f)$ correspond au spectre du signal *x* autour de t. *f* est le paramètre de fréquence.

L'inconvénient majeur de cette technique est que la taille de la fenêtre d'analyse fixe ne correspond pas nécessairement à la nature variable des signaux. [TOU13]

Egalement, on peut citer des travaux de classification des signaux EEG qui utilisent la STFT [SAM15][KOV14].

V.5.3.2. Transformée en ondelettes :

La transformée en ondelettes (WT) a été largement utilisée dans l'étude des signaux EEG [ORH11][VAV10][BAJ14]. WT a débuté au milieu du siècle dernier, toutefois, le concept d'ondelette ne fut introduit véritablement que durant les années 1980, par les travaux notamment de Morlet, Mayer, Daubechies et Haar. L'avantage principal de WT est de permettre d'analyser l'évolution du contenu fréquentiel du signal dans le temps.

L'analyse est réalisée au moyen d'une fonction Ψ appelée ondelette de base (ou ondelette mère) qui permette de spécifier les caractéristiques du signal que l'on souhaite à traiter. L'analyse en ondelette consiste à positionner, dans le domaine temporel, l'ondelette mère en regard de la partie du signal à traiter. L'ondelette mère ensuite délattée ou contractée, par l'utilisation du facteur d'échelle «*a*», permettant de concentrer l'analyse sur une gamme de données d'oscillations. Quand l'ondelette est délattée, l'analyse explore

les composantes du signal qui oscillent plus lentement, quand elle est contractée, l'analyse explore les oscillations rapides.

Par ce traitement d'échelle (contraction-dilatation d'une ondelette), la transformée en ondelette amène à une décomposition temporelle du signal. La transformée en ondelette d'un signal « *s* » peut s'écrire par la formule suivante [DUM01]:

$$C(a,b) = |a|^{-\frac{1}{2}} \int s(t) \Psi\left(\frac{t-b}{a}\right) dt. \tag{5.5}$$

Dans cette expression le terme *« a »* facteur d'échelle. C'est un terme qui représente l'inverse de la fréquence du signal. Le terme *« b »* est le coefficient de translation, il introduit le décalage temporel (le paramètre de localisation temporelle). $\Psi(t)$ représente l'ondelette mère. Il existe plusieurs ondelettes mères telles que les ondelettes de Daubechies, Coiflets, Haar, …etc.

Le coefficient d'ondelette *C(a,b)* dépend de la forme celui-ci au voisinage du temps *b.* quand *s(t)* varie peu dans le temps, son produit par l'ondelette Ψ engendre une petite aire. Quand au contraire, l'aire du produit entre le signal est l'ondelette est importante.

V.6. Conclusion

Dans ce qui précède, on déclare que les données EEG sont des données non-stationnaires, c'est-à-dire, les signaux EEG ne suivent pas une sinusoïde régulier. Pour cela, un système d'apprentissage automatique soit capable de s'adapter en prenant en compte ces évolutions. Dans ce mémoire, on souhaite extraire et modéliser de façon adaptative les connaissances afin de suivre les évolutions des signaux EEG, en particulier la détection des signaux EEG épileptiques, en utilisant les techniques d'aide multicritère à la décision (MCDA), les techniques d'apprentissage automatique et les technique avancées de traitement du signal.

Chapitre 06 : SYSTÈME DE FUSION DE CLASSIFIEURS POUR LA CLASSIFICATION DE DONNÉES ÉVOLUTIVES

VI.1. Introduction :

Au cours des premiers chapitres nous avons réalisé une étude détaillée d'état de l'art visant les domaines : d'aide multicritères à la décision (MCDA : MultiCriteria Decision Aiding), classification dynamique, les classes engendrées à cause des données non stationnaires et la fusion de données, l'outil de diagnostic médical l'ElectroEncéphaloGraphie (EEG). Elle nous a permis d'identifier les principaux algorithmes et méthodes de classification dynamique floues présentes dans la littérature ainsi les principaux algorithmes d'apprentissage incrémental.

A travers ce domaine de recherche nous souhaitons adapter ces méthodes de classification avec des données évolutives et multimodales (multicritères). L'approche que nous proposons dans ce chapitre s'applique bien à des données stationnaires et non-stationnaires qui varient en fonction de temps (*évolutives*), et provenons de sources homogènes ou hétérogènes.

Cette approche est dynamique et adaptative à l'évolution des classes et des données. Elle vise à optimiser les données aux classes, et à renforcer le mécanisme de la classification incrémentale et la détermination les phénomènes des classes évolutives (apparition brusque d'une nouvelle classe, fusion de deux classes homogènes, scission de classes, supprission d'une classe parasite,…etc).

La complexité de notre travail se situe au niveau de suivi du nombre et de la dimension des classes, ainsi que l'incrémentation du modèle de classification dès nouvelles données arrivent.

VI.2. Problématique :

De nouveaux domaines d'application émergent où les données ne sont plus persistantes mais "passagères" : gestion de réseaux, modélisation de profils web, expansion du cancer, ... Ces données arrivent rapidement, massivement, et ne sont visibles qu'une fois. Il est donc nécessaire de les apprendre au fur et à mesure de leurs arrivées. Les techniques d'apprentissage ont montré leurs capacités à traiter des volumétries importantes de données et ce sur des problèmes réels. Néanmoins le plus gros effort a été réalisé pour des analyses de données homogènes et stationnaires. La plupart des approches d'apprentissage statistique utilisent des bases d'apprentissage de taille finie et produisent des modèles statiques. De nouveaux domaines d'application de la fouille de données émergentes où les données ne sont plus sous la forme de tableaux de données persistants mais plutôt sous la forme de données "passagères", sous la forme de flux. Parmi ces domaines on citera : la gestion de réseaux de télécommunications, la modélisation des utilisateurs au sein d'un réseau social, le web mining, les systèmes d'aide au diagnostic médical, …etc.

Une fois qu'il y aura une disponibilité de nouvelles données, éventuellement, une évolution de données, le problème consiste à fournir une bonne affectation de ces données dans des classes d'appartenance. La problématique traité par notre méthode, vise à chercher la meilleure classe pour chaque donnée et non pas sa classes d'appartenance exacte, et cela en prenant en considération le phénomène de leur non-stationnarité ainsi la diversité de leur types. On focalise sur l'aspect collaboratif des classifieurs en se basant sur une des théories de fusion de données pour but de pouvoir fusionner les résultats des classificateurs, et avoir une meilleure décision finale.

VI.3. Objectif du système :

La classification dynamique floue est une discipline difficile à mettre en œuvre avec une maximisation des performances, où le système doit se doté d'une capacité d'auto-adaptation capable de faire face aux changements brusques de l'environnement d'exécution (évolution des classes), ainsi la non-stationnarité des données manipulées. Pour faciliter cette tâche nous nous sommes inspiré des travaux de recherches récents [AMA06][BOU07][HAR10][BAR00][BEL00]. D'où nous avons orienté notre conception vers un système d'apprentissage coopératif (Aide MultiCritère à la Décision, apprentissage automatique et les théories de fusion de données) entre plusieurs classificateurs maitrisant

l'ensemble des mécanismes suivants : apprentissage des connaissances (poids des critères, degrés d'appartenance des objets aux classes) à partir des données multicritères, ces degrés sont exploités par les différents classifieurs pour minimiser le temps d'exécution et le nombre d'itérations de ces classifieurs.

Dans le reste du chapitre, nous allons développer les points suivants :

- apprentissage des degrés d'appartenance des alternatives (objets ou instances) aux classes par la méthode *PROAFTN*. en prend en compte les préférences partielles (les préférences sur les critères). (cette étape a comme avantage de l'élimination l'inconvénient des méthodes de classification floue, où la matrice des degrés d'appartenance est initialisé aléatoirement dans les algorithmes de classification floues classiques);
- Choisir parmi les méthodes de classification les plus abordables pour notre problème.
- Une brève comparaison entre les théories de fusion et choisir un opérateur de fusion et une règle d'attribution de chaque objet à son meilleur classe.
- Utilisation de la technique de combinaison hybride (séquentielle et parallèle) de classifieurs.
- à cause de l'évolution de données, on construire une architecture pour suivi et détecter les évolutions des classes.
- Formalisation de l'incrémentation de notre architecture si des nouvelles données sont arrivées.

VI.4. Approche proposée :

VI.4.1. Acquisition des données :

Notre application traite le domaine de diagnostic médical par l'électroencéphalographie. Ce dernier est un appareil utilisé dans l'enregistrement des signaux, enregistre directement l'activité électrique du cerveau, grâce à son excellente résolution temporelle, de l'ordre de milliseconde, il permet de suivre la chronologie des opérations mentales et d'étudier la dynamique des phénomènes cérébraux, par exemple dans des cas pathologiques comme l'épilepsie. Une description détaillée sur la base de données utilisée dans ce travail est présentée dans le dernier chapitre.

VI.4.2. Extraction des caractéristiques :

Afin d'obtenir les meilleures performances possibles, il est nécessaire de transformer le signal traité aux valeurs qui décrivent certaines propriétés pertinentes des signaux. Ces valeurs sont connues *« caractéristiques »*. Ainsi, l'extraction de caractéristiques peut être définie comme une opération qui transforme un signal en un vecteur de caractéristiques. Le vecteur de caractéristiques, qui est composé de l'ensemble des valeurs utilisées pour décrire un signal. Pour cela, plusieurs caractéristiques utilisées pour représenter un signal comme les caractéristiques fréquentielles (l'énergie), les caractéristiques temporelle ou d'amplitude (minimum, maximum, l'écart-type,…etc), les caractéristiques de forme (coefficient d'aplatissement et le coefficient d'asymétrie) et les caractéristiques de similarité avec les signaux de référence (covariance). Vous trouverez une représentation plus détaillée sur ces caractéristiques dans le chapitre 07.

VI.4.3. Etape de classification des données :

VI.4.3.1. Choix de l'algorithme multicritère :

Les méthodes de classification développées dans [BEL00] ont pour objet de résoudre les problèmes d'affectation multidimensionnelle et multimodale. De ce fait, les actions seront représentées par une famille cohérente de critères ou un ensemble d'attributs ; $F = \{g_j / j = 1..n\}$ avec $j \geq 3$, et les catégories sont modélisées par les actions de référence qui représente les types d'actions liées à chaque catégorie. Ces actions sont connues sous le nom d'*actions de référence centrales* ou *prototypes.* Chaque catégorie $C^h, h = 1,..,k,$ est représentée par L_h actions de référence centrales, forment une famille B^h d'actions de référence centrales tel que : $B^h = \{b_1^h, b_2^h, ..., b_{Lh}^h\}$ avec $L_h \geq 1$ et $h = 1,..,k.$

Chaque action de référence centrale est définie par ses performances qui sont évaluées sur une famille de critères *F*. ces performances définissent un profil de référence qui est représenté par un vecteur de valeurs $g_j(b_i^h)$:

$\forall h \in \{1,...,k\}$ et $\forall i \in \{1,...,L_h\}$ $g(b_i^h) = (g_1(b_i^h), g_2(b_i^h), ..., g_n(b_i^h))$. Où $g_j(b_i^h)$ indique l'évaluation de l'action de référence *i* de la catégorie C^h selon le critère g_j. En général, les performances $g_j(b_i^h)$ sont définies à l'aide d'intervalles avec ou sans fonction d'appartenance.

En pratique, les performances des actions de référence centrales sont généralement données sous forme d'intervalles. Par exemple dans le diagnostic médical, les critères de classification sont donnés sous forme d'intervalles et non pas sous forme de valeurs précise. Comme en diagnostic par l'EEG, un des critères des actions de référence centrales peuvent être les fréquences des ondes, où il existe cinq ondes principaux dans ce contexte, la bande Delta δ (de 0,1 à 4Hz), la bande Thêta θ (de 4 à 8Hz), les ondes Alpha α (de 8 à 12Hz), les bandes Bêta β (de 12 à 30Hz) et les ondes Gamma γ (de 30 à 120Hz). Ces fréquences sont rencontrées dans tous les sujets d'EEG, mais avec formes différents, par exemple dans les signaux EEG épileptiques on trouve une accentuation des fréquences thêta. Egalement, l'amplitude de ces fréquences sont donnés sous forme d'intervalles non ordinales, voire tableau 8 pour plus de détaille.

Ainsi pour chaque critère g_j on associe à chaque action de référence centrale b_i^h l'intervalle $\left[S_j^1(b_i^h), S_j^2(b_i^h)\right]$ avec : $S_j^1(b_i^h) \leq S_j^2(b_i^h), j = 1 \dots n, h = 1, \dots, k \; et \; i = 1 \dots L_h$.

Les trois méthodes développées dans [BEL00] traitent les problèmes de classification multicritère qui consistent à affecter les objets à des catégories représentées par des actions de référence centrales (au contraire de la méthode de classification multicritère ELECTRE Tri, où les actions de référence représentées par les frontières inférieurs et supérieurs des classes). Ces algorithmes sont illustré en détaille dans L'annexe A. il est indispensable de citer que ces algorithmes se basent sur les notions des algorithmes de la famille ELECTRE Tri et PROMETHEE illustrés dans le premier chapitre.

VI.4.3.1.1. Quel type de méthode ?

Notre travaille opté sur la classification multicritère, ce domaine inclus dans le champ de classification semi-supervisé, où quelques paramètres sont connus *a priori* comme les actions de référence centrales et les paramètres de préférences partiels (les préférences sur les critères). Le résultat des méthodes de tri multicritères est la matrice des degrés d'appartenance des actions aux différentes classes. Cela constitue un premier avantage de notre méthode, puisque nous aurons considérer dans la suite cette matrice comme entrée de classifieurs non-supervisés choisis dans l'étape suivante.

Les méthodes de classification multicritère font partie au domaine d'aide multicritère à la décision. Elles consistent à élaborer des procédures d'affectation qui permettent d'affecter chaque individu à une classe prédéfinie et ceci à travers l'examen de sa valeur intrinsèque en

se référant à des normes préétablies. *Ces méthodes utilisent uniquement des comparaisons entre l'individu à affecter et les individus de référence par le biais d'un modèle de préférence. Ceci évite de recours à des distances et permet d'utiliser des critères quantitatifs et/ou qualitatifs.* En outre, elles permettent d'écarter les complications rencontrées lorsque les données sont exprimées dans différentes unités (la multi-modalité des données). Ces avantages constituent l'une des raisons qui nous a motivés à développer de nouvelle approche de classification dynamique utilisant le domaine d'aide multicritère à la décision.

En plus il n'existe pas de méthodes utilisant le domaine d'aide multicritère à la décision ont été appliquées dans le domaine d'aide au diagnostic par l'Electroencéphalographie, où les données sont évolués avec le temps. Ce fait à nous encouragé d'utiliser les méthodes de classification multicritère pour traiter les problèmes de classification médicale.

VI.4.3.1.2. Les méthodes de tri nominal ou ordinal ?

La problématique de classification en aide multicritère à la décision peut être soit ordinale ou nominal, la première dans le cas où les classes serait complètement ordonnées. La dernière dans le cas où il serait difficile d'établir un ordre entre les catégories. Tous les problèmes d'affectation qui seront caractérisés par des actions de référence centrales. Ainsi, si on est devant à un problème d'affectation dont les classes ont une signification ordinale et qu'on ne parvient pas à les cerner par les actions de référence limites, on considérera ce problème comme étant une problématique du tri nominal. Le problème de diagnostic par l'EEG sera traité comme une problématique de tri nominal puisqu'on ne peut pas représenter les classes par les actions de référence limites. Pour ces raisons, on a opté d'utiliser une méthode de classification multicritère floue PROAFTN (Annexe A) développée par BELACEL Nabil dans l'année 1990.

Pseudo code de l'algorithme multicritère utilisé (PROAFTN):

Entrées :

- Table des performances $A * F$ des éléments $g_j(a)$.
- poids d'importance des critères ou matrice de comparaison paire des critères.
- Ensemble de classes, $C = \{C_1, C_2, \dots, C_h, \dots, C_k\}$.
- Ensemble d'actions de référence centrales de chaque classe $B^h = \{b_1^h, b_2^h, \dots, b_i^h, \dots, b_{L_h}^h\}$

$$avec: L_h \geq 1 \; et \; h = 1 \dots k.$$

Pour chaque action à classifier faire:

- Calculer les indices de concordance partiels.
- Calculer l'indice de concordance global.
- Calculer les indices de discordance partiels.
- Calculer l'indice de discordance global.
- Définir les relations d'indifférence floues entre l'action à classifier et les actions de référence centrales.
- Calculer les degrés d'appartenances floues de l'action *a* à affecter aux différentes catégories.

Décision d'affectation.

Affecter l'action **a** selon la règle de maximum de possibilité.

VI.4.3. 2. Choix de l'algorithme en apprentissage automatique :

VI.4.3.2.1. La classification supervisées ou non supervisée ?

Dans [LAM13], l'auteur a illustré les différentes étapes pour construire un système de fusion de classifieurs. Nous devons, tout d'abord précisé si l'algorithme doit être supervisé ou non. L'emploi d'un algorithme supervisé nécessite, comme nous l'avons vu précédemment, une base d'apprentissage pour chaque classe et pour chaque patient faite par les experts. Cela constitue un inconvénient de ce type de méthodes.

A l'inverse, les méthodes non-supervisées présentent l'avantage d'être totalement indépendante de l'opérateur, de ne pas nécessiter d'apprentissage, d'être rapide et reproductibles. Pour toutes ces raisons, nous avons opté pour combiner une méthode semi-automatique (classification multicritère) et une méthode non-supervisée (machine learning).

VI.4.3.2.2. Classification floue on non floue ?

Les méthodes de classification floue utilisent une conception à base de règles beaucoup plus en langage naturel. Surtout dans le domaine de neurologie et diagnostic par EEG, qui examine différentes caractéristiques du signal EEG pour identifier correctement une crise épileptique. La règle floue d'autre part offre une conception plus simple du raisonnement approximatif qui peut imiter le raisonnement humain efficacement. Nous avons

développé notre méthode de telle manière à imiter le raisonnement des experts dans la détection des crises épileptiques à partir des données EEG. En plus, les règles floues peuvent être définies en utilisant les connaissances des experts pour prendre de décision qui plus simple à mettre en œuvre. Comme par exemple dans l'algorithme PROAFTN qui exige seulement d'identifier les signaux de référence par les experts. Dans cette étude, nous avons utilisé la méthode de classification floue multicritère combinée avec les méthodes de classification floue définies dans le cadre d'apprentissage automatique comme FCM, PCM,…

VI.4.3. 2.3. Choix les paramètres de l'algorithme :

Nous utilisons les algorithmes de classification floue pour classifier les signaux EEG. Pour cela, il nous faut définir les différentes paramètres gouvernant les algorithmes utilisés, à savoir les valeurs *m, c,* la distance, critère de convergence,….etc.

VI.4.3.2.3.1. Détermination du nombre de classes :

Nous nous plaçons ici dans une problématique de classification dynamique des signaux EEG. Notre objectif donc consiste à classifier les données et suivi les données et les classes évolutives. Dans le reste de manuscrit, le nombre de classes sera donc suivant le cas étudié. Alors le nombre de classes est varié d'un cas à un autre.

VI.4.3.2.3.2. Choix du paramètre m :

Le paramètre *m* contrôle le degré flou de la partition floue *U*. si **m** est proche de 1, la partition résultante est quasiment non floue, *i.e* chaque vecteur x_i est assigné à une et une seule classe *j* avec un degré d'appartenance u_{ij}. Alors que la croissance de **m** dans FCM tend à augmenter le degré de partage des vecteurs aux classes (les degrés d'appartenance de x_i à chacune des *c* classes sont égaux à $\frac{1}{c}$ lorsque **m** tend vers l'infini). Elle participe dans *PCM* à l'augmentation globale des appartenances de tous les vecteurs. Il n'existe pas de méthode pour optimiser de manière générale ce paramètre. Chaque problème appelle un choix dépendant de la nature des données. [BAR00]

Certains auteurs [BAR00] ont proposé d'attribuer à **m** une valeur comprise dans l'intervalle [1.5, 3]. Et autres [KHO98][CHE09][HAR10] ont proposé de prendre $m = 2$ afin d'assurer la convergence de l'algorithme.

VI.4.3.2.3.3. Choix de la distance :

Une fois l'algorithme de classification est déterminé, il faut choisir une mesure de ressemblance qui correspond à l'application recherchée. De nombreuses mesures de distance ont été proposées dans le troisième chapitre. Pour notre cas, nous avons choisi d'utiliser les algorithmes de classification floue avec la distance la plus usuelle et la plus rapide à calculer, qui est la distance euclidienne.

VI.4.4. Etape de la fusion de données :

Cette section est dédiée à la comparaison méthodologique entre les méthodes d'agrégation de données hétérogènes présentées dans le troisième chapitre, dont le but de trouver la plus adaptée à notre application de données évolutives. Nous cherchons donc, plus précisément, une méthode de fusion de données adaptée aux environnements non-stationnaires complexes qui permet de représenter l'imperfection inhérente aux données provenant de capteurs hétérogènes. Les classifieurs sont alors perçu comme différentes sources d'informations dont les imperfections sont facilement établies.

Les auteurs ont comparé entre les méthodes de fusion et montrer qu'il n'y a pas d'une méthode universelle de fusion. [MAR05][LEC05][LEL13][DUB99][BLO96].

Dans notre étude, les informations sont à la fois incertaines et imprécises. Par exemple, les fréquences dans un signal EEG sont des informations incertaines et imprécises. De plus, dans l'étape de modélisation plusieurs causes d'imprécision et d'incertitudes peuvent être existent dans le cas de traitement des signaux EEG. Citons par exemple :

Dans un système de diagnostic par EEG, les informations recueillies sont à la fois imprécises et incertaines. Cela est dû aux capteurs et différents instruments de mesure, au bruit et aux experts humains.

Un élément qui appartient à plusieurs classes avec des degrés d'appartenance dans l'intervalle [0,1], qui est une imprécision sur la localisation de l'information.

La dynamicité réelle du signal EEG, qui est une imprécision due au phénomène physique.

Le bruit qui participe à la dégradation de l'information et donc à l'incertitude sur le signal vrai mesuré.

Les paramètres numériques des algorithmes utilisés pour la classification sont des sources d'imprécision.

VI.4.4.1. Choix du cadre théorique de fusion :

VI.4.4.1.1. Les limites de la fusion probabiliste :

La théorie des probabilités, bayésienne essentiellement, a été souvent utilisée en fusion de classifieurs [KUN02-a][KUN02-b][XU92]. La fusion quant à elle est réalisée à l'aide de la règle de Bayes, et la décision est prise en fonction du maximum de vraisemblance. Cependant, cette théorie présente certains inconvénients qui limitent son utilisation dans le cas qui nous intéresse. Ces limites sont résumées ci-dessous :

Le problème majeur des probabilités est qu'elles représentent essentiellement l'incertitude et très mal l'imprécision, ceci entraîne souvent une confusion des deux notions. De plus, dans l'étape de modélisation probabiliste nous raisonnons sur des singletons qui représentent les différentes décisions. Ceci entraîne que le monde doit être fermé, ce qui ne correspond pas toujours à la réalité, en particulier les données évolutives (la base de données est enrichit au cours de temps).

De plus, le formalisme probabiliste, notamment introduit dans la règle de Bayes, requiert des connaissances *a priori* sur l'occurrence de chaque phénomène par l'intermédiaire des probabilités conditionnelles et des probabilités *a priori* des événements. L'estimation des probabilités *a priori* est beaucoup plus délicate et il est souvent nécessaire de faire appel à une connaissance plus subjective sur l'application. Des modèles peuvent représenter ces connaissances, mais imposent des hypothèses fortes sur l'étape de modélisation (densités, hypothèses d'indépendance, …etc).

VI.4.4.1.2. Comparaison entre les théories des possibilités et la théorie des croyances :

La théorie des ensembles flous représente un très bon outil pour représenter et manipuler les informations imprécises, sous la forme de fonctions d'appartenance. Ces fonctions ne sont pas soumises aux contraintes axiomatiques imposées aux probabilités et permettent donc une représentation plus souple de la connaissance. En revanche, cette souplesse peut être considérée comme un inconvénient puisqu'elle laisse facilement l'utilisateur démuni pour définir ces fonctions. L'inconvénient des ensembles flous est qu'ils représentent essentiellement le caractère imprécis des informations, l'incertitude

étant représentée de manière implicite et n'étant accessible que par déduction à partir des différentes fonctions d'appartenance.

La théorie des possibilités permet de représenter à la fois l'imprécision et l'incertitude, en utilisant des distributions de possibilités, des fonctions de possibilité et de nécessité qui caractérisent un événement. Comme la théorie des ensembles flous, les fonctions de distributions de possibilités permettent aussi une représentation plus souple de la connaissance par rapport à l'approche probabiliste. Néanmoins, la théorie des possibilités ne gère pas nativement le conflit entre sources mais elle peut être mise en œuvre par des règles adaptatives. De même, elle ne permet pas une modélisation de l'ignorance totale et l'hypothèse de monde ouvert.

La théorie des croyances détient une place importante dans le cadre des théories de l'imparfait, puisque elle apparait comme une généralisation de la théorie des probabilités et de la théorie des possibilités. En effet, la théorie de Dempster-Shafer fournit une modélisation très souple et très riche des connaissances imparfaites, en particulier de l'imprécision et de l'incertitude, mais aussi de l'inconsistance, de l'ambiguïté et de l'incomplétude, ceci en analysant les capacités de chaque source à donner une information sur chaque décision possible.

En outre, notons qu'une fonction de masse peut être vue comme une mesure de possibilité dans le cas particulier où les éléments focaux sont emboîtés. Ainsi que, les fonctions de plausibilité et de crédibilité calculées à partir d'une fonction de masse *m*, que nous définissons au quatrième chapitre. Un autre intérêt de la théorie réside dans l'utilisation de l'ensemble vide. Il permet d'envisager que l'état réel du système observé n'appartienne pas à l'ensemble des classes considérées. On dénote par monde ouvert et monde fermé respectivement les mondes acceptant et n'acceptant pas que l'ensemble vide comme solution du problème. [CAV12]

la théorie de l'évidence prend en compte les ensembles composés de plusieurs classes, ce qui permet de considérer le doute entre les classes.

Dans [LEC05], l'auteur a proposé un comparatif de ces deux théories illustré par le tableau suivant :

Etape de fusion	*Théorie des possibilités*	*Théorie de l'évidence.*
Cadre de discernement. $\Omega = \{C_1, C_2, ..., C_c\}$	x appartient à C_i	x appartient à A_k sous-ensemble de Ω où, $A_k=C_i$ ou $A_k = \bigcup C_i$.
Modélisation.	Degré d'appartenance d'un élément x à une classe C_i. $\pi_x(C_i) \in [0,1]$.	Masse m donnant la confiance qu'un élément x appartient à un sous-ensemble A_i du cadre de discernement, telle que : $\sum_{i=1}^{2^\Omega} m(A_i) = 1.$
Estimation.	Les degrés de possibilité peuvent être estimés à partir d'algorithme de classification automatique. Telle que : C-Moyennes-Possibiliste (PCM).	Utilise quelques techniques telles que :- les modèles probabiliste [BOU12][MAR05]. - Les modèles de distances [MAR05]. - Les méthodes des k-Plus-Proche Voisins classique [MAR05] et k-ppv floue [BOU07].
Fonctions utilisées	-Degré de Possibilité Π : $\Pi_j(\{x \in C_i\}) = \{\pi_j \in C_i\}$. -Degré de Nécessité N: Quantifie le degré de certitude de réalisation de l'hypothèse. $N_j(x) = \inf\{(1-\pi_j(x \in C_k)); C$	-Fonctions de masse m. -Fonctions de plausibilité Pl. $Pl(A_k) = \sum_{B \cap A_k \neq \phi} m(B)$ -Fonctions de Crédibilité Cr : $Cr(A_k) = \sum_{B \subseteq A_k} m(B)$
Relation	$\Pi_{C_i}(x) = N_{C_i}(x)$ $\Pi_{C_i}(x) = 1 - N_{C_i}(\bar{x})$	$Pl_x(A_k) \geq Cr_x(A_k)$. $Pl_x(A_k) = 1 - Cr_x(A_k)$.
Combinaison	Plusieurs opérateurs de combinaison existent, tels que : T-norms, T-conorms, intégrales de Choquet et de Sugeno,…etc	Plusieurs opérateurs de fusion existent dans cette théorie, on peut citer : -la somme orthogonale de Dempster-Shafer, la règle de Smets,…
Décision	- Max Π. - Max N.	-Max Pl. -Max Cr. -Intervalle de confiance : [$Cr(A)$, $Pl(A)$] -probabilité pignistique. -…etc.

Tableau 5: Comparaison de la théorie des possibilités et les la théorie des croyances

VI.4.4.1.3. Vers la théorie des croyances :

La théorie des croyances peut être appliquée dans un grand nombre de situations ce qui a poussé à l'introduction de fonction de croyance dans toute sorte de domaines. Par

exemple, segmentation d'images médicales [LEL13], reconnaissance des formes [DEN97], en classification [DEN06], …etc.

Nous reprenons ci-dessous les situations dans lesquelles la théorie des croyances est intéressante :

Le principe même de modélisation sur un espace d'hypothèses composées 2^{Ω} au lieu de Ω est un avantage certain de cette théorie. De plus, les fonctions de masse permettent d'obtenir une modélisation riche des imperfections des données.

La théorie des fonctions de croyance offre la possibilité d'extension le domaine de discernement par la notion du monde ouvert et monde fermé, cela correspond à notre problème de classes évolutives (ajout, suppression des classes).

La théorie de Dempster-Shafer est riche d'un point de vue d'opérateurs de fusion, où il existe l'opérateur de somme orthogonale, l'opérateur de Smets, l'opérateur de Yager,…

Le critère de prise de décision a une influence sur le résultat de la classification. L'un des avantages de la théorie de Dempster-Shafer est sa richesse en termes critères de décision. Les critères utilisés ont été Maximum de crédibilité, Maximum de plausibilité, Probabilités pignistiques, l'intervalle de confiance,…etc. Parmi ces critères, nous citons l'intervalle de confiance $[Cr(A), Pl(A)]$ plutôt qu'une grandeur unique pour représenter la connaissance de l'hypothèse A.

Notons qu'il existe un lien entre la théorie des croyances et celle des probabilités ainsi qu'avec la théorie des possibilités. En effet, nous avons vu que lorsque les éléments focaux sont des singletons, les fonctions de masse sont alors des probabilités. Dans ce cas, les fonctions de croyance et de plausibilité se retrouvent égales aux fonctions de masse.

VI.4.4.2. Les étapes de fusion de données.

Comme on le voit dans la section IV.4.2, le mécanisme de fusion de données est composé en quatre étapes successives illustrées comme suit :

VI.4.4.2.1. Modélisation :

L'étape de modélisation consiste à la représentation de l'information dans un cadre mathématique lié à une théorie particulière [ZOU07]. La modélisation dans la théorie des

croyances consiste à manipulation de *fonctions définies sur des sous-ensembles et non sur des singletons* comme dans la théorie des probabilités. Le formalisme mathématique repose tout d'abord sur la définition de masses accordée aux événements.

VI.4.4.2.2. Estimation les fonctions de masse :

Après la définition du cadre de discernement intervient la modélisation des croyances dans le cinquième chapitre. Le problème de l'affectation des masses de croyances aux différentes hypothèses est un problème majeur en fusion de données. Une fois cette étape est réalisée, la deuxième étape à réaliser consiste à affecter les fonctions de masse aux différentes hypothèses. Cette étape est un problème crucial où il n'y a pas de méthode générique. Selon le type d'application rencontrée, il existe différentes méthodes permettant d'élaborer les masses de croyances.

Des méthodes automatiques ont été proposées dans la littérature afin d'affecter les masses de croyance à partir de différents modèles tels que des modèles probabilistes, de distances et modèles basés sur l'algorithme EM (Expectation Maximization) [LEL13]. La construction du modèle dépend fortement de l'application et de la connaissance de données. Les deux modèles les plus connus et les plus utilisés sont : le modèle basé sur une approche probabiliste et modèle basé sur des informations de distance [LEL13][MAR05]. Dans le premier modèle, des masses non nulles sont affectées aux singletons, ainsi qu'à l'ensemble Ω. Dans ce modèle, les disjonctions autres que Ω ne sont pas modélisées, et l'on perd largement de l'apporte conceptuel de la théorie des fonctions de croyance. [LEL13]

Pour ces raisons, on choisi d'appliquer la méthode basée sur les informations de distances pour estimer les masses de croyance. Denoeux [DEN95][DEN06] a proposé une approche issue de l'algorithme *k-plus proche voisins*. Les fonctions de masses ici sont construites uniquement à partir de la représentation des vecteurs d'apprentissage. Mais cette méthode a souffert de plusiuers problèmes tels que, définir *a priori* le paramètre *k* de plus proches voisins et l'ensemble d'apprentissage ainsi que le temps de calcul est très important.

Soit X l'ensemble d'apprentissage composé de m points répartis en c classe. $\Omega = \{C_1, C_2, ..., C_c\}$, Ω : s'appel le domaine de discernement.

L'utilisation des deux concepts : degré d'appartenance et la distance, nous a permet d'obtenir une meilleur méthode d'estimation des fonctions de croyance.

L'algorithme de l'estimation des fonctions de masses pour chaque classifieur l est représenté comme suit :

$$m^{\Omega}\left(C_j^l, x_i\right) = \mu_{ijl}.\Phi_j\left\{d_i^l\right\}.\ \ j = 1..c; i = 1..m; l = 1..e. \tag{6.1}$$

Cette équation permet de calculer les masses de croyance des singletons C_j.

$$m^{\Omega}\left(A, x_i\right) = \min\left(m^{\Omega}\left(C_j^l, x_i\right)\right); où \quad C_j \in A \ ; A \subset 2^{\Omega} / \left\{C_j^l, \Omega\right\}; j = 1..c. \tag{6.2}$$

Ou, les masses des sous ensembles de domaine de discernement non singletons, peuvent être calculées par la règle λ *-Sugeno-measure qu'on a définie dans le premier chapitre.*

Cette deuxième équation nous a permet de calculer les masses de croyance des disjonctions A autre que le domaine de discernement.

$$m^{\Omega}\left(\Omega, x_i\right) = 1 - \left(\sum_{B \in 2^{\Omega} / \{\Omega\}} m^{\Omega}\left(B, x_i\right)\right) \tag{6.3}$$

Cette équation permet de calculer la masse de croyance de l'hypothèse Ω. Cette dernière équation a pour but de garder la propriété de la somme totale des fonctions de masse à 1.

Avec :

$\Phi_j\left\{d_i^l\right\}$: est la fonction de distance d_i entre le nouveau x et le centre de classe b_j. Selon [BOU07], il existe une infinité de fonctions de distance. Par exemple, la fonction de distance inverse est donnée par :

$$\Phi_j\left(d_j^l\right) = \frac{1}{1 - \lambda_j d_i^2} \tag{6.4}$$

Où :

λ_j : est un paramètre représentant la distribution des points de la classe C_j .

Cette approche qui permet de calculer les fonctions de masse est meilleure que l'approche proposée par Denoeux [DEN06] en temps de calcul, car cette dernière approche necessite une base d'apprentissage contient les k-plus proche voisins de l'élément à classifier. Ainsi que, l'approche basée sur les k-ppv permet de calculer les masses de croyance pour les classes singletons seulement, les disjonctions autres que le domaine de discernement ne sont

pas modélisées (l'équation 6.2). pour cela, on perd largement de l'apport conceptuel de la théorie des fonctions de croyance.

VI.4.4.2.3. Combinaison (choix de l'opérateur de fusion) :

L'étape d'agrégation est plus fondamentale pour une exploitation pertinente des informations issues des classifieurs. Plusieurs modes de combinaison ont été développés dans le cadre de la théorie des croyances. Comme on a vu, les opérateurs de combinaison dans la théorie des fonctions de croyance sont divisés en trois catégories, il y a des opérateurs disjonctifs, les opérateurs conjonctifs et les opérateurs de compromis. Tous les opérateurs de la théorie de croyance s'appuient sur l'opérateur conjonctif de Dempster-Shafer ou la somme orthogonale, il permet de combiner deux ou plusieurs fonctions de masse en une seule, mais L'inconvénient de l'opérateur de Dempster-Shafer est quelle masque le conflit existant entre les sources. Cet opérateur ne modélise pas l'hypothèse du monde ouvert ($m(\phi)=0$), il suppose que le domaine de discernement recense toutes les réponses possibles de la question considérée. Ce qui contredit avec notre problème, où les données ainsi que les classes sont évoluent au temps. La forme qui modélise l'hypothèse du monde ouvert a été développée plus tard par Smets [SME90]. Elle est donnée pour tout sous-ensemble de domaine de discernement par :

L'opérateur de Smets est donné pour tout sous-ensemble A de domaine de discernement par :

$$\begin{cases} m(A) = \sum_{B_1 \cap B_2 \cap \ldots \cap B_p = A} \prod_{k=1}^{p} m_j(B_j) \\ m(\phi) = \sum_{B_1 \cap B_2 \cap \ldots \cap B_p = \phi} \prod_{k=1}^{p} m_j(B_j) \end{cases} \tag{6.5}$$

VI.4.4.2.4. Prise de décision :

L'étape finale du processus de fusion, a pour objectif de la prise de décision quant à l'appartenance d'un élément à une classe. Cette étape dans la théorie de croyance est différemment des théories des probabilités et possibilités, l'un des avantages de la théorie de Dempster-Shafer est sa richesse en termes critères de décision. Parmi ces critères, nous citons l'intervalle de confiance [*Cr(A), Pl(A)*] plutôt qu'une grandeur unique pour représenter la

connaissance de l'hypothèse *A.* ces outils de décision sont illustrés en détaille dans la section IV.6.2.

Algorithme général de fusion :

- **Modélisation des données :**
  Pour ***i*** allant de 1 jusqu'à ***e***. (***e*** est le nombre de classifieurs flous utilisé).
  Calculer la matrice U_i des degrés d'appartenance.
  Fin pour.
- **Estimation :**
  Calculer les fonctions de masse par l'équation les formules (6.1) (6.2) (6.3) (6.4).
- **Fusion :**
  Fusion crédibiliste, en utilisant l'opérateur de Smets.
- **Décision :**
  Signal classifié (Application de la règle intervalle de confiance *[Cr(A), Pl(A)]).*

VI.4.5. Etape de détection de l'évolution de l'espace de classification et forcer le mécanisme d'incrémentation.

Le développement de méthodes d'analyse dynamique de l'information, comme le clustering incrémental et les méthodes de détection de nouveauté, devient une préoccupation centrale dans un grand nombre d'applications dont le but principal est de traiter de larges volumes d'informations variant au cours du temps. Ces applications se rapportent à des domaines très variés et hautement stratégiques, tels que l'exploration du Web et la recherche d'information [CUX11], l'analyse du comportement des utilisateurs et les systèmes de recommandation, la veille technologique et scientifique, ou encore, l'analyse de l'information génomique en bioinformatique, l'expansion du cancer, …etc.

Les classes peuvent être statiques ou dynamiques. Les classes statiques sont basées sur des données stationnaires. Ce qui signifie que les paramètres et la structure du modèle de chaque classe demeurent inchangés au cours de temps. Toutefois, comme l'environnement est en perpétuelle évolution, la majeure partie des données issues du monde réel est no-stationnaires. La non-stationnarité est définie comme une évolution dans le temps des paramètres et de la structure du modèle ou classifieur due par exemple au vieillissement de la machine ou encore aux changements temporels des caractéristiques du système. Les classes dans ce cas sont dynamiques. Les paramètres et la structure du modèle ou classifieur doivent être mis à jour afin de conserver sa performance au cours de temps.

Dans le domaine des méthodes de classification non supervisées, deux approches sont possibles (1) réaliser des classifications à différentes périodes de temps et analyser les évolutions entre les différentes phases (approche par pas de temps), (2) développer des méthodes de classification qui permettent de suivre directement les évolutions (les méthodes de classifications dynamiques incrémentales).

Dans cette section, le problème d'apprentissage des modèles ou classifieurs dans un environnement non-stationnaire est étudié. Une technique adaptée pour l'apprentissage des classifieurs efficace pour ce type d'environnement est présentée. La classification des signaux EEG et le suivi de leurs évolution au cours de temps est l'un des domaines d'application de ce type de méthodes.

La présente section traite les phénomènes importantes engendrées à cause les données non-stationnaires, où nous avons inspiré notre approche de travaux initié du [AMA06] [BOU07] [HAR10], ces travaux sont très riche sur ce domaine concernant la classification dynamique. Mais, les auteurs de ces travaux utilisent les algorithmes de classification supervisée comme les réseaux de neurones, l'algorithme des k-plus proche voisins, le modèle de mélange gaussien, les machines à vecteurs de support (SVM).

Le classifieur dynamique est élaboré avec un processus d'apprentissage lui conférant des aptitudes d'apprentissage en environnement des données non-stationnaires. Ce processus permet de construire et d'adapter le modèle de classification à travers l'intégration continue des nouvelles informations afin de prendre en compte les évolutions de classes. Pour permettre la prise en compte des phénomènes dynamiques décrits dans le deuxième chapitre, le processus d'apprentissage est décrit par les procédures suivantes :

- La procédure de création de nouvelles classes.
- La procédure de modification locale de classes.
- La procédure de fusion de classes.
- La procédure de scission de classes.
- La procédure d'élimination de bruits et suppression de classes inutiles.

VI.4.5.1. Apparition de nouvelles classes :

Lorsque des nouveaux données sont arrivées, une nouvelle classe peut être apparaître. En effet, les approches de classification des données statique ont prouvé leurs limites et

montré plusieurs lacunes où la plupart de ces approches ne propose pas d'apprendre au fur et à mesure du traitement, et se retrouvent généralement bloquées dès qu'un nouveau cas de figure se présente par exemple dans la classification des documents, peut être une classe de document jamais rencontrée. Dans ce cas un opérateur est obligé d'intervenir.

Afin de remédier à cela et pour détecter une ou des nouvelles classes inconnues précédemment nécessitent un nombre minimum de *points rejetés* et géographiquement proche les uns des autres. Cependant, il faut distinguer l'apparition d'une nouvelle classe représentative de l'apparition de classe parasite, crée à cause du bruit et de perturbations extérieures [BOU07]. Après la création de nouvelle classe, le modèle de classification sera incrémenté :

On peut modéliser les points rejetés dans notre approche, par le concept du monde ouvert dans la théorie des fonctions des croyances, où nous considérons l'ensemble vide comme un ensemble des points rejetés, dans ce cas :

Si : $m_{x_i}(\phi) > m_{x_i}(A); \forall A \in 2^{\Omega} - \{\phi\}$ alors, on effectue l'élément x_i à l'ensemble des éléments rejetés ϕ.

Si l'ensemble des éléments rejetés ϕ contient un nombre suffisant N_{min} points pour créer une nouvelle classe représentative alors, on applique l'algorithme FCM avec un critère de validité [HAR10] pour obtenir un nombre minimum de classes ajoutées.

Une fois, le classifieur dynamique détecte une nouvelle classe, il convient d'incrémenter le domaine de classification M où, $M_{t+1}=M_t+C_{new}$.

M_t: représente le modèle de classification de l'itération précédente.

VI.4.5.2. Modification locale de classe:

Dans cet exposé, nous présenterons les problématiques liées à la détection de changement et à l'adaptation de modèles. Nous présenterons différentes formulations et solutions proposées pour résoudre ces problèmes, en particulier, par des méthodes d'apprentissage incrémental.

La classe C_j qui reçoit l'action x est la seule qui doit être mise à jour. Les valeurs actuelles (matrice de covariance, centre de classe) de C_j sont alors mises à jour

incrémentalement. En appliquant l'algorithme IEM (Incremental Expictation Maximization) illustré dans [AMA06] :

L'ensemble des phases de fonctionnement de la procédure d'ajout de nouvelle classe et modification locale d'une classe sont représentées sur le schéma de la table ci-dessous :

Algorithme général d'ajout d'une classe et modification locale d'une classe.

Pour chaque nouvelle donnée **x**.

Si $m_x(\phi) > m_x(A); \forall A \in 2^{\Omega} - \{\phi\}$ Alors

x est un point rejeté {**x** est attribué à l'ensemble ϕ }

Incrémenter la cardinalité de ϕ ; $\{|\phi| \leftarrow |\phi| + 1\}$

Si $|\phi| = N_{\min}$ alors :

-Appliquer l'algorithme FCM avec un critère de validité sur les points

rejetés ϕ pour créer une ou des nouvelles classes représentatives C_{new}.

{Le critère de validité à pour objectif de calculer le nombre optimal de classes ajoutées}.

-Incrémenter le modèle de classification

$$M_{new} = M + C_{new}.$$

Fsi.

Sinon

-Affecter **x** à sa propre classe.

-Mettre à jour les caractéristiques de cette classe (centre de classe, matrice de covariance).

-Incrémenter le modèle de classification

$$M_{new} = M + C_{new}.$$

Fsi.

Fpour.

Nous proposons dans l'étape suivante une mesure de dissimilarité entre deux distributions représentés dans des instances différentes. L'objectif de cette mesure est de pouvoir comparer les distributions de deux ensembles de données décrites dans les mêmes espaces, de façons à détecter si leurs distributions sont identiques, similaires ou nettement différentes. Ce type de comparaison permet la détection de changement dans les classes dans le cas des nouvelles données arrivant constamment.

L'idée est de comparer les deux de distribution dans les instances *t* et *t+1*, par la matrice de covariance des classes dans l'état courant (t+1) et l'état précédent (t):

Si $A_{Cj}^{t+1} \neq A_{C_j}^{t}$ alors, il y a un changement local dans cette classe. Dans ce cas, on peut appliquer l'algorithme de fusion, de scission et de suppression d'une classe inutile.

Avec :

A_{Cj}^{t+1} : la matrice de covariance de la classe C_j dans l'instant courant.

$A_{C_j}^{t}$: la matrice de covariance de la classe C_j dans l'instant précédent.

Pour plus de simplification, on peut utiliser les valeurs propres des matrices de covariance des classes pour détecter l'évolution des classes par les comparer dans des instances successive (l'instant ***t*** et l'instant ***t+1***).

VI.4.5.3. Fusion de classes :

Les changements de caractéristiques des classes peuvent également conduire à la fusion de deux ou plusieurs classes, il est possible qu'à un moment donné les caractéristiques de plusieurs classes soient suffisamment proches pour permettre leur fusion. Pour décider de cette fusion, on applique la mesure de similarité utilisé dans [HAR10] qui est introduite précédemment par Frigui et al [FRI96]. Cette mesure de similarité permet de définir le degré de proximité entre deux classes, en se basant sur les valeurs d'appartenances de leurs points, elle est notée par σ_{iz} et définit comme suit :

$$\sigma_{iz} = 1 - \frac{\sum_{x \in C_i et x \in C_z} \left| \mu_i(x) - \mu_z(x) \right|}{\sum_{x \in C_i} \mu_i(x) + \sum_{x \in C_z} \mu_z(x)} \qquad (6.6)$$

μ_i et μ_z sont respectivement les valeurs d'appartenance de x à C_i et C_z. Plus σ_{iz} est proche de 1, plus les deux classes sont similaires.

Après plusieurs expérimentations sur ce degré de similarité, l'auteur a montré que, si $\sigma_{iz} \geq 0.2$ alors, on peut fusionner les deux classes C_i et C_z.

Une fois, la mesure de similarité de *Frigui* détecte la nécessité de fusionner deux classes, il faudrait de mettre à jour les caractéristiques de la classe résultante (la matrice de covariance, le nouveau centre). Pour cela, et pour éviter de recalculer ces caractéristiques dès le début, on peut appliquer l'algorithme EM Compétitive (Compétitive EM) introduit dans

[NEA98]. Cette algorithme nous a permet de calculer les caractéristiques de la classes résultante de l'opération de fusion à partir des caractéristiques des classes initiales.

Algorithme général de fusion de deux classes similaires.

Pour chaque paires de classes *i* et *z*:
  Calculer l'indice de similarité de **Frigui** σ_{iz} .
    Si $\sigma_{iz} \geq seuil$ alors :
      Fusionner les classes *i* et *z*.
      Calculer les nouvelles caractéristiques de la classe résultante (matrice de covariance, le nouveau centre).
      Mettre à jour le modèle de classification M.
    Fsi
Fpour.

VI.4.5.4. Scission de classe :

Au cours de la classification dynamique, suite à l'évolution de certaines classes, il peut une classe contient des données hétérogènes, ce qui contredit avec la notion de classification, où les données de la même classe doivent être homogène. Lorsque de tels phénomènes se produisent, la méthode de classification dynamique doit décider de la scission de classe en utilisant critère de détection de scission de classe. Peu de critères de scission de classes existent dans la littérature, parmi eux, on peut citer le critère de validité introduit dans [XIE91] qui est utilisé dans les travaux de Hatert [HAR10]. Une autre méthode qui permet de détecter une scission d'une classe contient des données hétérogène est celle qui décrit dans qui s'appelle (Compétitive EM) [AMA06], l'avantage de ce dernière est de calculer les caractéristiques des classes résultantes (les nouveaux centres des classes et ces matrices de covariance) après l'opération de scission, cela pour éviter de recalculer ces caractéristique dés le début.

VI.4.5.5. Suppression d'une classe parasite:

Les classes représentant les connaissances doivent être éliminées de l'espace de classification. De même, d'autres classes parasites créées à cause du bruit ou des perturbations extérieures, sont à supprimer. Dans ce cas, tous les algorithmes de classification dynamique développés jusqu'à présent basés sur critère de cardinalité de points (nombre d'élément dans cette classe) pour supprimer les classes créées à cause du bruit [AMA06] [HAR10] [BOU07],…etc.

Cependant, la notion de supprimer une classe obsolète ou parasite dépend fortement du contexte de l'application. De même, dans un espace de classification dynamique où la base d'apprentissage est enrichit au cours de temps, à quel moment peut on considérer qu'une classe est parasite ? [AMA06]

Dans [AMA06], l'auteur a proposé comme un critère de cardinalité pour supprimer une classe inutile, la valeur $N_{min}=d+1$ pour considérer une classe est représentative.

d: représente la dimension de données.

Dans [HAR10], l'auteur a proposé d'attribuer au paramètre N_{min} la valeur *k avec, k* représente les *k*- proches voisins définit dans ces algorithmes.

La condition de cardinalité d'une classe n'est pas suffisante pour supprimer une classe non représentative. Il faut distinguer l'apparition d'une nouvelle classe représentative de l'apparition de classe parasite, crée à cause du bruit et de perturbations extérieures, pour cela, [HAR10] a ajouté une deuxième condition pour distinguer les deux phénomènes (l'ajout et suppression d'une classe). Cette condition consiste à supprimer une classe si aucun élément classé dans une classe pendant qu'un nombre N_{max} *d'*éléments ont été classés dans d'autres classes. Pour supprimer une classe inutile, les deux conditions (cardinalité d'une classe N_{min}, nombre d'éléments classés dans d'autres classes N_{max} alors que aucun éléments classés dans cette classe). Après, la suppression d'une classe créées à cause de bruit, il faudrait de mettre à jour le modèle de classification *M*.

L'algorithme général de supprimer une classe crées à cause du bruit est comme suit :

Algorithme général pour supprimer une classe inutile :

Initialisation : N_{max}, N_{min}, $p \leftarrow 0$. {*p :nombre d'éléments dans la classe* C_i }.
Pour chaque classe C_i de domaine de classification M :
Si $|C_i| < N_{min}$ alors :
Pour chaque nouvel élément *x*
Si $x \notin C_i$ *alors : {si x n'est pas attribuer à la classe* C_i*}.*
$p \leftarrow p+1.$
Si $p = N_{\max}$ *alors*
Supprimer la classe C_i.
Mettre à jour le domaine de classification $M_{new}=M-\{C_i\}$.
.
Fsi.
Sinon
{*x est affecté à la classe* C_i }.
$p \leftarrow 0.$
Fsi.
Fpour.
Fsi.
Fpour.

Il est indispensable de citer que les approches développés par [HAR10] ne contient pas le mécanisme d'incrémentation dès fusion, scission des classes.

Nous pouvons résumer notre proposition d'approche de fusion de classifieurs appliquée à la classification de données évolutives (les signaux EEG), par l'architecture représentée ci-dessous :

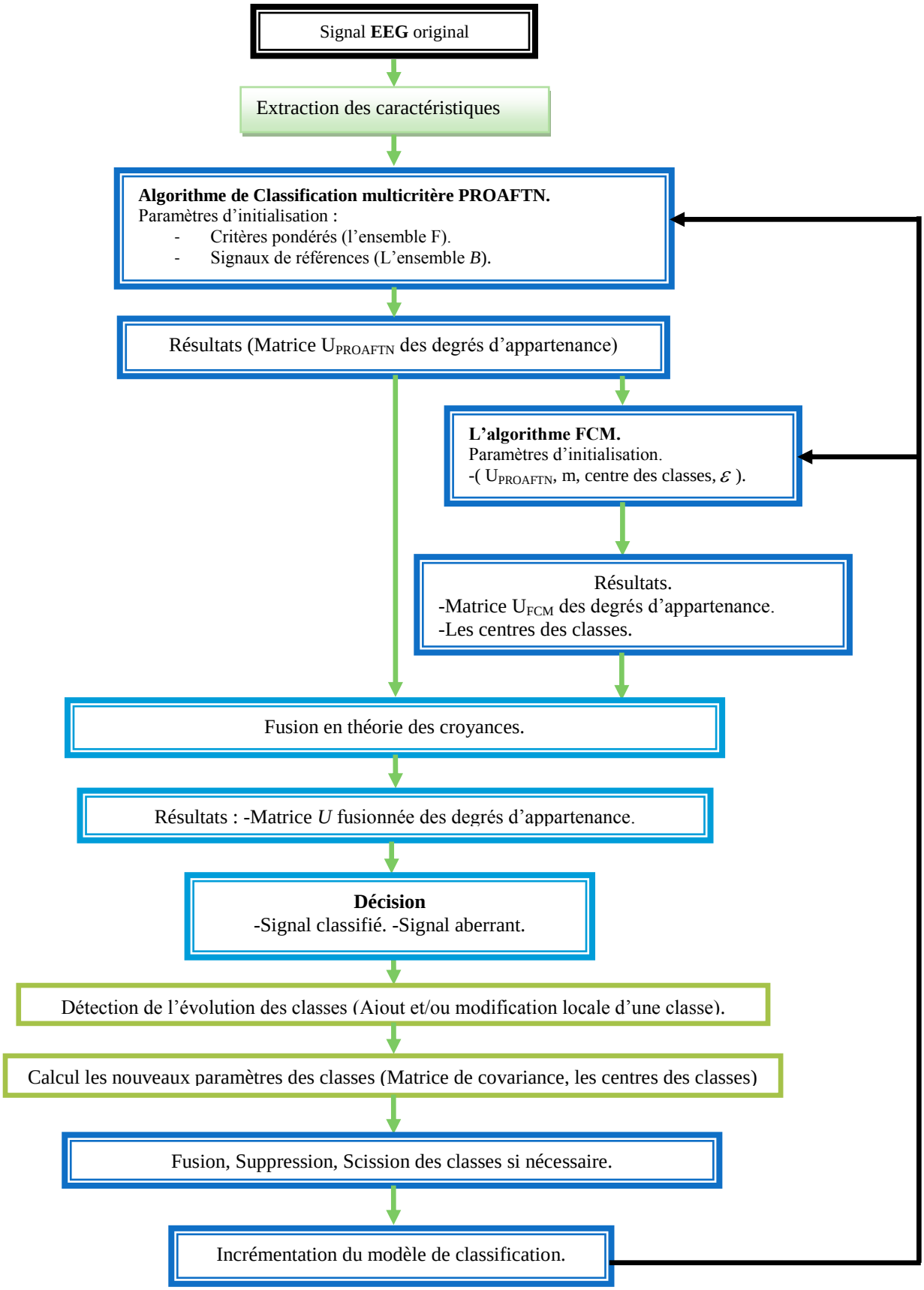


Figure 40: Architecture générale de l'approche proposée

VI.5. Conclusion :

Dans ce chapitre, nous avons explicité en détail les différentes étapes de construction un classifieur dynamique. Tout d'abord, nous avons présenté les méthodes de classification utilisées dans notre approche. Nous avons choisi d'utiliser la méthode de classification floue multicritère qui située dans le domaine d'aide multicritère à la décision et les méthodes de classification floue situées dans le domaine d'apprentissage automatique. Ce choix est motivé essentiellement par les avantages apportés par ces algorithmes en qualité de notre application.

Après la phase de modélisation (choix des algorithmes de classification), nous avons précisé le cadre théorique de fusion retenu, nous avons opté pour la théorie de Demster-Shafer, ou la théorie des fonctions des croyances qui offre plusieurs avantages concernant la modélisation des données évolutives. Ainsi que, cette théorie introduit plusieurs opérateurs de fusion à comportements différents (disjonctifs, conjonctifs et de compromis) et plusieurs règles de décisions permettant d'exploiter les résultats de la fusion telles que (maximum de plausibilité, maximum de crédibilité, probabilité pignistique, intervalle de confiance,…etc).

Notre approche permette de traiter toutes les différentes évolutions que des classes dynamiques peuvent rencontrer (nouvelle classe, modification locale de classes, fusion des classes, scission de classe, et l'élimination de classes parasites ou obsolètes).

Dans cette étape, on utilise les algorithmes d'apprentissage incrémental en prend en considération les itérations précédentes à chaque mise à jours de la base d'apprentissage. Le but de cette étape est extraire des connaissances sans avoir réaliser un apprentissage complet. Dans notre cas, les connaissances à extraire sont les centres des classes, la matrice de covariance de chaque classe. Cette approche convient aux problèmes de classification dans un environnement ouvert où *toutes* les classes ne sont pas connues *a priori.*

Dans le chapitre suivant, nous allons évaluer notre système proposé.

Chapitre 07 : IMPLÉMENTATION ET RÉSULTATS EXPÉRIMENTAUX

VII.1. Introduction :

Dans un démarche scientifique, chaque proposition doit-être évaluer empiriquement et comparée avec l'état de l'art. Pour cela, il est indispensable d'implémenter notre approche tout en effectue une série de tests expérimentaux.

Ce chapitre est consacré à l'application de méthode développée dans le domaine de l'aide au diagnostic médicale. Après une brève introduction générale sur les problèmes de classifications médicales. Une application de cette méthode dans le domaine de classification des signaux EEG. En particulier, détection des crises épileptiques.

Les classifications médicales des pathologies ont pour but de rassembler en classes les cas qui ont des similitudes biologiques fondamentales et qui sont susceptibles de partager certains facteurs étiopathologiques. L'identification de ces classes est importante car elle permet d'une part de comprendre le processus de la maladie et d'autre part d'instaurer l'approche thérapeutique adéquate. De plus, elle permet de dégager le pronostic global de la maladie. Plusieurs méthodes de classification ont été développées pour la classification des signaux EEG comprenant les statistiques, les réseaux de neurones, …, mais l'avantage de notre méthode est la prise en compte de fusion des classifieurs supervisés et non-supervisés et suivi des données et les classes évolutives ainsi que l'intervention d'un ou plusieurs décideurs.

VII.2. Environnement d'implémentation :

Nous avons choisi MATLAB comme environnement d'implémentation pour les facilités qu'il offre dans la manipulation des matrices et toutes les opérations élémentaires sur les matrices telles que la transposée, l'inverse, …etc sont prédéfinies dans le langage. Nous avons donc utilisé une matrice dont les colonnes contenant les exemples d'apprentissage et les lignes les critères (ou dimension).

VII.3. Base de données utilisée:

Après le choix de l'environnement d'implémentation de notre approche, nous choisissons la base de données pour les tests d'évaluation s'accorde bien avec le contexte de notre travail, pour cela nous avons cherché une base de données contenant des données évolutives, floues et multi-modales.

Le but principal de notre travail est l'affectation des données évolutives et multimodales dans ses classes optimales. Ainsi que, le suivi des données et les classes évolutives, enfin nous avons présenté le mécanisme d'incrémentation de notre système lors de l'arrivée de nouvelles données pour éviter de recalculer les paramètres des classes dès le début. Nous avons utilisé le domaine d'aide multicritère à la décision pour ces avantages qu'il offre par rapport à domaine de l'apprentissage automatique (Annexe A).

Pour cela, on utilise une approche de coopération de plusieurs classifieurs (plusieurs méthodes de classification).

L'ensemble de données utilisé dans ce mémoire est accessible au public en ligne [EPI01] à partir du Centre de l'épilepsie à l'Université de Bonn, Allemagne [AND01]. Il contenait cinq sous-ensembles de signaux EEG (Z, O, N, F et S) qui comprennent des enregistrements EEG sains et épileptiques.

Ces signaux ont été sélectionnés et découpés à travers des enregistrements continus de l'EEG multi-canal, après l'inspection visuelle des artefacts, par exemple en raison de l'activité musculaire, des mouvements oculaires et des bruits extérieurs.

Sous-ensembles Z et O ont été enregistrés à partir de cinq participants sains, avec les yeux ouverts et fermés respectivement avec le montage d'électrodes standard (le montage international 10-20).

Les ensembles F, N, S provenant de cinq patients épileptiques, les segments de la classe N ont été enregistrés à partir de la zone hippocampe du l'hémisphère opposé du cerveau. Les signaux des classes F et N sont mesurés pendant des intervalles inter-critiques, tandis que, les signaux de la classe S sont mesurés pendant la phase critique (pendant la crise). Tous les signaux ont été recueillis avec le même système d'amplificateur. Chaque classe contient 100 signaux et chacun de ce dernier est composé de 4097 échantillons successifs. Les 4097 échantillons successifs ont été divisés en huit (08) sous ensembles, chacun contient 512

échantillons (4096 en total) et le dernier a été supprimé. La figure 41 illustre des exemples de signaux de 05 classes de notre base de données. Le tableau 6 illustre les différentes classes utilisées pour l'expérimentation.

Classe	Description.
Z	État sain : relaxé (les yeux fermés)
O	État sain : non-relaxé (les yeux ouverts)
N	État épileptique : sans crise (région hippocampe)
F	État épileptique : sans crise (région épileptogène).
S	État épileptique : avec crise.

Tableau 6:la description des différentes classes des signaux EEG utilisées.

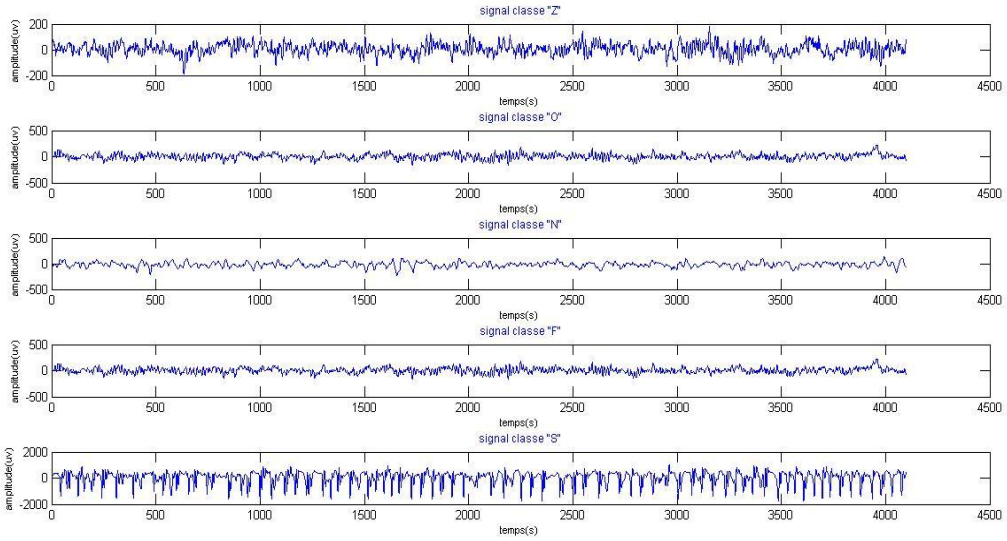


Figure 41:Des exemples des signaux EEG des classes Z, O, F et S.

VII.3.1.Critères de choix de la base de données :

Le choix de la base de données [AND01] a été basé principalement sur les raisons suivantes :

- Disponibilité d'un volume très important de données évolutives dans le temps représentant l'état de personne dans différentes instances.
- Multi-modalité des données caractérisant les signaux EEG (Statistiques, fréquentielles, temporelles, …).

- Consistance de la base, elle exploitée dans plusieurs travaux de recherche récents [SUB07][CHA09][GUE09], …etc.
- Disponibilité de l'ensemble des classes.

VII.3.2. Création de la base de données :

Pour chaque signal, nous avons transformé sous fichier texte en un fichier EXCEL, ensuite nous avons ordonnée et consolider tous ces fichiers dans une seule feuille EXEL afin de l'exporter dans notre environnement d'implémentation MATLAB.

VII.4. Extraction des caractéristiques (les critères):

Pour évaluer le bon fonctionnement du notre système, nous devons essayer d'examiner un ensemble de caractéristiques diverses. Les transformations mathématiques sont appliquées à des signaux afin d'obtenir plus d'informations à partir de ce signal. Dans ce travail, quelques caractéristiques sont extraites telles que les critères temporels ou d'amplitude, les critères fréquentiels, les critères de forme et les critères de similarité pour détecter les états mentaux.

VII.4.1. Les critères temporels :

- La valeur maximale et minimale représente la valeur d'amplitude des échantillons de données EEG collectés.
- La moyenne calcule la valeur d'amplitude moyenne des échantillons de données EEG recueillies entre les extrémités de la zone sélectionnée. L'équation suivante est utilisée pour extraire la valeur moyenne du signal EEG :

$$moyenne = \frac{1}{n_e - n_s} \sum_{i=n_s}^{n_e - n_s} x_{i\,EEG}.$$

Où :

n_s: Représente le point de départ de l'échantillon.

n_e: Représente le point final de l'échantillon. i : représente les valeurs des points à axe horizontale.

- Standard deviation (Ecart-Type) :

$$\sigma_x = \sqrt{\frac{1}{(n_e - n_s) + 1} \sum_{i=n_s}^{n_e - n_s} \left(x_{i_{EEG}} - \bar{x}_{i_{EEG}}\right)^2}$$

VII.4.2. Critère de similarité :

Le critère de similarité peut être mesuré en utilisant le coefficient de corrélation $\rho_{(x_{ref}, x_i)}$ avec les signaux de référence. Avec x_{ref} : le signal de référence.

Le coefficient de corrélation est écrit comme suit :

$$\rho_{(x_{ref}, x_i)} = \frac{\mathrm{cov}\left(x_{ref}, x_i\right)}{\sigma_{ref} . \sigma_{x_i}}$$

VII.4.3. Critères de forme :

- **Coefficient d'asymétrie (CAs):**

Ce coefficient (*Skewness* en anglais) montre le degré d'asymétrie dans la répartition (l'écart de distribution gaussienne normale) du signal EEG. Un coefficient positif indique une distribution décalée à gauche de la moyenne et un coefficient négatif indique une distribution décalée à droite de la moyenne. Il est calculé par la règle suivante :

$$CAs = \frac{\frac{1}{n_e - n_s + 1} \sum_{i=n_s}^{n_e} \left(x_{i_{EEG}} - \bar{x}_{i_{EEG}}\right)^3}{\left(\frac{\sqrt{\sum_{i=n_s}^{n_e} \left(x_{i_{EEG}} - \bar{x}_{i_{EEG}}\right)^2}}{n_e - n_s}\right)^3}$$

- **Coefficient d'aplatissement (CAp):**

Ce coefficient (*Kurtosis* en anglais) indique le degré d'aplatissement de la distribution du signal EEG. Si le coefficient d'aplatissement est supérieur à 0, la distribution possède une forme pointue au niveau de la moyenne. Si le coefficient d'aplatissement est inférieur à 0, le pic de la distribution est plus arrondi autour de la moyenne. L'équation suivante est utilisée pour extraire l'aplatissement :

$$CAp = \frac{\frac{1}{n_e - n_s + 1}\sum_{i=n_s}^{n_e}\left(x_{i\,EEG} - \bar{x}_{i\,EEG}\right)^4}{\left(\sum_{i=n_s}^{n_e}\frac{\left(x_{i\,EEG} - \bar{x}_{i\,EEG}\right)^2}{n_e - n_s}\right)}$$

VII.4.4. Les critères fréquentiels :

La transformée de Fourier est fréquemment utilisée pour le traitement du signal pour analyser le domaine fréquentiel, mais la transformée de Fourier n'est pas appropriée si le signal a une fréquence variable dans le temps comme on le voit dans le cinquième chapitre. Pour cela et afin de pallier cet inconvénient, on utilise la transformée on ondelette discrète (Discret Wavelet Transform, DWT). DWT décompose les signaux non-stationnaires aux différents intervalles de fréquences avec différents résolutions, le signal est décomposé en un ensemble de coefficients décrivant le contenu fréquentiel à des instants donnés.

La figure 42 Illustre les nivaux de décomposition des signaux de la classe S.

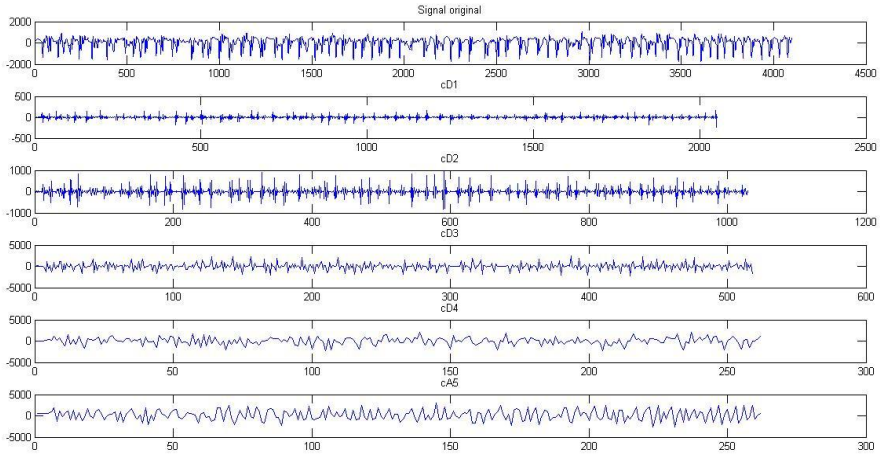


Figure 42: Décomposition d'un signal EEG épileptique par la transformée en ondelettes.

Dans la figure ci-dessus, un exemple du signal épileptique de la classe S a été décomposé en séries d'ondes. cD1 c'est le premier niveau de décomposition, il correspond aux fréquences d'onde (γ), cD2, cD2, cD3, cA5 correspond les fréquences d'ondes $(\beta, \alpha, \theta, \delta)$, respectivement.

- **L'énergie :**

L'énergie (*E*) représente la puissance du signal, l'énergie relative peut être considérée comme une caractéristique pertinente pour démontrer l'évolution des données. L'énergie relative a été calculée à partir des coefficients d'ondelette *C(a,b),* pour chaque instant *b* et fréquence *a.* L'haute énergie implique une activité de crise. L'énergie relative peut être donnée par la formule suivante :

$$E_a = \sum_{i=n_s}^{n_e} (C(a,b))^2$$

- **Entropie (EN):**

Est une mesure numérique utilisée pour quantifier l'information présente au niveau du signal. La présentation mathématique est donnée comme suit :

$$Entropy\ (EN) = -\sum_{i=n_s}^{n_e} P_{a,b}\ log P_{a,b}$$

Avec :

$P_{a,b}$: est le rapport d'énergie et a été calculée comme suit :

$$P_{a,b} = \frac{E_a}{E_{tot}}$$

E_{tot} est l'énergie totale et calculée par l'équation suivante :

$$E_{tot} = \sum_{a=1}^{z} \sum_{b=n_s}^{n_e} (C(a,b))^2 .$$

Avec z le nombre de bandes de fréquences.

VII.5. L'évolution des caractéristiques des classes :

Pour montrer l'évolution des signaux EEG, on calcule la variance et la moyenne de différents critères pour chaque période de temps.

	EN (Alpha)		EN (Bêta)		CAp		CAs	
	L'écart-type	Moyenne	L'écart-type	Moyenne	L'écart-type	Moyenne	L'écart-type	Moyenne
Classe Z	0,22	0,128	0,12	0,03	0,77	2,64	0,73	-0,11
Classe O	0,18	0,085	0,13	0,03	0,82	2,63	0,71	0,01
Classe F	0,24	0,3	0,22	0,3	0,69	2,75	0,72	-0,02
Classe N	0,26	0,27	0,17	0,12	0,74	2,92	0,59	-0,16
Classe S	0,17	0,33	0,20	0,15	1,35	4,76	0,63	-1,37

Tableau 7:L'écart-type et la moyenne des exemples des signaux EEG de cinq classes.

.D'après ce tableau, nous remarquons que l'écart-type varie entre 0.17 et 0.26 pour le critère entropie d'ondes alpha, 0.12 à 0.2 pour le critère d'ondes bêta, 0.69 à 1.35 pour le coefficient d'aplatissement et 0.59 à 0.73 pour le coefficient d'asymétrie. Cela signifie que les signaux EEG présentent un effet variable et dynamique.

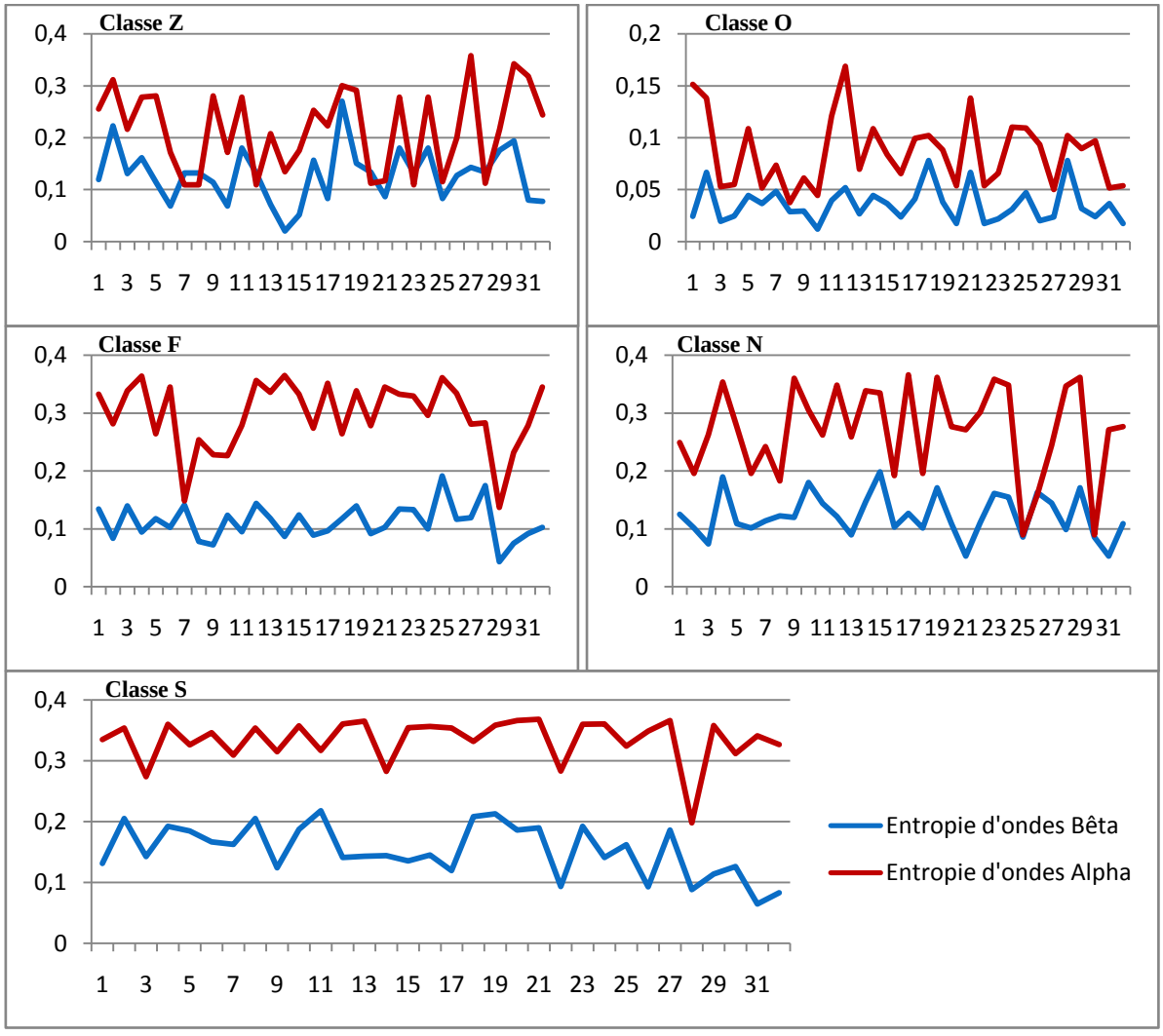


Figure 43: L'évolution des critères fréquentiels des exemples des signaux de chaque classe.

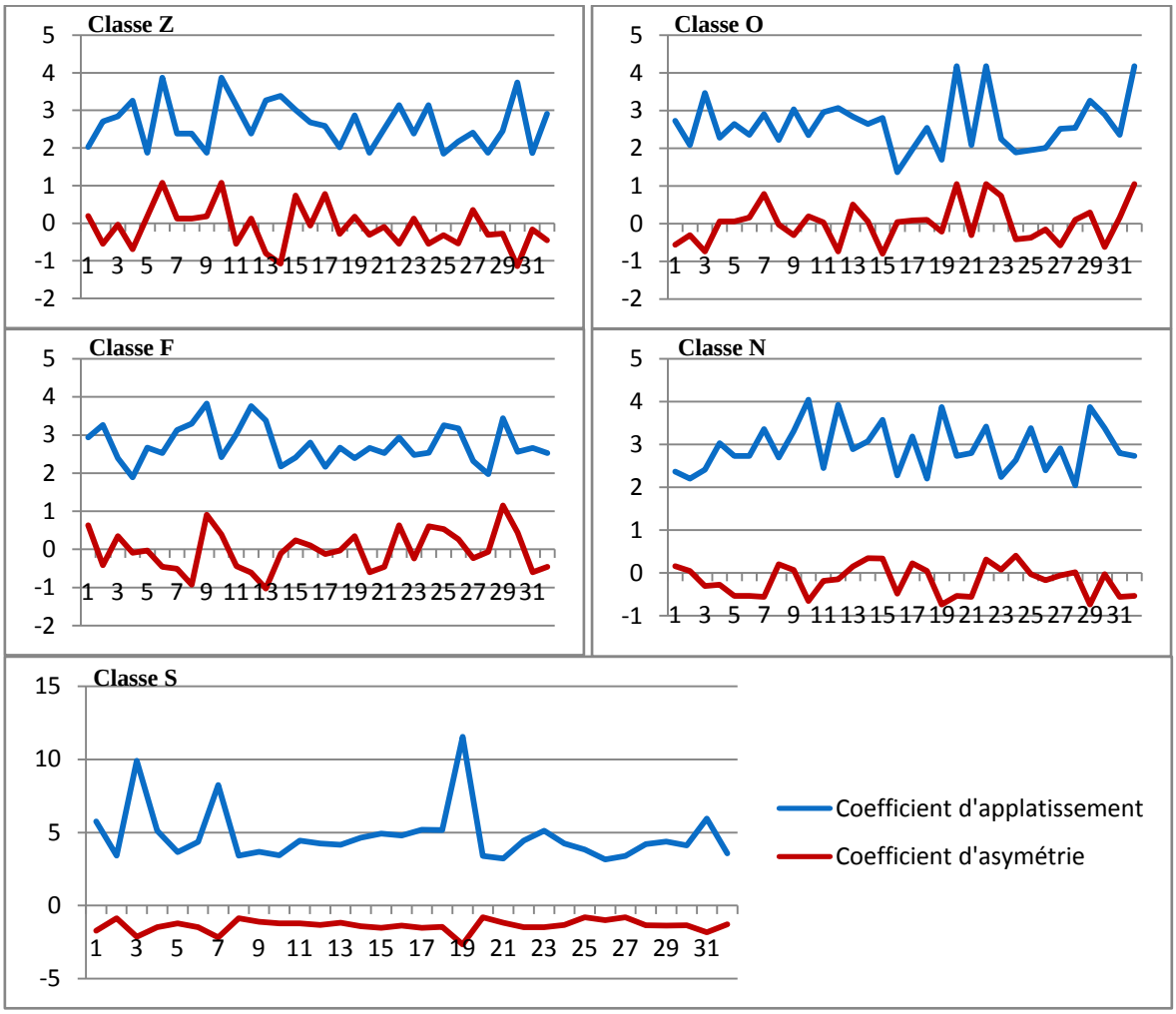


Figure 44:L'évolution des caractéristiques de forme des signaux EEG de chaque classe.

Pour démontrer la notion du critère dans l'application au diagnostic par l'Électroncéphalographie où il n'existe généralement pas d'alternative optimisant tous les critères simultanément. On traite les signaux EEG et à partir de l'étape d'extraction des caractéristiques, on extraite l'intervalle où il appartient chaque signal dans les classes définies ci-dessus :

	Entropie d'ondes Alpha		Entropie d'ondes Bêta		Coefficient d'aplatissement		Coefficient d'asymétrie	
	Min	Max	Min	Max	Min	Max	Min	Max
Classe Z	0,02	0,27	0,01	0,07	1,84	3,87	-1,14	1,07
Classe O	0,03	0,27	0,01	0,07	1,36	4,17	-0,8	1,04
Classe F	0,13	0,36	0,04	0,19	1,9	3,82	-1,03	1,15
Classe N	0,09	0,36	0,05	0,2	2,04	4,04	-0,7	0,4
Classe S	0,19	0,36	0,06	0,21	3,15	11,53	-0,2.6	-0,8

Tableau 8: les actions de référence centrales de toutes les classes.

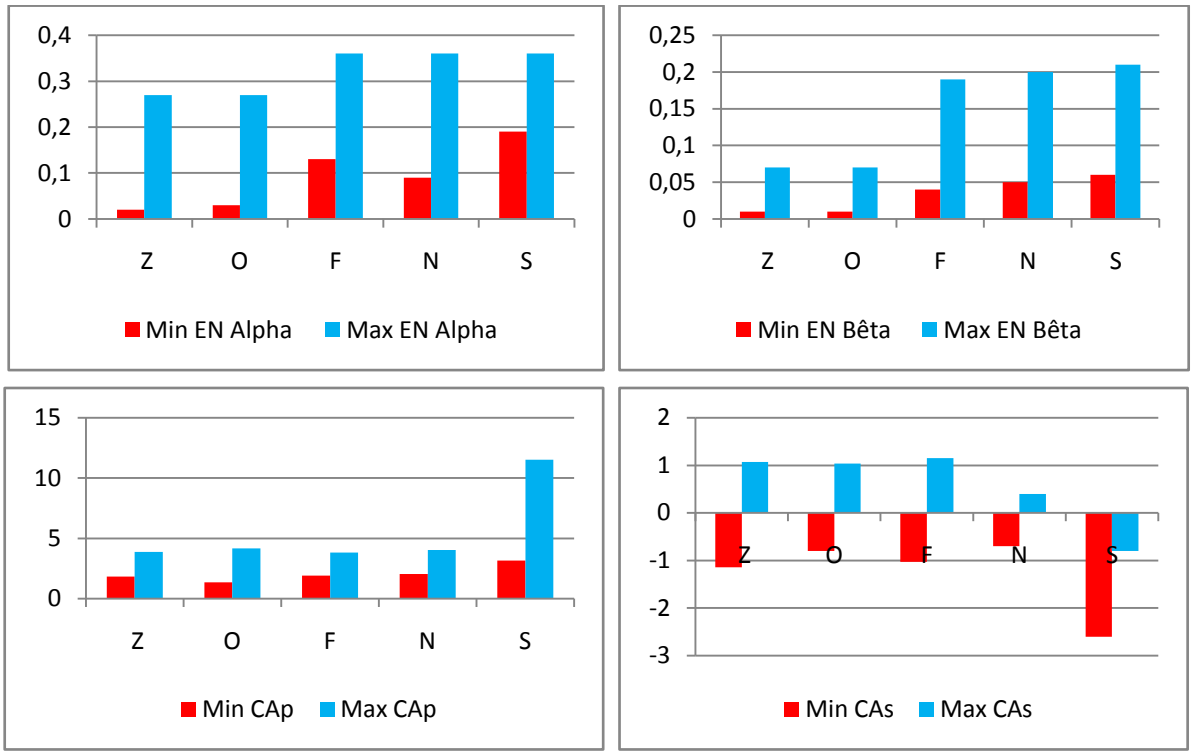


Figure 45:Comparaison de quelques critères extraits à partir des signaux de chaque classe.

A partir du tableau 8, on peut remarquer que les classes ne sont pas représentées par les actions de références limites mais, elles sont représentées par les actions de références centrales (voir, l'annexe A). Par exemple, l'entropie des ondes Alpha sont appartient aux intervalles [0.02, 0.27][0.03,0.27][0.13,0.36][0.09,0.36][0.19,0.36] pour les classes S, Z, O, F et N, respectivement. Si, on pose qu'il y a une alternative qui a une valeur d'entropie d'ondes Alpha de 0.2, où on peut affecter cette alternative ? C'est pour cela, on utilise une méthode d'affectation multicritère et pas monocritère, nominale et pas ordinale car on ne peut pas trouver un ordre entre les classes ou les frontières des classes. Même expression pour tous les critères précédents.

VII.6. Critères de validation :

L'évaluation quantitative des résultats joue un rôle important dans la classification. Pour tester la méthode développée de façon pertinente, nous jugeons la qualité de la classification obtenue par rapport à plusieurs estimateurs souvent utilisés dans la littérature [LAM13][ZOU07].

La performance de la classification des signaux a été faite à partir des critères classiques de sensibilité (Sen) et de Spécificité (Spé), d'exactitude (Acc), de recouvrement (RE) et de similarité (SI), calculés pour chaque classe selon les expressions suivantes :

- VP (Vrais Positifs) : représente le nombre de signaux qui appartiennent à la classe C avec un test positif (i.e. classés par les médecins et détectés par les méthodes automatiques dans la classe C).
- FN (Faux Négatifs) : représente le nombre de signaux qui appartiennent à la classe C avec un test négatif (i.e. classés par les médecins dans la classe C mais mal classés par les méthodes automatiques).
- VN (Vrais Négatifs) : représente le nombre de signaux qui n'appartiennent pas à la classe C avec un test négatif (i.e. non classés par les médecins dans la classe C et non classés par les méthodes automatiques).
- FP (Faux positifs) : représente le nombre de signaux qui n'appartiennent pas à la classe C avec un test positif (i.e. non classés par les médecins dans la classe C et mais reconnus par les méthodes automatiques comme C). Pour plus d'illustration sur les notions précédents, on défini le concept du matrice de confusion :
- Matrice de confusion :

Est un outil servant à mesurer la qualité d'un système de classification. Chaque colonne de la matrice représente le nombre d'occurrences d'une classe estimée, tandis que chaque ligne représente le nombre d'occurrences d'une classe réelle (ou de référence). Un des intérêts de la matrice de confusion est qu'elle permet d'identifier les erreurs de classification.

		Classe Estimée	
		positive	*Négative*
Classe réelle	*Positive*	**VP**	**FN**
	Négative	**FP**	**VN**

Tableau 9:Matrice de confusion.

- **La sensibilité (Sen):** est calculée comme étant le rapport entre les VP's et l'ensemble des signaux analysées. Elle illustre la capacité de l'algorithme à détecter des crises.

$$Sen = \frac{VP}{VP + FN}$$

- **La spécificité (Spé) :**

Est généralement définie par le rapport entre le nombre de VN's et la somme des VN's et FP's. Elle renseigne sur la capacité de l'algorithme à ne pas détecter des événements qui ne sont pas des crises d'épilepsie.

$$Spé = \frac{VN}{VN + FP}$$

- **Le taux de classification (*Accuracy*) :**

Le taux de classification (*Accuracy*) est une mesure qui détermine la capacité de l'algorithme d'affecter chaque donnée EEG à sa propre classe. Il donne la capacité de l'algorithme de classifier correctement les données EEG, il est donné comme suit :

$$Acc = \frac{VP + VN}{VP + VN + FP + FN}$$

L'approche décrite dans le cinquième chapitre a été programmée en langage MATLAB. Les résultats obtenus ont été comparés à ceux obtenus précédemment par les physiologistes afin de déterminer les taux de classification correcte pour chaque classe de l'EEG.

- **Le taux de recouvrement (RE):**

Il correspond à la proportion de vrais positifs par rapport à l'ensemble des signaux qui ont été ou devrait avoir été classifiés.

$$RE = \frac{VP}{VP + VN + FP}$$

Le taux de recouvrement (s'appel aussi indice de Tanimoto [LAM13]) tend vers 1 (resp. 0) s'il y a peu (resp. beaucoup) de faux positifs et de faux négatifs. Cette indicateur permet d'évaluer dans quelle mesure la structure recherchée correspond quantitativement et qualitativement à la classification.

- **Le taux de similarité (SI):**

Il correspond à la proportion de vrais positifs par rapport à l'ensemble des signaux qui ont été ou devraient avoir été classifiés. La similarité tends vers 1 (resp. 0) s'il ya peu (resp. beaucoup) de faux positifs et de faux négatifs. A l'instar de recouvrement, cet indicateur permet d'évaluer dans quelle mesure la structure recherchée correspond quantitativement et qualitativement à la classification.

$$SI = \frac{2*VP}{2*VP + VN + FP}$$

VII.7. Résultats :

Dans cette section, chaque expérimentation est considérée comme un cas, les cas 1 à 4 sont composés en deux classes, la classe épileptique S et une des autres classes. Le cas 5 est composé par la classe S et toutes les autres classes. Ces cas sont donnés comme suit :

Cas 1 : classe S vs classe Z

Cas 2 : classe S vs classe O

Cas 3 : classe S vs classe F

Cas 4 : classe S vs classe N

Cas 5 : classe S vs classes Z, O, F et N.

Notre étude est la classification des données qui variées au cours de temps pour cela, nous allons réaliser les expérimentations dans des instances de temps différents.

Les tableaux 10,11, 12, 13 et14 présentent respectivement la comparaison des critères de spécificité, de sensibilité, de taux de classification, le taux de recouvrement et le taux de similarité de la méthode proposée par rapport aux méthodes FCM et PROAFTN pour différents cas décrits ci-dessus.

Après introduction des critères, nous avons obtenu les résultats suivants :

- *Le critère de sensibilité :*

Algorithme	bi-classes	Sensibilité (Sen)								
		t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	Moy
PROAFTN	*S et Z*	0,97	0,98	0,99	0,99	0,98	0,99	0,91	0,96	*0,97125*
	S et O	0,97	0,98	0,99	0,99	0,98	0,99	0,99	0,96	*0,98125*
	S et F	1	1	1	1	1	1	1	1	*1*
	S et N	0,98	0,98	1	0,99	0,99	1	1	0,98	*0,99*
	S et Z,O,F,N	0,74	0,77	0,76	0,78	0,74	0,73	0,72	0,75	*0,74875*
FCM	*S et Z*	0,85	0,89	0,91	0,89	0,88	0,91	0,93	0,91	*0,89625*
	S et O	0,85	0,89	0,9	0,88	0,85	0,91	0,92	0,91	*0,88875*
	S et F	0,79	0,81	0,78	0,75	0,74	0,82	0,76	0,79	*0,78*
	S et N	0,86	0,87	0,92	0,84	0,93	0,94	0,9	0,93	*0,89875*
	S et Z,O,F,N	0,42	0,37	0,38	0,4	0,38	0,41	0,37	0,34	*0,38375*
PROAFTN & FCM	*S et Z*	0,94	0,98	0,96	0,96	0,98	0,97	0,97	0,96	*0,965*
	S et O	0,93	0,97	0,96	0,96	0,94	0,97	0,97	0,96	*0,9575*
	S et F	0,91	0,92	0,96	0,94	0,9	0,95	0,92	0,94	*0,93*
	S et N	0,99	0,97	0,99	0,99	1	1	1	1	*0,9925*
	S et Z,O,F,N	0,66	0,65	0,69	0,67	0,64	0,68	0,67	0,7	*0,67*

Tableau 10: les résultats de critère de sensibilité dans chaque période de temps.

Discussion :

La sensibilité est la capacité de l'algorithme de détecter les crises d'épilepsie. En observant les résultats obtenus dans le tableau 16 la sensibilité moyenne obtenue par l'algorithme PROAFTN est plus élevée que les deux autres méthodes par un taux de 93.83%, tandis que notre approche a un taux de sensibilité moyenne de 90.30% et la méthode FCM 76.95%. L'analyse par cas de critère de sensibilité obtenu par l'algorithme PROAFTN donne un pourcentage de 97.125% pour le cas 1 et 98.125% pour le cas 2, 100% pour le cas 3, 99% pour le cas 4 et 74.875% pour le dernier cas. Notre approche a donné un taux de sensibilité moyen 96.5% pour le cas 1, 95.75% pour le cas 2, 93% pour le cas 3, 99.25 % pour le cas 4 et 67% pour le cas 5. Au contraire de critère de spécificité, l'algorithme FCM a obtenu des faibles résultats pour le critère de sensibilité, cet algorithme a donné un taux de sensibilité moyen de 89.625% pour le cas 1, 88.875% pour le cas 2, 78% pour le cas 3, 89.875% pour le cas 4 et 38.375% pour le dernier cas.

- *Critère de spécificité :*

Algorithme	bi-classe	Spécificité (Spé)								
		t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	Moy
PROAFTN	*S et Z*	0,9	0,88	0,9	0,91	0,91	0,91	0,9	0,91	0,9025
	S et O	0,87	0,87	0,86	0,89	0,87	0,88	0,83	0,89	0,87
	S et F	0,68	0,69	0,72	0,63	0,74	0,65	0,67	0,7	0,685
	S et N	0,94	0,94	0,94	0,94	0,94	0,93	0,93	0,93	0,93625
	S et Z,O,F,N	0,88	0,88	0,88	0,8775	0,8825	0,88	0,8775	0,88	0,87969
FCM	*S et Z*	1	1	1	1	1	1	1	1	1
	S et O	0,99	0,99	0,99	0,98	0,98	0,98	0,98	0,99	0,985
	S et F	1	1	1	1	1	1	1	1	1
	S et N	1	1	1	1	1	1	1	1	1
	S et Z,O,F,N	0,9975	0,9975	0,9975	0,995	0,9975	0,995	0,995	0,9975	0,99656
PROAFTN & FCM	*S et Z*	1	1	1	1	0,97	1	1	1	0,99625
	S et O	0,99	0,99	1	0,98	0,99	0,98	0,98	0,99	0,9875
	S et F	1	1	1	1	1	1	1	1	1
	S et N	1	1	1	1	1	1	1	1	1
	S et Z,O,F,N	0,935	0,925	0,9275	0,9375	0,93	0,945	0,9175	0,9425	0,9325

Tableau 11:les résultats de critère de spécificité dans chaque période de temps.

Discussion :

Comme on le voit, le critère de validité *spécificité* représente la capacité de l'algorithme à ne pas détecter des événements qui ne sont pas des crises d'épilepsie. En observant les résultats obtenus dans le tableau 11, l'algorithme FCM a donné la meilleure précision pour le critère de spécificité, il a donné un taux de spécificité moyen de 100% pour les cas 1, 3, 4, 98.5% pour le cas 2 et 99.65% pour le dernier cas. Tandis que l'algorithme multicritère PROAFTN a donné un taux de spécificité moyen 90.25% pour le premier cas, 87% pour le cas 2, 68.5% pour le cas 3, 93.625% pour le cas 4 et 87.97% pour le dernier cas. Notre approche est la combinaison de ces deux algorithmes précédents, elle a donné un taux de spécificité moyenne de 99.625% pour le premier cas, 98.75% pour le deuxième cas, 100% pour le troisième et quatrième cas et 83.6% pour le cas 5. A travers ces résultats, l'approche FCM a une capacité maximale pour détecter des événements qui ne sont pas des crises d'épilepsie par un taux de spécificité moyen de 99.63%, tandis que notre approche a un taux de spécificité moyen de 96.39% et l'algorithme PROAFTN a un taux de spécificité moyen de 87.06% (voire le tableau 15 et la figure 47).

- *Taux de classification :*

Algorithme	bi-classe	Taux de Classification (Acc)								
		t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	Moy
PROAFTN	*S et Z*	0,935	0,93	0,945	0,95	0,945	0,95	0,905	0,935	*0,9368*
	S et O	0,92	0,925	0,925	0,94	0,925	0,935	0,91	0,925	*0,9256*
	S et F	0,84	0,845	0,86	0,815	0,87	0,825	0,835	0,85	*0,8425*
	S et N	0,96	0,96	0,97	0,965	0,965	0,965	0,965	0,955	*0,9631*
	S et Z,O,F,N	0,852	0,858	0,856	0,858	0,854	0,85	0,846	0,854	*0,8535*
FCM	*S et Z*	0,925	0,945	0,955	0,945	0,94	0,955	0,965	0,955	*0,9481*
	S et O	0,92	0,94	0,945	0,93	0,915	0,945	0,95	0,95	*0,9368*
	S et F	0,895	0,905	0,89	0,875	0,87	0,91	0,88	0,895	*0,89*
	S et N	0,93	0,935	0,96	0,92	0,965	0,97	0,95	0,965	*0,9493*
	S et Z,O,F,N	0,882	0,872	0,874	0,876	0,874	0,878	0,87	0,866	*0,874*
PROAFTN & FCM	*S et Z*	0,97	0,99	0,98	0,98	0,975	0,985	0,985	0,98	*0,9806*
	S et O	0,96	0,98	0,98	0,97	0,965	0,975	0,975	0,975	*0,9725*
	S et F	0,955	0,96	0,98	0,97	0,95	0,975	0,96	0,97	*0,965*
	S et N	0,995	0,985	0,995	0,995	1	1	1	1	*0,9962*
	S et Z,O,F,N	0,88	0,87	0,88	0,884	0,872	0,892	0,868	0,894	*0,88*

Tableau 12:les résultats de Taux de classification dans chaque période de temps.

Discussion :

Le taux de classification (*Accuracy*) donne la capacité de l'algorithme de classifier correctement les données EEG. Pour tous les cas, l'approche proposée permet d'atteindre une précision de classification plus élevée par rapport aux méthodes FCM et PROAFTN. Le taux de classification moyen est calculé en utilisant les valeurs de précision de tous les cas. La précision de classification moyenne de la méthode proposée est **95.89%**, tandis que la méthode PROAFTN obtient une précision de classification moyenne de 90.43% et la méthode FCM a donné un taux de précision moyen de 91.97%.

La figure suivante détermine le taux d'exactitude de l'approche proposée par rapport aux algorithmes FCM et PROAFTN pour tous les cas. On remarque que

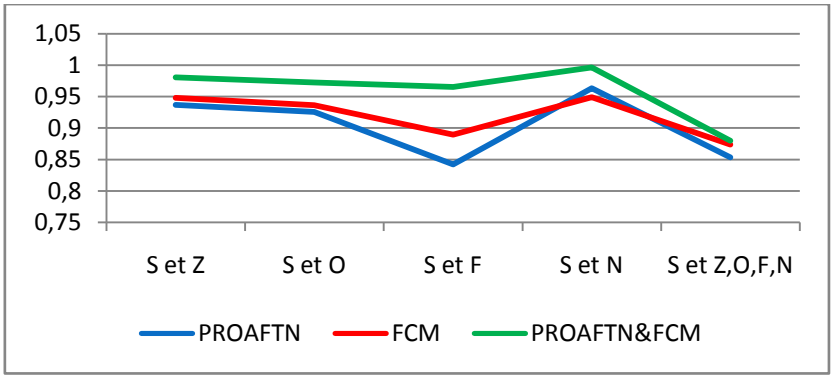


Figure 46: Comparaison de Taux de classification de notre approche et les algorithmes FCM et PROAFTN.

- *Taux de recouvrement :*

Algorithme	bi-classe	Taux De recouverement (RE)								
		t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	moy
PROAFTN	*S et Z*	0,8818	0,875	0,9	0,9082	0,8990	0,9083	0,8273	0,8807	0,8850
	S et O	0,8584	0,8672	0,8684	0,8918	0,8672	0,8839	0,8462	0,8649	0,8685
	S et F	0,7575	0,7633	0,7812	0,7299	0,7936	0,7407	0,7519	0,7692	0,7609
	S et N	0,9245	0,9245	0,94339	0,9339	0,9339	0,9346	0,9346	0,9159	0,9306
	S et Z,O,F,N	0,5	0,5202	0,5135	0,5234	0,5034	0,4932	0,4832	0,5068	0,5054
FCM	*S et Z*	0,85	0,89	0,91	0,89	0,88	0,91	0,93	0,91	0,8962
	S et O	0,8415	0,8811	0,8910	0,8627	0,8333	0,8922	0,902	0,901	0,8756
	S et F	0,79	0,81	0,78	0,75	0,74	0,82	0,76	0,79	0,78
	S et N	0,86	0,87	0,92	0,84	0,93	0,94	0,9	0,93	0,8987
	S et Z,O,F,N	0,4158	0,3663	0,3762	0,3921	0,3762	0,402	0,3627	0,3366	0,3785
PROAFTN & FCM	*S et Z*	0,94	0,98	0,96	0,96	0,9514	0,97	0,97	0,96	0,9614
	S et O	0,9207	0,9604	0,96	0,9411	0,9306	0,951	0,951	0,9505	0,9456
	S et F	0,91	0,92	0,96	0,94	0,9	0,95	0,92	0,94	0,93
	S et N	0,99	0,97	0,99	0,99	1	1	1	1	0,9925
	S et Z,O,F,N	0,5238	0,5	0,5348	0,536	0,5	0,5574	0,5038	0,5691	0,5281

Tableau 13:Comparaison de Taux de classification de notre approche et les algorithmes FCM et PROAFTN.

- *Taux de similarité :*

Algorithme	bi-classe	Taux De similarité (SI)								
		t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	Moy
PROAFTN	*S et Z*	0,9372	0,9333	0,9473	0,9519	0,9468	0,9519	0,9055	0,9366	0,9388
	S et O	0,9238	0,9289	0,9295	0,9428	0,9289	0,9384	0,9167	0,9275	0,9295
	S et F	0,8620	0,8658	0,8771	0,8438	0,8849	0,8511	0,8584	0,8696	0,8641
	S et N	0,9607	0,9607	0,9708	0,9658	0,9658	0,9662	0,9662	0,9561	0,9640
	S et Z,O,F,N	0,6667	0,6844	0,6785	0,6872	0,6696	0,6606	0,6516	0,6726	0,6714
FCM	*S et Z*	0,9189	0,9418	0,9528	0,9418	0,9361	0,9529	0,9637	0,9529	0,9451
	S et O	0,9139	0,9368	0,9424	0,9263	0,9090	0,943	0,9485	0,9479	0,9335
	S et F	0,8826	0,8950	0,8764	0,8571	0,8505	0,9011	0,8636	0,8827	0,8761
	S et N	0,9247	0,9304	0,9583	0,9130	0,9637	0,9691	0,9474	0,9637	0,9463
	S et Z,O,F,N	0,5874	0,5362	0,5467	0,5633	0,5467	0,5734	0,5324	0,5037	0,5487
PROAFTN & FCM	*S et Z*	0,9690	0,9899	0,9795	0,9795	0,9751	0,9848	0,9848	0,9796	0,9803
	S et O	0,9587	0,9798	0,9795	0,9697	0,9641	0,9749	0,9749	0,9746	0,9720
	S et F	0,9528	0,9583	0,9795	0,9690	0,9473	0,9744	0,9583	0,9691	0,9636
	S et N	0,9949	0,9847	0,9949	0,9949	1	1	1	1	0,9962
	S et Z,O,F,N	0,6875	0,6666	0,6969	0,6979	0,6667	0,7158	0,67	0,7254	0,6908

Tableau 14:Taux de similarité dans différents périodes de temps.

Discussion :

Les résultats obtenus après le calcul de taux de recouvrement et le taux de similarité donnent l'avantage de notre approche de fusion de classifieurs flous par rapport aux autres approches. L'approche proposée a obtenu un taux de recouvrement moyen de 87.15% et un taux de similarité moyen de 92.06%. L'algorithme PROAFTN donne 79.01% et 87.36% pour

taux de recouvrement et taux de similarité moyen, respectivement. L'algorithme FCM a obtenu 76.58% pour le recouvrement et 85% pour le taux de similarité moyen.

En résumé, la combinaison d'un algorithme de classification multicritère (PROAFTN) et l'algorithme de classification automatique (FCM) donne les meilleures résultats pour la classification des données évolutives. Á partir des résultats précédentes, l'algorithme PROAFTN a obtenu des meilleures résultats pour la classification des signaux épileptiques (mesure de sensibilité), et l'algorithme FCM augmente la mesure de spécificité c'est-à-dire, il permet de classifier mieux les signaux qui ne sont pas épileptiques. Á cause de fusion de ces deux algorithmes, on obtient un meilleur taux de classification. La figure et le tableau ci-après illustrent les mesures de validation de chaque approche.

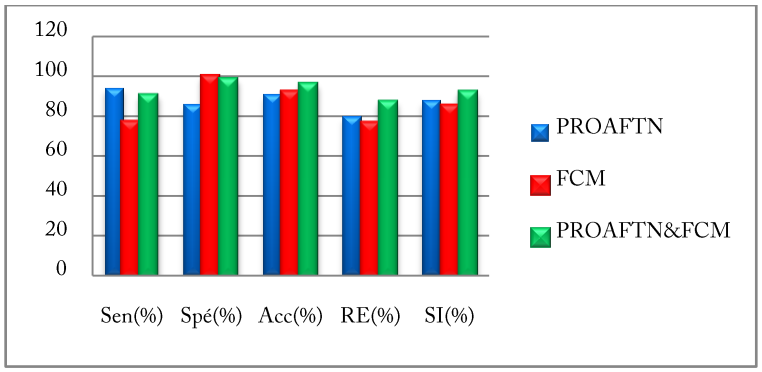


Figure 47: Comparaison entre l'approche proposée et les algorithmes FCM et PROAFTN.

	Sensibilité(%)	Spécificité(%)	Accuracy(%)	Recouvrement(%)	Similarité(%)
PROAFTN	***93,83***	85,47	90,43	***79,01***	***87,36***
FCM	76,95	***99,63***	***91,97***	76,58	85
PROAFTN&FCM	***90,3***	***98,33***	***95,89***	***87,15***	***92,06***

Tableau 15:Comparaison entre l'approche proposée et les algorithmes FCM et PROAFTN.

VII.8. Etude comparative et discussion :

Afin de disposer non seulement d'une évaluation absolue, mais aussi d'une évaluation relative par comparaison à des solutions existantes dans la littérature, les résultats obtenus par le système de fusion floue de données ont été comparés avec ceux fournies par trois approches de classification des signaux EEG. La première basée sur le classifieur SVM décrit dans [CHA09], la deuxième est un modèle de classification de réseau de neurones décrit dans [GUE09] et la troisième est une approche basée également sur une architecture neuronale décrit dans [SUB07]. Il est indispensable de citer que les auteurs ont développé ces approches à base des classifieurs statiques et non dynamiques où n'ont pas décrit le caractère

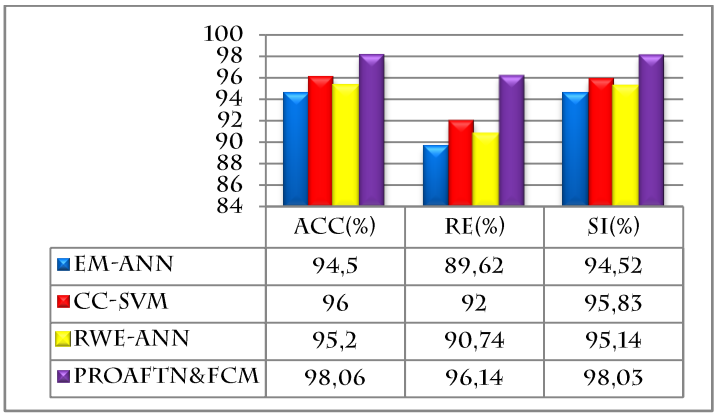
d'évolution des données et des classes (notion de partition dynamique) comme notre approche.

Dans [SUB07], l'auteur a décrit une architecture (EM-ANN) basée sur un réseau de neurones avec l'algorithme EM (Expectation Maximisation) pour détecter les crises épileptiques. Dans cette approche, les signaux EEG sont décomposés en sous bandes de fréquences $(\delta, \theta, \alpha, \beta, \gamma)$ à base de la transformée en ondelettes discrètes et chèque fréquence est utilisée comme entrée de ce réseau. Notons que cette approche est testée sur les classes S et Z seulement avec le taux de sensibilité de 95%, le taux de spécificité de 94% et le taux de classification de 94.5%.

Dans [CHA09], les auteurs ont proposé d'appliquer le classifieur SVM pour séparer les deux classes des signaux épileptiques (S) et non épileptiques (Z) avec la propriété de la Cross-Correlation entre les signaux de référence et le signal à classifié, on dénote cette approche par l'abréviation CC-SVM. Le taux d'exactitude de cette approche est 96% avec le taux de spécificité de 100% et taux de sensibilité est 92%.

Dans [GUO09], les auteurs ont développé une architecture neuronale basée sur la transformée en ondelettes avec l'énergie relative de différentes bandes de fréquences comme caractéristiques de ce réseau de neurones (RWE-ANN). Egalement, cette approche a été appliquée pour classifier les signaux de la classe S et la classe Z. les performances de cette approche sont données par taux de sensitivité de 98.17% et de taux de spécificité de 92.12% et taux de classification de 95.2%.

Le tableau 16 récapitule les résultats des travaux ci-dessus et les résultats de notre approche. En effet, on peut remarquer que les résultats obtenus par le système de fusion proposé sont meilleurs que ceux produits par les approches décrit dans [SUB07][GUE09][CHA09]. Notre système a un taux de classification de 98.06%, taux de recouvrement de 96.14% et taux de similarité de 98.03%. Par contre, les approches de classification précédentes fait passer la valeur du taux d'exactitude de 94.5% à 96%, avec un taux de recouvrement de 89.62% à 92% et taux de similarité de 94.52% à 95.83%.



	ACC(%)	RE(%)	SI(%)
EM-ANN	94,5	89,62	94,52
CC-SVM	96	92	95,83
RWE-ANN	95,2	90,74	95,14
PROAFTN&FCM	98,06	96,14	98,03

Tableau 16: Comparaison entre l'approche de fusion proposée et les approches décrit dans [SUB07][GUO09][CHA09].

VII.9. Conclusion :

La mise en œuvre d'une solution de classification dynamique basée sur une approche de combinaison de classifieurs nous a permis de faire coopérer et exécuter simultanément les deux types de méthodes de classification supervisées multicritère en aide multicritère à la décision et les méthodes de classification automatique en apprentissage automatique, afin d'augmenter les performances de système de classification en aide au diagnostic médical.

Dans ce chapitre nous avons présenté une série d'expérimentations pour valider le système de fusion de classifieurs proposé dans le chapitre 06 afin de classifier et suivi l'évolution les signaux EEG épileptiques et non épileptiques au cours de temps en vue d'améliorer la qualité de la classification. Une étude des caractéristiques évolutionnaires des signaux EEG est menée en premier temps. En second temps le modèle développé est comparé avec quelques travaux de classification de la littérature. Les résultats obtenus après le calcul des différents critères d'évaluations à savoir le taux de classification, le recouvrement et la similarité sont meilleurs et indiquent que le système proposé semble efficace d'avoir une classification de qualité à cause du modèle de fusion de classifieurs choisi. Ces résultats obtenus sont assez encourageants et soulignent le potentiel de la fusion de classifieurs dans le domaine médical.

« Le problème, avec le cerveau, c'est que le seul outil qui permette de l'étudier et d'améliorer son fonctionnement c'est… le cerveau lui-même.».Edmond.Wells.

Conclusion générale et perspectives

Conclusion :

Le développement de méthodes d'analyse dynamique de l'information, comme le clustering incrémental et les méthodes de détection de nouveauté, devient une préoccupation centrale dans un grand nombre d'applications dont le but principal est de traiter de larges volumes d'information variant au cours du temps. Ces applications se rapportent à des domaines très variés et hautement stratégiques, tels que l'exploration du Web et la recherche d'information, l'analyse du comportement des utilisateurs et les systèmes de recommandation, la veille technologique et scientifique, ou encore, l'analyse de l'information génomique en bioinformatique.

Le travail présenté dans ce mémoire s'insère dans le domaine de la classification dynamique floue à partir des données multimodales et évolutives. Au départ, nous avons évoqué une bonne synthèse bibliographique des méthodes existantes pour la classification dynamique de données. Il y a deux types de classification : classification supervisée en aide multicritère à la décision et en apprentissage automatique, et classification non-supervisée en apprentissage automatique. Ces derniers sont aussi classés selon deux catégories : les méthodes floues et les méthodes non-floues : cependant, en combinant deux types de méthodes de classification multicritère et de classification en apprentissage automatique. L'approche développée est composée en six phases : acquisition des données, prétraitement des données, extraction des caractéristiques des données, modélisation des données pour construire une matrice des degrés d'appartenance, fusion les matrices d'appartenance en utilisant la théorie des fonctions des croyances (théorie de dempster-Shafer), la détection des évolutions des données et des classes et l'incrémentation du modèle de classification.

Nous avons simulé le comportement de cette approche avec une base de données évolutive caractérisant la détection de l'état des personnes (sain ou malade en épilepsie) à partir des signaux d'EEG (ElectroEncéphaloGraphie), en coopérant l'ensemble des classifieurs multicritère (PROAFTN) et le classifieur non-supervisé flou FCM par une méthode de fusion de données (théorie de l'évidence) par une technique hybride. Cela a permis de réduire considérablement la complexité du processus décisionnel, par une séparation entre les phases de traitement, ainsi de réduire le temps d'obtention des résultats décisionnels, tout en garantissant de bonnes performances en classification suite à la maitrise de la non-stationnarité des données et classes.

Perspectives :

Nous nous somme placés dan une étude importante dans notre vie, c'est le diagnostic médical par l'ElectroEncéphaloGraphie pour détecter l'état du patient (épileptique ou autre) :

- En résumer, nous avons pu avec ce travail une nouvelle stratégie de classification dynamique pour détecter *les artéfacts* et les tumeurs dans l'EEG.
- Réalisation d'une application vidéo pour suivi les évolutions des données et classes.
- Suivi un système évolutive en utilisant le formalisme des équations différentielles continues (ces équations sont des meilleurs formalismes mathématiques pour suivi les trajectoires des données).
- Création une base de données de notre CHU de quelques patients et suivi l'état de chacun.
- Combiner les données d'IRM (Imagerie en Résonance Magnétique) et les données d'EEG, pour détecter mieux l'état du patient.
- Développement d'une nouvelle approche d'anticipation des crises épileptiques par combinaison de la théorie du Chaos et les méthodes de classification automatique.
- Utilisation des techniques d'optimisation à savoir, les algorithmes génétiques et l'algorithme des essaims de particules pour détecter les évolutions des données et les classes.

BIBLIOGRAPHIE

[ABB12] ABBAS. Mohamed Amir ; « *Système multi agents pour la classification dynamique floue de données multimodales et évolutives* » ; mémoire de magister ; université de Blida, Algérie ; 2012.

[ALL07] ALLILI Mohand Saïd, BOUGUILA Nizar, ZIOU. Djemel ; «*Finite generalized Mixture modeling and application to image and video foreground segmentation*»; Forth Canadian Conference on Computer and Robot Vision (CRV07); Canada; 2007.

[AMA06] AMADOU BOUBACAR. Habiboulay ; «*Classification des données non stationnaires, Apprentissage et suivi des classes évolutives* » ; Thèse de doctorat, université des sciences et technologies de LILLE, France ; soutenue le 28/06/2006.

[AMB09] AMBAPEUR. Samuel ; « *la théorie des ensembles flous : application à la mesure de la pauvreté au Congo.* » ; Bureau d'application des méthodes statistiques et informatiques BAMSI, Brazzaville, Congo, 2009.

[AND01] ANDRZEJAK. Ralph G, LHNERTZ. Klaus, MORMANN. Florian, RIEKE. Christophe, DAVID. Peter, ELGER. Christian; "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state"; PHYSICAL REVIEW, vol 64, n06, 061907;pp 01-08; published 20 November 2001.

[APS14] http://www.aps.dz/sante-sciences-tech/

[ASS10] ASSAGHIR. Zaineb ; « *Analyse formelle de concepts de fusion d'informations : Application à l'estimation et contrôle d'incertitude des indicateurs agro-environnementaux* » ; thèse de doctorat ; Institut national polytechnique de Lorraine, France ; Soutenue le 12/11/2010.

[AYU09] AYUB. Mohammad; *«Choquet and Sugeno Integrals»;* Thesis of the degree Master of science in Mathematical Modelling and simulation; Belking institute of Technologie School of Engineering; February 2009.

[BAJ14] BAJAJ. Varun; PACHORI. Ran. Bilas ; *« Humane motion classification from EEG signals using multivalet transforms* » ; International conference on Medical Biometrics ; pp : 125-130 ; 2014.

[BAR00] BARRA. Vincent ; « *Fusion d'images 3D du cerveau : Etude de modèles et applications » ;* Thèse de doctorat, Université d'Auvergne, France ; Soutenue le 10 Juillet 2000.

[BEL00] BELACEL. Nabil; *«Méthode de Classification Multicritère, Méthodologie et applications à l'Aide au Diagnostic Médical»;* Thèse de doctorat, Université Libre de Bruxelles, 2000.

[BEL08] BELER. Cédrick ; « *Modélisation générique d'un retour d'expérience cognitif : Application à la prévention des risques* » ; Thèse de doctorat, Université de Toulouse, France ; Soutenue le 14 Novembre 2008.

[BEN09] BENABBOU. Loubna; «Contribution à la classification supervisée multi-classes et multi-critères en aide à la décision » ; PhD thesis ; université de Laval, Québec, Canada ; 2009.

[BIS06] BISHOP. Cristopher M ; « *Pattern Recognition and machine learning* » ; Edition Springer ; 2006.

[BLO96] BLOCH. Isabelle; « *Information Combination Operators for Data Fusion: a comparative review with classification*»; IEEE Transaction on systems, man and cybernetics-part A: systems and humans; vol 26, n 01; January 1996.

[BLO04] BLOCH. Isabelle, MAÎTRE. Henri ; « les méthodes de raisonnement dans les images » ; école nationale supérieure des télécommunications, ENST ; Avril 2004.

[BLO05] BLOCH. Isabelle ; « *fusion d'informations numériques : panorama méthodologique* » ; Journées Nationales de la Recherche en Robotique, JNRR ; Guidel Morbihan, France ; pp 79-88 ; 5-7 octobre 2005.

[BOU06] BOUYSSOU. Didier. DUBOIS, PIRLOT. Marc, Henri. Prade; *«Concepts et méthodes pour l'aide à la décision 3, Analyse multicritère»;* Lavoisier, 2006.

[BOU07] BOUBOU. Mounzer; « *Contribution aux méthodes de classification non supervisée via des approches pré topologiques et d'agrégation d'opinions* » ; Thèse de doctorat, Université Claude Bernard, Lyon I, France ; Soutenue le 29 Novembre 2007.

[BOU12] BOULILA. Wadii ; « *Extraction de connaissances spatio-temporelles incertaines pour la prédiction de changements en imagerie satellitale* » ; thèse de doctorat ; université européenne de Bretagne ; Soutenue le 28 Juin 2012.

[BOU10] BOUKERMA. Hanane ; « *Combinaison de classifieurs flous pour la reconnaissance de l'écriture arabe manuscrite* » ; mémoire de magister ; université 20 aout 1955 Skikda, Algérie ; soutenu le 09 décembre 2010.

[BOU11] BOUKHAROUBA. Khaled ; «*Modélisation et classification des comportements dynamiques dans les systèmes hybrides* » ; thèse de doctorat, université de Lille 1, France ; Soutenue le 02 Juillet 2011.

[BOU13] BOUBERIMA. Wafia; «*modèles de mélange de Von Mises-Fisher*»; thèse de doctorat ; université Paris Descartes, France ; Soutenue le 15 novembre 2013.

[BOUG07] BOUGUEDID. Mohamed Saïd ; « *Contribution à l'application de la reconnaissance des formes et la théorie des possibilités au diagnostic adaptatif et prédictif des systèmes dynamiques*» ; Thèse de doctorat ; université de REIMS CHAMPAGNE ARDENNE, France ; Soutenue le 12 décembre 2007.

[BRA86] BRANS.J.P, VINCKE.Ph, MARESHAL.B; *«How to select and how to rank project: The PROMETHEE method»*, European Journal of Operational Research, vol 24, pp228-238, 1986.

[CAP06] CAPAROS. Mathieu ; « Analyse automatique des crises d'épilipsie du lobe temporel à partir des EEG de surface » ; thèse de doctorat ; Institut national polytechnique de Lorrain, Nancy, France ; Soutenue le 09 Octobre 2006.

[CAV12] CAVALCANTE. Aguilar ; « *Réseaux évidentiels pour la fusion de données multimodales hétérogènes, Application à la détection de chutes* » ; thèse de doctorat ; université d'Evry-val d'Essonne-Paris, France ; Soutenue le 22-10-2012.

[CHA95] CHAUVIN. Stéphane ; « *Evaluation des performances du modèle bayésien de fusion appliqué à l'imagerie satellitaire* » ; Quinzième colloque GREISI-Juan-les pins ; pp : 949-952 ; du 18 au 21 septembre 1995.

[CHA09] CHANDAKA. Suryannarayama, CHATTERJEE. Amitava, MUNSHI. Sugata ; « *Cross-Correlation aided support vector machine classifier for classification of EEG signals* » ; Expert Systems with Applications ; vol :36 ; pp :1329-1336 ; 2009.

[CHA11] CHATTOPADHYAY. Subhagata, PRATIHAR. Dilip Kumar, DESAKAR. Sunjib Chandra ; « *A comparative study of Fuzzy C-Means algorithm and entropy-based fuzzy clustering algorithms*» ; Journal of computing and Informatics, vol 30; pp 701-720;2011.

[CHE09] CHELLAKH. Hafida, MOUSSAOUI. Abdelouahab ; « *La segmentation des IRM cérébrales pathologiques par une combinaison floue possibiliste markovienne » ;* 1ière doctoriales STIC ; Université de M'sila, Algérie ; 07 et 08 décembre 2009.

[CHE12] CHERAGUI Mohamed Amine *«Une Analyse Morphologique de la Langue Arabe basées sur l'Aide Multicritère à la Décision»* ; CTIC, Université d'Adrar, Algérie, 21 Novembre 2012.

[CHI12] CHI. Yulun, LI. Haolin ; « *Simulation and analysis of grinding wheel based on Gaussian mixture model* »; International Journal Frontiers of Mechanical Engineering ; 7(4); pp 427-432; 2012.

[COR03] CORNEJOLS. Antoine, MICLET. Laurent, KODRATOFF. Yves ; « *Apprentissage artificiel : Concepts et Algorithmes* » ; Edition EYROLLES, France ; 2003.

[CUX11] CUXAC. Pascal, LAMIREL. Jean-Charles; «*Clustering incremental et methods de detection de nouveauté: Application à l'analyse intelligented'informations évoluant au cours du temps*»; Atelier CIDN ; 2011.

[DEM67] DEMPSTER.A.P; « *Upper and lower probabilities induced by multi-valued mapping*»; The Annals of Mathematical Statistics; vol38, N02; pp: 325-339; Apr 1967.

[DEN95] DENOEUX. Theirry; «*A k-nearest neighbor classification rule based on Dempster-Shafer Theory*»; IEEE Transactions on systems and cybernetics; vol 25, n°05; pp 804-813; May 1995.

[DEN97] DENOEUX. Theirry; «*Application du modèle des croyances transférables en reconnaissance des formes* » ;Traitement du signal ; vol 14, n°5 ;1997.

[DEN03] DENG.D, KASABOV.N ; « *On-Line pattern analysis by evolving self organizing maps*»; Neurocomputig; vol 55; pp87-103; 2003.

[DEN06] DENOEUX. Theirry, SMETS. Philippe; «*Classification using belief functions relationship between case-based and model-based approaches*»; IEEE Transactions on systems and cybernetics-Part B; vol 06, n°06; pp 1395-1406; December 2006.

[DES96] DESODT-LEBRUN. Anne-Marie ; « *Fusion de données* » ; Dossier Techniques de l'ingénieur ; 10-12 ;1996.

[DJE12] DJEFFAL. Abdelhamid ; « *Utilisation des méthodes Support Vector Machine (SVM) dans l'analyse des bases de données* » ; thèse de doctorat, Université Mahmoud KHIDER, Biskra, Algérie ; Année universitaire 2011-2012.

[DJI08] DJIKNAVORIAN. Pascal ; « *fusion d'information dans le cadre de raisonnement de Dezert-Smarandache appliqué sur des rapports de capteurs ESM sous le STANG 1241* » ; Mémoire de (M.sc) ; Université Laval, Québec ; 2008.

[DOU13] DOUMPOUS. Michael, GRIGOURDIS. Evangelos; *«Multicriteria decision aid. Links, theory and application»;* Wiley edition, 2013.

[DRO98] DROMIGNY-BADIN. Anne ; « *Fusion d'images par la théorie de l'évidence en vue d'applications médicales et industrielles* » ; Thèse de doctorat ; Institut National des sciences appliquées de Lyon ; France; Soutenue le 06 mai 1998.

[DOU04] DOUMPOS. Michael, Constantin. ZOPOUNIDIS; *«Multicriteria Decision Aid Classification Methods»;* Kluwer Academic Publishers, 2004.

[DUB88] DUBOIS. D, PRADE. H ; « *Possibility theory, an approach to the computerized processing of uncertainty*» ; Plenum Press; New-York; 1988.

[DUB99] DUBOIS. Didier, PRADE. Henri, YAGER. Ronald; «*Merging Fuzzy Information* »; Fuzzy Sets in Approximate Reasoning and Information Systems; Chapter 6; Kluwer Academic Publishers; pp 335-401; 1999.

[DUB01] DUBOIS. Didier, Jean Luc MARICHAL, Marc. ROUBENS, Régis.SABBADIN; *«The use of discrete Sugeno integral in decision-making : a survey»;* International Journal of uncertainty, Fyzziness and knowledge-base systems, vol 09, Issus 15. 2001.

[DUB06] DUBOIS. Didier, PRADE. Henri ; « *La théorie des possibilités* » ; REE ; vol 08 ; pp 42-50 ; 2006.

[DUM01] DUMAS. Jacky ; « *L'analyse temps-fréquence* » ; 01dB-Stell ; MVI technologies group ; Février 2001.

[EPI01] http://www.epileptologie bonn.de/cms/front_content.php?idcat=193&lang=3&changelang=3

[FAN02] FAN. Zhi-Ping, MA. Jian, ZHANG. Quan; «*An approach to multiple decision making based on fuzzy preference information on Alternatives»;* Fuzzy Sets and Systems; 131: 101-106. 2002.

[FRE97] FREIDMEIN. Jerome H. «*Data Mining and Statistics: What's the Connection*? »; 1997.

[FRE04] FRENOUX. Emanuelle ; « *Applications et validation de méthodologies de fusion de données en imagerie cérébrale » ;* thèse de doctorat ; 2004.

[FRI96] FRIGUI. Hichem, KRISHNAPURAM. Raghu; «*A robust algorithm for automatic extraction of an unkown number of clusters from noisy data*»; Pattern Recognition Letters; vol 17; pp: 1223-1232; 1996.

[GRA08] GRABISCH Michel, Ivan Kojadinovic, Patrick Meyer; *«A review of methods for capacity identification in choquet integral based multi-attribute utility theory, Application of the Kappalab R package»;* European Journal of operational research, vol 186: 766-785, 2008.

[GRA06] GRABISCH Michel; *«L'utilisation de l'intégrale de Choquet en aide multicritère à la décision»;* newsletter of the eurpean working group multicriteria aid for decisions ; 3(14), 2006.

[GRA96] GRABISCH.Michel, *«The application of fuzzy integrals in multicriteria decision making»;* European Journal of Operational Research, vol 89, 445-456, 1996.

[GRA06] GRANDIN. Jean-Franois ; « *Fusion de données – Théories et méthodes* » ; Dossier Techniques de l'ingénieur ; Systèmes et radars ; 10-03-2006.

[GUE09] GUENOUNOU Ouahib ; « *Méthodologie de conception de contrôleurs intelligents par l'approche génétique application à un bio-procédé* » ; thèse de doctorat ; université Paul Sabatier Toulouse III, France ; Soutenue le 22 avril 2009.

[GUO09] GUO. Ling, RIVERO. Daniel, SEOANE. Jose, PAZOS. Alejandro ; « *Classification of EEG signals using relative wavelet energy and artificial neural network* » ; Genetic and evolutionary Computation GEC; Shanghai, China; June 12-14, 2009.

[HAD14] HADDAD. Mohamed Tahar ; « *Anticipation des crises d'épilepsie temporels combinant des méthodes statistiques et non-linéaire d'analyse d'électroencéphalographie* » ; thèse de doctorat, Université de Québec; 2014.

[HAM08] HAMOU MAMAR. Zahra ; « *Analyse Temps-Echelle et reconnaissance des formes pour le diagnostic du système de guidage d'un tramway sur pneumatiques* »; thèse de doctorat, université de Blaise Pascal de Clermont II France ; Soutenue le 18.07.2008.

[HAR79] HARTIGANT. J. A, WONG. M. A ; « *A k-means clustering algorithm* » ; Journal of the royal statistical society. Series C (Applied statistics); pp: 99-108; 1979.

[HAR10] HARTERT. Laurent ; « *Reconnaissance des formes dans un environnement dynamique appliquée au diagnostic et au suivi des systèmes évolutifs* » ; thèse de

doctorat ; université de REIMS CHAMPAGNE ARDENNE, France ; Soutenue le 24 Novembre 2010.

[HEG04] HEGARAT-MASCLE. S. Le, SELTZ. R; «*Automatic change detection by evidential fusion of change indices*» ; Remote Sensing of Environment 91; pp:390-404; 2004.

[HER00] HERNIET Laurent, Thèse de doctorat, *«Systèmes d'Evaluation et de classification Multicritères pour l'Aide à la decision, Construction de Modèles et Procédures d'Affectation»;* Université PARIS DAUPHINE, Soutenue le 25 Janvier 2000.

[JOL03] JOMMOIS. François Xavier ; « *contribution à la classification automatique à la fouille de données* » ; thèse de doctorat ; université de Metz, France ; soutenue le 12 décembre 2003.

[HOC12] HICEIPED. Gatien ; « *Détection précoce de crises d'épilepsie à l'aide d'une modélisation du comportement oscillatoire neuronal* » ; thèse de doctorat ; école polytechnique de Bruxelles ; 2012.

[JAS58] JASPER. H. H ; « *the Ten-Twenty electrode system of the international Federation* » ; Electroencephalogram Clinical Neurophysiology ; vol 10 ; pp :367-380 ; 1985.

[JOS12] JOSE RAMON SAN CRISTOBAL MATEO; *«Multi-criteria Analysis in the Renewable Energy»;* British Library Cataloguing in Publication Data , springer, 2012.

[KEL85] KELLER. James M, GRAY.Michael R, GIVENS JR. James A ; « *AFuzzy K-Nearest Neighbor Algorithm* » ; IEEE Transaactions on systems, Man and Cybernetics, Vol15, No 4, PP 580-585, July/August 1985.

[KHO98] KHODJA. Lotfi; « *Contribution à la classification floue non supervisée*»; Thèse de doctorat ; Université de Savoie, France ; Soutenue le 21 décembre 1998.

[KIT98] KITTLER. Josef, HATEF. Mohamad, ROBERT. P.W. Duin, MATAS. Jiri ; « *On combining Classifiers* » ; IEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE ; Vol : 20, No : 3 ; pp : 226-239 ; MARCH 1998.

[KOC99] KOCZKODAJ.W.W, ORLOWSKI.M: *«Computing a consistent approximation to generalized pairwise comparison matrix»*; computers and Mathematics with Applications, Vol 37, 79-85, 1999.

[KOV14] KOVACS. Péter, Samiee. Kaveh ; GABBOUDJ. Moncef ; « *On Application of rational Discret Short Time Fourier Transform in epileptic seizure classification* » ; IEEE International on Acoustic and signal processing (ICASSP); pp: 5839-5843; 2014.

[KRI96] KRISHNAPURAM. Raghu, KELLER. James M; «*The possibilistic C-Means Algorithm: Insights and recommendations*»; IEEE Transactions on fuzzy systems, vol 4, n°3; August 1996.

[KUN02-a] KUNCHEVA. Ludmila I; «*A theoretical study on six classifiers fusion strategies*»; IEEE of transactions on pattern analysis and machine intelligence; vol 24, n02; pp 281-286; Febrauary 2002.

[KUN02-b] KUNCHEVA. Ludmila I; « *Switching between selection and fusion in combining classifiers: An Experiment*»; IEEE of transactions on systems, MAN and CYBERNETICS-Part B; vol 32, n02; pp 146-156; Avril 2002.

[LAM13] LAMICHE. Chaabane ; « *Fusion et fouille de données guidées par les connaissances : Application à l'analyse d'images* » ; thèse de doctorat ; université de Biskra, Algérie ; Soutenue le 16 juin 20013.

[LEC05] LECOMTE. Gwenaèl ; « *Analyse d'images radioscopiques et fusion d'informations multimodales pour l'amélioration de du contrôle de pièces de fonderie* » ; Thèse de Doctorat ; présentée devant l'institut national des sciences appliquées de Lyon, France ;soutenue le 20 décembre 2005.

[LEE14] LEE. Sang-Hong, LIM. Joon S, KIM. Jae-Kwon, Yang Junggi, LEE. Youngho; «*Classification of normal and epileptic seizure EEG signals using wavelet transform; phase- space reconstruction, and Euclidien diastance*»; Computer methods and programs in biomedicine; vol 116; pp 10-25; 2014.

[LEH07] LEHOUELLEUR. Jaques ; « *le potentiel d'action* » ; Cours neurosciences & comportement, 2ième partie : Neurologie moléculaire ; 2007.

[LEI92] LEI. Xu, XRZYAZAK. Adam, CHING Y. Suen ; « *Methods of combining multiple classifiers and their application to handwriting recognition*»; IEEE Transactions on systems, Man and Cybernetics; vol 22, N 03; pp: 418-435; May-June 1992.

[LEL13] LELANDAIS. Benoît ; « *fusion d'informations et segmentation d'images basées sur la théorie des fonctions de croyance : Application à l'imagerie médicale TEP multi-traceurs*» ; thèse de doctorat, université de Rouen, France ; soutenue le 23 avril 2013.

[LI07] LI. Hui-Fan, WANG Jian-Jun ; «An *improved ranking method for ELECTRE III*»; International conference on wireless communications, networking and mobile computing, IEEE WICOM 2007, Shanghai, China, pp 6659-6662, September 20-21, 2007.

[LIM04] LIMAYEM. Frej, YANNOU.B, *«Generalization of the RCGM and LSLR Pairwise Comparaison Methods»;* An international Journal of Computers & Mathematics with applications, (48): 539-548, 2004.

[LIM01] LIMAYAM. F ; « *Modèles de pondération par les méthodes de tri croisé pour l'aide à la décision collaborative en projet » ;* PhD thesis, Ecole centrale paris, France, Soutenue le 23 Novembre 2001.

[LON11] LONCA. Emmanuel; *«Utilisation de méthodes d'agrégation Multicritère appliqués à la gestion de dépendances»;* Thèse de Master de Recherche, Université D'Atrois de France, soutenue le 05 Juillet 2011.

[LOT08] Lotte. Fabien ; « *Study of electroencephalographic signal processing and classification techniques towards the use of Brain-Computer Interface in virtual reality applications* » ; thèse de doctorat ; Institut national des sciences appliquées de Rennes, France ; Soutenue le 04 Décembre 2008.

[LUR03] LURETTE. Cristophe ; « *développement d'une technique neuronale auto-adaptative pour la classification dynamique de données évolutives. Application à la supervision d'une presse hydraulique* » ; Thèse de doctorat, université des sciences et technologies de LILLE, France ; Soutenue le 17 septembre 2003.

[MAC10] MACHADO. Alexis; « *Etude des pointes épileptiques par acquisition simultanée en spectroscopie proche infrarouge et électroencéphalographie* » ; mémoire de maitrise ; université de Montréal ; Août 2010.

[MAR00] MARICHAL. Jean-Luc; *«On Sugeno Integral as an aggregation function, Fuzzy sets and Systems»;* vol 114, pp 347-365, 200.

[MAR05] MARTIN. Arnaud ; « *La fusion d'informations* » ; Polycopié de cours ENSEITA-Réf : 1484 ; Janvier 2005.

[MAU04] MAUCLAIR. Julie, PINQUIER. Julien ; « *Fusion of descriptors for speech / music classification*» ; In European Signal Processing Conference EUSIPCO; pp: 1285-1288; Vienna; Austria; September 2004.

[MAY94] MAYSTRE. Lucien Yves, Jacques. PICTEC, Jean. SIMOS ;*«Méthode Multicritères ELECTRE. Description, Conseils pratique et d'application à la gestion environnementale»;* Presses Polytechniques et Universitaires ROMANDES, Lausanne, SUISSE, 1994.

[MER06] MERCERE. Guillaume, BAKO. Laurent, AMADOU BOUBACAR. Habib, LECOEUCHE. Stephane ; « Suivi de systèmes non stationnaires basé sur une

approche de reconnaissance des formes » ; Conférence Internationale Francophone d'Automatique (CIF06) ; Bourdeaux, France, 30 Mai 2006.

[MIL03] MILISAVLEVIC. Nada, BLOCH. Isebelle, BROEK. Sebastiaan Van Den, ACHEROY. Marc ; « *Improving mine recognition through processing and Dempster-Shafer fusion of ground penetrating data*»; the journal of the pattern recognition society; vol 36, pp: 1233-1250; 2003.

[MIN06] MINGOTI. Sueli A, LIMA. Joab O; « *Comparing SOM neural network with fuzzy C-Means, K-Means and traditional hierarchical clustering algorithms* »; European Journal of Operational Research; Vol 174; pp 1742-1759; 2006.

[MOH06] MOHAMADALLY. Hasan, FORMANI.Boris ; « SVM : *Machines à Vecteurs de Support ou Séparateurs à Vastes Marges* » ; Université Varseilles St Quentin, France ; 16 Janvier 2006.

[MOU00] MOUSSEAU.V, SLOWINSKI. R, ZIELNIEWICZ.P; «*A user-oriented implementation of the ELECTRE-TRI method integrating preference elicitation support*»; Computers & Operations Research, (27): 757-777, 2000.

[MRA09] MRABTI. Fatima ; SERIDI Hamid ; « Comparaison des méthodes de classification réseau RBF, MLP et RVFLNN » ; Damascus University Journal ; vol 25, n 02 ; pp 119-129 ; 2009.

[MUR13] MURUGAPPAN. M ; MURUGAPPAN. Subbulakshmi; «*Humane motion recognition through short time electroencephalogram (EEG) signals using Fast Fourier Transform (FFT)* »; IEEE 9th International colloquium on Signal Processing and its Applications, Kuala Lumpur-Malaysia; 8-10 May. 2013.

[NAC09] NACERDDINE Nafâa, TABBONE Salvatore, ZIOU Djemel, HAMAMI Latifa; « *L'algorithme EM et modèle de mélange de gaussiennes généralisées pour la segmentation d'images, Application au contrôles des joints soudés par radiographie* » ; Conférence de Traitement et Analyse de l'Information : Méthodes et Applications (TAIMA) ; Hammamat, Tunisie ; de 04 au 09 mai 2009.

[NEA98] NEAL. Radford M, HINTON. Geoffrey E; « *A view of the EM algorithm that justifies incremental, Sparse, and other variants*» ; in M. I. Jordan (eds) Learning in Graphical Models Dordrecht: Kluwer Academic, pp:355-368.

[NOA99] NOOCHTAR. S, BINNIE. C, EBERSOLE. J, MAUGIERE. F, SCKAMOTO. A, WESTMORELAND. B ; « *A glossary of terms most commonly used by clinical*

electroencephalographers and proposal for the report form for the EEG findings»; Recommendation for the practice of clinical neurophysiology; chapter1.5; 1999.

[OLT08] OLTEANU. Ana-Maria ; « *fusion de connaissances imparfaites pour l'appariement de données géographiques, proposition d'une approche s'appuyant sur la théorie des fonctions de croyance* » ; thèse de doctorat, université Paris-Est, France ; Soutenue le 24 Octobre 2008.

[ORH11] ORHAN. Umut, HEKIM. Mahmut ; OZER. Mahmut ; « *EEG signals classification using K-means clustering and multi-layer perspectron neural network model*»; Expert systems with applications; vol 38; pp: 13475-13481; 2011.

[OUA12] OUAHABI. Abdeldjalil ; « *Analyse multi-résolution pour le signal et l'image* » ; édition LAVOISIER ; 2012.

[OUA09] OUAKKA. Hamid; «*Contribution à l'identification et la commande floue d'une classe de systèmes non linéaires* »; Thèse de doctorat, Université Sidi Mohamed Ben Abdellah, FES, Maroc, Soutenue le 27 Juillet 2009.

[OUK97] OUKHELLOU. Latifa ; « Paramétrisation des signaux et clasification de signaux en contrôle non destructif : Application à la reconnaissance des défauts de rails par courants de Faucault » ; thèse de doctorat, université de Paris-sud ; Soutenue le 04 Juillet 1997.

[OUS01] OUSSALAH. Mourad, MAAREF. Hichem, BARRET. Claude; « *New fusion methodology approach and application to mobile robotics: investigation in framework of possibility theory* »; Information Fusion; vol 02; pp: 31-48; 2001.

[PAA06] PAALANEN. Pekka, KAMAKAINEN. Joni Kritian, ILONEN. Jarmo ; « *Feature representation and discrimination based on Gaussian Mixture Model probability densities-Practices and Algorithms*»; The Journal of the pattern recognition society, vol 39; pp 1346-1358; 2006.

[PAG11] PAGANI. Clara, SADAOUI. Akim, CHABOT. Simon ; « *Introduction à la logique floue* » ; Université de Technologie Compiègne (UTC), France ; Automne 2011.

[PAR14] PARVEZ. Mohammad Zavid, Manoranjan Paul; «*Epileptic seizure detection by analyzing EEG signals using different transformation technique*»; Neurocomputing, vol 145; pp 191-200; 2014.

[PER04] PERRIN. Stéphane, DUFLOS. Emmanuel, VANHEGHE. Philippe, BIBAUT. Alain ; «*Multisensor Fusion in the Frame of Evidence Theory for Landmines Detection*»;

IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS-PART C: APPLICATIONS AND REVIEWS, VOL. 34, NO. 4; NOVEMBER 2004.

[RAJ15] RAJEDRAACHARYA. U, FUJITA. H, SUDERSHAN. V, BHAT. SKOHN JOE. E.W ; « *Application of entropies for automated diagnosis of epilepsy using EEG signals :A review* » ; Knowledge-Based Systems ; vol 88 ; pp :85-96 ; August 2015.

[RAK07] RAKUS-ANDERSON Elisabeth, JOUGREUS-Claes, *«The Choquet and Sugeno Integrals measures of total effectiveness of medicines»;* Theoretical Advances and Applications of Fuzzy Logic and Soft Computing, Advances in Soft Computing, vol 42, pp 253-262, 2007.

[RIC08] RICQUEBOURG. Vincent, DURAN. David, MENGA. David, DELAHOCHE. Laurent, MARHIC. Bruno, LOGE. Christophe, JOLLY-DESODT. Anne-Marie; «*La fusion dans l'habitat communicant: une approche non probabiliste*»;In Proceedings of the 4th French-speaking conference on Mobility and ubiquity computing (UbiMob '08); ACM; pp 9-16; May 28-30, 2008.

[ROG00] ROGERS. Martin, BRUEN. Michael, MAYSTRE. Lucient-Yves, *« ELECTRE and decision support: Methods and Application in engineering and infrastructure Investment» ;* springer science et business media, LLC, 1999.

[ROL12] ROLLAND Antoine, *«Aide à la Décision Multicritère et Apprentissage Automatique pour la Classification»;* Laboratoire ERIC-Université Lumière Lyon II, 2012.

[ROM10] ROMO VAZQUEZ. Rebeca del Carmen; « *Contribution à la détection et à l'analyse des signaux EEG épileptiques : débruitage et séparation de sources* » ; Thèse de doctorat, université de Nancy-France- ; soutenue le 24 février 2010.

[ROY93] ROY.Bernard, BOUYSSSOU. Denis; «*Aide multicritère à la décision: Méthodes et cas » ;* Economica, 1993.

[SAA77] SAATY. L Thomas; *«A scaling for priorities in hierarchical structures»;* Journal of mathematical psychology; University of pensilvania, Wharton school, Philadelphia, Pensilvania 19174; vol 15; 234-281, 1977.

[SAL12] SALPERWYCK. Cristophe ; « *Apprentissage incrémental en-ligne sur flux de données* », Thèse de doctorat, université de LILLE, France ; Soutenue le 30 Novembre 2012.

[SAM15] SAMIEE. Kaveh, KOVACS. Péter, GABBOUDJ. Moncef ; « *Epileptic-seizure classification of EEG time series using rational discret Short-Time Fourier*

Transform»; IEEE Transactions of Biomedical Engineering; vol 62 n:02; pp: 541-551; February 2015.

[SAY14] SAYED MOUCHAWECH. Moamar ; « *Apprentissage dynamique dans un environnement non-stationnaire* » Article techniques de l'ingénieur h3125 ; publié le : 10/08/2014.

[SEN02] SENTZ. Kari, FERSON. Scott; «*Combinaison of Evidence in Dempster-Shafer Theory*»; SAND 2002-0835; April 2002.

[SHA76] SAHAFER.G; « *A Mathematical Theory of evidence* »; Princeton University ;1976.

[SIM93] SIMPSON. Patrick K; «Fuzzy *Min-Max neural network-part2: Clustering*»; IEEE transactions on fuzzy systems; vol 1, n 1; pp32-45; February 1993.

[SIU12] SIULY; «*Analysis and Classification of EEG Signals* » ; PhD Thesis ; University of Southern Queensland-Australia; July 2012.

[SME90] SMETS. Philippe; « *the combinaison of evidence in the transferable belief model*»; IEEE Transactions on Pattern Analysis and Machine Intelligence; vol 05; N°05; pp447-458; 1990.

[SUB07] SUBASI. Abdelhamit; «*EEG signal classification using wavelet feature extraction and a mixture of expert model*»; Expert Systems with applications; vol 32; pp:1084-1093; 2007.

[THI15] THIEU. Thao Nguyen, YANG. Hyung-Jeong, « *Diagnosis of epilepsy in patients based on the classification of EEG signals using Fast Fourier Transform*»; Current approaches in applied artificial intelligence; pp: 493-500;2015.

[TOUI10] TOUIL. Walid; « *Mise en œuvre d'une technique mixte de cartographie-RNA pour l'analyse des signaux EEG quantifies* » ; Mémoire de Magister, Université de M'sila, Algérie ; 2010.

[TUF07] TUFFERY. Stéphane; "*Data Mining et Statistique décisionnelle, L'intelligence des données*" ; Editions TECHNIP ; Paris, France ; 2007.

[VAP95] VAPNIK. Vladimir, CORTES. Corina ; « *Support Vector Network* »; Journal of Machine Learning; vol 20; pp 273-297; 1995.

[VAV10] VAVADI. Hamed, AYATOLLAHI. Ahmad ; MIRZAE. I. Ahmad; « *Wavelet-Approximate entropy method for epileptic activity detection from EEG and its sub-bands*»; Journal of Biomedical science and Engineering; vol 03; pp:1182-1189; 2010.

[VIN89] VINCKE. Philipe, Préface de Bernard. ROY ; *«L'aide Multicritère à la Décision»;* Edition de l'université de Bruxelles et Editions Ellipses, Collection Statistiques et Mathématiques, 1989.

[XU92] XU. Lei, KRZYZAK. Adam, SUEN. Ching Y; «*Methods of combining multiple classifiers and their applications to hand writing recognition*»; IEEE of Transactions on systems, MAN and CYBERNETICS; vol 22, n°03; pp: 418-435; May/June 1992.

[ZAD65] ZADEH. L. A; « *Fuzzy sets* »; Information and control; vol8; pp: 338-353; 1965.

[ZAD78] ZADEH. L. A; « *Fuzzy sets as a basis for theory of possibility*»; International of fuzzy sets and systems; vol 01; pp 03-28; 1978.

[ZOP02] ZOPOUNIDIS Constantin, DOUMPOS Michael; *«Multicriteria Classification and Sorting methods: A literature review»; European Journal of Operational Research,* (138): 229-246, 2002.

[ZOU08] ZOUAOUI. Hakima ; « *Clustering par fusion de données appliqué à la segmentation d'images IRM* » ; Mémoire de magister ; Université M'hamed BOUGARA de Boumerdès, Algérie ; 2008.

Annexe : les méthodes de classification multicritère dans le cadre de tri nominal

Annexe :

Dans [BEL00], l'auteur a développé trois procédures en contexte de classification multicritère (PROCTN, PROAFTN et PROCFTN). Les algorithmes de classification multicritère présentés dans le premier chapitre se limitent par les classes ordinales, où les classes de domaine de classification se présentent par les actions de référence limites. En particulier, dans le problème de diagnostic par l'EléctroEncéphaloGraphie sera traité comme une problématique de tri nominal puisqu'on ne peut pas représenter les classes par les actions de référence limites. Pour cela, on utilise les procédures de tri multicritère nominal.

PROCTN :

Le principe de l'algorithme *PROCTN (PROcédure de Choix dans le cadre de Tri nominal)* est de déterminer un sous-ensemble aussi réduit que possible d'actions de référence centrale qui ont le meilleur écart avec l'action *a* à affecter. A partir de ce sous-ensemble d'actions de référence centrale la décision concernant l'affectation d'une action à une catégorie prise en utilisant la règle majoritaire *comme celle utilisée dans la méthode des k-Plus Proche Voisins*. La méthode PROCTN est exécutée en deux étapes comme toute méthode de classification multicritère à base de relation de sur-classement. L'étape une est la construction de la relation de sur-classement et la deuxième est l'exploitation de cette relation de sur-classement et affecter chaque action à sa catégorie.

PROAFTN : (PROcédure d'Affectation Floue dans le cadre de la problématique de Tri Nominal) :

Dans notre étude, nous avons focalisé d'utiliser les méthodes de classification flous. Pour cela, nous avons présenté en détaille la procédure de classification multicritère floue PROAFTN.

1- Construction de la relation de sur-classement I:

Comme on le voit dans la procédure PROCTN, cette procédure résoudre les problèmes de classification multicritère dans le cadre de tri nominal où les actions de référence centrales

b_i^h de l'ensemble *B* sont généralement données sous forme d'intervalle $\left[S_j^1(b_i^h), S_j^2(b_i^h)\right]$. En plus, dans cette procédure l'auteur a introduit deux seuils de discrimination ou seuils d'appartenance à l'intervalle $\left[S_j^1(b_i^h), S_j^2(b_i^h)\right]$. Ces seuils sont appelées seuil de préférence $p_j(b_i^h)$ et seuil d'indifférence $q_j(b_i^h)$. plus de détail, voir premier chapitre.

- **calcul de l'indice d'indifférence partiel $C_j(a, b_i^h)$:**

Dans cette section, on définir les relations d'indifférence partielles floues. Formellement, cette relation est définie comme suit :

$C_j : AxB \rightarrow [0,1],\ j = 1..n.$

Où : *A* : ensemble d'action à affecter

B : l'ensemble d'actions de référence centrales.

Chaque relation représente le degré de crédibilité de la situation suivante :

« *L'action a est indifférente à l'action de référence* b_i^h *selon le critère j»*

Formellement, l'utilisation des deux seuils de discrimination permet d'obtenir trois situations comparatives des actions *a* à affecter et b_i^h selon un critère *j* :

$$\textbf{Si}\ \ S_j^1(b_i^h) \leq g_j(a) \leq S_j^2(b_i^h),\ \ \textbf{alors}\ a\ est\ indifférente\ b_i^h, on\ note\ \ \boldsymbol{aIb_i^h}$$

Dans ce cas l'indice d'indifférence partiel $\boldsymbol{C_j(a, b_i^h) = 1.}$

Si $\left[S_j^1(b_i^h) - q_j(b_i^h) < g_j(a) < S_j^1(b_i^h)\right]$ ou $\left[S_j^2(b_i^h) < g_j(a) < S_j^2(b_i^h) + p_j(b_i^h)\right]$ alors :

$on\ a\ une\ indifférence\ faible\ entre\ a\ et\ b_i^h$, et $\boldsymbol{0 < C_j(a, b_i^h) < 1}, on\ note\ \ \boldsymbol{aQb_i^h}.$

Si $\left[g_j(a) \leq S_j^1(b_i^h) - q_j(b_i^h)\right]$ ou $\left[g_j(a) \geq S_j^2(b_i^h) + p_j(b_i^h)\right]$ alors : *a*

$n'estpas\ indifférente$ à b_i^h

Dans ce cas : $C_j(a, b_i^h) = 0$.

L'indice d'indifférence partiel est représenté par une fonction d'interpolation linéaire, dans deuxième cas où : $\boldsymbol{0 < C_j(a, b_i^h) < 1}$, cet indice est calculé comme suit :

$$C_j(a, b_i^h) = \frac{g_j(a) + q_j(b_i^h) - S_j^1(b_i^h)}{q_j(b_i^h)} \quad si \quad S_j^1(b_i^h) - q_j(b_i^h) < g_j(a) < S_j^1(b_i^h)$$

$$C_j(a, b_i^h) = \frac{S_j^2(b_i^h) + p_j(b_i^h) - g_j(a)}{p_j(b_i^h)} \quad si \quad S_j^2(b_i^h) < g_j(a) < S_j^2(b_i^h) + p_j(b_i^h)$$

- Calcul de l'indice d'indifférence globale :

Les relations d'indifférence permettent de déterminer le degré de crédibilité global, associé à la relation d'indifférence *I :*

Le degré de crédibilité $C(a, b_i^h)$ global est déterminé comme suit :

$$C(a, b_i^h) = \sum_{j=1}^{n} w_j^h . C_j(b_i^h).$$

Où : w_j^h est le poids de la classe h selon le critère *j*.

- Calcul de la relation d'indifférence de synthèse :

La relation d'indifférence de synthèse est fondée sur la règle de concordance et de non-discordance. Le principe général de cette règle est le suivant :

lorsqu'une action a est jugée indifférente avec une action de référence b_i^h selon une majorité suffisamment importante de critères «principe majoritaire » et qu'il n'existe aucun critère qui met son veto contre l'affirmation "a est indifférente à b_i^h" «principe du respect des minorités », alors l'action a est indifférente à l'action de référence b_i^h.

Par définition nous appellerons seuil de veto à droite $v_j^+(b_i^h)$ (resp. seuil de veto à gauche $\left(v_j^-(b_i^h)\right)$pour le critère g_j la valeur minimum de la différence $\left(g_j(a) - S_j^2(b_i^h)\right)$ (resp. $\left(S_j^2(b_i^h) - g_j(a)\right)$ qui est considérée comme étant incompatible avec la proposition aIb_i^h.

Les seuils de veto $v_j^+(b_i^h)$ et $v_j^-(b_i^h)$ doivent vérifier les conditions de cohésion suivantes :

$$\forall j \in \{1, \dots, n\}, \qquad v_j^-(b_i^h) \geq (b_i^h) \;\; p_j \, et \;\; v_j^-(b_i^h) \geq q_j(b_i^h)$$

Les deux indices de veto sont utilisés pour que l'action a est considérée non indifférente que l'action de référence centrale $\left(b_i^h\right)$ pour le critère g_j. D'après ce qui précède, il découle que :

$\exists j \in 1 \ldots n \; tel \; que$:

$$\left.\begin{array}{c} g_j(a) \geq S_j^2\left(b_i^h\right) + v_j^+\left(b_i^h\right) \\ ou \\ g_j(a) \leq S_j^2\left(b_i^h\right) - v_j^-\left(b_i^h\right) \end{array}\right\} \Rightarrow \neg\left(a \; I \; b_i^h\right)$$

- Indices de discordance :

  Indices de discordance partiels :

L'indice de discordance $D_j\left(a, b_i^h\right)$ du critère g_j vise à appréhender le fait que ce critère est plus au moins discordant avec la proposition $a \; I \; b_i^h$. cette discordance est maximum $\left(D_j\left(a, b_i^h\right) = 1\right)$ lorsque le critère g_j met son veto à l'indifférence. Elle est minimum $\left(D_j\left(a, b_i^h\right) = 0\right)$ lorsque le critère n'est pas en discordance avec l'indifférence. Si le critère g_j est en discordance $\left(C_j\left(a, b_i^h\right) = 0\right)$ avec l'indifférence mais qu'il ne met son veto à l'indifférence, alors on aura : $\left(0 < D_j\left(a, b_i^h\right) < 1\right)$, qui représente les zones intermédiaires entre la discordance et la non-discordance.

L'indice de discordance partiel $D_j\left(a, b_i^h\right)$ est généralement représenté entre les valeurs $S_j^2\left(b_i^h\right) + p_j\left(b_i^h\right)$ et $S_j^2\left(b_i^h\right) + v_j^+\left(b_i^h\right)$ d'une part et $S_j^1\left(b_i^h\right) - q_j\left(b_i^h\right)$ à $S_j^1\left(b_i^h\right) - v_j^-\left(b_i^h\right)$ d'autre part, par une fonction d'interpolation linéaire.

L'indice de discordance est composé de deux indices l'un à gauche $D_j^-\left(a, b_i^h\right)$ et l'autre à droite $D_j^+\left(a, b_i^h\right)$. D'après ce qui procède, les formules des indices de discordances sont représentées comme suit :

$$D_j^-\left(a, b_i^h\right) = \frac{g_j(a) - max\left\{g_j(a), S_j^1\left(b_i^h\right) - q_j\left(b_i^h\right)\right\}}{q_j\left(b_i^h\right) - max\left\{S_j^1\left(b_i^h\right) - g_j(a), v_j^-\left(b_i^h\right)\right\}}$$

$$D_j^+\left(a, b_i^h\right) = \frac{g_j(a) - min\left\{g_j(a), S_j^2\left(b_i^h\right) + p_j\left(b_i^h\right)\right\}}{-p_j\left(b_i^h\right) + max\left\{-S_j^2\left(b_i^h\right) + g_j(a), v_j^+\left(b_i^h\right)\right\}}$$

L'indice de discordance partiel est donnée par :

$$D_j\left(a, b_i^h\right) = max\left\{D_j^-\left(a, b_i^h\right), D_j^+\left(a, b_i^h\right)\right\}$$

Indice de discordance global :

L'indice de discordance global $D_I\left(a, b_i^h\right)$ dans la méthode PROATN est défini comme celle de la méthode ELECTRE Tri, cet indice est appelée relation de discordance globale avec la relation d'indifférence ***I.*** il est défini comme suit :

$$D_I\left(a, b_i^h\right) = 1 - \prod_{j=1}^{n}\left(1 - D_j\left(a, b_i^h\right)\right)^{w_j^h}$$

- Construction de la relation d'indifférence de synthèse :

Cette relation est une relation binaire floue synthétique d'indifférence notée *I* à partir d'une agrégation d'indices de concordance et de discordance, elle sera définie comme suit :

$$I\left(a, b_i^h\right) = \left(\sum_{j=1}^{n} w_j^h . C_j\left(a, b_i^h\right)\right) \mathrm{x} \left(\prod_{j=1}^{n}\left(1 - D_j\left(a, b_i^h\right)\right)^{w_j^h}\right)$$

2- Affectation des actions aux différentes catégories :

L'approche de classification PROAFTN affecte graduellement (affectation floue) les actions aux différentes catégories. A partir du degré d'appartenance floue d'une action a à la classe $C^h, h = 1, \ldots, k$, l'affectation nette d'une action a se fait par la règle suivante :

$$a \in C^h \quad \Leftrightarrow \quad I\left(a, b_i^h\right) = max\left\{I\left(a, b_i^h\right)\right\}$$

PROCFTN :

PROCFTN (PROcédure de Choix Flou dans le cadre de problématique du Tri Nominal) combine le principe de choix utilisé par la procédure *PROCTN* et les indices d'indifférence flous déterminés par la procédure *PROAFTN*. Pour déterminer les prototypes plus ressemblants à l'action *a.*