

Clustering par fusion floue de données appliqué à la segmentation d'images IRM cérébrales

Zouaoui Hakima¹, Moussaoui Abdelouahab²

¹Ecole doctorale nationale STIC Département d'Informatique -Université M'hamed Bougra de Boumerdes

²Ecole doctorale nationale STIC Département d'Informatique - Université Ferhat Abbas de Sétif

Hak-soraya@yahoo.fr Moussaoui_abdel@yahoo.fr

Résumé. Dans cet article nous proposons une approche originale basée sur les techniques de data mining pour l'extraction des connaissances par fusion floue de données. Cette approche présente une nouvelle architecture de fusion de données basée sur la théorie possibiliste pour la segmentation d'une cible à partir de plusieurs sources d'images. Le processus de fusion est décomposé en trois phases fondamentales : nous modélisons tout d'abord les informations dans un cadre théorique commun, le formalisme retenu consiste à faire la coopération entre l'algorithme FCM (Fuzzy C-Means) dont la contrainte d'appartenance d'un individu à une classe est gérée d'une manière relative et l'algorithme possibiliste PCM (Possibilistic C-means) pour les points aberrants. Nous agrégeons ensuite ces différentes informations par un opérateur de fusion, celui-ci doit affirmer les redondances, gérer les complémentarités et prendre en compte les conflits soulignant souvent la présence d'une pathologie. Nous construisons enfin une information synthétique permettant d'exploiter les résultats de la fusion.

Mots-clés. Data mining, Imagerie médicale, Fusion de données, Segmentation, C-moyennes floues, C-moyennes possibiliste.

1 Introduction

Malgré les avancées technologiques dans le domaine médical et en particulier dans l'imagerie médicale, l'étude du cerveau humain est un problème ardu et reste un sujet de recherche de forte actualité, du fait de l'évolution des techniques d'imagerie, la grande capacité de calcul des machines...etc. Une des modalités d'imagerie les plus couramment utilisées, à laquelle nous nous sommes intéressés dans le cadre de ce travail, est l'Imagerie par Résonance Magnétique (IRM), qui est devenue un outil indispensable pour tout examen clinique, elle présente l'avantage d'être non invasive et permet l'acquisition d'images bi ou tridimensionnelles sur lesquelles différents contrastes sont possibles. Cette modalité permet en outre la réalisation d'études du cerveau aussi bien anatomiques que fonctionnelles grâce aux techniques d'IRM et d'IRMf (IRM fonctionnelle) [1]

Ces dernières années, des modélisations formelles de cette attitude ont été construites, fondées pour la plupart sur des approches prenant en compte les redondances, les complémentarités et les ambiguïtés inhérentes aux données médicales. Regroupées sous l'appellation "*fusion*", ces modèles ont pour but de gérer au mieux ces différents aspects pour faire converger les connaissances et proposer une information synthétique la plus exploitable possible [2].

Notre travail concerne essentiellement le développement de nouveaux outils issues des techniques de data mining pour l'extraction des connaissances par fusion floue de données. Il s'agit essentiellement de contribuer au développement de systèmes de classification guidés par les connaissances *a priori* où l'aspect flou et possibiliste sont pris en considération lors du processus de classification. Afin de valider les algorithmes développés, une application a été développée pour la segmentation des images IRM.

Le processus de fusion tel que nous l'envisageons ici est composé de trois étapes. Dans la première, les informations disponibles sont modélisées dans un cadre théorique commun, permettant de prendre en compte les connaissances vagues et ambiguës. Dans la seconde, les modèles d'informations sont agrégés, en tenant compte des redondances et des conflits exprimés. Dans la troisième, enfin, une décision est prise en fonction de toutes les informations précédemment fusionnées [3].

Nous présentons à la deuxième section un état de l'art sur les méthodes de classification des tissus cérébraux. La fusion de données fait l'objet de la troisième section. Où nous proposons une étude des méthodes d'agrégation des informations. Dans la quatrième section, nous présentons les étapes de la fusion de données de l'approche proposée en réponse à la problématique de l'extraction des connaissances par fusion floue de données. Les résultats obtenus sur les images IRM sont présentés dans la cinquième section. Nous concluons au sixième paragraphe.

2 Méthodes de classification

Les problèmes de classification s'attachent à déterminer des procédures permettant d'associer une classe à un objet (individu). Ces problèmes se déclinent essentiellement en deux variantes selon Bezdek [4] : la classification dite supervisée et la classification dite non supervisée (automatique).

2.1 Méthodes supervisées

Ces sont des méthodes dans lesquelles les classes sont connues *a priori* avant d'effectuer l'opération d'identification des éléments de l'image. Elles demandent une première phase d'apprentissage sur l'échantillon représentatif dans le but d'apprendre les caractéristiques de chaque classe et une deuxième phase pour décider de l'appartenance d'un individu à telle ou telle classe. Les données segmentées de l'ensemble d'apprentissage proviennent d'un étiquetage manuel des images ou des régions d'intérêt en C classes de tissus ($C_1 \dots C_c$) par un ou plusieurs experts. Chaque classe C_i se voit donc affecter un ensemble d'apprentissage E_i , et les données de l'ensemble de test sont segmentées en fonction des E_i . Parmi ces méthodes on peut citer : la segmentation Bayésienne, la segmentation par les champs de Markov [5], réseaux de neurones [6], etc.

Pour notre cas, puisqu'il s'agit de segmenter des images médicales IRM cérébrales, la classification supervisée de ces images nécessite donc la création d'une base d'apprentissage pour chaque classe et pour chaque patient ce qui est en elle même une tâche très fastidieuse pour les experts [7]. C'est pour cette raison qu'on s'intéresse aux méthodes non supervisées.

2.2 Méthodes Non Supervisées (Automatiques)

L'intérêt des méthodes non supervisées est qu'elles ne nécessitent aucune base d'apprentissage et par là même aucune tâche préalable d'étiquetage manuel n'est requise. La seule intervention de l'expert se situe à la fin du processus pour identifier les tissus en comparant les classes calculées avec les classes biologiques.

Les algorithmes non supervisés les plus répandus tendent à minimiser une fonction coût, dépendant de la distance de chaque pixel aux prototypes (ou noyaux) des classes. Le prototype d'une classe étant un point connu dont l'appartenance à la classe est garantie et où chaque pixel est assigné à la classe qui lui est la plus proche. Selon la certitude de la classification que nous voulons obtenir, et la relation entre les classes, nous pourrions distinguer plusieurs méthodes de classification [8]:

2.2.1 Classification HCM (Hard C -Means)

Notons $X = (x_j, j = 1 \dots N)$ l'ensemble des vecteurs forme de R^p , avec $x_j = [x_{j1}, x_{j2}, \dots, x_{jp}]^T$, et $B = (b_1, \dots, b_c)$ un ensemble de vecteur prototypes inconnus, où b_i caractérise la classe i . Dans la méthode HCM un élément de X est attribué à une classe et une seule parmi les C proposées. Dans ce cas, la fonctionnelle à minimiser est :

$$J(B, U, X) = \sum_{i=1}^C \sum_{j=1}^N u_{ij} d^2(x_j, b_i) \quad (1)$$

Les solutions au problème s'écrivent :

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

$$b_i = \frac{\sum_{j=1}^N u_{ij} x_j}{\sum_{j=1}^N u_{ij}} \quad (3)$$

Dans une méthode comme HCM, les éléments sont classés de façon certaine comme appartenant à une classe et une seule. Quelle que soit la modalité d'imagerie, cette assertion ne reflète pas la réalité physique de l'échantillon étudié (bruit, volume partiel, hétérogénéité de champ, etc.) [1]. Les méthodes présentées dans les paragraphes suivants permettent d'obtenir une segmentation floue qui prend en compte ces aspects imprécis et incertains [2].

2.2.2 Classification floue FCM (Fuzzy C-Means)

La modélisation de l'imprécision s'effectue en considérant des frontières graduelles au lieu de frontières nettes entre les classes. L'incertitude s'exprime par le fait qu'un pixel possède aussi bien des attributs qui l'assignent à une classe qu'à une autre. La classification floue assigne donc, non pas à un pixel une étiquette relative à une classe unique, mais son degré d'appartenance à chacune des classes. Ces valeurs expriment l'appartenance incertaine d'un pixel à une région et sont appelées degrés d'appartenance. Le degré d'appartenance se situe dans l'intervalle [0, 1] et les classes obtenues ne sont pas forcément disjointes. Dans ce cas, les données x_j ne sont plus assignées à une classe unique, mais à plusieurs par l'intermédiaire de degrés d'appartenance u_{ij} du vecteur x_j à la classe i . Le but des algorithmes de classification est non seulement de calculer les centres de classe B mais aussi l'ensemble des degrés d'appartenance des vecteurs aux classes.

Si u_{ij} est le degré d'appartenance de x_j à la classe i , la matrice $U_{C \times N} [u_{ij}]$ est appelée matrice de C-partitions floues si et seulement si elle satisfait aux conditions :

$$\forall i \in \{1..C\}, \forall j \in \{1..N\} \left\{ \begin{array}{l} u_{ij} \in [0,1] \\ 0 < \sum_{j=1}^N u_{ij} < N \end{array} \right. \quad (4)$$

$$\forall j \in \{1..N\} \sum_{i=1}^C u_{ij} = 1 \quad (5)$$

la fonctionnelle à minimiser (6), et les solutions (7), (8), au problème du FCM sont décrites par les formules suivantes :

$$J(B, U, X) = \sum_{i=1}^C \sum_{j=1}^N (u_{ij})^m d^2(x_j, b_i) \quad (6)$$

$$b_i = \frac{\sum_{k=1}^n u_{ik}^m \cdot x_k}{\sum_{k=1}^n u_{ik}^m} \quad (7)$$

$$u_{ij} = \left[\sum_{k=1}^C \left(\frac{d^2(x_j, b_i)}{d^2(x_j, b_k)} \right)^{\frac{2}{m-1}} \right]^{-1} \quad (8)$$

2.2.3 Classification possibiliste PCM (Possibilist C-Means)

Krishnapuram et Keller [9] ont suivi les idées de Zadeh [10] selon lesquelles (5), ne doit pas contraindre les appartenances d'un vecteur quelconque x_j si les classes représentées par les nuages sont considérées comme ces

sous-ensembles flous sur le domaine $X = \{x_j, j=1..N\}$. Les degrés d'appartenance doivent seulement appartenir à l'intervalle $[0,1]$. Ainsi, un nouvel ensemble de contraintes est défini :

$$\forall i \in \{1..C\}, \forall j \in \{1..N\} \quad u_{ij} \in [0,1] \quad (9)$$

$$\forall j \in \{1..N\} \quad \left\{ \begin{array}{l} 0 < \sum_{j=1}^N u_{ij} < N \\ \max_i u_{ij} > 0 \end{array} \right. \quad (10)$$

La condition (10), assure simplement que la partition floue résultante de l'algorithme recouvre le domaine X . En toute rigueur, la matrice U résultante n'est plus une C -partition floue, puisque la contrainte (5), n'est plus satisfaite. La fonctionnelle à minimiser (11), et les solutions (12), (13), de l'algorithme PCM deviennent :

$$J(B, U, X) = \sum_{i=1}^C \sum_{j=1}^N u_{ij}^m d(x_j, b_i)^2 + \sum_{i=1}^C \eta_i \sum_{j=1}^N (1 - u_{ij})^m \quad (11)$$

$$\forall i \in \{1..C\} \quad b_i = \frac{\sum_{k=1}^n u_{ik}^m \cdot x_k}{\sum_{k=1}^n u_{ik}^m} \quad (12)$$

$$\forall i \in \{1..C\} \quad \forall j \in \{1..N\} \quad u_{ij} = \frac{1}{1 + \left(\frac{d^2(x_j, v_i)}{\eta_i} \right)^{\frac{1}{m-1}}} \quad (13)$$

où η_i est un réel positif déterminant la distance à laquelle le degré d'appartenance d'un vecteur à la classe i est égal à 0.5.

3 Fusion de données

Le concept de la fusion de données est facile à comprendre, mais il est difficile d'en trouver une définition qui rende compte de ce cadre formel et de ses multiples facettes.

La définition qui nous semble la plus adaptée dans notre travail est celle de [11]: «*la fusion de données consiste à combiner des informations issues de plusieurs sources afin d'améliorer la prise de décision*».

3.1 Classification des opérateurs de fusion

La notion de fusion s'applique au cas de n sources [12], [13].

On considère le problème de l'agrégation de deux informations, n_1 et n_2 , issues de deux capteurs différents pour un même phénomène. On cherche à agréger les informations fournies par n_1 et n_2 en exploitant au mieux l'ambiguïté et la complémentarité des données. Bloch propose dans [12] la définition suivante pour le comportement de F :

- *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 $F(n_1, n_2) \geq \max(n_1, n_2)$.

Il existe différents types d'opérateurs, suivant leur variabilité et leur dépendance au contexte. Les exemples cités dans la partie qui suit sont extraits de la classification effectuée dans [12].

3.1.1 Opérateurs à comportement constant et indépendant du contexte

Ce sont des opérateurs ayant le même comportement quelles que soient les valeurs de 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. Cette classe sera notée *CCIC* [2], [12].

3.1.2 Opérateurs à comportement variable et indépendant du contexte

Ce sont les opérateurs qui ne dépendent pas du contexte mais dont le résultat est fonction des valeurs de n_1 et n_2 . Par exemple, les sommes symétriques sont des opérateurs à comportement Variable et indépendant du contexte, elles sont de la forme : $\sigma = \frac{g(x, y)}{g(x, y) + g(1-x, 1-y)}$ où g est une fonction de $[0,1] \times [0,1]$ dans $[0,1]$. Cette classe sera notée *CVIC* [2], [12].

3.1.3 Opérateurs dépendants du contexte

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 sur le phénomène étudié. Il est ainsi possible de construire des opérateurs dont le comportement sévère (resp. indulgent) est une fonction croissante (resp. décroissante) de l'accord entre les deux capteurs. On a ici à faire à un problème d'accord entre les sources. Cette classe d'opérateurs, qui sera notée *CDC*, nous intéressera tout particulièrement dans la suite [2].

3.2 Théories utilisées pour la fusion

Divers cadres théoriques sont utilisés pour représenter les informations en fusion de données. Certains utilisent la théorie des croyances, la théorie des probabilités ou la théorie des possibilités.

3.2.1 Théorie des probabilités

Les probabilités, du fait de leur ancienneté, développées au 19^{ème} siècle pour la mécanique classique, sont encore la base théorique la plus utilisée pour la représentation des données incertaines.

Plusieurs distributions de probabilités peuvent ensuite être combinées à l'aide de la règle de Bayes : Soit $H = \{H_1, \dots, H_N\}$, l'ensemble des hypothèses. Les H_i sont mutuellement exclusives. La probabilité *a posteriori* d'un événement H_i parmi les N hypothèses connaissant l'information I_j est donnée par

$$P(H_i | I_j) = \frac{P(H_i) \cdot P(I_j | H_i)}{\sum_{k=1}^N P(H_k) \cdot P(I_j | H_k)} \quad (14)$$

où $P(H_i)$ est la probabilité *a priori* de l'hypothèse H_i , et $P(I_j | H_i)$ représente la probabilité d'observer l'information I_j lorsque l'hypothèse H_i est réalisée.

3.2.2. Théorie des croyances

La théorie de l'évidence fut historiquement introduite par Shafer [13]. Mais les origines de la théorie sont attribuables à Dempster [14], [15]. À partir du formalisme mathématique développé, Shafer [13] a montré l'intérêt des fonctions de croyance pour la modélisation de connaissances incertaines.

Soit X , un ensemble de N hypothèses H_i exclusives et exhaustives, 2^X désigne l'ensemble des 2^N sous-ensembles A_j de X .

Selon [16], une fonction de masse élémentaire, m , est définie de 2^X sur $[0,1]$ par :

$$m(\phi) = 0 \quad (15)$$

$$\sum_{j=1}^{2^N} m(A_j) = 1 \quad (16)$$

Une fonction de crédibilité, Cr , peut également être définie sur les mêmes ensembles par:

$$Cr(\phi) = 0 \quad (17)$$

$$Cr(X) = 1 \quad (18)$$

$$Cr(A_k) = \sum_{A_j \subset A_k} m(A_j) \quad (19)$$

A partir de la notion de masse élémentaire, on peut également introduire une fonction de plausibilité, Pl

$$Pl(A_k) = \sum_{A_j \cap A_k = \phi} m(A_j) \quad (20)$$

Cette fonction mesure à quel point les informations données par une source ne contredisent pas A_k .

3.2.3 Théorie des possibilités

La théorie des possibilités a été introduite en 1978 par Zadeh [17] puis développée par Dubois et Prade en France [18], [13]. De même que la théorie des croyances, elle constitue un cadre permettant de traiter des données à caractère *imprécis* et/ou *incertain*. Elle peut être vue indépendamment de toute interprétation probabiliste comme une approche ordinale de l'incertain dans $[0,1]$, exploitée à l'aide des mesures de possibilité et de nécessité.

3.2.3.1 Mesure et distribution de possibilité

Soit un ensemble de référence fini X . Une *mesure de possibilité* Π est définie sur l'ensemble des parties de X ($P(X)$) et prend ses valeurs dans $[0,1]$ telle que :

$$\Pi(\phi) = 0, \Pi(X) = 1 \quad (21)$$

$$\forall A_1 \in P(X), A_2 \in P(X), \dots \Pi(\cup_{i=1,2,\dots} A_i) = \sup_{i=1,2,\dots} \Pi(A_i) \quad (22)$$

On introduit une fonction appelée *distribution de possibilité*, π , qui associe à tout événement de X , un coefficient compris entre 0 et 1 reflétant le degré avec lequel cet événement est possible. Cette fonction doit vérifier la condition de normalisation suivante :

$$\sup_{x \in X} \pi(x) = 1 \quad (23)$$

Une distribution de possibilité est un ensemble flou qui peut être associé à la mesure de possibilité bijectivement, et l'on a :

$$\forall x \in X \quad \pi(x) = \Pi(\{x\})$$

3.2.3.2 Mesure de nécessiter

Une mesure de possibilité fournit une information sur l'occurrence d'un événement A relatif à un ensemble de référence X , mais elle ne suffit pas pour décrire l'incertitude existante sur cet événement.

Pour compléter l'information sur A , on indique le degré avec lequel la réalisation de A est certaine par l'intermédiaire d'une *mesure de nécessité*.

Une mesure de nécessité, N , est une fonction définie sur l'ensemble $P(X)$ des parties de X , à valeurs dans $[0,1]$, telle que :

$$N(\emptyset) = 0, N(X) = 1 \quad (24)$$

$$\forall A_1 \in P(X), A_2 \in P(X), \dots N(\bigcap_{i=1,2,\dots} A_i) = \inf_{i=1,2,\dots} N(A_i) \quad (25)$$

Sur un ensemble de référence X , une mesure de nécessité N peut être obtenue à partir de la mesure de possibilité Π correspondante, par l'intermédiaire du complémentaire A^c de toute partie A de X :

$$\forall A \in P(X) \quad N(A) = 1 - \Pi(A^c) \quad (26)$$

4. Approche proposée

La fusion d'images peut alors se décomposer en trois grandes étapes :

- Modélisation des informations dans un cadre théorique commun.
- Fusion des informations issues de la modélisation précédente.
- Prise de décision.

4.1 Modélisation

L'étape de modélisation consiste en la représentation de l'information dans un cadre mathématique lié à une théorie particulière.

Notre but ici n'est pas de déterminer la meilleure méthode, mais la méthode la plus adaptée dans le contexte de la fusion des tissus cérébraux

4.1.1 Choix des algorithmes utilisés

Nous devons, tout d'abord, préciser 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. Cela constitue un premier inconvénient de ce type de méthodes. De plus, la littérature rapporte que les méthodes supervisées en segmentation d'images médicales peuvent être très dépendantes de la base d'apprentissage [4]. Pour toutes ces raisons, nous avons opté pour une méthode non supervisée.

4.1.1.1 C-moyennes Floues ou Algorithme Possibiliste ?

Nous indiquant ici les raisons qui nous ont amené à effectuer une coopération entre l'algorithme flou FCM et l'algorithme possibiliste PCM. Cela a pour but de rendre l'algorithme de classification plus robuste face aux imprécisions et aux données aberrantes. Le PCM peut être utilisé dans un deuxième passage pour les points aberrants, après l'application d'un autre algorithme de regroupement, tel que les FCM, qui fournit une partition initiale pour les PCM. Ce dernier améliore cette partition obtenue suite à la première étape.

- *FCM et degrés d'appartenance relatifs* [2]

La contrainte de normalisation (5) Utilisée pour la minimisation de la fonctionnelle (6) est source d'erreur dans l'interprétation des degrés d'appartenance issus du *FCM*. Krishnapuram et Keller [9] donnent une série d'exemples simples qui illustrent les problèmes associés à cette contrainte, que nous résumons dans la figure 1.

Cette figure présente deux nuages de points avec deux points aberrants A et B . L'algorithme des C-moyennes floues appliqué à deux classes ($C=2$) avec $m=2$ et une distance euclidienne calcule les centres O_1 et O_2 . Les degrés d'appartenance de chaque point aux deux classes sont présentés dans le tableau 1.

Les points A et B , résultant, par exemple, d'une erreur de mesure ou d'un bruit, ne doivent pas avoir des degrés d'appartenance significatifs aux deux classes. De plus, les degrés d'appartenance de B doivent être plus faibles que ceux de A , ce point étant encore plus éloigné des classes. Néanmoins, en dépit de ces considérations intuitives, la contrainte (5) impose aux degrés d'appartenance de A et B aux deux classes d'être approximativement égaux à 0.5.

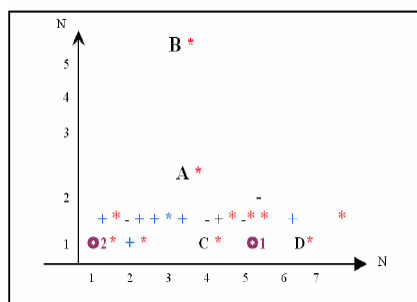


Fig. 1. Démonstration de la mauvaise interprétation des degrés d'appartenance du FCM [2].

Les degrés d'appartenance des points C et D , tels qu'ils sont calculés par FCM , sont significativement différents alors que ces deux points sont approximativement à la même distance du centre de la classe de gauche. Cela est encore une fois dû à la contrainte de normalisation (5) qui force C à partager un peu de son appartenance à la classe de droite, dont il est plus proche que D ne l'est.

Table 1. Comparaison des degrés d'appartenance générés par FCM et PCM [2].

Points	A		B		C		D	
$u_{ij(FCM)}$	0.53	0.47	0.52	0.48	0.62	0.38	0.82	0.18
$u_{ij(PCM)}$	0.26	0.18	0.09	0.06	0.77	0.13	0.76	0.07

- PCM et degrés d'appartenance absolus [2]

La valeur de u_{ij} dans (11) ne dépend maintenant que de la distance du vecteur x_j à la classe i , contrairement à (6). Les degrés d'appartenance générés par PCM ne sont ainsi plus des degrés relatifs ou de partage, ils deviennent des valeurs absolues reflétant la force avec laquelle chaque vecteur appartient à toutes les classes.

Les degrés d'appartenance générés par PCM reflètent de manière plus exacte la réalité de la distribution des points (A et B) se voient affecter des degrés d'appartenance faibles aux deux classes, (C et D) ont un degré d'appartenance à la première classe quasi identique.

Les images IRM sont affectées, comme on l'a déjà mentionné, par des artefacts que l'algorithme PCM les gère mieux. Pour toutes ces raisons, nous nous sommes orientés vers l'algorithme de classification possibiliste pour caractériser les tissus

4.1.1.2 Initialisation de l'algorithme

Les résultats de l'algorithme PCM vont évidemment dépendre de l'étape d'initialisation [1]. Ainsi, l'initialisation, si elle doit exister, ne doit pas être aussi précise que dans d'autres algorithmes comme FCM . Tout algorithme (flou ou non) peut donc être utilisé, et l'algorithme FCM constitue une manière adéquate d'initialiser les données [19], puisqu'il donne accès à une première estimation de U et de B .

4.1.1.3 Détermination du nombre de classes

Notre objectif consiste à segmenter l'encéphale, ce qui nous pousse à fixer à 3 le nombre de classe à identifier ($C=3$) correspondant aux trois tissus cérébraux présents dans l'encéphale à savoir la matière blanche (MB), la matière grise (MG) et le liquide céphalo-rachidien (LCR).

4.1.1.4 Choix des vecteurs forme

Le choix des vecteurs forme est fondamental puisque leur pertinence va permettre de discriminer les pixels entre eux. Ce choix est défini suivant le type de modalité. L'image anatomique que nous utilisons est une IRM. L'imagerie par résonance magnétique est une modalité d'imagerie multispectrale (multimodale) donnant accès à un grand nombre

de paramètres et donc de vecteurs forme. La première caractéristique qui peut être exploitée est le signal lui-même, principalement par l'intermédiaire d'images pondérées en T_1 , T_2 et en densité de protons. Le vecteur forme x_j d'un pixel j est alors formé des niveaux de gris de ce pixel dans toutes les modalités. Cette information est très largement utilisée en segmentation d'images [7].

4.2 Fusion

De nombreux auteurs ont comparé les théories des probabilités, des possibilités et des croyances, et ont détaillé les transformations permettant de passer d'un formalisme à l'autre.

4.2.1 Limitations de la fusion probabiliste

Notre choix ici est exclusivement orienté vers le domaine de l'imagerie. 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 :

- dans les images traitées dans la partie suivante, les informations sont à la fois incertaines et imprécises. Or la théorie des probabilités ne peut pas bien prendre en compte l'imprécision des données,
- le formalisme, 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. Des modèles peuvent représenter ces connaissances mais imposent alors des hypothèses fortes sur les étapes de modélisation et de fusion.

4.2.2 Vers la théorie des possibilités

Nous avons comparé les théories des croyances et des possibilités par rapport au problème spécifique de la fusion d'images médicales, cette comparaison permet de se rendre compte que des différences importantes existent, ces différences sont :

- La théorie de l'évidence prend en compte les ensembles composés de plusieurs classes, ce qui permet de considérer le doute entre les classes.
- L'étape de combinaison en théorie des croyances se résume la plupart du temps à l'application de l'opérateur orthogonal, dans le cas où toutes les ambiguïtés peuvent être introduites à l'étape de modélisation. Au contraire, la théorie des possibilités offre une grande variété d'opérateurs ayant des comportements différents suivant la situation présentée. Le fait de prendre en compte dans la combinaison le conflit et la fiabilité des sources nous a paru fondamental, en particulier pour préserver l'information pertinente pour le diagnostic [1]. Nous nous sommes donc plutôt orientés vers la construction d'opérateurs dépendants du contexte (CDC) et avons ainsi privilégié la représentation possibiliste des données.

L'analyse de ces modèles pour les fusions envisagées a révélé tout d'abord que la théorie des probabilités était mal adaptée, principalement en raison des effectifs d'expérimentation faibles dont nous disposons. La relative simplicité des informations à extraire dans les images et la pauvreté des modes de combinaison en théorie des croyances ont été ensuite pour nous des éléments décisifs dans le choix d'un cadre formel et nous ont fait préférer l'approche possibiliste à la théorie des croyances.

4.3 Décision

La dernière étape consiste à prendre la décision quant à l'appartenance d'un voxel de l'image I à une classe C_i . La règle de décision adoptée consiste à prendre une coupe de la carte de fusion en choisissant un seuil d'étiquetage.

$$\forall v \in I, v \in C_i \text{ Si } \pi_{\text{Fusionné}}(v) \geq \text{seuil}$$

5 Résultats

Afin de prouver l'efficacité de l'approche développée, nous avons implémenté les algorithmes FCM, PCM et l'approche développée. Il s'agit d'images pondérées en T1, T2 et en densité de proton (Dp) pour des patients de différents âges (taille pixel = 1mm, taille de matrice 512 X 512). Les images sont en format DICOM (*Digital Imaging and Communications in Medicine*).

5.1 Analyse des résultats

La figure 2 illustre les résultats obtenus après segmentation des différents tissus cérébraux d'images (a) et (b) pondérées en T1 et en T2 sur des coupes axiales. Les images (c), (d), (e) et (f) sont les résultats de la segmentation par FCM et PCM successivement.

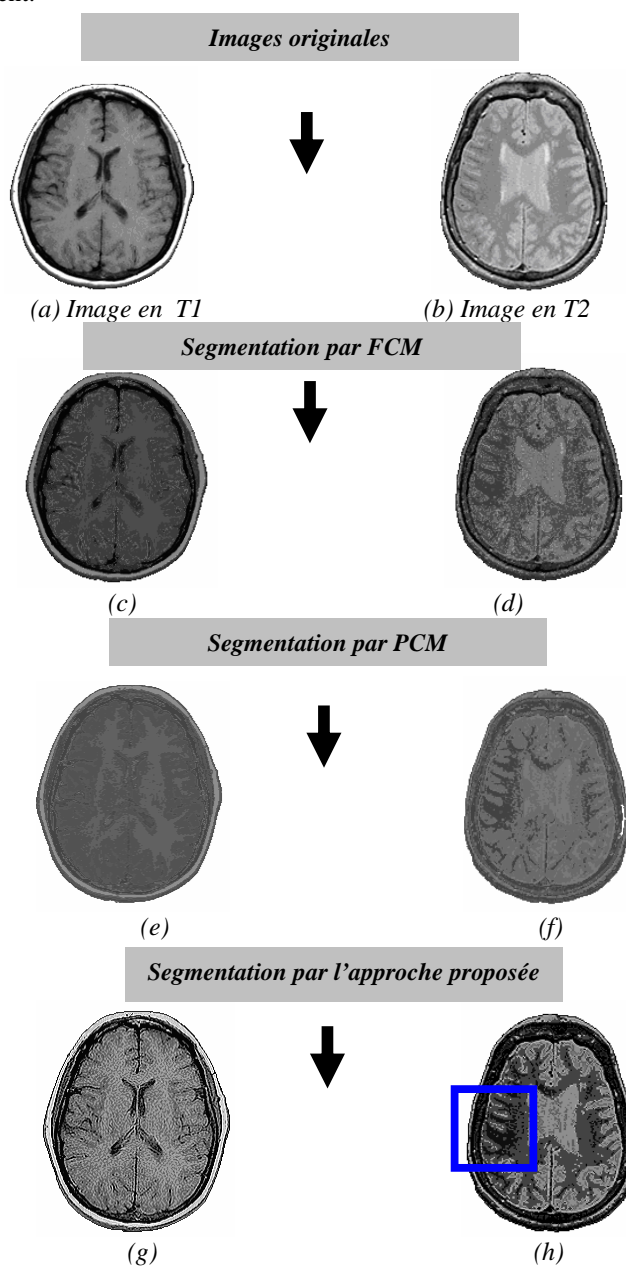


Fig. 2. Résultats de segmentation.

Table 2. Comparaison des taux de recouvrement obtenus par différents algorithmes.

	FCM	PCM	Approche proposée
MG	0,85	0,64	0,90
MB	0,9	0,52	0,92
LCR	0,63	0,65	0,91

Les résultats du tableau 2 confirment l'intérêt de l'approche de fusion par rapport aux approches prises indépendamment.

6 Conclusion

Dans cet article nous avons démontré l'importance et les difficultés de ces travaux de segmentation d'images IRM. Deux raisons expliquent ces difficultés :

- La première est qu'il existe une très grande variété de tissus anormaux qui diffèrent par leur taille, leur forme, leur position et leur composition.
- La seconde raison vient de ce que la donnée issue de l'acquisition IRM est sensible au bruit de fond et à l'échantillonnage.

Nous avons tout d'abord décomposé le processus de fusion de données en trois phases fondamentales.

En premier lieu, nous avons modélisé des informations, numériques ou symboliques, dans un cadre commun permettant de prendre en compte les ambiguïtés, les imprécisions et les incertitudes. Nous nous sommes pour cela placés dans le cadre de la coopération entre l'algorithme flou FCM et l'algorithme possibiliste PCM afin de rendre l'algorithme plus robuste face aux imprécisions et aux données aberrantes.

En second lieu, nous avons fusionné différentes données. Cette agrégation a été réalisée par des opérateurs de fusion qui modélisent l'analyse quotidienne du médecin confronté à des données cliniques hétérogènes. Nous avons explicité l'opérateur qui nous a semblé le plus adapté, en fonction de propriétés mathématiques souhaitées, de considérations intuitives sur la nature des données et de tests expérimentaux.

En dernier lieu, nous avons présenté ces informations fusionnées au clinicien. Ici encore, nous avons proposé pour chaque type de fusion une solution. Nous avons présenté les résultats de notre travail qui consiste à utiliser plus d'un algorithme pour segmenter des images médicales en vue d'améliorer la qualité de la segmentation. La bonne qualité de nos solutions tient au fait que :

1. C'est une méthode totalement automatique grâce aux modélisations des connaissances *a priori* des experts neuroradiologues. Elle ne dépend pas d'échantillons obligatoirement utilisés dans les méthodes d'apprentissage. La théorie floue est importante pour modéliser les connaissances humaines en utilisant les fonctions mathématiques et pour solutionner l'effet de volume partiel d'IRM.
2. Elle satisfait aux contraintes de l'application grâce à l'automatisme et aux différents résultats finaux pouvant être fournis par la reconstruction 3D floue.
3. Elle est robuste en présence des points aberrants.
4. Sa performance est meilleure que celle de la méthode supervisée.
5. C'est un système à base de la théorie floue et possibiliste.
6. Elle est efficace sur au moins 2 types de tissus.

7 Références

1. Semchedine, M., Toumi, L., Moussaoui, A. : Nouvelle Approche de Classification Multimodale Hybride d'Images IRM dans un SMA : Journées internationales sur l'informatique graphique .Tébessa 2006
2. Barra, V. : Fusion d'Images 3D du Cerveau : Etude de Modèles et Applications :, Ph.D. Thesis, Université d'Auvergne, Clermont-Ferrand (France), 2000.
3. Dou, W. : Segmentation d'images multi spectrales basée sur la fusion d'informations : application aux images IRM :, Thèse doctorat de l'université de Caen, Spécialité : Traitement du Signal et des Images, 2006
4. Bezdek, J., Hall, L., Clarke, L.: Review of MR Image Segmentation Techniques using Pattern Recognition:, Medical Physics, 20:1033-1048, 1993.

5. Held, K., Rota Kops, E., Krause, B., Wells, W.: Markov Random Field Segmentation of Brain MR Images: IEEE Transactions on Medical Imaging, 16:878-886. 1997
6. Raff, U., Scherzinger, A., Vargas, P., Simon, J.: quantitation of grey matter, white matter and cerebrospinal fluid from spin-echo magnetic resonance images using an artificial neural network: Technique, Medical Physics, 1994
7. Barra, V. : Segmentation floue des tissus cérébraux en IRM 3D: une approche possibiliste versus autres méthodes :. Thèse doctorat. Université Blaise Pascal 1999.
8. Semchedine, M., Toumi, L., Moussaoui, A.: Système Coopératif de Classification Floue Possibiliste avec Rejet d'Ambiguïté Application à la segmentation d'images IRM : 4th International Conference on Computer Integrated Manufacturing CIP'2007
9. Krishna, R.: The Possibilistic C-Means Algorithm: Insights and Recommendations: IEEE transactions on Fuzzy Systems 1996
10. Zadeh, L.: Fuzzy Sets as a Basis for Theory of Possibility:, International Journal of Fuzzy Sets and Systems, 1:3-28. 1978
11. Bloch, I. : Fusion de Données en traitement d'images : modèles d'information et décision :. Traitement du signal, 1994,vol.11, n°6, pp 435 - 446
12. Bloch, I.: Information Combination Operators for Data Fusion: A Comparative Review with Classification:, IEEE Transactions on Systems, Man, and Cybernetics, 1:52-67, 1996
13. Dubois, D., Prade, H.: Possibility Theory, an approach to the computerized processing of uncertainty:, Plenum Press, 1988
14. Shafer, G.: A Mathematical Theory of Evidence, Princeton University Press:, 1976
15. Dempster, A. P.: Upper and lower probability function in a context of uncertainty:, Annals of math. Statistics, vol 38, pp. 325-339, 1967
16. Dempster, A. P.: A generalization of bayesian inference", Jour. of the Royal Statistical Society, vol 30, pp. 205- 247, 1968.
17. Zadeh, L.: Fuzzy Sets as a Basis for Theory of Possibility:, International Journal of Fuzzy Sets and Systems, 1:3-28, 1978
18. Dubois, D., Prade, H.: Fuzzy Sets and Systems: Theory and applications:, Academic press, New york, 1980.
19. Ménard, M.: The fuzzy c+2 means: Solving The Extended Ambiguity Reject in Clustering:. In IEEE Transactions on fuzzy systems, vol. 1, N° .2, pages 195-203. 1998
20. Moussaoui, A.: Aneuro-fuzzy image segmentation :. International journal of soft computing 1(3):232-238 Medwell Online. 2006
21. Shufelt, J.A.: Performance Evaluation and Analysis of Monocular Building Extraction from Aerial Imagery:. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(4), pp. 311–326. 1999