



**HAL**  
open science

## Vers la classification de matériaux à partir de compositions chimiques incertaines

Arnaud Grivet Sébert, Jean-Philippe Poli

► **To cite this version:**

Arnaud Grivet Sébert, Jean-Philippe Poli. Vers la classification de matériaux à partir de compositions chimiques incertaines. LFA 2017 - 26e Rencontres francophones sur la Logique Floue et ses Applications, Oct 2017, Amiens, France. cea-01992352

**HAL Id: cea-01992352**

**<https://cea.hal.science/cea-01992352>**

Submitted on 24 Jan 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Vers la classification de matériaux à partir de compositions chimiques incertaines

## Towards material classification from uncertain chemical compositions

Arnaud Grivet Sébert

Jean-Philippe Poli

CEA, LIST

91191 Gif-Sur-Yvette cedex, France.

arnaud.grivet.sebert@cea.fr, jean-philippe.poli@cea.fr

### Résumé :

Dans cet article, une étude sur la classification de matériaux à partir de leur composition chimique est réalisée. La méthode utilisée se base sur un arbre de décision flou et l'accent est mis sur la gestion de l'incertitude des données, qui fait l'objet d'un traitement particulier que nous exposerons. L'algorithme complet est testé sur des jeux de données simulés, dans l'attente de données réelles.

### Mots-clés :

Arbre de décision flou, règles floues, incertitude, classification de matériaux.

### Abstract:

In this paper, a study about the classification of materials from their chemical composition is realized. The used method relies on a fuzzy decision tree and focuses on handling the uncertainty of the data by a technique we will explain. The complete algorithm is tested on simulated data sets, in the expectation of real data.

### Keywords:

Fuzzy decision tree, fuzzy rules, uncertainty, material classification.

## 1 Introduction

La sécurité dans les ports et aux frontières est un enjeu majeur en Europe : en effet, de nombreuses substances illicites ou dangereuses transitent à travers les douanes. Or, les fouilles systématiques des containers sont impossibles dans la pratique pour des raisons de coût et de temps. En effet, les volumes traités chaque jour dans les grands ports européens tels que Rotterdam ou Gdansk sont énormes : par exemple, 444 millions de tonnes de marchandises ont transité par le port de Rotterdam en 2014.

Notre travail s'inscrit dans le projet européen C-BORD visant à déterminer le contenu des containers à l'aide de diverses technologies.

L'une d'entre elles est le neutron marqué qui permet d'obtenir la composition chimique d'un volume du container. A partir de cette composition chimique, nous souhaitons déterminer le matériau présent dans le container. Afin que l'outil soit exploitable par les douaniers, nous souhaitons également fournir une justification de cette classification en utilisant un système expert flou.

Les proportions sont obtenues par différents traitements qui échappent à notre contrôle. Ainsi, les données que nous avons en entrée sont imprécises et sont accompagnées d'une mesure de cette imprécision. La logique floue semble ainsi tout indiquée à l'exploitation de ces données..

Compte tenu du temps nécessaire et des autorisations à obtenir pour utiliser un générateur de neutrons, nous disposerons d'un jeu de données assez restreint pour procéder à l'apprentissage, même si toutes les classes de matériaux seront bien évidemment représentées. L'idée est donc d'utiliser les arbres de décisions flous [3] qui ont été appliqués avec succès sur différents problèmes de classification [1, 7, 10]. La rareté des données pour l'apprentissage et l'incertitude intrinsèque aux traitements de celles-ci excluent les approches d'apprentissage statistique classiques telles que les réseaux de neurones, les SVM, etc. qui, de plus, ne justifient pas leurs décisions.

Nous présentons ainsi une modification des arbres de décision flous afin qu'à partir de ra-

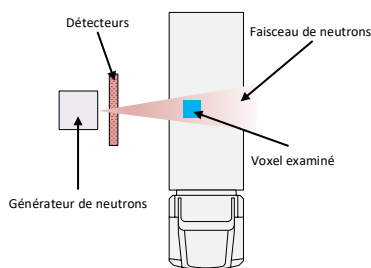


Figure 1 – Principe de l'inspection neutronique

tios stoechiométriques, exprimés en pourcentages, et une caractérisation de leur incertitude, nous puissions reconnaître les matériaux contenus dans la partie du container examinée.

L'article suit le plan suivant : nous exposerons dans un premier temps le contexte de notre travail avant de décrire de manière détaillée l'algorithme utilisé. Enfin, nous présenterons les premiers résultats obtenus à la suite des expérimentations sur des jeux de données simulées.

## 2 Contexte applicatif

### 2.1 Inspection neutronique pour la fouille des containers

Le projet H2020 C-BORD (*effective Container inspection at BORDER control points*) a pour but de faciliter la fouille des containers aux frontières en exploitant différentes technologies : nez artificiels, rayons X, photofission et neutrons marqués. Il s'agit pour cela de détecter des matières dangereuses (explosifs, armes, matières nucléaires, ...) ou illicites (drogues, contrebandes, ...).

Nous nous intéressons dans cet article à la classification des données issues de l'inspection neutronique. Comme le montre la figure 1, un dispositif produit un faisceau de neutrons permettant de se focaliser sur un certain volume (appelé voxel) du container. Les neutrons vont interagir avec les noyaux des atomes contenus dans le voxel, produisant ainsi de nouvelles particules qui peuvent être détectées

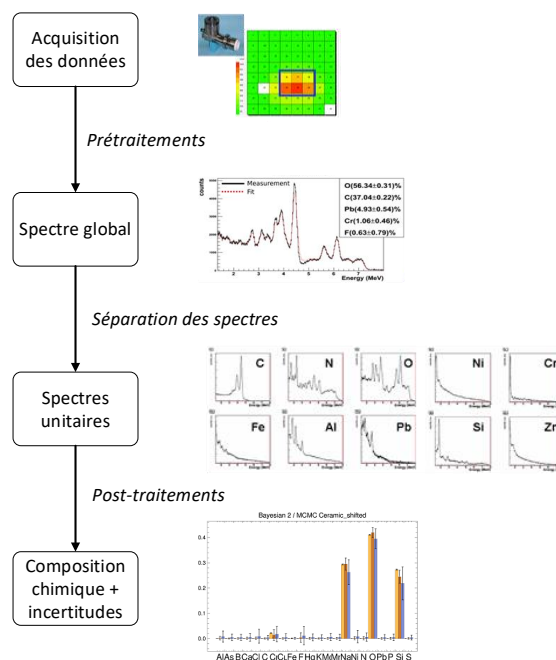


Figure 2 – Traitements à effectuer entre l'acquisition des données et la composition chimique

par les capteurs matriciels positionnés sur le côté du container. Ces particules sont ainsi caractéristiques des atomes rencontrés dans le voxel examiné.

Le traitement des données brutes ne fait pas l'objet de cet article mais nous en décrivons rapidement le principe dans la figure 2. Après différents prétraitements, un spectre global est obtenu. Par comparaison avec les spectres caractéristiques de chacun des atomes étudiés, ce spectre est décomposé en spectres individuels par un processus bayésien qui permet de déduire la composition chimique du voxel, exprimée en pourcentages. Le processus renseigne également une incertitude sur chacune des valeurs, exprimée sous la forme d'une variance. La figure 3 montre le résultat de ces traitements pour une exposition à la céramique, avec les incertitudes représentées par des "boîtes à moustache".

Notre travail consiste à exploiter ces histogrammes et incertitudes afin de reconnaître les matériaux contenus dans le voxel.

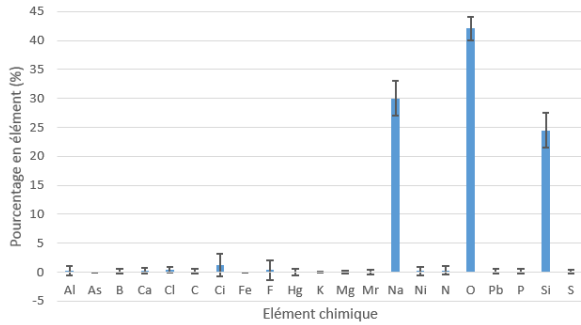


Figure 3 – Composition chimique calculée pour une simulation de céramique

## 2.2 Travaux antérieurs

La difficulté de la reconnaissance des matériaux à partir de la composition chimique réside dans plusieurs points. D'une part, les données sont rares car peu de campagnes d'acquisition peuvent être menées. D'autre part, l'incertitude sur la composition chimique rend la tâche difficile pour les modèles statistiques classiques. La rareté des données explique également le fait que peu de travaux s'intéressent à la reconnaissance des matériaux depuis leur composition chimique. Enfin, une difficulté non négligeable est que le dispositif est insensible aux atomes d'hydrogène.

Aussi, les travaux antérieurs proposent des méthodes d'analyse visuelle du contenu des voxels. Dans [2], les auteurs proposent notamment un diagramme de Voronoï permettant de mettre en évidence la proximité, en termes de composition chimique, du voxel courant avec des voxels précédemment inspectés et dont le contenu est connu (voir figure 4). Ainsi, il ne s'agit pas de reconnaître les matériaux présents dans le container mais d'afficher des containers visuellement proches et connus afin de déduire le contenu. Cette approche a l'avantage de ne pas nécessiter d'apprentissage ou de paramétrage puisqu'elle repose sur la sélection manuelle d'un voisinage. Conjointement à cette représentation, deux autres représentations ont été proposées. Il s'agit de projections du voxel courant sur deux triangles : un triangle dit

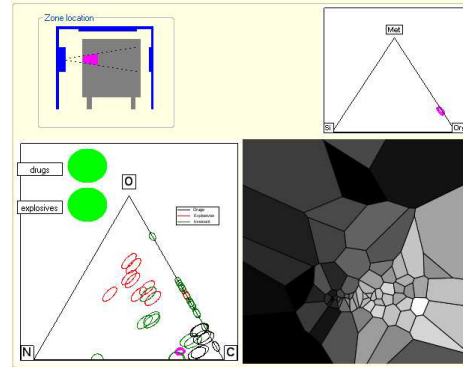


Figure 4 – Capture d'écran des visualisations proposées dans [2]

des matériaux indique la proximité du voxel avec les métaux, la céramique et les matières organiques, tandis qu'un triangle dit d'alerte présente les ratios entre carbone, azote et oxygène. Ce dernier triangle permet notamment de distinguer les drogues et les explosifs. L'opérateur doit ainsi être capable d'interpréter lui-même les différentes représentations.

En pratique, la prise en main de ces représentations, particulièrement le graphe de Voronoï, peut s'avérer difficile pour les opérateurs qui ne sont pas familiers avec ces techniques de visualisation. Dans le cadre du projet C-BORD, nous souhaitons aller plus loin et proposer une liste de matériaux, voire une explication de la décision, en utilisant un système expert flou.

## 3 Classification

### 3.1 Induction d'un arbre de décision flou

Pour pouvoir classifier efficacement les matériaux tout en générant des règles intelligibles par un humain, nous avons choisi d'utiliser un arbre de décision flou s'inspirant de celui défini par Janikow dans [6]. Chaque nœud est divisé, suivant un attribut qui n'a pas encore été utilisé, en  $N$  nœuds fils correspondant aux termes flous associés. L'arbre est donc d'arité le nombre de termes flous (on choisit pour le moment le même nombre pour tous

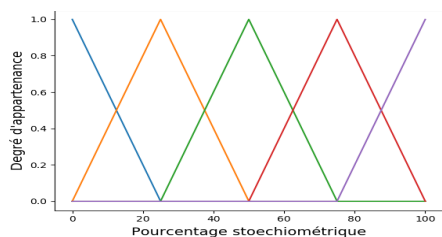


Figure 5 – Cinq termes flous triangulaires

les éléments) et de profondeur bornée par le nombre d'attributs. Dans notre cas, les attributs sont les différents atomes que le système est capable de cibler. La construction de notre arbre prend fin lorsque le gain d'entropie obtenu en divisant le nœud courant par tout attribut non utilisé est inférieur à un seuil fixé au préalable ; au niveau actuel d'avancement de notre travail, l'arbre n'est pas élagué.

Dans le but de générer des règles de classification des matériaux, nous devons dans un premier temps définir les termes des variables linguistiques. Nous avons choisi pour l'instant d'utiliser des termes flous triangulaires et de faire une partition forte du domaine de définition du pourcentage stoechiométrique de chaque élément, c'est-à-dire  $[0, 100]$  (voir figure 5). Ces termes sont donc fixés au préalable et ne sont pas appris ; cet apprentissage des termes pourra faire l'objet d'une amélioration ultérieure de notre méthode.

Les termes flous étant définis, nous devons pouvoir calculer le degré d'appartenance des exemples à chacun de ces termes. Dans [5], Duch expose une technique d'intégration de l'incertitude, qu'elle soit triangulaire, trapézoïdale ou gaussienne, avec des termes non flous, à valeurs dans  $\{0, 1\}$ , du type  $x \in [a, b]$ . Tsang et al. [9] ont plus récemment utilisé une technique d'intégration, pour des incertitudes de forme générale, mais toujours avec des termes non flous. Nous proposons une généralisation de ces travaux à des nombres flous. Pour construire l'arbre tout en tenant compte de l'incertitude des données, nous utilisons donc des distributions gaussiennes

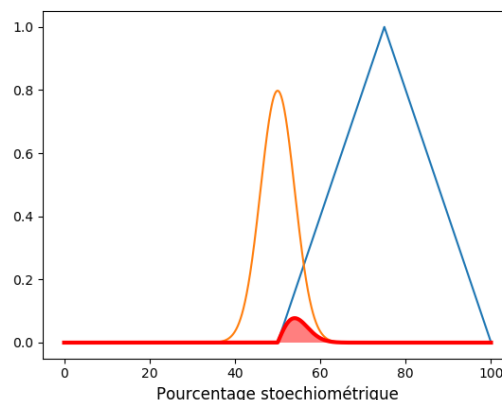


Figure 6 – Intégrale du produit incertitude-terme flou

à partir desquelles nous nous ramenons à des données ayant un degré d'appartenance unique et totalement connu par intégration du produit de la fonction d'incertitude et de la fonction d'appartenance au terme flou. En notant  $f$  la densité de probabilité représentant l'incertitude associée à la proportion en un élément  $e$  d'un exemple  $x$ , et  $\mu_v(t)$  le degré d'appartenance de la valeur  $t$  au terme flou  $v$ , on a donc l'expression suivante du degré d'appartenance de l'exemple incertain  $x$  au terme flou  $v$  :

$$\tilde{\mu}_v(x) = \int_0^{100} f(t)\mu_v(t)dt$$

Pour alléger l'écriture, nous utiliserons par la suite la notation  $\mu_v$  sans tilde pour le degré d'appartenance d'un exemple incertain.

Cette intégrale permet de prendre en compte tout le spectre continu de valeurs possibles pour la proportion de  $e$  dans  $x$ , et revient à faire une moyenne pondérée des degrés d'appartenance de ces valeurs possibles au terme flou  $v$ . Elle est illustrée dans la figure 6 : la gaussienne modélisant l'incertitude, dilatée d'un facteur 8 pour des raisons de visibilité, est en orange, le terme flou triangulaire en bleu et le produit des deux, ainsi que son intégrale représentée par l'aire sous la courbe, en rouge.

Afin que les distributions gaussiennes conservent leur sens probabiliste, on les

normalise sur le domaine de définition des données soit  $[0, 100]$ , les données étant des pourcentages en éléments chimiques. Pour ce faire, on divise simplement chaque gaussienne par son intégrale sur le domaine borné :

$$f(t) = norm(t) * \frac{\int_{-\infty}^{+\infty} norm}{\int_0^{100} norm} = \frac{norm(t)}{\int_0^{100} norm} \text{ où}$$

$norm$  est la densité de probabilité de la loi normale. Grâce à cette transformation, les termes flous extrêmes, qui ont une aire deux fois moins importante que les autres termes flous, ne sont pas pénalisés dans le calcul du degré d'appartenance d'un exemple incertain.

Pour le critère de choix de l'attribut, nous avons opté pour l'entropie floue utilisée par Peng et al. [8], qui est une généralisation de l'entropie de Shannon au cas flou. Cette entropie a l'avantage de prendre en compte les classes des objets analysés, ce qui est essentiel pour un problème de classification. Au contraire, l'entropie floue de De Luca et Termini [4] ne prend pas en compte les classes et évalue seulement le degré de "flou" d'un terme et non pas sa pertinence par rapport à la classification des objets. Dans l'entropie de Peng et al., la probabilité d'être dans une classe  $c$  donnée conditionnellement au nœud  $n$  qui apparaît dans l'entropie de Shannon est remplacée par le rapport :

$$fr_{c/n} = \frac{\sum_{x_i \in c} d_n(x_i)}{\sum_{x_i} d_n(x_i)} \quad (1)$$

qu'on nommera fréquence floue, où  $\{x_i, i \in [1, T]\}$  est le corpus d'entraînement de l'algorithme,  $d_n(x)$  est le degré d'appartenance d'un exemple  $x$  au nœud  $n$  de l'arbre. En faisant l'hypothèse d'une distribution de probabilité uniforme sur l'ensemble des exemples, on peut rapprocher ce rapport d'une probabilité conditionnelle floue au sens de Zadeh [11], mais en utilisant au numérateur l'intersection d'ensembles flous au lieu du produit :  $P(A|B) = \frac{P(A \cap B)}{P(B)}$ .

Nous avons choisi d'utiliser le produit comme T-norme car il donne de bien meilleurs résultats que la T-norme de Zadeh (dix à vingt points de plus en termes de taux de bonne classification).

Le degré d'appartenance d'un exemple  $x$  à un nœud  $n$  est donc le produit des degrés d'appartenance de  $x$  aux termes flous associés à  $n$  et aux ascendants de  $n$ . On a donc la définition récursive :

$$d_n(x) = d_N(x) \times \mu_v(x) \quad (2)$$

où  $N$  est le nœud père de  $n$ ,  $v$  est le terme flou associé au nœud  $n$  et  $\mu_v(x)$  est le degré d'appartenance de  $x$  au terme flou  $v$ .

On convient naturellement que, pour tout exemple  $x$  (d'entraînement ou de test),  $d_{n_0}(x) = 1$ , où  $n_0$  est la racine de l'arbre. On note de plus  $fr_n = \frac{\sum_{x_i} d_n(x_i)}{\sum_{n_j} \sum_{x_i} d_{n_j}(x_i)}$  la "fréquence d'appartenance" des exemples au nœud  $n$ , avec  $\{n_j, j \in [1, F]\}$  l'ensemble des nœuds "frères" de  $n$  et  $n$  lui-même.

On a donc finalement l'expression suivante de l'entropie floue utilisée à un nœud  $N$  ayant pour fils les nœuds  $\{n_j, j \in [1, F]\}$  relativement à un élément  $e$ , ces nœuds fils correspondant chacun à un des termes flous relatifs à  $e$  :

$$E = - \sum_{n_j} fr_{n_j} \sum_{c_k} fr_{c_k/n_j} \log(fr_{c_k/n_j})$$

où  $\{c_k, k \in [1, K]\}$  est l'ensemble des classes du problème.

Il convient de remarquer que dès la construction de l'arbre, les degrés d'appartenance des exemples aux termes flous sont bien calculés en considérant l'incertitude des données (via le calcul intégral exposé plus haut). L'incertitude des données d'entraînement est donc bien prise en compte, comme elle le sera avec les données de test par la même technique d'intégration.

La figure 7 montre un arbre de décision induit à partir d'un jeu de données simulé (voir section 4.1). Nous définissons le degré de confiance d'une règle  $r$  associée à une feuille  $f$  par :

$$conf(r) = \frac{\sum_{x_i \in c} d_f(x_i)}{\sum_{x_i} d_f(x_i)} = fr_{c/f}$$

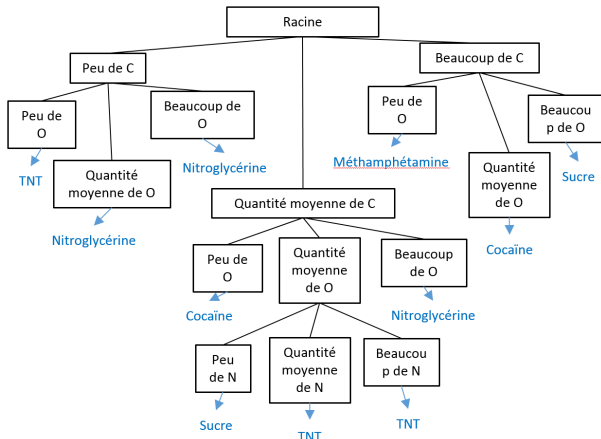


Figure 7 – Arbre de décision et règles induites

où  $c$  est la conséquence principale de  $r$ , c'est-à-dire la classe la plus représentée dans la feuille  $f$ .

### 3.2 Reconnaissance

Une fois l'arbre induit grâce aux données d'entraînement, le principe de classification des données de test est assez simple. Soit un exemple de test incertain  $x$  dont l'algorithme ignore la classe. On parcourt l'arbre depuis la racine en calculant les degrés d'appartenance de  $x$  aux noeuds de l'arbre, le degré d'appartenance à un noeud  $n$  se calculant par simple produit du degré d'appartenance au noeud père de  $n$  et du degré d'appartenance au terme flou associé à  $n$  (équation (2)).

On obtient donc le degré d'appartenance de  $x$  à chaque feuille  $f$  de l'arbre. Pour chaque classe  $c$ , on multiplie ce degré d'appartenance par la fréquence floue  $fr_{c/f}$  de  $c$  conditionnellement à  $f$  (voir définition (1)). On agrège alors ces résultats par classe sur toutes les feuilles à l'aide d'une somme, pour obtenir un degré de confiance  $conf(x \in c)$  de  $x \in c$ . On a ainsi, pour chaque classe  $c$  :  $conf(x \in c) = \sum_{f_i} d_{f_i}(x) \times fr_{c/f_i}$  où  $f_i, i \in \llbracket 1, F \rrbracket$  est l'ensemble des feuilles de l'arbre.

Pour les rendre plus intelligibles, on normalise ensuite, pour chaque exemple  $x$ , les degrés de confiance  $conf(x \in c)$  de manière à ce que leur

somme vale 1 :  $conf(x \in c) \leftarrow \frac{conf(x \in c)}{\sum_{c_k} conf(x \in c_k)}$ .

L'algorithme classe alors l'exemple  $x$  dans la classe  $c$  qui maximise  $conf(x \in c)$ . Nous pourrions néanmoins proposer au douanier qui utilise le détecteur les différentes hypothèses correspondant aux classes les plus probables, en affichant le degré de confiance de chacune.

## 4 Expérimentations

### 4.1 Constitution de jeux de données

Les expérimentations avec le générateur de neurons n'ayant pas pu être réalisées au moment où cet article est écrit, nous avons simulé des données. Pour cela, nous avons utilisé les formules brutes des matériaux (classes) à reconnaître, à partir desquelles nous déduisons les pourcentages stoechiométriques de chaque élément. Ces pourcentages nous donnent une valeur de référence pour chaque couple (classe, élément). Pour créer un exemple  $x$  d'une classe  $c$ , nous générons aléatoirement pour chaque élément  $e$  une valeur  $m$  dans un intervalle autour de la valeur de référence du couple  $(c, e)$ , cet intervalle étant d'amplitude proportionnelle à la valeur de référence. La valeur moyenne de la gaussienne représentant la proportion incertaine d'élément  $e$  dans  $x$  est fixée à  $m$ . Nous générons ensuite aléatoirement un écart-type pour cette gaussienne.

Nous créons ainsi, pour chaque classe, un ensemble d'exemples ayant pour chaque élément une proportion incertaine. On effectue alors une validation croisée sur le jeu complet de données obtenu, en divisant les ensembles correspondant à chaque classe en un nombre de plis donné, la distribution globale des classes est donc conservée dans les ensembles d'apprentissage et de test correspondant à chaque pli. Etant donné que les données réelles seront peu nombreuses en raison du coût financier et temporel des expérimentations, nous avons choisi de ne générer que dix exemples par classe, sur lesquels on effectue une validation croisée à cinq plis.



Pour simuler nos données, nous avons utilisé les formules brutes de 17 explosifs et 9 drogues. Nous travaillons donc sur 26 classes, représentées chacune par 10 exemples, soit 260 exemples. Il y a donc, pour chaque pli de la validation croisée, 208 exemples d'apprentissage et 52 exemples de test. Nous avons choisi de travailler avec un même nombre d'exemples pour toutes les classes car les différents types de matériaux feront l'objet du même nombre d'expérimentations, les classes seront donc représentées de manière équilibrée dans les données réelles.

## 4.2 Influence des paramètres

Le paramètre qui influence le plus clairement les résultats de classification de l'algorithme est le nombre de termes flous. Nous avons pour le moment utilisé le même nombre de termes flous pour chaque élément. Le graphique de gauche de la figure 8 représente l'évolution du taux de bonne classification en fonction du nombre de termes flous utilisés. Les "boîtes à moustaches" représentent les écart-types calculés sur les plis de la validation croisée. Il apparaît clairement que la performance de l'algorithme augmente lorsque le nombre de termes flous augmente. En revanche, le temps de calcul nécessaire pour l'apprentissage de l'arbre augmente beaucoup ; le nombre de calculs d'entropie est en effet polynomial par rapport au nombre de termes flous. Le temps d'apprentissage de l'arbre est multiplié par dix au passage de dix à douze termes. Nous présentons les prochains résultats de cet article en prenant dix termes flous.

Un autre paramètre important est le seuil de gain d'entropie à partir duquel on met fin à la construction de l'arbre. Afin de pouvoir étudier son influence de manière pertinente lorsque le nombre de termes varie, ce seuil est normalisé lors de l'analyse des résultats : on le divise par l'entropie maximale possible pour un nœud, c'est-à-dire au logarithme du nombre de termes flous utilisés. En faisant varier ce seuil, on obtient le graphique de droite de la figure 8 qui

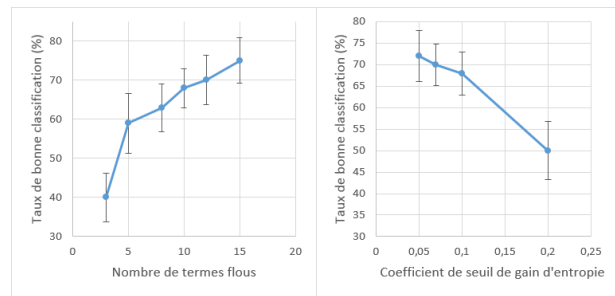


Figure 8 – Evolution du taux de bonne classification avec les paramètres

montre que la performance augmente lorsque le seuil de gain est diminué, ce qui correspond à un arbre plus profond.

## 4.3 Analyse des résultats

La figure 9 est une matrice de confusion résumant les résultats de l'algorithme pour dix termes flous et un coefficient de seuil de gain relatif (normalisé) de 0,1. On remarque que l'algorithme classe correctement les objets issus des premières classes (en partant du haut en ordonnée), qui correspondent à des explosifs, mais classe moins bien les drogues, qui correspondent aux dernières classes. Néanmoins, la grande majorité des drogues sont classées comme drogues ; en effet, notre algorithme discrimine très bien les explosifs des drogues, comme en témoigne le taux de bonne classification du problème binaire "explosif versus drogue" qui est de 96%. Cela peut s'expliquer par le fait que les drogues ont des teneurs nettement plus élevées en carbone que les explosifs, mais moindres en azote et en oxygène. Le taux moyen de bonne classification obtenu avec les 26 classes est de 70%. A titre de comparaison, l'algorithme CART de Breiman et al. donne un taux moyen de bonne classification de 56%.

Les branches de l'arbre obtenu ont une profondeur comprise entre 2 et 6, en moyenne 4, car parmi les 11 éléments chimiques disponibles, plusieurs sont peu discriminants.



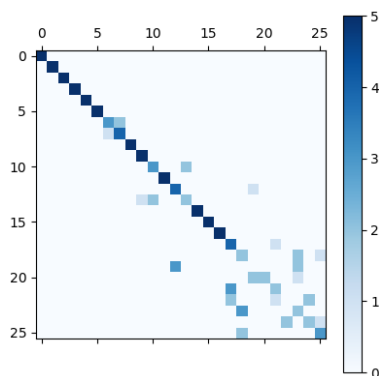


Figure 9 – Matrice de confusion

## 5 Conclusion

Dans cet article, nous avons proposé une méthode de reconnaissance des matériaux purs à partir de leur composition chimique. La particularité de la méthode que nous proposons est de prendre en compte l'incertitude liée aux données pendant la phase d'apprentissage et la phase de reconnaissance. En appliquant cette méthode, nous avons constaté qu'elle était bénéfique dans le cas où l'incertitude est assez importante.

Dans la suite de notre travail, nous pourrions éprouver notre méthode sur des données réelles (éléments purs, simulants de drogues, produits chimiques divers) issues des expérimentations qui permettront une évaluation plus précise de notre algorithme. Le cas traité par les données simulées est un cas simplifié par rapport aux données réelles prévues puisque, au sein d'un même voxel, plusieurs matériaux pourront être présents, ce qui augmente le nombre et la complexité des classes à apprendre, ainsi que leur proximité (par exemple du métal inséré dans une matrice de bois sera proche du bois pur en termes de composition chimique). Nous nous attendons de plus à des données bruitées, avec des incertitudes importantes, c'est d'ailleurs ce qui motive notre méthode de gestion de l'incertitude.

Par ailleurs, plusieurs pistes s'offrent à nous pour améliorer la performance de l'algorithme.

Nous pensons surtout à l'apprentissage des termes flous, qui est susceptible d'être mené par une technique de clustering. Il serait aussi judicieux de tester des techniques d'élagage de l'arbre, ainsi que d'employer un algorithme de forêt aléatoire. Enfin, nous devons également proposer des solutions pour les mélanges de matériaux.

### Remerciements :

Ces recherches sont financées par le projet H2020 C-BORD. Nous remercions Alix Sardet, Cédric Carasco, Bertrand Pérot, Guillaume Sannié et Vincent Picaud pour leurs travaux en amont des nôtres, pour leur disponibilité et leur expertise. Nous remercions plus généralement tous les partenaires du projet C-BORD.

### Références

- [1] T. Akiyama and H. Inokuchi. Application of fuzzy decision tree to analyze the attitude of citizens for wellness city development. In *2016 17th International Symposium on Advanced Intelligent Systems*, 2016.
- [2] M. Aupetit, L. Allano, I. Espagnon, and G. Sannié. Visual analytics to check marine containers in the eritr@c project. In *Proceedings of International Symposium on visual analytics science and technology*, pages 3 – 4, 2010.
- [3] R. L. P. Chang and T. Pavlidis. Fuzzy decision tree algorithms. *IEEE Transactions on Systems, Man and Cybernetics*, 7(1) :28 – 35, 1977.
- [4] A. de Luca and S. Termini. A definition of nonprobabilistic entropy in the setting of fuzzy sets theory. *Information and control*, 20, 1972.
- [5] W. Duch. Uncertainty of data, fuzzy membership functions, and multilayer perceptrons. *IEEE Transactions on Neural Networks*, 16(1) :10 – 16, 2005.
- [6] C. Z. Janikow. Fuzzy decision trees : issues and methods. *IEEE Transactions on Systems, Man, and Cybernetics - part B : Cybernetics*, 28(1), 1998.
- [7] C. Olaru and L. Wehenkel. A complete fuzzy decision tree technique. *Fuzzy sets and systems 138*, pages 221–254, 2003.
- [8] Y. Peng and P. A. Flach. Soft discretization to enhance the continuous decision tree induction. *Integrating Aspects of Data Mining, Decision Support and Meta-Learning*, 2001.
- [9] S. Tsang, B. Kao, K. Y. Yip, Wai-Shing, and S. D. Lee. Decision trees for uncertain data. *IEEE Transactions on Knowledge and Data Engineering*, 2009.
- [10] W. Yang, C. Marsala, A. Pinna, M. Rifqi, and P. Garda. Arbres de décision flous adaptatifs. In *25e rencontres francophones sur la logique floue et ses applications (LFA)*, 2016.
- [11] L. A. Zadeh. Probability measures of fuzzy events. *Journal of mathematical analysis and applications*, 23 :421–427, 1968.