



Fuzzy Rule Learning for Material Classification from Imprecise Data

Arnaud Grivet Sébert, Jean-Philippe Poli

► To cite this version:

Arnaud Grivet Sébert, Jean-Philippe Poli. Fuzzy Rule Learning for Material Classification from Imprecise Data. IPMU 2018: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Jun 2018, Cadiz, Spain. pp.62-73, 10.1007/978-3-319-91473-2_6 . cea-01838452

HAL Id: cea-01838452

<https://cea.hal.science/cea-01838452>

Submitted on 25 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.

Fuzzy rule learning for material classification from imprecise data

Arnaud Grivet Sébert and Jean-Philippe Poli

CEA, LIST, Data Analysis and System Intelligence Laboratory,
91191 Gif-sur-Yvette cedex, France.

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

Abstract. A complete method for classification of materials from imprecise chemical data is proposed. The method consists in a clustering method to find the suitable terms of the various linguistic variables and in a decision tree induction to learn rules for an explainable classification. Both methods are adaptations from classic methods to the case of imprecise data. Finally, results on simulated data are presented in the expectation of real data.

Keywords: Fuzzy partitioning, clustering, fuzzy decision tree, fuzzy rules, imprecise data, explainable material classification

1 Introduction

Customs and ports security is a major issue in Europe. Indeed, many illegal or dangerous substances such as drugs, weapons, explosives pass through customs. Unfortunately, systematic container inspections are impossible in practice because of the cost and time that would be required. The volumes passing daily through the major European ports such as Rotterdam, Antwerp or Hamburg are indeed enormous: for example, 461.2 million tons of goods passed through the port of Rotterdam in 2016.

In this paper, we use tagged neutrons to obtain the chemical composition of a volume of the container. From this chemical composition, we want to determine the materials present in the container. In order to bring more credibility to the final software used by the customs officers, we will also provide a justification for this classification. Fuzzy rules allow to avoid the “black box” effect since customs officers have access to a real explanation of the classification made by the software, and close to natural language : “the container may contain drug (confidence degree : x) because the quantity of carbon is high, the quantity of nitrogen is low and the quantity of oxygen is medium” for instance.

The proportions are obtained by different treatments that are beyond our control. Thus, the input data are imprecise and are accompanied by a measure of this inaccuracy. Fuzzy logic thus seems appropriate for the exploitation of such data.

Given the time required and the authorizations needed to use a neutron generator, we will have a small learning dataset, even if all classes of relevant

materials will obviously be represented. The idea is therefore to use fuzzy decision trees [4], which have been applied successfully on various classification problems [1, 13]. The scarcity of training data and the intrinsic inaccuracy due to the preprocessing preclude conventional statistical learning approaches such as neural networks, SVM, etc. which also do not provide an explanation to users.

In this article, we adapt to imprecise data the classic two steps workflow consisting in using clustering methods to create relevant terms from data and then in building a decision tree to get fuzzy classification rules.

The paper is structured as follows: section 2 describes the context of the application that motivates this work. Then, section 3 describes the method to induce rules from imprecise data. Section 4 presents the results of the different experiments we conducted. Finally, section 5 draws the conclusions of this paper.

2 Application context

2.1 Neutron inspection for container digging

The H2020 project C-BORD (effective Container inspection at BORDer control points) aims at facilitating the digging of containers at borders by exploiting different technologies: e-noses, X-rays, photo-fission and tagged neutrons. The goal is to detect dangerous (explosives, nuclear materials, ...) or illicit (drugs, contraband, ...) substances.

In this paper, we focus on the classification of materials with tagged neutron inspection. As shown in figure 1a, a device produces a neutron beam to focus on a certain volume (called voxel) of the container. The neutrons interact with the nuclei of the atoms contained in the voxel, producing new particles that can be detected by the matrix sensors which are positioned on the side of the container. These particles are thus characteristic of the atoms encountered in the examined voxel.

The processing of the raw data is not the topic of this article but we quickly describe the principle in figure 1b. After different pretreatments, a global spectrum is obtained. In comparison with the characteristic spectra of each of the studied atoms, this spectrum is decomposed into individual spectra by a Bayesian process which makes it possible to deduce the chemical composition of the voxel, expressed in percentages. This process is based on simulation and we can easily get the mean and the standard deviation for each proportion in order to characterize the inaccuracy of the reconstruction. Figure 2 shows the result of these treatments for an exposure to ceramics, and in which the inaccuracy is represented by “box plots”.

Our work consists in exploiting this information in order to recognize the materials contained in the voxel.

2.2 Previous work

The difficulty of recognizing materials from the chemical composition lies in several points :

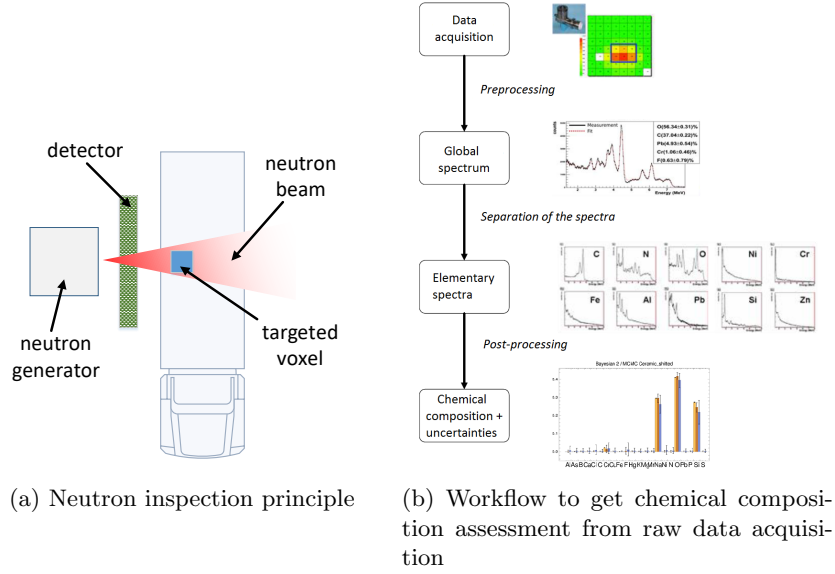


Fig. 1: Neutron inspection workflow

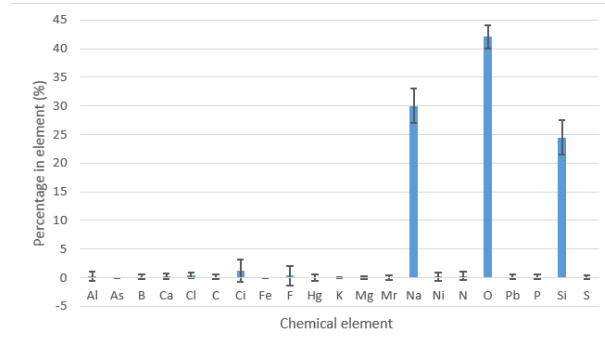


Fig. 2: Chemical composition obtained from ceramics simulation

- Data is scarce because few acquisition campaigns can be conducted. That also explains why few papers address the recognition of materials from their chemical composition.
- The inaccuracy of the chemical composition makes the task difficult for conventional statistical models
- The device is insensitive to hydrogen atoms.

For these reasons, previous works proposed visual analytics methods to represent the content of the voxels. In [2], the authors proposed a Voronoi diagram

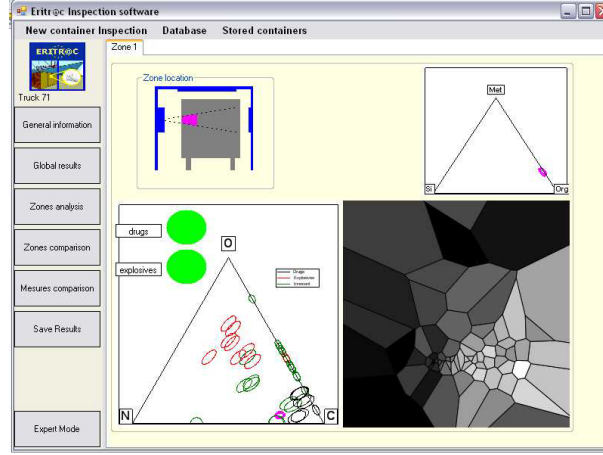


Fig. 3: Screenshots of the different visualizations introduced in [2]

to highlight the proximity, in terms of chemical composition, of the current voxel with voxels previously inspected and whose content is known (see figure 3). Thus, it is not a question of recognizing the materials present in the container but of displaying visually close and known containers in order to deduce the contents.

This approach has the advantage of not requiring learning or parameterization since it relies on the manual selection of a neighborhood. In figure 3, we can also see two classical representations that have been used in conjunction with this method. These are projections of the current voxel on two triangles: a triangle called “materials triangle” indicates the proximity of the voxel with metals, ceramics and organic materials, while a so-called “alert triangle” presents the ratios between carbon, nitrogen and oxygen. This last triangle makes it possible to distinguish between drugs and explosives. The main drawback of the visual analytics approach is that the operator must be able to interpret the different representations himself.

In practice, the mastering of these representations, particularly the Voronoi graph, can be difficult for operators who are not familiar with these visualization techniques. As part of the C-BORD project, we want to go further and propose a list of materials, and an explanation of the decision. It is to overcome these different difficulties that we want to use a fuzzy expert system.

3 Rule induction for classification

In order to effectively classify the materials while generating rules which are understandable by a human, we chose to use a fuzzy decision tree inspired by the one defined by C. Z. Janikow in [11]. Each node is split, by a feature which was not used yet, into N child-nodes corresponding to the associated fuzzy terms. The maximum depth of the tree is thus the number of available features if no

pruning is performed. In our case, the features are the different chemical elements the system can detect.

The rule induction follows a classic 2 steps workflow but the different algorithms have been adapted to take into account data imprecision: firstly, the method extracts from data the various terms that will be involved in the decision tree with clustering methods. Then, the tree is inducted. Finally, rules are created and used in a fuzzy expert system to perform the classification of new materials.

3.1 Fuzzy partition learning

To create rules which will provide the end-user with an explanation of the material classification, we have to define terms which will enable the building of the tree and will then appear in the premises of the rules. Nevertheless, crisp terms would not represent the physical reality for which strict boundaries between different compounds are irrelevant. That is why fuzzy terms are used. Except for the extreme terms which are trapezoidal, we opted for classical triangular overlapping terms that form a strong partition of the definition domains of the stoichiometric percentage of each element, namely $[0, 100]$. Triangular-shaped terms are used because more complex shapes, like trapezoidal or pentagonal shapes, reduce the performance and even the accuracy in some cases.

The best methods to learn linguistic variables terms turned out to be clustering approaches. For example, using entropy-minimization-based methods gave poorer results, due to the difficulty to minimize the fuzzy entropy (see 4) which is not a convex function of the fuzzy sets parameters.

The basic idea is to cluster the data, feature by feature (chemical element), as if they were one-dimensional data, and build fuzzy terms over the resulting clusters. Once the clusters are built, the mode of a fuzzy triangle is simply set to the mean of its corresponding cluster ; the spread is then induced by the other terms' modes through the constraint of the strong partition.

We tested existing algorithms and adapted them to the fuzzy case by the use of dissimilarities between distributions able to take into account the whole distribution of the data and not only an aggregated value. Distances or dissimilarities have already been defined and used with imprecise or uncertain data. A distance consisting in the sum of the center Euclidian distance and the spread Euclidian distance of the imprecise data was used in [5] and [7] but this implies a loss of the information provided by the whole distribution. Moreover, even though it might have been replaced by the deviation distance, the spread distance does not make a lot of sense in the case of normal distributions whose spreads are theoretically infinite, and in our case imposed by the domain boundaries. Gullo et al. [9] proposed an uncertain dissimilarity which corresponds, in the univariate case, to compute the double integral of the Euclidian distance of each pair $((x, f_1(x)), (x, f_2(x)))$ of points of the two distributions. We defined a simpler dissimilarity, in order to improve the performance.

The dissimilarity used is inspired by the Jaccard distance : $d(A, B) = 1 - \frac{|A \cap B|}{|A \cup B|}$, where A and B are two sets. The sets are here replaced by the probability

distributions representing two imprecise data, the cardinality by the integral of the distribution, the intersection and union respectively by the min and max. Thus, for two imprecise data x and y represented by the densities of probability f_x and f_y respectively, the dissimilarity becomes : $d(x, y) = 1 - \frac{\int_a^b \min(f_x(t), f_y(t))}{\int_a^b \max(f_x(t), f_y(t))}$, where a and b , elements of $R \cup \{-\infty, +\infty\}$, are the bounds of the definition domain of the data, 0 and 100 in our case. Since each distribution is built in such a way that its integral on $[0, 100]$ is equal to 1, we have :

$$\int_a^b \max(f_x(t), f_y(t)) + \int_a^b \min(f_x(t), f_y(t)) = \int_a^b f_x(t) + \int_a^b f_y(t) = 2$$

Thus, $d(x, y) = 1 - \frac{\int_a^b \min(f_x(t), f_y(t))}{2 - \int_a^b \min(f_x(t), f_y(t))}$. This formula is quite close and gives similar results to the simpler one :

$$d(x, y) = 1 - \int_a^b \min(f_x(t), f_y(t)) \quad (1)$$

Hence, we chose to use this second formula to define the dissimilarity between two imprecise data¹. Figure 4 shows the similarity between two distributions, drawn in red.

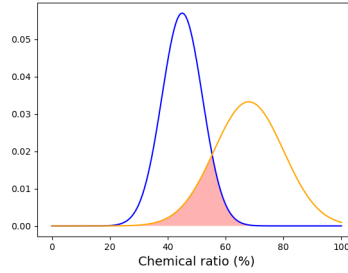


Fig. 4: Similarity between two imprecise data

We also tried to use the T-norm product instead of the min in the definition of the dissimilarity and it gave very similar results.

We then ran the k-medoids algorithm [12] using this dissimilarity instead of the traditional Euclidian distance which would have only used the information given by the aggregated value of each distribution - generally the mean.

3.2 Fuzzy decision tree induction

To build the tree while taking into account the imprecision of the data, we represent these data by Gaussian probability distributions. Since the definition

¹ This dissimilarity is actually a distance for continuous distributions but we will not show it in this paper for reasons of space.

domain of the data (proportion in each chemical element) is bounded ($[0, 100]$), the Gaussian distributions are transformed so that their integral on this bounded domain equals 1 and that the distribution thus keeps its probabilistic sense on the interval $[0, 100]$ beyond which the values, which are percentages, do not have any sense. To do so, each distribution is divided by its integral on $[0, 100]$: $f(t) = norm(t) * \frac{\int_{-\infty}^{+\infty} norm}{\int_0^{100} norm} = \frac{norm(t)}{\int_0^{100} norm}$ where $norm$ is the probability density of the normal law.

We also tried a possibilistic approach, representing the imprecise data by triangular possibilistic distributions but the accuracy was lower, that is why we chose to focus on the probabilistic approach in this paper.

Once the fuzzy terms are learnt, we have to define how to compute the membership degrees of the imprecise examples, modeled by distributions, to each of these terms, which represent the nodes of the tree. We adapted integration techniques for uncertain data and crisp terms ([6], [15]) to our imprecise data and fuzzy terms - the membership degree is defined as the integral of the product of the density of the imprecise distribution with the membership function of the fuzzy term. If f is the density of the distribution representing the imprecision associated with the proportion of an element e in an example x , and $\mu_v(t)$ the membership degree of the value t to the fuzzy term v , the membership degree of the imprecise example x to the fuzzy term v is :

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

For the sake of reading, we will write μ_v (without tilde) in the following to refer to the membership degree of an imprecise example.

This integral, illustrated in figure 5 where the imprecision distribution is dilated by a factor 10 for a better visibility, takes into account the whole continuous spectrum of probable values for the proportion of e in x , and is basically the weighted average of the membership degrees of these probable values to the fuzzy term n .

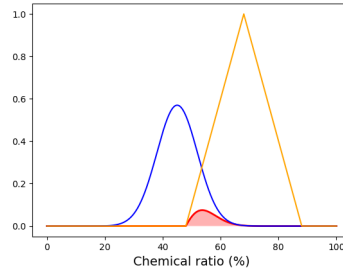


Fig. 5: Integral of the product imprecision-fuzzy term

Given the membership degree of an imprecise example x to the fuzzy terms corresponding to each chemical element, we can define the membership degree of x to each node of the tree. To do so, we chose the product as a T-norm, because it gave better results than other T-norms, like the minimum. Hence, the membership degree of an example x to a node n is the product of the membership degrees of x to the fuzzy terms associated to n and to the ascendants of n . Thus, we have the recursive definition :

$$\begin{cases} d_{n_0}(x) = 1 \\ d_n(x) = d_N(x) \times \mu_v(x) \end{cases} \quad (2)$$

where n_0 is the root of the tree, N is n 's father node, v is the fuzzy term associated to the node n and $\mu_v(x)$ is the membership degree of x to the fuzzy term v .

To select the next feature to use, we use the fuzzy entropy introduced by Peng et al. in [14].

We consider the "fuzzy frequency" of a class c in a node n :

$$fr_{c/n} = \frac{\sum_{x_i \in c} d_n(x_i)}{\sum_{x_i} d_n(x_i)} \quad (3)$$

where $\{x_i, i \in [1, T]\}$ is the training set and the "membership frequency" of the examples to the node n : $fr_n = \frac{\sum_{x_i} d_n(x_i)}{\sum_{n_j} \sum_{x_i} d_{n_j}(x_i)}$ where $\{n_j, j \in [1, F]\}$ is the set of the "siblings" nodes of n , including n .

The fuzzy entropy, according to an element e , used at a node N having the children $\{n_j, j \in [1, F]\}$, these nodes corresponding to each fuzzy term relative to e , is then defined as :

$$E = - \sum_{n_j} fr_{n_j} \sum_{c_k} fr_{c_k/n_j} \log(fr_{c_k/n_j}) \quad (4)$$

where $\{c_k, k \in [1, K]\}$ is the set of the classes of the problem.

When the tree is built, we create a rule for each of its leaves : the premise is the conjunction of the fuzzy terms associated with the nodes leading from the root to the leaf, and the consequence is the most represented class in the leaf.

3.3 Recognition of new samples

Once the tree has been constructed using the training data, the classification of the testing data is quite simple. Given an imprecise example x , we compute the membership degree of x to each leaf l of the tree, using the very same definition we used for the training data (equations 2). For each class c , this membership degree is multiplied by the fuzzy frequency $fr_{c/l}$ of c conditionally to l (see definition 3), using the same formula as in [3], $fr_{c/l}$ being in fact the certainty factor defined in [3]. Using the weighted voting method [10], these results are then summed up for each class, over all the leaves, to obtain the confidence degree $conf(x \in c)$ of $x \in c$. The algorithm finally classifies x in the class c maximizing $conf(x \in c)$.

4 Experiments

4.1 Data simulation

Since the physical experiments supposed to provide us with real data could not be realized at the moment of the writing of this paper, we generated simulated data. To do so, we used the theoretical stoichiometric percentages of each element in each material (class) which provided us with reference values for each pair (class, element). To create an example x of a class c , a value m is randomly generated for each element e in an interval around the reference value of the pair (c, e) , whose span is proportional to the reference value, the proportionality coefficient being called *degree of imprecision* in the following. The mean value of the Gaussian representing the imprecise proportion of e in x is set to m . A standard deviation is then randomly generated in an interval whose span is proportional to the one of the interval used to generate the mean.

To generate the data, we used the chemical formulæ of seventeen explosives and nine drugs, with a degree of imprecision of 15%. Since the real data will be few due to the financial and temporal costs of the physical experiments, we chose to generate only ten examples per class to create a data set. Then, a five-fold cross-validation is performed on the set. The following results are averaged on the tests of ten different data sets.

4.2 Comparison of fuzzy partitioning algorithms

One of the most influential parameters of the fuzzy partition in terms of accuracy of the whole classification process is the number of fuzzy terms constituting the partitions of each chemical ratio domain. Thus, we ran tests with k-medoids algorithm using the previously defined dissimilarity (see 1) over the grid of all the combinations of numbers of terms between 2 and 14, for elements C, N and O only, imposing five fuzzy terms for the other elements. Indeed, C, N and O ratios are the most discriminative features in our problem so we focused on them to run quicker tests. The graphic 6 shows the dependence of the correct classification rate of the tree on the numbers of fuzzy terms partitioning C, N and O chemical ratio domains. We can easily see that the accuracy increases when the number of fuzzy terms increases, until a certain threshold. For N and O, this threshold is about 12, whereas it is surprisingly low for C - around 6. The best results are actually obtained for number of terms between 5 and 7 for C, and between 11 and 14 for N and O - several of these combinations give the best classification rates, between 84.5% and 85%.

We then tested the clustering algorithm called affinity propagation [8], using the similarity associated with the previously defined dissimilarity : $s(x, y) = \int_a^b \min(f_x(t), f_y(t))$. Affinity propagation does not take the number of clusters (thus fuzzy terms) in parameter but a parameter called preference that increases the number of clusters when it is high. Giving the same preference to the clustering for the partitions of C, N and O ratios give worse results than the best obtained with k-medoids. But in the light of k-medoids tests, we tried to give

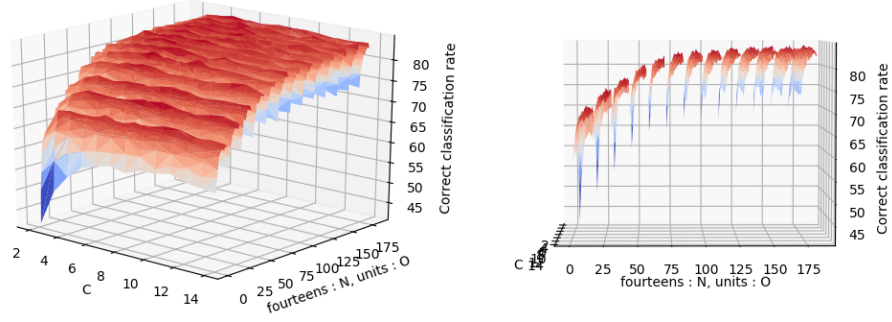


Fig. 6: Correct classification rate against number of fuzzy terms

smaller preference to C partition and greater for N and O partitions, and results similar to the best ones of k-medoids were obtained (between 84% and 85%).

We can conclude that k-medoids algorithm is preferable for this problem since it gives results as good as affinity propagation but the parameter to be optimized - the list of the numbers of fuzzy terms - is discrete whereas affinity propagation preference is continuous.

In order to improve the robustness of our method, it would have been interesting to split the training set in two parts in order to learn the fuzzy terms from different data from the ones used to build the tree. But since our data are very few, we decided to learn the fuzzy terms and the tree from the whole training set.

4.3 Results analysis

From the grid-based test over the combinations of numbers of fuzzy sets for C, N and O, we determined the best combination in terms of accuracy : respectively 5, 14 and 12 terms for C, N and O. Figure 8 is a confusion matrix that summarizes the results of the classification using k-medoids with this optimal combination of numbers of fuzzy sets regarding C, N and O, and 5 fuzzy terms for the other elements. We note that the algorithm classes quite correctly both the explosives (seventeen first classes from the top in ordinate) and the drugs (six last classes). This method does represent an improvement compared to the classic algorithms which do not take the imprecision of the data into account. As we can see on figure 7, the correct classification increases up to 14%, this difference being higher for more imprecise simulated data.

We ran the method with the same numbers of fuzzy terms on a more complex data set composed of 38 classes of drugs, explosives and benigns. The accuracy obtained is 75% (see the confusion matrix on figure 9).

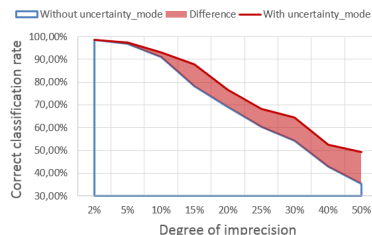


Fig. 7: Gain obtained when taking the imprecision of the data into account

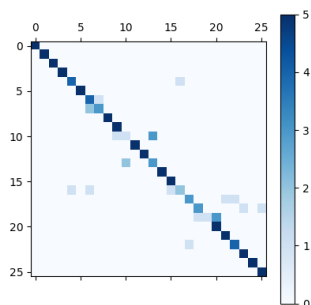


Fig. 8: Confusion matrix for drugs and explosives

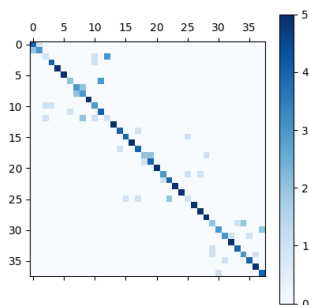


Fig. 9: Confusion matrix for drugs, explosives and benigns

5 Conclusion

In this article, we introduced a workflow for explainable material recognition from their chemical composition and based on both a decision tree induction and linguistic variables learning. The method takes into account the imprecision of data during the learning phase and the exploitation of the decision tree. It performs better for high imprecision: the classification rate is up to 84% for a classification into 26 classes.

We have several opportunities to enhance our method's accuracy in further work. We may improve our clustering technique taking the samples' labels into account. Using pruning techniques but also random forests could make the method more accurate. We will also study more sophisticated techniques than the sum we are using for aggregating the results of each leaf of the tree. As far as the rules are concerned, we might generate simpler rules corresponding to intern nodes of the tree. We shall also test the possibilistic approach with other measures and methods more adapted to it.

At this stage of our work, the method has been evaluated on simulated data while we are performing real data acquisition on pure elements, chemical products, drug and explosive simulants. This will be a first step since in real life

container voxels, several materials will be present. We will then have to deal with recognition of mixtures and not only mono-material samples.

Acknowledgment

This research has been funded by the project H2020 C-BORD. We warmly thank S. Moretto, C. Fontana, F. Pino, A. Sardet, C. Carasco, B. Prot and V. Picaud for their contributions before our work, for their availability and their expertise.

References

1. Akiyama, T., Inokuchi, H.: 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. Aupetit, M., Allano, L., Espagnon, I., Sannié, G.: Visual analytics to check marine containers in the eritr@c project. In: Proceedings of International Symposium on visual analytics science and technology. pp. 3 – 4 (2010)
3. Bounhas, M., Prade, H., Serrurier, M., Mellouli, K.: A possibilistic rule-based classifier. In: IPMU 2012 : Advances on computational intelligence (2012)
4. Chang, R.L.P., Pavlidis, T.: Fuzzy decision tree algorithms. *IEEE Transactions on Systems, Man and Cybernetics* 7(1), 28 – 35 (1977)
5. Coppi, R., d’Urso, P.: Fuzzy time arrays and dissimilarity measures for fuzzy time trajectories. In: Data analysis, classification, and related methods (2000)
6. Duch, W.: Uncertainty of data, fuzzy membership functions, and multilayer perceptrons. *IEEE Transactions on Neural Networks* 16(1), 10 – 16 (2005)
7. d’Urso, P., de Giovanni, L.: Robust clustering of unprecise data. *Chemometrics and Intelligent Laboratory Systems* (2014)
8. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. *Science* (2007)
9. Gullo, F., Ponti, G., Tagarelli, A.: Clustering uncertain data via k-medoids. In: Scalable Uncertainty Management: Second International Conference, SUM 2008, Naples, Italy, October 1-3, 2008, Proceedings (2008)
10. Ishibuchi, H., Nakasima, T., Morisawa, T.: Voting in fuzzy rule-based systems for pattern classification problems. In: Fuzzy sets and systems (1999)
11. Janikow, C.Z.: Fuzzy decision trees : issues and methods. *IEEE Transactions on Systems, Man, and Cybernetics - part B : Cybernetics* 28(1) (1998)
12. Kaufman, L., Rousseeuw, P.J.: Clustering by means of medoids. In: Statistical data analysis based on the L1 norm and related methods (1987)
13. Olaru, C., Wehenkel, L.: A complete fuzzy decision tree technique. *Fuzzy sets and systems* 138 pp. 221–254 (2003)
14. Peng, Y., Flach, P.A.: Soft discretization to enhance the continuous decision tree induction. *Integrating Aspects of Data Mining, Decision Support and Meta-Learning* (2001)
15. Tsang, S., Kao, B., Yip, K.Y., Wai-Shing, Lee, S.D.: Decision trees for uncertain data. *IEEE Transactions on Knowledge and Data Engineering* (2009)