



**HAL**  
open science

## Evidential classification of incomplete data via imprecise relabelling: Application to plastic sorting

Lucie Jacquin, Abdelhak Imoussaten, François Trouset, Jacky Montmain,  
Perrin Didier

### ► To cite this version:

Lucie Jacquin, Abdelhak Imoussaten, François Trouset, Jacky Montmain, Perrin Didier. Evidential classification of incomplete data via imprecise relabelling: Application to plastic sorting. SUM 2019 - The 13th international conference on Scalable Uncertainty Management, Dec 2019, Compiègne, France. 10.1007/978-3-030-35514-2\_10 . hal-02299144

**HAL Id: hal-02299144**

**<https://hal.science/hal-02299144>**

Submitted on 18 Feb 2020

**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.

# Evidential classification of incomplete data via imprecise relabelling: Application to plastic sorting

Lucie Jacquin<sup>1</sup>, Abdelhak Imoussaten<sup>1</sup>, François Troussset<sup>1</sup>, Jacky Montmain<sup>1</sup>,  
and Didier Perrin<sup>2</sup>

<sup>1</sup> LGI2P, IMT Mines Ales, Univ Montpellier, Ales, France

<sup>2</sup> C2MA, IMT Mines Ales, Univ Montpellier, Ales, France  
{firstname.name}@mines-ales.fr

**Abstract.** Besides the ecological issues, recycling of plastics involves economical matters that encourage industrial firms to invest in the field. Part of them have focused on the waste sorting phase by designing optical device able to discriminate on line among plastics categories. For achieving ecological and economical objectives, sorting errors must be minimized to avoid serious recycling problems and significant quality degradation of the final recycled product. Even with the most recent acquisition technologies based on spectra imaging, plastic recognition remains a tough task due to the presence of imprecision and uncertainty, e.g., variability in the measurement due to atmospheric disturbances, ageing of plastics, dark or black coloured materials etc. The enhancement of the recent sorting techniques based on classification algorithms leads to rather good performance results, however for such applications, the remaining errors have serious consequences. In this article, we propose an imprecise classification algorithm to minimize sorting errors of standard classifiers when dealing with incomplete data by both integrating the processing of classification's doubt and hesitation in the decision process and improving the classification performances. To this aim, we propose a relabelling procedure that allows to better represent the imprecision of the learning data and we introduce the belief functions framework to represent the posterior probability provided by a classifier. Finally, the performances of our approach compared to existing imprecise classifiers is illustrated on the sorting problem of four plastics categories from mid-wavelength infra-red spectra acquired in an industrial context.

**Keywords:** Machine Learning · Imprecise Classification · Reliable Classification · Belief Functions · Plastic Separation

## 1 Introduction

Plastic recycling is a promising alternative to landfills for facing up to the fastest growing waste stream in the world [7]. However for physico-chemical reasons related to the non-miscibility between plastics, most plastics must be recycled

separately. Plastic category identification is then a major challenge in the recycling process. With the emergence of hyperspectral imaging, some industrial firms have designed sorting devices able to discriminate several categories of plastics based on their absorption or transmittance spectra. The sorting process is generally performed with supervised classification that has been well developed with the emergence of computer sciences and data science [17, 38, 22]. The classification performance might be affected by several issues such as noises or overlapping regions in the feature space [34, 21]. The latter problem occurs when some samples from different classes share very similar characteristics. We are particularly faced with these problems when attempting to classify industrially acquired spectra. Indeed, within industrial context, the acquisition process is subject to technical and financial constraints in order to ensure throughput and financial competition. For this reason one cannot expect the same quality of data as for equivalent laboratory measures. Several issues imply the presence of imprecision and uncertainty in the acquired spectra : i) available spectral range might be insufficient ; ii) plastic categories to recycle are chemically close ; iii) atmospheric perturbations may cause noises ; iv) plastics ageing and plastics additives are known to change spectral information ; v) impurities like dust deposits or remains of tags will also produce spectral noises. As in many other decision problems solving, errors of classification may have serious consequences, *e.g.*, medical diagnosis applications. Regarding plastic sorting, errors of identification will cause serious recycling difficulties and significant degradation of the secondary raw material performances and thus quality degradation of the recycled products. Usually, the problem of plastic identification is treated using standard classification algorithms that are designed to produce point predictions, *i.e.*, a single plastic category. In the cases of imperfect data, standard classifiers get confused and commit inevitable errors. This brings us to consider alternative representation of the information that take into account imprecision and uncertainty to achieve more accurate classification. Modern theories of uncertainty such as fuzzy subsets [35], possibility theory [13], imprecise probabilities [33] or belief function [26] [30] offer better representations of the data imperfection of information. Several classification algorithms have been proposed in these frameworks. The majority of them are extensions of standard algorithms. We can cite the fuzzy version of the well known  $k$ -means algorithm [14], fuzzy and evidential versions of  $k$ -Nearest Neighbour ( $k$ -NN) [18, 9] or even some fuzzy and evidential revisions of neural networks algorithms [19, 10].

In this paper we consider the case where original imperfections come from data feature only. Available training examples labels are precise and considered trustworthy, *e.g.*, based on laboratory measures and expertise. In order to better represent all available information, we think that labels should conform with the feature imprecision. Say if an object of class  $\theta_1$  has its vector of features  $x$  in the overlapping region  $\theta_1$  and  $\theta_2$ , then the example should be relabelled by the set  $\{\theta_1, \theta_2\}$ . In order to achieve such representation we propose to relabel each training example in accordance with their discriminatory nature. Hence new labels are subsets of the original set of classes. This imprecise relabelling

would actually permit to better represent the learning data by mapping overlaps in the feature space. The resulting imprecise label can be naturally treated in the belief function theory context. Indeed the belief function theory [26] is an interesting framework to represent imprecise and uncertain data by allowing the allocation of a probability mass for imprecise data. Thus, imprecision is better captured in this framework compared to the probability framework where equiprobability and imprecision are confused. The recent growing interest to this theory has allowed to develop techniques to resolve a diverse panel of problems such as estimation [11, 16], standard classification [9, 32], or even hierarchical classification [1, 23].

Our proposed approach, called Evidential CLAssification of incomplete data via Imprecise RElabelling (ECLAIRE), is based on a relabelling procedure of the training examples that allows to better represent the missing information about some data features. Then a classifier is trained on the relabelled data producing a posterior belief function. With imprecise relabelling we try to quantify, using belief function, the extend to which a subsets of classes is reliable and relevant as output for a new data. In other words, we look for the set of classes which any more precise subset output would lead inevitably to an error. The resulting classification algorithm can enhance the classification accuracy as well as cope with difficult examples by allowing less precise but more reliable classification output which will optimize the recycling process.

The remainder of this paper is organized as follows: section 2 sets the main notations and provides a reminder on supervised classification and elements of belief function theory; in the section 3 we present the proposed approach ; the section 4 briefly describes the related works ; section 5 presents results of experimentation on the sorting problem of four plastics.

## 2 Theoretical background

Classification is a technique allowing to assign objects to categories from the observations of several of their characteristics. A classifier is a function that maps an object represented by its values of characteristics on a finite set of variables, to a category represented by a value of a categorical variable. More precisely, let us consider a set of  $n$  categories represented by a set  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , also refereed as a set of labels or classes. In the framework of belief function  $\Theta$  is called a frame of discernment. Each  $\theta_j$ ,  $j \in \{1, \dots, n\}$  denotes a singleton which represents the lowest level of discernible information in  $\Theta$ . Let us denote by  $X_1, X_2, \dots, X_p$ ,  $p$  variables where the taken values represent the characteristics, also called attributes or features, of the objects, to be classified. In the rest of the paper we refer to  $\Theta$  as a set of classes and to  $(X_1, X_2, \dots, X_p)$  as a vector of features where  $\forall i \in \{1, \dots, p\}$ ,  $X_i$  refers both to the name of the feature and to the space of the values taken by the feature, *i.e.*,  $X_i \subseteq \mathbb{R}$ . For an object  $x$  belonging to  $\mathcal{X} = \prod_{i=1}^p X_i \subseteq \mathbb{R}^p$ , let  $\theta(x) \in \Theta$  denote the unknown label that should be associated to  $x$ .

In this article, we focus on a supervised classification problem, i.e., the classifier maps new examples by inferring from labelled training examples. The specificity of the considered data, referred as *incomplete data*, is that some features are missing. The proposed classification approach qualified as *imprecise* integrates the incompleteness of the data in its process to predict subsets of classes comprising the true class when standard counterpart classifier would have predicted the wrong class. To this aim we diverted standard probabilistic classifiers from their natural use for computing probability on sets of classes. Such uncertain resulting information is then captured by belief functions. The following subsections, briefly recalls the notions discussed.

## 2.1 Supervised classification

To determine  $\theta(x)$  in a supervised classification manner, a standard classifier  $\delta_\Theta : \mathcal{X} \rightarrow \Theta$  is trained on a set of examples  $(x_i, \theta_i)_{1 \leq i \leq N}$  such that for all  $1 \leq i \leq N$ ,  $x_i$  belongs to  $\mathcal{X}$  and  $\theta_i$  to  $\Theta$ . By standard classifier we mean a classifier that assigns to  $x$  a single class  $\theta(x) = \theta_j, j \in \{1, \dots, n\}$ . In some cases when the input data is too large or redundant, it may be appropriate to perform some extraction features before the training of  $\delta_\Theta$ . By reducing the dimension of  $\mathcal{X}$ , i.e., working with a reduced feature space  $\mathcal{X}' \subseteq \mathbb{R}^{p'}$  with  $p' < p$ , the extraction such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) or Independent Component Analysis (ICA) facilitates the learning and may enhance the classification performance. When feature extraction is designed taking into account the labels of the training examples it is termed as supervised feature extraction, e.g., LDA also known as Fisher discriminant analysis reduces the number of features to  $n-1$  by looking for a linear combination of the variables maximizing the within-groups and minimizing between-groups variance.

## 2.2 Probabilistic classifier and decision rule

When  $\delta_\Theta$  can also provide for  $x$  a *posterior probability* distribution  $p(\cdot|x) : \Theta \rightarrow [0, 1]$ , it is called a probabilistic classifier. Many classifier algorithms base their decision only on  $p(\cdot|x)$  as follows:  $\theta(x) = \arg \max_{j=1, \dots, n} p(\theta_j|x)$ . For more sophisticated decision, one can use decision rule technique classically used in decision theory. Let  $\mathcal{A} = \{a_1, a_2, \dots, a_m\}$  be a finite set of actions that can be taken. In the case of a standard classifier, an action  $a \in \mathcal{A}$  corresponds to assign a class  $\theta \in \Theta$  to an object  $x$ . In such case, we simplify by setting  $\mathcal{A} = \Theta$ . In order to compare decisions in  $\mathcal{A}$  or to compare the classifier  $\delta_\Theta$  to another decision rule, two functions are introduced: *loss function* and *risk function*. A loss function  $L : \mathcal{A} \times \Theta \rightarrow \mathbb{R}$  is considered to quantify the loss  $L(a, \theta)$  incurred when choosing the action  $a \in \mathcal{A}$  while the true state of nature is  $\theta \in \Theta$ . A *risk function*  $r_{\delta_\Theta} : \mathcal{A} \rightarrow \mathbb{R}$  is defined as the following expectation:  $r_{\delta_\Theta}(a) = E_{p(\cdot|x)}(L(a, \theta))$ . In the case of discrete and finite probability distribution, we have  $r_{\delta_\Theta}(\theta_j) = \sum_{k=1}^n L(\theta_j, \theta_k) p(\theta_k|x)$ ,  $j \in \{1, \dots, n\}$ . Thus, considering the decision rule  $\delta_\Theta$ , the class  $\theta_j$  minimizing

the risk  $r_{\delta_{\Theta}}(\theta_j)$  over  $\Theta$  should be chosen. Note that the decision rule based only on  $p(\cdot|x)$  can be met if the loss function  $L$  is defined as  $L(\theta_j, \theta_i) = 1$ , if  $i \neq j$ , and  $L(\theta_j, \theta_j) = 0$ .

### 2.3 Elements of the belief function theory

Due to the additivity constraint inherent to the definition of a probability distribution, one cannot build a probability distribution when measures, observations, etc. are imprecise. The belief function theory as an extension of probability theory allows to assign masses to imprecise data. Two levels are considered when introducing belief functions: credal and pignistic levels. In the credal level, beliefs are captured and quantified by belief functions, while in the pignistic level or decision level, beliefs are quantified using probability distributions.

**Credal level** A *mass function*, also called *basic belief assignment* (bba), is a set function  $m : 2^{\Theta} \rightarrow [0, 1]$  satisfying  $\sum_{A \subseteq \Theta} m(A) = 1$ . For a set  $A \subseteq \Theta$ , the

quantity  $m(A)$  is interpreted as a measure of evidence committed exactly to the set  $A$  and not to any more specific subsets of  $A$ . The elements  $A \in 2^{\Theta}$  such that  $m(A) > 0$  are called focal elements and they form a set denoted  $\mathbb{F}$ .  $(m, \mathbb{F})$  is called body of evidence. The total belief committed to  $A$  is measured by the sum of all masses of  $A$ 's subsets. This is expressed by the *belief function*  $Bel : 2^{\Theta} \rightarrow [0, 1]$ ,  $Bel(A) = \sum_{B \subseteq \Theta, B \subseteq A} m(B)$ . Furthermore the *plausibility* of

$A$ ,  $Pl : 2^{\Theta} \rightarrow [0, 1]$ , quantifies the maximum amount of support that could be allocated to  $A$ ,  $Pl(A) = \sum_{B \subseteq \Theta, B \cap A \neq \emptyset} m(B)$ .

**Pignistic level** In the transferable belief model [29], the decision is made in the pignistic level. The evidential information is transferred into a probabilistic framework by means of the pignistic probability distribution  $betP_m$ , for  $\theta \in \Theta$ ,  $betP_m(\theta) = \sum_{A \subseteq \Theta, \theta \in A} m(A)/|A|$ , where  $|A|$  denotes the number of elements in  $A$ .

**Decision rule** The risk associated with a decision rule is adaptable for the evidential framework [8, 27, 12]. In the case of imprecise data, the set of actions  $\mathcal{A}$  is  $2^{\Theta} \setminus \{\emptyset\}$ . In order to decide among the elements of  $\mathcal{A}$  according to the chosen loss function  $L$ , it is possible to adopt different strategies. Two strategies are proposed in the literature: the optimistic strategy by minimizing  $r_{\delta_{\Theta}}$  or the pessimistic strategy by minimizing  $\bar{r}_{\delta_{\Theta}}$  which are defined as follows:

$$\underline{r}(A) = \sum_{B \subseteq \Theta} m(B) \min_{\theta \in B} L(A, \theta), \quad \bar{r}(A) = \sum_{B \subseteq \Theta} m(B) \max_{\theta \in B} L(A, \theta). \quad (1)$$

### 3 Problem statement and Proposed approach

#### 3.1 Imprecise supervised classification

For a new example  $x$ , the output of an imprecise classifier is a set of classes, all its elements are candidates for the true class  $\theta$  and the missing information prevent more precise output. In this case a possible output of the classifier is the information: " $\theta \in A$ ",  $A \subseteq \Theta$ . To perform an imprecise classification, two cases are to be distinguished related to the training examples: *case 1*) learning examples are precisely labelled, i.e., only a single class is assigned to each example; *case 2*) one or more classes are assigned to each training example. In the first case described in the subsection 2.1, standard classifiers give a single class as prediction to a new object  $x$  but some recent classifiers [36, 5, 6] give a set of classes as prediction of  $x$ . Some of these recent classifiers base their algorithm on the posterior probability provided by standard classifiers. More precisely, if we denotes by  $\mathbb{P}(\cdot|x)$  the probability measure associated to the posterior probability distribution  $p(\cdot|x)$ ,  $\mathbb{P}(A|x) = \sum_{\theta \in A} p(\theta|x)$ ,  $A \subseteq \Theta$  is used to determine the relevant subset of classes to be assigned to  $x$ . In the second case when the imprecision or doubt is explicitly expressed by the labels, [2, 37, 4], a classifier  $\delta_{2^\Theta} : \mathcal{X} \rightarrow 2^\Theta \setminus \{\emptyset\}$  is trained on a set of examples  $(x_i, A_i)_{1 \leq i \leq N}$  such that for all  $1 \leq i \leq N$ ,  $x_i$  belongs to  $\mathcal{X}$  and  $\emptyset \neq A_i \subseteq \Theta$ . This case is refereed in our paper as *imprecise supervised classification*.

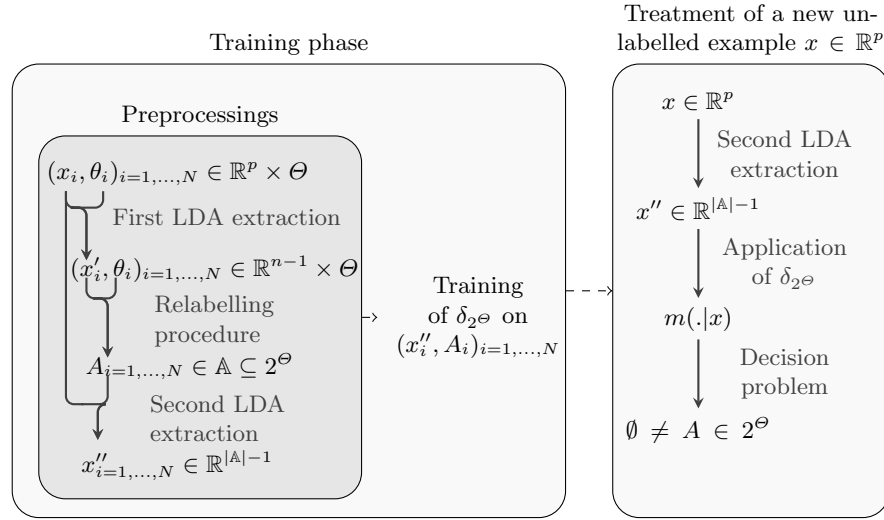
#### 3.2 Problem statement

Let us consider the supervised classification problem where the available training examples that are precisely labelled (*case 1*)  $(x_i, \theta_i)_{1 \leq i \leq N}$ ,  $x_i \in \mathcal{X}$  and  $\theta_i \in \Theta$  are such that i) the labels  $\theta_{i=1, \dots, N}$  are trusted e.g. they may derive from expertise on other features  $x_{i=1, \dots, N}^*$  which contain more complete information than  $x_{i=1, \dots, N}$ , ii) this loose of information induces overlapping on some examples, i.e.,  $\exists i, j \in \{1, \dots, N\}$  such that the characteristics of  $x_i$  are very close to those of  $x_j$  but  $\theta_i \neq \theta_j$ . When a standard classifier is trained on such data, it will commit inevitable errors. The problem that we handle in this paper is how to improve the learning step to better consider this type of data and get better performances and reliable predictions.

#### 3.3 The imprecise classification approach

The proposed approach of imprecise classification is constituted by three steps: i) the **relabelling step** which consists of analysing the training example in order to add to the class that is initially associated to an example the classes associated to the other examples having characteristics very close. Thus a new set of examples is built:  $(x_i, A_i)_{1 \leq i \leq N}$  such that for all  $1 \leq i \leq N$ ,  $x_i$  belongs to  $\mathcal{X}$  and  $\emptyset \neq A_i \subseteq \Theta$ ; ii) the **training step** which consists on the training of probabilistic classifier  $\delta_{2^\Theta} : \mathcal{X} \rightarrow 2^\Theta \setminus \{\emptyset\}$ .  $\delta_{2^\Theta}$  provides for a new object

$x \in \mathcal{X}$  a posterior probability distribution on  $2^\Theta$  which is also a mass function denoted  $m(\cdot|x)$ . The trained classifier ignores the existence of inclusion or intersection between subsets of classes. This unawareness of relations between the labels may seem counter intuitive, but is compatible with the purpose of finding a potentially imprecise label associated to a new incoming example; iii) the **decision step** which consists of proposing a loss function adapted for the case of imprecise classification that calculates the prediction that minimize the risk function associated to the classifier  $\delta_{2^\Theta}$ . Fig.1 illustrates the global process and the steps of relabelling, classification and decision are presented in details in the following.



**Fig. 1.** Steps of evidential classification of incomplete data

**Relabelling procedure** First we perform LDA extraction on the training examples (cf Fig.1) in order to reduce the complexity. The resulting features are  $x'_i \in \mathbb{R}^{n-1}, i = 1, \dots, N$  where  $n = |\Theta|$ . Then we consider a set of  $C$  standard classifiers  $\delta_\Theta^1, \dots, \delta_\Theta^C$  where on each classifier  $\delta_\Theta^c : \mathbb{R}^{n-1} \rightarrow \Theta, c \in \{1, \dots, C\}$  we compute leave-one-out (LOO) cross validation predictions for the training data  $(x'_i, \theta_i)_{i=1, \dots, N}$ .

The relabelling of the example  $(x'_i, \theta_i)$  is based on a vote procedure of the LOO predictions of the  $C$  classifiers. The vote procedure is the following: when more than 50% majority of the classifiers predict a class  $\theta_{maj_i}$ , the example is relabelled as the union  $A_i = \{\theta_i, \theta_{maj_i}\}$ . Note that when  $\theta_{maj_i} = \theta_i$  the original label remains, i.e.,  $A_i = \theta_i$ . If none of the predicted classes from the  $C$  classifiers gets the majority, then the ignorance is expressed for this example by relabelling it as  $A_i = \Theta$ . Remark that the new labels are consistent with the original classes that were trusted. The fact that several ( $C$ ) classifiers are used to express the imprecision permits a better objectivity on the real imprecision of the features,



i.e, the example is difficult not only for a single classifier. We denote by  $\mathbb{A}$  the set of different subsets  $A_i$  of  $\Theta$  considered in the relabelling examples.

Note that, we have limited the new labels cardinal to 2 elements except when expressing ignorance. This is done for avoiding too unbalanced training sets, but more general relabelling could be considered. Once all the training examples are relabelled, a classifier  $\delta_{2^\Theta}$  can be trained.

**Learning  $\delta_{2^\Theta}$**  As indicated throughout this paper,  $\delta_{2^\Theta}$  is learnt using the new labels ignoring the relations that might exist between the elements of  $\mathbb{A}$ . Reinforcing the idea of independence of treatment between the classes, LDA is applied to the relabelled training set  $(x_i, A_i)_{i=1, \dots, N}$ . This results to the reduction of the space dimension from  $p$  to  $|\mathbb{A}| - 1$  which better expresses the repartition of relabelled training examples. For the training example  $i \in \{1, \dots, N\}$ , let  $x_i'' \in \mathbb{R}^{|\mathbb{A}|-1}$  be the new projection of  $x_i$  on this  $|\mathbb{A}| - 1$  dimension space. The classifier  $\delta_{2^\Theta}$  is finally learnt on  $(x_i'', A_i)_{i=1, \dots, N}$ .

**Decision problem** As recalled in subsections 2.2 and 2.3, the decision to assign a new object  $x$  to a single class or a set of classes usually relies on the minimisation of the risk function which is associated to a loss function  $L : 2^\Theta \setminus \{\emptyset\} \times \Theta \rightarrow \mathbb{R}$ . As mentioned in the introduction of this paper, the application of our work concerns situations where errors may have serious consequences. It would then be legitimate to consider the pessimistic strategy by minimizing  $\bar{r}_{\delta_\Theta}$ . Furthermore, in the definition of  $\bar{r}_{\delta_\Theta}$ , equation (1), the quantity  $\max_{\theta \in B} L(A, \theta)$  concerns the loss incurred by choosing  $A \subseteq \Theta$ , when the true nature is comprised in  $B \subseteq \Theta$ . On the basis of this fact, we proposed a new definition of the loss function,  $L(A, B)$ ,  $A, B \subseteq \Theta$ , which directly takes into account the relations between  $A$  and  $B$ . This is actually a generalisation of the definition proposed in [6] that is based on F-measure, recall and precision for imprecise classification. Let us consider  $A, B \in 2^\Theta \setminus \{\emptyset\}$ , where  $A = \theta(x)$  is the prediction for the object  $x$  and  $B$  is its state of nature. The Recall is defined as the proportion of relevant classes included in the prediction  $\theta(x)$ . We define the recall of  $A$  and  $B$  as:

$$R(A, B) = \frac{|A \cap B|}{|B|}. \quad (2)$$

The precision is defined as the proportion of classes in the prediction that are relevant. We define the precision of  $A$  and  $B$  as:

$$P(A, B) = \frac{|A \cap B|}{|A|}. \quad (3)$$

Considering these two definition, the F-measure can be defined as follows:

$$F_\beta(A, B) = \frac{(1 + \beta^2)PR}{\beta^2P + R} = \frac{(1 + \beta^2)|A \cap B|}{\beta^2|B| + |A|}. \quad (4)$$

Note that  $\beta = 0$ , induce  $F_\beta(A, B) = P(A, B)$ , whereas when  $\beta \rightarrow \infty$ ,  $F_\beta(A, B) \xrightarrow{\beta \rightarrow \infty} P(A, B)$ . Let us comment on some situations according to the "true set"  $B$  and the predicted set  $A$ . The worse scenario of prediction is when there is no intersection between  $A$  and  $B$ . This would always be sanctioned by  $F_\beta(A, B) = 0$ . On the opposite, when  $A = B$ ,  $F_\beta(A, B) = 1$  for every  $\beta$ . Between those extreme cases, the errors of generalisation i.e.,  $B \subset A$  are controlled by the precision while the errors of specialisation i.e.,  $A \subset B$  are controlled by the recall. Finally, the loss function  $L_\beta : 2^\Theta \setminus \{\emptyset\} \times 2^\Theta \setminus \{\emptyset\} \rightarrow \mathbb{R}$  is extended:

$$L_\beta(A, B) = 1 - F_\beta(A, B). \quad (5)$$

For an example  $x$  to classify, which mass function  $m(\cdot|x)$  has been calculated by  $\delta_{2^\Theta}$ , we predict the set  $A$  minimizing the following risk :

$$\text{Risk}_\beta(A) = \sum_{B \subseteq \Theta} m(B) L_\beta(A, B). \quad (6)$$

## 4 Related Works

Regarding relabelling procedures, many research have been carried out to identify suspect examples with the intention to suppress or relabel them into a concurrent more appropriate class [15, 20]. This is generally done to enhance the performance. Other approach consist in relabelling into imprecise classes. This has been done for testing evidential classification approach on imprecise labelled data in [37]. But, as already stated, our relabelling serves a different purpose that is better mapping overlaps in the feature space. Concerning the imprecise classification, several works have been dedicated to tackle this problem. Instead of the term "imprecise classification" that is adopted in our article, authors use terms like "nondeterministic classification" [6], "reliable classification" [24], "indeterminate classification" [36] [5], "set-valued classification" [28] [31] or "conformal prediction" [3] (see [24] for a short state of the art). In [36], the Naive Credal Classifier (NCC) is proposed as the extension of Naive Bayes Classifier (NBC) to sets of probability distributions. In [24] the authors propose an approach that starts from the outputs of a binary classification [25] using classifier that are trained to distinguish aleatoric and epistemic uncertainty. The outputs are epistemic uncertainty, aleatoric uncertainty and two preference degrees in favor of the two concurrent classes. [24] generalizes this approach to the multi-class and providing set of classes as output. Closer to our approach are approaches of [4] and [6]. The approach in [6] is based on a posterior probability distribution provided by a probabilistic classifier. The advantage of such approach and ours is that any standard probabilistic classifier may be used to perform an imprecise classification. Our approach distinguishes itself by the relabelling step and by the way probabilities are allowed on sets of classes. To the best of our knowledge existing works algorithms do not train a probabilistic classifier on partially labelled data to quantify the body of evidence. Although we insisted for the use of standard probabilistic classifier  $\delta_{2^\Theta}$  unaware of relations between the sets, it

is possible to run our procedure with an evidential classifier as the evidential  $k$ -NN [4].

## 5 Illustration

### 5.1 Settings

We performed experiments on the classification problem of four plastic categories designated plastics A, B, C and D on the basis of industrially acquired spectra. The total of 11540 available data examples is summarized in Table 1. Each plastic example was identified by experts on the basis of laboratory measure of attenuated total reflectance spectra (ATR) which is considered as a reliable source of information for plastic category’s determination. As a consequence, original training classes are trusted and were not questioned. However data provided by the industrial devices may be challenged. These data consist in spectra composed of the reflectance intensity of 256 different wavelengths. Therefore and for the enumerated reasons in section 1, the features are subject to ambiguity. Prior to experiments, all the feature vectors, i.e., spectra, were corrected by the standard normal variate technique to avoid light scattering and spectral noise effects. We

**Table 1.** Number of spectra of each original class in learning and testing bases.

Classes	Category A	Category B	Category C	Category D
Learning base	1416	1412	1425	1434
Testing base	1469	1458	1454	1472

implemented our approach and compared it to the approaches in [4] and [6]. The implementation is made using R packages, using existing functions for the application of the following 8 classifiers naive Bayes classifier: (nbayes),  $k$ -Nearest Neighbour ( $k$ -NN), decision tree (tree), random forest (rf), linear discriminant analysis (lda), partial least squares discriminant analysis (pls-da), support vector machine (svm) and neural networks (nnet).<sup>3</sup>

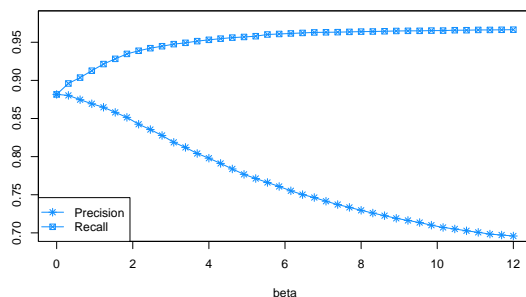
### 5.2 Results

In order to apply our procedure, we must primary choose a set of classifiers to perform the relabelling. These classifiers are not necessarily probabilistic but producing point prediction. Thus, for every experimentation, our algorithm ECLAIRE was performed with the ensemble relabelling using 7 classifiers: nbayes,  $k$ -NN, tree, rf, lda, svm, nnet<sup>4</sup>. Then, we are able to perform the ECLAIRE

<sup>3</sup> Experiments concerning these learning algorithm rely on the following functions (and R packages) : naiveBayes (e1071), knn3 (caret), rpart (rpart), randomForest (randomForest), lda (MASS), plsda (caret), svm (e1071), nnet (nnet).

<sup>4</sup> In order to limit unbalanced classes, we choose to exclude form the learning base examples which new labels count less than 20 examples

imprecise version of a selected probabilistic classifier. Fig. 2, shows the recall and precision scores of the probabilistic classifier nbayes to show the role of  $\beta$ . We see the same influence of  $\beta$  as mentioned in [6]. Indeed, (cf subsection 3.3), with small values of  $\beta$  we have good precision, traducing the relevance of prediction, i.e., the size of the predicted set is reasonable; while high values of  $\beta$  give good recall, meaning reliability, i.e., better chance to have true class included in the predictions. The choice of  $\beta$  should then result form a compromise between relevance and reliability requirement.



**Fig. 2.** Recall and precision of ECLAIRE using nbayes, i.e.  $\delta_{2\Theta}$  is nbayes, against  $\beta$ .

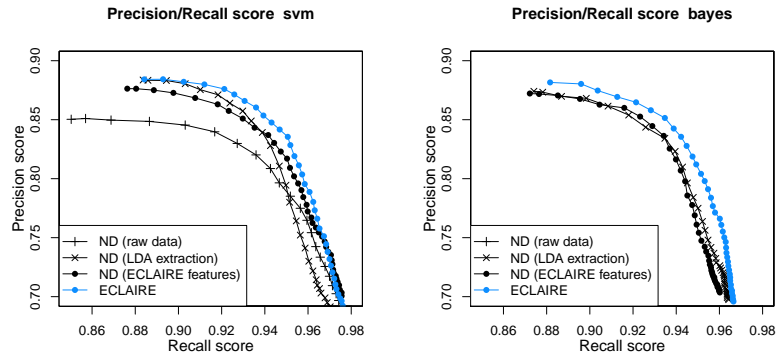
In order to evaluate the performances of ECLAIRE, we compared our results to the classifier proposed in [6] that is called here *nondeterministic* classifier. As *nondeterministic* classifier and ECLAIRE are set up for a parameter  $\beta$ , we decided to set  $\beta$ s such that global recalls equal to 0.90, and compare global precisions on a fair basis. For even more neutrality regarding the features used in both approach, we furnish to the *nondeterministic* classifier, the same reduced features  $x'_i, i = 1, \dots, N$ , that those used by ECLAIRE in the training phase (see Fig. 1). The 7 first columns of Table 2 shows the so obtained precisions for 7

**Table 2.** Precision P of ECLAIRE compared with *nondeterministic* with  $\beta$ s chosen such that recalls equal to 0.90.

	nbayes	<i>k</i> -NN	tree	rf	lda	pls-da	svm	evidential	<i>k</i> -NN
Nondeterministic	86.70	86.94	<b>85.00</b>	86.52	<b>83.41</b>	85.35	88.20		86.58
ECLAIRE	<b>87.78</b>	<b>87.89</b>	83.88	<b>87.45</b>	82.94	<b>86.33</b>	<b>88.31</b>		<b>86.69</b>

classifiers. These results show the competitiveness of our approach for most of the classifiers, especially nbayes, *k*-NN, rf and pls-da. However, these results are only partial since they do not show the general trend for different  $\beta$ s that are generally in favour of our approach. Therefore we present more complete results for nbayes and svm in Fig. 3., showing evaluation of precision score against recall score for several values of  $\beta$  varying in  $[0, 6]$ . On the same figure, we also present the results

of *nondeterministic* classifier with different input feature (in black): raw features, i.e.,  $x_i \in \mathbb{R}^p$ , LDA reduced features, i.e.,  $x'_i \in \mathbb{R}^{n-1}$  and the same features as those used for ECLAIRE, i.e.,  $x''_i \in \mathbb{R}^{|\mathbb{A}|-1}$  (see Fig. 1 for more details). Doing so, we show that the good performances of ECLAIRE are not only attributable to extraction phase. To facilitate the understanding of the results plotted in Fig. 3, one should understand that the best performances are those illustrated by points on the top right of the plots, i.e., higher precision and recall scores. We observe that ECLAIRE generally makes a better compromise between the recall and precision scores for the used classifiers. Regarding the special case when



**Fig. 3.** Precision vs recall of Nondeterministic (ND) and ECLAIRE

ECLAIRE is performed with an evidential classifier taking imprecise labelled training example (see section 4), the comparison is less straightforward. We considered the evidential  $k$ -NN [9] for imprecise labels by minimizing the error suggested in [39]. Using this evidential  $k$ -NN as a classifier  $\delta_{2\theta}$  in ECLAIRE procedure is straightforward. Concerning the application of *nondeterministic* classifier, we decided to keep the same parameter and turn the classifier into probabilistic by applying the pignistic transformation to the mass output of the  $k$ -NN classifier (see column of Table 2 ). ECLAIRE has a slightly better results.

## 6 Conclusion

In this article, an evidential classification of incomplete data via imprecise relabelling was proposed. For any probabilistic classifier, our approach proposes an adaptation to get more cautious output. The benefit of our approach was illustrated on the problem of sorting plastics and showed competitive performances. In the future works we plan to exploit our procedure to provide cautious decision for the problem of plastic sorting. This application requires high reliability of the decision for preserving the physicochemical properties of the recycle product. At the same time, the decision shall ensure reasonable relevance to guarantee financial interest, indeed the more one plastic category is finely sorted the more benefice the industrial gets.

## References

1. Alshamaa, D., Chehade, F.M., Honeine, P.: A hierarchical classification method using belief functions. *Signal Processing* **148**, 68–77 (2018)
2. Ambroise, C., Denoeux, T., Govaert, G., Smets, P.: Learning from an imprecise teacher: probabilistic and evidential approaches. *Applied Stochastic Models and Data Analysis* **1**, 100–105 (2001)
3. Balasubramanian, V., Ho, S.S., Vovk, V.: *Conformal prediction for reliable machine learning: theory, adaptations and applications*. Newnes (2014)
4. Côme, E., Oukhellou, L., Denoeux, T., Aknin, P.: Learning from partially supervised data using mixture models and belief functions. *Pattern recognition* **42**(3), 334–348 (2009)
5. Corani, G., Zaffalon, M.: Learning reliable classifiers from small or incomplete data sets: the naive credal classifier 2. *Journal of Machine Learning Research* **9**(Apr), 581–621 (2008)
6. Coz, J.J.d., Díez, J., Bahamonde, A.: Learning nondeterministic classifiers. *Journal of Machine Learning Research* **10**(Oct), 2273–2293 (2009)
7. Cucchiella, F., D’Adamo, I., Koh, S.L., Rosa, P.: Recycling of weees: An economic assessment of present and future e-waste streams. *Renewable and sustainable energy reviews* **51**, 263–272 (2015)
8. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. In: *Classic Works of the Dempster-Shafer Theory of Belief Functions*, pp. 57–72. Springer (2008)
9. Denoeux, T.: A k-nearest neighbor classification rule based on dempster-shafer theory. *IEEE transactions on systems, man, and cybernetics* **25**(5), 804–813 (1995)
10. Denoeux, T.: A neural network classifier based on dempster-shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* **30**(2), 131–150 (2000)
11. Denoeux, T.: Maximum likelihood estimation from uncertain data in the belief function framework. *IEEE Transactions on knowledge and data engineering* **25**(1), 119–130 (2013)
12. Denoeux, T.: *Logistic regression, neural networks and dempster-shafer theory: a new perspective*. Knowledge-Based Systems (2019)
13. Dubois, D., Prade, H.: *Possibility theory*. Springer (2012)
14. Dunn, J.C.: A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters (1973)
15. Kanj, S., Abdallah, F., Denoeux, T., Tout, K.: Editing training data for multi-label classification with the k-nearest neighbor rule. *Pattern Analysis and Applications* **19**(1), 145–161 (2016)
16. Kanjanatarakul, O., Kaewsompong, N., Sriboonchitta, S., Denoeux, T.: Estimation and prediction using belief functions: Application to stochastic frontier analysis. In: *Econometrics of Risk* (2015)
17. Kassouf, A., Maalouly, J., Rutledge, D.N., Chebib, H., Ducruet, V.: Rapid discrimination of plastic packaging materials using mir spectroscopy coupled with independent components analysis (ica). *Waste management* **34**(11), 2131–2138 (2014)
18. Keller, J.M., Gray, M.R., Givens, J.A.: A fuzzy k-nearest neighbor algorithm. *IEEE transactions on systems, man, and cybernetics* (4), 580–585 (1985)
19. Kruse, R.: Fuzzy neural network. *Scholarpedia* **3**(11), 6043 (2008)
20. Lallich, S., Muhlenbach, F., Zighed, D.A.: Improving classification by removing or relabeling mislabeled instances. In: *International Symposium on Methodologies for Intelligent Systems*. pp. 5–15. Springer (2002)

21. Lee, H.K., Kim, S.B.: An overlap-sensitive margin classifier for imbalanced and overlapping data. *Expert Systems with Applications* **98**, 72–83 (2018)
22. Leitner, R., Mairer, H., Kercek, A.: Real-time classification of polymers with nir spectral imaging and blob analysis. *Real-Time Imaging* **9**(4), 245–251 (2003)
23. Naeini, M.P., Moshiri, B., Araabi, B.N., Sadeghi, M.: Learning by abstraction: Hierarchical classification model using evidential theoretic approach and bayesian ensemble model. *Neurocomputing* **130**, 73–82 (2014)
24. Nguyen, V.L., Destercke, S., Masson, M.H., Hüllermeier, E.: Reliable multi-class classification based on pairwise epistemic and aleatoric uncertainty. In: *International Joint Conference on Artificial Intelligence* (2018)
25. Senge, R., Bösner, S., Dembczyński, K., Haasenritter, J., Hirsch, O., Donner-Banzhoff, N., Hüllermeier, E.: Reliable classification: Learning classifiers that distinguish aleatoric and epistemic uncertainty. *Information Sciences* **255**, 16–29 (2014)
26. Shafer, G.: *A mathematical theory of evidence*, vol. 42. Princeton university press (1976)
27. Shafer, G.: Constructive probability. *Synthese* **48**(1), 1–60 (1981)
28. Shafer, G., Vovk, V.: A tutorial on conformal prediction. *Journal of Machine Learning Research* **9**(Mar), 371–421 (2008)
29. Smets, P.: *Non-Standard Logics for Automated Reasoning*. Academic Press (1988)
30. Smets, P., Kennes, R.: The transferable belief model. *Artificial intelligence* **66**(2), 191–234 (1994)
31. Soullard, Y., Destercke, S., Thouvenin, I.: Co-training with credal models. In: *IAPR Workshop on Artificial Neural Networks in Pattern Recognition*. pp. 92–104. Springer (2016)
32. Sutton-Charani, N., Imoussaten, A., Harispe, S., Montmain, J.: Evidential bagging: Combining heterogeneous classifiers in the belief functions framework. In: *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. pp. 297–309. Springer (2018)
33. Walley, P.: Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning* **24**(2-3), 125–148 (2000)
34. Xiong, H., Li, M., Jiang, T., Zhao, S.: Classification algorithm based on nb for class overlapping problem. *Appl. Math* **7**(2L), 409–415 (2013)
35. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems* **1**(1), 3–28 (1978)
36. Zaffalon, M.: The naive credal classifier. *Journal of statistical planning and inference* **105**(1), 5–21 (2002)
37. Zhang, J., Subasingha, S., Premaratne, K., Shyu, M.L., Kubat, M., Hewawasam, K.: A novel belief theoretic association rule mining based classifier for handling class label ambiguities. In: *Proc. Workshop Foundations of Data Mining (FDM'04), Int'l Conf. Data Mining (ICDM'04)* (2004)
38. Zheng, Y., Bai, J., Xu, J., Li, X., Zhang, Y.: A discrimination model in waste plastics sorting using nir hyperspectral imaging system. *Waste Management* **72**, 87–98 (2018)
39. Zouhal, L.M., Denoeux, T.: An evidence-theoretic k-nn rule with parameter optimization. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **28**(2), 263–271 (1998)