Belief Classification Approach Based on Generalized Credal EM

Imene Jraidi and Zied Elouedi

LARODEC,
Institut Supérieur de Gestion Tunis,
41 Avenue de la liberté,
2000 Le Bardo, Tunisie
imene_jrd@yahoo.fr,zied.elouedi@gmx.fr

Abstract. The EM algorithm is widely used in supervised and unsupervised classification when applied for mixture model parameter estimation. It has been shown that this method can be applied for partially supervised classification where the knowledge about the class labels of the observations can be imprecise and/or uncertain. In this paper, we propose to generalize this approach to cope with imperfect knowledge at two levels: the attribute values of the observations and their class labels. This knowledge is represented by belief functions as understood in the Transferable Belief Model. We show that this approach can be applied when the data are categorical and generated from multinomial mixtures.

Keywords: Expectation Maximization, mixture models, Transferable Belief Model, partially supervised classification.

1 Introduction

Operating within an imperfect environment and facing imprecise, uncertain and even missing information is the real challenge in decision making. For instance, a doctor has to make a decision even if he is not able to identify the exact disease of his patient but he only knows that the patient has not such kind of diseases. On the other hand, a controller system must be able to integrate multiple sensors even when only a fraction may operate at a given time. In this context, most standard classification methods encounter a real problem to meet these real life situations which make them inappropriate to classify objects characterized by such imperfect information.

The idea is thus to combine classification methods with theories managing uncertainty as the belief function theory [11]. In the Transferable Belief Model's interpretation (TBM) [14], this theory provides a formalism for handling subjective and personal judgments and that can also deal with objective probabilities. So, this theory is able to handle partial knowledge and cope with partial and even total ignorance. Besides, this theory has provided a powerful tool to deal with uncertainty in classification problems. We notably mention belief decision trees [5], belief k-nearest neighbor [4], belief K-modes [2], etc.

K. Mellouli (Ed.): ECSQARU 2007, LNAI 4724, pp. 524–535, 2007.

On the other side, the Expectation-Maximization (EM) algorithm [3] is a generic approach for parameter estimation in incomplete data problems and has been widely used in supervised and unsupervised classification [7, 8]. In this context, data are assumed to be generated from a mixture model where each component of the mixture or class is identified by a probability distribution. In the supervised mode or discrimination, the class labels of the observations are known a priori and are used to classify new observations with unknown class labels. In the unsupervised mode or clustering, the class labels of the observations are unknown a priori and the goal is to find a partitioning of the observations by grouping similar observations together. Besides, when the class labels are only partially known that is the actual class of the observations can be imprecise or uncertain, the classification procedure becomes partially supervised.

Several works have been proposed in this uncertain context [1, 15]. In [1], the class labels can be imprecise and a probabilistic model relating the imprecise label to the true class is assumed. In [15], the class labels can be imprecise and/or uncertain and this knowledge is represented by belief functions. In both approaches, uncertainty occurs only at the class labels of the observations. However, uncertainty may also appear in the values of the attributes.

We propose then to treat a more general case where uncertainty can arise not only in the class labels but also in the values of the attributes characterizing the observations. This method is based on both the EM approach and the belief function theory as understood in the TBM.

The remainder of the paper is organized as follows. We start by presenting the EM algorithm for learning mixture models. Next, we outline the necessary background concerning the belief function theory and we describe the EM algorithm within this framework. Then, we develop our generalized approach that takes into account uncertainty in the attributes of the observations when data are categorical and generated from multinomial mixtures.

2 The EM Algorithm for Learning Mixture Models

In the mixture modeling approach [9], the data $X = \{x_1, ..., x_n\}$ are assumed to be identically and independently distributed (iid) according to a probability function given by:

$$f(x_i|\Theta) = \sum_{k=1}^{K} \pi_k f_k(x_i|\theta_k) \ , \tag{1}$$

where K is the number of components in the mixture, π_k are the mixing proportions that must be non negative and sum to one, f_k denotes a component, i.e. a probability function parameterized by θ_k, and $\Theta = \{(\pi_k, \theta_k), k = 1, ..., K\}$ are the parameters of the model to be estimated.

In this paper, we treat categorical data generated from multinomial mixture models where each observation x_i is described by D categorical attributes, with a respective number of categories $nb_1, ..., nb_D$. The data X can be represented by n binary vectors $(x_i^{dj}; d = 1, ..., D, j = 1, ..., nb_d)$ where $x_i^{dj} = 1$ if the attribute x_i^d

has the category j and 0 otherwise. In this model, each component k is identified by a D-dimensional multinomial distribution given by:

$$f_k(x_i|\theta_k) = \prod_{d=1}^{D} \prod_{j=1}^{nb_d} (p_k^{dj})^{x_i^{dj}} , \quad (2)$$

where the parameters θ_k are given by the probabilities $p_k^{dj}(d = 1, ..., D, j = 1, ..., nb_d)$, that the attribute x_i^d has the category j. In this model, the D variables are assumed to be independent given the component k [6].

2.1 Maximum Likelihood Estimation

To estimate the parameters Θ, we generally apply the Maximum Likelihood Estimation (MLE) principle: the parameters that have most likely generated the data, are those that maximize the likelihood (or the log-likelihood for the sake of simplicity) given by:

$$\mathcal{L}(\Theta|X) = \sum_{i=1}^{n} log(\sum_{k=1}^{K} \pi_k f_k(x_i|\theta_k)) . \quad (3)$$

Generally, the maximization of this equation cannot be obtained analytically. The classical approach to solve this problem is the EM algorithm [3] which provides an iterative procedure for computing MLE. In order to use the EM algorithm, the problem has to be reformulated as an incomplete data problem.

2.2 An Incomplete Data Problem

The idea is to introduce a set of "hidden" variables $Z = \{z_1, ..., z_n\}$ that indicate which component of the mixture has generated each observation. The problem would decouple then into a set of simple maximizations. More precisely, $z_i = (z_{i1}, ..., z_{ik}, ...z_{iK})$ where $z_{ik} = 1$ if x_i has been generated from the component k and 0 otherwise. The whole data $Y = \{y_1, ..., y_n\}$ where $(y_i = (x_i, z_i))$, is then the so-called augmented data or complete data. Using these indicator variables Z, the equation (3) can be reformulated as the complete log-likelihood:

$$\mathcal{L}_c(\Theta|Y) = \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik} log(\pi_k f_k(x_i|\theta_k)) . \quad (4)$$

2.3 The EM Algorithm

The EM algorithm can now be applied by considering the variables Z as the missing data. The algorithm provides a sequence of estimates $\Theta^{(t)}$[1], of the parameters Θ by the iteration of two steps: an E-step (for Expectation) and a M-step (for Maximization).

[1] We use the subscript (t) to denote the iteration t of the EM algorithm.

E-Step. The E-step computes the conditional expectation of the complete log-likelihood $\mathcal{L}_c(\Theta|Y)$ given the observed data X and the current parameters $\Theta^{(t)}$:

$$Q(\Theta|\Theta^{(t)}) = E[\mathcal{L}_c(\Theta|X, Z)|X, \Theta^{(t)}] \ , \tag{5}$$

which is a linear function of the missing data z_{ik}. So, at the iteration t, the E-step just requires to compute the conditional expectation of z_{ik} given X and $\Theta^{(t)}$:

$$E[z_{ik}|X, \Theta^{(t)}] = t_{ik}^{(t)} \ . \tag{6}$$

Actually, this quantity is nothing else then the posterior probability $p(z_{ik} = 1|X, \Theta^{(t)})$ that the observation x_i has been generated by the component f_k estimated at the iteration t. This probability measure is computed through the Bayes rule as follows:

$$t_{ik}^{(t)} = \frac{\pi_k^{(t)} f_k(x_i|\theta_k^{(t)})}{\sum_{l=1}^{K} \pi_l^{(t)} f_l(x_i|\theta_l^{(t)})} \ . \tag{7}$$

Using this result the equation (5) becomes:

$$Q(\Theta|\Theta^{(t)}) = \sum_{i=1}^{n} \sum_{k=1}^{K} t_{ik}^{(t)} log(\pi_k f_k(x_i|\theta_k)) \ . \tag{8}$$

M-Step. The M-step updates the current parameters $\Theta^{(t)}$ by maximizing $Q(\Theta|\Theta^{(t)})$ over Θ, so that to have an updated estimate $\Theta^{(t+1)}$. The mixing proportions π_k are computed independently of the component parameters θ_k:

$$\pi_k^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} t_{ik}^{(t)} \ . \tag{9}$$

The update of the parameters θ_k depends on the nature of the mixed components and can be obtained by analyzing the following equation:

$$\sum_{i=1}^{n} \sum_{k=1}^{K} t_{ik}^{(t)} \ \frac{\partial log f_k(x_i|\theta_k^{(t)})}{\partial \Theta} = 0 \ . \tag{10}$$

3 Belief Function Theory

Before we turn to the EM approach under the TBM framework, we shall sketch out some of the bases of the belief function theory. Details can be found in [11, 14].

3.1 Basic Concepts

Let Ω be a finite non empty set of mutually exclusive elementary events related to a given problem. Ω is generally called the frame of discernment. The set 2^{Ω} contains all the subsets of Ω: it is the power set of Ω.

The impact of a piece of evidence held by an agent (whatever is it: a sensor, a computer program, an expert, etc) among the propositions of Ω, is expressed by the so-called basic belief assignment (bba). The bba is a function $m^\Omega : 2^\Omega \rightarrow [0,1]$ that satisfies: $\sum_{A \subseteq \Omega} m^\Omega(A) = 1$.

The value $m^\Omega(A)$, called a basic belief mass (bbm), is the quantity of belief that supports exactly the proposition A and that due to the lack of information, does not support any strict subset of A.

The belief function $bel^\Omega : 2^\Omega \rightarrow [0,1]$, with $bel^\Omega(A) = \sum_{\emptyset \neq B \subseteq A} m^\Omega(B)$, expresses the total amount of belief assigned to the subsets implying A without implying $\overline{A}$.

The plausibility function $pl^\Omega : 2^\Omega \rightarrow [0,1]$, with $pl^\Omega(A) = \sum_{A \cap B \neq \emptyset} m^\Omega(B)$, quantifies the degree of belief committed to the propositions compatible with A.

Several special belief functions relative to particular states of uncertainty are defined. The vacuous belief function quantifies a state of total ignorance, in which no support is given to any particular subset of Ω. This function is defined as follows [11]:

$$m^\Omega(\Omega) = 1 \text{ and } m^\Omega(A) = 0, \forall A \subset \Omega \,. \tag{11}$$

A Bayesian belief function is a belief function where the belief is only allocated among elementary events of Ω [11].

A certain belief function is a Bayesian belief function where the whole belief is assigned to a unique elementary event A: it expresses a state of total certainty. This function is defined by:

$$m^\Omega(A) = 1 \text{ and } m^\Omega(B) = 0, \forall B \subseteq \Omega, A \in \Omega \text{ and } B \neq A \,. \tag{12}$$

3.2 Combination of Belief Functions

Let $m^\Omega_{E_1}$ and $m^\Omega_{E_2}$ be two bba's induced from two distinct information sources (E_1 and E_2) and defined on the same frame of discernment Ω. The joint impact of both pieces of evidence is given by the conjunctive rule of combination [12]:

$$(m^\Omega_{E_1} \bigcirc\!\!\!\!\cap\, m^\Omega_{E_2})(A) = \sum_{B \cap C = A} m^\Omega_{E_1}(B) m^\Omega_{E_2}(C) \,. \tag{13}$$

3.3 The Pignistic Transformation

In the TBM, beliefs can be held at two levels: a credal level where beliefs are entertained and quantified by belief functions, and a pignistic level, where decisions are made. At this level, beliefs are transformed into probability measures (denoted by $BetP^\Omega$) in order to choose the most likely hypothesis. The transformation rule is called the pignistic transformation defined for all $\omega_k \in \Omega$ as:

$$BetP^\Omega(w_k) = \sum_{A \ni \omega_k} \frac{m^\Omega(A)}{|A|} \frac{1}{(1 - m^\Omega(\emptyset))} \,. \tag{14}$$

3.4 The Generalized Bayesian Theorem

Smets [13] has generalized the Bayesian theorem (GBT), offering an interesting tool for inverting conditional belief functions within the TBM framework. Assume that we have a vacuous a priori belief on a frame Ω, and we know for each element $\omega_i \in \Omega$, what would be our beliefs on another frame X if this element happened. Suppose that we learn that the actual value of X is in $x \subseteq X$, then the GBT allows us to derive the conditional belief function over the frame Ω given the observation x. One has:

$$pl^{\Omega}[x](\omega) = 1 - \prod_{\omega_i \in \omega} (1 - pl^X[\omega_i](x)) \ . \tag{15}$$

4 The Credal EM Approach

The Credal EM (CrEM) [15] is a variant of EM for partially supervised learning. In this approach, the class label of the observations can be partially known. That is, it can be imprecise and/or uncertain. This knowledge is represented by belief functions as understood in the TBM.

The learning set is then given by: $L = \{(x_1, m_1^{\Omega}), ..., (x_n, m_n^{\Omega})\}$, where $X = \{x_1, ..., x_n\}$ are n iid observations derived from a mixture of K classes $\Omega = \{\omega_1, ..., \omega_K\}$, and $m_i^{\Omega} : 2^{\Omega} \to [0, 1]$ are the bba's representing the a priori beliefs of membership of the observations x_i into the subsets of Ω.

E-Step. In the classical approach, the algorithm computes the a posteriori probability $t_{ik}^{(t)}$ that x_i has been generated by the class k estimated at the current iteration. The CrEM computes the mass $m^{\Omega}[x_i, \Theta^{(t)}]$ that x_i has been generated by the class k with the current parameters $\Theta^{(t)}$ through the GBT from its corresponding plausibilities:

$$pl^{\Omega}[x_i](A) = 1 - \prod_{\omega_j \in A} (1 - pl^X[\omega_j](x_i)) \ , \forall A \subseteq \Omega \ . \tag{16}$$

These masses are then combined with the prior bba's through the conjunctive rule of combination. The resulting masses are given by:

$$\hat{m}^{\Omega}[x_i, \Theta^{(t)}](A) = \sum_{B \cap C = A} m^{\Omega}[x_i, \Theta^{(t)}](B) \ m_i^{\Omega}(C) \ , \forall \ A \subseteq \Omega \ . \tag{17}$$

M-Step. The M-step finds the most probable value of the mixture parameters. This comes down to determine the parameters $\theta \in \Theta$ that maximize the conditional plausibility of the data given θ. Under the iid assumption, this term is given: $\prod_{i=1}^{n} pl^X[\theta](x_i)$.

The likelihood function to be maximized is then given by [15]:

$$Q(\Theta|\Theta^{(t)}) = \sum_{i=1}^{n} \sum_{A \subseteq \Omega} \hat{m}^{\Omega}[x_i, \Theta^{(t)}](A) log(pl^X[A](x_i)) \ . \tag{18}$$

This equation is analogous to the equation (8) in the TBM framework.

5 Generalized Credal EM

The CrEM provides an alternative for learning in an uncertain environment that is more general than the one proposed in [1] which deals only with imprecise class labels.

However, this approach is not fitted to situations where the values of the attributes characterizing the observations are also partially known. This could involve missing data (some attribute values are missing), imprecise data (we only know that the value of such attributes belongs to a subset of possible values), or uncertain data (we only have some beliefs about the actual value of such attributes).

In this section, we develop a generalization of the CrEM approach that copes with these situations. We first introduce a method that takes into account missing data, then we propose a more general approach that integrates imprecise and uncertain knowledge. Hence, our approach deals with uncertainty in class and attribute values. Besides, we should note that our method deals only with categorical data.

5.1 Learning from Missing Data

In the previous sections, only one aspect of the EM algorithm has been highlighted: learning mixture models. Another important aspect of EM is to learn from data sets with missing values [3, 8]. In this section, we propose to combine this application of EM with that of learning mixture parameters in the TBM framework [15].

We assume that the data X are made up of two components: an observed component X^o and a missing component X^m. Each object x_i in the missing component is divided into (x_i^o, x_i^m) where x_i^o denotes the observed attribute values of x_i and x_i^m the missing attributes, and each x_i can have different missing attributes.

The conditional expected complete data likelihood given the observed data and the current parameters is then written as follows:

$$E[\mathcal{L}_c(\Theta|X^o, X^m, Z)|X^o, \Theta^{(t)}] \,. \tag{19}$$

So, there are two forms of incomplete data: the variables z_{ik} that indicate for each object, which class it comes from, and the missing data x_i^m. The E-step gives an estimation of both forms of missing data: $E[z_{ik}|X^o, \Theta^{(t)}]$ and $E[x_i^m|X^o, \Theta^{(t)}]$. The M-step uses then the completed data to update the mixture model parameters Θ.

E-Step. The first term to be estimated is given by $t_{ik}^{(t)}$, the probability that x_i has been generated from the class k. These probabilities are derived through the pignistic transformation from the masses $\hat{m}^\Omega[x_i, \Theta^{(t)}]$:

$$t_{ik}^{(t)} = \sum_{A \ni \omega_k} \frac{\hat{m}^\Omega[x_i, \Theta^{(t)}](A)}{|A|} \frac{1}{1 - \hat{m}^\Omega[x_i, \Theta^{(t)}](\emptyset)} \,. \tag{20}$$

The masses $\hat{m}^{\Omega}[x_i, \Theta^{(t)}]$ express the beliefs of membership of x_i into the classes of Ω computed as in the CrEM approach (see equation (17)) and measured only over the observed values x_i^o.

The second term gives for each missing attribute x_i^d in x_i^m, the probability that the attribute x_i^d takes the category j (for $j = 1, ..., nb_d$). Since we assume that within each class the attributes are independent, for each class k, we have a different estimation of x_i^{dj} given by $p_k^{dj(t)}$ computed at the current iteration.

M-Step. The M-step updates the current parameters using these expected values. The mixing proportions π_k are updated using the t_{ik} as in equation (9).

The parameters θ_k given by the probabilities p_k^{dj} are updated by:

$$p_k^{dj^{(t+1)}} = \frac{\sum_{i=1}^{n} t_{ik}^{(t)} x_i^{dj}}{\sum_{i=1}^{n} t_{ik}^{(t)}} \; , \tag{21}$$

where $\sum_{i=1}^{n} t_{ik}^{(t)} x_i^{dj}$ is the estimated number of objects in the class k in which the attribute x_i^d has the category j and $\sum_{i=1}^{n} t_{ik}^{(t)}$ is the total estimated number of objects in the class k. So, $t_{ik} x_i^{dj}$ has to be substituted by $t_{ik} p_k^{dj}$ for the missing components.

5.2 Learning from Partial Knowledge

In this subsection, we propose an approach that integrates imprecise and uncertain knowledge regarding the attribute values characterizing the objects of the learning set. As the prior knowledge about the class labels presented before, this knowledge is represented by belief functions.

The data X, are divided here into two components: a component known with certainty denoted by X^c and an uncertain component X^u. That is, each x_i of X^u is divided into (x_i^c, x_i^u) where x_i^c are the well defined attributes and x_i^u are the partially known attributes.

For uncertain attributes, we use a set of bba's $m_i^{\Omega^d} : 2^{\Omega^d} \rightarrow [0, 1]$ to express the a priori beliefs of the actual value of these attributes. 2^{Ω^d} denotes the power set corresponding to the set of possible values of the attribute d.

Example 1. Let us consider three attributes given by: the salary, the marital status, and the place of residence with respective possible categories:
$\Omega^{salary} = \{low, medium, high\}$,
$\Omega^{mariStat} = \{single, married, divorced, widowed\}$,
$\Omega^{placeRes} = \{apartment, house\}$.

One can have:
$x_1 = (\{low(0.2), medium(0.8)\}, married(1), house(1))$, where the attribute salary is uncertain (with $m_1^{\Omega^{salary}}(low) = 0.2$ and $m_1^{\Omega^{salary}}(medium) = 0.8$) and the remaining attributes are perfectly known.
$x_2 = (\{medium, high\}(1), married(1), \Omega^{placeRes}(1))$, where the attribute salary is imprecise, the attribute marital status is perfectly known and the attribute place of residence is totaly unknown.

This representation is then a generalization of the previous one since it covers the case where attribute values are missing. This is handled through vacuous belief functions. We also notice that the certain case, where all the attributes are perfectly known, can also be modeled here through certain belief functions.

E-Step. The E-step estimates both the variables z_{ik} and the uncertain values of x_i^u. The first values are again given by t_{ik} estimated over the certain component of x_i.

The second values denoted by $E[x_i^u|X^o, \Theta^{(t)}]$ are first estimated using the current parameters $p_k^{dj(t)}$. These probabilities which can be written in the form of Bayesian masses $m_{ik}^{\Omega^{d(t)}}$, are then combined with the a priori masses through the conjunctive rule of combination, to integrate our initial beliefs about the attribute values. The resulting masses denoted by $\hat{m}_{ik}^{\Omega^{d(t)}}$, are given by:

$$\hat{m}_{ik}^{\Omega^{d(t)}}(A) = \sum_{B \cap C = A} m_{ik}^{\Omega^{d(t)}}(B) m_i^{\Omega^d}(C) \ , \ \forall A \subseteq \Omega^d \ . \tag{22}$$

The updated estimation of x_i^{dj} expressing the probability that the attribute x_i^d has the category j, is denoted by $\hat{p}_{ik}^{dj(t)}$. These probabilities are derived from the resulting combined masses using the pignistic transformation:

$$\hat{p}_{ik}^{dj(t)} = \sum_{A \ni \omega_{dj}} \frac{\hat{m}_{ik}^{\Omega^{d(t)}}(A)}{|A|} \frac{1}{(1 - \hat{m}_{ik}^{\Omega^{d(t)}}(\emptyset))} \ , \tag{23}$$

where ω_{dj} denotes the category j of the attribute d.

M-Step. The M-step uses these estimations to update the current parameters as detailed in the previous subsection. The term $t_{ik}\hat{p}_{ik}^{dj}$ is used for the uncertain values in the equation (21).

Note that in both methods, the E and M steps are iterated until the likelihood function $Q(\Theta|\Theta^{(t)}) - Q(\Theta|\Theta^{(t-1)})$ becomes inferior to some threshold ε fixed a priori. This function is given in equation (8). As the classical EM approach [3], the proposed algorithms converge at a stationary point of the mixture parameters and provide a local maximum of the likelihood function.

6 Experimental Results

In order to evaluate our proposed method which consists in a partially supervised EM classification approach with imperfect knowledge at the attribute and class values, we have implemented two algorithms in Matlab V 7.0. Both algorithms deal with uncertain class labels. Besides, the former (GenCrEM_1) handles missing attribute values, whereas the latter (GenCrEM_2) deals with uncertain and/or imprecise attribute values.

We have then applied these algorithms on real databases obtained from the UCI Machine Learning Repository [10]. We have modified these databases in order to disturb their certainty: we have randomly eliminated some attribute values

Table 1. Description of databases

Database	#instances	#attributes	#classes
Balance scale	625	4	3
Wisconsin breast cancer	699	8	2
Car evaluation	1728	6	4

for the GenCrEM_1, and we have randomly introduced bba's in some attribute values for GenCrEM_2 by considering their initial certain values. Moreover, in both cases, we have randomly generated bba's on the class labels by taking into account the initial true labels. In Table 1, a brief description of these databases is given.

We have tested both algorithms for different percentages of missing and uncertain attribute values respectively for the GenCrEM_1 and the GenCrEM_2. We have then applied the CrEM [15] on the certain attribute part of the databases. Table 2 gives the percentages of correctly classified instances (PCC) compared with the initial classification for each database. The mean PCC's obtained from the three methods and measured over the considered databases are given in Figure 1.

It is found that the PCC's produced by GenCrEM_1 are higher than the PCC's given by CrEM for the three databases and for the three considered percentages of imperfect data (20%, 30% and 40%). For instance, in the Car evaluation database and with 40% of imperfect data, the PCC is equal to 70.61% for CrEM and 77.8% for GenCrEM_1. Besides, the results given by GenCrEM_2 are better than the ones given by GenCrEM_1 in all the test cases. For instance, the PCC is equal to 80.36% for GenCrEM_1 and $83,62\%$ for GenCrEM_2 in the Balance scale database and with 30% of imperfect data. So, GenCrEM_2 which is the generalized case, is very appropriate to integrate additional knowledge about the objects of the learning set even if this knowledge is uncertain. Furthermore, it is shown that while the PCC's of our method remain nearly constant and quite high (around 80% for GenCrEM_1 and 83% for GenCrEM_2) when the percentage of imperfect data increases, the PCC of CrEM shows considerable decrease from 75.3% (for 20% of imperfect data) to 69.89% (for 40% of imperfect data). So,

Table 2. Experimental results

	Balance scale			Wisconsin b.c.			Car evaluation		
percent_imperf_obj	20%	30%	40%	20%	30%	40%	20%	30%	40%
CrEM (in %)	75.25	71.67	69.5	73.87	70.33	69.56	76.8	73.23	70.61
GenCrEM_1 (in %)	80.67	80.36	80.62	81.26	82.5	82.69	78.73	78.26	77.8
GenCrEM_2 (in %)	82.4	83.62	82.85	84.53	83.82	84.68	82.13	81.75	82.28

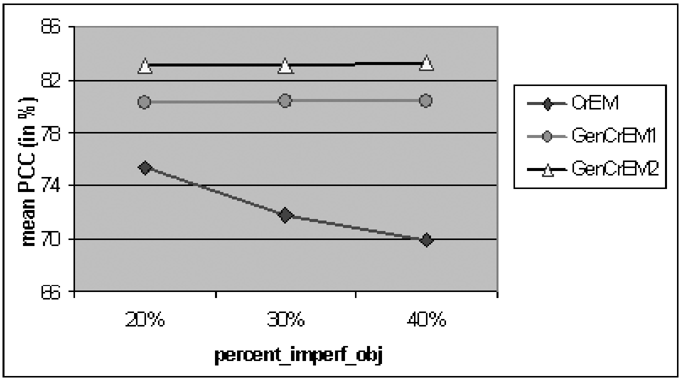

Fig. 1. Experimental results

our proposed approach is more appropriate to handle partially known attribute values.

We should mention that with our proposed method and if all the attribute bba's are certain, the results are equivalent to the CrEM. Besides, if both attribute and class values are perfectly known, that is when we are in a state of total certainty, the results are analogous to those obtained from the classical EM algorithm [3]. Note that when the class labels are imprecise, the CrEM produces very similar results than [1]. So our proposed approach is a generalization of these methods.

7 Conclusion

In this paper, we have proposed an EM approach for learning in an uncertain environment. The uncertainty is represented by belief functions as understood in the TBM. This approach is adapted for cases where not only the knowledge about the classes of the objects can be partial but also their characteristics. Our method provides a more flexible tool to deal with these situations. Future works are concerned with both continuous and mixed data. We will also focus on the model selection issue which notably includes the choice of the mixture components.

References

[1] Ambroise, C., Govaert, G.: EM algorithm for partially known labels. In: IFCS'2000, Namur, Belgium, vol. 1, pp. 161–166 (2000)

[2] Ben Hariz, S., Elouedi, Z., Mellouli, K.: Clustering approach using belief function theory. In: Euzenat, J., Domingue, J. (eds.) AIMSA 2006. LNCS (LNAI), vol. 4183, pp. 162–171. Springer, Heidelberg (2006)

[3] Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, series B 39, 1–38 (1977)
[4] 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)
[5] Elouedi, Z., Mellouli, K., Smets, P.: Belief decision trees: theoretical foundations. International Journal of Approximate Reasoning 28, 91–124 (2001)
[6] Everitt, B.: An introduction to latent variable models. Chapman and Hall, Sydney (1984)
[7] Figueiredo, M.A.T., Jain, A.K.: Unsupervised learning of finite mixture models. IEEE Transaction on Pattern Analysis and Machine Intelligence 24(3), 381–396 (2002)
[8] Ghahramani, Z., Jordan, M.I.: Supervised learning from incomplete data via an EM approach. In: Cowan, J.D., Tesauro, G., Alspector, J. (eds.) Advances in Neural Information Processing Systems, vol. 6, pp. 120–127. Morgan Kaufmann Publishers, San Francisco (1994)
[9] McLachlan, G.J., Basford, K.E.: Mixture models. Inference and application to clustering. Marcel Dekker, New York (1989)
[10] Newman, D.J., Hettich, S., Blake, C.L., Merz, C.J.: UCI Repository of machine learning databases (1998), http://www.ics.uci.edu/~mlearn/MLRepository.html
[11] Shafer, G.: A Mathematical Theory of Evidence. Princeton Univ. Press, Princeton (1976)
[12] Smets, P.: The combination of evidence in the transferable belief model. IEEE Pattern analysis and Machine Intelligence 12(5), 447–458 (1990)
[13] Smets, P.: Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem. International Journal of Approximate Reasoning 9, 1–35 (1993)
[14] Smets, P., Kennes, R.: The transferable belief model. Artificial Intelligence 66, 191–234 (1994)
[15] Vannoorenberghe, P., Smets, P.: Partially supervised learning by a credal EM approach. In: Godo, L. (ed.) ECSQARU 2005. LNCS (LNAI), vol. 3571, pp. 956–967. Springer, Heidelberg (2005)