International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:6, 2007

Selection Initial modes for Belief K-modes Method

Sarra Ben Hariz, Zied Elouedi, Khaled Mellouli

***Abstract*—**The belief K-modes method (BKM) approach is a new clustering technique handling uncertainty in the attribute values of objects in both the cluster construction task and the classification one. Like the standard version of this method, the BKM results depend on the chosen initial modes. So, one selection method of initial modes is developed, in this paper, aiming at improving the performances of the BKM approach. Experiments with several sets of real data show that by considered the developed selection initial modes method, the clustering algorithm produces more accurate results.

***Keywords*—**Clustering, Uncertainty, Belief function theory, Belief K-modes Method, Initial modes selection.

I. Introduction

CLustering techniques [16] are among the well known machine learning techniques to discover groups and identify interesting distributions in the considered data , and the K-modes method [14] is considered as one of the most popular of them. These techniques are used in many domains such as medicine, banking, finance, marketing, security, etc. They work under an unsupervised mode when the class label of each object in the training set is not known a priori.

The capability to deal with datasets containing uncertain information is very important due to the fact that this kind of datasets is common in real life data mining applications.

In fact, experts, in most real problems, may encounter several difficulties when expressing any classification parameters values and it may be more flexible to allow them providing their uncertainty relative to classification variables using a non-classical theory of uncertainty rather than exact values.

However, in such situations, standard methods cannot be applied for clustering such training objects. Thus and in order to overcome this drawback, many researches have been done to adapt standard methods to this kind of environment. The idea was to introduce theories that could represent and manage uncertainty. Several kinds of clustering techniques were developed, more precisely we mention the extensions of the K-modes method, : fuzzy K-modes method [15], [17], belief K-modes method (BKM)[3], etc . In our work, we will focus on belief K-modes method (BKM).

The belief K-modes method is a clustering approach adapted in order to handle uncertainty problem. Its main contributions are to provide one approach to deal with on one hand the construction of clusters where the attribute' values of training objects may be uncertain, and in the other hand the classification of new instances characterized also by uncertain values based on the obtained clusters. The uncertainty is represented by the Transferable Belief Model (TBM), one interpretation of the belief function theory. It is considered as a useful theory for representing and managing uncertain knowledge introduced by [27]. It permits to handle partial or even total ignorance concerning classification parameters, in a flexible way, and offers interesting means to combine several pieces of evidence. So, BKM is based on both the standard K-modes paradigm and belief function theory to handle uncertainty.

In fact, there are more developed classification techniques based on this theory to manage uncertainty, we mention, as supervised techniques, belief decision trees (BDT) [7], [11], belief k-nearest neighbor [6] which have provided interesting results. For unsupervised ones, belief function theory was also applied and that is the purpose of these clustering works [8], [9], [10], [22]. Contrary to these latter clustering approaches, the BKM method [3] deals with objects characterized by uncertain attributes either in the construction and the classification phases.

Like the standard hard version of this method, the BKM results depend on the chosen initial modes. Indeed, several works have been developed to deal with this latter problem: Huang [14] presented two methods of initialization for categorical data for K-modes method showing that if diverse initial modes are chosen then it could lead to better clustering results. Sun et al [28] developed an experimental study on applying Bradley and Fayyad's iterative initial-point refinement algorithm [5] to the k-modes clustering to improve the accurate and repetitiveness of the clustering results. Khan S.S. et al [18] proposed an algorithm to compute initial modes using Density based Multiscale Data Condensation. An other research work [19] extended the idea of Evidence Accumulation to categorical data sets by generating multiple partitions as different data organization by seeding K-modes algorithm, every time, with random initial modes. The resultant modes are then stored in a Mode Pool and the most diverse set of modes were computed, which were used as initial modes.

However, these latter techniques work in a certain context and fail to deal with this problem within uncertain environment.

To address the problem caused by the randomly choice of the initial modes under uncertainty, in this present study,

LARODEC, Institut Supérieur de Gestion de Tunis, 41 Avenue de la Liberté, cité Bouchoucha 2000 Le Bardo, Tunis, Tunisie.
Sarra Ben Hariz, email: sarra.benhariz@gmail.com
Zied Elouedi, email: zied.elouedi@gmx.fr
Khaled Mellouli, email: khaled.mellouli@ihec.rnu.tn

we develop one selection method for choosing initial modes instead of choosing them randomly using the dissimilarity matrix to provide the maximum of dissimilarity measure.

In this work, we focused on this initial step of the BKM process. So, we suggest to define selection method in order to improve clustering results based on the distance measure concept.

The remainder of this paper is organized as follows : Section 2 gives an overview of the belief function theory. In Section 3, we describe the BKM approach. As an improvement of this approach, in Section 4, we develop a new initial modes selection method instead of the randomly choice used in the initial version of BKM method. Finally, Section 5 presents and analyzes experimental results carried out on these two K-modes belief versions of the U.C.I. machine learning repository's data sets [20].

II. Belief function theory

A. Introduction

The belief function theory is appropriate to handle uncertainty in classification problems. So, In this section, the basic concepts of this theory as understood in the Transferable Belief Model (TBM), which is among the interpretations of this non-classical uncertainty's theory, are recalled briefly (for more details see [23], [24], [27]).

B. Background

Let Θ be a finite non empty set of elementary events to a given problem, called the frame of discernment. It also referred to as the universe of discourse or the domain of reference. This set contains hypotheses about some problem domain. All the subsets of Θ belong to the power set of Θ, denoted by 2^Θ and defined as follows:

$$2^\Theta = \{A : A \subseteq \Theta\} \tag{1}$$

Each element of 2^Θ is called a proposition or an event. The elements of Θ are called the elementary propositions.

The impact of a piece of evidence on the different subsets of the frame of discernment Θ is represented by the so-called basic belief assignment (bba), called initially [23] basic probability assignment. The bba is defined as follows:

$$m : 2^\Theta \mapsto [0,1]$$
$$\sum_{A \subseteq \Theta} m(A) = 1 \tag{2}$$

Each quantity m(A), named basic belief mass (bbm), is considered to be the part of belief that supports the event A, and that, due to the lack of information, does not support any strict subset of A.

The subsets A of the frame of discernment Θ such that m(A) is strictly positive, are called the focal elements of the bba m.

The pair (F,m) is called a body of evidence where F is the set of all the focal elements relative to the bba m.

The union of all the focal elements of m are named the core and are defined as follows:

$$\varphi = \bigcup_{A : m(A) > 0} A \tag{3}$$

A belief function, denoted bel, corresponding to a specific bba m, assigns to every subset A of Θ the sum of masses of belief committed to every subset of A by m [23]. It is defined as follows:

$$bel : 2^\Theta \mapsto [0,1]$$
$$bel(A) = \sum_{\phi \neq B \subseteq A} m(B) \tag{4}$$

The belief function bel represents the total belief that one commits to A without being also commited to $\overline{A}$. The bbm $m(\emptyset)$ is not included in bel(A) as $\emptyset$ is both a subset of A and $\overline{A}$.

The plausibility function pl associated with a mass function m quantifies the maximum amount of belief that could be given to a subset A of the frame of discernment. It is equal to the sum of the bbm's relative to subsets B compatible with A. The plausibility function pl is defined as follows [1]:

$$pl : 2^\Theta \mapsto [0,1]$$
$$pl(A) = \sum_{A \cap B \neq \phi} m(B) \tag{5}$$

A belief function is said to be vacuous belief function if Θ is its unique focal element [23]:

$$m(\Theta) = 1 \; and \; m(A) = 0 \; for \; all \; A \subset \Theta, A \neq \Theta \tag{6}$$

Such bba where Θ is the unique focal element, quantifies the state of *total ignorance* since there is no support given to any strict subset of Θ.

A certain belief function is a belief function such that it has only one focal element and which should be a singleton. Its corresponding bba is defined as follows:

$$m(A) = 1 \; and \; m(B) = 0 \; for \; all \; B \neq A \; and \; B \subseteq \Theta \tag{7}$$

where A is a singleton event of Θ. This function represents a state of total certainty as it assigns all the belief to a unique elementary event.

A belief function is said to be a simple support function (SSF) if it has at most one focal element different from the frame of discernment Θ. This focal element is called the focus of the SSF.

A SSF is defined as follows:

$$m(X) = \begin{cases} \omega & \text{if X = } \Theta \\ 1-\omega & \text{if X = A for some } A \subset \Theta \\ 0 & otherwise. \end{cases} \quad (8)$$

where A is the focus and $\omega \in [0, 1]$

C. Combination rules

Let m_1 and m_2 be two bba's defined on the same frame of discernment Θ. These two bba's are collected by two 'distinct' pieces of evidence and induced from two experts (information sources). These bba's can be combined either conjunctively or disjunctively [26].

- The Conjunctive Rule: when we know that both sources of information are fully reliable, then the bba representing the combined evidence satisfies [26]:

$$(m_1 \bigcap m_2)(A) = \sum_{B,C \subseteq \Theta; B \cap C = A} m_1(B) m_2(C) \quad (9)$$

- The Disjunctive rule of combination: when we only know that at least one of these sources of information is to be accepted, but we do not know which one. So, this rule is defined as follows [26]:

$$(m_1 \bigcup m_2)(A) = \sum_{B,C \subseteq \Theta; B \cup C = A} m_1(B) m_2(C) \quad (10)$$

Note that since the conjunctive and the disjunctive rules of combination are both commutative and associative, combining several pieces of evidence induced from distinct information sources (either conjunctively or disjunctively) may be easily ensured by applying repeatedly the chosen rule.

D. Decision Process

The TBM is based on a two level mental models:

- The credal level where beliefs are entertained and represented by belief functions.
- The pignistic level where beliefs are used to make decisions and represented by probability functions called the pignistic probabilities.

When a decision must be made, beliefs held at the credal level induce a probability measure at the pignistic measure denoted *BetP* [27].
The link between these two functions is achieved by the pignistic transformation.

$$BetP(A) = \sum_{B \subseteq \Theta} \frac{|A \cap B|}{|B|} \frac{m(B)}{(1 - m(\phi))}, \text{ for all } A \in \Theta \quad (11)$$

III. Belief K-modes Method

K-modes approach is considered as an efficient clustering method for classification problem. That's why, it is widely applied to a variety of problems in artificial intelligence. It extends the K-means [21] one by using a simple matching dissimilarity measure for categorical objects, modes instead of means for clusters, and a frequency-based method to update modes in the clustering process to minimize the clustering cost function.

Despite its accuracy when precise and certain data are available, the standard K-modes algorithm shows serious limitations when dealing with uncertainty. However, uncertainty may appear in the values of attributes of instances belonging to the training set that will be used to ensure the construction of clusters, and also in the classification of new instances which may be characterized by uncertain attribute values.

To overcome this limitation, a belief K-modes method (BKM) [3] was developed, a new clustering technique based on the K-modes method within the belief function framework. In this part of our paper, we expose the notations used within BKM framework. Next, we present the two major parameters of the belief K-modes method needed to ensure both the construction and the classification tasks, namely clusters' centers and the dissimilarity measure.

A. Notations

The following notations will be used in the following:

- T: a given data set of n objects.
- X_i: an object or instance, $i = 1, ..., n$.
- $A = \{A_1, ..., A_s\}$: a set of s attributes.
- Θ_j: the frame of discernment involving all the possible values of the attribute A_j related to the classification problem, $j = 1, ..., s$.
- D_j: the power set of the values of the attribute $A_j \in A$, with $|D_j| = 2^{|\Theta_j|}$
- $x_{i,j}$: the value of the attribute A_j for the object X_i, in the certain case.
- $m^{\Theta_j}\{X_i\}$: expresses the beliefs on the values of the attribute A_j corresponding to the object X_i, with $m^{\Theta_j}\{X_i\} = \{(c_{j,h}, m_i(c_{j,h}))|c_{j,h} \subseteq D_j\}$.
- $m_i(c_{j,h})$: denoted the bbm given to the category $c_{j,h} \subseteq D_j$ relative to the object X_i.
- $m_{Q_l}(c_{j,h})$: denoted the bbm given to the category $c_{j,h} \subseteq D_j$ relative to the mode Q_l of the cluster C_l.

B. The BKM parameters

As with standard K-modes method, building clusters within BKM needs the definition of its fundamental parameters, namely, cluster modes and the dissimilarity measure. These parameters must take into account the uncertainty encountered in the training set and that pervade the attribute values of training objects.

1) Cluster mode: Due to the uncertainty and contrary to the traditional training set where it includes only certain instances, the structure of considered training set is represented via bba's respectively to each attribute relative to each object, this training set offers a more generalized framework than the tradional one. Within this structure of training set, the belief K-modes cannot use the strategy used by the standard method which is the frequency-based method to update modes of clusters.

The idea is to apply the mean operator to this uncertain context since it permits combining bba's respectively to each attribute provided by all objects belonging to one cluster.
Note that using the mean operator offers many advantages since it satisfies these properties namely the associativity, the commutativity and the idempotency.

Using the mean operator solves also the non-uniqueness problem of modes encountered in the standard K-modes method.
Given a cluster $C = \{X_1, ..., X_p\}$ of objects, with $X_i = (x_{i,1}, ..., x_{i,s})$, $1 \leq i \leq p$. Then, the mode of C is defined by : $Q = (q_1, ..., q_s)$, with:

$$q_j = \{(c_j, m_{c_j}) | c_j \in D_j\} \tag{12}$$

where m_{c_j} is the relative bba of attribute value c_j within C.

$$m_{c_j} = \frac{\sum_{\mathrm{i=1}}^{\mathrm{p}} \mathrm{m_i(c_j)}}{|\mathrm{C}|} \tag{13}$$

with $C = \{X_1, X_2, ..., X_p\}$ and $|C|$ is the number of objects on C. m_{c_j} expresses the belief about the value of the attribute A_j corresponding to the cluster mode.

2) Dissimilarity measure: The dissimilarity measure has to take into account the bba's for each attribute for all objects in the training set, and compute the distance between any object and each cluster mode (represented by bba's). Many distance measures between two bba's were developed which can be charaterized into two kinds, namely the distance measures based on pignistic transformation [2], [12], [29], [30] and those between bba's defined on the power set [4], [13].

For the first category, one unavoidable step, which consists in the pignistic transformation of the bba's, may lose information given by the initial bba's. So, this kind of distance is not suitable within the BKM context.

While the second kind of belief distance measures are applied directly to bba's and not to the pignistic probabilities. These measures are defined on the power set. However, the measure developed by Fixen and Mahler [13] is a pseudo-metric since the condition of nondegeneracy of one distance metric is not respected. The other mentioned measure [4] within this category, verifies the basic properties for any distance measures namely, nonnegativity, nondegeneracy and symmetry. Thus, the BKM approach adapts this latter belief distance to this uncertain clustering context to compute the dissimilarity between any object and each cluster mode.

Indeed, this distance measure takes into account both the bba's distributions provided by the objects and one similarity matrix D which is based on the cardinalities of the subsets of the corresponding frame of one attribute and those of the intersection and union of these subsets.

Let m_1 and m_2 be two bba's on the same frame of discernment Θ_j, the distance between m_1 and m_2 is :

$$d(m_1, m_2) = \sqrt{\frac{1}{2}(\|\vec{m_1}\|^2 + \|\vec{m_2}\|^2 - 2 < \vec{m_1}, \vec{m_2} >)} \tag{14}$$

where $< \vec{m_1}, \vec{m_2} >$ is the scaler product defined by:

$$< \vec{m_1}, \vec{m_2} >= \sum_{w=1}^{2^{\Theta_j}} \sum_{z=1}^{2^{\Theta_j}} m_1(B_w) m_2(B_z) \frac{|\mathrm{B_w} \cap \mathrm{B_z}|}{|\mathrm{B_w} \cup \mathrm{B_z}|} \tag{15}$$

with $B_w, B_z \in D_j$ for w,z = 1,...,2^{Θ_j}, and $\| \vec{m} \|^2$ is then the square norm of $\vec{m}$: $\|\vec{m}\|^2 = < \vec{m}, \vec{m} >$

This scalar product is based on the bba's distributions (m_1, and m_2) and the elements of one similarity matrix D, which are defined as follows:

$D(B_w, B_z) = \frac{|\mathrm{B_w} \cap \mathrm{B_z}|}{|\mathrm{B_w} \cup \mathrm{B_z}|}$, where $B_w, B_z \in D_j$.

Thus, the dissimilarity measure between any object X_i and each mode Q can be defined as follows:

$$D(X_i, Q) = \sum_{j=1}^{m} d(m^{\Theta_j}\{X_i\}, m^{\Theta_j}\{Q\}) \tag{16}$$

where $m^{\Theta_j}\{X_i\}$ and $m^{\Theta_j}\{Q\}$ are the relative bba of the attribute A_j provided by respectively the object X_i and the mode Q.

C. The BKM algorithm

The BKM algorithm has the same skeleton as standard K-modes method. The different construction steps of this approach are described as follows:

(1) Giving K, the number of clusters to form.
(2) Select randomly K initial modes from the dataset objects.
(3) Assign all other objects to appropriate clusters based on the minimum of dissimilarity measure using Equation 16.
(4) Once all objects have been allocated to clusters, compute seed points as the clusters' modes of the current partition using the mean operator (see Equations 12 and 13).
(5) Update the partition of all objects according to their distances respectively to clusters' modes computed in the previous step.
(6) Go back to step 4, reiterate the step 4 and 5 and stop when no more new assignment. In other words, all objects have no changed clusters.

Once the clusters' construction is done, the classification of a new object that may be charaterized by uncertain attribute

values, we have to assign it to the most similar cluster based on its distance over the obtained clusters resulting from the construction phase, and using the distance measure (See Equation 16).

IV. Selection method for choosing initial modes

A. Principle

As mentioned above, clustering accuracy of BKM algorithm for categorical uncertain data depends upon the choice of initial data points (modes) which affect the clustering results.

The BKM method, like the standard version of the K-modes method, produces locally optimal solutions that are dependent on the initial modes by selecting randomly K initial cluster modes from the data set to cluster. Since, the algorithm is significantly sensitive to this choice, the BKM algorithm is run multiple times to reduce this effect.

As known, the aim of clustering is to group a set of objects into classes of similar objects by decreasing the dissimilarity within the same cluster and increasing this measure between distinct clusters.
So, in this section, we propose a new method for the selection of the K initial modes to improve our approach. Our method consists in choosing K objects from the data set which provide the maximum of dissimilarity measure between them. It means that the chosen K initial modes are the K objects which are the most far ones. The modes should be placed in a cunning way because different locations cause different results. However, to place them as much as possible far away from each other is the better manner to have the initial K modes.
The purpose of this method is to make the initial modes diverse, which can lead to better clustering results.

Since the selection method of the initial modes affect convergence of our BKM method, our intention is to study the impact of the new proposed selection method of the initial modes on the clustering results, in order to improve our initial approach results (BKM).

B. Algorithm for choosing initial modes

As shown in the section 3, the quality of clustering results depends on the initial values of the modes. This is a complex problem for which many approaches have been proposed within certain context [5], [14], [18], [19], [28].

In [14], the author includes two initial mode selection methods. The first one selects the first K distinct records from the dataset as the initial K modes. The second one is implemented based on the frequency concept. Indeed, it consists in calculating the frequencies of all categories for all attributes and assigning the most frequent values equally to the initial K modes. After that, each obtained mode must be replaced with the most similar record from the given dataset. In fact, within our framework, we have bba's representing the attribute values not the certain categories. So, we opt to use the distance concept instead of the frequency one.

The initial version of our BKM method starts with randomly chosen K objects as the initial modes which is the most popular way to start the algorithm.

In our case, and in order to overcome this drawback, a new method for choosing the initial modes was proposed, instead of the standard one.
The main idea is to select the K objects as K cluster modes based on the dissimilarity measure notion instead of the randomly choice.

To this end, we have to compute the dissimilarity matrix corresponding to n objects and assign the two objects which provide the highest dissimilarity measure to the two first cluster modes. When K is equal to two, the procedure is completed and the selection of our initial modes is done. Otherwise, for the other objects, we need to calculate their distance sum respectively to the already fixed modes and associate to the following mode the object with highest dissimilarity measure (the most far from the chosen modes).

After we have these modes, a new iteration has to be done for the remaining objects. We iterate until the selection of the K initial modes has been done.

This algorithm has as inputs a predefined number of modes (K) and the data set to cluster. It is composed of the following steps:

(1) Compute the dissimilarities matrix D, $n \times n$, where $D_{i,j}$ is the distance between the object X_i and the object X_j based on the distance measure defined by Equation 27 with $i, j \in \{1, .., n\}$.
(2) Allocate the two objects X_i and X_j which provide the highest dissimilarity measure ($D_{i,j} = max(D)$) to the two first clusters' modes.
(3) For the other objects (n - number of chosen cluster modes), compute their distances' sum respectively to the already chosen modes
(4) Assign the object with the highest distance measure to the following mode.
(5) Repeat 3 and 4 till the K initial modes have been chosen.

C. Example

Let us illustrate our method by a simple example. Assume that a firm wants to group its staff by taking into account their attributes.

Let T be a training set (see Table 1) composed of seven instances characterized by three categorical attributes:

- Qualification with possible values {A, B, C}.
- Income with possible values {High , Low, Average}.
- Department with possible values {Finance, Accounts, Marketing}.

For each attribute A_j for an object X_i belonging to the training set T, we assign a bba $m^{\Theta_j}\{X_i\}$ expressing beliefs

on its assigned attributes values, defined respectively on:

$\Theta_1 = \{A, B, C\}$
$\Theta_2 = \{High, Low, Average\}$
$\Theta_3 = \{Finance, Accounts, Marketing\}$.

If we consider that only the department attribute is known with uncertainty. The structure of the data set T is defined by the Table 1.

TABLE I
DATA SET T RELATIVE TO BKM

Objects	Qualification	Income	Department
X_1	A	High	$m^{\Theta_3}\{X_1\}$
X_2	B	Low	$m^{\Theta_3}\{X_2\}$
X_3	C	Average	$m^{\Theta_3}\{X_3\}$
X_4	C	Average	$m^{\Theta_3}\{X_4\}$
X_5	B	Low	$m^{\Theta_3}\{X_5\}$
X_6	A	High	$m^{\Theta_3}\{X_6\}$
X_7	B	Low	$m^{\Theta_3}\{X_7\}$

Where:

- $m^{\Theta_3}\{X_1\}(\{Finance$=0.5; $m^{\Theta_3}\{X_1\}(\{Finance, Accounts\})$ =0.3 $and\ m^{\Theta_3}\{X_1\}(\Theta_3)$=0.2
- $m^{\Theta_3}\{X_2\}(\{Finance\})$=0.8 $and\ m^{\Theta_3}\{X_2\}(\Theta_3)$=0.2
- $m^{\Theta_3}\{X_3\}(\{Marketing\})$=0.8; $m^{\Theta_3}\{X_3\}(\{Finance, Accounts\})$=0.1 $and\ m^{\Theta_3}\{X_3\}(\Theta_3) = 0.1$
- $m^{\Theta_3}\{X_4\}(\{Accounts\})$=0.8 $and\ m^{\Theta_3}\{X_4\}(\Theta_3)$=0.2
- $m^{\Theta_3}\{X_5\}(\{Marketing\})$=0.8 $and\ m^{\Theta_3}\{X_5\}(\Theta_3)$=0.2
- $m^{\Theta_3}\{X_6\}(\{Finance, Accounts\}) = 0.8$ $and\ m^{\Theta_3}\{X_6\}(\Theta_3) = 0.2$
- $m^{\Theta_3}\{X_7\}(\{Accounts\})$=0.8 and $m^{\Theta_3}\{X_7\}(\Theta_3)$=0.2

If we consider this training set T defined by Table 1.
Let us try to apply the proposed method for selection the 2 initial modes.
We have initially to define the dissimilarities matrix. To this end, for each object X_i, compute the dissimilarities $D(i, j)$, where $i, j \in \{1, .., 7\}$, using the dissimilarity measure defined by Equation 16.
The dissimilarities matrix is defined as follows:

$$\begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \end{array} \begin{array}{c} \begin{array}{ccccccc} X_1 & X_2 & X_3 & X_4 & X_5 & X_6 & X_7 \end{array} \\ \begin{pmatrix} 0 & 2.212 & 2.724 & 2.667 & 2.751 & 0.524 & 2.667 \\ 2.212 & 0 & 2.777 & \mathbf{2.8} & 0.8 & 2.566 & 0.8 \\ 2.724 & 2.777 & 0 & 0.777 & 2.063 & 2.768 & 2.777 \\ 2.667 & 2.8 & 0.777 & 0 & 2.8 & 2.566 & 2 \\ 2.751 & 0.8 & 2.063 & 2.8 & 0 & 2.8 & 0.8 \\ 0.524 & 2.566 & 2.768 & 2.566 & 2.8 & 0 & 2.566 \\ 2.667 & 0.8 & 2.777 & 2 & 0.8 & 2.566 & 0 \end{pmatrix} \end{array}$$

Note that $D(i, j) = 0$, if $i = j$.
$Max(D) = 2.8$. The first $D(i, j) = 2.8$ is obtained for $i = 2$ and $j = 4$.
So, we allocate the two objects X_2 and X_4 to the two initial modes.
If we have $K > 2$, we will iterate until obtaining the K initial modes as explained by the algorithm (see Section IV.B).

V. EXPERIMENTS

A. Framework

We have developed programs in Matlab V6.5 for the evaluation of the proposed selection method for the initial modes. Then, we have applied these programs to real databases obtained from the U.C.I repository of Machine Learning databases [20]. Since there are not available real datasets containing uncertainty within the belief function framework, we have modified these databases by introducing uncertainty in the attribute values of their instances in order to obtain uncertain data sets like within BKM framework.

We assume, as within BKM framework, the choice of numbers of clusters (K) to form is the same as the known classes number of actual datasets. A brief description of these databases is given in Table 4. #instances,#attributes, #classes denote respectively the total number of instances, the number of attributes and the number of classes.

TABLE II
DESCRIPTION OF DATABASES

Database	#intances	#attributes	#classes
Congressional voting records database	435	16	2
Balance scale database	625	4	3
Wisconsin breast cancer database	699	9	2
Solar Flare database	1389	10	3
Nursery database	12960	8	5

B. Uncertain data sets: Artificial creation

Attributes' values of training instances are perfectly known in the standard K-modes method and also in the belief clustering methods. However, in this work like the BKM framework, uncertainty is introduced in the values of attributes and it is presented through bba's concept given on the set of possible attribute values. So, the question is how construct these bba's to obtain uncertain data sets, since there is not real databases within the belief framework.

These bba's are created artificially by taking into account these following basic parameters:

- The real attributes' values of the training instances.
- Degree of uncertainty p (one for each attribute): it will vary in [0, 1] interval. The fixed value of p has a direct effect on the quality of results. In fact, for a large value of p, the number of the correctly classified instances will decrease.
  We will consider these four different intervals of p for our simulations:
  - Level 1: we take $0 < p \leq 0.25$
  - Level 2: we take $0.25 < p \leq 0.5$
  - Level 3: we take $0.5 < p \leq 0.75$
  - Level 4: we take $0.75 < p \leq 1$
- Percent of uncertain objects in a dataset: it represents the percent of generated uncertain objects for one given dataset.

The idea is to assign to each attribute, a bba over the set of remaining attributes of this object, based on the set of their possible values . After precise the uncertainty degree respective to each attribute, we affect $1 - f$, where f is a randomly number that must be less or equal to p (p is the

uncertainty degree) as bbm to the certain attribute's value and f to the frame of discernment corresponding to this attribute.

So, each bba has 2 focal elements as defined by the SSF (see Equation 8), where the focus is the real attribute value and $w = f$. These resulting bba's describe our belief about the value of the actual attributes' values which the object has.

C. Evaluation criteria

Huang [14] proposed a measure of clustering results called the clustering accuracy r computed as follows: $r = \frac{\sum_{i=1}^{k} a_i}{n}$, where n is the number of instances in the dataset, k is the number of clusters, a_i is the number of instances occurring in both cluster i and its corresponding labeled class. This criterion is equivalent to the ordinary PCC which represents the percent of correct classification of the instances which are classified according to the considered procedure (BKM methods). It is given by:

$$PCC = \frac{number\ of\ well\ classified\ instances}{total\ number\ of\ classified\ instances} \times 100 \quad (17)$$

The obtained cluster is considered as the class of the instance. Consequently, the number of well classified instances corresponds to the number of instances for which the cluster obtained by the BKM approaches is the same as their real class.

In our simulations, in order to obtain an unbiased estimation of the *PCC*, we have used a certain number of tests and after that we will calculate the final PCC as the average of all obtained ones. Since the initial modes affect the clustering results, we will use the following validation procedure which consists in randomly permutation of the n instances of a given data set to cluster, at each run, and the K first objects will be extracted to compute the initial clusters' modes. The procedure is repeated 10 times, each time using another K first instances as the initial clusters' modes.

Obviously, in each fold, we compute the corresponding *PCC* and the final *PCC* is given by the mean of the already computed ones. It is M.PCCs

D. Experimental results

Our simulations were performed for two cases, namely, the certain case and the uncertain case for a larger number of datasets. Note that the experimentations were performed using a Centrino 1.6 GHz PC with 1024 Mo of RAM running Windows XP. Note that all obtained PCC's values for both cases, showing in the following tables, are expressed in percentage.

1) Certain case: The first case tests the efficiency of our method (the initial version) and its improved version when there is no uncertainty in attributes' values, it means that each attribute is known with certainty and it has an unique value, and compares the results with ones obtained by applying the BKM method.

TABLE III
EXPERIMENTAL RESULTS : CERTAIN CASE

Database	M.PCC % (BKM1)	M.PCC % (BKM2)
Congressional voting records database	88.35	88.75
Balance scale database	79.20	80.25
Wisconsin breast cancer database	72.48	74.25
Solar flare database	87.85	88.25
Nursery database	90.75	91.25

Note that the second column (BKM1) is relative to results of the initial version of the BKM method which is based on the standard selection method for choosing the initial modes (random choice). However, the last one (BKM2) represents those of the improved BKM method, second version, by inducing the proposed method advanced in Section 5.

It can be seen that the clustering results have improved when the modes are chosen by our proposed method in comparison to the random selection of initial modes. Figure 1 represents these results graphically.

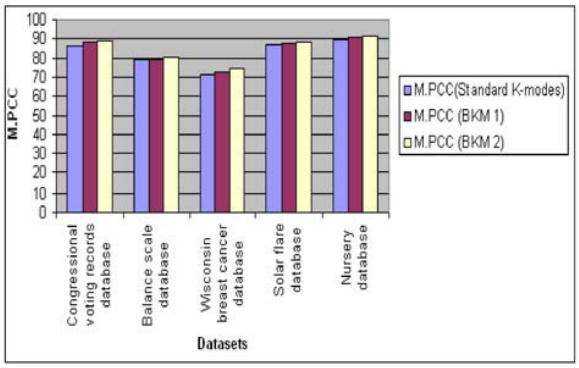

Fig. 1. M.PCC on the certain case

From this figure and Table 2, we can conclude that by applying the extended version (BKM2), based on the developed method for selection the initial modes, an improvement of the PCC's values is mentioned comparing to those obtained by the first version of our BKM approach.

2) Uncertain case: The following tables (Tables 4, 5, 6, 7 and 8) summarize different results carried out from testing the methods (the initial BKM in the first part of each column and the extended version using the developed selection initial modes method in the second one) in uncertain context to the same five datasets respectively to the four intervals of uncertainty degree p defined before and the different values of the uncertainty percent of dataset intances which are defined as follows : 25%, 50%, 75% and 100%.

To conduct experimental comparison and to verify the efficacy of our proposed approach (BKM2), let us move to

the following tables.

TABLE IV
EXPERIMENTAL RESULTS (CONGRESSIONAL VOTING RECORDS, M.PCC ON THE UNCERTAIN CASE)

Percent	25%		50%	
Degree	BKM1	BKM2	BKM1	BKM2
$0 < p \leq 0.25$	91.25	92.58	89.13	90.75
$0.25 < p \leq 0.5$	90.11	93.76	87.52	89.57
$0.5 < p \leq 0.75$	88.82	89.55	86.03	88.75
$0.75 < p \leq 1$	85.17	88.45	83.57	84.95
Means %	88.84	91.08	86.56	88.50
Percent	75%		100%	
Degree	BKM1	BKM2	BKM1	BKM2
$0 < p \leq 0.25$	88.15	88.80	86.23	87.95
$0.25 < p \leq 0.5$	87.48	88.45	81.13	83.75
$0.5 < p \leq 0.75$	85.57	87.54	76.29	80.25
$0.75 < p \leq 1$	84.12	86.44	71.63	75.56
Means %	86.33	87.80	78.82	81.87

TABLE V
EXPERIMENTAL RESULTS (BALANCE SCALE, M.PCC ON THE UNCERTAIN CASE)

Percent	25%		50%	
Degree	BKM1	BKM2	BKM1	BKM2
$0 < p \leq 0.25$	82.26	83.75	80.12	82.12
$0.25 < p \leq 0.5$	79.89	82.15	79.95	82.75
$0.5 < p \leq 0.75$	78.23	80.42	76.30	79.56
$0.75 < p \leq 1$	75.83	77.56	76.01	78.75
Means %	79.05	80.97	78.09	80.79
Percent	75%		100%	
Degree	BKM1	BKM2	BKM1	BKM2
$0 < p \leq 0.25$	79.29	81.35	77.22	79.75
$0.25 < p \leq 0.5$	78.56	80.78	77.15	79.45
$0.5 < p \leq 0.75$	75.59	77.75	74.35	77.65
$0.75 < p \leq 1$	74.44	76.65	71.23	73.95
Means %	76.97	79.13	74.98	77.7

TABLE VI
EXPERIMENTAL RESULTS (WISCONSIN BREAST CANCER, M.PCC ON THE UNCERTAIN CASE)

Percent	25%		50%	
Degree	BKM1	BKM2	BKM1	BKM2
$0 < p \leq 0.25$	75.24	78.56	75.13	77.55
$0.25 < p \leq 0.5$	73.55	75.55	73.21	75.75
$0.5 < p \leq 0.75$	71.03	73.42	70.81	73.55
$0.75 < p \leq 1$	70.99	72.45	69.18	71.15
Means %	72.70	74.99	72.08	74.5
Percent	75%		100%	
Degree	BKM1	BKM2	BKM1	BKM2
$0 < p \leq 0.25$	74.08	77.45	73.91	75.95
$0.25 < p \leq 0.5$	72.42	74.45	71.51	74.85
$0.5 < p \leq 0.75$	68.69	70.55	68.88	70.4
$0.75 < p \leq 1$	68.87	70.56	68.02	70.57
Means %	76.97	79.13	74.98	77.7

TABLE VII
EXPERIMENTAL RESULTS (SOLAR FLARE, M.PCC ON THE UNCERTAIN CASE)

Percent	25%		50%	
Degree	BKM1	BKM2	BKM1	BKM2
$0 < p \leq 0.25$	95.95	96.05	94.42	94.57
$0.25 < p \leq 0.5$	92.5	92.75	91.54	92.57
$0.5 < p \leq 0.75$	89.55	90.05	87.56	89.25
$0.75 < p \leq 1$	85.5	87.75	82.55	84.55
Means %	90.87	91.65	89.017	90.23
Percent	75%		100%	
Degree	BKM1	BKM2	BKM1	BKM2
$0 < p \leq 0.25$	92.75	93.55	87.85	88.05
$0.25 < p \leq 0.5$	88.75	90.45	86.4	87.75
$0.5 < p \leq 0.75$	81.7	83.45	79.75	81.05
$0.75 < p \leq 1$	78.90	80.95	75.45	77.05
Means %	85.52	87.1	82.36	83.47

TABLE VIII
EXPERIMENTAL RESULTS (NURSERY, M.PCC ON THE UNCERTAIN CASE)

Percent	25%		50%	
Degree	BKM1	BKM2	BKM1	BKM2
$0 < p \leq 0.25$	95.9	95.95	94.5	94.72
$0.25 < p \leq 0.5$	93.75	94.05	91.75	92.05
$0.5 < p \leq 0.75$	89.98	90.12	88.52	89.05
$0.75 < p \leq 1$	87.75	88.05	82.55	83.75
Means %	91.84	92.04	89.33	89.89
Percent	75%		100%	
Degree	BKM1	BKM2	BKM1	BKM2
$0 < p \leq 0.25$	94.0	94.05	89.9	90.05
$0.25 < p \leq 0.5$	90.55	91.15	88.75	89.05
$0.5 < p \leq 0.75$	86.75	87.25	83.95	84.45
$0.75 < p \leq 1$	79.56	80.35	74.51	75.75
Means %	87.71	88.2	84.27	84.82

It is found that the clustering results produced by the proposed method (initial version: BKM1) are very high in accuracy. However, including the developed selection initial modes method (BKM2) instead of the randomly choice used initially, as with the standard version, improves the obtained results and we mention an increase in the PCC's values.

In fact, these PCC's show that our method presented interesting results, these results confirm that our approach is well appropriate within the uncertain context.

Detailed results (PCC), for each data set, within this uncertain case and for different values of our uncertainty parameters, are given by figure 2.

As with the certain case, the results show that the proposed approach deals with uncertain instances as good as with certain instances. For instance, if we analyze the results shown in Figure 2, we remark that the PCC (uncertain case) remains high in average as in the certain case. These results, certify that our proposed approach is also well adapted to classify instances with uncertain attribute values.

Note that the PCC's decrease in average where uncertainty degree increases for a fixed uncertainty percent as shown in Figure 2.

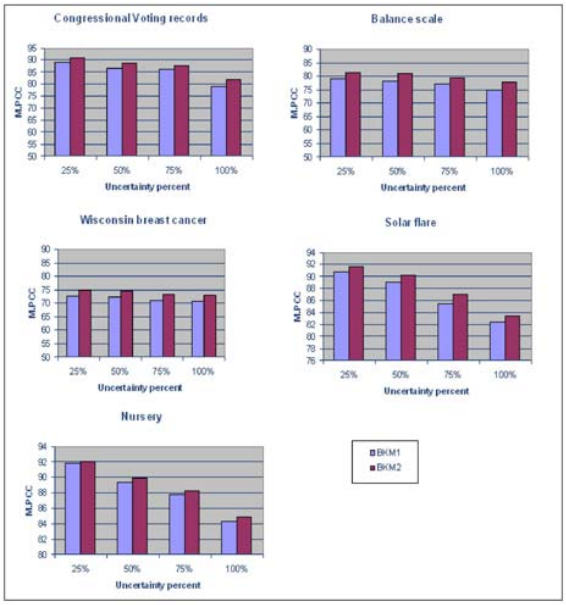

Fig. 2. M.PCC on the uncertain case(BKM1 compared with BKM2)

For example for the Congressional voting records database and if we consider that 100% of instances are uncertain, the PCC value becomes 71.63% when the uncertainty degree is in [0.75, 1[, which is considered as a high uncertainty, comparing to 86.23% with a low uncertainty (the uncertainty degree is as follow: $0 < p <= 0.25$) as shown in Table 4.

Figure 2 allows us to make a comparison between BKM1 clustering results and the BKM2 ones. It shows the effect of initial modes on clustering results. In fact, this improvement of results proves that the developed method for selection initial modes provides an interesting approach to choose them comparing to the standard one consisting in a random choice from the data sets objects.

As a conclusion, we have to note that these values are purely experimental and that depend on the used database and even on the used uncertainty degree and the considered uncertainty percent within a given database.

In fact, the clustering results produced by the proposed method are very high in accuracy. The PCC's show that our method presented interesting results. So, these results confirm that our approach is well appropriate within the uncertain context.

VI. Conclusion

Our objective through this work is to develop one initial modes selection method under uncertainty for the BKM approach in belief function theory framework as explained in TBM. Our method is developed to cope with the problem of of choosing random initial modes.

First, we have exposed the BKM approach, then we have improved it by developing a new selection method for choosing the initial modes based on the dissimilarity measure concept. The BKM method is the result of the combination between the standard K-modes method as clustering technique and the belief function theory to handle uncertainty problem. This uncertainty concerning attributes values can appear in both phases namely construction and classification one.

We have performed simulations on commonly used datasets obtained from the U.C.I repository [20] in order to evaluate and compare the performances of our belief clustering methods on the accuracy of the clustering results. Using the PCC as an evaluation criterion, the encouraging results of the experiments show the efficiency of our extended approach in both the certain and uncertain cases.

We plan to use our approach, using the developed initial modes selection approach, for real life problems especially in detection intrusion problems. In addition to the uncertainty on attribute values, another line of research will be to assume that each object in the training set may belong to more than one cluster, this uncertainty in the cluster membership can be represented via belief functions. However, this extension of the current work will allow us to compare our method with the belief clustering one proposed in [9].

An interesting future work is to make our method able to cluster datasets characterized by continuous attributes. Thus, the proposed method will be more flexible to handle mixed numerical and categorical databases.

References

[1] J.A. Barnett, Calculating Dempster-Shafer plausibility, IEEE Transactions on Pattern Analysis and Machine Intelligence, 13 (6), 1991, pp.599-602.
[2] M. Bauer, Approximation algorithms and decision making in the Dempster-Shafer theory of evidence - an empirical study, Int.J.Approx.Reason. 17 (2-3), 1997, pp.217.
[3] S. Ben Hariz, Z. Elouedi, and K. Mellouli, Clustering Approach using Belief Function Theory, Proceeding of the Twelfth International Conference on Artificial Intelligence: Methodology, Systems, Applications (AIMSA2006), 2006, pp.162-171.
[4] E. Bosse, D. Grenier and A.L. Jousselme, A new distance between two bodies of evidence, In Information Fusion 2, 2001, pp.91-101.
[5] P.S. Bradley, U.M. Fayyad, Refining Initial Points for K-Means Clustering, Proceedings of the 15th International Conference on Machine Learning (ICML98), San Francisco, Morgan Kaufmann, 1998.
[6] T. Denoeux, A k-nearest neighbor classification rule based on Dempster-Shafer theory, IEEE Transactions on Systems, Man and Cybernetics, 25 (5), 1995, pp.804-813.
[7] T. Denoeux and M. Skarstein-Bjanger, Induction of decision trees from partially classified data, Proceedings of SMC'2000,, Nashville, TN., 2000, pp.2923-2928.
[8] T. Denoeux and M. Masson, Clustering of proximity data using belief functions, Proceedings of IPMU'2002, Annecy, France., Vol I, 2002, pp.609-616.
[9] T. Denoeux and M. Masson, EVCLUS: Evidential Clustering of Proximity Data, IEEE Transactions on Systems, Man and Cybernetics, 34 (1), 2003, pp.95-109.
[10] T. Denoeux and M. Masson, Clustering Interval-valued Data using Belief Functions, Pattern Recognition Letters, 25 (2), 2004, pp.163-171.
[11] Z. Elouedi, K. Mellouli and P. Smets, Belief Decision trees: Theoretical foundations, International Journal of Approximat Reasoning, 28 (2-3), 2001, pp.91-124.
[12] Z. Elouedi, K. Mellouli and P. Smets, Assessing sensor reliability for multisensor data fusion within the transferable belief model, IEEE Trans.Syst.Man Cybern, 34 (1), 2004, pp.782-787.

[13] D. Fixen and R.P.S. Mahler, The modified Dempster-Shafer approach to classification, IEEE Trans.Syst.Man Cybern, 27 (1), 1997, pp.96-104.
[14] Z. Huang, Extensions to the k-means algorithm for clustering large data sets with categorical values, Data Mining Knowl.Discov., 2 (2), 1998, pp.283-304.
[15] Z. Huang and M.K. Ng, A fuzzy K-modes algorithm for clustering categorical data. IEEE Transaction on Fuzzy Systems,7(4), 1999, pp.446-452.
[16] A.K. Jain and R.C. Dubes, Algorithms for clustering data, Prentice-Hall, Englewood cliffs, NJ, 1988, pp.197-198.
[17] D-W. Kim and K.H. Lee, Fuzzy clustering of categorical data using fuzzy centroids. Pattern Recognition Letters, 25, 2004 , pp.1263-1271.
[18] S.S. Khan, A. Ahmad, Cluster center initialization algorithm for K-means clustering. Pattern Recognition Letters, 25 (11), 2004, 1293-1302.
[19] S.S. Khan, Dr.S. Kant, Computation of Initial Modes for K-modes Clustering Algorithm using Evidence Accumulation. IJCAI-07, 2007, 2784-2789.
[20] MP. Murphy and D.W. Aha, Uci repository databases.http://www.ics.uci.edu/mlearn, 1996.
[21] J. MacQueen, Some methods for classification and analysis of multivariate observations, Proceeding of the Fifth Berkeley Symposium on Math, Stat and Prob., 1, 1967, pp.281-296.
[22] J. Schubert, Clustering belief functions based on attracting and conflicting metalevel evidence, Intelligent Systems for Information Processing: From Representation to Applications, 2003.
[23] G. Shafer, A mathematical theory of evidence, Princeton Univ. Press. Princeton, NJ.30, 1976.
[24] P. Smets and R. Kennes, The transferable belief model, Artificial Intelligence, 66, 1994, pp.191-234.
[25] P. Smets, The combination of evidence in the transferable belief model, IEEE-Pattern analysis and Machine Intelligence, 12, 1990, pp.447-458.
[26] P. Smets, Belief functions: The disjunctive rule of combination and the generalized bayesian theorem, International Journal of Approximate Reasoning, 9, 1993, pp.1-35.
[27] P. Smets, The transferable belief model for quantified belief representation, In D.M. Gabbay and P. Smets (Eds.), Handbook of defeasible reasoning and uncertainty management systems, 1, 1998, pp.267-301.
[28] Y. Sun, Q. Zhu, Z. Chen: An Iterative initial points refinement algorithm for categorical data clustering, Pattern Recognition Letters, 23, 875-884, 2002.
[29] B. Tessem, Approximations for efficient computation in the theory of evidence, Artif.Intell, 61 (2), 1993, pp.315-329.
[30] L.M. Zouhal and T. Doeneux, An evidence-theory k-NN rule with parameter optimization, IEEE Trans.Syst.Man Cybern. 28 (2), 1998, pp.263-271.