



**HAL**  
open science

# Managing uncertainty during CBR systems vocabulary maintenance using Relational Evidential C-Means

Safa Ben Ayed, Zied Elouedi, Eric Lefevre

► **To cite this version:**

Safa Ben Ayed, Zied Elouedi, Eric Lefevre. Managing uncertainty during CBR systems vocabulary maintenance using Relational Evidential C-Means. 2020 International Multi-Conference on: “Organization of Knowledge and Advanced Technologies” (OCTA), Feb 2020, Tunis, France. pp.1-10, 10.1109/OCTA49274.2020.9151771 . hal-03643797

**HAL Id: hal-03643797**

**<https://hal.science/hal-03643797>**

Submitted on 16 Apr 2022

**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.

# Managing uncertainty during CBR systems vocabulary maintenance using Relational Evidential C-Means

Safa BEN AYED  
Université de Tunis, Institut Supérieur  
de Gestion, LARODEC, Tunis, Tunisia  
Univ. Artois, EA 3926, LGI2A,  
62400 Béthune, France  
Email: safa.ben.ayed@hotmail.fr

Zied ELOUEDI  
Université de Tunis, Institut Supérieur  
de Gestion, LARODEC, Tunis, Tunisia  
Email: zied.elouedi@gmx.fr

Eric LEFEVRE  
Univ. Artois, EA 3926, LGI2A,  
62400 Béthune, France  
Email: eric.lefevre@univ-artois.fr

**Abstract**—Due to the incremental learning of Case-Based Reasoning (CBR) systems, there is a colossal need to maintain their knowledge containers which are (1) the case base, (2) similarity measures, (3) adaptation, and (4) vocabulary knowledge. Actually, the vocabulary presents the basis of all the other knowledge containers since it is used for their description. Besides, CBR systems store real-world experiences which are full of uncertainty and imprecision. Therefore, we propose, in this paper, a new policy to maintain vocabulary knowledge using one of the most powerful tools for uncertainty management called the belief function theory, as well as the machine learning technique called Relational Evidential C-Means (RECM). We restrict the vocabulary knowledge to be the set of features describing cases, and we aim to eliminate noisy and redundant attributes by taking into account the correlation between them.

## 1. Introduction

Case-Based Reasoning (CBR) is a methodology based on solving new problems using past experiences. Actually, each new target problem triggers an entire cycle; In a nutshell, the CBR system (a) *retrieves* from the Case Base (CB) the most similar case(s). Then, it (b) *reuses* that case so as to propose an adapted solution. However, this solution can be rejected, so it should be (c) *revised*. Finally, the confirmed solution with its corresponding problem will be retained in the CB to serve, as a new case, for future problems resolution [1] (see Figure 1).

Since CBR systems proved over the years a widespread interest in several domains, and since they are designed to work over long time frames, their Knowledge Containers (KC) [2] are now the subject of a considerable maintenance targets. There are four KCs in CBR systems: (1) *CB*, (2) *Vocabulary*, (3) *Similarity measures* and (4) *Adaptation rules*. Their maintenance turns around the field of Case-Based Reasoner Maintenance (CBRM) [3]. In fact, research are mainly focused on Case Base Maintenance (CBM) [4], [5], [6], [7], [8] as well as models to estimate CBs competence in problem solving [9], [10]. However, there is no deep ex-

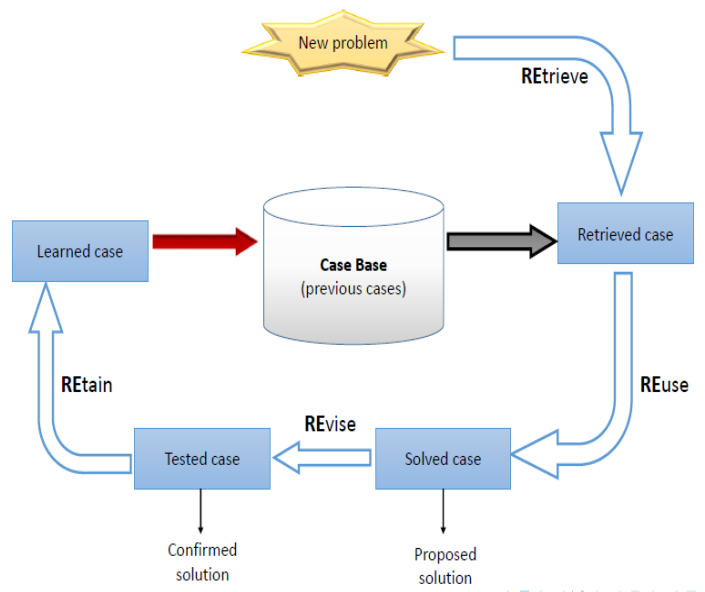


Figure 1. The CBR cycle

ploration on maintaining the vocabulary container although it presents the basis of all the other knowledge containers [2]. Actually, it needs maintenance for different reasons like context modification and domain overfitting. The most prevalent structure of vocabulary knowledge within CBR is the attribute-value representation [2]. Otherwise, other structures may be used such as predicates, set of words, and related constructed. In this work, we present the vocabulary knowledge as the set of features describing cases. Hence, it is relevant to maintain them by keeping only those that improve the competence of CBR systems in problem-solving. Generally, there are two kinds of knowledge that researchers aim to eliminate during maintenance: noisiness and redundancy. Noisy attributes present feature's knowledge that distorts problems solving operations. Redundant attributes have no added-value in offering accurate solutions, and their elimination will improve the performance of the

CBR system.

To select relevant features, *Attribute Clustering* concept [12], [23], [24] has proved high adequacy in CBR context since it allows to maintain the relation between features, as well as providing a flexibility in term of features substitution. Most of vocabulary maintenance policies are hard i.e they are not able to manage uncertainty within knowledge. Undoubtedly, there is a crucial need to handle imperfection within CBR systems since embedded knowledge refer to real-world experiences, which are generally uncertain and imprecise. Belief function theory or Evidence theory [25], [26] presents one of the most powerful tools for that matter. Some evidential clustering techniques such as *k-evclus* and *EVCLUS* [27], [28], have been used in [12] and [29] for features learning. However, other techniques [30] within this theory have been proposed and showed their effectiveness. Therefore, we aim, in this paper, to propose a new policy for vocabulary knowledge maintenance based on assessing the relation between features through correlation measurement and using the *Relational Evidential C-Means (RECM)* [30] as a machine learning technique. We also aim to show results offered by our policy and compare them to policies using other methods, as well as to the original non-maintained CBR.

The rest of the paper is organized as follows. The next Section defines the vocabulary container, overviews applicable notions for vocabulary maintenance, and present some related works. Section 3 presents tools and fundamental concepts of the belief function theory. During Section 4, we describe in details our proposal for maintaining vocabulary maintenance. Results will be discussed in Section 5. Finally, we end by concluding within Section 6.

## 2. Maintaining knowledge within CBR systems

As already mentioned, CBR systems contain different varieties of knowledge that are stored in four knowledge containers, which are Case Base, Similarity measures, Adaptation knowledge, and Vocabulary. These knowledge needs maintenance along the time on account of appearance of noisiness and redundancy as well because context change (see Figure 2).

Before moving and focusing on the literature of vocabulary knowledge and its maintenance to introduce our contribution, for the current work, let briefly present the other three knowledge containers within CBR systems.

**Case Base Maintenance.** Firstly, to maintain the CB knowledge container which defines the set of stored experiences, we find, in the literature, several policies that are based on different strategies to maintain cases such as:

- 1) The *selection based strategy* which contains CBM policies that revise CB's content through the selection of only representative cases that are able to cover the remaining set of cases problems. For instance, the *Condensed Nearest Neighbor (CNN)* [13] presents the baseline of data reduction methods.

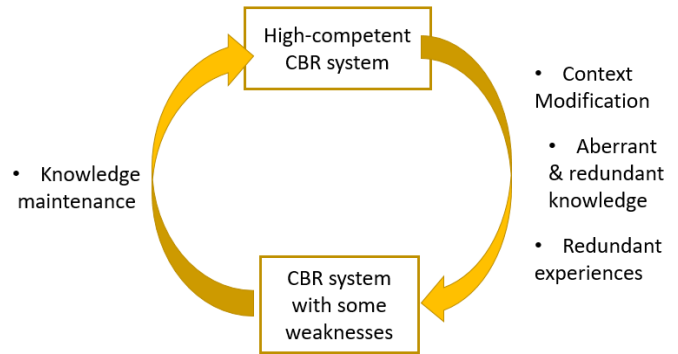


Figure 2. CBR knowledge maintenance

- 2) The *optimization based strategy* which contains policies that apply maintenance operations according to some evaluation criteria such that:
  - *The performance* which is quantified by the time spent to solve a target problem,
  - *The competence* which represents the range of problems that the CB can successfully solves [14]

Among that policies, *Iterative Case Filtering algorithm (ICF)* [15] is based on the competence criterion to make decision about cases deletion.

- 3) The *partitioning based strategy* which is characterized by giving the ability of treating the original CB in form of small ones, which makes it easier and more effective. Generally, the different subsets of cases are obtained using the clustering as an unsupervised machine learning technique. Actually, cases clustering is widely used within the CBM field due to its approved utility in detecting cases to be maintained. For instance, *Evidential Clustering and case Types Detection for case base maintenance method (ECTD)* [5] performs cases evidential clustering technique and, then, selects only cases that their deletion affects the whole CB quality. Some variants of ECTD policies have also been proposed [6], [8].

**Similarity Maintenance.** Secondly, Similarity measures are mainly used in CBR during the *Retrieve* step<sup>1</sup>. In the literature, there are several works focusing on the maintenance of similarity, and especially on learning features weighting in similarity measures [16], [17].

1. In the current research, we use the Euclidean Distance as a retrieve similarity measure inside the nearest neighbor algorithm (it is one of the most used algorithms in CBR systems).

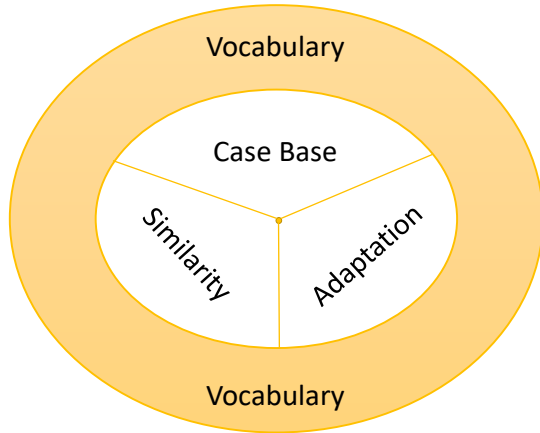


Figure 3. Illustrating the four knowledge containers within CBR systems

According to authors in [18], there are three main maintenance operations if we limit ourselves to weighted linear measures: modify a weight, modify a local measure, and stretch out a measure to a new attribute.

**Adaptation Maintenance.** Thirdly, within this adaptation maintenance area, research has to be so careful in adaptation changes. In fact, a low-quality maintenance may extremely influence the performance of the CB, and consequently the whole CBR system.

Handling adaptation rules is a very intensive knowledge. Consequently, we generally appeal domain-experts interventions for this matter. Nevertheless, there are different works in the context of adaptation learning and maintenance. For instance, an automated mechanism has been proposed in [19] to learn some adaptation rules from the different observations in the CB: If descriptions of two cases vary in just few number of features, the core of the adaptation rule is then formed from the differences in those features. Other works are also carried out in this area such that [20], [21]. Let us mention that, according to [22], there are four types of adaptations that can be applied inside CBR systems:

- 1) *Null Adaptation*<sup>2</sup>: It consists on returning the similar case solution without any modification,
- 2) *Transformational Adaptation*: It applies a modification in some parts of the solution description,
- 3) *Generative Adaptation*: It generates a solution from scratch.
- 4) *Compositional Adaptation*: It combines the three previous types to obtain the reused solution.

**Vocabulary Maintenance.** According to authors in containers [11], and as shown in Figure 3, the vocabulary is presented as the basis of all the other knowledge. During this Section, we define the vocabulary and the reason of maintenance, along with mentioning some related work and applicable concepts for maintenance.

2. This corresponds to the type of adaptation considered by the current research work.

## 2.1. Vocabulary container definition

To start, the vocabulary may be defined as the response to the question "Which element of the data structures are used to present fundamental notions?" [2]. In fact, the definition of this knowledge is highly depended on knowledge source nature. For object-oriented organization, for instance, attribute-value structure is generally used to define the vocabulary. However, if it consists on more complex types of data like text, sensor data or image, then the vocabulary may be defined differently using predicates, functions, set of words, or related constructs.

In the current work, we define the vocabulary of CBR systems by the set of features<sup>3</sup> describing cases.

## 2.2. The need of vocabulary knowledge maintenance

Obviously, there is a need to maintain all knowledge containers within CBR systems since they are designed to work for a long period of time. For instance, let assume that we possess a CBR system with high competence and performance. After a period of time, we may encounter noisy and redundant knowledge as well as some context changes. Hence, the system will know some weaknesses and degradation of competence and/or performance.

In this paper, we are focusing only on maintaining the set of features. In fact, every experience, in our life, can be described with unlimited number of features, where only some of them are considered as representative and able to direct the best solution. Generally, two kinds of features should be removed; noisy attributes which overwrite the smooth learning, and redundant ones which slow CBR system's operations.

## 2.3. Applied concepts for vocabulary maintenance

To maintain CBR vocabulary knowledge, different concepts within the machine learning field have been applied. Among the most used ones for this matter, we cite feature selection and attribute clustering.

Firstly, since we work with structured CBR systems and aim to retain only relevant attributes, the field of *Feature Selection* (FS) is suitable for such problem. FS is a NP-Hard problem that aims to select only relevant attribute, for some data, that do not contribute to the predictive model's accuracy. Hence, we note some FS methods that have been combined with CBR vocabulary maintenance context [31], [32], and others that select attributes through allocating weights according to features significance [33].

Secondly, the *Attribute Clustering* (AC) is a suitable way that could be used to maintain vocabulary within CBR context. In fact, AC allows to preserve relation between features which offers a high flexibility to CBR framework since each attribute could be substituted by another one belonging to the same cluster. Consequently, AC has been

3. *Feature* and *Attribute* terms are used exchangeably within this paper.

used in some works [34], [35] as a feature selection method for vocabulary maintenance [12].

The concept of AC is similar to objects clustering where similar objects should belong to the same cluster, and inversely. However, similarity between attributes is described in term of relation between them, such as correlation or dependency, which is generally depending on the task objective.

To manage uncertainty within knowledge, from the complete ignorance to the total certainty, we choose to use tools from the belief function theory [25], [26], as presented in the following Section.

### 3. Belief function theory

Since our work, elaborated in this paper, aims to manage knowledge imperfection which is naturally embedded within real-world experiences stored in CBR systems, some tools and techniques from the belief function theory are used, in this work, to allow a high-quality vocabulary maintenance. During this Section, the fundamental concepts of the belief function theory, which is named also Dempster-Shafer or Evidence theory [25], [26], are presented. Besides, the concept of credal partition within the evidential clustering is defined to model the doubt about features assignment to clusters.

#### 3.1. Fundamental Concepts

To model and quantify the evidence under the belief function framework, let consider  $\omega$  a variable that refers to  $C$  elementary events for some problem presented by  $\Omega = \{\omega_1, \omega_2, \dots, \omega_C\}$ . We call the  $\Omega$  set by the *frame of discernment* and the set of all the  $2^C$  possible subsets of the predefined events taking values in  $\Omega$  by the *power set*. The power set is therefore defined as follows:

$$2^\Omega = \{\emptyset, \{\omega_1\}, \dots, \{\omega_C\}, \{\omega_1, \omega_2\}, \dots, \{\omega_1, \omega_C\}, \dots, \Omega\} \quad (1)$$

The main key point of this theory is called the basic belief assignment (bba) which refers to the partial knowledge regarding the real value of  $\omega$ . It is defined by the following function:

$$\begin{aligned} m : 2^\Omega &\rightarrow [0, 1] \\ B &\mapsto m(B) \end{aligned} \quad (2)$$

where  $m$  satisfies the following constraint:

$$\sum_{B \subseteq \Omega} m(B) = 1 \quad (3)$$

We call an element  $B \subseteq \Omega$  as focal if  $m(B) > 0$ . The set of all the focal elements is called *Body of Evidence (BoE)*. If each element in BoE is a singleton, then  $m$  is called a *Bayesian bba*. Otherwise, if BoE contains only the frame of discernment  $\Omega$  as a focal element, then  $m$  is named a *vacuous belief function*. However, if it contains only one focal element of  $\Omega$  which is singleton, then  $m$  is a *Certain mass function*.

The basic belief mass  $m(B)$ , which presents the degree of belief assigned to the hypothesis " $\omega \in B$ ", can be attached to a subset of variables regardless any additive assumption. A bba corresponds to the open world assumption when we allow the assignment of evidence to the empty set partition. That means that  $\Omega$  is incomplete and the actual value may be taken outside the frame of discernment. This interpretation is meaningful especially in clustering when we aim to distinguish noisy knowledge [30], [37]. Contrariwise, if the bba is normalized ( $m(\emptyset) = 0$ ), then it corresponds to the closed-world assumption [26]. If we are interested to move on from unnormalized ( $m(\emptyset) \neq 0$ ) to normalized ( $m(\emptyset) = 0$ ) bba, then the *Dempster Normalization* can be applied, which consists on dispersing the belief's degrees assigned to the empty set over all the other focal sets such that:

$$m_*(B) = \begin{cases} \frac{m(B)}{1 - m(\emptyset)} & \text{if } B \neq \emptyset \\ 0 & \text{Otherwise} \end{cases} \quad (4)$$

Some useful functions are generally computed through the bba for some reason. For instance, for a given bba  $m$ , the corresponding belief (*bel*), plausibility (*pl*) and commonality (*q*) functions are from  $2^\Omega$  to  $[0, 1]$  and defined by [26] as follows:

$$bel(B) = \sum_{\emptyset \neq D \subseteq B} m(D) \quad \forall B \subseteq \Omega \quad (5)$$

$$pl(B) = \sum_{B \cap D \neq \emptyset} m(D) \quad \forall B \subseteq \Omega \quad (6)$$

and

$$q(B) = \sum_{B \subseteq D} m(D) \quad \forall B \subseteq \Omega \quad (7)$$

In other words, the belief function  $bel(B)$  represents the total belief that one commits to  $B$  without committed to  $\bar{B}$ , the plausibility function  $pl(B)$  quantifies the maximum amount of belief that could be given to a subset  $B$  and the commonality function  $q(B)$  represents the total mass that is free to move to every element of  $B$ .

The source of evidence providing bbas are often not fully reliable. Hence, a *discounting operation* [26] is necessary to update the bba according to the degree of trust assigned to the source. In fact, the discount rate, denoted  $\alpha \in [0, 1]$ , refers to the amount of belief that expert's information is reliable. The updated bba  ${}^\alpha m$  is defined as follows:

$${}^\alpha m(B) = \begin{cases} (1 - \alpha) m(B) & \text{for } B \neq \Omega \\ \alpha + (1 - \alpha) m(\Omega) & \text{for } B = \Omega \end{cases} \quad (8)$$

Let consider two mass functions  $m_1$  and  $m_2$  defined in the same frame of discernment  $\Omega$ . Hence, the degree of conflict between them may be computed with various ways. One of the most common methods is defined, in [26], as follows:  $\kappa = \sum_{A \cap B = \emptyset} m_1(A) m_2(B)$ . Authors, in [27],

demonstrated that if two mass functions  $m_1$  and  $m_2$  quantify evidence regarding two different questions that take values in the same possible answers  $\Omega$ , then the plausibility that both questions have the same answer is equal to  $1 - \kappa$ .

At the end, we generally need to make decision after handling beliefs' degrees. Hence, we present one of the most known tools for decision making called the pignistic probability transformation (*BetP*). It consists on choosing the hypothesis, regarding a normalized bba  $m$ , having the highest value. *BetP* is therefore defined as follows:

$$BetP(\omega) = \sum_{\omega \in B} \frac{m(B)}{|B|} \quad \forall \omega \in \Omega \quad (9)$$

To be able to apply the *BetP* transformation within the open world assumption ( $m(\emptyset) \neq 0$ ), a prefatory step of normalization (Equation 4) has to be performed.

### 3.2. Credal Partition

Handling imperfection using the evidence theory within the clustering problem has known a widespread interest in several work. The clustering is a machine learning technique that aims to organize data according to the similarity between instances. The more the objects are similar, the more they intend to share the same group.

The key point of the evidential clustering problem is known by the *Credal Partition*, which is close to the fuzzy partition concept but more general. The credal partition is created by assigning degrees of belief not only to singletons of the frame of discernment, but also to all possible subsets of  $\Omega$ . Within the evidential clustering, the frame of discernment refers to the set of  $C$  possible clusters. The uncertainty towards the membership of an object  $i$  to a partition of clusters  $B$  is presented by a mass function denoted  $m_i(B)$ . Its value supports the hypothesis saying "The actual cluster of instance  $i$  belongs to the partition  $B$ ". That bba quantifies the uncertainty regarding the membership of only one object. When there is  $n$  instances, the credal partition will be the set of  $n$ -tuple bbas  $(m_1, m_2, \dots, m_n)$ .

For a credal partition  $M = (m_1, m_2, \dots, m_n)$ , the two following particular cases are of interest [30]

- If every  $m_i$  is a certain bba, then  $M$  presents a crisp partition of the frame of discernment  $\Omega$ , which corresponds to a complete knowledge situation.
- If every  $m_i$  is a Bayesian bba, then  $M$  presents a fuzzy partition of the frame of discernment of  $\Omega$ .

## 4. Vocabulary maintaining process using Relational Evidential C-Means (RECM)

To remove noisy and redundant attributes from vocabulary knowledge, our proposed policy follows three principle steps as presented in Figure 4, and referred by three arrows. It is mainly based on the RECM machine learning technique that is able to handle relations between features and manage uncertainty within data.

### 4.1. Step 1: Generating relational matrix through studying the correlation between features

Let define the relation between features inversely proportional to the correlation between them. In fact, we aim to express this relation in term of dissimilarity, where the correlation may refer to the similarity between features. Actually, the more two given attributes are correlated, the more they offer the same information and considered as similar. To measure the linear association between two attributes  $A_i$  and  $A_j$ , we use the Pearson's correlation coefficient [36], which is defined such that:

$$r_{A_i A_j} = \frac{\sum_{l=1}^n (a_{il} - \bar{a}_i) (a_{jl} - \bar{a}_j)}{\sqrt{\sum_{l=1}^n (a_{il} - \bar{a}_i)^2} \sqrt{\sum_{l=1}^n (a_{jl} - \bar{a}_j)^2}} \quad (10)$$

where  $a_{il}$  and  $a_{jl}$  refer respectively to values of attributes  $A_i$  and  $A_j$  regarding case  $l$ , where  $\bar{a}_i$  and  $\bar{a}_j$  present their mean values.

Knowing that  $r_{A_i A_j}$  is bounded in  $[-1, 1]$ , three main situations arise to define the relational matrix.

- If  $r_{A_i A_j} \simeq 1$ , then there is a high correlation (positive)  $\Rightarrow$  similar provided information  $\Rightarrow$  high similarity.
- If  $r_{A_i A_j} \simeq -1$ , then there is a high correlation (negative)  $\Rightarrow$  similar provided information  $\Rightarrow$  high similarity.
- If  $r_{A_i A_j} \simeq 0$ , then there is no correlation  $\Rightarrow$  different provided information  $\Rightarrow$  high dissimilarity.

As a consequence, the Relational matrix  $R$  is defined as:

$$R = (1 - |r_{A_i A_j}|) \quad i, j = 1..p \quad (11)$$

where  $p$  presents the total number of features.

### 4.2. Step 2: Learning on attributes using Relational Evidential C-Means (RECM)

During this step, we aim to consider features as objects and cluster them according to their relational matrix previously generated in order to build clusters, where each one contains similar and correlated attributes. Since knowledge are never exact, we express clusters' membership through degrees of belief. To do, we use a powerful technique for this matter called the Relational Evidential C-Means (RECM) [30].

RECM [30] is a relational version of the basic Evidential C-Means (ECM) [37], where both are based on an alternate minimization scheme. However, ECM handles vectorial attribute data and RECM handle dissimilarity data.

Let start by presenting ECM and consider  $n$  objects described in  $p$  feature space,  $v_i$  is the center that represents a given cluster  $c_i$ , and  $\bar{v}_j$  is the barycenter that represents the partition of clusters  $C_j$  with  $C_j \subseteq \Omega$ . To derive the credal



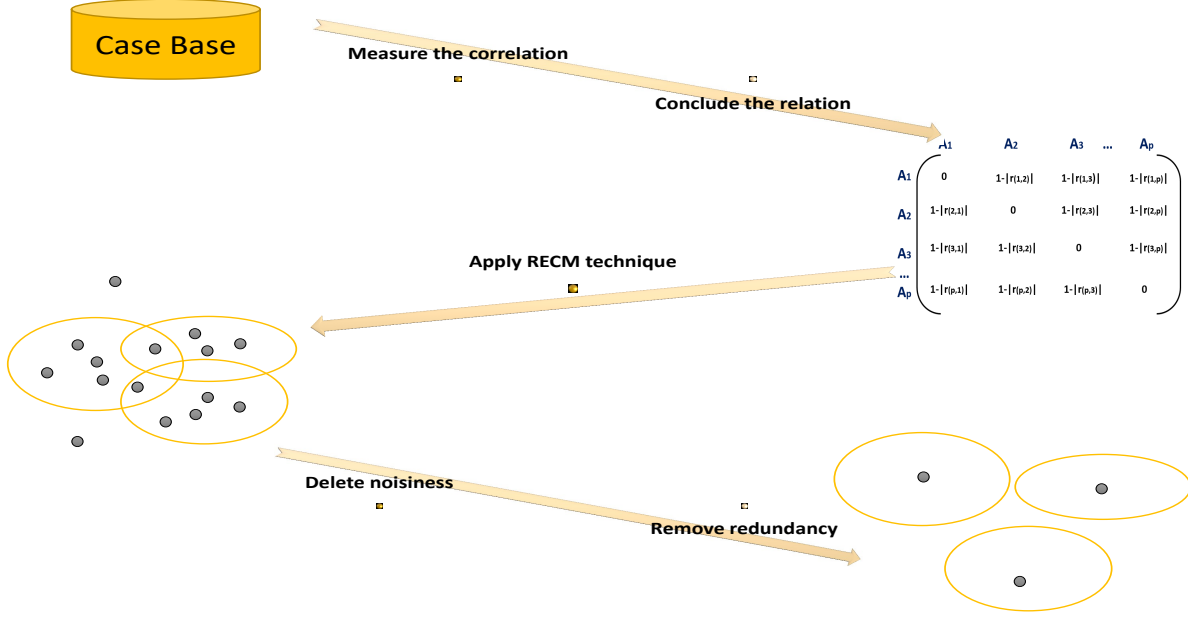


Figure 4. Steps of Evidential vocabulary maintenance based on RECM technique

partition  $M = (m_1, \dots, m_n)$  and the resulted clusters, which are presented by their centers  $V$ , ECM algorithm proceeds an alternate optimization scheme to optimize the following cost function:

$$J_{ECM}(M, V) = \sum_{i=1}^n \sum_{j/C_j \neq \emptyset, C_j \subseteq \Omega} |C_j|^\alpha m_{ij}^\beta d_{ij}^2 + \sum_{i=1}^n \delta^2 m_{i\emptyset}^\beta \quad (12)$$

subject to

$$\sum_{j/C_j \subseteq \Omega, C_j \neq \emptyset} m_{ij} + m_{i\emptyset} = 1 \quad \forall i = 1 \dots n \quad (13)$$

where  $m_{ij}$  denotes  $m_i(C_j)$  and  $\emptyset$  refers to noise cluster [30], [37] to be at a fixed distance  $\delta$  from every instance. The exponent  $\alpha$  aims at controlling the penalization degree for partitions having high cardinality, while  $\beta$  and  $\delta$  treat noisiness.

Firstly, as presented in [37],  $V$  is considered as fix and  $M$  is updated as follows:

$$m_{ij} = \frac{|C_j|^{-\alpha/(\beta-1)} d_{ij}^{-2/(\beta-1)}}{\sum_{C_k \neq \emptyset} |C_k|^{-\alpha/(\beta-1)} d_{ik}^{-2/(\beta-1)} + \delta^{-2/(\beta-1)}} \quad (14)$$

and

$$m_{i\emptyset} = 1 - \sum_{C_j \neq \emptyset} m_{ij} \quad \forall i = 1 \dots n \quad (15)$$

Secondly,  $M$  is considered fix, and we obtain an unconstrained optimization problem. The resulted  $V$  is defined as the solution of the following equation:

$$HV = UX \quad (16)$$

where  $X$  represents objects data, and  $H$  and  $U$  are defined, respectively, as follows:

$$H_{lk} = \sum_i \sum_{C_j \neq \emptyset, C_j \supseteq \{\omega_l, \omega_k\}} |A_j|^{\alpha-2} m_{ij}^\beta \quad k, l = 1 \dots C \quad (17)$$

and

$$U_{li} = \sum_{C_j \ni \omega_l} |A_j|^{\alpha-1} m_{ij}^\beta \quad l = 1 \dots C \quad i = 1 \dots n \quad (18)$$

with  $x_{iq}$  is the value of feature  $q$  regarding object  $o_i$ ,

For RECM, which is applied in our context, we handle dissimilarity data between features. Hence the notation  $n$  with ECM will refer here to the number of features  $p$ . Let, in this step, denote the already generated relational matrix  $R$  by  $\Delta = (\delta_{ii'})$  and consider the square matrix  $W = (w_{ii'})$  referring to dot products of features, which is defined as follows:

$$W = -\frac{1}{2} J \Delta J \quad (19)$$

where  $J = \frac{1}{n} e e^t - I$  with  $e = (1, \dots, 1)^t \in \mathbb{R}$  and  $I$  is  $n \times n$  identity matrix.

Last but not least, let consider the matrix  $Q = (q_{kk'})$  (respectively the matrix  $Z = (z_{kk'})$ ) the matrix of dot products of centers (respectively the matrix of doc products between centers and instances). As demonstrated in [30],  $Z$  and  $Q$  are obtained by solving systems of linear equations toward the following equations respectively:

$$HZ_{.i} = UW_{.i}, \quad (20)$$

and

$$HQ_i = UZ_i. \quad (21)$$

Ultimately, let present, in Algorithm 1, the different steps of RECM technique, as presented in [30], by making use of the previous defined Equations.

---

**Algorithm 1** RECM algorithm [30]

---

**Require:** - Dissimilarity data;  
- The number of clusters  $C$ ;  
- The parameters  $\alpha, \beta, \delta$ , and  $\epsilon$ ;  
**Ensure:** - The credal partition  $M$ ;  
- Clusters centers  $V$ ;

- 1: **BEGIN**
- 2: Randomly generate the initial credal partition  $M^{(0)}$ ;
- 3: Initialization:  $k \leftarrow 0$ ;
- 4: Calculate  $W$  using Equation 19;
- 5: **Repeat**
- 6:  $k \leftarrow k + 1$
- 7: Calculate  $H^{(k)}$  and  $U^{(k)}$  from  $M^{(k-1)}$  using Equations 17 and 18;
- 8: **For**  $i=1..n$
- 9: Calculate the  $i^{th}$  column of  $Z^{(k)}$  using Equation 20;
- 10: **End For**
- 11: **For**  $i=1..C$
- 12: Calculate the  $i^{th}$  column of  $Q^{(k)}$  using Equation 21;
- 13: **End For**
- 14: Calculate distances  $d_{ij}$  from  $R^{(k)}$  and  $Q^{(k)}$ ;
- 15: Update  $M^{(k)}$  using Equations 14 and 15;
- 16: **Until**  $\|M^{(k)} - M^{(k-1)}\| < \epsilon$
- 17: **END**

---

### 4.3. Step 3: Removing noisy and redundant attributes

To reach our objective, we aim, in this step, to keep only representative features by eliminating noisiness and redundancy. We call noisy features those that have a degree of belief to the empty set higher than all the other degrees. They, hence, should be removed since they may gravely reduce the competence of the CBR system to solve problems.

Eliminating redundant features consists on selecting only one representative attribute from every cluster and removing the others. We choose to select the nearest feature to the centre of cluster to which it belongs. The decision of attributes membership to clusters has been made using the pignistic probability transformation as defined in Equation 9.

### 4.4. Illustrative example

Let consider knowledge regarding vocabulary container is defined by the set attributes  $A_i$  with  $i$  goes from 1 to 4. Let assume now that the frame of discernment  $\Omega$  contains two clusters ( $\omega_1$  and  $\omega_2$ ) and the applied RECM algorithm offers a credal partition  $M = [m_1; m_2; m_3; m_4]$ , where its values are presented in Table 1.

TABLE 1. EXAMPLE OF CREDAL PARTITION VALUES

$M$	$\emptyset$	$\{\omega_1\}$	$\{\omega_2\}$	$\Omega$
$m_1$	0.05	0.75	0.15	0.05
$m_2$	0.65	0.1	0.1	0.15
$m_3$	0.1	0.05	0.8	0.05
$m_4$	0.2	0.1	0.5	0.2

TABLE 2. PIGNISTIC PROBABILITY TRANSFORMATION VALUES

	$\omega_1$	$\omega_2$
$BetP_1$	<b>0.8158</b>	<b>0.1842</b>
$BetP_2$	-0.5	-0.5
$BetP_3$	<b>0.0833</b>	<b>0.9167</b>
$BetP_4$	0.25	0.75

From values given in Table 1, we remark that  $m_2(\emptyset) > m_2(\{\omega_1\}) + m_2(\{\omega_2\}) + m_2(\{\omega_1, \omega_2\})$ . Hence, we flag the feature  $A_2$  as noisy according to specifications given by our policy in step 3. By this way, we refresh the vocabulary knowledge by eliminating  $A_2$ . Thereafter, we make a decision regarding the membership of attributes to clusters using Equation 9 and we obtain results as shown in Table 2.

We note that  $A_1$  belongs to  $\omega_1$ , and  $A_3$  and  $A_4$  belong to  $\omega_2$ . Ultimately, we only retain  $A_1$  as an attribute prototype of  $\omega_1$  and  $A_3$  as representative of  $\omega_2$  to build the new maintained vocabulary knowledge.

## 5. Experimentation and results

The main purpose of our experimentation part is to compare results with those offered when we use other evidential learning tools for vocabulary maintenance, as well as to show the effectiveness degree of our proposal that we call REVM for Evidential Vocabulary Maintenance based on RECM [30] technique.

### 5.1. Data and experimental settings

The different policies and tools, presented in this paper, have been implemented using the *R software*. The original and the maintained CBR systems are tested using six real-world datasets from UCI repository for machine learning<sup>4</sup>. The description of these datasets are presented in Table 3.

During the implementation of some policies, some parameters should be set. The number of clusters (or features  $p$ ) used by RECM, for every CB, is fixed similarly to those in [12] with the EVM policy. The initial credal partition is randomly set, and the parameter  $\alpha$  regarding the cardinality, within function  $J_{RECM}$ , is set to 1 which means that we do not penalize clusters' partitions with high cardinality.

4. <https://archive.ics.uci.edu>



TABLE 3. CASE BASES DESCRIPTION

	Case base	Ref	# instances	# attributes
1	Ionosphere	IO	351	34
2	Glass	GL	214	10
3	WDBC	BC	569	31
4	German	GR	1000	20
5	Heart	HR	270	13
6	Yeast	YS	1484	8

## 5.2. Testing strategy and evaluation criteria

During the evaluation, we make use of the most applied classification algorithm within the CBR context: k-Nearest Neighbor (k-NN)<sup>5</sup>.

Two evaluation criteria have been used, in our work, to assess our proposal efficiency. First, the accuracy criterion consists on measuring the competence of the system in term of the Percentage of Correct Classifications (PCC), and defined such as:

$$PCC(\%) = \frac{\# \text{ Correct classifications}}{\# \text{ Total classifications}} \times 100 \quad (22)$$

To obtain final estimation of accuracy, we make use of the 10-fold cross validation technique as shown in Figure 5.

Second, we use the retrieval time (RT) criterion which refers to the time spent to retrieve and classify problems by the CBR system.

## 5.3. Results and discussion

As shown in Tables 4 and 5, we compare results offered by our proposal that we call REVM, denoting *RECM based Evidential Vocabulary Maintenance policy*, to those given by the non-maintained CBR system (Original-CBR), by a feature selection method called ReliefF [33] (CBR-ReliefF), and by the EVM policy [29] which uses EVCLUS [27] [28] for learning.

In term of accuracy, we remark, from Table 4, that RECM method offers competitive results, especially comparing to the non maintained CBR system. In fact, we remark that it was able to improve their initial competence rates provided by Original-CBR with all the tested datasets. For instance, it improves the accuracy of "Yeast" dataset from 55.32 % (Original-CBR) to 98.98 % (RECM).

Comparing to the feature selection method called reliefF (ReliefF-CBR), we note, also, that RECM offers close results. Ultimately, we note that very competitive results are provided with both EVM and our current proposed method RECM. Actually, we note that they are slightly in favor with the Evidential Vocabulary Maintenance method (EVM)

5. We chose to set  $k$  to 5 in order to avoid the sensibility to noisiness.

TABLE 4. ACCURACY [PCC(%)]

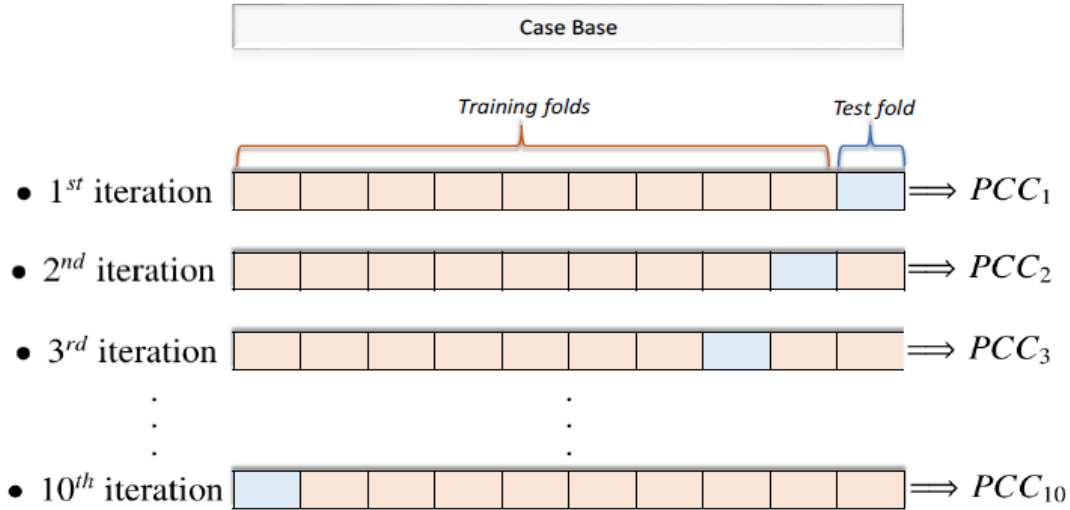
Case bases		Accuracy [PCC(%)]			
		Original-CBR	ReliefF-CBR	EVM	REVM
1	IO	85.48 %	84.88 %	88.33 %	86.98 %
2	GL	97.64 %	98.11 %	98.59 %	98.12 %
3	BC	60.16 %	96.33 %	96.46 %	96.18 %
4	GR	64.6 %	73.4 %	73.25 %	73.26 %
5	HR	57.5 %	62.45 %	62.98 %	60.91 %
6	YS	55.32 %	99.05 %	99.05 %	98.98 %

which uses the  $k$ -EVCLUS technique for learning. Although there is no high difference between these results (they offer almost same accuracy values for four datasets among six), we conclude that  $k$ -EVCLUS was more suitable. It may be explained by its high flexibility to handle different types of similarity or relational data, since our strategy considers them to be in term of correlations between features. However, we may tolerate this fact since our current proposal was able to retain or even largely improve the initial competence values.

TABLE 5. RETRIEVAL TIME [T(S)]

Case bases		Retrieval time [T(s)]			
		Original-CBR	ReliefF-CBR	EVM	REVM
1	IO	1.942	1.188	0.912	0.902
2	GL	0.967	0.882	0.762	0.604
3	BC	1.710	1.112	1.013	0.902
4	GR	1.812	1.213	1.211	1.222
5	HR	2.103	1.091	1.028	0.789
6	YS	0.954	0.722	0.724	0.776

In term of retrieval time, we note that the decreasing of the number of instances as well of features may conduct to decrease the research time, which is described in Table 5 in seconds. We remark that results offered after performing vocabulary maintenance are lower than those offered with the original non maintained CBR systems (Original-CBR). With "Ionosphere", "Breast Cancer", and "Heart" datasets, for instance, we note faster cases retrieval with values go, respectively, from 1.942s to 0.902s, from 1.710s to 0.902s, and from 2.103s to 0.789s. However, we also note very close results with the other ReliefF and EVM maintaining methods, which is logic since they offer data described with close number of features. In fact we would like to mention that, it could be affected by tasks executed in the



$$PCC = \frac{1}{10} \sum_{i=1}^{10} PCC_i$$

Figure 5. The principle of 10-fold cross validation

environment of development during experimentation.

## 6. Conclusion

In this paper, we used the Relational Evidential C-Means (RECM) as a machine learning technique at the aim to maintain vocabulary knowledge within CBR systems by keeping only representative features. The idea consists on studying, first of all, the relation between features through measuring the correlation between them and create the relational matrix. This matrix will then be used to learn on features with managing uncertainty using RECM. Ultimately, the generated credal partition has been analyzed and studied to retain only representative and relevant features to describe case knowledge. During the experimentation, we tested our proposal on six real datasets from different domains provided in UCI repository for machine learning. Offered results showed some improvement of problem-solving competence comparing to those offered by the non maintained CBR systems. Good results are also noted comparing to the feature selection method called ReliefF. However, we remarked very competitive results with the vocabulary maintenance policy, called Evidential Vocabulary Maintenance (ECM), that uses the EVCLUS technique for learning.

## References

[1] Aamodt, A., Plaza, E.: Case-based reasoning: Foundational issues, methodological variations, and system approaches. In *Artificial Intelligence Communications*, pp. 39-52 (1994).

[2] Richter, M. M., Michael, M.: Knowledge containers. *Readings in Case-Based Reasoning* (2003).

[3] Wilson, D. C., Leake, D. B.: Maintaining Case-Based Reasoners: Dimensions and Directions. *Computational Intelligence*, pp. 196-213 (2001).

[4] Smiti, A., Elouedi, Z.: SCBM: soft case base maintenance method based on competence model. *Journal of Computational Science*, 25, pp. 221-227 (2017).

[5] Ben Ayed, S., Elouedi, Z., Lefevre, E.: ECTD: evidential clustering and case types detection for case base maintenance. In *IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, pp. 1462-1469. IEEE (2017).

[6] Ben Ayed, S., Elouedi, Z., Lefevre, E.: DETD: Dynamic Policy for Case Base Maintenance Based on EK-NNclus Algorithm and Case Types Detection. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pp. 370-382 (2018).

[7] Smiti, A., Elouedi, Z.: WCOID-DG: An approach for case base maintenance based on Weighting, Clustering, Outliers, Internal Detection and Dbsan-Gmeans. *Journal of computer and system sciences*, 80(1), pp. 27-38 (2014).

[8] Ben Ayed, S., Elouedi, Z., Lefevre, E.: Exploiting Domain-Experts Knowledge Within an Evidential Process for Case Base Maintenance. In *International Conference on Belief Functions*, pp. 22-30 (2018).

[9] Ben Ayed, S., Elouedi, Z., Lefevre, E.: CEC-Model: A New Competence Model for CBR Systems Based on the Belief Function Theory. *International Conference on Case-Based Reasoning*, pp. 28-44 (2018).

[10] Smyth, B., McKenna, E.: Modelling the competence of case-bases. In: *Smyth, B., Cunningham, P. (eds.) EWCBR. LNCS, vol. 1488*, pp. 208-220 (1998).

[11] Roth-Berghofer, T. R. and Cassens, J.: Mapping goals and kinds of explanations to the knowledge containers of case-based reasoning systems. In *International Conference on Case-Based Reasoning*, pp. 451-464 (2005).

- [12] Ben Ayed, S., Elouedi, Z., Lefevre, E.: Maintaining case knowledge vocabulary using a new Evidential Attribute Clustering method. In *13th International FLINS Conference on Data Science and Knowledge Engineering for Sensing Decision Support*, pp. 347-354 (2018).
- [13] Hart, P.: The condensed nearest neighbor rule. *IEEE transactions on information theory*, 14(3), pp. 515-516 (1968).
- [14] Smyth, B., McKenna, E.: Competence models and the maintenance problem. *Computational Intelligence*, 17(2), pp. 235-249 (2001).
- [15] Brighton, H., Mellish, C.: On the consistency of information filters for lazy learning algorithms. In *Proceedings of european conference on principles of data mining and knowledge discovery*, pp. 283-288 (1999).
- [16] Stahl, A., Gabel, T.: Using evolution programs to learn local similarity measures. In *International Conference on Case-Based Reasoning*, pp. 537-551 (2003).
- [17] Stahl, A.: Learning similarity measures: A formal view based on a generalized CBR model. In *International Conference on Case-Based Reasoning*, pp. 507-521 (2005).
- [18] Richter, M. M., Weber, R. O.: Case-Based Reasoning. *Springer-Verlag Berlin Heidelberg* (2013).
- [19] Leake, D. B., Kinley, A., Wilson, D.: Acquiring case adaptation knowledge: A hybrid approach. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence, AAAI Press*, pp. 684-689 (1996).
- [20] Jarmulak, J., Craw, S., Rowe, R.: Using case-base data to learn adaptation knowledge for design. In *International joint conference on artificial intelligence, Vol. 17, No. 1*, pp. 1011-1020 (2001).
- [21] Craw, S., Jarmulak, J., Rowe, R.: Learning and applying case-based adaptation knowledge. In *International Conference on Case-Based Reasoning*, pp. 131-145 (2001).
- [22] Wilke, W., Bergmann, R.: Techniques and knowledge used for adaptation during case-based problem solving. In *Proceedings of the international conference on industrial, engineering and other applications of applied intelligent systems*, pp. 497-506 (1998).
- [23] Hong, T. P., Liou, Y. L.: Attribute clustering in high dimensional feature spaces. In *International Conference on Machine Learning and Cybernetics, Vol. 4*, pp. 2286-2289 (2007).
- [24] Maji, P.: Fuzzyrough supervised attribute clustering algorithm and classification of microarray data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 41(1), pp.222-233 (2011).
- [25] Dempster, A. P.: Upper and lower probabilities induced by a multi-valued mapping. In *The annals of mathematical statistics*, pp. 325-339 (1967).
- [26] Shafer, G.: A mathematical theory of evidence. *Princeton University Press, Princeton* (1976).
- [27] Deneux, T., Masson, M. H.: EVCLUS: Evidential clustering of proximity data. *IEEE Trans. on Systems, Man and Cybernetics B* 34 (1), pp. 95-109 (2004).
- [28] Deneux, T., Sriboonchitta, S., Kanjanatarakul, O.: Evidential clustering of large dissimilarity data. *Knowledge-based Systems* 106, pp. 179-195 (2016).
- [29] Ben Ayed, S., Elouedi, Z., Lefevre, E.: CEVM: Constrained Evidential Vocabulary Maintenance Policy for CBR Systems. *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pp. 579-592 (2019).
- [30] Masson, M. H., Deneux, T.: RECM: Relational evidential c-means algorithm. *Pattern Recognition Letters*, 30(11), pp. 1015-1026 (2009).
- [31] Arshadi, N., Jurisica, I.: Feature Selection for Improving Case-Based Classifiers on High-Dimensional Data Sets. In *FLAIRS Conference*, pp. 99-104 (2005).
- [32] Zhu, G., Hu, J., Qi, J., Ma, J., Peng, Y.: An integrated feature selection and cluster analysis techniques for case-based reasoning. In *Engineering Applications of Artificial Intelligence*, pp. 14-22 (2015).
- [33] Kononenko, I.: Estimating attributes: analysis and extensions of RELIEF. In *European conference on machine learning*, pp. 171-182 (1994).
- [34] Hong, T., Liou, Y.: Attribute clustering in high dimensional feature spaces. In *International Conference on Machine Learning and Cybernetics*, pp. 2286-2289 (2007).
- [35] Maji, P.: Fuzzy-rough supervised attribute clustering algorithm and classification of microarray data. In *Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, pp. 222-233 (2011).
- [36] Pearson, K.: Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. In *Philosophical Transactions of the Royal Society of London*, pp. 253-318 (1896).
- [37] Masson, M. H., Deneux, T.: ECM: an evidential version of the fuzzy c-means algorithm. *Pattern Recognition* 41 (4), pp. 1384-1397 (2008).