Evidential seed-based semi-supervised clustering

Violaine Antoine
Blaise Pascal University,
UMR 6158, LIMOS, F-63006,
Clermont-Ferrand, France
Email: violaine.antoine@univ-bpclermont.fr

Nicolas Labroche
Sorbonne Universities,
UPMC Paris 06, UMR 7606, LIP6,
F-75005, Paris, France
Email: nicolas.labroche@lip6.fr

Viet-Vu Vu
Electronics Faculty,
Thai Nguyen University of Technology,
Thai Nguyen, Viet Nam
Email: vuvietvu@tnut.edu.vn

***Abstract*—Evidential clustering algorithms produce credal partitions that enhance the concepts of hard, fuzzy or possibilistic partitions to represent all assignments ranging from complete ignorance to total certainty. This paper introduces the first semi-supervised extension of the evidential c-means clustering algorithm that can benefit from the introduction of a small set of labeled data (or seeds). Experiments conducted on real datasets show that the introduction of seeds can lead to a significant increase in clustering accuracy compared to a traditional evidential clustering algorithm as well as a decrease in the number of iterations to convergence.**

I. Introduction

Semi-supervised clustering algorithms rely on a small set of expert constraints to improve the quality of the output partition. This knowledge can be either provided as must-link (ML) or cannot-link (CL) pairwise constraints, or as a small set of labeled data [1]. Until now, almost every clustering methods have been improved to handle such expert supervision. One may cite k-means [2], [3], hierarchical [4], density-based [5], [6], spectral [7] and even stream [8] clustering algorithms. If we consider more specifically soft clustering approaches, an emphasis has been made on fuzzy semi-supervised algorithms [9], [10], [11], [12]. However, fuzzy clustering algorithms are known to produce sometimes poor results against noise and outliers. Thus, possibilistic methods [13], [14] and more recently evidential methods have been proposed [15], [16], [17]. These latter use the theoretical framework of belief functions, which enables to represents all kind of partial knowledge, in order to generate a new concept of partition, called *credal* partition. This partition enhances the existing concepts of hard, fuzzy and possibilistic partitions. Adding extra information in such unsupervised clustering algorithm enables thus to improve the accuracy as beneficing of the advantages of a model based on belief functions. Some recent works [18], [19] have already proposed a semi-supervised evidential clustering algorithms based on ML and CL constraints. However, in some applications, pairwise constraints are not available and only some pre-existing labeled data can be used as an expert knowledge.

This paper describes the first seed-based evidential clustering algorithm that benefits from the introduction of a small amount of labeled data to significantly improve the clustering accuracy. To achieve this goal, we propose a reformulation of the objective function that encourages the respect of the labels provided by the expert while building the credal clustering.

This paper is organized as follows: Section II first illustrates how seeds are used in semi-supervised clustering algorithms and second gives more details about belief function and evidential clustering algorithms. Then, Section III introduces our new seed-based evidential clustering algorithm. Section IV presents the results of experiments on real data sets from UCI Machine Learning Repository [20] and illustrates the benefit of introducing expert knowledge in the context of evidential clustering. Finally, Section V presents the conclusion and future works.

II. Background

A. Seed-based semi-supervised clustering algorithms

Traditional clustering methods such as k-means, fuzzy c-means (FCM) or DBSCAN are known to be simple and efficient [21]. However, they all have certain limitations due to their metric (spherical clusters for k-means), their random initialization (for k-means or fuzzy c-means) or the difficulty for an expert to directly set their internal parameters (for example the size of the neighborhood in DBSCAN). Moreover, even when these methods succeed in minimizing their objective function, still the obtained clustering can mismatch the primary objective of the analyst in real world applications.

One solution to all these problems consists in introducing some expert knowledge as pairwise Must-Link (ML) or Cannot-Link (CL) constraints or as seeds [2]. Pairwise constraints are known to be easier to produce for an expert and indicate that two points must or must not belong to the same cluster. Similarly, seeds correspond to data points for which the expert has provided a class label.

Two main approaches have been proposed so far to take into account these instance level constraints during the clustering process. First, in some methods, constraints are used to guide the clustering process. This can be done at the initialization step such as in [3], or during the iterations with either a strict enforcement of the constraints as in COP Kmeans [2] or a modification of the objective function to penalize solutions where constraints are not respected as in [22]. For instance, in [23], the authors propose an enhanced fuzzy c-means whose objective function includes a penalty term that considers the actual membership value of a point i to a cluster k, let say u_{ik}, and the membership as it should be according to the expert labels, let say $\tilde{u}_{ik}$, as shown in Equation 1.

$$J(\mathbf{U},\mathbf{V}) = \sum_{i=1}^{n}\sum_{k=1}^{c} u_{ik}^{\beta} d_{ik}^2 + \gamma \sum_{i=1}^{n}\sum_{k=1}^{c} (u_{ik} - \tilde{u}_{ik})^{\beta} d_{ik}^2 \quad (1)$$

where $\{\mathbf{x}_1, \dots \mathbf{x}_n\}$ is a set of vectors in $\mathbb{R}^p$ describing n objects to classify among c classes, $\mathbf{U}$, $\mathbf{V}$ respectively denote the membership matrix and the centers of the clusters. γ is a regulation parameter that allows to balance the relative importance of the label constraints in the objective function.

In some other algorithms such as SSDBSCAN [6], the seeds are used to identify the radius of dense regions while in spectral clustering, seeds can be used to reduce efficiently the similarity matrix and to speed-up the convergence [24], [7].

Second, in some other methods, the clustering process is not modified but the constraints are used to learn a metric that is compatible with the knowledge provided by the expert as in the MPC k-means algorithm [25]. For instance, in [26], the authors describe an adaptive metric extension of the algorithm introduced in [23], where the distance to each cluster is derived from a traditional Gustafson-Kessel distance [27].

In our work, we propose to modify the objective function of the evidential c-means clustering algorithm to take into account the constraints provided as labels with a regulation parameter similarly to [23].

B. Belief functions

The major strength of an evidential clustering algorithm is its capability to represent the doubt regarding the assignment of an object to a cluster. To that aim, evidential methods rely on the mathematical model referred to as the Dempster-Shafer theory of evidence or belief function theory [28], [29].

Let ω be a variable taking values in a finite set $\Omega = \{\omega_1, \dots, \omega_c\}$ called frame of discernment. Partial knowledge regarding the actual value of ω can be represented by a basic belief assignment (bba) m, which is an application from the power set of Ω in the interval $[0, 1]$, such that

$$\sum_{A \subseteq \Omega} m(A) = 1. \tag{2}$$

The subsets $A \subseteq \Omega$ such that $m(A) > 0$ are called the *focal sets* of m. The quantity $m(A)$ represents a fraction of a unit mass of belief allocated to A and which, for lack of information, cannot be allocated to any subset of A. Total ignorance corresponds to $m(\Omega) = 1$ whereas full certainty specifies a *certain* bba, i.e. the allocation of the whole mass of belief to a unique singleton of Ω. If all focal sets of m are singleton, then the mass function is equivalent to a probability function, and is called a *Bayesian* bba. The quantity $m(\emptyset)$ may be interpreted as the belief that the actual value of ω does not belong to Ω [30]. When $m(\emptyset) = 0$, the bba is said to be normalized.

Knowledge expressed by a bba can also be represented by the plausibility function $pl : 2^\Omega \to [0, 1]$ defined as following:

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \qquad \forall A \subseteq \Omega. \tag{3}$$

The quantity $pl(A)$ is interpreted as the maximal degree of belief which can be potentially assigned to the hypothesis that the actual value of ω belong to A.

When a decision has to be made regarding this actual value of ω, it can be interesting to transform a bba m into a probability distribution. This can be achieved using the pignistic transformation, which, for a normal bba, equally distributes each mass of belief among the elements of A [29]:

$$BetP(\omega) = \sum_{\omega \in A} \frac{m(A)}{|A|}, \qquad \forall \omega \in \Omega, \tag{4}$$

where $|A|$ denotes the cardinality of $A \subseteq \Omega$. When there exists $m(\emptyset) = 0$, a step of normalization must precede the pignistic transformation. The Dempster's normalization, which consists in dividing all the masses by $1 - m(\emptyset)$, is a classical method of normalization.

C. Evidential c-means algorithm

The belief function theory can be used in clustering to generate an informative partition, called *credal* partition. In this framework, partial knowledge regarding the class membership of an object i is represented by a bba m_i on the set Ω of possible classes. Thus, a degree of belief can be assigned not only to singletons, but also to any subsets of Ω. This representation enables to model a wide variety of situations, ranging from complete ignorance to total certainty.

The evidential version of k-means, called ECM, is a clustering algorithm that aims at deriving a credal partition from data. For each object $\mathbf{x}_i$, the bba m_i is computed by setting high (resp. low) quantity of belief on the subsets close (resp. far) in term of distance to $\mathbf{x}_i$. This distance d_{ij} is a metric function between an object $\mathbf{x}_i$ and a representation in $\mathbb{R}^p$ of a subset A_j. As the FCM algorithm, each class ω_k is modeled by a prototype $\mathbf{v}_k$. Then, for each subset $A_j \subseteq \Omega$, $A_j \neq \emptyset$, a centroid $\overline{\mathbf{v}}_j$ is calculated as the barycenter of the centers associated to the class composing A_j:

$$\overline{v}_j = \frac{1}{|A_j|} \sum_{k=1}^{c} s_{kj} \mathbf{v}_k \quad \text{ with } s_{kj} = \begin{cases} 1 & \text{if } \omega_k \in A_j, \\ 0 & \text{else.} \end{cases} \tag{5}$$

The distance d_{ij}^2 can now be defined by a Euclidean distance [16]. More recently, a variant has been proposed to take in account a Mahalanobis distance [18]. Similarly to Gutafson and Kessel [27], it enables to detect clusters with different geometrical shapes, thanks to a fuzzy covariance matrix $\mathbf{S}_k$ associated to each cluster ω_k. Then, similarly to the prototypes, for each non singleton subset A_j, a matrix $\overline{\mathbf{S}}_j$ is calculated by averaging the matrices included in A_j. The resulting distance between an object $\mathbf{x}_i$ and a center $\overline{\mathbf{v}}_j$ can be written as:

$$d_{ij}^2 = (\mathbf{x}_i - \overline{\mathbf{v}}_j)^T \mathbf{S}_j (\mathbf{x}_i - \overline{\mathbf{v}}_j). \tag{6}$$

The ECM algorithm searches to minimize the following objective function for the $\mathbf{M}$, $\mathbf{V}$ matrices and $\mathbf{S}$ matrices:

$$J_{ECM}(\mathbf{M}, \mathbf{V}, \mathbf{S}) = \sum_{i=1}^{n} \sum_{A_j \neq \emptyset} |A_j|^\alpha m_{ij}^\beta d_{ij}^2 + \sum_{i=1}^{n} \delta^2 m_{i\emptyset}^\beta, \tag{7}$$

subject to:

$$\sum_{j / A_j \subseteq \Omega, A_j \neq \emptyset} m_{ij} + m_{i\emptyset} = 1 \quad \forall i = 1 \dots n, \tag{8}$$

and

$$m_{ij} \geq 0, \quad \forall i = 1 \dots n, \quad \forall A_j \subseteq \Omega, \tag{9}$$

where $m_{ij} = m_i(A_j)$ and $m_{i\emptyset} = m_i(\emptyset)$. Since $m_{i\emptyset}$ corresponds to the belief that $\mathbf{x}_i$ is an outlier, it is separate from the rest of the subsets. The δ parameter denotes the distance of all objects to the empty set. Let us remark that a penalization of the subsets $A_j \in \Omega$ with high cardinality has been introduced with the weighting coefficient $|A_j|^\alpha$. The exponent α controls the degree of penalization.

As FCM, the evidential partition is carried out through an iterative optimization with the alternate update of **M**, **V** and **S**. Note that the complexity of an evidential clustering algorithm is linear with the respect to the number of samples and is exponential with the respect to the number of classes. Consequently, such approach remains limited to few hundreds of samples and a small number of classes.

III. SEED-BASED EVIDENTIAL C-MEANS (SECM)

A. Problem formalization

The main idea of the proposed algorithm is to add a term of penalty in the objective function of ECM, in order to take into account a set of labeled objects. Indeed, we consider that an expert can give for one object one or several classes, depending on his confidence about the final label of this object. Thus, an object with more than one label represents a doubt for its actual class, and not a multi-class assignment.

Let us now take an example to explain the expression of the labeled objects in the framework of belief functions. If $\mathbf{x}_i$ is an object associated to the class ω_k, then the *certain* bba $m_i(\omega_k) = 1$ fully respect the constraint. Inversely, a bba on an other subset of Ω express an uncorrect consideration of the constraint and must be penalized. However, this penalization must be gradual. Indeed, mass of belief on subsets containing (resp. not containing) the class ω_k partially respect (resp. totally unrespect) the constraint.

In order to detect mass functions which partially or fully respect a constraint on a label ω_k, one can compute its plausibility $pl(\omega_k)$. A 1 value will indicate a good consideration of the label and a 0 value a non-respect of the constraint.

The same whole reasoning can be made for an object constrained in several classes, and lead us to propose the following penalty term:

$$J_S = \sum_{i=1}^{n} \sum_{A_j \in \Omega, A_j \neq \emptyset} b_{ij}(1 - pl_i(A_j)), \tag{10}$$

with

$$b_{ij} = \begin{cases} 1 & \text{if } \mathbf{x}_i \in \omega_k \text{ and } \omega_k \in A_j. \\ 0 & \text{else.} \end{cases} \tag{11}$$

Note that $pl_i(A_j)$ represents the plausibility that the object $\mathbf{x}_i$ belongs to the subset A_j.

We propose then to minimize the following objective function:

$$J_{SECM} = (1-\gamma)\frac{1}{2^c n} J_{ECM} + \gamma \frac{1}{s} J_S, \tag{12}$$

such that the constraints (8) and (9) must be respected. The parameter s is the number of constraints. Consequently, when distances d_{ij} are normalized, the coefficients $\frac{1}{2^c n}$ and $\frac{1}{s}$ scale the two terms of J_{SECM} between 0 and 1. Finally, $\gamma \in [0, 1]$ is a coefficient controlling the tradeoff between the objective function of ECM and the constraints. The weighting coefficient $|A_j|^\alpha$ in J_{ECM} enables us to gradually penalize subsets $A_j \in \Omega$ which perfectly minimize J_S but have high cardinalities.

B. Main algorithm

The objective function J_{SECM} can be optimized as ECM, that is to say with an iterative refinement of the mass functions **M**, the prototypes **V** and the covariance matrices **S**. Since the penalty term J_S does not depend on **V** and **S**, their updates are equivalent to ECM, and the formulas can be found in [16]. The mass functions, by contrast, are present in J_S. If we set $\beta = 2$, the objective function becomes quadratic with the respect to the m_{ij}. Since there exists linear constraints, a classical optimization method [31] can be employed and the convergence is ensured in a reasonable time.

It is well-known that k-means and its variants may converge to a local optimum, depending on the initialization given to the prototypes. The resulting partition may then be counter-intuitive. The ECM and SECM algorithms deal with the same problematic. A classical solution is to run the clustering algorithm several times with different initializations of the prototypes, and to keep the solution giving the smallest objective function value.

The SECM algorithm handles a second type of local minimum problem: the first term of its objective function generates cluster labels on its own, mostly depending on the initialization of the prototypes. However, some class labels may have already been given as prior information and may not be coherent with the cluster labels found. Consequently, the objective function may reach a local minimum. In order to avoid such case, a permutation scheme has been introduced after the update of the mass functions: for all possible permutations of the labels on constraint objects, J_{SECM} is evaluated. Finally a permutation is validated if its value of J_{SECM} is the smallest one.

IV. EXPERIMENTS

Two main experiments are reported in this paper. First, a comparative study has been conducted with the unconstrained ECM evidential c-means clustering algorithm [16]. The objective of this experiment is to show to which extent the introduction of seeds impacts the clustering accuracy, the number of iterations before convergence or the running time of the algorithm. Second, a parameter study has been conducted to observe the influence of the γ parameter that balances the evidential clustering quality with the penalty term that forces the respect of the constraints.

A. Experimental protocol and evaluation

a) Datasets: for the ease of reproducibility of the experiments, several classical benchmark datasets from UCI Machine Learning Repository [20] have been used in our tests. Table I indicates for each dataset its number of objects, its number of attributes and its number of clusters as well as the

TABLE I. DESCRIPTION OF THE DATASETS FROM UCI MLR [20]

Name	# objects	# attributes	# clusters	Metric
Iris	150	4	3	Mahalanobis
Wine	178	13	3	Euclidean
LettersIJL	227	16	2	Mahalanobis
Ionosphere	351	34	2	Mahalanobis

metric that was used for clustering following [18]. Note that LettersIJL is the Letters data set transformed as [25].

b) Protocol: as c-means like algorithms are known to be heavily dependent on their initial partition, for each dataset and each parameter γ value, 25 runs have been conducted and for each of these runs, 5 random initializations have been performed. Then, for each run only the initialization that led to the minimum value of the objective function (i.e. the best value) is kept.

c) Parameters setting: for each experience and similarly to [16], we set $\alpha = 1$, $\beta = 2$ and $\delta^2 = 1000$. In order to give equivalent importance to the global structure of the clusters and to the constraints, the parameter γ is set by default to 0.5.

d) Evaluation: the evaluation of the clustering accuracy is performed with the Adjusted Rand Index (ARI) [32] which is a corrected version of the Rand index that takes into account the expected Rand index that a random clustering would score as follows:

$$ARI(X,Y) = \frac{\text{Rand Index} - \text{Expected Index}}{\text{Max Index} - \text{Expected Index}} \tag{13}$$

The ARI is measured for each dataset and each run both on all the dataset and on only the unconstrained data to evaluate the real impact of the introduction of labeled data on non labeled data. We also measure the number of iterations before convergence and the overall computing time in seconds. In this case, tests are conducted in the same experimental conditions.

B. Comparative results

1) Clustering accuracy: Figure 1 shows comparative results obtained in terms of Adjusted Rand Index (ARI) on our benchmark datasets for a ratio between 0 and 50% of labeled data. It is important to notice that, in the case where the percentage of constraints is set to 0, we use the original ECM algorithm as described in [16] and not our algorithm set with no constraint.

It can be seen from Figure 1 that, for all datasets, an increase in the number of labeled data improves the overall ARI scores. However, when looking into more details, two distinct behaviors can be observed.

On the one hand, for Iris and Letters datasets, constraints help the algorithm to reach the expected solution. Indeed, when we increase the number of seeds provided to the algorithm, this also improves the ARI for unconstrained data. This increase in the unconstrained ARI scores is directly related to the quality of the learning that our algorithm performs based on the seeds that are provided.

On the other hand, for Wine and Ionosphere datasets, although the global ARI score always increases, the ARI score

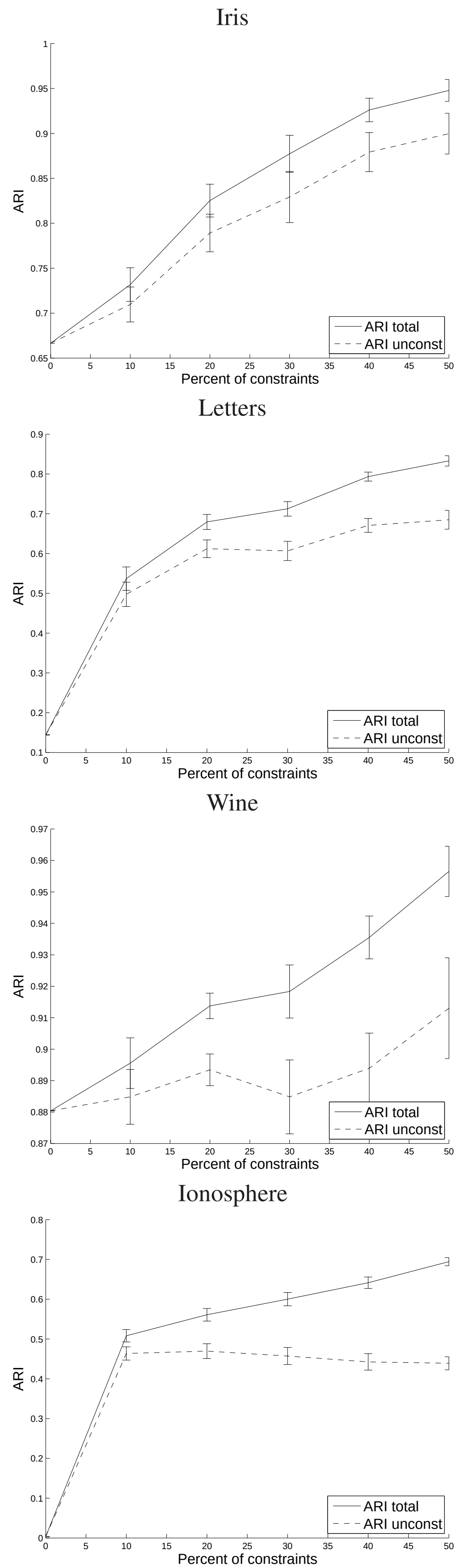

Fig. 1. Comparison of the ARI scores on benchmark datasets with parameter $\gamma = 0.5$ for several percentages of constraints ranging from 0% to 50%.

for unconstrained data is either stable (from 10% of constraints for Ionosphere) or eventually decrease punctually (between 20% and 30% constraints for Wine). In this case, the global structure of the class has already been found with no constraints and objects misclassified are located in the boundary of two or more clusters. Thus, constrained objects does not lead unconstrained object towards a better solution and moreover can be locally counter-productive. Further investigation should be conducted, but it is a well-known result in semi-supervised clustering that not all constraints are beneficial [33], [34].

2) Iterations and computation times: Figure 2 presents the average number of iterations and the average computation times (in seconds) for each dataset when parameter $\gamma = 0.5$.

It is interesting to notice as a general rule that the more seeds are provided to our algorithm, the less iterations are needed to converge. Only Iris and Ionosphere datasets show a slight increase in the number of iterations between the ECM algorithm with no constraint and our algorithm with 10% of constraints. In the case of Iris and according to previous results on clustering accuracy, this means that even if clustering accuracy is improved, more constraints are needed to reach the expected partition. In the case of Ionosphere, it seems that the constraints provided add some complexity to the algorithm but, in turn, help to improve significantly the accuracy compared to the ECM algorithm.

Finally, it is also interesting to notice that for all datasets, the computation time increases between the ECM algorithm and our constrained evidential algorithm. There are three reasons: first, as explained before, for some datasets there is an increase in the number of iterations. Second, for all datasets, our semi-supervised algorithm relies on a quadratic solver for the optimization of the mass functions while the ECM algorithm simply use Lagrange multipliers and find a direct formula for the update of $\mathbf{M}$. Finally, SECM is composed of an extra step consisting on the search of labels permutation for constrained objects.

3) Parameter setting: Experiments on all datasets have been conducted with values of parameter γ ranging from 0.2 to 0.6, that is to say from a configuration where the convergence is mainly guided by the ECM objective function to a configuration where most of the weight in the decision is given to the constraint penalty term. On our benchmark datasets, two behaviors can be observed. On the one hand, for Iris, Wine and Letters the evolution of γ has little or no influence on the clustering accuracy. On the other hand, for Ionosphere, giving more weight to the constraints force the constrained points to specific labels, even if it does not match with the global structure found by the first term of the objective function. As illustrated in Figure 3, it induces two particular results: (1) the γ coefficient should be lower so that the algorithm gives more importance to constraints leading to a structure that is compatible with the metric used and (2) accuracy on unconstrained object slightly decrease, which confirms that selecting smartly sets of constraints may lead to better results.

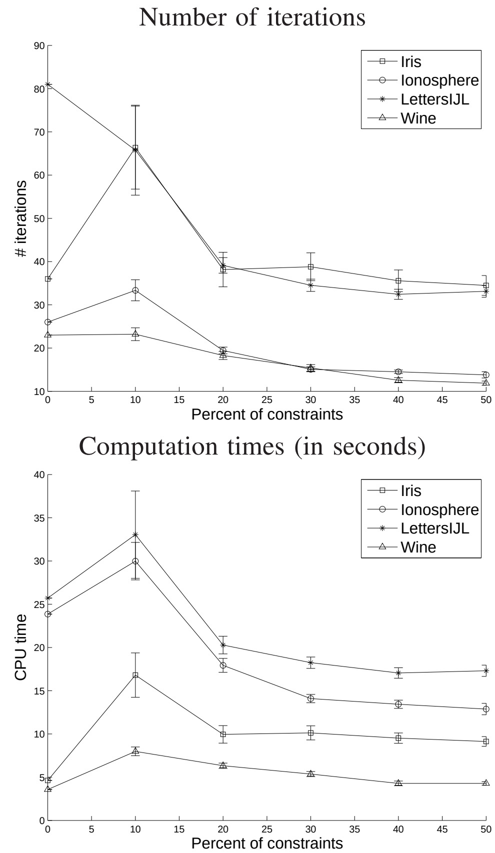

Fig. 2. Comparison of the average number of iterations (top) and computation times in seconds (bottom) on benchmark datasets with parameter $\gamma = 0.5$ for several percentages of constraints ranging from 0% to 50%.

V. Conclusion

This paper introduces a novel semi-supervised clustering algorithm that extends the evidential c-means with the introduction of expert knowledge as seeds. To that aim a penalty term based on the plausibility of the labeled data is proposed. One known difficulty in seed-based approaches is to link labels provided by the expert to those discovered by the algorithm automatically. To this aim we use a simple cluster label permutation procedure. Extensive experiments on real datasets from UCI Machine Learning repository show that the introduction of seeds in evidential clustering is possible and improves the clustering accuracy. Our tests also reveals that it is possible to reduce the number of iterations by adding more expert knowledge. Our results also highlight that our computation times are penalized by the exhaustive permutation procedure as expected, but that increasing the number of constraints can counter-balance this problem. Finally, our work illustrate the need for a utility function to determine sets of interesting constraints. Indeed, for some datasets, our experiments show that the clustering accuracy becomes stable or can decrease when the number of seeds increases. Future work includes: (1) extensive comparisons with other seed-based soft clustering algorithm such as those presented in [26] and experiences with objects labeled in more than one class, (2) the development of specialized utility function that use the information of the credal clustering to ask better queries to the expert and, (3) finally the search for a time-efficient clusters permutation procedure.

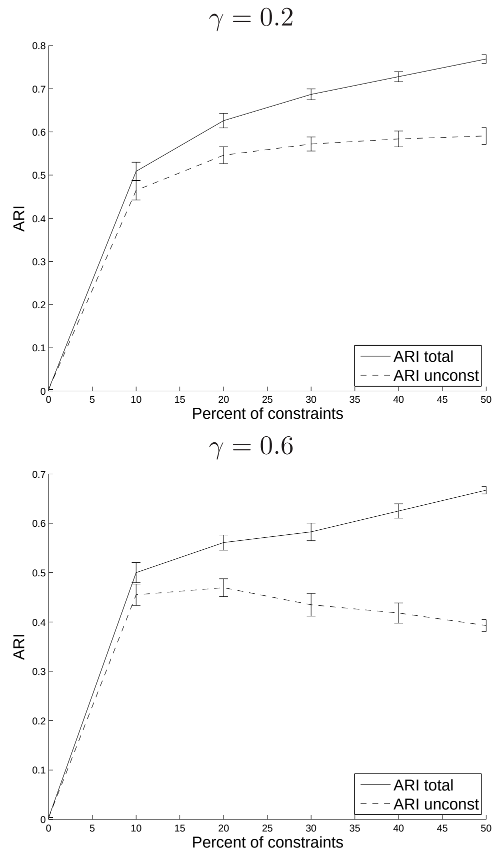

Fig. 3. Slight different impact of the γ parameter in terms of clustering accuracy (ARI) for the dataset Ionosphere for several percentages of constraints ranging from 0% to 50%. For the other datasets, no or little impact can be noticed in terms of average ARI.

REFERENCES

[1] K. L. Wagstaff and C. Cardie, "Clustering with instance-level constraints," in *In Proceedings of the 17th International Conference on Machine Learning, ICML*, 2000, pp. 1103–1110.

[2] K. Wagstaff, C. Cardie, S. Rogers, and S. Schroedl, "Constrained k-means clustering with background knowledge," in *In Proceedings of the 18th International Conference on Machine Learning, ICML-2001*. Morgan Kaufmann, 2001, pp. 577–584.

[3] S. Basu, A. Banerjee, and R. J. Mooney., "Semi-supervised clustering by seeding," in *In Proceeding of the 19th International Conference on Machine Learning (ICML)*, 2002, pp. 27–34.

[4] I. Davidson and S. Ravi, "Agglomerative hierarchical clustering with constraints: Theoretical and empirical results," in *Proceeding of European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, ECML PKDD-2005*, 2005, pp. 59–70.

[5] C. Ruiz, M. Spiliopoulou, and E. Menasalvas, "Density-based semi-supervised clustering," *Data Mining and Knowledge Discovery*, vol. 21, no. 3, pp. 345–370, 2010.

[6] L. Lelis and J. Sander, "Semi-supervised density-based clustering," in *Proceedings of the 2009 Ninth IEEE International Conference on Data Mining*, ser. ICDM '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 842–847.

[7] X. Wang and I. Davidson, "Flexible constrained spectral clustering," in *Proceeding of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD-2010*, 2010, pp. 563–572.

[8] C. Ruiz, E. Menasalvas, and M. Spiliopoulou, "C-denstream: Using domain knowledge on a data stream," in *DS 2009, LNAI 5808*, S.-V. B. Heidelberg, Ed., 2009, pp. 287–301.

[9] N. Grira, M. Crucianu, and N. Boujemaa, "Fuzzy clustering with pairwise constraints for knowledge-driven image categorization," *IEEE Vision, Image and Processing*, vol. 153, no. 3, pp. 299–304, 2006.

[10] W. Pedrycz, "Algorithm of fuzzy clustering with partial supervision," *Pattern Recognition Letters*, vol. 3, pp. 13–20, 1985.

[11] A. M. Bensaid, L. O. Hall, J. C. Bezdek, and L. P. Clarke, "Partially supervised clustering for image segmentation," *Pattern Recognition*, vol. 29, no. 5, 1996.

[12] W. Pedrycz and J. Waletzky, "Fuzzy clustering with partial supervision," *IEEE Transactions on systems, Man, and Cybernetics*, vol. 27, no. 5, pp. 787–795, 1997.

[13] R. Krishnapuram and J. Keller, "A possibilistic approach to clustering," *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 2, pp. 98–110, 1993.

[14] S. Sen and R. Davé, "Clustering of relational data containing noise and outliers," in *Fuzzy Systems Proceedings*, vol. 2, 1998, pp. 98–110.

[15] M.-H. Masson and T. Denœux, "Clustering interval-valued data using belief functions," *Pattern Recognition Letters*, vol. 25, no. 2, pp. 163–171, 2004.

[16] ——, "ECM: An evidential version of the fuzzy c-means algorithm," *Pattern Recognition*, vol. 41, pp. 1384–1397, 2008.

[17] ——, "RECM: Relational evidential c-means algorithm," *Pattern Recognition Letters*, vol. 30, pp. 1015–1026, 2009.

[18] V. Antoine, B. Quost, M.-H. Masson, and T. Denœux, "Cecm: Constrained evidential -means algorithm," *Computational Statistics & Data Analysis*, vol. 56, no. 4, pp. 894 – 914, 2012.

[19] ——, "Evidential clustering with instance-level constraints for proximity data," *Soft Computing*, vol. 56, pp. 894–914, 2013.

[20] A. Asuncion and D. Newman, "UCI machine learning repository – University of California, Irvine, School of Information and Computer Sciences," http://www.ics.uci.edu/$\sim$mlearn/MLRepository.html, 2007.

[21] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.

[22] W. Pedrycz and J. Waletzky, "Fuzzy clustering with partial supervision," *IEEE Transactions on systems, Man, and Cybernetics*, vol. 27, no. 5, pp. 787–795, 1997.

[23] A. Bouchachia and W. Pedrycz, "A semi-supervised clustering algorithm for data exploration," in *Proc. Internat. Fuzzy Systems Association World Congress IFSA'2003*, 2003, pp. 328–337.

[24] D. Mavroeidis, "Accelerating spectral clustering with partial supervision," *Data Mining and Knowledge Discovery*, vol. 21, pp. 241–258, 2010.

[25] M. Bilenko, S. Basu, and R. J. Mooney, "Integrating constraints and metric learning in semi-supervised clustering," in *Intl. Conference on Machine Learning, ICML 2004*, 2004, pp. 81–88.

[26] A. Bouchachia and W. Pedrycz, "Enhancement of fuzzy clustering by mechanisms of partial supervision," *Fuzzy Sets and Systems*, vol. 157, pp. 1733–1759, 2006.

[27] D. Gustafson and W. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," in *Proc. IEEE Conf. Decision and Control*, 1979, pp. 61–766.

[28] G. Shafer, *A Mathematical Theory of Evidence*. Princeton Univ. Press. Princeton, NJ, 1976.

[29] P. Smets and R. Kennes, "The transferable belief model," *Artificial Intelligence*, vol. 66, pp. 191–234, 1994.

[30] P. Smets, "The transferable belief model for quantified belief representation," in *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, vol. 1, 1998, pp. 267–301.

[31] Y. Ye and E. Tse, "An extension of Karmarkar's projective algorithm for convex quadratic programming," *Mathematical Programming*, vol. 44, no. 1, pp. 157–179, 1989.

[32] L. Hubert and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, no. 1, pp. 193–218, 1985.

[33] K. Wagstaff, "When is constrained clustering beneficial, and why?" in *Proceedings of the Twenty-first National Conference on Artificial Intelligence (AAAI)*, July 2006.

[34] K. L. Wagstaff, "Value, cost, and sharing: Open issues in constrained clustering," in *Proceeding of the 5th International Workshop on Knowledge Discovery in Inductive Databases, KDID-2007*, 2007, pp. 1–10.