Possibilistic clustering with seeds

Violaine Antoine*, Jose A. Guerrero†‡, Tanya Boone§, Gerardo Romero§
*Clermont Auvergne University, UMR 6158, LIMOS, F-63000,
Clermont-Ferrand, France
Email: violaine.antoine@uca.fr
†Irstea, 9 avenue Blaise Pascale, 63170 Aubire, France
‡Litis, EA 4108, INSA de Rouen,
76800 Saint Etienne du Rouvray, France
Email: jguerrer@ieee.org
§Electronic Department at U.A.M. Reynosa-Rodhe, UAT, México
Email:gromero@docentes.uat.edu.mx

***Abstract*—Clustering methods assign objects to clusters using only as prior information the characteristics of the objects. However, clustering algorithms performance can be improved when background knowledge is available. Such background knowledge can be incorporated in a clustering method as label constraints which results in a semi-supervised clustering algorithm. We propose to extend two possibilistic clustering algorithms to make use of available a priori information. The goal is twofold: to improve the accuracy of the clustering result by leading the method towards a desired solution and to detect outliers by taking advantage of the generated possibilistic partition. The proposed methods are called semi-supervised repulsive possibilistic c-means (SRPCM) and semi-supervised possibilistic fuzzy c-means (SPFCM). They correspond to possibilistic clustering algorithms that introduce label constraints. Experimental results show that the proposed algorithms using label constraints improve (1) the clustering result and (2) the outliers detection.**

I. Introduction

Clustering methods are part of exploratory data analysis techniques that aim to group unlabeled objects into clusters thanks to a similarity notion. Different approaches have been proposed in the literature such as hierarchical clustering and partition based clustering [1]. On one hand, hierarchical methods generate dendrograms based on a proximity matrix. On the other hand, partition based methods organize data into groups or clusters using either crisp (hard) partitions or soft partitions. Crisp methods divide data into groups by assigning each object of the dataset to a single cluster with total certainty. In the domain of partition based clustering, the most popular algorithms generating crisp partitions are k-means and density based algorithms such as DBSCAN. The k-means algorithm, which is an optimization based method, corresponds to a classical data analysis tool used in many topics [2], [3], [4]. On the contrary, soft clustering methods allow to express a degree of uncertainty for the membership of each object to each cluster. Fuzzy clustering methods, based on the fuzzy sets theory [5], are the most commonly used soft methods. On the contrary to crisp methods, soft clustering methods allow to express a degree of uncertainty for the membership of each object to each cluster. Fuzzy clustering methods, based on the fuzzy sets theory [5], are the most commonly used soft methods.

Fuzzy clustering algorithms have various application scenarios which range from biological sciences [6], [7], document and text processing [8], [9], to image processing [10], [11]. Amongst fuzzy clustering methods, Fuzzy C-Means (FCM) is a well known variant of k-means which assigns for each object a probability value to belong to each cluster. Its main drawback is a poor performance on noisy data. To overcome this weakness, [12] proposed a possibilistic clustering algorithm called Possibilistic C-Means (PCM) which is based on FCM with a relaxed membership degree constraint. However, as discussed in [13], PCM highly depends on initial conditions to produce good results. Indeed, based on initial conditions, PCM often generates coincident clusters. This problem has been addressed in [13], [14], [15]. In [14], the authors propose to add a repulsion based penalty term to the PCM objective function to avoid the coincident cluster problem resulting in an algorithm named Repulsive PCM (RPCM). In [15], the authors propose a new algorithm which is a linear combination of FCM and PCM named PFCM.

Simultaneously, it has been shown that the performance of hard and soft clustering algorithms can be improved by using a limited amount of background knowledge expressed as contraints [16], [17], [18]. These methods, called semi-supervised clustering methods, can be divided following the type of constraints employed: pairwise contraints [17], [19], label constraints [16], [18], [20], or others [21], [22].

In this paper, we are interested in the semi-supervised possibilistic clustering problem since the possibilistic framework enables to handle noisy data. Indeed, as explained in [15], a probabilistic partition forces an outlier to belong to one or more clusters with a high membership degree. Thus, this make impossible to mark it as an outlier. In contrast, a possibilistic partition allows for an object to be assigned with low membership degree to every clusters. When it happens, the object can be interpreted as an outlier. We propose to extend the RPCM and PFCM algorithms by incorporating a penalty term in their initial objective functions such that label based constraints are taken in account.

This work is organized as follows: in section II, possibilistic clustering preliminaries are introduced. Section III presents

a semi-supervised possibilistic clustering algorithm based on the repulsive PCM (RPCM) [14]. Section IV presents a semi-supervised possibilistic clustering algorithm based on the PFCM method [15]. In section V, the experimental results are discussed. Finally, the conclusions and future work are presented in section VI.

II. Possibilistic Clustering Preliminaries

Let us consider n objects represented by a set of feature vectors $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ in $\mathbb{R}^p$ and c centroids defined by a matrix $\mathbf{V} = (\mathbf{v}_k)$ such that each centroid $\mathbf{v}_k \in \mathbb{R}^p$ corresponds to a cluster. The PCM algorithm [12] iteratively updates the centroids $\mathbf{V}$ and a possibilistic partition $\mathbf{T} = (t_{ik})$ in order to minimize the J_{PCM} objective function:

$$J_{PCM}(\mathbf{T}, \mathbf{V}) = \sum_{i=1}^{n}\sum_{k=1}^{c} t_{ik}^m d_{ik}^2 + \sum_{k=1}^{c}\gamma_k \sum_{i=1}^{n}(1 - t_{ik})^m,$$

such that

$$0 \geq t_{ik} \geq 1, \quad \forall i = \{1 \ldots n\},\ k = \{1 \ldots c\}. \tag{1}$$

The first term of J_{PCM} corresponds to the objective function of FCM, where $m > 0$, usually set between 1.5 and 3 [23], is an exponent controlling the fuzziness of the partition and d_{ik} is the Euclidean distance of the object $\mathbf{x}_i$ to the cluster k. The second term of J_{PCM} was introduced in [12] to avoid the trivial solution consisting in a possibilistic partition with only null values. The weighting coefficients $\gamma_k > 0$ have a great influence on the clustering results and should be chosen with care. Default values can be computed by applying the following formula [12]:

$$\gamma_k = K \frac{\sum_{i=1}^{n} u_{ik}^m d_{ik}^2}{\sum_{i=1}^{n} u_{ik}^m}, \tag{2}$$

where $\mathbf{U} = (u_{ik})$ is a fuzzy partition obtained by applying the FCM algorithm and K, usually set to 1, is a weighting factor enabling to reduce or increase the overall size of the clusters.

By observation, it is easy to realize that while PCM helps to identify outliers it is also sensitive to initialization resulting in coincident clusters [13]. Such situation seems to appear when the objective function reaches a minimum, allowing to obtain satisfying results only with higher local minima [14]. Thus, variants of PCM have been proposed to solve this problem.

A. Repulsive Possibilistic c-means

In [14], the authors proposed a PCM variant which consists in introducing a third term in the objective function of PCM in order to penalize centroids too close from each other. This algorithm is referred to as Repulsive Possibilistic C-means (RPCM) due to the cluster repulsion role of the added term and its objective function is:

$$J_{RPCM}(\mathbf{T}, \mathbf{V}) = J_{PCM} + \sum_{k=1}^{c}\eta_k \sum_{l \neq k} \frac{1}{\|\mathbf{v}_k - \mathbf{v}_l\|^2}, \tag{3}$$

subject to constraint (1). The parameters $\eta_k \geq 0$, $\forall k \in \{1 \ldots c\}$ define specific degrees of repulsion for each cluster to the other ones.

The authors propose to minimize the objective function as PCM, i.e. by carrying out an iterative optimization of the possibilistic partition $\mathbf{T}$ and the centroids $\mathbf{V}$. Since the new term does not depend on $\mathbf{T}$, its update is identical to PCM:

$$t_{ik} = \frac{1}{1 + \left(\frac{d_{ik}^2}{\gamma_k}\right)^{\frac{1}{m-1}}}, \forall i \in \{1 \ldots n\}, k \in \{1 \ldots c\}.$$

The update of the centroids $\mathbf{V}$ composed of $\mathbf{v}_1 \ldots \mathbf{v}_k \ldots \mathbf{v}_c$ is more intricate. Indeed, in J_{RPCM} there exists a dependency between centroids. Thus, the optimization with the respect to $\mathbf{V}$ cannot be performed by updating each centroid $\mathbf{v}_k$ separately. However, in order to facilitate the optimization and to update prototypes one by one, [14] proposed to set as a constant $r_{kl} = ((\mathbf{v}_k - \mathbf{v}_l)^T(\mathbf{v}_k - \mathbf{v}_l))^{-2}$ $\forall k,\ l \in \{1 \ldots c\}$, making the hypothesis that such values are not significantly changing during the optimization process. The update formula is then obtained by setting the gradient $\frac{\partial J_{RPCM}}{\partial \mathbf{v}_k} = 0$:

$$\mathbf{v}_k = \frac{\sum_{i=1}^{n} t_{ik}^m \mathbf{x}_i - \eta_k \sum_{l \neq k} r_{kl}\mathbf{v}_l}{\sum_{i=1}^{n} t_{ik}^m - \eta_k \sum_{l \neq k} r_{kl}}, \forall k \in \{1 \ldots c\}. \tag{4}$$

Notice that the centroids update equation (4) does not always allow to minimize the objective function (3). Moreover, in [14], the authors state that if $\sum_{i=1}^{n} t_{ik}^m < \eta_k \sum_{l \neq k} r_{kl}$, concerned centroids have to be relocated at random positions. In [24], the authors propose to use a second order approximation method such as Newton, instead of a gradient based algorithm. Their choice is based on the fact that the gradient with respect to the centroids is non linear.

B. Possibilistic Fuzzy c-Means

A different approach to solve the problem raising with PCM is to combine possibility and probability membership values [15]. The objective function is then defined as follows:

$$\begin{aligned} J_{PFCM}(\mathbf{U}, \mathbf{T}, \mathbf{V}) \quad = \quad & \sum_{i=1}^{n}\sum_{k=1}^{c}(a u_{ik}^m + b t_{ik}^\eta) d_{ik}^2 \\ & + \sum_{k=1}^{c}\gamma_k \sum_{i=1}^{n}(1 - t_{ik})^\eta, \end{aligned} \tag{5}$$

where a, b and m, η are positive coefficients. The objective function, called PFCM for Possibilistic Fuzzy c-means, is subject to the constraints (1), (6) and (7).

$$\sum_{k=1}^{c} u_{ik} = 1 \qquad \forall k = \{1 \ldots c\}, \tag{6}$$

$$u_{ik} \geq 0 \quad \forall i \in \{1 \ldots n\},\ \forall k \in \{1 \ldots c\}, \tag{7}$$

In [15], the authors have demonstrated that J_{PFCM} is minimized by iteratively updating the fuzzy partition $\mathbf{U} = (u_{ik})$, the possibilistic partition $\mathbf{T}$ and the centroids $\mathbf{V}$ using equations (8)-(10) until convergence. Update formulas were obtained using the Lagrange multiplier method. Remark that the update formula of $\mathbf{U}$ is identical to FCM and the update formula of $\mathbf{T}$ is similar to PCM.

$$u_{ik} = \left(\sum_{l=1}^{c} \left(\frac{d_{ik}}{d_{il}} \right)^{\frac{2}{m-1}} \right)^{-1}. \tag{8}$$

$$t_{ik} = \left(1 + \left(\frac{b}{\gamma_k} d_{ik}^2 \right)^{\frac{1}{\eta-1}} \right)^{-1}. \tag{9}$$

$$\mathbf{v}_k = \frac{\sum_{i=1}^{n} (a u_{ik}^m + b t_{ik}^\eta) \mathbf{x}_i}{\sum_{i=1}^{n} (a u_{ik}^m + b t_{ik}^\eta)}. \tag{10}$$

III. Semi-supervised Repulsive PCM

In real applications, a priori information is available with various degrees of certainty. We propose to exploit soft label knowledge, i.e. objects labeled with a degree of membership to a cluster. Such information is retrieved thanks to an expert or automatically with the background knowledge.

Let $f_{ik} \in [0, 1]$ be the possibility known a priori that $\mathbf{x}_i$ belongs to the cluster k. This value, is equal to 0 when it is sure that the object i does not belong to the cluster k. Conversely, $f_{ik} = 1$ indicates that $\mathbf{x}_i$ has a strong possibility to belong to the cluster k, even if it also let the possibility to have degrees of belief for the other clusters.

A natural requirement for the Semi-supervised Repulsive PCM algorithm (SRPCM) is to obtain a possibilistic value t_{ik} the closest to f_{ik}. In particular cases and as in [16], constraints are softened to avoid sudden disturbance of the structure. Indeed, it can lead to inconsistent solutions.

Thus, we suggest to introduce a penalty term in the objective function of J_{rpcm} so that the label constraints are respected. The distance d_{ik} for a constrained object $\mathbf{x}_i$ on cluster k is employed to relax the constraint:

$$J_{SRPCM}(\mathbf{T}, \mathbf{V}) = J_{RPCM} + \alpha \sum_{i=1}^{n} \sum_{k=1}^{c} b_{ik} (t_{ik} - f_{ik})^m d_{ik}^2, \tag{11}$$

subject to constraint (1) and where $m > 1$ is even and $\alpha \geq 0$ is a tradeoff coefficient between the inherent structure unsupervisally retrieved and the consideration of the constraints. The variable b_{ik} enables to select only constrained values in the penalty term:

$$b_{ik} = \begin{cases} 1 & \text{if } \mathbf{x_i} \text{ and class } \mathbf{k} \text{ are constrained,} \\ 0 & \text{otherwise.} \end{cases}$$

The new objective function has to be minimized. As in RPCM, the optimization procedure follows an alternate scheme by fixing first the centroids $\mathbf{V}$ and second the possibilistic partition $\mathbf{T}$. We set $m = 2$ in order to ease the minimization.

A. Optimization with the respect to the possibilistic partition

The first step of the algorithm consists in fixing $\mathbf{V}$ to find an update formula of $\mathbf{T}$ that minimize $J_{srpcm}(\mathbf{T})$. Since each element t_{ik} of $\mathbf{T}$ are independent, we compute the derivative of the objective function (11) with the respect to t_{ik} and $m = 2$:

$$\frac{\partial J_{SRPCM}}{\partial t_{ik}} = 2 t_{ik} d_{ik}^2 - 2\gamma_k (1 - t_{ik}) + 2\alpha b_{ik} d_{ik}^2 (t_{ik} - f_{ik}).$$

Each value t_{ik} of the possibilistic partition minimizing J_{SRPCM} is obtained by setting $\frac{\partial J_{SRPCM}}{\partial t_{ik}} = 0$:

$$t_{ik} = \frac{\gamma_k + \alpha b_{ik} d_{ik}^2 f_{ik}}{\gamma_k + (\alpha b_{ik} + 1) d_{ik}^2}.$$

B. Optimization with the respect to the centroids

In a second step, the possibilistic partition $\mathbf{T}$ is fixed and the objective function J_{SRPCM} is minimized with the respect to the centroids. In order to choose a good optimization method, a study concerning the convexity of the objective function is first performed. The gradient is then calculated:

$$\begin{aligned} \frac{\partial J_{SRPCM}}{\partial \mathbf{v}_k} =& -2 \sum_{i=1}^{n} t_{ik}^2 (\mathbf{x}_i - \mathbf{v}_k) - 2\eta_k \sum_{l \neq k} \frac{\mathbf{v}_k - \mathbf{v}_l}{\|\mathbf{v}_k - \mathbf{v}_l\|^4} \\ & - 2\alpha \sum_{i=1}^{n} b_{ik} (t_{ik} - f_{ik})^2 (\mathbf{x}_i - \mathbf{v}_k). \end{aligned}$$

This gradient can be decomposed to obtain the value for a single element of $\mathbf{v}_k$, e.g. v_{k_j}:

$$\begin{aligned} \frac{\partial J_{SRPCM}}{\partial v_{k_j}} =& -2 \sum_{i=1}^{n} t_{ik}^2 (x_{i_j} - v_{k_j}) - 2\eta_k \sum_{l \neq k} \frac{v_{k_j} - v_{l_j}}{\|\mathbf{v}_k - \mathbf{v}_l\|^4} \\ & - 2\alpha \sum_{i=1}^{n} b_{ik} (t_{ik} - f_{ik})^2 (x_{i_j} - v_{k_j}). \end{aligned}$$

Then, elements composing the Hessian matrix $\mathbf{H}_k \in R^{p \times p}$ deduced from the second derivatives of J_{SRPCM} with the respect to the centroids are the following:

$$\begin{aligned} \frac{\partial J_{SRPCM}}{\partial^2 v_{k_j}} =& 2 \sum_{i=1}^{n} t_{ik}^2 - 2\eta_k \sum_{l \neq k} \frac{1}{d_{lk}^2} - \frac{4 (v_{k_j} - v_{l_j})^2}{\|\mathbf{v}_k - \mathbf{v}_l\|^6} \\ & + 2\alpha \sum_{i=1}^{n} b_{ik} (t_{ik} - f_{ik})^2, \\ \frac{\partial J_{SRPCM}}{\partial v_{k_j} \partial v_{k_{j'}}} =& 8\eta_k \sum_{l \neq k} \frac{(v_{k_j} - v_{l_j})(2 v_{k_{j'}} - 2 v_{l_{j'}})}{\|\mathbf{v}_k - \mathbf{v}_l\|^6}, \end{aligned}$$

where $v_{k_{j'}}$ corresponds the j'th element of the centroid $\mathbf{v}_k$ such that $j' \neq j$. Notice that a similar result, without the label constraints, is available in [24].

Finally, the Hessian matrix $\mathbf{H}_k$ can be rewritten as follows:

$$\begin{aligned}\mathbf{H}_k &= 2\left(\sum_{i=1}^{n} t_{ik}^2\right)\mathbf{I} + 8\eta_k \sum_{l\neq k} \frac{(\mathbf{v}_k-\mathbf{v}_l)(\mathbf{v}_k-\mathbf{v}_l)^T}{\|\mathbf{v}_k-\mathbf{v}_l\|^6} \\ &- 2\eta_k\left(\sum_{l\neq k}\frac{1}{d_{lk}^2}\right)\mathbf{I} + 2\alpha\sum_{i=1}^{n} b_{ik}(t_{ik}-f_{ik})^2\mathbf{I},\end{aligned}$$

where $\mathbf{I}$ is the identity matrix of proper dimension.

Let us remind that the sum of positive (semi)definite matrices results in a positive (semi)definite matrix. Since $\mathbf{I}$ is positive definite, $t_{ik}^2 \geq 0$ and $b_{ik}(t_{ik}-f_{ik})^2 \geq 0$ then the first and the last term of $\mathbf{H}_k$ are positive semidefinite. Similarly, the matrix $(\mathbf{v}_k-\mathbf{v}_l)(\mathbf{v}_k-\mathbf{v}_l)^T$ is positive semidefinite, $8\eta_k \geq 0$ and $\|\mathbf{v}_k-\mathbf{v}_l\|^6 \geq 0$ so the second term of $\mathbf{H}_k$ is also positive semidefinite. Finally, the third term is negative semidefinite. Indeed, the scalar coefficient applied for $\mathbf{I}$ is negative or equal to 0. Consequently, $\mathbf{H}_k$ is not guaranteed to be positive semidefinite.

Thus, in order to update the centroids, a standard trust-region method for non linear minimization is employed [25]. Although such method may just reach a local minimum, it assures the convergence of the clustering algorithm. Indeed, the objective function value after the centroids update is inferior or equal to the value before. Conversely the optimization methods employed in [14] and [24] might not follow a descent direction in case of a negative Hessian.

IV. Semi-supervised PFCM

The penalty term used to create SRPCM can also be integrated in PFCM in order to produce a new algorithm called SPFCM. It enables to handle soft label constraints. The objective function to be minimized is the following:

$$J_{SPFCM}(\mathbf{U},\mathbf{T},\mathbf{V}) = J_{PFCM} + \alpha\sum_{i=1}^{n}\sum_{k=1}^{c} b_{ik}(t_{ik}-f_{ik})^{\eta} d_{ik}^2, \tag{12}$$

subject to (1), (6) and (7).

Fixed parameters have the same limits as PFCM, except for η that should be positive and even. The α coefficient follows the same rule as SRPCM, i.e. $\alpha \geq 0$.

The objective function (12) is optimized using a heuristic method which consists in iteratively minimizing J_{SPFCM} with respect to $\mathbf{U}$, then $\mathbf{T}$, then $\mathbf{V}$ until convergence.

A. Update of the probabilistic partition

The optimization of J_{SPFCM} with respect to $\mathbf{U}$ is achieved by fixing $\mathbf{T}$ and $\mathbf{V}$ as constants. Since the penalty term incorporated for SPFCM does not contain any probabilistic partition values, the update of the membership degrees $\mathbf{U}$ are identical to PFCM and corresponds to the equation (8).

B. Update of the possibilistic partition

In order to minimize J_{SPFCM} with respect to $\mathbf{T}$, the variables $\mathbf{U}$ and $\mathbf{V}$ are fixed. The columns and rows of $\mathbf{T}$ are independent, letting us the possibility to update each value t_{ik} separately. By setting $\eta = 2$ to facilitate the optimization process, the problem becomes quadratic. The derivative is then calculated:

$$\frac{\partial J_{SPFCM}}{\partial t_{ik}} = 2bt_{ik}d_{ik}^2 - 2\gamma_k(1-t_{ik}) + 2\alpha b_{ik}d_{ik}^2(t_{ik}-f_{ik}).$$

Setting the derivative to 0 enables to obtain the following update formula:

$$t_{ik} = \frac{\gamma_k + \alpha b_{ik} d_{ik}^2 f_{ik}}{b d_{ik}^2 + \gamma_k + \alpha b_{ik} d_{ik}^2}.$$

Notice that when $b = 1$, the update of the possibilistic partition for SPFCM is identical to SRPCM.

C. Update of the centroids

Since $au_{ik}^m + bt_{ik}^{\eta} > 0$ and $(t_{ik}-f_{ik})^m > 0$, J_{SPFCM} is positive semidefinite with respect to $\mathbf{V}$. As a consequence, the minimum of the objective function corresponds to the value of $\mathbf{V}$ vanishing the derivative. Notice that each centroid $\mathbf{v}_k$ is independent to each other and can then be managed separately:

$$\begin{aligned}\frac{\partial J_{SPFCM}}{\partial \mathbf{v}_k} &= -2\sum_{i=1}^{n}(au_{ik}^m + bt_{ik}^2)(\mathbf{x}_i-\mathbf{v}_k) \\ &\quad -2\alpha\sum_{i=1}^{n} b_{ik}(t_{ik}-f_{ik})^2(\mathbf{x}_i-\mathbf{v}_k).\end{aligned}$$

Let z_{ik} be a scalar such that $z_{ik} = (au_{ik}^m + bt_{ik}^2) + \alpha b_{ik}(t_{ik}-f_{ik})^2$. Setting the derivative to 0 leads to the following result:

$$\mathbf{v}_k = \frac{\sum_{i=1}^{n} z_{ik}\mathbf{x}_i}{\sum_{i=1}^{n} z_{ik}}.$$

V. Experimental Results

A. Experimental protocol

We have run extensive experimental tests on three well known datasets from the UCI repository, i.e. Ecoli, Iris and Wine, and on a toy dataset called GaussK2. The characteristics of each dataset are presented in Table I.

TABLE I
Dataset characteristics

Dataset	# classes used	# classes	# attributes	# instances
Ecoli	5	8	7	336
Iris	3	3	4	150
Wine	3	3	13	178
GaussK2	2	2	2	400

Ecoli dataset contains eight classes from which three classes contain very few instances (2, 2, and 5 instances). We consider the instances from these classes as outliers, leading us to take into account five classes only.

The GaussK2 dataset is a two-dimensional space dataset generated by two gaussians with different covariance matrices.

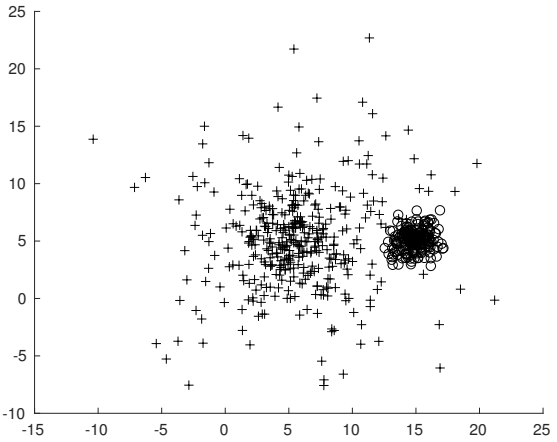

Fig. 1. GaussK2 dataset. Classes are represented with crosses and circles.

The result is that the clusters have different densities, as shown in Figure 1.

The parameters for SRPCM and SPFCM algorithms were defined as follows: $\alpha = 1$ and γ_k is retrieved as explained in [12] with equation (2) and $K = 1$. The parameters $\eta_k \; \forall k \in \{1 \dots c\}$ for SRPCM are identical and fixed manually for each dataset. Similarly, for SPFCM, we set $a = 1$ and b is manually chosen following each dataset. Details are presented Table II.

TABLE II
PARAMETERS EMPLOYED FOR SRPCM AND SPFCM

Dataset	η_k	b
Ecoli	0.8	2.2
Iris	2	3
Wine	0.2	1.6
GaussK2	0.1	0.2

In order to evaluate the proposed methods, a comparison with SKMEANS [26] and SFCM [16] has been performed. The SKMEANS algorithm corresponds the k-means method taking label constraints as background knowledge while the SFCM represents the fuzzy c-means version with label constraints.

Final acquired partitions are hard for SKMEANS, fuzzy for SFCM and possibilistic for SRPCM and SPFCM. To perform a comparison between these methods, fuzzy and possibilistic partitions are transformed into hard partition by selecting the cluster with the maximum of probability or possibility. Then, since the true classes of the datasets are known, we compute the ARI [27] to measure the performance of the clustering algorithms.

B. Clustering results

All clustering algorithms use the same initial conditions for centroids and constraints. The constraints correspond to totally certain labels. Experiments consist in 100 trials with a given percentage of constraints. Each trial corresponds to 5 executions of an algorithm with different centroid initializations. The partition with the minimum objective function value is then selected.

Figures 2, 3, 4 and 5 present the average ARI and its confidence interval against the proportion of labeled constraints for SKMEANS, SFCM and the proposed clustering algorithms, i.e. SRPCM and SPFCM.

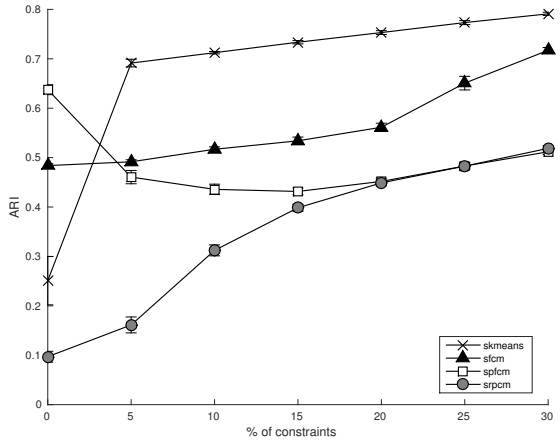

Fig. 2. Clustering performances against the proportion of constraints, Ecoli.

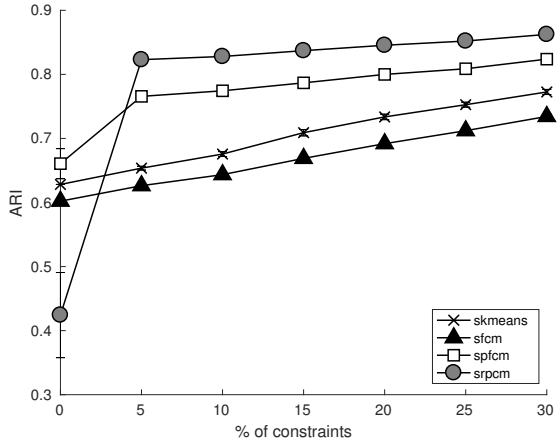

Fig. 3. Clustering performances against the proportion of constraints, GaussK2.

The results show that the addition of constraints on SRPCM clearly improves the partition compared to the initial partition found without constraints, i.e. with RPCM. For GaussK2 and Iris, the best performances are achieved by SRPCM. However, the same clustering method have low results on Wine and Ecoli. We can also observe that RPCM has always a low ARI. The reason is that the algorithm is more sensitive to local minima than the other clustering methods.

Conversely, the SPFCM algorithm gives more stable results than SRPCM. Then, it can be seen as a more secure choice than SRPCM when no expert assessment can be performed a posteriori. Indeed, it still obtains better results than SKMEANS and SFCM for Iris and GaussK2.

The results on the Ecoli dataset can be surprising on the first approach. Indeed, without constraints, PFCM outperforms the other algorithms. The performance of this algorithm is however inferior when constraints are taken into account. Such

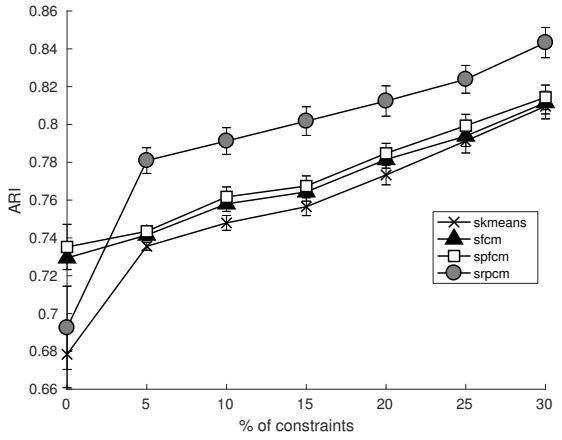

Fig. 4. Clustering performances against the proportion of constraints, Iris.

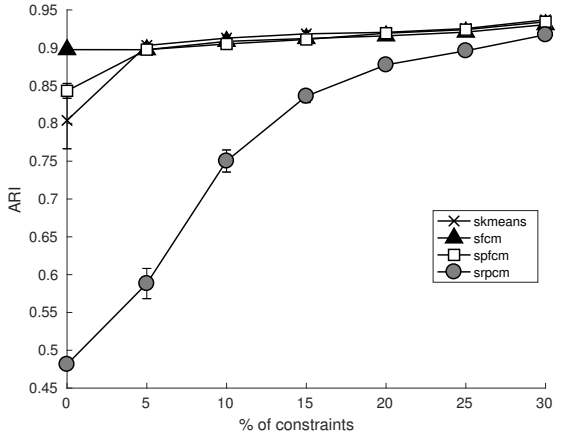

Fig. 5. Clustering performances against the proportion of constraints, Wine.

behavior can be explained by the inherent structure of Ecoli: over the 5 classes of the dataset, a couple of pairs of classes are quasi-overlapped. Thus, none of the clustering algorithms are able to find the real structure of Ecoli. Experimental results show that PFCM has the best ARI by finding coincident clusters. The addition of constraints in SPFCM tends to separate the clusters, thus decreasing its ARI value.

Inversely, few number of constraints enables SKMEANS to obtain the best accuracy for Ecoli. The reason is that SKMEANS forces the total respect of the constraints whereas the other algorithms, with the use of a penalty term, can let some constraints not respected to obtain a more coherent final structure. The SKMEANS algorithm has then the possibility to converge faster to the desired solution.

C. Outliers detection

An interesting feature of a possibilistic clustering algorithm is its ability to identify outliers. Consequently, an experiment on the Ecoli dataset has been performed to show how SRPCM and SPFCM handle outliers.

The same experimental protocol described above is used, i.e. 100 trials for each specific set of constraints. When a possibilistic partition is retrieved, a simple rule to detect outliers is applied: an object $\mathbf{x}_i$ is an outlier if $t_{ik} \leq 0.1$, $\forall k \in \{1 \dots c\}$. The average rates of good detection as well as their confidence interval are illustrated Figure 6. As it can be observed, labels constraints helps to the detection of outliers. The SRPCM algorithm detects better outliers than SPFCM. This can be explained by the fact that the optimization of the fuzzy partition for SPFCM has an impact on the possibilistic partition.

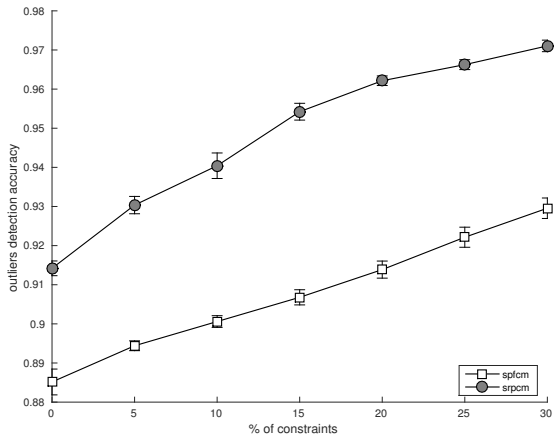

Fig. 6. Accuracy rate for outliers detection on Ecoli.

VI. Conclusions and Future Work

In this paper, two novel semi-supervised clustering algorithms have been presented: SRPCM and SPFCM. Both of them incorporate label constraints and handle a possibilistic partition. The use of a possibilistic framework enables to express various type of uncertainty. The advantage of such framework is twofold in SRPCM and SPFCM: first, constraints are represented in the form of possibility to belong to clusters. Such representation enables an expert to include partial information into the clustering algorithms more than any other semi-supervised variants of k-means and FCM. Second, the generation of a possibilistic partition by SRPCM and SPFCM enables to obtain rich information about the dataset and makes easy for instance the detection of outliers.

Encouraging results have emerged from experiments and comparisons with other constraint-based methods. It leads to consider several possible future works: first, the impact of labels having a degree of uncertainty can be study and an active learning scheme can be developed. Second, an investigation about the parameters that are currently manually fixed can be performed in order to acquire them automatically. Future works includes extensions to use other distance measures such as Mahalanobis distance as well as extending the proposed method to consider more complex uncertainty models.

Acknowledgment

This work has been partially supported by the Autonomous University of Tamaulipas through the grants No: PEI-2014-209955, PEI-2014-213516, P/PFCE-2016-28MSU0010B-22 and by Conacyt through the Master's Scholarships of Tanya Boone R.N 588607.

REFERENCES

[1] D. Xu and Y. Tian, "A comprehensive survey of clustering algorithms," *Annals of Data Science*, vol. 2, no. 2, pp. 165–193, 2015.

[2] O. Verevka and J. Buchanan, "The local k-means algorithm for colour image quantization.," in *Graphics Interface*, pp. 128–128, Canadian Information Processing Society, 1995.

[3] H. Hussain, K. Benkrid, H. Seker, and A. Erdogan, "Fpga implementation of k-means algorithm for bioinformatics application: An accelerated approach to clustering microarray data," in *NASA/ESA Conference on Adaptive Hardware and Systems (AHS)*, pp. 248–255, IEEE, 2011.

[4] R. Khan, C. Barat, D. Muselet, and C. Ducottet, "Spatial orientations of visual word pairs to improve bag-of-visual-words model," in *Proceedings of the British Machine Vision Conference*, pp. 89–1, BMVA Press, 2012.

[5] L. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.

[6] L. Tari, C. Baral, and S. Kim, "Fuzzy c-means clustering with prior biological knowledge," *Biomedical Informatics*, vol. 42, pp. 74–81, Feb 2009.

[7] S. Ghosh, S. Mitra, and R. Dattagupta, "Fuzzy clustering with biological knowledge for gene selection," *Applied Soft Computing*, vol. 16, pp. 102–111, 2014.

[8] M. Steinbach, G. Karypis, and V. Kumar, "A comparison of document clustering techniques," in *KDD Workshop on Text Mining*, vol. 400, pp. 525–526, Boston, 2000.

[9] V. Singh, N. Tiwari, and S. Garg, "Document clustering using k-means, heuristic k-means and fuzzy c-means," in *Computational Intelligence and Communication Networks (CICN)*, pp. 297–301, IEEE, 2011.

[10] M. Chacon and G. Ramirez, "Fuzzy clustering algorithms in subjective classification tasks," in *IEEE Intl. Conf. on Fuzzy Systems*, pp. 2309–2316, 2006.

[11] X. Wang and J. Bu, "A fast and robust image segmentation using fcm with spatial information," *Digital Signal Processing*, vol. 20, no. 4, pp. 1173–1182, 2010.

[12] R. Krishnapuram and J. Keller, "A possibilistic approach to clustering," *IEEE transactions on fuzzy systems*, vol. 1, no. 2, pp. 98–110, 1993.

[13] M. Barni, V. Cappellini, and A. Mecocci, "Comments on a possibilistic approach to clustering," *IEEE Trans. on Fuzzy Systems*, vol. 4, pp. 393–396, 1996.

[14] H. Timm, C. Borgelt, C. Döring, and R. Kruse, "An extension to possibilistic fuzzy cluster analysis," *Fuzzy Sets and systems*, vol. 147, no. 1, pp. 3–16, 2004.

[15] N. Pal, K. Pal, J. Keller, and J. Bezdek, "A possibilistic fuzzy c-means clustering algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 13, pp. 517–530, 2005.

[16] W. Pedrickz and J. Waletzky, "Fuzzy clustering with partial supervision," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 27, no. 5, pp. 787–795, 1997.

[17] K. Wagstaff, C. Cardie, S. Rogers, and S. Schroedl, "Constrained k-means clustering with background knowledge," in *Proceedings of the 18th International Conference on Machine Learning*, (Williamstown, MA, USA), pp. 577–584, 2001.

[18] V. Antoine, N. Labroche, and V. Vu, "Evidential seed-based semi-supervised clustering," in *Soft Computing and Intelligent Systems (SCIS), Joint 7th International Conference on and Advanced Intelligent Systems (ISIS), 15th International Symposium*, pp. 706–711, IEEE, 2014.

[19] V. Antoine, B. Quost, M. Masson, and T. Denoeux, "Cevclus: evidential clustering with instance-level constraints for relational data," *Soft Computing*, vol. 18, no. 7, pp. 1321–1335, 2014.

[20] A. Bensaid, J. Bezdek, and L. Clarke, "Partially supervised clustering for image processing," *Pattern Recognition*, pp. 859–871, 2007.

[21] S. Zhong and J. Ghosh, "Scalable, balanced model-based clustering," in *Proceedings of the 2003 SIAM International Conference on Data Mining*, pp. 71–82, SIAM, 2003.

[22] D. Gondek and T. Hofmann, "Non-redundant data clustering," *Knowledge and Information Systems*, vol. 12, no. 1, pp. 1–24, 2007.

[23] J. Bezdek, R. Ehrlich, and W. Full, "Fcm: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2, pp. 191 – 203, 1984.

[24] M. Ferraro and P. Giordani, "On possibilistic clustering with repulsion constraints for imprecise data," *Information Sciences*, vol. 245, pp. 63–75, 2013.

[25] Y. Yuan, "A review of trust region algorithms for optimization," in *ICIAM*, vol. 99, pp. 271–282, 2000.

[26] S. Basu, A. Banerjee, and R. Mooney, "Semi-supervised clustering by seeding," in *Proceedings of 19th International Conference on Machine Learning (ICML)*, pp. 19–26, 2002.

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