Semi-supervised fuzzy c-means variants: a study on noisy label supervision

Violaine Antoine[1] and Nicolas Labroche[2]

[1] Clermont Auvergne University, UMR 6158, LIMOS, F-63006,
Clermont-Ferrand, France
violaine.antoine@uca.fr
[2] University of Tours, EA 6300, LIFAT, France

Abstract. Semi-supervised clustering algorithms aim at discovering the hidden structure of data sets with the help of expert knowledge, generally expressed as constraints on the data such as class labels or pairwise relations. Most of the time, the expert is considered as an oracle that only provides correct constraints. This paper focuses on the case where some label constraints are erroneous and proposes to investigate into more detail three semi-supervised fuzzy c-means clustering approaches as they have been tailored to naturally handle uncertainty in the expert labeling. In order to run a fair comparison between existing algorithms, formal improvements have been proposed to guarantee and fasten their convergence. Experiments conducted on real and artificial data sets under uncertain labels and noise in the constraints show the effectiveness of using fuzzy clustering algorithm for noisy semi-supervised clustering.

Keywords: fuzzy clustering, label constraints, semi-supervised clustering, noise

1 Introduction

Semi-supervised clustering algorithms are part of exploratory data analysis. They intend to extract the underlying structure of datasets by grouping similar objects together with the help of some partial external knowledge usually provided as pairwise constraints [1], e.g. must-link/cannot-link constraints between pairs of objects that indicate if two objects must (or not) be in the same cluster, or labels constraints [2], that specify explicitly the class labels for some objects. These approaches can lead clustering algorithms towards a better definition of the existing structures in the data, or at least to a definition that better fits the needs of the final user. For clustering algorithms that are directly derived from the optimization of an objective function, like k-means and its variants, various methods have been proposed by adding a penalty term [2–4] or by learning a proper metric [2, 5] that adapts the topology so that less constraints are violated.

However, all these methods heavily depend on the quality of the provided expert knowledge. Even in the best case, where only correct constraints are provided to the algorithms, it has been shown that improperly chosen constraints can deteriorate performances [6]. Hence, solutions have been proposed to evaluate the quality or the utility of constraints prior to clustering to avoid such problem [7,8]. But, to the best of our knowledge, no work has directly tackled the problem of semi-supervised clustering when the expert does not provide relevant constraints.

This paper shows that, in this context of erroneous or uncertain expert labeling, it is possible to use the natural property of fuzzy clustering algorithm to handle uncertainty in constraints to maintain good clustering performances. For the sake of clarity, this paper is restricted to label constraints since they are more general than pairwise constraints. The study is also limited to variants of fuzzy c-means (FCM) that include a term to penalize the solution when label constraints are not respected. As such, we discard more complex FCM algorithms as the kernel-based [9] or those that determine the number of clusters [10,2].

Without loss of generality, we consider label constraints expressed as a fuzzy membership matrix $\tilde{\mathbf{U}} = (\tilde{u}_{ik})$ that indicates to which extent each object i is supposed to be assigned to the cluster k according to the expert. In this case, an object does not necessarily have constraints and these constraints may not be completely certain, ie. $0 \leq \sum_k \tilde{u}_{ik} \leq 1$. Table 1 illustrates such matrix $\tilde{\mathbf{U}}$ with 4 objects and 3 clusters and introduces the vocabulary that will be used in the experiments.

Table 1. Example of a constraint membership matrix. Object o_1 represents the traditional seed constraint with a crisp assignment to a single cluster. Object o_4 is not constrained. Objects o_2 and o_3 show the expressiveness brought by fuzzy representation of constraints with certain or uncertain / single or multi-labels.

	c_1	c_2	c_3	$\sum_k \tilde{u}_{ik}$	Explanations
o_1	1	0	0	1	Single-label and certain constraint
o_2	0	0.3	0	0.3	Single-label and uncertain constraint
o_3	0	0.5	0.5	1	Double-label and certain constraint
o_4	0	0	0	0	Unconstrained object

A comparative review on semi-supervised fuzzy c-means algorithms with label contraints has already been performed in [11]. However, their objective is not to evaluate the ability of the algorithms to deal with erroneous or noisy expert labels and the soundness of optimization techniques is not discussed, as a strict copy of the original algorithms is employed. In this paper, we consider modified algorithms to conduct a fair comparison that only involves penalty term employed in FCM for the constraints. To this aim, we ensure and fasten the convergence of the optimization and we introduce the Mahalanobis distance when it is not already achieved, as a specific and adaptive distance for each cluster is beneficial for some datasets.

The rest of the paper is then organized as follows. Semi-supervised clustering algorithms and their modifications are presented Sections 2 and 3. Experiments on raw, uncertain and noisy labels are introduced Section 4 and a conclusion is available Section 5.

2 Semi-supervised clustering algorithms

Let $\mathbf{X} = \{\mathbf{x}_i, \dots \mathbf{x}_n\}$ be a dataset composed of n objects such that $\mathbf{x}_i \in \mathbb{R}^p$ is the feature vector representing the object i. The clusters are defined by centroids $\mathbf{V} = \{\mathbf{v}_1, \dots \mathbf{v}_c\}$ and d_{ik}^2 corresponds to the squared Euclidean distance between the object $\mathbf{x}_i$ and the centroid $\mathbf{v}_k$. The standard fuzzy c-means algorithm minimizes the intraclass inertia by alternatively optimizing the degrees of membership $\mathbf{U} = (u_{ik})$ and the centroids $\mathbf{V}$ [12, 13]. The objective function is the following:

$$J_{FCM}(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^{n} \sum_{k=1}^{c} u_{ik}^m d_{ik}^2, \tag{1}$$

where $m > 1$ is a fixed value that controls the degree of fuzziness for the partition and u_{ik} should satisfy:

$$\sum_{k=1}^{c} u_{ik} = 1; \quad u_{ik} > 0 \quad \forall i \in \{1 \dots n\}, \forall k \in \{1 \dots c\}. \tag{2}$$

Gustafson and Kessel have proposed a variant of FCM that use a specific Mahalanobis distance for each cluster [13]. The distance between an object $\mathbf{x}_i$ and a cluster k becomes $d_{ik}^2 = (\mathbf{x}_i - \mathbf{v}_k)^T \mathbf{S}_k (\mathbf{x}_i - \mathbf{v}_k)$, where $\mathbf{S}_k$ is the norm-inducing matrix of the cluster k. The matrices $\mathbf{S}_1 \dots \mathbf{S}_c$ are defined as fuzzy covariance matrices and enable to detect the optimal geometrical shapes of the clusters.

sfcm is a famous algorithm that add a penalty term in the objective function of FCM to take into account uncertain labels [10] and for which an extension with Mahalanobis distance already exists [2]. The proposed objective function minimizes the following criteria such that constraints (2) are respected:

$$J_{sfcm}(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^{n} \sum_{k=1}^{c} u_{ik}^m d_{ik}^2 + \alpha \sum_{i=1}^{n} \sum_{k=1}^{c} (u_{ik} - \tilde{u}_{ik} \mathbf{b}_i)^m d_{ik}^2, \tag{3}$$

where $m > 1$ must be an even number, $\alpha \in \mathbb{R}^+$ is a coefficient controlling the tradeoff between the objective function of FCM and the constraints, $\tilde{\mathbf{U}} = (\tilde{u}_{ik})$ is a partition given by an analyst and $\mathbf{b}_i$ is such that $b_i = 1$ if $\mathbf{x}_i$ is constrained and $b_i = 0$ otherwise.

This paper proposes a simple correction of the update equation of the prototypes $\mathbf{V}$ that is similar to what is proposed in [2]:

$$\mathbf{v}_k = \frac{\sum_{i=1}^{n} \alpha \left(u_{ik}^m + (u_{ik} - \tilde{u}_{ik}\mathbf{b}_i)^m\right) \mathbf{x}_i}{\sum_{i=1}^{n} \alpha u_{ik}^m + (u_{ik} - \tilde{u}_{ik}\mathbf{b}_i)^m}, \forall k \in \{1 \ldots c\}. \tag{4}$$

ssfcm is the first of the two semi-supervised FCM algorithms proposed in [14]. It minimizes the following objective function:

$$J_{ssfcm}(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^{n} \sum_{k=1}^{c} |u_{ik} - \tilde{u}_{ik}|^m d_{ik}^2, \tag{5}$$

with $m \geq 1$ and such that constraints (2) are respected.

The algorithm ssfcm has no coefficient to set for some tradeoff between the inherent structure of the data and the constraints. Thus, the optimization is straightforward and the convergence ensured. However, it enforces a total respect of the constraints and consequently may not be able to deal efficiently with noisy or erroneous constraints.

In our test, we have proposed an extension of ssfcm with a Mahalanobis distance following the approach of Gustafson and Kessel [13] to make possible a fair comparison with the other algorithms when ellipsoidal clusters are to be found. Learning a Mahalanobis distance comes down to defining a matrix $(p \times p)$ $\mathbf{S}_k$ for each cluster k and minimizing the objective function (5) with the respect to $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{S} = (\mathbf{S}_1 \ldots \mathbf{S}_c)$. In order to avoid trivial solution consisting of $\mathbf{S}_k$ with only zeros that would minimize the objective function, a constant volume $\rho_k > 0$ is assigned to each cluster k:

$$det(\mathbf{S}_k) = \rho_k, \forall k \in \{1 \ldots c\} \tag{6}$$

The constrained optimization problem is solved by introducing c Lagrange multipliers λ_k in J_{ssfcm}:

$$\mathcal{L} = J_{ssfcm}(\mathbf{U}, \mathbf{V}, \mathbf{S}) - \sum_{k=1}^{c} \lambda_k(\rho_k - det(\mathbf{S}_k)). \tag{7}$$

Setting the derivative of the Lagragian function to 0 leads to the following result:

$$\begin{aligned} \mathbf{S}_k &= \rho_k \det(\mathbf{\Sigma}_k)^{\frac{1}{p}} \mathbf{\Sigma}_k^{-1}, \\ \mathbf{\Sigma}_k &= \sum_{i=1}^{n} \sum_{k=1}^{c} |u_{ik} - \tilde{u}_{ik}|^m (\mathbf{x}_i - \mathbf{v}_k)^T (\mathbf{x}_i - \mathbf{v}_k). \end{aligned}$$

esfcm is an entropy regularized FCM [14] with the following objective function:

$$J_{esfcm} = \sum_{i=1}^{n}\sum_{k=1}^{c} u_{ik} d_{ik}^2 + \lambda^{-1} \sum_{i=1}^{n}\sum_{k=1}^{c} (|u_{ik} - \tilde{u}_{ik}|) \log(|u_{ik} - \tilde{u}_{ik}|), \tag{8}$$

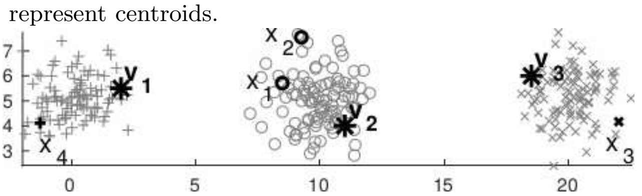
such that $\lambda \in \mathbb{R}^+$ and constraints (2) are respected. In order to minimize this objective function, the authors remove the absolute value and replace it by new constraints $u_{ik} \geq \tilde{u}_{ik}$ $\forall i \in \{1 \dots n\}$, $\forall k \in \{1 \dots c\}$ so that the function is still convex. As for ssfcm, the update equation of u_{ik} depends on $\tilde{u}_{ik}$ which may limit the way esfcm deals with erroneous constraints. Finally, enriching esfcm with a Mahalanobis distance is similar to what is performed for FCM.

3 Mapping function

One common problem when evaluating semi-supervised clustering algorithms based on random initial centers such as FCM, is that the label assigned randomly to these centers may not coincide with the labels used to express the constraints. The problem can be solved by taking as initial centers the barycenter computed with the constrained labeled objects [15]. However, this solution is inappropriate when there exists clusters without labels or when the constraints set is noisy.

As an exemple, let us consider a dataset with 4 objects and for each object the following constraints labels: x_1 and x_2 in cluster 1, x_3 in cluster 2 and x_4 in cluster 3. Figure 1 presents a dataset with the previous constraints and some initial centers named $v_{1,2,3}$. It is obvious to observe that there exists a mismatch between the clusters, more particularly their centers labels, and the labels of the constrained objects. For instance, $\mathbf{x}_4$ should be in the class 3 but is assigned to cluster 1. In this case, the convergence of the algorithm to a solution where the centroid $\mathbf{v}_3$ is close to $\mathbf{x}_4$ is too expensive compared to a solution where some constraints are violated which in turn leads to poor results.

Fig. 1. Dataset with three clusters. Symbols '+', 'o', 'x' correspond to the real classes whereas stars represent centroids.



To this aim, our mapping function simply considers all pairing of labels between the one provided by clusters centers and the one provided by the constraints and each time performs the complete clustering. The pairing that is finally kept is the one that minimizes the objective function.

4 Experiments

This section is devoted to the comparison of sfcm, esfcm, ssfcm as well as skmeans when possible for several real-world and synthetic datasets. The skmeans algorithm is a semi-supervised clustering method that uses labeled data to improve a traditional k-means algorithm [15]. We use it as a baseline to show the interest of using fuzzy approaches in the case of uncertain or noisy supervision. We also implicitly compare our approach to a traditional FCM as it corresponds to sfcm without constraints. First, a study of the λ parameter for esfcm is conducted as its behavior highly depends on this parameter. Next, experiments are carried out to represent different scenarios where expert annotation can induce errors in constrained algorithms. In the case of single constraint, the membership degree provided by the expert can either be wrong (error in the chosen class label), uncertain (low membership constraint while 1 was expected) or both at the same time. Finally, in our multi-label scenario, we deal with the case where the expert may hesitate between two class labels to annotate one object.

4.1 Experimental settings

We selected six well-known datasets from the UCI repository[3]: Glass, Ionosphere, Iris, Letters, Vehicle, Wine and a synthetic dataset generated with Gaussians: GaussK6. Characteristics are available in Table 2. For the Letters dataset, only the three letters I,J,L are kept as done in [16]. GaussK6 contains 2 overlapped classes. This dataset, as well as Wine, is suitable for a Euclidean distance whereas the other datasets offer better results with the Mahalanobis distance.

Table 2. Description of the datasets.

Name	GaussK6	Glass	Ionosphere	Iris	Letters	Vehicle	Wine
n	1200	214	351	150	227	846	178
p	2	8	33	4	16	18	13
c	6	2	2	3	3	4	3
Class sizes	200 / class	{163,51}	{126,225}	50 / class	{81,72,74}	{199,217,218,212}	{59,71,48}

In order to obtain a fair comparison between the algorithms, the same constraints and the same centers initializations have been tested at each experiment. An experiment consists in 100 trials where 1 trial executes 5 different initializations of the centers. The final result selected is the one with the minimal objective function.

In our experiments, our objective is to see how fuzzy clustering algorithms may help reaching better performances than crisp clustering algorithms when dealing with uncertain / noisy labels. However, in the end, we are interested in solving the crisp clustering problem since a decision has to be made about the class memberships of the objects. For this reason, the evaluation of the

[3] Available at http://archive.ics.uci.edu/ml

accuracy is calculated with the Adjusted Rand Index (ARI) [17] rather than a specific index related to fuzzy clustering. Moreover, ARI measures the similarity between two crisp partitions by taking into account the possibility that the obtained clustering is observed by chance. For fuzzy clustering algorithms, hard partition was determined by assigning objects to the cluster with the maximum membership value provided by the final fuzzy partition.

The modified partition coefficient (MPC) [18] has also been calculated to choose the λ parameter. This validity index measures the fuzziness of a partition: a crisp partition corresponds to a 1 value and a total fuzzy partition to a 0 value.

4.2 Choice of parameters

For all experiments the exponent m controlling the fuzziness of the final partition is set to 2.

The α parameter is set in such a way that two terms of the objective function have the same importance. Then, it gives a balance between the search for an underlying structure and the respect of the constraints. It is left to future work to study the influence of this parameter.

The λ parameter is more complicated to set, as it plays a key role on the behavior of esfcm even without constraints. Thus, experiments were conducted on esfcm with no partial supervision to set the value of λ. Various λ values have been tested and both MPC and ARI measures have been calculated.

As a result, we noticed that the MPC value is increasing as the λ value increases. This comportment is easily explained by the fact that MPC measures the fuzziness of a partition and λ behaves as a fuzzy controller of the final partition. Thus, setting a MPC value close to 0.8 assures us to obtain a partition neither too crisp nor too fuzzy. Nonetheless, we have also observed that the MPC and ARI measures are not totally correlated, particularly when a Mahalanobis distance is used. Experiments reported in Table 3 show that, in general, a good accuracy is reached when MPC is around 0.8.

Table 3. λ values used in esfcm for the average MPC measure around 0.8 and the average corresponding ARI.

	GaussK6	Glass	Ionosphere	Iris	LettersIJL	Vehicle	Wine
λ	0.14	3.24	2.35	4	0.125	2.5	0.31
MPC	0.80	0.81	0.80	0.80	0.81	0.79	0.80
ARI	0.81	0.37	0.08	0.67	0.22	0.16	0.90

4.3 Comparative experiments

Several experiments are reported in this section depending on the presence or not of constraints and on the quality of constraints ranging from single-label (un)certain constraints with added noise, to multi-label (un)certain constraints to simulate expert annotation errors.

No constraint First, the algorithms are executed without constraints to establish a comparative baseline for each dataset. Table 4 illustrates the average ARI and its 95% confidence interval for skmeans, sfcm, ssfcm and esfcm without constraints, i.e. k-means, FCM and FCM with an entropy regularization. Since skmeans has only the possibility to use a Euclidean distance, it cannot be compared to algorithms employing a Mahalanobis distance, hence the missing values in Table 4.

The sfcm and ssfcm algorithms without constraints, which correspond to FCM, outperform most of the time esfcm and skmeans. Low values of ARI are still visible, for example with the Vehicle dataset or the Ionosphere dataset. It means that the global structure of the data is difficult to detect and requires background knowledge to help its discovery.

Since we observed that the confidence interval remains stable when constraints are introduced, their values are not presented in the next tables.

Table 4. No constraint: average ARI and 95% confidence intervals for each algorithm and each dataset.

dataset	skmeans	sfcm	ssfcm	esfcm
GaussK6	0.80 ± 0.1	**0.91** ± 0.1	**0.91** ± 0.1	0.78 ± 0.1
Glass	/	**0.48** ± 0.1	**0.48** ± 0.1	0.41 ± 0.2
Ionosphere	/	**0.46** ± 0.0	**0.46** ± 0.0	0.10 ± 0.1
Iris	/	**0.75** ± 0.0	**0.75** ± 0.0	0.68 ± 0.1
Letters	/	0.21 ± 0.1	0.21 ± 0.1	**0.22** ± 0.1
Vehicle	/	0.06 ± 0.0	0.06 ± 0.0	**0.16** ± 0.0
Wine	0.82 ± 0.2	**0.90** ± 0.0	**0.90** ± 0.0	**0.90** ± 0.0

Single labels In this experiment, we assume that each constraint is expressed on a single cluster label with a specific membership value μ. Table 5 describes, for all the datasets, the performances of the algorithms when $\mu = 1$ (like in any traditional crisp seed-based semi supervised clustering) or $\mu = 0.5$. Figure 2(a) depicts the evolution of the ARI varying with the percentage of constraints. Results with $\mu = 0.2$ are similar to those with $\mu = 0.5$ and thus are not reported.

As expected, when provided contraints are correct, adding constraints enables the clustering algorithms to improve their accuracies and a membership on the constraint label equal to 1 achieves better results than a membership equal to 0.5.

As a general manner, esfcm and sfcm outperform the ssfcm algorithm although ssfcm holds better results than esfcm without constraints. As a matter of fact for ssfcm, constraints are not taken into account to compute the new centers, reducing indirectly its capacity to take into account an harmonious solution encompassing both constrained and unconstrained objects.

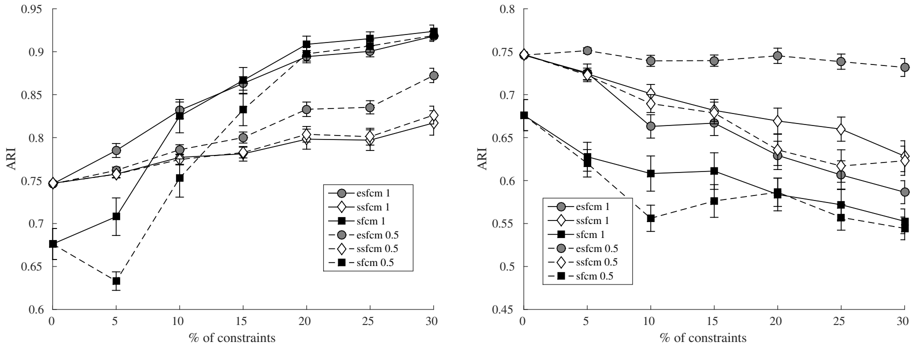
Single labels with noise Noise effect is studied by randomly modifying the labels of 20% of the constrained objects so as to produce erroneous annota-

Table 5. Single label constraints: average ARI for each algorithm and each dataset containing 30% of single label constraints with membership $\mu = 1$ or $\mu = 0.5$.

	$\mu = 1$				$\mu = 0.5$		
dataset	skmeans	sfcm	ssfcm	esfcm	sfcm	ssfcm	esfcm
GaussK6	0.99	0.99	0.99	0.99	0.99	0.99	0.99
Glass	/	**0.75**	0.56	0.74	0.61	0.47	**0.65**
Ionosphere	/	**0.59**	0.50	0.46	**0.59**	0.56	0.44
Iris	/	**0.92**	0.82	**0.92**	0.87	0.83	**0.92**
Letters	/	0.69	0.39	**0.73**	0.48	0.33	**0.69**
Vehicle	/	0.48	0.13	**0.53**	0.31	0.13	**0.42**
Wine	**0.93**	**0.93**	0.91	**0.93**	0.92	0.92	**0.93**

tions. In the end, 6% of the constraints are incorrect, 24% have the correct label and the rest is unconstrained. Table 6 and Figure 2(b) presents, with the same parametrization as before, the results with misconstrained objects.

Fig. 2. Average ARI and 95% confidence intervals on the Iris dataset as a function of the percentage of (a) not noisy (b) noisy constraints for sfcm, ssfcm and esfcm. Continuous lines represent constraints with membership $\mu = 1$ and dotted lines constraints with $\mu = 0.5$.



(a) (b)

These results show that as a general manner, noisy sets of labels generate lower quality solutions compared to labels constraints without noise. However, the sfcm algorithm is still able to reach a better accuracy than FCM (when there is no constraint). Indeed, sfcm can adjust to which extent it will respect the constraints. Thus, sfcm has a flexibility to ignore some constraints if it enables to keep a coherent overall structure. Inversely, esfcm and ssfcm force the total respect of the constraints, leading to a drop in performances in the presence of noise.

Table 6. Single label constraints with noise: average ARI for each algorithm and each dataset containing 30% of single label constraints with membership $\mu = 1$ or $\mu = 0.5$. Here 20% of the constraints are mislabeled.

	$\mu = 1$				$\mu = 0.5$		
dataset	skmeans	sfcm	ssfcm	esfcm	sfcm	ssfcm	esfcm
GaussK6	0.84	**0.85**	**0.85**	0.84	**0.98**	0.84	0.84
Glass	/	**0.55**	0.35	0.51	**0.67**	0.33	0.43
Ionosphere	/	**0.39**	0.34	0.23	**0.49**	0.38	0.26
Iris	/	0.59	**0.63**	0.55	**0.73**	0.62	0.54
Letters	/	**0.45**	0.28	0.43	0.38	0.25	**0.42**
Vehicle	/	0.31	0.09	**0.38**	0.21	0.09	**0.33**
Wine	**0.75**	**0.75**	0.74	**0.75**	**0.87**	0.75	0.75

The sfcm algorithm with noisy labels has a better accuracy than FCM in two situations. The first situation happens when the overall structure of a dataset is difficult to retrieve without constraints. It is for example the case for Vehicle or Letters, where the ARI without constraints is low. Consequently, the constraints, even a little noisy, enable to lead the algorithm towards a totally different solution, improving the accuracy. In the second situation, when constraints are uncertain (ie. with membership strictly below 1), it let sfcm more degrees of freedom to make a choice amongst the constraints in order to preserve a coherent overall structure.

Double labels In real-life use-case, an other source of erroneous annotations comes from an expert hesitating between two labels.The following experiment models such situation by setting for each object a pair of constraints on membership values for two classes. This pair of values indicates to some extent the degree of certainty of the expert for these two class labels. We simulate two distinct cases: one with membership values $\xi = (0.5, 0.5)$ where the expert is sure that one of the two labels is correct and $\xi = (0.2, 0.2)$ that indicates that the choice of the expert is not certain.As Glass and Ionosphere datasets only contains two classes, they are discarded from this experiment.

Table 7 and Figure 3 illustrate the results of both experimentations. Most of the time, the sfcm algorithm outperforms esfcm and ssfcm. While sfcm works better with membership values set to $\xi = (0.5, 0.5)$, esfcm and ssfcm often achieves higher accuracies with lower membership values. Indeed, sfcm has the ability to disrespect constraints when the solution moves too far away from a coherent choice for an overall structure, whereas esfcm and ssfcm are directly incorporating the constraints membership values in the fuzzy partition.

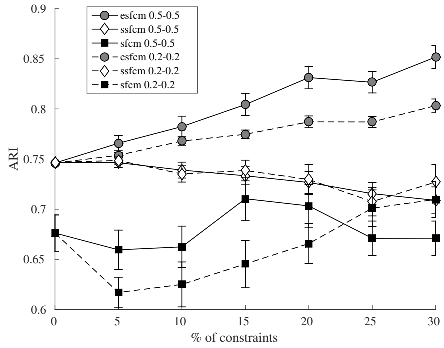
5 Conclusion

In this paper, we propose to use fuzzy algorithms to handle erroneous or uncertain expert annotations for the semi-supervised clustering problem. For the

Table 7. Double labels constraints: average ARI for each algorithm and each dataset with 30% of constraints with either $\xi = (0.5, 0.5)$ or $\xi = (0.2, 0.2)$.

	$\xi = (0.5, 0.5)$			$\xi = (0.2, 0.2)$		
dataset	sfcm	ssfcm	esfcm	sfcm	ssfcm	esfcm
GaussK6	**0.99**	0.89	0.88	**0.99**	0.95	0.98
Iris	**0.85**	0.71	0.67	**0.80**	0.73	0.71
Letters	**0.63**	0.32	0.55	0.38	0.30	**0.55**
Vehicle	0.43	0.10	**0.44**	0.22	0.11	**0.34**
Wine	**0.92**	0.79	0.81	**0.92**	**0.92**	0.91

Fig. 3. Double labels constraints: average ARI and 95% confidence intervals as a function of the percentage of constraints for sfcm, ssfcm and esfcm on the Iris dataset. Continuous lines represent constraints $\xi = (0.5, 0.5)$ while dotted lines corresponds to $\xi = (0.2, 0.2)$.



sake of clarity, we restrict our study to three main fuzzy semi-supervised algorithms. In order to make the comparison fair, each algorithm has been either corrected or improved with Mahalanobis distance to ensure comparable performances on all our test datasets. Moreover, we propose a first mapping function that solves the mismatch problem that may occur between labels defined by the initial cluster centers and labels defined in the constraints set. This mapping function although fully functional needs to be optimized, eventually based on a Hungarian algorithm.

Several scenarios are introduced to represent the variety of causes of annotation errors by an expert: either a wrong label, a low confidence in the chosen label or an hesitation between two labels.

We observed that sfcm reaches the more stable results with a good accuracy and esfcm obtains high accuracies only when labels constraints are certain. The ssfcm algorithm often does not achieve good performances. Such results can be explained by the fact that sfcm allows to violate constraints in the final solution whereas esfcm and ssfcm prohibit this behavior.

In our opinion, the major interest of fuzzy semi-supervised algorithms is their ability to handle constraints with a degree of certainty. In case of noise, lowering the labels confidence enables to keep a good improvement of the accuracy when compared to unsupervised clustering algorithm. A perspective is to investigate the addition of labels constraints in other soft clustering algorithms, that generates for instance possibilistic partitions.

References

1. Basu, S., Davidson, I., Wagstaff, K.: Constrained clustering: Advances in algorithms, theory, and applications. Chapman & Hall/CRC (2008)
2. Bouchachia, A., Pedrycz, W.: Enhancement of fuzzy clustering by mechanisms of partial supervision. Fuzzy Sets and Systems **157**(13) (2006) 1733–1759
3. Antoine, V., Quost, B., Masson, M.H., Denœux, T.: Evidential clustering with instance-level constraints for proximity data. Soft Computing **18**(7) (2014) 1321–1335
4. Basu, S., Banerjee, A., Mooney, R.: Active semi-supervision for pairwise constrained clustering. In: Proc. of the 2004 SIAM Inter. Conference on Data Mining. (2004) 333–344
5. Bilenko, M., Basu, S., Mooney, R.J.: Integrating constraints and metric learning in semi-supervised clustering. In: Proc. of the 21st ICML. (2004)
6. Wagstaff, K.L.: When is constrained clustering beneficial, and why. In: AAAI. (2006)
7. Vu, V., Labroche, N., Bouchon-Meunier, B.: Boosting clustering by active constraint selection. In: Proc. of the 2010 19th ECAI. (2010) 297–302
8. Vu, V., Labroche, N., Bouchon-Meunier, B.: An efficient active constraint selection algorithm for clustering. In: 20th ICPR. (2010) 2969–2972
9. Zhang, D., Tan, K., Chen, S.: Semi-supervised kernel-based fuzzy c-means. In: Neural Information Processing, Springer (2004) 1229–1234
10. Pedrycz, W., Waletzky, J.: Fuzzy clustering with partial supervision. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on **27**(5) (1997) 787–795
11. Lai, D., Garibaldi, J.: A comparison of distance-based semi-supervised fuzzy c-means clustering algorithms. In: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). (2011) 1580–1586
12. Bezdek, J.: Pattern recognition with fuzzy objective function algorithms. Advanced applications in pattern recognition (1981)
13. Gustafson, D., Kessel, W.: Fuzzy clustering with a fuzzy covariance matrix. In: IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes. (1979) 761–766
14. Endo, Y., Hamasuna, Y., Yamashiro, M., Miyamoto, S.: On semi-supervised fuzzy c-means clustering. In: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). (2009) 1119–1124
15. Basu, S., Banerjee, A., Mooney, R.: Semi-supervised clustering by seeding. In: Proc. of the 19th International Conference on Machine Learning (ICML). (2002) 27–34
16. Basu, S., Bilenko, M., Banerjee, A., Mooney, R. In: Probabilistic semi-supervised clustering with constraints. Cambridge, MA. MIT Press (2006) 71–98
17. Rand, W.: Objective criteria for the evaluation of clustering methods. Journal of the American Statistical association **66**(336) (1971) 846–850
18. Dave, R.: Validating fuzzy partitions obtained through c-shells clustering. Pattern Recognition Letters **17**(6) (1996) 613–623