

Semi-supervised collaborative clustering with partial background knowledge

Germain Forestier, Cédric Wemmert and Pierre Gañçarski
LSIIT - CNRS - University of Strasbourg - UMR 7005
Pôle API, Bd Sébastien Brant - 67412 Illkirch, France
Email: {forestier,wemmert,gancarski}@lsiit.u-strasbg.fr

Abstract—In this paper we present a new algorithm for semi-supervised clustering. We assume to have a small set of labeled samples and we use it in a clustering algorithm to discover relevant patterns. We study how our algorithm works against two other semi-supervised algorithms when the data are multimodal. Then, we study the case where the user is able to produce few samples for some classes but not for each class of the dataset. Indeed, in complex problems, the user is not always able to produce samples for each class present in the dataset. The challenging task is consequently to use the set of labeled samples to discover other members of these classes, but also to keep a degree of freedom to discover unknown clusters, for which samples are not available. We address this problem through a series of experimentations on synthetic datasets, to show the relevance of the proposed method.

I. INTRODUCTION

During the last years, an intensive work has been done in the field of data mining, especially on data classification [1]. Two main kinds of classification algorithms have emerged: the *supervised* classification algorithms [2] and the *unsupervised* classification algorithms [3]. In supervised classification, a fixed set of classes is available and examples for each class are used to build a classification function. This classification function is a predictive model used to find the class of new available patterns. In unsupervised classification (or *clustering*), an unlabeled dataset is partitioned into groups of similar items (i.e. *clusters*), generally by optimization of an objective function, which minimizes intra-class similarity and maximizes inter-class dissimilarity.

These two strategies for data classification have qualities and drawbacks. For supervised approaches, a large set of labeled samples is often needed to build an efficient classification function, but problems like overfitting can occur. Concerning unsupervised approaches, the determination of the number of clusters [4], the method used to build the clusters [5] and the capacity to find complex shaped clusters [6] are still intensive fields of research. Consequently, many clustering algorithms have been proposed in the literature and new clustering algorithms continue to appear every year.

To address the problems related to these two approaches, a new kind of algorithms has been investigated during the last ten years under the names of *semi-supervised* classification [7] (or *semi-supervised* learning [8]) and *semi-supervised* clustering [9]. In semi-supervised classification, the training step uses unlabeled samples in addition to the available labeled

samples, which generally leads to a better definition of the classification function. The function is consequently more generalizable. In semi-supervised clustering, the partitioning task is helped by constraints provided by the user or by a small set of labeled data. The constraints can be provided under the form of *must-link* and *cannot-link* constraints [10] giving respectively the information that two instances must be grouped together or be placed in different clusters. The use of a small set of labeled data has also been investigated [9], mainly to help the initialization of clustering algorithms.

In this paper, we are focused on this last kind of algorithms. We propose a new semi-supervised clustering algorithm using a collaborative clustering approach [11], which takes advantage of labeled samples. To the best of our knowledge, barely all the semi-supervised clustering algorithms assume a coherence between the available samples and the data space (i.e. the objects of a same class must all be placed in one area in the dataspace). Unfortunately, in complex data, this assumption is not always true. Indeed, when the user provides samples, he is not always aware of the complexity of the data space, and he does not know if the samples he provides are coherent. Two objects labeled of the same class can be localized in two different area in the data space, reflecting a multimodality, inherent to complex data. An advantage of the proposed method is its capacity to deal with such examples.

Moreover, in complex problems, the user is not always able to produce samples for each class of the dataset. All of the existing semi-supervised algorithms assume to have at least few samples for *each* expected class. We claim that this assumption is too strong and not realistic in many problems. Indeed, the user could be able to produce samples for certain classes, but can also expect to discover new classes into his data. By using a supervised or semi-supervised classification algorithm, the user loses the ability to understand his data and to discover new interesting patterns.

The outline of this paper is the following. In the Section 2, we review the main works of the literature about semi-supervised clustering. Then, in Section 3 we detail the proposed approach and in Section 4, we present some experiments to evaluate this method. Finally, conclusions and future works are drawn in Section 5.

II. CLUSTERING WITH BACKGROUND KNOWLEDGE

Many approaches have been investigated to use background knowledge to guide the clustering process. In constraint clus-

tering, knowledge is expressed as *must-link* and *cannot-link* constraints. A *must-link* constraint gives the information that two data objects should be in the same cluster, and *cannot-link* means the contrary. This kind of knowledge is sometime easier to obtain than a classical subset of labeled samples.

In [10], Wagstaff et al. present a constrained version of the k-means algorithm which uses such constraints to bias the affectation of the objects to the clusters. At each step, the algorithm tries to agree with the constraints given by the user.

The constraints can be used to tune the algorithm, but also to learn a distance function which is biased by the knowledge about the constraints between the data objects, as presented in [12]. The distance between two data objects is reduced for a *must-link* and increased for a *cannot-link*. Recent works on constraint clustering are focused on evaluating the utility (i.e the potential interest) of a set of constraints [13], [14].

In [9], a subset of data objects is used to *seed* (i.e to initialize) the clusters of the k-means algorithm. Two algorithms, *seeded kmeans* and *constrained kmeans*, are presented. In the first algorithm, the samples are only used to initialize the clusters and can eventually be affected to another class during the clustering process. In the second algorithm, the clusters are initialized with the samples, which stay affected to their initial cluster. The choice between these two approaches must be done according to the knowledge about noise in the dataset.

To tackle the problem of incorporating partial background knowledge into clustering, where the labeled examples have moderate overlapping features with the unlabeled data, Gao et al. [15] formulate a new approach as a constrained optimization problem. They propose two learning algorithms to solve the problem, based on hard and fuzzy clustering methods. An empirical study shows that the proposed algorithms improve the quality of clustering results, with limited labeled examples.

In [16], a pairwise constrained clustering framework is presented, as well as a new method for actively selecting informative pairwise constraints, to get improved clustering performance. Experimental and theoretical results confirm that this active querying of pairwise constraints significantly improves the accuracy of clustering, when given a relatively small amount of supervision.

Basu et al. [17] also propose a probabilistic model for semisupervised clustering, based on Hidden Markov Random Fields (HMRF), that provides a principled framework for incorporating supervision into prototype-based clustering. Experimental results on several text data sets demonstrate the advantages of the proposed framework.

A new method, to allow instance-level constraints to have space level inductive implications to improve the use of the constraints, is given in [18]. The method improves the previously studied constrained k-means algorithm, generally requiring less than half as many constraints to achieve a given accuracy on a range of real-world data, while also being more robust when over-constrained.

In [19], the authors propose a method which uses different clustering methods using different objectives. The final result is produced by selecting clusters among the results proposed

by the different methods. A resampling method is used to estimate the quality of the clusters.

Another approach called supervised clustering [20], uses the feature of the data objects as well as the class information to build clusters with a high class purity. The goal of supervised clustering is to identify class-uniform clusters that have high probability densities.

Different kinds of background knowledge are introduced in [21], namely partial supervision, proximity-based guidance and uncertainty driven knowledge hints. The authors discuss the ways to exploit and effectively incorporate these background knowledge (called *knowledge hints*) in the fuzzy c-means algorithm.

III. SEMI-SUPERVISED COLLABORATIVE CLUSTERING

The proposed approach is divided in two main steps: a collaborative clustering of the data based on an existing collaborative clustering method [11], and a cluster labeling process which affects a class to each cluster, thanks to a subset of available samples.

In the first step, different clustering algorithms are applied on the data, with different initializations and different parameters. These clustering results are then refined through a collaborative process to find an agreement about the classification of the data. This collaborative approach is used to discover the structure of the dataset. Indeed, by using different methods with different number of clusters, the methods will automatically find a good estimation of the number of clusters and the final classification will reflect the structure of the data. In the second step, a cluster labeling process is used to affect a class to each cluster.

In the next section, we describe the collaborative clustering process and then, the clusters labeling step.

A. Collaborative clustering

It is difficult to compute a consensual result from clustering results with different numbers of clusters or different structures (flat partitioning or hierarchical result), because of the lack of a trivial correspondence between the clusters of these different results. To address the problem, we present in this section a framework where different clustering methods work together in a collaborative way, to find an agreement about their proposals.

This collaborative process consists in an automatic and mutual refinement of the clustering results, until all the results have almost the same number of clusters, and all the clusters are statistically similar with a good internal quality. At the end of this process, as the results have comparable structures, it is possible to define a correspondence function between the clusters, and to apply a unifying technique, such as a voting method [22].

Before the description of the collaborative method, we introduce the correspondence function which is used in the system.

1) *Intercluster correspondence function*: There is no problem to associate classes of different supervised classifications, as a common set of class labels is given for all the classifications. Unfortunately, in the case of unsupervised classifications, the results may not have a same number of clusters, and no information is available about the correspondence between the different clusters of the different results.

To address the problem, we have defined a new intercluster correspondence function, which associates to each cluster from a result, a cluster from each of the other results.

Let $\{\mathcal{R}^i\}_{1 \leq i \leq m}$ be the set of results given by the different algorithms.

Let $\{C_k^i\}_{1 \leq k \leq n_i}$ be the clusters of the result \mathcal{R}^i .

The *corresponding cluster* $CC(C_k^i, \mathcal{R}^j)$ of a cluster C_k^i from \mathcal{R}^i in the result \mathcal{R}^j , $i \neq j$, is the cluster from \mathcal{R}^j which is the most similar to C_k^i :

$$CC(C_k^i, \mathcal{R}^j) = C_\ell^j \quad (1)$$

with $S(C_k^i, C_\ell^j) = \max(\{S(C_k^i, C_l^j), \forall l \in [1, n_j]\})$

where S is the *intercluster similarity* which evaluates the similarity between two clusters of two different results.

It is calculated from the recovery of the clusters in two steps. First, the intersection between each couple of clusters (C_k^i, C_l^j) , from two different results \mathcal{R}^i and \mathcal{R}^j , are calculated and stored in the *confusion matrix* $\mathcal{M}^{i,j}$:

$$\mathcal{M}^{i,j} = \begin{pmatrix} \alpha_{1,1}^{i,j} & \cdots & \alpha_{1,n_j}^{i,j} \\ \vdots & \ddots & \vdots \\ \alpha_{n_i,1}^{i,j} & \cdots & \alpha_{n_i,n_j}^{i,j} \end{pmatrix} \text{ where } \alpha_{k,l}^{i,j} = \frac{|\{C_k^i \cap C_l^j\}|}{|C_k^i|} \quad (2)$$

Then, the similarity $S(C_k^i, C_l^j)$ between two clusters C_k^i and C_l^j is evaluated by observing the relationship between the size of their intersection and the size of the cluster itself, and by taking into account the distribution of the data in the other clusters as follows:

$$S(C_k^i, C_l^j) = \alpha_{k,l}^{i,j} \alpha_{l,k}^{j,i} \quad (3)$$

2) *Collaborative process overview*: The entire clustering process is broken down in three main phases:

- 1) *Initial clusterings* - Each clustering method computes a clustering of the data using its parameters.
- 2) *Results refinement* - A phase of convergence of the results, which consists of conflicts evaluation and resolution, is iterated as long as the quality of the results and their similarity increase.
- 3) *Unification* - If needed, the refined results are unified using a voting algorithm.

a) *Initial clusterings*.: During this first step, each clustering method is initialized with its own parameters and a clustering is performed: all data objects are grouped into different clusters.

b) *Results refinement*.: The mechanism we propose for refining the results is based on the concept of distributed local resolution of conflicts, by the iteration of four phases:

- Detection of the conflicts by evaluating the dissimilarities between couples of results;
- Choice of the conflicts to solve;
- Local resolution of these conflicts;
- Management of the local modifications in the global result (if they are relevant).

Conflicts detection - The detection of the conflicts consists in seeking all the couples (C_k^i, \mathcal{R}^j) , $i \neq j$, such as $C_k^i \neq CC(C_k^i, \mathcal{R}^j)$. One conflict $\mathcal{K}_k^{i,j}$ is identified by one cluster C_k^i and one result \mathcal{R}^j .

We associate to each conflict a measurement of its importance, the *conflict importance coefficient*, calculated according to the intercluster similarity:

$$CI(\mathcal{K}_k^{i,j}) = 1 - S(C_k^i, CC(C_k^i, \mathcal{R}^j)) \quad (4)$$

Choice of the conflicts to solve - During an iteration of refinement of the results, several local resolutions are performed in parallel. A conflict is selected in the set of existing conflicts and its resolution is started. This conflict, like all those concerning the two results involved in the conflict, are removed from the list of the conflicts. This process is iterated, until the list of the conflicts is empty.

Different heuristics can be used to choose the conflict to solve, according to the conflict importance coefficient (Eq. 4). We choose to try to solve the most important conflict first.

Local resolution of a conflict - The local resolution of a conflict $\mathcal{K}_k^{i,j}$ consists of applying an operator on each result involved in the conflict, \mathcal{R}^i and \mathcal{R}^j , to try to make them more similar.

The operators that can be applied to a result are the following:

- *merging* of clusters: some clusters are merged together (all the objects are merged in a new cluster that replaces the clusters merged),
- *splitting* of a cluster in subclusters: a clustering is applied to the objects of a cluster to produce subclusters,
- *reclustering* of a group of objects: one cluster is removed and its objects are reclassified in all the other existing clusters.

The operator to apply is chosen according to the corresponding clusters of the cluster involved in the conflict. The *corresponding clusters* (CCS) of a cluster is an extension of the definition of the corresponding cluster (Eq. 1).

$$CCS(C_k^i, \mathcal{R}^j) = \{C_l^j | S(C_k^i, C_l^j) > p_{cr}, \forall l \in [1, n_j]\} \quad (5)$$

where p_{cr} , $0 \leq p_{cr} \leq 1$, is given by the user.

Having found the corresponding clusters of the cluster involved in the conflict, an operator is chosen and applied as follows:

Algorithm 1 Operator application

```

1: let  $n = |CCS(C_k^i, \mathcal{R}^i)|$ 
2: let  $\mathcal{R}^{i'}$  (resp.  $\mathcal{R}^{j'}$ ) be the result of the application of an
   operator on  $\mathcal{R}^i$  (resp.  $\mathcal{R}^j$ )
3: if  $n > 1$  then
4:    $\mathcal{R}^{i'} = \mathcal{R}^i \setminus \{C_k^i\} \cup \{split(C_k^i, n)\}$ 
5:    $\mathcal{R}^{j'} = \mathcal{R}^j \setminus CCS(C_k^i, \mathcal{R}^j) \cup \{merge(CCS(C_k^i, \mathcal{R}^j))\}$ 
6: else
7:    $\mathcal{R}^{i'} = reclustering(\mathcal{R}^i, C_k^i)$ 
8: end if

```

But the application of the two operators is not always relevant. Indeed, it does not always increase the similarity of the results implied in the conflict treated, and especially, the iteration of conflict resolutions may lead to a trivial solution where all the methods are in agreement. For example, they can converge towards a result with only one cluster including all the objects to classify, or towards a result having one cluster for each object. These two solutions are not relevant and must be avoided.

So we defined a criterion γ , called *local similarity criterion*, to evaluate the similarity between two results, based on the intercluster similarity S (Eq. 3) and a quality criterion δ :

$$\gamma^{i,j} = \frac{1}{2} \left(p_s \cdot \left(\frac{1}{n_i} \sum_{k=1}^{n_i} \omega_k^{i,j} + \frac{1}{n_j} \sum_{k=1}^{n_j} \omega_k^{j,i} \right) + p_q \cdot (\delta^i + \delta^j) \right) \quad (6)$$

where

$$\omega_k^{i,j} = \sum_{l=1}^{n_j} S(C_k^i, CC(C_k^i, \mathcal{R}^j)) \quad (7)$$

and, p_q and p_s are given by the user ($p_q + p_s = 1$).

At the end of each conflict resolution, the local similarity criterion enables to choose which couple of results are to be kept: the two new results, the two old results, or one new result with one old result.

Global management of the local modifications - After the resolutions of all these local conflicts, a global application of the modifications proposed by the refinement step, is decided if it improves the quality of the global result. The *global agreement coefficient* of the results is evaluated according to all the local similarity between each couple of results. It evaluates the global similarity of the results and their quality.

$$\Gamma = \frac{1}{m} \sum_{i=1}^m \Gamma^i \quad (8)$$

where

$$\Gamma^i = \frac{1}{m-1} \sum_{\substack{j=1 \\ j \neq i}}^m \gamma^{i,j} \quad (9)$$

Even if the local modifications decrease this global agreement coefficient, the solution is accepted to avoid to fall in a

local maximum. If the coefficient is decreasing too much, all the results are reinitialized to the best temporary solution (the one with the best global agreement coefficient). In this case, the conflicts that have not been solved are removed to avoid oscillations of the agreement coefficient. The global process is iterated until some conflicts can be solved.

c) *Unification.*: In the final step, all the results tend to have the same number of clusters, which are increasingly similar. Thus, we use a voting algorithm [22] to compute a unified result combining the different results. This multi-view voting algorithm enables to combine in one unique result, many different clustering results that have not necessarily the same number of clusters.

The basic idea is that for each object to cluster, each result \mathcal{R}^i votes for the cluster it has found for this object, C_k^i for example, and for the corresponding cluster of C_k^i in all the other results. The maximum of these values indicates the *best cluster* for the object, for example C_l^j . That means that this object should be in the cluster C_l^j according to the opinion of all the methods. After having done the vote for all objects, a new cluster is created for each *best cluster* found if a majority of the methods have voted for this cluster.

B. Clusters labeling

After the collaborative clustering and the unification of the different results, we obtain a clustering result \mathcal{R}^u composed of n_u clusters $\{C_k^u\}_{1 \leq k \leq n_u}$.

Let $\mathcal{S} = \{s_l\}_{1 \leq l \leq n_s}$ be the set of samples available for the dataset. For each sample s_l of \mathcal{S} , the known class (label) is given by $cl(s_l)$.

To label the clusters of \mathcal{R}^u , we apply a voting method as follow:

$$cl(C_k^u) = \arg \max_k \{|C_k^u \cap \mathcal{S}^k|\} \quad (10)$$

where

$$\mathcal{S}^k = \{s_l \in \mathcal{S} : cl(s_l) = k\} \quad (11)$$

If no sample is available in a cluster, that is

$$\forall k, C_k^u \cap \mathcal{S}^k = \emptyset$$

a new label is given to this cluster. This enables to discover new patterns in the data.

IV. EXPERIMENTAL RESULTS

In this section, we present different experiments carried out to illustrate how the method works. In the first experiment, we used three synthetic datasets (see Fig. 1), each containing two classes. In *Dataset 1*, the classes are well separable in the data space, while in *Dataset 2* one of the classes is bimodal and is composed of two clusters. Finally, in *Dataset 3*, the two classes are multimodal and are composed of separated groups of objects in the data space.

For this experiment we used the KMeans algorithm as the base method for the collaborative system. We set up the method with four different initializations of the KMeans

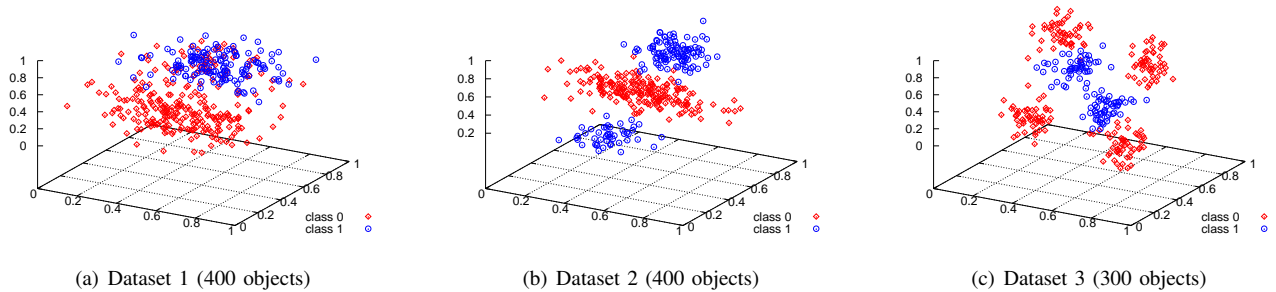


Fig. 1. The three dataset used for the experiments.

TABLE I
RESULTS ON THE FOUR DATASETS.

dataset	Proposed	SKMeans	CKMeans
<i>Dataset 1</i>	96.75(± 1.43)	97.75(± 0.00)	97.96(± 0.22)
<i>Dataset 2</i>	98.36(± 0.48)	61.81(± 7.21)	62.75(± 7.50)
<i>Dataset 3</i>	91.18(± 10.81)	55.56(± 0.69)	55.57(± 0.75)
<i>Dataset 4</i>	97.98(± 3.87)	70.3(± 5.09)	67.12(± 3.64)

algorithm with different numbers of clusters, randomly picked in [2, 10]. Then, the collaborative process presented in Section 3 is run to find an agreement among these different methods, which will reflect the structure of the dataset. Finally, the cluster labeling process (Section III-B) is used to affect a class to each cluster (if it is possible).

The Tab. I presents the accuracy obtained after the application of our clustering method. The values are the average and standard deviation of 100 runs. For comparison purpose, we also applied two different semi-supervised clustering algorithms, namely *seeded kmeans* (SKMeans) and *constrained kmeans* (CKMeans) [9]. These two algorithms were set up to find two clusters as they are set up according to the available samples. For this experiment, 5% of the dataset was given as background knowledge, randomly chosen at each run.

One can see on Tab. I that when the dataset is simple (*Dataset 1*), the methods are almost equivalent. However, when the data are more complex and multimodal, classical semi-supervised algorithm logically failed to discover the underlying structure of the classes. In these cases (*Dataset 2 and 3*), the proposed method clearly outperformed the two others.

The Fig. 2 shows the accuracy according to the percentage of available samples. As one can see, the proposed method shows a great stability, the accuracy being stable from 5% to 50% of available samples.

Another experiment has been made, when samples are available for some classes but not for all. The Fig. 3 presents the *Dataset 4*, which is similar to the *Dataset 2* with the addition of one unknown cluster. The challenge here, is to use the available samples to identify the two classes, but also to discover the third cluster, corresponding to a hidden pattern, where no background knowledge is available. The

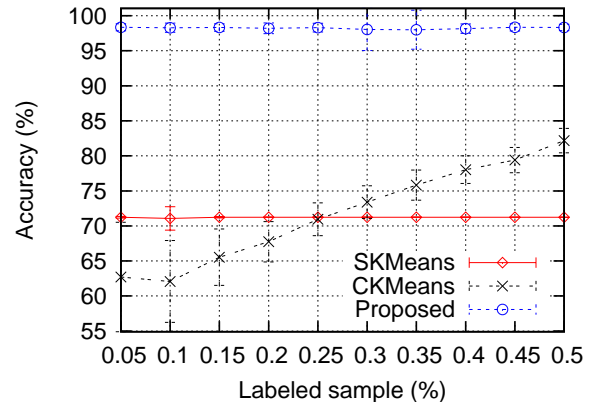


Fig. 2. The accuracy according to available sample on *Dataset 2*.

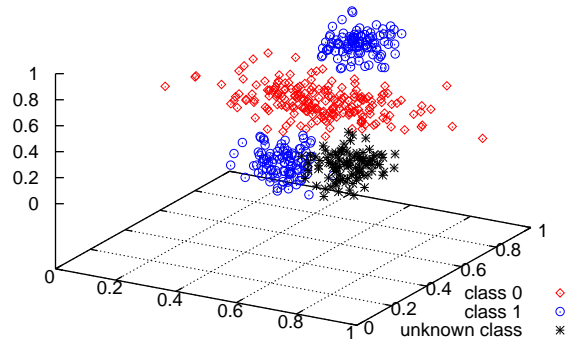


Fig. 3. The *Dataset 4* (500 objects) which contains one unknown cluster.

results for the *Dataset 4* are presented in Tab. I. Here again, the proposed method outperforms the others to discover interesting patterns for the user, but also to identify unknown underlying knowledge.

V. CONCLUSIONS

In this paper, we have presented a new semi-supervised clustering algorithm. This algorithm is decomposed in two

steps. Firstly a collaborative clustering of the data reveals the underlying structure of the dataset. Secondly, a clusters labeling, using a small subset of available samples, affects a class to each cluster.

The method is able to use samples provided by the user, which are not necessarily grouped together in the data space. This kind of knowledge can occur in mining complex multimodal data. The user is not always aware of the shape of the data space and can consequently provide sample placed in various places of the dataset for the same class. Furthermore, the method is able to use background knowledge provided by the user, but can keep a degree of freedom to discover hidden patterns. This feature allows the users to use the knowledge they have on their data but also to discover relevant unknown patterns.

In future works, we plan to improve the method through different modifications. For example, a problem occurs when no available sample is present in a cluster, consequently this cluster can not be labeled. This problem can be solved by setting to these clusters the labels of the nearest labeled cluster (if all the clusters need to be labeled). Another modification will consist to label the clusters of each result before the unification at the end of the collaboration process. A classical supervised voting method could then be used to affect a class to each object. This approach should be compare to the one presented in this paper.

Finally, more experiments will be conducted, especially on real life problems, as for example object-based remote sensing image interpretation.

REFERENCES

- [1] I. H. Witten and F. Eibe, *Data mining : practical machine learning tools and techniques*. Morgan Kaufmann, 2005.
- [2] Z. Fan-Zi and Q. Zheng-Ding, "A survey of classification learning algorithm," *International Conference on Signal Processing*, vol. 2, pp. 1500–1504, 2004.
- [3] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999. [Online]. Available: <http://www.cs.rutgers.edu/~mlittman/courses/lightai03/jain99data.pdf>
- [4] R. Kothari and D. Pitts, "On finding the number of clusters," *Pattern Recognition Letters*, vol. 20, no. 4, pp. 405–416, 1999.
- [5] R. Xu and D. Wunsch, "Survey of clustering algorithms," *Neural Networks, IEEE Transactions on*, vol. 16, no. 3, pp. 645–678, 2005.
- [6] A. Fred and A. Jain, "Combining multiple clusterings using evidence accumulation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 835–850, 2005.
- [7] B. Krishnapuram, D. Williams, Y. Xue, A. J. Hartemink, L. Carin, and M. A. T. Figueiredo, "On semi-supervised classification." in *Neural Information Processing Systems*, 2004.
- [8] X. Zhu, "Semi-supervised learning literature survey," Computer Sciences, University of Wisconsin-Madison, Tech. Rep. 1530, 2005.
- [9] S. Basu, A. Banerjee, and R. J. Mooney, "Semi-supervised clustering by seeding," in *International Conference on Machine Learning*, 2002, pp. 19–26.
- [10] K. Wagstaff, C. Cardie, S. Rogers, and S. Schroedl, "Constrained k-means clustering with background knowledge," in *International Conference on Machine Learning*, 2001, pp. 557–584.
- [11] P. Gancarski and C. Wemmert, "Collaborative multi-step mono-level multi-strategy classification," *Journal on Multimedia Tools and Applications*, vol. 35, no. 1, pp. 1–27, october 2007.
- [12] M. Bilenko, S. Basu, and R. J. Mooney, "Integrating constraints and metric learning in semi-supervised clustering," in *International Conference on Machine Learning*, 2004, pp. 81–88.
- [13] I. Davidson, K. L. Wagstaff, , and S. Basu, "Measuring constraint-set utility for partitional clustering algorithms," in *European Conference on Principles and Practice of Knowledge Discovery in Databases*, 2006, pp. 115–126.
- [14] K. L. Wagstaff, "Value, cost, and sharing: Open issues in constrained clustering," in *International Workshop on Knowledge Discovery in Inductive Databases*, 2007, pp. 1–10.
- [15] J. Gao, P.-N. Tan, , and H. Cheng, "Semi-supervised clustering with partial background information," in *SIAM International Conference on Data Mining*, 2006.
- [16] S. Basu, A. Banerjee, and R. J. Mooney, "Active semi-supervision for pairwise constrained clustering," in *SIAM International Conference on Data Mining*, 2004.
- [17] S. Basu, M. Bilenko, and R. J. Mooney, "A probabilistic framework for semi-supervised clustering," in *International Conference on Knowledge Discovery and Data Mining*, 2004, pp. 59–68.
- [18] D. Klein, S. Kamvar, and C. Manning, "From instance-level constraints to space-level constraints: Making the most of prior knowledge in data clustering," 2002.
- [19] M. Law, A. Topchy, and A. Jain, "Multiobjective data clustering," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2004, pp. 424–430.
- [20] C. F. Eick, N. Zeidat, , and Z. Zhao, "Supervised clustering - algorithms and benefits," in *International Conference on Tools with Artificial Intelligence*, 2004, pp. 774–776.
- [21] W. Pedrycz, "Fuzzy clustering with a knowledge-based guidance." *Pattern Recognition Letters*, vol. 25, no. 4, pp. 469–480, 2004.
- [22] C. Wemmert and Gancarski, "A multi-view voting method to combine unsupervised classifications," *Artificial Intelligence and Applications*, pp. 362–324, 2002.