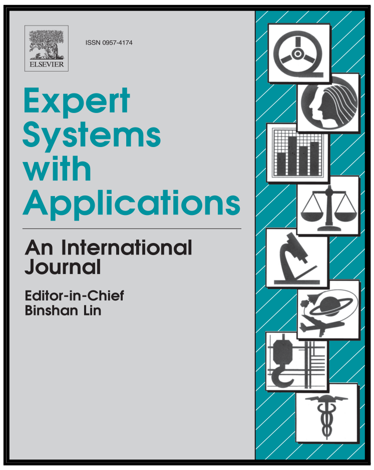
Accepted Manuscript



Local homogeneous consistent safe semi-supervised clustering

Haitao Gan, Yingle Fan, Zhizeng Luo, Qizhong Zhang

PII: S0957-4174(17)30868-0
DOI: 10.1016/j.eswa.2017.12.046
Reference: ESWA 11747

To appear in: *Expert Systems With Applications*

Received date: 9 April 2017
Revised date: 27 November 2017
Accepted date: 28 December 2017

Please cite this article as: Haitao Gan, Yingle Fan, Zhizeng Luo, Qizhong Zhang, Local homogeneous consistent safe semi-supervised clustering, *Expert Systems With Applications* (2017), doi: 10.1016/j.eswa.2017.12.046

Highlights

- We develop safe semi-supervised clustering which outperforms semi-supervised clustering.
- We construct a local homogeneous graph to safely exploit the risk prior knowledge.
- We can achieve the closed-form solution and obtain the promising results.

Local homogeneous consistent safe semi-supervised clustering

Haitao Gan[a,*], Yingle Fan[a], Zhizeng Luo[a,*], Qizhong Zhang[a]

[a]*School of Automation, Hangzhou Dianzi University, Hangzhou, China*

Abstract

Semi-supervised clustering generally assumes that prior knowledge is helpful to improve clustering performance. However, the prior knowledge may degenerate the clustering performance if one collects wrong information, such as wrong labels. Hence, it is meaningful to design a safe semi-supervised clustering method which never performs worse than the corresponding unsupervised and semi-supervised clustering methods. In this paper, we develop local homogeneous consistent safe semi-supervised clustering where class labels are given as the prior knowledge. To the best of our knowledge, it is the first time safe semi-supervised clustering has been studied. The basic idea is that the predictions of a labeled sample and its nearest homogeneous unlabeled ones should be similar when the labeled one is risky. In our algorithm, we firstly build a local graph to model the relationship between the labeled sample and its nearest homogeneous unlabeled ones through the results obtained by unsupervised clustering. A graph-based regularization term is then constructed to allow the predictions of the labeled samples to approach that of the local homogeneous neighbors. It is expected to reduce the risk of the labeled samples. Meanwhile, our algorithm positively exploits the labeled samples by restricting the corresponding outputs to be the given class labels when the labeled ones may be helpful. In this sense, the predictions of the labeled samples in our algorithm are a tradeoff between the given class labels and the predictions of local homogeneous neighbors. To verify the effectiveness of our algorithm, we conduct a series of experiments on

*Corresponding author. Telephone: +860571-86919130
E-mail addresses: htgan@hdu.edu.cn(Haitao Gan), fan@hdu.edu.cn(Yingle Fan), luo@hdu.edu.cn(Zhizeng Luo), zqz@hdu.edu.cn(Qizhong Zhang)

several UCI datasets. The results show that our algorithm outperforms the corresponding unsupervised and semi-supervised clustering methods even if the wrongly labeled ratio reaches 30%. In this sense, the proposed algorithm will not only enrich the theoretical knowledge in the machine learning field, but significantly improve the practicability of semi-supervised clustering in the expert and intelligent systems.

Keywords: Semi-supervised clustering, fuzzy c-means, safe mechanism, local homogeneous consistency

---

Response to Reviewers' comments on
Local homogeneous consistent safe semi-supervised clustering

We would like to thank the editor, reviewers and journal staffs for your time and thoughtful comments. Below are our responses to the reviewers' comments.

Authors' Response to reviewers' comments

——— *Reviewer 1* ———

This paper proposes a semi-supervised clustering method. The article is well organized;

Response: We appreciate the reviewer for the effect and suggestions. We have revised the paper according to the suggestions. Below are point-to-point responses to the reviewer's comments.

1. *however, there are many English mistakes and several parts of the text that are not clear. For example, in the abstract "it is the first time to study safe semi-supervised clustering" should be corrected.*

Response: We have carefully revised the paper including the abstract.

2. *The concept of safe semi-supervised should be better explained, even though this term have been used in semi-supervised learning, it is not common for clustering so a formal definition should be provided. Also the concept of risk.*

Response: We have given the definitions of safe semi-supervised clustering and risk. Prior knowledge is risky which means that it will degenerate the clustering performance, such as wrong labels. Therefore, safe semi-supervised clustering mainly focuses on how to reduce the risk of the prior knowledge. Safe semi-supervised clustering is expected to be superior to the corresponding unsupervised and semi-supervised clustering, especially when semi-supervised clustering performs worse than unsupervised clustering.

3. *In Section 3.1, revise the sentence "The traditional semi-supervised clustering assumes that the labeled samples are always benefit to the performance improvement," specifically in "always benefit", should be: "always beneficial "*

Response: We have revised the sentence.

4. *It is not clear why the uncorrected labeled sample cannot be evaluated in a pre-processing step to assure the quality of the data, instead of this it has to be done in the clustering method. A comparison between this approaches should be provided.*

Response: As pointed outed by the reviewer, one can assure the quality of the data by evaluating the uncorrected labeled samples. For example, the wrongly labeled samples can be identified and then removed as commonly used in supervised learning [1-4]. However, there are some drawbacks of this approach as follows: (1) The performance of the evaluation or identification approach generally relies on the amount of labeled samples. When we only have a few labeled samples, it will be not guaranteed that the identification approach can correctly identify the wrongly labeled samples. (2) If we employ the identification approach to evaluate the labeled samples, how to design an appropriate evaluation criterion is the key problem which we have to confront. (3) The result of the identification approach will directly affect the clustering performance. If the correctly labeled samples are identified to be the incorrect ones, it will further degenerate the clustering performance. (4) The identification approach will increase computation complexity of the algorithm. However, a graph-based regularization term is constructed to safely exploit the risky labeled samples and embedded into SSFCM_s in our algorithm. Hence, the safe exploitation of prior knowledge and clustering implementation are unified in a framework. Meanwhile, the optimization problem is easy to be solved. Overall, it is an interesting topic to employ an identification approach to evaluate the wrongly labeled samples and it will be our future work.

5. *Also, it is not clear when the proposed algorithm will perform better or worse than the compared semi-supervised methods? And why? An analysis of the results should be provided. For example, in the Balance-scale dataset, the proposed method is a lot better, but that is not true for the Ionosphere, so what is the reason for that? For the Mushroom dataset, the proposed method is also clearly superior for ratio 0%? Why? And the other method shows a flat performance in relation to the ratio(%). This more detailed analysis should be provided.*

Response: our algorithm can achieve much better clustering performance than SSFCM_h and SSFCM_s on Balance-scale and Mushroom. The reason may be that some correctly labeled samples are outliers and lie in the boundaries between different clusters in the two datasets. Additionally, SSFCM_h and SSFCM_s show a flat performance as the wrongly labeled ratio increases from 0%-30% on Mushroom and Waveform. The reason may be that the scale of the two datasets is larger than that of the other datasets. A relatively small number of the wrongly labeled samples have little effect on performance degeneration.

6. *In step 2 of the Algorithm 1 2. Contract the local graph w_{kr}, how this is done? And how this graph is used in the next steps?*

Response: We have revised the Algorithm 1. The graph is constructed through Eq.(10) and will be used to update u_{ik} and J_{sa}.

[1] Brodley, C. E., & Friedl, M. A. (1996). Identifying and eliminating mislabeled training instances. In Proceedings of the National Conference on Artificial Intelligence (pp. 799-805).

[2] Brodley, C. E., & Friedl, M. A. (1999). Identifying mislabeled training data. Journal of artificial intelligence research, 11, 131-167.

[3] Zeng, X., & Martinez, T. R. (2001). An algorithm for correcting mislabeled data. Intelligent data analysis, 5(6), 491-502.

[4]Venkataraman, S., Metaxas, D., Fradkin, D., Kulikowski, C., & Muchnik, I. (2004). Distinguishing mislabeled data from correctly labeled data in classifier design. In Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence (pp. 668-672). IEEE.

——————————— *Reviewer 2* ———————————

The topic is interesting. Even though the article is interesting in its current format, some aspects should be improved for possible publication and for a better understanding by the readers.

Response: We appreciate the review for the effect and suggestions. We have revised the paper according to the suggestions. Below are point-to-point responses to the reviewer's comments.

1. *The authors should give the readers some concrete information to get them excited about their work. The current abstract only describes the general purposes of the article. It should also include the article's main (1) impact and (2) significance on expert and intelligent systems.*

Response: We have modified the abstract to point out the significance and impact of the paper. Firstly, it is meaningful to design a safe semi-supervised clustering method which never performs worse than the corresponding unsupervised and semi-supervised clustering methods. To the best of our knowledge, it is the first time safe semi-supervised clustering has been studied. Secondly, the proposed algorithm will not only enrich the theoretical knowledge in the machine learning field, but significantly improve the practicability of semi-supervised clustering in the expert and intelligent systems.

2. *Please give a frank account of the strengths and weaknesses of the proposed research method. This should include theoretical comparison to other approaches in the field.*

Response: We have given a comparison to show the strengths and weaknesses of our algorithm in Section 1. Compared to the existing semi-supervised clustering methods, the advantages of our algorithm include: (1) The risky labeled samples are safely not freely exploited through local homogeneous consistency in our algorithm. (2) The proposed graph-based regularization term can be easily embedded into the other objective function-based semi-supervised clustering methods. The drawback of our algorithm is that a new regularization parameter λ_2 is introduced.

3. *In the related work section, a more rigorous investigation on the existing methods, such as comparison of previous approaches in terms of pros and cons, should be given. A summary table can be used in this regard.*

Response: In Section 2, we have presented a comparison of FCM, SSFCM_h and SSFCM_s in terms of pros and cons. In FCM, the optimization problem can be easily solved for $m > 1$. However, FCM is an unsupervised clustering method which does not make full use of the prior knowledge. SSFCM_h and SSFCM_s belong to semi-supervised clustering by considering the prior knowledge. In general, the two semi-supervised clustering methods can yield better clustering performance than FCM. In SSFCM_h, the labeled samples are exploited through a hard way which means that the partition matrix U^l is predefined and fixed in the iterative process. SSFCM_s employs a soft way to exploit the labeled samples by introducing a fidelity term. To obtain the closed-form solution, m is fixed to 2. It is worthy to point out that SSFCM_h can be considered as a special case of SSFCM_s with $\alpha \to +\infty$. Nevertheless, SSFCM_h and SSFCM_s do not consider the risk of the labeled samples.

4. *Moreover, I believe that it will make this paper stronger if the authors present insightful implications in at least one paragraph based on their experimental outcomes.*

Response: We have presented the implications in Section 4.3. From the above result analysis, one can see that it is necessary and meaningful to consider the risk of prior knowledge in semi-supervised clustering. Meanwhile, the proposed safe mechanism using local homogeneous consistency is effective to reduce the risk and can be used in some applications, such as document categorization. Furthermore, parameter setting (i.e., λ_1 and λ_2) is important for the practicability of our algorithm. As can be seen from the parameter analysis, the performance with respect to different parameter settings is related to the wrongly labeled ratio. Hence designing an approach to evaluate the quality of the data is a meaningful topic.

5. *What are future endeavors of your study? Please open a real window for future work in the conclusion section. The authors also need to clearly provide 4-5 solid future research directions in the Conclusion section. These directions should be written as at least a separate paragraph and such directions need to be insightful for most of ESWA community.*

Response: Our future work mainly focuses on the following research directions: (1) It is meaningful to design an approach to evaluate different risk of different labeled samples. (2) We will investigate the other forms of the prior knowledge, such as pair-wise constraints. (3) As discussed in section 4.2, the parameter setting may be related to the wrongly labeled ratio which presents the quality of the prior knowledge. Hence, it is very practical to measure the quality of the prior knowledge to guide the parameter setting. (4) We will try to employ the other intelligent optimization algorithms to solve the optimization problem where m is different from 2.

6. *Finally, the language and grammar also require some work, and I noted a large number of typographical errors. The paper needs a linguistic check, preferably by a native speaker.*

Response: We have revised the paper carefully and throughout.

1. Introduction

During the past decade, unsupervised clustering is well-known as an effective data mining tool and successfully applied in various tasks, such as image segmentation (Shi & Malik, 2000; Chuang et al., 2006; Hasnat et al., 2016; Yin et al., 2017), image categorization (Yang et al., 2010; Liu et al., 2017), document categorization (Mei et al., 2016; Xu et al., 2003) and bioinformatics (Jiang et al., 2017; Yeung et al., 2000), etc. Unsupervised clustering aims to partition a given dataset into several clusters according to a distance measure. Similar samples are expected to be clustered into the same cluster and dissimilar ones are clustered into different clusters. Many unsupervised clustering methods have been proposed, such as k-means (Hartigan & Wong, 1979), Gaussian Mixture Models (GMM) (Chen et al., 2011; Lu, 2006; Liu et al., 2010), Fuzzy c-Means (FCM) (Bezdek, 1981; Pal & Bezdek, 1995; Cannon et al., 1986), Non-negative Matrix Factorization (NMF) (Xu et al., 2003), spectral clustering (Ng et al., 2001), affinity propagation (Frey & Dueck, 2007), etc. Especially, k-means and FCM are the popular and practical unsupervised clustering methods because of their simple solution. More details about unsupervised clustering can be found in some excellent surveys (Xu & II, 2005; Bishop, 2006; Wu et al., 2007).

Although k-means and FCM can often achieve the promising clustering results, they are a kind of unsupervised clustering. In some practical applications, there are often different types of prior knowledge which may be collected by some experts and users. The common used prior knowledge includes class labels and pair-wise constraints (i.e., must-link and cannot-link). Consequently, semi-supervised clustering, which makes use of both the prior knowledge and unlabeled samples to aid the clustering, has become a meaningful and significant topic in the machine learning field. A lot of semi-supervised clustering methods (Gan et al., 2015; Zhang & Lu, 2009; Basu et al., 2002; Chen & Feng, 2012; Givoni & Frey, 2009; Bensaid et al., 1996; Pedrycz & Waletzky, 1997) are invented based on the traditional unsupervised clustering methods mentioned above. Generally speaking, semi-supervised clustering can be cast into the following three categories: (1) distance-based approach; (2) constraint-based approach; (3) hybrid approach.

The distance-based approach mainly focuses on how to learn a distance or metric measure. The learned distance measure is expected to satisfy the prior knowledge. Many researchers proposed the distance learning methods for semi-supervised clustering in the past years (Yin et al., 2010; Yan et al., 2012; Yin et al., 2012; de Amorim & Mirkin, 2012; Bar-Hillel et al., 2005; Xing et al., 2002). Yin et al. (2010) proposed an adaptive semi-supervised clustering kernel method based on metric learning(SCKMM) which utilized the pairwise constraints. de Amorim & Mirkin (2012) developed Minkowski metric weighted k-means which used the Minkowski metric to measure the distance between two samples. The Minkowski metric was learned from the samples in a semi-supervise manner. Yan et al. (2012) invented a novel search-based semi-supervised clustering method which learned the multi-viewpoint based similarity measure.

The constraint-based approach mainly studies how to revise the objective function or initialize the cluster centers to guide the clustering process. Basu et al. (2002) developed a semi-supervised version of k-means which used the labeled samples to estimate the initial cluster centers. Based on the FCM algorithm, Bensaid et al. (1996) proposed a semi-supervised FCM method with hard constraints (i.e., SSFCM_h in this paper) where the partition matrix of the labeled samples was fixed in the iterative process. After that, Pedrycz & Waletzky (1997) proposed another version, called semi-supervised FCM with soft constraints (i.e., SSFCM_s in this paper). SSFCM_s implemented a tradeoff between the unsupervised predictions and given class labels. Gan et al. (2015) introduced a semi-supervised locally consistent Gaussian mixture models (Semi-LCGMM) and successfully applied the method to image segmentation.

Furthermore, some researchers investigated the hybrid approach (Basu et al., 2004; Bilenko et al., 2004; Wei et al., 2017). Basu et al. (2004) proposed a hidden Markov random fields-based probabilistic framework for semi-supervised clustering which combined the distance-based and constraint-based approaches. In this method, Bregman divergence could be learned for the distance measure. Wei et al. (2017) developed a semi-supervised clustering ensemble approach which taken both pairwise constraints and metric measure into account. Image pixels clustering results confirmed the effectiveness of the proposed hybrid approach.

In the semi-supervised clustering methods, they generally assume that prior knowledge is helpful to improve the clustering performance. However, the prior knowledge may be risky which means that it will degenerate the clustering performance, such as wrong labels. Yin et al. (2010) have discussed the negative impact of noisy pair-wise constraints and pointed out that the wrong prior knowledge would yield the inferior clustering performance. Hence, it is meaningful to design safe semi-supervised clustering. Safe semi-supervised clustering mainly focuses on how to reduce the risk of the prior knowledge. Safe semi-supervised clustering is expected to be superior to the corresponding unsupervised and semi-supervised clustering, especially when semi-supervised clustering performs worse than unsupervised clustering.

In this paper, based on FCM and SSFCM_s, we present Local Homogeneous Consistent Safe SSFCM (LHC-S^{3}FCM) where the prior knowledge is given in the form of class labels. To the best of our knowledge, it is the first time safe semi-supervised clustering has been studied. The motivation is that the predictions of a labeled sample and its nearest homogeneous unlabeled ones should be similar when the labeled one is risky. In LHC-S^{3}FCM, we firstly build a local graph to model the relationship between the labeled sample and its nearest homogeneous unlabeled ones through the results obtained by unsupervised clustering. A graph-based regularization term is then constructed to allow the predictions of the labeled samples to approach that of the local homogeneous neighbors. It is expected to reduce the risk of the labeled samples. Meanwhile, our algorithm positively exploits the labeled samples by restricting the corresponding outputs to be the given class labels when the labeled ones may be helpful. In this sense, the predictions of the labeled samples in our algorithm are a tradeoff between the given class labels and the predictions of local homogeneous neighbors. Therefore LHC-S^{3}FCM achieves the goal of safe semi-supervised clustering.

Compared to the existing semi-supervised clustering methods, the advantages of our algorithm include: (1) The risky labeled samples are safely not freely exploited through local homogeneous consistency in our algorithm. (2) The proposed graph-based regularization term can be easily embedded into the other objective function-based semi-supervised clustering methods. The drawback of our algorithm is that a new regularization parameter λ_2 is introduced. To sum up, the main contributions can be listed as follows:

1. We develop a safe semi-supervised clustering method which is expected to outperform the corresponding unsupervised and semi-supervised clustering methods when there are wrong labels.

2. A local graph is constructed to model the relationship between the labeled sample and its nearest homogeneous unlabeled ones and the graph structure is preserved in our algorithm to safely exploit the risk prior knowledge.

3. We can achieve a closed-form solution of the optimization problem in our algorithm and obtain the promising results.

2. Unsupervised and semi-supervised fuzzy *c*-means

The traditional FCM is a kind of unsupervised clustering which is an effective tool in the data mining domain. Compared to k-means which belongs to hard clustering, FCM gives a soft conceptual enhancement by assigning each sample to different clusters to various membership degrees. FCM introduces a constrained optimization problem to partition a given dataset into several clusters. Formally, suppose a dataset $X = [x_1, x_2, \cdots, x_n]$, and the predefined number of clusters c, the following objective function is minimized in FCM:

$$J_m = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^m d_{ik}^2 \tag{1}$$

where m is a fuzzy degree with $m > 1$. $U = [u_{ik}] \in R^{c \times n}$ is a partition matrix and $d_{ik} = \|x_k - v_i\|_2$ denotes the distance between the kth sample x_k and the ith cluster center v_i.

Specifically, u_{ik} indicates the membership degree of x_k belonging to the ith cluster. It must satisfy the following constraints:

$$\sum_{i=1}^{c} u_{ik} = 1, \forall\ k = 1, \cdots, n$$
$$0 \le u_{ik} \le 1, \forall\ k = 1, \cdots, n$$

The constrained optimization problem (1) can be efficiently solved by an alternating iterative method. By minimizing the objective function (1), one can get the following solution:

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}},\ \forall i = 1, \cdots, c, k = 1, \cdots, n \tag{2}$$

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^m x_k}{\sum_{k=1}^{n} u_{ik}^m},\ \forall i = 1, \cdots, c \tag{3}$$

By iteratively computing u_{ik} and v_i based on Eq.(2) and Eq.(3), the optimal solution will be achieved when some convergence criterion is meet.

Although FCM can yield a promising performance in some cases and successfully applied in different tasks (Jiang et al., 2017; Chuang et al., 2006; Ahmed et al., 2002; Dembele & Kastner, 2003), it does not take into account the prior knowledge. More specially, when the prior knowledge was provided in the form of class labels, Bensaid et al. (1996) and Pedrycz & Waletzky (1997) respectively proposed two different semi-supervised clustering methods, called SSFCM_*h* and SSFCM_*s*. SSFCM_*h* exploited the prior knowledge through a hard way in the iterative process while SSFCM_*s* used them through a soft way in the revised objective function. Assuming that the first l samples in X are the labeled ones which means that each sample x_k will have a label $y_k \in \{1, \cdots, c\}$, the rest $u = n - l$ samples in X are the unlabeled ones.

In SSFCM_*h*, the partition matrix U is defined as:

$$U = [U^l | U^u] \tag{4}$$

where U^l is the partition matrix of the labeled samples in which $u_{ik}^l = 1$ if $i = y_k$ and $u_{ik}^l = 0$ if $i \neq y_k$. In the iterative process, U^l is fixed and u_{ik}^u in the U^u is computed as:

$$u_{ik}^u = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}}, \ \forall i = 1, \cdots, c, k = l+1, \cdots, n \tag{5}$$

The center v_i is calculated as:

$$v_i = \frac{\sum_{k=1}^{l} (u_{ik}^l)^m x_k + \sum_{k=l+1}^{n} (u_{ik}^u)^m x_k}{\sum_{k=1}^{l} (u_{ik}^l)^m + \sum_{k=l+1}^{n} (u_{ik}^u)^m}, \ \forall i = 1, \cdots, c \tag{6}$$

In SSFCM_*s*, the revised objective function is written as:

$$J_s = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^m d_{ik}^2 + \alpha \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik} - f_{ik} b_k)^m d_{ik}^2 \tag{7}$$

where α is a tradeoff parameter which controls a balance between the unsupervised component and given class labels. $B = [b_k]_{1 \times n}$ is a label indicator where $b_k = 1$ if x_k is labeled and $b_k = 0$ if x_k is unlabeled. $F = [f_{ik}]_{c \times n}$ denotes the fuzzy degrees of the labeled samples where $f_{ik} = 1$ if $i = y_k$ and $f_{ik} = 0$ otherwise.

m is set to 2 in the SSFCM_s to obtain a closed-form solution. In this case, one can obtain the value of u_{ik} as:

$$u_{ik} = \frac{1}{1+\alpha}\left\{\frac{1+\alpha\left(1-b_k\sum_{j=1}^{c} f_{jk}\right)}{\sum_{j=1}^{c}\frac{d_{ik}^2}{d_{jk}^2}} + \alpha f_{ik} b_k\right\}, \quad \forall i = 1,\cdots,c, k = 1,\cdots,n \tag{8}$$

The center v_i is computed as:

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^2 x_k + \alpha\sum_{k=1}^{n}(u_{ik} - f_{ik}b_k)^2 x_k}{\sum_{k=1}^{n} u_{ik}^2 + \alpha\sum_{k=1}^{n}(u_{ik} - f_{ik}b_k)^2}, \ \forall i = 1,\cdots,c \tag{9}$$

In FCM, the optimization problem can be easily solved for $m > 1$. However, FCM is an unsupervised clustering method which does not make full use of the prior knowledge. SSFCM_h and SSFCM_s belong to semi-supervised clustering by considering the prior knowledge. In general, the two semi-supervised clustering methods can yield better clustering performance than FCM. In SSFCM_h, the labeled samples are exploited through a hard way which means that the partition matrix U^l is predefined and fixed in the iterative process. SSFCM_s employs a soft way to exploit the labeled samples by introducing a fidelity term. To obtain the closed-form solution, m is fixed to 2. It is worthy to point out that SSFCM_h can be considered as a special case of SSFCM_s with $\alpha \to +\infty$. Nevertheless, SSFCM_h and SSFCM_s do not consider the risk of the labeled samples and we will discuss how to reduce the risk in the next section.

3. Local Homogeneous Consistent Safe SSFCM (LHC-S^{3}FCM)

In this section, we will introduce our algorithm in details.

3.1. Motivation

Semi-supervised clustering generally assumes that the labeled samples are always beneficial to the performance improvement. However, in some cases, the samples may be wrongly labeled or noise corrupted. In this case, the sample labels will be different from the true ones. For instance, we generate a synthetic dataset with wrong labels as shown in Fig. 1. The dataset is generated through a Gaussian distribution with an unit covariance matrix. Class 1 is a Gaussian distribution with mean $(2, -2)$ and Class 2 is a Gaussian distribution with mean $(-2,-2)$. The labeled samples are denoted as blue and the wrongly labeled ones are denoted as red. From Fig. 1b and 1c, one can see that semi-supervised clustering performs worse than FCM. It can be concluded that the wrong labels degenerate the performance of semi-supervised clustering. Meanwhile, the labeled samples are correctly clustered by FCM and our motivation is that the predictions of a labeled sample and its nearest homogeneous unlabeled ones in our algorithm should be similar if they have the same predicted labels in FCM. Hence it is expected to reduce the risk of the labeled samples and one can safely use the prior knowledge for clustering as shown in Fig. 1d.

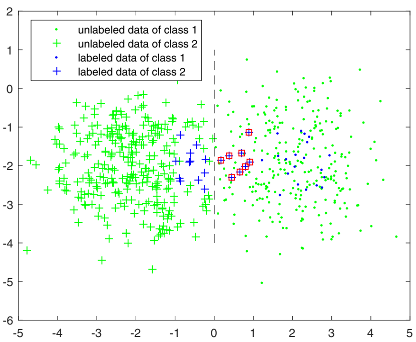
Since the predicted labels obtained by FCM are often inconsistent with the given class ones, the outputs of our algorithm can not be forced to directly approach that of FCM. Our idea is that we build a local graph to model the relationship between the labeled sample and its nearest homogeneous unlabeled ones which belong to the same cluster in FCM. The relationship is then preserved in our algorithm through a graph-based regularization term.

3.2. Method description

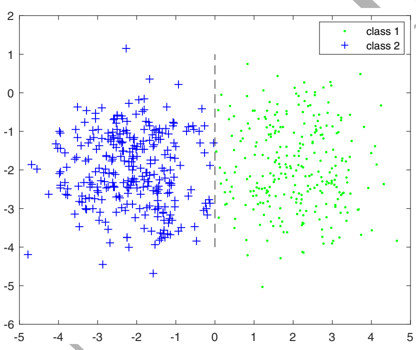
Based on the above analysis, we propose Local Homogeneous Consistent Safe SSFCM (LHC-S^{3}FCM). Firstly, we find the p nearest neighbors of the labeled samples according to the Euclidean distance and then find the homogeneous unlabeled samples in the p nearest neighbors using the clustering results of FCM. Assuming that the predicted labels obtained by FCM are denoted as $\widetilde{Y} = [\widetilde{y}_1, \widetilde{y}_2, \cdots, \widetilde{y}_n]$, the relationship between the labeled and unlabeled samples can be modeled as:

$$w_{kr} = \begin{cases} 1 & \text{if } x_r \in N_p(x_k) \text{ and } \widetilde{y}_k = \widetilde{y}_r \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

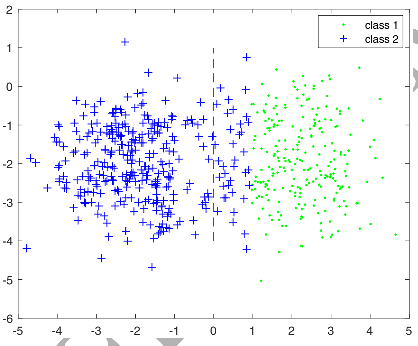
where $N_p(x_k)$ denotes the data sets of p nearest neighbors of x_k. x_k and x_r respectively represent the labeled and unlabeled samples.



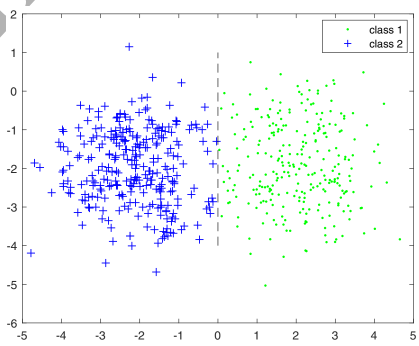
(a) A synthetic dataset with wrong labels



(b) Clustering results obtained by FCM



(c) Clustering results obtained by SSFCM



(d) Clustering results obtained by our algorithm

Figure 1: A plot of performance degeneration with a noisy synthetic dataset.

We then write the objective function of our algorithm as follows:

$$\begin{aligned} J_{sa} = & \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^{m} d_{ik}^{2} + \lambda_1 \sum_{k=1}^{l}\sum_{i=1}^{c} (u_{ik} - f_{ik})^{m} d_{ik}^{2} \\ & + \lambda_2 \sum_{k=1}^{l}\sum_{r=l+1}^{n} w_{kr} \sum_{i=1}^{c} (u_{ik} - u_{ir})^{2} \end{aligned} \tag{11}$$

$$\text{Subject to:} \quad \sum_{i=1}^{c} u_{ik} = 1, \forall k = 1, \cdots, n$$

where λ_1 and λ_2 are the regularization parameters which control the tradeoff between FCM and SSFCM. Specifically, the latter two terms respectively restrict the clustering predictions to be the given class labels and the outputs of the homogeneous neighbors.

3.3. Solution

In order to obtain the closed-form solution, we fix the fuzzy degree m to 2 as shown in (Pedrycz & Waletzky, 1997). In fact, one can give the numerical solution by some optimization algorithm when m is different from 2, such as genetic algorithm. Firstly, we try to find the solution of u_{ik} through the Lagrangian multiplier method. The Lagrangian multiplier function can be written as:

$$\begin{aligned} \mathcal{L} = & \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^{2} d_{ik}^{2} + \lambda_1 \sum_{k=1}^{l}\sum_{i=1}^{c} (u_{ik} - f_{ik})^{2} d_{ik}^{2} \\ & + \lambda_2 \sum_{k=1}^{l}\sum_{r=l+1}^{n} w_{kr} \sum_{i=1}^{c} (u_{ik} - u_{ir})^{2} - \gamma(\sum_{i=1}^{c} u_{ik} - 1) \end{aligned} \tag{12}$$

For the labeled sample x_k, the relevant part related to u_{ik} in the Lagrangian form can written as:

$$\begin{aligned} \mathcal{L}_1 = & \sum_{k=1}^{l}\sum_{i=1}^{c} u_{ik}^{2} d_{ik}^{2} + \lambda_1 \sum_{k=1}^{l}\sum_{i=1}^{c} (u_{ik} - f_{ik})^{2} d_{ik}^{2} \\ & + \lambda_2 \sum_{k=1}^{l}\sum_{r=l+1}^{n} w_{kr} \sum_{i=1}^{c} (u_{ik} - u_{ir})^{2} - \gamma(\sum_{i=1}^{c} u_{ik} - 1) \end{aligned} \tag{13}$$

By computing the derivative of $\mathcal{L}_1$ with respect to u_{ik} and setting it to 0, we have the following equation:

$$2u_{ik}d_{ik}^2 + 2\lambda_1(u_{ik} - f_{ik})d_{ik}^2 + 2\lambda_2 \sum_{r=l+1}^{n} w_{kr}(u_{ik} - u_{ir}) - \gamma = 0 \tag{14}$$

Therefore, we can achieve the solution of u_{ik} for the labeled sample x_k.

$$u_{ik} = \frac{p_{ik} + \frac{1-\sum_{j=1}^{c}\frac{p_{jk}}{q_{jk}}}{\sum_{j=1}^{c}\frac{1}{q_{jk}}}}{q_{ik}} \tag{15}$$

where $p_{ik} = 2\lambda_1 f_{ik}d_{ik}^2 + 2\lambda_2 \sum_{r=l+1}^{n} w_{kr}u_{ir}$ and $q_{ik} = 2d_{ik}^2 + 2\lambda_1 d_{ik}^2 + 2\lambda_2 \sum_{r=l+1}^{n} w_{kr}$.

For the unlabeled sample x_r, the relevant part related to u_{ir} in the Lagrangian form can be written as:

$$\begin{aligned} \mathcal{L}_2 = & \sum_{k=l+1}^{n}\sum_{i=1}^{c} u_{ik}^2 d_{ik}^2 + \lambda_2 \sum_{k=1}^{l}\sum_{r=l+1}^{n} w_{kr} \sum_{i=1}^{c} (u_{ik} - u_{ir})^2 \\ & - \gamma(\sum_{i=1}^{c} u_{ik} - 1) \end{aligned} \tag{16}$$

By computing the derivative of $\mathcal{L}_2$ with respect to u_{ir} and setting it to 0, we have the following equation:

$$2u_{ir}d_{ir}^2 - 2\lambda_2 \sum_{k=1}^{l} w_{kr}(u_{ik} - u_{ir}) - \gamma = 0 \tag{17}$$

By solving the above equation, we can achieve the following solution of u_{ir} for the unlabeled sample x_r.

$$u_{ir} = \frac{s_{ir} + \frac{1-\sum_{j=1}^{c}\frac{s_{jr}}{t_{jr}}}{\sum_{j=1}^{c}\frac{1}{s_{jr}}}}{t_{ir}} \tag{18}$$

where $s_{ir} = 2\lambda_2 \sum_{k=1}^{l} w_{kr}u_{ik}$ and $t_{ir} = 2d_{ir}^2 + 2\lambda_2 \sum_{k=1}^{l} w_{kr}$.

After we obtain the solution of u_{ik} for the labeled and unlabeled samples, we then compute the cluster center V. By taking the derivative of J_{sa} with respect to v_i based on the fact $d_{ik}^2 = \|x_k - v_i\|_2^2$, we can achieve the following equation:

$$\frac{\partial J_{sa}}{\partial v_i} = -2\sum_{k=1}^{n} u_{ik}^2 (x_k - v_i) - 2\lambda_1 \sum_{k=1}^{l} (u_{ik} - f_{ik})^2 (x_k - v_i) \tag{19}$$

By setting the derivative to 0, we can yield the following solution:

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^2 x_k + \lambda_1 \sum_{k=1}^{l} (u_{ik} - f_{ik})^2 x_k}{\sum_{k=1}^{n} u_{ik}^2 + \lambda_1 \sum_{k=1}^{l} (u_{ik} - f_{ik})^2} \tag{20}$$

By iteratively computing u_{ik} and v_i, we can obtain the optimal partition matrix U and cluster center V when a convergence criterion is reached. The implementation details can be seen in Algorithm 1.

Algorithm 1 LHC-S^{3}FCM

Input: Dataset $X = [x_1, x_2, \cdots, x_n]$ where the first l samples are labeled and the rest are unlabeled. The corresponding labels of the labeled samples are $Y = [y_1, y_2, \cdots, y_l]^T$, the parameters λ_1, λ_2, p, η, and $Maxiter$.

Output: The partition matrix U and clustering center V.

```
Perform FCM on the whole dataset X to yield the cluster result Ỹ;
Construct the local graph w_kr using Eq.(10);
Initialize the cluster center V^(0) by computing mean of the labeled samples in each class;
for t = 1 : Maxiter do
    Update u_ik^(t) using Eq.(10), Eq.(15) and Eq.(18);
    Update v_i^(t) using Eq.(20);
    Compute the value of J_sa^(t) using Eq.(10) and Eq.(11);
    if |J_sa^(t) − J_sa^(t−1)| < η then
        return U and V.
    end if
end for
```

4. Experimental analysis

In this section, we will report the experimental results to evaluate the performance of our algorithm. The experiments are carried out on several UCI datasets (Frank & Asuncion, 2010). The following state-of-the-art methods are used to compared to our algorithm:

- k-means (Jain, 2010)
- FCM (Bezdek, 1981)
- SSFCM_h (Bensaid et al., 1996)
- SSFCM_s (Pedrycz & Waletzky, 1997)

Twelve UCI datasets are used to verify the effectiveness of our algorithm and the details are described in Table 1. For each dataset, we randomly select 20% to form the labeled subset and the rest to form the unlabeled subset. In order to discuss the harm of the wrongly labeled samples, some labeled samples are labeled with wrong labels which are different from the true labels. The ratio of the wrongly labeled samples changes from 0%-30% with step size 5%. The parameter α in SSFCM_s is set to 1. λ_1, λ_2, and p in LHC-S^{3}FCM are respectively set to 0.1, 10 and 10 which will be discussed in section 4.2.

Table 1: Description of the experimental datasets

Dataset	#samples	#Features	#Classes
Australian	690	15	2
Balance-scale	625	4	3
Wdbc	569	30	2
Bupa	345	6	2
Heart	270	13	2
Ionosphere	351	34	2
Iris	150	4	3
liver-disorders	345	6	2
Mushroom	8124	112	2
Sonar	351	60	2
Vehicle	846	18	4
Waveform	5000	21	3

4.1. Result analysis

The results of different methods with respect to different wrongly labeled ratios are reported in Fig. 2 and 3. From these figures, we can have the following conclusions:

1. When the ratio is 0% (i.e., there are not wrong labels), semi-supervised clustering (e.g., SSFCM_h and SSFCM_s) can achieve better performance than FCM over the most datasets except Vehicle. It confirms the positive effect of the labeled samples.

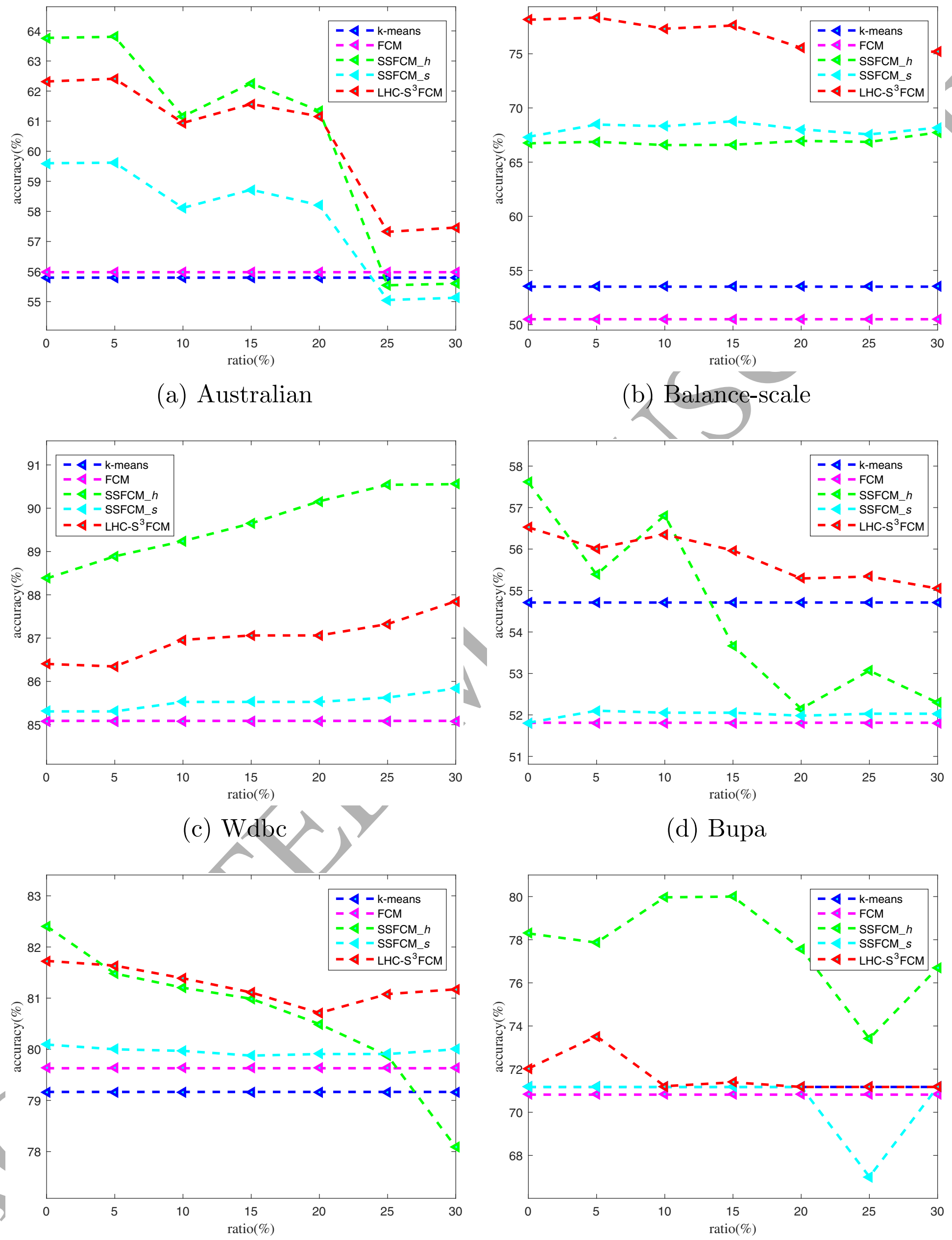
2. When the ratio is 0%, LHC-S^{3}FCM performs better than unsupervised clustering (i.e., k-means and FCM). It shows that our algorithm can be used for semi-supervised clustering.

3. As the wrongly labeled ratio changes from 0%-30%, the performance of different semi-supervised methods and our algorithm overall decreases over the most datasets except Wdbc. It meets our expectation that the wrong labels can degenerate the performance of semi-supervised clustering. In particular, our algorithm can achieve much better clustering performance than SSFCM_h and SSFCM_s on Balance-scale and Mushroom. The reason may be that some correctly labeled samples are outliers and lie in the boundaries between different clusters in the two datasets.

4. It is worthy to point out that SSFCM_h and SSFCM_s show a flat performance as the wrongly labeled ratio increases from 0%-30% on Mushroom and Waveform. The reason may be that the scale of the two datasets is larger than that of the other datasets. A relatively small number of the wrongly labeled samples have little effect on performance degeneration.

5. In some cases, SSFCM_s performs worse than FCM when the wrong labels reach a certain ratio, such as more than 20% on Australian and Ionosphere, 15% on IRIS. More specially, FCM outperforms SSFCM_s with different ratios on Vehicle. These experimental results demonstrate that the inappropriate prior knowledge can hurt the performance of semi-supervised clustering and it explains the necessary of designing safe semi-supervised clustering.

6. LHC-S^{3}FCM can achieve better performance than SSFCM_s in all cases. It shows that the mechanism proposed in our algorithm can safely exploit the labeled samples and verifies the effectiveness of our algorithm.

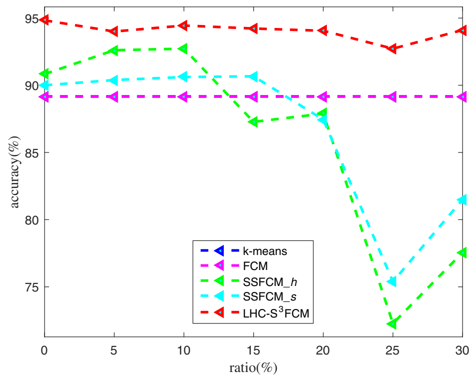


(a) Australian (b) Balance-scale

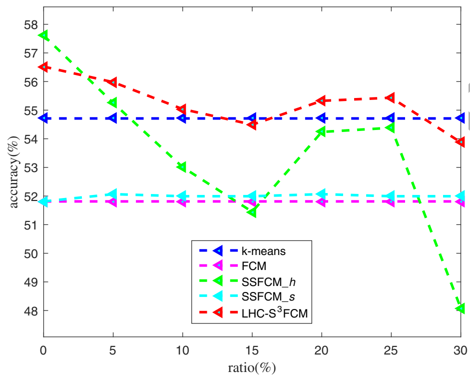
(c) Wdbc (d) Bupa

(e) Heart (f) Ionosphere

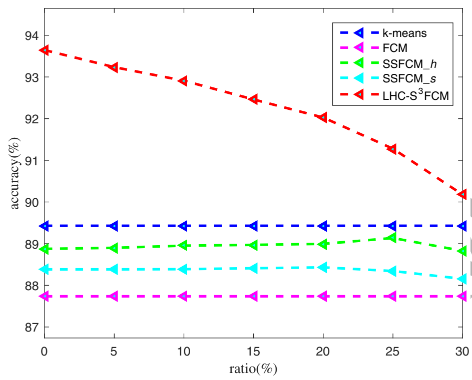
Figure 2: Performance comparison of different methods over the first six datasets.



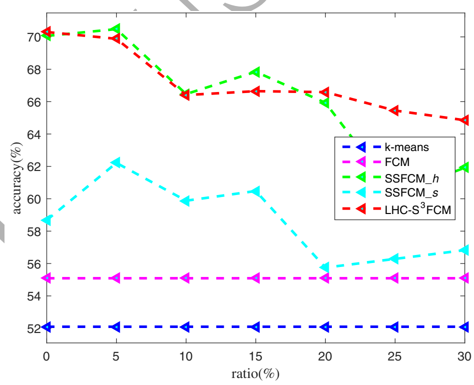
(a) Iris



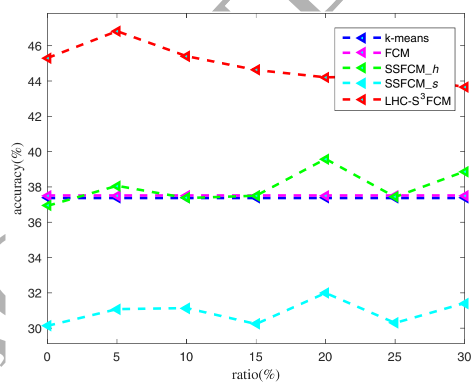
(b) liver-disorders



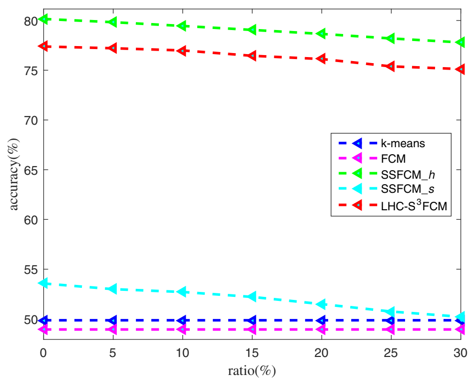
(c) Mushroom



(d) Sonar



(e) Vehicle



(f) Waveform

Figure 3: Performance comparison of different methods over the last six datasets.

4.2. Parameter analysis

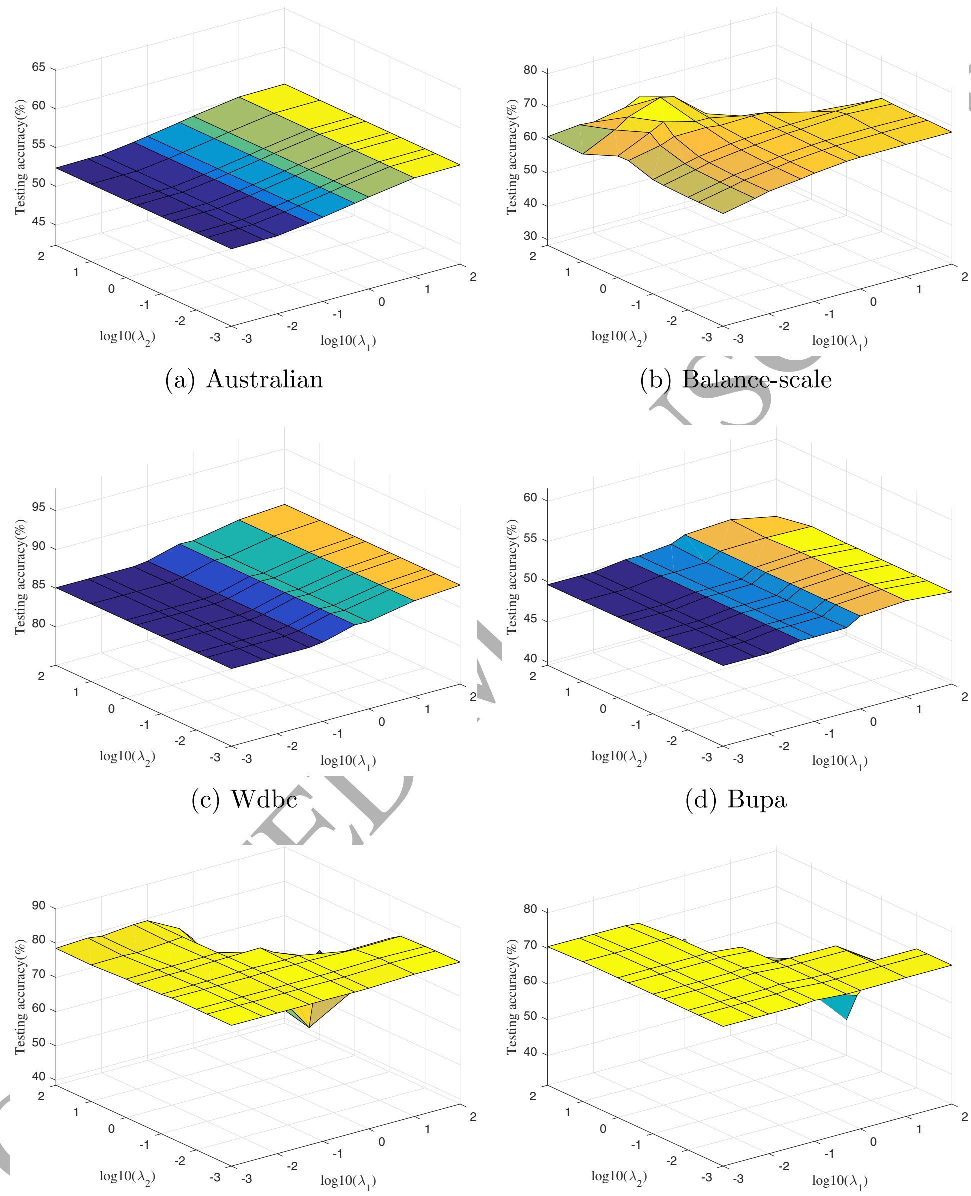
In our algorithm, there are two regularization parameters (i.e., λ_1 and λ_2) which play an important role in the clustering performance. In this section, we investigate the impact of λ_1 and λ_2 on the performance of our algorithm when the wrongly labeled ratio is set to 30% . The values of the parameters are selected in $\{10^{-3}, 5*10^{-3}, 10^{-2}, 5*10^{-2}, 10^{-1}, 0.5, 1, 10, 100\}$. Fig. 4 shows the plot on the first six datasets. From the plot, one can find that our algorithm generally obtains the best performance when λ_2 is large. It further explains that the proposed safe mechanism in our algorithm is effective and efficient.

4.3. Discussion

From the above result analysis, one can see that it is necessary and meaningful to consider the risk of prior knowledge in semi-supervised clustering. Meanwhile, the proposed safe mechanism using local homogeneous consistency is effective to reduce the risk and can be used in some applications, such as document categorization. Furthermore, parameter setting (i.e., λ_1 and λ_2) is important for the practicability of our algorithm. As can be seen from the parameter analysis, the performance with respect to different parameter settings is related to the wrongly labeled ratio. Hence designing an approach to evaluate the quality of the data is a meaningful topic.

5. Conclusion

Semi-supervised clustering does not consider the risk or harm of the labeled samples which may degenerate the performance. Hence, this paper tries to develop a safe semi-supervised clustering method based on FCM and SSFCM_*s*. The experimental results demonstrate that our algorithm can perform better than FCM and SSFCM_*s* which is the goal of safe semi-supervised clustering. Certainly, we can apply our algorithm to various practical applications. For example, in the document categorization, the documents can be easily collected through the Internet and some documents may be labeled by the users. However, some documents may be wrongly labeled because of the subjective experience of the users. The wrongly labeled samples will hurt the performance of semi-supervised clustering. Meanwhile, our study can alleviate the negative effect of the wrongly labeled samples to some extent and improve the practicability of semi-supervised clustering.



(a) Australian

(b) Balance-scale

(c) Wdbc

(d) Bupa

(e) Heart

(f) Ionosphere

Figure 4: Clustering performance with different regularization parameters over different datasets.

Although our algorithm has some advantages as presented in section 1, there are some limitations: (1) The parameter m is fixed to 2 in the paper; (2) We just consider the prior knowledge in the form of class labels and there are the other forms in the practical applications. Hence, our future work mainly focuses on the following research directions: (1) It is meaningful to design an approach to evaluate different risk of different labeled samples. (2) We will investigate the other forms of the prior knowledge, such as pair-wise constraints. (3) As discussed in section 4.2, the parameter setting may be related to the wrongly labeled ratio which presents the quality of the prior knowledge. Hence, it is very practical to measure the quality of the prior knowledge to guide the parameter setting. (4) We will try to employ the other intelligent optimization algorithms to solve the optimization problem where m is different from 2.

Acknowledgments

The work was supported by National Natural Science Foundation of China under grant No. 61601162, 61501154 and 61671197, and Zhejiang Provincial Natural Science Foundation of China under grant No. Y201328513.

References

Ahmed, M. N., Yamany, S. M., Mohamed, N., Farag, A. A., & Moriarty, T. (2002). A modified fuzzy c-means algorithm for bias field estimation and segmentation of mri data. *IEEE Transactions on Medical Imaging*, *21*, 193–199.

de Amorim, R. C., & Mirkin, B. (2012). Minkowski metric, feature weighting and anomalous cluster initializing in k-means clustering. *Pattern Recognition*, *45*, 1061–1075.

Bar-Hillel, A., Hertz, T., Shental, N., & Weinshall;, D. (2005). Learning a mahalanobis metric from equivalence constraints. *Journal of Machine Learning Research*, *6*, 937–965.

Basu, S., Banerjee, A., & Mooney, R. J. (2002). Semi-supervised clustering by seeding. In *Proceedings of the 19th International Conference on Machine Learning* (pp. 27–34). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Basu, S., Bilenko, M., & Mooney, R. J. (2004). A probabilistic framework for semi-supervised clustering. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 59–68). New York, USA: ACM.

Bensaid, A. M., Hall, L. O., Bezdek, J. C., & Clarke, L. P. (1996). Partially supervised clustering for image segmentation. *Pattern Recognition*, *29*, 859–871.

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, USA: Plenum Press.

Bilenko, M., Basu, S., & Mooney, R. J. (2004). Integrating constraints and metric learning in semi-supervised clustering. In *Proceedings of the Twenty-first International Conference on Machine Learning* (pp. 11–18). New York, NY, USA: ACM.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Cannon, R. L., Dave, J. V., & Bezdek, J. C. (1986). Efficient implementation of the fuzzy c-means clustering algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *8*, 248–255.

Chen, W., & Feng, G. (2012). Spectral clustering: A semi-supervised approach. *Neurocomputing*, *77*, 229–242.

Chen, X., Liu, X., & Jia, Y. (2011). Discriminative structure selection method of Gaussian Mixture Models with its application to handwritten digit recognition. *Neurocomputing*, *74*, 954 – 961.

Chuang, K.-S., Tzeng, H.-L., Chen, S., Wu, J., & Chen, T.-J. (2006). Fuzzy c-means clustering with spatial information for image segmentation. *Computerized Medical Imaging and Graphics*, *30*, 9 – 15.

Dembele, D., & Kastner, P. (2003). Fuzzy c-means method for clustering microarray data. *Bioinformatics*, *19*, 973–980.

Frank, A., & Asuncion, A. (2010). UCI machine learning repository.

Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. *Science*, *315*, 972–976.

Gan, H., Sang, N., & Huang, R. (2015). Manifold regularized semi-supervised gaussian mixture model. *Journal of the Optical Society of America A*, *32*, 566–575.

Givoni, I. E., & Frey, B. J. (2009). Semi-supervised affinity propagation with instance-level constraints. *Journal of Machine Learning Research - Proceedings Track*, *5*, 161–168.

Hartigan, J. A., & Wong, M. A. (1979). A K-means clustering algorithm. *Applied Statistics*, *28*, 100–108.

Hasnat, M. A., Alata, O., & Trmeau, A. (2016). Joint color-spatial-directional clustering and region merging (JCSD-RM) for unsupervised rgb-d image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*, 2255–2268.

Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, *31*, 651–666.

Jiang, Z., Li, T., Min, W., Qi, Z., & Rao, Y. (2017). Fuzzy c-means clustering based on weights and gene expression programming. *Pattern Recognition Letters*, *90*, 1 – 7.

Liu, J., Cai, D., & He, X. (2010). Gaussian mixture model with local consistency. In *The 24th AAAI Conference on Artificial Intelligence* (pp. 512–517). Georgia, USA: AAAI.

Liu, Q., Sun, Y., Wang, C., Liu, T., & Tao, D. (2017). Elastic net hypergraph learning for image clustering and semi-supervised classification. *IEEE Transactions on Image Processing*, *26*, 452–463.

Lu, Z. (2006). An iterative algorithm for entropy regularized likelihood learning on gaussian mixture with automatic model selection. *Neurocomputing*, *69*, 1674–1677.

Mei, J. P., Wang, Y., Chen, L., & Miao, C. (2016). Large scale document categorization with fuzzy clustering. *IEEE Transactions on Fuzzy Systems*, *PP*, 1–1.

Ng, A. Y., Jordan, M. I., & Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems* (pp. 849–856). MIT Press.

Pal, N. R., & Bezdek, J. C. (1995). On cluster validity for the fuzzy c-means model. *IEEE Transactions on Fuzzy Systems*, *3*, 370–379.

Pedrycz, W., & Waletzky, J. (1997). Fuzzy clustering with partial supervision. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *27*, 787–795.

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*, 888–905.

Wei, S., Li, Z., & Zhang, C. (2017). Combined constraint-based with metric-based in semi-supervised clustering ensemble. *International Journal of Machine Learning and Cybernetics*, (pp. 1–16).

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z.-H., Steinbach, M., Hand, D. J., & Steinberg, D. (2007). Top 10 algorithms in data mining. *Knowledge and Information Systems*, *14*, 1–37.

Xing, E. P., Ng, A. Y., Jordan, M. I., & Russell, S. (2002). Distance metric learning, with application to clustering with side-information. In *Advances in Neural Information Processing Systems* (pp. 505–512). MIT Press.

Xu, R., & II, D. W. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, *16*, 645–678.

Xu, W., Liu, X., & Gong, Y. (2003). Document clustering based on non-negative matrix factorization. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval* (pp. 267–273). New York, USA: ACM.

Yan, Y., Chen, L., & Nguyen, D. T. (2012). Semi-supervised clustering with multi-viewpoint based similarity measure. In *The 2012 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). IEEE.

Yang, Y., Xu, D., Nie, F., Yan, S., & Zhuang, Y. (2010). Image clustering using local discriminant models and global integration. *IEEE Transactions on Image Processing*, *19*, 2761–2773.

Yeung, K. Y., Haynor, D. R., & Ruzzo, W. L. (2000). Validating clustering for gene expression data. *Bioinformatics*, *17*, 309–318.

Yin, S., Qian, Y., & Gong, M. (2017). Unsupervised hierarchical image segmentation through fuzzy entropy maximization. *Pattern Recognition*, *68*, 245 – 259.

Yin, X., Chen, S., Hu, E., & Zhang, D. (2010). Semi-supervised clustering with metric learning: An adaptive kernel method. *Pattern Recognition*, *43*, 1320 – 1333.

Yin, X., Shu, T., & Huang, Q. (2012). Semi-supervised fuzzy clustering with metric learning and entropy regularization. *Knowledge-Based Systems*, *35*, 304 – 311.

Zhang, H., & Lu, J. (2009). Semi-supervised fuzzy clustering: A kernel-based approach. *Knowledge-Based Systems*, *22*, 477–481.