Received May 25, 2019, accepted July 7, 2019, date of publication July 16, 2019, date of current version August 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929307

Safe Semi-Supervised Fuzzy *C*-Means Clustering

HAITAO GAN
School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China
Anhui Key Laboratory of Detection Technology and Energy Saving Devices, Anhui Polytechnic University, Wuhu 241000, China

e-mail: htgan@hdu.edu.cn

This work was supported in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LY19F020040, in part by the National Natural Science Foundation of China under Grant 61601162, Grant 61771178, and Grant 61671197, in part by the Open Research Fund of Anhui Key Laboratory of Detection Technology and Energy Saving Devices, Anhui Polytechnic University under Grant 2017070503B026-A03, and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LY17F030021 and Grant LY18F030009.

ABSTRACT With the rapid increase in the number of collected data samples, semi-supervised clustering (SSC) has become a useful mining tool to find an intrinsic data structure with the help of prior knowledge. The common used prior knowledge includes pair-wise constraints and cluster labels. In the past decades, many relevant methods are proposed to improve clustering performance of SSC by mining prior knowledge. In general, the prior knowledge is assumed to be beneficial to yielding desirable results. However, one can gather inappropriate prior knowledge in some scenarios, such as wrong cluster labels. In this case, prior knowledge can result in degenerating clustering performance. Therefore, how to raise safe semi-supervised clustering (S3C) should be investigated. A main goal of S3C is that the corresponding result is never inferior to that of the corresponding unsupervised clustering part. To achieve the goal, we propose safe semi-supervised Fuzzy *c*-Means clustering (S^3FCM) which is extended from traditional semi-supervised FCM (SSFCM). In our algorithm, wrongly labeled samples are carefully explored by constraining the corresponding predictions to be those yielded by unsupervised clustering. Meanwhile, the predictions of the other labeled samples should approach to the given labels. Therefore the labeled samples are expected to be safely explored through a balance between unsupervised clustering and SSC. From the reported clustering results on different datasets, we can find that S^3FCM can yield comparable, if not the best, performance among different unsupervised clustering and SSC methods even if the wrong ratio achieves 20%.

INDEX TERMS Unsupervised clustering, semi-supervised clustering, fuzzy *c*-means, wrong labels.

I. INTRODUCTION

With rapid increase of the number of collected data samples, Semi-Supervised Clustering (SSC) has become an useful mining tool to find the intrinsic data structure with the help of prior knowledge. The common used prior knowledge includes pair-wise constraints and cluster labels. Up to now, many SSC methods [1]–[7] have been proposed which are extended from traditional unsupervised clustering methods, such as *k*-means [8], Gaussian Mixture Models (GMM) [9], [10], Fuzzy *c*-Means (FCM) [11], [12], Affinity Propagation (AP) [13], spectral clustering, and so on. In general, the SSC methods can be casted into the following two types: (1) metric-based approach; (2) constraint-based approach.

The metric-based approach aims to yield a distance metric which must satisfy the given prior information. There are many metric-based SSC methods proposed in the past years [4], [14], [15]. Yin *et al.* [4] used the pair-wise constraints to introduce an adaptive metric learning method for SSC. Yan *et al.* [15] proposed a semi-supervised clustering framework for multi-viewpoint based similarity measure in which the class labels were provided. Ding *et al.* [16] employed the prior knowledge to adaptively learn a similarity matrix and leaded to semi-supervised spectral clustering. Different from traditional SSC which used all the prior knowledge, Sanodiya *et al.* [17] tried to select the appropriate constraints to learn a distance through the Bregman projection and the obtained distance was used to help *k*-means label the datasets.

Different from the metric-based approach, the constraint-based approach concentrates on initializing cluster centers or revising objective function through the prior knowledge. Basu *et al.* [18] utilized the prior information to compute the initial cluster centers and further proposed semi-supervised *k*-means. Meanwhile, Pedryca and Waletzky [5] developed

The associate editor coordinating the review of this manuscript and approving it for publication was Malik Jahan Khan.

Semi-Supervised FCM (SSFCM) in which cluster labels of some samples were provided. SSFCM revised the objective function by adding a fidelity term between the outputs and given labels of the labeled samples. A semi-supervised version based on Gaussian Mixture Models (GMM) was proposed by Gan *et al.* [2] and applied in image segmentation. Ren *et al.* [19] developed a semi-supervised version (i.e., SSDC) of density-based clustering which was proposed by Rodriguez and Laio [20]. SSDC utilized the pairwise constraints to gradually merge the generated temporary clusters. Seeded FCM [21] which considered each labeled sample as a seed was used to detect regions of interest of medical images. This method could achieve promising detecting performance. Since deep technique has become an effective tool in the machine learning field, Ren *et al.* [22] extended deep embedded clustering to the semi-supervised framework which used the pairwise constraints to help learn the representations.

Among different SSC methods, it is generally assumed that prior information is beneficial to performance improvement. However, one can gather inappropriate prior knowledge in some scenarios, such as wrong labels. In this case, prior knowledge can result in degenerating clustering performance. This phenomenon has been verified by Yin *et al.* [4]. Harmful effect of noisy pair-wise constraints has been analyzed in the literature. Therefore, how to raise Safe Semi-Supervised Clustering (S3C) should be investigated. The goal of S3C is that the corresponding result is never inferior to that of the corresponding unsupervised clustering part. Gan *et al.* [23] firstly proposed a S3C method based on local homogeneous consistency. The method used the results of FCM to build a local graph and further constructed a regularizer to constrain the predictions of the labeled samples. From the reported results, the method could effectively reduce the risk of the labeled samples. However, the performance depended heavily on the graph quality. Hence, it is important to investigate the other strategies to achieve S3C.

To achieve the goal, we propose a novel S3C method, called Safe Semi-Supervised FCM clustering (S^3FCM), in which the sample labels are provided in this paper. In S^3FCM, risky samples which are wrongly labeled are carefully explored by unsupervised clustering not SSC. Based on this, the corresponding predictions should approach to those yielded by unsupervised clustering. Meanwhile, the helpful labeled samples should be positively mined and the corresponding predictions should approach to the given labels. To a certain extent, the predictions of the labeled samples in S^3FCM are a balance between those of unsupervised clustering and given labels. Therefore the labeled samples are safely explored and S3C is achieved.

In conclusion, the main work of this paper can be given as:

1) A novel S3L method is proposed which can enrich the S3C field and extend the applicability of SSC.
2) The exploration strategy will be easily extended to the other model-based SSC methods.

The remaining structure of the paper is as follows: We firstly give the background knowledge (i.e., FCM and SSFCM) in Section 2. We then give a detailed description of S^3FCM in Section 3. In Section 4, a series of experiments are carried out and clustering results are reported. Finally, we conclude the work and point out some future study directions in Section 5.

II. UNSUPERVISED AND SEMI-SUPERVISED FUZZY *C*-MEANS

Fuzzy *c*-Means (FCM) [11] which belongs to unsupervised clustering can be considered as an useful mining tool to explore the data structure in machine learning field. Different from *k*-means which is a kind of hard clustering, FCM attempts to assign different membership degrees to samples belonging to different clusters and thus is a kind of soft clustering. To cluster a given dataset, FCM intends to compute a partition matrix which denotes the membership degrees of different samples by solving an optimization problem. Formally, given a dataset $X = [x_1, x_2, \ldots, x_n]$ with the number of clusters c, FCM has the following objective function:

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

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

Meanwhile, u_{ik} should meet the following constraints:

$$\sum_{i=1}^{c} u_{ik} = 1, \quad \forall\, k = 1, \ldots, n$$
$$0 \le u_{ik} \le 1, \quad \forall\, k = 1, \ldots, n \tag{2}$$

One can employ the Lagrangian multiplier and alternating iterative method to solve the constrained optimization problem. By computing the derivative of the Lagrangian function, calculation formulas of u_{ik} and v_i can be given as:

$$u_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}}, \quad \forall i = 1, \ldots, c,\ k = 1, \ldots, n \tag{3}$$

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^m x_k}{\sum_{k=1}^{n} u_{ik}^m}, \quad \forall i = 1, \ldots, c \tag{4}$$

By computing u_{ik} and v_i through Eq.(3) and Eq.(4), the optimal solution of J_m can be yielded when the iteration process is converged.

In the past decades, FCM has acquired successfully applications in many domains since the procedure is simple and it can often achieve the desired performance [24], [25]. Nevertheless, FCM does not embed prior knowledge which can be collected and useful in some applications. More specifically, when cluster labels of a part of samples were provided, Pedrycz and Waletzky [5] developed a semi-supervised version of FCM (i.e., SSFCM). Given a dataset X as above mentioned, the first l samples with their cluster labels $y_k|_{k=1}^{l} \in \{1, \ldots, c\}$ constitute a labeled subset and the remaining $n - l$ ones constitute an unlabeled subset.

By embedding prior knowledge into FCM, SSFCM has the following objective function:

$$J_s = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^m d_{ik}^2 + \alpha \sum_{k=1}^{n}\sum_{i=1}^{c} (u_{ik} - f_{ik} b_k)^m d_{ik}^2 \quad (5)$$

where α is a parameter which reflects the importance of the fidelity term. $B = [b_k]_{1\times n}$ presents a label indicator in which a entry $b_k = 1$ for the labeled sample x_k and $b_k = 0$ otherwise. $F = [f_{ik}]_{c\times n}$ presents the fuzzy degrees in which a entry $f_{ik} = 1$ if $i = y_k$ for x_k and $f_{ik} = 0$ otherwise.

In order to achieve a simple and closed-form solution, m is set to 2 in SSFCM. Based on this, the formula of u_{ik} is shown as:

$$u_{ik} = \frac{1}{1+\alpha}\left\{ \frac{1 + \alpha\left(1 - b_k \sum_{j=1}^{c} f_{jk}\right)}{\sum_{j=1}^{c} \frac{d_{ik}^2}{d_{jk}^2}} + \alpha f_{ik} b_k \right\}, \quad \forall i = 1, \ldots, c,\ k = 1, \ldots, n \quad (6)$$

The formula of center v_i is shown as:

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^2 x_k + \alpha \sum_{k=1}^{n} (u_{ik} - f_{ik} b_k)^2 x_k}{\sum_{k=1}^{n} u_{ik}^2 + \alpha \sum_{k=1}^{n} (u_{ik} - f_{ik} b_k)^2}, \quad \forall i = 1, \ldots, c \quad (7)$$

By iteratively computing u_{ik} and v_i through Eq.(6) and Eq.(7), one can obtain the optimal solution of J_s.

III. SAFE SSFCM (S^3FCM)

Next, we will give a detailed description of S^3FCM.

A. FORMULATION

The traditional SSC methods generally make an assumption that label information is always helpful to improve clustering performance. However, due to the negligence and fatigue of experts, wrong labels of some samples may be collected in the collection procedure. These wrongly labeled samples can hurt the performance of SSC without consideration of the risk. We thus try to carefully explore the information of the labeled samples through unsupervised analysis in order to reduce the corresponding risk. In other words, the predictions of the wrongly labeled samples are restricted to be those of FCM. Based on this idea, we construct an unsupervised output-based regularization term to realize a safe exploration of the risky labeled samples.

Firstly, we perform FCM on X by ignoring the labels and partition the dataset into c clusters. Since the cluster labels yielded by FCM are often inconsistent with the given ones, we use a mapping algorithm [26] to map the predicted cluster labels to the equivalent given ones. We can then obtain a permuted partition matrix $\widehat{U} = [\widehat{u}_{ik}]_{c\times n}$ according to the corresponding relationships between the predicted cluster labels and given ones. After that, we build the objective function of S^3FCM as:

$$\begin{aligned} J_{sa} = &\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^2 d_{ik}^2 + \lambda_1 \sum_{k=1}^{n}\sum_{i=1}^{c} (u_{ik} - f_{ik} b_k)^2 d_{ik}^2 \\ &+ \lambda_2 \sum_{k=1}^{n}\sum_{i=1}^{c} (u_{ik} - \widehat{u}_{ik} b_k)^2 d_{ik}^2 \\ &\text{Subject to: } \sum_{i=1}^{c} u_{ik} = 1, \quad \forall k = 1, \ldots, n \\ &0 \le u_{ik} \le 1, \quad \forall\, k = 1, \ldots, n \end{aligned} \quad (8)$$

where λ_1 and λ_2 are the regularization parameters. Specifically, the latter two terms constrain the predictions of SSC to be the given labels and those of FCM, respectively.

In Eq.(8), it is expected to achieve the goal of safe exploration of the labeled samples through the last term which is the unsupervised output-based regularizer.

B. SOLUTION

1) When v_i is fixed, we employ the Lagrangian multiplier method to achieve the value of u_{ik}. The Lagrangian function is written as:

$$\begin{aligned} \mathcal{L} = &\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^2 d_{ik}^2 + \lambda_1 \sum_{k=1}^{n}\sum_{i=1}^{c} (u_{ik} - f_{ik} b_k)^2 d_{ik}^2 \\ &+ \lambda_2 \sum_{k=1}^{n}\sum_{i=1}^{c} (u_{ik} - \widehat{u}_{ik} b_k)^2 d_{ik}^2 - \gamma\left(\sum_{i=1}^{c} u_{ik} - 1\right) \end{aligned} \quad (9)$$

One can achieve the following equation by a derivation method and setting the derivative to 0.

$$2u_{ik} d_{ik}^2 + 2\lambda_1 (u_{ik} - f_{ik} b_k) d_{ik}^2 + 2\lambda_2 (u_{ik} - \widehat{u}_{ik} b_k) d_{ik}^2 - \gamma = 0 \quad (10)$$

The value of u_{ik} can be obtained as:

$$u_{ik} = \frac{1}{1+\lambda_1+\lambda_2}\left(\frac{1 + \lambda_1 + \lambda_2 - \sum_{j=1}^{c} \Delta_{jk}}{\frac{d_{ik}^2}{d_{jk}^2}} + \Delta_{ik} \right) \quad (11)$$

where $\Delta_{ik} = \lambda_1 f_{ik} b_k + \lambda_2 \widehat{u}_{ik} b_k$.

2) When u_{ik} is fixed, the value of v_i can be obtained based on the equation $d_{ik} = \|x_k - v_i\|_2$. The derivative of J_{sa} with respect to v_i is written as:

$$\begin{aligned} \frac{\partial J_{sa}}{\partial v_i} = &2\sum_{k=1}^{n} u_{ik}^2 (v_i - x_k) + 2\lambda_1 \sum_{k=1}^{n} (u_{ik} - f_{ik} b_k)^2 (v_i - x_k) \\ &+ 2\lambda_2 \sum_{k=1}^{n} (u_{ik} - \widehat{u}_{ik} b_k)^2 (v_i - x_k) \end{aligned} \quad (12)$$

By setting the derivative to 0, the value of v_i is obtained as (13), shown at the bottom of this page.

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^2 x_k + \lambda_1 \sum_{k=1}^{n} (u_{ik} - f_{ik} b_k)^2 x_k + \lambda_2 \sum_{k=1}^{n} (u_{ik} - \widehat{u}_{ik} b_k)^2 x_k}{\sum_{k=1}^{n} u_{ik}^2 + \lambda_1 \sum_{k=1}^{n} (u_{ik} - f_{ik} b_k)^2 + \lambda_2 \sum_{k=1}^{n} (u_{ik} - \widehat{u}_{ik} b_k)^2} \quad (13)$$

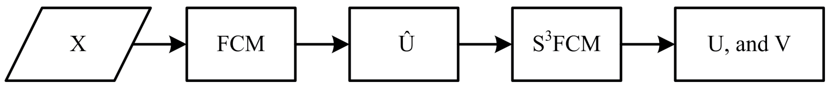


FIGURE 1. A flow chart of our algorithm.

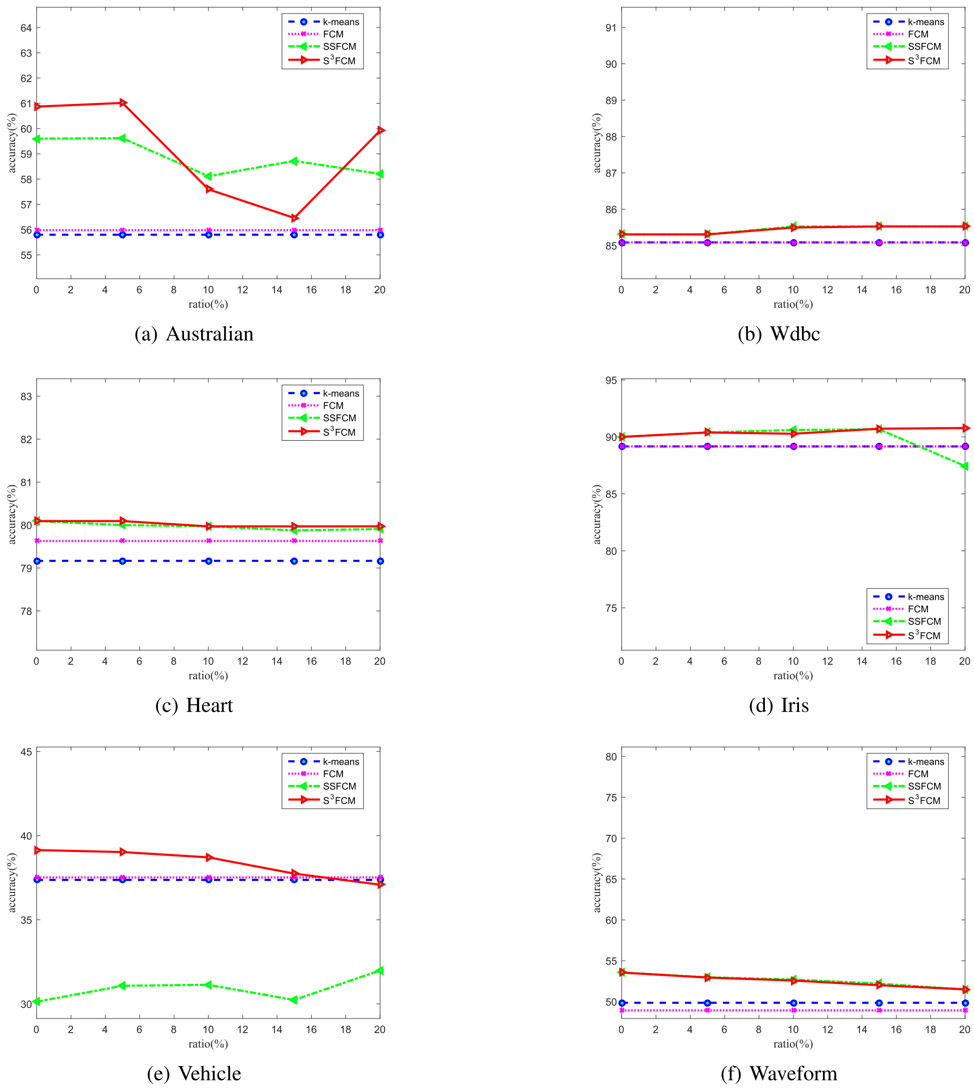


FIGURE 2. Performance comparison of the different methods over the six datasets.

By iteratively calculating u_{ik} and v_i, the optimal solution of U and V can be achieved when some convergence criterion is met, such as $|J_{sa}^{(t)} - J_{sa}^{(t-1)}| < \eta$ where t is the number of iterations and η is a predefined threshold. A flow chart of S^{3}FCM is shown in Fig. 1 and Algorithm 1 gives an implementation description.

IV. EXPERIMENTS

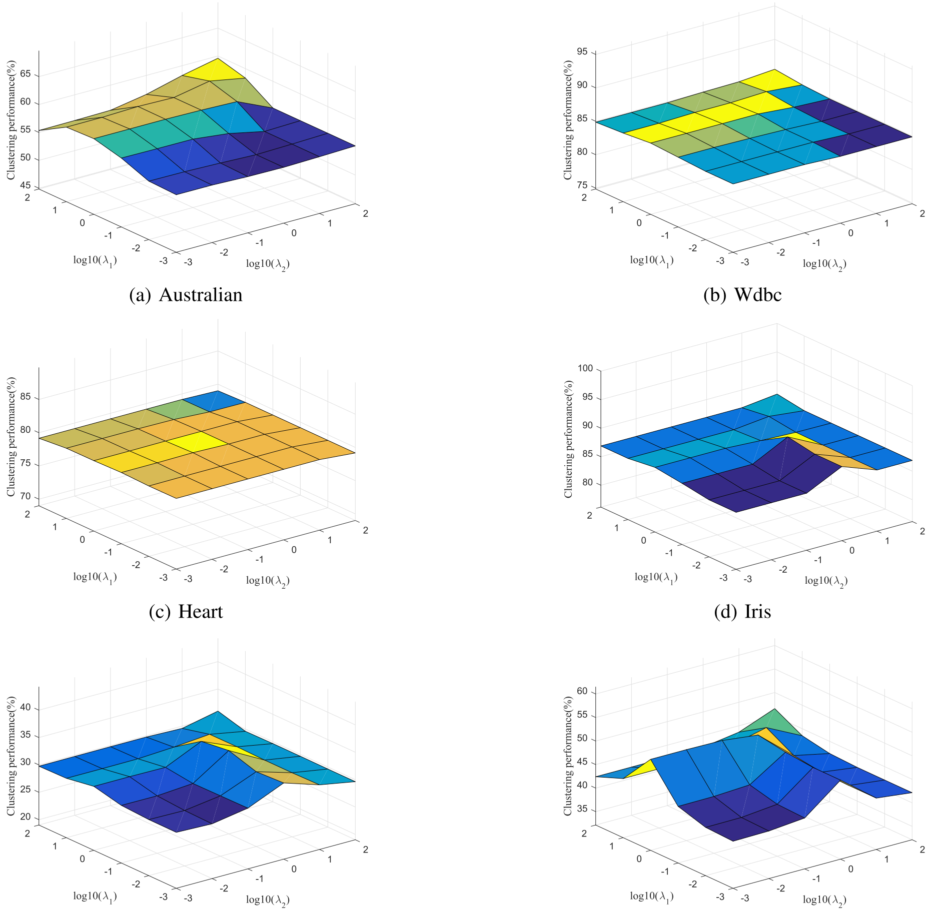
Next, we carry out our experiments on several UCI datasets [27]. In order to explain the usefulness of our algorithm, the following algorithms are used to be compared with our algorithm, including: (1) k-means [28]; (2) FCM [11]; (3) SSFCM [5].

A. EXPERIMENTAL SETTING

The information of the used UCI datasets is provided in Table 1. In the experimental setting of SSC, each dataset is divided into two subsets: (1) 20% of the samples are randomly

TABLE 1. Information of UCI datasets.

Dataset	#samples	#Features	#Classes
Australian	690	15	2
Wdbc	569	30	2
Heart	270	13	2
Iris	150	4	3
Vehicle	846	18	4
Waveform	5000	21	3



(a) Australian (b) Wdbc (c) Heart (d) Iris (e) Vehicle (f) Waveform

FIGURE 3. Clustering performance with respect to different values of λ_1 and λ_2 on the six datasets.

Algorithm 1 S^3FCM

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

Output: The partition matrix U and the center V.

Perform FCM on the whole dataset X to yield the cluster result $\widehat{U}$;
Initialize the cluster centers $V^{(0)}$ by calculating mean of the labeled samples in each cluster;
for $t = 1 : Maxiter$ **do**
  Update $u_{ik}^{(t)}$ using Eq.(11);
  Update $v_i^{(t)}$ using Eq.(13);
  Compute the value of $J_{sa}^{(t)}$ using Eq.(8);
  if $|J_{sa}^{(t)} - J_{sa}^{(t-1)}| < \eta$ **then**
    return U and V.
  end if
end for

---

chosen to constitute a labeled subset; (2) The remaining constitutes an unlabeled subset. Moreover, since our algorithm is designed to reduce the risk of prior knowledge, some samples are wrongly labeled which means that the given labels are different from the true ones. The wrong ratio increases gradually from 0%-20% with step length 5%. The parameter α in SSFCM is fixed to 0.1. λ_1 and λ_2 in S^3FCM are respectively set to 0.1, and 1.

B. RESULT DISCUSSION

The clustering performance of different algorithms under different wrong ratios are shown in Fig. 2. From the plot, the main conclusions are drawn as:

1) Overall, FCM can perform comparable, if not better, than *k*-means. It meets our expectation and explains that FCM is selected as the base clustering method.
2) Compared to FCM, SSFCM and S^3FCM can yield better clustering results on most datasets if the wrong ratio is set to 0%. It shows the usefulness of prior knowledge and S^3FCM can be used for semi-supervised clustering.
3) SSFCM which uses the wrongly labeled samples is inferior to FCM in the clustering performance if the wrong ratio increases to a certain value, such as 20% on the IRIS datasets. In particular, on the Vehicle dataset, FCM is always superior to SSFCM in the case of different wrong ratios. This phenomenon indicates that the wrongly labeled samples can result in performance

degradation of SSC and it verifies the importance of raising S^3C.

4) S^3FCM can achieve the best performance among k-means, FCM and SSFCM in most cases and perform slightly worse than unsupervised clustering with 20% on the Vehicle dataset. It illustrates the regularization approach employed in S^3FCM is feasible and can achieve the goal of safe exploration of the risky labeled samples.

Additionally, one can see that two regularization parameters λ_1 and λ_2 in Eq.(8) have important impacts on the clustering performance. Therefore, it is necessary to discuss the behaviours under different values of λ_1 and λ_2. Since S^3FCM tries to safely explore wrong label information, we fix the wrong ratio to 20%. The values of λ_1 and λ_2 are both selected from the set $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 100\}$. The clustering performance is shown in Fig. 3. As can be seen from this figure, one can find that the best performance of our algorithm is generally achieved when λ_2 is large. It is mainly because of the presence of the wrongly labeled samples. In this case, the risky prior knowledge should be explored through the last regularization term in J_{sa}. Therefore, the parameter λ_2 of the last term should be large. Meanwhile, in the case of the wrongly labeled samples, our algorithm generally obtains poor performance when λ_2 is small. It further explains the reason that we use the unsupervised output-based regularization term to safely explore the wrong label information and it shows the effectiveness and importance of the proposed exploration strategy in our algorithm.

V. CONCLUSION

This paper develops S^3FCM to safely explore the risky prior knowledge for improving the robustness of SSC. By building the unsupervised output-based regularization term, our algorithm can constrain the predictions of the labeled samples and reduce the corresponding risk. The experimental results show that S^3FCM can outperform unsupervised clustering and SSC even if SSFCM is inferior to FCM in some scenarios. It demonstrates the effectiveness of the used strategy in S^3FCM and the application scope of SSC can be extended. Certainly, our algorithm has its drawbacks. Therefore, we will aim to study the following directions: (1) It is important to develop novel S3C methods based on the other forms of prior knowledge (e.g., pair-wise constraints). (2) How to reduce the risk of both labeled and unlabeled samples for SSC is another interesting work.

REFERENCES

[1] N. Grira, M. Crucianu, and N. Boujemaa, "Unsupervised and semi-supervised clustering: A brief survey," *Rev. Mach. Learn. Techn. Process. Multimedia Content*, vol. 1, pp. 9–16, Jul. 2004.
[2] H. Gan, N. Sang, and R. Huang, "Manifold regularized semi-supervised Gaussian mixture model," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 32, no. 4, pp. 566–575, Apr. 2015.
[3] A. Martinez-Uso, F. Pla, and J. M. Sotoca, "A semi-supervised Gaussian mixture model for image segmentation," in *Proc. Int. Conf. Pattern Recognit.*, Los Alamitos, CA, USA, Aug. 2010, pp. 2941–2944.
[4] X. Yin, S. Chen, E. Hu, and D. Zhang, "Semi-supervised clustering with metric learning: An adaptive kernel method," *Pattern Recognit.*, vol. 43, no. 4, pp. 1320–1333, 2010.
[5] W. Pedrycz and J. Waletzky, "Fuzzy clustering with partial supervision," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 27, no. 5, pp. 787–795, Sep. 1997.
[6] I. E. Givoni and B. J. Frey, "Semi-supervised affinity propagation with instance-level constraints," in *Proc. 12th Int. Conf. Artif. Intell. Statist.*, Clearwater Beach, FL, USA, Apr. 2009, pp. 161–168.
[7] W. Chen and G. Feng, "Spectral clustering: A semi-supervised approach," *Neurocomputing*, vol. 77, no. 1, pp. 229–242, 2012.
[8] J. A. Hartigan and M. A. Wong, "A K-means clustering algorithm," *Appl. Stat.*, vol. 28, no. 1, pp. 100–108, 1979.
[9] C. A. Bouman. (1997). *Cluster: An Unsupervised Algorithm for Modeling Gaussian mixtures*. [Online]. Available: http://www.ece.purdue.edu/~bouman
[10] X. Chen, X. Liu, and Y. Jia, "Discriminative structure selection method of Gaussian Mixture Models with its application to handwritten digit recognition," *Neurocomputing*, vol. 74, no. 6, pp. 954–961, 2011.
[11] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York, NY, USA: Plenum, 1981.
[12] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Trans. Fuzzy Syst.*, vol. 3, no. 3, pp. 370–379, Aug. 1995.
[13] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.
[14] R. C. De Amorim and B. Mirkin, "Minkowski metric, feature weighting and anomalous cluster initializing in K-means clustering," *Pattern Recognit.*, vol. 45, no. 3, pp. 1061–1075, Mar. 2012.
[15] Y. Yan, L. Chen, and D. T. Nguyen, "Semi-supervised clustering with multi-viewpoint based similarity measure," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 2012, pp. 1–8.
[16] S. Ding, H. Jia, L. Zhang, and F. Jin, "Research of semi-supervised spectral clustering algorithm based on pairwise constraints," *Neural Comput. Appl.*, vol. 24, no. 1, pp. 211–219, Jan. 2014.
[17] R. K. Sanodiya, S. Saha, and J. Mathew, "A kernel semi-supervised distance metric learning with relative distance: Integration with a MOO approach," *Expert Syst. Appl.*, vol. 125, pp. 233–248, Jul. 2019.
[18] S. Basu, A. Banerjee, and R. J. Mooney, "Semi-supervised clustering by seeding," in *Proc. 19th Int. Conf. Mach. Learn.*. San Francisco, CA, USA: Morgan Kaufmann, 2002, pp. 27–34.
[19] Y. Ren, X. Hu, K. Shi, G. Yu, D. Yao, and Z. Xu, "Semi-supervised denpeak clustering with pairwise constraints," in *PRICAI: Trends in Artificial Intelligence*, X. Geng and B.-H. Kang, Eds. Cham, Switzerland: Springer, 2018, pp. 837–850.
[20] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.
[21] L. Santos, R. Veras, K. Aires, L. Britto, and V. Machado, "Medical image segmentation using seeded fuzzy C-means: A semi-supervised clustering algorithm," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–7.
[22] Y. Ren, K. Hu, X. Dai, L. Pan, S. C. H. Hoi, and Z. Xu, "Semi-supervised deep embedded clustering," *Neurocomputing*, vol. 325, pp. 121–130, Jan. 2019.
[23] H. Gan, Y. Fan, Z. Luo, and Q. Zhang, "Local homogeneous consistent safe semi-supervised clustering," *Expert Syst. Appl.*, vol. 97, pp. 384–393, May 2018.
[24] K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, and T.-J. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *Comput. Med. Imag. Graph.*, vol. 30, pp. 9–15, Jan. 2006.
[25] Z. Jiang, T. Li, W. Min, Z. Qi, and Y. Rao, "Fuzzy c-means clustering based on weights and gene expression programming," *Pattern Recognit. Lett.*, vol. 90, pp. 1–7, Apr. 2017.
[26] L. Lovasz and M. D. Plummer, *Matching Theory*. Budapest, Hungary: North Holland, 1986.
[27] A. Frank and A. Asuncion. (2010). *UCI Machine Learning Repository*. [Online]. Available: http://archive.ics.uci.edu/ml
[28] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010.