



A New Approach for Semi-supervised Fuzzy Clustering with Multiple Fuzzifiers

Tran Manh Tuan¹ · Mai Dinh Sinh² · Tran Dinh Khang³ · Phung The Huan⁴ · Tran Thi Ngan¹ · Nguyen Long Giang⁵ · Vu Duc Thai⁴

Received: 13 January 2022 / Revised: 19 May 2022 / Accepted: 4 July 2022
© The Author(s) under exclusive licence to Taiwan Fuzzy Systems Association 2022

Abstract Data clustering is the process of dividing data elements into different clusters in which elements in one cluster have more similarity than those in other clusters. Semi-supervised fuzzy clustering methods are used in various applications. The available methods are based on fuzzy C-Mean, kernel function, weight function and adaptive function. The fuzzification coefficient is an important factor that affects to the performance in these methods. In this paper, we propose the improvements of semi-supervised standard fuzzy C-Mean clustering (SSFCM) by using multiple fuzzifiers to increase clusters quality. Two proposed models, named as MCSSFC-P and MCSSFC-C, use different fuzzifiers for each data point and for each cluster, respectively, which are established in a form of optimal problems. The values of fuzzifiers are updated to get the best values of objective functions. Evaluations on different datasets are performed. The numerical results show the higher performance of our model than some related models.

Keywords Data clustering · Fuzzy clustering · Semi-supervised fuzzy clustering · Multiple fuzzifiers

1 Introduction

Nowadays, fuzzy theory is applied into many kinds of human life in various applications [32, 34]. Data clustering is the process of grouping objects into different groups, such that the objects in a group are more similar than those in other groups. Data clustering can be separated into two subgroups, including hard clustering and fuzzy clustering. In hard clustering, each data point belongs to a unique cluster, while in fuzzy clustering, each data point can belong to many different clusters with a probability of that data point to be in each cluster. In fuzzy clustering, Fuzzy C-means clustering combining with support vector machine is used in solving channel equalisation issue [19]. By integrating FCM, this model reduces the number of parameters based on the comparison to other models. In many cases, some additional information is provided by users to guide the clustering progress. In this case, the clustering algorithms are called the semi-supervised fuzzy clustering algorithms [27, 29]. Data clustering techniques have been applied in many areas such as document classification [11], logistics [25], dental radiographs segmentation [31], medical image segmentation [1, 4, 12], risk classification [44], etc. In satellite image processing, there are many separate application areas, for example, classification of vehicles participating in traffic [7], classification

✉ Tran Manh Tuan
tmtuan@tlu.edu.vn

¹ Faculty of Computer Science and Engineering, Thuyloi University, Hanoi, Vietnam

² Institute of Techniques for Special Engineering, Le Quy Don Technical University, Hanoi, Vietnam

³ School of Information and Communication Technology, Hanoi University of Science and Technology, Hanoi, Vietnam

⁴ University of Information and Communication Technology, Thai Nguyen University, Thainguyn, Vietnam

⁵ Institute of Information Technology, Vietnam Academy of Science and Technology, Hanoi, Vietnam

of satellite image to forecasting drought, flood, wildfire [22, 36], etc.

There are several kinds of additional information in semi-supervised fuzzy clustering: The constraints of must-link and cannot-link (1); The labels of data (2); The pre-defined membership degree (3). In the first one (1), the former requires that two elements must be in the same cluster. On the contrary, the latter shows that a couple of elements cannot belong to a same cluster [17]. In the second one (2), labels of a part of data are known but the remaining are not. In the last one (3), the additional information provides the membership degrees of data points in dataset to achieve better clustering quality and to focus on the objects that users need to orientate [30]. Thong [37] presented an overview of semi-supervised fuzzy clustering algorithms showing the use of three kinds of additional information in typical algorithms.

Being aware that fuzzification coefficient is an important factor that affects the performance of fuzzy clustering, this paper proposes the improvements of semi-supervised standard fuzzy C-Mean clustering (SSFCM) by using multiple fuzzifiers to increase clusters quality. Two proposed models, named as MCSSFC-P and MCSSFC-C, use different fuzzifiers for each data point and for each cluster, respectively, which are established in a form of optimal problems. The values of fuzzifiers are updated to get the best values of objective functions. Evaluations on different datasets are performed. The numerical results show the higher performance of our model than some related models.

2 Related Works

In this section, an overview of semi-supervised fuzzy clustering methods is given. Some related concepts are also presented in this part.

2.1 Researches of Semi-supervised Fuzzy Clustering Algorithms

Many researches of semi-supervised fuzzy clustering algorithms based on FCM [5] are introduced by different approaches. These methods combined some classic methods, such as kernel function, weight function, adaptive function, etc.

2.1.1 Semi-supervised Fuzzy Clustering Based on Multiple Kernel Function

Frigul et al. [13] examined the existing fuzzy kernel clustering algorithms. This research provided an overview of unsupervised algorithms and semi-supervised algorithms that incorporated partial supervision into the objective function to control the optimization process. Huang et al. [18] shown that most

efforts along with this direction have been used in supervised learning, support vector classification and regression.

In [28], an algorithm of semi-supervised fuzzy clustering based on multi-kernel function was proposed. This algorithm improved land classification performance by using a function that combines multiple kernels, in which each kernel had a different weight for each attribute. Kanzawa [20] also proposed a semi-supervised fuzzy clustering approach by modifying differential kernel matrices. This method improved FCM algorithm by using an unknown kernel function and controlling the entropy to use soft constraints. The experiments have shown the higher performance of this method comparing with other methods. Apart from the advantages of high clustering quality, the limitation of this algorithm is the process of parameter estimation of the kernels. The calculation of parameters depended on the initialization of original values. Salehi et al. [33] presented a semi-supervised multiple kernel fuzzy clustering based on entropy and relative entropy. In this method, the entropy coefficient and divergence measure of entropy are applied in place of the generator for the unsupervised part and the corresponding geometric distance measurement for the semi-supervised part by highlighting the combination of unsupervised and semi-supervised sections explicitly.

2.1.2 Safe Semi-supervised Fuzzy Clustering

Gan et al. [14] proposed a safe semi-supervised clustering method with weight function. First, the algorithm used unsupervised clustering method to partition the datasets and defined the difference matrix to calculate the safety reliability of each labelled sample. It is assumed that a correctly clustered sample must have a high level of certainty. Then, the algorithm built a graph-based regularization term to model the relationship among the labelled samples and unlabelled samples based on clustering results. This graph-based regularization term enabled the predictions of the labelled samples to approximate those of the local homogeneous neighbours. It is expected that the risk of labelled samples will be reduced. This algorithm actively takes advantage of labelled patterns by limiting the appropriate output to the specified class labels.

Another safe semi-supervised fuzzy clustering method based on FCM also was proposed by Gan [16]. This method has carefully uncovered trademark errors by tying the appropriate predictions to those due to the unsupervised subset. In the meantime, predictions of other marked patterns for certain labels may be accessible. Therefore, it is expected that labelled samples will be safely examined by the balance between the unattended subgroup and the semi-supervised subgroup. From the experimental results, the proposed algorithm can yield better results than FCM.

However, this method still has some limitations, including the inability to deal with pair constraints, and the high risk of labelled samples and unlabelled samples. Gan et al. [15] also proposed a method of safe semi-supervised clustering with confidence-weighted. This model aimed to determine the safety of each sample because in the dataset, each sample had a different effect on the performance of the model. It means that a sample with high accuracy clustering results is highly reliable. With 3 phases of implementation, the model determined that the labelled samples would be exploited safely. The performance of this method was compared with the unsupervised and semi-supervised clustering methods on selected datasets.

2.1.3 Semi-supervised Fuzzy Clustering Based on Adaptive Function

Casalino et al. [8] developed a semi-supervised fuzzy clustering algorithm based on FCM that integrated the adaptive function and data stream classification. The algorithm used a dynamic component to estimate the required number of clusters based on data distribution and used FCM to find optimal results, then. In another approach, Yu et al. [43] proposed the method of constraint propagation based on transitive closure using the bridging operator and affinity propagation to solve the first constraint. Then, a semi-supervised clustering framework based on a random subspace with a set of confidence coefficients was proposed to address the second constraint. Next, the adaptive semi-supervised clustering aggregation framework was introduced to address the third constraint. The optimal subspace set was found by a new adaptive process. The experimental results of this research direction are better than related researches. However, the computational volume of the algorithm is large with multiple steps. This increased processing time, especially for large datasets. Cai et al. [6] presented an adaptive method of object selection to apply the semi-supervised fuzzy clustering. Experimental results show that adaptive method of object selection applied in semi-supervised fuzzy clustering can improve clustering performance.

Some other methods using neural networks are suggested by Arshad et al. [2], Arshad et al. [3]. These methods mostly related to FCM and selective fusion techniques. At first, the algorithm divided the dataset into separated parts, including unlabelled data, properly labelled data, and incorrectly labelled data. Then, the FCM method was used to cluster all the data. Clustering results were included in the attribute selection process to give the final results. This method increased the accuracy of classification but it was not considering the relationship between attributes. Following the approach of Śmieja et al. [35], a neural network framework for semi-supervised

clustering with paired constraints (must-link or cannot-link) was proposed. In this framework, the semi-supervised clustering was consisted of two phases. In the first phase, two Siamese neural networks were used to identify pairs of unlabelled points. In second phase, based on a supervised neural network, clustering process was performed with the use of full paired data set generated in previous phase. Extensive experiments on different data sets showed the high efficiency of the proposed method. In [25], a method has been given to solve two problems in semi-supervised clustering, including (1) to improve the clustering process by discriminative feature; (2) to improve the clustering efficiency by using labelled data and unlabelled data. This method was based on deep learning combined with semi-supervised learning to improve the efficiency of the semi-supervised cluster. The experimental results showed the high efficiency of the proposed method.

Kumar et al. [23] developed a semi-supervised clustering method for dental radiographs segmentation. This approach used Otsu's method to divide the dental X-Ray image into background and foreground regions. Then used FCM algorithm and proposed algorithm to increase efficiency for clustering method. In the semi-supervised rough fuzzy Laplacian Eigenmaps [26], a set of semi-supervised fuzzy similarity particles was constructed to describe the similarity between samples based on the rules that homogeneous samples have higher similarity than heterogeneity sample. This method also evaluated the similarity of two samples on the same class. A Laplace proximity diagram and a class-based neighbourhood diagram were created to describe the topological structure among patterns and between each pattern and its prototype. This ensures that homogeneous samples in a space with lower dimensions were mapped closer and more compactly around the prototypes. The recommended method provided high classification accuracy and data visualization comparing to some related semi-supervised methods [6, 9, 10, 24, 27, 33, 38–41].

2.2 Semi-supervised Standard Fuzzy C-Mean clustering (SSFCM) algorithm

In the research of Semi-supervised Standard Fuzzy C-Mean clustering, SSFCM was introduced by Yasunori et al. [42] using pre-defined membership degree \bar{U} . In this method, the membership degree \bar{U} was integrated in the objective function of FCM to improve the clustering performance of the algorithm. The objective function is formed as below:

$$J(U, V) = \sum_{i=1}^N \sum_{j=1}^C \|u_{ij} - \bar{u}_{ij}\|^m \|X_i - V_j\|^2 \rightarrow \min$$

Table 1 Semi-supervised standard fuzzy C-mean clustering algorithm

Input	The dataset X including N data elements $\{X_1, X_2, \dots, X_N\}$; number of clusters C ; membership degree \bar{U} ; threshold ε and maximum iteration $MaxStep > 0$
Output	Matrices U and V
SSFCM	
1:	$t = 0$
2:	Initialize random value for V^t
3:	Repeat
3.1:	$t = t + 1$
3.2:	Calculate U^t with 2 cases: If $m = 1$: $u_{ij} = \begin{cases} \bar{u}_{ij} + 1 - \sum_{k=1}^C \bar{u}_{ik} & i = \arg(\min_k X_i - V_j ^2) \\ \bar{u}_{ij} & i \neq \arg(\min_k X_i - V_j ^2) \end{cases}, i = \overline{1, N}, j = \overline{1, C}$ If $m > 1$: $u_{ij} = \bar{u}_{ij} + \left(1 - \sum_{k=1}^C \bar{u}_{ik}\right) \frac{\left(\frac{1}{\ X_i - V_j\ }\right)^{\frac{2}{m-1}}}{\sum_{k=1}^C \left(\frac{1}{\ X_i - V_k\ }\right)^{\frac{2}{m-1}}}, i = \overline{1, N}; j = \overline{1, C}$
3.3	Calculate V^t by equation $V_j = \frac{\sum_{i=1}^N u_{ij} - \bar{u}_{ij} ^m X_i}{\sum_{i=1}^N u_{ij} - \bar{u}_{ij} ^m}, j = \overline{1, C}$
3.4:	Until $\ V^{(t)} - V^{(t-1)}\ \leq \varepsilon$ or $t > MaxStep$

The main steps of SSFCM are presented in Table 1 below.

2.3 Fuzzy C-Mean Clustering with Multiple Fuzzifiers Algorithm (MC-FCM)

Khang et al. [21] proposed a new fuzzy clustering algorithm using different fuzzifiers for each data point to increase the quality of the cluster. The main idea of this method is using multiple fuzzifiers instead of unique fuzzification coefficient in the FCM. Multiple fuzzifiers (m_i) can be defined by (1). When m_i are equal, the MCFCM algorithm is the FCM algorithm.

$$m_i = m_1 + (m_2 - m_1) \left(\frac{S_i - S_{min}}{S_{max} - S_{min}} \right)^\alpha; i = \overline{1, N}, \quad (1)$$

where m_1, m_2 is the lower and upper boundary of m_i ($1 \leq m_1 \leq m_2$); α is an exponent parameter; $S_i = \sum_{j=1}^{N/C} D_{ij}$; $D_{ij} = \|X_i - X_j\| (\forall i, j = \overline{1, N})$; $S_{max} = \max_{i \in N}(S_i)$. $S_{min} = \min_{i \in N}(S_i)$.

The main steps of MC-FCM are presented as in Table 2 below.

3 Methods

Based on the advantages of additional information in semi-supervised clustering from SSFCM and the advantages of multiple fuzzifiers from MC-FCM, in this paper, we propose two improvements of the semi-supervised standard fuzzy C-Mean clustering algorithm by using multiple fuzzifiers for data point and for cluster to increase clusters quality. The main idea of these methods is using SSFCM with multiple fuzzifiers instead of one fuzzifier. In this research, mathematical models in the optimal problem form are introduced. Two novel algorithms are named as the Semi-supervised fuzzy clustering algorithm with multiple point fuzzifiers (MCSSFC-P) and Semi-supervised fuzzy clustering model with multiple cluster fuzzifiers (MCSSFC-C). The details of two these algorithms are presented in Sections 3.1 and 3.2, respectively.

3.1 Semi-supervised Fuzzy Clustering Algorithm with Multiple Point Fuzzifiers (MCSSFC-P)

By applying SSFCM algorithm in the case of using multiple fuzzifiers, MCSSFC-P is formed as an optimal problem with the objective function (2):

$$J(U, V) = \sum_{i=1}^N \sum_{j=1}^C \|u_{ij} - \bar{u}_{ij}\|^{m_i} * \|X_i - V_j\|^2 \rightarrow \min \quad (2)$$

and the constraints:

Table 2 Fuzzy C-Mean Clustering with Multiple Fuzzifiers algorithm

Input	The dataset X including N data elements $\{X_1, X_2, \dots, X_N\}$; number of clusters C ; fuzzifiers m_i ($i = \overline{1, N}$); threshold ε and maximum iteration $MaxStep > 0$
Output	Matrices U and V
MCFCM	
1	$t = 0$
2	Initialize random value for V^t
3	Repeat
3.1	$t = t + 1$
3.2	Calculate U^t by equation $u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\ X_i - V_j\ }{\ X_i - V_k\ } \right)^{\frac{2}{m_i-1}}}, i = \overline{1, N}; j = \overline{1, C}$
3.3	Calculate V^t by equation $V_j = \frac{\sum_{i=1}^N u_{ij}^{m_i} X_i}{\sum_{i=1}^N u_{ij}^{m_i}}, j = \overline{1, C}$
3.4	Until $\ V^{(t)} - V^{(t-1)}\ \leq \varepsilon$ or $t > MaxStep$

$$\begin{aligned}
 u_{ij} \in [0, 1]; \sum_{j=1}^C u_{ij} = 1; \forall i \\
 &= \overline{1, N}, \bar{u}_{ij} \in [0, 1], \sum_{j=1}^C \bar{u}_{ij} \leq 1; \forall i \\
 &= \overline{1, N}
 \end{aligned}
 \tag{3}$$

Additional information is defined via the results of FCM algorithm. Based on that, the pre-defined membership degrees are calculated by using (4). If a part of data is labelled, based on the label of these samples, membership degrees are defined by using (5).

$$\begin{aligned}
 \bar{u}_{ij} = \begin{cases} u_{ij}^{(FCM)}, & \text{when } u_{ij}^{(FCM)} = \max_{j=\overline{1, C}} (u_{ij}^{(FCM)}) \\ 0, & \text{when } u_{ij}^{(FCM)} \neq \max_{j=\overline{1, C}} (u_{ij}^{(FCM)}) \end{cases} i \\
 = \overline{1, N}; j = \overline{1, C}
 \end{aligned}
 \tag{4}$$

$$\bar{u}_{ij} = \begin{cases} 1, & \text{when } X_i \text{ label } j \\ 0, & \text{otherwise} \end{cases}, i = \overline{1, N}; j = \overline{1, C}
 \tag{5}$$

In order to solve optimal problem (2)–(3), we use Lagrange multiplier method. Then, we have

$$L = \sum_{i=1}^N \sum_{j=1}^C |u_{ij} - \bar{u}_{ij}|^{m_i} \|X_i - V_j\|^2 - \sum_{i=1}^N \lambda_i \left(\sum_{j=1}^C u_{ij} - 1 \right)
 \tag{6}$$

Taking the derivation of L by V_j , we get

$$\frac{\partial J}{\partial V_j} = -2 \sum_{i=1}^N u_{ij}^{m_i} |u_{ij} - \bar{u}_{ij}| (X_i - V_j) = 0
 \tag{7}$$

From Eq. (7), cluster centres V_j are defined by

$$V_j = \frac{\sum_{i=1}^N |u_{ij} - \bar{u}_{ij}|^{m_i} X_i}{\sum_{i=1}^N |u_{ij} - \bar{u}_{ij}|^{m_i}}, j = \overline{1, C}
 \tag{8}$$

On the other hand, taking the derivation of L by u_{ij} , we get

$$\frac{\partial L}{\partial u_{ij}} = m_i (u_{ij} - \bar{u}_{ij})^{m_i-1} \|X_i - V_j\|^2 - \lambda_i = 0
 \tag{9}$$

From Eqs. (3) and (9), membership degrees u_{ij} are calculated using (10):

$$\begin{aligned}
 u_{ij} = \bar{u}_{ij} + \left(1 - \sum_{j=1}^C \bar{u}_{ij} \right) \frac{\left(\frac{1}{\|X_i - V_j\|} \right)^{\frac{2}{m_i-1}}}{\sum_{j=1}^C \left(\frac{1}{\|X_i - V_j\|} \right)^{\frac{2}{m_i-1}}}, i = \overline{1, N}; j \\
 = \overline{1, C}
 \end{aligned}
 \tag{10}$$

The general diagram of MCSSFC-P is shown in Fig. 1 below.

In MCSSFC-P algorithm, input data and parameters are used to define fuzzifiers for each data sample by Eq. (1). Using these fuzzifiers, SSFCM algorithm is implemented to determine centres of clusters and membership degree matrix. From general diagram, the specific steps of MCSSFC-P are given as in Table 3 below.

3.2 Semi-supervised Fuzzy Clustering Model with Multiple Cluster Fuzzifiers (MCSSFC-C)

In fuzzy clustering, the fuzzifier represents the uncertainty of the data. In fuzzy clustering algorithms and the mostly standard fuzzy semi-supervised clustering algorithms, the fuzzy parameter (called fuzzifier also) is a constant value for all data sample (denoted by m and usually chosen as 2). Normally, this parameter is chosen based on the experience and is constant to all data samples. However, this parameter represents for the uncertainty of data sample belonging

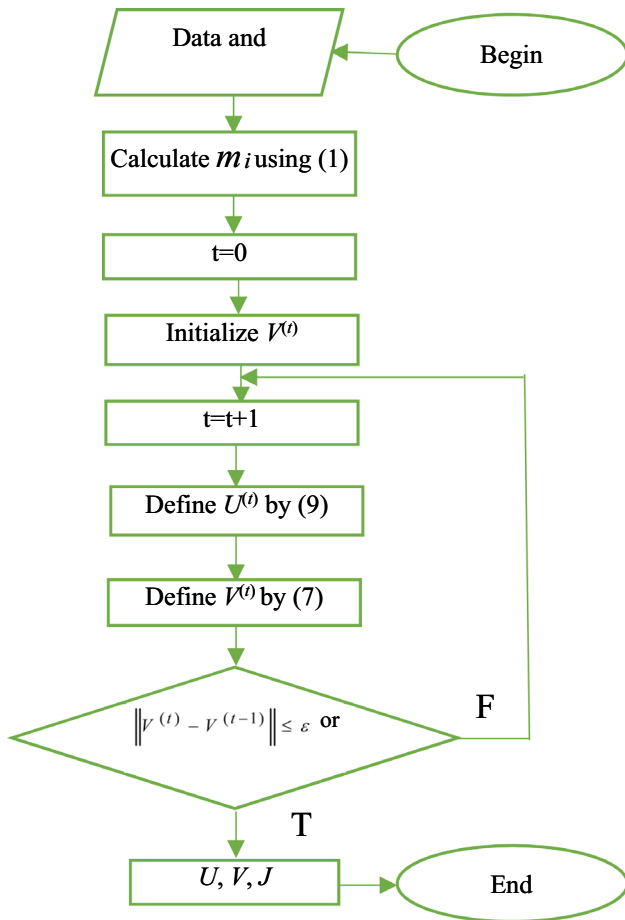


Fig. 1 General diagram of MCSSFC-P algorithm

to clusters. In fact, each data sample has the own uncertainty degree. The constant value of fuzzifier is not suitable in clustering process. In practice, however, each data sample exists a different degree of uncertainty. This affects to the uncertainty of the clusters. The more uncertainty data points a cluster has, the higher the uncertainty is. In the

case of using a constant value of fuzzifiers for all data samples, objective function is presented by

$$J(U, V) = \sum_{k=1}^N \sum_{j=1}^C \|u_{kj}\|^{m_i} \|X_k - V_j\|^2 \rightarrow \min \quad (11)$$

Because the data samples in a cluster have a certain similarity, we propose the membership function for the cluster (u_j) based on the membership function of the samples and the number of samples in that cluster at the time of consideration. Let C_j be the set of samples in the j th cluster, we have

$$u_j = \frac{\sum_{x_i \in C_j} u_{ij}}{|C_j|}, \quad j = \overline{1, C} \quad (12)$$

The objective function (11) achieves the minimum if and only if

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|X_i - V_j\|}{\|X_i - V_k\|} \right)^{\frac{2}{m_i-1}}}, \quad i = \overline{1, N}; j = \overline{1, C} \quad (13)$$

Thus, provided that the j th cluster has a membership function u_j , then from (12) and (13), we get

$$u_j = \frac{1}{|C_j|} \sum_{x_i \in X_j} \frac{1}{\sum_{k=1}^C \left(\frac{\|X_i - V_j\|}{\|X_i - V_k\|} \right)^{\frac{2}{m_i-1}}}; j = \overline{1, C} \quad (14)$$

Each cluster will also have a corresponding fuzzy parameter. Usually, the value of fuzzifier is chosen based on the experience without general formula. This is easy to see, the selection of m is only an approximation. Therefore, it is not feasible to find a formula to calculate the exact fuzzifier m .

In our research, we use different values of fuzzifier (called as multiple fuzzifiers) for each data point in MCSSFC-P or each cluster in MCSSFC-C. From the initial values, the multiple fuzzifiers may be updated in iterations to get the optimal values of objective function. From formula (14), considering as the approximate case, we have

Table 3 Semi-supervised fuzzy clustering algorithm with multiple point fuzzifiers

Input	The dataset X including N data elements $\{X_1, X_2, \dots, X_N\}$, number of clusters C , pre-defined membership degree \overline{U} , fuzzifiers m_i ($i = \overline{1, N}$); threshold ε , maximum iteration $MaxStep > 0$
Output	Matrices U and V
MCSSFC-P	
1	$t = 0$
2	Initialize random value for V^t
3	Repeat
3.1	$t = t + 1$
3.2	Calculate U^t using Eq. (9)
3.3	Calculate V^t using Eq. (7)
3.4	Until $\ V^t - V^{(t-1)}\ \leq \varepsilon$ or $t > MaxStep$

$$m_j = 1 + \frac{2}{\log u_j + \log |C_j|}, j = \overline{1, C} \tag{15}$$

The m_j computed by (15) satisfies the condition that the fuzzifier must be greater than or equal to 1. Apart from that, the fuzzifier needs to satisfy the condition inversely proportional to the value of the membership function. This is true of formula (15) because the larger the membership function of the cluster, the less uncertainty there is. Hence, the fuzzifier value is small and vice versa. With fuzzy cluster parameters, now the fuzzy membership function from formula (14) is replaced by formula (16) with fuzzifiers for each cluster.

$$u_j = \frac{1}{|C_j|} \sum_{X_i \in C_j} \frac{1}{\sum_{k=1}^C \left(\frac{\|X_i - V_j\|}{\|X_i - V_k\|} \right)^{2/(m_j-1)}}, j = \overline{1, C} \tag{16}$$

Based on the formula (12), membership degree of the j th cluster is $u_j \in \left(\frac{1}{c}, 1\right)$. Then, we have

$$\log \frac{|X_j|}{c} < \log u_j + \log |X_j| < \log |X_j|$$

The value of fuzzifier parameter of the j th cluster satisfies

$$1 + \frac{2}{\log |X_j|} < m_j < 1 + \frac{2}{\log \frac{|X_j|}{c}}$$

However, since the model is very sensitive when the fuzzifier changes, it is recommended to use one parameter to adjust the fuzzifier value. Moreover, increasing or decreasing the domain of values can give the highest accuracy. In our situation, the new objective function for the semi-supervised fuzzy algorithm with multiple fuzzifier values of clusters (m_j) will be

$$J(U, V) = \sum_{i=1}^N \sum_{j=1}^C \|u_{ij} - \bar{u}_{ij}\|^{m_j} * \|X_i - V_j\|^2 \rightarrow \min \tag{17}$$

Alpha coefficient is included to be flexible in adjusting the effects of fuzzy parameters. Because the algorithm is very sensitive to the change of the parameter m_j , we need to ensure to get the optimal target function after each loop. Hence, in each iteration, the values of the fuzzy parameters will be updated if the following condition is met:

$$\|J^{(t)} - J^{(t-1)}\| < \varepsilon \tag{18}$$

In the case of changing, the new fuzzy parameters m_j is used to update the values in U and V as below.

$$V_j = \frac{\sum_{i=1}^N |u_{ij} - \bar{u}_{ij}|^{m_j} * X_i}{\sum_{i=1}^N |u_{ij} - \bar{u}_{ij}|^{m_j}}, j = \overline{1, C} \tag{19}$$

When the fuzzifier m_j the value of the membership function is calculated by the following formula:

$$u_{ij} = \bar{u}_{ij} + \left(1 - \sum_{k=1}^C \bar{u}_{ik}\right) \frac{\left(\frac{1}{\|X_i - V_j\|}\right)^{\frac{2}{m_j-1}}}{\sum_{k=1}^C \left(\frac{1}{\|X_i - V_k\|}\right)^{\frac{2}{m_j-1}}}, i = \overline{1, N}; j = \overline{1, C} \tag{20}$$

In which, pre-defined membership degree (\bar{u}_{ij}) in this algorithm is calculated according to the data. If the data set has no labelled data, the values of \bar{u}_{ij} are computed as follows:

$$\bar{u}_{ij} = \begin{cases} u_{ij} & u_{ij} = \max_{k=\overline{1, C}}(u_{ik}) \\ 0 & u_{ij} \neq \max_{k=\overline{1, C}}(u_{ik}) \end{cases}, i = \overline{1, N}; j = \overline{1, C} \tag{21}$$

With labelled data, the values of \bar{u}_{ij} are defined as follows:

$$\bar{u}_{ij} = \begin{cases} 1 & i \text{ is labelled class } j \\ 0 & \text{otherwise} \end{cases}, i = \overline{1, N}; j = \overline{1, C} \tag{22}$$

From the above analysis, the paper proposes a semi-supervised fuzzy clustering algorithm with multiple fuzzifiers for clusters (MCSSFC-C, Fig. 2). In this approach, the initial values of fuzzifiers for all clusters are equal, $m'_j = 2$ for all $j = 1, \dots, C$. During clustering, this parameter will only change when condition (18) is satisfied. This means that only fuzzifier updates for clusters when clustering results in the next iteration are better than in the previous iteration. This is the key point of the proposed algorithm. In the worst case, all fuzzifier changes do not yield better clustering results. The proposed algorithm will become the SSFCM algorithm.

3.3 Numerical examples to illustrate the steps of the algorithm

To illustrate the proposed algorithm MCSSFC-C, we give an example with 10 data samples, 4 attributes and 2 clusters as shown in Table 4.

Numerical examples performed on the data set are shown in Table 4 with 10 data samples clustered in 2 classes.

The values obtained by this process are $\mathbf{m}_1 = 7.0616$; $\mathbf{m}_2 = 3.993$ based on formula (5) and $\mathbf{U}_1 = 0.9345$; $\mathbf{U}_2 = 0.9501$ based on formula (6). The process in Table 5 is repeated until the stop condition is satisfied.

Figure 3 describes the data distribution and clustering results in a 3-dimensional space.

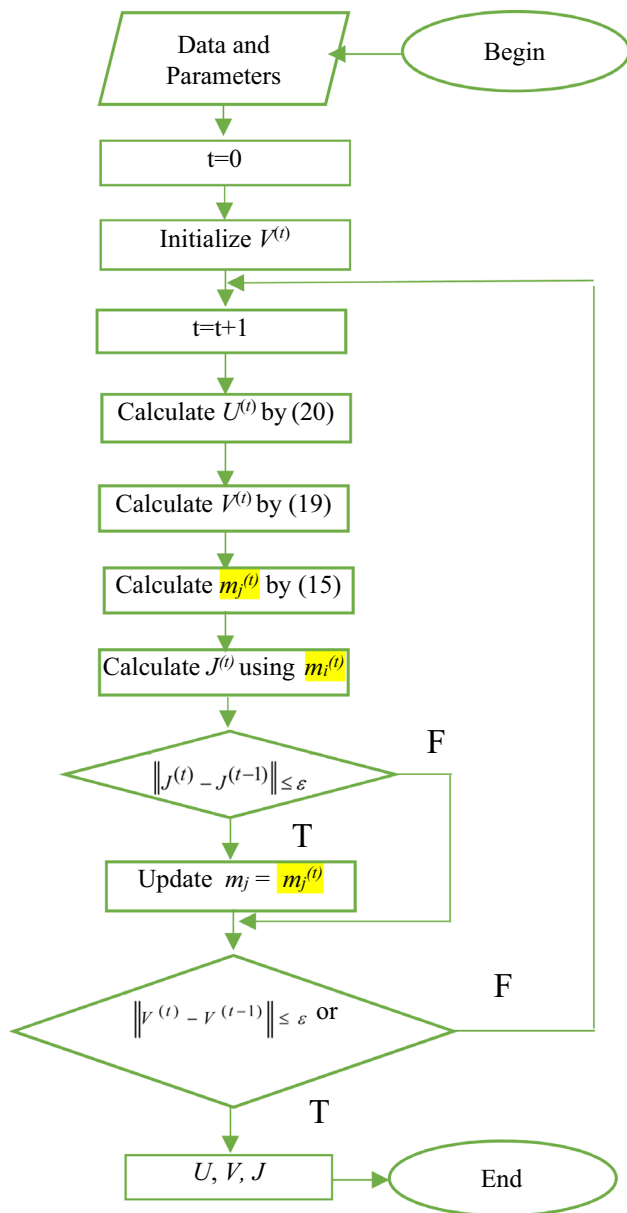


Fig. 2 Semi-supervised fuzzy clustering algorithm with multiple cluster fuzzifiers

3.4 Discussions

The goal of calculating fuzzifiers values for the cluster is to optimize the objective function. In the worst case, the fuzzifiers do not change. It means that fuzzifiers for the clusters are always equal to the initial initialization value $m_j = 2$. Then, the accuracy of MCSSF-C is equivalent to that of SSFCM. The advantage of the proposed algorithm is to find fuzzifiers for each cluster. This is more suitable than using only one fuzzifier for all data sets.

It can be seen that the computational complexity of the proposed algorithm includes the complexity of steps 1 to 7 in Table 3. Including the complexity of the FCM algorithm

Table 4 The data illustrate the proposed algorithm

No.	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	Label
1	6	2.2	5	1.5	1
2	7.2	3	5.8	1.6	1
3	6.3	2.8	5.1	1.5	1
4	6.1	2.6	5.6	1.4	1
5	6.7	2.5	5.8	1.8	1
6	6.3	2.3	4.4	1.3	2
7	5.6	3	4.1	1.3	2
8	5.5	2.5	4	1.3	2
9	5.5	2.6	4.4	1.2	2
10	6.1	3	4.6	1.4	2

Table 5 The process of define membership matrix and centres of clusters

V_1	6.0405	2.9237	4.5839	1.3916
V_2	6.2660	2.7426	5.2402	1.4885

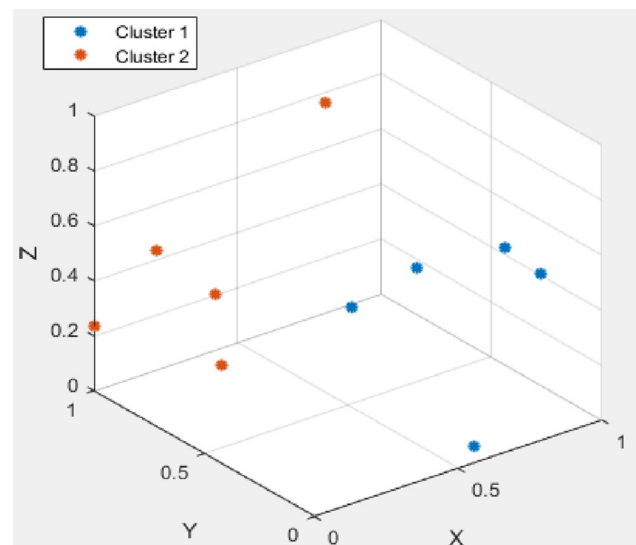


Fig. 3 Data distribution and clusters in 3-dimensional space

$O(T_1ndc^2)$ and in each loop: Update the fuzzy membership function U by Eqs. (10) or (11), update the V by Eq. (9), calculate $J(U, V)$ by Eq. (7), update the fuzzy parameters and check for stop condition has a computational complexity $O(ndc^2)$ and calculate potential fuzzy parameters for each cluster by Eq. (5) and cluster fuzzy membership by Eq. (6) has a computational complexity $O(2ndc^2)$. Therefore, the computational complexity of MCSSF-C algorithm is $O(3T_2ndc^2)$. T_1 and T_2 are the number of iterations of FCM and MCSSF-C, respectively.

Although when the number of samples approaches infinity, the computational complexity between the SSFCM algorithm and MCSSFC-C is equivalent. Compared with the SSFCM algorithm, the MCSSFC-C algorithm adds steps 4 and 6. These steps will allow the algorithm to try on many different fuzzifiers values, leading to slower convergence of this algorithm due to the need to calculate and test the optimal conditions for choosing fuzzy parameters. MCSSFC-C algorithm will have 3 times higher in computing complexity than SSFCM algorithms. Thus, in the actual test, the MCSSFC-C algorithm runs slower than the SSFCM algorithm.

In this research, we proposed two improvements of SSFCM model, including MCSSFC-P model and MCSSFC-C model with some main advantages as below:

- MCSSFC-P is an extension of SSFCM algorithm in the case of using fuzzifiers for each data sample. Additional information is provided by experts or based on the previous analyses. This information will increase the accuracy of clustering. Using additional information, the different values of fuzzifiers are flexibly defined for each sample based on its membership degree, instead of using a fixed fuzzifier for all samples in SSFCM.
- MCSSFC-C defines the values of fuzzifiers for each cluster. This reduces the time computing comparing with that of MCSSFC-P. Moreover, the samples in the same cluster having the same fuzzifiers will increase the performance of this model.

However, MCSSFC-C still have some limitations, including

- The values of fuzzifier parameters in MCSSFC-C are defined mainly based on the experiences of our research team.
- The analyses of types of data that are suitable to the proposals in this paper are not complete.

4 Experimental Results

We have been using the Dell laptop with Core i5 processor for implementing the algorithms in MATLAB 2014. In the evaluation, we use both two-label datasets and multi-label datasets. The two-label datasets are three benchmark datasets (Wine, Iris, Breast, Liver, Diabetes, Arrhythmia datasets) taken from the benchmark UCI Machine Learning Repository in Table 6.

The experiment of data classification on algorithms SSFCM [42], MCSSFC-P (taken according to each data point) and MCSSFC-C (MCSSFC taken according to each data cluster) in both cases: Using label data and not using label data to calculate the supplementary membership function values. Experimental parameters include $\epsilon = 0.001$, $\text{Maxstep} = 1000$ and $\alpha = 1.0$. Indexes Davies-Bouldin (DB), PBM, Partition Coefficient (PC) and Classification Entropy (CE) are used to evaluate the effectiveness of the algorithms. Furthermore, the accuracy (Acc) of classification results is also assessed by the percentage of the number of samples correctly classified to the total sample of the dataset. In which, Acc1: calculated on data know labels; Acc2: calculated on the all dataset.

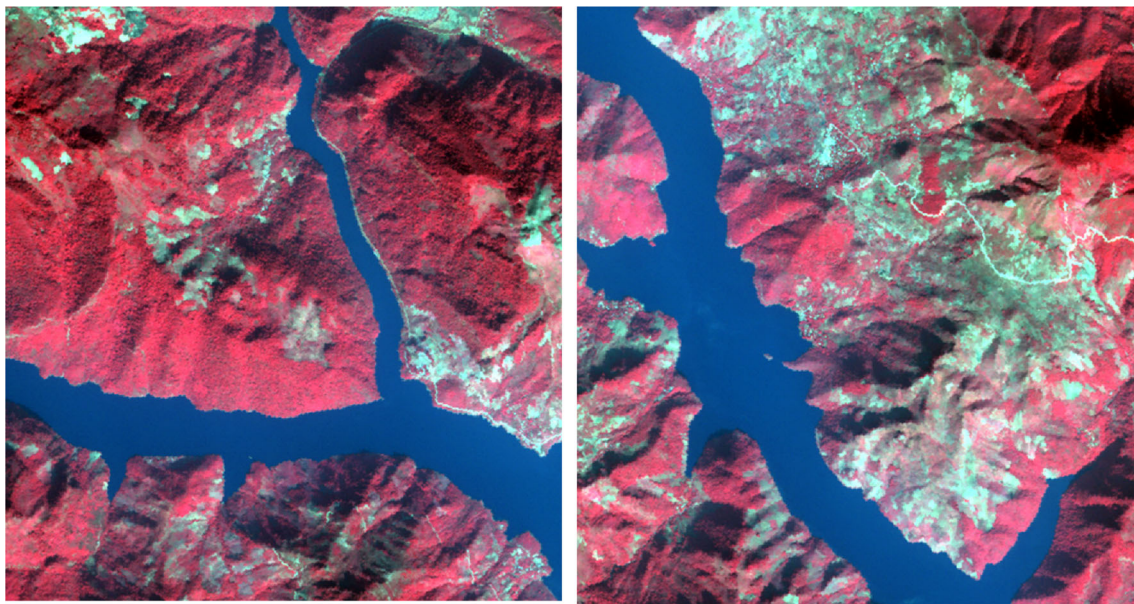
Remote sensing images in 2 regions (a and b) of BacDa, Hoabinh with size 515×515 are used for the experiment (see Fig. 4). With the parameters for the algorithms set up as above, the remote sensing image data are classified in 4 classes corresponding to 4 different types of land covers. Remote sensing image data are classified according to 4 classes.

4.1 Experimental Results on the Benchmark UCI Datasets

Label data were obtained at the rate of 10% on experimental data set. The supplementary membership function is calculated from the labelled data. Table 7 shows the values of the indexes used to evaluate cluster quality of the SSFCM, SSFC-FS [38], MCSSFC-P and MCSSFC-C

Table 6 Experimental datasets

No.	Dataset	No. of attributes	No. of instances	No. of labels
1	Wine quality	11	1599	6
2	Iris	4	150	3
3	Breast	9	680	2
4	Liver	10	4156	2
5	Diabetes	5	390	2
6	Arrhythmia	36	452	13



a Image 1

b Image 3

Fig. 4 Remote sensing image data

Table 7 Experimental results on datasets using labelled data

Dataset	Method	DB−	PBM+	CE−	PC+	ACC1+	ACC2+
Wine	SSFCM	1.34	13.75	0.5176	0.7397	0.95	0.89
	SSFC-FS	1.29	17.37	0.5175	0.7299	0.95	0.90
	MCSSFC-P	1.03	19.67	0.5173	0.7401	0.97	0.91
	MCSSFC-C	2.23	16.27	0.5174	0.7464	0.96	0.90
Iris	SSFCM	4.76	0.73	0.3632	0.7902	0.92	0.87
	SSFC-FS	4.47	0.78	0.3528	0.7827	0.89	0.88
	MCSSFC-P	4.79	0.85	0.3425	0.8037	0.97	0.89
	MCSSFC-C	3.27	1.27	0.3951	0.8009	0.98	0.89
Breast	SSFCM	7.76	2.48	0.2143	0.8927	0.97	0.90
	SSFC-FS	6.37	4.25	0.2093	0.8982	0.97	0.91
	MCSSFC-P	5.76	3.24	0.2049	0.9027	0.97	0.92
	MCSSFC-C	9.36	5.26	0.1973	0.9127	0.99	0.92
Liver	SSFCM	3.44	229.70	0.1726	0.9241	0.93	0.78
	SSFC-FS	3.27	528.93	0.1627	0.9536	0.94	0.81
	MCSSFC-P	2.99	707.26	0.1532	0.9673	0.96	0.83
	MCSSFC-C	4.26	607.27	0.1627	0.9325	0.95	0.87
Diabetes	SSFCM	2.43	456.38	0.2143	0.8937	0.93	0.86
	SSFC-FS	2.41	492.89	0.1982	0.8927	0.94	0.87
	MCSSFC-P	2.33	524.37	0.1726	0.9207	0.93	0.87
	MCSSFC-C	1.89	501.28	0.1827	0.9342	0.94	0.88
Arrhythmia	SSFCM	2.93	503.39	0.2479	0.0873	0.84	0.79
	SSFC-FS	2.89	512.93	0.2483	0.0903	0.88	0.81
	MCSSFC-P	3.43	533.76	0.2467	0.0892	0.90	0.82
	MCSSFC-C	2.37	502.36	0.2342	0.0932	0.89	0.83

Good value of each dataset on measure

algorithms using labelled data. SSFCM and SSFC-FS use a constant value of fuzzifier, while MCSSFC-P and MCSSFC-C use multiple fuzzifier in clustering progress.

By using DB and PBM indices, MCSSFC-C algorithm achieved the highest accuracy at 3/6 datasets, followed by MCSSFC-P algorithm with 3/6 datasets. With CE index, MCSSFC-P algorithm achieved the highest accuracy at 4/6 datasets, followed by MCSSFC-C algorithm with 2/6 datasets. With PC index, MCSSFC-C algorithm achieved the highest accuracy at 4/6 datasets, followed by MCSSFC-P algorithm with 2/6 datasets. With ACC1 index, MCSSFC-C algorithm achieved the highest accuracy at 3/6 datasets, followed by MCSSFC-P algorithm with 3/6 datasets. SSFC-FS gets the best result of ACC1 on Diabetes dataset. With ACC2 index, MCSSFC-C algorithm achieved the highest accuracy at 3/6 datasets, followed by MCSSFC-P algorithm with 1/6 datasets. The results of ACC2 obtained by applying two proposed algorithms are equal on 2/6 datasets, including Iris and Breast.

Based on results in Table 7 on 6 datasets (Wine, Iris, Breast, Liver, Diabetes and Arrhythmia) and 6 assessment indicators (DB, PBM, CE, PC, ACC1 and ACC2) of cluster quality with 36 comparisons, the best results are mainly obtained by using MCSSFC-P and MCSSFC-P. Both the proposed algorithms achieved the best 19/36 times. SSFC-FS gets the best result only once. It means that the

algorithms with multiple fuzzifiers get the higher cluster quality.

4.2 Experimental Results on the Remote Sensing Image Data

In these experiments, additional information is defined from the results of the FCM. The number of clusters is chosen as 4. Figures 5 and 6 show the two original images and clustered images by using SSFCM, MCSSFC-P and MCSSFC-C, respectively.

Apart from illustrating by the images above, the validity indices are also calculated and are given in Table 8 below.

Table 8 Experimental results on image datasets

Images	Method	DB-	PBM+	CE-	PC+
Image 1	SSFCM	38.5467	2.11E-06	0.7836	0.5855
	MCSSFC-P	38.1875	1.88E-06	0.1074	0.9287
	MCSSFC-C	37.9630	2.12E-06	0.1177	0.9316
Image 2	SSFCM	43.3027	1.84E-06	0.8002	0.5739
	MCSSFC-P	42.1076	1.94E-06	0.1654	0.9090
	MCSSFC-C	42.1704	2.02E-06	0.1260	0.9256

Good value of each dataset on measure

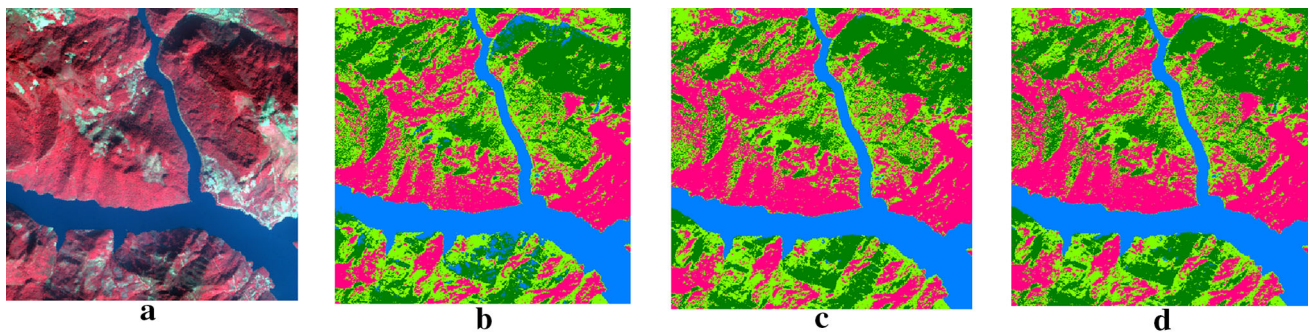


Fig. 5 Result of remote sensing image classification on image 1: a) Color image; b) SSFCM; c) MCSSFC-P and d) MCSSFC-C

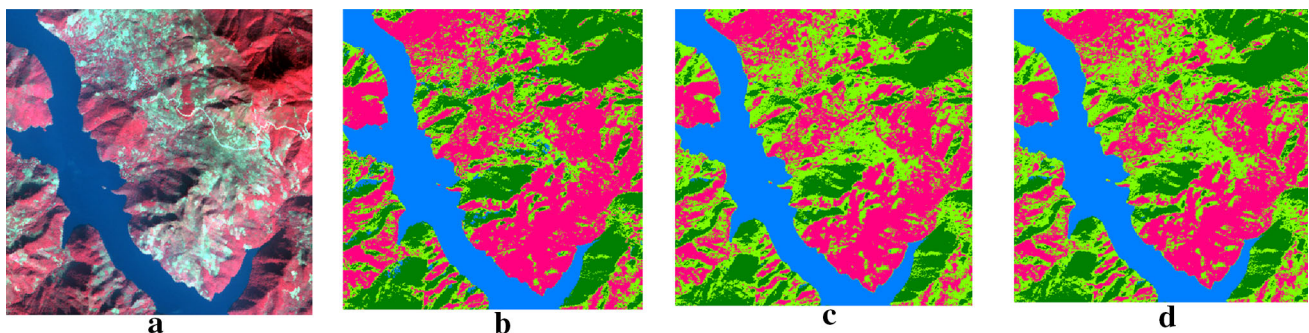


Fig. 6 Result of remote sensing image classification on images 2: a) Color image; b) SSFCM; c) MCSSFC-P and d) MCSSFC-C

Table 9 Fuzzy parameters achieved by algorithm MCSSFC-C in clusters

Images	m_1	m_2	m_3	m_4
Image 1	1.15483755685461	1.1771590526375	1.1601659897656	1.16236784549426
Image 2	1.17832035030204	1.1768389497498	1.1851293626568	1.18650726701434

Based on the values in Table 8, MCSSFC-P and MCSSFC-C give better results than SSFCM in most cases. The values of PBM, PC obtained by MCSSFC-C are highest on both selected images. With values DB, MCSSFC-C still gets the best result on Image 1 and MCSSFC-P gets the best result on Image 2. With values CE, MCSSFC-C still gets the best result on Image 2 and MCSSFC-P gets the best result on Image 1.

The values of fuzzifiers obtained by applying MCSSFC-C on these images are presented in Table 9.

In both Tables 8 and 9, the experimental results show that the use of multiple fuzzy parameters (MCSSFC-P and MCSSFC-C algorithms) can improve the accuracy of clustering results compared to using only one fuzzifier (SSFCM algorithm).

5 Conclusions

In this paper, we proposed two improvements of semi-supervised standard fuzzy C-Mean clustering by using multiple fuzzifiers. The contributions of this research include (i) to improve clustering quality in SSFCM by using different fuzzifiers for each data point (MCSSFC-P); (ii) to improve clustering quality and reduce the run time by using different fuzzifiers for each data cluster (MCSSFC-C); (iii) to implement two of these improvements on UCI datasets and remote sensing images to evaluate the performance of proposed models and compare these models to other related models.

The results evaluated on indicators (DB, PBM, CE, PC and Acc) show that the proposed method can give better clustering results in most of the indicators used. Meanwhile, algorithm MCSSFC-C reached, respectively, 19/36; MCSSFC-P algorithm reached 19/36, respectively. From the experimental results on remote sensing image data, it can be seen that MCSSFC-C algorithm gives better classification results than other algorithms based on PBM, PC indexes. With the above results, it can be confirmed that the use of many fuzzy parameters can improve the quality of data clustering.

From the experimental results, different clusters or data samples will have different levels of uncertainty, the advantage of the proposed algorithm can give fuzzy parameters for each cluster or point. This provides greater precision than just using one fuzzifier for all data sets.

However, the proposed algorithm with the addition of steps to find fuzzifier values leads to greater computational complexity and more time.

In the next studies, we will develop an algorithm to consider the effects of fuzzy coefficients for each cluster and data sample. Moreover, we also improve the algorithm for multiple fuzzifiers based on 3-dimensional functions, depending on U, V and X. Improving the algorithm for multiple fuzzifiers with mixed complementary information, including Label, Constraints, Membership, Fuzzifiers. Furthermore, the optimal parameters of the proposed algorithms can be determined by meta-heuristics algorithms such as GA, PSO, etc.

Acknowledgements This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant number. 102.05-2020.11.

Declarations

Conflict of interest No potential conflict of interest was reported by the authors.

References

- Abdar, M., et al.: BARF: a new direct and cross-based binary residual feature fusion with uncertainty-aware module for medical image classification. *Inf. Sci.* **577**, 353–378 (2021)
- Arshad, A., et al.: Semi-supervised deep fuzzy c-mean clustering for software fault prediction. *IEEE Access* **6**, 25675–25685 (2018)
- Arshad, A., Riaz, S., Jiao, L.: Semi-supervised deep fuzzy C-mean clustering for imbalanced multi-class classification. *IEEE Access* **7**, 28100–28112 (2019)
- Banerjee, A., Maji, P.: Segmentation of bias field induced brain MR images using rough sets and stomped-t distribution. *Inf. Sci.* **504**, 520–545 (2019)
- Bezdek, J.C., Ehrlich, R., Full, W.: FCM: the fuzzy c-means clustering algorithm. *Comput. Geosci.* **10**(2–3), 191–203 (1984)
- Cai, W. et al.: An adaptive approach of feature selection applied to semi-supervised fuzzy clustering. In: *Proceedings of the 2020 4th International Conference on Electronic Information Technology and Computer Engineering* (2020)
- Cao, L., Wang, C., Li, J.: Vehicle detection from highway satellite images via transfer learning. *Inf. Sci.* **366**, 177–187 (2016)
- Casalino, G., Castellano, G., Mencar, C.: Incremental adaptive semi-supervised fuzzy clustering for data stream classification. In: *2018 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*. IEEE (2018)
- Casalino, G., Castellano, G., Mencar, C.: Data stream classification by dynamic incremental semi-supervised fuzzy clustering. *Int. J. Artif. Intell. Tools* **28**(08), 1960009 (2019)

10. Casalino, G. et al.: Incremental semi-supervised fuzzy C-Means for bipolar disorder episode prediction. In: 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS). IEEE (2020)
11. Curiskis, S.A., et al.: An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. *Inf. Process. Manag.* **57**(2), 102034 (2020)
12. Ding, W., Abdel-Basset, M., Hawash, H.: RCTE: A reliable and consistent temporal-ensembling framework for semi-supervised segmentation of COVID-19 lesions. *Inf. Sci.* **578**, 559–573 (2021)
13. Frigui, H., Bchir, O., Baili, N.: An overview of unsupervised and semi-supervised fuzzy kernel clustering. *Int. J. Fuzzy Logic Intell. Syst.* **13**(4), 254–268 (2013)
14. Gan, H., et al.: Local homogeneous consistent safe semi-supervised clustering. *Expert Syst. Appl.* **97**, 384–393 (2018)
15. Gan, H., et al.: Confidence-weighted safe semi-supervised clustering. *Eng. Appl. Artif. Intell.* **81**, 107–116 (2019)
16. Gan, H.: Safe semi-supervised fuzzy C-means clustering. *IEEE Access* **7**, 95659–95664 (2019)
17. Grira, N., Crucianu, M., Boujemaa, N.: Active semi-supervised fuzzy clustering. *Pattern Recogn.* **41**(5), 1834–1844 (2008)
18. Huang, H.-C., Chuang, Y.-Y., Chen, C.-S.: Multiple kernel fuzzy clustering. *IEEE Trans. Fuzzy Syst.* **20**(1), 120–134 (2011)
19. Juang, C.-F., Hsieh, C.-D.: Fuzzy C-means based support vector machine for channel equalisation. *Int. J. Gen. Syst.* **38**(3), 273–289 (2009)
20. Kanzawa, Y.: Semi-supervised fuzzy c-means algorithms by revising dissimilarity/kernel matrices. In: *Fuzzy Sets, Rough Sets, Multisets and Clustering*, pp. 45–61. Springer, Cham (2017)
21. Khang, T.D., et al.: Fuzzy c-means clustering algorithm with multiple fuzzification coefficients. *Algorithms* **13**(7), 158 (2020)
22. Knorr, W., et al.: Combined use of weather forecasting and satellite remote sensing information for fire risk, fire and fire impact monitoring. *Comput. Ecol. Softw.* **1**(2), 112 (2011)
23. Kumar, A., Bhadauria, H.S., Singh, A.: Semi-supervised OTSU based hyperbolic tangent Gaussian kernel fuzzy C-mean clustering for dental radiographs segmentation. *Multimed. Tools Appl.* **79**(3), 2745–2768 (2020)
24. Li, X., et al.: Semi-supervised clustering with deep metric learning and graph embedding. *World Wide Web* **23**(2), 781–798 (2020)
25. Li, Z., et al.: Crowdsourcing logistics pricing optimization model based on DBSCAN clustering algorithm. *IEEE Access* **8**, 92615–92626 (2020)
26. Ma, M., et al.: Semi-supervised rough fuzzy Laplacian eigenmaps for dimensionality reduction. *Int. J. Mach. Learn. Cybern.* **10**(2), 397–411 (2019)
27. Mai, D.S., Ngo, L.T.: General semi-supervised possibilistic fuzzy c-means clustering for land-cover classification. In: 2019 11th International Conference on Knowledge and Systems Engineering (KSE). IEEE (2019)
28. Mai, D.S., Ngo, L.T.: Semi-supervised fuzzy C-means clustering for change detection from multispectral satellite image. In: 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). IEEE (2015)
29. Mai, D.S., Ngo, L.T., Hagrais, H.: A hybrid interval type-2 semi-supervised possibilistic fuzzy c-means clustering and particle swarm optimization for satellite image analysis. *Inf. Sci.* **548**, 398–422 (2021)
30. Maraziotis, I.A.: A semi-supervised fuzzy clustering algorithm applied to gene expression data. *Pattern Recogn.* **45**(1), 637–648 (2012)
31. Narayana, G.S., Kolli, K.: Fuzzy K-means clustering with fast density peak clustering on multivariate kernel estimator with evolutionary multimodal optimization clusters on a large dataset. *Multimed. Tools Appl.* **80**(3), 4769–4787 (2021)
32. Novák, V., et al.: Analysis of seasonal time series using fuzzy approach. *Int. J. Gen. Syst.* **39**(3), 305–328 (2010)
33. Salehi, F., Keyvanpour, M.R., Sharifi, A.: SMKFC-ER: Semi-supervised multiple kernel fuzzy clustering based on entropy and relative entropy. *Inf. Sci.* **547**, 667–688 (2021)
34. Sato-Ilic, M.: Dynamic fuzzy clustering using fuzzy cluster loading. *Int. J. Gen. Syst.* **35**(2), 209–230 (2006)
35. Śmieja, M., Struski, Ł., Figueiredo, M.A.T.: A classification-based approach to semi-supervised clustering with pairwise constraints. *Neural Netw.* **127**, 193–203 (2020)
36. Son, L.H., Thong, P.H.: Some novel hybrid forecast methods based on picture fuzzy clustering for weather nowcasting from satellite image sequences. *Appl. Intell.* **46**(1), 1–15 (2017)
37. Son, L.H., Thong, P.H.: An overview of semi-supervised fuzzy clustering algorithms. *Int. J. Eng. Technol.* **8**(4), 301 (2016)
38. Son, L.H., Tuan, T.M.: Dental segmentation from X-ray images using semi-supervised fuzzy clustering with spatial constraints. *Eng. Appl. Artif. Intell.* **59**, 186–195 (2017)
39. Son, L.H., Tuan, T.M.: A cooperative semi-supervised fuzzy clustering framework for dental X-ray image segmentation. *Expert Syst. Appl.* **46**, 380–393 (2016)
40. Tuan, T.M., Ngan, T.T., Son, L.H.: A novel semi-supervised fuzzy clustering method based on interactive fuzzy satisficing for dental X-ray image segmentation. *Appl. Intell.* **45**(2), 402–428 (2016)
41. Xiong, J., et al.: Semi-supervised fuzzy c-means clustering optimized by simulated annealing and genetic algorithm for fault diagnosis of bearings. *IEEE Access* **8**, 181976–181987 (2020)
42. Yasunori, E. et al.: On semi-supervised fuzzy c-means clustering. In: 2009 IEEE International Conference on Fuzzy Systems. IEEE (2009)
43. Yu, Z., et al.: Adaptive ensembling of semi-supervised clustering solutions. *IEEE Trans. Knowl. Data Eng.* **29**(8), 1577–1590 (2017)
44. Zhou, F., et al.: MetaRisk: semi-supervised few-shot operational risk classification in banking industry. *Inf. Sci.* **552**, 1–16 (2021)

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Fuzzy Computing.

Tran Manh Tuan received the Bachelor degree in Applied Mathematics and Informatics from Hanoi University of Science and Technology in 2003 and Master degree in Computer Science from Thainguyen University in 2007. He got Doctoral degree in July, 2017. Now, he is working as a lecturer at Faculty of Computer Science and Engineering, Thuyloi University. His researches focus on Artificial Intelligence, Data Mining, Soft Computing and



Mai Dinh Sinh received the Bachelor degree in GeoInformatics from Le Quy Don Technical University (LQDTU), Vietnam in 2009. He got the Master degree in Computer Science in 2013 and the Doctoral degree in Mathematical Foundations for Informatics in 2021 at LQDTU. He is also a lecturer at Institute of Techniques for Special Engineering, LQDTU. His research interests are Fuzzy logic, Interval type-2 fuzzy sets, Computational

Intelligence, Pattern Recognition and GeoInformatics.



Tran Dinh Khang is an Associated Professor of computer science at Hanoi University of Science and Technology (HUST), Vietnam. He received his Ph.D. in computer science from the Hanoi University of Science and Technology in 1999. He was a visiting scholar at the Institute of Artificial Intelligent, TU Dresden, in 2002, 2004. His current research interests are linguistic fuzzy logic using the hedge algebraic structure, hedge algebraic type-

2 fuzzy sets and fuzzy systems and decision support systems. He has published more than 100 refereed articles.



Phung The Huan is a researcher and a PhD student at University of Information and Communication Technology, Thai Nguyen University (ICTU). He received the Bachelor degree in Information Technology from ICTU, Vietnam in 2009. He got the Master degree in Computer Science from ICTU in 2012. He is also a lecturer at Faculty of Information Technology, ICTU. His research interests include Geographic information systems,

Pattern Recognition, Data Mining, Soft Computing and Fuzzy Computing.



Tran Thi Ngan received the Bachelor degree in Applied Mathematics and Informatics from VNU University of Science, Vietnam National University in 2003. She got the Master degree in Computer Science from Thainguyen University in 2007. She received PhD degree in August, 2014. She was promoted to Associate Professor in Information Technology in Vietnam in January 2021. Now, she is working at Faculty of Computer

Science and Engineering, Thuyloi University. Her research interests are Optimization, Machine learning, Data Mining and Fuzzy Computing.



Nguyen Long Giang obtained the phd degree in math fundamentals for informatics from Institute of Information Technology (IOIT), Vietnam Academy of Science and Technology (VAST) in 2012. He has been promoted to associate professor in information technology since 2017. Dr. Giang worked as senior researcher and head of department at the IOIT, VAST during 2009–2019. His major field includes artificial intelligence, data mining, soft computing, fuzzy machine learning, rough sets and fuzzy rough sets.

computing, fuzzy machine learning, rough sets and fuzzy rough sets.



Vu Duc Thai now works in Thai Nguyen University of Information and Communication Technology. He graduated from Hanoi University of Science and Technology, received a Master's degree from Thai Nguyen University and received a doctorate from the Institute of Information Technology—Vietnam Academy of Science and Technology. The main research is parallel computing on chip with Cellular Neural Network technology and optimal

problems.