An Overview of Semi-Supervised Fuzzy Clustering Algorithms

Pham Huy Thong and Le Hoang Son

***Abstract*—Fuzzy clustering plays an important role in data-mining, especially in decision making, pattern recognition, etc. There have been many approaches to improve fuzzy clustering performance and quality when it was first introduced by Bezdek. Recently, an approach related to data with sub-information has been most concerned. The idea of this approach combines the advantages of fuzzy C-means method with the benefits of additional information so-called the semi-supervised fuzzy clustering algorithms (SSFC). Through this report, a series of typical SSFC algorithms are presented in brief to give an overview of this approach.**

***Index Terms*—Clustering performance, data mining, fuzzy clustering, fuzzy c-means, semi-supervised fuzzy clustering algorithm.**

I. INTRODUCTION

Fuzzy clustering is a group of algorithms for clustering analysis, in which the data elements are distributed to the cluster is not "clear" (elements belong to only one cluster) that are "fuzzy" in the sense of similarity in fuzzy logic where each data element has a membership degree on a cluster. Fuzzy clustering is an important tool in pattern recognition and knowledge discovery from databases and it has been applied in many practical applications such as image segmentation [1]-[3], face recognition [4], [5], risk assessment [6], geographic demographic analysis [7], intrusion detection [8], [9], bankruptcy prediction [10], etc. In recent studies, some additional information provided by users is engaged in the input of fuzzy clustering to guide, monitor and control the process of clustering [6]-[10]. A group of fuzzy clustering algorithms that combine the advantages of fuzzy logic and the benefits of additional information is called the semi-supervised fuzzy clustering algorithm (SSFC).

There are several additional information in SSFC. The first one is the must-link and cannot-link constraints [11]. The former requires that two elements must be in the same cluster. On the contrary, the latter shows that a couple of elements cannot be in a cluster. The second one is the class labels when a part of data label is known and the rest of them are not [12], [13]. The last one is the dependence of the data [18], [19] where membership degrees are given to support fuzzy clustering algorithms to achieve better clustering quality and to focus on the object that user need to orientate. Consequently, the additional information consists of three categories:

1) The must-link and cannot-link constraints [11];
2) Class labels of a part of data [12], [13];
3) Pre-defined membership degrees [18], [19].

The rest of this paper is organized as follows. In Section II, we recall the FCM algorithm. From Sections III to Section VII, we give an overview of semi-supervised fuzzy clustering algorithms such as the active semi-supervised fuzzy clustering, the semi-supervised standard fuzzy clustering, the semi-supervised entropy regularized fuzzy clustering, the semi-supervised kernel fuzzy c-mean clustering and semi-supervised kernel fuzzy clustering. Section VIII draws the conclusion and further works.

II. FUZZY C-MEANS

The best well-known fuzzy clustering algorithm is Fuzzy C-means method. Bezdek [16] proposed the fuzzy clustering problem where the membership degree of a data element X_k to cluster j^{th} denoted by the term u_{kj} was appended to the objective function in equation (3). A data element could belong to some clusters depending on the membership degree.

$$J = \sum_{k=1}^{N}\sum_{j=1}^{C} {u_{kj}}^m \left\| X_k - V_j \right\|^2 \to \min \qquad (1)$$

- m is fuzzy number.
- C is the number of clusters, N is the number of data elements, r is the number of dimensions of data.
- u_{kj} is the membership degree of element X_k in cluster j.
- $X_k \in \Re^r$ is data k of $X = \{X_1, X_2, ..., X_N\}$.
- V_j is cluster center.

The constraints for (1) are,

$$\sum_{j=1}^{C} u_{kj} = 1 \quad ; u_{kj} \in [0,1] \quad , \forall k = \overline{1, N} \qquad (2)$$

By using the Lagranian method, the membership degrees and centers of the problem (1-1) are calculated as follows.

$$V_j = \frac{\sum_{k=1}^{N} {u_{kj}}^m X_k}{\sum_{k=1}^{N} {u_{kj}}^m} \qquad (3)$$

Manuscript received February 4, 2015; revised March 23, 2015.
the authors are with VNU University of Science, Vietnam National University, Vietnam (e-mail: thongph@vnu.edu.vn, sonlh@vnu.edu.vn).

DOI: 10.7763/IJET.2016.V8.902

$$u_{kj} = \frac{1}{\sum_{i=1}^{C}\left(\frac{\|X_k - V_j\|}{\|X_k - V_i\|}\right)^{\frac{2}{m-1}}} \quad (4)$$

The fuzzy C-means algorithm is as follow (see Table I):

TABLE I: FUZZY C-MEANS ALGORITHM.

Fuzzy C-means Algorithm	
I:	Data X whose number of elements (N) in r dimensions; Number of clusters (C); the fuzzifier m , threshold ε , maximum iteration *maxSteps > 0.*
O:	Matrices u and centers V ;
FCM:	
1:	$t = 0$
2:	$u_{kj}^{(t)} \leftarrow random$; ($k=\overline{1,N}$, $j=\overline{1,C}$) satisfy equation (2)
3:	Repeat
4:	$t = t + 1$
5:	Calculate $V_j^{(t)}$ ($j=\overline{1,C}$) by equation (3)
6:	Calculate $u_{kj}^{(t)}$ ($k=\overline{1,N}$; $j=\overline{1,C}$) by equation (4)
7:	Until $\left\|u^{(t)} - u^{(t-1)}\right\| \le \varepsilon$ or *maxSteps* has reached

III. Active Semi-Supervised Fuzzy Clustering

Class (1): The additional information is must-link and cannot-link constraints.

Grira *et al*. [11] used this sub-information to support the clustering with constraints in pair. Suppose that M is a set of must-link constraints $(x_i, x_j) \in M$, it means that x_i and x_j are belonging to the same cluster and Δ is a set of cannot-link constraints $(x_i, x_j) \in \Delta$, it means that x_i and x_j are belonging to different clusters.

The objective function of AFCC is as follows:

$$J(U,V) = \sum_{k=1}^{N}\sum_{j=1}^{C} u_{kj}^2 \left\|X_k - V_j\right\|^2 + \alpha \sum_{(x_k,x_h)\in M} \sum_{j=1}^{C} \sum_{l=1, l\ne j}^{C} u_{kj}u_{hl} + \alpha \sum_{(x_k,x_h)\in \Delta} \sum_{j=1}^{C} u_{kj}u_{hj} \to \min \quad (5)$$

Within condition (2) combining with (5) we have:

$$u_{kj}^{AFCC} = u_{kj}^{FCM} + u_{kj}^{Constrains} \quad (6)$$

$$V_j = \frac{\sum_{k=1}^{N} u_{kj}^2 X_k}{\sum_{k=1}^{N} u_{kj}^2} \quad ; j = \overline{1,C} \quad (7)$$

where:

$$u_{kj}^{FCM} = \frac{1}{\sum_{i=1}^{C}\left(\frac{\| X_k - V_j \|}{\| X_k - V_i \|}\right)^{\frac{2}{m-1}}} \quad ; k = \overline{1,N}, \quad j = \overline{1,C} \quad (8)$$

$$u_{kj}^{constrints} = \frac{\alpha}{2\left\|X_k - V_j\right\|^2} \frac{\sum_{i=1}^{C} \frac{\sum_{(X_k,X_h)\in M} \sum_{l=1, l\ne i}^{C} u_{hl} + \sum_{(X_k,X_h)\in\Delta} u_{hi}}{\left\|X_k - V_j\right\|^2}}{\sum_{i=1}^{C} \frac{1}{\left\|X_k - V_j\right\|^2}} + \frac{\alpha}{2\left\|X_k - V_j\right\|^2}\left(\sum_{(X_k,X_h)\in\Delta} u_{hi} - \sum_{(X_k,X_h)\in M} \sum_{l=1, l\ne i}^{C} u_{hl}\right) \quad (9)$$

$$k = \overline{1,N} \quad j = \overline{1,C}$$

$$\alpha = \frac{N}{M} \frac{\sum_{k=1}^{N}\sum_{j=1}^{C} u_{kj}^2 \left\|X_k - V_j\right\|^2}{\sum_{k=1}^{N}\sum_{j=1}^{C} u_{kj}^2} \quad (10)$$

$$k = \overline{1,N} \quad j = \overline{1,C}$$

M is the sum of must-link and cannot-link constraints.

The Active Semi-Supervised Fuzzy Clustering Algorithm (AFCC) is presented as follows (see Table III):

TABLE II: ACTIVE SEMI-SUPERVISED FUZZY CLUSTERING ALGORITHM.

I:	N data elements $X = \{X_1, X_2, ..., X_N\}$, Number of clusters (C), threshold ε , maximum iteration *maxSteps > 0.*
O:	Matrices u and centers V ;
AFCC:	
1:	t = 0
2:	$V_j^{(t)} \leftarrow random$; ($j=\overline{1,C}$)
3:	Repeat
4:	$t = t + 1$
5:	Calculate α by equation (10)
6:	Calculate $u_{kj}^{(t)}$ ($k=\overline{1,N}$; $j=\overline{1,C}$) by equation (6)
7:	Update $V_j^{(t+1)}$ ($j=\overline{1,C}$) by equation (7)
8:	Until $\left\|u^{(t)} - u^{(t-1)}\right\| \le \varepsilon$ or *maxSteps* has reached

IV. Semi-Supervised Standard Fuzzy Clustering

Class (3): The additional information is the pre-defined membership degree of data.

Yasunori *et al*. [14] presented a semi-supervised fuzzy clustering algorithm by integrating membership degree $\overline{u_{kj}}$ in objective function of FCM to improve the clustering performance of the algorithm.

The objective function in [14] is as follows:

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

Within condition (2), where $\overline{u_{kj}}\in[0,1]$ is a membership degree of X_k in cluster C_j and it is given as below:

$$\overline{U}=\left\{\overline{u_{kj}}\,|\,\overline{u_{kj}}\in[0,1],k=\overline{1,N},j=\overline{1,C}\right\},$$

$$\sum_{j=1}^{C}\overline{u_{kj}}\le 1,\ \left(\forall k=\overline{1,N}\right)$$

Solving the problem within equations (2) and (11) we have:

$$V_j=\frac{\sum_{k=1}^{N}\left|u_{kj}-\bar{u}_{kj}\right|^{m}X_k}{\sum_{k=1}^{N}\left|u_{kj}-\bar{u}_{kj}\right|^{m}}\quad ,j=\overline{1,\mathrm{C}} \tag{12}$$

And u_{kj} with 2 cases:

$m>1$:

$$u_{kj}=\overline{u_{kj}}+\left(1-\sum_{i=1}^{C}\overline{u_{kj}}\right)\frac{\left(\frac{1}{\left\|X_k-V_j\right\|}\right)^{\frac{2}{m-1}}}{\sum_{i=1}^{C}\left(\frac{1}{\left\|X_k-V_i\right\|}\right)^{\frac{2}{m-1}}} \tag{13}$$

$m=1$:

$$u_{kj}=\begin{cases}\overline{u_{kj}}+1-\sum_{j=1}^{C}\overline{u_{kj}}\ ,k=\arg\min_i\left\|X_k-V_i\right\|^2\\ \overline{u_{kj}}\qquad\qquad\qquad ,otherwise.\end{cases}$$

$$k=\overline{1,N},\ \ j=\overline{1,C}. \tag{14}$$

The Semi-Supervised Standard Fuzzy Clustering algorithm (SSSFC) is as follows (see Table III):

TABLE III: SEMI-SUPERVISED STANDARD FUZZY CLUSTERING ALGORITHM

I:	N data elements $X=\{X_1,X_2,...,X_N\}$, Number of clusters (C), membership degree given ($\overline{U}$). threshold ε, maximum iteration *maxSteps > 0.*
O:	Matrices U and centers V ;
SSSFC:	
1:	$t=0$
2:	$V_j^{(t)}\leftarrow random$; ($j=\overline{1,C}$)
3:	Repeat
4:	$t=t+1$
5:	Calculate u_{kj} ($k=\overline{1,N}$; $j=\overline{1,C}$) by equation (13) with $m>1$ or equation (14) with $m=1$.
6:	Calculate $V_j^{(t+1)}$ ($j=\overline{1,C}$) by equation (12)
7:	Until $\left\|V^{(t)}-V^{(t-1)}\right\|\le\varepsilon$ or *maxSteps* has reached

V. SEMI-SUPERVISED ENTROPY REGULARIZED FUZZY CLUSTERING

Class (3): The additional information is the pre-defined membership degree of data.

Yasunori *et al* [14] proposed the Semi-Supervised Entropy Regularized Fuzzy Clustering algorithm (eSFCM) using the given membership degree $\overline{u_{kj}}$ to increase the clustering performance. The objective function of eSFCM is as below:

$$J(U,V)=\sum_{k=1}^{N}\sum_{j=1}^{C}u_{kj}\left\|X_k-V_j\right\|^2$$
$$+\lambda^{-1}\sum_{k=1}^{N}\sum_{j=1}^{C}\left(\left|u_{kj}-\overline{u_{kj}}\right|\log\left|u_{kj}-\overline{u_{kj}}\right|\right)\rightarrow \min \tag{15}$$

Within constraints (2) we have:

$$u_{kj}=\overline{u_{kj}}+\frac{e^{-\lambda\left\|X_k-V_j\right\|^2}}{\sum_{i=1}^{C}e^{-\lambda\left\|X_k-V_i\right\|^2}}\left(1-\sum_{i=1}^{C}\overline{u_{kj}}\right)(15)$$

$$k=\overline{1,N},\ \ j=\overline{1,C}$$

$$V_j=\frac{\sum_{k=1}^{N}u_{kj}X_k}{\sum_{k=1}^{N}u_{kj}}\quad ;j=\overline{1,C} \tag{17}$$

The Semi-Supervised Entropy Regularized Fuzzy Clustering algorithm (eSFCM) is described below (see Table IV).

TABLE IV: SEMI-SUPERVISED ENTROPY REGULARIZED FUZZY CLUSTERING ALGORITHM

I:	N data elements $X=\{X_1,X_2,...,X_N\}$, Number of clusters (C), membership degree given ($\overline{U}$). threshold ε, maximum iteration *maxSteps > 0.*
O:	Matrices U and centers V ;
eSFCM:	
1:	t = 0
2:	$V_j^{(t)}\leftarrow random$; ($j=\overline{1,C}$)
3:	Repeat
4:	$t=t+1$

5:	Calculate u_{kj} ($k=\overline{1,N}$; $j=\overline{1,C}$) by equation (16).
6:	Calculate $V_j^{(t+1)}$ ($j=\overline{1,C}$) by equation (17)
7:	Until $\left\|V^{(t)}-V^{(t-1)}\right\|\le\varepsilon$ or *maxSteps* has reached

VI. Semi-Supervised Kernel Fuzzy C-Mean Clustering

Class (2): The additional information is class label of data.

Zhang *et al*. [12] presented fuzzy clustering algorithm using Kernel methods with the combination of learning labeled and unlabeled data to obtain best clusters. The idea of this algorithm is to use Φ function to map data elements in the original space into a feature space with high dimension. Suppose $\Phi(X_k)$, $\Phi(V_j)$ are images of elements X_k, X_j through Φ function then the objective function of FCM algorithm is modified as follows:

$$J(U,V)=\sum_{k=1}^{N}\sum_{j=1}^{C}u_{kj}^{m}\,\|\Phi(X_k)-\Phi(V_j)\|^2\rightarrow\min \quad (18)$$

The distance between $\Phi(X_k)$ and $\Phi(V_j)$ is calculated as below:

$$\left\|\Phi(X_k)-\Phi(V_j)\right\|^2=K(X_k,X_k)-2K(X_k,V_j)+K(V_j,V_j) \quad (19)$$

The kernel function $K(X_k,V_j)$ is defined as follows:

$$K(X_k,V_j)=e^{-\left(X_k-V_j\right)^2/2\sigma^2} \quad (20)$$

Zhang [12] used both labeled and unlabeled data on fuzzy clustering algorithm. Suppose labeled and unlabeled data is L and $\tilde{L}$; l and $\tilde{l}$ is labeled datum of L and unlabeled datum of $\tilde{L}$. For each $Y_k\in L$ satisfying the condition:

$$u_{kj}\in[0,1],\quad \sum_{j=1}^{C}u_{kj}=1\quad ,1\le k\le l \quad (21)$$

Integrating sub-information into the objective function (19) we have the objective function for Semi-Supervised Kernel Fuzzy c-mean Clustering:

$$J(L,\tilde{L})=w_l\sum_{k=1}^{l}\sum_{j=1}^{C}\left(u_{kj}^{m}-u_{kj,0}^{m}\right)\left(k\left(Y_k,V_{j,0}\right)-k\left(Y_k,V_j\right)\right) + w_{\tilde{l}}\sum_{k=1}^{\tilde{l}}\sum_{j=1}^{C}u_{kj}^{m}\left(2-2k\left(X_k,V_j\right)\right)\rightarrow\min \quad (22)$$

where: $w_l=\dfrac{l}{l+\tilde{l}}$ and $w_{\tilde{l}}=\dfrac{\tilde{l}}{l+\tilde{l}}$

Solving equations (21), (22) we have the membership degree equation for labeled data as follows:

$$u_{kj}=\left(\sum_{i=1}^{C}\left(\frac{k(Y_k,Y_k)-2(Y_k,V_j)+k(V_j,V_j)}{k(Y_k,Y_k)-2(Y_k,V_i)+k(V_i,V_i)}\right)^{\frac{1}{m-1}}\right)^{-1} \quad (23)$$

And for unlabeled data:

$$u_{kj}=\frac{u_{kj}k(Y_k,V_j)}{\sum_{i=1}^{C}u_{kj}k(Y_k,V_i)} \quad (24)$$

The σ parameter is calculated as follows:

$$\sigma=\sigma-\eta_\sigma\frac{\partial J(L,L)}{\partial\sigma} \quad (25)$$

$$J(L,\tilde{L})=Tanh^{-1}\left[\lambda_h\left\|u_{kj}-u_{kj,h}^{*}\right\|\right]\times Tanh^{-1}\left[\lambda_h\left\|\sigma-\sigma_k^{*}\right\|\right]\times J(L,\tilde{L}) \quad (26)$$

The Semi-Supervised Kernel Fuzzy Clustering algorithm (SSKFCM) is as follows (see Table V):

TABLE V: Semi-Supervised Kernel Fuzzy Clustering Algorithm

I:	N data elements $X=\{X_1,X_2,...,X_N\}$, Number of clusters (C),labeled dataset (L) and unlabeled dataset ($\tilde{L}$), threshold ε , maximum iteration *maxSteps > 0* and number of changing object function (p).
O:	Matrices u and centers V ;
SSKFCM:	
1:	$t=0$
2:	Repeat
3:	$u_{kj}^{(t)}\leftarrow random$ for labeled data. For unlabeled data, $(r_{1j},r_{2j},...,r_{Cj})\leftarrow random$ in $[0,1]$ and $u_{kj}^{(t)}=\dfrac{r_{kj}}{\sum_{i=1}^{C}r_{ki}}$; $\sigma^{(t)}\leftarrow random$ ($k=\overline{1,N}$, $j=\overline{1,C}$)
4:	$t=t+1$
5:	Calculate $u_{kj}^{(t)}$ ($k=\overline{1,N}$; $j=\overline{1,C}$) by equation (23) for labeled data and equation (24) for unlabeled data.
6:	Calculate $\sigma^{(t)}$ by equation (25)
7:	Calculate e ($k=\overline{1,N}$; $j=\overline{1,C}$) as follows: $e=\max\left\{\max_{kj}\left\|u_{kj}^{(t+1)}-u_{kj}^{(t)}\right\|,\left\|\sigma^{(t+1)}-\sigma^{(t)}\right\|\right\}$
8:	Until $t=p$
9:	Choose the best objective function which has minimum e .

VII. Semi-Supervised Kernel Fuzzy Clustering Algorithm

Class (1): The additional information is the must-link and cannot-link constraints.

Maraziotis [17] proposed an algorithm for Semi-supervised fuzzy clustering through must-link pair-constraints. The objective function is as follows:

$$J(U,V)=\sum_{k=1}^{N}\sum_{j=1}^{C}\left(nu_{kj}-(1-n)\beta_{\varphi(k,j)}\right)^{m}\left\|X_k-V_j\right\|^2 \quad (27)$$

Within constraint (2), where β_{kj} is defined:

$$\beta_{\varphi(k,j)}=\begin{cases}\beta_{kj} & ;khi:\arg(\max_i u_{ki})=j\\ 0 & ;nguoclai\end{cases} \quad (28)$$

where $t_{\max}$ is the maximum iteration. n_0, n_F is the initial and final of parameter n respectively. t is a current iteration. Solving equations (2) and (26) we have:

$$V_j=\frac{\sum_{k=1}^{N}\left(nu_{kj}-(1-n)\beta_{\varphi(k,j)}\right)^{m}X_k}{\sum_{k=1}^{N}\left(nu_{kj}-(1-n)\beta_{\varphi(k,j)}\right)^{m}} \quad (29)$$

$$V_j=\frac{\sum_{k=1}^{N}\left(nu_{kj}-(1-n)\beta_{\varphi(k,j)}\right)^{m}X_k}{\sum_{k=1}^{N}\left(nu_{kj}-(1-n)\beta_{\varphi(k,j)}\right)^{m}} \quad (30)$$

$$u_{kj}=u_{kj}^{FCM}+u_{kj}^{constr} \quad (31)$$

where:

$$u_{kj}^{FCM}=\frac{1}{\sum_{i=1}^{C}\left(\frac{\left\|X_k-V_j\right\|}{\left\|X_k-V_i\right\|}\right)^{\frac{2}{m-1}}} \quad (32)$$

$$u_{kj}^{constr}=\frac{1-n}{2n}\left(\frac{1}{\left\|X_k-V_j\right\|^2}\right)^{\frac{1}{m-1}}\left(\frac{\beta_{kj}}{\left(\frac{1}{\left\|X_k-V_j\right\|^2}\right)^{\frac{1}{m-1}}}-\frac{\sum_{i=1}^{C}\beta_{ki}}{\sum_{i=1}^{C}\left(\frac{1}{\left\|X_k-V_i\right\|^2}\right)^{\frac{1}{m-1}}}\right) \quad (33)$$

The Semi-Supervised Kernel Fuzzy Clustering algorithm (SSKFCM) is as follows (see Table VI):

TABLE VI: SEMI-SUPERVISED KERNEL FUZZY CLUSTERING ALGORITHM

I:	N data elements $X=\{X_1,X_2,...,X_N\}$, Number of clusters (C), threshold ε, maximum iteration *maxSteps > 0.*
O:	Matrices u and centers V ;
SSKFCM:	
1:	$t = 0$
	$u_{kj}^{(t)}\leftarrow random$; ($k=\overline{1,N}$, $j=\overline{1,C}$) satisfy equation (2)
2:	Repeat
3:	$t = t + 1$
4:	Calculate $V_j^{(t+1)}$ ($j=\overline{1,C}$) by equation (31)
5:	Calculate u_{kj} ($k=\overline{1,N}$; $j=\overline{1,C}$) by equation (32).
6:	Until $\left\|u_{kj}^{(t)}-u_{kj}^{(t-1)}\right\|\le\varepsilon$ or *maxSteps* has reached

VIII. CONCLUSION

In this paper, we have taken an overview of some typical semi-supervised fuzzy clustering algorithms that could be able to deal with additional information. Within each type of additional information, there have been some effective algorithms that could find good solutions to partition data elements into clusters. If the additional information is must-link and cannot-link constraints, Semi-supervised Kernel fuzzy clustering algorithm is seem to run faster than Active Semi-Supervised Fuzzy Clustering because it has not to calculate α parameter. However, Active Semi-Supervised Fuzzy Clustering algorithm is promised to be better in clustering quality because it has more targets on objective function. If the additional information is class labels of a part of data, Semi-Supervised Kernel Fuzzy c-mean Clustering algorithm is maybe the one to solve this problem. Finally, if the additional information is pre-defined membership degrees, *Semi-Supervised Standard Fuzzy Clustering algorithm* is more appropriate than *Semi-Supervised Entropy Regularized Fuzzy Clustering* algorithm. *Semi-Supervised Entropy Regularized Fuzzy Clustering* has more complicated objective function that causes high computational time.

Beside the advantages of these algorithms, they still remain some limitations related to fuzzy logic such as the uncertainty and vagueness that prevent reaching to the best clustering quality.

In the future, we will develop the semi-supervised fuzzy clustering algorithms on some advanced fuzzy sets to overcome the mentioned limitations. Moreover, some applications based on them are also our targets.

ACKNOWLEDGMENT

This work is sponsored by a VNU Project under contract No. QG.13.01.

REFERENCES

[1] M. R. Rezaee, P. M. V. D. Zwet, B. P. Lelieveldt, R. J. V. D. Geest, and J. H. Reiber, "A multiresolution image segmentation technique based on pyramidal segmentation and fuzzy clustering," *IEEE Transactions on Image Processing*, vol. 9, no. 7, pp. 1238-1248, 2000.

[2] K. S. Chuang, H. L. Tzeng, S. Chen, J. Wu, and T. J. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *Computerized Medical Imaging and Graphics*, vol. 30, no. 1, pp. 9-15, 2006.

[3] Y. W. Lim and S. U. Lee, "On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques," *Pattern Recognition*, vol. 23, no. 9, pp. 935-952, 1990.

[4] J. Lu, X. Yuan, and T. Yahagi, "A method of face recognition based on fuzzy c-means clustering and associated sub-NNs," *IEEE Transactions on Neural Networks*, vol. 18, no. 1, pp. 150-160, 2007.

[5] M. Agarwal, H. Agrawal, N. Jain, and M. Kumar, "Face recognition using principle component analysis, eigenface and neural network," in *Proc*. the *International Conference on Signal Acquisition and Processing*, 2010, pp. 310-314.

[6] J. Chen, S. Zhao, and H. Wang, "Risk analysis of flood disaster based on fuzzy clustering method," *Energy Procedia*, vol. 5, pp. 1915-1919, 2011.

[7] L. H. Son, P. L. Lanzi, B. C. Cuong, and H. A. Hung, "Data mining in GIS: A novel context-based fuzzy geographically weighted clustering algorithm," *International Journal of Machine Learning and Computing* , vol. 2, no. 3, pp. 235-238, 2012.

[8] G. Wang, J. Hao, J. Ma, and L. Huang, "A new approach to intrusion detection using artificial neural networks and fuzzy clustering," *Expert Systems with Applications*, vol. 37, no. 9, pp. 6225-6232, 2010.

[9] W. Chimphlee, A. H. Abdullah, N. M. Sap, S. Srinoy, and S. Chimphlee, "Anomaly-based intrusion detection using fuzzy rough clustering," in *Proc. the International Conference on Hybrid Information Technology*, 2006, vol. 1, pp. 329-334.

[10] A. Martin, V. Gayathri, G. Saranya, P. Gayathri, and P. Venkatesan, "A hybrid model for bankruptcy prediction using genetic algorithm, fuzzy C-means and mars," *International Journal on Soft Computing*, vol.2, no.1, February 2011.

[11] N. Grira, M. Crucianu, and N. Boujemaa, "Active semi-supervised fuzzy clustering," *Pattern Recognition*, vol. 41, no. 5, pp. 1834-1844, 2008.

[12] H. Zhang and J. Lu, "Semi-supervised fuzzy clustering: A kernel-based approach," *Knowledge-Based Systems*, vol. 22, no. 6, pp. 477-481, 2009.

[13] K. Li, Z. Cao, L. Cao, and R. Zhao, "A novel semi-supervised fuzzy c-means clustering method," in *Proc. the Control and Decision Conference*, 2009, pp. 3761-3765.

[14] E. Yasunori, H. Yukihiro, Y. Makito, and M. Sadaaki, "On semi-supervised fuzzy c-means clustering," in *Proc. the IEEE International Conference on Fuzzy Systems*, 2009, pp. 1119-1124.

[15] X. Yin, T. Shu, and Q. Huang, "Semi-supervised fuzzy clustering with metric learning and entropy regularization," *Knowledge-Based Systems*, vol. 35, pp. 304-311, 2012.

[16] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Kluwer Academic Publishers, 1981.

[17] I. A. Maraziotis, "A semi-supervised fuzzy clustering algorithm applied to gene expression data," *Pattern Recognition*, vol. 45, no. 1, pp. 637-648, 2012.

Pham Huy Thong is a researcher and PhD student at VNU University of Science, Vietnam National University. His research interests include geographic Information systems and molecular dynamics simulation.

Le Hoang Son obtained the PhD degree in mathematics from informatics at VNU University of Science, Vietnam National University. Currently, he is a researcher at the Center for High Performance Computing, VNU University of Science, Vietnam National University. His major field includes soft computing, fuzzy clustering, recommender systems, geographic information systems and particle swarm optimization. He is a member of IACSIT and also an associate editor of the International Journal of Engineering and Technology (IJET). He also served as a reviewer for various international journals and conferences such as PACIS 2010, ICMET 2011, ICCTD 2011, KSE 2013, BAFI 2014, NICS 2014, ACIIDS 2015, International Journal of Computer and Electrical Engineering, Imaging Science Journal, International Journal of Intelligent Systems Technologies and Applications, IEEE Transactions on Fuzzy Systems, Expert Systems with Applications, and International Journal of Electrical Power and Energy Systems. Dr. Son has got many publications in prestigious journals and undertaken some major research projects of Vietnam and international joint projects.