International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol. 2, No. 2, 2015, pp. 9-22.
ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

Evaluation of the intelligent hybrid methods to improve the performance of K-Means and C-Means algorithms

Behzad Radmehr[a] , Reza Ghaemi[b]

[a] *Department of Computer Engineering,Khorasan Razavi Science & Research Branch,Islamic Azad University,Neyshabur,Iran.*

[b] *Department of Computer Engineering,Quchan Branch,Islamic Azad University,Quchan,Iran*.

Abstract

In this paper , The intelligent hybrid methods are used for improving the performance of K-means and C-means algorithms. . To achieve this, these methods are explained in order to improve the performance of these two data mining algorithms. Some suggestions are provided for this aim. The methods used for explaining in relation to C-means algorithms are fuzzy C-means algorithm, combination of fuzzy markov model with evolution algorithms, combination of neural network with C-means algorighm.

Keywords: Data Mining, Cluster Analysis, Hybrid Intelligant Methods, K-Means Algorithm, C-Means Algorithm, Neural Networks, Fuzzy Algorithms, Evolution Algorithms

Introduction:

This paper is written for machine learing and datamining scopes. The writes want to analyze the hybrid intelligent methods that combine with C-means algorithm and suggest the researchers that combine this methods with K-means algorithms and explain this method what behaves with K-means algorithm.
First of all, we must explain that K-means algorithm, step by step. K-Means algorithms have three step.

- First step is define K points randomly for center of the clusters.
- In the second step every records is assigned as the element of the cluster with exact policy.
- In the third step, we elect the head cluster, again with some policy in this paper we want to improve this methods and suggest some methods to improve the performance of K-means algorithms. In the abstract of this paper, we said that there are some hybrid intelligent methods that improves the performance of
- C-Means algorithm , we want to explain this methods, this methods is explained in this paper is included:
    - The combaniation of evelutions algorithms with different fuzzy methods that combines with fuzzy hidden Markow model.
    - Talking about fuzzy type 2 combines with clustering methods in order to improve clustering accuracy , classification accuracy , pattern recognition accuracy.
    - Using from Performance evaluation of FMIG clustering with fuzzy validity indexes.
    - Using from data driven neural network combines with fuzzy methods that improves the clustering mechanism.

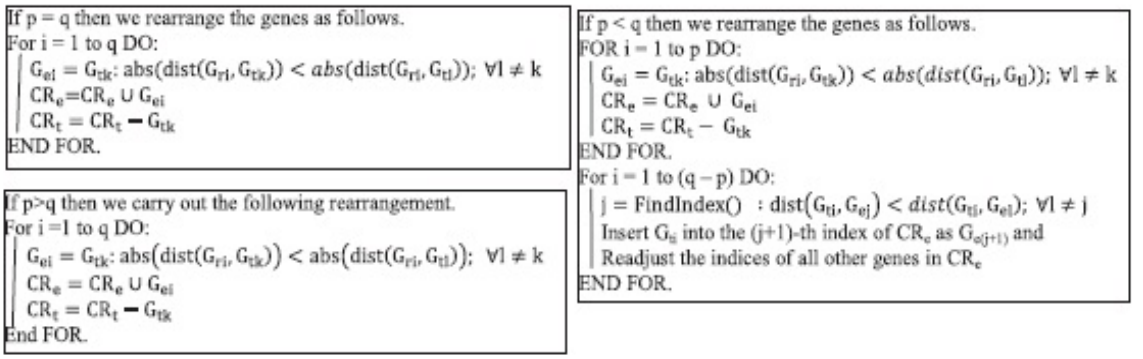


Figure 1(Gene rearrangement in this method)-[1]

- Using from Wang-Mendel method to generate fuzzy rules with classification methods that based on different fuzzy methods.

Using from fuzzy threashold that combines with genetic based algorithms with C-Means algorithms that improve the performance of cluster ensembles.

Explaining about combining Genetic algorithm with K-Means algorithm

In this part we want to combine genetic algorithm with K-Means algorithm in order to improve the Performance of the K-Means algorithm[1].

There are some important topics in this paper[1] that we want to review in this paper, One of them is gene rearrangement that said in this paper.The researchers want to improve the performance of this algorithm we think that , this item is one the best opportinuty for them, now we declare the method of the writters of this paper present it Figure 1.

Another important item in the genetic algorithm is combined with K-Means algorithm, is important is fitness function and membership degree of evey gene in this algorithm. So we want to review this important element of this paper[1]. Equation (1)is explained this item.

$$\text{Fitness} = \sum_{\forall j} (\text{Sep}_j - \text{Comp}_j)$$
$$\text{Comp}_j = \frac{\sum_{R_a \in D} \mu_{aj} \text{dist}(R_a, S_j)}{\sum_{R_a \in D} \mu_{aj}} \qquad (1)$$

When we talk about genetic algorithm, another element is very important that we want to review it is mutation operator that we want to explain in the Equation **Error! Reference source not found.**[1]

$$M_i = \begin{cases} K_1 * \frac{f_{\max} - f_i}{f_{\max} - \bar{f}}, & f_i > \bar{f} \\ K_1, & f_i \leqslant \bar{f} \end{cases} \qquad (2)$$

Explaining about data driven neural network combines with fuzzy methods that improves the clustering mechanism

In this method we want to review the way that improves the clustering mechanism.This neural network is explained in this paper has 5 layers [2] that we want to review this intelligent method step by steps. First layer that is used in explained in Equation **Error! Reference source not found.**[2]

$$f = u_i^{(1)} \text{ and } a^{(1)} = f \tag{3}$$

The second step of this neural network that review has this details that showed in the Equation **Error! Reference source not found.**[2].

$$f[u_{ij}^{(2)}] = -\frac{[u_{ij}^{(2)} - \mu_{ij}]^2}{\sigma_{ij}^2} \text{ and } a^{(2)} = e^f \tag{4}$$

The third layer of this neural network is showed in the Equation **Error! Reference source not found.** [2]. We want to explain this mechanism that improve the clustering methods.

$$f[u_i^{(3)}] = \prod_{i=1}^{n} u_i^{(3)} \text{ and } a^{(3)} = f \tag{5}$$

The fourth layer of this network is explained in Equation **Error! Reference source not found.**[2].

$$f[u_i^{(4)}] = \sum_{i=1}^{r} u_i^{(4)} \text{ and } a^{(4)}(f) = \frac{u_i^{(4)}}{f} \tag{6}$$

The last layer of this neural network is explained in Equation **Error! Reference source not found.**[2]. We want to focus on this method be cause this method improves clustering mechanism.

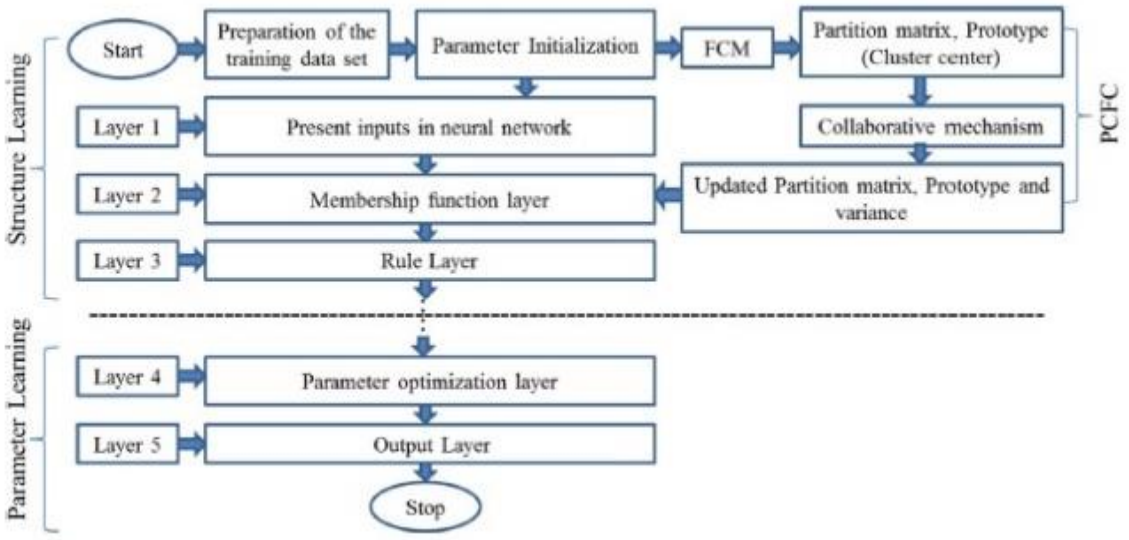


Figure 2(Frame Work of this Method)-[2]

$$f[u_i^{(5)}] = \sum_i w_i u_i^{(5)}, \; a^{(5)}(f) = f \tag{7}$$

Another important topic is said in this paper is Collaborative fuzzy clustering that we want to focus on it [2].
Equation **Error! Reference source not found.** is showed this Collaborative fuzzy clustering method[2]**.**

$$Q[l] = \sum_{l=1}^{N}\sum_{j=1}^{c} u_{ij}^{2}[l]d_{ij}^{2}[l] + \sum_{\substack{m=1 \\ m\neq l}}^{p} \beta[l,m] \sum_{l=1}^{N}\sum_{j=1}^{n} (u_{ij}[l]-u_{ij}[m])^{2} d_{ij}^{2}[l] \qquad \textbf{(8)}$$

By the way another scheme is helped in this paper that we want to review in our paper is showed in Figure 3[2]Figure 3**.**

At the end of this part of our paper we want to review another framework of this method to explain better in **Figure 2**[2]**.**

Explaining about fuzzy type 2 combines with clustering methods in order to improve clustering accuracy, classification accuracy, pattern recognition accuracy

In this part of this paper we want to review the importance of fuzzy type 2 and when combining with clustering methods to improve the accuracy of clustering , classification and pattern recognition.
First of all we want to explain about fuzzy type 2 in compareness with fuzzy type 1 and how they work in diffrenet situations.
In Equation **Error! Reference source not found.**[3] we want to formulize fuzzy type 2 , in our senario**.**

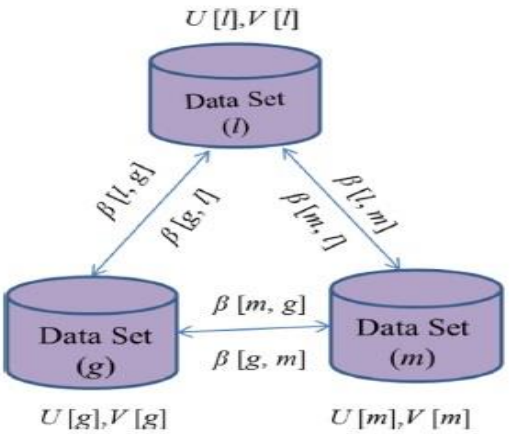


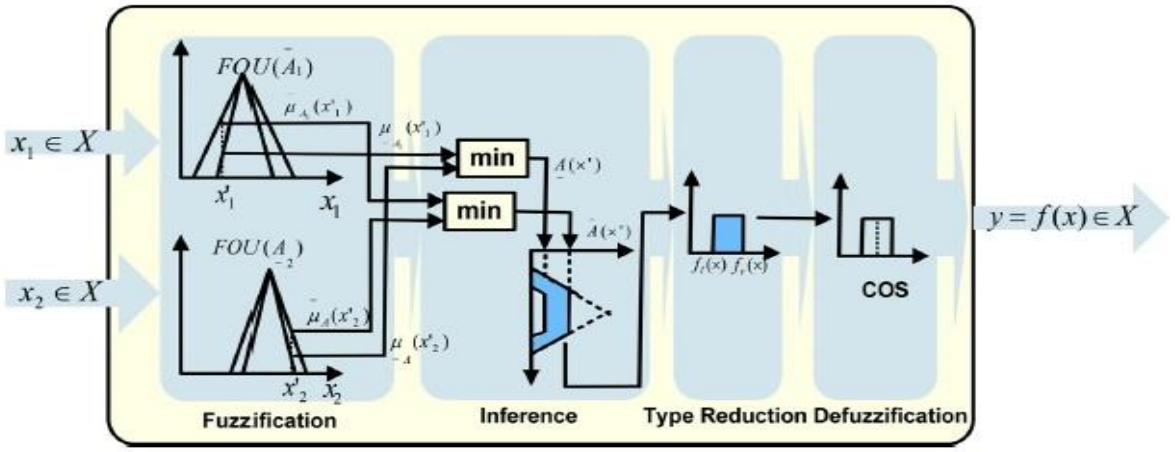
Figure 3(Collaborative Clustering Scheme)-[2]

$$\tilde{A} = \{((x,u), \mu_{\tilde{A}}(x,u)) | \forall x \in X, \forall u \in J_x \subseteq [0,1]\} \qquad (9)$$

In the Equation **Error! Reference source not found.**[3] we explain fuzzy type 2 in compare fuzzy type 1 and show how this methodwork.
In that paper[3] two table is included that we think is very help full for researchers works in this scopes if is necessary for them they can review this refrence[3].
They explain fuzzy type 2 how they work ,and how the fuzzy type 2 can be compared with fuzzy type 2 in clustering, classification and pattern recognition items.

Explaing about Improving Wang–Mendel method performance in fuzzy rules generation using the fuzzyC-means clustering algorithm

In this part of this paper we want to explain Prediction model based on FCAWM method [4] that named in this paper [4] .
with this name. Create a fuzzy system that help us to improve our method. The center-averaged defuzzifier is showed in Equation **Error! Reference source not found.** [4].and expalin about fuzzy system more in[4] howthis



fuzzy system work.

Figure 4(Structure of type 2 fuzzy system)-[3]

$$y = \frac{\sum_{l=1}^{M} y_c^l \cdot \omega^{(l)}}{\sum_{l=1}^{M} \omega^{(l)}}, \quad (10)$$

$$act(l) = \prod_{j=1}^{n} \mu_{A_{ij}^{(l)}}(x_{ij}), \quad (11)$$

$$f(x) = \frac{\sum_{l=1}^{M} y_c^l \cdot act(l)}{\sum_{l=1}^{M} act(l)},$$

About this paper we want to explain more about Flowchart of the FCAWM[4] that named in this paper with this name[4]. This flowchart explained about how this model work. Figure 5 is showed this model [4].

Explaing about fuzzy threashold that combines with genetic based algorithms with C-Means algorithms that improve the performance of cluster ensembles

In this part of the paper we want to talk about the fuzzy threashold how it help us to solve genetic based problem and how it help to cluster better in C-Means algorithm and improve the cluster ensembling [5].

This method is showed in[5] that explaind step by step in this figure.
In this paper u must focus on the fuzzy threashod how it solves the problem and how this algorithm solved the problem [5] help us to undesrstand this algorithm better.
Another flow chart is very important to understand this method better is showed [5].

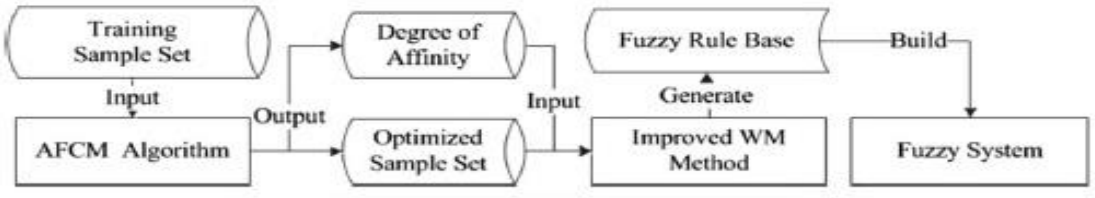


Figure 5(Flow chart of the FCAWM Model)-[4]

This flow chart is about the threshold fuzzy C-means clustering algorithm how it help to solve the problem [5].

Explaining about combaniation of evelutions algorithms with different fuzzy methods that combines with fuzzy hidden Markov model

In this part of this paper we want to review the combine of evolution algorithms with different fuzzy method with fuzzy hidden Markov model that improve the time series predictions.
Because this method is find out very good this flow chart is designed in this paper[6] is very useful ,So I review this flowchart in [6].
The flowchart of the hybrid model for generating fuzzy rules is showed in[6].Another Item is very bold in this paper is Fuzzy rule generation is showed in Equation **Error! Reference source not found.** [6].

$$M_{X_i} = e^{-(1/2)((x_i-\mu_{x_i})/\sigma_{x_i})^2} \qquad (12)$$

Explaining about design of hybrid radial basis function neural networks (HRBFNNs) realized with the aid of hybridization of fuzzy clustering method (FCM) and polynomial neural networks (PNNs)

In this part of the paper we want to explain about the Hybrid radial basis function neural networks and the idea is said here.

This formula is showed in Equation **Error! Reference source not found.** [7].This formula is explained how we should work in this scope.

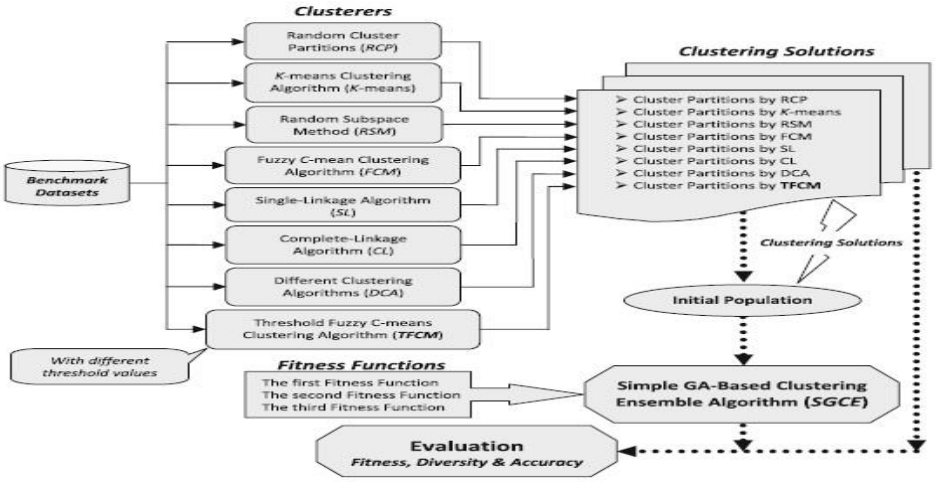


Figure 6(The methodology of the Method)-[5]

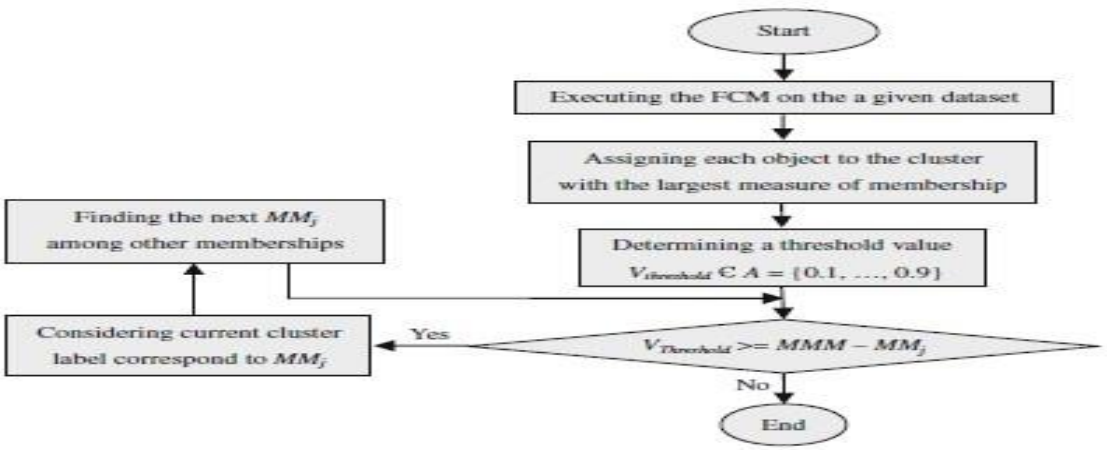


Figure 7(Flow chart of fuzzy threshold clustering algorithm)-[5]

$$y = f(x_1, x_2, \ldots, x_l) = a_0 + \sum_{i=1}^{l} a_i x_i + \sum_{i=1}^{l} \sum_{j=1}^{l} a_{ij} x_i x_j \qquad (13)$$

Equation **Error! Reference source not found.** is showed another pattern we must do it in this scenario Another item is important in HRBFNNs architecture is showed in

. That explain how we should behave with this network.
If we want to help the genetic algorithm to combine to this method to improve fuzzy clustering method. This explanation of the fitness function is showed in Equation **Error! Reference source not found.**[6].

$$\hat{y} = \hat{f}(x_1, x_2, \ldots, x_I) = a_0 + \sum_{i=1}^{I} a_i x_i + \sum_{i=1}^{I}\sum_{j=1}^{I} a_{ij} x_i x_j + \sum_{i=1}^{I}\sum_{j=1}^{I}\sum_{k=1}^{I} a_{ijk} x_i x_j x_k + \cdots. \quad (14)$$

$$f(\mathrm{PI}, \mathrm{VPI}) = \theta \times \mathrm{PI} + (1-\theta) * \mathrm{VPI} \quad (15)$$

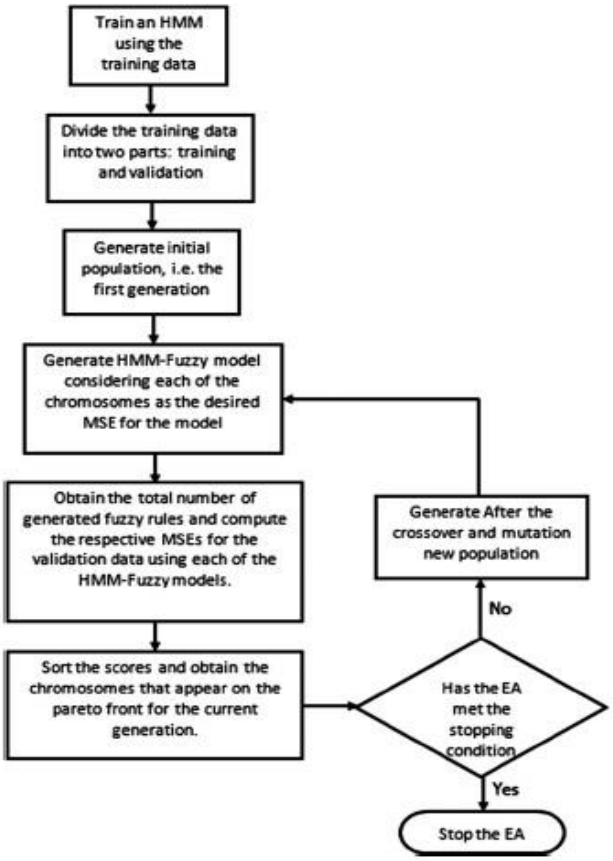
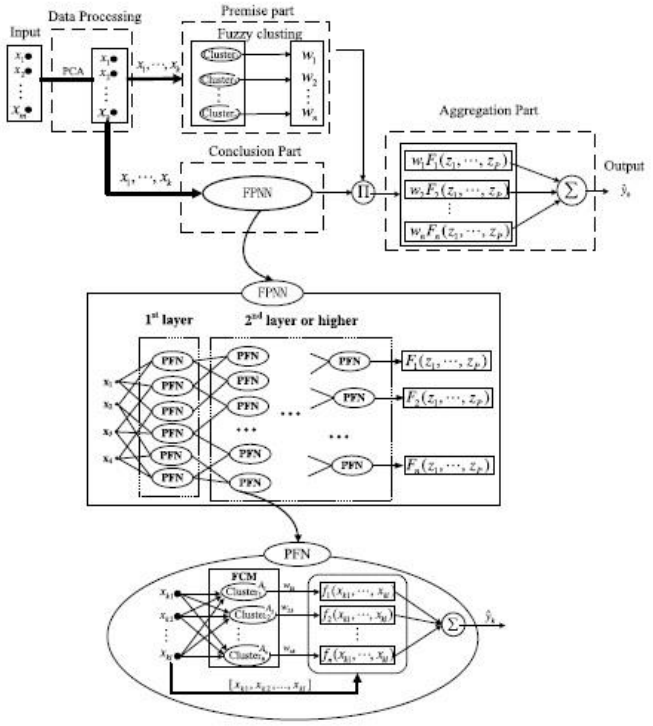


Figure 8(The flow chart of hybrid model for generating fuzzy rules)-[6]

Where θ is a weighting factor for this situation and behave for training, validation data, and testing data. Equation **Error! Reference source not found.**[7] is explaind this method and the situation.

$$\mathrm{PI(VPI\ or\ EPI)} = \begin{cases} \sqrt{\dfrac{1}{m}\sum_{i=1}^{m}(y_i - y_i^*)^2}, & \text{(RMSE)} \\ \dfrac{1}{m}\sum_{i=1}^{m}(y_i - y_i^*)^2. & \text{(MSE)} \end{cases} \tag{16}$$

In this scenario we want to review all item in 1 flow chart that designed in this paper [7] very good. We hope that is useful for researchers that work in this scopes. This flow chart is explained in the



[7].

Explaining about evaluation of Performance using fuzzy FMIG clustering validity indexes

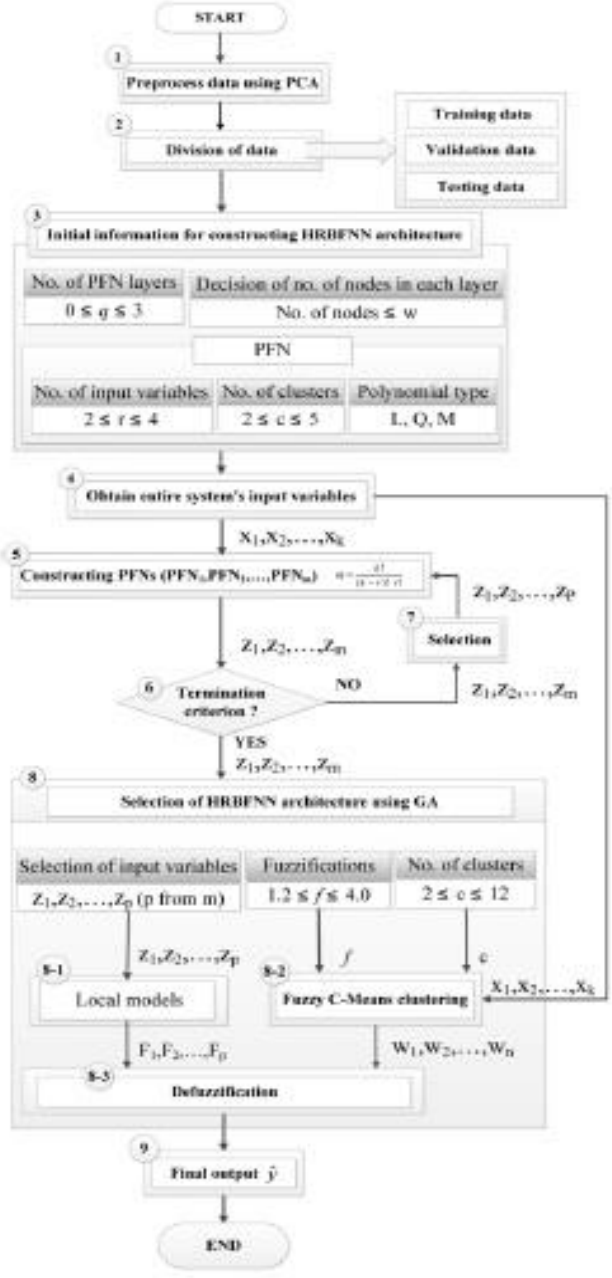
In this part of paper we want to talk about fuzzy multilevel interior GSOMs that called FMIG [8] and how work the algorithm. This method improve clustering mechanism by this steps First step is done by producing the best distribution of data by finding out the architecture correspond to the form of inputs.

Figure 9(An overal flow chart of Design of HRBFNNs)-[7]

The second step is in this method gives the minimum value of error quantization providing the best cluster structure [8].
In this two steps in this algorithms [8] is done correctly, the performance of clustering is getting better.

Figure 10(An overal flow chart of Design of HRBFNNs)-[7]

Recommanded part for the researcher want to work in this scope



In this part of the paper , we want to talk about how we improve the performance of this algorithms. In the introduction part of this paper is explained there steps of the K-means algorithms how we want to improve the performance of K-means algorithms [9] by combing this algorithms with fuzzy–evalition based algorithms [10].
The tools that optimised is defined . one of them is fuzzy type 2 methods [11], Genetic based algorithm[12, 13] and neural network. [14].
But the methods that the we explain in this paper for researchers wok in this scopes, is recommanded is fuzzy data driven neural networks that combines with K-means algorithm to improve the performance of the K-means accuracy algorithm.If the researchers wok on the neural science and neural network , we recommand the deep learning to improve the performance of K-Means algorithm.

Conclusion

In this paper we analyze some hybrid intelligent that improve C-means algorithms and explain th senario of this methods and suggest how thi method if combine with K-means algorithms what it done.
In this paper, The writers recommand some methods for researchers that want to work in this scope. Working on the combine of fuzzy data-driven network with K-means algorithms is recommanded more than the other methods that we explained in this paper.

References:

A. P. Rotshtein, and H. B. Rakytyanska, Fuzzy evidence in identification, forecasting and diagnosis: Springer, 2012.

F. Laurene, "Fundamentals of neural networks," Architecture, Algorithms, and Applications. Prentice Halls, 1994

J. Gou, F. Hou, W. Chen, C. Wang, and W. Luo, "Improving Wang–Mendel method performance in fuzzy rules generation using the fuzzy C-means clustering algorithm," Neurocomputing, vol. 151, pp. 1293-1304, 2015.

J. Gou, F. Hou, W. Chen, C. Wang, and W. Luo, "Improving Wang–Mendel method performance in fuzzy rules generation using the fuzzy C-means clustering algorithm," Neurocomputing, vol. 151, pp. 1293-1304, 2015.

J. Mendel, H. Hagras, W.-W. Tan, W. W. Melek, and H. Ying, Introduction to Type-2 Fuzzy Logic Control: Theory and Applications: John Wiley & Sons, 2014.

M. A. Rahman, and M. Z. Islam, "A hybrid clustering technique combining a novel genetic algorithm with K-Means," Knowledge-Based Systems, vol. 71, pp. 345-365, 2014.

M. A. Rahman, and M. Z. Islam, "A hybrid clustering technique combining a novel genetic algorithm with K-Means," Knowledge-Based Systems, vol. 71, pp. 345-365, 2014.

M. E. Celebi, Partitional Clustering Algorithms: Springer, 2015.

M. Prasad, Y. Lin, C. Lin, M. Er, and O. Prasad, "A New Data-Driven Neural Fuzzy System with Collaborative Fuzzy Clustering Mechanism," Neurocomputing, 2015.

M. Prasad, Y. Lin, C. Lin, M. Er, and O. Prasad, "A New Data-Driven Neural Fuzzy System with Collaborative Fuzzy Clustering Mechanism," Neurocomputing, 2015.

M. R. Hassan, B. Nath, M. Kirley, and J. Kamruzzaman, "A hybrid of multiobjective Evolutionary Algorithm and HMM-Fuzzy model for time series prediction," Neurocomputing, vol. 81, pp. 1-11, 2012.

M. Sugiyama, Statistical Reinforcement Learning: Modern Machine Learning Approaches: CRC Press, 2015.

M. Tlili, T. Ayadi, T. M. Hamdani, and A. M. Alimi, "Performance evaluation of FMIG clustering using fuzzy validity indexes," Soft Computing, pp. 1-14, 2014.

O. Castillo, and P. Melin, Fuzzy Logic Augmentation of Nature-Inspired Optimization Metaheuristics: Theory and Applications: Springer, 2014.

P. Melin, and O. Castillo, "A review on type-2 fuzzy logic applications in clustering, classification and pattern recognition," Applied soft computing, vol. 21, pp. 568-577, 2014.

P. Melin, and O. Castillo, "A review on type-2 fuzzy logic applications in clustering, classification and pattern recognition," Applied soft computing, vol. 21, pp. 568-577, 2014.

R. Ghaemi, M. N. Sulaiman, H. Ibrahim, and N. Mustapha, "A novel fuzzy C-means algorithm to generate diverse and desirable cluster solutions used by genetic-based clustering ensemble algorithms," Memetic Computing, vol. 4, no. 1, pp. 49-71, 2012.

W. Huang, S.-K. Oh, and W. Pedrycz, "Design of hybrid radial basis function neural networks (HRBFNNs) realized with the aid of hybridization of fuzzy clustering method (FCM) and polynomial neural networks (PNNs)," Neural Networks, vol. 60, pp. 166-181, 2014.