Journal of Computer Sciences

Original Research Paper

A Hybrid Clustering Process using a Genetic Fuzzy System for the Knowledge Base of a Fuzzy Rule-Based System

Hamedoun Lamiae, Attariuas Hicham and Ben Maati Mohamed Larbi

Laboratory LIROSA, Department of Computer Science, Faculty of Sciences, Abdelmalek Essaadi University, Mhannech II, BP: 2121, Tetouan, Morocco

Article history
Received: 04-04-2016
Revised: 29-11-2016
Accepted: 15-12-2016

Corresponding Author:
Hamedoun Lamiae
Laboratory LIROSA,
Department of Computer Science, Faculty of Sciences,
Abdelmalek Essaadi University, Mhannech II, BP: 2121, Tetouan, Morocco
Tel: +212 662 12 32 33
Email: lamiae.hamd@gmail.com

Abstract: The present paper proposes a new Hybrid clustering Process based on Fuzzy Genetic System. The proposed Approach consists of two steps: (1) Using a method called Fuzzy clustering, all data elements will be clustered into N groups; (2) utilizing a Fuzzy Genetic System, for every level the fuzzy rule of adhesion will be generated. If we compare our research to others that use the hard clustering, we will conclude that by using the fuzzy clustering we are able to raise the ingredient of each cluster and upgrade the accuracy of the offer target system and we will win in terms of complexity because the system is based on hybrid intelligent method and then we will not need to generate a new cluster every time we add a new data point. Experimental results on estimation models using clustering methods on synthetic data show that the proposed algorithm outperforms few commonly used clustering algorithms.

Keywords: Fuzzy Clustering, Genetic Fuzzy System, Back Propagation Network, Hybrid Intelligence Approach

Introduction

On our days, we can perceive that there is a greater movement for researchers to utilize clustering methods so as to rise the accuracy of their results. Thus, we can classify the clustering as the most important unattended apprenticeship and that is why every problem from this kind should be treat by located a system in a series of unlabeled input. They are enough famouns owing to their speed that is the higher. The results we obtain are spherical and the sensitive are very highly to initialization.

A hybrid intelligent clustering system was suggested (Oh and Han, 2001) it was based in ANN and change point detection. By changing the discovery item the staple construct of offer template is obtained. So we conclude that the proposed model is more exact that the traditional one.

Lately, some researchers have exposed that the use of the hybridization between fuzzy logic and Ga is principal to Genetic Fuzzy Systems (GFSs) (Cordón *et al*., 2001) is more performing than the traditional intelligent systems. Orriols-Puig *et al*. (2009; Martínez-López and Casillas, 2009; Esmin, 2007), employed GFS in several events Management. They have all got good results.

Recently, the consolidated intelligence technique employing fuzzy logic, Particle Swarm Optimization PSO and genetic algorithms proved that they are the best approach. Many studies practice the hybrid models because the sales input are nonlinear.

Hartigan (1975) has developed the K-means clustering algorithm. It's a simple method and the most famous. The principal of this process is to start with K2 cluster centers and divides into K subsets.

Our research is a comparative of K-means and others clustering methods (Dunham, 2002; Rakhlin and Caponnetto, 2007; Berkhin, 2002; Borah and Ghose, 2009; Han and Kamber, 2006; Xiong *et al*., 2009; Park *et al*., 2006).

This paper suggests a novel hybrid clustering approach utilizing a Gentic Fuzzy System. The article is organized as follows: Section 2, characterizes the proposed model which named Membership Cluster Genetic Fuzzy Systems (MCGFS). After all, in section 3, we finish the article with conclusions.

Materials and Methods

We are going to propose an architecture that consists of two stages (Fig. 1):

Stage1: By using the "fuzzy means" all the input are normalized into K clusters
Stage2: The difference from clusters centers (c_j) to all data (x_i) will be inserted into independent Membership Genetic Fuzzy Systems (MCGFS)

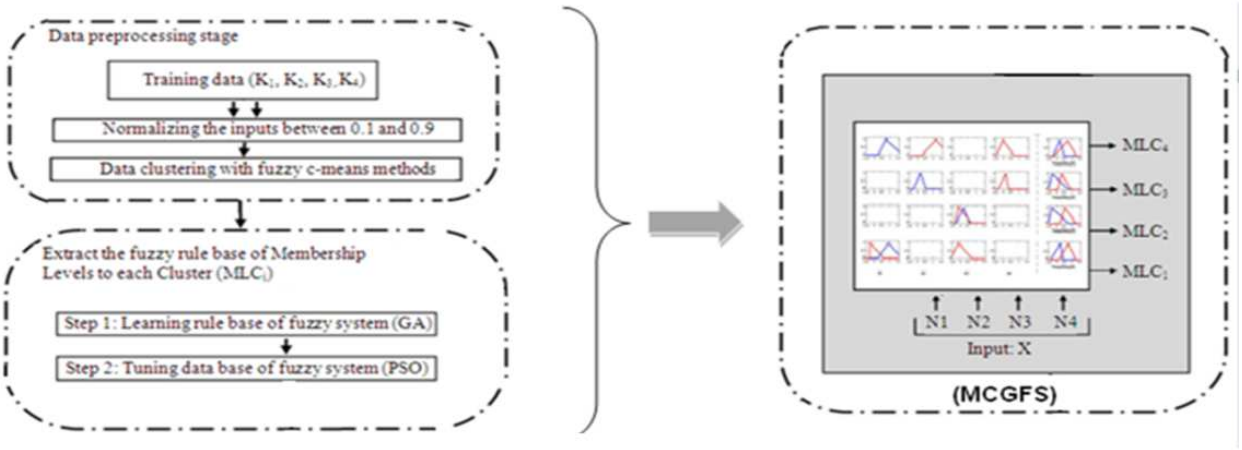

Fig. 1. The architecture of MCGFS model

The variable (K_1, K_2, K_3, K_4) of historical date of an company in Taiwan specialized on electronic is treated like an event of the clustering approach that has been used in différent studies.

Data Preprocessing Stage

This stage contain 2 steps in the first one, we are going to normalized all the records data and in the second one and by using the fuzzy method we are going to normalized records data into K clusters.

Data Normalization

In the interval [0.1, 0.9] all the input values (K_1, K_2, K_3, K_4) will ranged in order to meet property of neural networks.

The equation of the normalization can be expressed as follows:

$$N_i = \frac{0.1 + 0.8 * \left(K_i - \min\left(K_i\right)\right)}{\left(\max\left(K_i\right) - \min\left(K_i\right)\right)} \quad (1)$$

where, N_i a normalized input, K_i is a key variable, max (K_i) is the maximum of the key variables and min (K_i) minimum of the same Key variable.

Fuzzy C-Means Clustering

Input is divided into different clusters using hard clustering. Data elements can appertain to many clusters and joined with each element is a set of membership levels by employing the model Fuzzy C-Means (FCM) (developed by Dunn (1973)) and improved by Bezdek (1981)), it is founded on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2, 1 \le m < \infty$$

where, x_i present the i_{th} of measured data, u_{ij} present the grade of membership of x_i in the cluster j and c_j display the center of the j_{th} cluster. The algorithm is divided of 4 steps:

Step1: Initialize randomly the degrees of membership matrix $U = [u_{ij}]$, $U(0)$

Step2: Count the centroid for every cluster $C(k) = [c_j]$ with $U(k)$:

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

Step3: Update the coefficients for each point in the clusters ($U(k)$,$U(k+1)$):

$$u_{ij} = \frac{1}{\sum_{i=1}^{C}\left(\frac{\left\|x_i - c_j\right\|}{\left\|x_i - c_k\right\|}\right)^{2/(m-1)}}$$

Step4: If $\| U(k+1) - U(k)\| < \varepsilon$, $0 < \varepsilon < 1$. Then STOP; else return to step 2.

This process converges to a saddle point of J_m or a local minimum. The developed parameter combination of two factors (m and ε) are $m = 2$ and $\varepsilon = 0.5$ according to Bezdek (1981).

Using the model fuzzy c-means, we can const at (Table 1) that the use of four clusters is the best between all different clustering numbers.

Extract the Fuzzy IF-THEN Rules of Membership Levels to Each Cluster

The distance between the cluster center and the input record determine the degree of belonging to a cluster.

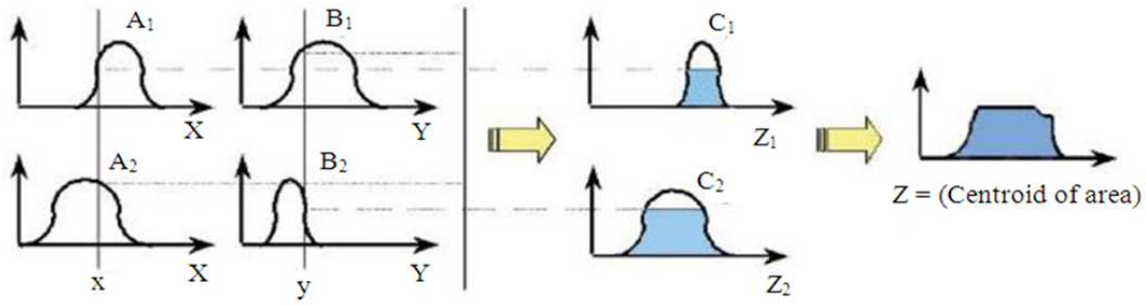

Fig. 2. Fuzzy inference system

Table 1. Similitude of dissimilar clustering algorithms

Clustering groups	Fuzzy c-means total distance
Cluster 2	Gr 23.7904
Cluster 3	Gr 20.2777
Cluster 4	Gr 18.1477

Therefore, there is a strong dependency between the position of cluster center and the degree of belonging to a cluster change the positon every time we add a new input. To avoid this dependency we should use the Genetic Fuzzy Systems (MCGFS) and we will have the rule that define the difference between the cluster and the input records. We can measure the difference between input and cluster by using the fuzzy rules generated.

In recent years, Fuzzy system become the most popular algorithms used to involve problems. The Principe of this process is to conserve in the form of fuzzy linguistic the applicable learning (Fig. 2). It is mixed of the rule base and the Data base (Casillas *et al*., 2004).

Clearly, the human experts found a lot of difficulties to demonstrate their knowledge in the form of fuzzy IF-THEN rules.

To disconcert this issue a lot of historical record had been suggested by using the fuzzy rules. In this way, Intelligent System, such as Genetic Algorithms (GA) (Casillas *et al*., 2004; Cordon and Herrera, 1997) or Particle Swam Optimization (Esmin, 2007) have been attested to be a efficient implement to execute assignment like generation of fuzzy IF-THEN rules. This approaches is called Genetic Fuzzy Systems (GFS) (Cordon and Herrera, 1997).

This stages uses Genetic Fuzzy Systole (GFS), it's a new type that have fuzzy IF-THEN rules of the degree of adhesion to clusters. MCGFS returns the best resukts (two) for each cluster, that offer the distance between centers of clusers and input. The MCGFS have two point.

The drivel RB is the best one:

Step1: The distance between training records and cluster center will be extract for every clusetrs using the genetic algorithm

Step 2: In order to fix data base of fuzzy system and to ameliorate the exactitude of results, we will use the particle swam optimization, changes the forms of appurtenance functions.

Genetic Rule Base Learning Process for FRBS (GA)

The goal of this part is to extract the two best fuzzy RB of the distance between training cluster center and training records, for each cluster. We could define each variable by using the rule defined by a fuzzy linguistic term (ex: None (00), medium (10), small (01) and large (11)).

Using the fuzzy rule, we will presented for each cluster the distance from each cluster center to each record input. The result will be the best chromosomes of the final population. The next steps will be the establishment for this stage (Fig. 3):

Step1: Encoding of chromosomes.

The triangular functions for output and the input variable for linguistic terms could be introduced by two genes and using many genes we can have a chromosome. Using 4 inputs and output variable we could have a specimen coded with a fuzzy rule base (Fig. 4).

Step2: Generating the initial values.

The chromosomes are randomly produced. The first population produced the first one.

Step3: Calculating the fitness values.

In order to have an estimation of the deviation of the training input, we will utilize the mean squared error as the objective function:

$$MSE\left(C_j^k\right)=\frac{1}{N}\sum_{i=1}^{N}\left(Dist_i^k-Out_i^k\right)$$

where, $Dist_i^k$ present the actual distance between the i_{th} training element x_i and k_{th} cluster center and Out_i^k, got from the FRBS utilizing the RB coded in j_{th}

chromosome $\left(C_j^k\right)$, present the output distance between the i_{th} training element x_i and k_{th} cluster center and N present the number of training input.

Step4: Reproduction and selection.

In this stage we will applied the roulette wheel selection (Goldberg, 1989). Without any transformation the best two results of every generation were reproduced in the next one. The binary contest is used for every process. Two individual will be randomly chosen and the best one is selected as a parent: Binary selection.

Step5: Crossover.

In this stage, two point crossover is applied after parameter design.

Step6: Mutation.

In this stage, one point crossover is applied after parameter design (Goldberg 1989).

Step7: Replacement.

The new population produced by the precedent steps updates the old population.

The old population is updated by the new one using the precedent stages:

Step8: Stopping criteria.

Stop, if the number maximum generation is the same as the number of generations else execute the stage 3.

Tuning Process of Fuzzy Rule Bases (PSO)

This sub-stage applies an adjustment process like the genetic adjustment process suggested by (Cordon and Herrera, 1997) In order to upgrade the exactitude of the two best fuzzy rules founds returned by the above generation method, the Particle Swarm Optimization method (PSO) is utilized by the proposed adjustment process in order to update the form of the appurtenance functions of the introductory the 2 RB of each Cluster.

The particle Swarm Optimization Algorithm (PSO) is a population founded on optimization method that discovers the optimal solution utilizing a population of particles (Eberhart and Kennedy, 1995). Every swarm is a solution in the solution space. PSO is fundamentally developed by simulation of bird flocking. PSO can efficiency faster convergence when compared to Genetic Algorithm (GA), because of the equilibrium between exploration and exploitation in the search space (Sivanandam and Visalakshi, 2009).

For each cluster fuzzy Rule Base (RP), we exercise Particle Swarm Optimization Algorithm (PSO) to adjust the parameters (shapes) of membership functions to upgrade the exactness of the asses distances between training records and cluster center. The proposed tuning process is introduced as follows:

Step1: Defining of the search space.

The PSO algorithms runs by having a search space (named a swarm) of candidate solutions (named particles). In our case, the search space is the ensemble of all possible three values performing the triangles of the membership functions. The dimensionality of the search space is 15. Each particle represented by:

$$P_i = \left(a_i^1, b_i^1, c_i^1, a_i^2, b_i^2, c_i^2, a_i^3, b_i^3, c_i^3, a_i^4, b_i^4, c_i^4, o_{1,i}^j, o_{2,i}^j, o_{3,i}^j\right)$$

where, $\left[a_i^k, b_i^k, c_i^k\right]$ presents the three parameters to specify the input triangle fuzzy membership function of the Kith variable (X_k) and $\left[o_{1,i}^j, o_{2,i}^j, o_{3,i}^j\right]$ presents three other parameters to specify the output triangle fuzzy membership function of fuzzy distance between cluster center c_j and normalized record data X(X1,X2,X3,X4).

Step2: Generating the initial population.

Initialization the positions of the particles $p_i = \left(p_i^1 \ldots p_i^{15}\right)$ where p_i^k is initialized with a similarly disturbed random $p_i^k \in U\left(p_i^{k,1}, p_i^{k,r}\right)$ where $p_i^{k,1}$ and $p_i^{k,r}$ presents the lower and upper limits of the kith dimension of the search-space. If t mod 3 = 1, then p_i^t is the left value of the support of a triangular fuzzy number. The triangular fuzzy number is defined by the three parameters $\left(p_i^t, p_i^{t+1}, p_i^{t+2}\right)$ and the intervals of performance are:

$$p_i^t \in \left[p_i^{t,1}, p_i^{t,r}\right] = \left[p_i^t - \frac{1}{2}, p_i^t + \frac{1}{2}\right]$$

$$p_i^{t+1} \in \left[p_i^{t+l,1}, p_i^{t+1,r}\right] = \left[p_i^{t+1} - \frac{1}{2}, p_i^{t+1} + \frac{1}{2}\right]$$

$$p_i^{t+2} \in \left[p_i^{t+2,1}, p_i^{t+2,r}\right] = \left[p_i^{t+2} - \frac{1}{2}, p_i^{t+2} + \frac{1}{2}\right]$$

Step3: For each particle calculate fitness value.

The fitness value for each particle is elaborated employing MSE over a training data set, which is calculated as:

$$MSE\left(P_i\right) = \frac{1}{N}\sum_{i=1}^{N}\left(Dist_i^k - Out_i^k\right)$$

where, $Dist_i^k$ present the actual distance between the l_{th} training element (x_l) and k_{th} cluster center (c_k) and Out_i^k, got from fuzzy rule coded in the particle P_i, present the

output distance between the i_{th} training element x_i and k_{th} cluster center and N present the number of training data.

Step4: Assign best particle's P_i^{best} value to gbest.

Assimilate each particle's P_i fitness evaluation with its P_i^{best}. If the present value is better than P_i^{best}, set the P_i^{best} value to the current value P_i. Compare the population's fitness evaluation with the population's global precedent best (gbest). If the present value is better than gbest, reset the gbest location to the current particle's location:

Step5: Calculate velocity for each particle.

The speed of each of the particles (P_i) for the next generation t+1 are updated as:

The Degree of Membership Levels (MLC_k)

Utilizing the two previous stages, we get six fuzzy rules as outcome (Fig. 6). Each pair of rules offers the distances between records data (X_i) and a cluster center (c_j).

In this level, the sigmoid function is used (Fig. 5) to ameliorate the exactness of results and to have a training process of neural netwok more faster. Then, the advanced fuzzy distance to cluster k ($AF\ D_k$) will be introduced like:

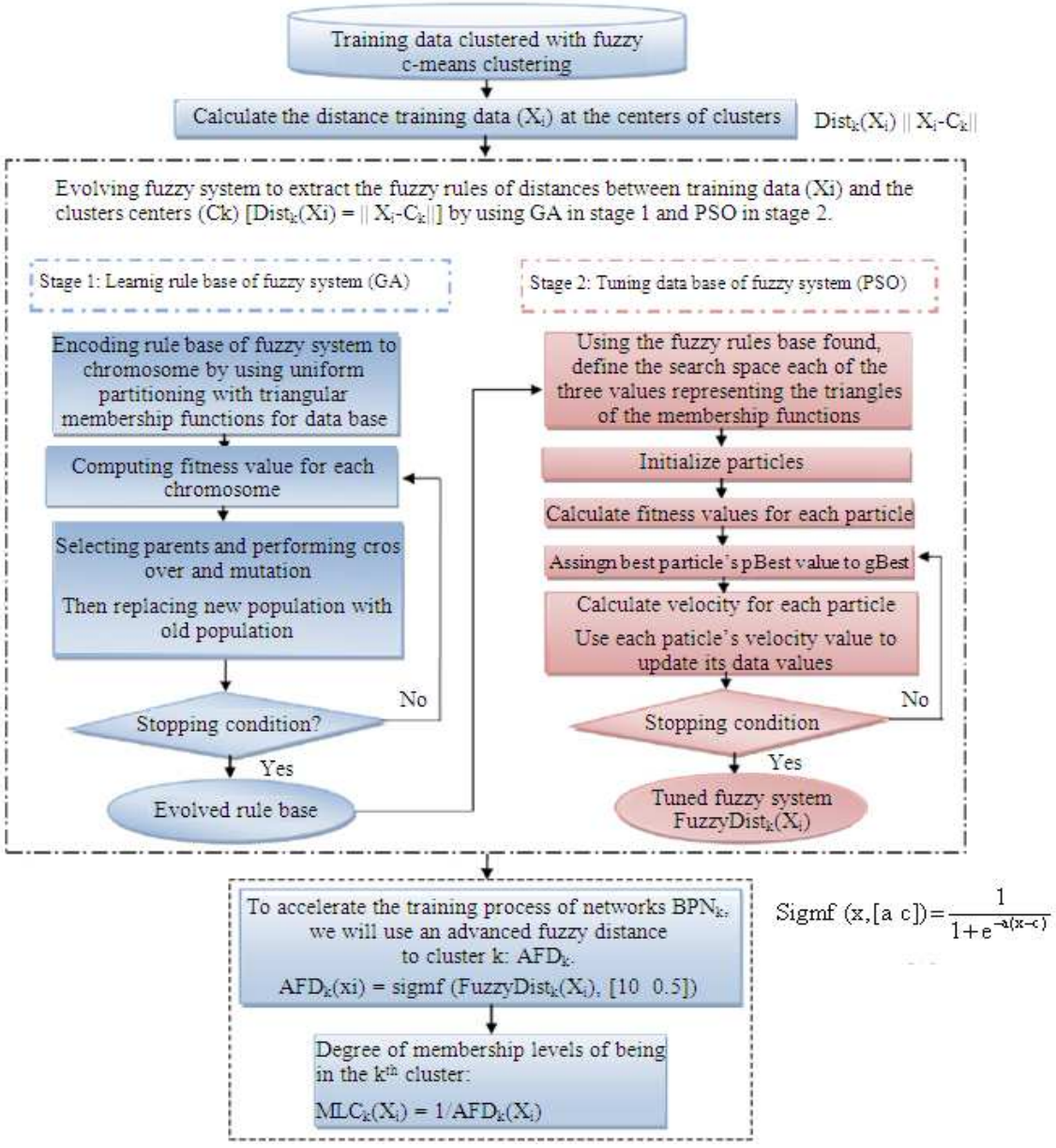

Fig. 3. Architecture of MCGFS

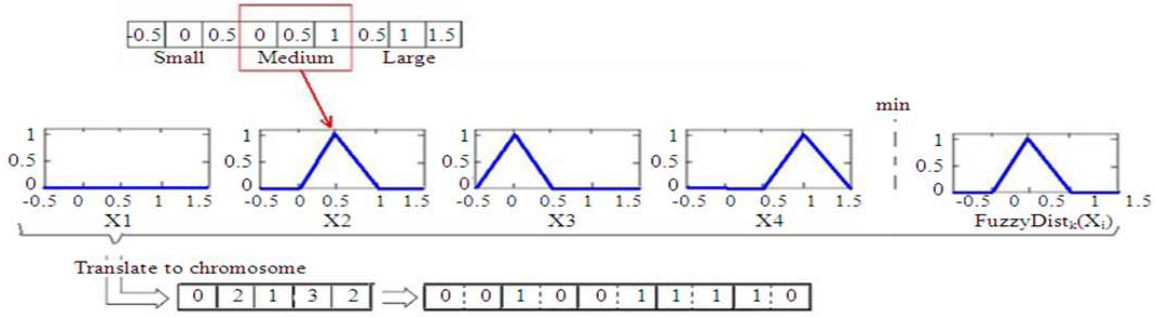

Fig. 4. Coding combination of fuzzy rule base as chromosomes

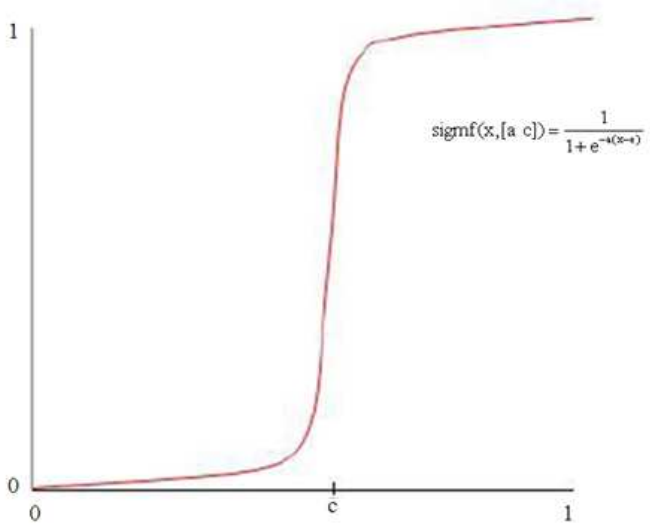

Fig. 5. Sigmoid function, a = 50 and c = 0,5

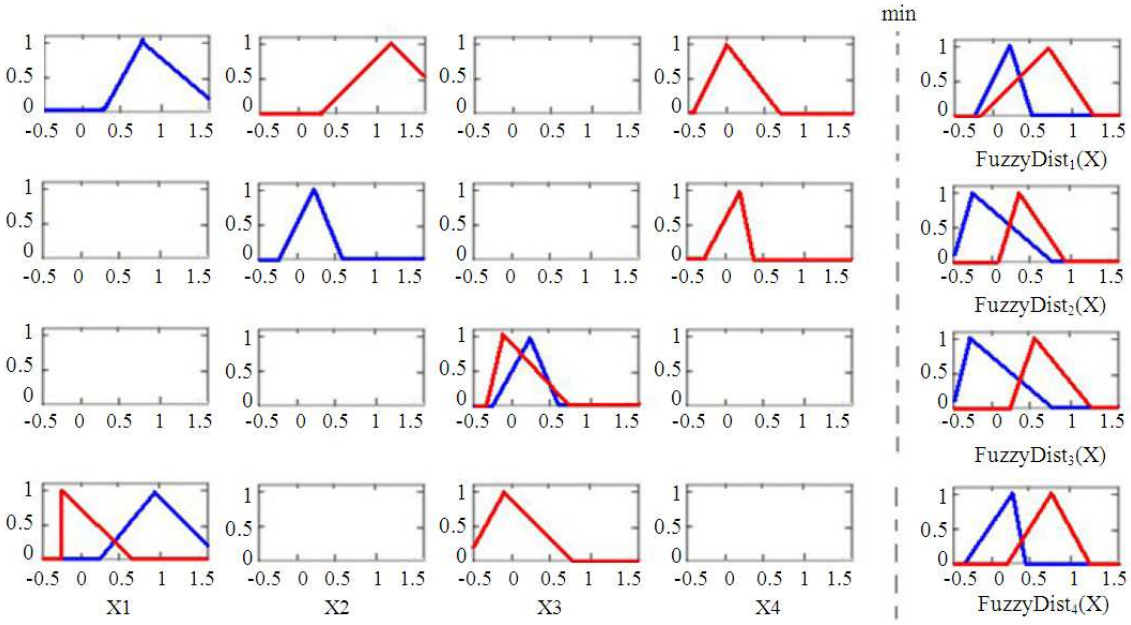

Fig. 6. The tuned membership functions of input and output variables for clusters GFS

$$AF\ D_k\left(X_i\right)=\text{sigmf}\left(FuzzyDist_k\left(X_i\right),[50,\ 0.5]\right)$$

$$\text{sigmf}\left(x,[a,c]\right)=\frac{1}{1+e^{-a(a..c)}}$$

The degree of appurtenance stage of a record X_i to kith cluster (MLC_k (X_i)) is related inversely to the distance from records data X_i to the cluster center c_k ($AF\ D_k$ (X_i)).

The grade of belonging stage of membership of a record X_i to kith cluster (MLC_k (X_i)) is related inversely to the distance from records data X_i to the cluster center ck ($AF\ D_k$ (X_i)):

$$MLC_k\left(X_i\right)=\frac{1}{AFD_k\left(X_i\right)}$$

Results

Constructing MCGFS Model

Our proposed system (MCGFS) has two stages: Steps: (1) using a method called Fuzzy clustering, all data elements will be clustered into N groups; (2) utilizing a Fuzzy Genetic System, the fuzzy rules of membership levels to each cluster will be generated.

The proposed MCGFS system was applied to forecast the sales data of the PCB. The results are in Table 2-4.

We chose BPN with clustering data as a forecast method. A parallel BP networks is trained with a learning rate adapted to the stage of cluster appurtenance of every record of training input We will compare the result of use of BPN with three clustering method:

- K-means
- Fuzzy c-means
- Membership Cluster Genetic Fuzzy Systems (MCGFS)

Comparisons of GFCBPN Model with Other Previous Models

Experimental comparison of outputs of GFCBPN with other methods show that the proposed model outperforms the previous approaches (Table 2-4 and Fig 7-9). We apply two different performance measures called Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE), to compare the BPN with MCGFS model with the previous methods, i.e., Fuzzy C-means and K-means:

$$MAPE=100\times\frac{1}{N}\sum_{t=1}^{N}\frac{\left|Y_t\ -P_t\right|}{Y_t}$$

where, P_t is the expected value for period t, Y_t is the actual value for period t and N is the number of periods.

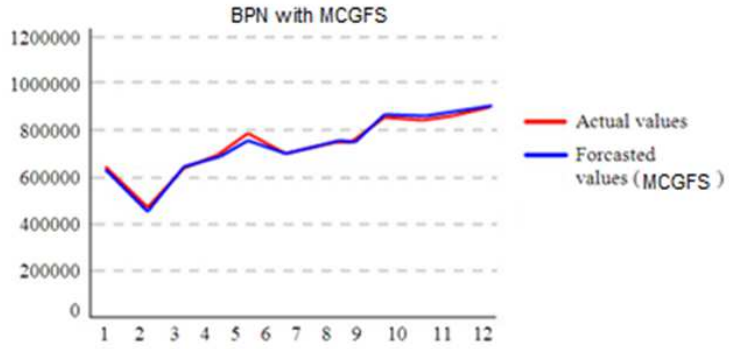

Fig. 7. The MAPE of BPN with MCGFS

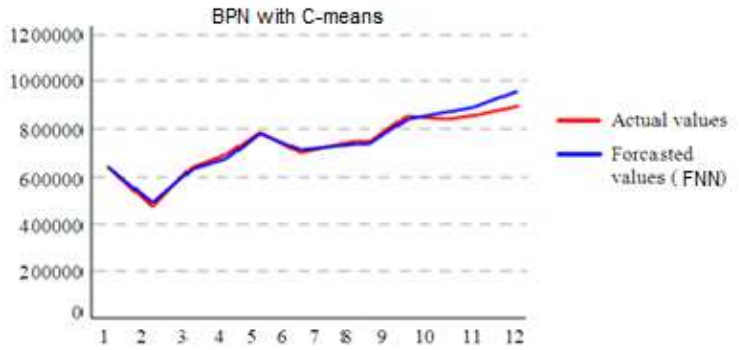

Fig. 8. The MAPE of FNN

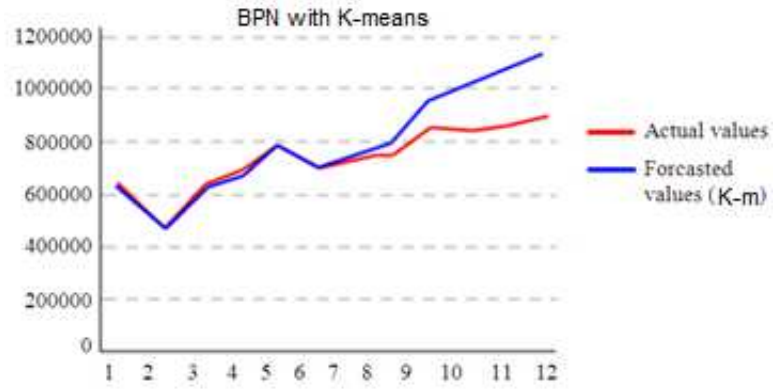

Fig. 9. The MAPE of BPN

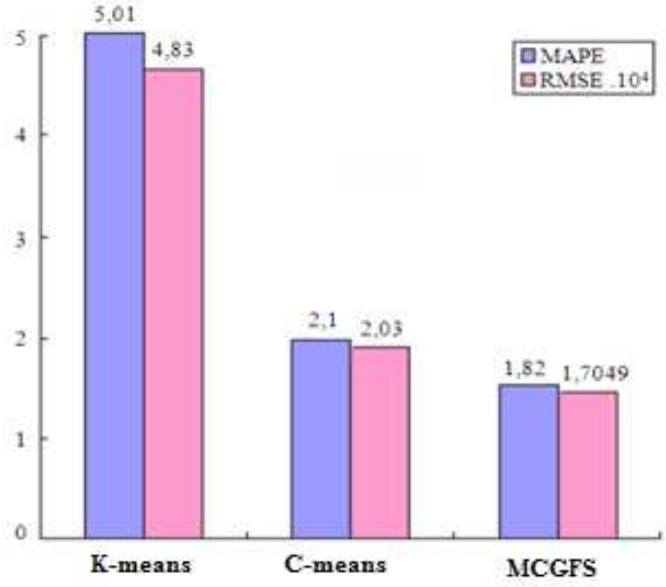

Fig. 10. The performance improvement of MCGFS after using GSF (MAPE and RMSE)

Table 2. The forecasted results by BPN with MCGFS method

Month	Forecasted values	Actual values
2003/1	638,749	649,066
2003/2	443,585	466,750
2003/3	633,837	633,615
2003/4	675,897	693,946
2003/5	747,220	785,838
2003/6	686,641	679,312
2003/7	724,807	723,914
2003/8	754,198	757,490
2003/9	826,618	836,846
2003/10	849,560	833,012
2003/11	874,510	860,892
2003/12	895,338	912,182

As shown in Fig. 10, to have a good precision is better to use the MCGFS than the fuzzy c-means clustering.

Table 3. The forecasted results by BPN with C-means method

Month	Forecasted values (FNN)	Actual values
2003/1	584,901.9	649,066
2003/2	483,872.3	466,750
2003/3	713,874.6	633,615
2003/4	711,356.1	693,946
2003/5	769,881.6	785,838
2003/6	684,634.5	679,312
2003/7	721,192.4	723,914
2003/8	770,609	757,490
2003/9	817,423.4	836,846
2003/10	851,827	833,012
2003/11	884,484.1	860,892
2003/12	912,129.1	912,182

It has made 1,7 as MAPE evaluation and 1820 as RMSE evaluation. The previous approaches regarding MAPE and RMSE evaluations Fig. 10 had performed by the forecasting accurance of GFCBPN.

Table 4. The forecasted results by BPN with K-means method

Month	Forecasted values	Actual values
2003/1	622,402.3	649,066
2003/2	456,226	466,750
2003/3	618,346	633,615
2003/4	669,445.5	693,946
2003/5	795,971.6	785,838
2003/6	682,646.4	679,312
2003/7	741,996.5	723,914
2003/8	789,756.8	757,490
2003/9	945,738.1	836,846
2003/10	1,006,899	833,012
2003/11	1,077,823	860,892
2003/12	1,141,621	912,182

Conclusion

This article offers a new hybrid system founded on genetic fuzzy clustering (MCGFS). Compared to others approach which tend to utilize the classical hard clustering methods (K-means clustering to separate data set into subgroups so as to minimize the noise and form more homogeneous clusters (Chang *et al*., 2009), the benefit of our proposal system (MCGFS) is that it employs a fuzzy clustering (fuzzy c-means clustering) which permits each data record to appertain to each cluster to some grade, which permits the clusters to be large which consequently raises the accuracy of forecasting system results.

Another benefit of our approach is with no dependencies of the positions of the cluster centers, the estimation of belonging degree of each input record to each cluster is calulated.

Author's Contributions

Hamedoun Lamiae: Participate in all experiments.

Attariuas Hicham: Designed the research plan.

Ben Maati Mohamed Larbi: Give final approval of the version.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved

References

Berkhin, P., 2002. Survey of clustering data mining techniques. Accrue Software, Inc.

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algoritms. 1st Edn., Plenum Press, New York, pp: 256.

Borah, S. and M.K. Ghose, 2009. Performance analysis of AIM-K-means and K-means in quality cluster generation. J. Comput., 1: 175-178.

Casillas, J., O. Cordón, F. Herrera and P. Villar, 2004. A hybrid learning process for the knowledge base of a fuzzy rule-based system. Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge based Systems, (UKS' 04), Perugia, Italy, pp: 2189-2196.

Chang, P.C., C.H. Liu and C.Y. Fan, 2009. Data clustering and fuzzy neural network for sales forecasting: A case study in printed circuit board industry. Knowl. Based Syst., 22: 344-355. DOI: 10.1016/j.knosys.2009.02.005

Cordon, O. and F. Herrera, 1997. A three-stage evolutionary process for learning descriptive and approximate fuzzy-logic-controller knowledge bases from examples. Int. J. Approximate Reason., 17: 369-407. DOI: 10.1016/S0888-613X(96)00133-8

Cordón, O., F. Herrera, F. Hoffmann and L. Magdalena, 2001. Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases. 1st Edn., World Scientific, Singapore, ISBN-10: 9810240171, pp: 462.

Dunham, M., 2002. Data Mining: Introductory and Advanced Topics. 1st Edn., Prentice Hall, USA., ISBN-10: 0130888923, pp: 315.

Dunn, J.C., 1973. A fuzzy relative of the ISODATA Process and its use in detecting compact well-separated clusters. J. Cybernet., 3: 32-57. DOI: 10.1080/01969727308546046

Eberhart, R.C. and J. Kennedy, 1995. A new optimizer using particle swarm theory. Proceedings of the 6th International Symposium on Micro Machine and Human Science, Oct. 4-6, IEEE Xplore Press, Nagoya, pp: 39-43. DOI: 10.1109/MHS.1995.494215

Esmin, A., 2007. Generating fuzzy rules from examples using the particle swarm optimization algorithm. Proceedings of the 7th International Conference IEEE Hybrid Intelligent Systems, Sept. 17-19, IEEE Xplore Press, Kaiserlautern, pp: 340-343. DOI: 10.1109/HIS.2007.52

Goldberg, D.A., 1989. Genetic Algorithms in Search, Optimization and Machine Learning. 1st Edn., Addison-Wesley.

Han, J. and M. Kamber, 2006. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers. 2nd Edn., New Delhi, ISBN-13: 978-81-312-0535-8.

Hartigan, J.A., 1975. Clustering Algorithms. 1st Edn., Wiley, New York, ISBN-10: 047135645X, pp: 351.

Martínez-López, F. and J. Casillas, 2009. Marketing intelligent systems for consumer behaviour modelling by a descriptive induction approach based on genetic fuzzy system. Industrial Marketing Manage., 38: 714-731. DOI: 10.1016/j.indmarman.2008.02.003

Oh, K.J. and I. Han, 2001. An intelligent clustering forecasting system based on change-point detection and artificial neural networks: Application to financial economics. Proceedings of the 34th Annual Hawaii International Conference on System Sciences, Jan. 3-6, IEEE Xplore Press, pp: 8-8.

Orriols-Puig, A., J. Casillas and F. Martínez-López, 2009. Unsupervised learning of fuzzy association rules for consumer behavior modeling. Mathware Soft Comput., 16: 29-43.

Park, H.S., J.S. Lee and C.H. Jun, 2006. A K-means like algorithm for K-medoids clustering and its performance.

Rakhlin, A. and A. Caponnetto, 2007. Stability of k-Means clustering. Adv. Neural Inform. Process. Syst., 12: 216-222.

Sivanandam and Visalakshi, 2009. Dynamic task scheduling with load balancing using hybrid particle swarm optimization. Int. J Bio-Inspired Comput., 1: 267-268.

Xiong, H., J. Wu and J. Chen, 2009. K-Means clustering versus validation measures: A data distribution perspective. IEEE Trans. Syst. Man Cybernet. Part B, 39: 318-331.