Journal of Information Systems and Telecommunication
Vol.12, No.1, January-March 2024, 62-71
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Proposing an FCM-MCOA Clustering Approach Stacked with Convolutional Neural Networks for Analysis of Customers in Insurance Company

Meisam Yadollahzadeh Tabari[1*], Motahare Ghavidel[1], Mehdi Golsorkhtabaramiri[2]

[1.] Department of Computer Engineering, Islamic Azad University, Babol Branch, Babol, Iran.

Received: 06 March 2023/ Revised: 20 Sep 2023/ Accepted: 11 Nov 2023

Abstract

To create a customer-based marketing strategy, it is necessary to perform a proper analysis of customer data so that customers can be separated from each other or predict their future behavior. The datasets related to customers in any business usually are high-dimensional with too many instances and include both supervised and unsupervised ones. For this reason, companies today are trying to satisfy their customers as much as possible. This issue requires careful consideration of customers from several aspects. Data mining algorithms are one of the practical methods in businesses to find the required knowledge from customer's both demographic and behavioral. This paper presents a hybrid clustering algorithm using the Fuzzy C-Means (FCM) method and the Modified Cuckoo Optimization Algorithm (MCOA). Since customer data analysis has a key role in ensuring a company's profitability, The Insurance Company (TIC) dataset is utilized for the experiments and performance evaluation. We compare the convergence of the proposed FCM-MCOA approach with some conventional optimization methods, such as Genetic Algorithm (GA) and Invasive Weed Optimization (IWO). Moreover, we suggest a customer classifier using the Convolutional Neural Networks (CNNs). Simulation results reveal that the FCM-MCOA converges faster than conventional clustering methods. In addition, the results indicate that the accuracy of the CNN-based classifier is more than 98%. CNN-based classifier converges after some couples of iterations, which shows a fast convergence in comparison with the conventional classifiers, such as Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighborhood (KNN), and Naive Bayes (NB) classifiers.

Keywords: Customer Clustering; Fuzzy C-Means; Cuckoo Optimization; Convolutional Neural Networks.

1- Introduction

Customers are one of the prominent parts of the commercial exchanges, and no business can succeed without having satisfied and faithful customers. This issue requires the study of customers from various aspects. To create a customer-based marketing strategy, several techniques, such as clustering and scoring, can be utilized for data analysis. Data mining algorithms offer several practical methods for businesses to extract the expected data.

This paper presents some solutions for clustering and classification of the customers. The Insurance Company (TIC) dataset is utilized which first introduced in Computational Intelligence and Learning (COIL) Challenge 2000 [1]. This challenge is a common data mining problem to predict the potential customers using the training and test datasets of customers. Identify potential purchasers is a powerful approach to advertising and market a product. If the company had a precise data of their clients, they can send fewer advertising emails and some expenses can be reduced in this way. For example, an insurance company often wants to know which customers are willing to buy a particular product, such as caravan insurance policy.

Several solutions have been recommended for this challenge based on data mining and computational intelligence. We generally tend to remove some features that are not effective. On the other hand, we tend to preserve client records, since removing clients may eliminate some important client groups. To handle this tradeoff, Fuzzy C-Means (FCM) clustering is a popular technique for selecting effective features and clients [2]. Using the k-means approach for clustering is another well-

✉ **Meisam Yadollahzadeh Tabari**
m_tabari@baboliau.ac.ir

known technique that uses an iterative approach to minimizing the sum of squared errors [3]. The purpose of FCM clustering is to group the data to pick out the groups with considerable numbers of potential purchasers. To identify the best centroid, the FCM algorithm employs the membership function. However, the number of features is generally large, and clustering is computationally difficult in such a condition. Moreover, the data are inherently sparse, and these sporadic data can be problematic for the clustering mechanisms. In summary, the clustering approaches shoulld combat with two main problems: the large number of clusters and imbalanced data. The former problem can be solved by merging similar clusters through an optimization process [4]. A popular and simple technique for reducinge features is selecting one feature each time. When the dimensions of the features are reduced, the basic form of the FCM clustering may lead to acceptable results [3].

The latter problem is that the distribution of the features is unbalanced, i.e., there are some classes with numerous members and the rest of classes have a few members. Existing classification techniques tend to find the classes with more members. There are several data-level approaches in the literature [5], [6], such as sampling techniques, that modify the distribution of the train data. These techniques have this power that are not dependent on the classifier types utilized. For example, the under-sampling technique ignores some data samples and provides a subgroup of the initial data. In addition to these data-level aproaches, imbalence data problem can effectively be solved by modifying well-known algorithms, such as the FCM algorithm

Customers are one of the prominent parts of the commercial exchanges, and no business can succeed without having satisfied and faithful customers. This issue requires the study of customers from various aspects. To create a customer-based marketing strategy, several techniques, such as clustering and scoring, can be utilized for data analysis. Data mining algorithms offer several practical methods for businesses to extract the expected data.

This paper presents some solutions for clustering and classification of the customers. The Insurance Company (TIC) dataset is utilized which first introduced in Computational Intelligence and Learning (COIL) Challenge 2000 [1]. This challenge is a common data mining problem to predict the potential customers using the training and test datasets of customers. Identify potential purchasers is a powerful approach to advertising and market a product. If the company had a precise data of their clients, they can send fewer advertising emails and some expenses can be reduced in this way. For example, an insurance company often wants to know which customers are willing to buy a particular product, such as caravan insurance policy.

Several solutions have been recommended for this challenge based on data mining and computational intelligence. We generally tend to remove some features that are not effective. On the other hand, we tend to preserve client records, since removing clients may eliminate some important client groups. To handle this tradeoff, Fuzzy C-Means (FCM) clustering is a popular technique for selecting effective features and clients [2]. Using the k-means approach for clustering is another well-known technique that uses an iterative approach to minimizing the sum of squared errors [3]. The purpose of FCM clustering is to group the data to pick out the groups with considerable numbers of potential purchasers. To identify the best centroid, the FCM algorithm employs the membership function. However, the number of features is generally large, and clustering is computationally difficult in such a condition. Moreover, the data are inherently sparse, and these sporadic data can be problematic for the clustering mechanisms. In summary, the clustering approaches shoulld combat with two main problems: the large number of clusters and imbalanced data. The former problem can be solved by merging similar clusters through an optimization process [4]. A popular and simple technique for reducinge features is selecting one feature each time. When the dimensions of the features are reduced, the basic form of the FCM clustering may lead to acceptable results [3].

The latter problem is that the distribution of the features is unbalanced, i.e., there are some classes with numerous members and the rest of classes have a few members. Existing classification techniques tend to find the classes with more members. There are several data-level approaches in the literature [5], [6], such as sampling techniques, that modify the distribution of the train data. These techniques have this power that are not dependent on the classifier types utilized. For example, the under-sampling technique ignores some data samples and provides a subgroup of the initial data. In addition to these data-level aproaches, imbalence data problem can effectively be solved by modifying well-known algorithms, such as the FCM algorithm.

2- Related Works

In [4], the authors suggested a new method for handling the CoIL Challenge 2000, where the goal of the clustering is to find a scoring table and select more important features. They perform FCM clustering for each feature and compute a response density, as a measure of the predictive capability for each cluster. The response density can be defined as the ratio of the number of caravan policy purchasers to the total number of people in the cluster. In other words, calculating the response density yields a criterion that permits us to order the clusters based on their predictive effectiveness. We compute a score for each client adopting their membership values and response densities. The clustering procedure can be executed many times, and the selection of useful features can be completed using these scores [7], [8].

In optimization-based clustering, the clustering objective is considered as the minimum sum of the square error and the optimization procedure is used in the related algorithm to solve the clustering objective. In this regard, most of the algorithms have utilized centroid sets as the solutions to generate optimal cluster centroids. In a fuzzy technique, a customer can belong to different clusters simultaneously, and the results highly depend on the initial cluster centers. Improper selection of centers causes that the algorithm fall into a local solution and this degrades the performance of the FCM clustering method [9], [10]. Several techniques, such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO) have been recommended to combat with this issue [11]. The authors in [27] have introduced a hybrid data clustering approach using GA and K-means algorithms. Their results demonstrate that this hybrid approach can improve the clustering performance. In [12-14], some hybrid optimization algorithms are suggested to obtain a better clustering performance. In order to achieve a faster convergence, in [15] and [16], the authors have suggested some hybrid methods based on the K-means and FCM clustering.

Cuckoo search algorithm is a powerful tool which can be utilized in many optimization problems [17], [18]. Cuckoo Optimization Algorithm is inspired by the behavior of a bird family called Cuckoo. Smart egg laying and breeding procedure of cuckoos is the essential of this algorithm. Instead of establishing their own nest, female cuckoos offspring in nests of host birds, and if these eggs are not destroyed by the host birds, they can grow into adult birds. Mature cuckoos immigrate to new habitats and this immigration lead to discover more food and better conditions for reproduction. Several groups are created in other locations due to the immigration, and the community with outstanding conditions is preferred as the target for other cuckoos. From an algorithmic point of view, the best environment is the global optimum of the objective function. We can form our groups, identify the best group, and find most proper target for immigration. Since the mature cuckoos are dispersed in an environment, understanding that which cuckoo is a member of which of the groups is a problematic task. The grouping is often accomplished with one of the well-known clustering techniques and their objective value is calculated. After some couples of iterations, all population immigrate to the best habitat and algorithm converges. The cuckoo search algorithm can work better than many other optimization techniques in solving common benchmark problems [19], [20].

2-1- Contributions

Innovative aspects of this work are summarized below:

- To improve the results of the clustering, a new hybrid clustering method based on FCM and Modified Cuckoo Optimization Algorithm (MCOA) is presented. The results show that FCM-MCOA converges faster than some existing methods, such as and Invasive Weed Optimization (IWO) and GA methods. Moreover, the final value of the cost function is less than the aforementioned conventional methods.
- Our ultimate goal is to identify potential caravan insurance purchasers. In order to predict who would be interested in purchasing a caravan policy, a classifier is also designed based on the principles of the Convolutional Neural Networks (CNNs). Simulation results show that our classifier is more accurate than some conventional classifiers, such as SVM and NB classifiers. We also showed that increasing the number of out-put neurons in the FC layers can improve the performance of the classifier.

The rest of this paper is organized as follows. The suggested FCM-MCOA approach and CNN-based classifier are represented in Section 2. Simulation results are described in Section 3. Finally, conclusions are represented in Section 4.

3- The Proposed Method

In this section, we first introduce the hybrid FCM-MCOA approach. Afterward, we define the designed CNN-based classifier in order to predict probable purchasers of a special type of insurance service.

3-1- Subheadings

In this paper, the MCOA approach is considered as a stochastic search technique to find new cluster centers. The cuckoo search procedure is to avoid the solutions from being captured into the local points and finding the global solutions. Another advantage of the MCOA is that a few parameters should be adjusted in this method. The proposed hybrid FCM-MCOA can be summarized as follows:

1. Some initial habitats are randomly allocated to the cuckoos.
2. Some eggs are assigned to each cuckoo. The minimum and maximum number of eggs should be defined.
3. For each cuckoo, an Egg Laying Radius (ELR) should be defined:

$$\mathrm{ELR} = \alpha \times \frac{\text{Number of current cuckoo's eggs}}{\text{Total number of eggs}} \times (\mathrm{var_h} - \mathrm{var_l}) \tag{1}$$

where α is an integer, var_h and var_i are the upper and lower bounds of the decision variables, respectively.

4. Each cuckoo lays its eggs within its ELR.
5. A number of eggs are detected and destroyed by the host bird.
6. Cuckoo eggs turn into young birds.
7. The habitat of the young birds is assessed. The habitat is actually a cost function of the problem.
8. The number of live cuckoos should be limited. The N_{max} indicates the maximum number of live cuckoos.
9. Cuckoos should immigrate to a better habitat in the search area. Therefore, we should first group the cuckoos. In this paper, grouping is done utilizing FCM method. Calculating the cluster centers by considering the membership values of each data sample as well as calculating the membership values are performed using the FCM. The objective function for data samples is defined as [10], [11].

$$J_m(U,V) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m D_{ik}(x_k, v_i) \tag{2}$$

In where n is the number of samples, $m > 1$ is the "fuzzifier", c is the number of clusters, U is the membership matrix, V is the cluster centers set, and $D_{ik}(x_k, v_i)$ represents squared distance between the k^{th} data sample and i^{th} cluster center. The membership values and cluster centers are given by

$$u_{ik} = \frac{D_{ik}(x_k, v_i)^{\frac{1}{1-m}}}{\sum_{j=1}^{c} D_{jk}(x_k, v_j)^{\frac{1}{1-m}}} \tag{3}$$

$$v_i = \frac{\sum_{j=1}^{n} (u_{ij})^m x_j}{\sum_{j=1}^{n} (u_{ij})^m} \tag{4}$$

1. The average cost function of each group is calculated and the best group is identified. This group of cuckoos then migrate to the best habitat.
2. In the MCOA, the ELR should be reduced in the next iteration. It can be done by reducing the α in Eq. (1).
3. If the stop condition of the algorithm is fulfilled, the optimization process ends. Otherwise, it starts again from the step 2.

Calculating the centers of clusters is often a difficult task. The optimization algorithm based on the FCM-MCOA is adopted to effectively calculate the centers of the clusters. For this purpose, the following cost function is minimized by the algorithm.

$$J = \sum_{j=1}^{k}\sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \tag{5}$$

where $\left\| x_i^{(j)} - c_j \right\|$ is a measure of the distance between the data points $x_i^{(j)}$ and the center of the cluster c_j .The parameter J defines the distance of n data points from their corresponding cluster centers. The flowchart of the proposed clustering technique is presented in Figure 1.

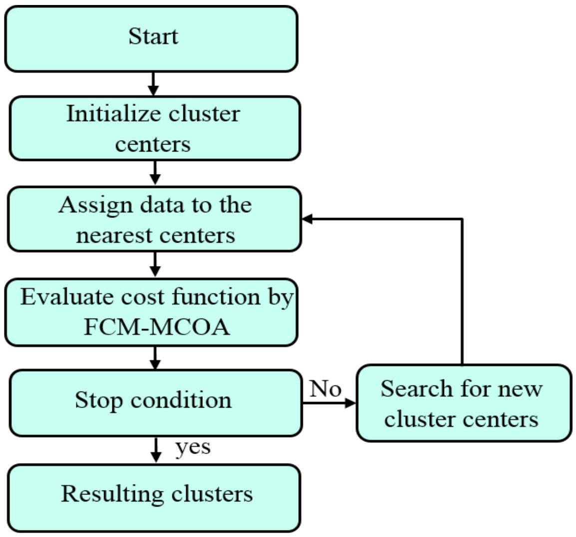


Figure 1 Flowchart of the proposed FCM-MCOA approach.

3-2- Proposed CNN-based Classifier

Several classification techniques are designed to solve the COIL challenge problem reducing the imbalanced data issue [25]. A strong classifier should properly identify two classes of the caravan policyholders and the rest of people (i.e., 0 and 1 classes).

CNNs have many applications in pattern recognition [21]-[23], and digital image processing [28]. CNN-based classifiers may have different architectures in different applications. After the input layer, there are one or more Fully Connected (FC) layers, such as the standard feed-forward neural networks. These layers are stacked to construct a deep model. The basic architecture of the CNN-based classifiers includes some other layers, such as convolutional filters and batch normalization layers. Finally, the last FC layer out-puts the class label. In addition to this basic architecture, several architecture variations have been suggested for modern applications. CNNs represent the input data in the form of multi-dimensional arrays, and each layer connects to the next layer by neurons. The effect of each neuron is different from other neurons. A typical CNN-based classifier includes the parts described below:

- Input layer: Consider the l^{th} layer of the network, whose inputs form an order 3 tensor as $x^l \epsilon R^{M^l \times N^l \times P^l}$. A function converts the input x^l to an out-put y. Note that the out-put of the l^{th} layer is the input to the layer $l+1$. In other words, y and x^{l+1} actually refer to one object. We assume that the out-put is of size $M^{l+1} \times N^{l+1} \times P^{l+1}$. Thus, an out-put element is indexed by a triplet $(i^{l+1}; j^{l+1}; k^{l+1})$, where $0 \le i^{l+1} < M^{l+1}$, $0 \le j^{l+1} < N^{l+1}$, and $0 \le k^{l+1} < P^{l+1}$.
- ReLU layer: This layer does not change the size of the input, and x^l and y have the same size. It means that the ReLU layer can be considered as a separate section for each element of the input, i.e.

$$y_{i,j,k} = max\{0, x^l_{i,j,k}\} \tag{6}$$

where, $0 \le i < M^l = M^{l+1}$, $0 \le j < N^l = N^{l+1}$, $0 \le k < P^l = P^{l+1}$.

- Convolutional layer: This layer extracts the features and learns them from the input data. Each neuron in the convolutional layer has a receptive field, and each receptive field is connected to the other neurons in the previous layer. Multiple convolution kernels are often used in a convolutional layer. Assume that D kernels are used, and each kernel is of spatial span $M \times N$. We show all these kernels by $\boldsymbol{h}$. The convolution process can be expressed as

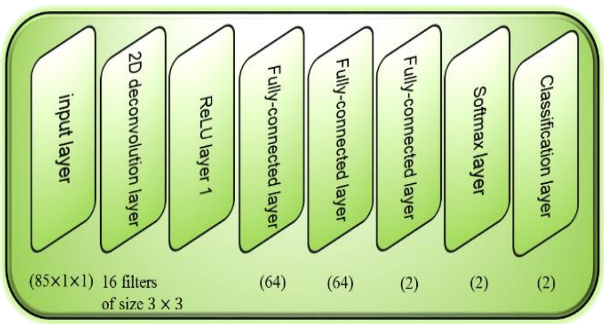
$$k = \sum_{i=0}^{M} \sum_{j=0}^{N} \sum_{k^l=0}^{P^l} h_{i,j,k^l,k} \times x^l_{i^{l+1}+i, j^{l+1}+j, k^l}, \tag{7}$$

where $x^l_{i^{l+1}+i, j^{l+1}+j, k^l}$ refers to an element of the x^l that indexed by the triplet $(i^{l+1}+i, j^{l+1}+j, k^l)$.

- Fully Connected (FC) Layer: The FC layer enhances the stability using several non-linear functions. Fully interconnected layers can follow these layers to extract and analyze the features, and perform the function of high-level information. Assume that the input of the l^{th} layer is of size $M^l \times N^l \times P^l$. Adopting convolution kernels with size $M^l \times N^l \times P^l$, using D kernels form an order 4 tensor of size $M^l \times N^l \times P^l \times D$, where the out-put is $\boldsymbol{y} \epsilon R^D$. To calculate the elements of the out-put, all elements of x^l should be used.
- Training: In order to gain the desired out-put, Deep Learning (DL) models typically utilize a learning algorithm, such as the back propagation algorithm, to adjust their parameters This algorithm controls the

network parameters by minimizing an objective function.

- Figure 2 illustrates the architecture of the suggested CNN-based classifier. The 2-D convolutional layer creates a layer with 16 filters of size 3×3. During the training process, the size of the padding can be calculated such that the out-put has the same size as the input. We consider two successive FC layers of size 64. The size of the final FC layer is 2, same as the number of out-put classes.



.

Figure 2 The proposed architecture of the CNN-based classifier

4- Simulation Result

In order to evaluate the performance of the proposed clustering and classification methods, some simulations are presented. All algorithms are implemented in MATLAB R2020a version, and tested on Intel(R) core TM i8-7000k CPU @ 5 GHz with 64 GB internal RAM.

4-1- The Insurance Company (TIC) Dataset

In this study, The Insurance Company (TIC) dataset is utilized for simulations [24]. TIC dataset comprises 86 features, product holders, and analytical information. The training dataset comprises more than 5000 reports about clients, including the information on having the caravan policy. The test dataset contains of 4000 clients. For the prediction task, the ultimate objective is to detect the clients that may purchase a caravan policy. The identified purchasers can be pulled out, and the other people receive the advertising emails. We organize the train and test data as the following:

- Train data: This dataset is used to train our prediction models, and comprises the information of 5822 clients. Each client record comprises 86 features, consisting of demographic data (features1-43), and purchasers of an insurance policy (features 44-86). All clients in the same cities have similar demographic features. Feature 86 includes the information about the holders of the caravan insurance policy, and their phone number. In other words, feature 86 can be utilized as the target variable.
- Test data: This dataset includes the information about 4000 clients. This dataset is adopted for prediction task and only the insurance company managers know that whether the clients have caravan insurance policy or not. It has the same form as train data, only the target variable is missing.
- Target values: Target values are used for the performance evaluation.

The suggested clustering and classification methods are tested on some informative features. The meaning of the utilized features and their values are listed in Table 1. The following illustrates the simulation results for data clustering and classification tasks.

Table 1: Margin specifications

Number	Name	Description	Domain
1	MOSTYPE	Customer subtype	1-41
2	MAANTUHI	Number of houses	1-10
5	MOSHFOOD	Customer main type	1-10
16	MOPLHOOG	High level education	-
68	APERSAUT	Number of car policies	-
82	APLEZIER	Num. of boat policies	-
86	CARAVAN	Target variable	0-1

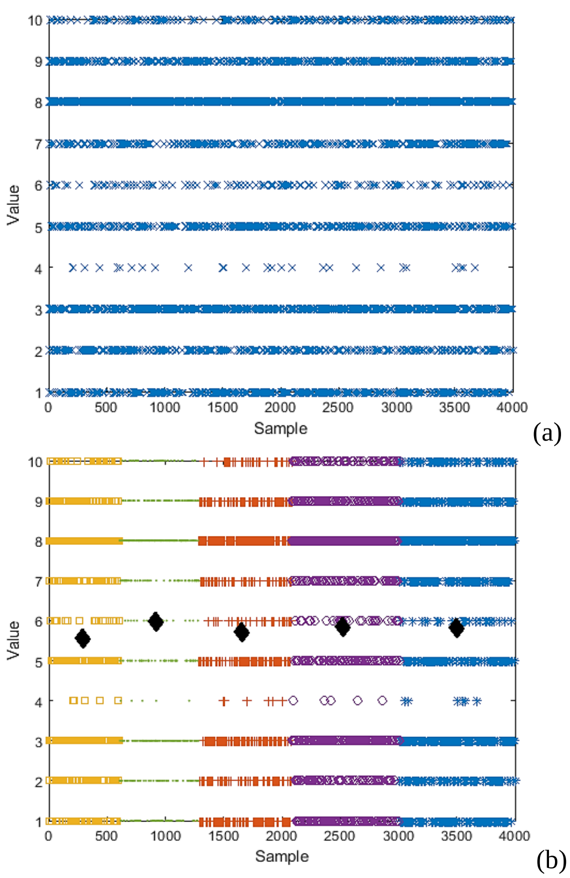
4-2- Clustering Results

The utilized parameters for FCM-MCOA are listed in Table 2. The results for 5 clusters of different features of the TIC dataset are shown in Figures 3 to 6. The selected features are MOSHOOF, MOPLHOOG, APERSAUT, and APLEZIER. Clusters are marked with different colors in each shape.

Figure 3 (a) displays the initial distribution of cuckoos/people in the environment. Figure 3 (b) shows 5 best habitat for 5 selected clusters. The centers of the clusters are marked with diamond in the figures. For example, for the first cluster of the MOSHOOF feature, the best habitat is (293.4, 5.6). Figure 3 (C) provides a comparison between the convergence curve of the FCM-MCOA, GA, and IWO methods. The initial population, mutation rate, and selection rate of GA are set to 100, 0.2, and 0.5, respectively. The convergence curves of IWO [26] and GA [27] reach to their best point after about 40 iterations, while the FCM-MCOA reaches to its best point only after 30 iterations. From the cost minimization curves, it can be concluded that the FCM-MCOA offers a great convergence. These advantages come from this fact that the cuckoo search process solves the local optimum points problem and converges rapidly to its global solution.

Table 2: The parameters of the FCM-MCOA

Parameter	Values
population Initial	10
Minimum number of eggs	5
Maximum number of eggs	10
Maximum number of iterations	100
Number of cluster centers reported by FCM	3
Lambda variable	2
N max	80
ELR	5
Population variance for process termination	10-15



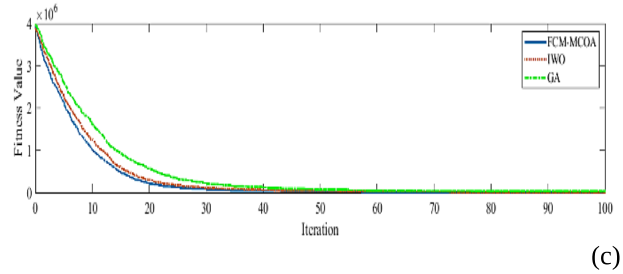


Figure 3 a) MOSHOOF feature, b) Clustering results, c) The cost functions of the FCM-MCOA, IWO and GA after 100 iterations.

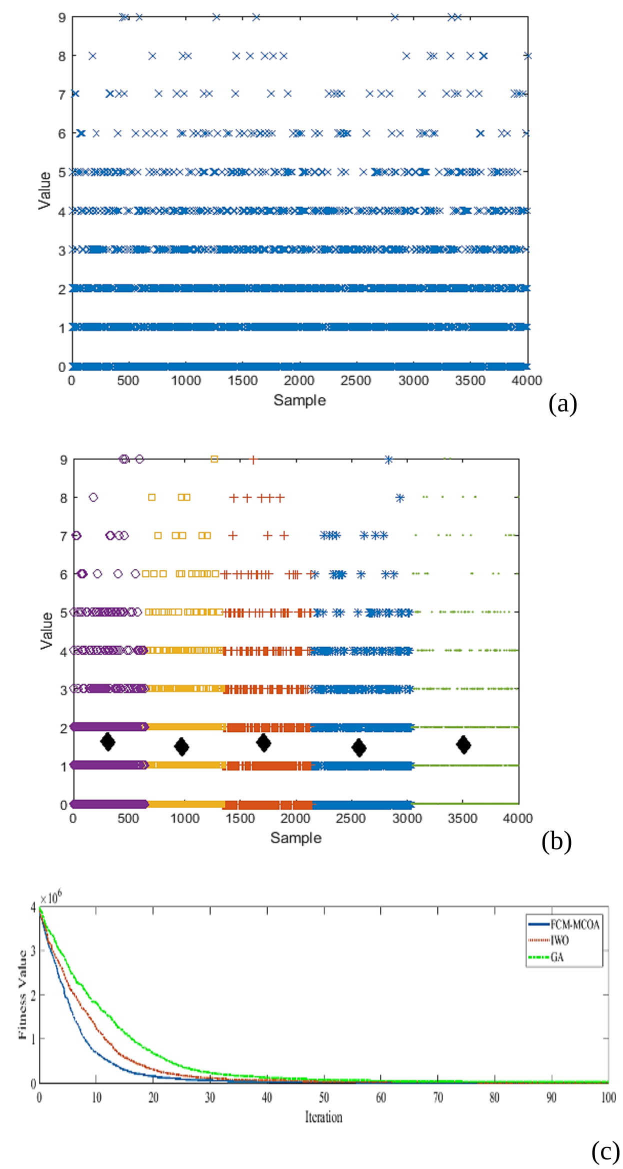


Figure 4 a) MOPLHOOG feature, b) Clustering results, c) The cost functions of the FCM-MCOA, IWO and GA after 100 iterations.

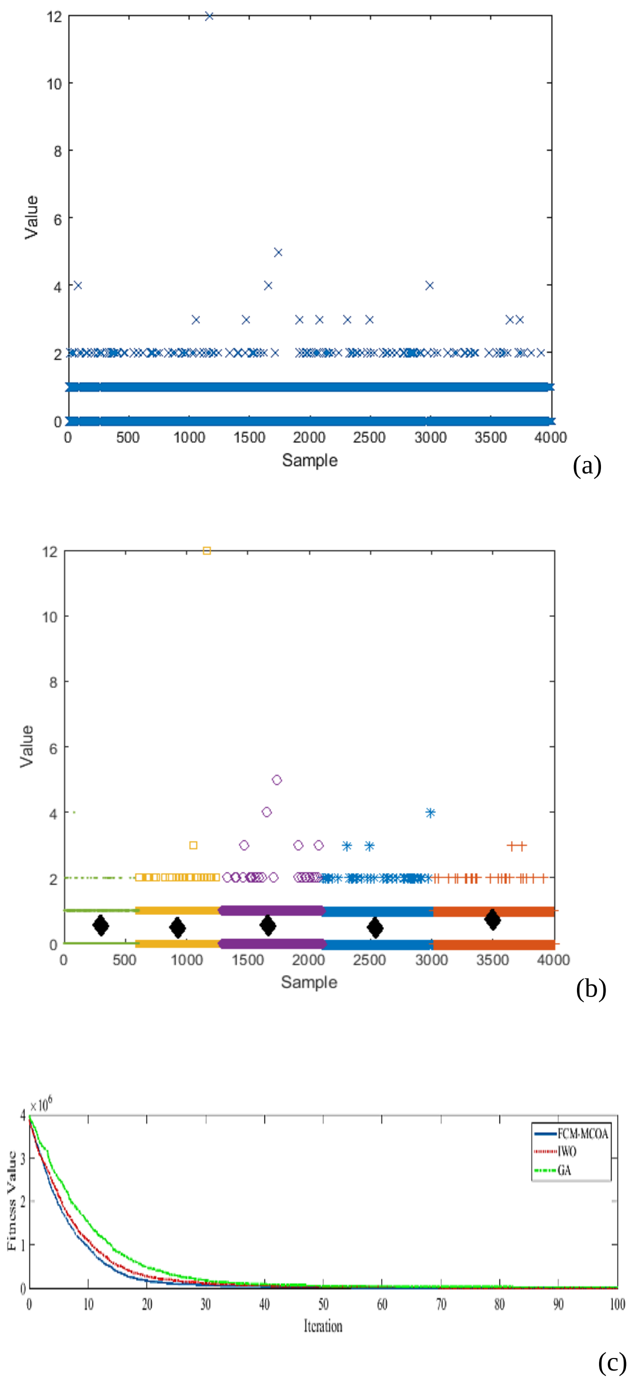


Figure 5 a) APERSAUT feature, b) Clustering results, c) The cost functions of the FCM-MCOA, IWO and GA after 100 iterations.

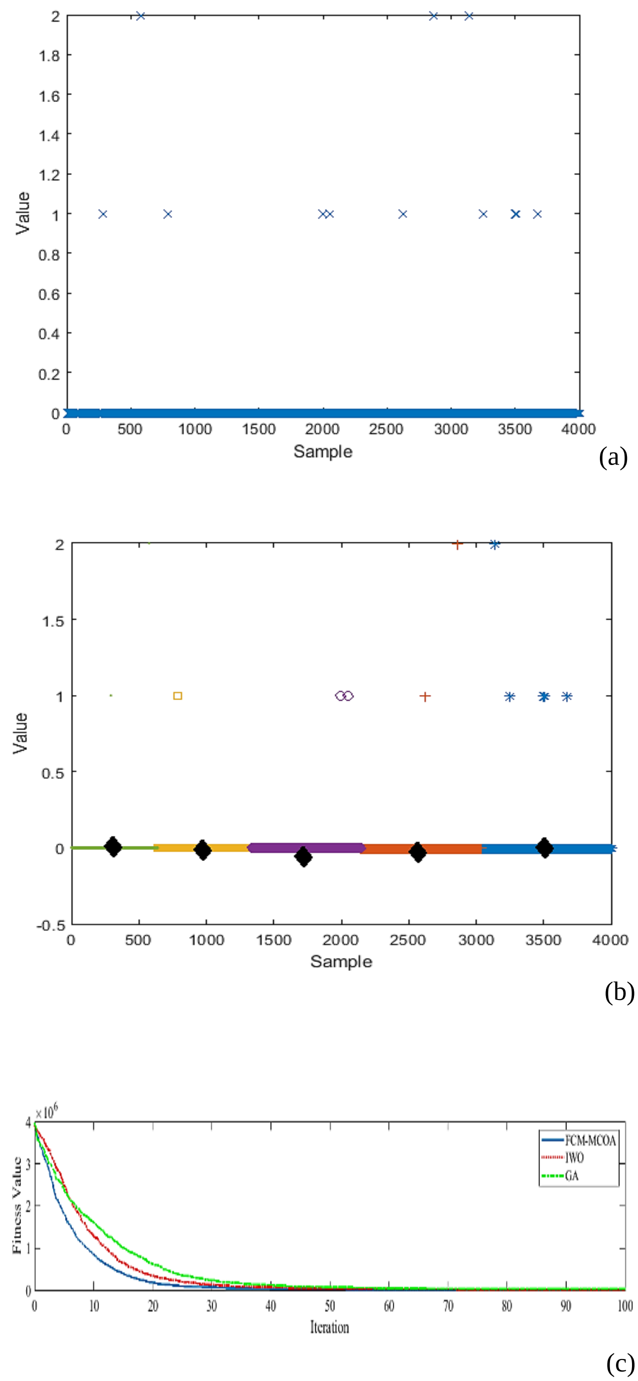


Figure 6 a) APLEZIER feature, b) Clustering results, c) The cost functions of the FCM-MCOA, IWO and GA after 100 iterations.

In Table 3, The values of the cluster centers obtained from the FCM-MCOA approach are reported. The final cost function values, for the FCM-MCOA, IWO [26], and GA [27] methods are also presented in Table 4 . In all scenarios, FCM-MCOA not only has faster convergence but also its final cost function values are considerably less than the above-mentioned conventional clustering methods.

Table 3: The cluster centers for some features.

Feature	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
MOSHOOF	(293.4, 5.6)	(920.5, 6)	(1653.4, 5.7)	(2522.6, 5.8)	(3493.4, 5.8)
MOPLHOOG	(3493.4, 5.8)	(3493.4, 5.8)	(3493.4, 5.8)	(3493.4, 5.8)	(3493.4, 5.8)
APERSAUT	(295.6,0.55)	(928.4,0.49)	(1667.6,0.53)	(2538.4,0.48)	(3499.6,0.71)
APLEZIER	(311.4,0.0046)	(970.6,-1.018)	(1715.4,-.062)	(2568.5,-.027)	(3511.5,0.0019)

Table 4: The performance of the FCM-MCOA, IWO, and GA.

Feature	Final cost function value of FCM-MCOA	Final cost function value of IWO	Final cost function value of GA
MOSHOOF	2635	10917	31365
MOPLHOOG	1798	10712	31072
APERSAUT	2428	9722	21761
APLEZIER	1783	26923	11015

4-3- Classification Results

We implement the suggested CNN-based classifier to find out who is interested in purchasing caravan policy. Pre-defined functions and objects of the MATLAB Neural Network Toolbox are employed to create the classifiers and define training options, such as learning rate, number of convolutional filters, and the number of neurons in the FC layers. To analyze the performance of the CNN-based classifier, the specificity and sensitivity values are computed. Sensitivity and specificity are used as the criteria that show the accuracy of the two classes, i.e., zero and one class, respectively. Due to imbalance data, a classifier with large sensitivity and specificity values achieves an appropriate score. In order to explore the effect of increasing the number of out-put neurons in the FC layers on the performance of the CNN-based classifier, we changed the number of neurons, except for the last FC layer. The obtained results are reported in Figure 7. The results show that the maximum accuracy for this dataset is 98.1%. The behavior of CNN-based classifier is compared with some popualr classifiers, such as the Naive Bayes (NB) and SVM classifiers [25].

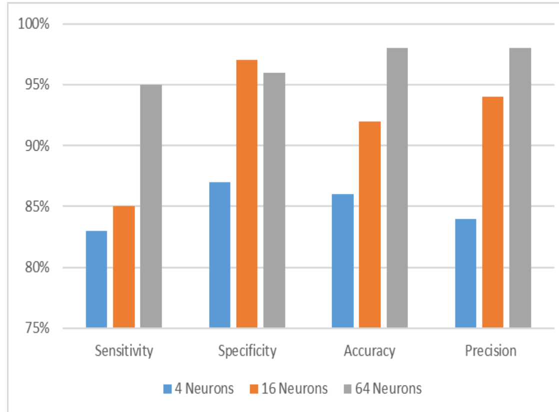


Figure 7 The effect of increasing the number of out-put neurons in the FC layers on the performance of CNN-based classifier.

Figure 8 provides a comparison between the accuracy rate of the CNN-based classifier and some conventional classifiers. The chart show that the proposed method leads to a better performance compared to the conventional methods in terms of the accuracy rate.

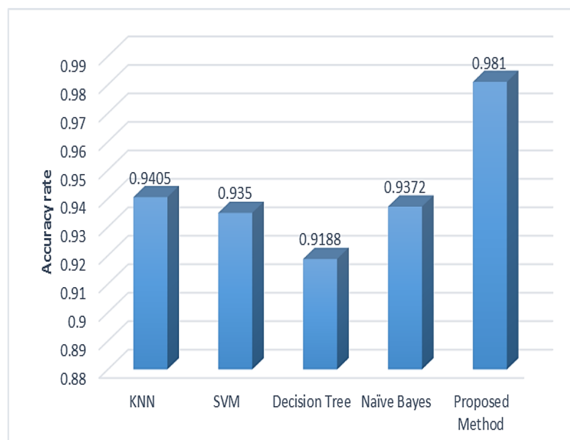


Figure 8 The accuracy rate of the CNN-based classifier compared to some conventional classifiers.

5- Conclusions

This study brings forward a hybrid FCM-MCOA approach in which the grouping task of cuckoos was performed by the FCM technique. A CNN-based classifier is also adopted for customer classification task, and the effect of the out-put neurons in the FC layers is evaluated. To investigate the effectiveness of our methods, we did some simulations, and the results showed that the FCM-MCOA scheme converges faster than conventional clustering methods. The results also showed that the CNN-based classifier is able to predict who wants to use the caravan insurance policy, providing an accuracy above 98%. It can be concluded that the CNN-based classifier outperforms the conventional classifiers, such as NB and SVM, in terms of the accuracy rate and convergence speed.

References

[1] A. Voulodimos, N. Doulamis, A. Doulami, and E. Protopapadakis, “Deep learning for computer vision: A brief review”, Computational intelligence and neuroscience, 2018.

[2] M. Jahangiri, and S. Ghavami, “Hybrid fuzzy c-means clustering algorithm and multilayer perceptron for increasing the estimate accuracy of the geochemical element concentration case study: eastern zone of porphyry copper deposit of Sonajil”, Iranian Journal of Geology, Vol. 48, No. 48, pp. 0, 2019.

[3] M. K. Pakhira, “A fast k-means algorithm using cluster shifting to produce compact and separate clusters”, Int J Eng, Vol. 28, No. 1, pp. 35-43, 2015.

[4] M. Setnes, and U. Kaymak, “Extended fuzzy c-means with volume prototypes and cluster merging”, In Proceedings of the 6th European Conference on Intelligent Techniques and Soft Computing (EUFIT’98), 1998, pp. 1360-1364.

[5] G. S. Budhi, R. Chiong, and Z. Wang, “Resampling imbalanced data to detect fake reviews using machine learning classifiers and textual-based features”, Multimedia Tools and Applications, Vol. 80, No. 9, pp. 13079-13097, 2021.

[6] S. Cateni, V. Colla, A.Vignali, and M.Vannucci, “Data Pre-processing for Efficient Design of Machine Learning-Based Models to be Applied in the Steel Sector”, In Impact and Opportunities of Artificial Intelligence Techniques in the Steel Industry: Ongoing Applications, Perspectives and Future Trends, pp. 13-27, 2021.

[7] Z. Abtahi, R. Sahraeian, and D. Rahmani, “A Stochastic Model for Prioritized Outpatient Scheduling in a Radiology Center”, International Journal of Engineering Transactions A: Basics, Vol. 33, No. 4, 2020.

[8] J. MacQueen, “Classification and analysis of multivariate observations”, In 5th Berkeley Symp. Math. Statist. Probability, pp. 281-297, 1967.

[9] T. Abukhalil, M. Patil, and T. Sobh, “A comprehensive survey on decentralized modular swarm robotic systems and deployment environments”, International Journal of Engineering (IJE), Vol. 7, No. 2, pp. 44, 2013.

[10] C. Li, L. Liu, X. Sun, J. Zhao, and J. Yin,” Image segmentation based on fuzzy clustering with cellular automata and features weighting”, EURASIP Journal on Image and Video Processing, Nom. 1, pp. 1-11, 2019.

[11] T. M. Silva filho, B. A. Pimentel, R. M. Souza, and A. L. Oliveira, “Hybrid methods for fuzzy clustering based on fuzzy c-means and improved particle swarm optimization”, Expert Systems with Applications, Vol. 42, No. 17-18, pp. 6315-6328, 2015.

[12] S. Das, A. Abraham, and A. Konar, “Automatic clustering using an improved differential evolution algorithm”, IEEE Transactions on systems, man, and cybernetics-part A: Systems and Humans, Vol. 38, No. 1, pp. 218-237, 2007.

[13] S. Paterlini and T. Krink, “Differential evolution and particle swarm optimization in partitional clustering”, Computational statistics & data analysis, Vol. 50, No. 5, pp. 1220-1247, 2006.

[14] T. Niknam, M. Nayeripour, and B. B. Firouzi, “Application of a new hybrid optimization algorithm on cluster analysis”, In Proceedings of world academy of science, engineering and technology, Vol. 36, pp. 599, 2008.

[15] K. Krishna, and M. N. Murty, “Genetic K-means algorithm”, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), Vol. 29, No. 3, pp. 433-439, 1999.

[16] H. Izakian, A. Abraham, and V. Snášel, “Fuzzy clustering using hybrid fuzzy c-means and fuzzy particle swarm optimization”, In 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), pp. 1690-1694, 2009.

[17] X. S. Yang, and S. Deb, “Cuckoo search: recent advances and applications”, Neural Computing and applications, Vol. 24, No. 1, pp. 169-174, 2014.

[18] X. S.Yang, and S. Deb, “Engineering optimisation by cuckoo search” , International Journal of Mathematical Modelling and Numerical Optimisation, Vol. 1, No. 4, pp. 330-343, 2010.

[19] R. Rajabioun, “Cuckoo optimization algorithm”, Applied soft computing, Vol. 11, No. 8, pp. 5508-5518, 2011.

[20] H. Kahramanli, “A modified cuckoo optimization algorithm for engineering optimization”, International Journal of Future Computer and Communication, Vol. 1, No. 2, pp. 199, 2012.

[21] M. Momeny, M. Agha Sarram, A.M. Latif, and R. Sheikhpour, “Improving the Architecture of Convolutional Neural Network for Classification of Images Corrupted by Impulse Noise”, Nashriyyah-i Muhandisi-i Barq va Muhandisi-i Kampyutar-i Iran, Vol. 76, No. 4, pp. 267, 2020.

[22] M. Rohanian, M. Salehi, A. Darzi, and V. Ranjbar, “Convolutional Neural Networks for Sentiment Analysis in Persian Social Media”, arXiv preprint arXiv:2002.06233, 2020.

[23] M. Mobini, G. Kaddoum, and M. Herceg, “Design of a SIMO deep learning-based chaos shift keying (DLCSK) communication system”, Sensors, Vol. 22, No. 1, pp. 333, 2022.

[24] D. Madurasinghe, and G. K. Venayagamoorthy, “LVQ neural network for online identification of power system network branch events”, In 2020 Clemson University Power Systems Conference, 2020, pp. 1-7.

[25] B. Zadrozny, and C. Elkan, “Transforming classifier scores into accurate multiclass probability estimates”, In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, 2002, pp. 694-699.

[26] A. R. Mehrabian, and C. Lucas, “A novel numerical optimization algorithm inspired from weed colonization”, Ecological informatics, Vol. 1, No. 4, pp. 355-366, 2006.

[27] K. J. Kim, and H. Ahn, “Using a clustering genetic algorithm to support customer segmentation for personalized recommender systems”, In International Conference on AI, Simulation, and Planning in High Autonomy Systems, 2004, pp. 409-415.

[28] C. Mouton, J. C. Myburgh, and M. H. Davel, “Stride and translation invariance in CNNs.”, In Southern African Conference for Artificial Intelligence Research, 2021, pp. 267-281.