IJCTA, 8(4), 2015, pp. 1301-1307

A Process-Mining Framework for Fraud and Abuse Detection with Combination of PSO and Fuzzy Clustering

Mohaddeseh Bayrami Zonoz[1], *Sohrab Khanmohammadi[2]

Abstract: There is no globally accepted definition of cheating. Wang et al., have defined cheating as,"A targeted attempt to gain unauthorized financial benefit which is against laws, rules and politics." Cheating is a kind of lying, fraud and trickery that gives an unfair advantage to cheater over others. It is a form of violation of law. The word "Cheating" is rarely used in law, and "fraud" and "corruption" are mentioned instead. Fraud includes a wide variety of approaches, incorrect data entry and illegal actions which can bring about very adverse financial and psychological effects on an organization. Even accusation of fraud is causes substantial damage to reputation and relationships between customer and employee. Considering these issues, it is important for any organization to have a fraud management system. Various technical methods have been designed and presented to prevent fraud among which the best presented method is to use process mining technique. Due to the fact that most evolutionary algorithms are derived from nature and natural motion of particles, and considering the fact that in real world, environment is constantly changing, there is a need for a method to implement these evolutionary algorithms in dynamic and changing environments. The aim of this paper is to achieve the optimal solution in fraud detection using the combination of particle swarm optimization and fuzzy clustering. To find a reasonable path in an environment, the use of the particlesswarm motion can be very effective. PSO algorithm is one of the algorithms that have been inspired by this natural fact.

***Keywords*:** PSO, fuzzy clustering, fraud

1. INTRODUCTION

Fraud is a kind of deception for personal or financial benefits. Generally, there are two types of fraud: External and internal fraud. External fraud is done by someone who is outside the organization and internal fraud is done by someone inside the organization. Internal fraud is similar to occupational fraud. Using job for personal enrichment by deliberate misuse or improper usage of the resources and assets of organization is called "Professional Fraud". The concept of fraud have various forms in which there are three main categories: asset misappropriation, corruption and fraud in financial statements [2]. These categories include several subcategories and the biggest imagine of fraud is doing it with undetectable and unpunished methods. More bigger and undetectable fraud paves the way for individual and organization for doing it and due to expensive costs associated with fraud, its detection and prevention is crucial. In 49% of organizations, any losses which stem from fraud could not be recovered. There are different ways for discovering fraud but the proposed method is the combination of PSO and fuzzy clustering. Clustering is an unsupervised learning task that partitions data objects into a certain number of clusters in a way that data in the same cluster should be similar to each other while data in different clusters should be dissimilar [1]. By this definition, clustering can be very useful in remote sensing data analysis because it can reveal useful information concerning the structure of the dataset. One of the most widely used clustering algorithms is fuzzy clustering algorithm.PSO is a population based on stochastic optimization technique inspired by social behavior of bird flock (and fish school,

[1] Software Engineering, Islamic Azad University of Shabestar Branch; Shabestar; IRAN, *E-mail: Mohaddeseh.bayramiz@gmail.com*

[2] Computer and Electrical Engineering University of Tabriz; Tabriz; IRAN, *E-mail: khan@tabrizu.ac.ir*

* *Corresponding author:* Sohrab Khanmohammadi.Computer and Electrical Engineering Universityof Tabriz; Tabriz; IRAN

etc.), and has been developed by Kennedy and Eberhart [6]. The main goal of this research is to find reasonable pathway by combination of PSO and fuzzy clustering.

2. SUGGESTING METHOD

In this study, for reducing of fraud, for example in healthcare of doctors, nurses, medical staff, labs and etc., the proposed method is that each node should be considered as a process that has some options for selecting[3]. The aim of this study was to find a way with the shortest time and the lowest cost and other paths are non-optimal. For example, for cancer treatment, a doctor can suggested multiple paths which have specific time and expenses. This project can be considered in several ways:

1) The process of detection based on time
2) The process of detection based on expense
3) The process of detection based on time and expense

The main goal of this paper is to detect fraud based on time and expense. Two fundamental parameters were added to the above features which included sidelong imposition and process risk that we traverse operations steps based on these four features. For more details see the graph in Figure 1.1. In this graph, each node represents one process and relationship between each node representing the time of each process to another process. Each node could be added by weight which reflects the expense of each process.

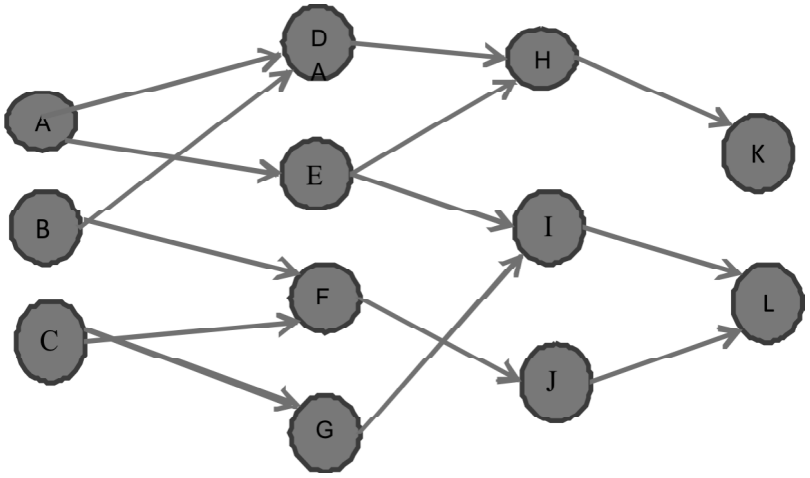

Figure 1.1: A Sample of Operational Graph

Based on Fig 1.1, dependency means that, for example for achieving process of D, the process of A or D should be finished. See the parameters, options and dependencies in Table 1.1.

According to Table 1.1, options is meant to pose this question that, "which of each these processes in regard to time and expense is applicable?" Now, if we want to move from A to K process, which way is reasonable in terms of cost, time and side effects? Any path except this reasonable path is unacceptable. Therefore, in order to discover the fact that whether the fraud is committed in these routs or not, by using of birds and fuzzy clustering, we classified the birds into the categories based on the data values. The purpose of each bird is the available routs. For example, the first path started in A process and ended in K process. The second path or the second bird started in B process and ended in L process and we will move the birds to their best generation. After several generations, the best bird of bad cluster will be replaced by one of the birds of good generation (if it is better) and we do this for other clusters. It should be noted that the best bird in each cluster will move to the best bird of good cluster and the birds of each cluster will move to the best individual of each cluster, the cluster will be changed during the program implementation and it is possible that the bad cluster would be converted to the medium and good cluster. The main

Table 1.1
The Operation Process Based on Four Parameters

Parameters	*Option2*	*Option1*	*Dependency*	*Diagnosis*
Time	2 days	3 days		1
Cost	100000 Rs	20000 Rs	_____	
Side Effects	2	10		
Risk	1	2		
Time		1 day	A	2
Cost		10000 Rs		
Side Effects	_____	3		
Risk		4		
Time	5 days	3 days	A, B	3
Cost	100000 Rs	250000 Rs		
Side Effects	6	5		
Risk	7	6		
Time		4days	D	4
Cost		5000000 Rs		
Side Effects	_____	5		
Risk		5		
Time	5 days	6 days	H	5
Cost	200000 Rs	100000Rs		
Side Effects	5	6		
Risk	9	8		

purpose of this study is to find the reasonable path during the shortest time and the lowest cost by adding two parameters: side effects and process risk. By this new method the fraud amount and misuse will be reduced. The proposed framework demonstrate the efficiency and accuracy.

THE STEPS OF THE OPERATION

The FCM algorithm tends to converge faster than PSO algorithm because it requires fewer function evaluations. But, it usually gets stuck in local optima. We integrate FCM with PSO to form a hybrid clustering algorithm called PSO-FCM which maintains the merits of FCM and PSO. More specifically, PSO-FCM will apply FCM with four iterations to the particles in the swarm every eight generations such that the fitness value of each particle is improved [4]. A particle is a vector of real numbers of dimension $k \_ d$, where k is the number of clusters and d is the dimension of data to be clustered[7].

The objective function of the FCM algorithmdefined in Eq. 1.1 is the fitness functions of the hybrid clustering algorithms:

$$j_m = \sum_{i=1}^{c} \sum_{j=1}^{N} u_{ij}^{m} \left| vi - Xj \right|, m \geq 1 \tag{1.1}$$

The hybrid PSO-FCM algorithm can be summarized as follows [5]:

1. Randomly generation of particles.
2. Calculation of cluster centers using Eq. 1.2.

$$V_j = \frac{\sum_{i=1}^{n} u_{ij}^m x_i}{\sum_{i=1}^{n} u_{ij}^m} \tag{1.2}$$

3. Calculation of fitness function using Eq. 1.1.
4. Updating of Pbest and Gbest according to fitness function of FCM
5. Updating velocity by Eq. 1.3.

$$v_{id}(t+1) = wv_{id}(t) + c_1 r_1(p_{id}(t) - x_{id}(t)) + c_2 r_2 (p_{gd}(t) - x_{id}(t)) \tag{1.3}$$

6. Updating position by Eq. 1.4.

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \tag{1.4}$$

7. Repeat step 2–6 until the stopping criteria is met.

3. THE GRAPH OF DIAGNOSIS AND TREATMENT STAGES OF ESOPHAGEAL CANCER

Based on the research method, the graph of Figure 1.2 has been suggested that the diagnosis and treatment of esophageal cancer, each of the nodes represented the activities like anesthesia and etc.Each activity has some options which are determined based on four parameters.

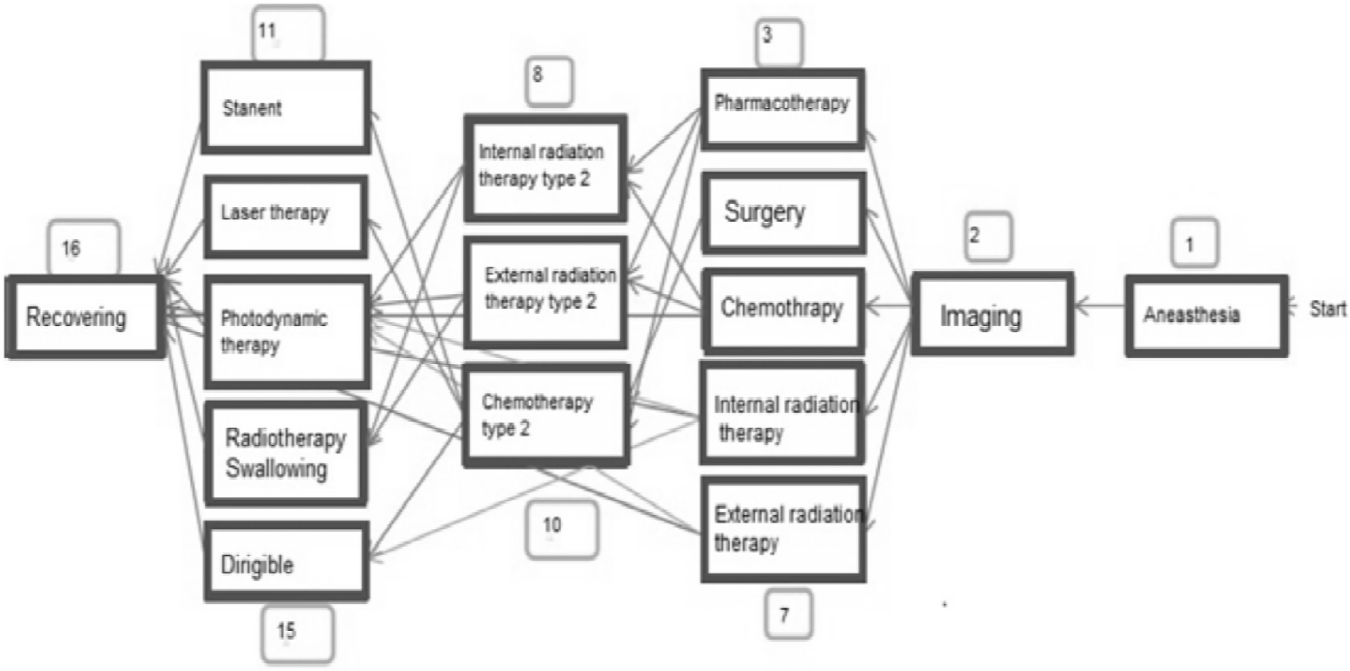

Figure 1.2: Diagnosis graph and esophageal cancer treatment

4. THE FINAL RESULTS BASED ON THE PROPOSED ALGORITHM

Based on the proposed algorithm, the final output for a time, side effects, risks and costs of three repetitive stages was shown in Table 1.2. The results were compared based on the FCM-FPSO, FPSO, FCM in accordance with Table 1.2. According to the results, FCM-FPCO works better than FCM in 80% of cases and FCM works better than FPSO.

Table 1.2
Final output for parameters in 3 repetitions

Parameters	*Results based on the FCM-FPSO*			*Results based on the FCM*		
The number of repetition	*The first iteration*	*The second iteration*	*The third iteration*	*The first iteration*	*The second iteration*	*The third iteration*
Time	110	69	69	110	106	106
The side effect	3.8	4	3.8	3.5	3.4	3.4
Risk	3.7	3.8	3.7	3.2	3	3
Cost	334450000	28740000	26250000	46640000	46750000	46750000

Results based on FPSO

Parametrs	*The first iteration*	*The second iterarion*	*The third iteration*
Time	133	70	133
The side effect	4	4	4.1
Risk	3.9	4	3
Cost	41152000	29592000	42172000

5. ASSESSMENT CHARTS

The charts related to the best population, the average population including (best, average and worst value of the cost), intra-cluster distance, the choice of the optimal cluster, the determination of the centers of clusters based on source program and data are listed below. These charts are related to the implementation of FCM-FPSO in the first repetition.

Totally, there are 45 charts for 3 repetitions but only those charts which are related to the first repetition are in this part. Chart 1.3, 1.4, 1.5 is related to the anesthesia process in the first repetition which evaluate the best population, the average population, the cost function, intra-cluster distance and the choice of optimum cluster.

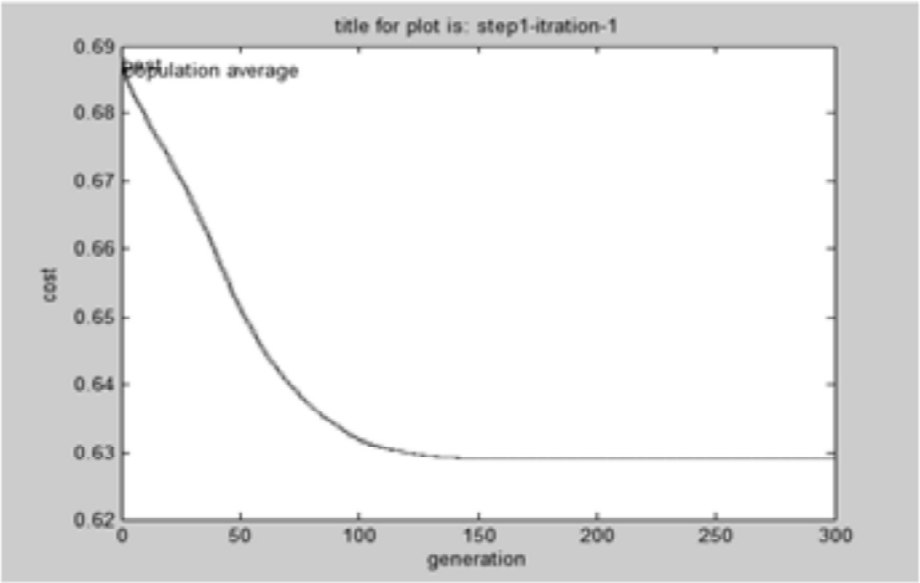

Figure 1.3: The best population and the average population in anesthesia process

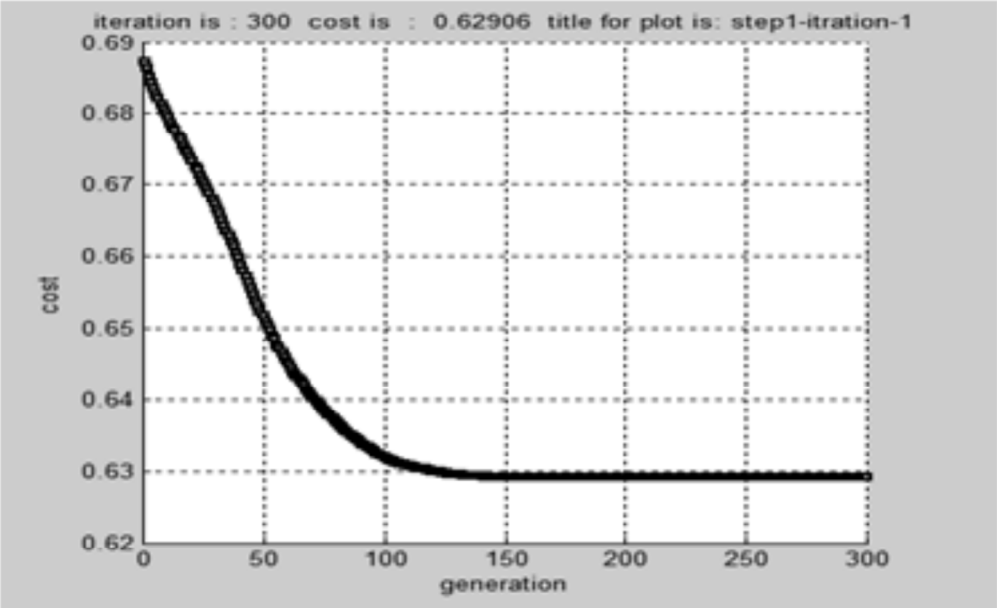

Figure 1.4: inter- cluster distance in anesthesia process

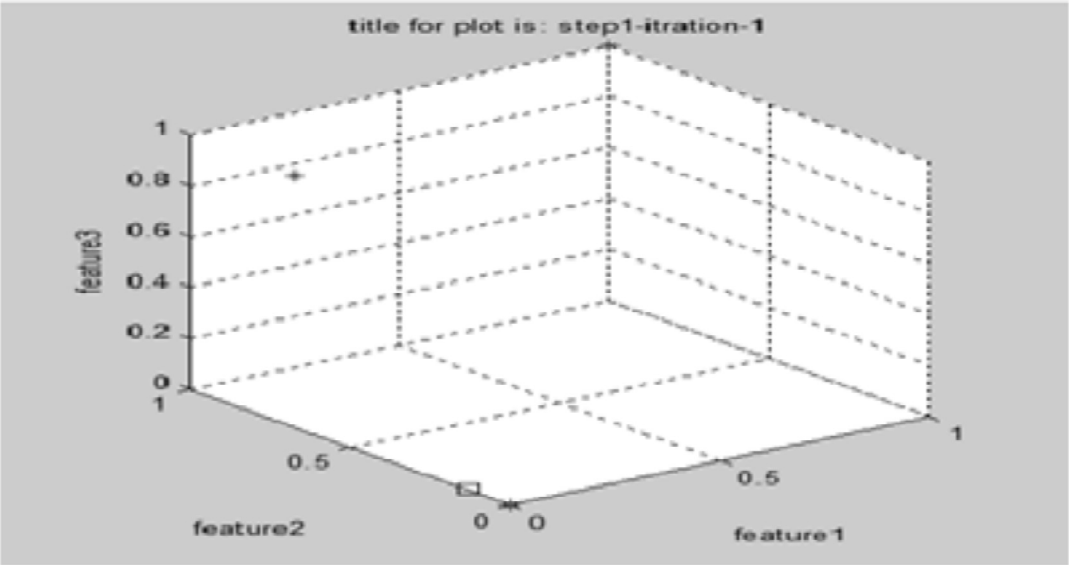

Figure 1.5: Determining the best cluster in anesthesia process and cluster centers

6. CONCLUSIONS

The main purpose is to find the best way and this can be achieved by using process graph and attached data given to the options. For example, the best way achieved by the proposed method on the basis of Figure 1.1 includes:

1. PET Scan Anesthesia, second option 2. X-ray scan, first option, 3. Internal radiation therapy –type one, third option, 4. Internal radiation therapy –type two, second option 5. Photodynamic therapy, first option. This way is achieved by cost, time, risk and side effect and it is an acceptable way. Other ways which are not the proposed algorithm output cannot be regarded as an acceptable way due to its high amount of its parameters and it is a fraud whether it is recommend by doctors or clinic personnel.

In each stage, a new acceptable way may be achieved by local and general optimums. In a comparison between the ways, those ways which have fewer amounts on the basis of patient priority are more acceptable and doctors and clinic personnel should recommend that in order to decrease the fraud.

It may be possible that all four parameters are not always optimal. In this case, the way is chosen based on the parameter which has the most importance in optimality. At each stage of the route of treatment that the patient cures, there is no need for the rest of the way. The method must also be approved by a specialist. On the basis of the analysis, FCM-FPSO algorithm have better results compared to FCM especially in cost and time and FCM have better results compared to FCM in clustering and FCM-FPSO is better in finding an acceptable way.

References

[1] WunschD,Xu R, (2008), Kernel-based clustering. In clustering, 1st ed. IEEE,New Jersey,pp 163-178.

[2] A.C., (2011), ²Fraud Detection Using Analytics in the Healthcare Industry, Anti-fraud materials including reports, case studies and on –demand webinars²

[3] Abraham A., Izakian. H., (2011), Fuzzy C-means and swarm for fuzzy clustering problem pp.1835-1838.

[4]. Yang F., Sun T., Zhang C. (2009), An efficient hybrid data clustering method based onK-harmonic means and particle swarm optimization. Expert Syst Appl 36, 9847–9852

[5] Li W., Yushu L., Xinxin Z., Yuanqing X., (2006), Particle swarm optimization for fuzzy c-meansclustering. Paper presented at the intelligent control and automation. WCICA 2006.The Sixth World Congress on Intelligent Control and Automation.

[6] Kennedy J., Eberhart R. (1995), Particle swarm optimization. Paper presented at the IEEE International Conference Neural Network (ICNN), Perth, WA

[7] Izakian H, Abraham A., Snasel V. (2009), Fuzzy clustering using hybrid fuzzy c-means and fuzzy particle swarm optimization. Paper presented at the World Congress on Nature &Biologically Inspired Computing, Coimbatore, India, 9–11 December.