Volume 3, No. 1, Jan-Feb 2012
ISSN No. 0976-5697
International Journal of Advanced Research in Computer Science
RESEARCH PAPER
Available Online at www.ijarcs.info

An Outlier Detection Method Based On Artificial Bee Colony Fuzzy Clustering

Poonam Chalotra *, Maitreyee Dutta
Deptt. Computer Science,
NITTTR Chandigarh,India
poonamchalotra@gmail.com*, d_maitreyee@yahoo.co.in

Abstract— There is a need for pre-processing of the raw data in many fields, such as data mining, information retrieval, machine learning and pattern recognition. Data Mining or Knowledge discovery refers to a variety of techniques that have developed in the fields of databases, machine learning and pattern recognition. Data pre-processing involves many tasks including detecting outliers, recovering incomplete data and correcting errors. Outlier detection is an important pre-processing task. Outlier detection is a task that finds objects that are dissimilar or inconsistent with respect to the remaining data. Outlier detection can be done using clustering methods. In this paper, an efficient outlier detection method has been proposed which is based on Fuzzy clustering using Artificial Bee Colony algorithm. The Fuzzy clustering based on Artificial Bee Colony algorithm is performed, and small clusters are calculated and considered as outlier clusters. Fuzzy clustering is used to choose the cluster heads and ABC to select the members of the clusters. Test result shows the effective results in finding the outliers on data sets in data mining literature.

Keywords— Artificial Bee Colony algorithm, Clustering, Data mining, Fuzzy C-Means clustering, Outliers.

I. INTRODUCTION

Outlier detection is a research problem in “small-pattern” mining in databases. It aims at finding a specific number of objects that are considerably dissimilar, exceptional and inconsistent with respect to the majority records in an input database. In many data analysis tasks, a large number of variables are being recorded or sampled. One of the first steps towards obtaining a coherent analysis is the detection of outlaying observations. Although outliers are often considered as an error or noise, they may carry important information. There is a need for pre-processing of the raw data in many fields, such as data mining, information retrieval, machine learning and pattern recognition. Zhang et. al [1] argue for the importance of data preprocessing and present the following reasons: (1) real world data is impure; (2) high performance data mining systems require high quality data and (3) quality data yields high quality patterns. Therefore, developing efficient data-preprocessing techniques is a critical task that requires considerable research efforts. Using clustering algorithms for outlier detection is a technique that is frequently used.

The clustering algorithms consider outlier detection only to the point they do not interfere with the clustering process. Several clustering-based outlier detection techniques have been developed, most of which rely on the key assumption that normal objects belong to large and dense clusters, while outliers form very small clusters. The Fuzzy C-Means algorithm (FCM), as one of the best known and the most widely used fuzzy clustering algorithms. However, FCM is an effective algorithm; the random selection in center points makes iterative process falling into the local optimal solution easily. To tackle this problem, evolutionary algorithms such as genetic algorithm (GA), differential evolution (DE), ant colony optimization (ACO), and particle swarm optimization (PSO) have been successfully applied. Recently a family of nature inspired algorithms, known as Swarm Intelligence (SI), has attracted several researchers from the field of pattern recognition and clustering. Inspired from this, in our work Artificial Bee Colony Algorithm is applied to Fuzzy clustering for outlier detection.

II. FUZZY C-MEANS CLUSTERING

Fuzzy c-means clustering involves two processes: the calculation of cluster centers and the assignment of points to these centers using a form of Euclidian distance. This process is repeated until the cluster centers stabilize. The algorithm is similar to k-means clustering in many ways but it assigns a membership value to the data items for the clusters within a range of 0 to 1. So it incorporates fuzzy set’s concepts of partial membership and forms overlapping clusters to support it. The objective function of the fuzzy clustering is to minimize the equation:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2 \qquad (1)$$

Where m is any real number greater than 1, it is set to 2.00 by Bezdek, u_{ij} is the degree of membership of x_i in the cluster j and $||xi - c_j||^2$ is the Euclidean distance from sample points x_i to cluster center c_j.

The algorithm needs a fuzzification parameter m in the range [1,n] which determines the degree of fuzziness in the clusters. When m reaches the value of 1 the algorithm works like a crisp partitioning algorithm and for larger values of m the overlapping of clusters is tend to be more.

The algorithm calculates the membership value µ with the formula,

$$\mu_j(x_i) = \frac{\left(\frac{1}{d_{ji}}\right)^{\frac{1}{m-1}}}{\sum_{k=1}^{p} \left(\frac{1}{d_{ki}}\right)^{\frac{1}{m-1}}} \qquad (2)$$

Where,

$\mu_j(x_i)$: is the membership of x_i in the j^{th} cluster
d_{ji} : is the distance of x_i in cluster C_j
m: is the fuzzification parameter

p: is the number of specified clusters
d_{ki}: is the distance of x_i in cluster C_k

This is a special form of weighted average. We modify the degree of fuzziness in x_i's current membership and multiply this by x_i. The product obtained is divided by the sum of the fuzzified membership. In this way new centroids are calculated with these membership values using equation (2) for clusters.

$$C_j = \frac{\sum_i \left[\mu_j \ \ x_i \ \right]^m x_i}{\sum_i \left[\mu_j \ \ x_i \ \right]^m} \quad (3)$$

Where
C_j : is the center of the j^{th} cluster
x_i : is the i^{th} data point
μ_j : the function which returns the membership
m: is the fuzzification parameter

The algorithm [2] starts with the calculations of membership values for data points in clusters and then recalculates the cluster centers using these membership values. When the cluster center stabilizes (when there is no change) the algorithm ends. The fuzzy c-means approach to clustering suffers from several constrains that affect the performance [4].

$$\sum_{j=1}^{p} \mu_j \ \ x_i \ = 1 \quad (4)$$

The main drawbacks are due to the restriction that the sum of membership values of a data point x_i in all the clusters must be equal to one as in expression (3). This restriction tends to give high membership values for the outlier points. So the algorithm has difficulty in handling outlier points.

III. ARTIFICIAL BEE COLONY FUZZY CLUSTERING

Artificial Bee Colony (ABC) algorithm is a new swarm intelligence method which simulates intelligent foraging behavior of honey bees. In the model of ABC algorithm, there are three groups of bees; employed bees, onlooker bees and scout bees in the colony of artificial bees [5]. Firstly, half of the colony consists of the employed bees and the second half consist the onlookers. Employed bees go to the food sources, and then they share the nectar and the position information of the food sources with the onlooker bees which are waiting on the dance area determine to choose a food source. The employed bee whose food source has been abandoned by the bees becomes a scout bee that carries out random search in the simulating model. The goal of bees in the ABC model is to find the best solution, the position of a food source represents a possible solution to the optimization problem and the nectar amount of a food source corresponds to the quality (fitness) of the associated solution. The detail pseudo code of ABC [6] fuzzy clustering (ABC-FC) is:

Step 1 Initialize the population of solutions x_{ij} and evaluate the population;
Step 2 Repeat;
Step 3 cycle=1;
Step 4 Produce new solutions (food source positions) v_{ij} in the neighborhood of x_{ij} for the employed bees using the formula

$$v_{ij} = x_{ij} + \Phi_{ij} (x_{ij} - x_{kj}) \quad (5)$$

Here k is a solution in the neighborhood of i, Φ is a random number in the range [-1, 1]. Evaluate the new solutions;
Step 5 Apply the greedy selection process between x_i and v_i;
Step 6 Calculate the probability values P_i for the solutions x_i by means of their fitness values using the equation:

$$P_i = \frac{f_i}{\sum_{i=1}^{SN} f_i} \quad (6)$$

Here SN denotes the number of solutions, and f denotes the fitness value;
Step 7 Normalize P_i values into [0, 1];
Step 8 Produce the new solutions (new positions) v_i for the onlookers from the solutions x_i, selected depending on P_i, and evaluate them;
Step 9 Apply the greedy selection process for the onlookers between x_i and v_i;
Step 10 Determine the abandoned solution (source), if exists, and replace it with a new randomly produced solution xi for the scout using the equation

$$x_{ij} = \min_j + \Phi_{ij} * (\max_j - \min_j) \quad (7)$$

Here Φ_{ij} is a random number in [0, 1];
Step 11 Memorize the best food source position (solution) achieved so far;
Step 12 cycle= cycle+1;
Step 13 cycle= MCN

It is clear from the above explanation that there are three control parameters in the basic ABC: the number of food sources which is equal to the number of employed or onlooker bees SN, the value of limit and the maximum cycle number MCN. The survival and progress of the bee colony are dependent upon the rapid discovery and efficient utilization of the best food resources. Similarly the successful solution of difficult engineering problems is connected to the relatively fast discovery of good solutions especially for the problems that need to be solved in real time. In a robust search process, exploration and exploitation processes must be carried out together. In the ABC algorithm, while onlookers and employed bees carry out the exploitation process in the search space, the scouts control the exploration process.

IV. RELATED WORK

Many approaches have been proposed to detect outliers. These approaches can be classified into four major categories based on the techniques used [7] which are: distribution-based, distance-based, density-based and clustering-based approaches. Distribution-based approaches develop statistical models from the given data and then apply a statistical test to determine if an object belongs to this model or not. Objects that have low probability to belong to the statistical model are declared as outliers. However, Distribution-based approaches cannot be applied in multidimensional scenarios. In the distance-based approach [8], outliers are detected as follows. Given a distance measure on a feature space, a point q in a data set is an outlier with respect to the parameters M and d, if there are less than M points within the distance d from q, where the values of M and d are decided by the user. The problem with this approach is that it is difficult to determine the values of M and d. Density-based approaches [9] compute the density of regions in the data and declare the objects in low dense regions as outliers. In [9], the authors assign an outlier score to any given data point, which is known as the Local Outlier Factor (LOF), depending on its distance from its local neighborhood.

Clustering-based approaches [10, 11] consider clusters of small sizes as clustered outliers. In these approaches, small clusters (i.e., clusters containing significantly less points than other clusters do) are considered outliers. The advantage of the clustering-based approaches is that they do not have to be supervised. Moreover, clustering-based techniques are capable of being used in an incremental mode (i.e., after learning the clusters, new points can be inserted into the system and tested for outliers).

A method is presented in [8] a method based on fuzzy clustering. In order to test the absence or presence of outliers, two hypotheses are used. However, the hypotheses do not account for the possibility of multiple clusters of outliers.

A two-phase method has been defined in [10] to detect outliers. In the first phase, the authors proposed a modified k-means algorithm to cluster the data, and then, in the second phase, an Outlier-Finding Process (OFP) is proposed. The small clusters are selected and regarded as outliers, using minimum spanning trees. In [14], clustering methods have been applied. The key idea is to use the size of the resulting clusters as indicators of the presence of outliers. The authors use a hierarchical clustering technique.

The PAM algorithm [13] is performed and followed by the Separation Technique (henceforth, the method will be termed PAMST). The separation of a cluster A is defined as the smallest dissimilarity between two objects; one belongs to Cluster A and the other does not. If the separation is large enough, then all objects that belong to that cluster are considered outliers. In order to detect the clustered outliers, one must vary the number k of clusters until obtaining clusters of a small size with a large separation from other clusters.

As mentioned in [13], the K-means is sensitive to outliers, and hence may not give accurate results. In [12], Al- Zoubi proposed an effective clustering-based method to detect outliers. First, the PAM algorithm is performed, producing a set of clusters and a set of medoids. To detect the outliers, the Absolute Distances between the Medoids, μ, of the current cluster and each one of the Points, pi, in the same cluster (i. e., $| pi - \mu |$) are computed. The produced value is termed (*ADMP*). If the *ADMP* value is greater than a calculated threshold, *T*, then the point is considered an outlier; otherwise, it is not. The value of T is calculated as the average of all *ADMP* values of the same cluster multiplied by (1.5).

V. PROPOSED OUTLIER DETECTION METHOD

A new clustering-based approach for outlier detection is proposed. First, we execute the ABC-FCM algorithm, producing an objective function. Small clusters are then determined and considered as outlier clusters. We follow [3] to define small clusters. A small cluster is defined as a cluster with fewer points than half the average number of points in the k clusters. To detect the outliers in the rest of clusters, we (temporarily) remove a point from the data set and re-execute the ABC-FCM algorithm. If the removal of the point causes a noticeable decrease in the objective function value, the point is considered an outlier; otherwise, it is not. The objective function represents the (Euclidean) sum of squared distances between the cluster centers and the points belonging to these clusters multiplied by the membership values of each cluster produced by the ABC-FCM.

The idea is based on the objective function calculated by performing Fuzzy clustering [16] using Artificial Bee Colony method. The Objective Function (OF) produced by the ABC-FCM algorithm represents the distances between the cluster centers and the points belonging to these clusters. Removing a point from the data set will cause a decrease in the OF value because of the total sum of distances between each point and the cluster center belonging to it. If this decrease is greater than a certain threshold, the point is then considered to be an outlier. The following parameters are used:

OF: the objective function of the whole cluster, OF_i: the objective function after removing point P_i from the cluster, DOF_i: $OF\text{-}OF_i$ and T= 1.5.

The basic structure of the proposed algorithm is as follows:

Step 1 Execute the ABC-FCM algorithm to produce an Objective Function (OF)

Step 2 Determine small clusters and consider the points that belong to these clusters as outliers.

Step 3 For the rest of the points:

```
Start
SUM = 0
For each point pi in the set
DO
        remove pi,from the set
        calculate OFi
        calculate DOFi
        SUM = SUM + DOFi
        return pi back to the set;
EndDO
AvgDOF = SUM / n
For each point, pi
DO
        if (DOFi > T(AvgDOF) then classify pi as
        an outlier
End DO
End
```

The following parameters are used for clustering using proposed method:

Table: 1 Parameters used in Clustering

Algorithm	Parameters	Values
Artificial Bee Colony algorithm	Colony size Maximum Cycle number Limit value	50 1000 1000

VI. RESULTS AND DISCUSSIONS

In this work, we implemented the proposed approach on data set1 which is the well known Iris data set [15]. The Iris data set has three classes of 50 instances each: Iris-setosa, Iris-versicolor and Iris-viginica, where each class refers to a type of iris plant. One class is linearly separable from the other two, the latter are not linearly separable from each other. The data set contain four attribute which are: Sepal length in cm, Sepal width in cm, Petal length in cm, Petal width in cm. The first step of our proposed approach i.e. ABC-FCM is implemented in MATLAB. The result after implementation of ABC-FCM is given below when cluster=3.

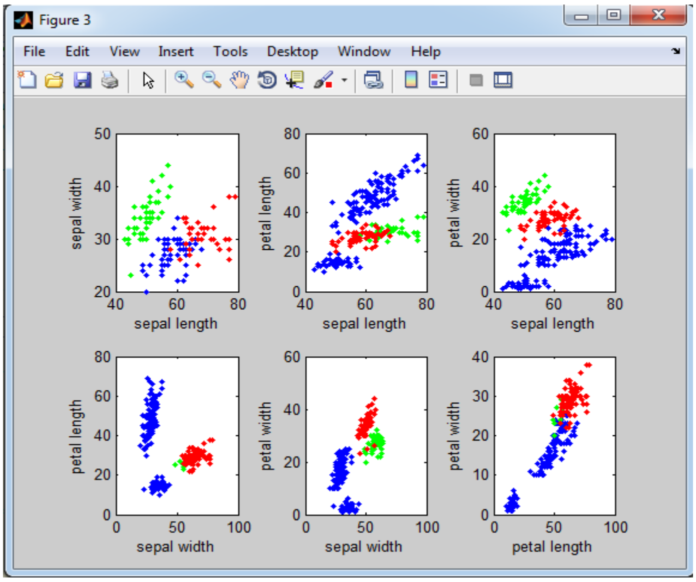

Figure 6.1: Clusters of Iris data when ABC-FCM implemented

It is known that the Iris data have some outliers. When defined number of clusters is three, then step 3 will classify the some points as outliers. After the forming clusters and calculating cluster centers, the step 2 of the proposed method is performed and calculated clusters are considered as outlier clusters. When defined number of clusters is three, then step 3 will classify the some points as outliers. Table II shows the result for the detected outliers in Class 1 of the Iris data set. The first column shows the data point no. and second column shows the OFi values when the current point is removed from the set. The third column shows the DOFi values. The value of the objective function for Iris data set, OF, is (6058.68). The AvgDOF is (28.6284). Multiplying the average values by (1.5) will give us the T(AvgDOF) is (42.9426), which is the threshold value used to determine the outliers. The fourth column shows whether the point is detected as outlier or not. Table III shows the result for the detected outliers in Class 2 of the Iris data set. The average value AvgDOF is (42.96) and T(AvgDOF) which is (64.44). Table IV shows the result for detected outliers in Class 3 of the Iris data set. The AvgDOF is (53.441) and T(AvgDOF) is (80.1615), which is the threshold value used to determine the outliers. By using proposed method eleven outliers detected in class 1, seven outliers in class 2 and nine outliers have detected in class 3 of the Iris data set.

T: Detected, F: Not Detected

Table: 2 The Data Points Detected by our Proposed Method from Class 1 of IRIS Data Set

Data Point No.	OFi	DOFi	Point Detected	Data Point No.	OFi	DOFi	Point Detected
1	6056.34	2.34	F	26	6039.98	18.7	F
2	6039.53	19.15	F	27	6055.08	3.6	F
3	6041.2	17.48	F	28	6053.71	4.97	F
4	6032.49	26.19	F	29	6053.77	4.91	F
5	6054.18	4.5	F	30	6043.18	15.5	F
6	6014.71	43.97	T	31	6042.99	15.69	F
7	6041.43	17.25	F	32	6040.9	17.78	F
8	6058.34	0.34	F	33	6007.71	50.97	T
9	5998.1	60.58	T	34	5978.32	80.36	F
10	6045.29	13.39	F	35	6047.34	11.34	F
11	6034.78	23.9	F	36	6045.73	12.95	F
12	6052.76	5.92	F	37	6030.18	28.5	F
13	6034.53	24.15	F	38	6051.01	7.67	F
14	5979.98	78.7	T	39	6003.82	54.86	T
15	5962.82	95.86	T	40	6057.4	1.28	F
16	5932.54	126.14	T	41	6054.32	4.36	F
17	6015.43	43.25	T	42	5926.55	132.13	T
18	6056.07	2.61	F	43	6015.24	43.44	T
19	5995.87	62.81	T	44	6044.57	14.11	F
20	6042.61	16.07	F	45	6024.61	34.07	F
21	6038.21	20.47	F	46	6036.62	22.06	F
22	6047.34	11.34	F	47	6041.24	17.44	F
23	6016.28	42.4	F	48	6036.9	21.78	F
24	6045.66	13.02	F	49	6041.45	17.23	F
25	6037.11	21.57	F	50	6056.36	2.32	F

Table: 3 The Data Points Detected by our Proposed Method from Class 2 of IRIS Data set

Data Point No.	OFi	DOFi	Point Detected	Data Point No.	OFi	DOFi	Point Detected
101	6004.86	53.82	F	119	5848.8	209.88	T
102	6008.25	50.43	F	120	6004.87	53.81	F
103	6040.97	17.71	F	121	6048.39	10.29	F
104	6031.25	27.43	F	122	6005.6	53.08	F
105	6046.51	12.17	F	123	5894.45	164.23	T
106	5924.42	134.26	T	124	6020.94	37.74	F
107	5969.02	89.66	T	125	6051.42	7.26	F
108	5987.87	70.81	F	126	6021.36	37.32	F
109	6023.25	35.43	F	127	6025.86	32.82	F
110	5959.39	99.29	T	128	6022.24	36.44	F
111	6027.02	31.66	F	129	6039.08	19.6	F
112	6026.82	31.86	F	130	6020.74	37.94	F
113	6055.9	2.78	F	131	5956.7	101.98	T
114	6000.18	58.5	F	132	5871.35	187.33	T
115	5986.17	72.51	F	133	6037.37	21.31	F
116	6028.64	30.04	F	134	6019.05	39.63	F
117	6043.06	15.62	F	135	5995.21	63.47	F
118	5858.26	200.42	T	136	5960.58	98.1	T

Table: 4 Data Points Detected by our Proposed Method from Class 3 of IRIS Data Set

Data Point No.	OFi	DOFi	Point Detected	Data Point No.	OFi	DOFi	Point Detected
51	5987.95	70.73	T	85	6029.03	29.65	F
52	6021.05	37.63	F	86	6019.67	39.01	F
53	6004.91	53.77	F	87	6008.3	50.38	F
54	6013.3	45.38	F	88	6024.38	34.3	F
55	6024.48	34.2	F	89	6037.61	21.07	F
56	6052.13	6.55	F	90	6024.97	33.71	F
57	6016.74	41.94	F	91	6037.79	20.89	F
58	5917.34	141.34	T	92	6043.67	15.01	F
59	6015.42	43.26	F	93	6038.71	19.97	F
60	5999.47	59.21	F	94	5919.79	138.89	T
61	5911.92	146.76	T	95	6046.31	12.37	F
80	5982.92	75.76	T	96	6043.22	15.46	F
81	5999.69	58.99	F	97	6049.51	9.17	F
82	5986.81	71.87	T	98	6046.1	12.58	F
83	6033.35	25.33	F	99	5920.13	138.55	T
84	6018.73	39.95	F	100	6047.01	11.67	F

Based on the simulation results, Table V summarizes the performance of Artificial Bee Colony Fuzzy Clustering and Fuzzy C-Means Clustering in context with finding the outliers points from the Iris Data set. It can be seen that the proposed method Artificial Bee Colony Fuzzy Clustering outperforms in finding the outliers in class 1 and class 2 of the Iris data set as it has not been detected by Fuzzy C-Means Clustering.

Table V: Analysis of FCM and ABC-FCM

Algorithm	Data		No. of Outlier Detected
Fuzzy C-Means Clustering Based	Iris data set	Class 1	ND
		Class 2	ND
		Class 3	8
Artificial Bee colony Fuzzy Clustering based	Iris data set	Class 1	11
		Class 2	7
		Class 3	9

VII. CONCLUSION

An efficient method for outlier detection is proposed in this paper. The proposed method is based on fuzzy clustering techniques. In this work, Artificial Bee Colony algorithm which is a recently introduced optimization algorithm is used to fuzzy clustering of Iris data. Fuzzy clustering is used to choose the cluster heads and ABC to select the members of the clusters. The ABC-FCM algorithm is first performed, and then small clusters are determined and considered as outlier clusters. Other outliers are then determined based on computing differences between objective functions values when points are temporally removed from the data set. If a noticeable change occurred on the objective function values, the points are considered outliers. Different experimentation have been conducted in [11] and showed that there are 8 outliers in class 3 of the Iris data set. Applying our proposed method, eleven outliers detected in class 1, seven outliers in class 2 and nine outliers have detected in class 3 of the Iris data set as results shown in Table 1, Table 2 and Table 3 respectively. The test results show that the proposed approach gave effective results. However, our proposed method is very time consuming. This is because the ABC-FCM algorithm has to be executed n times, where n is the number of data points in a set. We had applied proposed algorithm for small data set. Further investigation may include the use of the approach method for outlier detection with large data set.

VIII. REFERENCE

[1] Pham, D.T. and Afify, A.A. "Clustering techniques and their applications in engineering", Journal of Mechanical Engineering Science, 2006.

[2] Binu Thomas and Raju G.,"A Novel Fuzzy Clustering Method for Outlier Detection in Data Mining", International Journal of Recent Trends in Engineering, Vol. 1, No. 2, 2009.

[3] Berkhin P., "A Survey of Clustering Data Mining Techniques", http://citeseer.ist.psu.edu/berkhin02survey.Tavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd.

[4] D.T. Pham, S. Otri, A. Afify, M. Mahmuddin and H. Al-Jabbouli, "Data Clustering Using the Bees Algorithm", Proc. 40th CIRP International Manufacturing Systems Seminar 2007.

[5] Karaboga D., Ozturk C., "A Novel Clustering Approach: Artificial Bee Colony Algorithm", Applied Soft Computing pp. 652–657, 2011.

[6] Karaboga D., Ozturk C., "Fuzzy clustering with artificial bee colony algorithm", Scientific Research and Essays Vol. 5(14), pp. 1899-1902, 2010.

[7] Zhang, Q. and I. Couloigner, "A New and Efficient K-Medoid Algorithm for Spatial Clustering", in O. Gervasi et al. (Eds.):

[8] ICCSA 2005, Lecture Notes in Computer Science (LNCS) 3482, pp. 181 – 189, Springer-Verlag, 2005.

[9] Ramaswamy, S., R. Rastogi and K. Shim, "Efficient Algorithm for Mining Outliers from Large Data Sets", Proc. ACM SIGMOD, pp. 427-438, 2000.

[10] Breunig, M., H. Kriegel, R. Ng and J. Sander, "LOF: Identifying Density-Based Local Outliers", In Proceedings of ACM SIGMOD International Conference on Management of Data ACM Press, pp. 93–104, 2000.

[11] Jiang, M., S. Tseng and C. Su, "Two-phase Clustering Process for Outlier Detection", Pattern Recognition Letters, Vol. 22, pp. 691-700, 2001.

[12] Acuna E. and Rodriguez C., "A Meta Analysis Study of Outlier Detection Methods in Classification", Technical paper, Department of Mathematics, University of Puerto Rico at Mayaguez, available at academic. uprm. edu/~eacuna/paperout.pdf, in proceedings IPSI, 2004, Venice.

[13] Al- Zoubi, M. B., "An Effective Clustering-Based Approach for Outlier Detection", European Journal of Scientific Research, Vol. 28, No. 2, pp. 310-316, 2009.

[14] Laan, M., K. Pollard and J. Bryan, "A New Partitioning Around Medoids Algorithms", Journal of Statistical Computation and Simulation, Vol. 73, No. 8, pp. 575-584, 2003.

[15] Loureiro, A., L. Torgo and C. Soares, "Outlier Detection Using Clustering Methods: a Data Cleaning Application", in Proceedings of KDNet Symposium on Knowledge-based Systems for the Public Sector. Bonn, Germany, 2004.

[16] Blake,C.L. and C.J.Merz,"UCI Repository of Machine Learning Databases",

http://www.ics.uci.edu/mlearn/MLRepository, University of California, Irvine, Department of Information and Computer Sciences, 1998.

[17] Moh'd Belal, Al-Zoubi, Ali Al-Dahoud, Abdelfatah A. Yahya. "New Outlier Detection Method Based on Fuzzy Clustering", WSEAS TRANSACTIONS on Information Science and Applications, Volume 7, 2010.

