Journal of Scientific & Industrial Research
Vol. 71, September 2012, pp. 594-600

A study of digital mammograms by using clustering algorithms

S M Halawani[1], M Alhaddad[2] and A Ahmad[1]*
Department of Information Technology, Faculty of Computing and Information Technology, King Abdul Aziz University,
[1]Rabigh, [2]Jeddah, Saudi Arabia

Received 12 April 2012; revised 05 August 2012; accepted 06 August 2012

This study presents clustering algorithms to study digital mammograms. Probabilistic clustering algorithms performed better than hierarchical clustering algorithm. Clustering results are competitive with classification results, indicating that clustering algorithms can be used as an important tool to study digital mammograms. Probabilistic clustering algorithms can also be used by radiologists to improve their prediction accuracy.

Keywords: Classification, Clustering, K-mean clustering, Mammograms, Mixed datasets, Probabilistic

Introduction

Breast cancer is the most common form of cancer amongst women[1]. Early and accurate detection of breast cancer results in long survival of patients[2]. Machine learning techniques are being used to improve diagnostic capability for breast cancer[3-7]. Various classification techniques (decision trees[8], support vector machines[9], fuzzy- genetic algorithm[10-12] etc.) have been used to study breast cancer dataset. Mammography is a screening technique to detect breast cancer at an early stage[13,14]. However, radiologists not always give correct results. Generally, high false positive [benign (B) tumor is predicted as malignant (M) tumor] is related with these prediction. Hence, various computer added diagnostic tools have been suggested to help radiologists. Various data mining techniques (feature selection techniques, classification techniques and clustering techniques) have been used to study digital mammograms[15–17]. A breast cancer dataset[18], which has 4 attributes [3 features related with breast imaging reporting data systems (BI-RADS) and 1 age feature] has been used extensively to study various data mining techniques. A low number gives indication of B whereas, a high number is indicative of M. Various classification algorithms have been used to study this datasets[19,20]. However, only Polat[21] used few clustering techniques to improve classification results.

This study presents digital mammograms collected with various clustering algorithms to understand applicability of these algorithms for breast cancer detection.

Proposed Methodology

Digital mammograms were collected at University of Erlangen-Nuremberg between 2003 and 2006 with various clustering algorithms to apply for breast cancer detection. Experiments were carried out with various clustering algorithms and decision tree ensembles.

Clustering

Clustering involves partitioning a set of data points into non overlapping groups, or clusters of points where points in a cluster are more similar to one another than to points in other clusters[22]. In general, clustering algorithms are classified into two categories[22,23] (hard clustering algorithms and fuzzy clustering algorithms). In hard clustering, each data point belongs to one and only one cluster, whereas in fuzzy clustering, each data point is allowed to have membership functions to all clusters.

EM (Expectation-Maximization) Algorithm

This study used EM algorithm[22], an iterative procedure, which estimates parameters of multivariate probability density function in a form of Gaussian mixture distribution with a specified number of mixtures. Every iteration of EM algorithm includes two steps: i) Expectation-step or E-step, where a probability of each sample belongs to each mixture by using currently available mixture parameter estimates; and ii)

*Author for correspondence
E-mail: amirahmad01@gmail.com

Maximization-step or M-step, where mixture parameter estimates are refined using computed probabilities. These iterations end when a predefined condition for local minima is met. These algorithms come under probabilistic clustering algorithms.

Hierarchical Clustering

Hierarchical clustering[22] constructs a hierarchy of clusters by either repeatedly merging two smaller clusters into a larger one or splitting a larger cluster into smaller ones. First approach is agglomerative approach to build a binary tree of data that successively merges similar groups of data points. This process starts with *n* clusters, each consisting of a single point. These individual data points are leaf nodes. Similarity between merged clusters is monotone decreasing with the level of merge. Root node of tree is the cluster that has all data points. Second approach is divisive approach, in which a cluster is splitted into two smaller ones in a top-down fashion recursively. Hierarchical cluster structure provides a comprehensive description of data, which is quite useful for a number of applications. Final structure is a binary tree. However, because of high complexity, hierarchical clustering algorithms are not very useful for large datasets.

K-mean Clustering Algorithm for Mixed Datasets

K-mean clustering algorithm[24] has low complexity and is a variant of generalized EM algorithm. However, application of traditional k-mean type algorithm is limited to numeric data. Ahmad & Dey[25] proposed a clustering algorithm based on k-mean paradigm that works well for data with mixed numeric and categorical features, and also proposed a new cost function and distance measure (between two categorical attribute values) based on co-occurrence of values. This study used a mixed dataset, including algorithms proposed by Ahmad & Dey[25].

Classification

In classification problem, levels are given in training data. Model learns from given training data and predicts level of new data point. There are various classifiers (decision trees, naïve Bayes, neural networks etc.)[26,27]. Ensembles are a combination of multiple classifier models[28-31]; final classification results depend on combined outputs of individual models. Ensembles produce better results than single models, provided models are accurate and diverse. Bagging[32] and Boosting methods[33] are general and can be used with any classifiers. This study used decision tree ensembles.

Decision Trees

Decision trees are very popular tools for classification[34,35] as they represent rules. Rules can readily be expressed so that humans can understand them. A decision tree is in the form of a tree, where each node is either a leaf node (it indicates value of target class of examples) or a decision node (it specifies some test to be carried out on a single feature-value), with two or more than two branches and each branch has a sub-tree. A decision tree can classify an example by starting at the root of tree and moving through it until a leaf node, which provides rules for classification of example.

Bagging

Bagging (Bootstrap Aggregation)[32] generates different bootstrap training datasets from original training dataset and uses each of them to train one of the classifiers in ensemble. For example, to create a training set of *N* data points, it selects one point from training dataset, *N* times without replacement. Each point has equal probability of selection. In one training dataset, some of the points get selected more than once, whereas some of them are not selected at all. When different classifiers of ensemble are trained on different training datasets, diverse classifiers are created. Bagging does more to reduce the variance part of error of base classifier than bias part of error.

Adaboost.M1

Boosting[33] generates a sequence of classifiers with different weight distribution over training set. In each iteration, learning algorithm is invoked to minimize weighted error, and it returns a hypothesis. Weighted error of this hypothesis is computed and applied to update weight on training examples. Final classifier is constructed by a weighted vote of individual classifiers. Each classifier is weighted according to its accuracy on weighted training set that it has trained on. Key idea behind Boosting is to concentrate on data points that are hard to classify by increasing their weights so that probability of their selection in next round is increased. In subsequent iteration, Boosting tries to solve more difficult learning problems. Boosting reduces both bias and variance parts of error. As it concentrates on hard to classify data points, this leads to decrease in bias. At the same time, classifiers are trained on different training data sets so it helps in reducing variance. Boosting has difficulty in learning when dataset is noisy. In each iteration, weights assigned

Table 1—Information on attributes of dataset

Attribute	Type	Details on attribute
Age	Integer	Patient age is in integer
Mass shape	Categorical	round=1 oval=2 lobular=3 irregular=4
Mass margin	Categorical	circumscribed=1 microlobulated=2 obscured=3 ill-defined=4 spiculated=5
Mass density	Categorical	high=1 iso=2 low=3 fat-containing=4

to noisy data points increases so in subsequent iteration it concentrates more on noisy data points; it leads to over-fitting of data.

Experimental Setups

Elter[18] presented a dataset that is based on full field digital mammograms (Table 1) collected at Institute of Radiology, University Erlangen-Nuremberg between 2003 and 2006. This dataset has 4 attributes (1 patient's age and 3 BI-RADS attributes). This data set has 961 data points. Ground truth (B and M) for each data point is also given. There are 516 B and 445 M data points. Each data point has an associated BI-RADS assessment ranging from 1 (definitely B) to 5 (highly suggestive of M) assigned in a double-review process by physicians. One can compare performance of a computer aided design (CAD) system against prediction of radiologists. These can be an indication of how well a CAD system performs compared to radiologists. This dataset is available in UCI repository[36].

This study used Weka open source software (a collection of machine learning algorithms for data mining tasks) for experiments[37], and also hierarchical clustering and EM clustering modules from Weka software. Java program was taken from Ahmad & Dey[25]. In Hierarchical clustering, single link with euclidian was used as experimental setting. Two clusters were produced in all experiments. As EM clustering algorithms and Ahmad & Dey[25] algorithm had random initialization, 100 runs for each setting were carried out and average results with standard deviation were presented. In some experimental settings, experiments were also carried out with decision tree ensembles to compare classification results with clustering results. J48 decision tree (implementation of C4.5[35] decision tree in Weka) was used for experiments. Bagging and Adaboost.M1 modules of Weka were used in present experiments. Size of ensembles was fixed as 100. 10 repetitions of 10-cross validation were carried out and average results with statistical deviation were presented.

Parameters for Different Algorithms

Default parameters were used for all algorithms given in Weka software for present experiments. Parameters suggested in Ahmad & Dey[25] were used for Ahmad & Dey[25] algorithm. Default selection of parameters was done to remove any bias for any clustering or classification algorithm. EM algorithm and Ahmad & Dey algorithm have random initial points, hence 100 runs were carried out with different initial point. Average results for these algorithms were presented to remove any bias for any initial point. For classifier ensembles, default size is 10. Generally, performance of ensembles first improves with size[31] then becomes constant. It has been observed that for most of the experimental setups, ensemble reach optimal performance[30-32] with the size less than 100, hence ensemble size was selected as 100 so that ensembles achieve optimal performance.

Error Measure

Various parameters used to evaluate performances of various clustering techniques were as follows: Sensitivity = (TP/(TP + FN)) x 100%; Specificity = (TN/(FP + TN)) x 100%; and Accuracy = ((TP + TN)/(Total number of data points)) x 100%. Here, TP is number of true positive (M); TN, true negative; FP, false positive; and FN, false negative. High sensitivity is most desirable as any person having disease should be referred for biopsy.

Results and Discussion

Experiments were carried out in following three different settings:

a) Clustering of Complete Data and Comparing Results with Ground Truth

All data points (961) were clustered by three clustering algorithms, which created two clusters. Data has two classes, B and M (ground truth). Clustering results were compared with ground truth, and two clusters created by three algorithms, respectively, gave following results: i) EM algorithm (first cluster, 150 B data points & 392 M data points; and second cluster, 366

Table 2—Comparative study of different clustering algorithms and decision tree ensembles (classification algorithms) for complete dataset (average s.d.)

	Sensitivity %	Specificity %	Accuracy %
EM clustering algorithm	88.0 (2.1)	70.9 (1.9)	78.9 (2.4)
Ahmad & Dey clustering[25]	87.7 (2.8)	70.5 (2.3)	78.5 (2.9)
Bagging	82.1 (1.9)	78.5 (1.8)	80.7 (1.3)
Adaboost.M1	81.9 (1.7)	77.5 (2.1)	78.4 (1.5)

B data points & 53 M data points); ii) Hierarchical clustering algorithm (first cluster, 515 B data points & 445 M data points; and second cluster, only one B data point & no M data point; and iii) Ahmad & Dey[25] clustering algorithm (first cluster, 152 B data points & 390 M data points; and second cluster, 152 B data points & 392 M data points). Results shows that EM and Ahmad & Dey algorithms performed similarly, whereas hierarchical clustering algorithm put almost all data points in one cluster, suggesting that hierarchical clustering algorithm could not cluster data set. Performance of clustering algorithm depends on distance measure. This dataset is a mixed data set (3 categorical and 1 real attributes). In Weka, for categorical attributes, it uses Hamming distance (distance between two nominal values is 0 if same and 1 if different), whereas studies[38,39] uggested that this is not a very accurate distance measure, may be due to poor performance of hierarchical clustering algorithm.

To assess performance of EM and Ahmad & Dey[25] algorithms, performances were compared with the performance of classification algorithms (decision tree ensembles, Table 2). Though this is not a proper comparison as classification algorithms have access to training datasets, which have ground truth, however, this will give some indications of performance of these clustering algorithms. The best classification accuracy was obtained with Bagging algorithm (80.7%), followed by EM algorithm (78.9%) and Ahmad & Dey algorithm (78.5%). The best sensitivity was obtained with Bagging algorithm (82.1%), followed by EM algorithm (88.0%) and Ahmad & Dey algorithm (87.1%). Although accuracy of clustering algorithms are slightly low as compared to classification accuracy of classifiers, but sensitivity of clustering algorithms is better. Sensitivity is more important than overall accuracy, because better sensitivity means that more persons with M masses are predicted correctly or there is a high probability that persons with M masses will go for further tests.

In classification problem, levels are given in training data. Model learns from given training data and predicts the level of new data point. However, clustering algorithms have no access to levels. Hence, generally, performance of classification algorithms is better clustering algorithms. However, in this case, classification and clustering algorithms were observed to perform similarly. Sometimes medical datasets have class noise or attribute noise[40], that forces a classification algorithm to learn wrong model that may lead to poor accuracy. That may be the case in these experiments. Both clustering algorithms and classification algorithms learn decision boundary. Accuracy of a classification algorithm on a dataset gives indication about complexity of decision boundary of dataset. Linear decision boundaries are generally easy to learn, whereas highly non linear decision boundaries are difficult to learn[26,27]. In present experiments, classification accuracy was found 80%, which is reasonably low, suggesting that decision boundary is reasonably complex for dataset. Clustering results suggest that clustering algorithm performed well for complex decision boundary.

b) Clustering of Data Points with Prediction 4 (BI-RADS Assessment) by Radiologists and Comparing Results with Ground Truth

Most of the data points (892 out of 961) are found classified (BI-RADS assessments) by radiologists as 4 and 5. If boundary line between B and M is at 3 (if BI-RADS assessment is ≤ then data point is assigned as malign) then most of the points (892 out of 961) will be assigned as M and no real advantage of radiologist prediction. There are 547 data points with BI-RADS assessment value 4. Out of these 547 data points, 120 data points are M and 427 data points are B. If boundary is drawn at 4 then there will be another problem as 120 data points with BI-RADS assessments value 4 are M or a large number of persons with M masses will not be asked for further tests. It is suggest that clustering may

Table 3—Comparative study of different clustering algorithms for data points with prediction 4 (BI-RADS assessment) by radiologists (average s.d.)

	Sensitivity %	Specificity %	Accuracy %
EM clustering algorithm	75.0 (3.1)	74.2 (2.6)	74.4 (2.9)
Ahmad & Dey clustering[25]	73.3 (3.5)	75.6 (2.5)	75.1 (3.2)

be used to help radiologists to improve their prediction. Only those points were clustered that had BI-RADS assessments value 4. Clustering results were compared with ground truth, and two clusters created by three algorithms, respectively, gave following results: i) EM algorithm (first cluster, 317 B data points & 30 M data points; and second cluster, 110 B data points & 90 M data points; ii) Hierarchical clustering algorithm (first cluster, 426 B data points & 120 M data points; and second cluster, only one B data point & no M data point; and iii) Ahmad & Dey[25] clustering algorithm (first cluster, 323 B data points & 32 M data points; and second cluster, 104 B data points & 88 M data points). Ahmad & Dey algorithm[25] and EM algorithm achieved good results (Table 3). With Ahmad & Dey algorithm, 88 M data points (out of 120 data points) are in one cluster, whereas with EM algorithm, 90 data points (out of 120 data points) are in one cluster. With these algorithms 75% data points were correctly clustered. Hierarchical algorithm that put all data points in one cluster is not a proper clustering method for this problem. Clustering results suggest that if clustering is used with radiologist prediction, prediction accuracy can be improved.

c) Clustering of Data Points with Prediction 4 or 5 (BI-RADS Assessment) by Radiologists and Comparing Results with Radiologist Assessment

As most of the data points (892 out of 961) were predicted by radiologists in two classes on the basis of BI-RADS assessments (4 and 5), it is studied whether clustering algorithms can cluster these data points in two clusters. Clustering results were compared with prediction of radiologists (BI-RADS assessments; 4 and 5), and two clusters created by three algorithms, respectively, gave following results: i) EM algorithm (first cluster, 348 data points with BI-RADS assessment value 4 & 23 data points with BI-RADS assessment value 5; and second cluster, 199 data points with BI-RADS assessment value 4 & 322 data points with BI-RADS assessment value 5; ii) hierarchical clustering algorithm (first cluster, 547 data points with BI-RADS assessment value 4 & 342 data points with BI-RADS assessment value 5, and second cluster, only 3 data points with BI-RADS assessment value 5 & no data point with BI-RADS assessment value 4; and iii) Ahmad & Dey[25] clustering algorithm (first cluster, 390 data points with BI-RADS assessment value 4 & 32 data points with BI-RADS assessment value 5; and second cluster, 157 data points with BI-RADS assessment value 4 & 313 data points with BI-RADS assessment value 5). Comparative study of these algorithms (hierarchical clustering algorithm was not included because of its poor performance) indicated that Ahmad & Dey algorithm clustered 78.8% data points in correct clusters, whereas EM algorithm clustered 75.1% data points correctly. Hierarchical clustering algorithm clustered almost all data points in one cluster, suggesting that hierarchical clustering algorithm is not a proper clustering method for this problem. However, clustering results from Ahmad & Dey algorithm and EM algorithm are encouraging and suggest that these algorithms can be used for analysis of breast cancer data.

Ahmad & Dey algorithm is proposed for mixed datasets, and excellent clustering results are obtained by using this algorithm. It is therefore suggested to use various probabilistic clustering algorithm in general and Ahmad & Dey algorithm in particular to study breast cancer data. Fuzzy methods are better equipped to handle medical datasets[41-44]. Application of fuzzy clustering[44] for this dataset will be an interesting extension of this work.

Conclusions

This study applied various clustering algorithms to study digital mammograms. Probabilistic clustering algorithms performed well, however, in hierarchical clustering algorithm, almost all data points were clustered into one cluster, may be due to inappropriate choice of distance measure. Clustering algorithms can be combined with the decision of radiologists to improve prediction accuracy. Data has two classes (B & M), whereas there may be a case when there are some subclasses. In this study, dataset was divided into two classes and these classes (ground truth) were used to compute accuracy of clustering algorithms. It is suggested to use various probabilistic clustering algorithm in general and Ahmad & Dey algorithm in particular to study breast cancer data.

Acknowledgements

Authors thank King Abdulaziz University for its support. This work was carried out under project number 1431/830/524.

References

1 *World Cancer Report* (International Agency for Research on Cancer (Lyon, France) 2008.

2 Brenner H, Long-term survival rates of cancer patients achieved by the end of the 20th century: a period analysis, *Lancet*, **360** (2002) 1131-1135.

3 Ramakrishnan M M, Banerjee S, Chakraborty C & Ray A K, Statistical analysis of mammographic features and its classification using support vector machine. *Expert Syst Appl*, **37** (2010) 470-478.

4 Ren J, Wang D & Jiang, J, Effective recognition of MCCs in mammograms using an improved neural classifier, *Eng Appl AI*, **24** (2011) 638-645.

5 Chen H, Yang B, Liu J & Liu, D, A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis, *Expert Syst Appl*, **38** (2011) 9014-9022.

6 Ahmad A & Dey L, A k-means type clustering algorithm for subspace clustering of mixed numeric and categorical datasets, *Pattern Recogn Lett*, **32** (2011) 1062-1069.

7 Saritas I, Prediction of breast cancer using artificial neural networks, *J Med Syst*, online.

8 Quinlan J R, Improved use of continuous attributes in C4.5, *J Artif Intell Res*, **4** (1996) 77-90.

9 Bennet K P, Blue J & Math, R P I, *A Support Vector Machine Approach to Decision Trees, Report No. 97-100* (Rensselaer Polytechnic Institute, Troy, NY) 1997.

10 Pena-Reyes C A & Sipper M, A fuzzy-genetic approach to breast cancer diagnosis, *Artif Intell Med*, **17** (1999) 131-155.

11 Verma B & Panchal R, Neural networks for the classification of benign and malignant patterns in digital mammograms, in *Advances in Applied Artificial Intelligence*, edited by J Fulcher (Idea Group, USA) 2006.

12 Verma B, Novel network architecture and learning algorithm for the classification of mass abnormalities in digitized mammograms, *Artif Intell Med*, **42** (2008) 67-79.

13 Bjurstam N, Björneld L, Warwick J, Sala E, Duffy S W *et al*, The Gothenburg breast screening trial, *Cancer*, **97** (2003) 2387-2396.

14 Rijnsburger A J, van Oortmarssen G J, Boer R, Draisma G, Miler A B, Koning H J, Mammography benefit in the Canadian National Breast Screening Study-2: a model evaluation, *Int J Cancer*, **110** (2004) 756-762.

15 Verma B & Zakos J A, Computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction technique, *IEEE T Inf Technol Biomed*, **5** (2001) 46-54.

16 Acharya U R, Ng E Y K, Chang Y H, Yang J & Kaw G J L, Computer-based identification of breast cancer sing digitized mammograms, *J Med Syst*, **32** (2008) 499-507.

17 Rafayah M, Qutaishat M & Abdallah, M, Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural, *Expert Syst Appl*, **28** (2005) 713-723.

18 Elter M, Schulz-Wendtland R & Wittenberg T, The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process, *Med Phys*, **34** (2007) 4164-4172.

19 Luo S & Cheng B, Diagnosing breast masses in digital mammography using feature selection and ensemble methods, *J Med Syst*, DOI: 10.1007/s10916-010-9518-8, online.

20 Chakraborty S, Bayesian semi-supervised learning with support vector machine, *Stat Methodol*, **8** (2011) 68-82.

21 Polat K, Application of attribute weighting method based on clustering centers to discrimination of linearly non-separable medical datasets, *J Med Syst*, DOI: 10.1007/s10916-011-9741-y, online.

22 Jain A K & Dubes R C, *Algorithms for Clustering Data* (Prentice Hall, Englewood Cliff, New Jersey) 1988.

23 Everitt B, Sabine L, Morven L & Leese M, *Cluster Analysis* (Hodder, Arnold) 2001.

24 MacQuuen J B, Some methods for classification and analysis of multivariate observation, in *Proc 5th Berkley Symp on Mathematical Statistics and Probability* (Berkelely, CA) 1967, 281-297.

25 Ahmad A & Dey L, A k-mean clustering algorithm for mixed numeric and categorical data, *Data Knowl Eng*, **63** (2007) 503-527.

26 Mitchell T, *Machine Learning* (McGraw-Hill, New York) 1997.

27 Bishop C M, *Pattern Recognition and Machine Learning*, **1st edn** (Springer-Verlag, New York) 2006.

28 Tumer K & Ghosh J, Error correlation and error reduction in ensemble classifiers, *Connect Sci*, **8** (1996) 385-404.

29 Hansen L K & Salamon P, Neural network ensembles, in *IEEE Trans Patt Anal Mach Intell*, vol. **12** (Society of Industrial & ASpplied Mathematics, Philadelphia, USA) 1990, 993-1001.

30 Brown, G, Wyatt J & Tino P, Managing diversity in regression ensembles, *J Mach Learning Res*, **6** (2005) 1621-1650.

31 Kuncheva L I, *Combining Pattern Classifiers: Methods and Algorithms* (Wiley-IEEE Press, New York) (2004).

32 Breiman L, Bagging predictors, *Mach Learn*, **24** (1996) 123-140.

33 Freund Y & Schapire R E, A decision theoretic generalization of on-line learning and an application to boosting, *J Comput Syst Sci*, **55** (1997) 119-139.

34 Breiman L, Friedman J, Olshen R & Stone C, *Classification and Regression Trees*, **1st edn** (Chapman and Hall, London) 1984.

35 Quinlan J R, *C4.5: Programs for Machine Learning* (Morgan Kaufmann, San Mateo) 1993.

36 Newmann, D J, Hettich S, Blake C L & Merz C J, *UCI repository of machine learning database*, http://archive.ics.uci.edu/ml/datasets/ Mammographic+Mass, Irvine (University of Cali-

fornia, Department of Information and Computer Science, USA) 1998.

37 Witten I H & Frank E, *Data Mining: Practical Machine Learning Tools with Java Implementations* (Morgan Kaufmann, San Francisco) 2000.

38 Ahmad A & Dey L, A method to compute distance between two categorical values of same attribute in unsupervised learning for categorical data set, *Patt Recogn Lett*, **28** (2007) 110-118.

39 Boriah S, Chandola V & Kumar V, Similarity measures for categorical data: A comparative evaluation, in *Proc 8th SIAM Int Conf on Data Mining* 2008, 243-254.

40 Zhu X, & Wu X, Class noise *vs* attribute noise: a quantitative study, *Artif Intell Rev*, **22** (2004) 177-210.

41 Dunn J C, Some recent investigations of a new fuzzy algorithm and its application to pattern classification problems, *J Cybernetics*, **4** (1974) 1-15.

42 Bezdek J C, *Pattern recognition with fuzzy Objective Function algorithms (Plenum*, New York) 1981.

43 Innocent P R, John R I & Garibaldi J M, Fuzzy methods for medical diagnosis, *Appl Artif Intell*, **19** (2005) 69-98.

44 Ahmad A & Dey L, Algorithm for fuzzy clustering of mixed data with numeric and categorical attributes, *ICDCIT*, (2005) 561-572.