JOURNAL OF APPLIED QUANTITATIVE METHODS

DATA MINING TECHNIQUES IN PROCESSING MEDICAL KNOWLEDGE

Luminita STATE
PhD, University Professor, Department of Computer Science,
University of Pitesti, Pitesti, Romania

E-mail: radus@sunu.rnc.ro

Catalina COCIANU
PhD, University Professor, Department of Computer Science,
University of Economics, Bucharest, Romania

E-mail: ccocianu@ase.ro

Abstract: *Data mining is an evolving and growing area of research and development, both in academia as well as in industry. It involves interdisciplinary research and development encompassing diverse domains. In this age of multimedia data exploration, data mining should no longer be restricted to the mining of knowledge from large volumes of high-dimensional data sets in traditional databases only. The aim of the paper is to develop a new learning by examples PCA-based algorithm for extracting skeleton information from data to assure both good recognition performances, and generalization capabilities in case of large data set. The classes are represented in the measurement/feature space by continuous repartitions, that is the model is given by the family of density functions* $\left(f_h\right)_{h\in H}$*, where H stands for the finite set of hypothesis (classes). The basis of the learning process is represented by samples of possible different sizes coming from the considered classes. The skeleton of each class is given by the principal components obtained for the corresponding sample.*

Key words: *data mining; principal component analysis; fuzzy clustering; c-means algorithm; supervised learning; cluster analysis*

1. Introduction

The last decade has witnessed a revolution in interdisciplinary research where the boundaries of different areas have overlapped or even disappeared. New fields of research emerge each day where two or more fields have integrated to form a new identity. Examples of these emerging areas include bioinformatics (synthesizing biology with computer and information systems), data mining (combining statistics, optimization, machine learning, artificial intelligence, and databases), and modern heuristics (integrating ideas from tens of fields such as biology, immunology, statistical mechanics, and physics to inspire search techniques). These integrations have proved useful in substantiating problem-solving approaches with reliable and robust techniques to handle the increasing demand from practitioners to solve real-life problems.

In the old days, system analysts faced many difficulties in finding enough data to feed into their models. The picture has changed and since databases have grown exponentially, ranging in size into the terabytes within these masses of data being hidden information of strategic importance, the reverse picture becomes a daily problem-how to understand the large amount of data we have accumulated over the years. When there are so many trees, how do we draw meaningful conclusions about the forest? Research into statistics, machine learning, and data analysis has been resurrected. Unfortunately, with the amount of data and the complexity of the underlying models, traditional approaches in statistics, machine learning, and traditional data analysis fail to cope with this level of complexity. The need therefore arises for better approaches that are able to handle complex models in a reasonable amount of time.

The newest answer has been named data mining (sometimes data farming), to distinguish them from traditional statistics, machine learning and other data analysis techniques. In addition, decision makers were not interested in techniques that rely too much on the underlying assumptions in statistical models. The challenge is not to have any assumptions about the model and try to come up with something new, something that is not obvious or predictable (at least from the decision maker's point of view). Notwithstanding, models that are free from assumptions-or at least have minimum assumptions-are expensive to use because of the inherent complexity due to the high dimensionality of the processed data. The dramatic search space cannot be navigated using traditional searching techniques, and his has highlighted a natural demand for the use a special tailored methodology. Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid predictions.

According to the Gartner Group , "Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques".

There are also other definitions:

- "Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understanble and useful to the data owner (Hand, 2001)
- "Data mining is an interdisciplinary field bringing together techniques from machine learning, pattern recognition,statistics,databases,and visualization to address the issue of information extraction from large data bases" (Cabena,1998)

Data mining is an evolving and growing area of research and development, both in academia as well as in industry. It involves interdisciplinary research and development encompassing diverse domains. In this age of multimedia data exploration, data mining should no longer be restricted to the mining of knowledge from large volumes of high-dimensional data sets in traditional databases only. Researchers need to pay attention to the mining of different datatypes, including numeric and alphanumeric formats, text, images, video, voice, speech, graphics, and also their mixed representations. Fuzzy sets provide the uncertainty handling capability, inherent in human reasoning, while artificial neural networks help incorporate learning to minimize error. Genetic algorithms introduce effective parallel searching in the high-dimensional problem space.

2. Cluster analysis

Cluster analysis is a method of grouping data with similar characteristics into larger units of analysis. Since Zadeh, 1965, first articulated fuzzy set theory which gave rise to the concept of partial membership, based on membership functions, fuzziness has received increasing attention. Fuzzy clustering, which produce overlapping cluster partitions, has been widely studied and applied in various area (Bezdek, 1999).

So far, there have been proposed a relatively small number of methods for testing the existence/inexistence of a natural grouping tendency in a data collection, most of them being based on arguments coming from mathematical statistics and heuristic graphical techniques (Panayirci and Dubes,1983, Smith and Jain,1984,Jain and Dubes,1988, Tukey,1977, Everitt,1978).

The data are represented by *p*-dimensional vectors, $X=\left(x_1,...,x_p\right)^t$, whose components are the feature values of a specified attributes and the classification is performed against a certain given label set. The classification of a data collection $\aleph=\left\{X_1,...,X_n\right\}\subset\Re^p$ corresponds to a labelling strategy of the objects of $\aleph$.

In the fuzzy approaches, the clusters are represented as fuzzy sets $\left(u_i,1\le i\le c\right)$, $u_i:\aleph\to[0,1]$, where $u_{ik}=u_i\left(X_k\right)$ is the membership degree of X_k to the *i*-th cluster, $1\le i\le c$, $1\le k\le n$. A c-fuzzy partition is represented by the matrix $U=\left\|u_{ik}\right\|\in M_{c\times n}$. The number of labels c has to be selected in advance, the problem of finding the optimal c is usually referred as cluster validation.

The main types of label vectors are *crisp* N_c, *fuzzy* N_p, *and possibilistic* N_{poz}, defined as follows,

$$N_c=\left\{y\mid y\in\Re^c, y=\left(y_1,y_2,...,y_c\right), y_i\in\{0,1\}, 1\le i\le c, \sum_{i=1}^{c}y_i=1\right\}=\left\{e_1,e_2,...,e_c\right\}, \quad (1)$$

where $\left(e_i\right)_j=\delta_{ij}=\begin{cases}0, i\neq j\\ 1, i=j\end{cases}$ (2)

$$N_p=\left\{y\in\Re^c\mid y=\left(y_1,y_2,...,y_c\right), \forall i, y_i\in[0,1], \sum_{i=1}^{c}y_i=1\right\}, \quad (3)$$

$$N_{poz}=\left\{y\in\Re^c\mid y=\left(y_1,y_2,...,y_c\right), \forall i, y_i\in[0,1], \exists j, y_j\neq 0\right\}, \quad (4)$$

Obviously, $N_{pos}\supset N_p\supset N_c$. If we denote by $U=\left[U_1,...,U_n\right]=\left\|u_{ij}\right\|$ a partition of $\aleph$, then, according to the types of label vectors, we get the *c*-partition types M_{pos}, M_p and M_c,

$$M_{pos}=\left\{U\mid U\in M_{c\times n}, U=\left[U_1,...,U_n\right], \forall k, U_k\in N_{pos}, \forall i, \sum_{k=1}^{n}u_{ik}>0\right\} \quad (5)$$

$$M_p=\left\{U\,\middle|\,U\in M_{pos}, \forall k, U_k\in N_p\right\} \quad (6)$$

$$M_c=\left\{U\,\middle|\,U\in M_p, \forall k, U_k\in N_c\right\} \quad (7)$$

Note that $M_c\subset M_p\subset M_{pos}$.

3. C-MEANS model

The *C-MEANS* algorithm is the most popular non-hierarchical iterative clustering algorithm. When it is applied on a set of data $\aleph = \{X_1,...,X_n\} \subset \Re^p$ the c-means finds some of the most natural c-groups exiting in data.

C-Means Clustering Algorithm

(1) Choose c cluster centers to coincide with c randomly-chosen patterns or c randomly defined points inside the hypervolume containing the pattern set.

(2) Assign each pattern to the closest cluster center according to a certain pre-specified metric or dissimilarity measure.

(3) Recompute the cluster centers using the current cluster memberships.

(4) If a convergence criterion is not met, go to step 2.

else the computation is over and the current clusters correspond to the c- groups identified by the algorithm in the data.

Typical convergence criteria are: no (or minimal) reassignment of patterns to new cluster centers, or minimal decrease in squared error.

Several variants (Anderberg, 1973) of the c-means algorithm have been reported in the literature. Some of them attempt to select a good initial partition so that the algorithm is more likely to find the global minimum value.

In fuzzy clustering, the fuzzy c-means clustering algorithms are the best know and most powerful methods used in cluster analysis (Bezdek, 1981).

The variational problem corresponding to *c-means* model is given by

$$\min_{(U,V)}\{J_m(U,V;w)\} \tag{8}$$

where

$$J_m(U,V;w) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m D_{ik}^2 + \sum_{i=1}^{c} w_i \sum_{k=1}^{n}(1-u_{ik})^m$$

$U \in M_c / M_p / M_{pos}$, $V = (v_1,...,v_c) \in M_{c\times p}$, v_i is the centroid of the *i*-th cluster , $w = (w_1,..,w_c)^T$ is the penalties vector corresponding to the cluster system, $m \geq 1$ is the fuzzyfication degree, and $D_{ik}^2 = \|x_k - v_i\|^2$.

Let us denote by $(\hat{U},\hat{V})$ a solution of (8). Then,

1. The crisp model:

$$(U,V) \in M_c \times M_{c\times p}; w_i = 0, 1 \leq i \leq c\,,\ \hat{u}_{ik} = \begin{cases} 1, D_{ik} \leq D_{ij}, i \neq j \\ 0, \text{otherwise} \end{cases} \tag{9}$$

$$\hat{v}_i = \frac{\sum_{k=1}^{n} \hat{u}_{ik} x_k}{\sum_{k=1}^{n} \hat{u}_{ik}}; \quad 1 \leq i \leq c, 1 \leq k \leq n \tag{10}$$

2. The fuzzy model:

$$(U,V) \in M_p \times M_{c\times p}; m > 1, w_i = 0, 1 \leq i \leq c$$

$$\hat{u}_{ik} = \left[\sum_{j=1}^{c} \left(\frac{D_{ik}}{D_{jk}} \right)^{\frac{2}{m-1}} \right]^{-1} \quad (11)$$

$$\hat{v}_i = \frac{\sum_{k=1}^{n} u_{ik}^m x_k}{\sum_{k=1}^{n} u_{ik}^m}; \quad 1 \le i \le c, 1 \le k \le n \quad (12)$$

3. The possibilistic model:

$(U,V) \in M_{pos} \times M_{c \times p}; \forall i, w_i > 0$

$$\hat{u}_{ik} = \left[1 + \left(\frac{D_{ik}^2}{w_i} \right)^{\frac{1}{m-1}} \right]^{-1}$$

$$\hat{v}_i = \frac{\sum_{k=1}^{n} u_{ik}^m x_k}{\sum_{k=1}^{n} u_{ik}^m}; \quad 1 \le i \le c, 1 \le k \le n \quad (13)$$

The general scheme of a cluster procedure $\wp$ is ,

$t \leftarrow 0$

repeat

$t \leftarrow t+1$

$U_t \leftarrow F_{\wp}(V_{t-1})$

$V_t \leftarrow G_{\wp}(U_{t-1})$

$until\left(t = Tor \left\| V_t - V_{t-1} \right\| \le \varepsilon\right)$

$(U,V) \leftarrow (U_t, V_t)$

where c is the given number of clusters, *T* is upper limit on the number of iterations, *m* is the weight parameter, $1 \le m < \infty$, C is the terminal condition, w is the system of weights $\forall i, w_i > 0$, $V_0 = (v_{1,0}, ..., v_{c,0}) \in M_{c \times p}$ is the initial system of centroids and $F_{\wp}$, $G_{\wp}$ are the updating functions.

4. PCA-based algorithm for extracting skeleton information

In the following a new learning by examples PCA-based algorithm for extracting skeleton information from data to assure both good recognition performances, and generalization capabilities, is developed. Here the generalization capabilities are viewed twofold, on one hand to identify the right class for new samples coming from one of the classes taken into account and, on the other hand, to identify the samples coming from a new class. The classes are represented in the measurement/feature space by continuous repartitions, that is the model is given by the family of density functions $(f_h)_{h \in H}$, where *H* stands for the finite set of hypothesis (classes).

The basis of the learning process is represented by samples of possible different sizes coming from these classes. The skeleton of each class is given by the principal components obtained for the corresponding sample. The recognition algorithm identifies the class whose skeleton is the "nearest" to the tested example, where the closeness degree is expressed in terms of the amount of disturbance determined by the decision of allotting it to the corresponding class. The model is presented as follows. Let $X_1, X_2, ..., X_N$ be a series of *n*-dimensional vectors coming from a certain class C. The sample covariance matrix is

$$\Sigma_N = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \mu_N)(X_i - \mu_N)^T , \qquad (14)$$

where $\mu_N = \frac{1}{N}\sum_{i=1}^{N} X_i$.

We denote by $\lambda_1^N \geq \lambda_2^N \geq ... \geq \lambda_n^N$ the eigen values and by $\psi_1^N, ..., \psi_n^N$ a set of orthonormal eigen vectors of Σ_N .

If X_{N+1} is a new sample, then, for the series $X_1, X_2, ..., X_N, X_{N+1}$, we get

$$\Sigma_{N+1} = \Sigma_N + \frac{1}{N+1}(X_{N+1} - \mu_N)(X_{N+1} - \mu_N)^T - \frac{1}{N}\Sigma_N \qquad (15)$$

The computation can be carried out as follows.

$$\mu_{N+1} = \frac{1}{N+1}\sum_{i=1}^{N+1} X_i = \frac{N}{N+1}\mu_N + \frac{1}{N+1}X_{N+1}$$

$$\Sigma_{N+1} = \frac{1}{N}\sum_{i=1}^{N+1}(X_i - \mu_{N+1})(X_i - \mu_{N+1})^T =$$

$$= \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu_{N+1})(X_i - \mu_{N+1})^T + \frac{1}{N}(X_{N+1} - \mu_{N+1})(X_{N+1} - \mu_{N+1})^T$$

Obviously,

$$\frac{1}{N}(X_{N+1} - \mu_{N+1})(X_{N+1} - \mu_{N+1})^T = \frac{1}{N}\left(\frac{N}{N+1}\right)^2 (X_{N+1} - \mu_N)(X_{N+1} - \mu_N)^T$$

For each $1 \leq i \leq N$, we have

$$X_i - \mu_{N+1} = X_i - \frac{N}{N+1}\mu_N - \frac{1}{N+1}X_{N+1} = X_i - \mu_N + \frac{1}{N+1}\mu_N - \frac{1}{N+1}X_{N+1}$$

Hence, taking into account that $\sum_{i=1}^{N}(X_i - \mu_N) = 0$

$$\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu_{N+1})(X_i - \mu_{N+1})^T = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \mu_N + \frac{1}{N+1}\mu_N - \frac{1}{N+1}X_{N+1}\right)\left(X_i - \mu_N + \frac{1}{N+1}\mu_N - \frac{1}{N+1}X_{N+1}\right)^T =$$

$$= \frac{N-1}{N}\Sigma_N + \frac{1}{N(N+1)}\sum_{i=1}^{N}(X_i - \mu_N)(\mu_N - X_{N+1})^T + \frac{1}{N(N+1)}\sum_{i=1}^{N}(\mu_N - X_{N+1})(X_i - \mu_N)^T +$$

$$+\left(\frac{1}{N+1}\right)^2 (\mu_N - X_{N+1})(\mu_N - X_{N+1})^T = \frac{N-1}{N}\Sigma_N + \left(\frac{1}{N+1}\right)^2 (\mu_N - X_{N+1})(\mu_N - X_{N+1})^T$$

Therefore, we finally obtain.

$$\Sigma_{N+1} = \frac{N-1}{N}\Sigma_N + \left(\frac{1}{N+1}\right)^2 (\mu_N - X_{N+1})(\mu_N - X_{N+1})^T + \frac{1}{N}\left(\frac{N}{N+1}\right)^2 (X_{N+1} - \mu_N)(X_{N+1} - \mu_N)^T =$$

$$= \frac{N-1}{N}\Sigma_N + \frac{1}{N+1}(X_{N+1} - \mu_N)(X_{N+1} - \mu_N)^T$$

Lemma. In case the eigen values of Σ_N are distinct, the following first order approximations hold, (State, Cocianu & al., 2006)

$$\lambda_i^{N+1} = \lambda_i^N + \left(\psi_i^N\right)^T \Delta\Sigma_N \psi_i^N = \left(\psi_i^N\right)^T \Sigma_{N+1} \psi_i^N \quad (16)$$

$$\psi_i^{N+1} = \psi_i^N + \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{\left(\psi_N^j\right)^T \Delta\Sigma_N \psi_i^N}{\lambda_i^N - \lambda_j^N} \psi_j^N \quad (17)$$

The skeleton of C is represented by the set of estimated principal components $\psi_1^N,...,\psi_n^N$ .When the example X_{N+1} is included in C, then the new skeleton is $\psi_1^{N+1},...,\psi_n^{N+1}$. The skeleton disturbance induced by the decision that X_{N+1} has to be alloted to C is measured by

$$D = \frac{1}{n}\sum_{k=1}^{n} d\left(\psi_k^N, \psi_k^{N+/}\right) \quad (18)$$

where $d\left(\psi_k^N, \psi_k^{N+/}\right) = \left(\sum_{j=1}^{N}\left(\psi_{kj}^N - \psi_{kj}^{N+1}\right)^2\right)^{\frac{1}{2}}$

The crisp classification procedure identifies for each example the closest cluster in terms of the measure (18). Let $H = \{C_1, C_2,..., C_M\}$. In order to protect against misclassifications of samples coming from new classes not belonging to H, a threshold T>0 is imposed, that is the example X_{N+1} is alloted to one of C_i for which

$$D = \frac{1}{n}\sum_{k=1}^{n} d\left(\psi_{k,j}^N, \psi_{k,j}^{N+1}\right) = \min_{1\le p\le M} \frac{1}{n}\sum_{k=1}^{n} d\left(\psi_{k,p}^N, \psi_{k,p}^{N+1}\right) \quad (19)$$

and D<T, where the skeleton of C_i is $\psi_{1,j}^N,...,\psi_{n,j}^N$.

The classification of samples for which the resulted value of D is larger than T is postponed and the samples are kept in a new possible class CR. The reclassification of elements of CR is then performed followed by the decision concerning to either reconfigure the class system or to add CR as a new class in H.

In case of fuzzy classification, the value of the membership degree of X_{N+1} to each cluster of $HT = H \cup \{CR\}$ is computed as follows. Let $d(X_{N+1}, C_i) = \frac{1}{n}\sum_{k=1}^{n} d\left(\psi_{k,i}^N, \psi_{k,i}^{N+1}\right)$, $1 \le i \le M$ and

$d(X_{N+1}, CR) = \begin{cases} \frac{1}{n}\sum_{k=1}^{n} d\left(\psi_{k,R}^N, \psi_{k,R}^{N+1}\right), \text{ if } CR \neq \varnothing \\ \infty, \text{ otherwise} \end{cases}$, where the skeleton of CR is represented by the set of estimated principal components $\psi_{1,R}^N,...,\psi_{n,R}^N$.

$$\mu_{C_i}(X_{N+1}) = 1 - \frac{d(X_{N+1}, C_i)}{S}, \quad 1 \le i \le M \quad (20)$$

$$\mu_{CR}(X_{N+1}) = \begin{cases} 1 - \dfrac{d(X_{N+1}, CR)}{S}, \text{ if } d(X_{N+1}, CR) \neq \infty \\ 0, \text{ otherwise} \end{cases} \quad (21)$$

where

$$S = \begin{cases} \sum_{C \in HT} d(X_{N+1}, C), \text{ if } d(X_{N+1}, CR) \neq \infty \\ \sum_{C \in H} d(X_{N+1}, C), \text{ if } d(X_{N+1}, CR) = \infty \end{cases} \quad (22)$$

5. Experimental results and concluding remarks

Several tests were performed on simulated data and they pointed out very successful performance of the proposed classification strategy.

A series of tests were performed on 4-dimensional simulated data coming from 5 clases each of them having 50 examples. Each class consists of Gaussian data (State, Cocianu &al., 2006)

The classification criterion is: allote X_{N+1} to C_{j_i} if

$$D = \min_{1 \le l \le t} \frac{1}{m_{j_l}} \sum_{k=1}^{m_{j_l}} d\left(\psi_{j_l}^k, \psi_{j_l,N+1}^k\right) \quad (23)$$

In order to evaluate the generalization capacities, 100 new examples were generated for each distribution. The results are presented in Table 1. (State, Cocianu &al., 2006)

The evaluation of the generalization capacities in case of examples coming from new classes was performed on 1000 samples generated from $N(\mu, \Sigma)$, where μ = [0 11 -9 -9.5] and

$$\Sigma = \begin{bmatrix} 8.2725 & 3.1080 & 1.7925 & 1.3680 \\ 3.1080 & 6.8986 & 1.8390 & 2.5561 \\ 1.7925 & 1.8390 & 6.0422 & 1.6410 \\ 1.3680 & 2.5561 & 1.6410 & 5.2261 \end{bmatrix}.$$

The admissibility criterion for allotting a sample to a certain class is given by the maximum value of D corresponding to correct classifications. The results showed that about 975 examples were classified in CR, that is the algorithm managed to detect the intruded examples. (State, Cocianu &al., 2006)

Table 1. Results on new simulated examples

Class	C_1	C_2	C_3	C_4	C_5
Number of correct classified examples	100	100	96	99	100
Number of missclassified examples	0	0	4 - allotted to C_2	1 – allotted to C_1	0
The mean value of D in case of correct classifications	0.08	0.05	0.75	0.21	0.14
The maximum value of D in case of correct classified examples	0.41	0.19	1.85	0.55	0.53

References

1. Andemberg, M. R. **Cluster Analysis for Applications,** Academic Press, Inc., New York, NY, 1973
2. Al Sultan, K. S., and Selim, S. Z. **Global Algorithm for Fuzzy Clustering Problem,** Patt. Recogn. 26, 1993, pp.1375-1361
3. Bensaid, A., Hall, L. O., Bezdek, J. C., and Clarke, L. P. **Partially supervised clustering for image segmentation,** Patt. Recog., 29(5), 1996, pp. 859-871
4. Bezdek, J. C., Keller, J., Krisnapuram, R., and Pal, N. K. **Fuzzy Models and Algorithms for Pattern Recognition and Image Processing,** Springer Verlag, 1999.
5. Bezdek, J. C. **Pattern Recognition with Fuzzy Objective Function Algorithms,** Plenum Press, 1981
6. Bensaid, H., Bezdek, J. C., Clarke, L. P., Silbiger, M. L., Arrington, J. A., and Murtagh, R. F. **Validity-Guided (Re)clustering with applications to image segmentation,** IEEE Trans. Fuzzy Systems, 4, 1996, pp. 112-123
7. Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., and Zanasi, A. **Discovering Data Mining: From Concept to Implementation,** Prentice Hall, 1998
8. Clark, M., Hall, L. O., Goldgof, D. B., Clarke, L. P., Velthuizen, R. P., and Silbiger, M. S. **MRI Segmentation using Fuzzy Clustering Techniques,** IEEE Engineering in Medicine and Biology, 1994
9. Everitt, B. S. **Graphical Techniques for Multivariate Data,** North Holland, NY, 1978
10. Gath, J., and Geva, A. B. **Unsupervised optimal fuzzy clustering,** IEEE Trans. Pattern.Anal.Machine Intell, 11, 1989, pp. 773-781
11. Hyvarinen, A., Karhunen, J. and Oja, E. **Independent Component Analysis,** John Wiley &Sons, 2001
12. Hand, D, Mannila, H., and Smyth, P. **Principles of Data Mining,** MIT Press, Cambridge, MA, 2001
13. Huang, C., and Shi, Y. **Towards Efficient Fuzzy Information Processing,** Physica-Verlag, Heidelberg, 2002
14. Jain, A. K., and Dubes, R. **Algorithms for Clustering Data,** Prentice Hall, Englewood Cliffs, NJ, 1988
15. Jin, Y. **Advanced Fuzzy Systems Design and Applications,** Physica-Verlag, Heidelberg, 2003
16. Krishnapuram, R., and Keller, J. M. **A possibilistic approach to clustering,** IEEE Trans. Fuzzy Syst., 1(2), 1993
17. Larose, D. T. **Discovering Knowledge in Data: An Introduction to Data Mining,** Wiley-Interscience, 2005
18. Li, C., Goldgof, D. B., and Hall, L. O. **Automatic segmentation and tissue labelling of MR brain images,** IEEE Transactions on Medical Imaging, 12(4), 1993, pp.1033-1048
19. Pal, N. R., and Bezdek, J. C. **On Cluster validity for the Fuzzy c-Means Model,** IEEE Trans. On Fuzzy Syst., Vol. 3, no. 3, 1995
20. Smith, S. P., and Jain, A. K. **Testing for uniformity in multidimensional data,** IEEE Trans.`Patt. Anal.` amd Machine Intell., 6 (1), 1983, pp. 73-81
21. State, L., Cocianu, C., Vlamos, P., and Stefanescu, V. **PCA-based Data Mining Probabilistic and Fuzzy Approaches with Applications in Pattern Recognition,** Proceedings of ICSOFT Setubal, Portugal, September 11-14, 2006
22. Wu, K-L., and Yang, M-S, **A Cluster validity index for fuzzy clustrering,** Patt. Recog. Lett. 26, 2005, pp. 1275-1291
23. Zahid, N., Abouelala, O., Limouri, M., and Essaid, A. **Unsupervised fuzzy clustering,** Patt. Recog. Lett., 20,1, 1999
24. * * * **The Gartner Group**, www.garner.com