Mining of Fuzzy Association Rules using a Kernel-based Self-Organized Map for partitioning large databases

STERGIOS PAPADIMITRIOU[1] KONSTANTINOS TERZIDIS[2] SEFERINA MAVROUDI[3] SPIRIDON D LIKOTHANASSIS[4]

Department of Information Management
Technological Educational Institute of Kavala
65404 Kavala, GREECE

Pattern Recognition Laboratory, Department of Computer Engineering and Informatics, School of Engineering, University of Patras,
Rion, Patras, 26500, GREECE

Abstract: The extraction of fuzzy association rules for the description of dependencies and interactions from large data sets as those arising in gene expression data analysis applications perplexes very difficult combinatorial problems that depend heavily on the size of these sets. The paper describes a two stage approach to the problem that obtains computationally manageable solutions. The first stage aims to cluster transactions that more probably are associated. Thereafter, the second stage, the fuzzy association rule extraction follows, confronting a significantly reduced problem.
The clustering phase is accomplished by means of a Kernel Supervised Dynamic Grid Self-Organized Map (KSDG-SOM) utilizing the mutual information metric. This metric allows the formation of data clusters that maximize the mutual information for transactions of the same cluster and to minimize it between different clusters.
Subsequently, the fuzzy association rules are extracted locally on a *per cluster basis*. The paper applies the techniques for mining gene expression data.

Key-Words: *Fuzzy Association Rules, Mutual Information, Clustering, Self-Organized Maps, Entropy, Genome Data Mining, Gene Expression Analysis*

1 Introduction

Recently, the discovery of association rules from databases has become an important research topic

[25, 34]. Although traditionally these techniques have been developed for commercial applications, nowadays the genomic revolution presents another promising domain for their exploitation.
The whole genome studies of gene expression data of recent years produce an extraordinary large number of measurements. A great chalenge is to uncover dependencies between genes from these measurements by computational data mining techniques.
The extraction of association rules can provide useful insight to some aspects of the interactions between genes in an easily readable form [36]. However, a main difficulty arises from the extremely large size of whole genome gene expression data sets (and of many other data sets as well).
The present work attempts to cope with this difficulty by a two step divide-and-conquer approach:

1) A mutual information based clustering implemented with the KSDG-SOM model, that isolates the more probably related genes onto the same clusters.
2) Fuzzy association rules are effectively extracted for each KSDG-SOM node.

An important criterion that we impose for the KSDG-SOM growing phase, is the minimization of the number of genes in each cluster in order to restrict the search over large spaces. Thereafter, the mutual information metric with the provision for incorporating a priori functional classes constructs highly appropriate clusters for the detection of fuzzy association rules.

The paper proceeds as follows: Section 2 summarizes the KSDG-SOM algorithm, on which with a few modifications, we build the mutual information based clustering phase. Then at section 3 we describe the concept of the mutual information distance metric and its utilization for the formation of clusters. The next section, first briefly presents a background on association rules. Then we present the association rule extraction algorithms, that are applied locally for each KSDG-SOM formed cluster. Subsequently, the crisp association rule extraction framework is extended to the fuzzy domain. Section 5 presents and discusses the results obtained from this two phase fuzzy association rule extraction frawework. The paper concludes with hints on the many possible directions on which this research can be continued and extended.

2 Summary of the KSDG-SOM algorithm

This section summarizes the Kernel Based Self-Organizing Map (KSDG-SOM) that forms the basis of the mutual information based clustering framework. This clustering aims to separate small groups of related patterns which the association rules extraction algorithm will focus in order to reveal association rules.
Details of the algorithms of the KSDG-SOM approach are presented in [38].
The standard SOM algorithm has a number of properties, which render it to a candidate of particular interest as a basis framework for building more flexible and advanced algorithms for massive data analysis. SOMs can be implemented easily, are fast, robust and scale well to large data sets. They allow one to impose partial structure on the clusters and facilitate visualization and interpretation. In the case hierarchical information is required, it can be implemented on top of SOM, as in [14].
Recently, several dynamically extended schemes have been proposed that overcome the limitation of the fixed non-adaptable architecture of the SOM. Some examples are the Dynamic Topology Representing structures [15], the Growing Cell Structures [11,16], Self-Organized Tree Algorithms [17,8], the joint entropy maximization approach [18] and the Adaptive Resonance Theory [19,20].
The KSDG-SOM approach has some similarities to these dynamically extended schemes, from the point of view of its unsupervised component. However, one essential difference exists: all the forementioned schemes are purely unsupervised, lacking a design for the incorporation of problem domain knowledge. Instead, the KSDG-SOM focus on the design of such types of algorithms that aim to explore effectively existing *a priori supervised class labeling*, for *multi-class* and possibly *multi-labeled* data. The multiple labeling, i.e. the possible assignment of more than one class label at each pattern, perplexes the clustering and classification tasks. For many applications, e.g. the gene expression analysis, the multiple functional labeling of patterns is the rule and not the exception (e.g. most genes belong to more than one functional class).
Also, in contrast to the complexity of some of these schemes, the KSDG-SOM is based on simple algorithms that through the restriction of growing on a rectangular grid, can be implemented easily and the training of the models is very efficient. In addition, most of the benefits of the more complex alternatives of dynamical extension are still retained.
We call the proposed model KSDG-SOM from Kernel

Supervised Dynamic Grid SOM, since it is a model trained in kernel space and although it is SOM based it tightly integrates unsupervised and supervised learning components.

As a kernel the Gaussian one is used. The Gaussian kernel mapping implements more elaborate soft class separation boundaries than the hard separation onto Voronoi regions obtained by evaluating directly at the input space the inner products of the patterns and the prototype vectors. As with other kernel methods [35,39], we aim to exploit a nonlinear transformation of the input space onto a high-dimensional feature space. Intuitively, the SOM based learning constructs Voronoi regions over this high-dimensional space in which the extra dimensions enhance the possibilities of defining more elaborately the cluster boundaries.

The KSDG-SOM has been designed in order to automatically detect the *appropriate level of expansion*, so that the number of clusters is controlled by a properly defined measure of the algorithm itself, with no need for any a priori specification. This is quite important for many data mining applications where very little (or nothing) can be claimed about an a priori estimate of the number of clusters. To fullfill the needs of the association rule extraction framework we have performed slight modifications to the expansion phase, in order to obtain cluster sizes in the range 30 to 80 genes, that can be conveniently handled by the association rule extractor.

Details on the design and implementation of KSDG-SOM algorithms can be found in [38]. At the context of the present work, we augment the KSDG-SOM with powerful *mutual information* based distance metrics. These metrics automatically yield pattern clusters that are characterized by the *maximization of the mutual information between patterns of the same cluster* and at the same time *the maximization of the statistical independence between genes of different clusters*. In addition, the main advantages of the KSDG-SOM model, i.e. the dynamic adaptable growing and the potentiality to account for the a priori functional information are still retained.

3 Mutual Information

The mutual information metric has the capacity to measure a general dependence among random variables. We utilize it in order to identify sets of patterns that more probably are associated. The outcome is that the extraction of fuzzy association rules is performed on a much smaller space making the final problem computationally tractable even for large pattern sets.

The *entropy* of a pattern is a measure of the uncertainty information content in that pattern. For a random vector $\boldsymbol{X}$ with probability distribution $P(\boldsymbol{X}=x_i), i=1,\ldots,K_x$ with K_x the number of possible values of $\boldsymbol{X}$, the *Shannon entropy* is defined

$$H(\boldsymbol{X})=-\sum_{i=1}^{K_x} P(\boldsymbol{X}=x_i)\cdot \ln P(\boldsymbol{X}=x_i)$$

Higher entropy for patterns imply more uniform distribution. Similarly the *joint entropy* of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is a measure of the total uncertainty contained in $\boldsymbol{X}$ and $\boldsymbol{Y}$. It is defined as

$$H(\boldsymbol{X},\boldsymbol{Y})=-\sum_{i=1}^{K_x}\sum_{j=1}^{K_y} P(\boldsymbol{X}=x_i,\boldsymbol{Y}=y_j)\cdot \ln P(\boldsymbol{X}=x_i,\boldsymbol{Y}=y_j)$$

where K_x, K_y is the number of possible values of $\boldsymbol{X}$ and $\boldsymbol{Y}$ respectively. The *mutual information* between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is a measure of information about $\boldsymbol{X}$ (or $\boldsymbol{Y}$) contained in $\boldsymbol{Y}$ (or $\boldsymbol{X}$). It is given by:

$$I(\boldsymbol{X},\boldsymbol{Y})=H(\boldsymbol{Y})-H(\boldsymbol{Y}|\boldsymbol{X})=H(\boldsymbol{X})+H(\boldsymbol{Y})-H(\boldsymbol{X},\boldsymbol{Y})=\sum_{i=1}^{N_x}\sum_{j=1}^{N_y} P(\boldsymbol{X}=x_i,\boldsymbol{Y}=y_i)\ln\frac{P(\boldsymbol{X}=x_i,\boldsymbol{Y}=y_j)}{P(\boldsymbol{X}=x_i)P(\boldsymbol{Y}=y_j)} \quad (1)$$

The probabilities of equation (1) can be estimated with the empirical counts from the N training patterns as

$$P(\boldsymbol{X}=x_i,\boldsymbol{Y}=y_i)\approx\frac{\#(x_i,y_j)}{N}$$

$$P(\boldsymbol{X}=x_i)\approx\frac{\#(x_i)}{N}$$

$$P(\boldsymbol{Y}=y_j)\approx\frac{\#(y_j)}{N}$$

The KSDG-SOM partitions the set G of patterns into k disjoint subsets as $G=X_1\cup X_2\cup\ldots\cup X_k$.

The cost function that the KSDG-SOM learning minimizes is defined as the sum of pair-wise mutual information between their formed clusters over all the possible combinations, i.e.,

$$\text{cost(Partition)}=\sum_{i\neq j} I(\boldsymbol{X_i},\boldsymbol{X_j})$$

where Partition denotes a particular partition scheme. The motivation of this cost function is to derive statistically independent clusters.

This optimization is performed with the KSDG-SOM

learning rules with the mutual information as the distance metric.These rules are described in detail in [38].
We should note that the optimization is performed by the dynamic growing algorithm efficiently but approximately. In addition the cost function can easily account for a supervised bias in order to tend keeping patterns with a priori similar functional classes together onto the same cluster.

4 Fuzzy Association Rules

For a data set $D=\{t_1,t_2,...,t_n\}$ with attributes A and fuzzy sets associated with each attribute, the purpose of fuzzy association is to detect interesting and potentially useful regularities. These regularities are expressed in terms of fuzzy association rules of the form:

if $P=\{a_1,...,a_n\}$ is $V=\{f_1,...,f_m\}$ then $P'=\{a'_1,...,a'_n\}$ is $V'=\{f'_1,...,f'_m\}$

where $f_i,f_i^{'}$ are fuzzy sets related to attributes a_i,a'_i respectively and P,P' are disjoint itemsets in the sense that they do not share common attributes. The purpose is to detect the interesting rules, i.e. those that have enough support and high confidence value.
The *Apriori* algorithm [1] computes *frequent itemsets* from a set of patterns by performing multiple iterations. Each such iteration involves:

- candidate generation
- candidate counting and selection

Exploring the knowledge about infrequent itemsets, obtained from the previous iterations, the algorithm prunes *a priori* those candidate itemsets that cannot become frequent. After discarding every candidate itemset that has an infrequent subset, the algorithm enters the candidate counting step.
However, the crisp association rules as those extracted by the *Apriori* algorithm, require a coarse discretization of the attribute ranges to a few discrete "items". For example for the gene expression data application presented, for each gene we use a three-level discretization of the possible range of values:

a) Low expression values (i.e. underexpressed genes).
b) Insensitive (i.e. genes not affected across experiments).
c) High expression values (i.e. overexpressed genes).

It is evident that the loss of information is significant. In order to represent intervals with non-sharp boundaries, we can utilize a fuzzy set representation of items for generating fuzzy association rules. The assignment of meaningful linguistic terms to the fuzzy sets makes these rules very informative to the human expert. For the example application of gene expression analysis, a quantitative gene expression value v_g is mapped to a three dimensional vector $[\mu_L(v_g),\mu_I(v_g),\mu_H(v_g)]$ whose components refer to the membership value at the corresponding linguistic term (i.e. *Low*, *Insensitive*, *High*). We used *triangular membership functions* that consider each gene as *Underexpressed*, *Insensitive* and *Overexpressed* according to the two-fold and four-fold changes in expression levels.
For example in Figure 1, the *Overexpressed* linguistic variable starts assuming non negative values at the two-fold expression level increase relative to the normal conditions and takes the value 1 at the four-fold change. Therefore, genes that increase their expression by four fold or more are considered *Overexpressed* to a degree of one.

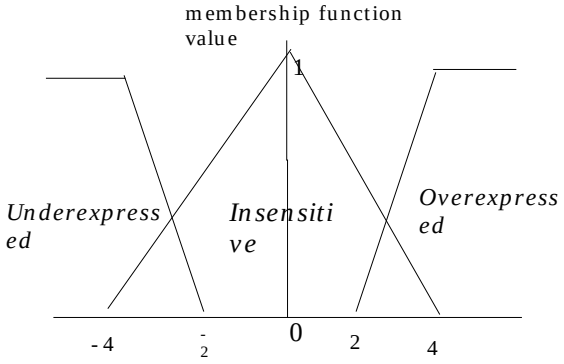

Figure 1 Illustration of the fuzzification of the gene expression values.

Suppose that we have N transactions each with n attribute (item) values, a set of membership functions, and a prespecified minimum support s and confidence c. The steps of the fuzzy association rule extraction algorithm are as follows:

1. Transformation of the quantitative values a_{ij} of each transaction $t^i,i=1,...,N$, for each attribute $a_j,j=1,...,n$, to its membership values at its corresponding linguistic partition. This is represented as

$$(\mu_{ij_1}(R_{j_1}), \mu_{ij_2}(R_{j_2}), \ldots, \mu_{ij_l}(R_{j_l}))$$

using the utilized membership functions. Here R_{j_k} denotes the k*th* linguistic variable (e.g. *Underexpressed*, *Insensitive*, *Overexpressed* etc.) of the attribute a_j, μ_{ij_k} is the fuzzy membership value for attribute a_j at the range R_{j_k}, and l is the number of fuzzy partitions used for the fuzzification of the quantitative values of the attribute a_j.

2. Computation of the *scalar cardinality* or *fuzzy count* for every region R_{j_k} of attribute a_j at the transaction data (consisting of *N* records) as $\text{count}_{j_k} = \sum_{i=1}^{N} \mu_{ij_k}$

3. Check whether count_{j_k} of each $R_{j_k}, j=1,\ldots,n$, is greater than or equal to the requirement for minimum support *s*, and if so, insert the item at the list of frequent 1-itemsets (F_1). Therefore $F_1 = \{R_{j_k} \mid \text{count}_{j_k} \geq s\}$ for $1 \leq j \leq n$.

   The meaning of the equation above is that we include the corresponding linguistic partition *k* for attribute *j*, at the frequent 1-itemsets if its fuzzy count count_{j_k} fullfils the minimum support requirement.

4. Set $r=1$, where *r* represents the number of items that are kept in the current frequent itemsets.

5. Generate the candidate set C_{r+1} from F_r in a manner similar to that of the *Apriori* algorithm.

6. For each newly formed *(r+1)*-itemset *R*, with items $(R_1, R_2, \ldots, R_{r+1})$ in C_{r+1}, do the following substeps.
   (a) Compute the fuzzy value of each transaction data t^i as $\mu_{iR} = \mu_{iR_1} \wedge \mu_{iR_2} \wedge \ldots \wedge \mu_{iR_{r+1}}$, where μ_{iR_j} is the membership value of t^i in region R_j.
   If the minimum operator is used for the intersection, we have

$$\mu_{iR} = \min_{j=1}^{r+1} \mu_{iR_j}$$

   At this point we should note that the fuzzy product operation is an alternative choice that is usually preferable since it better utilizes the available information:

$$\mu_{iR} = \prod_{j=1}^{r+1} \mu_{iR_j}$$

   (b) Computation of the scalar cardinality (i.e. fuzzy support) of *R* in the transaction data as

$$\text{count}_R = \sum_{i=1}^{N} \mu_{iR}$$

   (c) **if** count_R is larger than or equal to the support *s* **then** insert *R* in F_{r+1}

7. **if** F_{r+1} is **null then** go to the next step; **else** set $r=r+1$ and **repeat** steps 6-7.

8. Construct the fuzzy association rules for all the frequent *q*-itemsets *R* with items $(R_1, R_2, \ldots, R_q), q \geq 2$, using the following substeps:
   (a.) Form all possible association rules as

$$R_1 \wedge \ldots R_{k-1} \wedge R_{k+1} \wedge \ldots \wedge R_q \rightarrow R_k \quad \text{(1-itemset consequent)}$$

$$\wedge_{\text{for all i}} R_i \rightarrow \wedge_{\text{for all j}} R_j$$

$$i \in \text{Premise},\ j \in \text{Consequent}$$

$$(\text{k-itemset consequent}), k>1$$

   for $k=1,\ldots q$.
   (b.) Compute the confidence values) of all the fuzzy association rules as

$$\frac{\sum_{i=1}^{N} \mu_{iS}}{\sum_{i=1}^{N} (\mu_i \cdot R_1 \wedge \ldots \wedge \mu_i \cdot R_{k-1} \wedge \mu_i \cdot R_{k+1} \wedge \ldots \wedge \mu_i \cdot R_q)}$$

   The numerator of the former equation expresses the fuzzy support of the whole itemset *S* over each transaction i, μ_{iS}. The denominator corresponds to the fuzzy support of the *Premise* summed over all the transactions. Therefore, the equation quantifies the degree to which the fullfillment of the Premise condition associates with the fullfillment and of the *Consequent* condition. The expression for the *k*-itemset

consequent ($k>1$), is taken similarly by considering items of the premise of Equation 0 at the denominator of Equation 4.

9. Extract the rules with confidence values larger than or equal to the predefined confidence *c*.

We should note that we have based the fuzzy association rule extraction software on the implementation of the Apriori algorithm obtained from the WEKA [25] data mining packages implemented in the Java programming language. In addition to adapting the algorithm to the fuzzy case, we customized the code in order to extract the patterns for fuzzy association rules directly from the KSDG-SOM nodes.

5 Results and Discussion

This section first describes briefly the characteristics of the particular DNA microarray data analyzed. Then we proceed with the application of the KSDG-SOM for the analysis of these data and we discuss the extracted fuzzy association rules.

We have applied the KSDG-SOM to analyze microarray expression data from the budding yeast Saccharomyces cerevisiae. These data are public available from the Stanford web site. They were generated by studying this fully sequenced organism with microarrays, containing essentially every Open Reading Frame (ORF). The samples used were collected at various time points during the diauxic shift, the mitotic cell division cycle and sporulation. The data set consists of 80-element gene expression vectors for 6,221 genes.

The source of these profiles were eight different microarray experiments under different conditions. These conditions can be categorized into the following four types: 1. the mitotic cell division cycle, 2. sporulation, 3. temperature and reducing socks, 4. gene expression in the budding yeast during the diauxic shift.

For example, data for the last condition were obtained from [29]. With a fluorescence-ratio method, Derisi et. al. [29] measured the relative abundance of mRNA for the entire yeast genome, in yeast growing in a fresh medium to examine the changes in expression that take place with the metabolic shift from anaerobic to aerobic metabolism, with seven samples taken at 2-hour intervals. Measured levels of expression of genes in this experiment reflect metabolic reprogramming that occured during the diauxic shift.

Annotation for these genes was derived from the Functional Classification Catalogue of the Munich information center for protein sequences (MIPS) Comprehensive Yeast Genome Database (CYGD) available at *http://mips.gsf.de/proj/yeast/CYGD/db/index.html* Specifically, we present 5 from the 19 t*op-level functional categories* that include a total of 1974 genes:

1. Cell Fate (423 ORFs)
2. Cell Rescue, Defense And Virulence (273 ORFs)
3. Cellular Communication/Signal Transduction Mechanism (59 ORFs)
4. Cellular Transport And Transport Mechanisms (480 ORFs)
5. Metabolism (1059 ORFs)

The gene expression data is arranged in a table whose rows correspond to the genes and columns to the individual log-transformed gene expression values of each gene in a particular experimental condition represented by the column. The weighted K-nearest neighbors imputation method presented in [30] is applied in order to fill up systematically the missing values. This data imputation approach detects the *K* most similar genes in expression to the one with missing values and estimates them by weighting the values of these genes at the same columns with their similarity.

The format of the extracted rules displays the genes involved at the *Premise* and those at the *Consequence*. Therefore by examining them we can obtain evidence on some possible gene regulation relations at the presented application. However these rules should be further elaborated. Multiple conditions either at the Premise or the Precondition are in an implied conjunctive form. Next to the premise we present the *fuzzy Support Count*, i.e. the sum over the number of experimental conditions that support the association weighted by the degree of this support.

Finally, the confidence of the rule is displayed next to the rule. For example, at the rule

YAL001C=High and YAL003W=High ==> YAL002W=High conf:(1)

the overexpression of the genes YAL001C and YAL003W is associated with an overexpression of the gene YAL002W. This pattern is observed in at least 40 of the total 80 experimental conditions since we require a minimum fuzzy support of 40. Also the confidence is 1, i.e. for all the conditions for which YAL001C and YAL003W are both overexpressed, we observe also an overexpression of YAL002W.

Similarly at the rules:

YAL001C=Low ==> YAL003W=Low
YAL004W=Low conf:(1)
YAL003W=Low ==> YAL001C=Low
YAL004W=Low conf:(1)
YAL004W=Low ==> YAL001C=Low
YAL003W=Low conf:(1)

we observe that the underexpression of the corresponding genes at the Premises (e.g. YAL001C, YAL003W and YAL004W) is closely associated with the underexpression of two genes at the Consequences.

The KSDG-SOM growing phase, clusters together genes according to the mutual information metric, while at the same time it penalizes a large number of genes in order to avoid large cluster sizes. Therefore the search over large spaces for association rules is avoided and the standard implementation of the fuzzy Apriori algorithm with a provision for hashing items for fast access works very effectively. In particular, we have extracted rules for the whole 6221 gene set of yeast within a few minutes on a Pentium 4 machine.
The KSDG-SOM has also the provision for incorporating a priori functional classes. This advantage can be further explored for the restriction of the search space for the detection of fuzzy association rules. A comparison with the number of fuzzy association rules found with the Pearson Correlation metric reveals that the mutual information metric reveals more association rules. Although, we have not yet performed a detailed study, it seems there is a considerable overlap between rules extracted according to the mutual information metric and the Pearson correlation one, but the former succeeds at the discovery of about 20% more rules, that are probably characterized by *nonlinear associations* and thus remain hidden to the linear statistical dependence analysis tools (e.g. the Pearson correlation).
In addition, in order to consolidate the validity of our results, we constructed 'randomized' data sets, consisted of all the expression values for each transcript in the original data set being shifted with respect to the values of the other transcripts by a random number of experiments. The objective of using this randomized data set, is to study how many association rules would be produced in a data set with the same values as the original data set, but in which the data dependencies are destroyed by the randomization. We observed that at the randomized data, no association rules are produced with a large fuzzy support value (more than 60.0).

6 Conclusions

We have presented an approach to confront effectively the difficult computational problem of the extraction of fuzzy association rules for the description of pattern dependencies and interactions from large transaction data.
The association rules that we have discovered represent clearly a fraction of all the possible pattern-to-pattern interactions. However, the rules that we have mined, represent a considerable number of non-random patterns of interest. From those rules, new hypotheses can be stated that could ultimately be confirmed or rejected on the basis of specialized experiments for each application domain.
We described a two stage approach to the problem that obtains computationally manageable plausible solutions. The first stage clusters patterns that more probably are associated. Therefore our approach integrates the clustering machinery with that of the fuzzy association rule extraction. Thereafter, the second stage, the fuzzy association rule extraction algorithm follows, confronting a significantly reduced problem.
The clustering phase is accomplished by means of a Kernel Supervised Dynamic Grid Self-Organized Map (KSDG-SOM). We adapted the criteria for dynamic expansion, in order to obtain clusters of manageable size for the association rule extraction algorithms. The mutual information metric controls the development of the KSDG-SOM clusters. This metric allows the formation of pattern clusters that maximize the mutual information for patterns of the same cluster and to minimize it between different clusters.
In addition the KSDG-SOM is capable of incorporating a priori information for the known functional characteristics of patterns. This supervised bias on training can focus the model at the detection of more appropriate rules that exploit domain knowledge.
After this initial pattern clustering we concetrate on whether a pattern can be explained properly by means of the patterns of the same node. Fuzzy association rules are extracted for the patterns allocated at the same node.
An important criterion that we impose for the KSDG-SOM growing phase, is the minimization of the number of patterns in each cluster in order to restrict the search over large spaces. Thereafter, the mutual information metric with the provision for incorporating a priori functional classes constructs highly appropriate clusters for the detection of fuzzy

association rules. A comparison with the number of fuzzy association rules found with the Pearson Correlation metric reveals that the mutual information metric reveals more association rules.
Clearly, this "line" of research requires much more work. Also, the current work will be extended by considered more effective computationally algorithms for association rule extraction, e.g. the EClat algorithm [34] and by further tuning the interface between the mutual information clustering and the fuzzy association rule extraction machinery.

Acknowlegment
The authors wish to thank the Greek Ministry of Education, EPEAK administration office, with the support of this research through contract 86372, 04-3-001/5, *"Arximidis: Support of Research Groups of TEI Kavalas"* project: "Computational Intelligence techniques for the analysis of Gene Expression Data".

References

[1] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases", in *Proceedings of 20th International Conferenence on Very Large Databases*, pp. 478-499, September 1994
[2] Brown Michael P.S., Grundy William Noble, Lin David, Cristianini Nello, Sugnet Charles Walsh, Furey Terrence S., Ares Manuel, Haussler Jr., David, "Knowledge-based Analysis of Microarray Gene Expression Data By Using Support Vector Machines", Proceedings of the National Academy of Science, Vol 97, No 1, pp. 262-267, 1997
[3] H. Liu, L. Wong, "Data Mining Tools for Biological Sequences", Journal of Bioinformatics and Computational Biology, Vol. 1, No. 1, p. 139-168, April 2003
[4] Eisen Michael B., Spellman Paul T., Patrick O. Brown, and David Botstein,"Cluster analysis and display of genome-wide expression patterns", Proc. Natl. Acad. Sci. USA, Vol. 95, pp. 14863-14868, December 1998
[5] Mavroudi Seferina, Papadimitriou Stergios, Bezerianos Anastasios, "Gene Expression Analysis with a Dynamically ExtendedSelf-Organized Map that Exploits Class Information",
Bioinformatics, Vol. 18, no 11, 2002, p 1446-1453
[6] Papadimitriou S., Mavroudi S., Vladutu L., Bezerianos A.,"Ischemia Detection with a Self Organizing Map Supplemented by Supervised Learning",IEEE Trans. On Neural Networks, Vol. 12, No. 3, May 2001, p. 503-515
[7] V. Olman, D. Xu., Y. Hu, "CUBIC: Identification of Regulatory Binding Sites Through Data Clustering",Journal of Bioinformatics and Computational Biology", p. 21-40
[8] Herrero Javier, Valencia Alfonso, and Dopazo Joaquin, "A hierarchical unsupervised growing neural network for clustering gene expression patterns", Bioinformatics, (2001) Vol. 17, no. 2, pp. 126-136
[9] Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S.,Dmitrovsky, E., Lander, E.S. and Golub, T.R. (1999)"Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation", Proc. Natl. Acad. Sci., USA, 92, pp. 2907-2912
[10] Friedman, N., M. Linial, I. Nachman, and D'Peier, "Using Bayesian networks to analyze expression data", J. Comp. Bio. 7, 2000, 601-620,
[11] Fritzke Bernd, "Growing Grid - a self organizing network with constant neighborhood range and adaptation strength", Neural Processing Letters, Vol. 2, No. 5, pp. 9-13, 1995
[12] Van Hulle, N.M., "Kernel-Based Topographic Map Formation", Neural Computation,Vol. 14, No 7, p. 1560-1573, 2002
[13] Van Hulle, N.M., "Kernel-based equiprobabilistic topographic map formation, Neural Computation, Vol. 10, No. 7, p. 1847-1871, 2002
[14] Vesanto Juha Alhoniemi, Esa, "Clustering of the Self-Organized Map", IEEE Transactions on Neural Networks, Vol. 11, No. 3, May 2000, p. 586-600
[15] Si J., Lin S., Vuong M. A., "Dynamic topology representing networks", Neural Networks, 13, pp. 617-627, 2000
[16] Cheng Guojian and Zell Andreas, "Externally Growing Cell Structures for Data Evaluation of Chemical Gas Sensors", Neural Computing & Applications, 10, pp. 89-97, Springer-Verlag, 2001
[17] Campos Marcos M., Carpenter Gail A., "S-TREE: self-organizing trees for data clustering and online vector quantization", Neural Networks 14 (2001), pp. 505-525
[18] Van Hulle, N.M., "Joint Entropy Maximization in Kernel-Based Topographic Maps", Neural Computation, Vol. 14, No 8, p. 1887-1906, 2002
[19] Azuaje Franscisco, "A Computational Neural Approach to Support the Discovery of Gene Function and Classes of Cancer", IEEE Trans. Biomed. Eng., Vol 48, No. 3, March 2001, pp 332-339
[20] James R. Williamson, "Self-Organization of Topographic Mixture Networks Using Attentional Feedback", Neural Computation 13:563-593, 2001
[21] Janne Sinkkonen, Samuel Kaski, "Clustering Based on Conditional Distributions inan Auxiliary Space", Neural Computation, 14:217-239, 2001
[22] A. Sierra, F. Corbacho, "Reclassification as Supervised Clustering",Neural Computation 12:2537-2546, 2000

[23] Cheung Vivian G., Morley Michael, Aguilar Francisco, Massimi Aldo, Kucherlapati Raju, Childs Geoffrey, "Making and reading microarrays", Nature genetics supplement, Vol. 21, January 1999
[24] Bishop, C. M., Svensen, M., Williams, C. K., "GTM: The generative topographic mapping", Neural Computation, 10:215- 234, 1998
[25] Ian H. Witten, Eibe Frank, *Data Mining*, Morgan Kaufmann Publishers, 2000
[26] Alahakoon Damminda, Halgamuge Saman K., Srinivasan Bala, "Dynamic Self- Organizing Maps with Controlled Growth for Knowledge Discovery", IEEE Transactions On Neural Networks, Vol. 11, No. 3, pp 601- 614, May 2000.
[27] Kohonen T., Self- Organized Maps, Springer- Verlag, Second Edition,1997.
[28] Haykin S,, Neural Networks, Prentice Hall International, Second Edition, 1999.
[29] DeRisi, J. L., Iyer, V. R, and Brown, P.O, "Exploring the metabolic and genetic control of gene expression on a genomic scale", Science, 278, p. 680- 686, 1997
[30] Troyanskaya Olga, Cantor Michael, Shelock Gavin, Brown Pat, Hastie Trevor, Tibshirani Robert, Botstein David, Altman Russ B.,"Missing value estimation methods for DNA microarrays", Bioinformatics, Vol. 17, no 6, 2001
[31] Lawrence Hunter, Ronald C. Taylor, Sonia M. Leach and Richard Simon, "GEST: a gene expression search tool based on a novel Bayesian similarity metric", Bioinformatics, Vol. 17, Suppl. 1, p. S115- S122
[32] R. D' Haeseleer, X. Wen, S. Fuhrman, R. Somogyi, "Linear Modeling of mRNA expression levels during CNS development and injury", Pacific Symposium on Biocomputing, Vol. 4, 1999, pp. 41- 52
[33] Xiaobo Zhou, Xiaodong Wang, Edward R. Dougherts, "Construction of genomic networks using mutual information clustering and reversible- jump Markov- chain- Monte- Carlo predictor design",Signal Processing 83 (2003), p. 745- 761
[34] Sushmita Mitra, Tinku Acharya, "Data Mining: Multimedia, Soft Computing, and Bioinformatics", John Wiley & Sons, 2003
[35] Bernhard Scholkopf, Alexander J. Smola, "Learning with Kernels: Support Vector Machines,Regularization and Beyond", MIT Press 2002
[36] Chad Creighton, Samir Hanash, "Mining gene expression databases for association rules", Bioinformatics, Vol. 19, no. 1, 2003, pp. 79- 86
[37] Seiya Imoto, Sunyong Kim, Takao Goto, Satoru Miyano, "Bayesian Network and Nonparametric heteroscedastic regression for nonlinear modeling of genetic network",Journal of Bioinformatics and Computational Biology, Vol. 1, No. 2, (2003), 231- 252
[38] S. Papadimitriou, S.D. Likothanassis, "Kernel- Based Self- Organized Maps trained with Supervised Bias for Gene Expression Data Analysis", in print to Journal of Bioinformatics and Computational Biology, Journal of Bioinformatics and Computational Biology, Imperial College Press, January 2004 (in print)
[39] B. Scholkopf, A. J. Smola, R. C. Williamson, P. L. Bartlett, "New support vector algorithms", Neural Computation:2000