



Application of Datamining and Soft Computing in Bio informatics

¹N. Nandakumar, Dr. ^{2*}T. Balasubramaniyan

¹Sri Vidya Mandir Arts and Science college, Uthangarai, Krishnagiri (Dt), Tamil Nadu, India.

^{4*}Head of Computer Science, Sri Vidya Mandir Arts and Science College, Katteri, Uthangarai.

Abstract

The explosive growth of biological information generated by the scientific community all over the world has led to storage of voluminous data. This torrent of information has, in turn, led to an supreme requirement for computerized databases to store, categorize, and index the data and for specialized tools to view and analyze the data. Bio informatics is the analysis of biological information using computers and statistical techniques; the science of developing and utilizing computer databases and algorithms to accelerate and enhance biological research. Computers are used together, store, analyze and integrate biological and genetic information which can then be applied to gene-based drug discovery and development. Datamining or knowledge discovery from data (KDD), is used to extract interesting, nontrivial, implicit, previously unknown and potentially useful information from data. Soft computing differs from conventional (hard) computing in that, unlike hard computing, it is tolerant of imprecision, uncertainty, partial truth, and approximation. The guiding principle of soft computing is to exploit the tolerance for imprecision, uncertainty, partial truth, and approximation to achieve tractability, robustness and low solution cost for the emerging field of conceptual intelligence. Soft computing includes the techniques such as fuzzy logic, neural networks, genetic algorithms, etc. The methodologies of soft computing are complementary rather than competitive and they can be viewed as a foundation component. This paper will focus on issues related to data mining and soft computing and relevance of these in bio-environmental. Further the paper focuses on some of its current applications.

Keyword: Datamining, Soft Computing, Bioinformatics, Fuzzy Logic, Neural Networks, Genetic Algorithm

© 2016 Sri Vidya Mandir Arts & Science College, Uthangarai

I. INTRODUCTION

Bioinformatics, is a field committed to the interpretation and analysis of biological data using computational techniques, has evolved tremendously in recent years due to the explosive growth of biological information generated by the scientific community.

Bioinformatics is the science of managing, mining, integrating, and interpreting information from biological data at the genomic, proteomic, phylogenetic, cellular, or whole organism levels. The need for bioinformatics tools and expertise has increased as genome sequencing projects have resulted in an exponential growth in

*Corresponding author. Tel: +91 9486130457, E-mail address: balaeswar123@gmail.com

© 2016 Sri Vidya Mandir Arts & Science College, Uthangarai

complete and partial sequence databases. Data mining is the use of automated data analysis techniques to uncover previously undetected relationships among data items. Data mining often involves the analysis of data stored in a data warehouse. Data mining aids the scientists and the researchers by providing sophisticated techniques to extract the useful information from the huge amount of biological data at hand. Data mining refers to the extraction of useful information from a large set of data. It is a technique for the discovery of patterns hidden in large data sets, focusing on issues relating to their feasibility, usefulness, effectiveness, and scalability[2]. The term data mining refers to information elicitation. On the other hand, soft computing deals with information processing. If these two key properties can be combined in a constructive way, then this information can effectively be used for knowledge discovery in large databases.

Referring to this synergistic combination, the basic merits of data mining and soft computing paradigms are pointed out and novel data mining implementation coupled to a soft computing approach for knowledge discovery is presented. In this context, in the following two sections the properties of data mining and machine learning paradigms are pointed out. The present article provides an overview of the available literature on data mining, and its aspects. This is followed by Section-2 which discusses the state of art of soft computing and we discuss each of the soft computing methods in brief the following by section-4 which brings out the relevance of different soft computing methods in data mining. Finally we conclude with section -4, where the significance of soft computing in data mining is highlighted.[2][4]. The purpose of this paper is to provide an overall understanding of Data mining and soft computing techniques and their application and usage in bioinformatics.

II. Bioinformatics:

Bioinformatics is the field of science in which biology, computer science, and information technology merge to form a single discipline. The ultimate goal of the field is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be discerned. At the beginning of the "genomic revolution", a bioinformatics concern was the creation and maintenance

of a database to store biological information, such as nucleotide and amino acid sequences. Development of this type of database involved not only design issues but the development of complex interfaces whereby researchers could both access existing data as well as submit new or revised data. Ultimately, however, all of this information must be combined to form a comprehensive picture of normal cellular activities so that researchers may study how these activities are altered in different disease states. Therefore, the field of bioinformatics has evolved such that the most pressing task now involves the analysis and interpretation of various types of data, including nucleotide and amino acid sequences, protein domains, and protein structures. The actual process of analyzing and interpreting data is referred to as computational biology. Important sub-disciplines within bioinformatics and computational biology include the development and implementation of tools that enable efficient access to, and use and management of, various types of information; the development of new algorithms (mathematical formulas) and statistics with which to assess relationships among members of large data sets, such as methods to locate a gene within a sequence, predict protein structure and/or function, and cluster protein sequences into families of related sequences. The National Center for Biotechnology Information (NCBI 2001) defines bioinformatics as: "Bioinformatics is the field of science in which biology, computer science, and information technology merge into a single discipline. There are three important sub-disciplines within bioinformatics: the development of new algorithms and statistics with which to assess relationships among members of large data sets; the analysis and interpretation of various types of data including nucleotide and amino acid sequences, protein domains, and protein structures; and the development and implementation of tools that enable efficient access and management of different types of information." [ref-3]. The three terms bioinformatics, computational biology and bioinformatics infrastructure are of times used interchangeably. These three may be defined as follows:

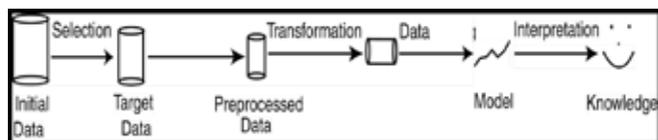
1. Bioinformatics refers to database-like activities, involving persistent sets of data that are maintained in a consistent state over essentially indefinite periods of time;

2. Computational biology encompasses the use of algorithmic tools to facilitate biological analysis.
3. Bioinformation infrastructure comprises the entire collective of information management systems, analysis tools and communication networks supporting biology. Thus, the latter may be viewed as a computational gibbet of the former two.

A. Biological Database

A biological database is a large, organized body of persistent data, usually associated with computerized software designed to update, query, and retrieve components of the data stored within the system. A simple database might be a single file containing many records, each of which includes the same set of information. For example, a record associated with a nucleotide sequenced at a base typically contains information such as contact name, the input sequence with a description of the type of molecule, the scientific name of the source organism from which it was isolated, and often, literature citations associated with the sequence. For researchers to benefit from the toxicodermatitides, two additional requirements must be met: easy access to the information and a method for extracting only that information needed to answer specific biological questions.

Datamining task : Figure:1



B. Importance of Bioinformatics

The rationale for applying computational approaches to facilitate the understanding of various biological processes includes: a more global perspective in experimental design, the ability to capitalize on the emerging technology of database-mining - the process by which testable hypotheses are generated regarding the function or structure of a gene or protein of interest by identifying similar sequences in better characterized organisms.

C. SCOPE AND USE OF BIOINFORMATICS:

Bioinformatics is used in analyzing genomes, protein sequences, three-dimensional modelling of bio molecules and biologic systems, etc. Different biological problems considered within the scope of bioinformatics fall into main tasks which are given below:

Alignment and comparison of DNA, RNA, and protein sequences.

Gene finding and promoter identification from DNA sequences.

Interpretation of gene expression and micro-array data.

Gene regulatory network identification.

Construction of phylogenetic trees for studying evolutionary relationships.

Protein structure prediction and classification.

Molecular design and molecular docking.

Therefore the aims of bioinformatics are:

- To organize data in a way that allows researchers to create and access information.
- To develop tools that facilitate the analysis and management of data.
- To use biological data to analyze and interpret the results in a biologically meaningful manner.

III. DATAMINING

Data mining in general is a term which refers to extracting or "mining" knowledge from huge amounts of records. Data Mining (DM) is the discipline of finding novel, remarkable patterns and liaison in vast quantities of data. Data mining is also sometimes called Knowledge Discovery in Databases (KDD)[4]. Data mining is not specific to any industry; it has its presence in most of the applications today. It requires sharp technologies and the compliance to discover the possibility of hidden knowledge that resides in the data. Data Mining approaches seem ideally suited for Bioinformatics, since it is data-rich, but lacks a comprehensive theory of life's

organization at the molecular level. The extensive data bases of biological information crafts both challenges and opportunities for development of novel KDD methods. Mining biological data helps to extract useful knowledge from massive data sets granduncles, and in other related life sciences are as such as medicine and neuroscience, etc. Knowledge discovery as a process is depicted in Figure 1 and consists of an iteratives equence of the following steps:

Selection: Obtain data from various sources.

Preprocessing: Cleanse data.

Transformation : Convert to common format.

Transform to new format.

Data Mining: Obtain desired results.

Interpretation/Evaluation: Present results to user in meaningful manner.

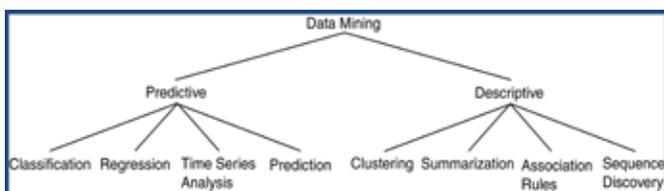


Figure 2:

A. Dataminingtasks:

The two "high-level" primary goals of data mining, in practice, are prediction and description. The main tasks well apt ford at a mining, all of which involves mining meaningful new patterns from the data, are: Learning from data falls into two categories: directed ("supervised") and undirected ("unsupervised") learning given in figure 2. The first three tasks– classification, estimation and prediction– are relationship among all the variables. Unsupervised learning attempts to find patterns with out the use of a particular target field. The development of new data mining and knowledged is covery tools is a subject of active research. One motivation behind the development of these tools is their potential application in modern biology.

Classification

Classification is learning a function that maps (classifies) a data item into one of several predefined

classes. 1. Classification[18], [19], [20], [21], [22]: classifies ad at a item into one of several predefined categorical classes. A classification problem is a supervised learning problem in which the output information is a discrete classification; i.e., given an object and its input attributes, the classification output is one oft he possible mutually exclusive classes of th ,[10],[11],[12]: maps ad at a item to areal valued prediction variable. Aureteroneopyelostomies supervised learning problem of building am ore or examples of supervised learning. The nextthree tasks – association rules, clustering and description&visualization – are examples of unsupervised learning. In unsupervised learning, no variable is singled out as the target; the goal is to establish some less transparent model in which the output information is a continuous numerical value ora vector of such values rather than a discrete class. Then, given an object, it is possible to periodontal its attributes by means oft he other attributes by using the built model. The prediction of numeric values maybe done by classical or more advanced statistical methods and by symbolic methods often used in the classification task.

Estimation

Given some input data, coming up with a value for some unknown continuous variable.

Prediction

Same as classification & estimation except that the records are classified according to some future be ha vi or ores ti mated future value).

Association rules

Determining which things go together, also called dependency mode ling. association rules[13], [14], [15],[16]: describes association relationship among different attributes. Given a collection of items and a set of records, each of which contain some number of items from the given collection, an association function is an operation against this set of records which return affinities or patterns that exist among the collection of items. These patterns can be expressed by rules such as "72% of all there cords that contain items A, B and C also contain items D and E." Association rule generators area powerful data mining technique used to search through an entire data set for rules revealing the nature

and frequency of relationships or association between data entities. The resulting associations can be used to filter the information for human analysis and possibly to define a prediction model based on observed behaviour.

Clustering

Segmenting a population into a number of subgroups or clusters. Clustering [17],[18],[19],[20],[21],[22],[23],[24]: maps an item into one of several clusters, where clusters are natural groupings of data items based on similarity metrics or probability density models

Rule generation

[25],[26],[27],[28],[29],[30],[31],[32]: extracts classification rules from the data. A clustering problem is an unsupervised learning problem that aims at finding clusters of similar objects sharing a number of interesting properties. It may be used in data mining to evaluate similarities among data, to bloodstained representative prototypes, to analyze correlations between attributes, or to automatically represent a data set by as small number of regions, preserving the topological properties of the original input space

Summarization

[33],[34],[35],[36]: provides a compact description for a subset of data. This task aims at producing compact and characteristic descriptions for a given set of data. It can take multiple forms: numerical (simple descriptive statistical measures like means, standard deviations), graphical forms (histograms, scatter plots), or the form of if-then rules. It may provide descriptions about objects in the whole database or in selected subsets of it. An example of summarization is "the minimum unitary price for all the transactions with energy is 70 price units"

Sequence analysis

[37],[38]: models sequential patterns, like time-series analysis. The goal is to model the states of the process generating the sequence or to extract and report deviation and trends overtime.

Dependency modelling

[37],[38]: describes significant dependencies among variables.

Description & visualization:

Representing the data using visualization techniques.

B.

C. Major Issues in Data Mining

The following are the major issues related with the usage of the data mining function activities:

1. Mining methodology and user interaction

Mining different kinds of knowledge in data bases. Interactive mining of knowledge at multiple levels of abstraction.

2. Performance Scalability

Efficiency and scalability of data mining algorithms. Parallel, distributed and incremental mining methods

3. Other Issues in Data Mining

Issues relating to the diversity of data types. Handling relational and complex types of data. Mining information from heterogeneous databases and global information systems (WWW). Issues related to applications and social impacts. Application of discovered knowledge. domain-specific data mining tools. Intelligent query answering. protection of data security, integrity, and privacy. Integration of the discovered knowledge with existing knowledge:

A knowledge fusion problem.

Data Mining techniques have been applied successfully in many areas from business to science to sports.

1. Science applications

Data mining techniques have been used in astronomy, bioinformatics, drug discovery and many more.

2. Business applications

Many organizations now employ data mining as a secret weapon to keep in pace or gain a competitive edge. Data mining has been used in advertising, CRM (Customer Relationship management), investments, manufacturing, sports / entertainment, telecom, e-Commerce, targeted marketing, health care, etc.

3. Other application

Data mining has been successfully used for various other applications such as Web : search engines, Government, law enforcement, profiling tax cheaters, anti-terror, credit approval, etc. Putting it together Data Mining is the step in the process of knowledge discovery in databases, that inputs predominantly cleaned, transformed data, searches the data using algorithms, and outputs patterns and relationships to the interpretation/evaluation step of the KDD process

IV Softcomputing:

Soft computing is a consortium of methodologies that works synergistically and provides, in one form or another, flexible information processing capability [10]. Its aim is to exploit the tolerance for imprecision, uncertainty, approximate reasoning, and partial truth in order to achieve tractability, robustness, and low-cost solutions [11]. Methodologies like fuzzy sets, neural networks, genetic algorithms, and rough sets are most widely applied in the data. An excellent survey demonstrating the significance of soft computing tools in data mining problem is provided by Mitra et al. [40]. These methodologies of soft computing are complementary rather than competitive and they can be viewed as a foundation

A. Importance of Soft Computing

The complementarity of fuzzy logic, neural networks, genetic algorithms and probabilistic reasoning has an important consequence in many cases. A problem can be solved most effectively by using fuzzy logic, neural networks, genetic algorithms and probabilistic reasoning in combination rather than using them exclusively. A typical example for such combination is neuro fuzzy systems.

Fuzzy Logic

Fuzzy logic is derived from fuzzy set theory dealing with reasoning that is approximate rather than precisely deduced from classical predicate logic. It can be thought of as the application side of fuzzy set. Fuzzy truth represents membership in vaguely defined sets, like likelihood of some event or condition and fuzzy sets are based on vague definitions of sets, not randomness [39]. Since, data mining is the process of extracting nontrivial relationships in the database, the

association that are qualitative are very difficult to utilize effectively by applying conventional rule induction algorithms. Since fuzzy logic model in a probability based modeling, it has many advantages over the conventional rule induction algorithms. The advantage is that it allows processing of very large data sets which require efficient algorithms. [ref book] The concept of fuzzy logic was conceived by Lotfi Zadeh. Fuzzy logic is an organized method for dealing with imprecise data. This data is considered to be as fuzzy sets [2]. Fig. 2 shows values for the continuous attribute income are mapped into the discrete categories low, medium, high, as well as how the fuzzy membership or truth values are calculated. Fuzzy logic systems typically provide graphical tools to assist users in this step.

Fuzzy logic can be introduced into the system to allow fuzzy thresholds or boundaries to be defined. Rather than having a precise cutoffs between categories or sets, fuzzy logic uses truth values between 0:0 and 1:0 to represent the degree of membership that a certain value has in a given category [2]. In general, the use of fuzzy logic in rule-based systems involves the following:

- 1) Attribute values are converted to fuzzy values. Above figure shows values for the continuous attribute income are mapped into the discrete categories low, medium, high, as well as how the fuzzy membership or truth values are calculated. Fuzzy logic systems typically provide graphical tools to assist users in this step.
- 2) For a given new sample, more than one fuzzy rule may apply. Each applicable rule contributes a vote for membership in the categories. Typically, the truth values for each predicted category are summed.
- 3) The sums obtained above are combined into a value that is returned by the system. This process may be done by weight in each category by its truth sum and multiplying by the mean truth value of each category. The calculations involved may be more complex, depending on the complexity of the fuzzy membership graphs.

1. Role of Fuzzy Sets

Modeling of imprecise/qualitative knowledge. Transmission and handling uncertainties at various stages. Supporting, to an extent, human type reasoning in natural form.

2. Fuzzy Sets in Data Mining

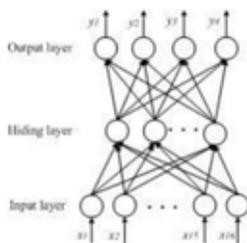
Classification/Regression/ Clustering. Discovering association rules (describing interesting association relationship among different attributes). Data summarization/condensation (abstracting the essence from a large amount of information).

Neural Networks

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain processes information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly inter connected processing elements (neurons) working in union to solve specific problems. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons.

NEURAL NETWORKS

An artificial neural network (ANN), [2][1], [4][1] often just called a "neural network" (NN), is a mathematical model or computational model based on biological neural networks, in other words, is an emulation of biological neural system. It consists of an inter connected group of artificial neurons and processes information using a connectionist approach to computation. The human brain is a highly complex structure viewed as a massive, highly inter connected network of simple processing elements called neurons. This behaviour of the neuron can be captured by a simple model which can be shown as:



In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the noteworthy component of the model bears a direct analogy to the actual constituents of a biological neuron and hence is termed as ANN. There are several classes of Neural Networks, classified according to their learning mechanisms.

- Single layer feed forward network
- Multi layer feed forward network:
- Recurrent networks:

During the learning phase

Role of ANN

Adaptively, robustness, parallelism, optimality. Machinery for learning and curve fitting (Learns from examples). Initially, thought to be unsuitable for - black box nature - no information available in symbolic form (suitable for human interpretation). Recently, embedded knowledge is extracted in the form of symbolic rules making it suitable for rule generation.

ANNs provide Natural Classifiers/prediction based models having

Crossover and mutation to produce a population of the next generation. This concept of survival of the fittest proposed by Darwinism has a in cause for the robust performance of GAs. Cross over helps in the exchange of discovered knowledge in the form of genes between individuals and mutation helps in restoring lost or unexplored regions in search space. [ref book] Genetic algorithm (GA) [1], belonging to a class of randomized heuristic and adaptive search techniques based on the principle of natural selection, is an attractive tool to find near optimal solutions for optimization problems. Genetic algorithms [2][1] attempt to incorporate ideas of natural evolution. In general, genetic learning starts as follows. An initial population is created consisting of randomly generated rules. Each rule can be represented by a string of bits. As a simple example, suppose that samples in a given training set are described by two Boolean attributes, A1 and A2, and that there are two classes, C1 and C2. The rule "IF A1 and not A2 THEN C2" can be coded as the bit string "100", where the two left most bits represent attributes A1 and A2,

respectively, and the rightmost bit represents the class. Similarly, the rule "if not A1 and not A2 then C1" can be coded as "001". If an attribute has k values where $k > 2$, then k bits may be used to encode the attribute's values. Classes can be encoded in a similar fashion. Based on the notion of

Genetic Algorithms

Genetic Algorithms have found a wide gamut of applications in data mining, where knowledge is mined from large databases. Genetic algorithms can be used to build effective classifier systems, mining association rules and other such data mining problems. Their robust search technique has given them a central place in the field of data mining and machine learning [42]. GA can be viewed as an evolutionary process where at each generation, from a set of offspring solutions, individuals or solutions are selected such that individuals with higher fitness have greater probability of getting chosen. At each generation, these chosen individuals undergo P evolution where each rule in P satisfies a prespecified fitness threshold. Genetic algorithms are easily parallelizable and have been used for classification as well as other optimization problems. In data mining, they may be used to evaluate the fitness of other algorithms.

1. Role of Genetic Algorithms

Robust, parallel, adaptive search methods - suitable when the search space is large. GAs: Efficient, Adaptive and robust Search Processes, Producing near optimal solutions Application of neural networks to biological data mining: A case study in protein sequence classification. In Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2000, pp. 305 - 309. neural networks method add definite advantage to bioinformaticians to solve many types of problems in the field of bioinformatics.

References:

[1] Data Mining Concepts and Techniques - Jiawei Han, Micheline Kamber
 [2] Principles of Soft Computing - S.N. Sivanandam and S.N. Deepa
 [3] bio ref in II

[4] K.R. Venugopal, K.G. Srinivasa and L.M. Patnaik
 soft computing for data Mining Application
 [5] T.M. Mitchell, "Machine learning and data mining,"
 Commun. ACM, vol. 42, no. 11, 1999.
 [6] U. Fayyad, G.P. Shapiro, and P. Smyth, "The KDD
 process for extracting useful knowledge from
 volumes of data," Commun. ACM, vol. 39, pp. 27-
 34, 1996.
 [7] neural networks, Fuzzy Logic and Genetic
 Algorithm Synthesis and Applications -
 S. Rajasekaran, G.A. Vijayalakshmi Pai.
 [8] "Bioinformatics World" magazine :
 www.bioinformaticsworld.info [21]
 [9]
 [10] V. Ciesielski and G. Palstra, "Using a hybrid neural/
 expert system for data base mining in market survey
 data," in Proc. Second International Conference on
 Knowledge Discovery and Data Mining (KDD-96),
 (Portland, OR), p. 38, AAAI Press, Aug. 2-4, 1996.
 [11] K. Xu, Z. Wang, and K.S. Leung, "Using a new
 type of non linear integral for multi-regression:
 an application of evolutionary algorithms in data
 mining," in Proceedings of IEEE International
 Conference on Systems, Man, and Cybernetics,
 (San Diego, CA), pp. 2326-2331, October 1998.
 [12] E. Noda, A.A. Freitas, and H.S. Lopes, "Discovering
 interesting prediction rules with a genetic
 algorithm," in Proceedings of IEEE Congress on
 Evolutionary Computation CEC 99, (Washington
 DC), pp. 1322-1329, July 1999.
 [13] R. Agrawal, T. Imielinski, and A. Swami,
 "Mining association rules between sets of
 items in large databases," in Proceedings of
 1993 ACM SIGMOD International Conference
 on Management of Data, (Washington D.C.),
 pp. 207-216, May 1993 [14] Q. Wei and G. Chen,
 "Mining generalized association rules with
 fuzzy taxonomic structures," in Proceedings of
 NAFIPS 99, (New York), pp. 477-481, June 1999.

- [15] W.H.Au and K. C.C.Chan, "An effective algorithm for discovering fuzzy rules in relational databases," in Proceedings of IEEE International Conference on Fuzzy Systems FUZZIEEE 98, (Alaska), pp.1314-1319, May 1998.
- [16] C. Lopes, M. Pacheco, M. Vellasco, and E. Passos, "Rule evolver: A evolutionary approach for data mining," in Proceedings of RSFDGrC'99, (Yamaguchi, Japan), pp.458-462, November
- [17] I.B.Turksen, "Fuzzy data mining and expert system development," in Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, (San Diego, CA), pp.2057-2061, October 1998.
- [18] S.Russell and W.Lodwick, "Fuzzy clustering in data mining for telco database marketing campaigns," in Proceedings of NAFIPS99, (New York), pp.720-726, June 1999.
- [19] W. Pedrycz, "Conditional fuzzy c-means," Pattern Recognition Letters, vol. 17, pp. 625-632, 1996.
- [20] D. Shalvi and N. De Claris, "Unsupervised neural network approach to medical data mining techniques," in Proceedings of IEEE International Joint Conference on Neural Networks, (Alaska), pp. 171-176, May 1998.
- [21] H.Kiem and D.Phuc, "Using rough genetic and Kohonen's neural network for conceptual cluster discovery in data mining," in Proceedings of RSFDGrC'99, (Yamaguchi, Japan), pp. 448-452, November 1999.
- [22] T. Kohonen, S. Kaski, K. Lagus, J. Salojärvi, J. Honkela, V. Paatero, and A. Saarela, "Self organization of a massive document collection," IEEE Transactions on Neural Networks, vol.11, pp. 574-585, 2000.
- [23] J. Vesanto and E. Alhoniemi, "Clustering of the self organizing map," IEEE Transactions on Neural Networks, vol.11, pp.586-600, 2000.
- [24] D. Alahakoon, S.K. Halgamuge, and B. Srinivasan, "Dynamic self organizing maps with controlled growth for knowledge discovery," IEEE Transactions on Neural Networks, vol.11, pp.601-614, 2000.