

An Optimized Algorithm for Biological and Environmental Problems

Yasmeen¹, Dr. Neetu Sharma²

¹M.Tech. Scholar, Computer Science & Engineering Ganga Institute of Technology and Management Kablana, Jhajjar, Haryana, India

²Associate Professor, Computer Science & Engineering Ganga Institute of Technology and Management Kablana, Jhajjar, Haryana, India

ABSTRACT

In the past(Earlier), most of the researchers used data mining techniques in many area. A lot of amounts of data have been collected from various scientific domains such as Geo sciences, Astronomy, Meteorology, Geology and Biological sciences. Data mining techniques and tools used by researchers and scientist in biological and environmental problems also. In biological science data mining used in sequences alignment is based on the fact that all living organisms are related by evolution. ntal science data mining concept used in predicting data such as earthquakes and landslide etc. This Research paper highlights on the wide survey of Protein sequences, (DNA, RNA) sequences, cancer prediction, Relational and semantic data mining for biomedical research area. Health care data, multiagent framework for bio data mining, predicting earthquakes, landslide and spatial data in distributed data mining algorithms and tools.

Keywords : Medical, Data Mining, Environmental, Problems of Data Mining.

I. INTRODUCTION

In Research, Researcher discussed protein sequences which protein who highly affiliated with each other and discovered a Hyperclique pattern approach for extracting functional modules. Gen Miner is preprocessing software tool and that receive data from protein data base and transform them in a form of suitable for input WEKA software. Decision tree model and WEKA tool used in this protein analysis researchers apply some classification techniques like Neural networks, Genetic algorithm Fuzzy ARTMAP, Rough set classifier. These are the new technologie such as computers etc. Protein analysis together these keywords are searched on genome sequences also. In genome sequences focused on DNA-descriptors, feature descriptors principal component analysis (PCA), and self-organizing

feature Map (SOFM),genomic data mining is an important problem in bioinformatics. We Introduced DNA-descriptor for a sample of the DNA sequence of mouse in the bar diagram form. Fuzzy association rule mining and fuzzy weighted associative classifier (FWAC) used in a predictive technique for health care data mining. In this bioinformatic related work used Association rule mining,weighted association rule mining etc. Genetic algorithm used for classification rules in data mining. Data mining researchers focused on the commercial sectors and applications only few researcher works in this particular area. Environment problems is also a big deal for data mining. But usually in this area data mining used such as predictive data like land slide, earthquakes, spatial data etc. So many tools, algorithms, concepts and techniques used of data mining. We can see that in any companies and any

other government and private sectors data traffic problem which is created, so researcher given the environmental tool also to solve these problems. Many type of keyword introduce in this research paper.

- ✓ Bioinformatics
- ✓ Earthquakes
- ✓ Land slide
- ✓ Spatial data
- ✓ Biological sequence
- ✓ Cancer prediction

In these keywords we distributed the data mining many type of area and improve the concepts and techniques of data mining. In earthquakes problem worked hierarchical and non hierarchical clustering method. Risk assessment and validation used for land slide and found the accuracy. In biological sequence is so difficult for analyzing the protein and genomic sequences. But researcher accepted these challenges and reduced these sequences problems something. So, Data mining used other biological terms also and these concepts highlighted any other term of bioinformatics and environment. It has used the spatial information technologies, Hazard assessments, monitor, decision making.

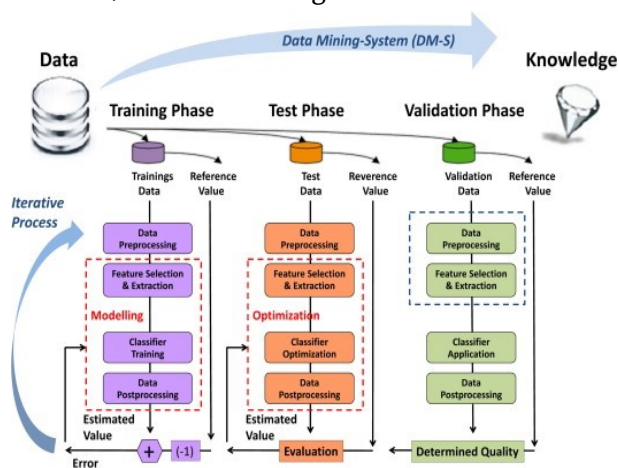


Figure 1. data mining presentation

II. A BRIEF REVIEW OF LATTER DATA MINING FOR BIOLOGICAL AND ENVIRONMENTAL PROBLEMS

Research on biological sequences in data mining: Hui xiong, Xiaofeng HE ,Chris ding ,Ya Zhang, Vipin kumar, Stephen r.holbrook worked on identifications of functional modules in protein complexes. Highly affiliated to each other and do not act isolated in cell but with function proteins work in cellular pathways in pair as a component. So researcher developed hyperclique pattern discovery from protein complexes it is type of association patterns.

- A. Hyper clique pattern discovery is only based on association rule.
- B. Protein complexes data and analysis tools(Gene ontology)
- C. Analysis of hyperclique pattern using Gene Ontology
- D. Hyper clique pattern as functional modules Proteins are So researchers have been developed in these pattern for protein complexes.

In this research described hyper clique pattern discovery approach to identification of proteins complexes. Many other researcher applied data mining tool for the analysis the proteins Gerasimos Hatzidamianos, Sotiris Diplaris, Ioannis Athanasiadis, Pericles A. Mitkas have been dealt with Gen miner. It is a pre-processing software tool, with the help of this tool we can receive data from three major protein databases and transform them in a form of suitable for input to the WEKA data mining suit and created the decision tree model. With used the derived training set and efficiency test was conducted in this research. In this research tried to solve problem “Given an amino acid database or training set that exists in proteins with known properties (that have been experimentally specified), we aim to create a tool that can classify a new, unknown proteins in some known to the training set

family of proteins, referred as protein class.” Researcher was used weka tool.with help of weka tool protein data sets and SQL manager also used in this research. From a functional point of view, GenMiner offered various services that are presented below:

1. Protein behavior discovering
2. Protein recognition
3. Decision tree building
4. Simple and functional user interface
5. Integration of multiple tools in one program

Most of the Researchers have been designed a scheme for automatic identification of a species from its genome sequence. For a DNA sequences there are many technologies was used of data mining. Shreyas Sen, Seetharam Narasimhan, and Amit Konar have been produced the Clustering Using Unsupervised Neural Learning for genomic sequence. It was used DNA-descriptors, Feature Descriptors, Principal Component Analysis (PCA), Self Organizing Feature Map (SOFM). That was a very challenging task for the researcher. But with the help of used the biological data mining it was implement in easy way. Bioinformatics is a new field of science. It is related by the science and engineering and Technology.The combination of statistics, molecular biology, and computational methods is used for analyzing and processing biological information like gene, DNA, RNA, and proteins and extract the other information.

III. CONCEPTS AND TECHNOLOGIES OF DATA MINING USED IN OTHERS BIOLOGICAL SEGMENTS

Researchers are used data mining techniques that used other biological factors. This phenomenon is popular and useful. Sunita Soni and O.P.V have been dealt with problem of classification using Fuzzy Association.International Journal of Scientific & Engineering Research, Volume And these proposed the concept of Fuzzy Weighted Associative Classifier

(FWAC). This research is only based on A PREDICTIVE TECHNIQUES FOR HEALTH CARE DATA MINING. In this research the word used keywords Associative Classifiers, Fuzzy Weighted Association Rule, FWAC, Fuzzy weighted support, Fuzzy weighted Confidence. In this research related work is Association Rule Mining, Weighted Association Rule Mining, Fuzzy Association Rule Mining (FARM), Fuzzy Weighted Association Rule Mining, Incorporating Weight in ARM, Utilizing Weight in Medical Domain, Fuzziness of Quantitative Attribute. Two important modification have been proposed (weight of an attribute and fuzzy fication of quantitative attributes). And these problem refused by Fuzzy Attribute Weight (FAW), Fuzzy Attribute set Transaction Weight (FASTW), Fuzzy Attribute Set Weight (FASW) Fuzzy Weighted Supported (FWS) and Fuzzy Weighted Confidence (FWC) for Fuzzy Weighted Associative Classifiers (FWAC). Techniques for Fuzzy Weighted Association Rule Mining is known as (FWARM).

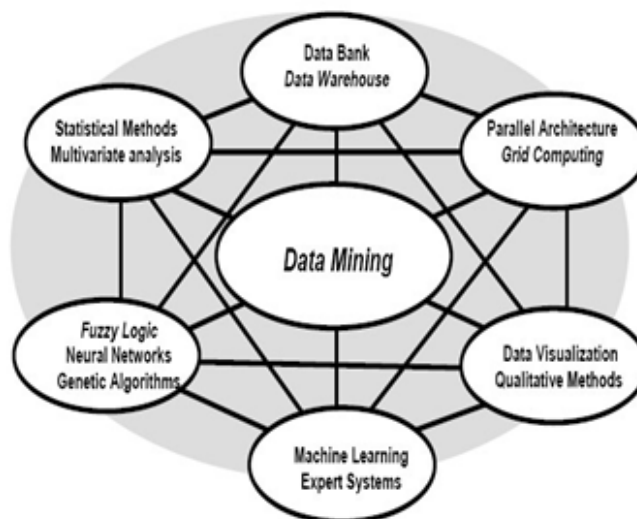


Figure 2. Data Mining Process

IV. INCLUDE THE ENVIRONMENT PROBLEMS ALSO IN DATA MINING

First of all what is environment..?

Environmentlism are the biotic surroundings which is of an organism, population, and includes particularly factors that have an influence in their

survival. And this is also have abiotic surrounding development and evolution.

Environmentalism (Environmental Problem) is a broad social and philosophical movement that, is a very big part, seeks to minimise and compensate the negative effect of human activity on the biophysical environment. The issues of concern for environmentalist usually related to the natural environments with the more important ones being climate change, species extinction, pollution, and old growth forest loss. Skilled environmental scientists have an important role that have to play in examining various environmental problems in a scientific manner and out Research And Development (R&D) activities for developing cleaning technologies and promoting development.

The Environmental problem given in below:-

- ✓ Earth quakes
- ✓ Land slide

IJSER(International Journal of Scientific & Engineering Research).

- ✓ Spatial Data
- ✓ Environmental tool

In above discussion, we can see that any researcher have been focused on environmental problem and improved the solution for reduced the problem. Data mining researcher focused on the only commercial data from environmental Problems. Only some few researchers focused on this field. K. Muralidharan researched on this particular area. He is introduced predicting earth quakes using data mining. His aims is study the scientific data. This research highlighted the various data mining techniques applied to mine for surface changes over time (e.g. Earthquake rupture) and with the help of the data mining Environment change in the intensity of volcan. Statistical model used in that area and highlighted observable space time earthquake patterns from unobservable dynamics using data mining techniques, pattern recognitions and resemble forecasting and

given that how data mining finding the why predict the earth quake and etc.

V. ACKNOWLEDGEMENT

I would specially like to thanks my worthy guide **Dr. Neetu Sharma**, who supervised me to complete this research paper. Her technical advice, ideas and constructive criticism contributed to the success of this paper. She suggested me many ideas and solved my puzzles. Her motivation and help has been of great inspiration to me. I would also like to thanks Mr. Ashish Sharma, respected faculty member of CSE department for providing me the necessary help to completed this research paper.

VI. CONCLUSION

We live in a world where various amounts of data are collected daily. such data is an important to need so data mining play the important role here. Data mining can meet this need by providing tools to discovery knowledge from data. Nowadays we can see that data mining use in many area. This paper presents researched on the data mining for biological and environment problem. So we observe that many kind of concepts and technique used in these problems. And try the removed complicated and hard type of data. This paper highlights on biological sequences problem such as protein and genomic sequences and other biological segments such as cancer prediction. In environment presents earth quakes, land slide, spatial data and environmental tool also discuss. Data mining algorithms, tools and concepts used in these problems Such as MATLAB, WEKA, SWIISPORT , Clustering , Biclstering and any other thing in this survey.

VII. REFERENCES

- [1]. Biological Data Mining edited By Jake Y. Chen and Stefano Lonardi.

- [2]. Scala For Data Science Leverage the power of scala tool for build to scalable , robust Data Science Applications.
- [3]. Next Generation Of Data Mining by Rajeev Motvani and Vipin Kumar.
- [4]. tutorial points for data mining problems of biological science and environmental science.