ISSN No. 0976-5697

Volume 4, No. 10, September-October 2013

International Journal of Advanced Research in Computer Science

REVIEW ARTICLE

Available Online at www.ijarcs.info

Overview of Emperical Data Mining Research

Lalitha Saroja Thota*
Research Scholar, Department of Computer Science,
Acharya Nagarjuna Univeristy, Guntur, India
lalithasarojathota@gmail.com

Allam Appa Rao
Director, CRRao AIMSCS,
University of Hyderabad campus, Hyderabad, India
apparaoallam@gmail.com

Abstract: Data and Information has a significant role on human activities. The explosive growth in databases has created a need to develop technologies that use information and knowledge intelligently. Data mining is the knowledge discovery process by analyzing the large volumes of data from various perspectives and summarizing it into useful knowledge. Due to the importance of extracting knowledge/information from the large data repositories, data mining has become an essential component in various fields of human life. Hence the data mining techniques has become an increasingly important research area. This paper discusses an overview of empirical data mining research from past to the present and reviews data mining techniques, research and their applications and development, through a survey of literature and the classification of articles and explores the future trends.

Keywords: data mining, supervised and unsupervised learning, knowledge discovery, hyper text, ubiquitous computing, multimedia, GIS, time series, sequences, business intelligence

I. INTRODUCTION

Although data mining is a relatively new term, the technology is not. Companies have used powerful computers to sift through volumes of supermarket scanner data and analyze market research reports for years. However, continuous innovations in computer processing power, disk storage, and statistical software are dramatically increasing the accuracy of analysis while driving down the cost.

The Emerging advances in computing power, communications networks, digital storage technologies, and high-throughput data-acquisition technologies, make it possible to gather and store unimaginable volumes of information in various fields of human life has lead to the large volumes of data storage in various formats like records, documents, images, sound recordings, videos, scientific data, and many new data formats.

The data collected from different applications require proper mechanism of extracting knowledge from large repositories for better decision making. Knowledge discovery in databases (KDD), often called data mining, aims at the discovery of useful knowledge from large collections of data.

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - knowledge that can be used to increase revenue, cuts costs, or both. The knowledge gained can be used for applications ranging from market analysis, fraud detection and customer retention, to production control and science exploration.

Thus data mining refers to extracting or mining knowledge from large amounts of data. Summarization, classification and prediction, regression, association, clustering etc are some of the data mining functionalities used to extract hidden knowledge. It's a useful software tool in several fields such as marketing, decision making, etc. It makes new opportunities for large-scale knowledge discovery from database. Data mining (DM) technology has emerged as a way of activity this discovery.

Before subjecting the data to data mining algorithms, a feature selection step is employed which might be performed on raw information or the detected peaks using unsupervised learning or supervised learning techniques. Data mining is a crucial part in databases which will be used to extract the hidden information by supervised or unsupervised learning methods.

A. Unsupervised learning methods:

Unsupervised approaches are simplest routine approach to visualize the distribution of data. These approaches include k-means clustering, principal component analysis (PCA), and hierarchical clustering which can be used a basis for feature selection. Principal Components Analysis (PCA) as a method of multivariate statistics [1] maps high dimensional knowledge by creating eigen values. Each linear combination or principal component is a weighted sum of the amplitude at each m/z value. The PCA approach was used to rank peak intensities within each spectrum and applied on cervical [2]. Hierarchical clustering (HC) is another powerful knowledge mining method for initial exploration of proteomic data. A very important category of clustering methods is hierarchical clustering. There are considerable research efforts which have been focused on algorithm-level improvements of the hierarchical clustering process [3]. When performing hierarchical clustering, some metric must be used to determine the similarity between the pairs of the clusters [4]. The HC begins by assigning each sample to its own cluster. It further calculates similarity scores or distance matrices between sample and places samples that are close to each other.

B. Supervised Learning methods:

Supervised learning techniques need class labels such that training can occur on knowledge obtained from a subset of the provided samples. This algorithms that reason from outwardly provided instances to provide general hypotheses that then build predictions about future instances. In alternative words, the goal of supervised learning is to create a concise model of the distribution of class labels in terms of predictor features [5].The two types of variables in this

exercise are predictor variables and a response variable (disease). Classification algorithms can be used for feature selection and classification. Such algorithms include genetic algorithms, decision trees, and neural networks. The aim of genetic algorithms (GA) is to extract a model by creating chromosomes of input variables and iteratively recombining chromosomes and mutating genes.

This paper reviews the various trends of data mining and its relative areas from past to present and explores the future areas of it.

II. THE CRAFT OF DATA MINING RESEARCH

Due to the increasing availableness and class of knowledge recording techniques, multiple data sources and distributed computing are getting the vital trends of modern data systems. Several applications like security informatics and social computing need a ubiquitous information analysis platform in order that decisions can be made quickly under distributed and dynamic system environments. Although data mining is currently been popularly utilized to achieve such outcomes, constructing a data mining system is, however, a nontrivial task, which require an entire understanding on varied data mining methods as well as solid programming skills [6].

There are many researchers and academicians engaged on coming up with designing efficient data mining techniques, methods, and algorithms [7]. Knowledge economy needs data mining is much goal-oriented in order that more tangible results are made. This demand implies that the semantics of the knowledge should be included into the mining process [8]. Monetary risks talk over with risks related with funding, like credit risk, business risk, debt risk and insurance risk, and these risks might place corporations in distress. Early detection of economic risks will facilitate credit grantors to cut back risk and losses, establish acceptable policies for various credit products and increase revenue. Because the size of financial databases grows, large-scale data mining techniques that can process and analyze huge amounts of electronic knowledge in a very timely manner become a key element of many financial risk detection methods and still be a theme of active research [9].

Data mining is a crucial technique in summarize and prediction in numerous industry. With the assistance of grid computer, the potential for knowledge storage and potency of data mining methods is extremely redoubled these days, most of studies concentrate on knowledge reproduction mechanism and replica choice method [10]. Data mining application to proteomic data from mass spectrometry has gained abundant interest in recent years. Advances created in proteomics and mass spectrometry have resulted in sizable amount of knowledge that can't be simply visualized or interpreted [11], and because of its advantage in speed and low price, we advice that MS proteomics may be a good candidate for an early primary screening method to disease diagnosis, identifying areas of risk and creating referrals for additional specific tests [12]. Knowledge points obtained from the information pre-processing step represent potential biomarkers. Several identification studies aim to seek out proteomic patterns which will discriminate between total different biological conditions.' proteomic patterns for disease diagnosis relies basically on the pattern of signals observed within a mass spectrum rather than the additional conventional identification and quantization of a biomarker [13]. Utilizing this technology, many clinical samples per day may be analyzed with the potential to be a new, highly sensitive diagnostic tool for the early identification of diseases or as a predictor of response to therapy [14]. Many profiling studies aim to seek out proteomic patterns which will discriminate between totally different biological conditions. Computational models of learning typically train on labeled input patterns (supervised learning), unlabeled input patterns (unsupervised learning), or a combination of the two (semi-supervised learning) [15].

Multi-relational data mining enables pattern mining from multiple tables. Multi-relational data mining algorithms can be used as practical proposal to overcome the deficiency of conventional algorithms. Deepak Meena and Hitesh Gupta [16] present an overview of multi relation data mining techniques and classification algorithms. Bhoj Raj Sharma et al [17] gives a review on data mining challenges, issues and applications

III. DATA MINING TRENDS

The current data mining methods capable of handling only a particular type of knowledge and limited amount of information, but as knowledge is expanding at a massive rate, there is a need to develop new data mining methods which are scalable and may handle different types of knowledge and large volume of information. Over recent years data mining has been establishing itself as one of the major disciplines in computer science with growing industrial impact. Undoubtedly, research in data mining will continue and even increase over coming decades [18].

There are lots of data mining trends which are currently being developed and researched. These trends include methods for analyzing more complex forms of data, as well as specific techniques and methods. They are distributed/collective mining, ubiquitous data mining, hypertext/hypermedia mining, multimedia mining, spatial data mining, time series and sequential data mining, constraint-based mining, visual mining, as well as business intelligence data mining etc. These are examined in detail in the upcoming sections

Not only do the application areas of data mining expand continuously, but also the utilized techniques keep up improving. In the rest of this article we take a closer look at above new methods:

A. Distributed /Collective Data Mining:

One area of data mining which is attracting a good amount of attention is that of distributed and collective data mining. Much of the data mining which is being done currently focuses on a database or data warehouse of information which is physically located in one place. However, the situation arises where information may be located in different places, in different physical locations. This is known generally as distributed data mining (DDM). Nowadays, the process of data mining is one of the most important topics in scientific and business problems. Grid Computing can be used as infrastructures to provide an effective computational support for distributed data mining applications [19]. Therefore, the goal is to effectively mine distributed knowledge that is found in heterogeneous sites. Samples of this embody biological information located in various databases, information that comes from the databases of different corporations, or analysis of

information from many branches of an organization, the adding of which would be an expensive and time-consuming process.

Distributed data mining (DDM) techniques become necessary for huge and multi-scenario datasets requiring resources that are heterogeneous and distributed. In the paper, it is focused on distributed data mining based in grid [20]. Distribution of knowledge and computation permits for finding bigger problems and executing applications that are distributed in nature.

B. Ubiquitous Data Mining:

Ubiquitous Computing poses the challenge of increased communication, context-awareness and functionality. In a highly dynamic and loosely connected ubiquitous environment, continuous access to the network (synchronous communication) is difficult. So it's necessary to go for tuple space which gives asynchronous communication without loss in information. Tuple space offers a coordination infrastructure for communication between autonomous entities by providing a logically shared memory along with knowledge persistence, transactional security as well as temporal and spatial decoupling properties that make it desirable for distributed systems such as Ubiquitous Computing [21].

With the advancement of laptops, palmtops, cell phones, and wearable computers is making ubiquitous access to large quantity of information possible. The emerging uses of "ubiquitous computing" have the potential of essentially altering the way we live and work. Most firms are in business to win, and trounce their competitors. They adopt novel technologies to avoid new competitors, reinforce an exciting competitive benefit, leapfrog competitors, or simply to make cash in new markets. Performance is important [22] Accessing and analyzing knowledge from a ubiquitous computing device offer several challenges. For instance, UDM introduces extra price owing to communication, computation, security and alternate factors. Several applications like security informatics and social computing need a ubiquitous information analysis platform in order that choices may be create rapidly under distributed and dynamic system environments.

C. Hyper Text and Hyper Media Data Mining:

Hypertext and hypermedia data mining can be characterized as mining data which includes text, hyperlinks, text markups, and various other forms of hypermedia information. As such, it’s closely associated to both web mining, and multimedia mining, which are covered independently in this section, but in reality are very close in terms of content and applications. While the World Wide Web is substantially composed of hypertext and hypermedia elements, there are other kinds of hypertext/hypermedia information sources which are not found on the web. The wealth of information available in the web or local corpora has grown while our ability to search and retrieve relevant information is being reduced. Hypermedia, being complex information objects, is hard to manage [23].

The importance of hypertext has been growing rapidly over past decade. The web and alternative information systems use hypertext format, with knowledge organized associatively instead of sequentially or relationally. A myriad of textual problems have been considered in the pattern matching field with several nontrivial results. Withal, amazingly very little work has been conducted on the natural combination of pattern matching and hypertext. In distinction to regular text, hypertext has nonlinear structure and therefore the methods of pattern matching for text cannot be directly applied to hypertext [24].

Hypertext has rapidly become an established paradigm within the design of information systems. The success of products within the software market, evident uses as reported by users and therefore the flowering of associated research activity all attest to the importance and staying power of hypertext rich information systems. Although standard hypertext has a number of unquestioned uses, the concept also has a variety of well-known issues and limitations. The text reviews the main issues and limitations of basic (standard) hypertext that constrain the utilization of hypertext in practical applications [25]. Net-based logs contain potentially helpful empirical knowledge with that World Wide Web (Web) designers and design theorists will assess usability and effectiveness of design choices [26].

D. Multimedia Data Mining:

Multimedia Data Mining is the mining and analysis of various types of data, including images, video, audio, and animation. As multimedia data mining incorporates the areas of text mining, similarly as hypertext/ hypermedia mining, these fields are closely connected. Abundant of the information describing these alternate areas also applies to multimedia data mining. Multimedia data mining is that the mining of high-level multimedia information and knowledge from massive multimedia databases. A multimedia data mining system paradigm, Multimedia Miner, has been designed and developed. It includes the development of a multimedia data cube that facilitates multiple dimensional analyses of multimedia data, based totally on visual content, and the mining of many kinds of knowledge, including generalization, discretization, classification, association, and clustering [27].

Multimedia Data Mining is that the mining and analysis of total different types of information, together with images, video, audio, and animation. The thought of mining knowledge which contains various kinds of information is that the main objective of multimedia data mining [27]. Because it is very common that the heterogeneous multimedia information of the same semantics always occur jointly in many domain and application specific databases, it is very useful to consider the location information when analyzing multimedia knowledge [28]. As multimedia data mining incorporates the areas of text mining, similarly as hypertext/ hypermedia mining, these fields are closely connected.

E. Spatial Data Mining:

A definition of spatial data mining is as follows: “the extraction of implicit knowledge, spatial relationships, or other patterns not explicitly stored in spatial databases.” Spatial Data Mining (SDM) is that the process of discovering fascinating, useful, non-trivial patterns information or knowledge from massive spatial datasets. Extracting fascinating and helpful patterns from spatial datasets should be more difficult than extracting the corresponding patterns from traditional numerical or categorical data owing to the complexity of spatial information types, spatial relationships, and spatial auto-correlation. Stressed overviewed the distinctive features that compare spatial data mining from

classical Data Mining, and presents major accomplishments of spatial Data Mining research. Extracting interesting patterns and rules from spatial datasets, like remotely sensed imagery and associated ground information, is vital in precision agriculture, community planning, resource finding and alternate areas [29]

Massive database of spatial information have been created within different applications such as Geographic Information Systems (GIS), environmental studies, banking, etc. The growing demand for knowledge housing inside these data-bases has attracted much attention to the field of Spatial Data Mining. Owing to the common complexity and massive size of spatial databases the aspect of efficiency is of the most considerations in spatial knowledge finding algorithms [30]. However, it is also important to consider information which is of an entirely different kind—spatial and geographic information which could contain information about astronomical data, natural resources, or even orbiting satellites and spacecraft which transmit images of earth from out in space.

F. Time Series and Sequence Data Mining:

Another important area in data mining centers on the mining of time series and sequence-based data. Simply put, this involves the mining of a sequence of data, which can either be referenced by time (time-series, such as stock market and production process data), or is simply a sequence of data which is ordered in a sequence. Sequence mining is a topic of data mining concerned with finding statistically relevant patterns between data. The concept of sequence Data Mining and discovering sequential patterns was first introduced by Rakesh Agrawal and Ramakrishnan Srikant in the year 1995 [31].

Sequence mining is a special case of structured data mining. In general, sequence mining problems can be classified as string mining which is typically based on string processing algorithms. String mining typically deals with a limited alphabet for items that appear in a sequence, but the sequence itself may be typically very long. Many interesting real-life mining applications rely on modeling knowledge as Sequences. In computational biology, DNA, RNA and protein information are all best modeled as sequences.

The knowledge produced from global different branch companies of one enterprise every day can be regarded as time series data. It is important and valuable for managers to make planning, decision, and control based on these time series data [32]. Time series is an important class of temporal data objects and it can be easily obtained from scientfic and financial applications. Tak-chung Fu [33] presents a review on time series data mining.

Another vital area in data mining canters on the mining of time series and sequence-based information. Simply put, this involves the mining of a sequence of knowledge, which might either be referenced by time (time-series, such as stock market and production process data), or is simply a sequence of knowledge which is ordered in a sequence. Data mining has been applied in DNA sequence analysis, which has got wide attention and rapid development. And considerable research results have emerged that provides an insight of research progress in DNA sequence data mining field [34].

G. Constraint-Based Data Mining:

Many of the data mining techniques which currently exist are very useful but lack the benefit of any guidance or user control. One method of implementing some form of human involvement into data mining is in the form of constraint-based data mining.

Constraint-based mining has been proven to be extremely useful. It has been applied not only too many pattern discovery settings but also, recently, on classification and clustering tasks. It appears as an important technology for an inductive database perspective on knowledge discovery in databases (KDD) and constraint-based mining is indeed an answer to vital data mining issues. However, some authors examine the nature of constraints and their semantics. Considering several samples of non important KDD processes, we describe the Hows, Whys, and When of constraints in a broader context than [35].

Data exceptions often reflect potential problems or dangers in the management of corporation. Analysts often need to find these exceptions from huge amount of data. A recent proposed approach automatically identifies and marks the exceptions for the user and reduces the reliance on manual discovery. However, the efficiency and scalability of this method are not so satisfying. According to these disadvantages, the optimization is investigated to improve it.

A new method that pushes several constraints into the mining process is proposed. By enforcing several user-defined constraints, this method first restricts the multidimensional space to a small constrained-cube and then mines exceptions on it. Experimental results show that this method is efficient and scalable [36].

Recent studies show that constraint pushing may substantially improve the performance of frequent pattern mining, and methods have been proposed to incorporate interesting constraints in frequent pattern mining. However, some popularly encountered constraints are still considered as "tough" constraints which cannot be pushed deep into the mining process [37] Data constraints. Due to the lack of research on methodological issues, the constraint-based pattern mining framework still suffers from many problems which limit its practical relevance [38].

H. Visual data mining:

It is justifiably aforesaid a picture is worth a thousand words. In our daily life we meet with several information media, which present us with facts and opinions, supported with some proofs, based, usually, on condensed data extracted from massive information. It’s common to speak such condensed data in a very visual kind - a static or animated, preferably interactive, visualization.

With the rise of multimedia contents in the net, individuals ought to handle a massive amount of multimedia contents in the internet as well as e-mail. Visual data mining is required to seek out appropriate visual data among massive multimedia system contents [39]. Visual data mining is a field of research which needs knowledge from several domains: statistics, data analysis, machine learning, artificial intelligence, human-machine interfaces, data or information visualization [40].

I. Business intelligence data mining:

After relational databases and data warehouses, the techniques used for data management have entered the next phase, Business Intelligence Tools. These tools provide enhanced business functionality by integrating data mining and advanced analytics into data warehouse systems to provide comprehensive support for the purposes of data

management, analysis and decision support [41]. Supported on synthetically processing the distributed application and data-stored environment about a few business intelligence system, a business intelligence application oriented distributed data mining system model with agents as basic parts is proposed.

IV. SOCIAL IMPACT OF DATA MINING TECHNIQUES

In the social sciences, non-utilization of information may be a major drawback. Several publications hold on in libraries or on the market on the net ought to be used more than they are currently. Typical approaches like providing abstracts and lists of keywords have proved to be not sufficient [42].

Arockiam et al [43] studied data clustering methods are discussed along with its two traditional approaches and their algorithms. Some applications of data clustering like data mining using data clustering and similarity searching in medial image databases are also discussed in the study.

It is known that in Information Technology (IT) and computer driven society, information is one of the most significant assets of any business firm. The information and communications technology (ICT), which discussed as a source for a novel knowledge based society, has played a vital role for creating a foundation for future society. It has been providing variety of knowledge based data services and compatibility enhancements of information society [44].

We exist in an environment of rapid change in which technology has an ever-increasing social relevance. The challenge now is to adapt our approaches to the application of new technologies like data mining, enabling us to use the tools technology provides wisely and with consideration for our society, its members, and its future.

V. DATA MINING AND ITS FUTURE

The area of data mining and knowledge discovery in databases (KDD) has been growing rapidly, and has shown great potential for the future. From the last two decades data mining and knowledge discovery applications have got a rich focus due to its significance in decision making and it has become an essential component in various organizations.

Emergency management is becoming more and more attractive in both theory and practice due to the frequently occurring incidents in the world. The objective of emergency management is to make optimal decisions to decrease or diminish harm caused by incidents. Nowadays the overwhelming amount of information leads to a great need of effective information analysis for the purpose of well informed decision. The potential of data mining has been demonstrated through the success of decision-making module in present-day emergency management systems [45].

Due to the enormous success of various application areas of data mining, the field of data mining has been establishing itself as the major discipline of computer science and has shown interest potential for the future developments. Ever increasing technology and future application areas are always poses new challenges and opportunities for data mining, the typical future trends of data mining includes Standardization of data mining languages, Data preprocessing Complex objects of data, Computing resources, Web mining, Scientific Computing and Business data.

What is the future of knowledge mining? Definitely, the area has made great strides in previous years, and many business analysts and experts in the field feel that the future will be bright. There is definite growth in the field of data mining.

In this article, we surveyed major challenges for data mining in the years ahead. We started with the classic definition of knowledge discovery and data mining. Although we believe that this definition still describes the essence of this important area of computer science, its interpretation has broadened over the last couple of years and will continue to do so in the future.

Although no human being can foretell the future, we believe that there are plenty of interesting new challenges ahead of us, and quite a few of them cannot be foreseen at the current point of time.

VI. CONCLUSIONS

In this paper we briefly reviewed and discuss an overview of empirical data mining research from past to the present and reviews data mining techniques, research and their applications and development, through a survey of literature and the classification of articles and explore the future trends.

This review would be helpful to researchers to focus on various issues of data mining. Finally, the ability to continually change and provide new understanding is the principle advantage of data mining techniques and methodologies, and will be at the core of data mining applications, in future.

VII. REFERENCES

[1] Maćkiewicz, A., Ratajczak, W, Principal components analysis (PCA), Computers and Geosciences 19 (3), pp. 303-342, 1993.

[2] Hellman, K., A. A. Alaiya, et al. (2004). "Protein expression patterns in primary carcinoma of the vagina." Br J Cancer 91(2): 319-26.

[3] Wu, J., Xiong, H., Chen, J., Towards understanding hierarchical clustering: A data distribution perspective, Neurocomputing 72 (10-12), pp. 2319-2330, 2009.

[4] Shi, S., Yang, G., Wang, D., Zheng, W., Potential-based hierarchical clustering, Proceedings - International Conference on Pattern Recognition 16 (4), pp. 272-275, 2002.

[5] Kotsiantis, S.B, Supervised machine learning: A review of classification techniques, Informatica (Ljubljana) 31 (3), pp. 249-268, 2007.

[6] Wu, X.-D., Zhu, X.-Q., Chen, Q.-J., Wang, F.-Y., Ubiquitous mining with interactive data mining agents, Journal of Computer Science and Technology 24 (6), pp. 1018-1027,2009.

[7] Wang, G., Domain-oriented data-driven data mining (3DM): Simulation of human knowledge understanding, Lecture Notes in Computer Science (including subseries Lecture

Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 4845 LNAI, pp. 278-290, 2007.

[8] Chen, Z, From data mining to behaviour mining, International Journal of Information Technology and Decision Making 5 (4), pp. 703-711, 2006.

[9] Peng, Y., Kou, G., Shi, Y., Knowledge-rich data mining in financial risk detection, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 5545 LNCS (PART 2), pp. 534-542,2009.

[10] Lin, H., Yi, F., Heping, H.,Dynamic service replica on distributed data mining grid, Proceedings - International Conference on Computer Science and Software Engineering, CSSE 2008 3, art. no. 4722367, pp. 390-393, 2008.

[11] Thomas, A., Tourassi, G.D., Elmaghraby, A.S., Valdes Jr., R., Jortani, S.A.,Data mining in proteomic mass spectrometry,Clinical Proteomics2(1-2), pp.13-32 ,2006

[12] McDonald, R.A., Skipp, P., Bennell, J., Potts, C., Thomas, L., O'Connor, C.D.,Mining whole-sample mass spectrometry proteomics data for biomarkers - An overview ,Expert Systems with Applications 36 (3 PART 1), pp. 5333-5340,2009.

[13] Conrads, T.P., Hood, B.L., Issaq, H.J., Veenstra, T.D., Proteomic patterns as a diagnostic tool for early-stage cancer: A review of its progress to a clinically relevant tool, Molecular Diagnosis 8 (2), pp. 77-85, 2004.

[14] Conrads, T.P., Veenstra, T.D, The utility of proteomic patterns for the diagnosis of cancer, Current Drug Targets: Immune, Endocrine and Metabolic Disorders 4 (1), pp. 41-50, 2004.

[15] Amis, G.P., Carpenter G.A, Self-supervised ARTMAP, Neural Networks 23 (2), pp. 265-282, 2010.

[16] Deepak Meena and Hitesh Gupta. Article: A Review: Data mining over Multi-Relations. International Journal of Computer Applications 62(8):42-45, Published by Foundation of Computer Science, New York, USA. January 2013.

[17] Bhoj Raj Sharma, Daljeet Kaur and Manju, A Review on Data Mining: Its Challenges, Issues and Applications, International Journal of Current Engineering and Technology, Vol.3,No.2 June- 2013

[18] Kriegel, H.-P., Borgwardt, K.M., Kröger, P., Pryakhin, A., Schubert, M., Zimek, A., Future trends in data mining, Data Mining and Knowledge Discovery 15 (1), pp. 87-97,2007.

[19] Wang, H., Nie, G., Fu, K.,Distributed data mining based on semantic web and grid, Proceedings of the 2009 International Conference on Computational Intelligence and Natural Computing, CINC 2009 (2), art. no. 5230988, pp. 232-234, 2009.

[20] Zhang, N., Bao, H., Research on distributed data mining technology based on grid , Proceedings - 2009 1st International Workshop on Database Technology and Applications, DBTA 2009 , art. no. 5207722, pp. 440-443, 2009.

[21] Sudha, R., Rajagopalan, M.R., Selvanayaki, M., Selvi, S.T., Ubiquitous semantic space: A context-aware and coordination middleware for ubiquitous computing, Proceedings of the 2007 2nd International Conference on Communication System Software and Middleware and Workshops, COMSWARE 2007, art. no. 4267986, 2007.

[22] Lee, H.J., Leem, C.S.,A study on value chain in a ubiquitous computing environment , Lecture Notes in Computer Science 3483 (IV), pp. 113-121,2005.

[23] Vassiliadis, B., Stefani, A., Drossos, L., Ioannou, K.,Some new algorithms for knowledge discovery in local hypertext corpora , WSEAS Transactions on Circuits and Systems 4 (11), pp. 1493-1500,2005.

[24] Amir, A., Lewenstein, M., Lewenstein, N., Pattern Matching in Hypertext, Journal of Algorithms 35 (1), pp. 82-99, 2000.

[25] Bieber, M.P., Kimbrough, S.O., On generalizing the concept of hypertext, MIS Quarterly: Management Information Systems 16 (1), pp. 77-93, 1992.

[26] Burton, M.C., Walther, J.B., The value of web log data in use-based design and testing, Journal of Computer-Mediated Communication 6 (3), 2001.

[27] Zaiane, Osmar R., Han, Jiawei, Li, Ze-Nian, Chee, Sonny H., Chiang, Jenny Y., Multimedia Miner: A system prototype for multimedia data mining, SIGMOD Record (ACM Special Interest Group on Management of Data)27 (2), pp. 581-583, 1998.

[28] Yang, Y., Zhuang, Y., Wang, W.,Heterogeneous multimedia data semantics mining using content and location context , MM'08 - Proceedings of the 2008 ACM International Conference on Multimedia, with co-located Symposium and Workshops , pp. 655-658,2008.

[29] Yang, T.-L., Bai, P., Gong, Y.-S.,Spatial data mining features between general data mining , 2008 International Workshop on Education Technology and Training and 2008 International Workshop on Geosciences and Remote Sensing, ETT and GRS 2008 2, art. no. 5070424, pp. 541-544, 2009.

[30] Zarnani, A., Rahgozar, M., Lucas, C., Nature-inspired approaches to mining trend patterns in spatial databases , Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 4224 LNCS, pp. 1407-1414,2006.

[31] Agrawal R. And Srikant R. , Mining Sequential Patterns, In Proc. of the 11th Int'l Conference on Data Engineering, Taipei,Taiwan, March 1995

[32] Jing, H., Junyi, R., Yanzhi, D.,Research on time series model for human resource management based on web technique, 2008 International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2008 , art. no. 4680876, 2008.

[33] Tak-chung Fu, A review on time series data mining, ELSEVIER Engineering Applications of Arti ficial Intelligence, 164–181, (24) 2011

[34] Zhu, Y.-Y., Xiong, Y.,DNA sequence data mining technique , Ruan Jian Xue Bao/Journal of Software 18 (11), pp. 2766-2781,2007.

[35] Boulicaut, J.-F, If constraint-based mining is the answer: What is the constraint? (invited talk), Proceedings - IEEE International Conference on Data Mining Workshops, ICDM Workshops 2008, art. no. 4733999, pp. 730, 2008.

[36] Li, C.-P., Li, S.-E., Wang, S., Du, X.-Y.,Constraint-based multi-dimensional data exception mining approach, Ruan Jian Xue Bao/Journal of Software 14 (9), pp. 1571-1577,2003.

[37] Pei, J., Han, J., Can we push more constraints into frequent pattern mining? , Proceeding of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 350-354, 2000.

[38] Bistarelli, S., Bonchi, F., Soft constraint based pattern mining, Data and Knowledge Engineering 62 (1), pp. 118-137, 2007.

[39] Bae, T.M., Kang, S.J., Ro, Y.M., Improving visual data mining using analyzing editing effect of visual data, Proceedings of SPIE - The International Society for Optical Engineering 5670, art. no. 11, pp. 88-98, 2005.

[40] Badjio, E.F., Poulet, F.,Usability of visual data mining tools , ICEIS 2004 - Proceedings of the Sixth International Conference on Enterprise Information Systems , pp. 254-258,2004.

[41] Basra, R.S., Lu, K.J.,Analysing multidimensional databases using data mining and business intelligence to provide decision support, ICEIS 2008 - Proceedings of the 10th International Conference on Enterprise Information Systems AIDSS, pp. 472-479,2008.

[42] Henk A. Becker a , Karin Sanders b, Innovations in meta-analysis and social impact analysis relevant for tech mining , Science Direct January 2006 Wu, X.-D., Zhu, X.-Q., Chen, Q.-J., Wang, F.-Y., Ubiquitous mining with interactive data mining agents , Journal of Computer Science and Technology 24 (6), pp. 1018-1027,2009.

[43] L.Arockiam, S.S.Baskar and L.Jeyasimman, Clustering Methods and Algorithms in Data Mining: Review, Asian Journal of Information Technology, 11:40-44, 2012

[44] Seok-Ji Park; An Analysis of R&D Policy Framework towards Information Society Technology Management for the Global Future, 2006. PICMET 2006 Page(s): 109 – 114, 2006.

[45] Chen, N., Chen, A., The role of data mining techniques in emergency management, ICEIS 2009 - 11th International Conference on Enterprise Information Systems, Proceedings AIDSS, pp. 118-123,2009.

Short Biodata for the Authors

Lalitha Saroja Thota received M.Tech in Software Engineering from Jawaharlal Nehru Technological University, Hyderabad, India in 2010 and M.Sc in Computer Science from Annamalai University, Chennai, Tamilnadu, India in 2008. She is pursuing Ph.D in Computer Science and Engineering from Acharya Nagarjuna Univeristy, Guntur, A.P, India. Her areas of interest are bioinformatics and data mining. She has more than 8 research papers in international journals and conferences to her credit.

Allam Appa Rao received Ph.D in Computer Engineering from Andhra University, Vishakapatnam, India in 1984. He has four decades of professional experience. His primary area of research is Bioinformatics. He has 60^{+} strong research team. 40 scholars were awarded Ph.D degrees under his guidance. He is currently Director, CRRao AIMSCS, University of Hyderabad campus, Hyderabad, A.P, India. He has more than 150 research papers in international journals and conferences to his credit.