DATA MINING IN EDUCATION : A REVIEW ON THE KNOWLEDGE DISCOVERY PERSPECTIVE

Pratiyush Guleria and Manu Sood

Department of Computer Science,
Himachal Pradesh University, Shimla, Himachal Pradesh, India

ABSTRACT

Knowledge Discovery in Databases is the process of finding knowledge in massive amount of data where data mining is the core of this process. Data mining can be used to mine understandable meaningful patterns from large databases and these patterns may then be converted into knowledge.Data mining is the process of extracting the information and patterns derived by the KDD process which helps in crucial decision-making.Data mining works with data warehouse and the whole process is divded into action plan to be performed on data: Selection, transformation, mining and results interpretation. In this paper, we have reviewed Knowledge Discovery perspective in Data Mining and consolidated different areas of data mining, its techniques and methods in it.

KEYWORDS

Decision, Knowledge, Mining, Selection, Transformation, Warehouse

1. INTRODUCTION

Knowledge Discovery in Databases (KDD) is the process of finding useful knowledge from large dataset.Data preparation, pattern search, knowledge evaluation and refinement are steps of KDD [1]. According to Han Jiawei, the process of Knowledge Discovery consists of Data Cleaning, Data Integration, Data Selection, Transformation, Data Mining and Pattern Evaluation Phases [2].

Data mining (DM) is the process where data is analysed and summarized into useful information. In short, data mining is process of deriving patterns from large databases [3].DM analyses large dataset to extract hidden patterns such as similar groups of data records using clustering technique.This data is used for machine learning and predictive analysis.DM works to analyze data stored in data warehouses and results in effective decision making [4]. Weiss and Indurkhya in [5] proposed that, "DM is the search for valuable information in large volumes of data". According to Technology Forecast [6], and Piatetsky-Shapiro et al. [7], it is the process of extracting previously unknown, useful information which include knowledge, association rules, pattern finding, statistical and mathematical techniques. Query languages or graphical user interface are required to express the DM requests and discovered information , so that results obtained from the DM Engine become understandable and usable for end users.

2. HISTORICAL TRENDS OF DATA MINING

Data Mining introduced in the year 1990's and it is the combination of many disciplines like database management systems (DBMS), Statistics, Artificial Intelligence (AI), and Machine Learning (ML) [8].Data Mining produce useful patterns by applying algorithmic methods on observational data.

2.1 Data mining trends

Data mining algorithms show best results for numerical data but with the emergence of Statis-tics and Machine Learning techniques, algorithms have been developed to mine non numerical data and relational databases [9].

Earlier most of the DM algorithms employed only statistical techniques [10],but now there are computing techniques like Artificial Intelligence, Machine Learning and Pattern Reorganization [9, 11] where huge heterogeneous data stored in data warehouses can be easily mined [2,12].

DM applications are successfully implemented in various fields like health care, finance, retail, telecommunication, fraud detection, risk analysis,education etc [13, 14, 15, 16]. Due to increasing complexities in various fields and improvements in technology, there are new challenges to DM which includes different data formats, distributed databases, networking resources etc.

3. KNOWLEDGE DISCOVERY IN DATABASES

Data mining and knowledge discovery in databases are related to each other and to other related fields such as machine learning, statistics, and databases. Data Mining is one of the steps in the overall process of KDD that consists of collection and preprocessing of data, data mining, inter-pretation, evaluation of discovered knowledge and finally post processing [17]. The KDD field's basic objective is to make data meaningful by developing methods and techniques of mining but problem being faced by the KDD process is to map huge and heterogeneous data into under-standable,more abstract and useful form [18, 19].

The phrase knowledge discovery in databases emphasizes that knowledge is the end product of a data-driven discovery [1, 18, 20, 21, 22].The data-mining step of KDD relies heavily on known techniques from machine learning, pattern recognition, and statistics to find patterns from data.

Data warehousing is one of the fields of databases [18, 21, 24, 25], which helps in business ana-lytics and decision support. Data warehousing helps set the stage for KDD in two ways: (1) data cleaning and (2) data access. Approach followed for analysis of data warehouses is called online analytical processing (OLAP) [23,26, 27, 28, 29].

3.1 The data mining step of the KDD process

Data mining step of KDD Process involves iterations for particular data-mining methods in ap-plication.There are two types of goals: (1) verification in which system is limited to verifying user's hypothesis and, (2) Discovery, in which system autonomously finds new patterns.

DM helps in determining patterns from observed data. Knowledge Inference is done from fitted models. Two primary mathematical formalisms [18] are used in model fitting: (1) statistical and (2) logical.

3.2 Data mining methods

Primary goals of data mining in practice are prediction and description.In prediction some variables and fields in the database are used to predict unknown values of other variables of interest, and description helps in finding human-understandable patterns describing the data [31, 32].

Weiss and Kulikowski in [33] proposed that, "Classification is learning a function that maps (classifies) a data item into one of several predefined classes". Apte and Hong in [34] suggested that classification methods of Data mining are used as part of knowledge discovery applications which includes classifying trends in financial markets, education and identifying objects of interest from large dataset of images.Regression is a predictive technique that maps data item to a prediction variable. Clustering is a descriptive task where we identify a finite set of categories or clusters to describe the data.E.g identifying those students who are short of attendance and shown poor performance in sessionals [35, 36, 37]. Cheeseman and Stutz in [38] suggested that examples of clustering applications in a knowledge discovery context include discovering similar groups. Summarization involves methods like calculating mean and standard deviations. There are some methods which involve deriving of abstract rules, visualization techniques, and the discovery of functional relationships between variables [18, 19]. Summarization techniques are often applied to interactive exploratory data analysis and automated report generation.

3.2.1 Decision trees and rules

Decision Trees are useful for multiple variable analyses. They split a data set into branch-like segments [39, 40].

3.2.2 Classification methods

These methods consist of techniques for prediction.Examples includes feed forward neural networks, adaptive spline methods, projection pursuit regression, Multi-Layer Perceptrons, Generalized Linear Models [41, 42, 43], Bayesian Networks, Decision Trees, and Support Vector Machines.

3.2.3 Example-based methods

In this, predictive analysis on new examples will be derived from those examples in the model for which predictions are known. Techniques include nearest neighbour classification and re-gression algorithms and case-based reasoning systems.

4. THE COMPONENTS OF DATA-MINING ALGORITHMS

One can identify three primary components [2, 10, 18] in any DM algorithm:

1. Model representation 2. Model evaluation 3. Search.

A model representation is used to describe or extract patterns whereas Model-evaluation criteria are statements which help in meeting the goals of Knowledge Discovery Process using particu-lar pattern or model.Predictive models are judged by the prediction accuracy on some dataset and descriptive models are evaluated along the dimensions of predictive accuracy, novelty, util-ity, and understandability of the model.

Search method consists of two components: (1) Parameter search and (2) Model search. Once the model representation and the model-evaluation criteria are fixed, then Data Mining problem left with optimization of task on observational dataset.

5. RESEARCH AND APPLICATION CHALLENGES

- Larger databases: There are databases with hundreds of fields, tables; millions of records and to derive some useful information from it is itself a challenge. Agrawal et al. [44], sug-gested methods for dealing with large data volumes using efficient algorithmic approaches because with increasing dataset there are chances of finding those patterns which are inva-lid.Solution to this problem is the use of prior knowledge to identify irrelevant variables.

- There are some issues related to prompt change, deletion of data that can make previously discovered patterns invalid [30, 45, 46]. Possible solutions are to discover methods for up-dating the patterns.

- Problem of missing and noisy data: This problem is related to business databases [47].and mostly happens when KDD methods and tools donot easily incorporate prior knowledge about a problem.

6. DATA MINING STEPS

- According to IBM report [6], three main steps in DM are preparing the data, reducing the data and, finally, looking for useful information.

- Predictive modeling [50] uses inductive reasoning techniques and algorithms like neural networks.

- Database segmentation [51] use statistical clustering techniques to partition data into clusters.

- Link analysis [3] identifies useful associations between data.

- Deviation detection [3] detects and explains why certain records cannot be put into specific segments.

- Fayyad et al. [18],proposed following steps of Data Mining:

  - Retrieving the data from a large database.

- Selecting the relevant subset to work with.

- Deciding appropriate sampling system, transformations, cleaning the data and to deal with missing fields and records.

- Fitting models to the pre-processed data.

7. DM TECHNIQUES

There are different data mining techniques which are used to extract information from a data set and transform it into an understandable format for further use. Table I shows different Data Mining Techniques and their roles.

7.1 Statistics

Statistics is a vital component in data selection, sampling, Data Mining, and knowledge evaluation.In data cleaning process, statistics offer the techniques to detect outliers to simplify data when necessary, and to estimate noise, it deals with missing data using estimation techniques [52, 53].

7.2 Classification and prediction

One of the most useful data mining techniques for e-learning is classification. Classification maps data into predefined group of classes. Classification is supervised learning approach be-cause the classes are determined before examining the data. The prediction of student's perfor-mance with high accuracy is more beneficial for identifying low academic performance of the students at the beginning.

Classification [54] is the processing of finding a set of models which describe and distinguish data classes or concepts. The derived results may be represented in various forms, such as classification (IF-THEN) rules, decision trees, or neural networks. Models then can be used for predicting the class label of data objects. In many applications, there is need to predict some missing data values rather than class labels. E.g. Case when the predicted values are numerical data, and is often specifically referred to as prediction.

7.3 Clustering

Clustering groups the data, which is not predefined and it can identify dense and sparse regions in object space. Clustering algorithm groups the data. Unlike classification and prediction, which analyse class labelled data objects, clustering analyses data objects without consulting a known class label. The class labels are not present in the training data and clustering can be used to generate such labels.Clusters of objects are formed so that objects within a cluster have high similarity in comparison to one another, but are very dissimilar to objects in other clusters. Each cluster formed can be viewed as a class of objects, from which rules can be derived [8]. Application of clustering in education can help in finding academic trends, student's performance analysis in class
.

7.4 Association

Association rule mining is to find set of binary variables that occurs in the transaction database repeatedly. Apriori measures are the association rule mining algorithm [52, 55]. Association analysis is the discovery of association rules showing attribute-value conditions that occur frequently together in a given set of data. The association rule A=>B shows those database tuples that satisfy the conditions in A as well as in B.

8. TECHNIQUES FOR MINING TRANSACTIONAL/RELATIONAL DATABASE

8.1 Artificial intelligence (AI) techniques

AI techniques consists of pattern recognition, machine learning, and neural networks.Other techniques in AI such as knowledge acquisition, knowledge representation, and search, are relevant to the various processes in DM.

8.2 Decision tree approach

Decision trees are non-linear data structures which start from root node and end with leaf node.Decision Trees represent sets of decisions. This approach can generate rules for the classification of a data set. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID) [56].These techniques are used for classification of a data set. They provide a set of rules that is applied to an unclassified data set to predict results. CART typically requires less data preparation than CHAID.

8.3 Visualization

Visual DM techniques are helpful in exploratory data analysis, and mining the large database. This approach requires integration of human in the DM process. There are examples of visualization techniques that work on large data sets and produce interactive displays [57].

There are various techniques for visualizing multidimensional data like scatter plot matrices, coplots, matrices, parallel coordinates, projection matrices, and other geometric projection techniques such as icon-based techniques, hierarchical techniques, web-based techniques, graph-based techniques, and dynamic techniques.

TABLE 1 DATA MINING TECHNIQUES AND THEIR ROLES

Techniques	Roles
Classification	Pre-Defined Examples
Clustering	Identification of similar classes of objects.
Prediction	Regression Technique.
Association Rules	Find frequent item set findings among large data sets.
Neural Networks	Derive meaning from complex or imprecise data and can be used to extract patterns and detect trends that are complex.

Decision Trees	Represent set of decisions using CART (Classification and Regression Trees) and CHAID (Chi Square Automatic Inte-raction and Detection), C4.5, ID3.
Nearest Neighbour method	Classify each record in a dataset Based on a combination of the classes of the K-records which are most similar in his-torical dataset.

9. THE VARIOUS DATA MINING AREAS

9.1 Web Mining

Web mining is the application of data mining to discover the patterns from the Web in the form of data collected from online information databases, hyperlinks, and digital data. Data mining technique used in web mining are Classification (supervised learning), Clustering (unsupervised learning) [59, 60].

9.2 Ubiquitous data mining

Increasing computational capacity and emergence of latest electronic devices leads to ubiquitous or pervasive computing paradigm [61]. The Ubiquitous computing environments give rise to Ubiquitous Data Mining (UDM).

9.3 Data mining using multimedia

The multimedia data includes images, video, audio, and animation. Data mining techniques fol-lowed in multimedia data are rule based decision tree classification algorithms like Artificial Neural Networks, Instance-based learning algorithms, Support Vector Machines, Association rule mining, clustering methods [63].

9.4 Spatial data mining

The spatial data includes astronomical and data related to space technology. It includes the use of spatial warehouses, spatial data cubes, spatial OLAP, and clustering methods [64].

9.5 Emergence of Data mining in other fields

Other data mining areas include visualisation, medical, pattern, wireless networks, association rule based mining.

10. PERFORMANCE IMPROVEMENT IN EDUCATION SECTOR

10.1 Data mining techniques in education

Applying data mining techniques to educational data for knowledge discovery is significant to educational organizations as well as students. Knowledge driven data supports educational deci-sion support system. Educational data mining enhance our understanding of learning by finding educational trends which includes improving student performance, course selection, in-house

trainings and faculty development.Using linear regression analysis [11], some factors are correlated to students academic performance like mother's education and student's family income.Data mining techniques helps in increasing student's retention rate, increase educational improvement ratio, and increase student's learning outcome.Thus, data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationship which help in effective decision making [53].

According to Han and Kamber [2] data mining softwares should be developed in such a manner that it allows the users to analyze data from different dimensions, enable to categorize it and summarize the derived results. Data mining can be applied on traditional as well as distance education.There are many general data mining tools that provide mining algorithms, filtering and visualization techniques.Some examples of data mining tools are DBMiner, Clementine, Intelligent Miner, RapidMiner and Weka etc [11] .DM combines machine learning, statistics and visualization techniques to discover and extract knowledge. Questionnaires and feedback forms are often used to collect data related to students's approach towards educational patterns or trends,interest towards technologies,teaching methodologies followed and data collected is to be analyzed using techniques like decision tree, neural networks etc.

There are different Mining models like Decision Trees, Naive Bayes,SupportVector Machines, Linear Regression, Minimum Description Length,K-means,O-Cluster.By using these models, one can get Student Behaviour Patterns, Course Behaviour Patterns, Predict Student Retention, Predict Course Suitability, and Personalized Intervention Strategy [65].

10.1.1 Statistics and visualization

According to Tsantis & Castellani [69], Student's log history and usage statistics are helpful in evaluation of an e-learning system.Information visualization techniques [66] can be used to graphically represent student data like his maximum interest towards which technologies or interest which he has shown in solving questionnaires etc are collected by web-based educational systems.Visualization techniques involves conversations among online groups, social networking websites etc. These techniques are also helpful for instructors which can manipulate the graphical representations generated and get the understanding and interest of their learners.

10.2 Web mining

Srivastava et al., [70] proposed, "Web mining is used to extract knowledge from web data". In web mining useful information is extracted from the contents of web documents and web usage mining is another technique to discover meaningful patterns from data generated by client-server transactions on one or more web localities.

10.2.1 Clustering, classification and outlier detection

Clustering and classification are both classification methods. Clustering is unsupervised and classification is supervised. Classification and prediction are also related techniques. Classification predicts class labels, whereas prediction predicts continuous-valued functions and outlier is an observation that is unusually large or small relative to the other values in a dataset.

According to Chen, Liu [72] decision tree i.e C5.0 algorithm and data cube technology are used for managing classroom processes. Induction analysis helps in identifying potential student groups having similar characteristics.Talavera and Gaudioso [73] propose mining student data using clustering to discover patterns reflecting user behaviours.

10.2.2 Adaptive and intelligent web-based educational systems

Tang et al. [74] gives concept of data clustering for web based learning and to help in solving learner based problems. They find clusters of students with similar learning characteristics based on the sequence and the contents of the pages they visited.

10.3 Association rule mining

Association rule mining is popular mining method used between set of items in large databases. Here one or more attributes of a dataset are associated with each other using IF-THEN statements.

10.3.1 Particular web-based courses

Ha et al. [75] perform web page navigational structure analysis from web-based virtual classrooms, e-learning portals and web pages navigated by learners.

10.3.2 Adaptive and intelligent web-based educational systems

Lu uses association fuzzy rules in a personalized e-learning material recommender system. Fuzzy matching rules are used to discover associations between student's requirements and a list of learning materials [75]. Romero et al. [76] propose to use grammar-based genetic program-ming with optimization techniques for providing a feedback to authors who designed courses and derived relationships from student's usage information.

10.4 Text mining

In text mining, mining is done on text data and is related to web content mining. It is an inter-disciplinary area involving machine learning and data mining, statistics, information retrieval and natural language processing[66,77].Text mining can work with unstructured or semi-structured datasets such as full-text documents, HTML files, emails, etc.

10.5 Web-based educational systems

Data mining and text mining technologies are used in Web-based educational systems for shared learning.Text mining is used for discussion board for expanded correspondence analysis. Learners select the relevant category which represents his/her comment and the system provides evaluations for learner's comments between peers.

10.5.1 Well-known learning content management systems

Dringus and Ellis [78, 79] propose to use text mining as a strategy for assessing conversations among irregular discussion forums. Text mining techniques also helps in evaluating the progress

of a thread or user group discussions. Data can be retrieved from pdf interactive multimedia productions for helping the evaluation of multimedia presentations for statistics purpose and for extracting relevant data [75, 76].Web-based educational systems collect large amount of student data from web log history which can be further analysed for deriving meaningful patterns [80].

10.5.2 Adaptive and intelligent web-based educational systems

Tang et al. [74] propose to construct a personalized web based application by which mining can be done on both framework and structure of the courseware. Keyword-driven text mining algorithms are used to select articles for distance learning students.

11. CONCLUSION

Educational Data Mining is an upcoming field related to several well-established areas of research including e-learning, web mining, text mining etc. Data Mining Techniques are used to analyze Educational data and extract useful information from large amount of data. This paper presents review of the KDD and basic data-mining techniques so as to integrate research in this area.

The KDD field is related to development of methods and techniques which make the data relevant.In Educational Sector softwares and visualization techniques can be developed using Data Mining Techniques which not only predict student's performance in examinations as well as helps us to cluster those students who need special attention in their studies. Knowledge Discovery in Databases results in better decision-making related to latest technologies useful in classroom teaching as well as faculty enhancement programs and in-house trainings etc.Using data mining techniques we can achieve refined data from distributed databases. Data Mining is an efficient tool for improving institutional effectiveness and student learning. Knowledge ac-quired by Educational Data Mining not only help teachers to manage their classes, improves their teaching skills, students learning processes but also provide feedback to institutions to im-prove their infrastructures and quality. For making this approach successful and to increase its scope, more data can be collected from Educational Institutions and queries can be performed on it.

Using Techniques like Decision Tree we can predict the Class Result of students based on the attributes taken. Decision tree classifiers are used on student's data to predict the student's performance in class result. These techniques will help in identifying those students who are below attendance and shown poor performance in Sessionals. The main finding of using these techniques is the gathering of knowledge from student's academic performance.

Another helpful technique is K-Means Clustering through which we can cluster the students based on some attributes like their Class Performance, sessionals and Attendance in class. Centroids are calculated from the educational data set taking K-clusters. It enhances the decision-making approach to monitor the performance of students. On increasing the value of K clusters, the accuracy becomes better with huge dataset and Kmeans can find the better grouping of the data. It also helps us to clusters those students who need special attention. This review on the Knowledge Discovery perspective in Data Mining would be helpful to find useful patterns related to Educational Data Sets.

REFERENCES

[1] Fan Jianhua, Li Deyi, "An Overview of Data Mining and Knowledge" Discovery, J. of Comput. Sci. & Technol., Vol.13 No.4, Jul. 1998

[2] Han Jiawei, Micheline Kamber, "Data Mining: Concepts and Technique". Morgan Kaufmann Publishers,2000

[3] Sang Jun Lee, Keng Siau, "A review of data mining techniques, Industrial Management & Data Systems", 101/1 [2001] 41-46.

[4] Padhraic Smyth, "Data Mining: Data Analysis on a Grand Scale", July 6,2000

[5] Weiss S. & Indurkhya N, "Predictive Data Mining: A Practical guide", Morgan Kauf-. mann, 1998.

[6] Technology Forecast: 1997 (1997), Price Waterhouse World Technology Center, Menlo Park, CA

[7] William J. Frawley, Gregory Piatetsky-Shapiro, and Christopher J. Matheus, "Knowledge Dis-covery in Databases: An Overview", AI Magazine Volume 13 Number 3 (1992) (© AAAI)

[8] Piatetsky-Shapiro, Gregory. 2000. "The Data-Mining Industry Coming of Age". IEEE Intelligent Systems.

[9] Venkatadri.M, Dr. Lokanatha C. Reddy, "A Review on Data mining from Past to the Future", International Journal of Computer Applications (0975 – 8887), Volume 15– No.7, February 2011

[10] Joyce Jackson, "Data Mining: A Conceptual Overview, Communications of the Association for Information Systems" ,Volume 8, 2002,pp.267-296

[11] Brijesh Kumar Bhardwaj, Saurabh Pal, " Mining Educational Data to Analyze Students Perfor-mance", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011

[12] S.Hameetha Begum, "Data Mining Tools and Trends – An Overview", International Journal of Emerging Research in Management &Technology, ISSN: 2278-9359, Feb 2013.

[13] Dharminder Kumar , Deepak Bhardwaj , "Rise of Data Mining: Current and Future Application Areas", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, September 2011 ISSN (Online): 1694-0814

[14] Annan Naidu Paidi, "Data Mining: Future Trends and Applications", International Journal of Modern Engineering Research (IJMER) Vol.2, Issue.6, ISSN: 2249-6645,Nov-Dec. 2012 pp.4657-4663

[15] Lokendra Singh, "Data Mining: Review, Drifts and Issues" ,International Journal of Advance Research and Innovation, Volume 2,,ISSN 2347 – 3258, 2013,pp.44-48

[16] Bharati M. Ramageri, Dr. B.L. Desai, "Role of Data Mining in Retail sector", International Jour-nal on Computer Science and Engineering (IJCSE), Vol. 5 No. 01 ,ISSN : 0975-3397, Jan 2013

[17] C. Romero, S. Ventura, "Educational data mining: A survey from 1995 to 2005", Expert Systems with Applications, Volume 33 Issue 1, July, 2007, 135–146

[18] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth," From Data Mining to Know-ledge Discovery in Databases", Copyright © 1996, American Association for Artificial Intelli-gence. All rights reserved. 0738-4602-1996

[19] Arabinda Nanda, Saroj Kumar Rout, "Data Mining & Knowledge Discovery in Databases: An AI Perspective", Proceedings of national Seminar on Future Trends in Data Mining (NSFTDM-2010):-10th may, 2010

[20] Yas A. Alsultanny, "Database Preprocessing and Comparison between Data Mining Methods", International Journal on New Computer Architectures and Their Applications (IJNCAA) 1(1): 61-73, The Society of Digital Information and Wireless Communications, ISSN 2220-9085, 2011.

[21] Ms. Chhavi, "Knowledge Discovery and Data Mining for the Future", Proceedings of the 3rd National Conference; INDIA COM, Computing For Nation Development, 2009, Bharati Vidya-peeth's Institute of Computer Applications and Management, New Delhi.

[22] Tho Manh Nguyen, A Min Tjoa, Juan Trujillo, "Data Warehousing and Knowledge Discovery: A Chronological View of Research Challenges", A Min Tjoa and J. Trujillo (Eds.): DaWaK 2005, LNCS 3589, 2005, pp. 530 – 535, © Springer-Verlag Berlin Heidelberg 2005

[23] Jose Samos, Felix Saltor, Jaume Sistac, Agustí Bardés, "Database Architecture for Data Ware-housing: An Evolutionary Approach"

[24] Tarun Dhar Diwan, Kamlesh Lehre , Vertika Kashyap , "An Evolutionary Approach for Disco-vering Changing Frequent Pattern in Data Mining", International Journal For Advance Research in Engineering and Technology , Vol. 1, Issue VII, ISSN 2320-6802, Aug 2013

[25] Matteo Golfarelli, Stefano Rizzi, "A Survey on Temporal Data Warehousing", International Journal of Data Warehousing & Mining, 5(1), 1-17, January-March 2009

[26] Eya Ben Ahmed, Ahlem Nabli and Faïez Gargouri, "A Survey of User-Centric Data Warehouses: From Personalization to Recommendation".

[27] G.Satyanarayana Reddy, Rallabandi Srinivasu, M. Poorna Chander Rao, Srikanth Reddy Rikku-la, "Data Warehousing, Data Mining, OLAP and OLTP Technologies are essential elements to support decision-making process in industries", (IJCSE) International Journal on Computer Science and Engineering,Vol. 02, No. 09, 2010, 2865-2873

[28] Surajit Chaudhuri, Umeshwar Dayal, "An Overview of Data Warehousing and OLAP Technolo-gy", Appears in ACM Sigmod Record, March 1997

[29] Muhammad Saqib, Muhammad Arshad, Mumtaz Ali, Nafees Ur Rehman, Zahid Ullah, "Improve Data Warehouse Performance by Preprocessing and Avoidance of Complex Resource Intensive Calculations", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 2, January 2012

[30] Usama Fayyad, Gregory Piatetsky - Shapiro, Padhraic Smyth, "The KDD Process for Extracting Useful Knowledge from Volumes of Data", Communications of the ACM, November 1996/Vol. 39, No. 1111

[31] Samir Farooqi, "Data Mining: An Overview", I.A.S.R.I.,Library Avenue,Pusa,New Delhi-110012

[32] Khalid Raza, "Application of Data Mining in Bioinformatics", Indian Journal of Computer Science and Engineering, Vol 1 No 2, 114-118

[33] Weiss S.M. and Kulikowski C.A, Computer Systems that learn. Morgan Kaufman Publishers, 1991.

[34] Apte, C., and Hong, S. J. 1996. "Predicting Equity Returns from Securities Data with Minimal Rule Generation". In Advances in Knowledge Discovery and Data Mining, eds. U. Fayyad, G. Piatetsky-Shapiro, P.Smyth, and R. Uthurusamy, 514–560. Menlo Park, Calif.: AAAI Press.

[35] Jain, Anil K., M. Narasimha Murty, and Patrick J. Flynn. "Data clustering: a review." ACM com-puting surveys (CSUR) 31, no. 3 (1999): 264-323.

[36] Sami Ayramo,Tommi Karkkainen, "Introduction to partitioning-based clustering methods with a robust example", ISBN 951392467X,ISSN 14564378

[37] Karl-Heinrich Anders and Monika Sester, "Parameter-Free Cluster Detection in Spatial Databas-es and its application to typification", International Archives of Photogrammetry and Remote Sensing. Vol. XXXIII, Part B4. Amsterdam 2000

[38] Cheeseman, P., and Stutz, J. 1996. "Bayesian Classification (AUTOCLASS): Theory and Re-sults". In Advances in Knowledge Discovery and Data Mining, eds. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, 73–95. Menlo Park, Calif.: AAAI Press.

[39] http://support.sas.com/publishing/pubcat/chaps/57587.pdf

[40] http://iasri.res.in/ebook/win_school_aa/notes/Decision_tree.pdf

[41] J.Elder, NonLinear Classification and Regression, CSE 4404/5327 Introduction to Machine Learning and Pattern Recognition

[42] Irene Kouskoumvekaki,Non-linear Classification and Regression Methods, September 29, 2011

[43] Peter Filzmoser, "Linear and Nonlinear Methods for Regression and Classification and applica-tions in R", Forschungsbericht CS-2008-3, Juli 2008

[44] Agrawal, R.; Mannila, H.; Srikant, R.; Toivonen, H.; and Verkamo, I. 1996. "Fast Discovery of Association Rules". In Advances in Knowledge Discovery and Data Mining, eds. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, 307–328. Menlo Park, Calif.: AAAI Press.

[45] http://digital.cs.usu.edu/~xqi/DataMining.html

[46] N.K. Sharma, Dr. R.C. Jain, Manoj Yadav, "A Survey on Data Mining Algorithms and Future Perspective", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (5) , ISSN:0975-9646, 2012,5149 – 5156

[47] Padhraic Smyth, David Heckerman, and Michael Jordan, "Probabilistic Independence Networks for Hidden Markov Probability Models", Massachusetts Institute of Technology, 1996

[48] Edgar Casasola, Susan Gauch, "Intelligent Information Agents for the World Wide Web", Information and Telecommunication Technology Center, Technical Report: ITTCFY97-11100-1.

[49] Venkat N. Gudivada, Vijay V. Raghavan, William I. Grosky, Rajesh Kasanagottu, "Information Retrieval on the World Wide Web", 1089-7801/97/$10.00 ©1997 IEEE

[50] Predictive Analytics 101: : "Next-Generation Big Data Intelligence", Intel IT Center, MARCH 2013

[51] J.Kishore Kumar,Dr A. Ravi Prasad,S.Ramakrishna, "Data Mining Techniques for Maintenance of Instances for Universities", (IJCSIT) International Journal of Computer Science and Informa-tion Technologies, Vol. 4 (3) , 2013, 475-476

[52] M. Sukanya, S. Biruntha, Dr.S. Karthik and T. Kalaikumaran, "Data Mining: Performance Improvement in Education Sector using Classification and Clustering Algorithm", International Conference on Computing and Control Engineering (ICCCE 2012), 12 & 13 April, 2012.

[53] Manoj Bala, Dr. D.B. Ojha, "Study of applications of Data Mining Techniques in Education", International Journal of Research in Science and Technology, (IJRST) 2012, Vol. No. 1, Issue No. IV, Jan-Mar ISSN: 2249-0604

[54] Pankaj Kumar Deva Sarma, Rahul Roy, "A Data Warehouse for Mining Usage Pattern in Library Transaction Data", Assam University Journal of Science &Technology : Physical Sciences and Technology Vol. 6 Number II,125-129, 2010

[55] Ankit Bhardwaj, Arvind Sharma, V.K. Shrivastava, "Data Mining Techniques and Their Implementation in Blood Bank Sector –A Review", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622,Vol. 2, Issue4, July-August 2012, pp.1303-1309

[56] Sumit Garg,Arvind K. Sharma , "Comparative Analysis of Data Mining Techniques on Educa-tional Dataset" ,International Journal of Computer Applications (0975 – 8887),Volume 74– No.5, July 2013

[57] JL Wesson, PR Warren, "Interactive Visualization of Large Multivariate Datasets on the World-Wide Web", Copyright 2001, Australian Computer Society, Inc.

[58] Jeffrey Hsu, "Data Mining Trends and Developments :The Key Data Mining Technologies and Applications for the 21st Century", Proceedings of the 19th Annual Information Systems,2002

[59] Soumen Chakrabarti, "Data Mining for hypertext: A tutorial survey", SIGKDD Explorations, Vol 1,Issue 2,Jan 2000

[60] Ming-Syan Chen,Jiawei Han,Data Mining: "An Overview from a Database Perspective,IEEE Transactions on Knowledge and Data Engineering,Vol 8,No.6,December 1996.

[61] Hsu, J. 2002. "Data Mining Trends and Developments: The Key Data Mining Technologies and Applications for the 21st Century", The Proceedings of the 19th Annual Conference for Informa-tion Systems Educators (ISECON 2002), ISSN: 1542-7382. Available Online: http://colton.byuh.edu/isecon/2002/224b/Hsu.pdf

[62] Shonali Krishnaswamy. 2005. "Towards Situation awareness and Ubiquitous Data Mining for Road Safety: Rationale and Architecture for a Compelling Application (2005)", Proceedings of Conference on Intelligent Vehicles and Road Infrastructure 2005, pages-16, 17.Available at: http://www.csse.monash.edu.au/~mgaber/CameraReady.

[63] Kotsiantis, S., Kanellopoulos, D., Pintelas, P. 2004. "Multimedia mining. WSEAS Transactions on Systems", No 3, s. 3263-3268.

[64] Abdulvahit, Torun. , Ebnem, Düzgün. 2006. "Using spatial data mining techniques to reveal vulnerability of people and places due to oil transportation and accidents: A case study of Istanbul strait", ISPRS Technical Commission II Symposium, Vienna. Addison Wesley, 1st edition.

[65] Ying Zhang, Samia Oussena, Tony Clark, Hyeonsook Kim, "Use Data Mining to improve stu-dent retention in higher education – A CASE STUDY"

[66] Cristobal Romero, Sebastian Ventura, Enrique Garcia, " Data mining in course management sys-tems: Moodle case study and tutorial", Computers & Education xxx (2007) xxx–xxx,Available Online at www.sciencedirect.com

[67] Lukasz A. Kurgan and Petr Musilek, "A survey of Knowledge Discovery and Data Mining process models", The Knowledge Engineering Review, Vol. 21:1, 1–24, 2006

[68] Sachin, R.B, Vijay, M.S, "A Survey and Future Vision of Data Mining in Educational Field, published in 2012",Second International Conference on Advanced Computing & Communication Technologies (ACCT), Rohtak, Haryana, ISBN 978-1-4673-0471-9, 7-8 Jan. 2012, pp 96 – 100.

[69] Tsantis, L. & Castellani, J. (2001). "Enhancing Learning Environments through Solution-based Knowledge Discovery Tools: Forecasting for Self-Perpetuating Systemic Reform", Journal of Special Education Technology, 16(4), 39-52. February 18, 2014

[70] Jaideep Srivastava, Prasanna Desikan, Vipin Kumar, "Web Mining - Concepts, Applications & Research Directions", AHPCRC Technical Report, Chapter 3,pp.51-53

[71] Haixun Wang,Wei Wang,Jiong Yang, Philip S. Yu, "Clustering by Pattern Similarity in Large 13 Data Sets", In Proceedings of the 2002 ACM SIGMOD international conference on Management of data,pp. 394-405,2002,ACM

[72] Liu, Chen-Chung. "Knowledge discovery from web portfolios: tools for learning performance assessment". Diss. 2001.

[73] Talavera, Luis, and Elena Gaudioso. "Mining student data to characterize similar behavior groups in unstructured collaboration spaces." In Proceedings of the Artificial Intelligence in Computer Supported Collaborative Learning Workshop at the ECAI 2004, pp. 17-23. 2004.

[74] Tang, Tiffany Ya, and Gordon McCalla. "Student modeling for a web-based learning environ-ment: a data mining approach." In AAAI/IAAI, pp. 967-968. 2002

[75] Ha, S., Bae, S., & Park, S. (2000). "Web mining for distance education". In IEEE international conference on management of innovation and technology (pp. 715–719)

[76] Cristobal Romero, Sebastian Ventura and Paul De Bra, "Knowledge Discovery with Genetic Programming for Providing Feedback to Courseware Authors, User Modeling and User-Adapted Interaction" (2004) 14: 425–464 © Springer 2005

[77] Vishal Gupta,Gurpreet S. Lehal, "A Survey of Text Mining Techniques and Applications", Jour-nal of Emerging Technologies in Web Intelligence, vol. 1, no. 1, August 2009

[78] Laurie P. Dringus , Timothy Ellis, "Using data mining as a strategy for assessing asynchronous discussion forums", Computers & Education 45 (2005) 141–160

[79] M'hammed Abdous, Wu He and Cherng-Jyh Yen, "Using Data Mining for Predicting Relation-ships between Online Question Theme and Final Grade", Educational Technology & Society, 15 (3),2012, 77–88,Available Online at www.sciencedirect.com.

[80] Agathe Merceron, KalinaYacef, "Educational Data Mining: a Case Study, Supporting Learning through Intelligent and Socially Informed Technology". Proceedings of the 12th International Conference on Artificial Intelligence in Education, AIED 2005, July 18-22, 2005, Amsterdam, The Netherlands

AUTHORS

Pratiyush Guleria is pursuing PhD in Computer Science from Himachal Pradesh University Shimla, INDIA. He has done Mtech in Computer Science with a Gold Medal from Hi-machal Pradesh University, Shimla, INDIA. He has received his MBA in Operation Research from Indira Gandhi National Open University (IGNOU) and Btech in Information Technology from I.E.E.T Baddi, Distt Solan, Himachal Pradesh University. He has more than 6 Years of Experience in IT Industry and Academics. His research interests in-clude Data Mining and Web Technologies.

Prof. Manu Sood is currently working as a Professor in the Department of Computer Science at Himachal Pradesh University Shimla, India. He has completed his Ph.D. in Computer Engineering under the Facul-ty of Technology from University of Delhi, Delhi, India. He completed his M. Tech. in Information Systems with a Gold Medal from Netaji Subhash Institute of Technology, Delhi, India. He has received his B.E. degree in Electronics and Telecommunication from Government Engineering College, Jabalpur, Madhya Pradesh, India. Prof. Sood has over 25 years of extensive experience in IT Industry and Academics in India at various positions. His research interests include Software Engi-neering, Model Driven Software Development, Model Driven Architec-ture, Aspect Oriented Software Development, E-learning, Service Oriented Architecture, MANETs and VANETs