Middle-East Journal of Scientific Research 24 (10): 3131-3138, 2016
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2016.3131.3138

Multi Attribute Implicit Inference Model Based EDS Using Data Warehousing Techniques

[1]R. Sundara Bharathi and [2]R. Mala

[1]Research Scholar, Alagappa University College of Arts and Science, Paramakudi, India
[2]Department of Computer Science,
Alagappa University College of Arts and Science, Paramakudi, India

Abstract: Warehousing information of any domain has great deal in information mining to solve many problems. To support in obtaining intelligence about any problem, there are many data warehousing approaches has been discussed earlier. In this paper, we consider the application of data warehousing technique to support generating intelligence from the warehouse data in such a way to develop the educational sectors. We propose a multi attribute implicit inference model based Educational Development system (EDS) using data warehousing techniques. The method clusters the vast information about educational sectors into different categories and performs inference from the clustered information in many ways. The framework supports the administering people to get knowledge about any factor from the student to the professors and also in the side of business intelligence. The proposed method presents a learning framework for the students with more accurate and efficient information. In case of administration the model presents various influencing factors of education and for the business side the model presents the intelligence about growing factors.

Key words: Data Warehousing · Educational Development Systems · Implicit Inference Model · Data Mining

INTRODUCTION

The data warehousing is the process of organizing information of different categories in a meaningful way such that it could be extracted in efficient manner. In earlier days the warehousing is performed by storing information in files in form of papers when the computerization comes and the growth of information technology helps storing information under different names in electronic manner which can be extracted whenever necessary. The problem of warehousing is about how good and how fast the required information can be fetched from the warehouse. If the data presented is stored in meaningful manner then the process of extracting the information also becomes easier. So the kind of indexing speaks the efficiency of warehousing. There are numerous data warehousing techniques are available and will be discussed in the next section. For example, the information about any educational organization can be grouped or stored based on many factors. The information of student details can be stored in different form, whereas the administering information can be presented in another ways. Similarly the course details and course materials also can be stored in efficient manner to present the data efficiently [1].

Data mining is the process of extracting information from the large knowledge base. In our case, if the data is organized in the data warehouse, then extracting information from the data warehouse is called data mining. The information present in the warehouse will be huge in volume and identifying and extracting required information also difficult. For example, if the user requires the detail about the students studied in particular year in any branch or course then the data mining process can perform search process to findout the list of students from the huge knowledge base. In order to produce the result for such a query, the warehousing should be efficient and the indexing scheme must be efficient [2].

There may be lot of peoples being studying in any educational organization and they would be belonging to many courses and years. Also there will be many lecturers would be working in the organization and takes classes to

Corresponding Author: R. Sundara Bharathi, Research Scholar, Alagappa University College of Arts and Science, Paramakudi, India.

the students. Not all the students like the teaching method of all the lecturers and the acceptance of lecturers teaching is depend on the students interest and how the student understand the teaching of the lecturer. So the data mining tools and warehousing applications can be used to predict the interest of the students in different subjects and the inference techniques can be used to identify how the interest of students changes between courses and methodologies. Because each lecturer have their own way of teaching and some of the students like particular way of teaching. For example, in the session of one lecturer there may be more examples with real time constraints and in some others that will be missing. Also some of the lecturers speak many general things between the seminar which will be interesting and keep the mind of students in the topic and does not get boring.

How this can be identified is, by warehousing the online lecture of different staffs according to many factors. The simple case is organizing them according to the topic of lecture. That may be the root case; further the lectures can be grouped under the name of authors or based on other factors like informatics, more explanatory, more realistic and so on. The factors are not limited like this, but can be grouped into many ways. On the other side, the lecturers can be grouped into several categories and warehoused according to many factors like by voting from students, by experience, by research ethics and so on. This kind of warehousing might be useful in identifying the more efficient lecturers in any topic to use them in many ways. Similarly the information about the students and their records can be organized in many categories, so that from the information about rank holders in any subject and how their interest changes in the online tutors can be obtained. This kind of systems would be helpful in developing the educational systems in many ways.

We limit our research in warehousing the student records, lecturer's records, admission details due to many factors and scope in this paper. This paper speaks about how the data is warehoused and how it has been used to support educational systems.

Related Works: There are number of data warehousing techniques has been discussed for the development of educational systems. We discuss some of the methods here in this section.

Data mining techniques for teaching result analysis using rough set theory [1]: proposes a concept map for each student and staff and finds the result of the subjects and also recommending for sequence of remedial teaching. This paper uses rough set theory for dealing with uncertainty in the hidden pattern of data. For each competence the lower and upper approximations are calculated based on the brainstorm.

Analysis of New Teaching Model for Undergraduates in University Based on the Environment of Information Technology - A Case Study [2], for Specialty of Information Management & Information System: proposes and explains three new teaching models in detail. Thirdly, based on analysis of the ability system for the specialty of information Management & Information System, this paper tries to design the new teaching models and construct a framework for platform of teaching for undergraduate based on networks. Finally, a design concept of the platform is discussed detailed.

Experience from teaching performance analysis of object-oriented systems [3]: reports the experience from teaching "Performance Analysis of Object-Oriented Systems" which was offered for the first time in the spring of 2004. The class was designed for juniors/seniors and graduate students majoring in Computer Science and Computer Engineering. The main focus of this course is on the implementations of class loading, Just-In-Time compiler, threading and garbage collection in virtual machines supporting Object-Oriented languages such as Java and C#. We adopted Microsoft Shared Source Common Language Infrastructure (SSCLI) as the main experimental platform and the "Shared Source CLI Essentials" as one of the main accompanying textbooks. We find that the combination of SSCLI and the book provides a very effective means to deliver the course contents. In this paper, a complete documentation of the course design, the evaluation of students' work and the instructor's reflection is presented.

Impact of E-learning system using Rank-based Clustering Algorithm [4], implemented through this work suggests the instructors to use the combination of E-learning System Using Rank-Based Clustering Algorithm (ESURBCA) was designed. The main aim of the model developed is to get consistency in content delivery, quality content in learning materials, students self-learning concept and performance improvement in their examination. A study has been conducted During June 2013 to September 2013, the author collected samples of 1631 from final year and Second year of BCA, B. SC and B. Sc-IT students were trained through e-learning system architecture and the objectives of this study is 1. To measure the effectiveness of E-learning System Using Rank-Based Clustering Algorithm (ESURBCA) among the

students of Mercury College of arts and science And Ankara arts and Science College in concepts of Programming in JAVA Course. The newly designed E-learning System using Rank-Based Clustering Algorithm (EUSRBCA) shows an improvement over the existing systems with better results. From the various evaluations carried out, the performance of the system found to be good comparatively to other systems in e-learning domain.

A Clustering Methodology of Web Log Data for Learning Management Systems [5], proposed a methodology for analyzing LMS courses and students' activity. This methodology uses a Markov Clustering (MCL) algorithm for clustering the students' activity and a Simple KMeans algorithm for clustering the courses. Additionally we provide a visualisation of the results using scatter plots and 3D graphs. We propose specific metrics for the assessment of the courses based on the course usage. These metrics applied to data originated from the LMS log files of the Information Management Department of the TEI of Kavala. The results show that these metrics, if combined properly, can quantify quality characteristics of the courses. Furthermore, the application of the MCL algorithm to students' activities provides useful insights to their usage of the LMS platform.

Cluster Analysis of Behavior of E-learners [6], proposes the analysis of students' behavior using data mining tools and techniques. Classification and clustering techniques are used to analyze the relationship between usage of gcourses and performance of students. Students' performance depends upon their grades, how much time they spend in learning, usage of courses as well as richness of course quality. The study uses data from previous approach, E-learning data from Greek University. This paper uses same approach with different data mining tools and techniques.

Predicting Students Performance Using Data Mining Technique with Rough Set Theory Concepts [7, 8], discusses the basic approach and concepts of the rough set theory in the field of academic domain for the performance prediction of students in course works.

A generic agent based Cloud Computing architecture for E-Learning [9-11], proposes Agent based e-learning which is helpful in managing the information overload, it can serve as academic expert and manages creation of programming environment for learners. There are many characteristics that an E-learning environment has to support; they are Interaction, Data Security, User Personalization, Adaptability, Intelligence, Interoperability, Accessibility and User Authentication. E-learning must also support a few other features like cost effectiveness, reusability, storage capacity, powerful computing and virtualization which can be provided by Cloud computing.

Cloud e-Learning: A New Challenge for Multi-Agent Systems [12] proposes that Cloud e-Learning, a new approach to e-learning, will open opportunities for learners, by allowing personalisation, enhancing self-motivation and collaboration. The learners should be able to choose what to learn, what sources to use, with and by whom, how and in what pace, what services and tools to use, how to be assessed, whether to get credits towards a degree etc. In such a dynamic environment, the need for Multi-Agents Systems is necessary. Actors in Cloud e-Learning would need automated facilitation in all services involved. We outline few indicative scenario for Cloud E-Learning in which smart agents will act on behalf of the learners, teachers and institution in order to maximize the benefit of the proposed concept.

Application of Data Warehouse Technique in Educational Decision Support System [13-15], introduces techniques of data warehouse and online analyzing processing and provides a solution to decision-making support system based on an education-related case study. The paper also discusses how to collect business data, transform and integrate these data to analyze the state of education based on OLAP and how the end user can get the information easily.

A Prototype System for Educational Data Warehousing and Mining [16], present the design and development of the proposed data warehouse solution, which facilitates better and more thorough analysis of departmentpsilas data. The proposed system constitutes an integrated platform for a thorough analysis of departmentpsilas past data. Analysis of data could be achieved with OLAP operations. Moreover, we propose a thorough statistical analysis with an array of data mining techniques that are appropriate for the examined tasks.

Data warehouse based evaluation of students' achievements in information systems education [17], present data analysis and evaluation of results of students' knowledge and skills assessment in the field of information systems education. As assignment methods it was organized that students are assigned to projects or they do practical partial exam at laboratory tests. This way they fulfill their pre-exam requirements. Assessment results data from the period of fifteen years is analyzed by using data warehouse technology. Trends are recognized in students' achievements data and they show students' preferences regarding assessment methods.

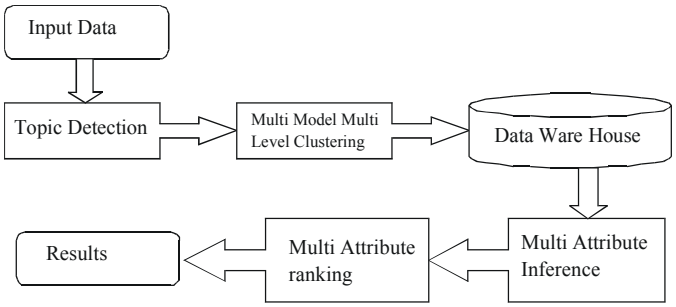


Fig. 1: Architecture of proposed system

Analysis of Data Warehousing and Data Mining in Education Domain [18] provides an option to build data warehouse and extract useful information using data warehousing and data mining open source tools. In this paper we have explored the need of data warehouse / business intelligence for an educational institute, the operational data of an educational institution has been used for experimentation. The study may help decision makers of educational institutes across the globe for better decisions.

Big data mining in education sector: Critical Issues and Current Status in Indian Context [19], discuss critical issues that are not allowing India to leverage the result of educational data analytics. We also discuss initiatives taken by government and corporate, current status and stakeholders in current digital learning revolution age.

From the above discussion we have identified the following problems and there is no such analysis model available for the evaluation of quality of education. It has to be identified and designed to provide good quality education. The earlier approaches are proposed to cluster the documents and phone calls. There is a method to cluster the texts for Russian language. The earlier method lacks with accuracy because of the manual text transcript is required to perform clustering.

Multi Attribute Implicit Inference Model for EDS: The educational system has many requirements in the area of future development and need various tools to provide intelligence about how the educational system can be developed. The data warehousing techniques can be applied in efficient manner to improve the efficiency of educational systems. In that sense, we proposed a multi attribute implicit inference model for educational development system which includes multi model multi level topical clustering to warehouse the huge information in such a way to retrieve in an efficient manner. Also the method proposes a multi attribute inference model which generates intelligence according to user requirements. The entire process can be split into different steps namely multi model multi level topical clustering, Implicit Inference, Multi Attribute Ranking. We explain each of the functional stages in detail in this section.

The Figure 1 shows the architecture of the proposed multi attribute inference model.

Topic Detection: The method is given with pure data and the Meta data as the input. From the input, the method identifies the list of topics as classes and the sub classes of each topic. The Meta data has various information about the data, as mentioned in the section 1 of this paper. From the method the method generates various hierarchical taxonomy of data which will be used to perform clustering in the next stage. For example, if the Meta data has the topic namely tutor then the method identifies the list of sub classes belongs to the topic tutor. The names of subclasses are extracted from the list of feedback traces available in the data base.

Algorithm:

Input: Meta Data Md, Data Set Ds.

Output: Topic Set ts.

Start

Read data set Ds.

Initialize Topic Set Ts.

For each data Mi from Md

For each Data point Di from Ds

Identify the presence of Mi in Di.

If Di.Topic==Mi.Topic Then

Add to Topic set Ts.

$$Ts = \int_{i=1}^{size(Mi)} \int_{j=1}^{size(Di)} \Sigma topic(Di = Mi) \nexists Ts$$

End

End

End

Stop

The above discussed algorithm identifies the set of all topics present in the data set which has to be clustered. The identified topics and their sub classes are used to perform clustering in the next stage.

Multi Model Multi Level Clustering: In this stage, using the topics being identified the method reads the data set. For each topic and their sub classes identified the method extracts the data points from the data set. According to the topics and their subspace identified, the method groups the records to form the cluster. This will be iterated for each of the level and will be performed for all the topics being identified. In this method, using the topics identified, the method constructs the tree and once the method constructs the tree, then the data points are assigned in any one of the leaf according to the topic identified.

Algorithm:
Input: Topic Set Ts, Data Set Ds.
Output: Cluster C
Start
Initialize Tree T.
For each topic Ti from Ts
If Ti==Root then
Create Node Ni and add to T.
Else
Traverse the tree to identify the parent
Create Node under the parent Pi.
End
End
For each data point Di from Ds
Identify the topic Ti.
Ti = Di.Topic.
Traverse the tree to identify the topical node.
Ni = $\int_{i=1}^{size(T)} Ti, Topic = Ti$
Create Node Na.
Add to Ni.
Ni = $\Sigma\, Nodes(Ni) \cup Na$
End
Stop.

The above discussed algorithm performs the clustering of data points to form the warehouse where the data points are grouped to form a tree structure.

Multi Attribute Inference: The multi attribute inference is the process of selecting the result from the warehouse according to the query being submitted. For example, the query "lists the tutorials on data mining" the method has to produce number of relevant results. In this stage, the method considers various attributes under the class data mining and tutorials. In the class of tutorials there will be number of sub classes namely, informatics, explanatory, realistic. Based on all these classes the method has to produce results of data mining tutorials which satisfy the above conditions. If the users includes or specify any one of the category then it has to produce the result from the particular class.

Algorithm:
Input: Query Q, Data Tree T
Output: Results Rs.
Start
Read query Q.
Identify the topic and classes.
TC = $\sum_{i=1}^{size(Ts)} Ts(i) \in Q$
For each class C from Tc
Extract results from T.
Rs= $\int_{i=1}^{size(TC)} \int_{j=1}^{size(T)} \Sigma\, Ds(i).\, Topic == C$
End
Stop.

The above discussed algorithm performs multi attribute inference which selects set of results from the data set or data warehouse.

Multi Attribute Ranking: The ranking is the original process which supports the development of the educational system. From the result being returned by the inference system, the method computes ranking for each of the material, lecturer, or any factor based on the feedback result stored in the data warehouse. The method collects all the results related to each distinct result and from that it computes the multi user feedback weight based on which the ranking is performed to produce the result to the user or administrator or anybody.

Algorithm:
Input: Data Set Ds, Result R.
Output: Re-ranked Result RRS.
Start
For each result Ri from R
Collect all the records belongs to Ri.
TR = $\sum_{i=1}^{size(Dz)} Di == Ri$
Identify all positive feedbacks.

$$NPF = \int_{i=1}^{size(TR)} \Sigma TR(i).F == P$$

Compute multi user feedback weight.

$$MUFW = NPF \Big/ size(TR)$$

End

Sort Results according to MUFW.

Return results

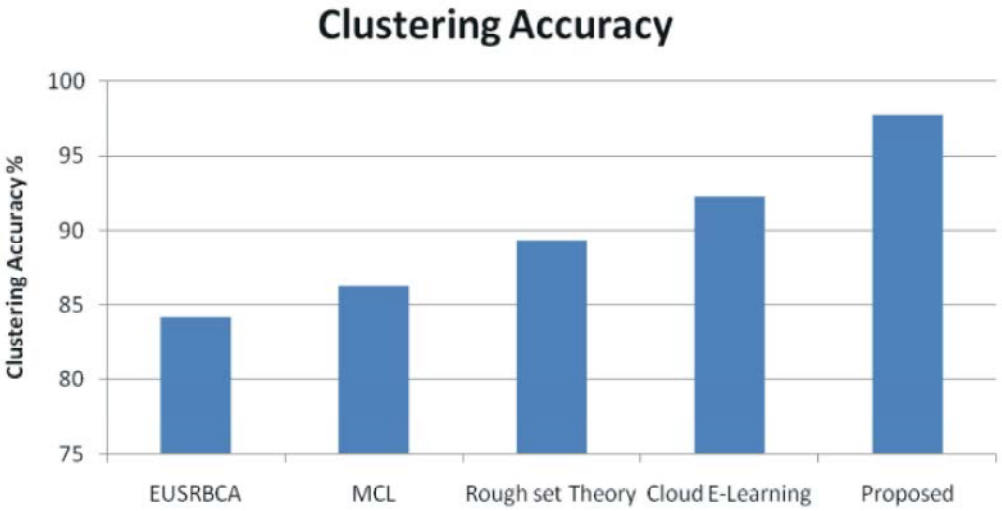
Stop.

The multi attribute ranking algorithm computes the multi user feedback weight based on various attributes of feedback for each tutor or anything. Using computed weight the results are re-ranked to produce more efficient results.

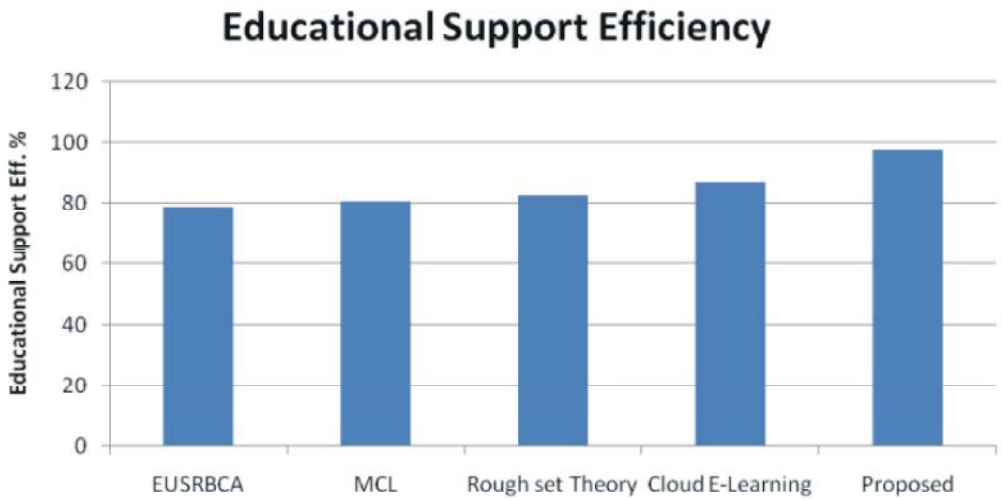
RESULTS AND DISCUSSION

The proposed method has been implemented and evaluated for its efficiency. The result shows that the proposed method has produced efficient results in all the factors of data warehousing and the method improves the performance of educational systems. The method can be easily adapted to rank and support the online education and E-learning approaches. Also the method has been focused to support various other solutions like business intelligence, Administrative support and many more.

The Graph 1, shows the comparison of clustering accuracy produced by different methods on warehousing the data. It shows clearly that the proposed method has produced more efficient result than other methods.



Graph 1: Comparison of clustering accuracy produced



Graph 2: Comparison of educational support efficiency

The Graph 2, shows the comparison of educational support efficiency produced by different methods and it shows clearly that the proposed method has produced more support than other methods.

CONCLUSION

The growing educational institutions needs more strategic tools for their development and to provide more efficient education to the students. In the motivation to develop a strategic support tool, we proposed a multi model multi level clustering based multi model inference framework which groups the data records using the MMML clustering and the input query has been identified for the presence of various topics and their sub classes. Based on the presence of topics the method extracts results and computed multi user feedback weight for each of the result entry. The re-ranked result will be produced to the user from which the user can choose the required class of result and further he can extend his selection according to many factors as like the tree framed in the topic identification. The proposed method has produced efficient result than other methods and produced more efficient educational support than other tools.

REFERENCES

1. Ramasubramanian, 2008. Data mining techniques for teaching result analysis using rough set theory, International Computing and Communication Conference.
2. Ping Lu-Zhi, 2009. Analysis of New Teaching Model for Undergraduates in University Based on the Environment of Information Technology - A Case, for Specialty of Information Management & Information System, Education Technology and Computer Science.
3. Srisaan, 2005. Experience from teaching performance analysis of object-oriented systems, Frontier Education.
4. Suresh, D. and S. Prakasam, 2013. Article: The Impact of E-learning system using Rank-based Clustering Algorithm (ESURBCA). International Journal of Computer Applications, 83(7): 13-18.
5. Valsamidis, S., S. Kontogiannis, I. Kazanidis, T. Theodosiou and A. Karakos, 2012. A Clustering Methodology of Web Log Data for Learning Management Systems. Educational Technology & Society, 15 (2): 154-167.
6. Kaur, Mandeep and Kewal Krishan, 2013. Cluster Analysis of Behavior of E-learners, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, 3(2).
7. Abdous, M. and W. Yen, 2012. Predicting Relationship between Online Questions Theme and Final Grade.
8. Namdeo Jyoti, 2014. Predicting Students Performance Using Data Mining Technique with Rough Set Theory Concepts, International Journal of Advance Research in Computer Science and Management Studies, 2(2).
9. Prof. Dr. Aytekin İŞMAN, 2011. instructional design in education: new model, TOJET: The Turkish Online Journal of Educational Technology, 10(1).
10. Samitha R. Babu, Krutika G. Kulkarni and K. Chandra Sekaran, 2014. A generic agent based Cloud Computing architecture for E-Learning , ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India- Vol I, Advances in Intelligent Systems and Computing, 248: 523-533.
11. Pireva Krenare, Petros Kefalas, Dimitris Dranidis and Thanos Hatziapostolou, Anthony Cowling, Cloud e-Learning: A New Challenge for Multi-Agent Systems, Agent and Multi-Agent Systems: Technologies and ApplicationsAdvances in Intelligent Systems and Computing, 296: 277-287.
12. Bennane, A., 2010. Tutoring and multi-agent systems: Modeling from experiences. Informatics in Education-An International Journal, 9(2): 171-184.
13. Chatzara, K., C. Karagiannidis and D. Stamatis, 2010. Students attitude and learning effectiveness of emotional agents. In: 2010 IEEE 10th International Conference on Advanced Learning Technologies (ICALT), pp: 558-559.
14. Dong Ping, 2006. Application of Data Warehouse Technique in Educational Decision Support System, IEEE Transaction on Service Operations and Logistics.
15. Dimokas, A., 2008. Prototype System for Educational Data Warehousing and Mining, IEEE Conference on Informatics.
16. Kazi, L., 2012. Data warehouse based evaluation of students' achievements in information systems education, IEEE Transaction on MIPRO.
17. Salwa Mohammed, 2015. Analysis of Data Warehousing and Data Mining in Education Domain, d Nejres, International Journal of Advances in Computer Science and Technology, 4(4):

18. Manjunath, T.N. Ravindra, S. Hegadi and G.K. Ravikumar, 2010. Analysis of Data Quality Aspects in DataWarehouse Systems, (IJCSIT) International Journal of Computer Science and Information Technologies, 2(1): 477-485.
19. Sha Bhavan, D., 2015. Big Data Mining in Education Sector: Critical Issues and Current Status in Indian Context, International Journal of Advanced Research in Computer Science and Software Engineering, 5(4).