**EDUCATIONAL DECISION SUPPORT SYSTEM AND INFORMATION RETRIVAL**

Abhijit Kumar, Himanshu Sati, Saurabh Tandan, Arunit Gupta, Sandeep Kumar

Department of Computer Science, Amity School of Engineering and Technology

Bijwasan, New Delhi

***Abstract—* Education always plays an important role in building up every country around the world. Hence, educational decision making support is significant to students, educators, and educational organizations. The support will be more valuable if a lot of relevant data and knowledge mined from data are available for educational managers in their decision making process. An academic credit system is nowadays very widely-used in many educational organizations. We propose a knowledge-driven educational decision support system for education with a semester credit system by taking advantage of educational data mining. Our resulting system can provide educational managers with actionable knowledge discovered from educational data. Such knowledge-driven decision support is helpful for educational managers to make more appropriate and reasonable decisions about student’s study. Above all, a waste of effort, time, and money can be avoided accordingly for both students and educators. This project will help students and managers to take appropriate decision for selecting electives subjects in final year of their studies. They are provided with a Decision Support System for selecting elective subjects and they can have access to Online Reference Material which depends on the knowledge level of the student, this will ultimately help students to take appropriate decision for their future.**

***Keywords—*** *Educational Decision Support System. PageRank, ODP Biasing.*

1. **INTRODUCTION**

Educational decisions made by educational managers are important and have a strong impact on not only individual students and educators but also our society. A Decision Support System (DSS) is a practical application that is tightly associated with the characteristics of the environment where the system is deployed and run. Implementation of this decision support system (DSS) in educational application domain is called Educational Decision Support System (EDSS). This implies that an EDSS is specific for each educational organization. Educational Decision Support System takes advantage of educational data mining for regular undergraduate students with a semester credit system. The proposed system is expected to provide the educational managers with problem-relevant data and knowledge prepared by means of data management and visualization, classification, clustering, and association analysis. Using the proposed system, the educational managers can have data and knowledge relevant to the problem they are considering as to student’s study. Together with their own experiences and knowledge, these data and knowledge will be useful evidences that support them in producing more appropriate and reasonable decisions in educational management. Once the decisions are effective, a lot of effort, time, and money can be saved for both students and educators.

As one of the first EDSSs, presented the advantages and disadvantages of an educational decision support system as well as of the development of such a system. There are a few types of educational decisions including programmed to nonprogrammed decisions. Moreover, several types of educational information can also be determined. An EDSS has provided some student’s result analysis utilities with tables and statistical charts. This system helped determining whether learning objectives had been achieved through data analysis on student’s learning results including entrance records, the ratio between the number of required semesters and the number of real semesters that students had taken, and the grade point averages. Another EDSS was developed to support strategic planning about educational resources, about distribution and usage of those resources at universities. Educational resources mentioned in [19] consisted of services about courses, supervision, etc. that students used in their study. The result of this system included reports used for decisions making support containing input data and temporary outputs simulating strategic plans about educational resources.

1. **BACKGROUND KNOWLEDGE**
2. *Background Knowledge of DSS* - Information Systems researchers and technologists have built and investigated computerized Decision Support Systems (DSS) for approximately 40 years. The journey begins with building model-driven DSS in the late 1960s, theory developments in the 1970s, and implementation of financial planning systems, spreadsheet-based DSS and Group DSS in the early and mid 1980s. Data warehouses, Executive Information Systems, OLAP and Business Intelligence evolved in the late 1980s and early 1990s. Finally, the chronicle ends with knowledge-driven DSS and the implementation of Web-based DSS beginning in the mid-1990s. The field of computerized decision support is expanding to use new technologies and to create new applications.
3. *Background of PageRank -* PageRank is a link analysis algorithm that assigns a numerical weighting to each element of a hyperlinked set of documents. The algorithm may be applied to any collection of entities with reciprocal quotations and references. A PageRank results from a mathematical algorithm based on the webgraph, created by all World Wide Web pages as nodes and hyperlinks as edges. The rank value indicates an importance of a particular page. A hyperlink to a page counts as a vote of support. The PageRank of a page is defined recursively and depends on the number and PageRank metric of all pages that link to it ("incoming links"). A page that is linked to by many pages with high PageRank receives a high rank itself. If there are no links to a web page, then there is no support for that page. PageRank is a probability distribution used to represent the likelihood that a person randomly clicking on links will arrive at any particular page. PageRank can be calculated for collections of documents of any size. It is assumed in several research papers that the distribution is evenly divided among all documents in the collection at the beginning of the computational process. The PageRank computations require several passes, called "iterations", through the collection to adjust approximate PageRank values to more closely reflect the theoretical true value.A probability is expressed as a numeric value between 0 and 1. A 0.5 probability is commonly expressed as a "50% chance" of something happening. Hence, a PageRank of 0.5 means there is a 50% chance that a person clicking on a random link will be directed to the document with the 0.5 PageRank.
4. **PROPOSED WORK**

Every semester there is a list of students to be warned for poor study performance or to be forced to stop their study for many various reasons one of which is also poor study performance. The rationale behind this policy is that the entire study of each student is viewed as a cumulative process of obtaining knowledge from all courses specified in an educational programme. Poor study performance at some point of time in the process implies that the student has not yet got enough knowledge to move forward to their successful graduation and the remaining knowledge the student needs to study more is too much for the student to be able to obtain in the rest of permitted time. Therefore, the educational managers have to make a decision about who could continue the study from the list. For those who could continue their study, the educational managers will give them another chance in one extra semester so that the students can improve their study and then get out of the list next semester. If a student is capable of keeping their study, the chance given to the student is really appreciated for the student to successfully accomplish the entire educational programme for graduation. Otherwise, the chance given to the student will be wasted in terms of student and educator’s effort, time, and money.

****

Figure 1. Architecture of the proposed knowledge-driven Educational decision support system

1. **PROJECT OVERVIEW**

In this project we are doing the following:

1. *Courses opted*

Student will go through an online registration in which he/she will be given choices for the course registration; choices will be given to them for choosing their elective subjects for their final year. This project deals with choosing elective subjects, which will be selected based on their previous performance. Final year students of computer science branch have following subjects in their syllabus:

Theory Subjects:

Advanced Computer Networks

Advanced Computer Architecture

Elective Subjects (any two):

Compiler Construction

Mobile Computing

Requirements & Estimation Techniques

Digital Image Processing

Theory Subjects are compulsory, but for electives only two choices are to be made from above options, so a fruitful decision is to be made, which will be made based on their previous performance in the subjects related to electives, which will help students to filter out best options for themselves.

1. *Student categorization*

Students are categorized according to the marks obtained in previous semesters into three categories:-

Beginner, Intermediate, Advance.

Student categorization will help educators to classify a huge strength of students into the above categories according to the performance of each student in previous semesters, which will help them to provide adequate decision to the students, so that they can perform well in future and get proper guidance from their mentor.

1. *Information retrieval*

We have used Google search engine for retrieving the learning material. We filter the retrieved web pages and divide them into three categories**:**

* Beginner , Intermediate , Advanced
* We can categorize each kind of learning material based on its content. Each web page contains information about its category. Apart from classification, document clustering is also a very popular approach to organize information in an efficient way. Clustering organizes information by grouping similar or related entities together. In classification, categories are first determined and documents are assigned to them according to characteristics of those objects.

1. **TOPIC SENSITIVE PAGERANK**
2. *OUTLINE OF TOPIC SENSITIVE PAGERANK*

In our approach to topic-sensitive PageRank, we pre-compute the importance scores online, as with ordinary PageRank. However, we compute multiple importance scores for each page; we compute a set of scores of the importance of a page with respect to various topics. At query time, these importance scores are combined based on the topics of the query to form a composite PageRank score for those pages matching the query. This score can be used in conjunction with other IR-based scoring schemes to produce final rank for the result pages with respect to the query. As the scoring functions of commercial search engines are not known, in our work we do not consider the effect of these other IR scores. We believe that the improvements to PageRank's precision will translate into improvements in overall search rankings, even after other IR-based scores are factored in.

1. *ODP Biasing*

The first step in our approach is to generate a set of biased PageRank vectors using a set of basis topics. This step is performed once, online, during the preprocessing of the Web crawl. For the personalization vector p, we use the URLs present in the various categories in the ODP. We create 16 different biased PageRank vectors by using the URLs present below each of the 16 top-level categories of the ODP as the personalization vectors. In particular, let Tj be the set of URLs in the ODP category cj. Then when computing the PageRank vector for topic cj, in place of the uniform damping vector p = [1/ N],we use the non-uniform vector p = vj where

The PageRank vector for topic cj will be referred to as PR (α, vj). We also generate the single unbiased PageRank vector (denoted as No-Bias) for the purpose of comparison.

We also compute the 16 class term-vectors Dj consisting of the terms in the documents below each of the 16 top-level categories. Djt simply gives the total number of occurrences of term t in documents listed below class cj of the ODP.

1. *QUERY TIME IMPORTANCE SCORE*

The second step in our approach is performed at query time. Given a query q, let q0 be the context of q. In other words, if the query was issued by highlighting the term q in some Web page u, then q0 consists of the terms in for ordinary queries not done in context, let q0 = q. Using a unigram language model, with parameters set to their maximum-likelihood estimates, we compute the class probabilities for each of the 16 top-level ODP classes, conditioned on q0.

Let q0 i be the ith term in the query (or query context) q0. Then given the query q, we compute for each cj the following:

Using a text index, we retrieve URLs for all documents containing the original query terms q. Finally, we compute the query-sensitive importance score of each of these retrieved URLs as follows. Let rankjd be the rank of document d given by the rank vector PR(α,vj ) (i.e., the rank vector for topic cj ). For the Web document d, we compute the query-sensitive importance score as follows.

1. *SIMILARITY MEASURES FOR INDUCED RANKING*

We use two measures when comparing rankings. The first measure, denoted

*OSim(T1,T2):* Indicates the degree of overlap between the top n URLs of two rankings. The overlap measure OSim gives an incomplete picture of the similarity of two rankings, as it does not indicate the degree to which the relative orderings of the top n URLs of two rankings are in agreement. Therefore, we also use a variant of the Kendall's T distance measure.

*KSim(T1,T2):* Indicates the degree to which the relative orderings of the top n URLs of two rankings are in agreement. We define our similarity measure KSim as follows:

1. **RESULT**

We performed Educational Decision Support System to provide decision support for choosing elective subjects based on student’s past performance and to categorize students on the basis of their marks into 3 categories - Beginner, Intermediate, Advanced

This system also provided the students with a Learning material/Information Retrieval Module which will help them in extracting the exact and accurate information from the WWW (World Wide Web) in an optimized manner.

1. **CONCLUSION**

Educational Decision Support System and Information Retrieval is an efficient asset in the institutions for maintaining the various activities and can be developed and increased to an unlimited extend. The application has met its goal of providing a well-organized structure of a DSS for choosing elective subjects and provides an information retrieval module for efficient content searching which depends on the student’s knowledge level and administrator or educational manager can easily relate it with the manual system used before. The Educational Decision Support System once developed will adhere very closely to the present needs of the institutions to deliver the institution related information to the educational managers. Student’s commitment is increased with the opportunity to discover one’s capability and potential in one’s work.

1. **FUTURE SCOPE**

The system can be modified and upgraded in various senses ranging from implementation level to user interface and functionality. Student categorization can be done by conducting an online aptitude examination for the entire registered student. Input taken from the student can be increased like hobbies of the student, favorite subjects, work experience details, industrial training and academic projects done; this will ultimately enhance the DSS capability and will provide a wide variety of decisions.Information retrieval or content searching can be enhanced or can be customized to a greater extend by implementing better algorithms.

1. **ACKNOWLEDGEMENT**

With profound respect and gratitude, we would like to take this opportunity to convey our deep gratitude to **Prof. Dr. Rekha Aggarwal (Director)** and **Prof. M. N. Gupta** for providing the infrastructure and labs in ASET. We would like to thank our guide **Mr. Abhijit Kumar** his continuous help and support throughout the making of this project. Without his help, the project could not have been accomplished. We also acknowledge our sincere thanks to our Lab Assistants and the faculty of ASET for providing us with the guidance to steer our efforts in right direction. Last but not the least; we are thankful to our parents without whose efforts and blessing we could not come this far in our pursuit for success.

1. **REFERENCES**

[1] Vo Thi Ngoc Chau, Nguyen Hua Phung , “A Knowledge-Driven Educational Decision Support System”, in IEEE RIVF International conference, University of Technology Ho Chi Minh City, Vietnam 2011, pp. 1-6.

[2] Taher H. Haveliwala ,“TopicSensitive PageRank”, in the Proceedings of the 11th international conference on World Wide Web, Standford University,Standford, 2003, pp.517-526.

[3] M. Zorrilla, D. Garcia, and E. Alvarez, “A Decision Support System to Improve e-Learning Environments,” in Proc. of EDBT 2010, 2010.

[4] D. Z. Deniz and I. Ersan, “An Academic Decision-Support System Based on Academic Performance Evaluation for Student and Program Assessment,” Int. J. Engng Ed., vol. 18, pp. 236-244, 2002.