Recommender System for Stimulating the Learning Skill of Slow Learner in Higher Educational Institution using EDM.

S. JothiLakshmi, M.Thangaraj

Abstract Educational data mining is a prominent area to predict the student learning behaviour. A regular analysis of student learning behaviour is used to improve the quality of the academic institution. This paper focuses on the identification of slow learner in the higher educational institution and showing it by educational data mining model. The student performance data from higher educational institution is taken and applied on the classification algorithm using WEKA tool. A recommender system is developed in this research work to stimulate the learning skills of slow learner.

Key words- Data mining, classification, Slow learner, Recommender, EDM.

I. INTRODUCTION

Data mining has fascinated interest from researcher in many different field including design of the database, statistical analysis, pattern recognition technique, machine learning technique and data visualization. Knowledge discovery process is the method of analyzing the information from diverse points of view and summarizing the result as valuable data. Educational data mining combined the three major areas such as as computer science, education and statistics EDM is the primary utility of the data mining techniques.

The EDM process begins with gathering the various kinds of data from the educational environment. The collected raw educational data required to pre-process the data. Once the data is pre processed, the corresponding EDM algorithm is applied. Educational Data mining research work involved in the variety of areas, including Student modelling, Course modelling, Decision support system, generating recommendation. EDM methods applied in improvement of student models. Its application is used to predict slow learner, dropout student, under performer etc.

Various kind of research work involved to improve the overall higher educational system. There are group of factors which affects the education system like the student performances, infrastructures, teacher, teaching methodology, basic amenities, transport facilities etc. The EDM techniques are performed on these factors and find out the factors which improve the higher educational system.

The purpose of the research work is to categorize the various factors associated with student performance of higher educational institutional institution and develop the quality by identifying the slow learner.

In this research work, classification technique is used to analyze the student performance, identification of slow learner and design a recommender system to stimulate the learning skills of slow learner which is used to get better performance in the higher educational institution.

II. RELATED WORK

This section provides the detailed study of previous research work on student performance. Han and Kamber describes data mining process is a knowledge discovery in Databases. Educational data mining is the one of the major application of data mining. Educational data mining is an emerging field in the educational environment which can be applied in various areas like sports, accounts transport etc.[12] This paper reviews the CHAID prediction model to predict the student performance of school. [12]

Amajad,Abu applied ID3,C4.5 classification algorithm to construct a tree which is used to predict the student academic performance in the educational institution[5].Nauyen designed a model to improve the student result. They analyzed class imbalance problem by three methods and achieved satisfactory results.

Brijesh Kumar designed an EDM model using the classification technique to extract knowledge from student database that describes the student performance in the end semester exams [8]. Dorina Kabakchieva used clustering and classification technique to design a model to analyze the student overall performance in learning management system [10].FP tree mining was applied by Arockiam et al to study the performance of rural and urban students programming skills. This study shows that urban student skills were good compared to rural students.

Parmeet Kaur conducted a study on classification algorithms to predict slow learner in the educational institution. They used various classification algorithms in their study such as Multilayer perception. Navie Bayes, SMO,J48.REP Tree applied on 152 student records. It was reported that Multilayer algorithm had the highest predictive accuracy of 75%, compare to the other classification algorithm. These various research works motivates to design a recommender system to stimulate the learning skills of slow learner as a proposed system.[19]

III. PROPOSED FRAME WORK

This section gives the detailed description of the proposed system Slow Learner Recommender-Educational Data Mining (SLR-EDM) is given with important components. The proposed system is composed of three main layers named as 1 data layer, Pre-process Layer Pattern discovery layer, and...
Recommendation layer. The system comprises two separate phases i) Construction of Decision tree ii) generation of recommendation. The result of the recommendations will be used to stimulate the learning skill of slow learner in the higher educational institution. For this work, higher educational student performance data is identified from higher educational institution. The function of this architecture is explained as follows.

A. Data Layer
The Data layer collects and stores the student result data from various higher educational institutions. Here data layer integrates various data bases in the area of various educational institutions.

B. Pre-processor Layer
The Pre-processor Layer involves Data pre-processing work. Data is originated from different databases. It is susceptible to noise, missing value and inconsistency. Pre-processing technique is used to get better quality of the data and get correct mining result. So the data needs to be pre-processed by using certain pre-processing component.

C. Pattern Discovery Layer
The Pattern Discovery layer generates patterns using the C4.5 decision tree algorithm. It uses the training data (pre-processing and filtering prepares the training data) as input data for generating the decision tree. This tree is used by pattern discovery component to generate rules for recommendations.

Procedure:C4.5(V)

begin
N=empty; /* Tree Node initially empty */
S_best=empty; /* Attribute for storing max. Information gain */
if V is “pure” OR other stopping criteria met then
terminate
end if
for all attribute s £ V do /* calculate information criteria */
Calculate information-gain if we split on s.
end for
S_best=attribute with MAX_INFO gain;
X=MIN_INFO gain;
If(S_best>X) then

N= S_best; /* create S_best as root node */
V=Induced sub-datasets from V based on S_best
/* partitioning the data set */
for all V_i do
N_i=C4.5(V_i)
Attach N_i to the corresponding branch of N
end for
return N
end

C. Recommendation Layer
The main function of Recommendation layer is to generate recommendation rules and provides the possible solution to stimulate the learning skills of slow learner in the higher education institution.

IV. EXPERIMENTAL SETTINGS
SLR-EDM has been implemented with JAVA language using Net Bean version 7.3 as JAVA environment. The Experiments were done Intel core i3 2.1GH 4GB RAM, running on windows 8. The student result data have been collected from The experimental evaluation aims analysing whether the recommendation module of the SLR-EDM model is useful to the higher educational institution. The student result data of various distances learning study centre are used in this model. Table I shows the Attribute set selected by SLR-EDM system and which is used to create a Decision tree.

Fig a shows the student result data set and Fig, b show the decision tree

Table I. Higher educational Student related variable

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description of the variable</th>
<th>Domain of the variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Student’s sex</td>
<td>[Male,Female]</td>
</tr>
<tr>
<td>Area of the school</td>
<td>Area of the school</td>
<td>[urban,Rural]</td>
</tr>
<tr>
<td>Fa_Income</td>
<td>Father Income</td>
<td>[high,low,medium]</td>
</tr>
<tr>
<td>Pat_Edu</td>
<td>Parent Education</td>
<td>[yes,No]</td>
</tr>
<tr>
<td>HSC_class</td>
<td>Higher Secondary Result</td>
<td>[First,Second]</td>
</tr>
<tr>
<td>HSC_Medium</td>
<td>Higher Secondary Medium</td>
<td>[Tamil,English]</td>
</tr>
<tr>
<td>HSC_School</td>
<td>Higher Secondary School</td>
<td>[Govt/Private]</td>
</tr>
<tr>
<td>HSC_Brd</td>
<td>Higher Secondary Board</td>
<td>[State,CBSE,ICSE]</td>
</tr>
<tr>
<td>Int_ASS</td>
<td>Internet Access</td>
<td>[yes,No]</td>
</tr>
<tr>
<td>Class Att</td>
<td>Class Attendance</td>
<td>[Poor,Avg,Good]</td>
</tr>
<tr>
<td>IQ level of the student</td>
<td>Student IQ level</td>
<td>[poor,good]</td>
</tr>
<tr>
<td>Writing_Skill</td>
<td>Student Writing Skill</td>
<td>[Poor,Good]</td>
</tr>
<tr>
<td>Int_Test Grade</td>
<td>Internal Test Grade</td>
<td>[Pass,Fail]</td>
</tr>
<tr>
<td>Class</td>
<td>Wheather pass or fail</td>
<td>[pass,fail]</td>
</tr>
</tbody>
</table>
A. Recommendation rules

Production rules are generated from the Decision trees that are used by recommendation agent to provide the recommendation.

1. If IQ level of the student=good then ITG=good then result=pass
2. If IQ level of the student =poor and int_Ass=no and ITG=poor result=fail
3. If Att=poor then result =fail
4. if Int_ass =no and Att= poor then result=fail
5. If att=good result=pass
6. If ITG="AVG" then Atten=AVG result=fail
7. If att="avg" and ITG="AVG" then result fail
8. If att= avg and ITG="poor" then result=fail

V. EXPERIMENTAL EVALUATION

The feature of a recommender system can be assessing using various kinds of dimension. Accuracy is the fraction of correct recommendations out of total possible recommendations The proposed system is assessed with Decision support accuracy that are popularly used are Precision, Recall and F-measure. These metrics help the user in selecting items that are very high quality out of the available set of items. The evaluation has been done by the measuring of the parameters Precision, Recall and F-measures.

A. Experiment 1: Precision Measure

Fig 5.1 summarizes the student result data versus precision for SLR-EDM and SLRM. X-axis denotes the various size of result data set and Y-axis denotes the corresponding Precision value. From the graph it is proven that the precision value of proposed SLR-EDM system is improved by 17% compared to the existing system. This result shows that the SLR-EDM precision is better than SLR.

<table>
<thead>
<tr>
<th>Result Data set</th>
<th>500</th>
<th>1000</th>
<th>1500</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLRM(Precision)</td>
<td>0.761</td>
<td>0.768</td>
<td>0.775</td>
<td>0.782</td>
</tr>
<tr>
<td>SLR-EDM(Precision)</td>
<td>0.898</td>
<td>0.954</td>
<td>0.958</td>
<td>0.975</td>
</tr>
</tbody>
</table>

B. Experiment 2: Recall Measure

The recall refers to the percentage of total relevant result correctly classified. The Fig 5.2 summarizes the student result data set versus recall values for SLR-EDM system. The student result data is shown in X-axis and corresponding recall value in Y-axis. The highest recall value is obtained with the data set of 2000. Fig 4.6 shows the comparison of Recall and mean Recall value The Recall value is increased by 11% in SLR-EDM system.

<table>
<thead>
<tr>
<th>Result Data set</th>
<th>500</th>
<th>1000</th>
<th>1500</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLRM(Recall)</td>
<td>0.819</td>
<td>0.824</td>
<td>0.829</td>
<td>0.835</td>
</tr>
<tr>
<td>SLR-EDM(Recall)</td>
<td>0.938</td>
<td>0.926</td>
<td>0.905</td>
<td>0.934</td>
</tr>
</tbody>
</table>
C. Experiment 3.F-measure

F-Measure is a single measure that trades off precision versus recall, which is the weighted harmonic mean of precision and recall. Fig 5.3 summarizes the student result data and corresponding F-measure value. The X-axis consists of student result data and Y-axis consists of F-measure value. The system achieved improvement by way of 14% in F-Measure.

<table>
<thead>
<tr>
<th>Result Data set</th>
<th>SLRM (F-Measure)</th>
<th>SLR-EDM (F-Measure)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.789</td>
<td>0.918</td>
</tr>
<tr>
<td>1000</td>
<td>0.794</td>
<td>0.94</td>
</tr>
<tr>
<td>1500</td>
<td>0.801</td>
<td>0.93</td>
</tr>
<tr>
<td>2000</td>
<td>0.807</td>
<td>0.954</td>
</tr>
</tbody>
</table>

Fig f. F-measure

DISCUSSION

The primary goal of this work has been architected a SLR-EDM frame work which is a recommender system that uses educational data mining technique to stimulate the learning skills of slow learner in the higher educational institution. The Decision tree classification mining techniques is implemented by the SLR-EDM systems. The experimental results are observed for various size of result data set showing different level of performance of the proposed system. The evaluation of the result has been done for the precision, recall and F-measure which compare the precision values of the test data and it is observed the precision value of proposed system is improved by 13% compare to the existing system. The recall value is increased by 11% and improvement of F-measure value is 14% compare to the existing system. The experimental result shows that the SLR-EDM outperforms the existing system and it provides recommendation rules to improve the learning skills of slow learner in the higher educational institution. The followings factors are recommended by the SLR-EDM model to improve learning skills of the slow learners in the higher education institution.

1. Teachers should be made aware of the importance of reading.
2. Teachers guide the students to the important points of lessons and internal test.
3. Teachers give class lessons on study skills.
4. Teachers use audio and video aids for teaching the subjects.
5. Teachers give more assignments on the important lessons.

CONCLUSION

The purpose of this research work is to stimulate the learning skills of slow learner in higher educational institution. In this paper classification technique is used to design a model for prediction on the result data of 2000 students to predict the slow learners. This research work is used to identify slow learners in the higher educational institution and provides the recommendation to improve the skills of slow learners. A model performance chart is plotted. Integration of educational data mining techniques on the different data set is the future work.

REFERENCES


AUTHOR PROFILE

S. Jothi Lakshmi working as assistant professor in the Department of computerscience Mannar Tirumalai Naicker College, Madurai She received the Master degree in computer science from R.V.S college of Engineering, Dindigul & M.Phil(CS) from Annamalai University. Her interest include, Data Mining, Big data Analytics, wireless sensor networks. She has published more than 20 papers in journals and conference proceedings. She can be reached at jothi2008@gmail.com

Dr. M Thangaraj, Professor & Head, Department of Computer Science, School of Information Technology, Madurai Kamaraj University, India. M Thangaraj received his post-graduate degree in Computer Science from Alagappa University, Karaikudi, M.Tech. degree in Computer Science from Pondicherry University and Ph.D. degree in Computer Science from Madurai Kamaraj University, Madurai, TN, South India in 2006. He is now the Head of Computer Science Department at M.K.University. He is an active researcher in Big Data Analytics, Social Media Analytics, Wireless Sensor Networks and has published more than 100 papers in Journals and Conference Proceedings.