LEARNING RESOURCES RECOMMENDATION
FRAMEWORK FOR MOODLE BASED ON ANALYSIS OF MOSTLY ACCESSED RESOURCES BY GOOD STUDENTS

Sivakumar Tharsan
128232U

Department of Computer Science and Engineering
University of Moratuwa
Sri Lanka
June 2014
LEARNING RESOURCES RECOMMENDATION
FRAMEWORK FOR MOODLE BASED ON ANALYSIS OF
MOSTLY ACCESSED RESOURCES BY GOOD
STUDENTS

Sivakumar Tharsan

128232U

Department of Computer Science and Engineering
University of Moratuwa
Sri Lanka

June 2014
DECLARATION

“I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).”

--------------------------------     ------------------------------
Signature         Date

(Sivakumar Tharsan)
MSc in Computer Science / 128232U

The above candidate has carried out the research for the Masters thesis under my supervision.

--------------------------------     ------------------------------
Signature         Date

Dr. Shehan Perera (Supervisor)

Department of Computer Science and Engineering

University of Moratuwa.
Learning resources recommendation framework for Moodle based on analysis of mostly accessed resources by good students

ABSTRACT

We live in an e-learning era, where the fast growth of e-learning around the world is inspiring many educational and business institutions. The rapid growth of e-learning has changed the traditional learning behavior and presented a new situation to both lecturers and students. Lecturers find it harder to guide students to select suitable learning materials due to overdosed learning resources. In the other hand learners spend more time on browsing and filtering learning resources to identify the suitable materials, rather than focusing on learning all the materials. Hence, the learning material recommendation is an essential requirement for e-learning system, which would take the e-learning, to the next level.

Though e-learning brings many benefits in comparison with the conventional learning paradigm, with the rapid increase of learning contents on the web, the e-learning lacks proper feedback or the guidance, which is a key in the traditional teaching process to identify the relevant resources. The students who do not have knowledge to find out the most suitable resources, links and references for their studies and the assignments, may waste most of their time on browsing in search of the relevant material, without any guidance. Some of the “good students” may indirectly act as good guides to their fellow friends. The fellow average learners could follow the methods adapted by good students in learning or accessing relevant learning material. They may refer to the mostly accessed “learning materials” by the “good learners”. The lack of feedback in e-learning systems could be overcome by evolving an intelligent feedback system that would recommend the “mostly accessed resources” by the good students to the average students and that is the aim of this project.

The research collected the Moodle log data of the courses that contain a rich set of electronic course contents from Department of Computer Science and Engineering, University of Moratuwa. This audited data of the collected courses have been used to construct “student classification” and “resource recommendation” models. The former classifies the students as “Good students” and “Average Students” and the latter recommends the “mostly accessed resources” by good students.

The results show that the resources that are mostly accessed by good students are more probable to be recommendable resources. The learning resource recommendation framework helps the average students who fail to choose the relevant materials for their studies from the heaps of learning resources. By making the website interactive, the communication between faculty and students could be made more effective, hence that would promote active learning.

Key words: e-learning, recommendation framework, access pattern, learning analytics, Moodle
ACKNOWLEDGEMENT

I take pleasure to acknowledge the active support and valuable advice provided by my project supervisor Dr. Shehan Perera. The constant guidances and frequent reviews had enabled the successful completion of this research project. His continuous support was available via meetings, email and telephone communications throughout the project duration. Whenever I experienced constraints, this gentleman had the generosity in providing the necessary advice and suggestions that helped me a lot in completing this task.

Also I am glad to express my sincere gratitude to Dr. Shantha Fernando for granting the access to the Moodle backups which were used for data analytics and validate the research models. In addition to that that Mr. Pradeep Manoj, the course assistant was much cordial with regard to obtaining the necessary facts and figures through the course backups and final grades. Last but not least I thank to all who helped me in numerous ways in order to make this project a success.
TABLE OF CONTENTS

DECLARATION .......................................................................................................................... I

ABSTRACT ................................................................................................................................... II

ACKNOWLEDGEMENT ............................................................................................................... III

TABLE OF CONTENTS ............................................................................................................... IV

LIST OF FIGURES ..................................................................................................................... VI

LIST OF TABLES ........................................................................................................................ VIII

LIST OF EQUATIONS ................................................................................................................ IX

LIST OF ABBREVIATIONS .......................................................................................................... X

LIST OF APPENDICES ............................................................................................................. X

1 INTRODUCTION .................................................................................................................. 1

1.1 MOTIVATION FOR THE RESEARCH .............................................................................. 1

1.2 RESEARCH OBJECTIVES ............................................................................................... 1

1.3 RESEARCH DESIGN ......................................................................................................... 2

1.3.1 The scope of the research ............................................................................................. 2

1.3.2 The research method ...................................................................................................... 2

1.3.3 The data collection and analysis ................................................................................... 3

1.4 THESIS OUTLINE ........................................................................................................... 3

2 LITERATURE REVIEW .......................................................................................................... 4

2.1 E-LEARNING SYSTEMS .................................................................................................... 4

2.2 THE DATA MINING IN E-LEARNING .............................................................................. 5

2.3 PREVIOUS RESEARCHES IN LEARNING ANALYTICS ...................................................... 6

2.4 LEARNING RESOURCE RECOMMENDATIONS ................................................................ 7

2.4.1 Recommendation Based on Good Learners’ Rating .................................................... 7

2.4.2 Recommendation Based on Good learners’ Access Pattern .................................... 10

2.4.3 Recommendation Based on Collaborative Filtering .................................................. 10

2.4.4 Recommendation Based on Content Filtering ............................................................ 11

3 METHODOLOGICAL FRAMEWORK .................................................................................. 14

3.1 COLLECT DATA ............................................................................................................. 16

3.1.1 Remove the user identity from the course data ........................................................... 19

3.1.2 Extraction of moodle data and store it in the database .............................................. 23

3.1.3 Extraction of moodle resource data ............................................................................ 29

3.2 TRANSFORM DATA ........................................................................................................... 31

3.2.1 Dropout students .......................................................................................................... 32

3.2.2 The Computed attributes of student classification model .......................................... 33

3.2.3 Update the good students in the student table............................................................ 60
3.2.4 The computed attributes of resource recommendation model ............................................... 61
3.3 REDUCE DATA.............................................................. 69
  3.3.1 RapidMiner ............................................................................................................................ 70
  3.3.2 The reason for RapidMiner ................................................................................................... 70
  3.3.3 Data reduction for student classification model ...................................................................... 71
3.4 PARTITION DATA................................................................................................................................ 74
3.5 BUILDING MODELS ................................................................................................................................ 75
  3.5.1 Why classification .................................................................................................................. 75
  3.5.2 Student classification model .................................................................................................... 79
  3.5.3 The resource recommendation model .................................................................................. 91
3.6 EVALUATE AND CHOOSE MODELS............................................................................................. 97
  3.6.1 The performance factors used to evaluate the models ........................................................ 98
  3.6.2 Data quality issues ................................................................................................................. 99
  3.6.3 Evaluate student classification model ................................................................................... 99
  3.6.4 Evaluate resource recommendation model ........................................................................ 103
4 ANALYSIS OF THE RESULTS ................................................................................................................. 106
  4.1 THE ANALYSIS OF STUDENT CLASSIFICATION MODEL .................................................. 106
    4.1.1 The accuracy analysis of student classification model ....................................................... 106
    4.1.2 RMSE analysis of student classification model .................................................................... 111
    4.1.3 The best student classification model .................................................................................... 112
  4.2 THE ANALYSIS OF RESOURCE RECOMMENDATION MODEL ........................................ 114
    4.2.1 The accuracy of resource recommendation model .......................................................... 114
    4.2.2 The RMSE analysis of resource recommendation model ............................................. 117
    4.2.3 The best resource recommendation model .......................................................................... 118
5 CONCLUSION AND RECOMMENDATION .............................................................................................. 121
  5.1 DISCUSSION................................................................................................................................ 121
    5.1.1 The analytical models ............................................................................................................ 121
    5.1.2 The student model ................................................................................................................. 123
    5.1.3 The resource recommendation model ................................................................................ 123
  5.2 RECOMMENDATION .................................................................................................................... 124
  5.3 LIMITATIONS OF THE RESEARCH ......................................................................................... 124
  5.4 FUTURE RESEARCH ..................................................................................................................... 125
6 REFERENCES ........................................................................................................................................... 126
APPENDIX A: PROGRAMS ....................................................................................................................... 130
APPENDIX B: RESULTS ............................................................................................................................ 132
LIST OF FIGURES

Figure 2-1 Rating based recommendation ([4]) ................................................................. 9
Figure 2-2 PLRS Framework ([2]) .................................................................................. 12
Figure 3-1 High level design of proposed recommendation system ..................... 14
Figure 3-2 The modeling process .................................................................................. 16
Figure 3-3 Folder structure of an archived course ......................................................... 18
Figure 3-4 The intermediate steps of data set preparation .......................................... 19
Figure 3-5 Anonymize the students' private data to lessen confidential leakage ........ 21
Figure 3-6 Solution for encoding issues in jdom ........................................................... 22
Figure 3-7 Grab data from moodle.xml and transfer it to the database ..................... 24
Figure 3-8 Insert the grabbed data into the database ..................................................... 25
Figure 3-9 Sets up the connection to the SQL server .................................................... 26
Figure 3-10 Get the connection object and executes the query string ....................... 26
Figure 3-11 SQL server error in Windows authentication mode .................................... 27
Figure 3-12 The code snippet for the data extraction of Moodle resources ............... 30
Figure 3-13 Distinguishes the dropped students ............................................................ 33
Figure 3-14 Eliminates the students with NULL grades ............................................... 35
Figure 3-15 Computes the Average Ratio of each student .......................................... 36
Figure 3-16 Collects all of the opened discussions of each course into a temp table .... 39
Figure 3-17 Assess the percentage of open discussions of each student .................... 40
Figure 3-18 Picks all of the replies by the student per course ....................................... 41
Figure 3-19 Updates the percentage of open discussion of student ............................ 42
Figure 3-20 The percentage of wiki entries per student ................................................. 44
Figure 3-21 Quantifies the percentage of assignment of completion of each student ... 46
Figure 3-22 Assess percentage of visit of course ........................................................... 49
Figure 3-23 Accumulates the total number of resources exist in each course ............. 52
Figure 3-24 Quantifies the percentage of resource access .......................................... 52
Figure 3-25 The number of late accesses by a student is getting composited .......... 54
Figure 3-26 Acquire the percentage of last access of a student .................................... 55
Figure 3-27 The variation of skewness (source: www.pertrac.com) .............................. 57
Figure 3-28 The skewness of access of some of the students selected for this research.. 58
Figure 3-29 The selection of StudentResourceLogPivot table ..................................... 59
Figure 3-30 Update the skewness of student in the Students table ............................. 60
Figure 3-31 Classify the students based on the rule learnt by the best model ............ 61
Figure 3-32 The percentage of access of a single resource .......................................... 63
Figure 3-33 Percentage of access of a resource by good students ............................. 64
Figure 3-34 The position of resource in week ............................................................... 66
Figure 3-35 The position of section in the course ......................................................... 67
Figure 3-36 Percentage of good students accessed a resource .................................... 68
LIST OF TABLES

Table 3-1  List of Courses selected for the research.................................................................17
Table 3-2  List of tables in the Moodle database filled by moodle.xml..................................27
Table 3-3  A list of tables which were derived from tables in Table 3.2.................................31
Table 3-4  The mean of the averages obtained in all assignments for each course ..........36
Table 3-5  Number of open discussions per course...............................................................39
Table 3-6  The number of replies posted per course...............................................................41
Table 3-7  Number of wiki entries to each course.................................................................43
Table 3-8  The number of assignments given in every course..............................................46
Table 3-9  Some of the students’ resource access volume.....................................................57
Table 3-10 The boundary of each Student label..................................................................90
Table 3-11 The percentage of access of a resource by good students..............................96
Table 3-12 The accuracies of student models with manual data fragmentation for 2 labels ........................................................................................................................................100
Table 3-13 The accuracies of student models with random data fragmentation for two class labels ........................................................................................................................................102
Table 3-14 The RMSEs of student models with manual data fragmentation for 2 class labels ........................................................................................................................................103
Table 3-15 The accuracies of resource models with manual split for two class labels........104
Table 3-16 The accuracies of resource models with random split for two class labels........104
Table 3-17 The RMSEs of resource models with random data fragmentation for two class labels ........................................................................................................................................105
Table 4-1  The number of highly accessed resources at each different splitting point........115
LIST OF EQUATIONS

Equation 2-1  Average rating of learning material ................................................................. 7
Equation 2-2  The weight of a document j respective to the term i ................................. 8
Equation 3-1  Ratio of average marks (Average ratio) .......................................................... 35
Equation 3-2  The percentage of open discussion .............................................................. 38
Equation 3-3  The percentage of replies to the forum discussion ..................................... 41
Equation 3-4  The percentage of wiki entries ........................................................................ 44
Equation 3-5  The percentage of assignment completion .................................................. 45
Equation 3-6  The percentage of visits of student ............................................................... 48
Equation 3-7  The percentage of resource access ............................................................... 51
Equation 3-8  The percentage of last minute access ............................................................ 53
Equation 3-9  Microsoft Excel’s skew function ................................................................. 59
Equation 3-10 The percentage of access of a resource ..................................................... 63
Equation 3-11 The percentage of access of a resource by good students ....................... 65
Equation 3-12 The percentage of access of a resource by average students ................. 65
Equation 3-13 The percentage of access by good students .............................................. 68
Equation 3-14 The accuracy of prediction ............................................................................ 98
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOODLE</td>
<td>Modular Object Oriented Dynamic Learning Environment</td>
</tr>
<tr>
<td>LMS</td>
<td>Learning Management System</td>
</tr>
<tr>
<td>PLRS</td>
<td>Personalized Learning Recommendation System</td>
</tr>
<tr>
<td>PAWS</td>
<td>Perdue Early Warning System</td>
</tr>
<tr>
<td>CMS</td>
<td>Course Management System</td>
</tr>
<tr>
<td>VLE</td>
<td>Virtual Learning Environment</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machines</td>
</tr>
<tr>
<td>K-NN</td>
<td>K-Nearest Neighbour</td>
</tr>
<tr>
<td>SQL</td>
<td>Structured Query Language</td>
</tr>
<tr>
<td>CSV</td>
<td>Comma Separated Values</td>
</tr>
<tr>
<td>XML</td>
<td>Extensive Markup Language</td>
</tr>
</tbody>
</table>

**LIST OF APPENDICES**

Appendix A: PROGRAMS ................................................................................................................................... 130
Appendix B: RESULTS ....................................................................................................................................... 132