Discovery of Course Enrollment Patterns Using Sequential Data Mining

Yei-Sol Woo
Indiana University - Purdue University Fort Wayne

Follow this and additional works at: http://opus.ipfw.edu/stu_symp2014

Part of the Computer Sciences Commons

Recommended Citation
http://opus.ipfw.edu/stu_symp2014/6

This Poster is brought to you for free and open access by the IPFW Student Research and Creative Endeavor Symposium at Opus: Research & Creativity at IPFW. It has been accepted for inclusion in 2014 IPFW Student Research and Creative Endeavor Symposium by an authorized administrator of Opus: Research & Creativity at IPFW. For more information, please contact admin@lib.ipfw.edu.
Discovery of Course Enrollment Patterns Using Sequential Data Mining

Yei-Sol Woo
Department of Computer Science, Indiana University - Purdue University Fort Wayne

Introduction

Motivation
- Universities collect a lot of data on students' information and course enrollment history inside databases.
- This study is focused on finding the most common course taken patterns prior to students finishing their degree or students dropping out of their major.
- Analyzing course enrollment history data and discovering patterns using sequential data mining methods could help education industries in many ways:
  1. Predicting which courses to offer for the next semester.
  2. Identifying frequent courses that students took before they drop out of their degree successfully.
- Research Questions
  1. Identify frequent course enrollment patterns until students finish their degree successfully.
  2. Identify frequent courses that students took before they drop out of their major (e.g., CS).

Methodology and Procedure

Preprocessing Step:
- I wrote an application to parse the exported CSV data for students who graduated or dropped the CS degree, and export each student's course enrollment history into a data file for SPMF to analyze.

Data Mining Step:
- To solve sequential pattern mining problem, I used SPMF (Sequential Pattern Mining Framework) which is an open-source data mining library written in Java by Philippe Fournier-Viger.
- I used PrefixSpan algorithm proposed by Pei, Han, et al. PrefixSpan algorithm discovers sequential patterns in sequence databases and it has an outstanding performance compared to other algorithms.

Data

- Data Source: The data was extracted from Oracle database Banner student information system at IPFW into a CSV file.
- Sequential Data Terminology: [Figure 1]
- Sequence: Course enrollment history of a given student
  - Term Code: Unique code for each semester that IPFW has (i.e., 201410 -> 2013 Fall)
- Data Scope:
  - **: 1 million rows
  - *: 0.1 million rows
- Data Attributes: [Table 1]

Results

Result for Question 1: [Table 3] shows Top-K frequent course enrollment patterns from students who majored or minored CS degree.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Term Code</th>
<th>Course Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS160 → CS161</td>
<td>Java I</td>
<td></td>
</tr>
<tr>
<td>CS161</td>
<td>Java II</td>
<td></td>
</tr>
<tr>
<td>CS260</td>
<td>Data Structures</td>
<td></td>
</tr>
</tbody>
</table>

Conclusion and Future Work

- From Result 2 we can interpret CS180 is the most important course for students who want to major in CS. Since such a large percentage of CS drop outs occur after a single course then the course curriculum may need to be updated to increase student retention.

Future Work:
- Course Recommendation Based On The Student Characteristics:
  - By including students' information such as gender, age, and program (B.S. or B.A.), we can find interesting course enrollment patterns between different gender and different program.
- Course Recommendation System Using Bayesian Rules:
  - By using department curriculum and student enrollment data who successfully finished their degree, we can build a Bayesian based recommendation system.