Academic Performance Prediction Using Chance Discovery from Online Discussion Forums

COMPSAC 2016 - Preliminary Results Presentation

By Dr. Gary. K. W. Wong and Simon Y. K. Li
Outline

- Introduction and Motivation
- Project Objectives
- Prior Works
- Experiment Design
- Preliminary Results
- Conclusion and Discussion
- Potential Contribution
- Follow-up
Introduction and Motivation

- E-learning systems have become more popular recently in the higher education sector producing a large volume of data created by students and teachers inside via collaborative online discussion forums, which are difficult to analyze without using data mining technologies.

- The forums contain a lot of useful data for educational data mining to extract insightful patterns so that educators can better understand the thinking patterns of students during the learning process.
Project Objectives

- Identify / develop innovative **analytical methods and tools** facilitating teachers’ assessment to **evaluate the students learning in the online learning environment** to determine if students are able to meet **learning objective** and / or **generate new knowledge** beyond the expected learning activities (serendipitous / accidental learning).

- Present **our preliminary research findings** on educational data mining using student discussion forums for future research planning.

- **Adaptive pedagogical design** can be made for teaching and learning.
Prior Works - KeyGraph

- The black nodes and black links represent the **frequent items and their co-occurrences**, implying an **established trend in the data**.
- Black nodes and links form clusters representing **concepts**.
- The red nodes and red links **connect multiple clusters** or some phenomena such as transition of events from one to other clusters.
- Red nodes can be regarded as **chances**.
- Keygraphs are generated to visualize **scenarios** to understand the situation.

Prior Works - Chance Discovery (CD)

- Three stages: Scenario Generation, IMG, and AHP.
- KeyGraph is the algorithm used to **generate scenarios** for finding chances from document to (a) **valuable result(s)**.
- IMG aims at evaluating the chances using human intervention (with expert / prepared minds) to identify values, which is enhanced by using the third stage AHP.
Experiment Design

- In this project, a total 24 undergraduate students in the HKIEd from the General Education course called “Technology, Entertainment and Mathematics” have been sampled for this preliminary experiment.
- One of the course requirements was to complete a reflective posting on an online discussion forum in Schoology.
- They were asked to watch a BBC documentary film called “Beautiful Equations” and other selected movies.
- Afterwards they posted their reflections in the forum. Each student was also required to comment on three self-selected peers, which were extracted in our experiment as text files for analysis.
- There was a total of 110 posts.
- We copied and pasted the posts to text files for storage and further analysis.
Experiment Design

- A software tool called “Polaris” from Ohsawa Laboratory was used for mining text from the following sources.

- Sources of Data (text format):
  - Online reflective discussion forum, etc.
  - Students’ final grade

- Group the discussion posts by their final grades of the module, e.g. A, B, C, D, etc.

- Performed analysis using KeyGraph to generate the visual patterns to identify:
  - The formulation of **key concepts** from black nodes and links
  - **chances** (red nodes and links) for the purposes of decision making and planning in the associated areas above.
  - See whether there are patterns of back / red nodes / links across different grades for future predictive applications
Input Text Files
Preliminary results (Black nodes / links - concepts)

No. of Black Nodes vs. Academic Grades

A declining trend of having less black nodes was observed.

No. of Black Links vs. Academic Grades

A rising trend of having more black links.
Preliminary results (Red nodes / links - chance)

No. of Red Nodes vs. Academic Grades

The number of red nodes for the grade D dropped to zero while for other grades the differences were minimal.

No. of Red Links vs. Academic Grades

The number of red links was generally on the declining trend across the grades although the number grew sharply for grade C.
Preliminary Results – Student with A grade
Preliminary Results – student with D grade
Conclusion and Discussion

- The above results can potentially provide insights into the performance prediction of students.

- The students who can score better grades (e.g. “A”, “A-“, “B”, etc.) would usually have the tendency to contribute more in-depth contents in their posts and hence creating more “chances” and linkages to other concepts from their posts, as identified as the red nodes and links in the above KeyGraphs shown.

- While those of the students who scored bad grades (e.g. “D”) contributed isolated facts with little / no smooth connection or transition from a concept to another.

- The numbers of “chances” and linkages would have the potential to be used as an indicator to predict future student performance in some general education modules.

- Having a lot of black nodes and links alone may not ensure a good and presentable organization of contents, which is also an important factor to deliver a good assay.

- Similarly, a lot of red nodes and links may not really mean that a lot of chances have been explored. Instead, this may just indicate to have noise in the posts. Poor written presentation may be deduced. The writing style may be an important area for the markers to determine grades of a written assay.

- if the student is going to study some other non-liberal modules, the situation may look completely different since some more concrete (rather than liberal or lateral) viewpoints would be needed. The above patterns explored may have totally different interpretations.

- Therefore, merely depending on the numbers of the black / red nodes and links may not be sufficient enough without considering the context in which the contents are produced.

- However, once those patterns can be observed, they can be accumulated and further studied to become generalized rules.

- Machine learning algorithms (e.g. classification, association, clustering, etc.) can then be applied in our project later to explore the combinations of those patterns to formulate rules.
Potential Contribution

- Teachers can **better understand the patterns of thinking of students** during the learning process.

- The **assessment of students** can be facilitated through a **systematic approach** with effective **pedagogic changes** for particular students through the learning process to optimize their learning outcome.

- There seems a **huge potential of unexpected learning discoveries – serendipity**, in collaborative online learning.
Follow-up

- Explore different perspectives to improve the analysis like
  - gender,
  - age,
  - sample sizes,
  - multiple academic subjects,
  - writing styles / language competence,
  - and many other qualitative factors.
- Revise the stop words list to filter noise.
- Further cleanse the sources of text to fix some conversion errors.
- Introduce semantics technologies into matching (dis)similarity
- Perform the KeyGraph analysis again or select more powerful text mining tools.
- Proceed to the IMG and AHP stages of CD for evaluation of the chances already spotted.
References


- Buchem, I. (2011). Serendipitous learning: Recognizing and fostering the potential of microblogging. Form@ re-Open Journal per la formazione in rete, 11(74), 7-16.


References


References


End of Presentation – Thank You