Student Performance Analysis

Shreyas Ugemuge (sugemuge2014@my.fit.edu)
Shiru Hou (shou2015@my.fit.edu)
Yaqeen Al Khatiri (yalkathiri2013@my.fit.edu)
Mohammed Al Habsi (malhabsi2013@my.fit.edu)
Faculty Sponsor and Client: Dr. Philip Chan (pkc@cs.fit.edu)

Goal and Motivation

The client has plenty of raw data available for different classes taken online by students. Data for each class will Include three documents which are the syllabus, final grades for each student and a log file containing the 'clicks' performed by the user. The syllabus will contain information like test dates, assignment due dates, class structure etc. The term 'clicks' refers to URLs that the students access over the course of the class. This information and the final grades are kept anonymous.

The goal of the project is to use this data and examine the 'behaviors' that would or would not lead to a better grade. The raw data does not directly provide any insight on the said 'behaviors'. One of the first and key aspects of this project is to conceptualize these behaviors. Over time the definition of what the term loosely means here will be refined. A very abstract and simple example of a behavior would be "How often does a student access course relevant information before a deadline". After these behaviors are computed from the data, we will determine correlations between the behaviors and the grades.

The next part is where the motivation lies. Our vision is that using the information about behaviors and the correlations to the grades our system should be able to recommend behaviors that would help students achieve better performance. It will essentially be able to tell a student what behaviors they are missing that generally lead to a better grade in their course.

Approach

There are three primary parts to this project. Conceptualizing the behaviors, extracting these behaviors from the data and correlating them to the grades.

The first part doesn't involve technology per se. Conceptualizing the behaviors means that the team will brainstorm and model what we want the behaviors to be. In the simplest sense the data is just a bunch of URLs, timestamps and grades. The team will determine what we want to really extract from the plethora of data we have available. We will use 'common sense' to determine and observe what kind of behaviors students exhibit in real life to achieve a good grade. For instance, 'Studying frequently' is a well known behavior, once we have it identified what we need we move on to how we are going to get it.

The second part deals with finding ways to extract the information that we need. Here is where we find ways to detect the behaviors from the previous part in the data given. Here is where we will write programs to extract the data.

The third part will begin once we have our desired results from the previous two parts. We will try to correlate these set of behaviors to the grades that the students achieved. More details will become clear as we start to extract behaviors.

Features

The entire concept of using ‘clicks’ to determine student performance is a new concept. Internet is an essential part of learning and millennials seem to spend a significant amount of time on the internet. It can be agreed upon that a person's activity on the internet that give you a very good idea about their behaviors and also reflect their approach towards a certain course. Information will be extracted using data mining from raw data that at a glance provides no relevant information. The recommender system will also encourage better work practice.

Technical Challenges

Data mining that will be used extensively is fairly unfamiliar concept for some group members.Even though the team has experience with machine learning, AI and decision trees more tools have to be explored (specified in the task matrix). There are countless libraries and tools that can be used and it is crucial that we research and choose one that best fits the purpose. Discussion reveal that we might use python which the majority of the team will have to learn. There will be slight emphasis on math in the third part of the project.

Milestone 1 (February 24)

- Examine Data
- Set up collaboration tools
- Completely familiarize ourselves with data mining technologies, 2 tools per person
- Investigate data mining packages and tools to be used and suitable programming language for the same
- Try small programs in assigned data mining tools/packages
- Create Requirement Document
- Create Design Document
- Create Test Plan

Milestone 2 (March 24)

- Finalize Conceptualizing the behaviors be used

- Finalize packages, programming languages and tools to be used
- Finalize and set up source control methods
- Implement and test framework to get input from the syllabi
- Select Data mining methods to extract information from the data
- Provide a report explaining all behaviors that are going to be extracted and the methods to be used
- Implement and extract data for at least 2 behaviors

Milestone 3 (April 21)

- Finish extracting information using the behaviors
- Begin correlation of information with grades
- Find appropriate ways to present the information
- Design and start implementing simple GUI

Task Matrix for milestone 1

Task	Shreyas	Shiru	Mohammed	Yaqeen
Select collaboration tools	Create Git Repository on Github.com with webpage branch for project	Set up team communication on slack. Including Calendar and To do list plugin	Set up Google docs	Set up google slides
Investigate packages, programming languages and tools to be used	Research AdvancedMiner and GhostMiner; provide report	Research GNOME and Explore feasability of SPAD GUI; provide report	Research ESTARD data miner and statistica; provide report	Explore Tanagra and compare to SPAD. Research Rapid Miner; Provide report
Completely familiarize ourselves with all required technologies	Learn selected programming language and show simple test program in one of the assigned tools	Learn GIT, Learn selected programming language and show simple test program in one of the assigned tools	Learn selected programming language and show simple test program in one of the assigned tools	Learn selected programming language and show simple test program in one of the assigned tools

Create Requirement Document	Write 13%	Write 13%	Write and organize 50%	Write 13%
Create Design Document	Write 20%	Write 20%	Write and organize 40%	Write 20%
Create Test Plan	Write 20%	Write 20%	Write and organize 40%	Write 20%

Approval from Faculty Sponsor

"I have discussed with the team and approve this project plan. I will evaluate the progress and assign a grade for each of the three milestones."
Signature: _____ Date: _____