Identifying At-Risk Students For Early Interventions– A Time- Series Clustering Approach

B. Bala Krishna[1], Dr G. V.Ramesh Babu [2]
[1]Dept of computer science
[2]Assistant Proffesor
[1, 2] SV University, Tirupati -INDIA

Abstract- *The purpose of this study is to identify at-risk online students earlier, more often, and with greater accuracy using time-series clustering. The case study showed that the proposed approach could generate models with higher accuracy and feasibility than traditional frequency aggregation approaches. The best performing model can start to capture at-risk students from week 10. In addition, the four phases in student's learning process detected holiday effect and illustrates at-risk students' behaviors before and after a long holiday break. The findings also enable online instructors to develop corresponding instructional interventions via course design.*

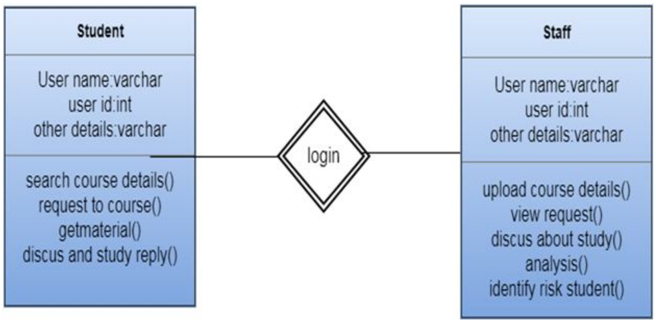
I. EXISTING SYSTEM

For more than a decade, time-series analysis has been an active research topic and been applied to various areas such as finding similar time-series, prediction/forecasting, classification, and segmentation. In online courses, student demo graphics and prior academic record scan be regarded as static data. Students' learning patterns are the only dynamic data that reflect individual learning processes and engagement levels. Analyzing student behavioral time-series can capture variances of individual students on learning, that will allow researchers to generate more accurate predictions. In addition, classifying time series data based on pattern similarity can classify students where "within-group-student" has maximum pattern similarity and a "between-group-student" has maximum dissimilarity.

II. PROPOSED SYSTEM

This study has the following assumptions: First, studentswith similar profiles and learning patterns should resultin similar learning performances. Because time-seriesdata represent variances of online learning behaviors,incorporating time-series data can improve accuracy ofpredictive modeling. Second, one major application oftime-series analysis is prediction/forecasting. Therefore, if a student's shorter time-series (such as a student's patternfrom weeks 1 to 8 in a 16 week semester) matches with oneof the at-risk patterns, the student can be regarded as at-risk at an early time point (early warning).

III. CLASS DIAGRAM



Module Implementation

1. Student Engagement

Generally speaking, at-risk signals identified by previous literatures canbe classified into two categories: (1) student profiles(2) student's engagement level in the online courses, suchas frequency of logins, frequency of course material accessed, number of discussions posted, percentages of grade earned, and total time spent. However, results of previousstudies are difficult to generalize as common profile signals

Because findings really depend on what profile variables can be collected and analyzed. On the other hand, studies on student's engagement level in online courses have ledto pretty consistent findings. The literature shows that, in general, student's performance is highly related to their engagement level in any given course. Almost all related studies found higher engagement level usually leads to higher performance. The engagement levels can reflect on some common learning behaviors across all online courses, such as total clicks, frequency of logins, the number of discussions posted, thenumber of questions posted, and total time spent.

2. Time Serious Clustering

The increasing use of temporal data, especially time-series data, has attracted various research and development efforts. A time-series is a collection of observations made chronologically. The nature of time-series data includes: large data size, high dimensionality, and continuous update. Generally speaking, time-series methods can be divided into the following categories: forecasting/prediction, clustering; classification, and segmentation. Time-series forecasting is theuse of a model to predict future values based on previously observed values. Time-series clustering aims to group time-series data based on similarity/dissimilarity measures.Time-series classification categorize an unlabeled time series to one of the predefined classes. Time-series segmentation divides a time-series into a sequence of discrete segments in order to reveal the underlying properties.

3. Student Data Analysis

In this case study, data was collected from an online graduate program in the United States. The program offers approximately 20 graduate-level courses, hosted in Moodle. Dynamic data was collected from Moodel logs that contain12 courses with 25 course sections and 509 enrollments inthe semester of Spring 2014. Some courses might incorporate eunique learning activities, such as Blog, Glossary, and Wiki.These unique activities were filtered out and obtained the following four common course behaviors for analysis(1) frequency of course material accessed, (2) frequency of forumread, (3) number of discussions posted, and (4) number of replies posted. After initial cleaning, the dynamic dataset contains 427,382 logs in the time period of 16 weeks. Studentstatic data contains student demographics retrieved from the institution's data warehouse. Student's final grade is anominal variable which contains multiple levels A+, A, A-,B+, B, B-, C+, C, C-, and F (failed). To avoid the curse of dimensionality, the final grade was consolidated into three levels A (A+, A, and A-), B (B+, B, and B-), and F (C+, C, C-, and F).

4. At-Risk Student Detection

Both B and F students decreased their participation levels on FR, DP, and RP during the spring break. However, B students kept relatively higher CMA than F studentsduring the spring break. After spring break, F students showed some catch-up behaviorson FR, DP, and RP During the period, F student decreasedtheir FR, DP, and RP behaviors due to the long vacation, then showed high-peak FR, DP, and RP right after the long vacation represented weeks whose eigenvector values were higheror lower than 0.2 or -0.2 respectively. The results showed that at-risk students had unstable discussion Overall, the misclassification rate is10.74% and it can capture up to 85.45% at-risk students. The results indicate that the best model has high practical value to capture at-risk students while the course is still in progress.

IV. SOFTWARE ENVIRONMENT

Java Technology

Java technology is both a programming language and a platform.

The Java Programming Language

The Java programming language is a high-level language that can be characterized by all of the following buzzwords:

- Simple
- Architecture neutral
- Object oriented
- Portable
- Distributed
- High performance
- Interpreted
- Multithreaded
    - Robust
    - Dynamic
    - Secure

With most programming languages, you either compile or interpret a program so that you can run it on your computer. The Java programming language is unusual in that a program is both compiled and interpreted. With the compiler, first you translate a program into an intermediate language called Java byte codes —the platform-independent codes interpreted by the interpreter on the Java platform. The interpreter parses and runs each Java byte code instruction on the computer. Compilation happens just once; interpretation occurs each time the program is executed. The following figure illustrates how this works.

You can think of Java byte codes as the machine code instructions for the Java Virtual Machine (Java VM). Every Java interpreter, whether it's a development tool or a Web browser that can run applets, is an implementation of the Java VM. Java byte codes help make "write once, run anywhere" possible. You can compile your program into byte codes on any platform that has a Java compiler. The byte codes can then be run on any implementation of the Java VM. That means that as long as a computer has a Java VM, the same program written in the Java programming language can run on Windows 2000, a Solaris workstation, or on an iMac.

V. SYSTEM REQUIREMENTS

H/W SYSTEM CONFIGURATION:-

PROCESSOR - PENTIUM –III
SPEED - 1.1 GHZ
RAM - 256 MB(MIN)
HARD DISK - 20 GB
KEY BOARD - STANDARD WINDOWS KEYBOARD
MOUSE - TWO OR THREE BUTTON MOUSE
MONITOR - SVGA

S/W SYSTEM CONFIGURATION

OPERATING SYSTEM :WINDOWS95/98/2000/XP /7
APPLICATION SERVER : TOMCAT5.0/6.X /8.X
FRONT END : HTML, JAVA, JSP
SCRIPTS : JAVASCRIPT, JQUERY, AJAX
SERVER SIDE SCRIPT : JAVA SERVER PAGES.
DATABASE CONNECTIVITY : MYSQL.

VI. CONCLUSION

As the first article utilizing time-series clustering in online education, this study contributes to the literature of learning analytics. The daily learning behavioral patterns of individual online students comprise the only dynamic data source available during an online course. Analyzing daily time-series patterns can generate models with higher accuracyand feasibility than aggregated models constructed by Machine Learning or Time-series analysis algorithms alone. Furthermore, at-risk situations can occur suddenly and atany point during the semester. Thus, the proposed early warning system tracks daily learning activities to identify at risk students earlier, more often, and with greater accuracy. This study shows the potentials of time-series clustering in teaching, learning, and research. The best performing model identifies one successful and one at-risk pattern. In addition, the four phases in student's learning processes

Identifies holiday effect and illustrates at-risk students' behaviors before and after a long holiday break. The LD Aanalysis confirms visual observation findings. The results also enable online instructors to develop corresponding instructional interventions via course design or student teacher interactions. After confirming potential applications of time-series clustering in education. Another study which compares different distance measures and clustering approaches is on-going.