Motivation of the edX InSight Project

Arinjoy Basak, Dewang Palav, Sagar Agarwal, Ashwani Pandey, Apoorva Agarwal

May 25, 2015

What is edX InSight?

- InSight is a BI (Business Intelligence) platform that was built by edX for the purpose of communicating the details of the courses, and analytical data concerning the student activity on the various courses to the administrators and course instructors.
- The primary aim of this platform is not only to monitor the student performances and validation of the choices made in designing the course, but also allows the designers to re-think the choices, and improve the course in its different aspects for creating a better experience for the students.

What is edX InSight?

- Insight is essentially a Django application that allows instructors and administrators to know about their students with respect to.
  - course activity
  - demographic distributions
  - location based distributions, and
  - educational qualifications
- The unstructured data in the form of events are stored in Amazon S3 as JSON objects, while the processed data is then stored in MySQL. Insight uses MySQL 5.1 for its purpose. The data is then transferred to Insights via a REST API.

What our project is

"Knowledge is power"

-Francis Bacon

The project aims to extract maximum usable data from the huge amount of logs running into terabytes per day, and providing a concise outlook to the educators using edX platform. Using the available technologies, we want to create a workable product which presents the data in a form easy to analyse.

What our project is

- Our main project tasks are to study the edX InSight system and contribute towards its development and final integration with the IITBombayX MOOC Platform.
- In particular, our aim will be to extend and implement the edX InSight services so that it appropriately suits to the Indian education environment.

Why do we need InSight?

- The study by Ma, Han, Yang and Chen analysed the the impact of an instructor on a students engagement in an online learning environment, by building an interaction activity model for the teaching and learning process.
- It was showed that learning data analytics could be used to capture authentic, timely and objective evidence regarding online learning behaviour, with a focus towards college level online learning environments.
- The course was the unit of analysis, not the student.

Why do we need InSight?

edX Insight provides access to various graphs, metrics and reports for analysing the student behaviour and activity in different aspects, such as :

- Course Enrollment data : daily student enrollment chart, enrollment metric, and enrollment over time metrics.
- Engagement Data : weekly engagement charts, content engagement breakdown report.
- Demographic Data : analytics on age bands and educational backgrounds.
- Location/Geographic Data : enrollment geography.
- Data on graded and ungraded contents : Find points of difficulty, and attempts to solve the problems.

Project Application Ideas to present

The following ideas were developed by the respective members of the group

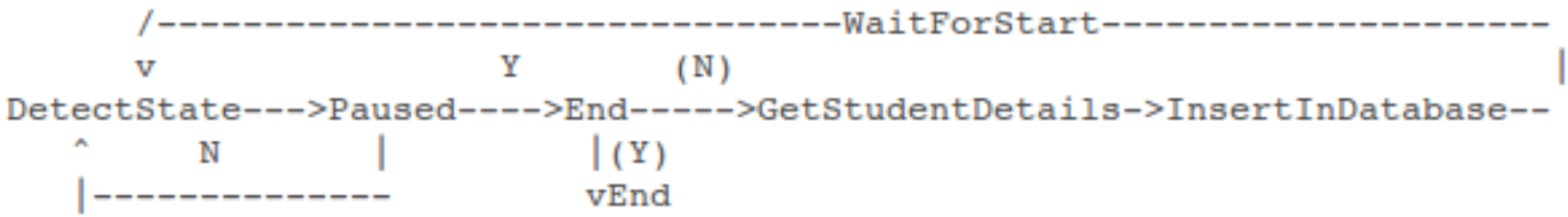
- Arinjoy Basak
- Dewang Palav
- Sagar Agarwal
- Ashwani Pandey
- Apoorva Agarwal

Project application idea 1 (Arinjoy Basak)

- Analyse the student learning behaviour and activities through the pauses in the lecture videos.
- The locations of the pauses could be collected as events, and metrics could be developed and appropraite visualizations made to notify or warn the instructors about the difficulty faced by students in particular parts of materials.
- This could then be met by appropriate measures on the teacher's end, such as
  - Adding more explanatory material
  - Release of an expansion video
  - More quizzes ad practice exercises.

Project application idea 1 (Arinjoy Basak)

- API exists for such tracking, and we can have a possible flow such as the following:



Project application idea 2 (Arinjoy Basak)

- Create dashboard system for students based on their own course activity and performance
- Allow the students to understand whether they are either falling behind the others on their progress, at par with the other students, or ahead of the other students in learning
- This could be implemented as a mobile application too, refreshed with the analysed data and visualizations after a certain time period.
- Inspired by Erik Duval's speech at TEDxUHowest

Project application idea 3 (Arinjoy Basak)

- The communication and interaction between the students could be collected
- We can then use this data to visualize the communication and interaction between the students, to find he students/student groups who are in close communication, and those who aren't in close communication and are slightly outside this conglomerate, residing on the edges.
- Inspired by Erik Duval's speech at TEDxUHowest

Project application idea 4 (Arinjoy Basak)

- A spectral band image system for determining the student's performance
  - —) Because people understand images better than text or data! (:D)
- The spectrum would have different colours for different areas (examinations, lessons, mood of students, activity overall)
- Intensity of the colours depicts performance.
- The degree of transitions between the individual regions determines the balance of the activities.

Project application idea 5 (Arinjoy Basak)

- For each pair of students chosen at a time, show a graphical measure to determine interaction among any two students chosen at a time.
- For example, a colour spectrum with the degree of transition showing the amount of interaction between the students (may even extend to student and teacher/instructor)
- Can be used for determining homogeneity in student population and points of disturbances (less interaction, ineffective activity, behind others in progress, and so on).

Project application ideas (Dewang Palav)

Applications:

- The consistency graph of a student's time of attempting a test gives an insight into the level of seriousness of a student towards a course.
- A certain extent of dynamic behaviour can be implemented by real time time-per-question statistics.
- The courses can be personalized to a greater extent by the instructor with the data about the student at hand.

Project application ideas (Sagar Agarwal)

Applications:

- Analysis of time required and performance on particular test can give information about the understanding level of students and the difficulty level of test.
- It can also give information about difficulty level of course and hence instructor can enhance course content.

Project application ideas (Ashwani Pandey)

Applications:

- Many students hesitate in asking questions or discussing their doubts.
- In a classroom scenario, this will help the teachers see these visual analytics, and help such studetns.
- Ultimately, this will make sure that no one is left behind in a classroom that has students with different abilities.

When provided with incentives, students will show increased interest, leading to increased learning.

What answers can InSight provide? (Ashwani Pandey)

- How much time students spent in viewing the videos ?
- How much time student spent thinking on a particular question, and how many times did he try to attempt a difficult question before giving up ?
- How active is a student on discussion boards ?
- Are the answers and questions of the students being applauded in the discussion board ?
- Is the student also spending equal amount of time with ungraded quizzes, and extra problems that are given to those who want to go a little bit further ?

The analytics for these questions will help figure out the interest shown by the students for a particular course.

Project application ideas (Apoorva Agarwal)

- Student Wise performance availaible to the teacher
  - Weak Students
  - Strong Students
- Generic Analytics of the course content availaible to the teacher.
  - Average duration of answering a question for each student.
  - The strong and weak parts of course:
    - Which part was most clear or unclear to the students.
    - Which part of the material needs to be upgraded or changed
  - The hardest or the easiest questions in the course.

Technologies to learn and apply

Our motivation for this project is also to learn the use of the following software and applications, which besides being beneficial to IITBombayX, is also useful in our attempt to gain knowledge in the sphere of modern Big Data Analytics.

- EdX Platform: Edx is a recent MOOC platform developed by MIT University for offering of quality education to the students on a large scale, through the web service. It has the potential to become the future of Education Technology.
- Hadoop:Hadoop is an open-source software framework written in Java for distributed storage and distributed processing of very large data sets on computer clusters built from commodity hardware.

Technologies to learn and apply

- Hive: Apache Hive is a data warehouse infrastructure built on top of Hadoop for providing data summarization, query, and analysis.
- Python: Python is a widely used general-purpose, high-level programming language, emphasizing code readability, and to express concepts in fewer lines of code than in C++ or Java.The language provides constructs intended to enable clear programs on both a small and large scale.
- Django: A free and open source web application framework, written in Python, which follows the model-view-controller (MVC) architectural pattern.

Conclusion

- The concept of e-learning is gaining precedence. With IITBombayX, an indigenous implementation of the edX is pioneering this field. The need for a robust Insight mechanism which analyses the data generated is being felt at large.
- The ultimate motivation is to meet this need with technology at hand and a contribution to the open source community.

THANK YOU