Just-in-Time Intervention

Hosuk Choi, Jianan Jin, Anusha Prasad, Manley Roberts, Ania Thomas, Matt Yang

Reminder of Motivation

- We would like to use data about a student (everything we know about them up until now) in order to predict their future success in a course
  - Success: Final grade, dropout prediction
  - What we know: concrete course progress up until now (assignment grades, assessment grades, participation with learning management tools)

Reminder of Motivation

- Primarily goal is to expand upon analysis done last semester.
  - Gaining more semesters of data
  - Adjusting timeframe of each snippet of data to be cumulative
  - Gaining more attributes
    - Selecting the most *effective* attributes
  - Improving predictions

Current Status Check

- Developing familiarity with tools
- Collecting data
- Transforming data
- Building a dataframe

Mongo DB team?

- Disbanded -> all SQL now
- Found that all the Mongo data was already aggregated by the server onto the PostgreSQL server (weekly)
- Now just one subteam again (PostgreSQL)
- Started learning PostgreSQL
  - Learned basic queries
  - Still need to learn pivots

Mongo DB Team(?)

- Moving forward
- Currently all of our compiled event data is in Postgres, in a table with the columns
  - course_id, user_id, week, event_type, count
- Next step is to write a query that finds the cumulative sum of each event type by week, of our most important event types (i.e. a “pivot”)
- Then our table will look like

user_id	course_id	relative_week	Event_1 sum	Event_2 sum	… event_k sum
234233	ISYE	0			
234234	ISYE	1			

- Then we are ready for modeling (removing parameters and running algorithms)

SQL Team

- Overarching goal: Modify/create queries to identify a students grade at the end of the week
- Currently working on identifying how many of a specific assignment is submitted a week
  - Submission_score
  - Submission_score_summary
  - Submission_studentItem
  - Submission_submission

- Find the weight of the assignments submitted during the week
  - Length in weeks
  - Average Raw Score
  - Raw Points Per Submission
  - Change in weekly average

New Data

- Some additional forum data!
  - Gives us access to the *EDx* forum data (posts, replies, etc.)
  - Goal of our subteam is not complex text analysis
  - We may start by using the forum data almost like clickstream data - simply looking at if students are using it and what usage patterns they follow.
  - Traditionally, these forums are underused for the courses we are examining

Challenges

- Laying the framework for a project to be successful in the long run.
  - Sometimes, an initial semester can be a hackathon-style scope
  - We are moving beyond that stage and want to make decisions which reflect that this project is aimed to scale up.
- Sometimes we are not immediately sure *where* our data will reside, long-term.
  - Currently, data is exported and manipulated by cronjobs running on the C21U database
  - If one of our objectives includes making this tool relevant to any potential future user at Georgia Tech/beyond, we may need to identify every element of this pipeline so that they are replicable elsewhere.
- Building familiarity with tools, process of data science.

Next Steps

- Construct stage of data pipeline which will transform data from server-manipulated into locally-manipulated and connected dataframe
- Identify all moving pieces along this way so we can create a roadmap for setting up any course at GT or beyond for this analysis process
- Develop models and evaluating attribute success