

# Predicting CS course demand and topics to automate course creation at SJSU

Mayuri Wadkar and Samanvitha Basole and Pratik Surana \*

October 2, 2017

## Abstract

Every semester, approximately 5-10 students are waitlisted in popular computer science courses. Additionally, not a lot of students are confident about what they hope to accomplish by completing a course they register for. As the so-called "hot technologies" are constantly changing, students are seeking to enroll in such courses to help keep up with the technology. Thus, it would be useful to the students and the CS department at SJSU if they could get information about the top technologies as well as the predicted demand for future courses. Our method is technology-based, and we propose a solution to the above problems by building an intelligent agent that can predict demand for CS graduate courses, suggest workshop topics, and list topics for CS 286 or CS 185C. We plan to collect information from professors, department chair, students, online course syllabi, education and training data from the Bureau of Labor and Statistics, and Glassdoor. If successful, our product could help the CS department predict course demand, get an insight into popular CS topics, and it could help CS graduate students make better course choices and help prepare for their career. In order to measure our agent's performance, we will ask our experts to gauge the effectiveness of the agent's output.

## 1 Introduction

Long waitlists for CS graduate courses is the main motivation behind this proposal. With limited supply of professors and classrooms, it becomes a challenge to estimate a demand for courses mainly due to the fact that many students hope to register for courses that teach technologies currently in demand. Thus, we hope to automate the task of predicting demand based on past waitlists and current technologies.

In the second paragraph, describe any research or products that deal with similar questions or problems. Make sure to cite all primary references, for example [1].

---

\*CS256 Section 2 Fall 2017: Technology Driven Project

In the third paragraph, explain how your project will differ from work described in the preceding paragraph.

## 2 Materials and Methods

Web scraping for gathering information about job market and industry trends in Computer Science will give us broader idea of the curriculum and courses to be offered at San Jose State University for MSCS. All project related tasks will be done in Python.

Environment will be responsible for:

- i) Gathering information about job market, industry trends, and hot research topics in Computer Science
- ii) Gathering and analyzing information regarding popularity of a course based on the number of waitlisted students in the past.
- iii) Gathering top trending courses offered in top 20 universities in the US.
- ii) Outcomes of above steps i.e., trending technologies or topics (which could be possible new courses), will have a corresponding set of skills associated with it.
- iv) Environment will create a mapping of these trends and their respective skills, which will act as a training dataset for the actuator.
- v) Trending topics will be clustered based on the similarity of skill sets.
- vi) Each cluster will give rise to a potential new course topic for CS 286 or CS 185C.

### 2.1 Environment

The expected environment in which the agent will operate will be San Jose State University. Specifically, the agent will operate on CS grad courses offered by the CS department at SJSU.

### 2.2 Sensors

Web scraped and cleaned information will be turned into meaningful data which will be stored in lookup tables or datasets. Agent will be served percepts from the Environment by fetching the data stored in data structures or datasets.

### 2.3 Percepts

To find topics with similar skill sets, agent will receive skill sets mapped to trending topics as a percept. To find N high demand courses, agent will receive percepts as average number of wait-listed students for a particular course and course's value in market. Percepts will be passed to the agent, and these percepts represent the information regarding CS course topics or workshop topics. Example percepts include: {'information retrieval', 'decision analysis', 'machine learning'}

## 2.4 Actuators

Actuators will send back the list of courses along with their weightage (demand percentage) back to the Environment. It will also send back the workshops, in case it finds out the need for a workshop based on the demand.

## 2.5 Performance Measures

Describe at least two performance measures that you will use to analyze your agent. One performance measure will be based on the input from our experts - professors. We plan to go to the professors who usually teach the courses predicted by the agent, and we will ask those professors if the agent is accurately producing output. Another measure we will do is to compare similarity measure with universities such as Stanford, Cal, and Caltech.

## 3 Experiments

We will thoroughly test our application during the last two weeks before our final project deliverable. During this test phase, our team will test the project after which our users will test it. These users will consist of students and professors, and we will collect feedback on our application. Along with surveys, we will use student reviews to improve the accuracy of our prediction. In our application, we plan to include a feedback survey after the user uses the application.

## References

- [1] A. Einstein, “Zur Elektrodynamik bewegter Körper. (German) [On the electrodynamics of moving bodies],” *Annalen der Physik*, vol. 322, no. 10, pp. 891–921, 1905.
- [2] D. Knuth, “Knuth: Computers and typesetting.”