Oregon State University Computer Science Senior Design Fall 2016

Progress Report

By Alex Hoffer, Jake Smith, and Chen Chongxian

Team Name: Stat Champs

Abstract

The application of machine learning to Biochemistry and Biophysics has enabled researchers in this field to make remarkable discoveries, such as the generation of new DNA sequences. However, students of Biochemistry and Biophysics do not get the opportunity to learn machine learning. Dr. Victor Hsu of the Oregon State University Biochemistry and Biophysics department has commissioned the Stat Champs to produce an instructional module to give his students the chance to familiarize themselves with machine learning. The software product the Stats Champs have agreed to develop is a web page that allows students to train a machine learning model based on the college basketball statistics and machine learning algorithm of their choosing in order to produce a March Madness bracket. This will help students understand how machine learning algorithms produce models and how inclusion or exclusion of certain data can influence such models. Over the course of Fall term 2016, the Stat Champs have developed materials such as design documents and technology reviews in order to prepare for the engineering of this module. This report comprehensively describes the progress the Stat Champs have made thus far.

Table of Contents

1   Introduction

This report chronicles the progress the Stat Champs have made on developing the machine learning instructional tool. As of the end of Fall term 2016, the progress that has been made consists of a project statement, a requirements document, a design document, a technology review, and this progress report. The remaining sections of this report are devoted to each of the individual members of the Stat Champs to describe their experiences this term as well as their contributions to these documents.

2   Progress Descriptions

2.1   Alex Hoffer

In week one, we were assigned to the "Machine Learn Your Way to March Madness Glory" project proposed by our client Dr. Victor Hsu. During week two, we met with Dr. Hsu to get clarity on the specific details of the project. A synopsis of the project can be found in the Abstract of this document.

After meditating on what we learned from our exchange with Dr. Hsu, we wrote a rough draft of our project statement. In week three, we refined this document and sent it to Dr. Hsu for his approval. The project statement consisted of an abstract, a problem definition, a proposed solution, and performance metrics. We were all relatively new to LaTex and so we had some issues formatting this document correctly. Dr. Hsu felt that we needed to change several things in the document, so in week four, we made these changes and re-submitted it.

During week five, we began developing our project requirements document. I made individual progress by writing the document in Word. This document was formatted using the IEEE standard and thus consisted of an introduction, an overall description, specific requirements, and a Gantt chart. My individual contribution included writing the introduction, an overall description, and the specific requirements. The next week, we adapted my writing to a LaTex file developed by Jake and added a Gantt chart that Chongxian made. Then, we submitted the final product to Dr. Hsu and got his approval.

When week seven came, we revised and re-submitted our project requirements document because Dr. Winters felt it was too vague. Individually, we made progress on our sections of the technology review document. In week eight, we each individually completed our technology review sections and compiled them into one large LaTex formatted document. We faced some difficulty in coming up with responsibilities for each of us and realizing what technologies would be best suited to satisfy said responsibilities. My responsibilities were centered around the usability and appearance of the system. They were 1) the graphical user interface of the web page 2) a presentation of instructions to the user of the module and 3) the presentation of the machine learned bracket. After careful consideration, I landed on using PyGUI for 1), Webix for 2), and VTK for 3).

For week nine, we discussed as a team what approach we should take to the design document, and each individually landed on different software design templates. We also discussed how we would record ourselves for the progress report. One solution we considered was renting out a microphone from the library. As usual, we had some issues with LaTex formatting.

In the final two weeks, we finished and submitted our design document. I used sequence diagrams to describe the functionality of the technologies I selected. In my rationale, I included the average use case for each of my responsibilities. It was difficult getting LaTex to properly generate the sequence diagrams, particularly because the packages I installed were being stubborn. Then, we got together and recorded a progress report. Finally, we each wrote our section of this document and compiled it together.

Retrospective:

Positives	Deltas	Actions
Finished project statement and requirements document	Need to learn basic Python	Will take Python tutorials over winter break
Finished design document and progress report	Need to learn basic JavaScript	Will take JavaScript tutorials over winter break
We all learned LaTex	Need to learn machine learning algorithms	Will check online to understand the machine learning algorithms Chongxian wants to use
Prepared for engineering phase	Need to learn my selected technologies	Will read the documentation for my technologies over winter break

2.2 Chongxian Chen

At beginning of the term, we were selected for Machine Learning to March Madness Glory project. After confirming the members, we quickly set up meeting with our client.

During week 3, we wrote our problem statement this week after meeting with our client Dr.Hsu. After a few revisions our client and our team all signed the paper and submitted a hard copy. The difficulty we meet is figuring out how to convert our statement to LaTex without previous experience of LaTex. But we worked it out. Next week we plan to go into details about where to start our project. Maybe from the interface fist as we shortly discussed this week.

In week 4 we get feedback from the instructor about our problem statement from last week and we revise it. After communicating with our client again with our new problem statement, we submitted again. Revising our problem statement goes smoothly. Next week we will work on the requirement document and prepare for the career fair.

In week 5 we completed our problem statement and beginning writing our requirement document. We met difficulties when writing the requirement document. We don't know what specific guidelines, format and contents to follow for the requirement and we didn't discuss about this during lecture. Next week we will talk with our client and complete our requirement document.

In week 6 we completed our problem statement and beginning writing our requirement document. We met difficulties when writing the requirement document. We don't know what specific guidelines, format and contents to follow for the requirement and we didn't discuss about this during lecture. Next week we will talk with our client and complete our requirement document.

In week 7 we revised our requirement document. We elaborate the details and add more detailed job specification to our gantt chart. We also separate the labors between our teammates for the tech review. Besides we are researching Machine Learning Algorithms to use in our project. So far we have found Scikit to be a good start. Difficulties we met are formatting LaTeX and researching for Machine Learning Algorithm. Next week we will be working on the design document and continue researching

in week 8 we are focusing on writing our tech review. We first talked about dividing labor between group members. After reaching agreement, we begin researching our own technology. We finished writing the tech review on Wednesday and submitted the PDF. The difficulties I met are coming up with three technologies for the parts I am responsible for. I am responsible generally for designing machine algorithm thus I have to explore machine algorithm options. Since none of us are familiar with machine learning, it took me a while for researching. And three technologies are kind hard for me since my part is more a mathematical solution than using some technology. But finally I come up with tools/library I am very likely to use and finished the tech review. Next week we will be working on the design document.

In week 9 we are discussing our design document and preparing for our term progress term. Difficulties we met including setting up the environment we need for voice over presentation. Next week we will be implementing design document and writing our progress report.

In week 10 we are doing our design document. We wrote our own individual three pieces and then work together to integrate them into one design document. We also add the introduction, glossary section, etc. Difficulties we met including researching the machine learning models and statistics models. I spent a lot of time reading documents

about which model to use in the SciKit-Learn. Finally I found supervised learning regression model best fits our need. We will be finishing the progress report and presentation later.

During the fianls week we finished our voice over presentation and summarized our term progress report. The recording and editing goes smoothly. We are looking forward to working on this project next term.

2.3   Jake Smith

In week's one and two we were assigned to the project "Machine Learn Your Way to March Madness Glory" lead by Dr. Victor Hsu. We also wrote and submitted our abstract after sitting down with Dr. Hsu to make sure we fully understood what he envisioned for this project.

At the beginning of week three we wrote a rough draft of the problem statement and sent it to Dr. Hsu to get approved. This document was intended to provide as much clarity as possible to what our projects functions and goals will be. The paper included a Problem Definition, Proposed Solutions and Performance Metrics. In week four we made the changes Dr. Hsu felt necessary to add. we also reformatted the document which was difficult due to no one in the group being comfortable with Latex.

In week five we started on our requirements document which it outlined the Functional, Technical and Usability requirements for the project. Alex wrote a lot of this document while I adapted it to Latex and Chongxian made the Gantt chart. We then submitted it at the end of the week to Dr. Hsu for approval and comments. Week 6 we made a lot of changes to the format and added more details to the document as well as the Gantt chart. Then resubmitted the final product to the client with the changes to get his approval.

During week seven we again revised our project requirements document because it was too vague and didn't nail down the details that needed to be clear. After that we got together and came up with our technologies for the technical review document. In week 8 we all completed our technical review sections and combined them to make one cohesive document. I ran into trouble coming up with all my technologies, which were associated with gathering and storing the data. I decided to we are going to start off by using a multitude of free data websites then store the data in an AWS database to easily work with the AWS machine learning algorithms.

In the final two weeks we did our design document which described the functionality of the technologies we chose in the technical review. For the design document I ran into issues on how to format and make the glossary look good I also could not get the UML diagrams to load correctly in Latex. We also wrote and recorded the progress report.

Retrospective:

Positives	Deltas	Actions
Got more comfortable with Latex	Need to learn Python	Will take Python tutorials on code academy
Got a good understanding of the project	Need to learn more about Amazon AWS	Read AWS documentation and play around over winter break
Re learned SQL	Need to learn machine learning algorithms	Will check online to understand the machine learning algorithms Chongxian wants to use
Finished all documents	Need to verify stats	Average all duplicate stats