

EE/CPR E/ SE Weekly Report II

Feb 11, 2019 - Feb 15, 2019

Group Number: 08

Project Title: Artificial Intelligence for Requirement Analysis Tools

Client & Advisor: Collins Aerospace & Dr. Simanta Mitra

Team Members & Roles:

Ryan Cerveny: Scrum Master, Meeting Scribe/Facilitator, Project Lead, Communicator

Apurva Patel: Project Lead, Report Manager, Technical Support, Communicator, AI Training Lead

Jonathan Murphy: Testing Engineer, Researcher, Requirement Lead

Takao Shibamoto: Chief Engineer, Researcher, UI Lead

Weekly Summary:

The focus of the previous weeks was mostly on project planning, and creating task for teammates for the upcoming weeks.

Past Week Accomplishments:

Project-Planning: Gantt Chart & Proofreading(Apurva & Ryan):

Apurva and Ryan worked on creating the gantt chart on the project planning. Apurva and Ryan created the timeline and curriculum for the year long project with the concern of the team. Moreover, at the end of the Project planning document, Apurva proofread it and inserted any missing pieces

Implementation on Word2Vec (Ryan & Apurva):

Ryan implemented the word2vec using python and committed it to the server. Apurva experimented on the Word2Vec using python and some other libraries in referenced to the kaggle example available online.

Research (Ryan Cerveny & Apurva Patel):

Ryan and Apurva Patel discussed on researching more about word2vec, pros and cons of doc2vec, word2vec and flock algorithm. Apurva research a bit on GloVE algorithm and is understanding its concept.

Setting up python 3.7 on server (Apurva)

Apurva set up python3.7 version on the server, and install all the required libraries for the project, provided access to Ryan to implement the word2vec on the server.

Meeting with Advisor(ALL TEAM MEMBERS):

See summary of advisor meeting below.

Worked on project plan (ALL MEMBERS)

Doc2Vec Experiment (Takao)

- Experimented with Doc2Vec
- Working on making the first prototype

Individual Contributions:

Name	Individual Contributions	Hours this Week	Hours Cumulative
Jonathan Murphy	<ul style="list-style-type: none">● Researched Word2Vec and Word Movers Distance● Project plan● Developed potential designs for front-end GUI	Approx. 6	Approx. 14
Apurva Patel	<ul style="list-style-type: none">● Server Configuration of python● Task upload on jira● Implementation of word2vec using other python libraries● Research on GloVE algorithm● Advisor meeting	11	27
Ryan Cerveny	<ul style="list-style-type: none">● Implemented Word2Vec● Project plan● Gantt chart● Advisor meeting● Researched Word Mover's Distance● Researched FastText	13	28
Takao Shibamoto	<ul style="list-style-type: none">● Researched doc2vec and RNN● Worked on timeline for Project Plan● Experimented with doc2vec and wrote code● Met with advisor and asked the questions mentioned below● Worked on the project plan with other members	10	24

Plans for Upcoming Week:

- Start Experiment/Develop GUI
- Implement Word Mover Distance
- Implement GloVE algorithm or complete entire research
- Look at real requirements data at research park
 - Use data to train models implemented so far

Summary of Weekly Advisor Meeting:

All the team members met our advisor, Dr. Simanta Mitra, for discussion and progress of the project. We discussed the question we had and has decided to ask collins aerospace representative the following question:

- 1) Recommend some book or website that contains large amount of aerospace related english corpus. We won't need to worry about problem of carefully choosing between common english text data and requirement data (which is significantly smaller)
- 2) Retraining model is not as easy as we think. If the program finds unknown vocabulary the machine learning algorithm needs significant amount of other text that contains the unknown word, otherwise the algorithm won't find any statistical relationships with other words. Is it OK if the tool is a black box (neural networks etc)?
- 3) If we use neural networks or doc2vec or whatever, we cannot guarantee the prediction will work for new datasets at the certain accuracy. Is that OK?