

EE / CprE / SE 492 – Sddec18-16

Use machine learning to predict relevant support content based on historical user interactions.

Week 1 Report

August 27, 2018 – September 9, 2018

Team members

Erin Elsbernd: Machine Learning Lead

Ram Luitel: Project Manager and Software Architect

Faizul Jasmi: Communication coordinator

Taizhong Huang: QA Lead

Christian Chiang: Cloud Tech Lead

Khoa Bui: Webmaster and DB lead

Summary of Progress the Past Two Weeks.

- Random Forest Classification - Erin:
 - Continued to work on classification model for help article prediction. The best model is currently a random forest, but it is not good at predicting individual articles high accuracy. To overcome this, grouping similarly related articles and predicting the groups seems to lead to better model performance.
- Collaborative filtering - Ram:
 - Review some machine learning algorithm and try to pick up from where I left last semester. Continued on classification model for help article prediction. I am trying to use collaborative filtering and trying to recommend article according to behavior, I am on the process of it.

```
-----INPUT Article-----  
Article ID: 2916585  
Printing article ID: 2916585  
-----RECOMMENDATIONS-----  
Article ID: 6547360811843584  
Printing article ID: 6547360811843584  
Article ID: 458076148334592  
Printing article ID: 458076148334592  
Article ID: 1334566227  
Printing article ID: 1334566227  
Article ID: 5286801054629888  
Printing article ID: 5286801054629888  
Article ID: 12964181  
Printing article ID: 12964181
```

- Review code - Taizhong:
 - Review some of my previous semester machine learning code and trying to refactor it to come up with some better prediction.
 - Researching different learning models.
- Changing/porting .ipynb files to .py - Faizul:
 - Chose our best models on Jupyter Notebook and made adjustments to make it executable as a Python file.
 - An example of .py file that is converted from .ipynb is showed by Christian to further clarify the process.
 - Made sure that the files chosen can compile correctly for further improvements to the model.
- Review AWS requirement - Christian:
 - Reviewing the AWS requirement materials for our project and setting up the accounts to use this services
 - Created 'Hello World' program for Chalice
 - Researched more about limitation and timeout between API Gateway and Pure lambdas
- Project website update - Khoa:
 - Install FileZilla and set up credential to remote connection. Download the current code that is on our site and wrote some code to update the site for SE492. The site will update when we upload this report.

Pending Issues

Our model's performance is not where it needs to be in terms of article-group prediction accuracy. We are hoping that given more data, our models will be able to perform at a higher level. Our current block is lack of more data. However, our client is providing more data this week so hopefully we will have more data to play with after this week.

Plans for Upcoming Reporting Period

- Article Grouping - Erin:
 - Continue to work on ways to optimally cluster/group related help articles for prediction.
 - Given more data from client, see if the model can be optimized for better prediction metrics with more data points to learn from.
- Try more machine algorithm - Ram:
 - Once our client provide more data, which is probably the week of September, 10th, I will keep trying different machine learning model to come up with better prediction model.
- Faizul:

- Work closely and consult with Christian about the porting process.
- Make sure models are working correctly with the new data
- Clean models and comment it
- New Features - Taizhong:
 - Once I get new data from our client, I will try to come up with different features and try different machine learning to come up with better prediction.
- Christian:
 - Keep working on the cloud architecture and data pipeline
 - Integrate the data cleansing script in it's own lambda
 - Tailor the chalice endpoints
- Update project website - Khoa:
 - Keep updating website and contribute and collaborate with Ram on machine learning models.

Individual Contributions

Team Member	Contribution	Biweekly hours	Total hours
Erin Elsbernd	Worked on help article grouping for predictions.	5	5
Ram Luitel	Booked room for meeting. Created and modified slides for presentation. Continue working on different machine learning algorithm to come up with better prediction	4	4
Faizul Jasmi	Joined advisor meeting and group meeting. Created and modified slides for presentation. Porting .ipynb to .py files	5	5

Taizhong Huang	Joined advisor meeting and group meeting. Created and modified slides for presentation.	4	4
Christian Chiang	Started the cloud architecture to host the models and scripts. Created all the account needed for the AWS suite. Launched a 'Hello World' Chalice and pura lambda app.	6	6
Khoa Bui	Joined Client meeting and group meeting. Set up the project's website. Work on team website and update team page	4	4