EE / CprE / SE 492 – Sddec18-16

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

Week 2 Report
September 10, 2018– September 22, 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:
    - Tried better grouping strategies with the random forest model with minimal success in terms of great classification accuracy, precision, sensitivity, etc. Right now we can achieve relatively high accuracy, but the numbers are inflated due to predicting the most dominant (commonly occuring) groups. When we get more data from the client hopefully we can create equally balanced groups for better prediction accuracy.
- Collaborative filtering - Ram:
    - Since we have not received any additional data from our client I try to rewrite some of my code and for classification model for help article prediction.I also rewrite collaborative filtering and trying to recommend article according to behavior.
- Review code - Taizhong:
    - I reviewed some branches on github and also did some research about cloud.
- Changing/porting .ipynb files to .py - Faizul:
    - Continue working on porting.
    - Analyzed the .ipynb files with help from Christian.
    - Researched on better ways to port the files.
    - Decided to which models to prioritize in porting to .py
    - Contacted Alex about acquiring more data
- Review AWS requirement - Christian:
    - Successfully connected chalice application with other lambdas

    - Ported part of the data to an S3 bucket
    - Analyzed performance on cleaning script
- Project website update - Khoa:
    - Update the website and redesign it to give new look.

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 date. We have not yet received data from our client yet.

Plans for Upcoming Reporting Period

- Article Grouping - Erin:
    - Integrate new data with model, and continue on creating optimal groups for random forest classification.
    - Work on having the model output a ranked confidence of predictions, e.g. 85% Group 1, 10% Group, 5% Group 3 etc.
    - Refactor code so it is more readable and reusable in terms of the data processing and classification functions.
- Try more machine algorithm - Ram:
    - Once our client provide more data, which is probably this week I will keep trying different machine learning model to come up with better prediction model.
- Faizul:
    - Finish up porting selected models and make sure it runs as intended.
    - Finalize with Alex about getting the new set of datas and distribute it to team members.
- New Features - Taizhong:
    - Our client said that he is going to send us more data in very soon. Once we receive the data, we will update more code to gitlab, and I will review them and help my teammates debugging.
- Christian:
    - There is a time limit in AWS lambdas, making the cleaning script unfeasible with bad data will definitely make the lambdas hit the limit
    - Architecture might come with a redesign to accommodate this time constraints.
    - Integrate scripts into lambdas
- 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.	4	8
Ram Luitel	Created and present the presentation. Continue working on different machine learning algorithm to come up with better prediction	4	8
Faizul Jasmi	Joined advisor meeting and group meeting. Created and modified slides for presentation. Porting .ipynb to .py files Selected models to port Researched on porting methods Communicate with Alex on acquiring more datas.	4	9
Taizhong Huang	Reviewed other people's code. Researched AWS.	1	5
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.	5	11

Khoa Bui	Work on team website and update team page	2	6