Proposal for MAI-Computation Intelligence Final Project

November 2nd, 2016

Bárbara Garza

Cole MacLean

Miguel Martínez

Suren Oganesian

A Neural Computation Based Subreddit Recommendation System

Reddit is a social network used for sharing and discussing topics of interest. Reddit's registered community can submit content as links or text, upvote, downvote and make comments. These entries are organized by areas of interest called "subreddits", which serve as sub-communities within the overall reddit community. Reddit has 234 million unique users from 217 different countries and generates 8.19 billion views a month, making it the 9th most visited site in the United States and the 27th in the world. With over 846,336 subreddits, it can be difficult for users to find and explore communities that will interest them. We propose to build a subreddit recommendation system using neural computational models trained on users' historical data to suggest likely subreddits of interest to the user.

Objectives

- Build a system that proposes recommendations whose performance metric is comparable to current recommendation systems [1]
    - "Performance metric" defined as the systems accuracy in correctly predicting users subreddit subscriptions on a holdout validation dataset
- Show that neural computation techniques are suitable for developing recommendation systems for forum category suggestions

Project Stages

The project is proposed in 2 stages: Stage 1 will encompass the experimental results and discussion of the project, where optional Stage 2 will involve building and deploying a user interface for interacting with the final product. Stage 2 will be accomplished if time allows, otherwise it will be discussed as future work.

Data Sources

- https://www.reddit.com/r/datasets/comments/3bxlg7/i_have_every_publicly_available_reddit_comment/
- http://files.pushshift.io/reddit/comments/
- https://praw.readthedocs.io/en/stable/

Potential Tooling

- Python – data-munging and exploring
- https://www.tensorflow.org/ - NN model development
- mongoDB - database
- flask – webapp
- d3.js – visualization
- PRAW reddit python API

References

[1] http://essay.utwente.nl/59711/1/MA_thesis_J_de_Wit.pdf
[2] https://cs224d.stanford.edu/reports/LiuSingh.pdf
[3] https://arxiv.org/pdf/1606.07792v1.pdf
[4] http://benanne.github.io/2014/08/05/spotify-cnns.html