

ParaTweet: A Twitter Content Based Recommendation Engine

Sanketh Katta

Anthony Salgado

Rohit Turumella

Jamie Turley

Mentor : Shai Haim



Motivation

- Lots of users on Twitter but it's hard to find new users that are similar to people I already follow
 - Similar to You Feature
- Current approaches are centered around the concept of triadic closure
- Follow and List Recommendations help generate growth, increase user engagement, and reduces churn



Our Solution

- Generate recommendations based on the content that a given user consumes
 - This allows us to do textual analysis to get a better idea of the type of content that the user prefers
- Web Application that allows users to enter a Twitter Handle and get a list of personalized recommendations for users that they should follow



Similarity Algorithm

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

- Other approaches include TF-IDF, Jaccard Coefficient
 - Noticed better performance and recommendations with Cosine Similarity
- Easy to implement hard to compute with given vector size
- Recommendation Engine is meant not to run in real time (background process)



The Stack

- Frontend: Twitter Bootstrap, JQuery
- Backend: Python-Flask, MongoDB,
- Scaling: Green Unicorn, Nginx, Supervisor
- Scraping: Python-Twitter, Stemming (Porter Stemming)



How would this work at Twitter scale?

- Every user would have pre-generated recommendations which are periodically refreshed
 - The algorithm would be run as a Cron Job
- Ability to Parallelize Algorithm with access to a cluster
 - Currently running on one machine
- Not limited by the REST API
 - Getting the data through the REST API was a bottleneck



Future Work

- Parallelize and Optimize Scraping and Recommendation Algorithms
- Breadth-First Traversal of all Twitter users to generate reports on the fly
- Introduce support for foreign language users
 - Enhance corpus of data (ground truth users) with notable individuals in different countries
 - Add a signal to facilitate location-based recommendations

