

Book Review Based on Twitter Data

Rejul James

Dayana M Sunny

A20345321

A20346386

Steps

- Twitter data regarding set of books will be collected for next 6 weeks
- Twitter data include the tweets made by the user and geographical position of the users.
- Using Machine Learning algorithms, We analyze the tweets made by the users and based on analysis we rate the books
- We will validate credibility of the rating from twitter by comparing with amazon book reviews
- Plot the geographical position of the user in a map.
- Like to extend the project by analyzing the genre of the books which a user like to read and recommend similar books to the user

Approach for the project

- **Twitter API** : Twitter data regarding set of books will be collected by making stream connection with twitter using Twitter python API and we will write the data to a file. Continue this process for next 6 weeks
- **Naïve Bayes Classifier** : We will use Naïve Bayes Classifier or Maximum Entropy classifier for text classification and analysis. Using this we can distinguish positive reviews and negative reviews for a set of books
- **Amazon API** : User ratings from the twitter feed can be validated against the reviews from Amazon API
- **Graph plot** : To show the geographical spread of users of the book

Issues Anticipated

- Noise removal
- Identifying relevant locations associated with tweets
- Identify the tweets not related to any book's company or not using any company profiles.

Timeline of the project

- 10/25/2015 - Initial data collection
- 11/2/2015 - Data cleansing
- 11/16/2015 - Machine learning algorithm implementation
- 11/20/2015 - Geographical position mapping of the readers
- 11/27/2015 - Validate the accuracy of the model against amazon book review

Q&A