Epoch - Social Media Product Analysis

Surendar Nambirajan  
M.S. (Computer Science)  

Saurabh Dixit  
M.S. (Robotics)
Introduction

- This project is to analyse people's polarity towards product on a given timeline using data from social media

- Sentiment analysis using Naïve Bayes Classifier on Twitter data using textblob framework.

- A Web Application to view results and gain useful insights
Motivation and Objective

- **Motivation**
  - There is hype before every product launch and we should be able to quantify that hype by using people’s sentiments.

- **Objective**
  - To devise a web based application for learning insightful informations and trends by analysing users sentiment on social media.
Model Selection

- Naive Bayes Classifier
  - Naive Bayes Classifier of Textblob, a wrapper of NLTK framework is used for classification

- Support Vector Machine
  - Support vector machine of LibShortText library is used. Built upon LIBLINEAR classifier.
Model Selection

- With labeled data from Stanford's course website and manually labeled data, gathered from Twitter.

- Prediction results of classifiers on testing data
  - Naive Bayes Classifier: 62.319%
  - Support Vector Machine: 61.025%

- Though test results are marginal, Naive Bayes Classifier is selected, since SVM framework had implementation issues.
Frameworks used

- Django
- jqPlot
- MySQL
- TextBlob
- Bootstrap
- jQuery
- SQLite
- Python
- Natural Language Analysis with NLTK
- Apache HTTP Server
- Tweepy
Application Architecture
Data Source

- Twitter is primary data source for analysis.
- How Twitter works?
  - Users posts tweet, a 140 characters long tweet with text and url with image or video labeled with #HASHTAG.
  - User's followers can see this tweets, retweet it, favorite it and reply back to this tweet
- Why Twitter?
  - Because of wealthy information about tweet and user unlike other social media has separate API calls for individual informations
Data Collection

- Twitter APIs
  - Streaming APIs: To get twitter's global stream of Tweet data
  - Search API: For searching old tweets, users
- Search based on keyword to get relevant tweets by GET REST API call
- API result
  - JSON format data with tweet and user information
- Data Filtering
  - Tweets with unicodes and non textual content are ignored
Sentiment Prediction

- Collected tweet with the timestamp is predicted with Naive Bayes Classifier using Textblob library
- Labeled tweet is stored in database with tweet creation timestamp.
- These labeled data gives insightful information on product at a particular intervals of time
- Data collected on HTC, ONEPLUS, company and recently launched FARCRY 4 game.
## Insights

### HTC

<table>
<thead>
<tr>
<th>Total Tweet Count</th>
<th>241720</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most positive tweets on</td>
<td>December 1st 2014</td>
</tr>
<tr>
<td>Most Negative tweets on</td>
<td>December 1st 2014</td>
</tr>
</tbody>
</table>

### Oneplus

<table>
<thead>
<tr>
<th>Total Tweet Count</th>
<th>87797</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most positive tweets on</td>
<td>December 2nd 2014</td>
</tr>
<tr>
<td>Most Negative tweets on</td>
<td>December 2nd 2014</td>
</tr>
</tbody>
</table>
Future Developments

- Predicting emotions by combining NLP techniques
- To use advanced data visualization techniques for gathering better insights
- To use data from other social mediums like facebook, flickr and google+
- To identify spam from useful data.
References:

- http://www.csie.ntu.edu.tw/~cjlin/libshorttext/
- http://cs.stanford.edu/people/alecmgo/trainingandtestdata.zip
- http://www.nltk.org/