Kushagra Singh (2014056), Lohitaksh Parmar(2014059)
Dr. Chetan Arora
Advanced Programming
02 December 2015

Predicting popularity of a Facebook status

Objective

In this project, we attempted to predict how popular a Facebook status was, based on the user's personal attributes, and the post content.

Approach

We attempted to predict the number of likes by using a Multiple Linear Regression algorithm. This classifier tries to predict the likes using a set of features, listed below. These are derived not only from the Facebook status under consideration, but also the author of the status.

- Age of author
- Gender of author
- Number of Facebook friends of author
- Time(hour) of posting the status
- Number of Shares of post
- Privacy Setting of post (Public, Friends, Friends of Friends)
- Type of post (event, link, image)
- `**Power**` of the status (Using TF IDF)

The power of a status is calculated as a cumulative score of the words of the status, using the TF IDF algorithm. The feature that we try to predict is the number of likes on the post.

Data

We collected the around 200 Facebook statuses. We did so by asking friends to generate temporary access tokens with particular user data permissions so that we could extract the

required information. We trained our classifier on 160 of these statuses, and tested it on the remaining 40.

Reference: K Chen, B Huang, B Lee, *Facebook Like Predictor Within Your Friends*,