Public Abstract
First Name:Abhishek
Middle Name:
Last Name:Shah
Adviser's First Name:Dr. Wenjun
Adviser's Last Name:Zeng
Co-Adviser's First Name:
Co-Adviser's Last Name:
Graduation Term:FS 2015
Department:Computer Science
Degree:MS
Title:Classification of Twitter Trends using Feature ranking and Forward Feature Selection

316 million active users and 500 million tweets per day. These are some of the numbers of Twitter, one of the most popular social networks in the world. Twitter users tend to use a more natural as opposed to a more formal language in their tweets. These tweets giving rise to twitter's trending topics. This tendency of using natural language makes it very difficult to make sense out of twitter's data programmatically.
In our research we try to solve this problem of understanding natural language by categorizing Twitter's trending topics using various machine-learning techniques. This thesis includes a new approach to classifying twitter trends by adding a layer of forward feature selection and rigorous data pre-processing techniques. A variety of feature ranking techniques like TF-IDF and Bag of Words are used to facilitate this feature selection process, in order to surface the important features while reducing the feature space and making the classification process more efficient. Four Naive Bays Text Classifiers (one for each class), backed by these sophisticated feature ranking and feature selection techniques, is used to successfully classify the twitter trends. Our approach provides an average improvement (in terms of class precision) over the current State of the Art of 33.14% using Bag of Words approach and 28.67% using TF-IDF.