Insult Detection in Hindi

Chetan Dalal(11218) and Shivyansh Tandon(11690)
Prof. Amitabha Mukerjee
IIT Kanpur

March 11, 2014

1 Introduction

With the advent of social networking and blogging, a lot of social interaction is being done on Internet. Since the Internet is uncensored, hence people might sometimes use it for wrong purposes intentionally or unitentionally. The usage can also be be personal insults. This may hurt the ones on the receiver end of the insults. Hence ample research has been done to identify if a given statement is an insult or not. However since much of the data is in English, most research has been for the English crpus only and not for resource poor lanuage Hindi. There is no classifier yet which aims at identifying the insults in Hindi. We propose to build one such classifier.

2 Related Works

There has been research on Insult Detection in English. One such work was Detecting Offensive Tweets via Topical Feature Discovery over a Large Scale Twitter Corpus by using Latent Dirichlet Allocation model. The work by Razavi, Inkpen have used a Insulting and Abusing language dictionary along with features like frequency to run the three-level classification algorithm. For Hindi, there has been a work on sentiment analysis in movie reviews by using a Semi-Supervised approaches to train a Deep Belief Network.

3 Dataset

No labeled dataset is available for Hindi. We plan to create a small dataset manually and then train and test our algorithms on it.

4 Proposed Approach

We propose to use supervised learning algorithms for this problem. First we wish to clean the data by removing the noise like words with high frequency, numbers etc. Then we plan to apply the algorithms like logistic regression, SVM, random forests etc to classify the data. Creating a labelled data will also be one of the major goals of the project.

5 References

1. Xiang, G., Hong, J., & Rosé, C. P. (2012). Detecting Offensive Tweets via Topical Feature Discovery over a Large Scale Twitter Corpus, Proceedings of The 21st ACM Conference on Information and Knowledge Management, Sheraton, Maui Hawaii, Oct. 29- Nov. 2, 2012.

2. Amir H. Razavi, Diana Inkpen, Sasha Uritsky, and Stan Matwin. 2010. Offensive language detection using multi-level classification.

3. Sentiment Analysis For Hindi Language, MS Thesis IIIT-H 2013, Piyush Arora.

4. Insult Detection in Social Media. Amit Roy, Nisha Ramesh, Nivedita Viswnath, Sayan Dey