Insult Detection in Hindi - Supervised Approach

Chetan Dalal
Shivyansh Tandon

Motivation

- Social Networking
- Message Filtering

Dataset

- Will use Hindi SentiWordnet for identifying positive and negative sentiments of a word
- We will develop a small labelled dataset of Insults manually from various forums.
- Also we will use Google Translate for an already established English dataset

Our approach

- Clean the input
- Features
- Feature Selection
- Apply Classifiers
- Combine the classifiers

Cleaning the data

- Removing words which come only once
- Removing punctuations, numbers
- Removing words which come very frequently

Features

- Sentiment value of a word
- Tf-idf score
- Special cases of Second-Person Narrative (Proximity)
- Taking the special case of negation

Feature Selection

- Mutual Information / Chi-Square
- We will remove all the features with less values of MI and Chi-Square

Applying Classifiers

- Logistic Regression
- SVM
- Random Forests

Combining classifiers

- Add the scores according to the weightage of the classifiers
- Weightage decided by individual performance of the classifiers on the training set

References

- Insult Detection in Social Media. Amit Roy, Nisha Ramesh, Nivedita Viswnath, Sayan Dey
- Semi-Supervised Sentiment Analysis in Hindi. Naman Bansal, Umair Z. Ahmed
- http://home.iitk.ac.in/~prasant/HindiCorpus/corpus.html dataset for most frequent words
- http://www.cfilt.iitb.ac.in/wordnet/webhwn/ for stemming the data

Any Questions??

Mutual Information

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)\,p(y)} \right)$$

Chi Squared Test

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

where

χ^2 = Pearson's cumulative test statistic, which asymptotically approaches a χ^2 distribution.

O_i = an observed frequency;

E_i = an expected (theoretical) frequency, asserted by the null hypothesis;

n = the number of cells in the table.

source: http://en.wikipedia.org/wiki/Pearson's_chi-squared_test