Stigma Detection

Characterising Stigma Against Cognitive Decline on Twitter

CS534F14 - Machine Learning - Term Project
Amin Alipour, Aishwarya Ashok, & Nels Oscar
Purpose

Stigma against a disease is often associated with a reduced likelihood of early treatment.

We aim to examine the prevalence of stigma against cognitive decline on Twitter, a service primarily used by younger individuals, to gauge attitudes towards dementia, cognitive decline and Alzheimer’s Disease.
Purpose

- Twitter is a massive source of data.
- Similar work in this area has relied on hand coding [McNeil et al. 2012]
  - With very small sample sizes (<1700 total tweets).
- Machine learning techniques are a promising route to magnifying the human efforts.

Alicia Keys ✔️ @aliciakeys
Follow
Did I mention that u get 2 bring a friend w/u 2 meet me in Rio?! Ohh weeee! Gonna b crazy & helps @keepachildalive
at http://prizeo.com/AliciaNqw
Data

Tweets were collected and tagged by the IGERT students.

31150 total tweets (ignoring explicit retweets)

49 examples of ‘ridicule’ out of the 144 tweets in our example set.
Methods

Feature Engineering

Naive Bayes

Logistic Regression

LogReg and Naive Bayes were compared using the best featureset found in the process of selecting features.

We used Naive Bayes as a baseline to test different methods of constructing the featureset.
In total, 1728 individual experimental conditions.

Features of a tweet are presence or absence of its tokens in the bag of tokens that occur in the training examples.

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stemmer</td>
<td>None, Porter Stemmer, Lancaster Stemmer</td>
</tr>
<tr>
<td>Stop Words</td>
<td>Without, With</td>
</tr>
<tr>
<td>Tokenizers</td>
<td>Naive Whitespace, Punkt English Tokenizer, Twokenize</td>
</tr>
<tr>
<td>Feature Extractors</td>
<td>Positive features, Negative features, Intersection features</td>
</tr>
<tr>
<td>Local Normalization</td>
<td>none, lowercase, strip whitespace, lower + strip</td>
</tr>
<tr>
<td>Global Normalization</td>
<td>none, remove control chars, replace url, remove control chars + replace url</td>
</tr>
<tr>
<td>Include creator</td>
<td>Yes, No</td>
</tr>
</tbody>
</table>
Results Naive Bayes and Feature Selection

Comparison of Feature Selection Choices

Training Accuracy

Testing Accuracy
Results  Logistic Regression
Next Steps

Augment the existing human driven tagging process with an online version of our selected algorithm

Determine appropriate levels of tagged data for the desired level of generality

Incorporate emoticon/emoji in analysis
Feature selection