Q. how can one leverage large-scale, socio-technical systems toward a smarter society?

A. utilize sentiment analysis to improve the identification of influential actors within social networks with the creation of a new sentiment classification scheme.

Methods

Classification Schema
Looking at the needs for identifying users/tweets for social causes we found the along side classification of 2 orthogonal classes, to be best suited. This classification schema allows us to move beyond the positive and negative sentiment classification of tweets to a more audience identification-centric approach.

Codebook Generation
We created a codebook which will help us in generating a training corpus for our classifier for each of the following classes. The codebook was used to hand code the 1500 tweet corpus along the 2 orthogonal classification schemas. We avoided using context based knowledge for getting the coding done so as to remove personal opinions from the coding scheme. For Non-Supportive class we considered the case where the tweets were either directly against the cause or just spreading negative information about the cause. We merged these two cases to build the Non-supportive class as the corpus had very few tweets which were directly against the cause.

Classifier Training
Once we had the training corpus we decided to train a Linear Support Vector Machine (SVM) based classifier. The classifier was trained using the following features. We used 10 fold cross validation to train the classifier and report the accuracies.

Results

Table:

<table>
<thead>
<tr>
<th>Category</th>
<th>Inter Coder Reliability</th>
<th>Accuracy (SVM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enthusiastic v/s Passive</td>
<td>93 %</td>
<td>79.0749 %</td>
</tr>
<tr>
<td>Supportive v/s Non-Supportive</td>
<td>85 %</td>
<td>76.652 %</td>
</tr>
</tbody>
</table>

coded as negative

Now coded as Enthusiastic & Non-Supportive

Now coded as Positive & Supportive

Tweet Corpus

1500 Tweets collected using the following social causes as query terms. The corpus didn’t have duplicate tweets and had only tweets with length greater than 3 words.

Lesbian Gay Bisexual Transgender [LGBT]

Concussions in National Football League [CTE in NFL]

Cyberbullying

Web Tool

Our web tool allows users to search for tweets for their topic of interest and then show our classification with confidence scores. The tool also allows users to see the aggregated count of tweets for each class.

Acknowledgement

Faculty Advisor: Jana Diesner
Social Media Expo Team (iConference 2014, Berlin)
Microsoft Fuse Labs

Website

sentinets-smexyyweby.rhcloud.com/coded