SentiNets: User Classification Based on Sentiment for Social Causes within a Twitter Network

Agarwal, Sneha sagarwal8@illinois.edu
Guo, Jinlong jguo24@illinois.edu
Mishra, Shubhanshu smishra8@illinois.edu
Phelps, Kirstin kphelps@illinois.edu
Picco, Johna picco2@illinois.edu

Academic Advisor: Diesner, Jana [jdiesner@illinois.edu]
Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign

Introduction

Available sentiment classifiers typically describe statements as either positive or negative. While helpful for consumer products or marketing initiatives, this sort of binary classification is limiting for other types of sentiments, particularly those related to social causes. Our research contribution is the creation of new orthogonal sentiment classifiers unique to social causes. This new classification helps capture a more nuanced sentiment along level of *support* (enthusiastic/passive) and the degree of *enthusiasm* (enthusiastic/passive) toward a cause. Twitter data is noisy and content specific, making it difficult for any topic-specific approach. However, our findings show that Enthusiastic and Supportive tweets were more densely present in tweets about social causes in Twitter.

Our research takes a computational approach to address how social media data, with a better classification of sentiment analysis for social causes, can be maximized by individuals and agencies. With a more nuanced classifier, users within social networks more receptive to social causes can be more easily identified for collective action and advocacy.

Background

Since its conception, Twitter has redefined the way social activities are discussed, coordinated and executed. Of particular interest to social media research is the detection of influence within Twitter, with a focus on the understanding of influence based upon the number of followers, retweets and/or mentions (Cha, Haddadi, Benevenuto, & Gummadi, 2010). Understanding influential users on Twitter has a number of important real-world applications, including implications for marketing/advertising costs (Bakshy, Hofman, Mason & Watts, 2011) and consumer feedback, where Twitter has been show to detect users' opinions toward a product and/or brand (Jansen, Zhang, Sobel, & Chowdury, 2009). However, such outcomes are not due to influence alone. Cha et al (2010) found that looking only at indegree (number of followers) reveals little about the influence of a user. Tinati, Carr, Hall, & Bentwood (2012) categorize Twitter users by specific roles and identify key users based on their dynamic communication behavior (such as URLs and hashtags), which have been found to improve re-tweetability. Given these findings, influence cannot be fully understood by looking solely at indegree or tweet characteristics.

Social media sentiment analysis explores emotive aspect(s) as necessary considerations within social networks. Sentiment extracted from Twitter is found to be predictive of real-world outcomes, including box office success (Asur & Huberman, 2010), stock market values (Gilbert et al, 2010) and trending topics (Thelwall, Buckley & Paltoglou, 2011). Yet, the majority of the sentiment analysis on Twitter involves labeling tweets according to polarity or a scale of positive-neutral-negative (Pang, Lee & Vaithyanathan, 2002; Go, Bhayani, & Huang, 2009). Much of the existing literature of sentiment analysis focuses on classifying emotions along these sorts of scales, which work for product markets or popular topics. However, such classifications are limiting to the study of social science. Available research that captures more nuanced dimensions of emotions within social network interactions and Twitter communication is lacking. Thus, in contributing to the theme for the Social Media Expo, we addressed the following research questions:

1. *Is there a better classification system for sentiment analysis in Twitter specific to social causes?*
2. *How can we improve the identification of influential users in Twitter via sentiment analysis with a classifier that captures the degree of enthusiasm and level of support toward a social cause?*

Methods

We created a research workflow [Fig 1] that included three social causes: Chronic traumatic encephalopathy (CTE) in the National Football League (NFL), cyber bullying and lesbian, gay, bisexual and transgender (LGBT) topics. In addition to crosscutting a range of sectors and user demographics, our team also chose these topics as they were representative of the types of social causes for which this social analysis aligns; both of which help normalize the classifier.

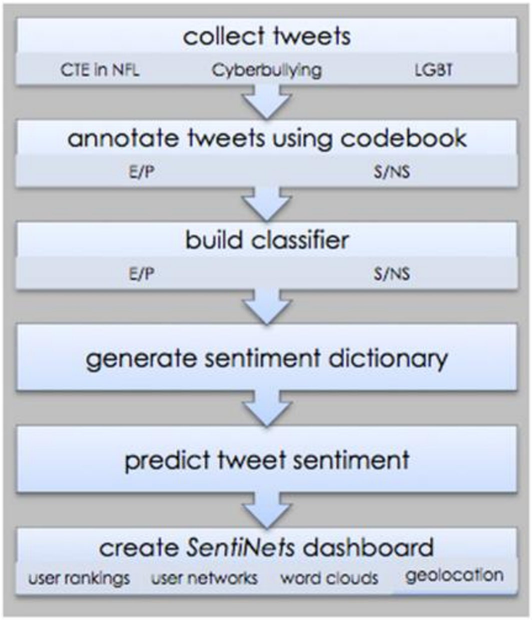

Figure 1 Research Workflow

Next, we collected tweets for each cause using our tool SentiNets, designating a new set of labels appropriate for the purpose of detecting influential users as defined by two orthogonal classes: **supportive/non-supportive** (appropriate to measure the level of *support* toward particular social cause) and **enthusiastic/passive** (appropriate to measure the degree of *enthusiasm* toward a social cause) [Fig 2].

- **Supportive (S)** is defined as *'actively showing favor of cause/subject through use of outright statements/words of support'*
- **Non-Supportive (NS)** '*actively showing favor against the cause/subject.'*
- **Enthusiastic (E)** '*sender includes some personal expression of emotion or call to action for others*

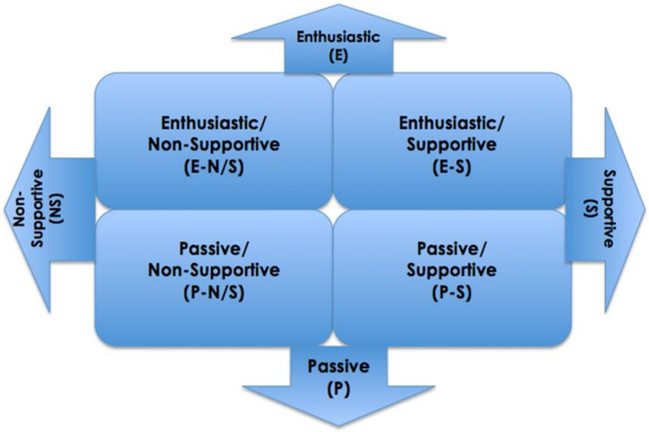

Figure 2 Orthogonal Sentiment Classes

regarding the subject/cause'

- **Passive (P)** '*lack of clear emotive content or lack of call to action.'*

Examples of tweets for each dimension include:

- **S-E**: *Hope he starts his own church! Methodist pastor, suspended for performing gay marriage, weighs options* http://t.co/65K2JCPCwU
- **S-P**: *Here's a link to League of Denial. http://t.co/HHedjKg6JP*
- **NS-E**: *I LOVED the movie cyberbully. I couldn't stop laughing.*
- **NS-P**: *Cyberbully is a gay ass movie*

Using these labels, a detailed codebook was created for manual annotation and hand-coding of tweets. 'Meta'-features of the tweets were also considered, including features such as: *tweet length, number of quotes, number of hashtags, number of mentions, number of emoticons, number of URLs.*

Results

We created a codebook for our classification scheme which was used to manually code a total of more than 1000 tweets for both sentiment categories, for which we achieved a strong inter-coder reliability score for both classes. Each tweet was then parsed to extract the meta-feature and replace the strings like mentions, URL, hashtags in the tweet with a placeholder string so as to normalize all tweets. The meta-features and word vector of each tweet were used to train two Support Vector Machines (SVM) classifiers. The SVM classifier accuracies for each class were as follows:

Category	Inter Coder Reliability	Accuracy (SVM)
Enthusiastic v/s Passive	**93 %**	**79.0749 %**
Supportive v/s Non - Supportive	**85 %**	**76.652 %**

Once trained, we used our classifier to obtain a confidence score for every predicted class, which was assigned to each tweet and displayed in our dashboard called *SentiNets*.

The *SentiNets* dashboard gives users a unified platform to fetch tweets and see a cumulative and individual assessment of each tweet based on the predictions of the two classifiers. *SentiNets* synthesizes our sentiment classifier with aspects of user influence in social networks to deliver a sentiment score (total and average sentiment score) of each tweet, as well as a ranked list of users according to these scores.

Future Contributions

Our work contributes to ongoing research that aims to expand our understanding of language and sentiment within social media. While still limited by sarcasm and other jargon, it moves the conversation forward beyond simple positive-negative scales. Additionally, *SentiNets* will benefit organizations and individuals working with social causes by providing a way to classify tweets according to the level of support and degree of enthusiasm of users within Twitter. *SentiNets* will also feature visualizations such as word clouds, geo-location and network mapping.

Conclusion

Social media has the potential to build relationships, raise awareness and support advocacy of important issues facing our communities. However, social causes are nuanced, at times controversial and infused with opinions that are often lost when evaluated along a scale of simple positive/negative polarity. Our work provides a solution for this problem by creating a way to classify tweets according to the level of support and degree of enthusiasm of users within Twitter. Even with unstructured data, our classifier was able to successfully predict sentiments on social causes within acceptable industry standards. However, our classifiers are limited by the use of language on social media and may not be able to classify context based tweets or sarcastic tweets.

Through *SentiNets,* our user dashboard, we are able to combine our sentiment classifier with aspects of user influence in social networks to help those working toward important social causes to make sense of social media data for positive social outcomes.

Acknowledgements

We would like to thank our academic advisor Dr. Jana Diesner, Microsoft's FUSELab, Dr. Jon Gant, The Graduate School of Library and Information Science at the University of Illinois and the Social Media Expo planning committee.

References

Asur, S., & Huberman, B. A. (2010, August). Predicting the future with social media. *In Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on* (Vol. 1, pp. 492-499). IEEE.

Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011, February). Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 65-74). ACM.

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, P. K. (2010). Measuring User Influence in Twitter: The Million Follower Fallacy. ICWSM, 10, 10-17.

Chung, J. E., & Mustafaraj, E. (2011, April). Can collective sentiment expressed on twitter predict political elections?. In *AAAI*.

Gilbert, E., & Karahalios, K. (2010, May). Widespread Worry and the Stock Market. In ICWSM (pp. 59-65).

Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1-12.

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, *60*(11), 2169-2188.

Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10* (pp. 79-86). Association for Computational Linguistics.

Sinha, S., Dyer, C., Gimpel, K., & Smith, N. A. (2013). Predicting the NFL using Twitter. *arXiv preprint arXiv:1310.6998*.

Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology*,*62*(2), 406-418.

Tinati, R., Carr, L., Hall, W., & Bentwood, J. (2012, April). Identifying communicator roles in twitter. In *Proceedings of the 21st international conference companion on World Wide Web* (pp. 1161-1168). ACM