

Community Threat Analysis on Twitter: Combining Sentiment, Agreement, and size

Johnny W Carter Jr, Dr Abdelzaher, Tarek

Abstract

Background: Twitter is a social networking site. Users follow other users based on interactions and affiliations. Members' cluster into social sub-networks due to a shared-mutual view. The communication protocol on for twitter is called tweeting. A tweet consist of 140 characters that allow users to post messages onto the site. A tweet stores behavioral information such as sentiment on topics. Sentiment analysis of tweets will identify the sentiment of social sub-networks. Retweeting another user's tweet suggest a high degree of agreement with a view. Re-tweeting help identify members of social sub-networks. Social networking provides a new approach into behavioral surveillance of social sub-networks. The aim of this research is to investigate the feasibility of clustering and sentiment analysis for behavioral surveillance. **Methods:** Clustering features with sentiment analysis assist in identifying large social groups who exhibit different views. Views are categorized by positive and negative values. Views are validated with clustering along with running against the sentiment database. Per tweet, the classifier from the set shows the greatest confidence. **Findings and Conclusion:** Sentiment analysis in combination with clustering shows improvement in result.

Introduction

Social networking sites provide a new approach into behavioral surveillance of social sub-networks. Re-tweeting may help identify members of social sub-networks. Members may cluster into social sub-networks due to a shared-mutual view. Sentiment analysis of tweets may help identify the sentiment of social sub-networks. Apollo Is a tool designed to find reliable real time information on social media. Events are separated in to smaller groups with a higher value of content.

Methods

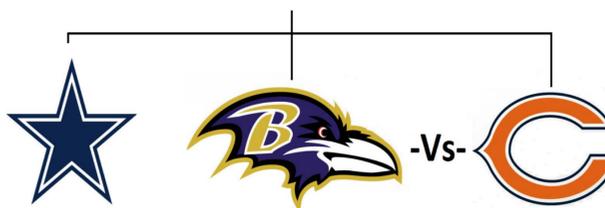
Data Collection: Enter 'NFL' into Apollo to identify a large network of 32 teams for our data. Enter 'Dallas Cowboys' to identify a sub-network of members. Enter 'Baltimore Ravens vs Chicago Bears' to identify a polarized subnets . The Dallas Cowboy are the non polarized subnets.

Data Processing: Preferred sentiment analysis on each data set. The term are extracted from text. Label the sentiment of each terms found in tweet text using a scientifically accepted sentiment database. Using the labeled sentiment of tweet terms, train a set of classifiers .Per tweet, the classifier from the set that shows the greatest accuracy is chosen as sentiment classifier for tweet

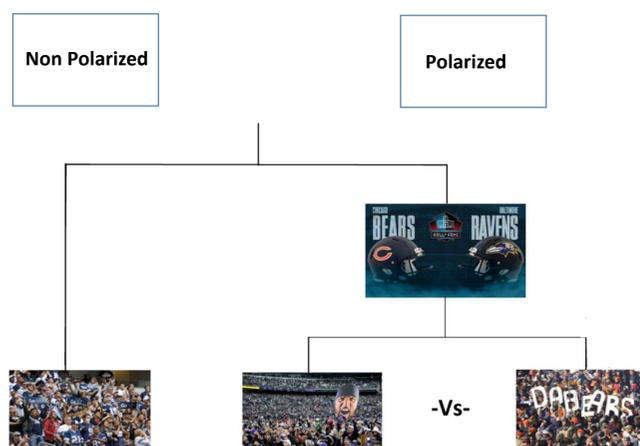
Enter National Football League



Sentiment of the NFL, Dallas Cowboys and Baltimore Ravens vs. Chicago Bears



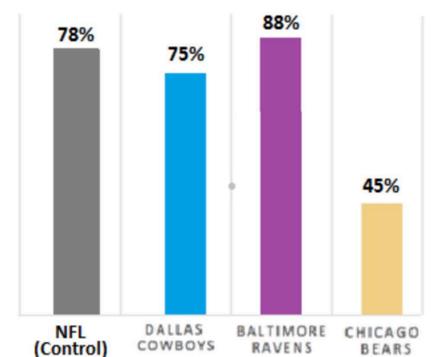
Sentiment after the Baltimore Ravens vs. Chicago Bears games.



Results

The NFL shows the Sentiment of the larger network. It's value was at 78% positive sentiment. The nonpolarized Dallas Cowboys showed a 75% positive sentiment. The polarized Baltimore Ravens vs Chicago Bears showed the Baltimore Ravens with an 88% positive sentiment and the Chicago Bears with a 45% positive sentiment

SENTIMENT ANALYSIS & CLUSTERING



Discussion & Conclusion

Preliminary findings shows that sentiment analysis in combination with clustering is feasible for behavioral surveillance.

References

1. S. Wang et al., "StoryLine: Unsupervised Geo-event Demultiplexing in Social Spaces without Location Information," 2017 IEEE/ACM Second International Conference on Internet-of-Things Design and Implementation (IoTDI), Pittsburgh, PA, 2017, pp. 83-94.
2. D. Wang et al., "Using humans as sensors: An estimation-theoretic perspective," IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks, Berlin, 2014, pp. 35-46. doi: 10.1109/IPSIN.2014.6846739
3. Amin, M. T. A., Aggarwal, C., Yao, S., Abdelzaher, T., & Kaplan, L. (2017). Unveiling polarization in social networks: A matrix factorization approach. In INFOCOM 2017 - IEEE Conference on Computer Communications [8056959] Institute of Electrical and Electronics Engineers Inc.. DOI: 10.1109/INFOCOM.2017.8056959

Acknowledgments

Dr. Abdelzaher Tarek