TITLE Format[*]

[Extended Abstract] [†]

Author1[‡]
Affiliation
affliation addr
affliation address
email

Author2[§]
Affiliation
affliation addr
affliation address
email

Author3[¶]
Affiliation
affliation addr
affliation address
email

Author4
Affiliation
affliation addr
affliation address
email

Author5
Affiliation
affliation addr
affliation address
email

Author6
Affiliation
affliation addr
affliation address
email

ABSTRACT

ABSTRACT

Categories and Subject Descriptors

H.4 [**Crowdsourcing Systems and Social Media**]: Miscellaneous

General Terms

Theory

Keywords

ACM proceedings, LaTeX, text tagging

1. INTRODUCTION

Sensors are devices used for measuring some aspect of an environment and converting it into continuous or discrete value for use in information or control systems. For example, thermocouple as a sensor senses the temperature and converts it to continuous output voltage or a fire detector outputs a boolean value as a result of sensing smoke. Expanding on this definition, social media can be used as a sensor to detect important news e.g., death of Michael Jackson, real-time events e.g., earthquake or epidemics, sentiments and opinions e.g., people's political alignment towards political debates, preferences and traits e.g., stock market prediction or purchase behavior prediction.

However, due to the complexity of information in social media, designing sensors to detect specific phenomena requires highly specialized research to extract targeted information. Designing methodologies and extracting features that are robust to highly dynamic changes in social media, with various forms of expression e.g., informal, short, unstructured texts, written by individuals from different educational levels, and with large volumes of extraneous material mixed in is a difficult task in need of extensive research.

This paper surveys existing work to explore social media as a sensor, methodology and features developed. This paper concludes by examining open areas for future research.

2. SOCIAL SENSORS

Should we mention definition of Social Sensors in our context?

2.1 Feature Analysis

What is useful in Twitter?

- Mentions
- From Users
- Terms
- Hashtags

2.2 Tables/Scatter Plots/Box Plots

2.2.1 Box Plots for Topics/Locations

2.2.2 Tables

3. METHODOLOGY

- too hard to label tweets
- goal is to generalize from seed set of hashtags (Snowball : LOOK AT NEWMAN)

3.1 Learning Social Sensors

Logistic Regression

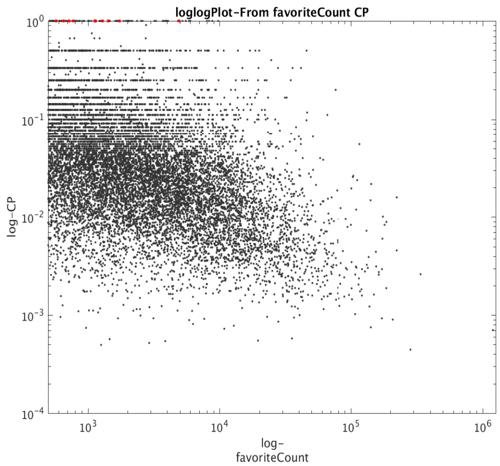

Figure 1: A sample scatter plot for Conditional Property of $From_user$ vs. Favorite Count (.eps format) that has been resized with the `epsfig` command.

3.2 Baselines

- random
- top 100 tweeters of topical tweets in train
- top 100 mentions of topical tweets in train
- top 100 terms of topical tweets in train
- top 100 hashtags of topical tweets in train

Consider them as weights of 1, add up, rank the tweets

4. EVALUATION

4.1 Data Description

1.4 TB of Twitter over 2 years

4.2 Data Annotation

- snowball
- 10 categories

A *Caveat* for the TEX Expert

Because you have just been given permission to use the `\newdef` command to create a new form, you might think you can use TEX's `\def` to create a new command: *Please refrain from doing this!* Remember that your LATEX source code is primarily intended to create camera-ready copy, but may be converted to other forms – e.g. HTML. If you inadvertently omit some or all of the `\def`s recompilation will be, to say the least, problematic.

5. RELATED WORKS

social media event can be defined as an occurrence at a certain time interval and geographical region. It can be planned or unexpected e.g, concert vs. death of a celebrity, man-made or natural e.g., parade vs. earthquake, local or global e.g., concert vs. World Peace Day. Events can further be categorized based on their target users, including individuals, government agencies concerned about natural disasters and health epidemics, marketing companies, and news websites. Historically, event detection has been studied extensively in text mining, NLP, and IR to find events from conventional media sources such as news streams [20]. With the growth of social media sites such as Facebook, Twitter and other microblogs, social media sites have become known as powerful communication tools for sharing and exchanging information about such events. However, event detection on social media sites is more challenging due to features such as unstructured and informal text, highly length restricted, generated by novice reporters compared to journalism-trained news editors.

Nevertheless, it is important to investigate event detection in social media because in comparison to traditional news blogs, social media has faster response time to events and time is money (marketing), lives (disasters), or simply relevance (new).

To see how different use cases address the aforementioned technical difficulties, we focus on the three highly studied types of event detections:

- **Natural Disaster Detection**

  Predictive studies on disaster [10] studied the network of users and focused on choosing the best groups of users in order to achieve lead-times i.e. faster detection of disastrous event (following the concept of "friendship paradox"[1]). On the other hand, [16] used SVM classifier for detecting earthquake and employed location estimation method such as Kalman Filtering for localizing it. [16] extracted statistical features e.g., the number and position of words in a tweet, keyword features and word context features. These studies investigated the real-time nature of Twitter and provided promising results.

  Descriptive studies on disaster Related works discuss the behavior of Twitter users during crisis [18, 5, 17] but do not address exploiting detection of crisis events. They investigated the use of social media during crisis in order to identify information propagation properties, social behavior of users e.g. retweeting behavior, information contributing to situational awareness, and active players in communicating information. However, this behavioral information could be exploited in development of sensors.

- **Health Epidemic Detection**

  Content-based methods [7] and [2] both tried to identify influenza-related tweets and find correlations of these tweets to CDC statistics. Both works extracted bag-of-words as features. As for methodology, the former used single and multiple linear regression showing that multiple linear regression works better, while the latter employed SVM. Results showed high correlation of their estimation of influenza in early stages with values from U.S CDC and Japan's Infection Disease Surveillance Center.

[1] On average, most people have fewer friends than their friends have

Structure-based method In contrast to previous methods, [9] focused on a structure-related technique and developed a model for contagious information diffusion in a social network. They provided a method for choosing sensor groups from friends of random sets of users to find more central individuals in order to enforce early detection.

Hybrid method [14] exploited both content and structure information. They employed a semi-supervised approach to learn a robust SVM classifier by training two classifiers on a labeled set and then applying them to non-labeled tweets . This enabled them to model the interplay of social activity, human mobility, and the spread of infectious disease in a large real-world population. Further, [**?**] identified sick individuals from the content of online communication.

6. CONCLUSIONS

7. ACKNOWLEDGMENTS

[6]

8. ADDITIONAL AUTHORS

9. REFERENCES

[1] M.-D. Albakour, C. Macdonald, and I. Ounis. Identifying local events by using microblogs as social sensors. In *Proceedings of the 10th Conference on Open Research Areas in Information Retrieval*, OAIR '13, pages 173–180, Paris, France, France, 2013. LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE.

[2] E. Aramaki, S. Maskawa, and M. Morita. Twitter catches the flu: Detecting influenza epidemics using twitter. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 1568–1576, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.

[3] C. Budak, D. Agrawal, and A. El Abbadi. Structural trend analysis for online social networks. *PVLDB*, 4(10):646–656, 2011.

[4] CDC. Principles of epidemiology, second edition. *Centers for Disease Control and Prevention*.

[5] F. Cheong and C. Cheong. Social media data mining: A social network analysis of tweets during the 2010-2011 australian floods. In *Pacific Asia Conference on Information Systems, PACIS 2011: Quality Research in Pacific Asia, Brisbane, Queensland, Australia, 7-11 July 2011*, page 46, 2011.

[6] A. Cui, M. Zhang, Y. Liu, S. Ma, and K. Zhang. Discover breaking events with popular hashtags in twitter. In *21st ACM International Conference on Information and Knowledge Management, CIKM'12, Maui, HI, USA, October 29 - November 02, 2012*, pages 1794–1798, 2012.

[7] A. Culotta. Towards detecting influenza epidemics by analyzing twitter messages. In *Proceedings of the First Workshop on Social Media Analytics*, SOMA '10, pages 115–122, New York, NY, USA, 2010. ACM.

[8] S. Eubank, H. Guclu, V. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, 2004.

[9] M. García-Herranz, E. M. Egido, M. Cebrián, N. A. Christakis, and J. H. Fowler. Using friends as sensors to detect global-scale contagious outbreaks. *PloS one*, abs/1211.6512, 2012.

[10] Y. Kryvasheyeu, H. Chen, E. Moro, P. V. Hentenryck, and M. Cebrián. Performance of social network sensors during hurricane sandy. *PLoS one*, abs/1402.2482, 2014.

[11] M. Mathioudakis and N. Koudas. Twittermonitor: trend detection over the twitter stream. In *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2010, Indianapolis, Indiana, USA, June 6-10, 2010*, pages 1155–1158, 2010.

[12] J. Nichols, J. Mahmud, and C. Drews. Summarizing sporting events using twitter. In *17th International Conference on Intelligent User Interfaces, IUI '12, Lisbon, Portugal, February 14-17, 2012*, pages 189–198, 2012.

[13] S. Petrovic, M. Osborne, and V. Lavrenko. Streaming first story detection with application to twitter. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 2-4, 2010, Los Angeles, California, USA*, pages 181–189, 2010.

[14] A. Sadilek, H. A. Kautz, and V. Silenzio. Modeling spread of disease from social interactions. In *Proceedings of the Sixth International Conference on Weblogs and Social Media, Dublin, Ireland, June 4-7, 2012*, 2012.

[15] T. Sakaki, Y. Matsuo, T. Yanagihara, N. Chandrasiri, and K. Nawa. Real-time event extraction for driving information from social sensors. In *Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), 2012 IEEE International Conference on*, pages 221–226, May 2012.

[16] T. Sakaki, M. Okazaki, and Y. Matsuo. Tweet analysis for real-time event detection and earthquake reporting system development. *Knowledge and Data Engineering, IEEE Transactions on*, 25(4):919–931, April 2013.

[17] K. Starbird and L. Palen. *Pass it on?: Retweeting in mass emergency.* International Community on Information Systems for Crisis Response and Management, 2010.

[18] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, Atlanta, Georgia, USA, April 10-15, 2010*, pages 1079–1088, 2010.

[19] J. Weng and B. Lee. Event detection in twitter. In *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011*, 2011.

[20] Y. Yang, T. Pierce, and J. G. Carbonell. A study of retrospective and on-line event detection. In *SIGIR '98: Proceedings of the 21st Annual International (ACM) (SIGIR) Conference on Research and Development in Information Retrieval, August 24-28*

1998, Melbourne, Australia, pages 28–36, 1998.

[21] S. Zhao, L. Zhong, J. Wickramasuriya, and V. Vasudevan. Human as real-time sensors of social and physical events: A case study of twitter and sports games. *Technical Report TR0620-2011, Rice University and Motorola Mobility*, abs/1106.4300, 2011.