SafetyLit Editorial
-David Lawrence

Editorial: Brazen, Unseemly, Multiple Acts of Plagiarism

Background

Preparation for each SafetyLit Weekly Update Bulletin includes identifying recently-published literature (journal articles, technical reports, conference presentations, theses, etc.) that meet the SafetyLit inclusion criteria listed on our FAQ page. This is a labor-intensive process involving the examination of metadata sent by FTP from major publishers and from some smaller publishers, examining the product of several hundred RSS feeds from publishers and other services, searches of several open access databases operated by government agencies, and searching Google Scholar. Each week I identify between 4000 and 6000 potential items. These are screened and 500 to 800 new relevant publications are identified. Of these about 80 percent are included in the Weekly Update Bulletin and the others are simply added to the database. Sometimes the search process also identifies reports and journal articles that were published several years ago. These older items are also directly added to the database.

An Unpleasant Surprise – Identification of Similarities and Duplicates

While conducting the literature searches for the Update to be released on 4 September, a Google Scholar search using one of my standard search strings (that I have used weekly for many years) returned several journal articles with the same (or almost the same) title. Upon further examination, it became clear that several of the articles are essentially word-for word duplicates of one another. Placed in order of publication date, the journal article that was released first was by Eleonore D'Andrea et al and published in IEEE Transactions on Intelligent Transportation Systems.

I focused a second search using the textwords “(Twitter OR Tweet) AND (Traffic OR Congestion)” and didn't include a time interval restriction. This identified additional items – conference proceedings and theses. If I could find these items with a simple (free) Google search, what did the authors of these items find/cite?

At the end of this commentary I will include an annotated bibliography with reference lists of the publications on this Twitter/Traffic topic.

Reader Response – Another Unpleasant Surprise

While I was pleased to learn that the Weekly Update has a readership that is considerably wider than the mailing list subscribers; the first editorial precipitated a flurry of email messages to me from university professors who acted as thesis supervisors, university deans, journal editors, journal publishers, and many others who were only peripherally (if that) involved. In almost every case the writer was upset that: 1) their publication or that of an acquaintance had been listed among items that potentially involved plagiarism; 2) I had not assigned blame; or 3) my definition of plagiarism was too broad. I replied with both an apology and an explanation to those directly involved as authors, editors, or supervisors.

With only four positive messages, the remaining 60+ messages were polite complaints or angry threats. I was accused of slander, liable, being a busybody, having insulted naïve and overworked students, being gutless because I wouldn't name who I thought was a plagiarist and who was a victim, and more. There were quite a few researchers who said that some copying was OK because the later articles were being published in open access journals and most of the world could not afford to read expensive journals. Others said that there were only a few ways to state something in the English and that coincidences of similar phrasing were likely to occur because of the expanding body of literature.

My response to the criticisms

In my first two commentaries (now removed from the SafetyLit Update Bulletins and replaced by this document) I pointed out that some of the articles were clearly word-for-word plagiarism by any definition. Further evidence of overt copying is: while the authors of the D'Andrea article used tweet-words that were in their native Italian language, the later articles also contained the same terms and weights assigned to the terms. Authors of these later-published articles (judging from their authors' names and their institutions) do not seem to have any connection to Italy.

While I did point out which journal article was published first, I did not accuse any author of plagiarism. Upon reflection, I realize that the date of publication (who was first to print) is only one indicator in assessing plagiarism from original thought and that word-for-word copying, clearly the most egregious form, is not the only way to plagiarize another's intellectual product. Presentations at conferences, the release of drafts to ArXive and similar preprint repositories, etc. make determination of first release beyond my ability and outside my role and responsibilities. Determining plagiarist from victim isn't my duty. Investigating such things is the role of journal editors and graduate thesis supervisors. As curator of the SafetyLit database and update service it is my role to identify literature relevant to all safety issues and to index it. In my first editorial in 19 years I commented about my surprise and annoyance with having found multiple nearly identical articles.

Further thoughts

I believe that this is an important issue – much larger than the articles and theses discussed here. It is an issue that must be addressed. The fact that plagiarism is accepted or even encouraged by some journal publishers will require an effort by the publishers of reputable journals, their editors, universities, and others to solve this problem. Consumer product patents and book, music, and video copyrights are all addressed (with uneven success) by international treaties. While there are very serious attempts to address patent and copyright theft, there is little organized effort to handle academic intellectual property theft and, I suggest, as a result, the theft of academic reputation.

What can be done about Plagiarism?

- How thoroughly should an author conduct a literature search to learn what came before?
- When can an author *not* cite earlier works on similar topics?
- Does an author's knowledge of earlier, *un-cited* work falsely suggest that the recent author's idea is original?
- What about the uncredited “borrowing” of an approach to a problem or a methodology?
- The use of plagiarism detection software is not universally used by editors and professors. Why?
- How severely should undergraduate students be penalized when they ignore simple rules of proper citation?
- How severely should graduate students be penalized when they omit key publications from the reference lists of their theses but use ideas presented in earlier publications?
- Should thesis supervisors require that students present a comprehensive listing of the product of their literature search and a rationale for items not referenced?
- What penalty should a university assign to academic authors who are found to have used without credit the intellectual property of another?
- There are journals with policies that restrict the number of sources allowed in a reference list. How does this affect proper crediting of the intellectual products of others?

I believe that even with the most obscure topics, someone else deserves credit for inspiring later work.

In my own work I have been pleased to liberally use the brilliance of those who came before me but I acknowledge and reference those earlier ideas and I quote when the earlier, eloquent word use is too good to paraphrase. Perhaps, I could be accused of taking my citation philosophy too far because my own thesis contains more than 400 references.

I welcome comments on this topic. davidl@safetylit.org

The following annotated bibliography may be interesting to those who would like to further examine the genealogy of scholarly literature on traffic and Twitter. Please pay particular attention to the first item from 2015 and the first four items from 2016.

2010

Real-time sensing of traffic information in Twitter messages

Type	Thesis
Author	Sara Filipa Lemos de Carvalho
URL	https://repositorio-aberto.up.pt/bitstream/10216/58384/1/000143724.pdf
Place	Porto, Portugal
Date	2010
Library Catalog	www.safetylit.org
Type	Masters
Language	en
University	Universidade do Porto, Faculdade de Engenharia

Abstract

Traffic issues affect the mobility of many people and the dynamics of the big urban centers. The study of the traffic urban networks is of great importance for the improvement routes and traffic flow. This document intends to introduce a new source of information that can be helpful in the traffic scene analysis: microblogging messages. This new source surpasses some of the disadvantages of traditional traffic sensors, like area coverage and the costs of installation and maintenance. The problem in focus in this study is to investigate whether information can be found, in microblogging messages, that is relevant to the traffic study. The microblogging platform used to collect the messages was Twitter and the objective was to retrieve messages shared by individual users, in opposite to official sources like news agencies. The approach followed to solve the problem was to address it as a text classification problem using SVMs. The solution was divided into two main iterations of built and improvement of two classification models (bootstrapping strategy) to capture traffic messages. In each phase, for each model, the results achieved were registered and the evaluation measures were calculated and compared. Also, the improvements made from one phase to the other were registered. In the end, the two classification models were able to capture a generic traffic message with a precision of more than 80% and a traffic message shared by an individual user with a precision of approximately 50%. In conclusion, the objectives were achieved and the results were considered satisfactory in capturing messages from individual users, although in the capture of general traffic messages the results were a success.

Keywords: Twitter-Traffic-Status

References

Eugene Agichtein, Carlos Castillo, Debora Donato, Aristides Gionis, and Gilad Mishne. Finding high-quality content in social media. In WSDM '08: Proceedings of the international conference on Web search and web data mining, pages 183–194, New York, NY, USA, 2008. ACM.

Sitaram Asur and Bernardo A. Huberman. Predicting the future with social media. Mar 2010.

ANACOM. Number of mobile broadband users surpasses number of fixed broadband users for first time. ANACOM, 2008.

Twitter Blog. Top twitter trends of 2009. http://blog.twitter.com/2009/12/top-twitter-trends-of-2009.html. Last visited: 26-06-2010.

Andy Bloxham. Facebook more effective than emergency services in a disaster. The Daily Telegraph, December 20, 2008.

Avrim Blum and Tom Mitchell. Combining labeled and unlabeled data with co-training. In COLT '98: Proceedings of the eleventh annual conference on Computational learning theory, pages 92–100, New York, NY, USA, 1998. ACM. Chih C. Chang and Chih J. Lin. LIBSVM: a library for support vector machines, 2001.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. In Machine Learning, pages 273–297, 1995.

Jesse Davis and Mark Goadrich. The relationship between precision-recall and roc curves. In In ICML '06: Proceedings of the 23rd international conference on Machine learning, pages 233–240. ACM Press, 2006.

Eurostat. Energy, transport and environment indicators, 2009 edition. European Comission, 2009.

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The WEKA data mining software: an update. SIGKDD Explorations, 11(1):10–18, 2009.

Jiawei Han and Micheline Kamber. Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems). Morgan Kaufmann, 2nd edition, 2006.

T.Huang, D.Koller, J.Malik, G.Ogasawara, B.Rao, S.Russell, and J. Weber. Automatic symbolic traffic scene analysis using belief networks. In Proceedings 12th National Conference in AI, pages 966–972. AAAI Press, 1994.

Bernardo A. Huberman, Daniel M. Romero, and Fang Wu. Social networks that matter: Twitter under the microscope. 2008.

Jaiku. http://www.jaiku.com/. Last visited: 26-06-2010.

Thorsten Joachims. Text categorization with support vector machines: learn- ing with many relevant features. In Claire Nédellec and Céline Rouveirol, editors, Proceedings of ECML-98, 10th European Conference on Machine Learning, number 1398, pages 137–142, Chemnitz, DE, 1998. Springer Verlag, Heidelberg, DE.

Steve Johnson. How twitter will change the way we live. Time, June 05, 2009.

Brian Slack Jean-Paul Rodrigue, Claude Comtois. The Geography of Transport Systems, chapter 4. Routledge, 2009.

Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. Why we twitter: understanding microblogging usage and communities. In WebKDD/SNA- KDD '07: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, pages 56–65, New York, NY, USA, 2007. ACM.

Boris S. Kerner. Introduction to Modern Traffic Flow Theory and Control: The Long Road to Three-Phase Traffic Theory, chapter 1. Springer, 2009.

S. Keshav. Why cell phones will dominate the future internet. SIGCOMM Comput. Commun. Rev., 35(2):83–86, 2005.

David D. Lewis. Representation quality in text classification: An introduction and experiment. In In: Proceedings of Workshop on Speech and Natural Language. Hidden, pages 288–295. Morgan Kaufmann, 1990.

David D. Lewis. Feature selection and feature extraction for text categorization. In HLT '91: Proceedings of the workshop on Speech and Natural Language, pages 212–217, Morristown, NJ, USA, 1992. Association for Computational Linguistics.

Christopher D. Manning and Hinrich Schuetze. Foundations of Statistical Natural Language Processing. The MIT Press, 1st edition, June 1999.

Peter Guest, Mike Slinn, Paul Matthews. Traffic Engineering Design: Principles and Practice, chapters 1 and 2. Elsevier Butterworth-Heinemann, 2nd edition, 2003.

Mike Musgrove. Twitter freedom's only link in Iran. The Washington Post, July 09, 2009.

Kamal Nigam, Andrew Kachites Mccallum, Sebastian Thrun, and Tom Mitchell. Text classification from labeled and unlabeled documents using em. In Machine Learning, pages 103–134, 1999.

Foster Provost, Tom Fawcett, and Ron Kohavi. The case against accuracy estimation for comparing induction algorithms. In Proceedings of the Fifteenth International Conference on Machine Learning, pages 445–453. Morgan Kaufmann, 1997.

Plurk. http://www.plurk.com/. Last visited: 26-06-2010. Jason Pontin. From many tweets, one loud voice on the internet. The New York Times, April 22, 2007.

J. Rocchio. Relevance Feedback in Information Retrieval, pages 313–323. 1971.

Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In WWW '10: Proceedings of the 19th international conference on World wide web, pages 851–860, New York, NY, USA, 2010. ACM.

Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner, and Isabell M. Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. In AAAI-10: Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, Atlanta, USA, 2010. AAAI.

Jen-Chao Tai, Shung-Tsang Tseng, Ching-Po Lin, and Kai-Tai Song. Real-time image tracking for automatic traffic monitoring and enforcement applications. Image and Vision Computing, 22(6):485 – 501, 2004.

Tumblr. http://www.tumblr.com/. Last visited: 26-06-2010. Twitter. http://twitter.com/. Last visited: 26-06-2010.

Vladimir N. Vapnik. The nature of statistical learning theory. Springer-Verlag New York, Inc., 1995.

E. Veiszadeh. Twitter freedom's only link in Iran. The Australian, July 16, 2009.

Emma Young. Crisis puts a new face on a social networking. The Sydney Morning Herald, June 07, 2009.

Harry Zhang. The optimality of naive bayes. In Valerie Barr, Zdravko Markov, Valerie Barr, and Zdravko Markov, editors, FLAIRS Conference. AAAI Press, 2004.

Zhigang Zhu, Guangyou Xu, Bo Yang, Dingji Shi, and Xueyin Lin. Visatram: a real-time vision system for automatic traffic monitoring. Image and Vision Computing, 18(10):781 – 794, 2000.

# of Pages 59

Real-time sensing of traffic information in Twitter messages

Type	Conference Paper
Author	Sara Filipa Lemos de Carvalho
Author	Luís Sarmento
Author	Rosaldo J. F. Rossetti
Place	Madeira Island, Portugal
Publisher	IEEE
Date	2010
Library Catalog	www.safetylit.org
Conference Name	IEEE-ITSC Workshop on Artificial Transportation Systems and Simulation
Language	en

Abstract

This paper presents an initial attempt to use micro-blogging messages posted on Twitter (by users in transit) to perform real-time sensing of traffic-related information. We propose a text classification approach to the problem: we wish to automatically identify traffic-related messages posted on Twitter, among the millions of unrelated messages posted by users. Given that the fraction of relevant traffic messages on Twitter is extremely low (< 0.05%), the main challenges involved at this stage are (i) creating a suitable training set for setting up the classifier, and (ii) driving the classifier to a reasonable level of precision in identifying relevant messages. We opted for a dual-stage bootstrapping strategy for tackling both these problems simultaneously. First, we used short message reports that are automatically posted on Twitter by certain official news source to compile an initial set of training messages that are comparable in contents to the user-generated messages we wish to identify. Then, using a classifier trained on such robot-sent messages, we process a large collection of Twitter messages to identify traffic-related messages set by users, which are then added to the training set and are used to train a second version of the classifier. Results show that, despite the highly-unbalanced example distribution, we are able to almost double the performance of the classifier from the first iteration to the second, and that F-measures above 23% in identifying relevant traffic-related Twitter messages can be achieved, with little human effort involved in creating a training set.

SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based.

References

[1] J. Pontin, "From many tweets, one loud voice on the internet," The New York Times, April 22, 2007.

[2] S. Johnson, "How twitter will change the way we live," Time, June 05, 2009.

[3] E. Veiszadeh, "Twitter freedom's only link in Iran," The Australian, July 16, 2009.

[4] M. Musgrove, "Twitter is a player in Iran's drama," The Washington Post, July 09, 2009.

[5] A. Bloxham, "Facebook more effective than emergency services in a disaster," The Daily Telegraph, December 20, 2008.

[6] E.Young,"Crisis puts a new face on social networking,"The Sydney Morning Herald, June 07, 2009.

[7] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: real-time event detection by social sensors," in WWW '10: Proceedings of the 19th international conference on World wide web. New York, NY, USA: ACM, 2010, pp. 851–860.

[8] S. Asur and B. A. Huberman, "Predicting the future with social media," Mar 2010. [Online]. Available: http://arxiv.org/abs/1003.5699

[9] A.Tumasjan, T.O.Sprenger, P.G.Sandner, and I.M. Welpe,"Predicting elections with twitter: What 140 characters reveal about political sentiment," in AAAI-10: Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media. Atlanta, USA: AAAI, 2010.

[10] E. Agichtein, C.Castillo, D.Donato, A.Gionis, and G.Mishne,"Finding high-quality content in social media," in WSDM '08: Proceedings of the international conference on Web search and web data mining. New York, NY, USA: ACM, 2008, pp. 183–194.

[11] C. Cortes and V. Vapnik, "Support-vector networks," in Machine Learning, 1995, pp. 273–297.

[12] V. N. Vapnik, The nature of statistical learning theory. Springer-Verlag New York, Inc., 1995.

[13] T. Joachims, "Text categorization with support vector machines: learning with many relevant features," in Proceedings of ECML-98, 10th European Conference on Machine Learning, C. Ne ́dellec and C. Rouveirol, Eds., no. 1398. Chemnitz, DE: Springer Verlag, Heidelberg, DE, 1998, pp. 137–142. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.21.6982

[14] C. C. Chang and C. J. Lin, LIBSVM: a library for support vector machines, 2001. [Online]. Available: http://www.csie.ntu.edu.tw/cjlin/libsvm

[15] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," SIGKDD Explorations, vol. 11, no. 1, pp. 10–18, 2009. [Online]. Available: http://dx.doi.org/10.1145/1656274.1656278 Keywords: Twitter-Traffic-Status

Proceedings Title Proceedings: IEEE-ITSC Workshop on Artificial Transportation Systems and Simulation

2011

Identification and characteriaztion of events in social media

Type	Thesis
Author	Hila Becker
URL	http://www.cs.columbia.edu/~hila/hila-thesis-distributed.pdf
Date	2011
Accessed	9/28/2016, 7:55:07 PM
Library Catalog	www.safetylit.org

Abstract

Millions of users share their experiences, thoughts, and interests online, through social media sites (e.g., Twitter, Flickr, YouTube). As a result, these sites host a substantial number of user-contributed documents (e.g., textual messages, photographs, videos) for a wide variety of events (e.g., concerts, political demonstrations, earthquakes). In this dissertation, we present techniques for leveraging the wealth of available social media documents to identify and characterize events of different types and scale. By automatically identifying and characterizing events and their associated user-contributed social media documents, we can ultimately offer substantial improvements in browsing and search quality for event content.

To understand the types of events that exist in social media, we first characterize a large set of events using their associated social media documents. Specifically, we develop a taxonomy of events in social media, identify important dimensions along which they can be categorized, and determine the key distinguishing features that can be derived from their associated documents. We quantitatively examine the computed features for different categories of events, and establish that significant differences can be detected across categories. Importantly, we observe differences between events and other non-event content that exists in social media. We use these observations to inform our event identification techniques.

To identify events in social media, we follow two possible scenarios. In one scenario, we do not have any information about the events that are reflected in the data. In this scenario, we use an online clustering framework to identify these unknown events and their associated social media documents. To distinguish between event and non-event content, we develop event classification techniques that rely on a rich family of aggregate cluster statistics, including temporal, social, topical, and platform-centric characteristics. In addition, to tailor the clustering framework to the social media domain, we develop similarity metric learning techniques for social media documents, exploiting the variety of document context features, both textual and non-textual.

In our alternative event identification scenario, the events of interest are known, through user-contributed event aggregation platforms (e.g., Last.fm events, EventBrite, Facebook events). In this scenario, we can identify social media documents for the known events by exploiting known event features, such as the event title, venue, and time. While this event information is generally helpful and easy to collect, it is often noisy and ambiguous. To address this challenge, we develop query formulation strategies for retrieving event content on different social media sites. Specifically, we propose a two-step query formulation approach, with a first step that uses highly specific queries aimed at achieving high-precision results, and a second step that builds on these high-precision results, using term extraction and frequency analysis, with the goal of improving recall. Importantly, we demonstrate how event-related documents from one social media site can be used to enhance the identification of documents for the event on another social media site, thus contributing to the diversity of information that we identify.

The number of social media documents that our techniques identify for each event is potentially large. To avoid overwhelming users with unmanageable volumes of event information, we design techniques for selecting a subset of documents from the total number of documents that

we identify for each event. Specifically, we aim to select high-quality, relevant documents that reflect useful event information. For this content selection task, we experiment with several centrality-based techniques that consider the similarity of each event-related document to the central theme of its associated event and to other social media documents that correspond to the same event. We then evaluate both the relative and overall user satisfaction with the selected social media documents for each event.

The existing tools to find and organize social media event content are extremely limited. This dissertation presents robust ways to organize and filter this noisy but powerful event information. With our event identification, characterization, and content selection techniques, we provide new opportunities for exploring and interacting with a diverse set of social media documents that reflect timely and revealing event content. Overall, the work presented in this dissertation provides an essential methodology for organizing social media documents that reflect event information, towards improved browsing and search for social media event data.

Bibliography

[ACD+98] James Allan, Jaime Carbonell, George Doddington, Jonathan Yamron, and Yiming Yang. Topic detection and tracking pilot study final report. In *Proceedings of Broadcast News Transcription and Understanding Workshop*, 1998.

[AGAV08] Enrique Amigó, Julio Gonzalo, Javier Artiles, and Felisa Verdejo. A comparison of extrinsic clustering evaluation metrics based on formal constraints. *Information Retrieval*, 2008.

[AH10] Sitaram Asur and Bernardo A. Huberman. Predicting the future with social media. Technical Report HPL-2010-53, HP Laboratories, 2010.

[Ahn06] David Ahn. The stages of event extraction. In *Proceedings of the COLING/ACL 2006 Workshop on Annotating and Reasoning about Time and Events (ARTE '06)*, 2006.

[All02] James Allan, editor. *Topic Detection and Tracking: Event-based Information Organization*. Kluwer Academic Publisher, 2002.

[APL98] James Allan, Ron Papka, and Victor Lavrenko. On-line new event detection and tracking. In *Proceedings of the 21st ACM International Conference on Research and Development in Information Retrieval (SIGIR '98)*, 1998.

[BBS05] Mikhail Bilenko, Sugato Basu, and Mehran Sahami. Adaptive product normalization: Using online learning for record linkage in comparison shopping. In *Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM '05)*, 2005.

[BCI+11] Hila Becker, Feiyang Chen, Dan Iter, Mor Naaman, and Luis Gravano. Automatic identification and presentation of Twitter content for planned events. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM '11)*, 2011.

[Ber02] Pavel Berkhin. Survey of clustering data mining techniques. Technical report, Accrue Software, 2002.

[BHB11] Edward Benson, Aria Haghighi, and Regina Barzilay. Event discovery in social media feeds. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT '11)*, 2011.

[BING11] Hila Becker, Dan Iter, Mor Naaman, and Luis Gravano. Identifying content for planned events across social media sites. Submitted for publication, 2011.

[BKM06] Mikhail Bilenko, Beena Kamath, and Raymond J. Mooney. Adaptive blocking: Learning to scale up record linkage. In *Proceedings of the Sixth IEEE International Conference on Data Mining (ICDM '06)*, 2006.

[BM98] Avrim Blum and Tom Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the 11th Annual Conference on Computational Learning Theory (COLT '98)*, 1998.

[BM03] Mikhail Bilenko and Raymond J. Mooney. Adaptive duplicate detection using learnable string similarity measures. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '03)*, 2003.

[BMC07] Hila Becker, Christopher Meek, and David Maxwell Chickering. Modeling contextual factors of click rates. In *Proceedings of the 22nd AAAI Conference on Artificial Intelligence (AAAI '07)*, 2007.

[BNG10] Hila Becker, Mor Naaman, and Luis Gravano. Learning similarity metrics for event identification in social media. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining (WSDM '10)*, 2010.

[BNG11a] Hila Becker, Mor Naaman, and Luis Gravano. Beyond trending topics: Real-world event identification on Twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM '11)*, 2011.

[BNG11b] Hila Becker, Mor Naaman, and Luis Gravano. Selecting quality Twitter content for events. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM '11)*, 2011.

[BNJ03] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[BP98] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the 7th International World Wide Web Conference (WWW7)*, 1998.

[BSR+05] Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. Learning to rank using gradient descent. In *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)*, 2005.

[BW03] Susanne Boll and Utz Westermann. MediAEther: An event space for context-aware multimedia experiences. In *Proceedings of the 2003 ACM SIGMM Workshop on Experiential Telepresence (ETP '03)*, 2003.

[BXNG10] Hila Becker, Bai Xiao, Mor Naaman, and Luis Gravano. Exploiting social links for event identification in social media. In *Proceedings of the Third Annual Workshop on Search in Social Media (SSM '10)*, 2010.

[CKM09] Zhaoqi Stella Chen, Dmitri V. Kalashnikov, and Sharad Mehrotra. Exploiting context analysis for combining multiple entity resolution systems. In *Pro-*

ceedings of the 2009 ACM International Conference on Management of Data (SIGMOD '09), 2009.

[CL09] Marc Cheong and Vincent Lee. Integrating Web-based intelligence retrieval and decision-making from the Twitter trends knowledge base. In *Proceeding of the Second ACM Workshop on Social Web Search and Mining (SWSM '09)*, 2009.

[CMS09] W. Bruce Croft, Donald Metzler, and Trevor Strohman. *Search Engines: Information Retrieval in Practice*. Addison-Wesley Publishing Company, 2009.

[CP11] Deepayan Chakrabarti and Kunal Punera. Event summarization using tweets. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM '11)*, 2011.

[CR02] William W. Cohen and Jacob Richman. Learning to match and cluster large high-dimensional data sets for data integration. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '02)*, 2002.

[CR09] Ling Chen and Abhishek Roy. Event detection from Flickr data through wavelet-based spatial analysis. In *Proceedings of the 18th ACM International Conference on Information and Knowledge Management (CIKM '09)*, 2009.

[CS01] Koby Crammer and Yoram Singer. Pranking with ranking. In *Advances in Neural Information Processing Systems 14 (NIPS '01)*, 2001.

[CS08] Riley Crane and Didier Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences*, 105(41):15649–15653, 2008.

[DAR09] Carlotta Domeniconi and Muna Al-Razgan. Weighted cluster ensembles: Methods and analysis. *ACM Transactions on Knowledge Discovery from Data*, 2(4):1–40, 2009.

[DDGR07] Abhinandan S. Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. Google news personalization: Scalable online collaborative filtering. In *Proceedings of the 16th International World Wide Web Conference (WWW '07)*, 2007.

[Dem06] Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.

[DI08] Wisam Dakka and Panagiotis G. Ipeirotis. Automatic extraction of useful facet hierarchies from text databases. In *Proceedings of the IEEE 24th International Conference on Data Engineering (ICDE '08)*, 2008.

[Diw03] Urmila M. Diwekar. *Introduction to Applied Optimization*. Springer, 2003.

[DK92] Daniel Dayan and Elihu Katz. *Media Events: The Live Broadcasting of History*. Harvard University Press, 1992.

[DKJ+07] Jason V. Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S. Dhillon. Information-theoretic metric learning. In *Proceedings of the 24th International Conference on Machine Learning (ICML'07)*, 2007.

[DLSL09] Bertrand De Longueville, Robin S. Smith, and Gianluca Luraschi. "OMG, from here, I can see the flames!": A use case of mining location based social networks to acquire spatio-temporal data on forest fires. In *Proceedings of the 2009 International Workshop on Location Based Social Networks (LBSN '09)*, 2009.

[DNKS10] Nicholas Diakopoulos, Mor Naaman, and Funda Kivran-Swaine. Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '10)*, 2010.

[DS10] Nicholas Diakopoulos and David A. Shamma. Characterizing debate performance via aggregated Twitter sentiment. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI '10)*, 2010.

[ER04] Günes Erkan and Dragomir R. Radev. Lexrank: Graph-based lexical centrality as salience in text summarization. *JAIR*, 22(1):457–479, 2004.

[Eve02] Events, 2002. In Stanford Encyclopedia of Philosophy. Retrieved June 2nd, 2010 from http://plato.stanford.edu/entries/events/.

[FH03] Elena Filatova and Vasileios Hatzivassiloglou. Domain-independent detection, extraction, and labeling of atomic events. In *Proceedings of Recent Advances in Natural Language Processing (RANLP '03)*, 2003.

[FHM06] Elena Filatova, Vasileios Hatzivassiloglou, and Kathleen McKeown. Automatic creation of domain templates. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (ACL-COLING '06)*, 2006.

[FISS03] Yoav Freund, Raj Iyer, Robert E. Schapire, and Yoram Singer. An efficient boosting algorithm for combining preferences. *Journal of Machine Learning Research*, 4:933–969, 2003.

[GGLNT04] Daniel Gruhl, R. Guha, David Liben-Nowell, and Andrew Tomkins. Information diffusion through blogspace. In *Proceedings of the 13th International World Wide Web Conference (WWW '04)*, 2004.

[GHT04] Natalie S. Glance, Matthew Hurst, and Takashi Tomokiyo. BlogPulse: Automated trend discovery for Weblogs. In *Proceedings of the WWW 2004 Workshop on the Weblogging Ecosystem*, 2004.

[GMT05] Aristides Gionis, Heikki Mannila, and Panayiotis Tsaparas. Clustering aggregation. In *Proceedings of the 21st International Conference on Data Engineering (ICDE '05)*, 2005.

[Gri10] Ralph Grishman. The impact of task and corpus on event extraction systems. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC '10)*, 2010.

[GSFW94] Deborah J. Gerner, Philip A. Schrodt, Ronald A. Francisco, and Judith L. Weddle. Machine coding of event data using regional and international sources. *International Studies Quarterly*, 38(1):91–119, 1994.

[HBB10] Matthew Hoffman, David Blei, and Francis Bach. Online learning for latent Dirichlet allocation. In *Advances in Neural Information Processing Systems 23 (NIPS '10)*, 2010.

[HCL07] Qi He, Kuiyu Chang, and Ee-Peng Lim. Analyzing feature trajectories for event detection. In *Proceedings of the 30th ACM International Conference on Research and Development in Information Retrieval (SIGIR '07)*, 2007.

[HD10] Liangjie Hong and Brian D. Davison. Empirical study of topic modeling in Twitter. In *Proceedings of the First Workshop on Social Media Analytics (SOMA '10)*, 2010.

[HGM00] Vasileios Hatzivassiloglou, Luis Gravano, and Ankineedu Maganti. An investigation of linguistic features and clustering algorithms for topical document clustering. In *Proceedings of the 23rd ACM International Conference on Research and Development in Information Retrieval (SIGIR '00)*, 2000.

[HHSC11] Brent Hecht, Lichan Hong, Bongwon Suh, and Ed H. Chi. Tweets from Justin Bieber's heart: The dynamics of the "location" field in user profiles. In *Proceedings of the 29th International Conference on Human Factors in Computing Systems (CHI '11)*, 2011.

[HS95] Mauricio A. Hernández and Salvatore J. Stolfo. The merge/purge problem for large databases. In *Proceedings of the 1996 ACM International Conference on Management of Data (SIGMOD '96)*, 1995.

[JG08] Heng Ji and Ralph Grishman. Refining event extraction through cross-document inference. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT '08)*, 2008.

[JZSC09] Bernard J. Jansen, Mimi Zhang, Kate Sobel, and Abdur Chowdury. Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60(11):2169–2188, 2009.

[KA04] Giridhar Kumaran and James Allan. Text classification and named entities for new event detection. In *Proceedings of the 27th ACM International Conference on Research and Development in Information Retrieval (SIGIR '04)*, 2004.

[KAKS97] George Karypis, Rajat Aggarwal, Vipin Kumar, and Shashi Shekhar. Multilevel hypergraph partitioning: Application in VLSI domain. In *Proceedings of the 34th ACM Conference on Design Automation (DAC '97)*, 1997.

[KGP+04] April Kontostathis, Leon Galitsky, William M. Pottenger, Soma Roy, and Daniel J. Phelps. A survey of emerging trend detection in textual data mining. In Michael W. Berry, editor, *A Comprehensive Survey of Text Mining - Clustering, Classification and Retrieval*. Springer, 2004.

[Kle03] Jon Kleinberg. Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery*, 7:373–397, 2003.

[KLPM10] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is Twitter, a social network or a news media? In *Proceedings of the 19th International World Wide Web Conference (WWW '10)*, 2010.

[KMC05] Reiner Kraft, Farzin Maghoul, and Chi Chao Chang. Y!Q: Contextual search at the point of inspiration. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management (CIKM '05)*, 2005.

[LBK09] Jure Leskovec, Lars Backstrom, and Jon M. Kleinberg. Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*, 2009.

[LBW07] Ping Li, Chris Burges, and Qiang Wu. McRank: Learning to rank using multiple classification and gradient boosting. In *Advances in Neural Information Processing Systems 20 (NIPS '07)*, 2007.

[LPS+08] Sophia Liu, Leysia Palen, Jeannette Sutton, Amanda Hughes, and Sarah Vieweg. In search of the bigger picture: The emergent role of on-line photo-sharing in times of disaster. In *Proceedings of the Information Systems for Crisis Response and Management Conference (ISCRAM '08)*, 2008.

[LS99] M. LeCompte and J. Schensul. *Designing and conducting ethnographic research*. Altamira Press, Walnut Creek, CA, 1999.

[LXQ+07] Tie-Yan Liu, Jun Xu, Tao Qin, Wenying Xiong, and Hang Li. LETOR: Benchmark dataset for research on learning to rank for information retrieval. In *Proceedings of SIGIR 2007 Workshop on Learning to Rank for Information Retrieval*, 2007.

[Mak03] Juha Makkonen. Investigations on event evolution in TDT. In *Proceedings of HLT-NAACL Student Workshop of Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, Edmonton, Canada, 2003.

[Mak09] Juha Makkonen. *Semantic Classes in Topic Detection and Tracking*. PhD thesis, University of Helsinki, Department of Computer Science, 2009.

[MAMS04] Juha Makkonen, Helena Ahonen-Myka, and Marko Salmenkivi. Simple semantics in topic detection and tracking. *Information Retrieval*, 7(3–4):347–368, 2004.

[MBN07] Fabian Moerchen, Klaus Brinker, and Claus Neubauer. Any-time clustering of high frequency news streams. In *Proceedings of the ACM SIGKDD Data Mining Case Studies Workshop (DMCS '07)*, 2007.

[MNU00] Andrew McCallum, Kamal Nigam, and Lyle H. Ungar. Efficient clustering of high-dimensional data sets with application to reference matching. In *Pro-*

ceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '00), 2000.

[MRS08] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge Univ. Press, 2008.

[NBG11] Mor Naaman, Hila Becker, and Luis Gravano. Hip and trendy: Characterizing emerging trends on Twitter. *Journal of the American Society for Information Science and Technology*, 62(5):902–918, 2011.

[NBL10] Mor Naaman, Jeffrey Boase, and Chih-Hui Lai. Is it really about me?: Message content in social awareness streams. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*, 2010.

[NGS+09] Meenakshi Nagarajan, Karthik Gomadam, Amit P. Sheth, Ajith Ranabahu, Raghava Mutharaju, and Ashutosh Jadhav. Spatio-temporal-thematic analysis of citizen sensor data: Challenges and experiences. In *Proceedings of the 10th International Conference on Web Information Systems Engineering (WISE '09)*, 2009.

[OBRS10] Brendan OConnor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM '10)*, 2010.

[OKA10] Brendan OConnor, Michel Krieger, and David Ahn. TweetMotif: Exploratory search and topic summarization for Twitter. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM '10)*, 2010.

[Osb02] Jason Osborne. Notes on the use of data transformations. *Practical Assessment, Research and Evaluation*, 8(6), 2002.

[PC98] Jay M. Ponte and W. Bruce Croft. A language modeling approach to information retrieval. In *Proceedings of the 21st ACM International Conference on Research and Development in Information Retrieval (SIGIR '98)*, 1998.

[PMS09] Owen Phelan, Kevin McCarthy, and Barry Smyth. Using Twitter to recommend real-time topical news. In *Proceedings of the Third ACM Conference on Recommender Systems (RecSys '09)*, 2009.

[POL10] Saša Petrović, Miles Osborne, and Victor Lavrenko. Streaming first story detection with application to Twitter. In *Proceedings of the 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL '10)*, 2010.

[RCD+11] Timo Reuter, Philipp Cimiano, Lucas Drumond, Krisztian Buza, and Lars Schmidt-Thieme. Scalable event-based clustering of social media via record linkage techniques. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM '11)*, 2011.

[RDL10] Daniel Ramage, Susan Dumais, and Dan Liebling. Characterizing microblogs with topic models. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM '10)*, 2010.

[RGN07] Tye Rattenbury, Nathaniel Good, and Mor Naaman. Towards automatic extraction of event and place semantics from Flickr tags. In *Proceedings of the 30th ACM International Conference on Research and Development in Information Retrieval (SIGIR '07)*, 2007.

[RH07] Andrew Rosenberg and Julia Hirschberg. V-Measure: A conditional entropy-based external cluster evaluation measure. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP '07)*, 2007.

[RJ05] Filip Radlinski and Thorsten Joachims. Query chains: Learning to rank from implicit feedback. In *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '05)*, 2005.

[RW99] Stephen E. Robertson and Steve Walker. Okapi/Keenbow at TREC-8. In *Proceedings of the 14th Text REtrieval Conference (TREC-8)*, 1999.

[SGC02] Alexander Strehl, Joydeep Ghosh, and Claire Cardie. Cluster ensembles – A knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3:583–617, 2002.

[SHK10] Beaux Sharifi, Mark-Anthony Hutton, and Jugal K. Kalita. Experiments in microblog summarization. In *Proceedings of the 2010 IEEE Second International Conference on Social Computing (SOCIALCOM '10)*, 2010.

[Sin84] Roger W. Sinnott. Virtues of the Haversine. *Sky and Telescope*, 68:159, 1984.

[SJ10] Vivek K. Singh and Ramesh Jain. Structural analysis of the emerging event-web. In *Proceedings of the 19th International World Wide Web Conference (WWW '10)*, 2010.

[SKC10] David A. Shamma, Lyndon Kennedy, and Elizabeth Churchill. Statler: Summarizing media through short-message services. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*, 2010.

[SOM10] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes Twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International World Wide Web Conference (WWW '10)*, 2010.

[SPHV10] Kate Starbird, Leysia Palen, Amanda L. Hughes, and Sarah Vieweg. Chatter on the red: What hazards threat reveals about the social life of microblogged information. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*, 2010.

[SPS08] Irina Shklovski, Leysia Palen, and Jeannette Sutton. Finding community through information and communication technology in disaster response. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW '08)*, 2008.

[SST+09] Jagan Sankaranarayanan, Hanan Samet, Benjamin E. Teitler, Michael D. Lieberman, and Jon Sperling. Twitterstand: News in tweets. In *Proceedings of the 17th ACM International Conference on Advances in Geographic Information Systems (GIS '09)*, 2009.

[TBW95] Stelios C. A. Thomopoulos, Dimitrios K. Bougoulias, and Chin-Der Wann. Dignet: An unsupervised-learning clustering algorithm for clustering and data fusion. *IEEE Transactions on Aerospace Electronic Systems*, 31:21–38, 1995.

[TdRW11] Manos Tsagkias, Maarten de Rijke, and Wouter Weerkamp. Linking online news and social media. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining (WSDM '11)*, 2011.

[WF94] Stanley Wasserman and Katherine Faust. *Social network analysis: Methods and applications*. Cambridge University Press, 1994.

[WF05] Ian H. Witten and Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2005.

[WL11] Jianshu Weng and Francis Lee. Event detection in Twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM '11)*, 2011.

[WZHS07] Xuanhui Wang, ChengXiang Zhai, Xiao Hu, and Richard Sproat. Mining correlated bursty topic patterns from coordinated text streams. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '07)*, 2007.

[XLW+08] Fen Xia, Tie-Yan Liu, Jue Wang, Wensheng Zhang, and Hang Li. Listwise approach to learning to rank: Theorem and algorithm. In *Proceedings of the 25th International Conference on Machine Learning (ICML '08)*, 2008.

[XNJR02] Eric P. Xing, Andrew Y. Ng, Michael I. Jordan, and Stuart Russell. Distance metric learning, with application to clustering with side-information. In *Advances in Neural Information Processing Systems 15 (NIPS '02)*, 2002.

[Yb10] Sarita Yardi and danah boyd. Tweeting from the town square: Measuring geographic local networks. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM '10)*, 2010.

[YCB+99] Yiming Yang, Jaime Carbonell, Ralf Brown, Thomas Pierce, Brian T. Archibald, and Xin Liu. Learning approaches for detecting and tracking news events. *IEEE Intelligent Systems Special Issue on Applications of Intelligent Information Retrieval*, 14(4):32 – 43, 1999.

[YPC98] Yiming Yang, Thomas Pierce, and Jaime Carbonell. A study on retrospective and on-line event detection. In *Proceedings of the 21st ACM International Conference on Research and Development in Information Retrieval (SIGIR '98)*, 1998.

[Zhu05] Xiaojin Zhu. Semi-supervised learning literature survey. Technical Report 1530, Computer Sciences Department, University of Wisconsin-Madison, 2005.

[ZJH+11] Wayne Xin Zhao, Jing Jiang, Jing He, Yang Song, Palakorn Achananuparp, Ee-Peng Lim, and Xiaoming Li. Topical keyphrase extraction from Twitter. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT '11)*, 2011.

[ZRL96] Tian Zhang, Raghu Ramakrishnan, and Miron Livny. BIRCH: An efficient data clustering method for very large databases. In *Proceedings of the 1996 ACM International Conference on Management of Data (SIGMOD '96)*, 1996.

[ZT01] Jeffrey M. Zacks and Barbara Tversky. Event structure in perception and conception. *Psychological Bulletin*, 127, 2001.

[ZZW07] Kuo Zhang, Juan Zi, and Li Gang Wu. New event detection based on indexing-tree and named entity. In *Proceedings of the 30th ACM International Conference on Research and Development in Information Retrieval (SIGIR '07)*, 2007.

2012

Harvesting real time traffic information from Twitter

Type	Journal Article
Author	Raymondus Kosala
Author	Erwin Adi
Rights	Elsevier Publishing
Volume	50
Pages	1-11
Publication	Procedia engineering
ISSN	1877-7058
Date	2012
DOI	[10.1016/j.proeng.2012.10.001](https://doi.org/10.1016/j.proeng.2012.10.001)
Library Catalog	www.safetylit.org
Language	en

Author Abstract: The flow of information on Twitter holds a wealth of information that can be extracted for specific uses. This study attempted to extract road traffic information from Twitter timelines for real time mapping. Given the worsening traffic conditions in Jakarta, a real time map of traffic information could be very beneficial for motorists as they deal with everyday road conditions. With the virtually indeterminate sources of data on Twitter, we proposed an algorithm to measure the traffic information confidence level for the real time traffic mapping system. Based on the test results of this study, the system could be used to serve its intended purpose. Copyright © 2012 The authors.

ELSEVIER NOTICE: Further to an investigation the International Conference on Advances Science and Contemporary Engineering 2012, Procedia Engineering, Volume 50, has been removed by the publisher due to insufficient assurances by the programme organisers that the professional ethical codes of publishing and standards were applied consistently. Where the copyright in individual papers was transferred to the publisher, these transfers are now considered null and void, and revert back to the authors.

Abstract Nonetheless this paper is available online at: https://www.researchgate.net/publication/232906167_Harvesting_Real_Time_Traffic_Information_from_Twitter

SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based.

References

[1] Triyono H. Pertumbuhan Kendaraan di Jakarta Capai 1.172 per Hari. Available: http://www.tempointeraktif.com/hg/layanan_publik/2010/07/28/brk,20100728-267039,id.html, 2010.

[2] Kusumaputra R.A. Wow! Kerugian akibat Kemacetan di Jakarta Rp 28 Triliun Setahun. Available: http://megapolitan.kompas.com/read/2009/09/10/14592832/wow.kerugian.akibat.kemacetan.di.jakarta.rp.28.triliun.setahun, 2009.

[3] Kumar P., Singh V., and Reddy D., Advanced Traveler Information System for Hyderabad City, IEEE Trans. Intell. Transp. Syst., vol. 6, no. 1, March, 2005.

[4] Ezell S. Explaining International IT Application Leadership: Intelligent Transportation Systems. Report of The Information Technology and Innovation Foundation (ITIF). Available: www.itif.org/files/2010-1-27-ITS_Leadership.pdf, 2010.

[5] IEEE-USA Board of Directors. U.S. Economic Competitiveness and Intelligent Transportation Systems Technology. Available: www.ieeeusa.org/policy/positions/ITS.pdf, 2010.

[6] Wu C. H., Su D. C., Chang J., Wei C.C., Ho J.M., Lin K.J., and Lee D.T. An Advanced Traveler Information System with Emerging Network Technologies, In the Proceedings of the 6th Asia-Pacific Intelligent Transportation Systems Forum, 2003.

[7] Lotan G., Graeff E., Ananny M., Gaffney D., Pearce I., and Boyd D., The Revolutions Were Tweeted: Information Flows During the 2011 Tunisian and Egyptian Revolutions, International Journal of Communication, vol. 5, 2011.

[8] A. Leavitt, E. Burchard, D. Fisher, S. Gilbert. The Influentials: New Approaches for Analyzing Influence on Twitter. Webecology Project. Available: http://www.webecologyproject.org/wp-content/uploads/2009/09/influence-report-final.pdf, 2009.

[9] B. Heil and M. piskorski. New Twitter Research: Men Follow Men and Nobody Tweets. Harvard Business School. Available: http://blogs.hbr.org/cs/2009/06/new_twitter_research_men_follo.html, 2009.

[10] H. Kwak, C. Lee, H. Park and S. Moon. What is Twitter, a Social Network or a News Media?, In the Proceedings of the International World Wide Web Conference, 2010.

[11] Wanichayapong N., Pruthipunyaskul W., Pattara-Atikom W., Chaovalit P. Social-based traffic information extraction and classification. In the Proceedings of the 11th International Conference on ITS Telecommunications, 2011.

[12] Endarnoto S.K., Pradipta S., Nugroho A.S., Purnama J. Traffic Condition Information Extraction & Visualization from Social Media Twitter for Android Mobile Application. In the Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, July 2011.

2013

Improving traffic prediction with tweet semantics.

Type	Journal Article
Author	Jingrui He
Author	Wei Shen
Author	Phani Divakaruni
Author	Laura Wynter
Author	Rick Lawrence
URL	http://www.safetylit.org/citations/index.php?fuseaction=citations.viewdetails&citationIds[]=citconferenceproceeding_394_34
Pages	1387-1393
Publication	AAAI Press
Date	2013
Library Catalog	www.safetylit.org

Abstract

Road traffic prediction is a critical component in modern smart transportation systems. It provides the basis for traffic management agencies to generate proactive traffic operation strategies for alleviating congestion. Existing work on near-term traffic prediction (forecasting horizons in the range of 5 minutes to 1 hour) relies on the past and current traffic conditions. However, once the forecasting horizon is beyond 1 hour, i.e., in longer-term traffic prediction, these techniques do not work well since additional factors other than the past and current traffic conditions start to play important roles. To address this problem, in this paper, for the first time, we examine whether it is possible to use the rich information in online social media to improve longer-term traffic prediction. To this end, we first analyze the correlation between traffic volume and tweet counts with various granularities. Then we propose an optimization framework to extract traffic indicators based on tweet semantics using a transformation matrix, and incorporate them into traffic prediction via linear regression. Experimental results using traffic and Twitter data originated from the San Francisco Bay area of California demonstrate the effectiveness of our proposed framework.

Available: http://www.aaai.org/ocs/index.php/IJCAI/IJCAI13/paper/view/6776

Keywords: Twitter-Traffic-Status

SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based.

References
[Ahmed and Cook, 1979] M. S. Ahmed and A. R. Cook. Analysis of freeway traffic time-series data by using box-jenkins techniques. Transportation Research

Board, 722:1-9, 1979.

[Al-Deek et al., 2001] H. Al-Deek, S. Ishak, and M. Wang. A new short-term traffic prediction and incident detection system on i-4, vol. i. final research report. Technical report, Transportation Systems Institute(TSI), Department of Civil and Environmental Engineering, University of Central Florida, 2001.

[Bollen et al., 2010] Johan Bollen, Huina Mao, and XiaoJun Zeng. Twitter mood predicts the stock market. CoRR, abs/1010.3003, 2010.

[Clark et al., 1993] S. D. Clark, M. S. Dougherty, and H. R. Kirby. The use of neural networks and time series models for short-term traffic forecasting: a comparative study. Proceedings of the PTRC 21st Summer Annual Meeting, 1993.

[Clark, 2003] S. Clark. Traffic prediction using multivariate nonparametric regression. Journal of Transportation Engineering, 129:161-168, 2003.

[Eisenstein et al., 2010] Jacob Eisenstein, Brendan O'Connor, Noah A. Smith, and Eric P. Xing. A latent variable model for geographic lexical variation. In EMNLP, pages 1277-1287, 2010.

[Friedman et al., 2010] J. H. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software, 33(1):1-22, 2 2010.

[Go et al., 2011] Alec Go, Richa Bhayani, and Lei Huang. Twitter sentiment classification using distant supervision. Technical report, Stanford University, 2011.

[Guo and Williams, 2010] J. Guo and B. M. Williams. Realtime short-term traffic speed level forecasting and uncertainty quantification using layered kalman fiters. Transportation Research Record, 2175:28-37, 2010.

[Huang and Sadek, 2009] S. Huang and A. W. Sadek. A novel forecasting approach inspired by human memory: The example of short-term traffic volume forecasting. Transportation Research Part C, 17:510-525, 2009.

[Kamarianakis and Prastacos, 2003] Y. Kamarianakis and P. Prastacos. Forecasting traffic flow conditions in an urban network: comparison of multivariate and univariate approaches. Transportation Research Record, 1858:74- 84, 2003.

[Khosravi et al., 2011] A. Khosravi, E. Mazloumi, S. Nahavandi, D. Creighton, and J. Van Lint. A genetic algorithmbased method for improving quality of travel time prediction intervals (in press). Transportation Research Part C, 2011. [Levin and Tsao, 1980] M. Levin and Y.-D. Tsao. On forecasting freeway occupancies and volumes. Transportation Research Record, 773:47-49, 1980.

[Mahmassani et al., 2009] H. S. Mahmassani, J. Dong, J. Kim, R. B. Chen, and B. Park. Incorporating weather impacts in traffic estimation and prediction systems. Technical Report FHWA-JPO-09-065, Northwestern University, 2009.

[Maze et al., 2006] T.H. Maze, M. Agarwal, and G. Burchett. Whether weather matters to traffic demand, traffic safety, and traffic operations and flows. Transportation Research Record, 1948:170-176, 2006. [Min and Wynter, 2011] W. Min and L. Wynter. Real-time road traffic prediction with spatio-temporal correlations. Transportation Research Part C, 19:606-616, 2011.

[Okutani and Stephanedes, 1984] I. Okutani and Y. J. Stephanedes. Dynamic prediction of traffic volume through kalman filtering theory. Transportation Research Part B, 18:1-11, 1984.

[Petrovic et al., 2010] Sasa Petrovic, Miles Osborne, and Victor Lavrenko. Streaming first story detection with application to twitter. In HLT-NAACL, pages 181-189, 2010.

[Sakaki et al., 2010] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In WWW, pages 851- 860, 2010. [

Sankaranarayanan et al., 2009] Jagan Sankaranarayanan, Hanan Samet, Benjamin E. Teitler, Michael D. Lieberman, and Jon Sperling. Twitterstand: news in tweets. In GIS, pages 42-51, 2009.

[Smith and Demetsky, 1996] B. L. Smith and M. J. Demetsky. Multiple-interval freeway traffic flow forecasting. Transportation Research Record, 1554:136-141, 1996.

[Smith and Demetsky, 1997] B. L. Smith and M. J. Demetsky. Traffic flow forecasting: comparison of modelling approaches. Journal of Transportation Engineering, 123(4):261-266, 1997.

[Smith et al., 2002] B. L. Smith, B. M. Williams, and R. K. Oswalsd. Comparison of parametric and nonparametric models for traffic flow forecasting. Transportation Research Part C, 10:303-321, 2002.

[Sriram et al., 2010] Bharath Sriram, Dave Fuhry, Engin Demir, Hakan Ferhatosmanoglu, and Murat Demirbas. Short text classification in twitter to improve information filtering. In SIGIR, pages 841-842, 2010.

[van Lint et al., 2005] J. van Lint, S. Hoogendoorn, and H. van Zuylen. Accurate freeway travel time prediction with state-space neural networks under missing data. Transportation Research Part C, 13:347-369, 2005.

[Vlahogianni et al., 2005] E. Vlahogianni, M. G. Karlaftis, and J. C. Golias.

Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach. Transportation Research Part C, 13:211-234, 2005.

[Vythoulkas, 1993] P. C. Vythoulkas. Alternative approaches to short-term traffic forecasting for use in driver information systems. In Transportation and Traffic Theory, Proceedings of the 12th International Symposium on Traffic Flow Theory and Transportation, Berkeley, CA, July 1993.

[Williams et al., 1998] B. M. Williams, P.K. Durvasula, and D. E. Brown. Urban traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models. Transportation Research Record, 1644:132-144, 1998.

[Yun et al., 1998] S.-Y. Yun, S. Namkoong, J.-H. Rho, S.-W. Shin, and J.-U. Choi. A performance evaluation of neural network models in traffic volume forecasting. Mathematical Computer Modelling, 27:293-310, 1998

Language: en

Monitoring urban traffic status using Twitter messages

Type	Thesis
Author	Fatima Amin Elsafoury
URL	https://www.itc.nl/library/papers_2013/msc/gfm/elsafoury.pdf
Place	Eschede, Netherlands
Date	2013
Type	Masters
Language	en
University	University of Twente, Faculty of Geo-Information Science and Earth Observation
Abstract	Traffic congestion is a worldwide problem. All large cities and capitals around the world suffer from this problem. Traffic congestions cost money and time. The existing tools used to help in collecting information about traffic are expensive and have limitations. Nowadays, micro-bloggers are being used widely. It allows people to share information, opinions and stories in short messages. Twitter is a very popular micro-blogger. It allows people to share whatever they want in 140 characters. Twitter offers a new source of information for variety of topics. This research proposes a system to use traffic information shared by Twitter messages (tweets) in a real time manner. It uses a customized Part of speech (POS) tagging method for extracting information from the tweets. POS is also used for Geo-locating the tweets with custom developed locations' dictionaries. Google Geo-Code API is also used in the geo-locating task. It also follows the traffic information sent by @TfLTrafficNews which is an official Twitter account used for reporting about traffic in London. A prototype of the proposed system was implemented. This prototype contains implementation of the proposed POS algorithm and also the implementation of the system work flow. The result of the system is a map showing a highlighted route. This route is the location (road) mentioned in the tweet. The highlight colour depends on the traffic status which is also mentioned in the tweet. The results were tested by comparing it against Google Maps traffic feature. The results could be helpful for future work. Keywords: Twitter, Twitter Streaming API, Traffic congestion, POS, Google Geo-Code API, Google Maps API Twitter-Traffic-Status References Abel, F., Hauff, C., Houben, G. J., Stronkman, R., & Tao, K. (2012). Twitcident: fighting fire with information from social web streams. Ahmad, S. (2007). Tutorial on Natural Language Processing. Artificial Intelligence, 810(161). . Arnott, R., & Small, K. (1994). The economics of traffic congestion. American Scientist, 446-455. Bikel, D. M., Miller, S., Schwartz, R., & Weischedel, R. (1997). Nymble: a high-performance learning name-finder. Paper presented at the Proceedings of the fifth conference on Applied natural language processing. Bontcheva, K., Tablan, V., Maynard, D., & Cunningham, H. (2004). Evolving GATE to meet new challenges in language engineering. Natural Language Engineering, 10(3-4), 349-373. doi: doi:10.1017/S1351324904003468 Brill, E. (1992). A simple rule-based part of speech tagger. Bruns, A., & Burgess, J. E. (2011). # Ausvotes: how Twitter covered the 2010 Australian federal election. Communication, Politics and Culture, 44(2), 37-56.

Bryant, M. a. B., M. (2010). Britain's most congested street is in south London Retrieved 10-02, 2013, from http://www.standard.co.uk/news/britains-most-congested-street-is-in-south-london- 6519888.html

Carvalho, S. F. L. (2012). Real-time sensing of traffic information in twitter messages.

Cheng, Z., Caverlee, J., & Lee, K. (2010). You are where you tweet: a content-based approach to geo-locating twitter users. Paper presented at the Proceedings of the 19th ACM international conference on Information and knowledge management, Toronto, ON, Canada.

Collobert, R., Weston, J., #233, Bottou, o., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural Language Processing (Almost) from Scratch. J. Mach. Learn. Res., 12, 2493-2537.

Endarnoto, S. K., Pradipta, S., Nugroho, A. S., & Purnama, J. (2011). Traffic Condition Information Extraction & Visualization from Social Media Twitter for Android Mobile Application. .

The Google Geocoding API. (2012) Retrieved 08-01-2013, 2013

Green, A. (2013a). 140dev Twitter Framework Retrieved 15-12, 2013, from http://140dev.com/free- twitter-api-source-code-library/

Green, A. (2013b). Free Source Code – Twitter Database Server: Code Architecture Retrieved 15-12, 2013, from http://140dev.com/free-twitter-api-source-code-library/twitter-database-server/code-architecture/

Guduru, N. (2006). Text Mining with Support Vector Machines and Non-negative Matrix Factorization Algorithms. University of Rhode Island.

Guo, J. (1997). Critical tokenization and its properties. Computational Linguistics, 23(4), 569-596.

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we twitter: understanding microblogging usage and communities.

MacEachren, A. M., Jaiswal, A., Robinson, A. C., Pezanowski, S., Savelyev, A., Mitra, P., . . . Blanford, J. (2011). SensePlace2: GeoTwitter analytics support for situational awareness.

Mansouri, A., Affendey, L. S., & Mamat, A. (2008). Named entity recognition approaches. International Journal of Computer Science and Network Security, 8, 339-344.

Merialdo, B. (1994). Tagging English text with a probabilistic model. Comput. Linguist., 20(2), 155-171.

. MYSQL. (2013) Retrieved 15-12, 2013, from http://www.mysql.com/ Paradesi, S. (2011). Geotagging tweets using their content. Proceedings of the Twenty-fourth International Florida Artificial Intelligence Research Society Conference, 355-356.

Patell, P. (2011). Experiments on phrasal chunking in nlp using exponentiated gradient for structured prediction. . The Penn Treebank Tag Set. (1998) Retrieved 15-02, 2013, from http://www.ims.uni- stuttgart.de/projekte/CorpusWorkbench/CQP-HTMLDemo/PennTreebankTS.html

Ritter, A., Clark, S., Mausam, & Etzioni, O. (2011). Named Entity Recognition in Tweets: An Experimental Study.

Robinson, A. C. (2009a). MARKOV MODELS Retrieved 12-01, 2013, from http://language.worldofcomputing.net/pos-tagging/markov-models.html

Robinson, A. C. (2009b). MAXIMUM ENTROPY Retrieved 12-01-2013, 2013, from http://language.worldofcomputing.net/pos-tagging/maximum-entropy.html

Robinson, A. C. (2009c). PARTS-OF-SPEECH TAGGING Retrieved 12 january, 2013, from http://language.worldofcomputing.net/pos-tagging/parts-of-speech-tagging.html

Robinson, A. C. (2009d). Transformation Based Learning Retrieved 15-01, 2013, from http://language.worldofcomputing.net/pos-tagging/transformation-based-learning.html#

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection by social sensors.

Sang, E. F. T. K., & Meulder, F. D. (2003). Introduction to the CoNLL-2003 shared task: language-independent named entity recognition. Paper presented at the Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003 - Volume 4, Edmonton, Canada.

Sekine, S. (1998). NYU: Description of the Japanese NE System used for MET-2. Paper presented at the Proc. of the Seventh Message Understanding Conference (MUC-7).

. The Streaming APIs. (2012) Retrieved 15-09, 2013, from https://dev.twitter.com/docs/streaming-apis

.Traffic conditions on Google Maps. (2009) Retrieved 15-01, 2012, from . Transport for London. (2013), from http://www.tfl.gov.uk/

Wallach, H. M. (2004). Conditional random fields: An introduction. Technical Reports (CIS), 22.

Wanichayapong, N., Pruthipunyaskul, W., Pattara-Atikom, W., & Chaovalit, P. (2011). Social-based traffic information extraction and classification. Paper presented at the ITS Telecommunications (ITST), 2011 11th International Conference on.

Yerva, S. R., Miklós, Z., & Aberer, K. (2010). It was easy, when apples and blackberries were only fruits. Paper presented at the Third Web People Search Evaluation Forum (WePS-3), CLEF. Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010).

Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. World Medical & Health Policy, 2(2), 7.

# of Pages 46

2014

Citybeat: real-time social media visualization of hyper-local city data

Type	Conference Paper
Author	Chaolun Xia
Author	Raz Schwartz
Author	Ke Xie
Author	Adam Krebs
Author	Andrew Langdon
Author	Jeremy Ting
Author	Mor Naaman
Place	New York, NY, USA
Publisher	ACM
Pages	167-170
ISBN	978-1-4503-2745-9
Date	2014
DOI	10.1145/2567948.2577020
Library Catalog	www.safetylit.org
Conference Name	23rd International Conference on World Wide Web; Seoul, South Korea, April 7-11, 2014
Language	en

Abstract

With the increasing volume of location-annotated content from various social media platforms like Twitter, Instagram and Foursquare, we now have real-time access to people's daily documentation of local activities, interests and attention. In this demo paper, we present CityBeat, a real-time visualization of hyper-local social media content for cities. The main objective of CityBeat is to provide users -- with a specific focus on journalists -- with information about the city's ongoings, and alert them to unusual activities. The system collects a stream of geo-tagged photos as input, uses time series analysis and classification techniques to detect hyper-local events, and compute trends and statistics. The demo includes a visualization of this information that is designed to be installed on a large-screen in a newsroom, as an ambient display.

DOI: 10.1145/2567948.2577020
ISBN: 978-1-4503-2745-9

SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based.

Note: OCR errors may be found in this Reference List extracted from the full text article. ACM has opted to expose the complete List rather than only correct and linked references.

References
1 Hila Becker , Mor Naaman , Luis Gravano, Learning similarity metrics for event identification in social media, Proceedings of the third ACM international conference on Web search and data mining, February 04-06, 2010, New York, New York, USA [doi 10.1145/1718487.1718524]

2 Ling Chen , Abhishek Roy, Event detection from flickr data through wavelet-based spatial

analysis, Proceedings of the 18th ACM conference on Information and knowledge management, November 02-06, 2009, Hong Kong, China [doi 10.1145/1645953.1646021]

3 Yuheng Hu , Shelly D. Farnham , Andrés Monroy-Hernández, Whoo.ly: facilitating information seeking for hyperlocal communities using social media, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, April 27-May 02, 2013, Paris, France [doi 10.1145/2470654.2466478]

4 Michael Mathioudakis , Nick Koudas, TwitterMonitor: trend detection over the twitter stream, Proceedings of the 2010 ACM SIGMOD International Conference on Management of data, June 06-10, 2010, Indianapolis, Indiana, USA [doi 10.1145/1807167.1807306]

5 Joaquin Quionero-Candela , Masashi Sugiyama , Anton Schwaighofer , Neil D. Lawrence, Dataset Shift in Machine Learning, The MIT Press, 2009

6 Takeshi Sakaki , Makoto Okazaki , Yutaka Matsuo, Earthquake shakes Twitter users: real-time event detection by social sensors, Proceedings of the 19th international conference on World wide web, April 26-30, 2010, Raleigh, North Carolina, USA [doi 10.1145/1772690.1772777]

7 Raz Schwartz, Mor Naaman, and Ziad Matni. Making sense of cities using social media: Requirements for hyper-local data aggregation tools. In Proceedings of the International AAAI Conference on Weblogs and Social Media, 2013.

8 Maximilian Walther , Michael Kaisser, Geo-spatial event detection in the twitter stream, Proceedings of the 35th European conference on Advances in Information Retrieval, March 24-27, 2013, Moscow, Russia [doi 10.1007/978-3-642-36973-5_30]

9 Andreas Weiler , Marc H. Scholl , Franz Wanner , Christian Rohrdantz, Event identification for local areas using social media streaming data, Proceedings of the ACM SIGMOD Workshop on Databases and Social Networks, p.1-6, June 22-27, 2013, New York, New York [doi 10.1145/2484702.2484703]

10 Ke Xie , Chaolun Xia , Nir Grinberg , Raz Schwartz , Mor Naaman, Robust detection of hyper-local events from geotagged social media data, Proceedings of the Thirteenth International Workshop on Multimedia Data Mining, p.1-9, August 11-11, 2013, Chicago, Illinois [doi 10.1145/2501217.2501219] Keywords: Twitter-Traffic-Status

Language: en

Proceedings Title Proceedings 23rd International Conference on World Wide Web

A search and summary application for traffic events detection based on Twitter data

Type	Conference Paper
Author	Meiling Liu
Author	Kuang Fu
Author	Chang-Tien Lu
Author	Guangsheng Chen
Author	Huiqiang Wang
Place	Dallas, TX, USA
Publisher	ACM Press
Pages	549-552
Date	2014
DOI	10.1145/2666310.2666366
Library Catalog	www.safetylit.org
Conference Name	22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information
Language	en
Abstract	As a form of social media, Twitter records real life events in our cities as they happen. Huge numbers of tweets under the heading of transportation or metro are published every day. This paper presents an application for Traffic Events Detection and Summary (TEDS) based on mining representative terms from the tweets posted when anomalies occur. The proposed ensemble application contains an efficient TEDS search engine with multiple indexing, ranking, and scoring schemes. Spatio-temporal analysis and a novel wavelet analysis model are applied for traffic event detection. This application could benefit both drivers and transportation authorities. Users can search transportation status and analyze traffic events in specific locations of interest. Utilizing the proposed signal processing technology, we demonstrate the system's effectiveness by examining traffic and metro travel in the Washington D.C. area. As the collaboration between a citizen's life and social media becomes ever greater, this could have a significant impact on the prediction of traffic flow, travel selection, and other city computing functions. Available: DOI 10.1145/2666310.2666366 Keywords: Twitter-Traffic-Status Language: en SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based. References Note: OCR errors may be found in this Reference List extracted from the full text article. ACM has opted to expose the complete List rather than only correct and linked references. 1 Bertrand De Longueville , Robin S. Smith , Gianluca Luraschi, "OMG, from here, I can

see the flames!": a use case of mining location based social networks to acquire spatio-temporal data on forest fires, Proceedings of the 2009 International Workshop on Location Based Social Networks, November 03-03, 2009, Seattle, Washington (doi 10.1145/1629890.1629907)

2 Shiels, Mark. BBC News: Twitter co-founder Jack Dorsey rejoins company, http://www.bbc.co.uk/news/business-12889048 (2011), {Online; accessed 01-November-2011}

3 Kosala, Raymondus, and Erwin Adi. "Harvesting real time traffic information from Twitter." Procedia Engineering 50 (2012): 1-11.

4 Yiming Yang , Tom Pierce , Jaime Carbonell, A study of retrospective and on-line event detection, Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, p.28-36, August 24-28, 1998, Melbourne, Australia (doi 10.1145/290941.290953)

5 Jon Kleinberg, Bursty and hierarchical structure in streams, Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, July 23-26, 2002, Edmonton, Alberta, Canada (doi 10.1145/775047.775061)

6 Weng, Jianshu, and Bu-Sung Lee. "Event detection in Twitter." ICWSM 11 (2011): 401--408.

7 Gerard Salton , Christopher Buckley, Term-weighting approaches in automatic text retrieval, Information Processing and Management: an International Journal, v.24 n.5, p.513-523, 1988 (doi 10.1016/0306-4573(88)90021-0)

8 Liu, Mei-Ling, et al. "Research on dynamic multi-document summarization by topic detection and tracking technology." Journal of Harbin Institute of Technology 11 (2010): 020.

9 Bingsheng Wang , Haili Dong , Arnold P. Boedihardjo , Chang-Tien Lu , Harland Yu , Ing-Ray Chen , Jing Dai, An integrated framework for spatio-temporal-textual search and mining, Proceedings of the 20th International Conference on Advances in Geographic Information Systems, November 06-09, 2012, Redondo Beach, California (doi 10.1145/2424321.2424418)

10 Sumit Shah , Fenye Bao , Chang-Tien Lu , Ing-Ray Chen, CROWDSAFE: crowd sourcing of crime incidents and safe routing on mobile devices, Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, November 01-04, 2011, Chicago, Illinois (doi 10.1145/2093973.2094064)

Proceedings Title Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information

Discovering social events through online attention

Type	Journal Article
Author	Dror Y. Kenett
Author	Fred Morstatter
Author	H. Eugene Stanley
Author	Huan Liu
Rights	Public Library of Science
Volume	9
Issue	7
Pages	e102001
Publication	PLoS one
ISSN	1932-6203
Date	2014
DOI	10.1371/journal.pone.0102001
Library Catalog	www.safetylit.org
Language	en

Abstract

Twitter is a major social media platform in which users send and read messages ("tweets") of up to 140 characters. In recent years this communication medium has been used by those affected by crises to organize demonstrations or find relief. Because traffic on this media platform is extremely heavy, with hundreds of millions of tweets sent every day, it is difficult to differentiate between times of turmoil and times of typical discussion. In this work we present a new approach to addressing this problem. We first assess several possible "thermostats" of activity on social media for their effectiveness in finding important time periods. We compare methods commonly found in the literature with a method from economics. By combining methods from computational social science with methods from economics, we introduce an approach that can effectively locate crisis events in the mountains of data generated on Twitter. We demonstrate the strength of this method by using it to locate the social events relating to the Occupy Wall Street movement protests at the end of 2011.

Language: en

Keywords: Twitter-Traffic-Status

SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based.

References Tsukayama H (2013) Twitter turns 7: Users send over 400 million tweets per day. The Washington Post. Available" http://www.washingtonpost.com/business/technology/twitter-turns-7-users-send-over-400-million-tweets-per-day/2013/03/21/2925ef60-9222-11e2-bdea-e32ad90da239_story.html. Accessed 2014 Jul 4.

Kumar S, Morstatter F, Liu H (2014) Twitter Data Analytics. Springer.

Lazer D, Pentland AS, Adamic L, Aral S, Barabasi AL, et al. (2009) Life in the network: the coming age of computational social science. Science (New York, NY) 323: 721. doi: 10.1126/science.1167742

Conte R, Gilbert N, Bonelli G, Cioffi-Revilla C, Deffuant G, et al. (2012) Manifesto of computational social science. The European Physical Journal Special Topics 214: 325–346. doi: 10.1140/epjst/e2012-01697-8

Rybski D, Buldyrev SV, Havlin S, Liljeros F, Makse HA (2012) Communication activity in a social network: relation between long-term correlations and inter-event clustering. Scientific reports 2.

Gallos LK, Rybski D, Liljeros F, Havlin S, Makse HA (2012) How people interact in evolving online affiliation networks. Physical Review X 2: 031014. doi: 10.1103/physrevx.2.031014

Ciulla F, Mocanu D, Baronchelli A, Gonçalves B, Perra N, et al. (2012) Beating the news using social media: the case study of American Idol. EPJ Data Science 1: 1–11. doi: 10.1140/epjds8

Gonzalez MC, Hidalgo CA, Barabasi AL (2008) Understanding individual human mobility patterns. Nature 453: 779–782. doi: 10.1038/nature06958 Eagle N, Pentland AS, Lazer D (2009) Inferring friendship network structure by using mobile phone data. Proceedings of the National Academy of Sciences 106: 15274–15278. doi: 10.1073/pnas.0900282106

Havlin S, Kenett DY, Ben-Jacob E, Bunde A, Cohen R, et al. (2012) Challenges in network science: Applications to infrastructures, climate, social systems and economics. European Physical Journal-Special Topics 214: 273. doi: 10.1140/epjst/e2012-01695-x

Gao J, Hu J, Mao X, Perc M (2012) Culturomics meets random fractal theory: insights into long-range correlations of social and natural phenomena over the past two centuries. Journal of The Royal Society Interface 9: 1956–1964. doi: 10.1098/rsif.2011.0846

Kenett DY, Portugali J (2012) Population movement under extreme events. Proceedings of the National Academy of Sciences 109: 11472–11473. doi: 10.1073/pnas.1209306109

Moat H, Curme C, Avakian A, Kenett DY, Stanley HE, et al. (2013) Quantifying wikipedia usage patterns quantifying wikipedia usage patterns before stock market moves. Scientific Reports 3: 1801. doi: 10.1038/srep01801

Preis T, Moat HS, Stanley HE (2013) Quantifying trading behavior in financial markets using google trends. Scientific Reports 3: 1684. doi: 10.1038/srep01684

Moat HS, Preis T, Olivola CY, Liu C, Chater N (2014) Using big data to predict collective behavior in the real world. Behavioral and Brain Sciences 37: 92–93. doi: 10.1017/s0140525x13001817

De Longueville B, Smith RS, Luraschi G (2009) "OMG, from here, I can see the flames!": A use case of mining location based social networks to acquire spatio-temporal data on forest fires. In: Proceedings of the 2009 International Workshop on Location Based Social Networks. New York, NY, USA: ACM, LBSN'09, pp. 73–80. doi:10.1145/1629890.1629907. URL http://doi.acm.org/10.1145/1629890.1629907.

Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World wide web. New York, NY, USA: ACM, WWW'10, pp. 851–860. doi:10.1145/1772690.1772777. Available: http://doi.acm.org/10.1145/1772690.1772777. Accessed 2014 Jul 4.

Morstatter F, Lubold N, Pon-Barry H, Pfeffer J, Liu H (2014) Finding eyewitness tweets during crises. In: Association of Computational Lingustics Workshop on Language Technologies and Association of Computational Lingustics Workshop on Language Technologies and Computational Social Science.

Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. Phys Rev Lett 86: 3200–3203. doi: 10.1103/physrevlett.86.3200

Nagarajan M, Purohit H, Sheth A (2010) A qualitative examination of topical tweet and retweet practices. In: Fourth International AAAI Conference on Weblogs and Social Media. AAAI.

Kwak H, Lee C, Park H, Moon S (2010) What is twitter, a social network or a news media? In: Proceedings of the 19th international conference on World wide web. New York, NY, USA: ACM, WWW'10, pp. 591–600. doi:10.1145/1772690.1772751.

Boyd D, Golder S, Lotan G (2010) Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In: System Sciences (HICSS), 2010 43rd Hawaii International Conference on. pp. 1–10. doi:10.1109/HICSS.2010.412.

Mendoza M, Poblete B, Castillo C (2010) Twitter under crisis: can we trust what we RT? In: Proceedings of the First Workshop on Social Media Analytics. New York, NY, USA: ACM, SOMA '10, pp. 71–79. doi:10.1145/1964858.1964869. Available: http://doi.acm.org/10.1145/1964858.1964869. Accessed 2014 Jul4.

Yang L, Sun T, Zhang M, Mei Q (2012) We know what @you #tag: does the dual role affect hashtag adoption? In: Proceedings of the 21st international conference on World Wide Web. New York, NY, USA: ACM, WWW'12, pp. 261–270. doi:10.1145/2187836.2187872. Available: http://doi.acm.org/10.1145/2187836.2187872. Accessed 2014 Jul 4.

Romero DM, Meeder B, Kleinberg J (2011) Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In: Proceedings of the 20th international conference on World wide web. New York, NY, USA: ACM, WWW'11, pp. 695–704. doi:10.1145/1963405.1963503. URL.

EfronM(2010) Hashtag retrieval in a microblogging environment. In: Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. New York, NY, USA: ACM, SIGIR '10, pp. 787–788. doi:10.1145/1835449.1835616.

Weng J, Lim EP, He Q, Leung CK (2010) What do people want in microblogs? measuring interestingness of hashtags in twitter. In: Data Mining (ICDM), 2010 IEEE 10th International Conference on. pp. 1121–1126. doi:10.1109/ICDM.2010.34.

Yin Z, Cao L, Han J, Zhai C, Huang T (2011) Geographical topic discovery and comparison. In: Proceedings of the 20th international conference on World wide web. New York, NY, USA: ACM, WWW'11, pp. 247–256. doi:10.1145/1963405.1963443.

Pozdnoukhov A, Kaiser C (2011) Space-time dynamics of topics in streaming text. In: Proc. of the 3rd ACM SIGSPATIAL Int'l Workshop on Location-Based Social Networks. New York, NY, USA: ACM, LBSN'11, pp. 1–8. doi:10.1145/2063212.2063223.

Morstatter F, Pfeffer J, Liu H, Carley KM (2013) Is the sample good enough? comparing data from twitters streaming api with twitters firehose. In: International Conference on Weblogs and Social Media. pp. 400–408. Cheng Z,

Caverlee J, Lee K (2010) You Are Where You Tweet: A Content-Based Approach to Geo-locating Twitter Users. In: Proceedings of The 19th ACM International Conference on

Information and Knowledge Management. Toronto, Ontario, Canada: International Conference on Information and Knowledge Management, pp. 759–768. doi:10.1145/1871437.1871535.

Li R, Wang S, Deng H, Wang R, Chang KCC (2012) Towards social user profiling: unified and discriminative influence model for inferring home locations. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY, USA: ACM, KDD '12, pp. 1023–1031. doi:10.1145/2339530.2339692.

Bennett S (2011). Twitter: Faster than earthquakes. Media Bistro. Available: http://www.mediabistro.com/alltwitter/twitter-earthquake-video_b13147. Accessed 2014 Jul 4.

Mourtada R, Salem F (2011) Civil movements: The impact of facebook and twitter. Arab Social Media Report 1. Huang C (2011) Facebook and twitter key to arab spring uprisings: report. The National Abu Dhabi Media 6.

Campbell DG (2011) Egypt Unshackled: Using Social Media to @#:) the System. Amherst, NY: Cambria Books.

Berrett D (2011) Intellectual roots of wall st. protest lie in academe. The Chronicle of Higher Education. Available: http://chronicle.com/article/Intellectual-Roots-of-Wall/129428/. Accessed 2014 Jul 4.

Chappell B (2011). Occupy wall street: From a blog post to a movement. http://www.npr.org/2011/10/20/141530025/occupy-wall-street-from-a-blog-post-to-a-movement. Accessed 2014 Jul 4.

..(2011) Occupy wall street gets union support. United Press International. Available: http://www.upi.com/Top_News/US/2011/09/30/Occupy-Wall-Street-gets-union-support/UPI-89641317369600/. Accessed 2014 Jul 4.

Kumar S, Barbier G, Abbasi MA, Liu H (2011) Tweettracker: An analysis tool for humanitarian and disaster relief. In: Fifth International AAAI Conference on Weblogs and Social Media, ICWSM.

Swets JA (1996) Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers. Lawrence Erlbaum Associates Mahwah, NJ.

Rhoades SA (1993) The Herfindahl-Hirschman index. Fed Res Bull 79: 188.

Cover TM, Thomas JA (2012) Elements of information theory. Wiley-Interscience. 44.

McClelland CA (1961) The acute international crisis. World Politics 14: 182–204. doi: 10.2307/2009561

McClelland CA (1968). Access to berlin: the quantity and variety of events, 1948-1963. Available: http://www.econbiz.de/Record/access-to-berlin-the-quantity-and-variety-of-events-1948-1963-mcclelland-charles/10002418818. Accessed 2014 Jul 4.

Boydstun AE, Bevan DS, Thomas HF (2014) The importance of attention diversity and how to measure it. Public Policy and Administration 42(2): 173–196. doi: 10.1111/psj.12055

Date Added 9/27/2016, 2:27:42 PM

Modified 9/28/2016, 5:09:26 PM

2015

Real-time detection of traffic from Twitter stream analysis

Type	Journal Article
Author	Eleonora D'Andrea
Author	Pietro Ducange
Author	Beatrice Lazzerini
Author	Francesco Marcelloni
Rights	IEEE (Institute of Electrical and Electronics Engineers)
Volume	16
Issue	4
Pages	2269-2283
Publication	IEEE transactions on intelligent transportation systems
ISSN	1524-9050
Date	2015
DOI	10.1109/TITS.2015.2404431
Library Catalog	www.safetylit.org
Language	en

Abstract

Social networks have been recently employed as a source of information for event detection, with particular reference to road traffic congestion and car accidents. In this paper, we present a real-time monitoring system for traffic event detection from Twitter stream analysis. The system fetches tweets from Twitter according to several search criteria; processes tweets, by applying text mining techniques; and finally performs the classification of tweets. The aim is to assign the appropriate class label to each tweet, as related to a traffic event or not. The traffic detection system was employed for real-time monitoring of several areas of the Italian road network, allowing for detection of traffic events almost in real time, often before online traffic news web sites. We employed the support vector machine as a classification model, and we achieved an accuracy value of 95.75% by solving a binary classification problem (traffic versus non-traffic tweets). We were also able to discriminate if traffic is caused by an external event or not, by solving a multi-class classification problem and obtaining an accuracy value of 88.89%.

SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based.

References

1. F. Atefeh and W. Khreich, "A survey of techniques for event detection in Twitter", Comput. Intell., vol. 31, no. 1, pp. 132-164, 2015.

2. P. Ruchi and K. Kamalakar, "ET: Events from tweets", Proc. 22nd Int. Conf. World Wide Web Comput., pp. 613-620, 2013.

3. A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel and B. Bhattacharjee, "Measurement and analysis of online social networks", Proc. 7th ACM SIGCOMM Conf. Internet Meas., pp. 29-42, 2007.

4. G. Anastasi, "Urban and social sensing for sustainable mobility in smart cities", Proc. IFIP/IEEE Int. Conf. Sustainable Internet ICT Sustainability, pp. 1-4, 2013.

5. A. Rosi, "Social sensors and pervasive services: Approaches and perspectives", Proc. IEEE Int. Conf. PERCOM Workshops, pp. 525-530, 2011.

6. T. Sakaki, M. Okazaki and Y. Matsuo, "Tweet analysis for real-time event detection and earthquake reporting system development", IEEE Trans. Knowl. Data Eng., vol. 25, no. 4, pp. 919-931, Apr. 2013.

7. J. Allan, Topic Detection and Tracking: Event-Based Information Organization, 2002, Kluwer.

8. K. Perera and D. Dias, "An intelligent driver guidance tool using location based services", Proc. IEEE ICSDM, pp. 246-251, 2011.

9. T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, and K. Nawa, "Real-time event extraction for driving information from social sensors", Proc. IEEE Int. Conf. CYBER, pp. 221-226, 2012.

10. B. Chen and H. H. Cheng, "A review of the applications of agent technology in traffic and transportation systems", IEEE Trans. Intell. Transp. Syst., vol. 11, no. 2, pp. 485-497, Jun. 2010.

11. A. Gonzalez, L. M. Bergasa and J. J. Yebes, "Text detection and recognition on traffic panels from street-level imagery using visual appearance", IEEE Trans. Intell. Transp. Syst., vol. 15, no. 1, pp. 228-238, Feb. 2014.

12. N. Wanichayapong, W. Pruthipunyaskul, W. Pattara-Atikom and P. Chaovalit, "Social-based traffic information extraction and classification", Proc. 11th Int. Conf. ITST, pp. 107-112, 2011.

13. P. M. d'Orey and M. Ferreira, "ITS for sustainable mobility: A survey on applications and impact assessment tools", IEEE Trans. Intell. Transp. Syst., vol.
15, no. 2, pp. 477-493, Apr. 2014.

14. K. Boriboonsomsin, M. Barth, W. Zhu and A. Vu, "Eco-routing navigation system based on multisource historical and real-time traffic information", IEEE Trans. Intell. Transp. Syst., vol. 13, no. 4, pp. 1694-1704, Dec. 2012.

15. J. Hurlock and M. L. Wilson, "Searching twitter: Separating the tweet from the chaff", Proc. 5th AAAI ICWSM, pp. 161-168, 2011.

16. J. Weng and B.-S. Lee, "Event detection in Twitter", Proc. 5th AAAI ICWSM, pp. 401-408, 2011.

17. S. Weiss, N. Indurkhya, T. Zhang and F. Damerau, Text Mining: Predictive Methods for Analyzing Unstructured Information, 2004, Springer-Verlag.

18. A. Hotho, A. Nürnberger and G. Paaß, "A brief survey of text mining", LDV Forum-GLDV J. Comput. Linguistics Lang. Technol., vol. 20, no. 1, pp. 19-62, May 2005.

19. V. Gupta, S. Gurpreet and S. Lehal, "A survey of text mining techniques and applications", J. Emerging Technol. Web Intell., vol. 1, no. 1, pp. 60-76, Aug. 2009.

20. V. Ramanathan and T. Meyyappan, "Survey of text mining", Proc. Int. Conf. Technol. Bus. Manage., pp. 508-514, 2013.

21. M. W. Berry and M. Castellanos, Survey of Text Mining, 2004, Springer-Verlag.

22. H. Takemura and K. Tajima, "Tweet classification based on their lifetime duration", Proc. 21st ACM Int. CIKM, pp. 2367-2370, 2012.

23. No reference is included for item 23

24. A. Schulz, P. Ristoski and H. Paulheim, "I see a car crash: Real-time detection of small scale incidents in microblogs" in The Semantic Web: ESWC 2013 Satellite Events, vol. 7955, pp. 22-33, 2013, Springer-Verlag.

25. M. Krstajic, C. Rohrdantz, M. Hund and A. Weiler, "Getting there first: Real-time detection of real-world incidents on Twitter", Proc. 2nd IEEE Work Interactive Vis. Text Anal.—Task-Driven Anal. Soc. Media IEEE VisWeek, 2012.

26. C. Chew and G. Eysenbach, "Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak", PLoS ONE, vol. 5, no. 11, pp. 1-13, Nov. 2010.

27. B. De Longueville, R. S. Smith and G. Luraschi, "OMG from here I can see the flames!: A use case of mining location based social networks to acquire spatio-temporal data on forest fires", Proc. Int. Work. LBSN, pp. 73-80, 2009.

28. J. Yin, A. Lampert, M. Cameron, B. Robinson and R. Power, "Using social media to enhance emergency situation awareness", IEEE Intell. Syst., vol. 27, no. 6, pp. 52-59, Nov./Dec. 2012.

29. P. Agarwal, R. Vaithiyanathan, S. Sharma and G. Shro, "Catching the long-tail: Extracting local news events from Twitter", Proc. 6th AAAI ICWSM, pp. 379-382, Jun. 2012.

30. F. Abel, C. Hauff, G.-J. Houben, R. Stronkman and K. Tao, "Twitcident: fighting fire with information from social web streams", Proc. ACM 21st Int. Conf. Comp. WWW, pp. 305-308, 2012.

31. R. Li, K. H. Lei, R. Khadiwala and K. C.-C. Chang, "TEDAS: A Twitter-based event detection and analysis system", Proc. 28th IEEE ICDE, pp. 1273-1276, 2012.

32. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, "The WEKA data mining software: An update", SIGKDD Explor. Newsl., vol. 11, no. 1, pp. 10-18, Jun. 2009.

33. M. Habibi, Real World Regular Expressions with Java 1.4., 2004, Springer-Verlag.

34. Y. Zhou and Z.-W. Cao, "Research on the construction and filter method of stop-word list in text preprocessing", Proc. 4th ICICTA, vol. 1, pp. 217-221, 2011.

35. W. Francis and H. Kucera, "Frequency analysis of English usage: Lexicon and grammar", J. English Linguistics, vol. 18, no. 1, pp. 64-70, Apr. 1982.

36. M. F. Porter, "An algorithm for suffix stripping", Program: Electron. Library Inf. Syst., vol. 14, no. 3, pp. 130-137, 1980.

37. G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval", Inf. Process. Manage., vol. 24, no. 5, pp. 513-523, Aug. 1988.

38. L. M. Aiello, "Sensing trending topics in Twitter", IEEE Trans. Multimedia, vol. 15, no. 6, pp. 1268-1282, Oct. 2013.

39. C. Shang, M. Li, S. Feng, Q. Jiang and J. Fan, "Feature selection via maximizing global information gain for text classification", Knowl.-Based Syst., vol. 54, pp. 298-309, Dec. 2013.

40. L. H. Patil and M. Atique, "A novel feature selection based on information gain using WordNet", Proc. SAI Conf., pp. 625-629, 2013.

41. M. A. Hall and G. Holmes, "Benchmarking attribute selection techniques for discrete class data mining", IEEE Trans. Knowl. Data Eng., vol. 15, no. 6, pp. 1437-1447, Nov./Dec. 2003.

42. H. Uğuz, "A two-stage feature selection method for text categorization by using information gain principal component analysis and genetic algorithm", Knowl.-Based Syst., vol. 24, no. 7, pp.

1024-1032, Oct. 2011.

43. Y. Aphinyanaphongs, "A comprehensive empirical comparison of modern supervised classification and feature selection methods for text categorization", J. Assoc. Inf. Sci. Technol., vol. 65, no. 10, pp. 1964-1987, Oct. 2014.

44. J. Platt, B. Schoelkopf, C. J. C. Burges and A. J. Smola, "Fast training of support vector machines using sequential minimal optimization" in Advances in Kernel Methods: Support Vector Learning, pp. 185-208, 1999, MIT Press.

45. G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers", Proc. 11th Conf. Uncertainty Artif. Intell., pp. 338-345, 1995.

46. J. R. Quinlan, C4.5: Programs for Machine Learning, 1993, Morgan Kaufmann.

47. D. W. Aha, D. Kibler and M. K. Albert, "Instance-based learning algorithms", Mach. Learn., vol. 6, no. 1, pp. 37-66, Jan. 1991.

48. E. Frank and I. H. Witten, "Generating accurate rule sets without global optimization", Proc. 15th ICML, pp. 144-151, 1998.

49. C. Cortes and V. Vapnik, "Support-vector networks", Mach. Learn., vol. 20, no. 3, pp. 273-297, Sep. 1995.

50. T. T. Cover and P. E. Hart, "Nearest neighbour pattern classification", IEEE Trans. Inf. Theory, vol. IT-13, no. 1, pp. 21-27, Jan. 1967.

51. W. W. Cohen, "Fast effective rule induction", Proc. 12th ICML, pp. 115-123, 1995.

52. G. Pagallo and D. Haussler, "Boolean feature discovery in empirical learning", Mach. Learn., vol. 5, no. 1, pp. 71-99, Mar. 1990.

53. J. Derrac, S. Garcia, D. Molina and F. Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms", Swarm Evol. Comput., vol. 1, no. 1, pp. 3-18, Mar. 2011.

54. F. Wilcoxon, "Individual comparisons by ranking methods", Biometrics Bull., vol. 1, no. 6, pp. 80-83, Dec. 1945.

55. H. Becker, M. Naaman and L. Gravano, "Beyond trending topics: Real-world event identification on Twitter", Proc. 5th AAAI ICWSM, pp. 438-441, 2011.

56. H. Kwak, C. Lee, H. Park and S. Moon, "What is Twitter a social network or a news media?", Proc. ACM 19th Int. Conf. WWW, pp. 591-600, 2010.

Keywords: Twitter-Traffic-Status

Traffic detection using tweets on Twitter social network

Type	Journal Article
Author	Supriya Bhosale
Author	Sucheta Kokate
Rights	IJSR
Volume	4
Issue	12
Pages	2319-7064
Publication	International journal of science and research (Nagpur)
ISSN	2319-7064
Date	2015
Library Catalog	www.safetylit.org
Language	en

Abstract

Social networks can be employed as a source of information for event detection such as road traffic congestion and car accidents. Existing system present a real-time monitoring system for traffic event detection from twitter. The system fetches tweets from twitter and then; processes tweets using text mining techniques. Lastly performs the classification of tweets. The aim of the system is to assign the appropriate class label to each tweet, whether it is related to a traffic event or not. System employed the support vector machine as a classification model. The proposed system uses the system based on semi-supervised approach, which gives training using traffic related dataset. We propose a clustering approach for classification of the tweets in traffic related and non- traffic related tweets. We employ a Euclidean distance to calculate the similarity between the tweets. Keywords: Tweet classification, Traffic event detection, Data mining, text mining, and social sensing.

Available at: http://www.ijsr.net/archive/v4i12/NOV152452.pdf

Keywords: Twitter-Traffic-Status

SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based.

References: [1] Eleonora D'Andrea, Pietro Ducange, Beatrice Lazzerini, Member, IEEE, and Francesco Marcelloni, Member, IEEE ,Real-Time Detection of Traffic

[2] Twitter Stream Analysis, IEEE Transactions on intelligent transportation systems, 16(4), August 2015

[3] Rui LI, Kin Hou Lei, Ravi Khadiwala, Kevin Chen- Chuan Chang ―TEDAS: a Twitter Based Event Detection and Analysis System, IJCSIT 2014.

[4] Harshita Rajwani, Srushti Somvanshi, Anuja Upadhye,Dynamic Traffic Analyzer Using Twitter ―, International Journal of Science and Research (IJSR) 2014.

[5] Vikram Singh and Balwinder Saini ―AnEffective Tokenization Algorithm for Information Retrieval System‖ CS & IT-CSCP 2014

[6] Maximilian Walther and Michael Kaisser, Geo-spatial Event Detection in the Twitter Stream‖, P. Serdyukov et al. (Eds.): ECIR 2013, LNCS 7814, pp. 356367, 2013.cS pringer VerlagBerlinHeidelberg 2013.

[7] T. Sakaki, M. Okazaki, and Y.Matsuo, —Tweet analysis for real-time event detection and earthquake reporting system development, IEEE Trans. Knowl. Data Eng., vol. 25, no. 4, pp. 919–931, Apr. 2013.

[8] M. Krstajic, C. Rohrdantz, M. Hund, and A. Weiler, —Getting there first: Real-time detection of real-world incidents on Twitter‖ in Proc. 2nd IEEE Work Interactive Vis. Text Anal.—Task-Driven Anal. Soc. Media IEEE VisWeek, Seattle, WA, USA, 2012.

[9] A. Schulz, P. Ristoski, and H. Paulheim, —Isee a car crash: Real-time detection of small scale incidents in microblogs, in The Semantic Web: ESWC 2013 Satellite Events, vol. 7955. Berlin, Germany: Springer- Verlag, 2013, pp. 22–33.

[10]J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power, —Using social media to enhance emergency situation awareness, IEEE Intell. Syst., vol. 27, no. 6, pp. 52–59, Nov./Dec. 2012.

Steds: social media based transportation event detection with text summarization

Type	Journal Article
Author	Kaiqun Fu
Author	Chang-Tien Lu
Author	Rakesh Nune
Author	Jason Xianding Tao
Volume	ITSC-2015
Pages	1952-1957
Publication	IEEE Explore
ISSN	2153-0009
Date	2015
DOI	10.1109/ITSC.2015.316
Library Catalog	www.safetylit.org
Language	en

Abstract

Ubiquitous user-input contents on social media and online services have generated a tremendous amount of informa- tion. Such information has great potential applications in various areas such as events detection and text summarization. In this paper, a social media based traffic status monitoring system is established. The system is initiated by a transportation related keyword generation process. Then an association rules based iterative query expansion algorithm is applied to extract real time transportation related tweets for incident management purpose. We also confirm the feasibility of summarizing the redundant tweets to generate concise and comprehensible textual contents. Comparison results show that our query expansion method for tweets extraction outperforms the previous ones. Analysis and case studies further demonstrate the practical usefulness of our tweets summarization algorithm. Index Terms--component, social media analysis, data mining, intelligent transportation systems, tweets extraction. DOI: 10.1109/ITSC.2015.316 Published in: IEEE Explore - Proceedings of the Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference held 15-18 Sept. 2015, Gran Canaria, Spain.

ISBN 978-1-4673-6595-6

Keywords: Twitter-Traffic-Status

SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based.

References [1] Y. Xie, Z. Chen, Y. Cheng, K. Zhang, A. Agrawal, W.-K. Liao, and A. Choudhary, "Detecting and tracking disease outbreaks by mining social media data," in Proceedings of the Twenty-Third international jointconferenceonArtificialIntelligence. AAAIPress,2013,pp.2958– 2960.

[2] R. S. Jafarzadeh, "Emergency management 2.0 Integrating social media in emergency communications," Journal of Emergency Management, vol. 9, no. 4, pp. 15–18, 2011.

[3] A. Schulz, J. Ortmann, and F. Probst, "Getting user-generated content structured: Overcoming information overload in emergency management," in Global Humanitarian Technology Conference (GHTC), 2012 IEEE. IEEE, 2012, pp. 143–148.

[4] J. Cui, R. Fu, C. Dong, and Z. Zhang, "Extraction of traffic information from social media interactions: Methods and experiments," in Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on. IEEE, 2014, pp. 1549–1554.

[5] K. Fu, R. Nune, and J. X. Tao, "Social media data analysis for traffic incident detection and management," in Transportation Research Board 94th Annual Meeting. TRB, 2015.

[6] G. Erkan and D. R. Radev, "Lexrank: graph-based lexical centrality as salience in text summarization," Journal of Artificial Intelligence Research, pp. 457–479, 2004.

[7] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." 1999.

[8] K. Fu, Y.-C. Lu, and C.-T. Lu, "Treads: A safe route recommender using social media mining and text summarization," in Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, 2014, pp. 557–560.

[9] K. Makice, Twitter API: Up and running: Learn how to build applications with the Twitter API. " O'Reilly Media, Inc.", 2009.

[10] R. Agrawal, R. Srikant et al., "Fast algorithms for mining association rules," in Proc. 20th int. conf. very large data bases, VLDB, vol. 1215, 1994, pp. 487–499.

[11] I. Twitter. (2013) Rest apis — twitter developers. [Online]. Available: https://dev.twitter.com/rest/public<p /> <p>Language: en</p>

Mining complaints for traffic-jam estimation: a social sensor application

Type	Conference Paper
Author	Theodore Georgiou
Author	Amr El Abbadi
Author	Xifeng Yan
Author	Jemin George
Place	New York, NY, USA
Publisher	Association for Computing Machinery (ACM Press)
Pages	330-335
Date	2015
DOI	10.1145/2808797.2809404
Library Catalog	www.safetylit.org
Conference Name	IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015
Language	en
Abstract	Physical events in the real world are known to trigger reactions and then discussions in online social media. Mining these reactions through online social sensors offers a fast and low cost way to understand what is happening in the physical world. In some cases, however, further study of the affected population's emotional state can improve this understanding. In our study we analyzed how car commuters react on Twitter while stuck in heavy traffic. We discovered that the online social footprint does not necessarily follow a strict linear correlation with the volume of a traffic jam. Through our analysis we offer a potential explanation: people's mood could be an additional factor, apart from traffic severity itself, that leads in fluctuations of the observed reaction in social media. This finding can be important for social sensing applications where external factors, like sentiment, also contribute on how humans react. We propose a novel traffic-congestion estimation model that utilizes the volume of messages and complaints in online social media, based on when they happen. We show through experimental evaluation that the proposed model can estimate, with higher accuracy, traffic jam severity and compare the results with several baselines. The model achieves at least 38% improvement of absolute error and more than 45% improvement of relative error, when compared with a baseline that assumes linear correlation between traffic and social volume. To support our findings we combined data from the California Department of Transportation (CALTRANS) and Twitter, for a total of 6 months, and focused on a major traffic-heavy freeway in Los Angeles, California. DOI: 10.1145/2808797.2809404 SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based. References Note: OCR errors may be found in this Reference List extracted from the full text article. ACM has opted to expose the complete List rather than only correct and linked references. 1 D. Boyd and K. Crawford. Critical questions for big data. Information, Communication and Society, 15(5):662--679, 2012.

2 Caltrans performance measurement system (pems). http://pems.dot.ca.gov.

3 R. Cervero. Road expansion, urban growth, and induced travel: A path analysis. University of california transportation center, working papers, University of California Transportation Center, 2001.

4 C. Chen, A. Skabardonis, and P. Varaiya. Systematic identification of freeway bottlenecks. In Proceedings of 83rd Transportation Research Board Annual Meeting, 2004.

5 Text messaging law, california, 2008. https://www.dmv.ca.gov/cellularphonelaws.

6 Twaffic - will twitter and tweets about traffic change the way we drive? http://www.slate.com/articles/life/transport/2011/04/twaffic.html.

7 Elizabeth M. Daly , Freddy Lecue , Veli Bicer, Westland row why so slow?: fusing social media and linked data sources for understanding real-time traffic conditions, Proceedings of the 2013 international conference on Intelligent user interfaces, March 19-22, 2013, Santa Monica, California, USA [doi>10.1145/2449396.2449423]

8 T. Golob and W. Recker. Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions. Journal of Transportation Engineering, 129(4):342--353, 2003.

9 Jingrui He , Wei Shen , Phani Divakaruni , Laura Wynter , Rick Lawrence, Improving traffic prediction with tweet semantics, Proceedings of the Twenty-Third international joint conference on Artificial Intelligence, August 03-09, 2013, Beijing, China

10 Kai Nagel and Michael Schreckenberg. A cellular automaton model for freeway traffic. J. Phys. I France, 2(12):2221--2229, 1992.

11 W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg. Human behavior as origin of traffic phases. Phys. Rev. E, 65:015101, Dec 2001.

12 J. Kwon, B. McCullough, K. Petty, and P. Varaiya. Evaluation of PeMS to improve the congestion monitoring program. Technical report, Final Report for PATH TO 5319, UC Berkeley, Berkeley, CA, 2006.

13 Microsoft azure helps researchers predict traffic jams. http://blogs.msdn.com/b/msr_er/archive/2015/04/02/microsoft-azure-helps-researchers-predict-traffic-jams.aspx.

14 Anna Izabel João Tostes Ribeiro , Thiago Henrique Silva , Fátima Duarte-Figueiredo , Antonio A.F. Loureiro, Studying traffic conditions by analyzing foursquare and instagram data, Proceedings of the 11th ACM symposium on Performance evaluation of wireless ad hoc, sensor, & ubiquitous networks, September 21-26, 2014, Montreal, QC, Canada [doi>10.1145/2653481.2653491]

15 Sílvio S. Ribeiro, Jr. , Clodoveu A. Davis, Jr. , Diogo Rennó R. Oliveira , Wagner Meira, Jr. , Tatiana S. Gonçalves , Gisele L. Pappa, Traffic observatory: a system to detect and locate traffic events and conditions using Twitter, Proceedings of the 5th ACM SIGSPATIAL International Workshop on Location-Based Social Networks, November 06-06, 2012, Redondo Beach, California [doi>10.1145/2442796.2442800]

16 R. Singleton, B. Straits, and M. Straits. Approaches to social research. Oxford University Press, 1993.

17 Arvind Thiagarajan , Lenin Ravindranath , Katrina LaCurts , Samuel Madden , Hari Balakrishnan , Sivan Toledo , Jakob Eriksson, VTrack: accurate, energy-aware road traffic delay estimation using mobile phones, Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems, November 04-06, 2009, Berkeley, California [doi>10.1145/1644038.1644048]

18 Waze: Community-based traffic app. https://www.waze.com. 19 C. Winston. On the performance of the u.s. transportation system: Caution ahead. Journal of Economic Literature, 51(3):773--824, 2013.

Keywords: Twitter-Traffic-Status

Proceedings Title	Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015

Modeling the impacts of inclement weather on freeway traffic speed: exploratory study with social media data

Type	Journal Article
Author	Lei Lin
Author	Ming Ni
Author	Qing He
Author	Jing Gao
Author	Adel W. Sadek
Rights	Transportation Research Board, National Research Council, National Academy of Sciences USA
Volume	2482
Pages	82-89
Publication	Transportation research record
ISSN	0361-1981
Date	2015
DOI	10.3141/2482-11
Library Catalog	www.safetylit.org
Language	en
Abstract	Recently, there has been increased interest in quantifying and modeling the impact of inclement weather on transportation system performance. One problem that the majority of research studies on the topic have faced was the great dependence on weather data merely from atmospheric weather stations, which lack information about road surface condition. The emergence of social media platforms, such as Twitter and Facebook, provides a new opportunity to extract more weather-related data from such platforms. This study had two primary objectives: (a) examine whether real-world weather events can be inferred from social media data and (b) determine whether including weather variables extracted from social media data can improve the predictive accuracy of models developed to quantify the impact of inclement weather on freeway traffic speed. To achieve those objectives, weather data, Twitter data, and traffic information were compiled for the Buffalo-Niagara, New York, metropolitan area as a case study. A method called the Twitter Weather Events Observation was then applied to the Twitter data, and the sensitivity and false alarm rate for the method was evaluated against real-world weather data. Then, linear regression models for predicting the impact of inclement weather on freeway speed were developed with and without the Twitter-based weather variables incorporated. The results indicated that Twitter data have a relatively high sensitivity for predicting inclement weather (i.e., snow), especially during the daytime and for areas with significant snowfall. The results also showed that the incorporation of Twitter-based weather variables could help improve the predictive accuracy of the models. SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based. References Road Management Program. How Do Weather Events Impact Roads? FHWA. http://www.ops.fhwa.dot.gov/weather/q1_roadimpact.htm. Accessed July 16, 2014.

Bartlett, A., W. Lao, Y. Zhao, and A. W. Sadek. Impact of Inclement Weather on Hourly Traffic Volumes in Buffalo, New York. Presented at 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013.

Stern, A. D., V. Shah, L. Goodwin, and P. Pisano. Analysis of Weather Impacts on Traffic Flow in Metropolitan Washington, DC. Presented at 19th International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology, American Meteorological Society, Long Beach, Calif., Feb. 12, 2003.

Vlahogianni, E. I., and M. G. Karlaftis. Comparing Traffic Flow Time-Series under Fine and Adverse Weather Conditions Using Recurrence-Based Complexity Measures. Nonlinear Dynamics, Vol. 69, No. 4, 2012, pp. 1949–1963.

Zhao, Y., A. W. Sadek, and D. Fuglewicz. Modeling the Impact of Inclement Weather on Freeway Traffic Speed at Macroscopic and Microscopic Levels. In Transportation Research Record: Journal of the Transportation Research Board, No. 2272, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 173–180.

Rakha, H., M. Arafeh, and S. Park. Modeling Inclement Weather Impacts on Traffic Stream Behavior. International Journal of Transportation Science and Technology, Vol. 1, No. 1, 2012, pp. 25–48.

Omer, R., and L. Fu. An Automatic Image Recognition System for Winter Road Surface Condition Classification. Proc. of 13th International IEEE Conference on Intelligent Transportation Systems, Institute of Electrical and Electronics Engineers, 2010, pp. 1375–1379.

Jonsson, P. Remote Sensor for Winter Road Surface Status Detection. Sensors 2011, IEEE, 2011, pp. 1285–1288.

Hyvärinen, O., and E. Saltikoff. Social Media as a Source of Meteorological Observations. Monthly Weather Review, Vol. 138, No. 8, 2010, pp. 3175–3184.

Cox, J., and B. Plale. Improving Automatic Weather Observations with the Public Twitter Stream. Indiana University School of Informatics and Computing, 2011.

Ni, M., Q. He, and J. Gao. Using Social Media to Predict Traffic Flow Under Special Event Conditions. Presented at 93rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2014.

Grant-Muller, S. M., A. Gal-Tzur, E. Minkov, S. Nocera, T. Kuflik, and I. Shoor. The Efficacy of Mining Social Media Data for Transport Policy and Practice. Presented at 93rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2014.

Twitter Statistics. Statistic Brain Research Institute. http://www.statisticbrain.com/twitter-statistics. Accessed July 18, 2014.

Maze, T. H., M. Agarwal, and G. Burchett. Whether Weather Matters to Traffic Demand, Traffic Safety, and Traffic Operations and Flow. In Transportation Research Record: Journal of the Transportation Research Board, No. 1948, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 170–176.

Historical Weather. Weather Underground, Inc., San Francisco, Calif. http://www.wunderground.com/history. Accessed May 15, 2014.

Mass, C., and C. F. Mass. Nowcasting: The Next Revolution in Weather Prediction. Bulletin of the American Meteorological Society, Vol. 93, No. 6, 2012, pp. 797–809.

The Streaming APIs. Twitter, Inc., San Francisco, Calif. https://dev.twitter.com/docs/api/streaming. Accessed May 15, 2014.

Lin, L., Q. Wang, S. Huang, and A. W. Sadek. On-Line Prediction of Border Crossing Traffic Using an Enhanced Spinning Network Method. Transportation Research Part C: Emerging Technologies, Vol. 43, Part 1, 2014, pp. 158–173.

Lin, L., A. W. Sadek, and Q. Wang. Multiple-Model Combined Forecasting Method for Online Prediction of Border Crossing Traffic at Peace Bridge. Presented at 91st Annual Meeting of the Transportation Research Board, Washington, D.C., 2012. Keywords: Twitter-Traffic-Status<p />

The movement in real-time analysis to find the twitter stream

Type	Journal Article
Author	K. Jaya Shree
Author	Ch. Prudvini
Rights	Yuva Engineers
Volume	2
Issue	12
Pages	263-265
Publication	International journal and magazine of engineering, technology, management and research
ISSN	2320-3706
Date	2015
Library Catalog	www.safetylit.org
Language	en

Abstract

Social network congestion and vehicle road traffic accident with reference to the source of information for the detection of the event, as has been used in recent years. In this paper, we analyze the traffic situation and to identify the Twitter stream to provide real-time monitoring system. The system according to different search criteria, brings tweets from Twitter. Tweets are text-mining operations through the application of techniques, and ultimately lead Rating tweets. The traf c incident or not, appropriate class tag is set for each Tweet. The traffic on the Internet news sites, often in different areas of the Italian road network traffic in real-time monitoring system for detection work, and to allow for the real-time traffic incident detection. We support vector machine is used as a model site, and we are the site of the binary problem (non-commercial traffic tweets v) solution has achieved 95.75% accuracy standards. We have the distinction of traffic through an external event or not, multiclass classification problems and were able to get the value of precision.

This article may be viewed at:
http://www.ijmetmr.com/oldecember2015/KJayaShree-ChPrudvini-47.pdf

SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based.

References

[1] F. Atefeh and W. Khreich, "A survey of techniques for event detection in Twitter," Comput. Intell., vol. 31, no. 1, pp. 132–164, 2015.

[2] P. Ruchi and K. Kamalakar, "ET: Events from tweets," in Proc. 22nd Int. Conf. World Wide Web Comput., Rio de Janeiro, Brazil, 2013, pp. 613–620.

[3] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of on- line social networks," in Proc. 7th ACM SIGCOMM Conf. Internet Meas., San Diego, CA, USA, 2007, pp. 29–42.

[4] G. Anastasi et al., "Urban and social sensing for sus- tainable mobility in smart cities," in Proc. IFIP/IEEE Int. Conf. Sustainable Internet ICTSustainability, Palermo, Italy, 2013, pp. 1–4.

[5] A. Rosi et al., "Social sensors and pervasive services: Approaches and perspectives," in Proc. IEEE Int. Conf. PERCOM Workshops, Seattle, WA, USA, 2011, pp. 525– 530.

[6] T. Sakaki, M. Okazaki, and Y.Matsuo, "Tweet analy- sis for real-time event detection and earthquake reporting system development," IEEE Trans. Knowl. Data Eng., vol. 25, no. 4, pp. 919–931, Apr. 2013.

[7] J. Allan, Topic Detection and Tracking: Event-Based Information Organization. Norwell, MA, USA: Kluwer, 2002.

[8] K. Perera and D. Dias, "An intelligent driver guid- ance tool using location based services," in Proc. IEEE ICSDM, Fuzhou, China, 2011, pp. 246–251.

[9] T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, and K. Nawa, "Real-time event extraction for driving in- formation from social sensors," in Proc. IEEE Int. Conf. CYBER, Bangkok, Thailand, 2012, pp. 221–226.

[10] B. Chen and H. H. Cheng, "A review of the appli- cations of agent technology in traf c and transportation systems," IEEE Trans. Intell. Transp. Syst., vol. 11, no. 2, pp. 485–497, Jun. 2010.

[11] A. Gonzalez, L. M. Bergasa, and J. J. Yebes, "Text de- tection and recognition on traf c panels from street-level imagery using visual appearance," IEEE Trans. Intell. Transp. Syst., vol. 15, no. 1, pp. 228–238, Feb. 2014.

[12] N. Wanichayapong, W. Pruthipunyaskul, W. Pattara- Atikom, and P. Chaovalit, "Social-based traf c informa- tion extraction and classi cation," in Proc. 11th Int. Conf. ITST, St. Petersburg, Russia, 2011, pp. 107–112.

[13] P. M. d'Orey and M. Ferreira, "ITS for sustainable mobility: A survey on applications and impact assessment tools," IEEE Trans. Intell. Transp. Syst., vol. 15, no. 2, pp. 477–493, Apr. 2014.

[14] K. Boriboonsomsin, M. Barth, W. Zhu, and A. Vu, "Eco-routing navigation system based on multisource historical and real-time traf c information," IEEE Trans. Intell. Transp. Syst., vol. 13, no. 4, pp. 1694–1704, Dec. 2012.

[15] J. Hurlock and M. L. Wilson, "Searching twitter: Separating the tweet from the chaff," in Proc. 5th AAAI ICWSM, Barcelona, Spain, 2011, pp. 161–168.

Keywords: Twitter-Traffic-Status

Citywide traffic congestion estimation with social media

Type	Conference Paper
Author	Senzhang Wang
Author	Lifang He
Author	Leon Stenneth
Author	Philip S. Yu
Author	Zhoujun Li
URL	http://www.safetylit.org/citations/index.php?fuseaction=citations.viewdetails&citationIds[]=citconferenceproceeding_390_34
Place	New York, NY, USA
Publisher	ACM Press
Pages	e34
Date	2015
DOI	10.1145/2820783.2820829
Accessed	9/27/2016, 2:27:42 PM
Library Catalog	www.safetylit.org
Conference Name	23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems
Language	en
Abstract	Conventional traffic congestion estimation approaches require the deployment of traffic sensors or large-scale probe vehicles. The high cost of deploying and maintaining these equipments largely limits their spatial-temporal coverage. This paper proposes an alternative solution with lower cost and wider spatial coverage by exploring traffic related information from Twitter. By regarding each Twitter user as a traffic monitoring sensor, various real-time traffic information can be collected freely from each corner of the city. However, there are two major challenges for this problem. Firstly, the congestion related information extracted directly from real-time tweets are very sparse due both to the low resolution of geographic location mentioned in the tweets and the inherent sparsity nature of Twitter data. Secondly, the traffic event information coming from Twitter can be multi-typed including congestion, accident, road construction, etc. It is non-trivial to model the potential impacts of diverse traffic events on traffic congestion. We propose to enrich the sparse real-time tweets from two directions: 1) mining the spatial and temporal correlations of the road segments in congestion from historical data, and 2) applying auxiliary information including social events and road features for help. We finally propose a coupled matrix and tensor factorization model to effectively integrate rich information for Citywide Traffic Congestion Eestimation (CTCE). Extensive evaluations on Twitter data and 500 million public passenger buses GPS data on nearly 700 mile roads of Chicago demonstrate the efficiency and effectiveness of the proposed approach. DOI: 10.1145/2820783.2820829 SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based.

References

Note: OCR errors may be found in this Reference List extracted from the full text article. ACM has opted to expose the complete List rather than only correct and linked references.

Rakesh Agrawal , Ramakrishnan Srikant, Fast Algorithms for Mining Association Rules in Large Databases, Proceedings of the 20th International Conference on Very Large Data Bases, p.487-499, September 12-15, 1994

Huiping Cao , Nikos Mamoulis , David W. Cheung, Mining Frequent Spatio-Temporal Sequential Patterns, Proceedings of the Fifth IEEE International Conference on Data Mining, p.82-89, November 27-30, 2005 [doi 10.1109/ICDM.2005.95]

Po-Ta Chen , Feng Chen , Zhen Qian, Road Traffic Congestion Monitoring in Social Media with Hinge-Loss Markov Random Fields, Proceedings of the 2014 IEEE International Conference on Data Mining, p.80-89, December 14-17, 2014 [doi 10.1109/ICDM.2014.139]

Elizabeth M. Daly , Freddy Lecue , Veli Bicer, Westland row why so slow?: fusing social media and linked data sources for understanding real-time traffic conditions, Proceedings of the 2013 international conference on Intelligent user interfaces, March 19-22, 2013, Santa Monica, California, USA [doi 10.1145/2449396.2449423]

E. D'Andrea, P. Ducange, B. Lazzerini, and F. Marcelloni. Real-time detection of traffic from twitter stream analysis. Intelligent Transportation Systems, PP(99):1--15, 2015.

C. de Fabritiis, R. Ragona, and G. Valenti. Traffic estimation and prediction based on real time floating car data. In ITCS, 2008.

S. K. Endarnoto, S. Pradipta, A. S. Nugroho, and J. Purnama. Traffic condition information extraction and visualization from social media twitter for android mobile application. In ICEEI, 2011.

D. Helbing. Traffic and related self-driven many-particle systems. Reviews of modern physics, 2001.

Alexandros Karatzoglou , Xavier Amatriain , Linas Baltrunas , Nuria Oliver, Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering, Proceedings of the fourth ACM conference on Recommender systems, September 26-30, 2010, Barcelona, Spain [doi 10.1145/1864708.1864727]

Meiling Liu , Kaiqun Fu , Chang-Tien Lu , Guangsheng Chen , Huiqiang Wang, A search and summary application for traffic events detection based on Twitter data, Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, November 04-07, 2014, Dallas, Texas [doi 10.1145/2666310.2666366]

L. Muñoz, X. Sun, R. Horowitz, and L. Alvarez. Traffic density estimation with the cell transmission model. In the 2003 American Control Conference, 2003.

C. Ozkurt and F. Camci. Automatic traffic density estimation and vehicle classification for traffic surveillance systems using neural networks. Mathematical and Computational Application, 14(3):187--196, 2009.

W. Pattara-Atikom, P. Pongpaibool, and S. Thajchayapong. Estimating road traffic congestion using vehicle velocity. In ITS Telecommunications, 2006.

F. Porikli and X. Li. Traffic congestion estimation using hmm models without vehicle tracking. In Intelligent Vehicles Symposium, 2004.

Badrul Sarwar , George Karypis , Joseph Konstan , John Riedl, Item-based collaborative filtering recommendation algorithms, Proceedings of the 10th international conference on World Wide Web, p.285-295, May 01-05, 2001, Hong Kong, Hong Kong [doi 10.1145/371920.372071]

D. Schrank, B. Eisele, and T. Lomax. 2012 URBAN MOBILITY REPORT Powered by INRIX Traffic Data. 2012.

A. Schulz, P. Ristoski, and H. Paulheim. I see a car crash: Real-time detection of small scale incidents in microblogs. In ESWC, 2013.

Jingbo Shang , Yu Zheng , Wenzhu Tong , Eric Chang , Yong Yu, Inferring gas consumption and pollution emission of vehicles throughout a city, Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, August 24-27, 2014, New York, New York, USA [do 10.1145/2623330.2623653]

Sílvio S. Ribeiro, Jr. , Clodoveu A. Davis, Jr. , Diogo Rennó R. Oliveira , Wagner Meira, Jr. , Tatiana S. Gonçalves , Gisele L. Pappa, Traffic observatory: a system to detect and locate traffic events and conditions using Twitter, Proceedings of the 5th ACM SIGSPATIAL International Workshop on Location-Based Social Networks, November 06-06, 2012, Redondo Beach, California [doi 10.1145/2442796.2442800]

S. Tao, V. Manolopoulos, S. Rodriguez, and A. Rusu. Real-time urban traffic state estimation with a-gps mobile phones as probes. Journal of Transportation Technologies, 2(1):22--31, 2012.

Ilias Tsoukatos , Dimitrios Gunopulos, Efficient Mining of Spatiotemporal Patterns, Proceedings of the 7th International Symposium on Advances in Spatial and Temporal Databases, p.425-442, July 12-15, 2001

Y. Wang, Y. Zhu, and Z. He. Challenges and opportunities in exploiting large-scale gps probe data. In Technical Report, HPL-2011-109, 2011.

Jing Yuan , Yu Zheng , Chengyang Zhang , Wenlei Xie , Xing Xie , Guangzhong Sun , Yan Huang, T-drive: driving directions based on taxi trajectories, Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, November 02-05, 2010, San Jose, California [doi 10.1145/1869790.1869807]

Yu Zheng , Licia Capra , Ouri Wolfson , Hai Yang, Urban Computing: Concepts, Methodologies, and Applications, ACM Transactions on Intelligent Systems and Technology (TIST), v.5 n.3, p.1-55, September 2014 [doi 10.1145/2629592]

Yu Zheng , Tong Liu , Yilun Wang , Yanmin Zhu , Yanchi Liu , Eric Chang, Diagnosing New York city's noises with ubiquitous data, Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, September 13-17, 2014, Seattle, Washington [doi 10.1145/2632048.2632102]

Proceedings Title Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems

2016

Estimation of traffic with accuracy through Twitter stream analysis

Type	Journal Article
Author	Kunigiri Eswar Kumar
Author	H. Ateeq Ahmed
Rights	IJI Tech
Volume	4
Issue	8
Pages	1317-1324
Publication	International journal of innovative technologies
ISSN	2321-8665
Date	2016
Library Catalog	www.safetylit.org
Language	en
Abstract	Internet sites are source of info for event detection, with specific mention of the road traffic activity blockage and accidents or earth-quack sensing system. In this paper, we present a real-time monitoring system intended for traffic occasion detection coming from Twitter stream analysis. The system fetches tweets coming from Twitter as per a several search criteria; methods tweets, by applying textual content mining methods; last but not least works the classification of twitter posts. The goal is to assign suitable class packaging to every tweet, because related with an activity of traffic event or perhaps not. The traffic recognition system or framework was utilized for real-time monitoring of various areas of the street network, taking into account detection of traffic occasions just almost in actual time, regularly before on-line traffic news sites. All of us employed the support vector machine like a classification unit; furthermore, we accomplished a great accuracy value of ninety five. 75% by attempting a binary classification issue. All of us were also capable to discriminate if traffic is triggered by an external celebration or not, by resolving a multiclass classification issue and obtaining accuracy worth of 88. 89%. Keywords: Traffic Event Detection, Tweet Classification, Text Mining, Twitter-Traffic-Status; Social Sensing. Available at: ijitech.org/uploads/543216IJIT9948-234.pdf SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based. References [1] Eleonora D'Andrea, Pietro Ducange, Beatrice Lazzerini, Member, IEEE, and Francesco Marcelloni, Member, IEEE, "Real-Time Detection of Traffic From Twitter Stream Analysis", IEEE Transactions on Intelligent Transportation Systems, Vol. 16, No. 4, August 2015. [2] F. Atefeh and W. Khreich, "A survey of techniques for event detection in Twitter," Comput. Intell., vol. 31, no. 1, pp. 132–164, 2015. [3] P. Ruchi and K. Kamalakar, "ET: Events from tweets," in Proc. 22nd Int. Conf. World Wide Web Comput., Rio de Janeiro, Brazil, 2013, pp. 613–620.

[4] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, “Measurement and analysis of online social networks,” in Proc. 7th ACM SIGCOMM Conf. Internet Meas., San Diego, CA, USA, 2007, pp. 29–42.

[5] G. Anastasi et al., “Urban and social sensing for sustainable mobility in smart cities,” in Proc. IFIP/IEEE Int. Conf. Sustainable Internet ICT Sustainability, Palermo, Italy, 2013, pp. 1–4.

[6] A. Rosi et al., “Social sensors and pervasive services: Approaches and perspectives,” in Proc. IEEE Int. Conf. PERCOM Workshops, Seattle, WA, USA, 2011, pp. 525–530.

[7] T. Sakaki, M. Okazaki, and Y.Matsuo, “Tweet analysis for real-time event detection and earthquake reporting system development,” IEEE Trans. Knowl. Data Eng., vol. 25, no. 4, pp. 919–931, Apr. 2013.

[8] J. Allan, Topic Detection and Tracking: Event-Based Information Organization. Norwell, MA, USA: Kluwer, 2002.

[9] K. Perera and D. Dias, “An intelligent driver guidance tool using location based services,” in Proc. IEEE ICSDM, Fuzhou, China, 2011, pp. 246–251.

[10] T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, and K. Nawa, “Real-time event extraction for driving information from social sensors,” in Proc. IEEE Int. Conf. CYBER, Bangkok, Thailand, 2012, pp. 221–226.

[11] B. Chen and H. H. Cheng, “A review of the applications of agent technology in traffic and transportation systems,” IEEE Trans. Intell. Transp. Syst., vol. 11, no. 2, pp. 485–497, Jun. 2010.

[12] A. Gonzalez, L. M. Bergasa, and J. J. Yebes, “Text detection and recognition on traffic panels from street-level imagery using visual appearance,” IEEE Trans. Intell. Transp. Syst., vol. 15, no. 1, pp. 228–238, Feb. 2014.

Monitoring and analysis of real time detection of traffic from Twitter stream analysis

Type	Journal Article
Author	Anusha Jalaparthi
Author	A. Suraj Kumar
Rights	International Scientific Research Organization for Science, Engineering and Technology
Volume	4
Issue	3
Pages	33-36
Publication	International journal of scientific research in computer science and engineering
ISSN	2320-7639
Date	2016
Library Catalog	www.safetylit.org
Language	en

Abstract Social networks have been, at a recent time utilized as a source of information for event detection, with particular referral to road traffic congestion and for the car accidents. In this paper, we present a real-time monitoring system for traffic event detection from the twitter stream analysis. The system fetches tweets from the twitter according to several search criteria; processes tweets, by applying the text mining techniques; and finally performs the classification of tweets. The aim is to assign the suitable class label to each tweet, as relevant to a traffic event or not. The traffic detection system was exploited for real-time monitoring of several areas of the Italian road network, allowing for detection of traffic events almost in real time, generally before online traffic news web sites. We employed the support vector machine as a classification model, and we attain an accuracy value of 95.75% by solving a binary classification problem (traffic versus non-traffic tweets). We would also able to discriminate if traffic is caused by an external event or not, by solving a multiclass classification issue and obtaining accuracy value of 88.89%.

Keywords- Twitter; Traffic event detection; tweet classification; text mining, social sensing; Twitter-Traffic-Status

This article may be viewed at: http://www.isroset.org/pub_paper/IJSRCSE/9-ISROSET-IJSRCSE-2018-2.pdf

SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based.

References

[1]. F. Atefeh and W. Khreich, "A survey of techniques for event detection in Twitter," Comput. Intell., vol. 31, no. 1, pp. 132– 164, 2015

[2]. P. Ruchi and K. Kamalakar, "ET: Events from tweets," in Proc. 22nd Int. Conf. World Wide Web Comput., Rio de Janeiro, Brazil, 2013, pp. 613–620.

[3]. A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in Proc. 7th ACM SIGCOMM Conf.

Internet Meas., San Diego, CA, USA, 2007, pp. 29–42.

[4]. The Smarty project. [Online]. Available: http://www.smarty.toscana.it/

[5]. T. Sakaki, M. Okazaki, and Y.Matsuo, "Tweet analysis for real- time event detection and earthquake reporting system development," IEEE Trans. Knowl. Data Eng., vol. 25, no. 4, pp. 919–931, Apr. 2013.

[6]. M. Krstajic, C. Rohrdantz, M. Hund, and A. Weiler, "Getting there first: Real-time detection of real-world incidents on Twitter" in Proc. 2nd IEEE Work Interactive Vis. Text Anal.— Task-Driven Anal. Soc. Media IEEE VisWeek," Seattle, WA, USA, 2012.

[7]. J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power, "Using social media to enhance emergency situation awareness," IEEE Intell. Syst., vol. 27, no. 6, pp. 52–59, Nov./Dec. 2012.

[8]. T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, and K. Nawa, "Real-time event extraction for driving information from social sensors," in Proc. IEEE Int. Conf. CYBER, Bangkok, Thailand, 2012,pp. 221– 226.

[9]. N. Wanichayapong, W. Pruthipunyaskul, W. Pattara-Atikom, and P. Chaovalit, "Social-based traffic information extraction and classification," in Proc. 11th Int. Conf. ITST, St. Petersburg, Russia, 2011, pp. 107– 112.

[10]. A. Schulz, P. Ristoski, and H. Paulheim, "I see a car crash: Real- time detection of small scale incidents in microblogs," in The Semantic Web: ESWC 2013 Satellite Events, vol. 7955. Berlin, Germany: SpringerVerlag, 2013, pp. 22–33.

[11]. P . Agarwal, R. V aithiyanathan, S. Sharma, and G. Shro, "Catching the long-tail: Extracting local news events from Twitter," in Proc. 6th AAAI ICWSM, Dublin, Ireland, Jun. 2012, pp. 379–382.

Real-time detection of traffic from Twitter stream analysis

Type	Journal Article
Author	Sweety Kumari
Author	Firdos Khan
Author	Shekh Sultan
Author	Ruchita Khandge
Rights	Fast Track Publications
Volume	3
Issue	4
Pages	2350-2354
Publication	International research journal of engineering and technology
ISSN	2395-0072
Date	2016
Library Catalog	www.safetylit.org
Language	en

Abstract

Internet sites are source of info for event detection, with specific mention of the road traffic activity blockage and accidents or earth-quack sensing system. In this paper, we present a real-time monitoring system intended for traffic occasion detection coming from Twitter stream analysis. The system fetches tweets coming from Twitter as per a several search criteria; methods tweets, by applying textual content mining methods; last but not least works the classification of twitter posts. The goal is to assign suitable class packaging to every tweet, because related with an activity of traffic event or perhaps not. The traffic recognition system or framework was utilized for real- time monitoring of various areas of the street network, taking into account detection of traffic occasions just almost in actual time, regularly before on-line traffic news sites. All of us employed the support vector machine like a classification unit, furthermore, we accomplished a great accuracy value of ninety five. 75% by attempting a binary classification issue. All of us were also capable to discriminate if traffic is triggered by an external celebration or not, by resolving a multiclass classification issue and obtaining accuracy worth of 88. 89%.

Key Words: Twitter-Traffic-Status ; Twitter, Social media; Traffic detection; Text mining; Privacy; Service Oriented Architecture (SOA), machine learning, Twitter stream analysis, Traffic event detection, tweet classification, text mining, social sensing.

SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based.

References

Takeshi Sakaki, Makoto Okazaki, Yutaka Matsuo, "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors," IEEE Trans. Services Comput., vol. 5, no. 4, pp. 564–577, Fourth Quarter 2012.

AminaMadani, Omar Boussaid, Djamel Eddine Zegour and *Algiers, Algeria [sic]*, "What's Happening: A Survey of Tweets Event Detection," J. Internet Services Appl., vol. 1, no. 1, pp. 7–18, 2010.

Parikh, Kamalakar Karlapalem., "ET: Events from Tweets," Future Generat. Comput. Syst., vol. 25, no. 6, pp. 599–616, Jun. 2009.

Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, Bobby Bhattacharjee., "Measurement and Analysis of Online Social Networks," Proc. IEEE, vol. 99, no. 1, pp. 149–167, Jan. 2011. B.

Chen and H. H. Cheng, "A review of the applications of agent technology in traffic and transportation systems," IEEE Trans. Intell. Transp. Syst., vol. 11, no. 2, pp. 485–497, Jun. 2010.

T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, and K. Nawa, "Real-time event extraction for driving Programming, Bangalore, India, Jan 2010, pp. 147-158.

M. Yuriyama and T. Kushida, "Sensor-cloud infrastructure Physical sensor management with virtualized sensors on cloud computing," in Proc. 13th Int. Conf. Netw.,-Based Inf. Syst., Sep. 2010, pp. 1–8

Chen and Harry H., "A Review of the Applications of Agent Technology in Traffic and Transportation Systems," Wireless Commun. Mobile Comput, vol. 14, no. 1, pp. 19–36, Jan. 2014.

Álvaro González, Luis M. Bergasa and J. Javier Yebes, "Chen and Harry H.," ACM Trans. Sensor Netw., vol. 5, no. 2, Mar. 2009, Art. ID 10. Keywords:<p /> <p>Language: en</p>

Twitter stream analysis for traffic detection and earthquake

Type	Journal Article
Author	Pavan S. Sonune
Author	Sameer K. Shaikh
Author	Rushikesh L. Sonavame
Author	R. S. Shishupal
Rights	Amoghsiddhi Education Society
Volume	4
Issue	1
Pages	705-709
Publication	Journal of information, knowledge, and research in computer engineering
ISSN	0975-6760
Date	2016
Library Catalog	www.safetylit.org
Language	en

Abstract

Social networks have been recently empployed as a source of information for event detection, with specific reference to road traffic activity congestion and accidents or earthquack reporting system. In our paper, we present a real-time monitoring system for traffic occasion detection from Twitter stream analysis. The system fetches tweets from Twitter as per a several search criteria; procedures tweets, by applying text mining methods; lastly performs the classification of tweets. The aim is to assign suitable class label to every tweet, as related with an activity of traffic event or not. The traffic detection system or framework was utilized for real-time monitoring of several areas of the Italian street network, taking into consideration detection of traffic events just almost in real time, regularly before online traffic news sites. We employed the support vector machine as a classificatio model, furthermore, we accomplished an accuracy value of 95.75% by tackling a binar classification issue (traffic versus nontraffic tweets). We were also able to discriminate if traffic is caused by an external event or not, by solving a multiclass classification problem and obtaining accuracy value of 88.89%.

SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based.

References

[1] F. Atefeh and W. Khreich, "A survey of techniques for event detection in Twitter," Comput. Intell., vol. 31,no. 1, pp. 132–164, 2015.
[2] P. Ruchi and K. Kamalakar, "ET: Events from tweets," in Proc. 22nd Int. Conf. World Wide Web Comput., Rio de Janeiro, Brazil, 2013, pp. 613–620.

[3] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in Proc. 7th ACM SIGCOMM Conf. Internet Meas., San Diego, CA, USA, 2007, pp. 29–42.

[4] The Smarty project. [Online]. Available: http://www.smarty.toscana.it/

[5] T. Sakaki, M. Okazaki, and Y.Matsuo, "Tweet analysis for real-time event detection and earthquake reporting system development," IEEE Trans. Knowl. Data Eng., vol. 25, no. 4, pp. 919–931, Apr. 2013.

[6] M. Krstajic, C. Rohrdantz, M. Hund, and A. Weiler, "Getting there first: Real-time detection of real-world incidents on Twitter" in Proc. 2nd IEEE Work Interactive Vis. Text Anal.—Task-Driven Anal. Soc. Media IEEE VisWeek," Seattle, WA, USA, 2012.

[7] J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power, "Using social media to enhance emergency situation awareness," IEEE Intell. Syst., vol. 27, no. 6, pp. 52–59, Nov./Dec. 2012.

[8] T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, and K. Nawa, "Real-time event extraction for driving information from social sensors," in Proc. IEEE Int. Conf. CYBER, Bangkok, Thailand, 2012,pp. 221– 226.

[9] N. Wanichayapong, W. Pruthipunyaskul, W. Pattara-Atikom, and P. Chaovalit, "Social-based traffic information extraction and classification," in Proc. 11th Int. Conf. ITST, St. Petersburg, Russia, 2011, pp. 107– 112.

[10] A. Schulz, P. Ristoski, and H. Paulheim, "I see a car crash: Real-time detection of small scale incidents in microblogs," in The Semantic Web: ESWC 2013 Satellite Events, vol. 7955. Berlin, Germany: Springer- Verlag, 2013, pp. 22–33.

[11] P. Agarwal, R. Vaithiyanathan, S. Sharma, and G. Shro, "Catching the long-tail: Extracting local news events from Twitter," in Proc. 6th AAAI ICWSM, Dublin, Ireland, Jun. 2012, pp. 379–382.

Utilising location based social media in travel survey methods: bringing Twitter data into the play

Type	Conference Paper
Author	Alireza Abbasi
Author	Taha Hossein Rashidi
Author	Mojtaba Maghrebi
Author	S. Travis Waller
Place	New York, NY, USA
Publisher	ACM Press
Pages	e1
ISBN	978-1-4503-3975-9
Date	2016
DOI	10.1145/2830657.2830660
Library Catalog	www.safetylit.org
Conference Name	8th ACM SIGSPATIAL International Workshop on Location-Based Social Networks; Seattle, WA, USA; NOvember 3-6, 2015
Language	en

Abstract

A growing body of literature has been devoted to harnessing the crowd-sourcing power of social media by extracting knowledge from the huge amounts of information available online. This paper discusses how social media data can be used indirectly and with minimal cost to extract travel attributes such as trip purpose and activity location. As a result, the capacity of Twitter data in complementing other sources of transport related data such as household travel surveys or traffic count data is examined. Further, a detailed discussion is provided on how short term travellers, such as tourists, can be identified using Twitter data and how their travel pattern can be analysed. Having appropriate information about tourists/visitors - such as the places they visit, their origin and the pattern of their movements at their destination - is of great importance to urban planners. The available profile information of users and self-reported geo-location data on Twitter are used to identify tourists visiting Sydney as well as also those Sydney residents who made a trip outside Sydney. The presented data and analysis enable us to understand and track tourists' movements in cities for better urban planning. The results of this paper open up avenues for travel demand modellers to explore the possibility of using big data (in this case Twitter data) to model short distance (day-to-day or activity based) and long distance (vacation) trips.

DOI: 10.1145/2830657.2830660

Keywords: Twitter-Traffic-Status

SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based.

References

[1] Rashidi, T., J. Auld, and A. Mohammadian, Effectiveness of Bayesian Updating Attributes in Data Transferability Applications. Transportation Research Record: Journal of the Transportation Research Board, 2013(2344): p. 1-9.
[2] Miller, E., et al., A Framework for Urban Passenger Data Collection, , in 10th International Conference on Transport Survey Methods. 2014: Leura, Australia.
[3] Rashidi, T., A. Mohammadian, and Y. Zhang, Effect of Variation in Household Sociodemographics, Lifestyles, and Built Environment on Travel Behavior. Transportation Research Record: Journal of the Transportation Research Board, 2010(2156): p. 64-72.

[4] Beirão, G. and J.S. Cabral, Understanding attitudes towards public transport and private car: A qualitative study. Transport policy, 2007. 14(6): p. 478-489.

[5] Francis, R.C., et al., Object tracking and management system and method using radio-frequency identification tags. 2003, Google Patents.

[6] Zhang, Y. and A. Mohammadian, Bayesian updating of transferred household travel data. Transportation Research Record: Journal of the Transportation Research Board, 2008(2049): p. 111-118.

[7] Cramer, H., M. Rost, and L.E. Holmquist. Performing a check-in: emerging practices, norms and'conflicts' in location-sharing using foursquare. in Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services. 2011. ACM.

[8] Maghrebi, M., et al., Complementing Travel Diary Surveys with Twitter Data: Application of Text Mining Techniques on Activity Location, Type and Time, in 17th International Conference on Intelligent Transportation Systems (ITSC). 2015, IEEE: Las Palmas, Spain.

[9] Williams, S. and E. Currid-Halkett, Industry in Motion: Using Smart Phones to Explore the Spatial Network of the Garment Industry in New York City. PloS one, 2014. 9(2): p. e86165.

[10] Byon, Y.-J., B. Abdulhai, and A. Shalaby, Real-time transportation mode detection via tracking global positioning system mobile devices. Journal of Intelligent Transportation Systems, 2009. 13(4): p. 161-170.

[11] Yin, J., et al., Using social media to enhance emergency situation awareness. IEEE Intelligent Systems, 2012. 27(6): p. 52-59.

[12] Cameron, M.A., et al. Emergency situation awareness from twitter for crisis management. in Proceedings of the 21st international conference companion on World Wide Web. 2012.

[13] Gao, H., J. Tang, and H. Liu. Exploring Social-Historical Ties on Location-Based Social Networks. in ICWSM. 2012.

[14] Gao, H., et al. Exploring temporal effects for location recommendation on location-based social networks. in Proceedings of the 7th ACM conference on Recommender systems. 2013.

[15] Fu, K., R. Nune, and J.X. Tao. Social Media Data Analysis for Traffic Incident Detection and Management. in Transportation Research Board 94th Annual Meeting. 2015.

[16] Mai, E. and R. Hranac. Twitter Interactions as a Data Source for Transportation Incidents. in Proc. Transportation Research Board 92nd Ann. Meeting. 2013. [

17] Steur, R., Twitter as a spatio-temporal source for incident management. 2015.

[18] Hasan, S., X. Zhan, and S.V. Ukkusuri. Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. in Proceedings of the 2nd ACM SIGKDD international workshop on urban computing. 2013. ACM.

[19] Hasan, S. and S.V. Ukkusuri, Urban activity pattern classification using topic models from online geo-location data. Transportation Research Part C: Emerging Technologies, 2014. 44: p. 363-381.

[20] Xiang, Z. and U. Gretzel, Role of social media in online travel information search. Tourism management, 2010. 31(2): p. 179-188.

[21] Girardin, F., et al., Digital footprinting: Uncovering tourists with user-generated content. Pervasive Computing, IEEE, 2008. 7(4): p. 36-43.

[22] Popescu, A. and G. Grefenstette. Mining social media to create personalized recommendations for tourist visits. in Proceedings of the 2nd International Conference on Computing for Geospatial Research & Applications. 2011.

[23] Majid, A., et al., A context-aware personalized travel recommendation system based on geotagged social media data mining. International Journal of Geographical Information Science, 2013. 27(4): p. 662-684.

[24] Sun, Y., et al., Analyzing human activities through volunteered geographic information: Using Flickr to analyze spatial and temporal pattern of tourist accommodation, in Progress in Location-Based Services. 2013, Springer. p. 57- 69.

[25] Pozdnoukhov, A. and C. Kaiser. Space-time dynamics of topics in streaming text. in Proceedings of the 3rd ACM SIGSPATIAL international workshop on location-based social networks. 2011.

[26] Kling, F. and A. Pozdnoukhov. When a city tells a story: urban topic analysis. in Proceedings of the 20th International Conference on Advances in Geographic Information Systems. 2012.

[27] De Choudhury, M., et al. Automatic construction of travel itineraries using social breadcrumbs. in Proceedings of the 21st ACM conference on Hypertext and hypermedia. 2010.

[28] Ichimura, T. and S. Kamada. A generation method of filtering rules of Twitter via smartphone based Participatory Sensing system for tourist by interactive GHSOM and C4. 5. in Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on. 2012.

[29] Lee, J.H., A.W. Davis, and K.G. Goulias, Activity Space Estimation with Longitudinal Observations of Social Media Data, in Paper submitted for presentation at the 95th Annual Meeting of the Transportation Research Board. Washington, D.C., January 10-14, 2016.

[30] Lee, J.H., et al., Can Twitter data be used to validate travel demand models?, in 14th International Conference on Travel Behaviour Research. 2015: Windsor, UK.

[31] Blei, D.M., A.Y. Ng, and M.I. Jordan, Latent dirichlet allocation. the Journal of machine Learning research, 2003. 3: p. 993-1022.

[32] Coffey, C. and A. Pozdnoukhov. Temporal decomposition and semantic enrichment of mobility flows. in Proceedings of the 6th ACM SIGSPATIAL International Workshop on Location-Based Social Networks. 2013. ACM.

[33] Hasan, S. and S.V. Ukkusuri, Social contagion process in informal warning networks to understand evacuation timing behavior. Journal of Public Health Management and Practice, 2013. 19: p. S68-S69.

[34] Kaigo, M., Social media usage during disasters and social capital: Twitter and the Great East Japan earthquake. Keio Communication Review, 2012. 34: p. 19-35.

Proceedings Title	Proceedings: 8th ACM SIGSPATIAL International Workshop on Location-Based Social Networks

Real-time traffic detection by analyzing Twitter's tweet

Type	Journal Article
Author	Hanmant Telange
Rights	Aspire Publishers
Volume	1
Issue	2
Pages	e37
Publication	International journal of emerging technology and computer science
ISSN	2455-9954
Date	2016
Library Catalog	www.safetylit.org
Language	en

Abstract Physical events in the real world are known to trigger reactions and then discussions in online social media. Mining these reactions through online social sensors offers a fast and low cost way to understand what is happening in the physical world. In some cases, however, further study of the affected population's emotional state can improve this understanding. In our study we analyzed how car commuters react on Twitter while stuck in heavy traffic. We discovered that the online social footprint does not necessarily follow a strict linear correlation with the volume of a traffic jam. Through our analysis we offer a potential explanation: people's mood could be an additional factor, apart from traffic severity itself, that leads in fluctuations of the observed reaction in social media. This finding can be important for social sensing applications where external factors, like sentiment, also contribute on how humans react. Ignoring the existence of such factors can lead in reduced quality and accuracy of a regression analysis. We propose a novel traffic-congestion estimation model that utilizes the volume of messages and complaints in online social media, based on when they happen. We show through experimental evaluation that the proposed model can estimate, with higher accuracy, traffic jam severity and compare the results with several baselines. The model achieves at least 38% improvement of absolute error and more than 45% improvement of relative error, when compared with a baseline that assumes linear correlation between traffic and social volume. To support our findings we combined data from the California Department of Transportation (CALTRANS) and Twitter, for a total of 6 months, and focused on a major traffic-heavy freeway in Los Angeles, California.

Keywords: Twitter-Traffic-Status

Available: www.aspirepublishers.com/index.php/ijetcs/article/download/37/44

SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based.

References

[1] P. Ruchi and K. Kamalakar, "ET: Events from tweets," in Proc. 22nd Int. Conf. World Wide Web Comput., Rio de Janeiro, Brazil, 2013, pp. 613–620.

[2] G. Anastasi et al., "Urban and social sensing for sustainable mobility in smart cities," in Proc. IFIP/IEEE Int. Conf. Sustainable Internet ICT Sustainability, Palermo, Italy, 2013, pp.

1–4.

[3] A. Rosi et al., "Social sensors and pervasive services: Approaches and perspectives," in Proc. IEEE Int. Conf. PERCOM Workshops, Seattle, WA, USA, 2011, pp. 525– 530.

[4] T. Sakaki, M. Okazaki, and Y. Matsuo, "Tweet analysis for real-time event detection and earthquake reporting system development," IEEE Trans. Knowl. Data Eng., vol. 25, no. 4, pp. 919–931, Apr. 2013. [7] J. Allan, Topic Detection and Tracking: EventBased Information Organization. Norwell, MA, USA: Kluwer, 2002.

[5] K. Perera and D. Dias, "An intelligent driver guidance tool using lo- cation based services," in Proc. IEEE ICSDM, Fuzhou, China, 2011, pp. 246–251.

[6] T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, and K. Nawa, "Real-time event extraction for driving information from social sensors," in Proc. IEEE Int. Conf. CYBER, Bangkok, Thailand, 2012, pp. 221–226.

[7] B. Chen and H. H. Cheng, "A review of the applications of agent tech- nology in traffic and transportation systems," IEEE Trans. Intell. Transp. Syst., vol. 11, no. 2, pp. 485–497, Jun. 2010.

[8] S. Weiss, N. Indurkhya, T. Zhang, and F. Damerau, Text Mining: Predic- tive Methods for Analyzing Unstructured Information. Berlin, Germany: Springer-Verlag, 2004.

[9] V. Gupta, S. Gurpreet, and S. Lehal, "A survey of text mining tech- niques and applications," J. Emerging Technol. Web Intell., vol. 1, no. 1, pp. 60–76, Aug. 2009.

[10] M. W. Berry and M. Castellanos, Survey of Text Mining. New York, NY, USA: Springer-Verlag, 2004.

[11] H. Takemura and K. Tajima, "Tweet classification based on their life- time duration," in Proc. 21st ACM Int. CIKM, Shanghai, China, 2012, pp. 2367–2370.

[12] J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power, "Using social media to enhance emergency situation awareness," IEEE Intel. Syst., vol. 27, no. 6, pp. 52–59, Nov./Dec. 2012.

[13] R. Li, K. H. Lei, R. Khadiwala, and K. C.-C. Chang, "TEDAS: A Twitter- based event detection and analysis system," in Proc. 28th IEEE ICDE, Washington, DC, USA, 2012, pp. 1273– 1276.

(References numbered 14 and 15 were not included in the reference list nor were there citations with these numbers in the body of the paper)

[16] J. Weng and B.-S. Lee, "Event detection in Twitter," in Proc. 5th AAAI ICWSM, Barcelona, Spain, 2011, pp. 401–408. Keywords: Twitter-Traffic-Status<p /> <p>Language: en</p>

Exploring the diffusion of tweets designed to raise the road safety agenda in Saudi Arabia

Type	Journal Article
Author	Yousef A. Albalawi
Author	Jane Sixsmith
Rights	Sage Publications, International Union for Health Promotion and Education
Volume	ePub
Issue	ePub
Pages	ePub
Publication	Global health promotion
ISSN	1757-9759
Date	2016
DOI	10.1177/1757975915626111
Library Catalog	www.safetylit.org
Language	en

Abstract

This study demonstrates the importance of understanding the diffusion process in social media such as Twitter as an example of the relationship between new media platforms and health promotion interventions. Evidence-informed tweets were developed, pilot tested and distributed to all followers of the Ministry of Health's Twitter account with the aim of influencing the agenda on road safety in Saudi Arabia. The dissemination pattern and influence of this health communication was assessed. We collected 70 tweets into two groups (29 intervention tweets and 41 additional supported tweets) extracted from the Tweetreach data set and then analysed them using Microsoft Excel and SPSS.Using the concept of innovation/imitation as defined in the Bass Model, we classified retweeting by direct followers as innovation and retweeting by users who were not followers as imitation. In the study, we identify an informative indicator of successful diffusion and propose a novel procedure to measure innovation/imitation coefficients (p and q). We also provided a statistical procedure for evaluating tweet adoption by innovators (influentials) and imitators. In addition, we also assessed the use of message design tools for new media messages. The resulting information can be used to improve public health and health promotion interventions at the levels of planning, design, implementation and evaluation.

Keywords: Twitter-Traffic-Status

Real-time detection of traffic from Twitter stream analysis

Type	Journal Article
Author	Amol Jadhao
Author	Sweety Kumari
Author	Firdos Khan
Author	Shekh Sultan
Author	Ruchita Khandge
Rights	IJISET
Volume	2
Issue	9
Pages	685-691
Publication	International journal of innovative science, engineering and technology
ISSN	2348-7968
Date	2016
Library Catalog	www.safetylit.org
Language	en

Abstract

Twitter has received much thoughtfulness recently.In this paper, we present a real-time monitoring system for traffic event detection from Twitter stream analysis. An important characteristic of Twitter is its real-time nature. The system fetches tweets from Twitter by using many search criteria; processes tweets, by usinging text mining techniques and then performs the classification of tweets. To detect a target event, we devise a classifier of tweets based on features like keywords in a tweet, the number of words, and their context. Users are using Twitter to report real-life events. It focuses on detecting those events by analyzing these text stream in Twitter.The characteristics of Twitter make it a non-trivial task.The traffic detection system was employed for real-time monitoring of many areas of the road network, that allow for detection of traffic events almost in real time.

Keywords: Twitter, Traffic event detection, tweet classification, text mining, social sensing.

SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based.

References

[1] F. Atefeh and W. Khreich, "A survey of techniques for event detection In Twitter, Comput. Intell., vol. 31, no. 1, pp. 132–164, 2015.

[2] P. Ruchi and K. Kamalakar, "ET: Events from tweets," in Proc. 22ndnt Conf. World Wide Web Comput., Rio de Janeiro, Brazil, 2013, pp. 613–620.

[3] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee,"Measurement and analysis of online social networks," in Proc. 7th ACM SIGCOMM Conf. Internet Meas., San Diego, CA USA, 2007, pp. 29–42.

[4] G. Anastasi et al., "Urban and social sensing for sustainable mobility in smart cities," in Proc. IFIP/IEEE Int. Conf. Sustainable Internet ICT Sustainability, Palermo, Italy, 2013, pp.

1–4.

[5] A. Rosi et al., "Social sensors and pervasive services: Approaches and perspectives," in Proc. IEEE Int. Conf.PERCOM Workshops, Seattle,WA, USA, 2011, pp. 525–530.

[6] T Sakaki, M. Okazaki, and Y.Matsuo, "Tweet analysis for real-time event detection and earthquake reporting system development," IEEE Trans.Knowl. Data Eng., vol. 25, no. 4, pp. 919–931, Apr. 2013.

[7] J. Allan, Topic Detection and Tracking: Event-Based Information Organization. Norwell, MA, USA: Kluwer, 2002.

[8] K. Perera and D. Dias, "An intelligent driver guidance tool using Location based services," in Proc. IEEE ICSDM, Fuzhou, China, 2011, pp. 246–251.

[9] T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, and K. Nawa, "Real-time event extraction for driving information from social sensors," In Proc. IEEE Int. Conf. CYBER, Bangkok, Thailand, 2012, pp. 221–226.

[10] V. Gupta, S. Gurpreet, and S. Lehal, "A survey of text mining techniques and applications," J. Emerging Technol. Web Intell., vol. 1, no. 1, pp. 60–76, Aug. 2009.

Traffic detection using tweets on Twitter social network

Type	Journal Article
Author	Supriya Bhosale
Author	Sucheta Kokate
Rights	IJARCSMS
Volume	4
Issue	7
Pages	84-88
Publication	International journal of advance research in computer science and management studies
ISSN	2347-1778
Date	2016
Library Catalog	www.safetylit.org
Language	en

Abstract

Now a days social networking sites are becoming the real time information channel over time. With the help of portable devices many users are able to share the real life events on the internet. This feature of the social networking sites made them more popular and valuable. Basically these social networking sites are used for the maintaining social relationship, finding friends and users with the similar interest .The text message shared by user in social networks that message is called Status Update Message. In that SUM the text, meta-information like timestamp, name of the user, geographic coordinates, links of the other resources are present. The SUM's considered in a specific area may provide the proper information. Social networks is a source of information for event detection such as road traffic congestion and car accidents. Existing system present a real-time monitoring system for traffic event detection from twitter. The system fetches or collect tweets from twitter and then processes tweets using text mining techniques. Lastly performs the classification of tweets. The system aim is to assign the appropriate class label to each tweet, whether it is related to a traffic event or not. System used the support vector machine as a classification model. The proposed system uses the semi-supervised approach, which gives training using traffic related dataset. We propose a clustering approach for classification of the tweets in traffic related and non- traffic related tweets. We employ a Euclidean distance to calculate the similarity between the tweets.

Keywords: Tweet classification, Traffic event detection, Data mining, text mining, and social sensing; Twitter-Traffic-Status

SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based.

References

1. Eleonora D' Andrea, Pietro Ducange, Beatrice Lazzerini, Member, IEEE, and Francesco Marcelloni, Member, IEEE ,"Real-Time Detection of Traffic FromTwitter Stream Analysis",IEEE transaction on intelligent transportation system, VOL. 16, NO. 4, AUGUST 2015.

2. Rui LI, Kin Hou Lei, Ravi Khadiwala, Kevin Chen-Chuan Chang "TEDAS: a Twitter Based Event Detection and Analysis System", IJCSIT 2014.

3. Harshita Rajwani, Srushti Somvanshi, AnujaUpadhye," Dynamic Traffic Analyzer Using Twitter", International Journal of Science and Research (IJSR) 2014.

4. Vikram Singh and Balwinder Saini "An Effective Tokenization Algorithm for Information Retrieval System" CS and IT-CSCP 2014

5. Maximilian Walther and Michael Kaisser,"Geo-spatial Event Detection in the Twitter Stream", P. Serdyukov et al. (Eds.): ECIR 2013, LNCS 7814, pp. 356367, 2013 .springer Verlag Berlin Heidelberg 2013.

6. T. Sakaki, M. Okazaki, and Y.Matsuo, "Tweet analysis for real-time event detection and earthquake reporting system development," IEEE Trans. Knowl. Data Eng., vol. 25, no. 4, pp. 919-931, Apr. 2013.

7. M. Krstajic, C. Rohrdantz, M. Hund, and A. Weiler, "Getting there first: Real-time detection of real-world incidents on Twitter" in Proc. 2nd IEEE Work Interactive Vis. Text Anal.-Task-Driven Anal. Soc. Media IEEE VisWeek," Seattle, WA, USA, 2012.

8. A. Schulz, P. Ristoski, and H. Paulheim, " see a car crash: Real-time detection of small scale incidents in microblogs," in The Semantic Web: ESWC 2013 Satellite Events, vol. 7955. Berlin, Germany: Springer- Verlag, 2013, pp. 22-33.

9. J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power, "Using social media to enhance emergency situation awareness," IEEE Intell. Syst., vol. 27, no. 6, pp. 52-59, Nov./Dec. 2012.

10. K. Boriboonsomsin, M. Barth, W. Zhu, and A. Vu, "Ecorouting navigation multisource historical and real-time traffic information," IEEE Trans. Intell. Transp. Syst., vol. 13, no. 4, pp. 1694-1704, Dec. 2012.

UNIETD - assessment of third party data as information source for drivers and road operators

Type	Journal Article
Author	Daniel Elias
Author	Friedrich Nadler
Author	Ian Cornwell
Author	Susan Grant-Muller
Author	Thomas Heinrich
Rights	Elsevier Publications
Volume	14
Pages	2035-2043
Publication	Transport research procedia
ISSN	2352-1465
Date	2016
DOI	10.1016/j.trpro.2016.05.171
Library Catalog	www.safetylit.org
Language	en

Abstract The paper deals with the assessment of third party data such as crowd sourced/social media and floating vehicle data as information source for road operators in addition to traditional infrastructure-based techniques. For purposes of quality assessment of different types of data and available ground truths existing test/evaluation methodologies have been assessed. A new methodology has been designed for assessment of speeds and travel times using normalized (between 0 and 1) quality indicators that can distinguish between "detection rate" and "false alarm rate" concepts. In terms of harvesting social media the relevance of social media content has been assessed against a range of traffic management requirements. Furthermore the level of content that will be available has been estimated as well as commercial sources and business models for road authorities. Analyses cover unstructured data from Twitter and Facebook both historical data and three months of contemporary data. In addition surveys are conducted in England and Austria to retrieve information from the public in terms of which social media platforms are commonly used to share information about traffic related incidents.

SafetyLit note: We believe that the inclusion of references in this case falls under fair use. Why? 1) This article was published as open access and the journal and copyright owner is acknowledged; 2) a link to full text is provided; 3) reproduction of this reference list is relevant to a commentary on what is and is not original thought and 4) the items in this reference list are part of the data upon which an investigation are based.

References

K. Bogenberger Qualität von Verkehrsinformationen Straßenverkehrstechnik, 47 (2003), p. 10

K. Bogenberger, M. Hauschild 2009. QFCD – A microscopic model for measuring the individual quality of traffic information. ITS World Congress 2009. Stockholm, Sweden.

A. Gal-Tzur, S.M. Grant-Muller, T. Kuflik, E. Minkov, S. Nocera, I. Shoor The potential of social media in delivering transport policy goals Transport Policy, 32 (2014), pp. 115–123

M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten The WEKA data mining software: An update SIGKDD Explorations, 11 (1) (2009)

Huber, G., Bogenberger, K., Bertini, R., 2014. New Methods for Quality Assessment of Real-Time Traffic Information, Paper Number 14-2918. 93rd Annual Meeting of Transportation Research Board.

Huber, G., Bogenberger, K., 2013. A Quality Evaluation Model for Real-Time-Traffic-Information. ITSC 2013, (S. 2126-2131). ISO TR25100:2012 Intelligent transport systems – Systems architecture – Harmonization of ITS data concepts.

Lux, C., 2011. QBench – Evaluation of Traffic Flow Quality. Proceedings of BASt colloquium "Quality of on-trip road traffic information".

van Vuren, T., Cornwell, I., Finer, M., van Hinsbergen, C., Zuurbier, F., 2013. Using Short Term Traffic Predictors in Traffic Management Centres, 9th ITS European Congress, Dublin.

An approach for detecting traffic events using social media

Type	Journal Article
Author	Carlos Gutiérrez
Author	Paulo Figueiras
Author	Pedro Oliveira
Author	Ruben Costa
Author	Ricardo Jardim-Goncalves
Rights	Springer Science+Business Media
Volume	647
Pages	61-81
Publication	Studies in computational intelligence
Date	2016
DOI	10.1007/978-3-319-33353-3_4
Library Catalog	www.safetylit.org
Language	en
Abstract	Emerging Trends and Advanced Technologies for Computational Intelligence Nowadays almost everyone has access to mobile devices that offer better processing capabilities and access to new information and services, the Web is undoubtedly the best tool for sharing content, especially through social networks. Web content enhanced by mobile capabilities, enable the gathering and aggregation of information that can be useful for our everyday lives as, for example, in urban mobility where personalized real-time traffic information, can heavily influence users' travel habits, thus contributing for a better way of living. Current navigation systems fall short in several ways in order to satisfy the need to process and reason upon such volumes of data, namely, to accurately provide information about urban traffic in real-time and the possibility to personalize the information presented to users. The work presented here describes an approach to integrate, fuse and process tweet messages from traffic agencies, with the objective of detecting the geographical span of traffic events, such as accidents or road works. Tweet messages are considered in this work given their uniqueness, their real time nature, which may be used to quickly detect a traffic event, and their simplicity. We also address some imprecisions ranging from lack of geographical information, imprecise and ambiguous toponyms, overlaps and repetitions as well as visualization to our data set in the UK, and a qualitative study on the use of the approach using tweets in other languages, such as Greek. Finally, we present an application scenario, where traffic information is processed from tweets massages, triggering personalized notifications to users through Google Cloud Messaging on Android smartphones. The work presented here is still part of on-going work. Results achieved so far do not address the final conclusions but form the basis for the formalization of a domain knowledge along with the urban mobility services. Keywords: Twitter-Traffic-Status Copyright © 2016 Springer International Publishing Switzerland

From Twitter to detector: real-time traffic incident detection using social media data

Type	Journal Article
Author	Yiming Gu
Author	Zhen (Sean) Qian
Author	Feng Chen
Rights	Elsevier Publishing
Volume	67
Pages	321-342
Publication	Transportation research part C: emerging technologies
ISSN	0968-090X
Date	2016
DOI	10.1016/j.trc.2016.02.011
Library Catalog	www.safetylit.org
Language	en
Abstract	The effectiveness of traditional incident detection is often limited by sparse sensor coverage, and reporting incidents to emergency response systems is labor-intensive. We propose to mine tweet texts to extract incident information on both highways and arterials as an efficient and cost-effective alternative to existing data sources. This paper presents a methodology to crawl, process and filter tweets that are accessible by the public for free. Tweets are acquired from Twitter using the REST API in real time. The process of adaptive data acquisition establishes a dictionary of important keywords and their combinations that can imply traffic incidents (TI). A tweet is then mapped into a high dimensional binary vector in a feature space formed by the dictionary, and classified into either TI related or not. All the TI tweets are then geocoded to determine their locations, and further classified into one of the five incident categories. We apply the methodology in two regions, the Pittsburgh and Philadelphia Metropolitan Areas. Overall, mining tweets holds great potentials to complement existing traffic incident data in a very cheap way. A small sample of tweets acquired from the Twitter API cover most of the incidents reported in the existing data set, and additional incidents can be identified through analyzing tweets text. Twitter also provides ample additional information with a reasonable coverage on arterials. A tweet that is related to TI and geocodable accounts for approximately 5% of all the acquired tweets. Of those geocodable TI tweets, 60-70% are posted by influential users (IU), namely public Twitter accounts mostly owned by public agencies and media, while the rest is contributed by individual users. There is more incident information provided by Twitter on weekends than on weekdays. Within the same day, both individuals and IUs tend to report incidents more frequently during the day time than at night, especially during traffic peak hours. Individual tweets are more likely to report incidents near the center of a city, and the volume of information significantly decays outwards from the center. SafetyLit Keywords: Twitter-Traffic-Status

Analysis of the characteristics of expressway traffic information propagation using Twitter

Type	Journal Article
Author	Hwan Pil Lee
Author	Doo-Pyo Hong
Author	Eum Han
Author	Soo Hee Kim
Author	Ilsoo Yun
Rights	Springer Science+Business Media
Volume	ePub
Issue	ePub
Pages	ePub
Publication	KSCE journal of civil engineering
ISSN	1226-7988
Date	2016
DOI	10.1007/s12205-016-0781-1
Library Catalog	www.safetylit.org
Language	en
Abstract	The purpose of this study was to analyze the characteristics of traffic information propagation via Twitter using a keyword analysis and a network analysis. For the keyword analysis, the main contents of Twitter messages were identified using a TF-IDF (Term Frequency - Inverse Document Frequency) model. For the network analysis, the network connectivity among Twitter users, including Traffic Information Producers (TIPs), Opinion Leaders (OLs), and their followers were measured by estimating the densities and mean distances in their follow networks. Based on the keyword analysis result, the words representing traffic conditions were revealed as the most influential keywords. In addition, the information regarding traffic accident occurrences was found to be most frequently retweeted. As a result of the network analysis, MBC news which is one of the biggest newscasts in Korea showed the greatest connectivity among TIPs. OLs proved more powerful in information propagation than TIPs. Conclusively, there is an apparent demand for establishing strategies to propagate traffic information based on the characteristics of Twitter in a more efficient manner. SafetyLit Keywords: Twitter-Traffic-Status

Real-time incident detection using social media data

Type	Report
Author	Zhen (Sean) Qian
Place	Pittsburgh, PA, USA
Pages	46
Date	2016
Institution	Carnegie Mellon University / Pennsylvania Department of Transportation
Library Catalog	www.safetylit.org
Language	en

Abstract

The effectiveness of traditional incident detection is often limited by sparse sensor coverage, and reporting incidents to emergency response systems is labor-intensive. This research project mines tweet texts to extract incident information on both highways and arterials as an efficient and cost- effective alternative to existing data sources. This research report presents a methodology to crawl, process and filter tweets that are accessible by the public for free. Tweets are acquired from Twitter using the REST API in real time. The process of adaptive data acquisition establishes a dictionary of important keywords and their combinations that can imply traffic incidents (TI). A tweet is then mapped into a high dimensional binary vector in a feature space formed by the dictionary, and classified into either TI related or not. All the TI tweets are then geocoded to determine their locations, and further classified into one of the five incident categories. We apply the methodology in two regions, the Pittsburgh and Philadelphia Metropolitan Areas. Overall, mining tweets holds great potentials to complement existing traffic incident data in a very cheap way. A small sample of tweets acquired from the Twitter API cover most of the incidents reported in the existing data set, and additional incidents can be identified through analyzing tweets text. Twitter also provides ample additional information with a reasonable coverage on arterials. A tweet that is related to TI and geocodable accounts for approximately 10% of all the acquired tweets. Of those geocodable TI tweets, the majority are posted by influential users (IU), namely public Twitter accounts owned by public agencies and media, while a small number is contributed by individual users. There is more incident information provided by Twitter on weekends than on weekdays. Within the same day, both individuals and IUs tend to report incidents more frequently during the day time than at night, especially during traffic peak hours. Individual tweets are more likely to report incidents near the center of a city, and the volume of information significantly decays outwards from the center. We develop a prototype web application to allow users extract both real-time and historical incident information and visualize it on the map. The web application will be tested in PennDOT transportation management centers.

Pear Analytics (2009), ‘Twitter study–august 2009’, San Antonio, TX: Pear Analytics. Available at: www.pearana-lytics.com/blog/wp-content/uploads/2010/05/Twitter-Study-August-2009.pdf.

Gelernter, J. and Balaji, S. (2013), ‘An algorithm for local geoparsing of microtext’, GeoInformatica 17(4), 635–667

Report Number FHWA-PA-2016-004-CMU WO 03

Real time monitoring system for traffic from Twitter stream analysis

Type	Journal Article
Author	S. K. Sameera
Author	S. K. Karimulla
Rights	Rishi Educational Society
Volume	3
Issue	5
Pages	27-31
Publication	International journal of computational science, mathematics and engineering
ISSN	2349-8439
Date	2016
Library Catalog	www.safetylit.org
Language	en

Abstract

We present a real-time monitoring system for traffic event detection from Twitter stream analysis. The system fetches tweets from Twitter according to several search criteria; processes tweets. The traffic detection system was employed for real-time monitoring of several areas of the road network, allowing for detection of traffic events almost in real time, often before online traffic news web sites. We were also able to discriminate if traffic is caused by an external event or not. Event detection from social networks analysis is a more challenging problem than event detection from traditional media like blogs, emails, etc., where texts are well formatted. SUMs are unstructured and irregular texts, they contain informal or abbreviated words, misspellings or grammatical errors. SUMs contain a huge amount of not useful or meaningless information. In our project, we focus on a particular small-scale event, i.e., road traffic, and we aim to detect and analyze traffic events by processing users' SUMs belonging to a certain area and road traffic. To this aim, we propose a system able to fetch, elaborate, a road traffic event. Tweets are up to 140 characters, enhancing the real-time and news-oriented nature of the platform. In fact, the life-time of tweets is usually very short, thus Twitter is the social network platform that is best suited to study SUMs related to real-time events. It detects the traffic events in real-time; and it is developed as an event-driven infrastructure, built on an SOA architecture. Keywords: Social media; Traffic detection; Text mining; Privacy; Service Oriented Architecture (SOA), machine learning, Twitter stream analysis

References

[1] F. Atefeh and W. Khreich, "A survey of techniques for event detection in Twitter," Comput. Intell., vol. 31, no. 1, pp. 132–164, 2015.

[2] . Ruchi and K. Kamalakar, "ET: Events from tweets," in Proc. 22nd Int. Conf. World Wide Web Comput., Rio de Janeiro, Brazil, 2013, pp. 613–620

[3] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in Proc. 7th ACM SIGCOMM Conf. Internet Meas., San Diego, CA, USA, 2007, pp. 29–42.

[4] G. Anastasi et al., "Urban and social sensing for sus- tainable mobility in smart cities," in Proc. IFIP/IEEE Int. Conf. Sustainable Internet ICT Sustainability, Palermo, Italy, 2013, pp. 1–4.

[5] A. Rosi et al., "Social sensors and pervasive services: Approaches and perspectives," in Proc. IEEE Int. Conf. PERCOM Workshops, Seattle, WA, USA, 2011, pp. 525–530.

[6] T. Sakaki, M. Okazaki, and Y. Matsuo, "Tweet analysis for real-time event detection and earthquake reporting system development," IEEE Trans. Knowl. Data Eng., vol. 25, no. 4, pp. 919–931, Apr. 2013.

[7] J. Allan, Topic Detection and Tracking: Event-Based Information Organization. Norwell, MA, USA: Kluwer, 2002.

[8] K. Perera and D. Dias, "An intelligent driver guid- ance tool using location based services," in Proc. IEEE ICSDM, Fuzhou, China, 2011, pp. 246–251.

[9] T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, and K. Nawa, "Real-time event extraction for driving information from social sensors," in Proc. IEEE Int. Conf. CYBER, Bangkok, Thailand, 2012, pp. 221–226.

[10] N. Wanichayapong, W. Pruthipunyaskul, Pattara Atikom, and P. Chaovalit, Social based traffic infor- mation extraction and classification in Proc. 11th Int. Conference IT ST, St. Petersburg, Russia, 2011.pp. 107–112.

Real-time road traffic information detection through social media

Type	Thesis
Author	Chandra P. Khatri
Place	Atlanta, GA, USA
Date	2016
Library Catalog	www.safetylit.org
Type	Masters
Language	en
University	Georgia Institute of Technology, Department of Civil and Environmental Engineering

Abstract

In current study, a mechanism to extract traffic related information such as congestion and incidents from textual data from the internet is proposed. The current source of data is Twitter, however, the same mechanism can be extended to any kind of text available on the internet. As the data being considered is extremely large in size automated models are developed to stream, download, and mine the data in real-time. Furthermore, if any tweet has traffic related information then the models should be able to infer and extract this data. To pursue this task, Artificial Intelligence, Machine Learning, and Natural Language Processing techniques are used. These models are designed in such a way that they are able to detect the traffic congestion and traffic incidents from the Twitter stream at any location. Currently, the data is collected only for United States. The data is collected for 85 days (50 complete and 35 partial) randomly sampled over the span of five months (September, 2014 to February, 2015) and a total of 120,000 geo-tagged traffic related tweets are extracted, while six million geo-tagged non-traffic related tweets are retrieved. The classification models for detection of traffic congestion and incidents are trained on this dataset. Furthermore, this data is also used for various kinds of spatial and temporal analysis. A mechanism to calculate level of traffic congestion, safety, and traffic perception for cities in U.S. is proposed. Traffic congestion and safety rankings for the various urban areas are obtained and then they are statistically validated with existing widely adopted rankings. Traffic perception depicts the attitude and perception of people towards the traffic. It is also seen that traffic related data when visualized spatially and temporally provides the same pattern as the actual traffic flows for various urban areas. When visualized at the city level, it is clearly visible that the flow of tweets is similar to flow of vehicles and that the traffic related tweets are representative of traffic within the cities. With all the findings in current study, it is shown that significant amount of traffic related information can be extracted from Twitter and other sources on internet. Furthermore, Twitter and these data sources are freely available and are not bound by spatial and temporal limitations. That is, wherever there is a user there is a potential for data.

Keywords: Twitter-Traffic-Status

References:

1. Apple Maps, https://www.apple.com/ios/maps/

2. Artificial Neural Network, Wikipedia: http://en.wikipedia.org/wiki/Artificial_neural_network

3. Aftabuzzaman, M.; "Measuring Traffic Congestion-A Critical Review." Australasian Transport Research Forum (ATRF), 30TH, 2007, Melbourne, Victoria, Australia, Vol 30. 2007.

4. Allstate Best Drivers Report: The Allstate America's Best Drivers Report, https://www.allstate.com/resources/Allstate/attachments/tools-and- resources/ABD-Report-2014.pdf

5. Autoencoder, Wikipedia: http://en.wikipedia.org/wiki/Autoencoder

6. BeatTheTraffic: http://www.beatthetraffic.com/

7. Bengio, Y.; Goodfellow, I.; Courville, A., "Deep Learning", Book in preparation for MIT Press.

8. Bing Maps, https://www.bing.com/maps/

9. Blei, D., "Probabilistic topic models". Communications of the ACM CACM Homepage archive Volume 55 Issue 4, April 2012 Pages 77-84

10. Blei, D.; Ng, A.; Jordan, M., "Latent Dirichlet Allocation". Journal of Machine Learning Research 3 (2003) 993-1022

11. Bottou, L.; "Stochastic Gradient Descent Tricks". Lecture Notes in Computer Science Volume 7700, 2012, pp 421-436

12. Burks, L.; Miller, M.; Zadeh, R. "Rapid Estimate of Ground Shaking Intensity By Combining Simple Earthquake Characteristics With Tweets". Tenth U.S. National Conference on Earthquake Engineering, Frontiers of Earthquake Engineering, July 21- 25, 2014

13. Burton, S.; Tanner, K.; Giraud, C.; West, J.; Barnes, M. ""Right Time, Right Place". Health Communication on Twitter: Value and Accuracy of Location Information", JMIR Publications, Vol 14, No 6 (2012)

14. CartoDB http://cartodb.com/

15. Chen, P.; Chen, F.; Qian, Z., "Road Traffic Congestion Monitoring in Social Media with Hinge-Loss Markov Random Fields". Data Mining (ICDM), 2014 IEEE

16. Choudhury, M.; Counts, S.; Horvitz, E., "Predicting Postpartum Changes in Emotion and Behavior via Social Media". Proceeding CHI '13 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Pages 3267-3276. ACM New York, NY, USA 2013

17. Chowdhury, S.; Amer-Yahia, S.; Castillo, C.; Imran M.; Asghar, M., "Tweet4act: Using Incident-Specific Profiles for Classifying Crisis-Related Messages". The 10th International Conference on Information Systems for Crisis Response and Management (ISCRAM), 2013.

18. CNET, Apple wins patent for crowd-sourced traffic navigation. http://www.cnet.com/news/apple-crowdsourced-traffic-system-would-account-for- stop-lights/

19. Collins, C.; Hasan, S.; Ukkusuri, S., V., "A novel transit rider satisfaction metric: Riders sentiment measured from online social media data." Journal of Public Transportation Vol. 16(2), pp. 21-45, 2013.

20. Cross-validation, Wikipedia http://en.wikipedia.org/wiki/Cross- validation_%28statistics %29

21. Crowdsourcing, Wikipedia http://en.wikipedia.org/wiki/Crowdsourcing

22. Curse of Dimensionality, Wikipedia, http://en.wikipedia.org/wiki/Curse_of_dimensionality

23. Dejoy, D. "The optimism bias and traffic accident risk perception". Accident Analysis & Prevention, Volume 21, Issue 4, August 1989, Pages 333–340

24. Dejoy, D. "An examination of gender differences in traffic accident risk perception". Accident Analysis and Prevention, Volume 24, Issue 3., 1992

25. Feature Selection, Wikipedia: Feature Selection, http://en.wikipedia.org/wiki/Feature_selection

26. Freifeld, C.; Brownstein, J.; Menone, C.; Bao, C.; Filice, R.; Kass-Hout, T.; Dasgupta, N. "Digital Drug Safety Surveillance: Monitoring Pharmaceutical Products in Twitter". Drug Safety 2014; 37(5): 343–350.

27. Gensim https://radimrehurek.com/gensim/

28. Georgia 511, http://www.511ga.org/

29. Grauwin, S., Sobolevsky, S., Szell, M., Ratti, C. "Signature of Humanity", MIT Senseable City Lab, 2012.

30. Google Word2Vec Library, https://code.google.com/p/word2vec/

31. Guyon, I.; Elisseeff, A., "An Introduction to Variable and Feature Selection." Journal of Machine Learning Research 3 (2003) 1157-1182

32. Hasan, S.; Ukkusuri, S.,V., "Urban activity pattern classification using topic models from online geo-location data." Transportation Research Part C: Emerging Technologies Volume 44, pp. 363-381, 2014.

33. Here, https://www.here.com/?x=ep&map=33.7763,-84.398,10,normal

34. Ibanez, R.; Muresan, S.; Wacholder, N., "Identifying Sarcasm in Twitter: A Closer Look." HLT '11 Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers - Volume 2

35. IBM, Big Data, for better or worse: 90% of world's data generated over last two years, SINTEF, May 22, 2013

36. INRIX: http://intrix.com/

37. INRIX Ranking Methodology: http://www.inrix.com/scorecard/methodology-en/

38. Intermodal Transportation Database, http://www.rita.dot.gov/bts/data_and_statistics/intermodal_transportation_databa se.html

39. Kohavi, R.; Provost, K. (1998). "Glossary of terms". Journal of Machine Learning Research (JMLR) 30: 271–274

40. Kosala, R.; Adi, E.; Steven, "Harvesting Real Time Traffic Information from Twitter". International Conference on Advances Science and Contemporary Engineering (ICASCE), 2012

41. Kostkova, P.; Szomszor, M.; Luis, C. "#swineflu: The Use of Twitter as an Early Warning and Risk Communication Tool in the 2009 Swine Flu Pandemic"

42. Latent Dirichlet Allocation, Wikipedia.
http://en.wikipedia.org/wiki/Latent_Dirichlet_allocation

43. Lecue, F.; Tucker, R.; Bicer, V.; Tommasi, P.; Tallevi-Diotallevi, S.; Sbodio, M. "Predicting Severity of Road Traffic Congestion Using Semantic Web Technologies." The Semantic Web: Trends and Challenges Lecture Notes in Computer Science Volume 8465, 2014, pp 611-627

44. Levinson, H.S.; Lomax, T.J., "Developing a travel time congestion index," Transportation Research Record: Journal of the Transportation Research Board, (1996) No. 1564, pp. 1-10.

45. Li, X.; Lord, D.; Zhang, Y.; Xie, Y., "Predicting motor vehicle crashes using Support Vector Machine models". Accident Analysis & Prevention, Volume 40, Issue 4, July 2008, Pages 1611–1618

46. Lund, I.; Rundmo, T., "Cross-cultural comparisons of traffic safety, risk perception, attitudes and behaviour". Safety Science, 47, 547-553, 2009.

47. Mahmud, J.; Nichols, J.; Drews, C. "Home Location Identification of Twitter Users", ACM Transactions on Intelligent Systems and Technology (TIST) - Special Section on Urban Computing archive, Volume 5 Issue 3, September 2014, Article No. 47

48. Manning, C.; Schutze, H., "Foundations of Statistical Natural Language Processing", 1999.

49. Manning, C.; Raghavan, P.; Schutze, H., "An Introduction to Information Retrieval", 2008.

50. Mapquest: http://www.mapquest.com/

51. Marcus, A.; Bernstein, M.; Badar, O.; Karger, D.; Madden, S.; Miller, R., "Processing and Visualizing the Data in Tweets" (201, Pages 3267-32763), 2013.

52. Maynard, D.; Greenwood, M. "Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis." Proc. of LREC 2014, Reykjavik, Iceland, May

53. Microblogs, Wikipedia. http://en.wikipedia.org/wiki/Microblogging

54. Mikolov, T.; Yih, W.; Zweig, G., "Linguistic regularities in continuous space word representations." Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT-2013)

55. Miley Legal Group, Fatal Car Accident Statistics Based on Time of Day, http://www.hg.org/article.asp?id=29836

56. Mirowski, P.; Ranzato, M.; LeCun, Y., "Dynamic Auto-Encoders for Semantic Indexing". Proceedings of the NIPS 2010 Workshop on Deep Learning, 2010

57. Mitchell, T., "Machine Learning." McGraw-Hill, New York, 1997

58. Naïve Bayes Classifier, StatSoft, Dell Inc. http://www.statsoft.com/textbook/naive- bayes-classifier

59. Natural Language Processing, Wikipedia, http://en.wikipedia.org/wiki/Natural_language_processing

60. Natural Language Toolkit http://www.nltk.org/

61. Natural Language Toolkit, http://www.nltk.org/book/ch07.html, http://www.nltk.org/

62. N-gram, Wikipedia: http://en.wikipedia.org/wiki/N-gram

63. Nordfjaern, T.; Jorgensen, S.; Rundmo, T., "Cultural and socio-demographic predictors of car accident involvement in Norway, Ghana, Tanzania and Uganda, Safety Science". 2012, DOI: http://dx.doi.org/10.1016/j. ssci.2012.05.003, 50(9): 1862-1872

64. OMB Bulletin, "Revised Delineations of Metropolitan Statistical Areas, Micropolitan Statistical Areas, and Combined Statistical Areas, and Guidance on Uses of the Delineations of These Areas", Executive Office of the President Office of Management and Budget, OMB Bulletin No. 13-01 , 2013

65. Open Street Map https://www.openstreetmap.org/#map=5/51.500/-0.100

66. Oxford dictionary, Artificial Intelligence http://www.oxforddictionaries.com/us/definition/american_english/artificial- intelligence

67. Palen, L.; Anderson, K.; Mark, G.; Martin, J.; Sicker, D.; Palmer, M.; Grunwald, D., "A vision for technology- mediated support for public participation & assistance in mass emergencies & disasters." In Proceedings of the 2010 ACM-BCS Visions of Computer Science Conference, 2010

68. Pereira, F.; Rodrigues, F.; Ben-Akiva, M., "Text analysis in incident duration prediction." Transportation Research Part C 37 177–192, 2013.

69. Pew Research, Social Networking Fact Sheet, Pew Research Internet Project

70. Pickle Python Library, https://docs.python.org/2/library/pickle.html 71. Postgre SQL http://www.postgresql.org/ 72. Prudhomme, R.; Lee, C.; "Sprawl, Speed and the Efficiency of Cities." Urban Studies October 1999 vol. 36 no. 11 1849-1858

73. Quora, Twitter usage. (2014) http://www.quora.com/How-much-or-what-percent- of-data-can-I-stream-using-the-twitter-streaming-API-I-want-to-analyse-some- tweets-on-my-Hadoop-cluster

74. Rehurek, R., Gensim Topic Modeling Library, https://radimrehurek.com/gensim/about.html

75. Ribeiro, S.; Davis, C.; Oliveira, D.; Meira, W.; Goncalves, T.; Pappa, G., "Traffic observatory: a system to detect and locate traffic events and conditions using Twitter." 2012.

76. Rios, M.; Lin, J., "Visualizing the "Pulse" of World Cities on Twitter." Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media

77. Rouse, M. , Microblogging." http://searchmobilecomputing.techtarget.com/definition/microblogging

78. Russell, Stuart J.; Norvig, Peter (2003), Artificial Intelligence: A Modern Approach

79. Saabas, A.; Microsoft, 2014, "Selecting good features." A blog on machine learning, data mining and visualization, http://blog.datadive.net/selecting-good-features-part- iii-random-forests/

80. Sakaki, T.; Matsuo, Y.; Yanagihara, T.; Chandrasiri, N.; Nawa, K., "Real-time Event Extraction for Driving Information from Social Sensors." Proceedings of the 2012 IEEE

International Conference on Cyber Technology in Automation, Control and Intelligent Systems May 27-31, 2012, Bangkok, Thailand

81.Sang, E.; Meulder, F., "Introduction to the CoNLL-2003 shared task: language- independent named entity recognition." CONLL '03 Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003 - Volume 4. Pages 142- 147

82. Sentiment Analysis, http://en.wikipedia.org/wiki/Sentiment_analysis

83. Schulz, A.; Ristoski, P.,"The Car that Hit The Burning House: Understanding Small Scale Incident Related Information in Microblogs." AAAI Technical Report WS-13-0, 2013.

84. Scikit Learn http://scikit-learn.org/stable/

85. Scipy http://www.scipy.org/

86. Sheila Stewart Traffic Accident, 2013, http://www.nbcwashington.com/news/local/Sheila-Stewart-Radio-One-Personality- Killed-in-Car-Accident-229100381.html

87. Sheppard, K., "15 Green Cities". Environmental News and Commentary. Retrieved June 23, 2010.

88. Simsekoglu, O.; Nordfaern , T.; Rundmo , T.,"Traffic risk perception, road safety attitudes, and behaviors among road users: a comparison of Turkey and Norway." Volume 15, Issue 7, 2012.

89. Stanford Named Entity Recognition Library, http://nlp.stanford.edu/software/CRF-NER.shtml

90. Stanford Parts of Speech Tagger, http://nlp.stanford.edu/software/tagger.shtml

91. Stanford CoreNLP http://nlp.stanford.edu/software/corenlp.shtml

92. Stochastic Gradient Descent, Wikpedia. http://en.wikipedia.org/wiki/Stochastic_gradient_descent

93.Tang, J.; Alelyani, S., Liu, H., "Feature Selection for Classification: A Review." http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.409.5195&rep=rep1&ty pe=pdf

94. Texas A&M Transportation Guides, http://guides.library.tamu.edu/content.php?pid=207995&sid=1736740

95.Texas Transportation Institute, Schrank, D., Eisele, B., Lomax, T., "Texas Transportation Institute 2012 Urban Mobility Report", 2012

96. TfIdf, Wikipedia, http://en.wikipedia.org/wiki/Tf%E2%80%93idf

97. Thesaurus dictionary www.thesaurus.com

98. Transportation Data Sources, https://www.tdsource.com/

99. Tolosi, L.; Lengauer, T., "Classification with correlated features: unreliability of feature ranking and solutions." National Center for Biotechnology Information, Journal of Bioinformatics. 2011 Jul 15;27(14):1986-94

100. Topic Modeling: http://en.wikipedia.org/wiki/Topic_model

101. TomTom: http://www.tomtom.com/en_us/

102. Tostes, A.; Silva, T.; Duarte-Figueiredo, F., "Studying Traffic Conditions by Analyzing Foursquare and Instagram Data." PE-WASUN '14 Proceedings of the 11th ACM symposium on Performance evaluation of wireless ad hoc, sensor, & ubiquitous networks. Pages 17-24

103. Turner, M., "Examination of indicators of congestion level." Transportation Research Record: Journal of the Transportation Research Board, (1992) No. 1360, pp.150-157.

104. Tweepy Python Library, http://www.tweepy.org/

105. Twitter: About Twitter: https://about.twitter.com/company

106. Twitter: APIs: https://dev.twitter.com/streaming/overview

107. Twitter for Business, Twitter https://business.twitter.com/

108. Twitter Research Grants, https://blog.twitter.com/2014/introducing-twitter- data-grants https://dev.twitter.com/streaming/overview/processing

109. Urban Mobility Corporation .The "Smart Growth" Debate Continues". May–June 2003. Retrieved November 7, 2006.

110. United States Census Bureau (http://www.census.gov/compendia/statab/cats/transportation.html)

111. Verma, S.; Vieweg, S.; Corvey, W.; Palen, L.; Martin, J.; Palmer, M.; Schram, A.; Anderson, K., "Natural Language Processing to the Rescue? Extracting "Situational Awareness" Tweets During Mass Emergency." AAAI Publications, Fifth International AAAI Conference on Weblogs and Social Media, 2010

112. Wanichayapong, N. ; Nat. Electron. & Comput. Technol., "Social-based traffic information extraction and classification", ITS Telecommunications (ITST), 2011 11th International Conference

113. Waze: https://www.waze.com/

114. Wongsuphasawat, K., "A Narrative Display for Sports Tournament Recap." Workshop on Sports Data Visualization in conjunction with IEEE VIS 2013

115. Wu, S.; Boston, W., University of Southern California 20% Tweets geotagged, https://pressroom.usc.edu/twitter-and-privacy-nearly-one-in-five-tweets-divulge- user-location-through-geotagging-or-metadata/, 2013

116. Yu, R., Abdel-Aty, M., "Utilizing support vector machine in real-time crash risk evaluation." Accident Analysis & Prevention, Volume 51, March 2013, Pages 252–259