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](mailto: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

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


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


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,


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


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,


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.


Real-time sensing of traffic information in Twitter messages

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


Proceedings Title

Proceedings: IEEE-ITSC Workshop on Artificial Transportation Systems and Simulation
2011
Identification and characteriaztion of events in social media

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.

Copyright © Hila Becker
Bibliography


[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.


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.


[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.


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.


2012
Harvesting real time traffic information from Twitter

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.

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


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

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.


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


Language: en
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


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


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


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


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]


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

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

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)


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)


Discovering social events through online attention

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

Abstract

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


2015
Real-time detection of traffic from Twitter stream analysis

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


23. No reference is included for item 23


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.


Keywords: Twitter-Traffic-Status
Traffic detection using tweets on Twitter social network

Abstract

Social networks can be employed as a source of information for event detection such as road traffic congestion and car accidents. Existing systems 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, it 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. The system employed the support vector machine as a classification model. The proposed system uses a system based on a semi-supervised approach, which gives training using traffic-related datasets. We propose a clustering approach for classification of the tweets into 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


Ubiquitous user-input contents on social media and online services have generated a tremendous amount of information. 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.

Copyright © 2015 by The Institute of Electrical and Electronics Engineers, Inc.


Language: en
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.


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]


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


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

<table>
<thead>
<tr>
<th>Type</th>
<th>Journal Article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author</td>
<td>Lei Lin</td>
</tr>
<tr>
<td>Author</td>
<td>Ming Ni</td>
</tr>
<tr>
<td>Author</td>
<td>Qing He</td>
</tr>
<tr>
<td>Author</td>
<td>Jing Gao</td>
</tr>
<tr>
<td>Author</td>
<td>Adel W. Sadek</td>
</tr>
<tr>
<td>Rights</td>
<td>Transportation Research Board, National Research Council, National Academy of Sciences USA</td>
</tr>
<tr>
<td>Volume</td>
<td>2482</td>
</tr>
<tr>
<td>Pages</td>
<td>82-89</td>
</tr>
<tr>
<td>Publication</td>
<td>Transportation research record</td>
</tr>
<tr>
<td>ISSN</td>
<td>0361-1981</td>
</tr>
<tr>
<td>Date</td>
<td>2015</td>
</tr>
<tr>
<td>DOI</td>
<td>10.3141/2482-11</td>
</tr>
<tr>
<td>Library Catalog</td>
<td><a href="http://www.safetylit.org">www.safetylit.org</a></td>
</tr>
<tr>
<td>Language</td>
<td>en</td>
</tr>
</tbody>
</table>

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**


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 traffic 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:

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


Keywords: Twitter-Traffic-Status
Citywide traffic congestion estimation with social media

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 Estimation (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.


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]


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]


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]


2016
Estimation of traffic with accuracy through Twitter stream analysis

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


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


Real-time detection of traffic from Twitter stream analysis

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


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.


Twitter stream analysis for traffic detection and earthquake

Social networks have been recently employed as a source of information for event detection, with specific reference to road traffic activity congestion and accidents or earthquake reporting systems. 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 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 classification 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


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

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.
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


Real-time traffic detection by analyzing Twitter's tweet

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


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


(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)

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.
Real-time detection of traffic from Twitter stream analysis

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 using 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


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.


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


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

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


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)


**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
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

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
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
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.

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.


Report Number FHWA-PA-2016-004-CMU WO 03
Real time monitoring system for traffic from Twitter stream analysis

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),

References


Real-time road traffic information detection through social media

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.

Copyright © 2016, Khatri, Chandra P.

Keywords: Twitter-Traffic-Status

References:


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


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


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


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/


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. 


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


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


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


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

66. Oxford dictionary, Artificial Intelligence


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

70. Pickle Python Library, https://docs.python.org/2/library/pickle.html
71. PostgreSQL http://www.postgresql.org/


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


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


97. Thesaurus dictionary www.thesaurus.com

98. Transportation Data Sources, https://www.tdsource.com/


105. Twitter: About Twitter: https://about.twitter.com/company

106. Twitter: APIs: https://dev.twitter.com/streaming/overview


110. United States Census Bureau (http://www.census.gov/compendia/statab/cats/transportation.html)


113. Waze: https://www.waze.com/


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