Visually Contextualizing Social Media within Spatial, Temporal and Thematic Constraints for Disaster Situation Awareness

Brian Tomaszewski, Katherine Dickens, Aishwarya A. Sawant, Pawan Kumar Ogireddy, Qi Yu, and Ed Holden

Abstract—Social media such as Twitter is an increasingly important "big data" source for disaster situation awareness. However, social media often lacks context in terms of relating social media to heterogeneous indicator documents such as maps, reports and images also vital to disaster situation awareness. In this paper, we present our research in progress examining how social media can be contextualized through indicator documents within spatial and thematic constraints. Specifically, we present pilot experiment results where we relevance ranked Federal Emergency Management Agency (FEMA) national situation reports using query terms derived from machine learning-based classification of 3.4 million Tweets. We also present our initial work on developing two visual interfaces to support analytical reasoning and contextualization of machine-classified social media with indicator documents. The first is a graph-based interface to examine thematic relationships between machine-classified terms derived from social media and ranked indicator documents. The second is a geographic map interface that visually displays relationships between social media point densities and indicator document location references. We also outline ideas for future work in the temporal dimension. Our quantitative results and ongoing visual interface work indicate that the approach we are investigating has promise and can be improved with further research.

Index Terms— Social Media, Visual Analytics, Geographic Information Retrieval, Information Fusion, Disaster Analytics

INTRODUCTION

The problem our research program is investigating is how to manage, query, and analyze massive social media data sets in order to visually contextualize heterogeneous structured and unstructured data during disasters**.** Social media such as Twitter is an increasingly important source of information for disaster situation awareness [1]. Existing research on Twitter analytics for disaster situation awareness has primarily focused on examining Twitter content in isolation such as reports of earthquake shock locations [2] and tweet classification and clustering from non-news sources to group news-related tweets by location and time [3]. Most existing works in this area rely solely on social media data without referring to other information sources to establish a proper context and content verification. Since social media (such as tweets) come from autonomous users, many of them may be highly noisy and lack authenticity. Solely relying on social media sources to deal with disaster situations may lead to biased decisions. Furthermore, during a disaster, heterogeneous structured and unstructured indicator data such as maps, reports and images are also vital disaster situation awareness sources [4, 5]. Our team is thus interested in understanding how social media can visually contextualize indicator data sources to provide disaster situation awareness within space, time and thematic constraints as these are fundamental contextual dimensions [6] (Figure 1). In the following section, we outline important relevant literature for situating our research.

1 LITERATURE REVIEW

Visual Analytics has emerged in past 10 years as an interdisciplinary field focused on integrating computational data processing and transformation methodologies with interactive, visual interfaces to support human analytical reason and capacities associated with human vision [7]. Visual Analytics could in fact be seen as a pre-cursor to the now almost ubiquitous focus in the computing world on "big data" as harvesting, structuring, analysing, visual representing and making sense of massive, *unstructured* data was a hallmark of what distinguished Visual Analytics as a new discipline. Geographic Information Scientists were very quick to adopt geographic and spatial perspectives on Visual Analytics leading to the development of Geovisual Analytics [8]. Disaster Management was also quickly identified as a highly germane Geovisual Analytics application domain [9] with several research-oriented systems, validated by practitioner input, such as the *Visual Analytic Globe* [10] and *SensePlace 2* [11] being established in the literature. Our research draws upon this rich tradition of applied Geovisual Analytics grounded in real world need for disaster management to advance calls in the Geovisual Analytics and GIScience community for additional research on diverse data, data integration, and lightweight and scalable Geovisual Analytic systems [12]. In the following section we outline our social media contextualization experiments that underlie our visual contextualization research ideas.

2 THE EXPERIMENTS

2.1 The Scenario

Our simple, yet realistic hypothetical social media contextualization disaster situation awareness scenario for our experiment is as follows: Kate is an information officer working on flood mitigation activities. She needs to know the "who, what, where and when" to achieve situation awareness about flooding over the past summer to make future flood management decisions. She wants to know what citizens were saying about floods through twitter and contextualize those tweets within the context of official government reports. (Scenario based on [13]).

2.2 Data Sets

2.2.1 FEMA Situation Reports

We captured 89 FEMA daily situation reports during summer 2013 to serve as indicator documents for our experiment where the FEMA daily situation reports would be contextualized based on keywords derived from flood tweets.

- *Brian Tomaszewski, Katherine Dickens, Aishwarya Sawant, Pawan Kumar Ogireddy, Qi Yu, and Ed Holden are with the Department of Information Sciences and Technologies, Rochester Institute of Technology, USA.*
- *E-mail: {bmtski, kjd1073, aas8000, pxo6019, qyuvks, edward.holden}@rit.edu.*

Manuscript received 28 Apr. 2014; accepted x xxx. 2014; date of publication xx xxx 2014; date of current version xx xxx 2014.
For information on obtaining reprints of this article, please send e-mail to: bmtski@rit.edu

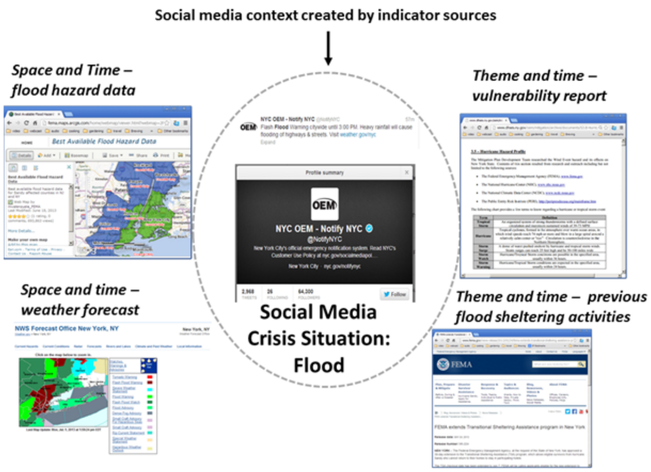

Fig. 1. Overall research concept. In the center of Figure 1, a tweet about a flood situation in New York City has been reported. To put the flood tweet in context, indicator sources have been retrieved based on space, time and thematic dimensions derived from the flood tweet. For example, at the bottom right, a FEMA report about previous flood sheltering activities that are related to the flood tweet in terms of theme (flood) and time (within the past six months) is shown. The bottom left shows how flood hazard data is related to a weather forecast and the top left how the flood tweet is related in terms of space (same location as the flood tweet) and time (recently available flood data). The top right shows how a flood vulnerability report is related to flood tweet in terms of theme (floods) and time (a recent report created after the last flooding incident).

FEMA daily situation reports were a particularly good source as they often contain numerous disaster types such as floods, tornadoes, wildfires and thus will have varying degrees of relevance to our flood scenario described previously and have been used in our previous disaster visual analytics research [14].

2.2.2 Twitter

As Twitter data are available on a massive scale, we used initial filtering mechanisms to remove potentially irrelevant data. Specifically, we used the keywords "hurricane", "flood" in Twitter as a filter string to capture approximately 3.4 million tweets with these terms or hashtags. MongoDB was used to store the retrieved tweets as JavaScript Object Notation (JSON) objects. Tweets collected were first pre-processed using common free-form text processing steps such as tokenization (generating individual terms from a bigger text string), stop word removal (removing words like "the" or "and"), and stemming (reducing different forms of a term into a common root form, for example "flooding" becomes "flood"). Other data cleaning steps were also run such as removing user tags, hash tags, special characters, numeric, alphanumeric and special words such as rt, ft, and timezone details (e.g., PST, MST, EST, CST).

2.2.3 Classifying and Querying Twitter Data with Machine Learning and Document Retrieval

We used machine learning algorithms to automate the process of separating semantically relevant tweets from others that only contained synthetically relevant terms but with different semantics. For example, "earthquake" and "shake" are both relevant terms to an earthquake disaster. Nonetheless, people may tweet about "attending a conference about earthquake research" or "shaking hands with some celebrity". In particular, we adopted a supervised learning strategy to construct a classifier from a small (400) set of training tweets to then automatically classify all remaining tweets as either flood or noise (i.e., those relevant to the floods are expected to convey disaster information while the remaining tweets are "noise"). We used the Support Vector Machine (SVM) algorithm to construct the classifier over the small training dataset as SVM achieves the best accuracy for classifying textual data [15], making it a suitable tweets classification choice. The top-k occurring terms within tweets assigned to the flood (i.e., non-noise) category were then used as query inputs for contextual document retrieval and ranking with Apache Lucene.

2.3 Gold Standard and Experiment Evaluation Metrics

A document ranking Gold Standard evaluation dataset was built from the 89 FEMA documents to find similarities and differences between document rankings done by the machine compared to those done by a human reviewing the same document set. Documents were ranked based on key pieces of information derived from the FEMA situation documents that were very useful for determining document relevancy. For example, clear mentioning of a flood in specific areas. As this was a small pilot study, documents were ranked by a single reviewer with fifteen years of disaster management experience and independently from the machine classifying aspects of this work to mitigate any bias. In addition to ordinal relevance ranks, documents were also assigned a code of 3 = very relevant, 2 = somewhat relevant, 1 = slightly relevant, or 0 = completely irrelevant to our scenario's hypothetical flood query needs. Of the 89 documents classified, 11 were assigned a 3 code, 34 were assigned a 2 code, 38 were assigned a 1 code and 5 were assigned a 0 code. These relevance codes were an important factor for our experiment evaluation metric - normalized discounted cumulative gain (nDCG). DCG works on the assumptions that (1) highly relevant documents are of greater use than marginally relevant documents and (2) the further down a relevant document is in a ranked list, the less likely it is to be examined and hence of less use [16]. DCG scores can be normalized to produce a [0..1] value. The normalized DCG value, referred to as nDCG is calculated by diving real DCG values by ideal DCG values, or DCG values derived if documents were ranked perfectly by relevance. nDCG was a particularly relevant metric to use as many of FEMA documents contain some, if even small, references to floods and would thus all be considered relevant (and misleading) if using metrics such as Average Precision and/or F-measure were used that cannot distinguish between "shades" of relevance vital to contextualization.

2.3.1 Preliminary Results – Document Ranking

We ran three machine document ranking queries. The first query used the top five terms derived from the machine twitter data classification, the second query used the top ten terms, and the third query used the top twenty terms. We varied the query term numbers to examine variation in document rankings based on the number of terms and to examine how good the machine classified twitter data was at deriving discriminating terms. The top 20 words were: - **flood flash warning hurricane issued july august nws watch advisory rain county relief people weather areal tl warnings heavy water.** All the terms were entered using an OR clause (i.e., "flood" OR "flash"). We calculated nDCG scores based on the top 20 documents returned by machine document retrieval based on our experience with search tools like Google that most people will only look for the first 10 (first page) and perhaps one additional results page. Query 1 used the first five terms resulting in an nDCG score of 0.68**;** Query 2 used the first ten terms resulting in an nDCG score of 0.68; Query 3 used all twenty terms resulting in a nDCG score of 0.71. Table 1 shows Query 3 document rankings. Query 3 proved to have the best NDCG score most likely due to the fact that the additional 10 terms such as "rain", "weather" and "warnings" provided good discriminating power for higher document ranking. Five out of eleven documents considered "very relevant" appeared within the top ten machine returned documents (even though they were not in the exact rank orders created by a human).

Table 1: Query 3 document rankings. Each row represents a single FEMA document - the machine rank column is the relevance rank given by Apache Lucene and the human rank column is the corresponding relevance rank of the document assigned by a human.

Machine Rank	Human Rank	Human Assigned Relevance
1	11	3 - very relevant
2	2	3 - very relevant
3	13	2 - somewhat relevant
4	3	3 - very relevant
5	21	2 - somewhat relevant
6	77	1 - slightly relevant
7	4	3 - very relevant
8	7	3 - very relevant
9	75	1 - slightly relevant
10	16	2 - somewhat relevant

3 Visual Interfaces

The following sections outline our ongoing visual interface developments to support analytical reasoning and contextualization with machine learning classified social media terms and machine-ranked indicator documents.

3.1 Classified Terms/ Ranked Documents Interface

We are utilizing a graph visual interface to support analytical reasoning about relationships between top-k query terms derived from machine learning classified tweets and Lucene-ranked documents returned based on machine term queries (Figure 2).

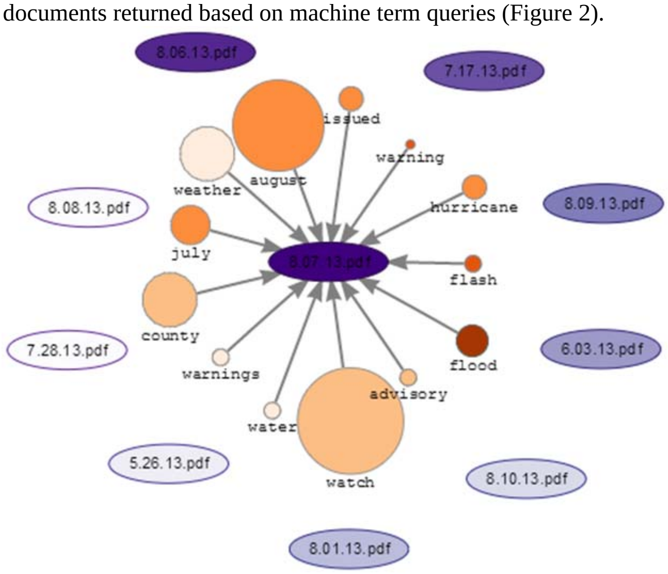

Fig. 2. Graph-based visual interface showing relationship between top machine-classified twitter query terms for a selected document (8.07.13.pdf) and the top 10 Lucene-ranked document.

We are using the vis.js (http://visjs.org/) open-source javascript library as it is a lightweight, scalable, easy to program visualization library. The machine classified twitter terms/lucene-ranked documents visual interface is currently using the following visual representation design strategy. The top 10 lucence ranked documents are added as elliptical graph nodes with the document file name added as a central label. Ranked document node visual representations use a single-hue scheme with changing lightness using ColorBrewer™[1] values and based on the document's term frequency–inverse document frequency (TF-IDF) Lucene score[2]. For example, in the center of Figure 2, 8.07.13.pdf is the top-ranked document as seen by the intense purple color, followed 8.06.13.pdf as the 2nd rank, 7.1.7.13.pdf as the 3rd ranked and so forth. Terms from the top 20 occurring terms assigned to the flood category from machine learning outputs (as discussed in section 2.2.3) are rendered as dot-shape graph nodes with the specific term added as label beneath the node. Term node size is proportional to the number of times the term was found inside a selected document. Term node colors represent the overall frequency by which the term was found from the machine classified tweets using a 5-class natural breaks classification with ColorBrewer™ single hue sequential colors. For example, in Figure 2, although the term "flood" is the most overall frequent term from the machine classified-tweets (as seen by the darkest orange color), it is not the most common term in the selected document (8.07.13.pdf). Rather, the term "watch" as seen by the largest dot-shape node is. Thus, an analyst can gain a dual sense of context by seeing the how the most frequently occurring terms compare with specific document terms, In the case of Figure 2, the term "watch" can alert an analyst to specifically look closer at hurricane or flood watches issued during the document's time period. Furthermore, graph edges are rendered between the machine classified query terms and documents containing the query terms with the idea that an analyst can contextualize the terms by seeing specific documents where the machine-classified terms occurred.

3.2 Geographic Term/Document Visual Interface

Given that very few twitter users reveal their geographic location via their user profile, to support geographical contextualization, we currently are examining machine-classified twitter term geographies using the following approach. First, we run parts-of-speech (POS) tagging on the entire group of unique terms machine-classified as flood using the GATE ANNIE[3] system with customized gazetteer lists for finding US counties (as these are common in FEMA reports). Locations found are sent to Geonames[4] for geocoding. The top five places associated with a location term, relevance ranked by population for potential location name disambiguation (i.e, multiple "Rochester" or "Springfield), are stored in XML and .CSV for ease of import into our mapping tools. We create a point-density representation of rank 1 locations extracted from tweets and compare those with manually geocoded locations found in the FEMA documents location (Figure 3). In Figure 3, rank 1 tweet locations are shown using a 9 class equal interval representation. Locations from three of the top machine ranked FEMA documents are shown as black square symbols. In its current state, an analyst can use the twitter-point density representations for making visual comparisons with locations found in FEMA documents. For example, the detail map in Figure 3 shows an area in Kansas, USA where several instances of flooding events are referenced in a FEMA report. These locations appear inside a Tweet point density area. Thus, the intent with the geographic term/document locations visual interface is that an analyst can visually explore and contextualize social media with the geographic context of locations derived from indicator documents. Our longer term goal is to publish result outputs from creating tweet point densities as RESTful map services for incorporation into scalable, light-weight, web-based mapping tools.

[1] http://colorbrewer2.org/
[2] See: http://lucene.apache.org/core/3_0_3/api/all/org/apache/lucene/search/Similarity.html for details
[3] http://gate.ac.uk/sale/tao/splitch6.html#chap:annie
[4] http://www.geonames.org/

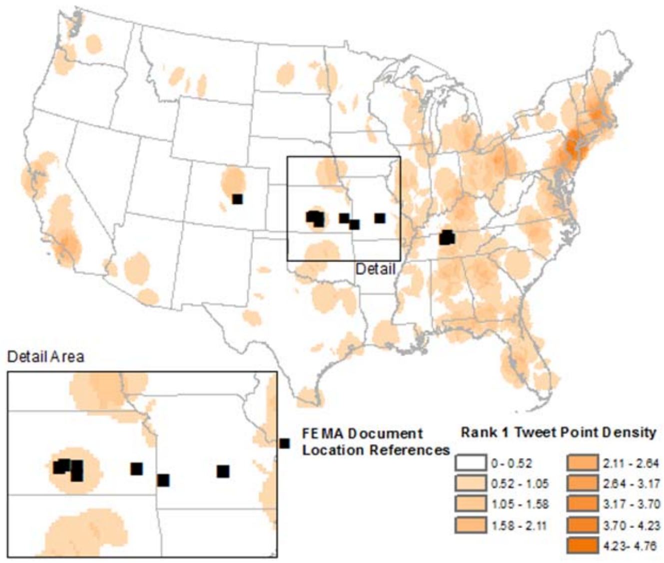

Fig 3. Visual interface for comparing Rank 1, machine classified tweet location point densities with locations referenced inside a small (3) sample of FEMA reports.

4 Future work

The primary future work area is to develop social media contextualization analytics and subsequent visual interfaces in the temporal dimension such as examining cyclical time for contextualizing the seasonality of natural disasters such as flooding and hurricanes. We also plan to address several limitations with our current approach, Namely, (1) improving the precision and accuracy of our information retrieval tools for improved nDCG scores through improved text engineering, (2) enhanced filtering options in the machine classified twitter terms/ranked documents graph visual interface, and (3) statistical clustering of twitter and indicator point locations for improved data analysis reliability and to move beyond frequency counts for improved spatial analysis rigor and analytical insight.

5 Summary and Conclusions

In this paper, we have presented our research in progress on examining how social media can be visually contextualized through indicator documents within space, time and thematic constraints. We presented results from a pilot experiment where we relevance ranked FEMA situation reports using query terms derived from machine learning-based classification of Tweets. We also presented initial results of developing visual interfaces to support analytical reasoning in thematic dimensions through a machine classified-ranked document graph interface and geographic dimensions through a tweet-point density, indicator document interface. Ideally with additional research the approach we are developing can be used by disaster management practitioners to harness the power of social media to make better-informed decisions during disasters by contextualizing social media with the vast variety of other data sources that are brought to bear during a disaster.

Acknowledgments

This research was supported by a research seed development grant from the B. Thomas Golisano College of Computing and Information Sciences at the Rochester Institute of Technology.

References

[1] Harvard Humanitarian Initiative, "Disaster Relief 2.0: The Future of Information Sharing in Humanitarian Emergencies," Washington, D.C. and Berkshire, UK2011.

[2] M. Mendoza, B. Poblete, and C. Castillo, "Twitter Under Crisis: Can we trust what we RT?," in *Proceedings of the First Workshop on Social Media Analytics (SOMA '10)*, 2010, pp. 71-79.

[3] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling, "Twitterstand: news in tweets," in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2009, pp. 42-51.

[4] National Research Council, Successful Response Starts With a Map: Improving Geospatial Support for Disaster Management. Washington, D.C.: National Academies Press, 2007.

[5] B. Tomaszewski, "Situation Awareness and Virtual Globes: Applications for Disaster Management," *Computers and Geosciences,* vol. 37, pp. 86-92, 2011.

[6] D. Peuquet, *Representations of Space and Time*: Guilford Press, 2002.

[7] J. J. Thomas and K. A. Cook, Illuminating the Path: The Research and Development Agenda for Visual Analytics. Los Alametos, CA: IEEE, 2005.

[8] G. Andrienko, N. Andrienko, P. Jankowski, D. Keim, M. Kraak, A. MacEachren, and S. Wrobel, "Geovisual analytics for spatial decision support: Setting the research agenda," *International Journal of Geographical Information Science,* vol. 21, pp. 839-858, 2007.

[9] B. Tomaszewski, A. C. Robinson, C. Weaver, M. Stryker, and A. M. MacEachren, "Geovisual Analytics and Crisis Management," in *Proceedings of the 4th International Information Systems for Crisis Response and Management (ISCRAM) Conference,* Delft, the Netherlands, 2007, pp. 173-179.

[10] B. Tomaszewski and A. MacEachren, "Geovisual Analytics to Support Crisis Management: Information Foraging for Geo-Historical Context," *Information Visualization,* vol. 11, pp. 339–359, 2012.

[11] A. M. MacEachren, A. Jaiswal, A. C. Robinson, S. Pezanowski, A. Savelyev, P. Mitra, X. Zhang, and J. Blanford, "Senseplace2: Geotwitter analytics support for situational awareness," in *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, 2011, pp. 181-190.

[12] G. Andrienko, N. Andrienko, U. Demsar, D. Dransch, J. Dykes, S. I. Fabrikant, M. Jern, M.-J. Kraak, H. Schumann, and C. Tominski, "Space, time and visual analytics," *International Journal of Geographical Information Science,* vol. 24, pp. 1577-1600, 2010.

[13] Federal Emergency Management Agency (FEMA). (2013, August 15, 2013). *What is Mitigation?* Available: http://www.fema.gov/what-mitigation

[14] B. Tomaszewski, C. Pan, P. Mitra, and A. M. MacEachren, "Facilitating Situation Assessment through GIR with Multi-scale Open Source Web Documents," in *Proceedings of the 4th ACM workshop on Geographical Information Retrieval (GIR '07),* Lisbon, Portugal, 2007, pp. 95-96.

[15] E. Leopold and J. Kindermann, "Text categorization with support vector machines. How to represent texts in input space?," *Machine Learning,* vol. 46, pp. 423-444, 2002.

[16] B. Croft, D. Metzler, and T. Strohman, *Search Engines: Information Retrieval in Practice*: Addison Wesley., 2009.