Visualizing Social Media Sentiment in Disaster Scenarios

Yafeng Lu
Arizona State University
lu.yafeng@asu.edu

Xia Hu
Arizona State University
xia.hu@asu.edu

Feng Wang
Arizona State University
fwang49@asu.edu

Shamanth Kumar
Arizona State University
shamanth.kumar@asu.edu

Huan Liu
Arizona State University
huan.liu@asu.edu

Ross Maciejewski
Arizona State University
rmacieje@asu.edu

ABSTRACT

Recently, social media, such as Twitter, has been successfully used as a proxy to gauge the impacts of disasters in real time. However, most previous analyses of social media during disaster response focus on the magnitude and location of social media discussion. In this work, we explore the impact that disasters have on the underlying sentiment of social media streams. During disasters, people may assume negative sentiments discussing lives lost and property damage, other people may assume encouraging responses to inspire and spread hope. Our goal is to explore the underlying trends in positive and negative sentiment with respect to disasters and geographically related sentiment. In this paper, we propose a novel visual analytics framework for sentiment visualization of geo-located Twitter data. The proposed framework consists of two components, sentiment modeling and geographic visualization. In particular, we provide an entropy-based metric to model sentiment contained in social media data. The extracted sentiment is further integrated into a visualization framework to explore the uncertainty of public opinion. We explored Ebola Twitter dataset to show how visual analytics techniques and sentiment modeling can reveal interesting patterns in disaster scenarios.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; H.5 [**Information Interfaces and Presentation**]: User Interfaces; I.2 [**Artificial Intelligence**]: Natural Language Processing

General Terms

Sentiment Analysis, Social Media Visual Analytics

1. INTRODUCTION

Social media data, encapsulating knowledge chunks about events and people's opinion, is sensitive to disasters and hu-

manitarian activities [21]. For instance, in an emergency situation [19], some users generate information either by providing first-person observations or by bringing relevant knowledge from external sources. Twitter, with its real-time nature, has been successfully used as a sensor of earthquakes [12] and wildfires [17]. Furthermore geo-located Twitter data has been shown to be a reliable source for detecting disasters and investigating response [7]. While social media mining has been widely used in different disaster scenarios, one of the most important aspects to understand social responses is to gauge people's opinion for improved disaster management [10, 13, 20].

To understand public sentiment during disasters, an accurate sentiment classifier is required. While sentiment analysis has been extensively studied for some domains, such as product reviews [8, 11], the performance on social media data is still unsatisfactory due to the distinct data characteristics [5, 6]. First, social media posts are always short and unstructured. For example, Twitter allows no more than 140 characters and uses many informal words such as "cooool" and "OMG" . The short texts can hardly provide sufficient statistical information for learning based models. Second, it is labor intensive and time consuming to obtain ground truth for training data, which is needed to build an effective supervised learning model. In this paper, we study this problem from a novel aspect with visual analytics.

Visual analytics is widely used in social media data analysis and contributes in many areas of exploratory data analysis, such as geographical analysis [2], information diffusion [22] and business prediction [9]. Besides showing the data intuitively, visual analytics enables users to navigate through the data, compare different metrics or datasets, and interactively explore patterns. In this paper, we propose a visual analytics framework to explore geo-located Twitter data in disaster scenarios specifically focusing on sentiment. This framework enables us to observe the distribution of Tweets, compare between positive and negative sentiment, and study the sentiment predictions from multiple models. The research questions motivating our visual analytics framework can be described as follows.

RQ1 How can we reveal disagreements among multiple sentiment classifiers?

RQ2 Does positive sentiment exist in a disaster scenario? If yes, can we compare the patterns between the distribution of positive and negative sentiment.
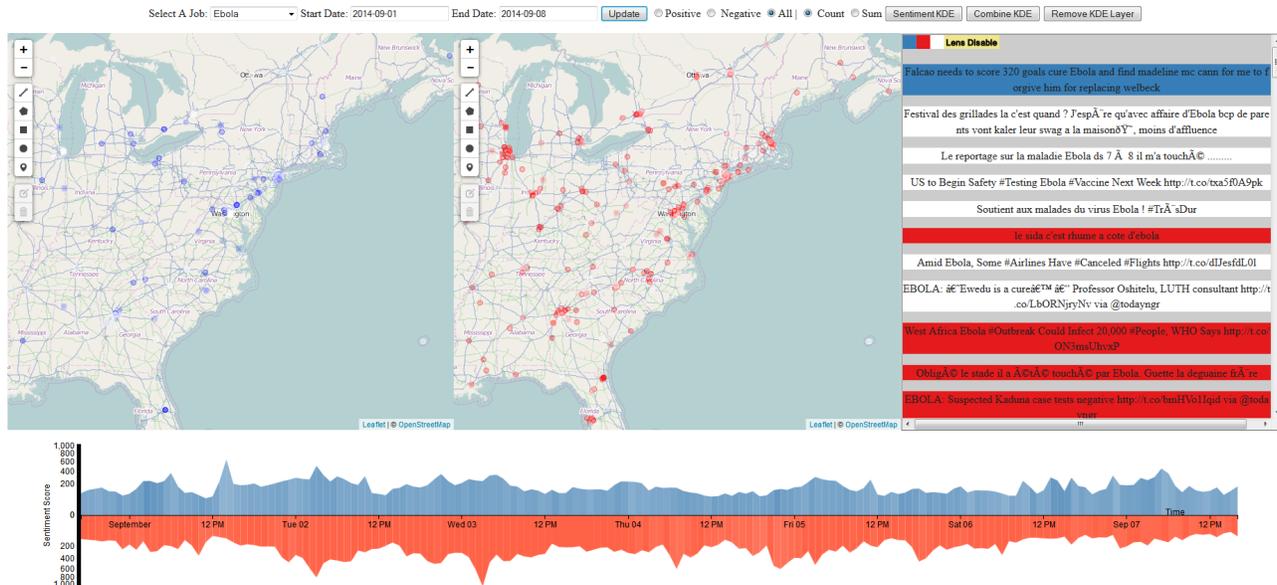
Figure 1: Sentiment analysis and visualization overview of sentiment analysis on our Ebola Twitter dataset. The two maps make up the geo-comparison view. The list on the right contains the top Tweets ordered by their retweet count. The bottom view shows our entropy sentiment river.

Table 1: Statistics of the ebola dataset

Classifier	#Positive	#Negative	#Neutral
CoreNLP	24089	529848	138630
SentiStrength	67506	247643	377418
SentiWordNet	39335	51243	601989
Committee Vote	25673	261500	405394

Table 2: Statistics of geo-located ebola dataset

Classifier	#Positive	#Negative	#Neutral
CoreNLP	849	10002	3474
SentiStrength	2447	4750	7128
SentiWordNet	1018	1512	11795
Committee Vote	935	5517	7873

2. MULTI-CLASSIFIER SENTIMENT ANALYSIS AND VISUALIZATION

The goal of our work is to develop a visual analytics framework for sentiment analysis on social media data relating to disasters, particularly Twitter data, so that users can find disagreements among multiple sentiment classifiers, observe the uncertainty of sentiment predictions, and investigate interesting sentiment distribution patterns, as well as compare between the distributions. To achieve this goal and answer RQ1 and RQ2, we propose a visual analytics framework consisting of an entropy-based sentiment model and a geographical visualization.

To test our sentiment analysis method and visual analytics framework, we carried out an experiment on Ebola Twitter dataset, which has been collected using the Twitter Search API with keyword "ebola". From September 1st to September 8th, this dataset has 567,015 Tweets among which 5,338 have geographical locations. The statistics of this dataset in regard to sentiment classifiers are shown in Table 1 (the whole dataset) and Table 2 (only geo-located Tweets).

2.1 Sentiment Modeling

The disagreement among multiple sentiment classifiers is shown from a primary study on the Ebola dataset using three well-know sentiment analysis classifiers [1, 16, 18].

Table 1 and Table 2 show the supporting statistical results. To answer the first research question (RQ1), we propose a metric to evaluate the inconsistency between sentiment classes and then use a committee vote method to decide on a Tweet's sentiment class.

Since entropy is a well defined metric for measuring the level of disagreement, we define an uncertainty measure using vote entropy [3] to gauge the disagreement among multiple classifiers. Our uncertainty is defined as:

$$UC = 1 - normalize(Max\_Entropy - Entropy) \quad (1)$$

where $Entropy$ is defined as follows.

$$Entropy = -\sum_{i=1}^{K} \frac{V(y_i)}{C} \log \frac{V(y_i)}{C} \quad (2)$$

Here $V(y_i)$ is the number of "votes" that a class (K_i) receives from among the committee members' prediction, K denotes the number of classes, and C is the committee size. $Max\_Entropy$ is the highest possible entropy given C and K. There are two situations, $C \leq K$ and $C > K$. When $C \leq K$, the entropy is maximized when no two votes go to the same class, thus $Max\_Entropy = \log C$. When $C > K$, the entropy is maximized when the difference of the number of votes between any two classes is no larger than 1. Without loss of generality, assume $C = Kt + d$ with t and d being

Figure 2: Glyph of sentiment representation for Tweets. The left most one is a Tweet with high uncertainty represented by a blurred circle (threshold = 0.6). Others are relatively certain Tweets classified into negative. The deep red area indicates multiple negative Tweets that overlap each other (opacity = 0.5).

positive integers and $d < K$, then the entropy is maximized when there are d labels having $t + 1$ votes for each while the other $K - d$ labels having t votes for each, and thus

$$Max\_Entropy = -\left(\sum_{i=1}^{d} \frac{t+1}{C} \log \frac{t+1}{C} + \sum_{i=d+1}^{K} \frac{t}{C} \log \frac{t}{C} \right).$$

Having uncertainty described with regard to entropy, we can reveal the disagreement among multiple classifiers. Previous works have shown that even a small committee can improve the performance of prediction in practice [4, 14]. In this paper,we take the majority of the committee's predictions as the final label to increase confidence.

Regarding RQ1, in our visualization design, we create a Tweet sentiment glyph and an entropy sentiment river to represent the uncertainty. On the map view, the sentiment of Tweets is labeled as the majority vote from the committee for a confident sentiment class representation. In our pilot experiment, we used the following three sentiment classifiers for our committee: SentiWordNet [1], SentiStrength [18] and CoreNLP [16]. SentiWordNet generates a decimal score from -1 to 1, with -1 being the most negative, 1 being the most positive and 0 being neutral. SentiStrength (trinary) assigns integer scores from -4 to 4 to each Tweet, with 0 being the neutral. And CoreNLP classifies each Tweet into 5 classes scored from 0 to 4, with 2 being neutral.

2.2 Visual Analytics Framework

To explore the underlying sentiment of our Twitter dataset, we have developed a visual analytics framework, in which our sentiment model is used to show the uncertainty of sentiment prediction among multiple classifiers and enable sentiment distribution analysis. The proposed framework consists of three linked views: the geo-comparison dual map, the top Tweets list, and the entropy sentiment river(Figure 1). The top Tweet list is linked with time and area selection to show the most popular Tweets.

The geo-comparison dual map is designed for displaying the geographical sentiment distribution of Tweets. It has two maps centering on the same region and displaying positive and negative Tweets synchronously. Tweets are displayed as translucent color coded circle glyphs (Figure 2) to show the sentiment, uncertainty and density. Positive sentiment is colored blue, negative sentiment is colored red,

and neutral sentiment is colored white. While setting opacity, the dense area can be identified by the deeper color. A Tweet whose UC is above a threshold will be represented by the blurred glyph.

Regarding RQ2, this view shows a kernel density estimation (KDE) map and implements a sentiment comparison lens. The sentiment KDE is obtained by first splitting the sample Tweets into positive and negative groups and then calculating the fixed bandwidth KDE [15].

$$\hat{f}_h(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{h^d} K\left(\frac{|\mathbf{x} - \mathbf{x_i}|}{h} \right) \tag{3}$$

Here, h is the bandwidth, d is the data dimension, in our case $d = 2$ for spatial data, N is the total number of samples. $|\mathbf{x} - \mathbf{x_i}|$ is the Sphere Mercator projection distance between two locations, and the kernel function is:

$$K(x) = \frac{2}{\pi}(1 - (x^2))I_{(x^2 \leq 1)} \tag{4}$$

where the indicator function $I_{(x^2 \leq 1)}$ is evaluated as 1 when $(x^2 \leq 1)$, and 0 otherwise.

When a user clicks on the "Combine KDE" button on the top-right side of the overview, a kernel density estimation based on positive Tweets and negative Tweets will be calculated and visualized on the dual map, as shown in the left two maps in Figure 3. The density map pair shows the distribution patterns of the sentiment, as well as the similar and different hot spots.

To enable quick density distribution comparison, we propose a novel sentiment comparison lens to show the contrast of a positive sentiment distribution and a negative sentiment distributions by alpha blending the images (the right map in Figure 3). In blending the KDE images, we blend the S(source) over the D(destination), e.g. positive KDE over negative KDE. The alpha blending algorithm used in our system can be described as:

$$O_A = S_A + D_A(1 - S_A)$$

$$O_{RGB} = \begin{cases} 0, & \text{if } O_A = 0 \\ (S_{RGB}S_A + D_{RGB}D_A(1 - S_A))/O_A, & \text{otherwise.} \end{cases}$$

In this expression, O is the output color, S is the source color and D is the destination color with subscript A representing the alpha channel and RGB representing the RGB color channel. Users can move the lens to investigate both the positive and negative sentiment on one map to find overlaps, exclusions, and differences in distribution patterns.

Our geo-comparison view also supports circle, rectangle and polygon selection by which only Tweets in the user defined area are displayed. This selection is linked with the top Tweets list so that the list will update to the most retweeted Tweets posted in the given time range from the selected area.

Our third view, the entropy sentiment river, is designed to reveal the uncertainty of sentiment classification over time. Additionally, when setting the class label by a single classifier, it shows the prediction bias of the classifier under analysis. The entropy sentiment river is developed based on sentiment river [9] by adding uncertainty information. In Figure 1, the bottom view shows the entropy sentiment river with blue representing the volume of positive Tweets and red representing the volume of negative Tweets. A lower opacity is used when the Tweets have a high average uncertainty in a particular time chunk. The volume in each polarity refers to
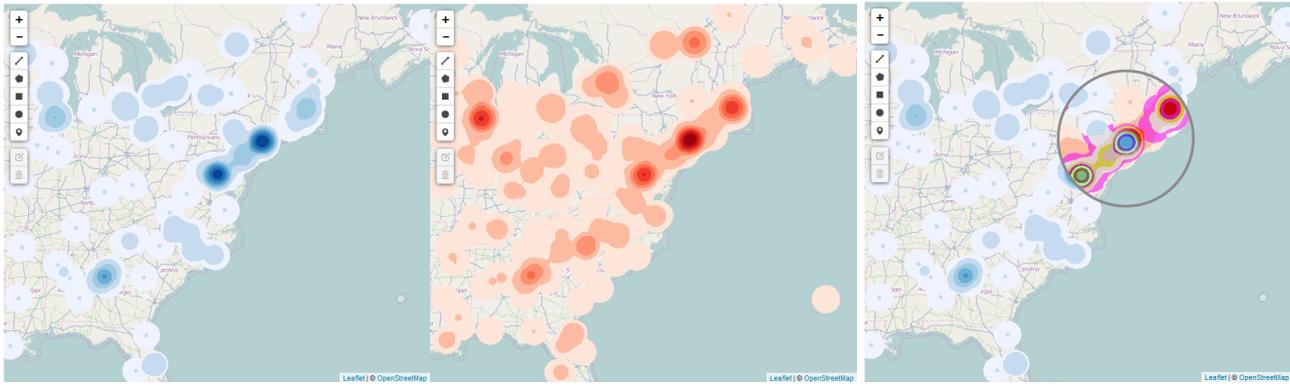
Figure 3: Geo-comparison view with kernel density estimation on positive and negative Tweet sentiment and the sentiment comparison lens blending negative sentiment distribution over positive sentiment distribution. This KDE is calculated using the Ebola dataset with fixed bandwidth of 55 miles.

SentiWordNet in Figure 1. It can change to other classifiers or the majority vote from a committee.

3. CASE STUDY - SENTIMENT IN EBOLA DATASET

In our experiment, we loaded the Ebola dataset, described in section 2. The geo-comparison dual map shows the positive Tweets vs. negative Tweets. It is obvious that negative Tweets (represented by red) has a higher volume than positive Tweets (represented by blue). However, from this view, users can find some positive hot spots in the disaster scenario with non-negligible magnitude. The sentiment label for this view is decided by the majority of the committee. In contrast, the entropy sentiment river uses a single classifier, SentiWordNet, in this example. From the entropy sentiment river, the magnitude of positive and negative sentiment trends similarly, which is different from the impression gained from the maps. It also shows many low opacity chunks along the river, especially on the positive side. This indicates that SentiWordNet is likely to provide a positive-biased label for the Ebola disaster dataset. This result can also be confirmed by evaluating the polarity proportion from Table 1 and Table 2. The inconsistency in the conveyed sentiment volume from the map and the entropy sentiment river and the uncertainty visualization both reveal the disagreement problem among multiple sentiment classifiers.

The geo-comparison dual map also shows that in general, the east coast cities, such as the areas around Washington DC and Boston, have more Tweets. For negative Tweets, we observe other possible hot spots, such as Chicago and Atlanta. To see the density distribution accurately, we generated the density maps shown in Figure 3. Now the magnitude is normalized and the distribution is more clear. From the density maps, we confirmed our hypothesis of Washington DC and Boston being hot spots of negative Tweets; however it shows that these two cities are the hot spots for positive Tweets too. Additionally, we can see that there is a slight hot spot around New York City. To compare these two density distributions, we used our sentiment comparison lens to look at the mixture of these two maps, and this is shown on the right map in Figure 3. Now with the lens, we can clearly see that the New York City hot spot

in the negative distribution is denser than the one in the positive distribution. By comparing with the Boston hot spots, Washington DC has a higher percentage of positive Tweets posted because the color code in the combined image contains more blue in that hot spot. From exploring the sentiment distribution patterns, we may assume that there are some positive opinions related to Ebola starting from Washington DC. This may provide a clue for the users to compare the effects of local activities in different places.

4. CONCLUSION AND FUTURE WORK

We have presented a sentiment visual analytics framework for social media. This framework consists of sentiment modeling and geographical visualization components, and answers our research questions. The uncertainty under multiple classifiers' prediction is measured by means of entropy and further visualized in the Tweets' glyphs on the map and the chunks of the entropy sentiment river. Through this visualization design, users are able to detect places with high or low sentiment confidence and the change of sentiment polarity and uncertainty. The sentiment distributions can be analyzed through our KDE maps, and compared via using the sentiment comparison lens. By analyzing the sentiment distribution, users can locate hot spots and reveal similarities and differences between the distributions. From our Ebola case study, we demonstrated the usage of the framework and explained the sentiment investigation.

Our future work includes (1) extending the committee vote method by involving more classifiers and evaluating the accuracy between voted prediction and single classifier's prediction, (2) generalizing our density map comparison lens to other differential metrics and evaluating the effects on users' perception, and (3) applying the comparison lens to other kind of data measures besides sentiment.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] S. Baccianella, A. Esuli, and F. Sebastiani. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the International Conference on Language Resources and Evaluation*, volume 10, pages 2200–2204, 2010.

[2] J. Chae, D. Thom, H. Bosch, Y. Jang, R. Maciejewski, D. S. Ebert, and T. Ertl. Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 143–152, 2012.

[3] I. Dagan and S. P. Engelson. Committee-based sampling for training probabilistic classifiers. In *ICML*, volume 95, pages 150–157, 1995.

[4] X. Hu, J. Tang, H. Gao, and H. Liu. Actnet: Active learning for networked texts in microblogging. In *SDM*, pages 306–314, 2013.

[5] X. Hu, J. Tang, H. Gao, and H. Liu. Unsupervised sentiment analysis with emotional signals. In *Proceedings of the 22nd international conference on World Wide Web*, pages 607–618, 2013.

[6] X. Hu, L. Tang, J. Tang, and H. Liu. Exploiting social relations for sentiment analysis in microblogging. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 537–546, 2013.

[7] S. Kumar, X. Hu, and H. Liu. A behavior analytics approach to identifying tweets from crisis regions. In *Proceedings of the 25th ACM conference on Hypertext and social media*, pages 255–260, 2014.

[8] B. Liu. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167, 2012.

[9] Y. Lu, F. Wang, and R. Maciejewski. Business intelligence from social media: A study from the vast box office challenge. *IEEE Computer Graphics and Applications*, pages 58–69, 2014.

[10] B. Mandel, A. Culotta, J. Boulahanis, D. Stark, B. Lewis, and J. Rodrigue. A demographic analysis of online sentiment during hurricane irene. In *Proceedings of the Second Workshop on Language in Social Media*, LSM '12, pages 27–36, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.

[11] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008.

[12] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, pages 851–860, 2010.

[13] A. Schulz, T. Thanh, H. Paulheim, and I. Schweizer. A fine-grained sentiment analysis approach for detecting crisis related microposts. *ISCRAM 2013*, 2013.

[14] B. Settles. Active learning literature survey. *Computer Sciences Technical Report 1648, University of Wisconsin, Madison*, 52:55–66, 2010.

[15] B. W. Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.

[16] R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1631–1642, 2013.

[17] J. Sutton, L. Palen, and I. Shklovski. Backchannels on the front lines: Emergent uses of social media in the 2007 southern california wildfires. In *Proceedings of the 5th International ISCRAM Conference*, pages 624–632. Washington, DC, 2008.

[18] M. Thelwall, K. Buckley, and G. Paltoglou. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173, 2012.

[19] S. Vieweg. Microblogged contributions to the emergency arena: Discovery, interpretation and implications. *Computer Supported Collaborative Work*, pages 515–516, 2010.

[20] B. Vo and N. Collier. Twitter emotion analysis in earthquake situations. *International Journal of Computational Linguistics and Applications*, 4(1):159–173, 2013.

[21] D. Yates and S. Paquette. Emergency knowledge management and social media technologies: A case study of the 2010 haitian earthquake. *International Journal of Information Management*, 31(1):6–13, 2011.

[22] J. Zhao, N. Cao, Z. Wen, Y. Song, Y.-R. Lin, and C. Collins. # fluxflow: Visual analysis of anomalous information spreading on social media. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2014.