Discover Trends in Public Emotion Using Social Sensing

Maryam Hasan, Elke Rundensteiner, Xiangnan Kong AND Emmanuel Agu
Worcester Polytechnic Institute

Social networks, such as Twitter and Facebook, are increasingly used by individuals to share their opinions and feelings on current issues and events in the form of text messages. This results in massive amounts of text stream data rich with emotional content. Such data provides a great opportunity for identifying and analyzing people's emotions in response to various public events, such as epidemics, terrorist attacks and political elections. However, the high volume and fast pace of social data make it challenging to analyze public emotions in social networks in real-time. In this paper we propose an online method to measure public emotion and detect emotion-intensive moments during real-life events. We first classify emotions expressed in text stream messages using a supervised learning approach. Then we aggregate each emotion class to discover emotion-evolving patterns over time and detect emotion-intensive moments. Our emotion analysis method is shown to present a fast and robust approach of analyzing online streams of text messages.

DOI: 10.1145/3065953.3065955 http://doi.acm.org/10.1145/3065953.3065955

1. INTRODUCTION

The massive streams of text messages in social networks such as Twitter often contain a rich diversity of emotions. This makes them valuable data sources for behavioral studies, especially for studying emotions of individuals as well as the general crowd. Moreover, social networks have been used as one of the prevalent communication channels for spreading news. They allow users to share their opinions and feelings about events (e.g., sport games, political elections, or social issues) as they occur. For example, the Ebola virus epidemic burst in West Africa in late 2013 were extensively reported by Twitter users. Analyzing these events can provide valuable information about reactions and emotions of people regarding the events reaching beyond what is feasible using traditional media only. The growth of social networks such as Twitter now empowers us to identify the influence of a social event on a large group of people in near real-time.

Studying public emotion promises to be of great value in different fields including social science, political science, public health research, and market research, which are interested in aggregate emotion instead of individual cases. It could assist government agencies in recognizing growing public fear or anger associated with a particular decision or event or in helping them to understand the public's emotional response toward controversial issues or international affairs. Public emotion analysis can aid public health researchers by providing them with (1) a low-cost method to measure potential risk across different sub populations; (2) useful knowledge for identifying at-risk populations; and (3) a method to formulate new hypotheses about the impact of real-time events on populations.

We study public emotion through analyzing emotion trends driven by external events. We first classify emotions expressed in text stream messages using a supervised learning approach [Hasan et al. 2014]. Then we aggregate each emotion class to discover emotion-evolving patterns over time and detect emotion-intensive moments using an online method [Hasan et al. 2017]. Our methodology is shown to present a fast and robust way of analyzing online streams of text messages.

2. PROPOSED METHOD TO MEASURE PUBLIC EMOTION AND DETECT IMPORTANT MOMENTS IN LIVE STREAMS OF TEXT MESSAGES

Instead of simply detecting the frequency of similar messages, we are looking for the percentage of people in a geographic location experiencing certain emotions and correlate this with the current event. The goal is to explore temporal distributions of aggregate emotion during events and detect temporal bursts in public emotion. For this purpose, we first apply our Emotex system [Hasan et al. 2014] to automatically detect the emotion of people from their text messages. It learns an emotion classification model from a large dataset of emotion-labeled messages. Then, the model is deployed to classify the live streams of text messages posted about an event using our EmotexStream framework [Hasan et al. 2017]. Then we aggregate text messages at the emotion class level to analyze public emotion trends driven by social events and discover emotion-evolving patterns over time. We propose an online approach to measure public emotion and predict emotion-intensive moments during social events. Our approach is able to analyze real-life events, in spite of noise, linguistic diversity, and the fast-evolving nature of live streams of tweets.

Let $e_1, \cdots e_i, \cdots e_n$ denote the emotion class E_{c1} of the tweets posted within a temporal window of length W in the tweet stream (n is the number of tweets posted within W). We can model $e_1, \cdots e_i, \cdots e_n$ as independent 0-1 random variables (e_i=0 means tweet message M_i doesn't belong to the emotion class E_{c1}, and e_i=1 means tweet message M_i belongs to the emotion class E_{c1}). In order to estimate the value of a specific emotion class E_{c1} among the people in a geographic location L within the temporal window W, we define a function as described below:

$$E_{public}(T_c - W, T_c, L, E_{c1}) = \sum_{i=1\dots n} F(M_i, E_{c1}) \tag{1}$$

where T_c is the current time and n is the number of tweets posted within window W, M_i is a tweet message, $E_{c1} \in E_{Class}$, and $F(M_i, E_{c1})$ is an indicator function of E_{c1} defined as described below:

$$F(M_i, E_{c1}) = \begin{cases} 1 & \text{if } M_i \in E_{c1} \\ 0 & \text{O}therwise. \end{cases} \tag{2}$$

During real life events, we can track public emotion to detect temporal bursts in public emotion. These sudden bursts are characterized by a change in the fractional presence of messages in particular emotion classes. Formally, we define such changes as "emotion bursts", which can point towards important moments during events.

DEFINITION 1 EMOTION BURST. *An emotion burst over a temporal window of length W at the current time T_c is said to have occurred in a geographic region L, if the presence of a specific class emotion E_{c1} during a time period $(T_c - W, T_c)$ is less than the lower threshold α or greater than the upper threshold β.*

In other words, we should have either

$$E_{public}(T_c - W, T_c, L, E_{c1}) \leq \alpha \tag{3}$$

or

$$E_{public}(T_c - W, T_c, L, E_{c1}) \geq \beta. \tag{4}$$

Now we need to define the upper bound α and lower bound β of public emotion for each emotion class during a temporal window. If our algorithm is applied offline (i.e., tweets of the entire event are available), the thresholds for the entire event can be estimated from the average sum of the overall event duration. However in the online approach the tweets of the entire event are not available. Therefore, in the online approach, we compute the thresholds from the tweets in a temporal sliding window, where the size of the moving window is a parameter.

As we know Hoeffding's inequality [Hoeffding 1963] provides an upper bound on the probability that the sum of random variables deviates $\lambda > 0$ from its expected value as shown by Equation 5:

$$Pr[|X - \mu| >= \lambda] <= 2e^{-2\lambda^2/n} \tag{5}$$

where X is the sum of independent random variables $X_1, X_2, \cdots, X_n$, with $E[X_i] = p_i$, and the expected value $E[X] = \sum\limits_{i=1...n} p_i = \mu$.

We can use Hoeffding's inequality to define an upper bound on the probability that the public emotion E_{c1} deviates from its expected value. Using the Hoeffding bound, for any small $\lambda > 0$ we have:

$$Pr[|E_{public}(T_c - W, T_c, L, E_{c1}) - \mu_e| >= \lambda] <= 2e^{-2\lambda^2/n} \tag{6}$$

where μ_e is the expected number of tweets that belong to the emotion class E_{c1} in window W and n is the number of tweets posted within W. Given that n is large in a stream of tweet, the emotion class E_{c1} can be approximated using a normal distribution:

$$\mu_e = n \times P_e$$

where P_e is the expected rate of the emotion class E_{c1}.

We use the historical average rate of each emotion class as expected rate P_e for that emotion class. For example, a weekly window can be used to average the rate of each emotion class based on all tweets in general. Therefore, other than a sliding detection window over the recent tweets posted about the event, we also utilize a larger reference window to summarize the past information about the tweets posted in general. In fact, our emotion-burst detection methodology utilizes two sliding windows. One small window W_{event} that keeps the rate of each emotion class based on the most recent tweets posted about the event. Another large reference window $W_{general}$ that keeps the average rate of each emotion class based on all the past tweets posted in general.

3. EXPERIMENTAL RESULTS USING REAL-LIFE EVENTS

We select the death of Eric Garner in New York [1] which stirred public protests and rallies with charges of police brutality as a social event with significant concern to the public for our case study. We utilize the Twitter search API to search for tweets containing the specified set of hashtags. We collected 4K tweets containing the hashtag "Garner" from November 24 2015 until January 5 2015. After collecting tweets we classify them using our model [Hasan et al. 2017]. Then, the emotion-classified tweets are aggregated into a daily-based histogram. Finally, using our methodology described in Section 2 we analyze public emotion and detect emotion-critical moments.

Figure 1 presents the temporal changes of different classes of emotion in New York during the selected event. The important moments of this event are also indicated in the figure. The distribution shows a predominance of sad and angry emotions over happy emotion in many days during the event. In order to predict the important moments as emotion bursts, we apply a sliding window W_{event} of length one day over the emotion stream of tweets aggregated in daily bins, as described in Section 2. Also a reference weekly window $W_{general}$ is applied over the general stream of tweets to calculate the average rate of each emotion class. Then, we continuously monitor the frequency rate $E_{public}(Tc - W_{event}, Tc, L, E_{c1})$ over time for each emotion class E_{c1}. Whenever this rate for an emotion class exceeds the upper threshold of β or falls beneath the lower limit α, an emotion burst is reported. Table I presents the days of abrupt changes in happiness. The second row shows the frequency rate of emotion bursts which are out of range. The last row shows the low and high boundaries. Comparing the results of this table with the important moments specified in Figure 1 confirms that our method is able to detect emotion-critical moments.

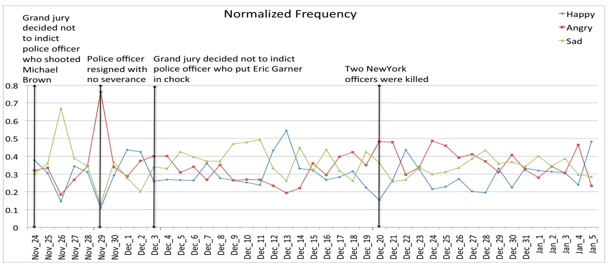

Fig. 1. Changes of emotions about selected sad events in New York

Date	Nov 26	Nov 29	Dec 19	Dec 20	Dec 27	Dec 28	Dec 30
Happy Rate	210	175	576	462	463	360	503
Boundary(α, β)	(360,936)	(400,1040)	(641,1668)	(753,1957)	(573,11491)	(461,1199)	(561,1459)

Table I. Detected burst changes in happiness

4. CONCLUSION

We analyze public emotion trends driven by social events and investigate its temporal distributions using microblogs on twitter. We propose an online approach to measure public emotion and detect important moments during social events. We deploy our approach in live stream of tweets to predict public emotion during real-life events. From the daily

[1] https://en.wikipedia.org/wiki/Death_of_Eric_Garner

tweets we were able to observe interesting temporal changes in public positive and negative emotion and also identified major moments when public emotional tweets are intensive.

REFERENCES

HASAN, M., RUNDENSTEINER, E., AND AGU, E. 2014. Emotex: Detecting emotions in twitter messages. In *Proceedings of the Sixth ASE International Conference on Social Computing (SocialCom 2014)*. Academy of Science and Engineering (ASE), USA.

HASAN, M., RUNDENSTEINER, E., KONG, X., AND AGU, E. 2017. Using social sensing to discover trends in public emotion. In *Proceedings of the 11th IEEE International Conference on Semantic Computing (ICSC) 2017*. IEEE.

HOEFFDING, W. 1963. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association 58,* 301, 13–30.