EmoTrend: Emotion Trends for Events

Yi-Shin Chen, Carlos Argueta, and Chun-Hao Chang(✉)

Intelligent Data Engineering and Applications Laboratory, Institute of Information Systems and Applications, National Tsing Hua University , No. 101, Section 2, Kuang-Fu Road, Hsinchu 30013 , Taiwan, Republic of China
ccha97u@gmail.com

Abstract. In this demo paper we present EmoTrend, a web-based system that supports event-centric temporal analytics of the global mood, as expressed in Twitter. Given a time range, and optionally a set of keywords, the system relies on peak frequencies, and the social graph, to identify relevant events. Subsequently, by performing sentiment analysis on related tweets, the global impact and reception of the events are presented by a visualization of the overall mood trend in the time range.

1 Introduction

The widespread use of connected portable devices has allowed users to spread news as they happen, making microblog services the most rapid news outlets in the world. In this demo paper, we present EmoTrend, a system that automatically identifies events from Twitter streams and summarizes the impact on society in a meaningful way. The system first clusters keywords with frequency peaks as candidate event descriptors. It then uses evolving social graphs to identify meaningful events. Batches of tweets related to the detected events are sentiment-analyzed based on emotion-bearing patterns. Finally, the user is presented with a timeline of events. By selecting an event, the evolution through time of the global mood towards the event is summarized based on six basic emotions—anger, fear, hope, joy, sadness, surprise. Such tool can help a user to better understand how an event has evolved and how it has impacted society.

2 Methodology

The system presented in this demo paper has two major components: the Event Detection component and the Mood Summarization component.

2.1 Event Detection

Good keywords in tweets must satisfy two special criteria: meaningfulness and burstiness. To detect meaningful keywords, profanity is first filtered out. Then, an adaptation of the m-function used by the Porter Stemmer [1] is used to remove words without meaning. To determine whether or not a keyword is bursty

M. Renz et al. (Eds.): DASFAA 2015, Part II, LNCS 9050, pp. 522–525, 2015.
DOI: 10.1007/978-3-319-18123-3_32

(i.e. it has been frequently used during a specific period of time), we calculate its frequency inside a sliding window. Using z-scores, we obtain a probability statistic determining how likely the word frequency in the past window is less than or equal to the frequency in the current timeframe. If that value is big enough, then the word is considered bursty and with high temporal use.

The filtered keywords are then grouped into event candidates. To achieve this, an event graph is constructed for each time frame. The vertices correspond to keywords and each edge connects co-occurrent keywords within a same tweet. A weight for each directed edge is computed based on frequency and co-occurrence statistics. PageRank [2] is then used to rank the keywords. The top keywords and their strongest neighbors are grouped as event candidates.

Concept-Based Evolving Graph Sequences (cEGS) [3] is a sequence of directed graphs representing information propagation within social streams. In this system, one such sequence of graphs is built for each event candidate that is monitored. Given a cEGS for a specific event, a directed graph is built for every day, with its vertices being the users that mention one or more event keywords on that day, and its edges representing a "following" relationship between two users.

2.2 Mood Summarization

After the events in the time frame have been identified, related tweets are fed, in order of publication, to the sentiment analyzer. Using emotion-bearing patterns, the analyzer summarizes the overall mood of the global community towards the event, and how it evolves in time. The steps performed are described in the following subsections.

By monitoring word frequencies in tweets, the system separates words into high-frequency (HW) and low-frequency words. Infrequent words that appear in a dictionary obtained from LIWC's [4] [5] psychological categories are deemed as psychological-words (PW). Subsequences of words pertaining to a combination of HW (e.g. "this") and PW (e.g. "hate", "love", "beach"), and appearing frequently in tweets, are grouped together based on matching HWs (e.g. "hate this weather" and "love this beach"). Subsequently, by having their PWs replaced by a wildcard (e.g. ".+ this .+"), the subsequences form emotion-bearing patterns. Patterns that fall below a frequency threshold are discarded.

An adaptation of the term frequency-inverse document frequency (tf-idf) statistic is used. By viewing patterns as terms, and collections of tweets per emotion as documents (6 documents total), the tf-idf is adapted to include a third score based on how many PWs in a collection can be captured by a pattern with its wildcard. The result is one ranking per emotion class, where top patterns are both more relevant to the class and bear a high level of emotion.

For each tweet fed to the sentiment analyzer, two different classifiers are used, a bag-of-words style classifier, and a Neural Networks classifier. The top two emotions obtained from the classifiers are used for each tweet. Finally, the overall mood state towards an event is computed for every day based on the individual emotions expressed in tweets related to that event in specific days.

3 Demonstration Overview

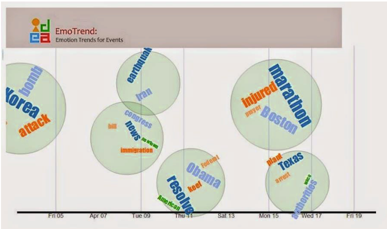

Fig. 1. Main page of the system The Timeline

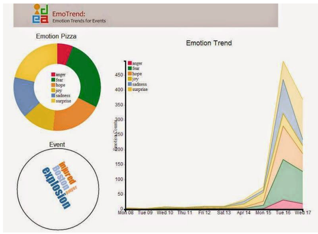

Fig. 2. Detailed view for event Boston Marathon Bombing

In this section the functionality of the proposed system is at http://www.cs.ccu.edu.tw/~ccha97u/emotrend/. When opening the system's main page, the user will be presented with the screen in Figure 1. The screen presents a timeline showing the different detected events. Scrolling horizontally on the timeline the user can move forward or backward in time. As the user moves, the timeline will update to show the corresponding events. Each event is represented by a circle which extends across the timeline according to its duration. Inside the circle the user is presented with the most important keywords representing the event. As can be seen in Figure 1, the event for the Marathon bombing contains relevant keywords, such as *marathon*. The size of the font for each keyword is proportional to the frequency with which it was used in related tweets. Clicking a circle brings the detailed view for the specific event.

The detailed view for each event has different components. In this demo scenario let's assume a user clicks on the circle representing the Boston Marathon bombing event. The detailed view, presented in Figure 2, shows three main components. The Emotion Pizza, like a regular pie chart, presents the proportion of tweets expressing each of the six emotions detected by the emotion detection algorithm. The Event circle is the same circle representing the event in the timeline and contains the most important keywords for the event. The Emotion Trend is an area chart used to represent the emotion trend over time.

The detailed view allows a user to inspect the impact of the event in society by observing how the global emotion evolves. Considering the Boston bombing event, from this view it can be seen at a glance that the predominant emotions are fear and sadness as direct consequences of such a tragedy, but also hope as expressed in tweets wishing the survivors to recover soon.

Additionally, the interface allows the user to restrict the summary to a specific emotion or subinterval of time. For instance, by clicking on an emotion name, or corresponding portion in the Emotion Pizza, the Emotion Trend will show only the area over time for that specific emotion. The resulting peaks give the user an idea of when the event caused specific emotional reactions on the society. Clicking on other emotions disable or enable them, updating the Emotion Trend accordingly. Additionally, if the user clicks and drags the mouse along the Emotion Trend graph, the Emotion Pizza gets updated to show the proportions of tweets with corresponding emotions within the selected time interval only. This is particularly useful to better understand how the emotions evolved by observing the predominant reactions from people at specific times.

4 Conclusions

In this paper we presented EmoTrend, a system that provides temporal summarization of the global mood towards interesting events mined from Twitter. For the regular user, EmoTrend constitutes an interesting tool to understand how a society is affected, and its mood evolves, during and after events happen (e.g. natural disasters, royal weddings). For corporations, politicians, and stars, EmoTrend represents a clear chance to grasp the overall reception of their events as they happen (product releases, campaign speeches), giving the possibility of better decision-making in order to improve image and products. We have demonstrated the ability of our system to extract events, and summarize their impact in a comprehensive way, an ability that traditional news articles and encyclopedia entries lack.

References

1. Porter, M.F.: An algorithm for suffix stripping. Program: Electronic Library and Information Systems 14(3), 130–137 (1980)
2. Page, L., et al.: The PageRank citation ranking: Bringing order to the web (1999)
3. Kwan, E., et al.: Event identification for social streams using keyword-based evolving graph sequences. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. ACM (2013); APA
4. Pennebaker, J.W., et al.: The development and psychometric properties of LIWC 2007, Austin, TX, LIWC. Net (2007)
5. Tausczik, Y.R., Pennebaker, J.W.: The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and Social Psychology **29**(1), 24–54 (2010)