What Causes Different Emotion Distributions of a Hot Event? A Deep Event-Emotion Analysis System on Microblogs

Yanyan Zhao[1(✉)], Bing Qin[2], Zhenjiang Dong[3], Hong Chen[3], and Ting Liu[2]

[1] Department of Media Technology and Art, Harbin Institute of Technology, Harbin, China
yyzhao@ir.hit.edu.cn
[2] Department of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
{bqin,tliu}@ir.hit.edu.cn
[3] ZTE Corporation, Shenzhen, China
{dong.zhenjiang,chen.hong}@zte.com.cn

Abstract. Current online public opinion analysis systems can explore lots of hot events and present the public emotion distribution for each event, which are useful for the governments and companies. However, the public emotion distributions are just the shallow analysis of the hot events, more and more people want to know the hidden causation behind the emotion distributions. Thus, this paper presents a deep Event-Emotion analysis system on Microblogs to reveal what causes different emotions of a hot event. We here use several related sub-events to describe a hot event in different perspectives, accordingly these sub-events combined with their different emotion distributions can be used to explain the total emotion distribution of a hot event. Experiments on 15 hot events show that the above idea is reasonable to exploit the emotion causation and can help people better understand the evolution of the hot event. Furthermore, this deep Event-Emotion analysis system also tracks the amount treads and emotion treads of the hot event, and presents the deep analysis based on the user profile.

Keywords: Emotion causation · Emotion distribution · Hot event · Event-emotion analysis · Microblog

1 Introduction

Nowadays, the social media (e.g., Twitter, Facebook, Sina Microblog, Dianping) significantly influence our life. A large number of users are involved in the activities on the social media, such as posting a tweet or image, commenting on a hot event and so on. Thus the Web contains a considerable amount of user-generated content that describes the opinions of customers regarding the events in the form of reviews, blogs, tweets, and so on. These reviews are valuable for

J. Li et al. (Eds.): NLPCC 2015, LNAI 9362, pp. 453–464, 2015.
DOI: 10.1007/978-3-319-25207-0_42

assisting governments in social management, helping customers in making purchasing decisions and for guiding companies in pursuing their business activities. However, browsing the extensive collection of available reviews to search for useful information is a time-consuming and tedious task. Therefore, automatically analyzing and processing the reviews, such as detecting and tracking the hot event, analyzing the public emotion distributions of a hot event, is necessary.

To date, many online public opinion analysis systems are built to exploit the hot events that people focus on and to present the public emotion distributions for each hot event. For example, a search-based demonstration system, called MemeTube [7], is presented to summarize the sentiments of microblog messages in an audiovisual manner. C-Feel-It [6] is another system which can categorizes tweets pertaining to a search string as positive, negative or objective and gives an aggregate sentiment score that represents a sentiment snapshot for a search string. Social Event Radar [5] is a new social networking-based service platform, that aim to alert as well as monitor any merchandise flaws, food-safety related issues, unexpected eruption of diseases or campaign issues towards to the Government.

Although so many applications can provide the hot events and their related emotion distributions, there is no system showing the reasons what causes the emotion distributions. Apparently, exploiting the emotion causations hidden behind the hot event is more valuable and important for the governments and companies, compraring to just exploring the emotion distributions of a hot event. For example, Figure 1 is an example representation of the emotion distributions for a hot event "埃博拉病毒" (Ebola virus). Many people want to know why some people are "喜悦" (happy) about "Ebola virus" which people actually feel fear and sad about. In this paper, we propose a new and important task, **emotion causation analysis**, which aims to reveal the causations that are hidden behind the total emotion distribution (such as the distribution in Figure 1) of a hot event.

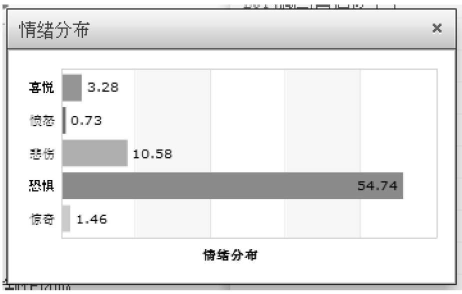

Fig. 1. Example of the emotion distribution for a hot event "埃博拉病毒" (Ebola virus).

What can be used to represent the emotion causation is the key question for emotion causation analysis. We use the sub-events extracted from the Hashtags of the hot event to represent the emotion causation. For example, we can use the sub-event "埃博拉病毒出现" ("Appearance of Ebola virus") to explain why so many people feel fear, and use the sub-event "疫苗研制成功" ("success of vaccine study") to explain why there is "happy" emotion appearing in the emotion distribution. To extract the sub-events of the hot event, a hierarchical clustering method is applied. Then, we can classify all the hot event related microblogs into these sub-events. The extracted sub-events combined with their new emotion distributions can be used to reveal the reasons that causes the total emotion distribution of the hot event. For example, the two different emotion distributions in Figure 2 corresponds to two sub-events of the hot event "Ebola virus". The left one is the "Appearance of Ebola virus" sub-event, in which the most prominent emotion is "fear". Thus this sub-event can explain the "fear" emotion in the total emotion distributions in Figure 1. The right one in Figure 2(b) is the "success of vaccine study" sub-event, in which the most prominent emotion is "happy". So this sub-event can explain why there is "happy" emotion for the event "Ebola virus".

Fig. 2. Different emotion distributions for two sub-event of the hot event "埃博拉病毒"(Ebola virus).

We analyze 15 hot events (e.g., "MERS virus", "QingAn shooting"), including more than one million microblogs. We observe that using the sub-events and their emotion distributions is a good way to exploit the hidden causation of the public emotion distributions of a hot event. Experiments on the 15 hot events show that the hierarchical clustering based sub-event extraction method can achieve an accuracy of 66.7%, which can reveal most of the reasons. Besides, the deep Event-Emotion analysis system also tracks the hot events over time and displays various results.

The remainder of this paper is organized as follows. Section 2 introduces the framework of the Event-Emotion analysis system. Section 3 presents the experiments and results. Section 4 introduces the related work. Finally we conclude this paper in Section 5.

2 Framework of Event-Emotion Analysis System

Event-Emotion analysis system can be divided into several steps, namely, microblog fetcher, hot event detection, emotion predictor, emotion causation analysis, other analysis result display. Since microblog fetcher, hot event detection and emotion predictor are the techniques that are usually used in the related systems shown in Section 4, we will briefly introduce them. Specially, emotion causation analysis related steps, including emotion causation detection and causation related microblog collection, are new proposed tasks. We will introduce them in detail.

2.1 Microblog Fetcher

We crawl and store about 120 messages each second because the Sina Weibo API allows only the retrieval of a subset of messages. Thus we can store about 4,000,000 messages each day. The format for each microblog we record is shown in Table 2. We mainly use the microblog contents to exploit the hot events and their emotion distributions. The other information, such as the “User ID” or the “Sex”, can be summarized as the user profile, which can be used to display various analysis results. For example, we can use the sex information to observe the different emotion distributions towards a hot event when the users are all female or male.

Table 1. Microblog format.

User ID	Sex	Post Time	Province	Location	Microblog Content
2054436031	男	Tue Apr 30 00:01:29 CST 2013	黑龙江	哈尔滨	为什么受伤的总是我

2.2 Hot Event Detection

Because of the characteristics of microblogging, detecting hot event in social media is different from detecting them from news, which has been studied in the previous work [1,3]. Messages that report such event are usually teemed with meaningless “babbles.” Moreover, event detection algorithm should be scalable given the sheer amount of messages [13]. Considering the above works, the task of hot event detection in this study contains the following steps:

- **Event Detection:** Considering that hashtags can cover almost all events, hashtags that appear in the messages can be extracted after a filtering model as the event set. The biggest advantage of using hashtags as events is that hashtags themselves are perfectly organized, topic-related phrases, which are short, simple, and easy-to-understand. They are much more easy and accurate than the phrases extracted from the main body of the messages.
- **Event Clustering:** The events (hashtags) are sponsored by different users. Thus, some events describe the same issues. Event clustering aims to solve this problem by clustering the events into different clusters. And each cluster is factually a meaningful event.
- **Event Popularity Ranking:** Each event cluster has different popularity. To exploit the hot events, the event clusters should be ranked according to their popularity. Then we can extract the top events as the hot events.

2.3 Emotion Predictor

Five kinds of emotions are used to tag the sentiments of the messages, namely, "happy," "sad," "surprise," "angry," and "fear". This task aims to classify each message into one of the five emotions. Many researchers focused on the emotion classification task [9], which is a typical task in the sentiment analysis.

Among all the algorithms on this task, machine learning based method is the most effective one. In our Event-Emotion analysis system, we use the features of the best system in the "Sentiment analysis in twitter" task in SemEval [9, 12] and the word embedding features. We trained a SVM model based on these features to predict the emotion for each microblog.

Figure 3 shows the total emotion distribution for the hot event "QingAn Shooting", in which different color refers to different emotion. We can find that many people feel "angry" and "sad" about this event. This is reasonable. However, there are still some people who feel "happy" about this event. This is very intriguing and worth exploring.

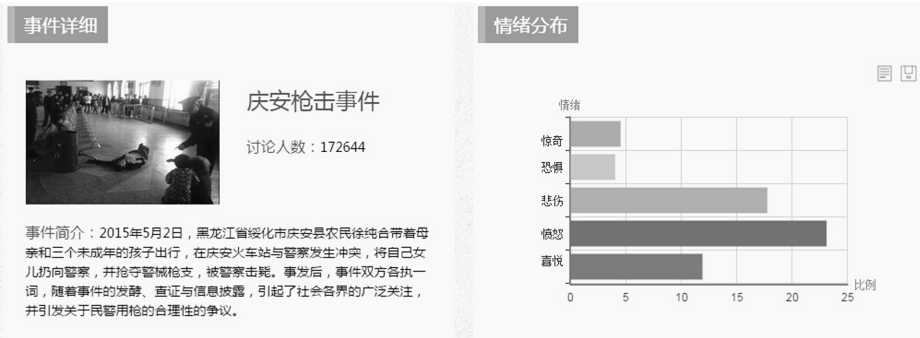

Fig. 3. The total emotion distribution for the hot event "QingAn Shooting".

2.4 Emotion Causation Analysis

What is Used as the Emotion Causation

The first necessary question that needs to be solved is what should be used as the emotion causation. One obvious way is to use some key words to represent the emotion causation. For example, MoodLens system [15] extracted the top 5 bi-gram terms of high frequency to detect the abnormal event, which is a little similar to the emotion causation task. However, the key words are hard to be understood by users, if the users do not know the background of the events. Thus, using the key words as the emotion causation is not a good way.

Instead, we observe that the different emotions for a same hot event are caused by some sub-events. In detail, most of the hot events can last for several days and can derive several related sub-events. The public emotion for the hot event can change with the evolution of the sub-events, such as the examples in Figure 1 and Figure 2. Based on the above, using the sub-events as the emotion causation looks like a good way. Therefore, the emotion causation analysis can be converted to exploit the sub-events and obtain all the related microblogs for each sub-event. Finally, the sub-events combined with their emotion distributions can be used to reveal the hidden emotion causation of the hot event.

In summary, the emotion causation analysis can be divided into two subtasks: Sub-event Detection and Sub-event related Microblogs Collection.

Sub-event Detection

Similar to the event detection task in Section 2.2, we also extract some typical hashtags to represent the sub-events, since the hashtag is user-generated content and easy to understand. Thus the sub-event detection can also be considered as a clustering task. We can use various kinds of clustering algorithms. This paper employs hierarchical clustering algorithm as a case of study. Similarity computation between two hashtags is the key technique during the clustering process. We can initially segment the hashtag and use a word vector to represent each hashtag. For example, we can use the words "埃博拉" (Ebola), "病毒" (virus) and "出现" (appearance) to represent the hashtag "埃博拉病毒出现" (Appearance of Ebola virus). Then, we use the cosine similarity computation method to compute the similarity between two hashtags. Finally, we conduct hierarchical clustering algorithm based on the similarity matrix.

However, the mere use of words that appear in the hashtag to compute similarities is insufficient because the hashtag contains only a few words, which are so sparse for clustering. We therefore introduce the background knowledge for each hashtag to alleviate these problems. This means that we can use more knowledge to compute the similarity between two hashtags besides their literal similarity. This idea is based on the hypothesis that the background knowledge of two similar hashtags is similar. This paper expands the background knowledge for each hashtag by importing all the messages that contain the hashtag.

Fig. 4. Sub-events extracted for the hot event "QingAn Shooting".

The hierarchical clustering algorithm is conducted based on this similarity. First, we suppose each hashtag as a cluster, noted as c_1, ..., c_i, ..., c_n. Next, we compute the similarity between each pair of clusters. If the similarity between c_i and c_j is the maximum and greater than a threshold θ, c_i and c_j are merged into a new cluster. The process is repeated until the amount of the clusters remains constant. Finally, each cluster can be considered as a sub-event. We simply rank the clusters according to their messages and capture the top n clusters as the typical sub-events. We use the longest hashtag in the cluster to represent the sub-event.

Figure 4 shows the five extracted sub-events for the hot event "QingAn Shooting", in which different color refers to different sub-event.

Sub-event Related Microblogs Collection

When we get the top clusters as the sub-events of the hot event, there are also many clusters left. Since lots of them are related to the sub-events, we need to classify these clusters into the sub-events. In this step, we adopt a simple method of training a classifier based on bag-of-words to classify the left clusters. Then we can obtain all the related microblogs for each sub-event.

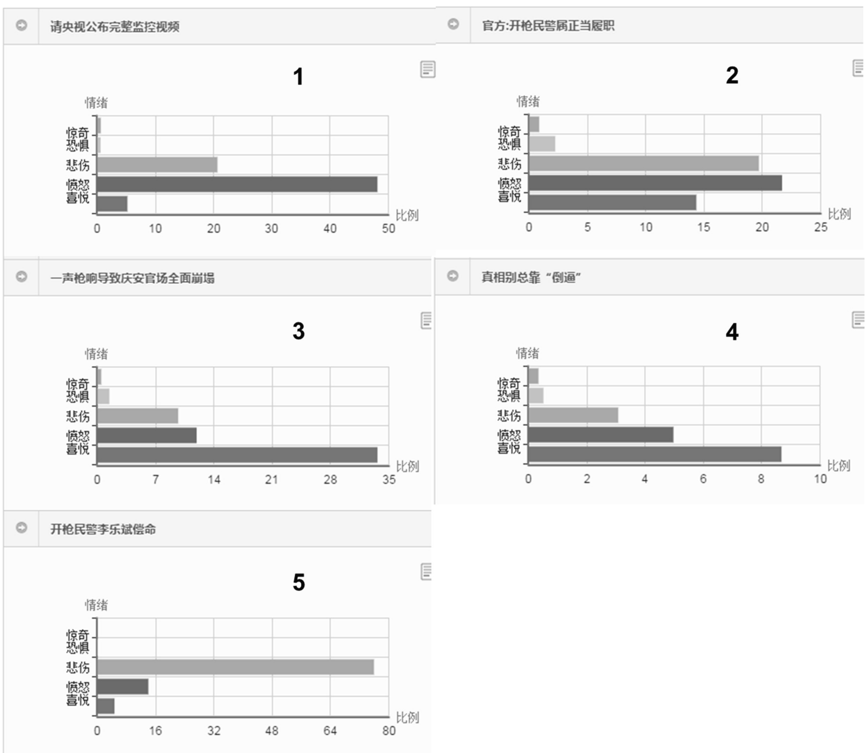

Fig. 5. Emotion causations for the hot event "QingAn Shooting".

Table 2. Statistics of 15 hot events.

ID	Hot Event	# of microblogs	# of hashtags
1	毕福剑不雅视频事件	121,204	1,988
2	云南女导游辱骂游客	18,338	605
3	四川女司机被当街暴打	324,775	4,091
4	男子暴打扫地小男孩	111,080	1,611
5	庆安枪击事件	172,644	1,930
6	王思聪范冰冰骂战	42,252	1,408
7	吴镇宇发飙	76,366	1,049
8	携程被黑	6,137	508
9	故宫女模裸照	7,198	255
10	MERS入侵广东	237,520	1,341
11	东方之星游轮倾覆	556,008	7,396
12	边策吸毒坠楼身亡	14,896	500
13	肃宁县特大枪击案	22,128	578
14	微博炫腹大赛	18,533	720
15	贵州自尽男童	30,848	1,930
	Sum	1,219,351	18,715

After the above two steps, we can analyze the emotion distribution for each sub-event using the algorithm in Section 2.3. Thus the sub-events combined with their emotion distributions can help the public realize the causations hidden behind each emotion of the hot event. Figure 5 shows the five emotion distributions of the five sub-events to reveal emotion causations hidden in the hot event "庆安枪击案" (QingAn Shooting). For example, the third emotion distribution in Figure 5 denotes that most people feel "happy" towards the sub-event "一声枪响导致安庆官场全面崩塌" (destroy the officialdom). This can explain why the "happy" emotion takes on not small proportion in the total emotion distribution of "庆安枪击案" (QingAn Shooting) in Figure 3.

2.5 Other Analysis Result Display

The Event-Emotion analysis system can also provide other deep analysis results. For example, it can track the trends of the microblog accounts, the trends of emotion distributions, the province ranking according to microblog accounts, some typical examples for each emotion, and some interesting analysis based on the user profile and event keywords. These visualization results can be found in http://qx.8wss.com/zkmap.com.

3 Experiments

In this part, we just have conducted the experiments on the emotion causation analysis task. The reason is that the hot event detection and emotion analysis are both popular tasks in natural language processing, lots of researchers have focused on these tasks and lots of algorithms are studied and compared. We just adopt the-state-of-the-art method to complete the two steps. In comparison, the emotion causation analysis task is firstly proposed in this paper.

3.1 Dataset

We collect 15 hot events in the recent half years, the statistics of which are listed in Table 1. We can find that these events are all discussed and commented by millions of people. And meanwhile, each hot event contains thousands of hashtags, it is possible to detect the sub-events from these hashtags.

3.2 Results of the Emotion Causation Analysis

We use two steps, sub-event detection and sub-event related microblogs collection to analyze the emotion causation hidden in a hot event.

Table 3. Performance of sub-event detection.

Method	P@5
Sub-event detection	66.7%

A hierarchical clustering based method is used to detect the sub-events. Table 3 shows the performance, which can reach an *accuracy* of 66.7%. Table 4 illustrates two hot events and their detected sub-events, where the standard sub-events are manually annotated. We can observe that most detected sub-events are correct causations.

In the sub-event related microblogs collection step, a simple classifier is used. Experiments show that we can obtain an average *F-score* of 68.74%, which is not high enough.

Table 4. Examples of detected sub-events.

Hot event	Standard sub-events	Detected sub-events
庆安枪击事件	请央视公布完整监控视频 官方:开枪民警属正当履职 一声枪响导致庆安官场全面崩塌 真相别总靠“倒逼” 开枪民警李乐斌偿命	庆安县副县长董国生被停职 官方:庆安开枪民警属正当履职 监控还原庆安枪击案全程 视频还原庆安枪击案全过程:徐纯合夺棍连砸警察 “黑龙江庆安火车站事件”调查结果公布
东方之星游轮倾覆	载客458人已救起8人 沉船内部有生命迹象 乘客家属收到诈骗短信 “东方之星”客船翻沉事件幸存游客口述 沉船存在生还者可能性已十分渺茫	载客458人已救起8人 沉船内部有生命迹象 乘客家属收到诈骗短信 “东方之星”客船翻沉事件幸存游客口述 大圣炽盛光如来拥护咒轮

It should be noted that this paper aims to try some simple algorithms for each step to validate the idea of using sub-events combined with their emotion distributions to explain the emotion causation of a hot event. Although the performance for each step is not very ideal, the treads or the ratios of the five emotions for a hot event are reasonable. We will still polish every step of our framework in future.

4 Related Work

A major contribution of Web 2.0 is the explosive rise of user-generated content. The importance of social media is constantly growing. Lots of systems or platforms are constructed to share mainstream media news and comment on the related events. In this manner, readers can follow the main events happening in the world, both from the perspective of mainstream as well as social media and the public's perception on them. Balahur et al. [4] build a system that links the main events detected from clusters of newspaper articles to tweets related to them, detects complementary information sources from the links they contain and subsequently applies sentiment analysis to classify them into positive, negative and neutral. Hsieh et al. [5] propose a Social Event Radar platform, which is used as a realtime risk control management technology to assist monitoring huge amount of new media related information and giving a warning for utility users' sake in efficiency way. Joshi et al. [6] build a C-Feel-It platform, which is a web-based system predicting sentiment in microblogs on Twitter. Li et al. [7] build a sentiment-based audiovisual system, which is called MemeTube, to recognize the sentiments of messages and display related music melody. Zhao et al. [14] present a Social Sentiment Sensor (SSS) system on Sina Weibo to detect daily hot topics and analyze the sentiment distributions toward these topics. Osborne et al. [10] introduce ReDites, a system for realtime event detection, tracking, monitoring and visualization.

From all the related systems, we can observe that topic detection/tracking and sentiment analysis are two key techniques. Topic detection and tracking (TDT) [2, 16] refers to the automatic techniques for finding topically related material in streams of data (e.g., newswire or social media data), in which topic detection involves detecting the occurrence of a new event such as a plane crash, a murder, a jury trial result, or apolitical scandal in a stream of news stories from multiple sources, and topic tracking is the process of monitoring a stream of news stories to find those that track (or discuss) the same event as one specified by a user. Sentiment analysis (also known as opinion mining) [8, 11] aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document.

For both techniques, lots of related algorithms have been exploited. They are indeed very useful and important in many social media applications. However, to date, almost all the public opinion analysis results are the emotion or sentiment orientation ("like" or "dislike") distributions for the hot events, which can be considered as kind of shallow analysis of the hot events. More and more people want to know the hidden causation behind the emotion distributions. Unfortunately, no system is built to automatically analyze the causation of the hot event emotion distribution. Thus in our knowledge, we are the first to propose the emotion causation analysis task and focus on this task.

5 Conclusion and Future work

In this paper, we present an Event-Emotion analysis system. Different from other public opinion analysis systems, our Event-Emotion system can not only show the hot events and their public emotion distributions like other common systems, but also reveal what causes these emotion distributions. The main contributions of this paper can be concluded as follows:

– We are the first to propose the emotion causation analysis task, which can help the users deeply analyze the social media.
– We propose to use the sub-events combined with their emotion distributions to represent the emotion causation. Experiments show that this idea is reasonable.
– We built a Chinese public opinion analysis system: Event-Emotion system, which fused the deep emotion causation analysis.

Although we have designed and built the architecture of the Event-Emotion system, the performance for each technique in this system is not ideal enough. As future work, we would like to polish each technique to make the system more practical.

Acknowledgments. We thank the anonymous reviewers for their helpful comments. This work was supported by National Natural Science Foundation of China (NSFC) via grant 61300113, 61133012 and 61273321, the Fundamental Research Funds for the Central Universities via grant No.HIT.NSRIF.2013090 and ZTE Research Project.

References

1. Allan, J., Carbonell, J., Doddington, G., Yamron, J., Yang, Y.: Topic detection and tracking pilot study: final report. In: Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop, Lansdowne, VA, USA, pp. 194–218, February 1998
2. Allan, J.: Introduction to topic detection and tracking. In: Topic Detection and Tracking, pp. 1–16. Kluwer Academic Publishers, Norwell (2002). http://dl.acm.org/citation.cfm?id=772260.772262
3. Allan, J. (ed.): Topic Detection and Tracking: Event-based Information Organization. Kluwer Academic Publishers, Norwell (2002)
4. Balahur, A., Tanev, H.: Detecting event-related links and sentiments from social media texts. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations. Association for Computational Linguistics, Sofia, Bulgaria, pp. 25–30, August 2013. http://www.aclweb.org/anthology/P13-4005

5. Hsieh, W.T., Wu, C.M., Ku, T., Chou, S.C.T.: Social event radar: a bilingual context mining and sentiment analysis summarization system. In: Proceedings of the ACL 2012 System Demonstrations. Association for Computational Linguistics, Jeju Island, Korea, pp. 163–168, July 2012
6. Joshi, A., Balamurali, A.R., Bhattacharyya, P., Mohanty, R.: C-feel-it: a sentiment analyzer for micro-blogs. In: Proceedings of the ACL-HLT 2011 System Demonstrations. Association for Computational Linguistics, Portland, Oregon, pp. 127–132, June 2011. http://www.aclweb.org/anthology/P11-4022
7. Li, C.T., Wang, C.Y., Tseng, C.L., Lin, S.D.: Memetube: a sentiment-based audiovisual system for analyzing and displaying microblog messages. In: Proceedings of the ACL-HLT 2011 System Demonstrations. Association for Computational Linguistics, Portland, Oregon, pp. 32–37, June 2011. http://www.aclweb.org/anthology/P11-4006
8. Liu, B.: Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers (2012)
9. Mohammad, S., Kiritchenko, S., Zhu, X.: NRC-canada: building the state-of-the-art in sentiment analysis of tweets. In: Proceedings of the Seventh International Workshop on Semantic Evaluation Exercises (SemEval-2013), Atlanta, Georgia, USA, June 2013
10. Osborne, M., Moran, S., McCreadie, R., Von Lunen, A., Sykora, M., Cano, E., Ireson, N., Macdonald, C., Ounis, I., He, Y., Jackson, T., Ciravegna, F., O'Brien, A.: Real-time detection, tracking, and monitoring of automatically discovered events in social media. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations. Association for Computational Linguistics, Baltimore, Maryland, pp. 37–42, June 2014. http://www.aclweb.org/anthology/P14-5007
11. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Found. Trends Inf. Retr. **2**(1–2), 1–135 (2008)
12. Rosenthal, S., Ritter, A., Nakov, P., Stoyanov, V.: Semeval-2014 task 9: sentiment analysis in twitter. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Association for Computational Linguistics and Dublin City University, Dublin, Ireland, pp. 73–80, August 2014. http://www.aclweb.org/anthology/S14-2009
13. Weng, J., Lee, B.S.: Event detection in twitter (2011). http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2767
14. Yanyan, Z., Bing, Q., Ting, L., Duyu, T.: Multimedia tools and applications, pp. 1–18, August 2014
15. Zhao, J., Dong, L., Wu, J., Xu, K.: Moodlens: an emoticon-based sentiment analysis system for chinese tweets. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD 2012, pp. 1528–1531. ACM, New York (2012). http://doi.acm.org/10.1145/2339530.2339772
16. Zhao, X., Shu, B., Jiang, J., Song, Y., Yan, H., Li, X.: Identifying event-related bursts via social media activities. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Association for Computational Linguistics, Jeju Island, Korea, pp. 1466–1477, July 2012. http://www.aclweb.org/anthology/D12-1134