ImgWordle: Image and Text Visualization for Events in Microblogging Services

Chong Kuang, Jiayu Tang[*], Zhiyuan Liu, Maosong Sun
State Key Laboratory of Intelligent Technology and Systems
Tsinghua National Laboratory for Information Science and Technology
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
{kuangchong07, tjy430, lzy.thu}@gmail.com, sms@tsinghua.edu.cn

ABSTRACT

With the wide usage of microblogging services, microposts grow at high rate and provide a rich source of information related to important social events and trends. However, analyzing microposts is challenging due to its high complexity and large volume. In this paper, we present ImgWordle, an interactive visualization prototype to help people perceive and analyze the image and text information in microblogging services. The prototype extends tag cloud by involving images as well as words, and provides multiple coordinated views for a given event, including keywords of various topics, representative images, geographic sentiments and amounts of microposts, and popular microposts over time. We implement ImgWordle on Sina Weibo, the most popular microblogging service in China, and illustrate the usefulness of this visual interface.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces

General Terms

Design, Human Factors

Keywords

Information visualization, microblogging, tag cloud, visual analysis

1. INTRODUCTION

In microblogging services, users generate small content by the so-called *microposts* in various forms, such as text, image, video, and geotagged data, within a certain length, e.g., 140 English or Chinese characters. With the wide usage of microblogging services, microblogging has become a new

---

[*]indicates equal contribution with the first author

form of media and had more and more powerful influence on the general public. Therefore, microblogging services have attracted many attentions from various communities, such as business companies, policymakers, public health officials, social scientists, and other researchers.

However, leveraging microposts is challenging because of its large volume, heterogeneity, and high generated frequency. To reduce the size of data to be perceived, it is straightforward to automatically detect patterns in microposts by using natural language processing and machine learning methods. Nevertheless, the results are usually uncertain and still hard for humans to discover interesting patterns.

To tackle the abovementioned challenge, we introduce ImgWordle, a multifaceted and interactive visualization prototype to help people explore online discussions about major events. In order to visually summarize what an event is, we introduce three visual primitives which are linked via filtering, highlighting, and brushing: (1) spatiotemporal tag cloud consisting of representative images and keywords extracted from online discussions; (2) map integrating geotagged information, the sentiments, and amounts of discussions; and (3) time-based popular micropost stream.

2. RELATED WORK

In recent years, researchers have proposed several visualization approaches to help people analyze and understand microblog-based events. Visual Backchannel [4] provides a visual summary of large-scale events in Twitter. TwitInfo [5] automatically discovers and labels peaks of high tweet activity. Whisper [2] traces the dynamics of information diffusion on Twitter. ScatterBlogs2 [1] supports the detection of low-profile events. Chae et al. [3] presented a system for exploring and examining abnormal topics and events.

Nevertheless, these systems mainly focus on analyzing and visualizing text information on social media. By contrast, ImgWordle aims at highlighting the role of image information on microblogging streams by showing what the popular images are, how these images related to keywords and topics, and when these images appear.

3. THE IMGWORDLE PROTOTYPE

We build our prototype on Sina Weibo, the largest microblogging platform in China. ImgWordle automatically retrieves structured microposts from Sina Weibo by using its APIs. Fig 1 shows an example of the interface. (In practice, the language of the interface is Chinese). It consists of three major parts, e.g., georeferenced summarization, image

Figure 1: The ImgWordle interface representing the microposts about the event that highway toll fees are cancelled during festivals in Beijing from September 29, 2013 to October 1, 2013.

and word cloud, and time-based micropost stream, to provide both macro and micro perspectives for a given event. In order to allow people to view a subset of the discussion stream, the coordinated-view interface supports responsive and interactive cross-filtering along regions and time. The timeline displayed in the top of the interface supports to change time span and temporal granularity, accompanying a search box used for filtering the event of viewer's interest.

3.1 Georeferenced Summarization

Fig. 1a shows the geo-information for the microposts during the filtered time. The map view presents two kinds of information: the sentiment of each region, and the amount of the microposts in each region. As colors have been used to illustrate various topics in other two parts of the interface, we use textures to illustrate sentiments to avoid confusion. The positive sentiment is designated with horizontal lines and negative with dots. Besides, the more microposts a region has, the deeper the color.

3.2 Image and Word Cloud

Fig. 1b shows the topic related keywords and representative images after spatiotemporal filtering. Keywords related to the same topic are in the same color. The image and word cloud provides interactive brushing to help viewers find how the terms are related to each other. When viewers hover over a word or an image, co-occurrence words and images are highlighted with distinct shadows.

3.3 Time-based Micropost Stream

Fig. 1c shows the time-based stream of popular microposts for a closer examination of event facts after spatiotemporal filtering. The stream is visualized in two lines to better exploit available space of the interface. This area is occupied by interactive squares that visualize the dynamics of popular discussions about a given event over 24 hours. Viewers can scroll the stream to the left and right to explore more microposts during the filtered time. Each square represents a micropost, and detailed information of the micropost will be shown after clicking the square. The background color of the square indicates the micropost's topic.

4. CONCLUSIONS AND FUTURE WORK

ImgWordle provides a convenient way to visually perceive events in microblogging services. Context, such as images, keywords, geographic and temporal information, and original text, for a given event is provided by the interface.

Though our first users have shown great interest in the representative images of the microblog-based events, a formal user study is needed to understand how effective the image and word cloud is, how we can improve it, and how integrated images support the visual analysis of microblogging data.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] H. Bosch, D. Thom, F. Heimerl, E. Puttmann, S. Koch, R. Kruger, M. Worner, and T. Ertl. Scatterblogs2: Real-time monitoring of microblog messages through user-guided filtering. *IEEE TVCG*, 19(12):2022–2031, 2013.

[2] N. Cao, Y.-R. Lin, X. Sun, D. Lazer, S. Liu, and H. Qu. Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE TVCG*, 18(12):2649–2658, 2012.

[3] 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 *Proc. IEEE VAST*, pages 143–152, 2012.

[4] M. Dork, D. Gruen, C. Williamson, and S. Carpendale. A visual backchannel for large-scale events. *IEEE TVCG*, 16(6):1129–1138, 2010.

[5] A. Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, and R. C. Miller. Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proc. ACM CHI*, pages 227–236, 2011.