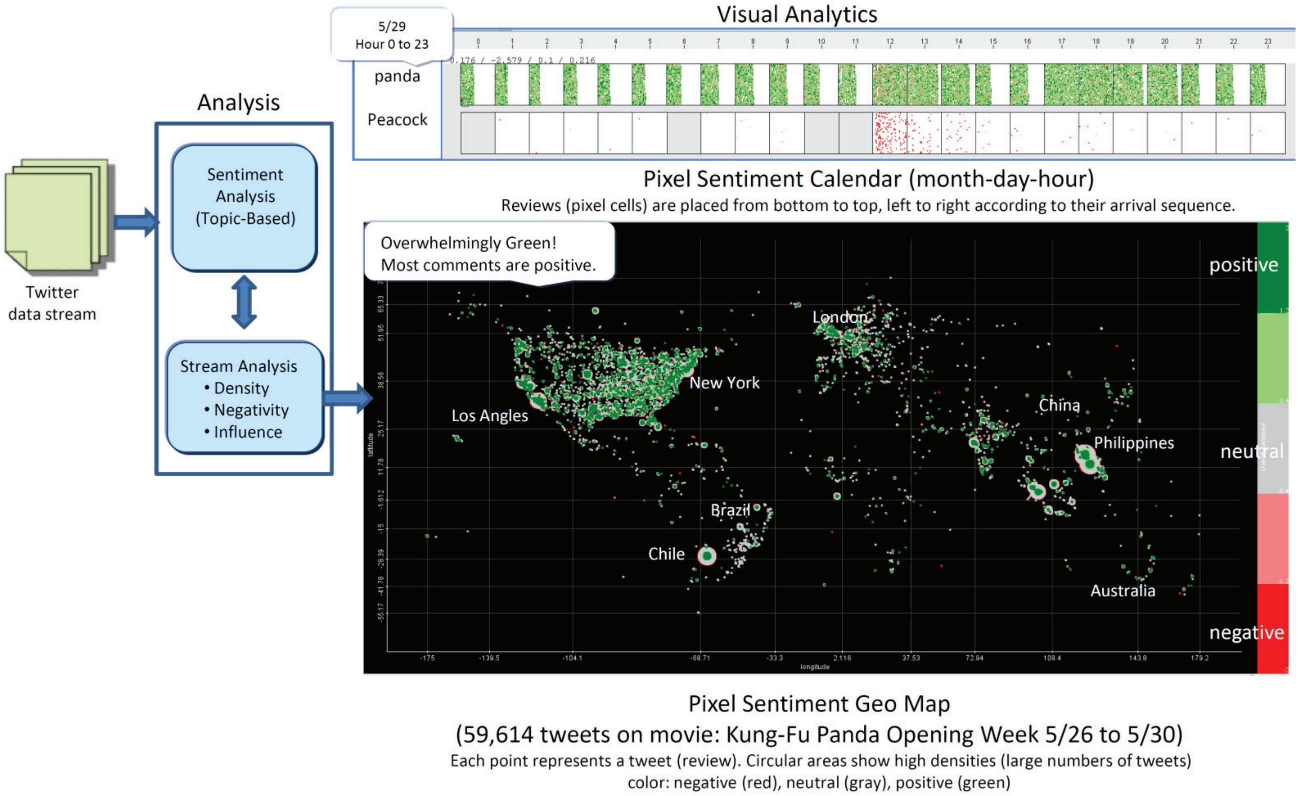
Visual Sentiment Analysis on Twitter Data Streams

Ming Hao, Christian Rohrdantz*, Halldór Janetzko*, Umeshwar Dayal
Daniel A. Keim*, Lars-Erik Haug, Mei-Chun Hsu
Hewlett-Packard Labs and *University of Konstanz



Pixel Sentiment Calendar (month-day-hour)

Reviews (pixel cells) are placed from bottom to top, left to right according to their arrival sequence.

Pixel Sentiment Geo Map

(59,614 tweets on movie: Kung-Fu Panda Opening Week 5/26 to 5/30)

Each point represents a tweet (review). Circular areas show high densities (large numbers of tweets)
color: negative (red), neutral (gray), positive (green)

Figure 1. A pipeline for visual analytics of live twitter data streams

Abstract

Twitter currently receives about 190 million tweets (small text-based Web posts) a day, in which people share their comments regarding a wide range of topics. A large number of tweets include opinions about products and services. However, with Twitter being a relatively new phenomenon, these tweets are underutilized as a source for evaluating customer sentiment. To explore high-volume twitter data, we introduce three novel time-based visual sentiment analysis techniques: (1) topic-based sentiment analysis that extracts, maps, and measures customer opinions; (2) stream analysis that identifies interesting tweets based on their density, negativity, and influence characteristics; and (3) pixel cell-based sentiment calendars and high density geo maps that visualize large volumes of data in a single view. We applied these techniques to a variety of twitter data, (e.g., movies, amusement parks, and hotels) to show their distribution and patterns, and to identify influential opinions.

Keywords: Visual Opinion Analysis, Sentiment Analysis, Twitter Analysis, Topic Extraction.
Index Terms: I.7.5 [Document and Text Processing]: Document Capture - Document Analysis; I.5.2 [Pattern Recognition]: Design Methodology - Feature evaluation and selection.

1 Introduction

With the rapid growth of social media, the amount of customer feedback available to corporations, business owners, and IT services managers interested in obtaining customer input is greater than ever. Figure 1 shows comments on the movie Kung-Fu Panda on Twitter. During the five days from 5/26 to 5/30, there were 59,614 tweets related to this movie. The total volume of twitter comments was much larger than that of normal reviews on the web site. The tweets came in by the minute in contrast to normal web reviews that come in on a daily basis. However, the language of the tweets is more casual than that of web reviews, and twitters comments were much shorter (maximum 140 characters) and contained more abbreviations. The enormous size of the data stream, the diversity of the comments, and the uneven distribution of opinions over time make the analysis of twitter data very challenging.

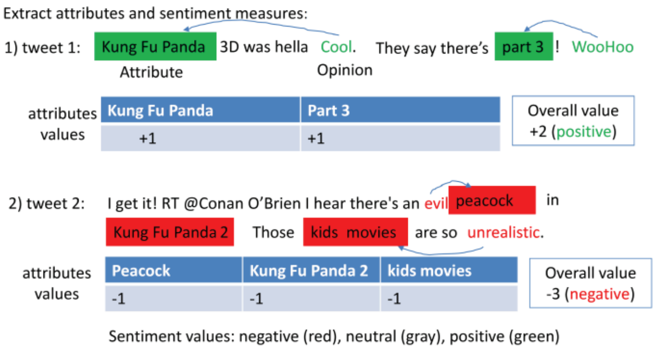
2 Our Approach

In this poster, we present an approach to exploring twitter data, as shown in Figure 1. Our approach attempts to automatically analyze large volumes of twitter comments with respect to what was commented on positively or negatively. To achieve our goal, we developed a novel topic-based text stream analysis technique that automatically detects which attributes were frequently commented on in tweets, based on their density distribution, negativity, and influence characteristics. Furthermore, we present two novel visualization techniques that help analysts explore the data in different ways from the previous methods using tag-cloud and radial visualization [3].

IEEE Symposium on Visual Analytics Science and Technology
October 23 - 28, Providence, Rhode Island, USA
978-1-4673-0013-1/11/$26.00 ©2011 IEEE

2.1 Sentiment Analysis

Using natural language processing techniques [1], we determine topics, extract attributes of the topics, detect opinions about the attributes, and measure the sentiment value as shown in Figure 2.


Extract attributes and sentiment measures:

1) tweet 1: Kung Fu Panda 3D was hella Cool. They say there's part 3 ! WooHoo

Attribute | Opinion

attributes	Kung Fu Panda	Part 3	Overall value
values	+1	+1	+2 (positive)

2) tweet 2: I get it! RT @Conan O'Brien I hear there's an evil peacock in Kung Fu Panda 2 Those kids movies are so unrealistic.

attributes	Peacock	Kung Fu Panda 2	kids movies	Overall value
values	-1	-1	-1	-3 (negative)

Sentiment values: negative (red), neutral (gray), positive (green)

Figure 2. Methods to extract attributes and to measure values

2.2 Stream Analysis

To analyze large live data streams, three key factors need to be considered:

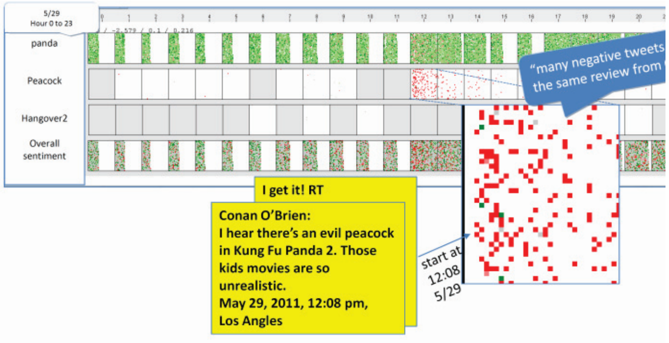
1. Density distribution: Do the tweets arrive together? Are they from different locations or from very close by locations geographically?
2. Negativity: Do the tweets have the same negative opinion?
3. Influence: What are the characteristics of an influential review? How many followers does it have? And from where?

2.3 Visual Analytics

To enable a more powerful exploration of the Twitter feeds, we combine the automated sentiment analysis with the following interactive visual analytics techniques:

2.3.1 Pixel Sentiment Calendar

Figure 1 top right shows a pixel cell-based sentiment calendar arranged in a row (topic) and column (time interval, i.e., hours) format. Each opinion is represented by a cell. The cell color is the sentiment value (green: positive, gray: neutral, red: negative). Cells are arranged from bottom left to top right and ordered according to the input time. Analysts can quickly grasp the distribution of customer feedback on specific features over time. As illustrated in Figure 3, many negative reviews are often influenced by the comment given by a celebrity, in this case Conan O'Brien.



"many negative tweets the same review from

I get it! RT

Conan O'Brien:
I hear there's an evil peacock in Kung Fu Panda 2. Those kids movies are so unrealistic.
May 29, 2011, 12:08 pm,
Los Angles

Figure 3. Pixel Sentiment Calendar

2.3.2 Pixel Sentiment Geo Map

The pixel sentiment geo map allows the visualization of large amounts of data fitting entirely into the display window without overlap [2]. The basic idea is to enable the user to optimize the degree of overlap and distortion to generate the best possible view as shown in Figure 1. Overlap often causes the loss of important information. Figure 4 is a visual representation of the map with the data-induced, high degree of overlap. Figure 5 shows each single tweet as one pixel without any overlap. However, the pixel geo map in Figure 1 shows the better geographical distribution of tweets. Figure 5 reveals dense areas, from which a large volume of reviews was posted on Twitter. The reviews with the same positions are placed around the original position based on their sentiment value. For example, the red border around the circular area represents negative tweets which have the same positions as the positive tweets (green) in the center of a circular area. Users are able to navigate through the dense areas for further analysis.

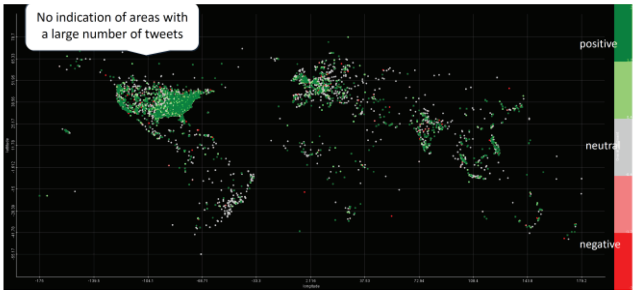

No indication of areas with a large number of tweets

Figure 4. Sentiment geo map with high degree of overlap

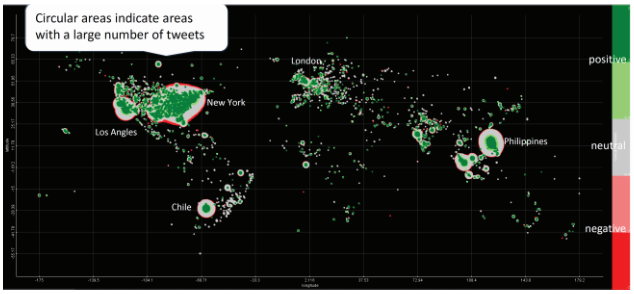

Circular areas indicate areas with a large number of tweets

Figure 5. Pixel Sentiment geo map without overlap

3 Conclusion

In this poster, we presented a visual analysis of Twitter time-series, which combines sentiment and stream analysis with geo- and time-based interactive visualizations for the exploration of real-world Twitter data streams. In addition to applying the above visual sentiment techniques to movie tweets, we have successfully applied them to post purchase web survey data and amusement park Twitter data identifying interesting patterns of customer feedback.. Today's visual analysis tools (e.g., SAS JMP, Vivisimo, Polyanalyst, etc.) mainly provide feedback on reviews using yes/no questions, numeric ratings, and direct comments. Our future work will incorporate information about opinion associations to find related features and visualize them appropriately.

[1] D. Oelke, et. al., Visual opinion analysis of customer feedback data. VAST, 2009.

[2] Keim, D. A., Hao, M. C. Dayal, U., Janetzko, H., and Bak, P., Generalized Scatter Plots. Information Visualization Journal (IVS09),

[3] Wu, Yingcai, et al., OpinionSeer: Interactive Visualization of Hotel Customer Feedback. IEEE InfoVis2010.