Sentiment Stream: A Visualization Design Study

Lynne Burks, Danny Cochran, Mahalia Miller, and Calvin Wang
{lynne09, ccochran, mahalia, thecw}@stanford.edu
Stanford University

ABSTRACT
Text data, such as customer satisfaction reports, personal micro-blogs, or emails, is being produced like never before; but tools for visualizing this growing data source are lacking. Typical text visualizations, like word clouds and word trees, tend to leave out relevant context, remove dimensionality of the underlying data, and restrict access to source text. In this paper, we present the Sentiment Stream, a proposed form of text visualization. Sentiment Stream aggregates text documents by time and space bins, preserving temporal and spatial context, and finds the most emotional content in each bin using part of speech tagging and sentiment analysis. The most relevant word from each bin is displayed in a grid encoding time and space, sentiment is encoded using color, and objectivity is encoded using font weight. We demonstrate the application of Sentiment Stream on a dataset of 20 million Tweets following Hurricane Sandy in the week after landfall in October 2012 in New Jersey, USA. Because of the unique spatial, temporal, and emotional context provided by Sentiment Stream, we are able to make conclusions about the sentiment surrounding this natural disaster that would not have been possible using other text visualizations.

Author Keywords
word clouds, text visualization, trend analysis

INTRODUCTION
Whether customer satisfaction reports, personal micro-blogs, or logs of unstructured text data, each second text is being produced like never before. The sheer quantity of text, however, is prohibitively large to visualize. Researchers have proposed strategies for exploring text and a popular tool is the word cloud. However, some key shortcomings include not providing access to the original source text, removing higher dimensions of the text such as spatial, temporal, and emotional context, and not allowing user interaction once the words are displayed.

These shortcomings become particularly striking when considering text with particularly important context coming from higher dimensions, such as geo-tagged micro-blogs about an emotional international sporting event, such as the Olympics or World Cup, or a natural disaster. As researchers in natural disasters, we have found a dearth of tools that allow users to explore the temporal and spatial variations of text, particularly with regards to emotion. We would expect higher negativity in areas most affected by a natural disaster. In addition, since affected people might express more common needs than those people responding to the news from afar, we might expect less variance in the topics discussed near a disaster as compared to text coming from locations farther away. However, we need a visualization tool to explore these hypotheses. Unfortunately, there is a lack of visualization tools that aggregate text in a meaningful way while preserving general context and allowing access to source text.

In this paper we present Sentiment Stream, a design study of visualizing text with a key spatial, temporal, and emotional context. Our tool offers new insight to a particular class of domain problems with text data situated in a rich context, which we demonstrate through a case study of over 20 million geo-tagged and often emotionally-charged Tweets in the week after Hurricane Sandy made landfall in New Jersey, USA on October 29, 2011 around 8pm local time (EDT). Despite recent advances in visualizing text data, our paper stands out, to the best of our knowledge, through the following contributions:

A text visualization tool that displays the spatial and temporal context of the text. We use time and distance bins to group the text and to extract the salient features, not in general, but for different points in time and space. This approach enables users to discover trends not visible with other aggregate methods.

A framework for exploring the emotional context of text. For all source text in the form of micro-blogs, we perform sentiment analysis. This analysis enables the functionality of looking at important words not simply by themselves (in their temporal and spatial context), but also in their emotional context. We show the positivity, negativity and objectivity of each time and distance bin as well as the most interesting word and corresponding source text.

Functionality to explore sentiment at various levels of detail while maintaining context. Through brush and linking methods, we enable the user to explore the text at various temporal scales. This approach places the text in the dynamic context of the changing emotional tide of an event.
The rest of the paper is organized as follows. First, we review related work and the tools from which we base our design study. Then, we crisply state the design requirements and explain our design choices and methods. We next demonstrate how our visualization tool is a suitable solution for visualizing the text and its emotional, spatial, and temporal context of Tweets following Hurricane Sandy. Finally, we discuss new insights garnered from the data that are enabled by our new visualization tool and conclude with a summary and future work.

RELATED WORK

*Sentiment Stream* draws upon considerable work in text visualization as well as on algorithmic tools for identifying either topics or emotions over time.

### Text analysis tools

Algorithmic and computational tools have enabled new ways to understand textual data. For example, term frequency-inverse document frequency weights terms to give insight into relevance [7, 16]. Other researchers have focused on identification of topics over time [8], sentiment over time [11], and memes over time [9]. However, despite the interesting insights gained from these approaches, they lack a corresponding visualization tool for exploring the data.

### Text visualization

Text visualization is an active field. Some researchers have focused on identifying key words and showing connections between entities such as people and locations in phone logs or dissertations and departments [4, 3]. Others have focused on word frequencies and where they have occurred in a text, i.e. the document concordances [5, 12]. An expansion of this work has been the word or tag clouds, which is arguably the most popular text visualization tool at this time [14]. Tag clouds show frequently occurring words where the size of the words indicate the relative frequency. Recently researchers have demonstrated a visualization tool that goes beyond word frequency to visualizing the topics in text identified by topic models, such as from a LDA topic model [2]. However, these solutions do not address the need in particular domains for also visualizing the temporal context of the text nor do they provide access to the original text.

Some popular tools that capture some of the context such as the temporal dimension, but still lack access to the original text, include works by [6, 15]. *Sentiment Stream* builds upon these ideas by expanding text visualization that showed relations over time to include temporal and emotional dimensions.

### METHODS

In this section, we describe the techniques and algorithms we use for our design study of *Sentiment Stream*.

### Text analysis

Our tool differs from most text visualization tools in that we bin our data along two main dimensions of interest, time and space, and then perform a text analysis of all text in each bin. To aggregate the text, we aimed to determine a most salient word to represent the text in the corresponding time-distance bin. We determined this by combining two techniques: part-of-speech tagging and term frequency-inverse document frequency for each adjective (tf–idf).

Depending on the application, different parts of speech could be displayed. A common method to classify parts of speech is using the maximum entropy treebank part-of-speech tagger, implemented in Python’s Natural Language Toolkit (nltk).

We compute the tf–idf by defining a document as all the text in a single time-distance bin. For each document, we compute the normalized term frequency of each chosen word as shown in Equation 1:

\[
 tf(t, d) = \frac{f(t, d)}{\max\{f(w, d) : w \in d\}}
\]

where \( f(t, d) \) is the frequency of term \( t \) in document \( d \), \( \max\{f(w, d) : w \in d\} \) is the maximum term frequency for all terms, and \( tf(t, d) \) is the normalized term frequency. The inverse document frequency is a measure of how frequently the term appears in other documents, and is computed using Equation 2.

\[
 idf(t, D) = \log \frac{|D|}{\max\{|d \in D : t \in d\}|}
\]

where \(|D|\) is the total number of documents, \( \max\{|d \in D : t \in d\}| \) is the number of documents where the term \( t \) appears, and \( idf(t, D) \) is the inverse document frequency. Since the final tf–idf reflects how relevant a chosen word is to its time-distance bin, we display the word with the largest tf–idf score in each bin.

In addition to displaying relevant words, our visualization tool encodes sentiment: color for positivity/negativity. For each time-distance bin, we determine the sentiment using an established sentiment lexical resource, SentiWordNet 3.0 [1], which offers real-valued scores for word positivity and objectivity. For example, *good, elysian,* and *amiable* have very high positivity scores [1]. We use an augmented bag-of-words model based on work by [13] where we flip the polarity of the sentiment scores for words between a negation (*not, *n’t, no, never*) and the next end-of-sentence punctuation as discussed by [11]. While sentiment extraction using this approach can be difficult if each document is short and contains only a few words in a standard lexicon, we found through an internal validation that we achieve good results when we extract sentiment from all documents in an entire time-distance bin.

### Visualization techniques

Our proposed visualization requires techniques for which there were few existing public templates in the java-script library D3.

\[\text{nltk.org}\]
visualizations, we eventually discovered Swimlane Chart. Though our final visualization differs dramatically in function, Swimlane proved to be a valuable foundation upon which our project is now built. Below are example snippets of code revealing how color and font size are determined:

```javascript
colorScale = d3.scale.quantize()
    .domain(positivityDomain)
    .range([0,75,150,225]);
// quantize font size
wordSizeScale = d3.scale.quantize()
    .domain(frequencyDomain)
    .range([12,16,18,20]);
```

The code includes not just the D3 library, but J-query (for the automatic Play) and Tipsy (for hover states).

The final structure of our case study code is organized and commented in such a fashion that someone with even minimal exposure to D3 could append their own data set to create *Sentiment Stream*.

**RESULTS**

We have implemented a proof-of-concept of *Sentiment Stream* using a dataset of Tweets about Hurricane Sandy, which caused significant damage in the Northeastern United States in the Fall of 2012. In this section, we discuss in detail the underlying analysis and the final results from this case study.

**Text Data**

We obtained data from Twitter by scraping for all geo-tagged Tweets containing the keyword “hurricane” or “sandy” that occurred within the week after landfall. Our resulting dataset consists of over 20 million Tweets from across the globe, each with a unique ID, time stamp, latitude and longitude, and text. A sample raw Tweet is presented below before any text analysis was applied.

```json
{
    "tweet_id": "262973175988629505",
    "user_id": "14228560",
    "name": "Alex",
    "screen_name": "alexgoetz",
    "latitude": "42.931396484375",
    "longitude": "-85.64969635009766",
    "tweet_text": "RT @vuberrybie: What if Gangnam Style is actually a rain dance and we have brought this hurricane on ourselves? #Sandy",
    "created_at": "2012-10-29 17:44:25"
}
```

We also defined the origin of our data to be the time and place where Hurricane Sandy made landfall, which was Monday, October 29, at 8 p.m. EDT just outside Atlantic City, NJ at 39.4 degrees North latitude and 74.5 minutes West longitude. The upper left corner the *Sentiment Stream* will be represented by this point in space and time.

2 [http://bl.ocks.org/1962173](http://bl.ocks.org/1962173)

3 Special thanks to Reza Zadeh for his help with data collection and analysis.

Figure 1. Map of where Hurricane Sandy made landfall, as indicated by the red hurricane symbol, and distance ranges of radius 50 miles, 110 miles, 200 miles, 500 miles, and 1000 miles indicated by red circles.

**Text Analysis**

The first step is to pre-process the raw Tweets by removing non-printable characters such as foreign language symbols and irrelevant symbols like the newline and tab characters. In this step, we also compute the distance between each Tweet and the origin point by computing the time, in hours, and the distance, in miles, that the Tweet occurred from the landfall of Hurricane Sandy.

We aggregate the Tweets using time and distance from landfall by creating a bin for each hour in the week following the hurricane and a bin for the following distances, as shown in Figure 1:

- 0–50 miles: southern New Jersey
- 50–110 miles: includes Manhattan
- 110–200 miles: includes Washington DC
- 200–500 miles: East Coast of the US
- 500–1000 miles: Midwest, including Chicago
- 1000–3000 miles: continental US
- ≥3000 miles: rest of the world

The main visual feature of *Sentiment Stream* is a grid that displays the time bin on the horizontal axis and the distance bin on the vertical axis. For each time and distance bin, we find and display the adjective with the highest tf–idf score from all Tweets that occurred in the bin. We also determine the positivity sentiment scores as described in the Methods section.
Text Visualization and Encodings

*Sentiment Stream* is designed to be an interactive text visualization, with the following design requirements:

- Provide spatial, temporal, and emotional context
- Aggregate text in a meaningful way for visual display
- Allow access to source text
- Allow interactive exploration of data

In this section, we describe the visualization and the encodings we used to meet these design goals. See Figure 2 for a screen shot of *Sentiment Stream*.

The main component of *Sentiment Stream* is a grid of words, where the horizontal axis represents time and the vertical axis represents distance from the origin (the upper left corner of the visualization). For the presented case study, the origin is the time and location where Hurricane Sandy made landfall, which was Monday, October 29, at 8 p.m. EDT just outside Atlantic City, NJ. Each grid point represents one hour of time and an approximate logarithmic scale of distance. For each distance bin, we list the number of miles and provide a qualitative description upon mouse-over (like “southern New Jersey”, or “continental US”) to provide geographic context for the user, since a simple number of miles may be too abstract. Spatial and temporal context is encoded as location within the grid structure of the visualization.

We provide emotional context by encoding the results of the sentiment analysis as the background color of each time-distance bin. The color scale ranges from black to light blue, where black encodes a low positivity score and light blue encodes a high positivity score. We chose a black and blue color scale because those colors inherently represent sombre emotions, which is appropriate for a natural disaster. Furthermore, hue has been shown to be a more justified choice for encoding an ordered variable than, for example, an arbitrary rainbow of colors [10]. For Hurricane Sandy, we discover that people closest to landfall are more negative consistently across time while people farthest away are more positive across all time. This indicates a strong correlation of positivity with distance rather than time, which was unexpected. This is just one example of new insights offered by our visualization.

*Sentiment Stream* aggregates text by displaying the adjective with the highest tf–idf score in each time-distance bin. This underlying language model is meaningful for conveying emotion because adjectives are generally the most expressive words, as opposed to nouns or verbs. The use of the tf–idf score ensures that the displayed word is Tweeted frequently in its time-distance bin but rarely in others, meaning that each word is representative of the bin in which it is displayed.

The principle interaction method of *Sentiment Stream* is brushing and linking between the main grid, which displays relevant words for up to ten hour time periods, and a small overview of the entire grid at the bottom of the visualization.

![Figure 3. A sample Tweet containing the keyword is displayed when the mouse is hovered over a word.](image)

The small overview grid shows the color-encoded positivity score for all time, providing a summary graph of how positivity changes over time and space following landfall of the hurricane. The user can also highlight a region, which is then displayed in the main grid with the most relevant words. The highlighted region can be expanded or dragged through time and the main grid display will adapt to reflect these changes. This interactive method allows the user to explore the data at their own pace and zoom in on time periods that they find interesting. A play button is also made available that will slowly scroll through the data with time.

We also implement the interaction method of hover, which displays additional information upon hovering over with the mouse. The user can gain access to the original source text by hovering over a word, and seeing a sample Tweet containing that keyword displayed (see Figure 3). The legend is also interactive in that the name of encodings are displayed, like font type, color, and word frequency, and details about each encoding are provided on hover (see Figure ??). We chose to provide details when hovering to keep the visualization clean and reduce the displayed text that does not come directly from the data.

The final interactive technique that we implement is the ability to click on a word and see all places that it appears in the grid. When clicked, each bin containing that word is highlighted in green and displayed in the overview grid at the bottom of the visualization, as shown in Figure 2. This way, the user can see when and where certain topics are trending.

Design Iterations

The initial design for *Sentiment Stream* involved placing the most relevant words on a map to show geographic context of the underlying text data. But once we started aggregating Tweets to get the most relevant word in a time-distance bin, the idea of a single location for a relevant word was quickly lost because the most relevant word appears in many Tweets and therefore many locations. We came up with the grid design for displaying words to keep as much temporal and spatial context without sacrificing text aggregation.

We also added the mouse-over interaction technique after receiving preliminary feedback that users would like more access to the underlying data, including an example Tweet where each keyword came from. To keep the visual design clean and free of cluttering text, we implemented hover for the legend and the distance labels. We also added an overt marker above the legend instructing the user to hover their
In the first week after the event. The encoding of objectivity shows us that the most emotional Tweets tend to occur close to the origin in both time and space, while Tweets farther away tend to be more objective.

One challenge that we encountered with our dataset is that traditional natural language processing techniques are not all well suited for Twitter data because people tend to Tweet in slang and symbols rather than proper English. Tweets typically have poor or nonexistent grammar, frequently contain terms that are shorthand or not real words, like “omg”, “lol”, and “askin”, and contain uniquely relevant symbols, like @ or #. The @ and # symbols require customized tokenizing of the tweet text into a bag of words. Similarly, poor grammar and slang terms can confuse the part of speech tagger, particularly for nouns and verbs, and result in poorer coverage by the sentiment lexicon than with more traditional texts. To overcome these challenges, we customized our tokenizer, focused on a part of speech that was more robust (adjectives) to noisy text, and aggregated our Tweets into bins before analyzing the sentiment.

CONCLUSION AND FUTURE WORK

This paper introduces Sentiment Stream, a design study of visualizing text with a strong temporal, spatial, and/or emotional context. We demonstrate how our tool shows the spatial, temporal, and emotional context of micro-blog text and allows users to interactively explore the textual data and context at different scales. Through a case study of over 20 million geo-tagged Tweets in the week following landfall of
Hurricane Sandy in the United States, we illustrate how new insights can be gained from the text by using our visualization tool.

This work suggests two main directions for further research. First, through user studies we could create an even more powerful visualization. Allowing the user to see more than one word per distance-time bin while still maintaining a crisp design is an area ripe for future work. Furthermore, we could increase the interactivity of the visualization through different filters of the text. We also might allow users to choose new axes depending on the contextual dimensions of the data in which they are most interested.

Second, while we have demonstrated our visualization on Tweets following Hurricane Sandy, this framework suggests other applications. For example, our approach would provide great power for understanding the changing pulse of emotions in different locations related to a major sports event, such as the World Cup. While we could provide a retrospective tool for exploring the data as we have done here, another area of future work is to update the visualization in real time. Such a tool could provide new insights in a variety of applications such as for real-time understanding of brand and customer satisfaction in different regions and how it’s changing or for an augmented version of presidential opinion polls. Future work could also enable users to independently include their own data and thus provide a robust system for users themselves to visualize textual data and its rich context.

REFERENCES


