Collaborative Visual Analysis of Sentiment in Twitter Events

Michael Brooks, John J. Robinson, Megan K. Torkildson, Sungsoo (Ray) Hong, and Cecilia R. Aragon

University of Washington, Seattle, WA, USA
{mjbrooks,soco,mtorkild,rayhong,aragon}@uw.edu

Abstract. Researchers across many fields are increasingly using data from social media sites to address questions about individual and group social behaviors. However, the size and complexity of these data sets challenge traditional research methods; many new tools and techniques have been developed to support research in this area. In this paper, we present our experience designing and evaluating Agave, a collaborative visual analysis system for exploring events and sentiment over time in large tweet data sets. We offer findings from evaluating Agave with researchers experienced with social media data, focusing on how users interpreted sentiment labels shown in the interface and on the value of collaboration for stimulating exploratory analysis.

Keywords: Collaboration; visual analytics; social media; sentiment; Twitter.

1 Introduction

Every day, activity on social media sites like Twitter and Facebook generates real-time data about human interaction at an unprecedented scale, capturing the ideas, opinions, and feelings of people from all over the world. These new data sources present an exciting research opportunity, and researchers across fields have approached many aspects of social media, e.g. the structure or dynamics of the social graph, the transmission of information and ideas, or users' emotions, affect, or sentiment [5,3].

Social media data sets can be quite large, and have complex network structures that shift and change over time; their main substance is text communication between users. These characteristics make them challenging to work with, and even getting a good overall sense of a social data set can be quite difficult.

As in other areas of data-intensive science, data visualization can enable researchers to reach new insights. Numerous examples of visualization systems for social data such as *Vox Civitas* [4] and twitInfo [13] have demonstrated the potential for visual analysis to support research on social media. Because these projects are often interdisciplinary, collaborative visual analysis [11] may be particularly useful. However, so far, there has been little research on collaborative visual analysis tools for social data.

We present Agave, a tool we developed to support collaborative exploration of events in large Twitter data sets, with a particular focus on sentiment. Timeline visualizations of trends and spikes in sentiment help teams of users find relevant events, which can be examined in greater detail through filtered lists of tweets. Annotation and discussion features allow users to collaborate as they explore the data set.

We recruited a group of researchers to evaluate Agave by exploring a data set of almost 8 million tweets from the 2013 Super Bowl, a major sports event. We contribute the findings of our qualitative study, discussing the usefulness of collaboration for exploratory analysis of difficult social media data sets, and implications for the design of sentiment visualizations. Agave and its source code are publicly available[1], to encourage further development and research on collaborative social media analysis tools and sentiment visualization.

2 Background and Related Work

We review examples of Twitter research focused on emotion and sentiment to provide context for how Agave might be used. Following this, we discuss related work on visual analysis of Twitter data and collaborative visual analysis.

2.1 Emotion in Twitter

Tweets are often explicitly emotional or carry emotional connotations, giving rise to a variety of research on emotion, affect, or sentiment in Twitter.

Dodds et al. demonstrate how Twitter can be used to calculate a metric for social happiness, and analyze temporal fluctuations in happiness on Twitter over days, months, and years [5]. In a similar vein, Quercia et al. calculated a gross community happiness metric based on tweets originating from different census communities in the UK, finding that their metric correlated with socio-economic status at the community level [14]. Mood extracted from Twitter has also been associated with daily changes in the stock market [1].

At a personal scale, a study of individual tweeting behavior has associated sharing emotion in tweets with having larger, sparser follower networks [12]. De Choudhury et al. have used mood patterns in the social media activities of individuals to understand behavior changes related to childbirth [2], and to recognize signs of depression [3]. All of these projects involve analysis of large numbers of tweets. Our goal in this paper is to explore how collaborative visualization can support data exploration in such projects.

2.2 Collaborative Visual Social Media Analytics

Visualization and visual analytics are promising tools for tackling complex, dynamic social media data sets, and *collaborative* visual analysis can enable research teams to reach greater insight in interdisciplinary projects.

[1] http://depts.washington.edu/sccl/tools

Researchers in the visualization and visual analytics communities have explored visual analysis of Twitter data. The "Visual Backchannel" system developed by Dork et al. presents a stream graph of Twitter topics over time, as well as a display of relevant Twitter usernames, tweets, and photos [6]. The *Vox Civitas* [4] and twitInfo [13] systems use temporal visualizations and sentiment analysis to support journalists exploring tweets about specific events. However, these projects do not address collaborative visual analysis.

Over the past 10 years, there has also been interest in collaborative visual analytics in general. The NameVoyager system was an early example of large-scale public data analysis [15]. Heer et al. followed up on this idea with additional exploration of collaborative features such as graphical annotation, view sharing, and threaded comments [11]. Focusing more on analyst teams, Heer & Agrawala presented a summary of design considerations for collaborative visual analytics systems [10]. However, there is little related work at the intersection of visualization, collaborative data analysis, and social media studies, especially focusing on sentiment.

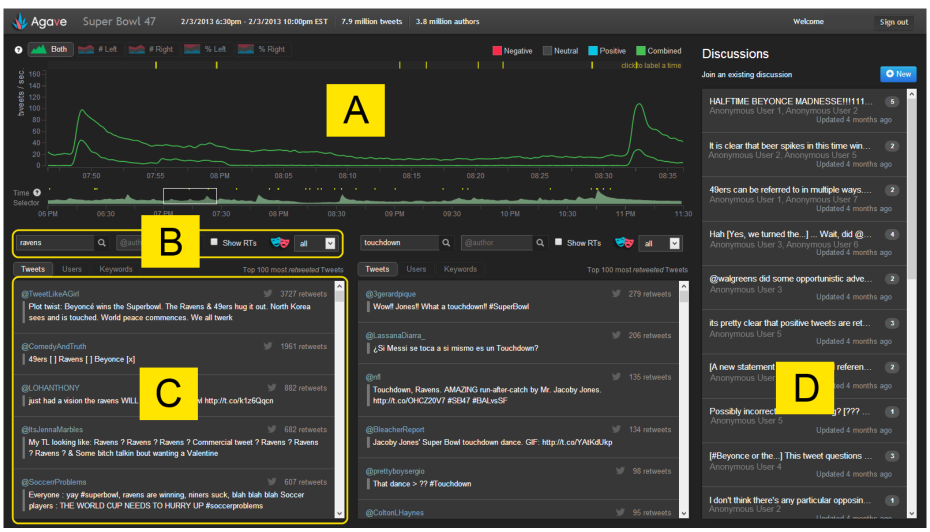

Fig. 1. A screenshot of Agave with its four main features: (A) timeline visualizations displaying different representations of the data, (B) data filters to refine searches, (C) data details showing lists of tweets, users, and keywords, and (D) threaded discussions to communicate with other users.

3 Design

The interface is illustrated in Figure 1, and a demo is available[2]. A prominent timeline visualization highlights temporal trends. In the default mode (Figure 1,

[2] http://depts.washington.edu/sccl/tools

A), the timeline shows the rate of tweets over time as a line graph. Other modes show changes in *sentiment* over time using a streamgraph of positive, negative, and neutral layers representing tweet counts (Figure 2, top) or percents (Figure 2, bottom).

The tabbed panels below the timeline (Figure 1, C) display tweets, users, and "burst keywords" [9] to provide a snapshot of the activity within the selected time range and filters. Zooming and panning on the timeline updates the contents of the detail panels by time range. The tweets displayed can also be filtered by keyword and author, or by sentiment label. Users can define two parallel sets of filters for compare and contrast tasks.

To facilitate shared context between users, we implemented an annotation system for labeling events on the timeline. A bar just above the timeline can be clicked to add an annotation. The user provides a brief description and the annotations are shared between all users. Furthermore, to help users document and share their findings, we also provide threaded discussions (Figure 1, D). New posts are automatically tagged with the current state of the users' view, promoting common ground for discussions. Users may also attach interactive references to specific tweets and annotations in their posts.

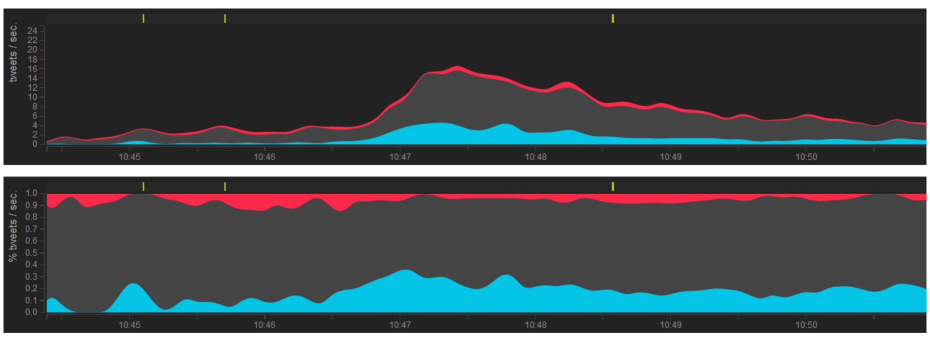

Fig. 2. Sentiment streamgraphs for the keyword search "Flacco", the Super Bowl MVP. Negative is red, neutral is gray, and positive is blue. **Top:** overall frequency of tweets, divided by sentiment type. **Bottom:** sentiment as percent of overall volume.

4 Evaluation

We evaluated Agave to investigate how collaborative features and sentiment visualizations could support exploratory analysis. Below, we discuss the Twitter data we used in the study, and the study procedure.

4.1 Twitter Data Collection

We collected a set of almost 8 million tweets during Super Bowl XLVII, the annual championship football game of the US National Football League. This event was selected based on an expectation of high Twitter volume with emotional content. Data was collected from the Twitter Streaming API, using a tracking list of 142 terms including team and player names, coaches, and entertainers, from Friday, February 1st at 22:30 EST until 20:30 EST on Tuesday, February 5th.

For sentiment analysis, we used the Sentiment140 API [8], which categorizes each individual tweet as positive, negative, or neutral. Based on 500 randomly selected tweets manually labeled by two of the authors, Sentiment140 achieved an overall accuracy of 71%, with a Cohen's kappa of 0.26 (between the two researchers Cohen's kappa was 0.57).

4.2 Procedure

Because of the open-ended nature of exploratory data analysis, our evaluation used a qualitative approach. Participants explored real Twitter data in Agave and posted comments and annotations.

We recruited 7 participants experienced with Twitter-based research. Participants' prior grounding and interest supports the validity of their interpretations of the data set and their usage of Agave. After a 5-10 minute tutorial on Agave, participants explored the tool freely for 20-30 minutes, either in our lab or remotely by video conference. We used a think-aloud protocol and observer notes to monitor these open-ended sessions, similar to the lab study used in [11].

We then allowed participants 3-4 days to revisit Agave to continue exploring the data set on their own. After this out-of-lab session, participants completed a questionnaire about discoveries, problems encountered, and attitudes about Agave. Log data were also collected by the system to determine how often each feature was used. Finally, post-study interviews focused on how participants used the visualizations and collaborative features to explore the data.

5 Findings and Discussion

Below, we discuss the importance of indirect collaboration for exploring the data, and the interpretation of sentiment visualizations.

5.1 Indirect Collaboration

Collaboration we observed in Agave consisted of mostly indirect interactions between people; participants only occasionally addressed other users directly. Because participants did not use the tool at the same time, there was little reason for them to expect a response to a direct interaction. Threads in the study were short (with fewer than 5 posts), and some of these discussions had only one participant. For example, one participant appropriated discussion posts as a way of bookmarking interesting tweets. Of course, these posts were still public, so it was possible for any other user to view and interact with them. As a result, even posts created without any expectation that others would respond were often read by other participants:

> I am looking at a discussion post [in a thread about misclassified sentiment]: "#IRegretNothing is a positive hashtag..." I wonder how many people used that. [Searches for the hashtag] Now I'm Looking at RTs for #IRegretNothing. (Peter)

This indirect interaction through the discussion threads was useful for suggesting and extending lines of analysis. Participants read comments made by other users and sometimes restored other users' views so that they could easily focus on a potentially interesting section of the data, often adapting and extending the filters created by others.

More than discussion posts, annotations created by previous users provided "jumping off points" for exploration of the data. Allowed to explore the tool freely, a few participants were initially unsure how to get started with the Twitter data, and began by opening the discussions or examining the annotations that had been left by other participants. Many participants seemed to prefer the annotations:

> The discussions I found harder to look at and understand than just the yellow tags [annotations], which were fairly straight forward, and I looked at a bunch of those. (Peter)

Participants used existing annotations, which were shown in context as part of the timeline, as a way of choosing periods of time to focus on (i.e. social navigation [7]). Just as participants noticed and zoomed in on sharp changes in the timeline, several participants focused on times surrounding annotations in order to understand what the annotations were referring to. The following excerpt shows how Peter used annotations of others to understand how the 49ers fans reacted to their team's defeat:

> Peter focuses the timeline on an interval surrounding the end of the game, where another user's annotation has drawn his attention. He then uses the sentiment filter to limit the results to negative sentiment. Browsing through the details pane, he notices multiple tweets about the 49ers losing.

In this case, another participant's annotation at the end of the football game was useful for highlighting and explaining a shift in sentiment that is visible in the timeline display. Sometimes, instead of zooming in on annotations, participants sought out relatively unexplored areas of the data set to focus on, an example of "anti-social navigation" [10,15]:

> I started by kind of looking in the trough of the blackout period because that hadn't had a lot of annotations, so I thought it would be interesting to look there. In that dip during the blackout, I saw that [a company] was tweeting and got retweeted a bunch of times. (Hannah)

Participants expressed appreciation for the annotations feature in particular; as a low-risk way of indirectly collaborating with other users, annotations were value for helping people start to explore the data. These observations support Heer et al. who argued that "doubly-linked" discussions (linked to and from the relevant data or view) enable users to not only tell what views others were talking about, but to find existing discussions about their current view of the data [11]. Our system did not fully close the "doubly-linked" loop by connecting these annotations to discussions where they were referenced, but our findings suggest that this is a promising idea for future work:

> My goal was very data driven, starting with the data, and there's no real way to end up in the discussion if you start with the data. [...] I think being able to maintain links to the data, narrative centric, vs. marking the data itself (data centric) is really great. (Krista)

5.2 Exploration of Sentiment

Emotion, affect, and sentiment are important facets of social data sets [4,13,6]. However, the implications of incorporating sentiment in visualizations have not been thoroughly considered. Agave presents information about sentiment in the timeline visualizations and beside tweets in the details panels (Figure 1, C).

Although the visualization of total tweets over time was useful for identifying events of interest, some events were more easily identified using the visualizations of positive, neutral, and negative sentiment (Figure 2). Peter reflected on his use of the sentiment timeline: *"I think I was most interested in visually seeing where the peaks of positive and negative were and what was clustered around those areas."*

The sentiment timelines successfully facilitated insight about particularly emotional topics or events such as touchdowns or advertisements. However, the sentiment indicators that we attached to individual tweets in the tweet list played a more ambiguous role because they provoked doubt and suspicion in the validity of the sentiment data:

> I saw a few more tweets in there that were in the positive or negative column which were questionably positive or negative and which I felt were more neutral. (Allison)

Sentiment classifications were not merely "wrong"; they were ambiguous and subjective. In our study, the presentation of individual sentiment labels was more problematic to participants than the aggregated sentiment timelines. Additional research is needed to understand trust and validity in visual analytics systems that rely on questionable data.

5.3 Limitations and Future Work

Future studies should address how people interact with collaborative social media visualization systems in the context of their own projects, e.g. with their own data and with their actual colleagues. The breadth of collaborative interactions we were able to observe in our study was limited by the unfamiliarity of the data set and short-term involvement with the system. More research is needed to understand how the presentation of sentiment in visualization and visual analytics tools affects the way people interpret and interact with sentiment data.

6 Conclusion

Agave is a collaborative visual analytics tool supporting exploratory analysis of events in large Twitter data sets, with a focus on sentiment. We conducted a qualitative evaluation to find out how researchers used Agave to develop insights about an 8 million tweet data set. Participants found the timeline visualizations, particularly displays of sentiment changes over time, useful for finding and understanding interesting events. Annotation and discussion were used to share findings, but also served as as jumping off points, enabling indirect collaboration that stimulated broader and deeper exploration of the data.

References

1. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. Journal of Computational Science 2(1), 1–8 (Mar 2011)
2. De Choudhury, M., Counts, S., Horvitz, E.: Major life changes and behavioral markers in social media: case of childbirth. In: Proc. CSCW (2013)
3. De Choudhury, M., Gamon, M., Counts, S., Horvitz, E.: Predicting Depression via Social Media. In: Proc. ICWSM (2013)
4. Diakopoulos, N., Naaman, M., Kivran-Swaine, F.: Diamonds in the rough: Social media visual analytics for journalistic inquiry. In: Proc. VAST (2010)
5. Dodds, P.S., Harris, K.D., Kloumann, I.M., Bliss, C.a., Danforth, C.M.: Temporal patterns of happiness and information in a global social network: hedonometrics and Twitter. PloS one 6(12), e26752 (Jan 2011)
6. Dörk, M., Gruen, D., Williamson, C., Carpendale, S.: A visual backchannel for large-scale events. In: Proc. InfoVis (2010)
7. Dourish, P., Chalmers, M.: Running out of space: Models of information navigation. In: Proc. HCI (1994)
8. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. Tech. rep., Stanford University (2009)
9. Guzman, J., Poblete, B.: On-line relevant anomaly detection in the Twitter stream: an efficient bursty keyword detection model. In: Proc. KDD Workshop on Outlier Detection and Description (2013)
10. Heer, J., Agrawala, M.: Design considerations for collaborative visual analytics. Information Visualization 7(1), 49–62 (Feb 2008)
11. Heer, J., Viégas, F.B., Wattenberg, M.: Voyagers and voyeurs: supporting asynchronous collaborative information visualization. In: Proc. CHI (2007)
12. Kivran-Swaine, F., Naaman, M.: Network properties and social sharing of emotions in social awareness streams. In: Proc. CSCW (2011)
13. Marcus, A., Bernstein, M.S., Badar, O., Karger, D.R., Madden, S., Miller, R.C.: TwitInfo: aggregating and visualizing microblogs for event exploration. In: Proc. CHI (2011)
14. Quercia, D., Ellis, J., Capra, L., Crowcroft, J.: Tracking ”gross community happiness” from tweets. In: Proc. CSCW (2012)
15. Wattenberg, M., Kriss, J.: Designing for social data analysis. IEEE TVCG 12(4), 549–557 (2006)