

Tweetgeist: Can the Twitter Timeline Reveal the Structure of Broadcast Events?

David A. Shamma
Yahoo! Research
4301 Great America Pkwy
Santa Clara, CA USA
aymans@acm.org

Lyndon Kennedy
Yahoo! Research
4301 Great America Pkwy
Santa Clara, CA USA
lyndonk@yahoo-inc.com

Elizabeth F. Churchill
Yahoo! Research
4301 Great America Pkwy
Santa Clara, CA USA
churchill@acm.org

ABSTRACT

We explore applications for enriching experiences around live visual media events by leveraging conversational activity on short messaging services. We investigate the application of existing methods to discover the structure and content of media events and develop methods for exposing the discovered informational cues to users. We demonstrate these approaches using video footage and Twitter activity during two broadly watched media events: the first 2008 USA Presidential Debate and the Inauguration of Barack Obama. For the debates, we demonstrate a method for segmenting and annotating the media via conversational activity for the purpose of watching the video after the event has already passed. For the inauguration, we demonstrate approaches for exposing an awareness of the current topics of discussion on Twitter and the apparent levels of interest via a real-time feedback display. The primary contribution of this CSCW Horizons note is an initial exploration of approaches for applying cues mined from community conversation towards enhanced experiences around multimedia events and an invitation for a discussion of further approaches for applying these techniques to a growing number of domains and applications.

Author Keywords

Experience mining, TV, Twitter, inauguration, conversation, commentary, broadcast, event, participation, community, follower, social media

ACM Classification Keywords

H.5.3 Information Interfaces and Presentation: Group and Organization Interfaces—*Synchronous interaction; Collaborative computing*; H.5.1 Information interfaces and presentation: Multimedia Information Systems—*Video*

General Terms

Human Factors

INTRODUCTION

As digital video recorders become more prevalent and more television shows become available for on-demand streaming, many people are effectively watching television on a delay. However, there are a number of live media events, like award shows or championship games, where real-time consumption of the action is a compelling experience—people often gather together to share in the experience or communicate with each other online. Recently, the emergence of services like Twitter and Facebook, allows users to broadcast short messages in response to events that they're watching. In aggregate, this textual communication can convey a great deal about the content and structure of on-going live media events as usage patterns swell and recede and the textual content shifts to respond to the content coming through the television.

In this CSCW Horizon exploration, we investigate applications for mining activity on short messaging services during live media events. In particular, we explore existing methods for mining Twitter usage and develop metrics and applications in which the results of these approaches might yield richer experiences around video events. To do this, we leverage the structure and content conversation on Twitter to enable methods for archiving and revisiting media as well as surfacing emerging topics and trends in real-time.

We illustrate our approach with Statler [6]—for a full demo see: <http://bit.ly/statler>—a tool for identifying video content and in-situ commentary from community annotation. Instead of just inspecting or following a tweet stream that relates to a TV show, Statler examines the semantics and structure of the Twitter messages, in addition to the messages' content. Statler automatically, through the prescribed metrics we propose in this note, finds chapters, segments, important moments, and topics of momentary and sustained conversation.

RELATED WORK

Unlike traditional work that examines news content for segmentation [1], Statler explores the relationship between the news media and community annotation. Recently Shamma et al. [5, 4] demonstrated several methods that use Twitter messages about a media event to segment and topically identify trends. Statler presents these findings in a live application. While a few of these systems have been built, most notably the MTV Video Music Awards Twitter visualiza-

tion [2], Statler’s goal is not to show overall volume, but rather identify interesting moments within the tweet stream itself.

STATLER

Statler currently explores the first U.S. presidential debate of 2008 as well as the inauguration of Barack Obama. The debate tweets, ≈ 3000 , came from Twitter’s search interface and the inauguration tweets, ≈ 53000 , were pulled from Twitter’s constant rate data mining feed. Each sample represented 90 minutes of event footage in total. The two feeds varied in structure, we explore different features visible to each particular data set.

Presidential Debate

The debate interface consists of two main components, a table of contents (ToC) and topic trends, see Figure 1. The table of contents is generated automatically by examining tweet volume [5] and is wired to the video player. Since people often want at ToC but use the video player to find segments [7], Statler connects the two—scrubbing on the video moves the ToC’s highlighted region. For each segment, the ToC shows key terms, from the tweets and from the debate’s closed captioning. This gives insight into the content of the segment.

The Obama Inauguration

For the inauguration, Statler displays two aggregate metrics, trending terms and a world map of tweet locations, see Figure 2. Importance and Chatness are measurements to indicate event onset (is something really compelling happening right now?) and attention (are people watching the event or chatting amongst themselves). We also display the salient terms being used on Twitter for any given minute. At a given point in time, it calculates the frequency of each term during the previous two minutes and weighs that by the typical term frequency. These terms, which are being used more frequently at that time stamp, are then shown to capture the topics of discussion emerging on Twitter around each point in time. A small graph showing frequency over time can be displayed for any currently trending term.

METRICS, METHODS, AND DATA

A significant amount of information is encoded in simply *how* users are using the messaging service. This includes the rate at which users are sending messages, how long those messages are, or how conversational those messages are. Additionally, the content in aggregate provides insight into the media event itself.

Usage Cues

The raw volume of messaging activity at a particular point can reflect a varying level of collective interest in or attention to the media event. In particular, we observe a tendency for the number of messages regarding an event to either peak or valley at the points where topics change, which may reflect a propensity to discuss and react to the media in the lulls during changes in topic. To automatically detect these changes in topic, we obtain a signal of the total number of

messages related to the event sent during each minute. We find local extrema within this signal by applying Newton’s Method, which simply identifies the roots of the function, or points at which the slope of the function changes from positive to negative (for maxima) or from negative to positive (for minima). Only extrema greater than one standard deviation ($\mu \pm \sigma$) are observed. We further smooth the signal to remove spurious low-magnitude extrema to yield a more robust extrema detection.

Another usage-driven indicator is the actual length of messages. During each minute, we can calculate the average number of characters contained in each message sent. During the onsets of moments of interest, we observe a significant decrease in message lengths, which may reflect an increased attentiveness to the media event, and thus a desire to spend less time typing. Another indicator, which is correlated with message length, is the number of messages containing @ mentions (a Twitter convention, wherein users conversationally address each other by placing the @ symbol in front of the addressee’s user name). We similarly find that there is a significant decrease.

From the @ characters and the overall number of characters per Tweet, we aggregate metrics two metrics: Importance and Chatness. Each score is normalized between 0.0 and 1.0. The Importance metric is a function of how many @mentions were used on any given minute:

$$Importance = 1 - \frac{count_{minute}(@)}{count_{max}(@)}$$

More @ symbols means more conversation and, hence, people are paying less attention to the event. Similarly, how many characters typed per tweet per minute is used as a metric of how much time people spent typing their tweet that minute. More chatty means more to discuss or announce.

$$Chatness = 1 - \frac{count_{minute}(chars) - count_{min}(chars)}{count_{max}(chars) - count_{min}(chars)}$$

At Obama’s swearing in, we find few @ symbols in the stream where as later, people were more chatty, see Figure 3. Above the bar graphs, Statler displays a world map that highlights the origin countries of the tweets for that minute.

Content Cues

Beyond the usage practices, there is also knowledge that can be mined from what people are saying in their messages. To exploit this signal, we employ a simple term scoring approach similar to the well-known tf-idf model [3]. In tf-idf, the salience of a term in a particular document is given as a function of the number of times the term appears within the document (term frequency, or tf) normalized by the total number of documents in which the term appears (inverse document frequency, or idf).

Statler

Debate Inauguration



#	Time	Terms
1	0:00	Tweets: drinking, candidates, wins, minute, lehrer, letting, tv, mocalins, tweet, moderator CC: presidential, debates, minutes, eisenhower, financial, direct, policy, news, mississippi, university
2	2:21	Tweets: kennedy, plan, hope, hand, comment, moderator, tie, home, audience, republicans CC: street, greed, main, how's, house, package, accountable, wall, rewarded, crisis
3	13:15	Tweets: cut, earmarks, compared, tax, joke, bear, dollars, problems, personal, american CC: \$18, requests, earmarks, gateway, loopholes, \$5,000, pork-barrel, employer, business, tax
4	25:09	Tweets: screen, bailout, energy, economic, power, condulo, nuclear, giving, freeze, looked CC: programs, medicare, cost, eliminate, \$700, hard, decisions, agency, rescue, tom
5	38:07	Tweets: senate, difference, winning, strategy, iraq, festooned, fact, war, giving, hand CC: funding, winning, leave, timetable, violence, lessons, baghdad, surge, succeed, started
6	49:20	Tweets: pulling, pacistan, understand, drinking, republican, strategy, sounds, coming, telling, story CC: border, talban, prepared, supported, muddle, u.s., bombing, pakistan, qaeda, army
7	1:02:12	Tweets: lousy, democracies, lol, government, story, wars, league, times, iran, -5 CC: henry kissinger, contacts, preparation, legitimize, table, iranians, sanctions, ahmadinejad, precondition
8	1:14:56	Tweets: georgia, video, blog, league, russia, ha, lot, condescending, issue, oil CC: georgia, international, putin, ukraine, russia, world's, offshore, aggression, resurgent, nato
9	1:24:33	Tweets: security, experience, 9/11, safe, country, tactics, attack, video, pakistan, bringing CC: restore, knowledge, missile, safer, veterans, terms, focused, earth, billions, challenges

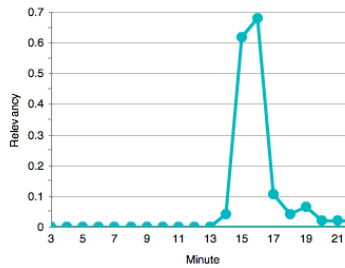
Figure 1. A screen shot of Statler watching the first U.S. Presidential Debate. The video is played (left) while an interactive Table of Contents (ToC) is displayed (right). The ToC contains automatically determined segments and shows terms from the tweets as well as from the debate's closed captioning. Seeking on the video marks the current segment in the ToC.

Statler

Debate Inauguration

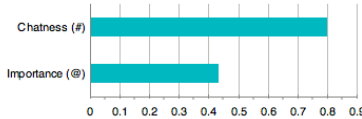


remaking's TF/IDF score over the inauguration
Current Minute: 13



Click on a term to see its frequency graph
(The top term's graph is displayed to the left.)

- remaking
- dust
- scripture
- greatness
- pick
- childish
- begin
- senior
- equal
- measure
- passed
- dogs
- tailed
- sooooo
- blogs
- chosen
- data
- stucco
- capacity



Above are two scores to show how chatty people were on twitter at this minute in the video and how important is this video minute. Chatness is a normalized score from 0.0 to 1.0 showing the average number of characters per tweet on that minute. Important is the inverse proportion of @ symbols in the tweets at the corresponding minute in the video.

Figure 2. A screen shot of Statler watching the inauguration of Barack Obama. The video is played (left) while a meter if importance and chatness is displayed (bottom middle) along with a rank list of tweeted terms (far right). The trend for the top ranking term of the current second is displayed next to the term list (center top). Below the video is a world map highlighting the countries tweeting at that minute.

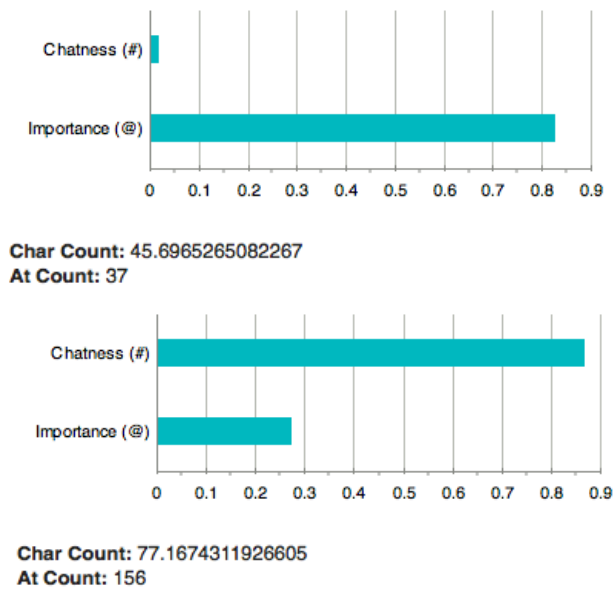


Figure 3. Top: Minute 12:05 of the inauguration was the swearing in of Barack Obama; it features a significant drop in @mentions ($> 2\sigma$). Bottom: Towards the end of Obama’s inaugural address, people were typing more and chatting amongst each other more.

In our data source, the notion of a document-model is quite different. We essentially construct make-shift documents out of temporally co-occurring messages. In some cases, we may have predefined temporal boundaries. For example, in the Presidential Debates, we may have boundaries occurring at topic changes, which might be annotated by hand or discovered automatically. We can then construct a document out of all of the messages sent during a given time frame. Terms can then be ranked according to their frequency within these pseudo documents normalized against their overall frequency.

More typically, though, no such topical boundaries are available. Here, we have applied a sliding window to gather messages within a few minutes of a given point in time, which allows us to identify terms that are instantaneously salient. To conduct such a calculation in real-time, we might construct a window of only the past few minutes or seconds, which might yield the most recently trending terms.

The changes in salient terms over time can also yield important cues about the content and structure of the media event. In particular, the occurrence of unique events of momentary interest may be reflected by a quick increase in saliency of a particular term (or a brief appearance by it in the top-salient terms), which quickly fades away. We can automatically detect such points by measuring the evolution of a single term’s saliency score over time. When a particular point in time has a term with a high-magnitude saliency score, we can infer that the point may be of interest and direct users towards it. The salient term, itself, may also be leveraged to give the user an understanding of why that moment is interesting. For example, during the inaugural address, terms used by

Obama in the speech are frequently echoed instantly in the salient terms from Twitter activity and then quickly drop off in interest.

Other topics may have a more sustained interest over time. This might be reflected by an increase in a term’s saliency that is sustained over a longer duration (i.e., it does not instantaneously drop back off). For example, during the inauguration, terms like “flubbed” and “messed” gain saliency in Twitter activity around the time of the oath of office, during which the orders of a few words were switched. These terms remained salient over a longer period of time as people continued to discuss the event. We can detect such occurrences automatically by finding the original onset peak of the event and then comparing the level of saliency after the onset against its saliency before. For an event of passing interest, we expect the saliency to dip back down to its previous level. For sustained interest events, we expect saliency to remain high.

APPLICATIONS FOR VISUALIZATION AND NAVIGATION

Given these tools for analyzing user messaging behavior during media events, we move on to exploring approaches for leveraging the results to better navigate, visualize, and consume video. We broadly categorize these approaches into two types. First, there is real-time feedback, where users might want to monitor the unfolding conversation in real-time while they’re still watching the event. Second, there is post-event consumption, where users return at a later time to view the video. Here, we can leverage knowledge of conversational activity to retroactively segment and annotate the video in order to enable users to navigate into areas of interest.

Real-time feedback

While watching a media event, users might be interested in gaining an ambient awareness of what conversations are unfolding online. To provide such an experience, we apply a number of the methods discussed in the previous section to the inauguration of Barack Obama. We simulate a real-time experience by mining only messaging activity that would be available at the moment as the event unfolds. During each second, we rank terms according to their saliency as calculated by the method, as previously discussed, over a window composed of the previous two minutes of activity. The top-20 terms are then updated and shown to the user as a reflection of topical interest on Twitter. A graph shows the saliency of the current top term over the duration of the event. Users can also click on any given term and see a similar graph of saliency versus time, which allows for term comparisons across the overall video, see Figure 4. A bar graph shows the normalized Importance and Chatness scores by the minute as the video plays. Additionally, a world map showing the origin of the tweets for the current video minute is displayed.

Post-event consumption

Revisiting a piece of media after the event, itself, has passed offers some opportunities for different types of interactions. There is no need to watch the whole video sequentially and

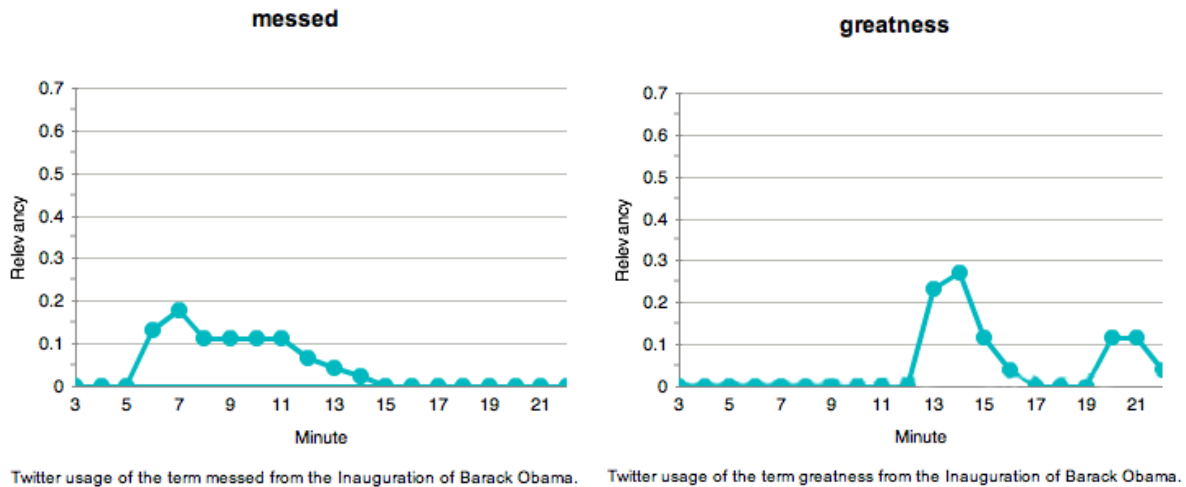


Figure 4. The trend from two terms from the swearing in and inauguration speech. Left: The conversation around Messed sustained and decayed from the onset of the fumbling of words during the swearing in of President Obama. Right: Usage of the term Greatness spiked at minute 12 and again at minute 20.

users may skip to any point in time. Also, the conversation on Twitter that emerged during the debate might be of less interest, since the event has passed and the user is unable to also join in the discussion.

For long pieces of media with clear topical segments, it may be useful to provide users with suggested entry points and some summarization of the content of various segments. The presidential debates are a good example of such an event. The entire debate is just over 94-minutes long, which is an atypical length of video to be consumed online. Also, the debate is structured into logical segments, where the topic of discussion is changed by the moderator. We present the video alongside an automatically-generated table of contents, where the video is segmented according to the discussed volume-based method. For each segment, we also provide the most salient terms. We provide two sets of terms, one found by mining terms on Twitter and another based on frequent terms from within the closed caption transcript of the event.

FUTURE WORK

Using Statler, we demonstrate how to understand content semantics derived from implicit community annotation. We have begun to explore metrics, based on current methodologies [5, 4], and application designs which can reveal these insights. Our metrics are derived from several existing approaches, such as tf-idf, as well as newer features, like Importance and Chatness. While we optimize these metrics for live applications, we will explore their applicability in genres outside of the news domain. This CSCW Horizons note invites conversation and research to address new domains, users, and communities (from journalists, to sports fans, to Drama TV watchers); different users may wish to see different trends or have new feature sets which point to event onset. Future research will allow for more synchronous, live, and meaningful interactions in this emerging space.

REFERENCES

1. C. Y. Low, Q. Tian, and H. Zhang. An automatic news video parsing, indexing and browsing system. In *MULTIMEDIA '96: Proceedings of the fourth ACM international conference on Multimedia*, pages 425–426, New York, NY, USA, 1996. ACM Press.
2. E. Rodenbeck. Kanye west is an a**hole and other twitter moments from the mtv video music awards. http://content.stamen.com/kanye_west_is_an_a__hole_and_other_twitter_moments, September 2009. Accessed October 12th, 2009.
3. G. Salton and M. McGill. *Introduction to modern information retrieval*. McGraw-Hill, Inc. New York, NY, USA, 1986.
4. D. A. Shamma, E. F. Churchill, and L. Kennedy. Shouting while watching: Describing a live media event through short message conversations. In *CHI 2010*. ACM, 2010. Submitted.
5. D. A. Shamma, L. Kennedy, and E. F. Churchill. Tweet the debates: Understanding community annotation of uncollected sources. In *WSM '09: Proceedings of the international workshop on Workshop on Social Media*, Beijing, China, 2009. ACM. *In press*.
6. D. A. Shamma, L. Kennedy, and E. F. Churchill. Statler: Summarizing media through short-messaging services. In *CSCW 2010*. ACM, 2010. Demonstration, *Submitted*.
7. E. G. Toms, C. Dufour, J. Lewis, and R. Baecker. Assessing tools for use with webcasts. In *JCDL '05: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital loibraries*, pages 79–88, New York, NY, USA, 2005. ACM Press.