Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media

Dancing with the Stars, NBA Games, Politics: An Exploration of Twitter Users' Response to Events

Ana-Maria Popescu and **Marco Pennacchiott**
Yahoo! Labs
Sunnyvale, USA
amp,pennac@yahoo-inc.com

Abstract

Microblogging services such as Twitter offer great opportunities for analyzing the reactions of a wide audience with respect to current events. In this paper, we explore the correlation between types of *user engagement* and events centered around *celebrities* (e.g., personal or professional events involving Actors, Musicians, Politicians, Athletes).

1 Introduction

The explosion of social media is allowing unprecedented access to the thoughts and reactions of a large audience in response to breaking news, from political developments to pop culture events. In this paper, we focus our attention on Twitter, one of the most successful social media services. Our goal is to study Twitter users' behaviors, in order to get a better understanding of the Twitter content and, support the development of effective tools for the enrichment of the social media experience. Our contributions are the following:

1. We provide a detailed study of a set of 509 celebrity events with respect to a set of event types (*life*, *work*, *social*, *media*) and a set of user engagement measures (tweet volume, mixed sentiment, explicit controversy, intensity). **2.** We describe a low-cost method for extracting and annotating user opinions about celebrities.

2 Corpus of Twitter Events

Event mining has a long history in the context of news collections and it has recently moved to the realm of social streams (Zhao, Mitra, and Chen 2007; Sayyadi, Hurst, and Maykov 2009; Popescu and Pennacchiotti 2010). In this work we adopt the definitions introduced by (Popescu and Pennacchiotti 2010), as follows.

Given a target entity, an **event** is defined as an activity or action with a clear, finite duration in which the entity plays a key role.

A **snapshot** is a triple $s = (e_s, \delta_s, tweets_s)$ where: e_s is the target entity, i.e. any type of concept or named entity (e.g. 'Barack Obama', 'earthquake'); δ_s is a time period (e.g. one hour, one day, one week); $tweets_s$ is the set of tweets from the target period, which refer to the target entity.

A snapshot s is **buzzy** if in the given time period the target entity is discussed more than in the recent past. A buzzy snapshot may indicate that an event involving the entity is happening, or it could be the result of spam, generic discussions or hoaxes. We therefore define two types of snapshots: **event snapshot** describes a specific event involving the target entity; **non-event snapshot** does not describe any event or only marginally involves the target entity.

Event Corpus. We build our event corpus from a Twitter firehose spanning from July 2009 to February 2010 and focus on tweets containing target entities from a list of about 100K Actors, Musicians, Politicians and Athletes scraped from Wikipedia (we remove entities whose names have less than 3 characters).

We first mine 738,045 buzzy snapshots. Then, we discard *irrelevant* snapshots by eliminating (1) snapshots with less that 10 tweets; (2) snapshots with more than 80% of tweets not in English; (3) snapshots where, on average, tweets have a percentage of overlapping tokens higher that 80%. The application of these filters leaves us with a final set of 73,368 snapshots. Finally, we apply the GBDT machine learning model in (Popescu and Pennacchiotti 2010) to classify the buzzy snapshots as being either *event* or *non-event* snapshots. The model is built by using a training set of 5040 buzzy snapshots manually annotated by a pool of human experts: 2249 events and 2791 non-events.

We evaluate the accuracy of the classification step by performing a 10-fold cross validation experiment over the training set of 5040 buzzy snapshots. Our model obtains a precision of 70.2% and a recall of 64.1.

3 Celebrity Events Analysis

This section seeks to answer the following questions:

1. What types of celebrities and events do people tend to discuss more in Twitter? (see Section 3.1)
2. What types of engagement with celebrity events do users exhibit? (see Section 3.2)
3. What comments do Twitter users have about celebrities? (see Section 3.3)

Previously, (Thelwall, Buckley, and Paltoglou 2011) tried to answer the question: "Why do particular events resonate with the population ?". While the resulting analysis is interesting, the paper focused on only 30 coarse-grained events (e.g. "Oscars"), while we analyze more than 500 fine-grained ones.

In order to derive the dataset for our study, we randomly sampled 720 snapshots from the automatically derived event corpus and asked two human experts to label them using one of these 4 *event types*:

- **Life**: a major personal life event, e.g. marriage, death.
- **Work**: a paid job-related activity, e.g. an actor selected for a movie.
- **Media**: a promotional media appearance, e.g. an actor giving an interview).
- **Social**: any type of other social activity, e.g. being arrested or contributing to a charity event.

The editors first performed a calibration annotation on a shared set of 20 snapshots, with overlap agreement of 92%. They then independently annotated the rest of 700 snapshots, 509 of these being actual event snapshots (confirming that the accuracy of our classifier is about 0.70). Results in the following are derived from the set of 509 event snapshots.

3.1 Entity and event types

Table 1 reports overall statistics of the annotation task.

Entity types. Actors are the most discussed celebrities (50% of all events), while Politicians are the least discussed (12%). Users are interested in everything Actors do, from professional developments to charity work. While Politicians also have high media exposure, we found that Twitter users are less interested in commenting on political developments (with the exception of non-US English speaking users).

Event types. Celebrities are mainly discussed in Twitter because of work-related events, followed by media appearances. Life and social events together constitute a mere 21% of the event snapshots. We had expected Twitter users to talk more about celebrity life events (e.g., Brad Pitt and Angelina Jolie getting married) or social activities. Instead, users seem highly engaged with *micro-events*, i.e. events of a smaller scope which do not always warrant significant news coverage (live performance on a tv show, walking on a red carpet).

We also asked editors for a short event description. In most cases, *work* events take the form of paid participation in a TV show. An example is 'Dancing with the Stars', which attracts a vast audience of both TV viewers and people commenting on Twitter. Other work-related events include: being nominated for or winning an award (especially for Actors), live performances (mostly for Musicians and stand-up comedians) and performing an actual job-related activity (the majority of work-related events for Athletes and Politicians). Most of *media* events (especially for Actors) are red carpet appearances, which are highly popular with Twitter users interested in celebrities' appearance and fashion choices. As for *life* events, most are birthdays, followed by engagement and death announcements. Finally, we find a large variety of *social* events, especially crimes and allegations (e.g., driving under the influence, assaults) or humanitarian endeavours (e.g. participating in a telethon). Surprisingly, very few gossip items are discussed (e.g., celebrities dating), suggesting that users may follow such stories in other venues (e.g., entertainment news sites).

	work	**media**	**life**	**social**	*total*
Actor	122	80	26	29	*257*
Athlete	79	1	2	8	*90*
Musician	61	11	17	14	*103*
Politician	35	17	4	3	*59*
Total	*297*	*109*	*49*	*54*	*509*

Table 1: Overall results of the 509 annotations.

Event type distribution per entity type. Actors and Politicians are mostly discussed in conjunction with work and media events (e.g., red carpet appearances for Actors; tv interviews for Politicians). Musicians also have a high percentage of work events (e.g., musical performances), while athletes are predominantly discussed in the context of an athletic performance (e.g., in real-time, during a game) rather than in the context of publicity appearances.

3.2 User Engagement with Events

In this section we use 4 content-based *user engagement measures* in order to analyze the user response to celebrity-centered events:

Volume: The number of tweets in the snapshot.

Intensity index: Estimates the level of user emotional involvement. First, we compute 2 tweet-level intensity measures for each tweet t in the snapshot. $C(t)$ is the % of characters which are question marks, exclamation points, part of an emoticon or a sequence of repeated characters which typically indicate emphasis or excitement (e.g. "fiiine"). $O(t)$ is the % of tweet tokens which are strongly positive or strongly negative subjective entries in OpinionFinder 1.5 (Wiebe and Cardie 2005)(e.g., "awful", "awesome"). Given the set S of tweets in a snapshot, the intensity index is $\sum_{i=0}^{|S|} \frac{C(t_i)+O(t_i)}{2}$.

Sentiment index: Estimates how *mixed* the sentiments in the users' reactions are. The index is derived by first computing the polarity of each tweet in the snapshot with respect to the target entity. Given a window of 4 words on the right and the left of the target entity, we identify positive and negative words using the OpinionFinder 1.5. sentiment lexicon (and handling negation tokens). If the window contains only positive words, the tweet is positive wrt the target entity (negative tweets are similarly identified). Given the sets of tweets with positive (P), negative (N) and neutral (U) sentiment, the index is: $\frac{Min(P,N)}{Max(P,N)} \cdot \frac{P+N}{P+N+U}$.

Controversy index: A measure that estimates how controversial the event is. The index is computed as the % of tweets in the snapshot containing at least one term from a controversy dictionary in (Popescu and Pennacchiotti 2010) containing 750 controversy terms (e.g., 'scandal' and 'outrageous').

entity	date	volume	event type	entity type	event description
alec baldwin	2010/03/08	1421	work	actor	Presents the Oscars
david tennant	2010/01/01	1301	work	actor	Quits TV series
chris brown	2009/08/25	1189	social	musician	Sentenced to community service
christoph waltz	2010/03/08	972	work	actor	Wins an Oscar prize
barack obama	2009/10/09	846	work	politician	Wins Nobel prize
barack obama	2010/01/28	820	work	politician	Speech at the State of the Union

Table 2: Highest volume event snapshots in our dataset.

Results. Table 2 shows the highest **volume** snapshots in our dataset. The top 20 highest volume snapshots are all *macro-events* (general interest events reported in conventional media) with at least one related news article, while only 7 among the 15 lowest ranking events are *macro-events*. The average volume per snapshot is 83 tweets. The rank of the events mostly discussed in Twitter does not necessarily match the order of relevance based on more conventional media coverage: 'David Tennant' quitting the 'Doctor Who' TV series barely makes the news, but it has great resonance with the Twitter audience.

Analysis by entity types. Figures 1-2-3 report the average values of the intensity, sentiment and controversy indexes for the different entity types. Specifically, the plots report the percentage of snapshots whose index value exceeds various thresholds. High levels of user reaction **intensity** (Figure 1) are mainly the province of Actors and Musicians in the context of birthdays and interviews on talk shows. The emotional intensity for Politicians is the lowest (people discussing political issues tend to be less excited than those reacting to an idol's birthday).

As seen in Figure 2, the **sentiment** index is similar across the 4 celebrity classes, with a boost for Athletes: e.g., 41% of the Athletes' snapshots have a sentiment index of 0.2 or higher, compared to 29% of the Actors. This is usually due to fans on opposite sides of a game or a match involving two teams. We also examined the % of tweets with positive and negative polarity for each class. Surprisingly, users tend to positively comment on events much more than negatively. We would have expected more balanced sentiments, especially for Politicians, since people tend to have clear opinions on political issues. Our hypothesis is that Twitter is skewed towards positive sentiments: people with negative opinions may tend to remain 'silent' instead than posting a negative comment. To investigate, we collected positive and negative emoticons from millions of tweets collected from Aug. 2009 and Dec. 2010. We found that positive emoticons occur 6 times more frequently than negative ones. Twitter is, in general, an example of 'positively skewed' media.

The **controversy** index indicates that Politician events tend to be highly controversial - Politicians seem to be commented upon when they do something unbecoming (e.g., senator Charles Schumer getting into a verbal altercation with a flight attendant).

Analysis by event types. Figures 4-5-6 report the index values by event type. Social events are characterized by both a high sentiment and high controversy index, but by a lower intensity score. We attribute this to social events involving celebrities being arrested or charged and therefore being in-

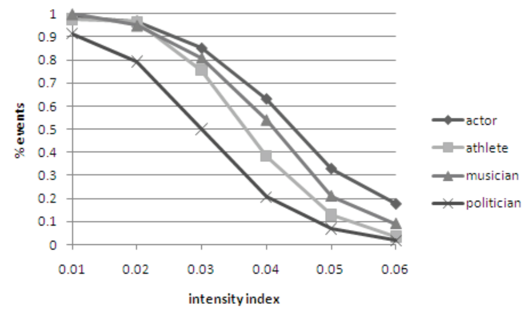

Figure 1: Percentage of events for each celebrity class, varying the intensity score.

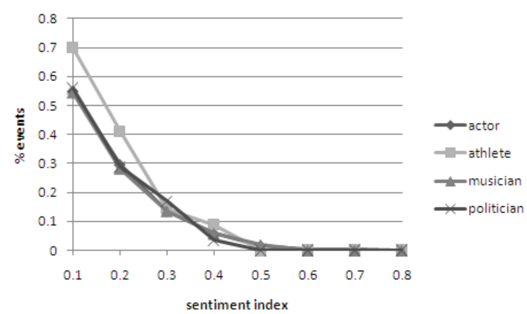

Figure 2: Percentage of events for each celebrity class, varying the sentiment score.

trinsically polarizing. However, the low intensity score suggests that while disagreeing about such events, Twitter users are not overly preoccupied by them (see Table 2). Work events also attract slightly mixed sentiments, but have low percentages of controversy terms, indicating that any implicit controversy manifested in the mixed sentiment score for musical performances or actor cameos is short-lived, rather than a lasting controversy which is labeled as such.

Finally, life events (especially birthdays) and media events (especially talk show interviews and red carpet appearances) lead to the highest intensity responses. Life events - birthdays, deaths - are also characterized by either outpourings of positive or negative sentiment (leading to low mixed sentiment).

3.3 User Comments about Celebrities

We use a low-cost but effective solution to the problem of automatically extracting user opinions about the celebrity at the core of the event. First, tweets are annotated with

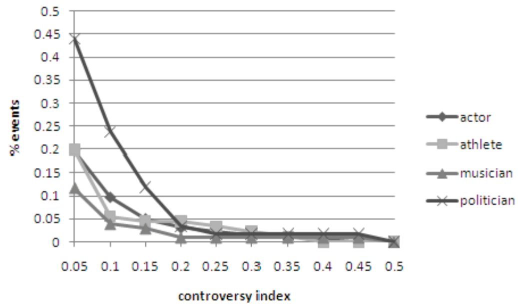

Figure 3: Percentage of events for each celebrity class, varying the controversy score.

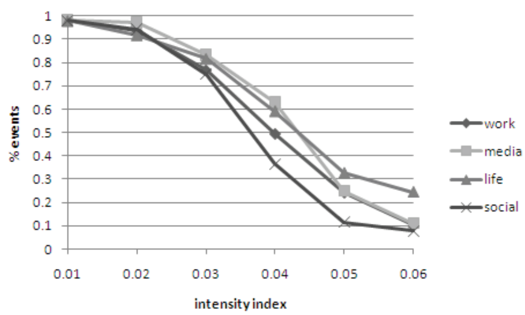

Figure 4: Percentage of events for each event type, varying the intensity score.

POS tags by using the off-the-shelf Brill tagger (Brill 1995). Opinion extraction is then performed by using two types of regular expressions:

(1) the verbs *be*, *look* and *seem* preceded by a target entity, and followed by either a noun or adjective phrase representing a candidate opinion, e.g. 'Barack Obama is *my hero*'.

(2) the pronoun *I* followed by a verb phrase representing the candidate opinion, and then the target entity, e.g. 'I *hate* Julia Roberts'. We allow interleaved particles in the sequence to improve recall.

We then classify each potential opinion phrase using a sentiment-dictionary lookup (Wiebe and Cardie 2005). If it contains a sentiment word, we classify it accordingly as positive or negative; otherwise it remains neutral. To improve coverage, edit distance is used to map misspelled words to dictionary entries (e.g. 'prettay' to 'pretty').

Evaluation Opinion extraction is evaluated by collecting 600 random opinions from the corpus, and checking the accuracy of the positive/negative classification; we also check if the extracted opinion is grammatically sound, so to assess the reliability of the POS tagger and the regular expressions. Results show that 85% opinion phrases are grammatically sound. Out ot these, 78% are correctly spotted by the dictionary, with an accuracy of 84%. In some cases, Twitter phenomena such as misspellings which mimic conversational style (e.g., "perrty") cannot be handled by our simple edit-distance metric, event though the stem word is in the dictionary ("pretty", 'faded"). In other cases, comparisons ("like a

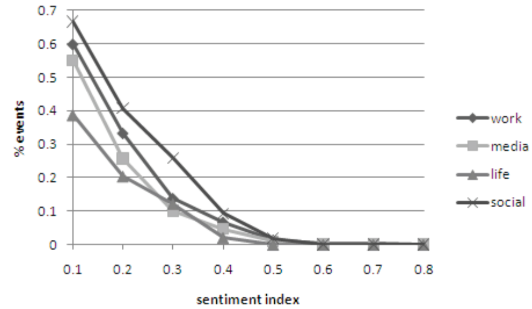

Figure 5: Percentage of events for each event type, varying the sentiment score.

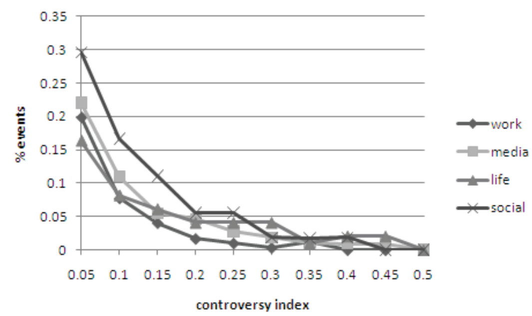

Figure 6: Percentage of events for each event type, varying the controversy score.

frog") or choice collocations ("a hot mess") need improved handling.

In conclusion, this paper presented a detailed study of how Twitter users react to different types of celebrity events. We find that *micro-events* (e.g, performance in a game) are an important subset of celebrity events, underscoring the "watercooler" aspect of Twitter discussions. While *work* events represent the bulk of the developments in our set, *life* and *social* events elicit the highest degree of user engagement.

References

Brill, E. 1995. Transformation-based error-driven learning and natural language processing: A case study in part of speech tagging. *Computational Linguistics* 21:543–565.

Popescu, A.-M., and Pennacchiotti, M. 2010. Detecting Controversial Events from Twitter. In *Proc. of CIKM*.

Sayyadi, H.; Hurst, M.; and Maykov, A. 2009. Event Detection and Tracking in Social Streams. In *Proc. of ICWSM*.

Thelwall, M.; Buckley, G.; and Paltoglou, G. 2011. Sentiment in Twitter events. In *J.Am.Soc.Inf.Sci.*, 406–418.

Wiebe, J., and Cardie, C. 2005. Annotating expressions of opinions and emotions in language. In *Language Resources and Evaluation*.

Zhao, Q.; Mitra, P.; and Chen, B. 2007. Temporal and Information Flow Based Event Detection from Social Text Streams. In *Proc. of WWW*.