Extracting Events and Event Descriptions from Twitter

Ana-Maria Popescu
Yahoo! Labs
Sunnyvale, CA, 94089
amp@yahoo-inc.com

Marco Pennacchiotti
Yahoo! Labs
Sunnyvale, CA, 94089
pennac@yahoo-inc.com

Deepa Arun Paranjpe
Yahoo! Labs
Sunnyvale, CA, 94089
deepap@yahoo-inc.com

ABSTRACT

This paper describes methods for automatically detecting events involving known entities from Twitter and understanding both the events as well as the audience reaction to them. By using natural language processing techniques, we show that we can extract events with encouraging results, and reliably detect the main entities involved in the events and the audience reactions.

Categories and Subject Descriptors

I.2.6 [**Artificial Intelligence**]: Learning—*knowledge acquisition*

General Terms

Algorithms

Keywords

social media, information extraction, opinion mining, twitter

1. INTRODUCTION

Social media has become in recent years an attractive source of up-to-date information and a great medium for exploring the types of developments which most matter to a broad audience. Recent work has included sentiment analysis in social media [2], mining coherent discussions on particular topics between social actors [5] and mining controversial events centered around specific entities [4]. This paper builds on the work in [4] by focusing on detecting events involving known entities from Twitter and understanding both the events as well as the audience reaction to them. More specifically we show that: (1) *Events* centered around specific entities can be extracted with encouraging results (70% precision and 64 % recall); (2) *Main entities* for the event can be reliably extracted and good quality *entity actions* for these entities can be found, providing a good initial summary for the event; (3) A simple lexicon-based model for *opinion identification* performs well for understanding the audience response to a target entity and to the event. In the following we delve into each of these areas in more detail.

2. EVENT EXTRACTION

Definitions. Following [4], we focus on detecting events involving known entities in Twitter data. Let a snapshot denote a triple $s = (e, \delta_t, tweets)$, where e is an entity, δ_t is a time period and $tweets$ the set of tweets from the time period which refer to the target entity. Events are defined as activity or action with a clear, finite duration in which the target entity plays a key role.

Task and methods. Given a snapshot s of an entity e, our task is to decide whether the snapshot describes a single central event involving the target entity or not (e.g., is a generic discussion, or refers to many events with no clear main one). Following [4], we formulate this problem as a supervised Machine Learning (ML) problem and use the Gradient Boosted Decision Trees framework to solve it. We investigate two learning models:

EventBasic is a supervised classification method which represents each potential event snapshot using the large set of Twitter-based and external features described in [4] (e.g., number of action verbs, entity buzziness in Twitter on the given day, entity buzziness in news on the given day, etc.)

EventAboutness is a supervised classification method which augments the feature set of *EventBasic* as follows: we use a highly scalable *document aboutness* system [3] (see Section 3 for a brief description) in order to rank the entities in a snapshot with respect to their relative importance to the snapshot. We construct additional features based on such entities' importance scores in order to capture commonsense intuitions about *event* vs. *non-event* snapshots: most *event* snapshots have a small set of important entities and additional minor entities while *non-event* snapshots may have a larger set of equally unimportant entities (e.g. in the case of spam tweets which simply list unrelated entity names, etc.). Feature include the mean and std.dev. of the top 3 importance scores, the std.dev. of the target entity score from the mean of the top 3 scores.

Evaluation. We use a gold standard of 5040 snapshots which have been manually classified as *events* (2249) or *non-events* (2791). As a result, a baseline which would classify all snapshots as events would give a 0.45 precision. Table 2 summarizes the performance of the 2 versions of our event detection pipeline. While the systems are close in performance, *EventAboutness* performs best, with 0.70 precision and 0.60 recall. When inspecting the features ranked by importance by the GBDT framework, 1 aboutness feature appears in the top 10 (the st.dev. of the top 3 scores) and 2 additional ones in the top 20 (the standard dev. of the top score and the average of the top 3 scores). The most useful feature for both *EventBasic* and *EventAboutness* is the % of snapshots tweets which contain an action verb, while other useful features include the buzziness of an entity in the news on the given day and the number of reply tweets.

3. MAIN ENTITY EXTRACTION

In order to identify main entities, we use an off-the-shelf, highly scalable, document aboutness system [3]: The system relies on a large dictionary (27 million phrases) for identifying potential entities, and contains inflectional and lexical variants so that surface

WWW 2011, March 28–April 1, 2011, Hyderabad, India.
ACM 978-1-4503-0637-9/11/03.

Snapshot	*Julia Roberts, 2010-01-28, Golden Globes attendance*	*Jyoti Basu, 2010-01-17, Death*
Example Tweets	"julia roberts looks absolutely stunning! .." "lol julia roberts is faddeddddd" "I may have had one too many white russians but doesn't julia roberts look like madge?" "#goldenglobes julia roberts presenting the best picture award 2 avatar. me sooo sad" "tom hanks and julia roberts bustin on nbc and their late night decisions - nice!"	"@BDUTT:jyoti basu died at 96,so sad,he missed 100." "comrade jyoti basu died.. donated his whole body ...." "#news jyoti basu has passed away. biman bose made the announcement at AMRI Hospital Kolkata" "BDUTT is jyoti basu no more?" "@ItsProHere who is jyoti basu??"
Main entities	julia roberts, golden globes	jyoti basu, biman bose,amri hospital,kolkata
Audience opinions	+ julia roberts : absolutely stunning - julia roberts : faddeddddd - julia roberts : like madge + julia roberts : so kool	+ jyoti basu : great personality + jyoti basu : pioneer = jyoti basu : communist + jyoti basu : true example
Main entities' actions	julia roberts : presenting : best picture award julia roberts : bustin : on nbc julia roberts : sitting by : sir paul	jyoti basu : died : at 96 jyoti basu : donated : his body biman bose : made : the announcement

Table 1: Examples of event snapshot descriptions output of our system.

System	P	R	F-1	Avg P	AROC
EventBasic	0.691	0.632	0.66	0.751	0.791
EventAboutness	**0.702**	**0.641**	**0.67**	**0.752**	**0.788**

Table 2: Performance of event detection from Twitter.

System	MRR	Prec@1	Prec@3	Prec@5
TF-IDF	0.956	0.676	0.826	0.873
ML Aboutness	0.965	0.682	0.836	0.882

Table 3: Performance of main entity extraction.

forms corresponding to the same entity are grouped together. The *aboutness* computation system solves the classic term relevance problem defined as follows:

Let $T = t_1, t_2, t_3...$ be the set of terms in the Twitter snapshot s. The *aboutness* of the snapshot is the set A of (term t_i, score sc_i) tuples s.t.:

$$A = \{(t_i, sc_i) \mid t_{i-1} \succ t_i, sc_i > sc_{i-1}, t_i \in T\} \qquad (1)$$

where $x \succ y$ represents x is more relevant than y and x should be ranked higher than y. We acquire the snapshot's *aboutness* description by using a ML approach that learns to score and rank snapshot terms based on implicit user feedback available in search engine click logs. The feature set includes relative positional information (e.g. offset of term in snapshot), term-level information (term frequency, Twitter corpus IDF), snapshot-level information (length of snapshot, category, language), etc.

Evaluation We evaluate the performance of the system for extracting the main entities, using a gold standard of 200 snapshots with an average of 30 tweets, each annotated by editors with their set of main entities. We use two measures for evaluating the entities ranked by the system: mean reciprocal rank (MRR) and average precision at several ranks. MRR helps in finding out how early in the system's ranked list of entities, we capture the first main entity provided by the editors. However, most snapshots have more than one main entity. We then use a version of average precision that computes the fraction of the entities in the gold standard per snapshot covered in the top k terms in the ranked list. Results are reported in Table 3, showing that our system improves over a baseline tf-idf system.

4. EXTRACTING ACTIONS AND OPINIONS

In a final step, for each snapshot we extract relevant actions performed by main entities, and we extract and classify audience opinions about these entities. The system takes as input the tweets of an event snapshot and its list of main entities, and performs Part of Speech (PoS) tagging on the tweets, using an off-the-shelf tagger [1]. Then, it applies regular expressions over the obtained PoS-sequences to extract entities' related information. Our approach is deliberately shallow, to reduce execution time and because the noisy, short and sparse nature of tweets discurages the use of more advances approaches. **Action extraction** is performed by extracting sequences where the entity is followed by a verb and then by a noun phrase (e.g. 'david duchovny showed up at the globes'). All such sequences are retained as entities'action, without performing any frequency-based filtering, as the signal in Twitter is to small to apply such techniques. Yet, in the future we plan to focus on specific verb semantic classes to improve accuracy, and on nominalizations to improve coverage (e.g. 'Joe Biden *appearance* at State of Union'). Our method extracts in average 8 actions per snapshot. We evaluate by providing editors a snapshot and the extracted actions, and asking them if the actions are grammatical and if they summarize appropriatly the main happenings in the event. Results over a sample of 50 snapshots show that 77% actions are grammatical, and that for 68% of snapshots they provide an appropriate summarization.

Audience opinion extraction is performed by using two types of regular expressions: (1) the verbs *be*, *look* and *seem* preceeded by a main entity, and followed by either a noun or adjective phrase representing the user's opinion, e.g. 'Barack Obama is *my hero*'. (2) the pronoun *I* followed by a verb phrase representing the opinion, and then a main entity, e.g. 'I *hate* Julia Roberts'. We allow interleaved particles in the sequence to improve recall. We then classify each opinion by a sentiment-dictionary lookup [4]: if an opinion contains a sentiment word, we classify it accordingly as positive or negative polarity; otherwise neutral. For example 'Judie Law is *quite gorgeous*' is classified as a positive opinion since 'gorgeous' is a positive word in the dictionary. To improve coverage, edit distance is used to map mispelled words to dictionary entries (e.g. 'prettay' to 'pretty'). Opinion extraction is evaluated by collecting 600 random opinions from the corpus, and manually checking if the sentiment classification is correct; we also check if the extracted opinion is grammatically sound, so to assess the reliability of the tagger and the regular expressions. Results show that 85% of opinion are grammatical. Out ot these 78% are correctly spotted by the dictionary, with an accuracy of 0.84.

5. REFERENCES

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

[2] M. Choudhury, H. Sundaram, A. John, and D. Seligmann. Multi-scale characterization of social network dynamics in the blogosphere. In *Proc. of CIKM*, pages 1515–1516, 2008.

[3] D. Paranjpe. Learning document aboutness from implicit user feedback and document structure. In *Proc. of CIKM*, 2009.

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

[5] Q. Zhao, P. Mitra, and B. Chen. Temporal and information flow based event detection from social text streams. In *Proc. of WWW*, 2007.