

Cross-genre Event Extraction with Knowledge Enrichment

Hao Li and Heng Ji

Computer Science Department
Rensselaer Polytechnic Institute
jih@rpi.edu

Abstract

The goal of Event extraction is to extract structured information of events that are of interest from unstructured documents. Existing event extractors for social media suffer from two major problems: lack of context and informal nature. In this paper, instead of conducting event extraction solely on each social media message, we incorporate cross-genre knowledge to boost the event extractor performance. Experiment results demonstrate that without any additional annotations, our proposed approach is able to provide 15% absolute F-score improvement over the state-of-the-art.

1 Introduction

The rapid development of social media and social networks since 2000s has made it an important channel of information dissemination. Because of its real-time nature, social media can be used as a sensor to gather up-to-date information about the state of the world. Effective automatic detection and extraction of events from the media will be an extremely important contribution. Recently there has been increasing interests in event extraction from social media (Yang et al., 1998; Kleinberg, 2003; He et al., 2007; Weng and Lee, 2011; Benson et al., 2011; Ritter et al., 2012).

Identifying and extracting events in social media is more challenging than traditional event extraction due to two major reasons: (1). **Lack of Context:** compared with traditional genres (e.g., news articles), social media context is usually short and incomplete (e.g., each tweet has a length limitation of 140 char-

acters). Lacking of context, a single tweet itself usually cannot provide a complete picture of the corresponding events. For example, for the tweet “*Pray for Mali - the situation is coming to light, and it isn’t pretty.*”, an event extraction/discovery system (e.g. (Ji and Grishman, 2008)) fails to discover that it is about the same *war* event in *Mali* as mentioned in the news article “*State military forces on Friday retook a key town in northern Mali after intense fighting that included help from French military forces, a defense ministry spokesman said.*” (2). **Informal Nature:** social media messages are written in an informal style, which causes the poor performance of event extractors designed for formal genres. For example, the tweet “*#AaronSwartz, Dead @ 26, #CarmenOrtiz “pushed him to exhaustion” don’t let her get away with this! #scandal.*” includes an “*Die*” event with *Aaron Swartz* as the *Victim*. However, the person name “*AaronSwartz*” appears in the hashtag “*#AaronSwartz*” and “*Dead at 26*” is written as “*Dead @ 26*”. Existing supervised name taggers and event extractors fail to identify the same “*Die*” event mentioned in the news article “*Internet activist Aaron Swartz dead at 26*”.

Based on the intuition that news articles contain more detailed and formal information than tweet messages, we apply an unsupervised knowledge enrichment algorithm to link each tweet to its most relevant news article. By incorporating the cross-genre knowledge to tweets, we are able to formulate the task of event extraction on tweets as the task of cross-genre extraction for tweets and news articles. Thus we can alleviate the previous mentioned challenges in single-genre event extraction for tweets to

t_1 : Crowds rally in Belfast for flag protest#thetruth
n_1 : Protesters march in Northern Ireland: Under a gray, overcast sky, more than 1,000 protesters gathered Saturday in the Northern Ireland city of Belfast carrying large Union flags, some wrapped around their shoulders.
te_1 : {EventPhrase = [rally, protest], EventArgument = Belfast, Time = 2013-01-14}
ne_1 : {EventType = Conflict, EventPhrase = march, EventArgument = [protesters, Northern Ireland, Belfast]}
e_1 : {EventType = Conflict, EventPhrase = [rally, protest, march], EventArgument = [protesters, Northern Ireland, Belfast], Time = 2013-01-14}.

Table 1: Cross-genre Event Extraction Example

some extent. To the best of our knowledge, it is the first work to conduct cross-genre event extraction through unsupervised knowledge enrichment.

2 Problem Definition

Given a tweet t_i , our cross-genre event extraction framework first discovers its most relevant news article n_i , then identifies event tuples (event phrases and event arguments) for the tweet (te_i) and the news article (ne_i) respectively, and finally conducts merging on the event extraction outputs from both genres to produce the cross-genre event extraction result (e_i). For example in Table 1, given the following tweet t_1 , n_1 is retrieved as its most relevant news article. te_1 and ne_1 are the extracted event tuples for the tweet and the news article respectively and e_1 is the final cross-genre event extraction output after merging. To evaluate the performance of an event extractor, the precision, recall and f-measure of the extracted event phrases and event arguments will be measured using the following criteria: an event phrase is correctly labeled if it matches a reference trigger; an argument is correctly labeled if it matches a reference argument.

3 Approach

3.1 Baseline Event Extraction Systems

We use two state-of-the-art event extraction systems (Ritter et al., 2012; Li et al., 2013) to extract events from tweets and news articles respectively. The tweet event extractor TwiCal-Event (Ritter et al., 2012) is able to extract open-domain significant

events from Twitter. It is a supervised system that identifies event phrases and event participants with tailored part-of-speech tagging and shallow parsing for tweets. In addition, it is also able to discover event categories and classify extracted events based on latent variable models. It takes tweets as input and outputs a four-tuple representation of events which includes event participants, event phrase, calendar date, and event type. The news event extractor (Li et al., 2013) is a joint framework based on structured prediction which extracts triggers and arguments simultaneously while incorporating diverse lexical, syntactic, semantic and global features. It takes raw documents as input, distinguish events from non-events by classifying event triggers and identifying and classifying argument roles.

3.2 Knowledge Enrichment Approach

To produce the latent vector representations for the whole dataset, we follow the same procedure in (Guo et al., 2013): represent the dataset in a matrix X , where each cell stores the TF-IDF values of words. Word vectors P and tweet vectors Q are optimized by minimizing the following objective function:

$$\sum_i \sum_j W_{ij} (P_{\cdot,i} \cdot Q_{\cdot,j} - X_{ij})^2 + \lambda \|P\|_2^2 + \lambda \|Q\|_2^2 + \delta \cdot \left(\frac{Q_{\cdot,j_1} \cdot Q_{\cdot,j_2}}{|Q_{\cdot,j_1}| |Q_{\cdot,j_2}|} - 1 \right)^2$$

$$W_{i,j} = \begin{cases} 1, & \text{if } X_{ij} \neq 0 \\ w_m, & \text{if } X_{ij} = 0 \end{cases} \quad (1)$$

where λ is a regularization term, Q_{\cdot,j_1} and Q_{\cdot,j_2} are linked pairs connected by text-to-text relations, $|Q_{\cdot,j}|$ denotes the length of vector $Q_{\cdot,j}$ and the coefficient δ denotes the importance of the text-to-text links. we follow the same optimization procedure as (Steck, 2010) by alternating Least Square [ALS] is used for inference on P and Q .

After obtaining the vector representations for the whole dataset, for each tweet, we retrieve its cross-genre knowledge by finding the news article with the highest cosine similarity.

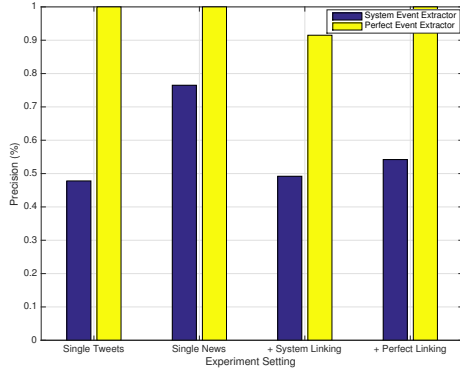


Figure 1: Event extraction precision

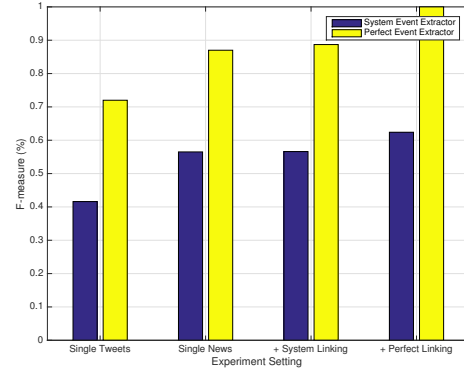


Figure 3: Event extraction f-measure

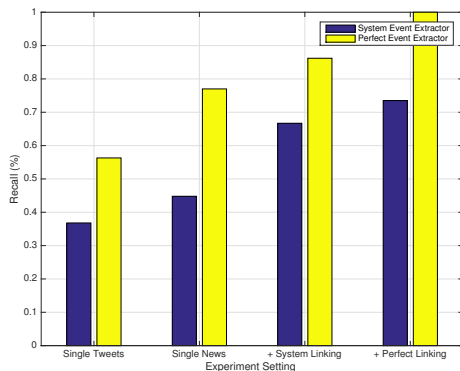


Figure 2: Event extraction recall

4 Experiments

4.1 Data Description

We use the same dataset as (Guo et al., 2013) which contains 34,888 tweets and 12,704 news articles. For each tweet, we consider the url-referred news article as its gold standard cross-genre knowledge – the most relevant news document. As the news event extractor is designed for a closed set of 33 event types (ace, 2005) while the tweet event extractor is for open domain, in this paper we only focus on the tweet-news pairs that the news event extraction output is not empty. We randomly selected 50 tweet-news pairs for the cross-genre event extraction annotation and evaluation.

4.2 Experiment Results

Figure 1, 2 and 3 present the Precision (P), Recall (R) and F-measure (F) of the overall event extraction performance for different settings respectively.

From these experiment results, we have the following four observations:

1. Event Extraction solely on tweet messages achieves the lowest precision, recall and f-measure. It exactly confirms our motivation of conducting knowledge enrichment for event extraction in tweet messages. Because of the informal nature of tweet messages, the event extractor misidentified 54.16% of the events thus the precision is low. Take the ill-formatted tweet in Section 1 as an example, the named entity “AaronSwartz” is in the hashtag “#AaronSwartz” and “Dead at 26” is written as “Dead @ 26”. It makes the automatic event extractor extremely difficult to identify the “Die” event for the person “Aaron Swartz”.

The low recall is mainly caused by the “lack of context” problem. Due to the length limitation of tweets, users tend to use recapitulate languages to describe an event. For Example, the following tweet “Well. That sucks. *Deepening Crisis for the Boeing 787*” actually refers to an “emergency landing” event made by “All Nippon Airways” in “western Japan”. The user only mentioned the summary “Deepening Crisis for the Boeing 787” to refer to the actual event. Therefore, the single-genre event extractor missed the event trigger “landing” and the event arguments “All Nippon Airways” and “western Japan”.

2. Both single-genre event extractors can contribute to the cross-genre event extractor through the cross-genre linking process. Even with the automatic linking output, Figure 2 and Figure 3 show 29.9% recall and 14.0% f-measure improvement

over single-genre event extractor for tweets. It is because that in most cases, news articles cover more information than tweets as they are produced by professional news agencies while tweets are written by individuals with a 140-character length limitation. In the following example, although referring to the same events, the news article covers more event information than the tweet, thus the cross-genre event extractor can surpass the single-genre event extractor for tweets.

News: *Deepening Crisis for the Dreamliner The two largest Japanese airlines said they would ground their fleets of Boeing 787 aircraft after one operated by All Nippon Airways made an emergency landing in western Japan.*

Tweets: *Well. That sucks. "Deepening Crisis for the Boeing 787"*

For some certain cases, the cross-genre event extractor is also able to benefit from single-genre event extractor for tweets. For the following example, the event extractor for news articles missed the "Attack" event as "Halt an ... Advance" is an unusual phrase to describe an "Attack" event. However, in the related tweet, "Battling" is a strong indicator of an "Attack" event and the event extractor for tweets is able to catch it. As a result, we are able to extract the overall event tuples {EventPhrase=[Advance,Battling], EventParticipant=[France,Islamist,Mali]}.

News: *French Troops Help Mali Halt an Islamist Advance France answered an urgent plea from the government of its former colony to help blunt an advance into the center of the country by Islamist extremist militants.*

Tweets: *France Battling Islamists in Mali #rebels tarnish W. African #Islam Western #intervention*

3. Either improving single-genre event extractors or achieving better cross-genre linking performance is able to boost the overall event extraction performance. From Figure 1, 2 and 3, we can observe that higher quality of single-genre event extractors will significantly enhance the precision while better linking performance will mainly contribute to a higher recall.

4. Compared with cross-genre linking accuracy, the quality of single-genre event extractors is more important. Figure 3 shows that "Gold Event Extractor + System Linking" achieved 26.3% higher F-measure score than "System Event Extractor + Gold

Linking". It is mainly because of two reasons: on one hand, the errors of single-genre event extractors will be propagated to the final event output; on the other hand, the current linking system is able to provide reasonable linking results thus the use of perfect linking will not have too much gain.

5 Remaining Challenges

Linking Errors: mistakenly linking tweets to irrelevant news articles. For example, the tweet "*The lack of investigative movement - his return and his flippant attitude is what is insulting. Not his new placement.*" is about a "Movement" event that Assemblyman returns to Albany after scandal. However, the tweets express the event so implicitly that the automatic linking system is not able to discover its corresponding news article.

Extraction Errors: single-genre event extractors failed on both the tweet side and the news article side. For the following example, both single genre event extractors missed the "Threaten" event between the vice president Hugo Chavez and those questioning the legitimacy of Chavez's government.

Tweet: *#Venezuela VP warns those questioning the legitimacy of #Chavez's government: "Watch your words and your actions."*

News: *The vice president threatened action against any who question the legality of delaying the swearing-in of President Hugo Chavez, who is still in Cuba.*

6 Conclusions and Future Work

In this paper we study the bottlenecks of event extraction for tweets. We have two observations: (1) Because of the "lack of context" and "informal nature" characteristics of tweets, conducting event extraction solely on tweet messages cannot produce satisfactory results; (2) The events embedded in tweets and news articles are often complementary. Based on these observations, we proposed to link each tweet to its most relevant news article, and further incorporated this cross-genre knowledge to conduct cross-genre event extraction. Experiment results showed that without any additional annotation, our proposed cross-genre event extractor is able to outperform state-of-the-art tweet event extraction. Our future research will focus on joint modeling

of cross-genre event extraction in the training stage through cross-genre knowledge enrichment.

Acknowledgment

This work was supported by the U.S. DARPA DEFT Program No.FA8750-13-2-0041, ARL NS-CTA No. W911NF-09-2-0053, NSF Award IIS-1523198, AFRL DREAM project, and a gift award from Bosch. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

2005. ACE (automatic content extraction) english annotation guidelines for events, Jan. [Online]. Available: <https://www ldc.upenn.edu/collaborations/past-projects/ace>, (Date Last Accessed, Nov 4th, 2015).
- Edward Benson, Aria Haghighi, and Regina Barzilay. 2011. Event discovery in social media feeds. In *ACL*, pages 389–398.
- Weiwei Guo, Hao Li, Heng Ji, and Mona T. Diab. 2013. Linking tweets to news: A framework to enrich short text data in social media. In *Proc. Annu. Meeting of the Assoc. for Comput. Linguist.*, pages 239–249, Sofia, Bulgaria.
- Qi He, Kuiyu Chang, and Ee-Peng Lim. 2007. Analyzing feature trajectories for event detection. In *Proc. Int. ACM SIGIR Conf.*, pages 207–214, Amsterdam, Netherland.
- Heng Ji and Ralph Grishman. 2008. Refining event extraction through cross-document inference. In *Proc. Annu. Meeting of the Assoc. for Comput. Linguist.*, pages 254–262, Columbus, OH.
- Jon M. Kleinberg. 2003. Bursty and hierarchical structure in streams. *Data Min. Knowl. Discov.*, 7(4):373–397.
- Qi Li, Heng Ji, and Liang Huang. 2013. Joint event extraction via structured prediction with global features. In *Proc. Annu. Meeting of the Assoc. for Comput. Linguist.*, pages 73–82, Sofia, Bulgaria.
- Alan Ritter, Mausam, Oren Etzioni, and Sam Clark. 2012. Open domain event extraction from twitter. In *Proc. ACM SIGKDD Conf. on Knowl. Discovery and Data Mining*, pages 1104–1112, Beijing, China.
- Harald Steck. 2010. Training and testing of recommender systems on data missing not at random. In *Proc. ACM SIGKDD Conf. on Knowl. Discovery and Data Mining*, pages 713–722, Washington, DC.
- Jianshu Weng and Bu-Sung Lee. 2011. Event detection in twitter. In *Proc. Int. AAAI Conf. on Web and Social Media*, pages 401–408, Barcelona, Catalonia, Spain.
- Yiming Yang, Thomas Pierce, and Jaime G. Carbonell. 1998. A study of retrospective and on-line event detection. In *SIGIR*, pages 28–36.