Extracting Subevents via an Effective Two-phase Approach

Allison Badgett and **Ruihong Huang**
Texas A&M University
{allisonbadgett, huangrh}@cse.tamu.edu

Abstract

We present our pilot research on automatically extracting subevents from a domain-specific corpus, focusing on the type of subevents that describe physical actions composing an event. We decompose the challenging problem and propose a two-phase approach that effectively captures sentential and local cues that describe subevents. We extracted a rich set of over 600 novel subevent phrases. Evaluation shows the automatically learned subevents help to discover 10% additional main events (of which the learned subevents are a part) and improve event detection performance.

1 Introduction

General and abstract event expressions and event keywords are often used for detecting events. To detect civil unrest events for example, common event expressions "took to the streets" and "staged a protest", and event keywords "rally" and "strike" are usually considered as the first option. Subevents, events that occur as a part of the main event and therefore are useful to instantiate the main event, widely exist in event descriptions but they are rarely used for detecting the main event.

In this paper, we focus on learning subevent phrases that describe physical actions composing an event. Such subevents are the evidence that an event is occurring. For example, if a person were to explain how he knew that the crowd gathered in the street was rioting, he might point to the shouting of political slogans or tires lit on fire. In this instance, the riot would be the event. The slogan-shouting and tire-burning would be the subevents. Because subevent detection requires an understanding of levels of abstraction, this task can even be difficult for humans to perform. Furthermore, subevent phrases and general event phrases often have the same grammatical structure and may share common words, which makes automatically differentiating between event phrases and subevent phrases within a document even more difficult.

Additionally, events and subevents are not disjoint classes. There are some subevents that are unambiguous. For example, "burning tires" is a concrete phrase that would not fall into the category of more abstract events. However, "gathered at site" is certainly more ambiguous. Even human analysts would not necessarily agree on the appropriate class for this phrase. In many cases, the categorization would be context-dependent. Because of this, our research focused on identifying the less ambiguous, and therefore more concrete, cases.

Instead of separating subevent phrases from general event phrases, we explicitly acquire subevent phrases by leveraging both sentential and local cues in describing subevents. We observed that subevents of the type in which we are interested, as important facts presented in news stories, are commonly mentioned in sentences that refer to the source of information or simply in quotation sentences. These sentences usually start or end with characteristic phrases such as "media reports" and "witness said". Furthermore, we observed that subevent phrases often occur in conjunction constructions as a sequence of subevent phrases, as shown in the following examples:

Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 906–911,
Austin, Texas, November 1-5, 2016. ©2016 Association for Computational Linguistics

(1) *State* <u>*television broadcast*</u> *the event live, offering sweeping aerial views that showed the sea of people* **waving banners, blew whistles, and shouted slogans**.

(2) *They also* **set fires, stoned civilian vehicles, taunted the police and hurled stones** *at them,* <u>*witnesses said*</u>.

where the subevents are shown in bold, and sentential cues are underlined.

Inspired by these observations, we propose a novel two-phase approach to automatically extract subevents, which consists of a sentence classifier that incrementally identifies sentences mentioning subevents and a subevent extractor which looks for a sequence of subevent phrases in a conjunction structure. Our sentence classifier is trained in a weakly supervised manner and only requires a small set of subevent phrases as a guide. The classifier was initially trained with sentences containing eight provided subevent seeds, then it proceeded to label more sentences that mention new subevent phrases.

This two-phase subevent extraction approach can successfully identify 610 diverse subevent phrases from a domain-specific corpus. We evaluate our automatically learned subevent phrases by using them to detect events. Experimental results show that the learned subevent phrases can recover an additional 10% of event articles and improve event detection F-1 score by 3%.

2 Related Work

While it is generally agreed that subevents are an important type of information in event descriptions, they are seldom considered in decades of event extraction research (Appelt et al., 1993; Riloff, 1993; Soderland et al., 1995; Sudo et al., 2003; Li et al., 2005; Yu et al., 2005; Gu and Cercone, 2006; Maslennikov and Chua, 2007; S. and E., 2009; Liao and Grishman, 2010; Huang and Riloff, 2011; Chambers and Jurafsky, 2011; Huang and Riloff, 2012; Huang et al., 2016). Subevents as a theme has been discussed in the past three Event workshops (Eve, 2013), (Eve, 2014), (Eve, 2015). However, despite the great potential of using subevents to improve event detection and extraction (Hakeem and Shah, 2005), and event coreference resolution (Araki et al., 2014), there is little existing research on automatically learning subevent phrases, partially because researchers have not agreed upon the definition of subevents. Much recent research in event timeline generation (Huang and Huang, 2013) suggests the usefulness of subevents in improving quality and completeness of automatically generated event summaries. However, they often focus on a different notion of subevents that broadly covers pre-condition events and consequence events and is temporally-based.

Subevents have been studied for event tracking applications (Shen et al., 2013; Meladianos et al., 2015). However, most current research is specifically related to social media applications, like Twitter, in terms of both its definition of subevents and methodologies. For example, in previous research by (Shen et al., 2013), a subevent is defined as a topic that is discussed intensively in the Twitter stream for a short period of time before fading away. Accordingly, the subevent detection method relies on modeling the "burstiness" and "cohesiveness" properties of tweets in the stream. We instead aim to provide a more general definition of subevents as well as present a method for identifying subevent at the article level.

3 A Two-phase Approach for Subevent Extraction

As illustrated in Figure 1, We use a two-phase algorithm to identify subevent phrases from our domain-specific corpus. For the first stage, we implemented a bootstrapped artificial neural network in order to identify sentences that are likely to contain a subevent phrase. In the second stage, we identify phrases fitting a predetermined conjunction pattern within the sentences classified by the first-stage neural network.

3.1 Phase 1: Identifying Subevent Sentences

3.1.1 Domain-specific Corpus

Thanks to previous research on multi-faceted event recognition by (Huang and Riloff, 2013), we compiled our own domain-specific corpus that describes civil unrest events. Using civil unrest events as an example, (Huang and Riloff, 2013) demon-

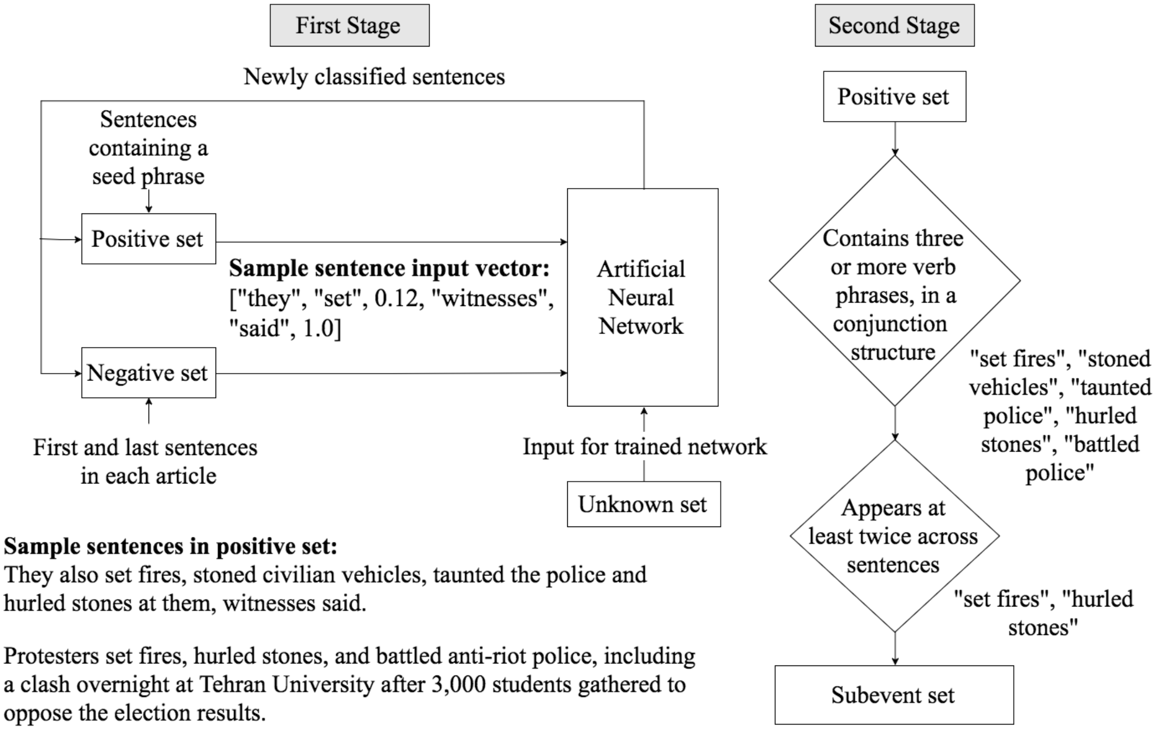
First Stage
Second Stage
Newly classified sentences
Sentences containing a seed phrase
Positive set
Negative set
First and last sentences in each article
Sample sentence input vector:
["they", "set", 0.12, "witnesses", "said", 1.0]
Artificial Neural Network
Input for trained network
Unknown set
Sample sentences in positive set:
They also set fires, stoned civilian vehicles, taunted the police and hurled stones at them, witnesses said.
Protesters set fires, hurled stones, and battled anti-riot police, including a clash overnight at Tehran University after 3,000 students gathered to oppose the election results.
Positive set
Contains three or more verb phrases, in a conjunction structure
"set fires", "stoned vehicles", "taunted police", "hurled stones", "battled police"
Appears at least twice across sentences
"set fires", "hurled stones"
Subevent set

Figure 1: The Two-step Subevent Learning Paradigm

strated that we can use automatically-learned event facet phrases (event agents and purposes) and main event expressions to find event articles with a high accuracy. We first obtained their learned event facet phrases and event expressions, most of which refer to general events. Then we followed their paper and identified two types of news articles that are likely to describe a civil unrest event by matching the obtained phrases to the English Gigaword fifth edition (Parker et al., 2011)

Specifically, we first consider news articles that contain a civil unrest keyword such as "strike" and "protest"[1], then we identify an article as relevant if it contains a sentence where either two types of facet phrases or one facet phrase together with an event expression are found. In addition, we consider news articles that do not contain a civil unrest keyword; we require an article to contain a sentence where three types of event information are matched. Overall, we get a civil unrest corpus containing 232,710 news articles.

3.1.2 Context Feature Selection

We hypothesized that the first and last noun/verb phrases and their positions in the sentence were likely to be good indicators that the sentence might contain a subevent phrase. Because our document corpus was composed of news articles, we determined that concrete subevents would require a level of substantiation that abstract, non-specific events would not. For example, a reporter would not usually cite a source in a sentence informing the reader that a riot occurred but would likely choose to quote a source when reporting that rioters burned tires in the streets. Because of this, sentences containing subevent phrases often begin or end with phrases such as "he witnessed" or "she told the press." To represent the nouns and verbs, we used the 50-dimention Stanford GloVe (Pennington et al., 2014) word embeddings pre-trained on Wikipedia 2014 and Gigaword5 (Parker et al., 2011).

3.1.3 Seeds and Training Sentence Generation

To form a training set for the neural network, we used eight seed subevent phrases (as shown in Table 1) to identify a set of positive sentences that con-

[1] We used the same set of keywords as used by (Huang and Riloff, 2013)

waved banners
shouted slogans
chanted slogans
burned flag
burned flags
blocked road
clashed with police
clashed with government

Table 1: First Stage Classifier Seed Subevent Phrases

tain one of the seed phrases. In total, we obtained around 5000 positive sentences and bounded this to 3500 for use with the classifier. Finding a sufficient number of negative sentences was a more challenging task. After reviewing the corpus, we determined that the first and last sentences of an article are unlikely to contain subevent phrases. These sentences often function as general introductions and conclusions that refer to the main event of the article. We selected 7000 of these sentences to form the negative set. The rest of the sentences not classified as positive or negative remain unknown, amounting to almost 1 million.

3.1.4 Artificial Neural Networks for Sentence Classification

We implemented an artificial neural network with a single hidden layer composed of 500 nodes. In order to facilitate faster training, we used tanh as the activation function of the hidden layer. Softmax was used as the output layer activation function. In order to train the network, we provided an initial set of positive and negative vectors representing sentence data from the corpus as described in Section 3.1.3. These input vectors were then divided into a training set, validation set and testing set. The training set was comprised of 70% of the full dataset, the validation set of 20% and the testing set of 10%. The neural network was trained for 1000 epochs and used the validation set and test set to measure performance in order to avoid overfitting.

Because we began with a limited number of seed phrases to create the positive set, we chose to use a bootstrapping approach to expand our data set and improve the classifier. After training, the entire unknown data set would be classified, and sentences determined to be positive with 0.90 certainty or greater would be added to the positive set. Sentences classified as negative with 0.90 certainty or greater would be added to the negative set. However, in order to maintain the 2:1 ratio of negative to positive vectors, the number of negative vectors that could be added was capped at twice the number of positive additions for each iteration. After the new sentences were added to the positive and negative sets, the neural network was retrained with this data and classified additional previously unknown sentences. The process repeated for five iterations, then bootstrapping ended because not enough newly identified positive sentences were found (<3000 in the last iteration). Table 2 shows the number of sentences that were added after each bootstrapping iteration.

Iteration	Positives	Negatives
1	13223	26446
2	12611	25222
3	9411	18822
4	6076	12152
5	2842	5684

Table 2: Number of Sentences Added after Each Iteration

3.2 Phase 2: Subevent Extraction

After accumulating a large set of sentences likely containing subevents from the first phase of the system, the second step identifies the subevent phrases within these sentences. We observe that subevent phrases often occur in lists and we focus on leveraging such local cues to extract subevents. Specifically, we identify conjunction constructions that contain three or more verb phrases, each verb phrase obeys one of the following two forms: verb + direct object or verb + prepositional phrase. We extract the sequence of verb phrases, each as a subevent candidate. We only included subevents with frequency greater than or equal to two in the final evaluation. Through the two-stage extraction procedure, we identified 610 unique subevents. Table 3.2 shows some of the learned subevents.

Clearly, this second phase suffers from low recall. However, because subevents are identified at the corpus level as opposed to the document level, per-sentence recall is not a significant concern as long as a sufficient number of subevents are identi-

threw stones, hurled rocks, pounded in air
smashed through roadblocks, detained people
forced way, fired in air, threw at police
smashed windows, set fire, burned tires
threw bombs, opened fire, blocked road
pelted with stones, appeared on balcony
arrested for vandalism, threw bombs
burned cars, carried banners, lit candles
detained people, planted flag, wore masks
stoned police, converged on highway
chanted against authorities, chanted in city
broke through barricade, blocked traffic
broke windows, screamed outside palace
torched cars, ransacked office, smashed shop
shouted in unison, sang songs, planted flags
runs alongside shrines, chanted for democracy

Table 3: Subset of learned subevents

	Recall	**Precision**	**F1-score**
(Huang and Riloff, 2013)	71	**88**	79
+Subevents	**81**	83	**82**

Table 4: Event Recognition Performance Before/After Incorporating Subevents

fied across the whole corpus. As we demonstrate in the evaluation section, corpus-level recall was high enough to produce noticeable results.

4 Evaluation

We show that our acquired subevent phrases are useful to discover articles that describe the main event and therefore improve event detection performance.

For direct comparisons, we tested our subevents using the same test data and the same evaluation setting as the previous multi-faceted event recognition research by (Huang and Riloff, 2013). Specifically, they have annotated 300 new articles that each contains a civil unrest keyword and only 101 of them are actually civil unrest stories. They have shown that the multi-faceted event recognition approach can accurately identify civil unrest documents, by identifying a sentence in the documents where two types of facet phrases or one facet phrase and a main event expression were matched. The first row of Table 4 shows their multi-faceted event recognition performance.

We compared our learned subevent phrases with the event phrases learned by (Huang and Riloff, 2013) and found that 559 out of our 610 unique phrases are not in their list. We augmented their provided event phrase list with our newly acquired subevent phrases and then used the exactly same evaluation procedure. Essentially, we used a longer event phrase dictionary which is a combination of main event expressions resulted from the previous research by (Huang and Riloff, 2013) and our learned subevent phrases. Row 2 shows the event recognition performance using the extended event phrase list. We can see that after incorporating subevent phrases, additional 10% of civil unrest stories were discovered, with a small precision loss, the F1-score on event detection was improved by 3%.

5 Conclusion

We have presented a two-phase approach for identifying a specific type of "subevents", referring to physical actions composing an event. While our approach is certainly tailored to the civil unrest domain, we believe that this method is applicable to many other domains within the scope of news reports, including health, economics and even politics, where reporters overwhelmingly rely on outside opinion to present the facts of the story and provide the summary themselves. However in more casual domains where this is not necessarily the case, this approach will suffer. For instance, in sports writing, a reporter giving a play-by-play of a basketball game will not need to call upon witnesses or field experts to present concrete subevents.

Furthermore, we have shown the great potential of using subevents to improve event detection performance. In addition, distinguishing between events and subevents develops an event hierarchy and can benefit multiple applications such as text summarization and event timeline generation.

Acknowledgments

We want to thank our anonymous reviewers for providing useful comments.

References

D. Appelt, J. Hobbs, J. Bear, D. Israel, and M. Tyson. 1993. FASTUS: a Finite-state Processor for Information Extraction from Real-world Text. In *Proceedings* *of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI)*.

Jun Araki, Zhengzhong Liu, Eduard H Hovy, and Teruko Mitamura. 2014. Detecting subevent structure for event coreference resolution. In *LREC*, pages 4553–4558.

Nathanael Chambers and Dan Jurafsky. 2011. Template-Based Information Extraction without the Templates. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-11)*.

2013. The 1st Workshop on EVENTS: Definition, Detection, Coreference, and Representation. In *https://sites.google.com/site/cfpwsevents/home*.

2014. The 2nd Workshop on EVENTS: Definition, Detection, Coreference, and Representation. In *https://sites.google.com/site/wsevents2014/home*.

2015. The 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation. In *https://sites.google.com/site/wsevents2015/home*.

Z. Gu and N. Cercone. 2006. Segment-Based Hidden Markov Models for Information Extraction. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 481–488, Sydney, Australia, July.

Asaad Hakeem and Mubarak Shah. 2005. Multiple agent event detection and representation in videos. In *AAAI*, pages 89–94.

Lifu Huang and Lian'en Huang. 2013. Optimized event storyline generation based on mixture-event-aspect model. In *EMNLP*, pages 726–735.

Ruihong Huang and Ellen Riloff. 2011. Peeling Back the Layers: Detecting Event Role Fillers in Secondary Contexts. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-11)*.

Ruihong Huang and Ellen Riloff. 2012. Modeling Textual Cohesion for Event Extraction. In *Proceedings of the 26th Conference on Artificial Intelligence (AAAI-12)*.

Ruihong Huang and Ellen Riloff. 2013. Multi-faceted Event Recognition with Bootstrapped Dictionaries. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-13)*.

L. Huang, T. Cassidy, X. Feng, H. Ji, C. Voss, J. Han, and A. Sil. 2016. Liberal event extraction and event schema induction. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-16)*.

Y. Li, K. Bontcheva, and H. Cunningham. 2005. Using Uneven Margins SVM and Perceptron for Information Extraction. In *Proceedings of Ninth Conference on Computational Natural Language Learning*, pages 72–79, Ann Arbor, MI, June.

Shasha Liao and Ralph Grishman. 2010. Using Document Level Cross-Event Inference to Improve Event Extraction. In *Proceedings of the 48st Annual Meeting on Association for Computational Linguistics (ACL-10)*.

M. Maslennikov and T. Chua. 2007. A Multi-Resolution Framework for Information Extraction from Free Text. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*.

P. Meladianos, G. Nikolentzos, F. Rousseau, Y. Stavrakas, and M. Vazirgiannis. 2015. Degeneracy-based real-time sub-event detection in twitter stream. In *Proceedings of the 9th AAAI international conference on web and social media (ICWSM)*, pages 248–257.

Robert Parker, David Graff, Junbo Kong, Ke Chen, and Kazuaki Maeda. 2011. English Gigaword. In *Linguistic Data Consortium*.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

E. Riloff. 1993. Automatically Constructing a Dictionary for Information Extraction Tasks. In *Proceedings of the 11th National Conference on Artificial Intelligence*.

Patwardhan S. and Riloff E. 2009. A Unified Model of Phrasal and Sentential Evidence for Information Extraction. In *Proceedings of 2009 the Conference on Empirical Methods in Natural Language Processing (EMNLP-2009)*.

Chao Shen, Fei Liu, Fuliang Weng, and Tao Li. 2013. A participant-based approach for event summarization using twitter streams. In *HLT-NAACL*, pages 1152–1162.

S. Soderland, D. Fisher, J. Aseltine, and W. Lehnert. 1995. CRYSTAL: Inducing a Conceptual Dictionary. In *Proc. of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1314–1319.

K. Sudo, S. Sekine, and R. Grishman. 2003. An Improved Extraction Pattern Representation Model for Automatic IE Pattern Acquisition. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL-03)*.

K. Yu, G. Guan, and M. Zhou. 2005. Resumé Information Extraction with Cascaded Hybrid Model. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 499–506, Ann Arbor, MI, June.