

Employing Event Inference to Improve Semi-Supervised Chinese Event Extraction

Peifeng Li, Qiaoming Zhu, Guodong Zhou

School of Computer Science & Technology

Soochow University, Suzhou, 215006, China

{pfli, qmzhu, gdzhou}@suda.edu.cn

Abstract

Although semi-supervised model can extract the event mentions matching frequent event patterns, it suffers much from those event mentions, which match infrequent patterns or have no matching pattern. To solve this issue, this paper introduces various kinds of linguistic knowledge-driven event inference mechanisms to semi-supervised Chinese event extraction. These event inference mechanisms can capture linguistic knowledge from four aspects, i.e. semantics of argument role, compositional semantics of trigger, consistency on coreference events and relevant events, to further recover missing event mentions from unlabeled texts. Evaluation on the ACE 2005 Chinese corpus shows that our event inference mechanisms significantly outperform the refined state-of-the-art semi-supervised Chinese event extraction system in F1-score by 8.5%.

1 Introduction

An event is a specific occurrence involving arguments (participants and attributes) of the specific roles. In an event, trigger is the main word which most clearly expresses its occurrence, so recognizing an event can be recast as identifying a corresponding trigger. An event may have several arguments, which are entity mentions (e.g., person name, time, location, etc.) and must fulfill the corresponding roles. Take the following sentence as an example:

S1: On the 25th Dec. (A1: *Artifact*), peacekeepers (A2: *Artifact*) **returned** (E1: *Transport*) to Amman (A3: *Place*) by flight (A4: *Vehicle*).

For this example, an event extraction system should identify one event mention E1, which is triggered by verb “returned” whose event type is *Transport*, with four arguments, “peacekeepers”, “25th Dec.”, “flight”, and “Amman”, fulfilling the roles of *Artifact*, *Time*, *Vehicle*, and *Place*, respectively.

Automatically extracting events from free texts is a higher-level Information Extraction (IE) task, which is still a challenge due to the complexity of natural language and the domain-specific nature, especially in Chinese for its specific characteristics. In particular, most of previous studies have focused on English event extraction, while only a few concern Chinese.

Currently, supervised learning models have dominated event extraction. To reduce the labeled data required, a few semi-supervised models have been applied to English event extraction (e.g., Riloff 1996; Yangarber et al., 2000; Stevenson and Greenwood, 2005; Huang and Riloff, 2012). Since classifier-based model needs dozens of annotated documents to train model, most of previous semi-supervised models focused on pattern-based approach, which only needed a few seed (event) patterns. In those pattern-based approaches, frequent event patterns, which occur in many documents, were chosen as relevant patterns to match event mentions in unlabeled texts. However, the order of words in a Chinese sentence is rather agile for its open and flexible structure, and different orders might express the same meaning due to the semantics-driven nature of the Chinese language. This results in the diversity of Chinese event patterns and numerous infrequent patterns, even some event mentions having no matching patterns. Hence, it is an issue to extract the event mentions with infrequent patterns.

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Page numbers and proceedings footer are added by the organisers. Licence details: <http://creativecommons.org/licenses/by/4.0/>.

In this paper, we first implement a pattern-based semi-supervised model for Chinese event extraction as a baseline, following the state-of-the-art system as described in (Liao and Grishman, 2010a) and then refine this model to suit Chinese event extraction. Moreover, we propose various kinds of novel linguistic knowledge-driven event inference mechanisms to address the above issue and recover missing event mentions. These event inference mechanisms can capture the linguistic knowledge from semantics of argument role, compositional semantics of trigger, consistency on coreference events and relevant events. Evaluation on the ACE 2005 Chinese corpus shows that our event inference mechanisms dramatically outperform the baseline.

The rest of this paper is organized as follows. Section 2 overviews related work. Section 3 presents the refined semi-supervised model for Chinese event extraction. Section 4 proposes several linguistic knowledge-driven event inference mechanisms. Section 5 reports and analyzes the experimental results. Finally, we conclude our work in Section 6.

2 Related Work

Almost all previous semi-supervised models focus on English event extraction, which can be subdivided into pattern-based models (e.g., Riloff, 1996; Yangarber et al., 2000; Liao and Grishman, 2010a; Chambers and Jurafsky, 2011; Balasubramanian et al., 2013) and classifier-based models (e.g., Chieu et al., 2003; Maslennikov and Chua, 2007; Patwardhan and Riloff, 2009; Liu and Strzalkowski, 2012; Wang et al., 2013). Classifier-based models normally require a small set of annotated data (e.g., 100 annotated documents), while pattern-based models need dozens of high quality seed patterns.

Riloff (1996) first divided unlabeled documents into irrelevant and relevant documents, and the latter was much likely to contain further relevant patterns. Then event patterns from relevant documents were generated by using an annotated data and a set of heuristic rules. Yangarber et al. (2000) proposed a document-centric view to boost a semi-supervised event extraction system, which assumes relevant documents always contain some shared patterns. Yangarber (2003) further introduced multiple learners into the bootstrapping procedure to make the final decision on the combination of multiple learners on distinct event types. Huang and Riloff (2012) employed role-identifying nouns, which proposed by Phillips and Riloff (2007), as seed terms to extract patterns from relevant documents and then generated the labeled instances to train three classifiers in their event extraction system.

As an alternative, Stevenson and Greenwood (2005) proposed a pattern similarity-centric view and selected relevant patterns on similarity scores. Normally, bootstrapping on the document-centric view tends to accept the irrelevant patterns with a high occurrence frequency in relevant documents. To address this problem, Liao and Grishman (2010a) introduced a pattern similarity metric into the document-centric view as a filter to eliminate those irrelevant patterns. Liao and Grishman (2011) further applied an information retrieval mechanism to detect relevant documents and proposed a self-training strategy for bootstrapping.

In addition, several studies focused on the event pattern representation, such as pairwise (e.g., Subject-Verb, Verb-Object) (Chambers and Jurafsky, 2008, 2009), SVO (Subject-Verb-Object) (Yangarber, 2000; Balasubramanian et al., 2013), chain (Sudo et al., 2001), subtree (Sudo et al., 2003) and complex pattern (Liu and Strzalkowski, 2012).

In the literature, only one paper concerns semi-supervised Chinese event extraction. Chen and Ji (2009a) applied various kinds of cross-lingual features in the bootstrapping procedure to extract Chinese event. With the help of over 500 annotated seed event mentions in 100 documents, they only achieved 35% in F1-score. This indicates the critical challenge in semi-supervised Chinese event extraction.

Only a few studies concern event inference mechanisms. Ji and Grishman (2008) employed a rule-based approach to propagate consistent triggers and arguments across topic-related documents. Liao and Grishman (2010b) employed cross-event consistent information to improve sentence-level event extraction. Hong et al. (2011) regarded entity type consistency as a key feature to predict event mentions and adopted an information retrieval mechanism to promote event extraction. Li et al. (2013) proposed a global argument inference model on Chinese argument extraction to explore specific relationships among relevant event mentions to recover those inter-sentence arguments in the sentence, discourse and document layers. Li et al. (2014) also introduced Markov Logic Network (MLN) to capture the discourse-level consistency between Chinese trigger mentions to further recover those poor-

context event mentions. In a word, all of above mechanisms focus on supervised event extraction and no literature involves in the event inference of semi-supervised event extraction.

3 Semi-supervised Model for Chinese Event Extraction

In this section, we refine a semi-supervised model for Chinese event extraction as a baseline, which includes two views, the document-centric view and pattern similarity-centric view.

3.1 Semi-supervised Model

Liao and Grishman (2010a) proposed a state-of-the-art semi-supervised event extraction system, which was a pattern-based approach and adopted bootstrapping mechanism to extract relevant patterns. Besides, two distinct views, the document-centric view and the pattern similarity-centric view as described in Subsection 3.2 and 3.3, are incorporated in the bootstrapping procedure to rank event patterns on different metrics. In each iteration, the candidate patterns, which extracted from unlabeled texts as the candidates of relevant patterns, are ranked following the document-centric view, then the candidate patterns with pattern similarity scores below a similarity threshold (0.9 in (Liao and Grishman, 2010a)) will be removed; only top 3 candidate patterns in the ranking scores of the document-centric view will be accepted as relevant patterns. In addition, if no pattern is found in the current iteration, the threshold will be reduced by 0.1 until new relevant patterns are extracted.

As we mentioned earlier, the open and flexible structure of Chinese sentences results in the diversity of Chinese event patterns. Moreover, the syntax or semantic path is often used to represent event patterns, but the performance in Chinese syntactic parsers and Semantic Role Labeling (SRL) tools is lower than that in English. Therefore, we refine this semi-supervised model to suit Chinese event extraction in three aspects as follows, due to the above characteristics of Chinese language.

Firstly, we construct a refined event pattern representation of Chinese events. Liao and Grishman (2010a) used semantic roles to represent the relationship between the trigger and its arguments. Due to the wide spread of ellipsis (especially entities) and the relatively low performance of Chinese SRL, pairwise (trigger-entity) representation and dependency path are introduced to represent Chinese event pattern in our refined model. Hence, the event pattern in this paper is a triple-style template as follows.

<trigger, entity type, their dependency path >

A pattern is formed by a trigger, the entity type of its argument¹ and the dependency path from the trigger to the argument. For example, trigger “returned” and its argument “peacekeepers” (entity type: PER) in sentence S1 can be described as a pattern *<returned, PER, nsubj>*.

Secondly, we introduce a novel mechanism to extract candidate patterns. Since verb and noun dominate in triggering an event in Chinese and they are chosen as candidate triggers to create candidate patterns. Besides, since different event types may have different roles and different roles are fulfilled by entities with different types, the entities whose types can fulfil the core roles of a specific event are chosen as candidate entities. For example, *Attacker* and *Target* are the core roles of event *Attack* and entity types *PER/ORG/GPE*² can fulfil above two roles, so we only accept those entities, whose types belong to *PER/ORG/GPE*, to form candidate patterns. For each sentence in the unlabeled data, all candidate trigger-entity pairs and their dependency path are enumerated as candidate patterns.

Finally, we present a new mechanism to generate seed patterns based on seed triggers. Considering the relatively large number of Chinese triggers and the flexibility of Chinese sentences, an instance-based approach is adopted by enumerating a few high-quality seed triggers with explicit meaning and high probability to trigger a specific event. Instead of dozens of predefined patterns required in previous studies, only one seed trigger is given to each event type or subtype without any predefined patterns. Hence, all patterns consisting of a seed trigger in the candidate patterns are accepted as seed patterns for their high probability to trigger a specific event.

3.2 Document-centric View

The document-centric view regards those documents containing the patterns always identified as relevant to a specific event as relevant documents and concludes that they are likely to contain additional

¹ All event arguments must be entity mentions following the ACE 2005 annotation guidelines of events.

² PER/ORG/GPE refers to person, organization and geo-political entity respectively, which are annotated in the ACE 2005 corpus. These helpful information can be seen as ontological classes.

relevant patterns. Hence, those candidate patterns occurring in the relevant documents frequently will be extracted as relevant ones. Following Yangarber et al. (2000) and Liao and Grishman (2010a), we also employ the disjunctive voting scheme to calculate the ranking scores $R_{score}(p)$ of pattern p as follows.

$$R_{score}(p) = \frac{\sum_{d \in L(p)} Rel(d)}{|L(p)|} * \log \sum_{d \in L(p)} Rel(d) \quad (1)$$

where $L(p)$ is the set of documents, which contain candidate pattern p , and $Rel(d)$ is the relevance score of document d as follows.

$$Rel(d) = 1 - \prod_{p \in P} \left(1 - \frac{\sum_{d \in L(p)} Rel'(d)}{|L(p)|} \right) \quad (2)$$

where $Rel'(d)$ is the relevance score of document d in the previous iteration. Initially the relevance score of document d is set to n if document d has n relevant patterns in the set of extracted patterns P .

3.3 Pattern Similarity-centric View

The similarity-centric view tries to find the candidate patterns who are similar to those seed patterns. The similarity scores derive from two aspects, lexical similarity and syntactic similarity, while the former is based on the trigger and entity type in a pattern and the latter is based on the relation between the trigger and the entity. Especially, we realize the pattern similarity view following the lexical and syntactic similarity, and refine the similarity ranking score $I_{score}(p)$ of candidate pattern p as follows:

$$I_{score}(p) = \text{Max}_{s \in P} (WSim(t_p, t_s) \times ESim(e_p, e_s) \times DSIM(d_p, d_s)) \quad (3)$$

where t , e and d represent the trigger, entity type and dependency path in candidate pattern $p(t_p, e_p, d_p)$ or seed pattern $s(t_s, e_s, d_s)$ in the set of extracted patterns P , respectively; $ESim$ identifies whether two entities have the same type, and assigned 1 if two entities have the same entity type and otherwise a small number 0.1; $DSim$ calculates the similarity between two dependency paths in edit distance. Finally, $WSim$ is to obtain the trigger similarity in lexical semantics, using HowNet (Dong and Dong, 2006) following Liu and Li (2002):

$$WSim(t_p, t_s) = \frac{\phi}{Dis(t_p, t_s) + \phi} \quad (4)$$

where $Dis(t_p, t_s)$ is the distance between the sememes of triggers t_p and t_s , in HowNet's sememe hierarchical architecture, with parameter ϕ assigned 0.75 following Liu and Li (2002).

4 Event Inference

The pattern-based semi-supervised model cannot extract those event mentions matching infrequent patterns or without matching patterns. The knowledge from linguistic aspect (e.g., definition of events, compositional semantics of Chinese words, coreference events and relevant events, etc.) is helpful to further recover missing event mentions or filter pseudo event mentions. In this section, various kinds of event inference mechanisms based on linguistic knowledge are proposed to improve the performance of semi-supervised Chinese event extraction.

We unify the semi-supervised model and the event inference mechanisms into one model as follows: In each iteration, after the top 3 patterns have been chosen following the document-centric view and event mentions in the unlabeled data have been extracted by pattern matching, all event inference

mechanisms are applied to recover missing event mentions,. Due to our inference mechanisms are trigger-based and each inferred event mention may have more than one pattern while most of them are noisy, we do not add those patterns in the set of relevant patterns for bootstrapping.

4.1 Event Inference on Role Semantics

The core of an event can be expressed as “*Who do What to Whom*” in which “Who” and “Whom” are the core roles³ to participate in an event, while “What” often refers to event trigger. The relationship between the verbal trigger and its core roles are the key clues to express event semantics. Since the subject or object always play the core roles in an event mention, SVO (Subject-Verb-Object) is a better representation of event pattern. However, ellipsis is a widespread phenomenon in Chinese language and many sentences do not have an overt subject or object, so lots of event mentions cannot be represented as SVO pattern. In this paper, we only use the trigger-entity pair to represent event pattern and one of the disadvantages of this representation is its loose constraint on events, which will extract lots of pseudo event mentions.

In most cases in Chinese, the object is often the most important core role to identify a specific event and it is more helpful than the subject to distinguish true event mentions from pseudo ones. Take following two sentences as examples:

S2: 老师(PER) 打(hit)了 这个学生(PER)。 (The teacher hit this student.)

S3: 老师(PER) 打(call)了 电话 给 这个学生(PER)。 (The teacher made a phone call to this student.)

The relation between verb 打 (hit) and object 这个学生 (this student) is clear to indicate sentence S2 is an *Attack* event mention since the object is a person, while object 电话 (phone) in sentence S3 is not a person and it indicates this sentence is not an *Attack* event mention following the sense of verb 打 (call). Therefore, the object is an effective evidence to indicate event mentions and it is incorporated in our model to remove pseudo event mentions as follows.

Role Semantics: If the object of a candidate verbal trigger mention is not an entity or its entity type cannot fulfil the object roles (e.g., *Victim* in events *Injure* and *Die*) in a specific event, this candidate trigger mention⁴ will be inferred as pseudo one.

For example, core role *Target* of event *Attack* often acts as the object of a verbal trigger and entity types PER, ORG and GPE can fulfill this role according to be definition of event *Attack* in the ACE 2005 corpus. Hence, a candidate trigger mention of event *Attack* will be regarded as pseudo one when this mention has an object which is not an entity or whose entity type is not PER, ORG or GPE.

4.2 Event Inference on Compositional Semantics

In Chinese language, a word is composed of one or more characters. Almost all Chinese characters have their own meanings and are morpheme (or single-morpheme word), the minimal meaningful unit. If a Chinese word contains more than one character, its meaning can often be derived from its composite morphemes. This more fine-grained semantics is compositional semantics of Chinese words. Actually, it is also a normal way for a native Chinese speaker to understand a new Chinese word.

Two-morpheme words are used widely in Chinese language and almost all Chinese triggers contain one or two morphemes. The compositional semantics of a two-morpheme word comes from both its morphemes and morphological structure. Besides morphological structure *Coordination*, all other morphological structures (e.g., *Modifier-Head*, *Predicate-Object*, *Predicate-Complement* (Li and Zhou, 2012)) always have one head morpheme, the morpheme as the governing semantic element, to express the meaning of a word. Commonly, there are two head morphemes in a two-morpheme word of *Coordination* structure. In particular, a two-morpheme word triggers an event if its two head morphemes are homogeneous (e.g., 攻(attack)击(attack), 死(die)亡(die)). Otherwise, it may refer to more than one event and this means that two triggers are within a word whose morphological structure is *Coordination*. Take the following sentence as an example:

³ We select core roles following the ACE Chinese annotation guidelines of events. *Agent/Victim* are the core roles of events *Die/Injure* while *Attacker/Target* are the core roles of event *Attack*.

⁴ Recognizing a trigger mention can be recast as identifying a corresponding event mention, since trigger is the main word which most clearly expresses the occurrence of an event.

S4: 一名少年刺(E2: *Attack*)死(E3: *Die*)一名妇女。(A younger **stabbed** (E2: *Attack*) a woman to **death** (E3: *Die*)).

In S4, two-morpheme word 刺死 (stab a person to death) is a trigger with the *Coordination* structure. There are two event mentions in sentence S4, one *Attack* (E2) and one *Die* (E3), while morpheme 刺 (stab) triggers an *Attack* event and 死 (die) refers to a *Die* one.

Almost all event extraction systems assigned only one event type to a trigger and this will lead to that the other event type does not have any patterns to match and then cannot be identified. To address this issue, we first identify those triggers who refers to two distinct events as follows: for each two-morpheme candidate trigger in the candidate patterns whose morphemes are m_1 and m_2 , it will be identified as candidate trigger with two event types and split into two single-morpheme word to generate two candidate trigger mentions when the following three conditions are satisfied:

- 1) $verb \in POS(m_1) \wedge verb \in POS(m_2)$
- 2) $Max_{s_1 \in seeds} (Wsim(m_1, s_1)) = 1 \wedge Max_{s_2 \in seeds} (Wsim(m_2, s_2)) = 1 \wedge Etype(s_1) \neq Etype(s_2)$
- 3) $Morph(m_1 m_2) = Coordination$

where $POS(m)$ returns all possible parts of speech of morpheme m in Hownet and $Etype(s)$ is to obtain the event type of seed trigger s ; $Wsim(m, s)$ is defined in Subsection 3.3 and returns 1 when one word m is the synonym of the other word s ; $Morph(w)$ is to obtain the morphological structure of word w following Li and Zhou (2012).

Since there is a strong trigger consistency in those two-morpheme words of *Coordination* structure which refers to two distinct events, we propose an event inference mechanism as follows.

Compositional semantics: For each two-morpheme word identified by the above three conditions, if one of its morphemes has been extracted as an trigger mention of a specific event type, the other morpheme in the same word will refer to an a relevant event type.

4.3 Event Inference on Coreference Events

To mine more event mentions, we use the simple trigger-entity pair to represent event pattern in this paper. However, lots of event mentions still cannot be extracted due to the ellipsis of arguments. Take following sentences as examples:

S5: 美国与北韩在吉隆坡结束会谈(E4: *Meeting*)。(The US and DPRK finished **talking** (E4: *Meeting*) in Kuala Lumpur.)

S6: 会谈(E5: *Meeting*)的气氛严肃。(The **talks** (E5: *Meeting*) are serious.)

Obviously, more than one pattern of event mention E4 can be generated from sentence S5, since it contains more than one entity. On the contrary, no pattern can be extracted from S6 and this leads to event mention E5 cannot be extracted in our pattern-based semi-supervised model.

Within a document, almost all event mentions are around a topic and there is a strong trigger consistency: if one mention of a word triggers a specific event, its other mentions in the same document will refer to the same event type. Besides, similar words (e.g., 炸 (bomb), 爆炸 (bomb), 轰炸 (bomb)), which contains the same head morpheme, always express the same or similar meaning following the principle of compositional semantics. Similarly, there is a strong trigger consistency on those similar words: If one mention of a word refers to a specific event, the mentions of its similar words in the same document will trigger events of the same type.

Since the mentions of the same word or similar words are often coreference ones and always refer to the same event type, we propose an event inference mechanism on coreference events to recover missing event mentions based on head morpheme as follows. In particular, head morphemes are also identified following Li and Zhou (2012).

Coreference events: 1) if a mention of a candidate trigger refers to a specific event, all its other mentions in the same document will trigger the same type event; 2) if one mention of a candidate trigger refers to a specific event, all the mentions of its similar words in the same document will trigger the same type event too.

4.4 Event Inference on Relevant Events

The bootstrapping procedure of the document-centric view selects frequent patterns in relevant docu-

ments and ignores those infrequent patterns both in relevant or irrelevant documents. However, the number of infrequent patterns in Chinese is larger than that in English, due to its open and flexible sentence structure, as mentioned in Subsection 3.1.

Besides the pattern-based semi-supervised model, we propose a trigger-based mechanism as a supplement to recover those missing event mentions concerning infrequent patterns following this assumption: if a trigger mention refers a specific event in a document, there is a high probability that its relevant events occur in the same document. Take the following sentence as an example:

S7: 在冲突(E6: *Attack*)中, 有 1名阿拉伯人死亡(E7: *Die*)。 (An Arabian was **dying** (E7: *Die*) in this **conflict** (E6: *Attack*)).

In sentence S7, there is an extracted *Die* event mention E7 triggered by 死亡 (die) and 冲突 (conflict) is a candidate trigger mention. If there is an evidence that 冲突 (conflict) triggers an *Attack* event in the other documents, it is possible to identify 冲突 (conflict) as a trigger mention of *Attack* event in S7 for the high probability that events *Die* and *Attack* occur in the same document. We propose an inference mechanism on relevant events as follows.

Relevant Events: If a trigger mention is identified in a document, each candidate trigger mention in the same document will be recognized as true ones when it satisfies the following condition: this candidate trigger occurs in the other documents as an event trigger and refers to the relevant events of this identified trigger mentions.

Since the seed triggers have a high probability to trigger a specific event, to further explore those missing event mentions, we expand this inference mechanism following compositional semantics in Chinese and expand the condition as follows: This candidate trigger occurs in the other documents as an event trigger or contains one of the seed triggers, which refers to the relevant events of this identified trigger mentions.

5 Experimentation

In this section, we systematically evaluate our event inference mechanisms on the ACE 2005 Chinese corpus and provide the analysis.

5.1 Experimental Setting

The ACE 2005 Chinese corpus is the only available corpus in Chinese event extraction and it is used in all our experiments. This corpus contains 633 documents annotated with 33 predefined types. Due to evaluation on all 33 types is a hard work for the time-consuming bootstrapping procedure and the diversity of distinct event types, most of previous works selected part of event types for evaluation. In this paper, 3 event types (i.e. *Die*, *Injure* and *Attack*) are selected for evaluation, because they reflect the relevance of different event types and occur at different frequencies in the corpus. While events *Die* and *Injure* are easy to define, event *Attack* is rather complicated and can be divided into several subtypes. In the ACE 2005 Chinese corpus, almost one third of the annotated event mentions belong to the above three event types. Moreover, we report the experimental results on all 33 event types to further verify the effectiveness of our inference mechanisms in Subsection 5.2.

Unlike MUC shared task, which only distinguishes whether a sentence contains a specific event mention or not, we follow previous studies on the ACE 2005 corpus and report the performance of trigger-based event extraction: a trigger is correctly identified if its position and event type match a reference trigger. As for evaluation, we use the ground truth entities, time and values annotated in the ACE 2005 Chinese corpus, and report the micro-average Precision (P), Recall (R) and F1-score (F1).

Table 1 shows the seed triggers for the three event types. For example, only one seed trigger is provided for either the *Die* or *Injure* event, while three seed triggers are given for event *Attack*. Since the *Attack* event contains several distinct event subtypes, we assign one seed trigger to each of its major subtypes. Thus, all patterns whose triggers belong to the set of seed triggers are accepted as seed patterns automatically.

Type	Die	Injure	Attack
Seed triggers	死(die)	伤(injure)	攻击(attack), 冲突(conflict), 打(hit)

Table 1. Seed triggers of *Die*, *Injure* and *Attack* event types

Besides, all the sentences in the corpus are divided into words using a Chinese word segmentation tool (*ICTCLAS*) with all entities annotated in the corpus kept. We use *Berkeley Parser* and *Stanford Parser* to create the constituent and dependency parse trees.

5.2 Experimental Results

To verify the performance of our event inference mechanisms, it is compared with the refined baseline, a supervised model for Chinese event extraction. Table 2 shows the results of our event inference mechanisms with peak recall, precision and F1-score, following Liao and Grishman (2010a). Compared with the baseline, Table 2 shows that our event inference mechanisms improve the F1-score of Chinese event extraction by 8.5%, largely due to the improvement of 11.8% in recall. These results confirm the effectiveness of our event inference mechanisms in recovering missing event mentions. The disadvantage of our event inference mechanisms is the fact that it will also introduce some pseudo event mentions into our model and harm the precision. Additionally, there is still a big performance gap between our model and the supervised model and this leaves much room for future research.

Approach	Attack			Injure			Die			All (micro-average)		
	P(%)	R(%)	F1	P(%)	R(%)	F1	P(%)	R(%)	F1	P(%)	R(%)	F1
Baseline	71.4	36.6	48.4	93.2	41.7	57.6	90.1	44.0	59.3	79.7	39.4	52.7
+Event inference	70.9	47.5	56.9	83.2	54.6	65.9	80.8	57.2	67.0	75.5	51.2	61.2
Supervised model	70.4	72.5	71.4	85.3	78.4	81.7	83.9	92.9	88.1	77.2	78.4	77.8

Table 2. Performance of event inference mechanisms in Chinese event extraction (*Attack/Injure/Die*).

Table 2 also indicates the performance difference of our inference mechanisms for distinct event types. Among all event types, event *Attack* achieves the highest improvement (8.5%) in F1-score, with a dramatic improvement of 10.9% in recall and a less loss of 0.5% in precision. Event *Die* and *Injure* also gain a significant improvement of 7.7% and 8.3% in F1-score respectively, largely due to the increase in recall, while their precisions reduce rapidly due to those pseudo event mentions inferred by our inference mechanisms. However, the loss of precision of event *Attack* is much less than these of events *Die* and *Injure*. The reason is that the inference on role semantics mainly impacts on *Attack* events to remove pseudo event mentions.

To well evaluate different approaches, it is better to compare them on different corpora. Since the ACE 2005 Chinese corpus is the only available corpus in Chinese event extraction, we divide it into three sub-corpora according to data sources, i.e. Broadcast News, Newswire and WebLog, which are much different in various aspects, such as quality, length and style. Figure 1 compares the performance of different models on different sub-corpora. It indicates that our event inference mechanisms perfect better than the baseline in all three sub-corpora and that results confirm the huge influence of the event inference mechanisms. It also shows that the WebLog sub-corpus reports the worst F1-score due to the low document quality and the low percentage of relevant documents, and that the Newswire sub-corpus reports significantly better performance than the Broadcast News sub-corpus due to its spoken nature.

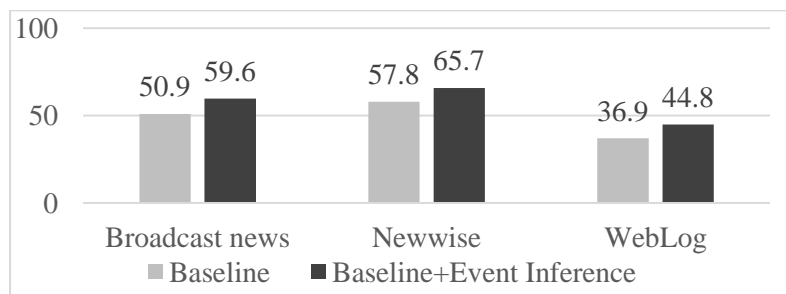


Figure 1. Performance comparison (F1-score) on different data sources.

To further verify the effectiveness of our event inference mechanisms, we evaluate them on all 33 event types. Due to event extraction is a domain-specific task, distinct event types have the different seed triggers and different pro-process procedures. In this paper, we just report the final results for the

sake of brevity. Table 3 shows the experimental results on all 33 event types and it ensures that our mechanisms are effective on extracting all event types. Compared with the baseline, our approach improves the F1-score by 7.6%, which is less than that reported in Table 2. Among all 33 event types, the performances of almost all event types associated with justice are higher than other event types for their unambiguous definitions and high coverage of seed triggers while event *Transport* achieves the lowest performance for its complexity and low coverage of seed triggers. Besides, the performance on all event types is lower than that on 3 event types and this result comes from the low performance of the *Transport* event which occupies almost 20% of all annotated event mentions in the ACE 2005 Chinese corpus.

Approach	P(%)	R(%)	F1
Baseline	70.7	34.2	46.1
+Event inference	65.2	45.7	53.7

Table 3. Performance of event inference mechanisms in Chinese event extraction (All 33 event types).

5.3 Analysis on Event Inference Mechanisms

Table 4 shows the contributions of the different event inference mechanisms. It is worthy to mention that an event mentions may be identified by both the semi-supervised model and the event inference mechanisms. In this paper, we attribute those extracted event mentions to the former and the contribution of our inference mechanisms is greater than those in Table 4.

Inference	P(%)	R(%)	F1
Baseline	79.7	39.4	52.7
+Inference on role semantics (RS)	87.5(+7.8)	39.1(-0.3)	54.1(+1.4)
+Inference on compositional semantic (CS)	85.7(+6.0)	43.7(+4.3)	57.8(+3.7)
+Inference on coreference events (CE)	83.0(+3.3)	45.8(+6.4)	59.0(+1.2)
+Inference on relevant events (RE)	75.7(-4.0)	51.3(+11.9)	61.2(+2.2)

Table 4. The contribution of event inference on Chinese event extraction.

Actually, inference mechanism **RS** is a filter to remove those pseudo event mentions and it can improve the precision (+7.8%), with a less lost (-0.3%) in recall. Moreover, it can also help the seed pattern generation to generate high quality seed patterns. Table 5 shows the contribution of **RS** on seed pattern generation and we report the result of Chinese event extraction which only uses the seed patterns⁵. It improves the accuracy from 75.8% to 82.5%, largely due to the decline (-30) in the set of pseudo event mentions. These results indicate that the object is a key clue to identify event mentions.

Method	#True event mentions	#Pseudo event mentions
w/o RS	273	87
w/ RS	269	57

Table 5. The contribution of **RS** on seed pattern generation.

Chen and Ji (2009b) have reported that almost 13% of Chinese triggers are in-word or cross-words and this figure ensures it is an important issue. Inference mechanism **CS** gains the highest improvement (+3.7%) in F1-score and this result indicates that compositional semantics is an effective way to solve such issue. The accuracy of this inference mechanism is very high (~92%) and most of the exceptions need the help of deep semantics since these instances are also hard to be distinguished by humans without the context.

Inference mechanisms **CE** and **RE** improve the F1-scores by 1.2% and 2.2% respectively. **CE** assumes all mentions of a word in a document only have one sense and it will introduce lots of pseudo event mentions to reduce precision. The experimental results also show that **RE** is an effective supplement of the document-centric view to mine event mentions. Although they derive from the similar

⁵ Since sometimes a pattern can infer both true event mentions and pseudo event mentions, it is hard to identify whether a pattern is relevant or irrelevant without the test data. Hence, we compare their extracted event mentions in this paper.

principle of occurrence of relevant events, they focus on different perspectives where **RE** is trigger-based and the document-centric view is pattern-based. **RE** ignores the difference on patterns and identifies event mentions on the occurrence of their relevant event mentions. In addition, sense shifting of Chinese words in different contexts is the main factor to extract lots of pseudo event mentions and then reduce the precision rapidly.

It's obvious that these inference mechanisms interact with others. In particular, almost 20% event mentions can be inferred by both **CE** and **RE** for the transitivity of event inference on coreference and relevant events. Besides, **RS** is not only beneficial to the semi-supervised model, but also helpful to the other inference mechanisms to further remove pseudo event mentions.

6 Conclusion

This paper proposes various kinds of novel linguistic knowledge-driven event inference mechanisms as a supplement of the semi-supervised Chinese event extraction to recover missing event mentions. The experimental results verify their effectiveness to extract the event mentions with infrequent patterns or without matching pattern. Although this paper focuses on Chinese language, most of the event inference mechanisms are language-independent and can be applied to other languages. Our future work will focus on how to apply our event inference mechanisms to other languages and introduce more effective inference mechanisms to further improve the performance of semi-supervised event extraction.

Acknowledgments

The authors would like to thank three anonymous reviewers for their comments on this paper. This research was supported by the National Natural Science Foundation of China under Grant No. 61331011 and No. 61272260, the National 863 Project of China under Grant No. 2012AA011102.

Reference

- Niranjan Balasubramanian, Stephen Soderland, Mausam and Oren Etzioni. 2013. *Generating Coherent Event Schemas at Scale*. In Proc. EMNLP 2013, pages 1721-1731, Seattle, WA.
- Nathanael Chambers and Dan Jurafsky. 2008. *Unsupervised Learning of Narrative Event Chains*. In Proc. ACL-HLT 2008, pages 787-797, Hawaii.
- Nathanael Chambers and Dan Jurafsky. 2009. *Unsupervised Learning of Narrative Schemas and Their Participants*. In Proc. ACL 2009, pages 602-610, Columbus, OH.
- Nathanael Chambers and Dan Jurafsky. 2011. *Template-Based Information Extraction without the Templates*. In Proc. ACL 2011, pages 976-986, Portland, OR.
- Hai Leong Chieu, Hwee Tou Ng and Yoong Keok Lee. 2003. *Closing the Gap: Learning-based Information Extraction Rivaling Knowledge-Engineering Methods*. In Proc. ACL 2003, pages 216-230, Sapporo, Japan.
- Zheng Chen and Heng Ji. 2009a. *Can One Language Bootstrap the Other: A Case Study on Event Extraction*. In Proc. NAACL-HLT 2009 Workshop on Semi-supervised Learning for Natural Language Processing, pages 66-74, Boulder, CO.
- Zheng Chen and Heng Ji. 2009b. *Language Specific Issue and Feature Exploration in Chinese Event Extraction*. In Proc. NAACL-HLT 2009, pages 209-212, Boulder, CO.
- Zhengdong Dong and Qiang Dong. 2006. *HowNet and the Computation of Meaning*. World Scientific Pub Co. Inc.
- Yu Hong, Jianfeng Zhang, Bin Ma, Jianmin Yao, Guodong Zhou and Qiaoming Zhu. 2011. *Using Cross-Entity Inference to Improve Event Extraction*. In Proc. ACL 2011, pages 1127-1136, Portland, OR.
- Ruihong Huang and Ellen Riloff. 2012. *Bootstrapped Training of Event Extraction Classifiers*. In Proc. EACL 2012, pages 286-295, Avignon, France.
- Heng Ji and Ralph Grishman. 2008. *Refining Event Extraction through Cross-Document Inference*. In Proc. ACL-HLT 2008, pages 254-262, Columbus, OH.

- Peifeng Li and Guodong Zhou. 2012. *Employing Morphological Structures and Sememes for Chinese Event Extraction*. In Proc. COLING 2012, pages 1619-1634, Mumbai, India.
- Peifeng Li, Qiaoming Zhu, and Guodong Zhou. 2013. *Argument Inference from Relevant Event Mentions in Chinese Argument Extraction*. In Proc. ACL 2013, pages 1477-1487, Sofia, Bugaria.
- Peifeng Li, Qiaoming Zhu, Guodong Zhou. 2014. *Using Compositional Semantics and Discourse Consistency to Improve Chinese Trigger Identification*. Information Processing and Management, 50: 399-415.
- Shasha Liao and Ralph Grishman. 2010a. *Filtered Ranking for Bootstrapping in Event Extraction*. In Proc. COLING 2010, pages 680-688, Beijing, China.
- Shasha Liao and Ralph Grishman. 2010b. *Using Document Level Cross-Event Inference to Improve Event Extraction*. In Proc. ACL 2010, pages 789-797, Uppsala, Sweden.
- Shasha Liao and Ralph Grishman. 2011. *Can Document Selection Help Semi-supervised Learning? A Case Study On Event Extraction*. In Proc. ACL 2011, pages 260-265, Portland, OR.
- Qun Liu and Sujian Li. 2002. *Word Similarity Computing Based on How-net*. In Proc. 3th Chinese Lexical Semantic Workshop, Taibei, Taiwan.
- Ting Liu and Tomek Strzalkowski. 2012. *Bootstrapping Events and Relations from Text*. In Proc. EACL 2012, pages 296-305, Avignon, France.
- Mstislav Maslennikov and Tat-Seng Chua. 2007. *A Multi-resolution Framework for Information Extraction from Free Text*. In Proc. ACL 2007, pages 592-599, Prague, Czech Republic.
- Siddharth Patwardhan and Ellen Riloff. 2009. *A Unified Model of Phrasal and Sentential Evidence for Information Extraction*. In Proc. EMNLP 2009, pages 151-160, Singapore.
- William Phillips and Ellen Riloff. 2007. *Exploiting Role-Identifying Nouns and Expressions for Information Extraction*. In Proc. RANLP 2007, pages 468-473, Borovets, Bulgaria.
- Ellen Riloff. 1996. *Automatically Generating Extraction Patterns from Untagged Text*. In Proc. AAAI 1996, pages 1044-1049, Portland, OR.
- Mark Stevenson and Mark Greenwood. 2005. *A Semantic Approach to IE Pattern Induction*. In Proc. ACL 2005, pages 379-386, Ann Arbor, MI.
- Kiyoshi Sudo, Satoshi Sekine, Ralph Grishman. 2001. *Automatic Pattern Acquisition for Japanese Information Extraction*. In Proc. HLT 2001, pages 1-7, San Diego, CA.
- Kiyoshi Sudo, Satoshi Sekine, Ralph Grishman. 2003. *An Improved Extraction Pattern Representation Model for Automatic IE Pattern Acquisition*. In Proc. ACL 2003, pages 224-231, Tokyo, Japan.
- Roman Yangarber, Ralph Grishman, Pasi Tapanainen and Silja Huttunen. 2000. *Automatic Acquisition of Domain Knowledge for Information Extraction*. In Proc. COLING 2000, pages 940-946, Hong Kong.
- Roman Yangarber. 2003. *Counter-Training in Discovery of Semantic Patterns*. In Proc. ACL 2003, pages 343-350, Sapporo, Japan.
- Jian Wang, Qian Xu, Hongfei Lin, Zhihao Yang, Yanpeng Li. 2013. *Semi-supervised Method for Biomedical Event Extraction*. Proteome Science, 11(Suppl 1): S17.