

Argument Inference from Relevant Event Mentions in Chinese Argument Extraction

Peifeng Li, Qiaoming Zhu, Guodong Zhou*

School of Computer Science & Technology
Soochow University, Suzhou, 215006, China

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

Abstract

As a paratactic language, sentence-level argument extraction in Chinese suffers much from the frequent occurrence of ellipsis with regard to inter-sentence arguments. To resolve such problem, this paper proposes a novel global argument inference model to explore specific relationships, such as *Coreference*, *Sequence* and *Parallel*, among relevant event mentions to recover those inter-sentence arguments in the sentence, discourse and document layers which represent the cohesion of an event or a topic. Evaluation on the ACE 2005 Chinese corpus justifies the effectiveness of our global argument inference model over a state-of-the-art baseline.

1 Introduction

The task of event extraction is to recognize event mentions of a predefined event type and their arguments (participants and attributes). Generally, it can be divided into two subtasks: trigger extraction, which aims to identify trigger/event mentions and determine their event type, and argument extraction, which aims to extract various arguments of a specific event and assign the roles to them. In this paper, we focus on argument extraction in Chinese event extraction. While most of previous studies in Chinese event extraction deal with Chinese trigger extraction (e.g., Chen and Ji, 2009a; Qin et al., 2010; Li et al., 2012a, 2012b), there are only a few on Chinese argument extraction (e.g., Tan et al., 2008; Chen and Ji, 2009b). Following previous studies, we divide argument extraction into two components, argument identification

and role determination, where the former recognizes the arguments in a specific event mention and the latter classifies these arguments by roles.

With regard to methodology, most of previous studies on argument extraction recast it as a Semantic Role Labeling (SRL) task and focus on intra-sentence information to identify the arguments and their roles. However, argument extraction is much different from SRL in the sense that, while the relationship between a predicate and its arguments in SRL can be mainly decided from the syntactic structure, the relationship between an event trigger and its arguments are more semantics-based, especially in Chinese, as a paratactic (e.g., discourse-driven and pro-drop) language with the wide spread of ellipsis and the open flexible sentence structure. Therefore, some arguments of a specific event mention are far away from the trigger and how to recover those inter-sentence arguments becomes a challenging issue in Chinese argument extraction. Consider the following discourse (from ACE 2005 Chinese corpus) as a sample:

D1: 巴勒斯坦自治政府否认和加沙走廊 20 号清晨造成两名以色列人**丧生(E1)**的**炸弹攻击(E2)**事件有关...表示将对这起**攻击(E3)**事件展开调查。(The Palestinian National Authority denied any involvement in the **bomb attack (E2)** occurred in the Gaza Strip on the morning of the 20th, which **killed (E1) two Israelites**. ... They claimed that they will be investigating this **attack (E3)**.) - From CBS20001120.1000.0823

In above discourse, there are three event mentions, one *kill* (E1) and two *Attack* (E2, E3). While it is relatively easy to identify 20 号清晨 (morning of 20th), 加沙走廊 (Gaza Strip) and 炸弹 (bomb) as the *Time*, *Place* and *Instrument* roles in E2 by a sentence-based argument

extractor, it is really challenging to recognize these entities as the arguments of its corefered mention E3 since to reduce redundancy in a Chinese discourse, the later Chinese sentences omit many of these entities already mentioned in previous sentences. Similarly, it is hard to recognize 两名以色列人 (two Israelites) as the *Target* role for event mention E2 and identify 炸弹 (bomb) as the *Instrument* role for event mention E1. An alternative way is to employ various relationships among relevant event mentions in a discourse to infer those inter-sentence arguments.

The contributions of this paper are:

- 1) We propose a novel global argument inference model, in which various kinds of event relations are involved to infer more arguments on their semantic relations.
- 2) Different from Liao and Grishman (2010) and Hong et al. (2011), which only consider document-level consistency, we propose a more fine-grained consistency model to enforce the consistency in the sentence, discourse and document layers.
- 3) We incorporate argument semantics into our global argument inference model to unify the semantics of the event and its arguments.

The rest of this paper is organized as follows. Section 2 overviews the related work. Section 3 describes a state-of-the-art Chinese argument extraction system as the baseline. Section 4 introduces our global model in inferring those inter-sentence arguments. Section 5 reports experimental results and gives deep analysis. Finally, we conclude our work in Section 6.

2 Related Work

Almost all the existing studies on argument extraction concern English. While some apply pattern-based approaches (e.g., Riloff, 1996; Califf and Mooney, 2003; Patwardhan and Riloff, 2007; Chambers and Jurafsky, 2011), the others use machine learning-based approaches (e.g., Grishman et al., 2005; Ahn, 2006; Patwardhan and Riloff, 2009; Lu and Roth, 2012), most of which rely on various kinds of features in the context of a sentence. In comparison, there are only a few studies exploring inter-sentence information or argument semantics (e.g., Liao and Grishman, 2010; Hong et al., 2011; Huang and Riloff, 2011, 2012).

Compared with the tremendous work on English event extraction, there are only a few studies (e.g., Tan et al., 2008; Chen and Ji, 2009b;

Fu et al., 2010; Qin et al., 2010; Li et al., 2012) on Chinese event extraction with focus on either feature engineering or trigger expansion, under the same framework as English trigger identification. In addition, there are only very few of them focusing on Chinese argument extraction and almost all aim to feature engineering and are based on sentence-level information and recast this task as an SRL-style task. Tan et al. (2008) introduce multiple levels of patterns to improve the coverage in Chinese argument classification. Chen and Ji (2009b) apply various kinds of lexical, syntactic and semantic features to address the special issues in Chinese argument extraction. Fu et al. (2010) use a feature weighting scheme to re-weight various features for Chinese argument extraction. Li et al. (2012b) introduce more refined features to the system of Chen and Ji (2009b) as their baseline.

Specially, several studies have successfully incorporated cross-document or document-level information and argument semantics into event extraction, most of them focused on English.

Yangarber et al. (2007) apply a cross-document inference mechanism to refine local extraction results for the disease name, location and start/end time. Mann (2007) proposes some constraints on relationship rescoring to impose the discourse consistency on the CEO's personal information. Chambers and Jurafsky (2008) propose a narrative event chain which are partially ordered sets of event mentions centered around a common protagonist and this chain can represent the relationship among the relevant event mentions in a document.

Ji and Grishman (2008) employ a rule-based approach to propagate consistent triggers and arguments across topic-related documents. Liao and Grishman (2010) mainly focus on employing the cross-event consistency information to improve sentence-level trigger extraction and they also propose an inference method to infer the arguments following role consistency in a document. Hong et al. (2011) employ the background information to divide an entity type into more cohesive subtypes to create the bridge between two entities and then infer arguments and their roles using cross-entity inference on the subtypes of entities. Huang and Riloff (2012) propose a sequentially structured sentence classifier which uses lexical associations and discourse relations across sentences to identify event-related document contexts and then apply it to recognize arguments and their roles on the relation among triggers and arguments.

3 Baseline

In the task of event extraction as defined in ACE evaluations, an event is defined as a specific occurrence involving participants (e.g., *Person*

flexible sentence structure. Hence, different word orders can often express the same semantics. For example, a *Die* event mention “Three person died in this accident.” can be expressed in many different orders in Chinese, such as “在事故中三人死亡。”, “事故中死亡三人。”, “三人在事故中死亡。”, etc.

In a word, above two issues indicate that syntactic feature-based approaches are limited in identifying Chinese arguments and it will lead to low recall in argument identification. Therefore, employing those high level information to capture the semantic relation, not only the syntactic structure, between the trigger and its long distance arguments is the key to improve the performance of the Chinese argument identification. Unfortunately, it is really hard to find their direct relations since they always appear in different clauses or sentences. An alternative way is to link the different event mentions with their predicates (triggers) and use the trigger as a bridge to connect the arguments to the trigger in another event mention indirectly. Hence, the semantic relations among event mentions are helpful to be a bridge to identify those inter-sentence arguments.

4.2 Relations of Event Mention Pairs

In a discourse, most event mentions are surrounding a specific topic. It’s obvious that those mentions have the intrinsic relationships to reveal the essential structure of a discourse. Those relevant semantics-based relations are helpful to infer the arguments for a specific trigger mention when the syntactic relations in Chinese argument extraction are not as effective as that in English. In this paper, we divide the relations among relevant event mentions into three categories: *Coreference*, *Sequence* and *Parallel*.

An event may have more than one mention in a document and *coreference* event mentions refer to the same event, as same as the definition in the ACE evaluations. Those *coreference* event mentions always have the same arguments and roles. Therefore, employing this relation can infer the arguments of an event mention from their *Coreference* ones. For example, we can recover the *Time*, *Place* and *Instrument* for E3 via its *Coreference* mention E2 in discourse D1, mentioned in Section 1.

Li et al. (2012a) find out that sometimes two trigger mentions are within a Chinese word whose morphological structure is *Coordination*.

Take the following sentence as a sample:

D2: 一名 17 岁的少年劫持一辆巴士, **刺(E4)死(E5)** 一名妇女。(A 12-year-old younger hijacked a bus and then **stabbed (E4)** a woman to **death (E5)**.) - From ZBN20001218.0400.0005

In D2, 刺死 (stab a person to death) is a trigger with the *Coordination* structure and can be divided into two single-morpheme words 刺 (stab) and 死 (die) while the former triggers an *Attack* event and the latter refers to a *Die* one. It’s interesting that they share all arguments in this sentence. The relation between those event mentions whose triggers merge a Chinese word or share the subject and the object are *Parallel*. For the errors in the syntactic parsing, the second single-morpheme trigger is often assigned a wrong tag (e.g., NN, JJ) and this leads to the errors in the argument extraction. Therefore, inferring the arguments of the second single-morpheme trigger from that of the first one based on *Parallel* relation is also an available way to recover arguments.

Like that the topic is an axis in a discourse, the relations among those relevant event mentions with the different types is the bone to link them into a narration. There are a few studies on using the event relations in NLP (e.g., summarization (Li et al., 2006), learning narrative event chains (Chambers and Jurafsky, 2007)) to ensure its effectiveness. In this paper, we define two types of *Sequence* relations of relevant event mentions: *Cause* and *Temporal* for their high probabilities of sharing arguments.

The *Cause* relation between the event mentions are similar to that in the Penn Discourse TreeBank 2.0 (Prasad et al., 2008). For example, an *Attack* event often is the cause of an *Die* or *Injure* event. Our *Temporal* relation is limited to those mentions with the same or relevant event types (e.g., *Transport* and *Arrest*) for the high probabilities of sharing arguments. Take the following discourse as a sample:

D3: 这批战俘离开(E6)阿尔及利亚西部城市廷杜夫前往(E7)摩洛哥西南部城市阿加迪尔。(These prisoners left (E6) Tindouf, a western city of Algeria, and went (E7) to Agadir, a southwestern city of Morocco.) - From Xin20001215.2000.0158

In D3, there are two *Transport* mentions and it is natural to infer 阿加迪尔 (Agadir) as the *Destination* role of E6 and 廷杜夫 (Tindouf) as the *Origin* role of E7 via their *Sequence* relation.

4.3 Identifying Relations of Event Mention Pairs

Currently, there are only few studies focusing on such area (e.g., Ahn, 2006; Chamber and Jurafsky, 2007; Huang and Riloff, 2012; Do et al., 2012) and their approaches cannot be introduced to our system directly for the language nature and the different goal. We try to achieve a higher accuracy in this stage so that our argument inference can recover more true arguments.

Inspired by Li and Zhou (2012), we also use the morphological structure to identify the *Parallel* relation. Two parallel event mentions with the adjacent trigger mentions w_1 and w_2 must satisfy follows two conditions:

1) $Morph(w_1, w_2)$ is *Coordination*

2) $HM(w_1) \in T_i, HM(w_2) \in T_j \quad i \neq j$

where $Morph(w_1, w_2)$ is a function to recognize the morphological structure of joint word $w_1 w_2$, $HM(w_i)$ is to identify the head morpheme¹ in word w_i and T_i is the set of the head morphemes with i th event type. These constraints are enlightened by the fact that only Chinese words with *Coordination* structure can be divided into two new words and each word can trigger an event with the different event type². The implementation of $Morph(w_1, w_2)$ and $HM(w)$ are described in Li and Zhou (2012).

The *Coreference* relation is divided into two types: *Noun-based Coreference (NC)* and *Event-based Coreference (EC)* while the former always uses a verbal noun to refer to an event mentioned in current or previous sentence and the latter is that an event is mentioned twice or more actually. For example, the relation between E2 and E3 in D1 is *NC* while the trigger of E3 is only a verbal noun without any direct arguments and it refers to E2.

We adopt a simple rule to recognize those *NC* relations: for each event mention whose trigger is a noun and doesn't act as the subject/object, we regard their relation as *NC* if there is another event mention with the same trigger in current or previous sentence.

Inspired by Ahn (2006), we use the following conditions to infer the *EC* relations between two event mentions with the same event type:

- 1) Their trigger mentions refer to the same trigger;
- 2) They have at least one same or similar

¹ It acts as the governing semantic element in a Chinese word.

² If they have the same event type, they will be regarded as a single event mention.

subject/object;

- 3) The score of cosine similarity of two event mentions is more than a threshold³.

Finally, for the *Sequence* relation, instead of identifying and classifying the relations clearly and correctly, our goal is to identify whether there are relevant event mentions in a long sentence or two adjacent short sentences who share arguments. Algorithm 1 illustrates a knowledge-based approach to identify the *Sequence* event relation in a discourse for any two trigger mentions tri_1 and tri_2 as follows:

Algorithm 1

```

1: input:  $tri_1$  and  $tri_2$  and their type  $et_1$  and  $et_2$ 
2: output: whether their relation is Sequence
3: begin
4:    $hm_1 \leftarrow HM(tri_1); hm_2 \leftarrow HM(tri_2)$ 
5:    $MP \leftarrow FindAllMP(hm_1, et_1, hm_2, et_2)$ 
6:   for any  $mp_i$  in  $MP$ 
7:     if  $ShareArg(mp_i)$  is true then
8:       return true // Sequence
9:     end if
10:  end for
11:  return false
12: end

```

In algorithm 1, $HM(tri)$ is to identify the head morpheme in trigger tri and $FindAllMP(hm_1, et_1, hm_2, et_2)$ is to find all event mention pairs in the training set which satisfy the condition that their head morphemes are hm_1 and hm_2 , and their event types are et_1 and et_2 respectively. Besides, $ShareArg(mp_i)$ is used to identify whether the event mention pair mp_i sharing at least one argument. In this algorithm, since the relations on the event types are too coarse, we introduce a more fine-gained *Sequence* relation both on the event types and the head morphemes of the triggers which can divide an event type into many subtypes on the head morpheme. Li and Zhou (2012) have ensured the effectiveness of using head morpheme to infer the triggers and our experiment results also show it is helpful for identifying relevant event mentions which aims to the higher accuracy.

4.4 Global Argument Inference Model

Our global argument inference model is composed of two steps: 1) training two sentence-based classifiers: argument identifier (AI) and role determiner (RD) that estimate the score of a candidate acts as an argument and belongs to a

³ The threshold is tuned to 0.78 on the training set.

specific role following Section 3. 2) Using the scores of two classifiers and the event relations in a sentence, a discourse or a document, we perform global optimization to infer those missing or long distance arguments and their roles.

To incorporate those event relations with our global argument inference model, we regard a document as a tree and divide it into three layers: document, discourse and sentence. A document is composed of a set of the discourses while a discourse contains three sentences. Since almost all arguments (~98%) of a specific event mention in the ACE 2005 Chinese corpus appear in the sentence containing the specific event mention and its two adjacent sentences (previous and next sentences), we only consider these three sentences as a discourse to simplify the process of identifying the scope of a discourse.

We incorporate different event relations into our model on the different layer and the goal of our global argument inference model is to achieve the maximized scores over a document on its three layers and two classifiers: AI and RD. The score of document D is defined as

$$\hat{D} = \arg \max_{X,Y} (\alpha \sum_{I_i \in D} \sum_{S_{\langle i,j \rangle} \in I_i} \sum_{T_{\langle i,j,k \rangle} \in S_{\langle i,j \rangle}} \sum_{A_Z \in T_{\langle i,j,k \rangle}} (f_I(E_Z)X_Z + (1-f_I(E_Z))(1-X_Z)) \quad (1)$$

$$+ (1-\alpha) \sum_{I_i \in D} \sum_{S_{\langle i,j \rangle} \in I_i} \sum_{T_{\langle i,j,k \rangle} \in S_{\langle i,j \rangle}} \sum_{A_Z \in T_{\langle i,j,k \rangle}} \sum_{m \in R} (f_D(E_Z, R_m)Y_{\langle Z,m \rangle} + (1-f_D(E_Z, R_m))(1-Y_{\langle Z,m \rangle}))$$

$$s.t. \quad X_Z \in \{0,1\} \quad (2)$$

$$Y_{\langle Z,m \rangle} \in \{0,1\} \quad (3)$$

$$X_Z \geq Y_{\langle Z,m \rangle} \quad \forall m \in R \quad (4)$$

$$X_Z = \sum_{m \in R} Y_{\langle Z,m \rangle} \quad (5)$$

where I_i is the i th discourses in document D ; $S_{\langle i,j \rangle}$ is the j th sentences in discourse I_i ; $T_{\langle i,j,k \rangle}$ is the k th event mentions in sentence $S_{\langle i,j \rangle}$; $A_{\langle i,j,k,l \rangle}$ is the l th candidate arguments in event mention $T_{\langle i,j,k \rangle}$; Z is used to denote $\langle i,j,k,l \rangle$; $f_I(E_Z)$ is the score of AI identifying entity mention E_Z as an argument, where E_Z is the l th entity of the k th event mention of the j th sentence of the i th discourse in document D . $f_D(E_Z, R_m)$ is the score of RD assigning role R_m to argument E_Z . Finally, X_Z and $Y_{\langle Z,m \rangle}$ are the indicators denoting whether entity E_Z is an argument and whether the role R_m is assigned to entity E_Z respectively. Besides, Eq. 4 and Eq. 5 are the inferences to enforce that:

- 1) if an entity belongs to a role, it must be an argument;
- 2) if a entity is an argument of a specific event mention, it must have a role.

Parallel relation: Sentence-based optimization is used to incorporate the *Parallel* relation of two event mentions into our model and they share all arguments in a sentence. Since different event type may have different role set, each role in a specific event should be mapped to the corresponding role in its *Parallel* event when they have the different event type. For example, the argument “一名 17 岁的少年” (A 12-year-old younger) in D2 acts as the *Attacker* role in the *Attack* event and the *Agent* role in the *Die* event. We learn those role-pairs from the training set and Table 2 shows part of the role relations learning from the training set.

Event type pair	Role pair
Attack-Die	Attacker-Agent; Target-Victim;...
Injure-Die	Agent-Agent; Victim-Victim;...
Transport-Demonstrate	Artifact-Entity; Destination-Place;...

Table 2. Part of role-pairs for those event mention pairs with *Parallel* relation.

To infer the arguments and their roles on the *Parallel* relation, we enforce the consistency on the role-pair as follows:

$$Y_{\langle i,j,k,l,m \rangle} = Y_{\langle i,j,k',l',m' \rangle} \quad (6)$$

$$\forall I_i \in D \wedge S_{\langle i,j \rangle} \in I_i \wedge T_{\langle i,j,k \rangle} \in S_{\langle i,j \rangle} \wedge T_{\langle i,j,k' \rangle} \in S_{\langle i,j \rangle} \wedge A_{\langle i,j,k,l \rangle} \in T_{\langle i,j,k \rangle} \wedge A_{\langle i,j,k',l' \rangle} \in T_{\langle i,j,k' \rangle} \wedge \langle m,m' \rangle \in RP_{et_h \times et_{h'}} \wedge E_{\langle i,j,k,l \rangle} = E_{\langle i,j,k',l' \rangle}$$

where $RP_{et_h \times et_{h'}}$ is the set of role-pairs between two *Parallel* event mention et_h and $et_{h'}$ and $E_{\langle i,j,k,l \rangle} = E_{\langle i,j,k',l' \rangle}$ means they refer to the same entity mention. With the transitivity between the indicators X and Y , Eq. 6 also enforces the consistency on $X_{\langle i,j,k,l \rangle}$ and $X_{\langle i,j,k',l' \rangle}$.

Coreference relation: Since the *NC* and *EC* relation between two event mentions are different in the event expression, we introduce the discourse-based optimization for the former and document-based optimization for the latter.

For two *NC* mentions, we ensure that the succeeding mentions can inherit the arguments from the previous one. To enforce this consistency, we just replace all $f_I(E_Z)$ and $f_D(E_Z, R_m)$ of the succeeding event mention with that of the previous one, since the previous one have the more context information.

As for two *EC* event mentions, algorithm 2 shows how to create the constraints for our

global argument inference model to infer arguments and roles.

Algorithm 2

1: **input:** two event mentions T, T' and their arguments set A and A'
2: **output:** the constraints set C
3: **begin**
4: **for** each argument a in A **do**
5: $a' \leftarrow \text{FindSim}(a)$
6: **if** $a' \neq \emptyset$ **then**
7: $C \leftarrow C \cup \text{Consistency}(Y_a, Y_{a'})$
8: **end if**
9: **end for**
10: **end**

In algorithm 2, the function $\text{FindSim}(a)$ is used to find a similar candidate argument a' in A' for a . If it's found, we enforce the consistency of argument a and a' in the role by using $\text{Consistency}(Y_a, Y_{a'})$ where Y_a and $Y_{a'}$ are the indicators in Eq. 1. To evaluate the similarity between two candidates a and a' , we regard them as similar ones when they are the same word or in the same entity coreference chain. We use a coreference resolution tool to construct the entity coreference chains, as described in Kong et al (2010).

Sequence relation: For any two event mentions in a discourse, we use the event type pair with their head morphemes (e.g., *Attack: 炸 (burst) - Die: 死 (die)*, *Trial-Hearing: 审 (trial) - Sentence: 判 (sentence)*) to search the training set and then obtain the probabilities of sharing the arguments as mentioned in algorithm 1. We denoted $\text{Pro}_{\langle et, et', HM(tri), HM(tri'), R_m, R_m' \rangle}$ as the probability of the trigger mentions tri and tri' (their event types are et and et' respectively.) sharing an argument whose roles are R_m and R_m' respectively. We propose following discourse-based constraint to enforce the consistency between the roles of two arguments, which are related semantically, temporally, causally or conditionally, based on the probability of sharing an argument and the absolute value of the difference between the scores of RD:

$$\begin{aligned}
Y_{\langle i, j, k, l, m \rangle} &= Y_{\langle i, j', k', l', m' \rangle} \\
\forall i \in D \wedge S_{\langle i, j \rangle}, S_{\langle i, j' \rangle} \in I_i \wedge T_{\langle i, j, k \rangle} \in S_{\langle i, j \rangle} \\
&\wedge T_{\langle i, j', k' \rangle} \in S_{\langle i, j' \rangle} \wedge m, m' \in R \wedge E_{\langle i, j, k, l \rangle} = E_{\langle i, j', k', l' \rangle} \\
&\wedge \text{Pro}_{\langle et, et', HM(tri), HM(tri'), R_m, R_m' \rangle} > \delta \wedge \\
&\left| f_D(E_{\langle i, j, k, l \rangle}, R_m) - f_D(E_{\langle i, j', k', l' \rangle}, R_m') \right| > \lambda
\end{aligned} \tag{7}$$

where δ and λ are the thresholds learned from the

development set; tri and tri' are triggers of k th and k' th event mention whose event types are et and et' in $S_{\langle i, j \rangle}$ and $S_{\langle i, j' \rangle}$ respectively.

4.5 Incorporating Argument Semantics into Global Argument Inference Model

We also introduce the argument semantics, which represent the semantic relations of argument-argument pair, argument-role pair and argument-trigger pair, to reflect the cohesion inside an event. Hong et al. (2011) found out that there is a strong argument and role consistency in the ACE 2005 English corpus. Those consistencies also occur in Chinese and they reveal the relation between the trigger and its arguments, and also explore the relation between the argument and its role. Besides, those entities act as non-argument also have the consistency with high probabilities.

To let the global argument inference model combine those knowledges of argument semantics, we compute the prior probabilities $P(X_{\langle i, j \rangle} = 1)$ and $P(Y_{\langle i, j, m \rangle} = 1)$ that entity en_j occurs in a specific event type et_i as an argument and its role is R_m respectively. To overcome the sparsity of the entities, we cluster those entities into more cohesive subtype following Hong et al. (2011). Hence, following the independence assumptions described by Berant et al. (2011), we modify the $f_I(E_Z)$ and $f_D(E_Z, R_m)$ in Eq. 1 as follows:

$$f_I(E_Z) = \log \frac{P(X_Z = 1 | F_Z)P(X_Z = 1)}{(1 - P(X_Z = 1 | F_Z))P(X_Z = 0)} \tag{8}$$

$$f_D(E_Z, R_m) = \log \frac{P(Y_{\langle Z, m \rangle} = 1 | F_{\langle Z, m \rangle})P(X_{\langle Z, m \rangle} = 1)}{(1 - P(X_{\langle Z, m \rangle} = 1 | F_{\langle Z, m \rangle}))P(X_{\langle Z, m \rangle} = 0)} \tag{9}$$

where $P(X_Z = 1 | F_Z)$ and $P(Y_{\langle Z, m \rangle} = 1 | F_{\langle Z, m \rangle})$ are the probabilities from the AI and AD respectively while F_Z and $F_{\langle Z, m \rangle}$ are the feature vectors. Besides, $P(X_{\langle Z, m \rangle} = 1)$ and $P(X_Z = 1)$ are the prior probabilities learning from the training set.

5 Experimentation

In this section, we first describe the experimental settings and the baseline, and then evaluate our global argument inference model incorporating with relevant event mentions and argument semantics to infer arguments and their roles.

5.1 Experimental Settings and Baseline

For fair comparison, we adopt the same experimental settings as the state-of-the-art event extraction system (Li et al. 2012b) and all the

evaluations are experimented on the ACE 2005 Chinese corpus. We randomly select 567 documents as the training set and the remaining 66 documents as the test set. Besides, we reserve 33 documents in the training set as the development set and use the ground truth entities, times and values for our training and testing. As for evaluation, we also follow the standards as defined in Li et al. (2012b). Finally, all the sentences in the corpus are divided into words using a Chinese word segmentation tool (*ICTCLAS*)¹ with all entities annotated in the corpus kept. We use *Berkeley Parser*² and *Stanford Parser*³ to create the constituent and dependency parse trees. Besides, the ME tool (*Maxent*)⁴ is employed to train individual component classifiers and *lp_solver*⁵ is used to construct our global argument inference model.

Besides, all the experiments on argument extraction are done on the output of the trigger extraction system as described in Li et al. (2012b). Table 3 shows the performance of the baseline trigger extraction system and Line 1 in Table 4 illustrates the results of argument identification and role determination based on this system.

Trigger identification			Event type determination		
P(%)	R(%)	F1	P(%)	R(%)	F1
74.4	71.9	73.1	71.4	68.9	70.2

Table 3. Performance of the baseline on trigger identification and event type determination.

5.2 Inferring Arguments on Relevant Event Mentions and Argument Semantics

We develop a baseline system as mentioned in Section 3 and Line 2 in Table 4 shows that it slightly improves the F1-measure by 0.9% over Li et al. (2012b) due to the incorporation of more refined features. This result indicates the limitation of syntactic-based feature engineering.

Before evaluating our global argument inference model, we should identify the event relations between two mentions in a sentence, a discourse or a document. The experimental results show that the accuracies of identifying *NC*, *EC*, *Parallel* and *Sequence* relation are 80.0%, 72.4%, 88.5% and 87.7% respectively. Those results ensure that our simple methods are

effective. Our statistics on the development set shows almost 65% of the event mentions are involved in those *Coreference*, *Parallel* and *Sequence* relations, which occupy 63%, 50%, 9% respectively⁶. Most of the exceptions are isolated event mentions.

System	Argument identification			Argument role determination		
	P(%)	R(%)	F1	P(%)	R(%)	F1
Li et al.(2012b)	59.1	57.2	58.1	55.8	52.1	53.9
Baseline	60.5	57.6	59.0	55.7	53.0	54.4
BIM	59.3	60.1	59.7	54.4	55.2	54.8
BIM+RE	60.2	65.6	62.8	55.0	60.0	57.4
BIM+RE+AS	62.9	66.1	64.4	57.2	60.2	58.7

Table 4. Performance comparison of argument extraction on argument identification and role determination.

Once the classifier AI and RD are trained, we would like to apply our global argument inference model to infer more inter-sentence arguments and roles. To achieve an optimal solution, we formulate the global inference problem as an Integer Linear Program (ILP), which leads to maximize the objective function. ILP is a mathematical method for constraint-based inference to find the optimal values for a set of variables that maximize an objective function in satisfying a certain number of constraints. In the literature, ILP has been widely used in many NLP applications (e.g., Barzilay and Lapata, 2006; Do et al., 2012; Li et al., 2012b).

For our systems, we firstly evaluate the performance of our basic global argument inference model (BIM) with the Eq. 2–5 which enforce the consistency on AI and RD and then introduce the inference on the relevant event mentions (RE) and argument semantics (AS) to BIM. Table 4 shows their results and we can find out that:

- 1) BIM only slightly improves the performance in F1-measure, as the result of more increase in recall (R) than decrease in precision (P). This suggests that those constraints just enforcing the consistency on AI and RD is not effective enough to infer more arguments.
- 2) Compared to the BIM, our model BIM+RE enhances the performance of argument identification and role determination by 3.1% and 2.6% improvement in F1-measure respectively. This suggests the effectiveness

¹ <http://ictclas.org/>

² <http://code.google.com/p/berkeleyparser/>

³ <http://nlp.stanford.edu/software/lex-parser.shtml>

⁴ <http://mallet.cs.umass.edu/>

⁵ <http://lpsolve.sourceforge.net/5.5/>

⁶ 20% of the mentions belongs to both *Coreference* and *Sequence* relations.

of our global argument inference model on the relevant event mentions to infer inter-sentence arguments. Table 5 shows the contributions of the different event relations while the *Sequence* relation gains the highest improvement of argument identification and role determination in F1-measure respectively.

Constraint	Argument identification			Argument role determination		
	P(%)	R(%)	F1	P(%)	R(%)	F1
BIM	59.3	60.1	59.7	54.4	55.2	54.8
+Parallel	+0.6	+0.7	+0.6	+0.4	+0.6	+0.5
+NC	+0.0	+0.8	+0.4	-0.2	+0.6	+0.2
+EC	+0.6	+1.2	+0.9	+0.5	+1.0	+0.7
+ Sequence	-0.3	+2.8	+1.2	-0.2	+2.6	+1.1

Table 5. Contributions of different event relations on argument identification and role determination. (Incremental)

- 3) Our model BIM+ER+AS gains 1.6% improvement for argument identification, and 1.3% for role determination. The results ensure that argument semantics not only can improve the performance of argument identification, but also is helpful to assign a correct role to an argument in role determination.

Table 3 shows 25.6% of trigger mentions introduced into argument extraction are pseudo ones. If we use the golden trigger extraction, our exploration shows that the precision and recall of argument identification can be up to 78.6% and 88.3% respectively. Table 6 shows the performance comparison of argument extraction on AI and RD given golden trigger extraction. Compared to the Baseline, our system improves the performance of argument identification and role determination by 6.4% and 5.8% improvement in F1-measure respectively, largely due to the dramatic increase in recall of 10.9% and 10.4%.

System	Argument identification			Argument role determination		
	P(%)	R(%)	F1	P(%)	R(%)	F1
Baseline	76.2	77.4	76.8	70.4	72.0	71.2
Model2	78.6	88.3	83.2	72.3	82.4	77.0

Table 6. Performance comparison of argument identification and type determination. (Golden trigger extraction)

5.3 Discussion

The initiation of our paper is that syntactic features play an important role in current machine learning-based approaches for English

event extraction, however, their effectiveness is much reduced in Chinese. So the improvement of our model for English event extraction is much less than that of Chinese. However, our model can be an effective complement of the sentence-level English argument extraction systems since the performance of argument extraction is still low in English and using discourse-level information is a way to improve its performance, especially for those event mentions whose arguments spread in complex sentences.

Moreover, our exploration shows that our global argument inference model can mine those arguments within a long distance which are unannotated as arguments of a special event mention in the corpus since the annotators just tagged arguments in a narrow scope or omitted a few arguments. Actually, they are the true ones to our knowledge and are more than 30.6% of those pseudo arguments inferred by our model. This ensures that our global argument inference model and those relations among event mentions is helpful to argument extraction.

6 Conclusion

In this paper we propose a global argument inference model to extract those inter-sentence arguments due to the nature of Chinese that it is a discourse-driven pro-drop language with the wide spread of ellipsis and the open flexible sentence structure. In particular, we incorporate various kinds of event relations and the argument semantics into the model in the sentence, discourse and document layers which represent the cohesion of an event or a topic. The experimental results ensure that our global argument inference model outperforms the state-of-the-art system.

In future work, we will focus on introducing more semantic information and cross-document information into the global argument inference model to improve the performance of argument 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. 61070123, No. 61272260 and No. 61273320, the National 863 Project of China under Grant No. 2012AA011102. The co-author tagged with “*” is the corresponding author.

References

- David Ahn. 2006. *The Stages of Event Extraction*. In Proc. COLING/ACL 2006 Workshop on Annotating and Reasoning about Time and Events. Pages 1-8, Sydney, Australia.
- Regina Barzilay and Miralla Lapata. 2006. *Aggregation via Set Partitioning for Natural Language Generation*. In Proc. NAACL 2006, pages 359-366, New York City, NY.
- Jonathan Berant, Ido Dagan and Jacob Goldberger. 2011. *Global Learning of Typed Entailment Rules*. In Proc. ACL 2011, pages 610-619, Portland, OR.
- Mary Elaine Califf and Raymond J. Mooney. 2003. *Bottom-up Relational Learning of Pattern Matching rules for Information Extraction*. Journal of Machine Learning Research, 4:177-210.
- Nathanael Chambers and Dan Jurafsky. 2008. *Unsupervised Learning of Narrative Event Chains*. In Proc. ACL 2008, pages 789-797, Columbus, OH.
- Nathanael Chambers and Dan Jurafsky. 2011. *Template-based Information Extraction without the Templates*. In Proc. ACL 2011, pages 976-986, Portland, OR.
- 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, Colorado.
- 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, Colorado.
- Zhengdong Dong and Qiang Dong. 2006. *HowNet and the Computation of Meaning*. World Scientific Pub Co. Inc.
- Quang Xuan Do, Wei Lu and Dan Roth. 2012. *Joint Inference for Event Timeline Construction*. In Proc. EMNLP 2012, pages 677-687, Jeju, Korea.
- Jianfeng Fu, Zongtian Liu, Zhaoman Zhong and Jianfang Shan. 2010. *Chinese Event Extraction Based on Feature Weighting*. Information Technology Journal, 9: 184-187.
- Ralph Grishman, David Westbrook and Adam Meyers. 2005. *NYU's English ACE 2005 System Description*. In Proc. ACE 2005 Evaluation Workshop, Gaithersburg, MD.
- 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.

- for Information Extraction*. In Proc. EMNLP 2009, pages 151-160, Singapore.
- Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi and Bonnie Webber. 2008. *The Penn Discourse Treebank 2.0*. In Proc. LREC 2008, pages 2961-2968, Marrakech, Morocco.
- Bing Qin, Yanyan Zhao, Xiao Ding, Ting Liu and Guofu Zhai. 2010. *Event Type Recognition Based on Trigger Expansion*. Tsinghua Science and Technology, 15(3): 251-258, Beijing, China.
- Ellen Riloff. 1996. *Automatically Generating Extraction Patterns from Untagged Text*. In Proc. AAAI 1996, pages 1044-1049, Portland, OR.
- Hongye Tan, Tiejun Zhao, Jiaheng Zheng. 2008. *Identification of Chinese Event and Their Argument Roles*. In Proc. 2008 IEEE International Conference on Computer and Information Technology Workshops, pages 14-19, Sydney, Australia.
- Nianwen Xue and Yaqin Yang. 2010. *Chinese Sentence Segmentation as Comma Classification*. In Proc. ACL 2010, pages 631-635, Uppsala, Sweden.
- Roman Yangarber, Clive Best, Peter von Etter, Flavio Fuart, David Horby and Ralf Steinberger. 2007. *Combining Information about Epidemic Threats from Multiple Sources*. In Proc. RANLP 2007 Workshop on Multi-source, Multilingual Information Extraction and Summarization, pages 41-48, Borovets, Bulgaria.