Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence

Joint Modeling of Argument Identification and Role Determination in Chinese Event Extraction with Discourse-Level Information

Peifeng Li, Qiaoming Zhu and Guodong Zhou*
School of Computer Science & Technology
Soochow University, Suzhou, 215006, China
{pfli, qmzhu, gdzhou}@suda.edu.cn

Abstract

Argument extraction is a challenging task in event extraction. However, most of previous studies focused on intra-sentence information and failed to extract inter-sentence arguments. This paper proposes a discourse-level joint model of argument identification and role determination to infer those inter-sentence arguments in a discourse. Moreover, to better represent the relationship among relevant event mentions and the relationship between an event mention and its arguments in a discourse, this paper introduces various kinds of corpus-based and discourse-based constraints in the joint model, either automatically learned or linguistically motivated. Evaluation on the ACE 2005 Chinese corpus justifies the effectiveness of our joint model over a strong baseline in Chinese argument extraction, in particular argument identification.

1 Introduction

Generally, events can be defined as "Who did What to Whom Where, When and How". As a classic task of Information Extraction (IE), event extraction recognizes event mentions of a predefined event type and their arguments[1] in text and is always divided into two subtasks: event trigger extraction and argument extraction. In the literature, most of previous studies aimed at trigger extraction (e.g., Chen and Ji, 2009a; Qin et al., 2010; Li et al., 2012a; Li and Zhou 2012) and only a few ones on argument extraction (e.g., Chen and Ji, 2009b; Hong et al., 2011).

In this paper, we focus on argument extraction in Chinese event extraction. Following previous studies, we divide it into two components, argument identification and role determination, where the former recognizes the arguments of a specific event mention and the latter classifies these arguments by roles. Most of previous studies on argument extraction recast it as a Semantic Role Labeling (SRL) task and focused 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 discourse-driven language with the wide spread of ellipsis. For example, Kim (2000) compared the use of overt subjects in English and Chinese and found that overt subjects occupy over 96% in English, while this percentage drops to only 64% in Chinese. Take following discourse as an example:

S1: *10日凌晨 (A1:Time)* ，*恐怖分子 (A2:Attacker)* 在*拿撒勒镇 (A3:Place)* 制造了一起汽车*炸弹 (A4:Instrument)* **爆炸 (E1)**事件，……有 *1名以色列籍阿拉伯人(A5:Victim)* **死亡(E2)**，……有 *3名路人(A6:Victim)***受伤(E3)**。

(At *daybreak on the 9th (A1)*, the *terrorists (A2)* set off a truck *bomb (A3)* **attack (E1)** in *Nazareth (A4)*, …… An *Israeli Arabian (A5)* was **killed (E2)**, …… *Three pedestrians (A6)* were **injured (E3)**.)

Table 1 shows both the intra-sentence and inter-sentence arguments of event mentions E1~E3, and we can find out that more than half of the arguments would be missing if only intra-sentence information is considered. Obviously, it is hard to capture those long distance inter-sentence information in a syntactic way. An alternative way is to employ the relationship between event mentions in a discourse. For example, since the event mention E1 is the cause of E2 and E3, it is natural to infer E1's arguments (e.g., *Victim* A5 and A6) from E2 and E3, or infer E2 and E3's arguments (e.g., *Agent* A2, *Time* A1) from E1.

Event Mention	Trigger	Intra-sentence Arguments	Inter-sentence Arguments
E1	爆炸	A1, A2, A3, A4	A5, A6
E2	死亡	A5	A1, A2, A3, A4
E3	受伤	A6	A1, A2, A3, A4

Table 1 Intra-sentence and inter-sentence arguments in event mentions E1, E2 and E3

Above observation is consistent with Li et al. (2012a), who noticed the significant difficulty of argument identification over role determination in argument extraction. From this regard, this paper proposes a novel discourse-level joint model of argument identification and role determination so that argument identification can benefit from role determination to improve the performance of argument

* Corresponding author
[1] The mentions that are involved in an event, such as participants and attributes.

extraction as a whole. In particular, various kinds of discourse-based constraints and corpus-based constraints, either automatically learned or linguistically motivated, are explored into the joint model.

The rest of this paper is organized as follows. Section 2 is the related work. Section 3 describes a state-of-the-art Chinese argument extraction system as the baseline. Section 4 proposes a sentence-level joint model with only intra-event constraints. Sections 5 further explores a discourse-level joint model with various kinds of corpus-based and discourse-based constraints. Section 6 reports experimental results and analysis. We conclude our work in Section 7.

2 Related Work

In the literature, almost all the existing studies on argument extraction are concerned with English. While some apply pattern-based approaches (e.g., Patwardhan and Riloff, 2007; Chambers and Jurafsky, 2011), the others use machine learning-based approaches (e.g., 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 discourse-level information (e.g., Hong et al., 2011; Huang and Riloff, 2011). For argument extraction in Chinese event extraction, we only know two studies, Chen and Ji (2009b) and Li et al. (2012a), which follow the English tradition.

While a pipeline model may suffer from the errors propagated from upstream tasks, a joint model can benefit from the close interaction between two or more tasks. Recently, joint modeling has been widely attempted in various NLP tasks, such as joint named entity recognition and syntactic parsing (Finkel and Manning, 2009), joint syntactic parsing and semantic role labeling (Li et al., 2010), joint anaphoricity and coreference determination (Denis and Baldridge, 2007; Song et al., 2012), joint entity and event coreference resolution (Lee et al., 2012).

As for event extraction, only a few studies are concerned with joint modeling, mostly in the bio-molecular domain and in the English language. Riedel et al. (2009) used Markov Logic as a general framework for joint modeling of the bio-molecular event structure for a given sentence. Poon and Vanderwende (2010) also adopted Markov Logic to jointly predict bio-molecular events and their arguments. Riedel and McCallum (2011) presented a joint model of extracting triggers and their arguments, which explicitly captures the interaction of various arguments in the same bio-molecular event. Li et al. (2012b) introduce a joint model of trigger identification and type determination in Chinese event extraction with effective constraints.

Inspired by Li et al. (2012b) on joint modeling of trigger identification and type determination, this paper proposes a discourse-level joint model of identifying arguments and determining their roles for Chinese event extraction with the help of various kinds of effective discourse-level information. As far as we know, there are no successful models for jointly solving argument identification and role determination at discourse level. Besides, our work is different from those joint models on the bio-molecular domain due to different task definitions and different issues addressed.

3 Baseline

In the task of event extraction, an event is defined as a specific occurrence involving participants and attributes. Normally, an event mention is triggered via a word in a phrase or sentence which clearly expresses the occurrence of a specific event. Therefore, extracting an event consists of four basic steps, identifying an event trigger, determining its event type, identifying involved arguments and determining their roles.

As the baseline, we choose a state-of-the-art Chinese event extraction system, as described in Li et al. (2012b), which consists of four typical components (trigger identification, event type determination, argument identification and role determination) where the later two components, argument identification and role determination, are processed in a pipeline way. Besides, the Maximum-Entropy (ME) model is employed to train individual component classifiers for above four components.

To provide a stronger baseline, we introduce more refined features in argument identification and role determination, besides those adopted in Li et al. (2012b). Following is a list of features adopted in our baseline.

1) Basic features: trigger, POS of the trigger, event type, head word of the entity, entity type, entity subtype;
2) Neighbouring features: left neighbouring word of the entity + its POS, right neighbour word of the entity + its POS, left neighbour word of the trigger + its POS, right neighbour word of the trigger + its POS;
3) Dependency features: dependency path from the entity to the trigger, depth of the dependency path;
4) Syntactic features: path from the trigger to the entity, difference of the depths of the trigger and entity, place of the entity (before trigger or after trigger), depth of the path from the trigger to the entity, siblings of the entity;
5) Semantic features: semantic role of the entity tagged by an SRL tool (e.g., *ARG0*, *ARG1*, etc.) (Li et al., 2010), sememe of trigger in Hownet (Dong and Dong, 2006).

4 Sentence-level Joint Model

In our sentence-level joint model, an ILP (Integer Logic Programming) - based inference framework is introduced to integrate two ME-based classifiers, the argument identifier and the role determiner. 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) in combining multiple classifiers, where the traditional pipeline architecture is not appropriate. Currently, most joint models on event extraction and other information extraction tasks (e.g., SRL, entity relation extraction) are sentence-based (e.g., Roth and Yih, 2004; Poon and Vanderwende, 2010; Srikumar and Roth, 2012). In this section, we propose a sentence-level joint model of argument identification (AI) and role determination (RD) with some intra-event constraints.

We assume $p_{AI}(em_i \mid a_{<i,j>})$ the probability of AI identifying entity mention $a_{<i,j>}$ as an argument of event mention em_i

with a specific event type, where $a_{<i,j>}$ is the *jth* entity mention in the sentence containing the *ith* event mention em_i, and $p_{RD}(r_k \mid a_{<i,j>})$) is the probability of RD assigning role r_k to entity mention $a_{<i,j>}$. Like Roth and Yih (2004), we define following assignment costs with $-log$:

$$c^{AI}_{<i,j>} = -\log(p_{AI}(em_i \mid a_{<i,j>})) \tag{1}$$

$$\overline{c}^{AI}_{<i,j>} = -\log(1 - p_{AI}(em_i \mid a_{<i,j>})) \tag{2}$$

$$c^{RD}_{<i,j,k>} = -\log(p_{RD}(r_k \mid a_{<i,j>})) \tag{3}$$

$$\overline{c}^{RD}_{<i,j,k>} = -\log(1 - p_{RD}(r_k \mid a_{<i,j>})) \tag{4}$$

where $c^{AI}_{<i,j>}$ and $\overline{c}^{AI}_{<i,j>}$ are the cost of entity mention $a_{<i,j>}$ as or not as an argument of event mention em_i respectively; $c^{RD}_{<i,j,k>}$ and $\overline{c}^{RD}_{<i,j,k>}$ are the cost of assigning or not assigning role r_k to entity mention $a_{<i,j>}$ respectively.

Besides, we use indicator variable $x_{<i,j>}$ which is set to 1 if entity mention $a_{<i,j>}$ is an argument of event mention em_i, and 0 otherwise. Similar to $x_{<i,j>}$, we use another indicator variable $y_{<i,j,k>}$ which is set to 1 if entity mention $a_{<i,j>}$ is an argument with role r_k, and 0 otherwise. Finally, the objective function for event mention em_i can be represented as follows, where S_i is the set of the entities in the sentence containing event mention em_i and R is the set of roles for a specific event.

$$\begin{aligned} Min\ \ \lambda & \sum_{a_{<i,j>} \in S_i} (c^{AI}_{<i,j>} x_{<i,j>} + \overline{c}^{AI}_{<i,j>} (1 - x_{<i,j>}) \\ + (1-\lambda) & \sum_{A_{<i,j>} \in S_i} \sum_{r_k \in R} c^{RD}_{<i,j,k>} y_{<i,j,k>} + \overline{c}^{RD}_{<i,j,k>} (1 - y_{<i,j,k>}) \end{aligned} \tag{5}$$

Subject to

$$x_{<i,j>} \in \{0,1\} \qquad \forall a_{<i,j>} \in S \tag{6}$$

$$y_{<i,j,k>} \in \{0,1\} \qquad \forall a_{<i,j>} \in S \ \wedge r_k \in R \tag{7}$$

We use single parameter λ to balance the overall contribution of two components AI and RD. Constraints (6) and (7) are used to make sure that $x_{<i,j>}$ and $y_{<i,j,k>}$ are binary values. Besides, in order to reduce the complexity, we assume the event type of event mention em_i is pre-determined from the previous step – event type determination. Therefore, we just compute the costs of an entity acted as a role in a pre-determined event.

Furthermore, we enforce following constraints on the consistency between AI and RD:

(C1) Argument role type constraints: if entity mention $a_{<i,j>}$ has role r_k, it must be an argument.

$$x_{<i,j>} \geq y_{<i,j,k>} \qquad \forall a_{<i,j>} \in S \wedge r_k \in R \tag{8}$$

(C2) True argument constraints: if entity mention $a_{<i,j>}$ is an argument, it must have a role.

$$x_{<i,j>} = \sum_{r_k \in R} y_{<i,j,k>} \qquad \forall a_{<i,j>} \in S \tag{9}$$

5 Discourse-level Joint Model

Previous learning-based classifiers and joint models enable us to make reliable extraction of correct arguments of a specific event in a single sentence or clause. However, as we mentioned in Section 1, arguments of an event mention may appear in different clauses or sentences, especially in Chinese due to its discourse-driven nature. This will lead to a considerable decrease in recall. Besides, considering more entities will definitely introduce many more pseudo arguments and harm the precision.

In this paper, we propose a discourse-level joint model to reliably infer those inter-sentence arguments. In particular, our discourse-level joint model is based on our sentence-level one and the objective function for the discourse D can be re-written as follows, enforcing the same intra-event constraints (6) - (9) as mentioned in Section 4.

$$\begin{aligned} Min\ \ \lambda & \sum_{em_i \in D} \sum_{a_{<i,j>} \in S_i} (c^{AI}_{<i,j>} x_{<i,j>} + \overline{c}^{AI}_{<i,j>} (1 - x_{<i,j>}) + (1-\lambda) \\ & \sum_{em_i \in D} \sum_{a_{<i,j>} \in S_i} \sum_{r_k \in R} c^{RD}_{<i,j,k>} y_{<i,j,k>} + \overline{c}^{RD}_{<i,j,k>} (1 - y_{<i,j,k>}) \end{aligned} \tag{10}$$

where all notations are defined as same as that in Section 4.

Besides, the discourse-level joint model is enhanced with the help of various discourse-based constraints derived from the relationship among relevant event mentions. Moreover, instead of hand-written constraints reflecting one's linguistic knowledge, we incorporate a series of probabilistic constraints estimated from the training data (so called corpus-based constraints) to constrain probabilistic inference. In this way, flexibility and adaptability are achieved.

Finally, since the discourse-level joint model attempts to infer additional arguments from other sentences in the same discourse and 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 joint inference.

5.1 Corpus-based Constraints

This paper proposes some corpus-based constraints represented as a template and learned from the training data. That is, a constraint will be added to the joint model only when its associated template occurs in the training data with a high probability. Here, three types of corpus-based constraints are introduced into our joint model as follows.

The first one is a constraint on the event mention:

(C3) Minimal argument number constraints: an event mention must have at least one participant if it has candidate entities. Participants play a key role in an event and cannot be empty, like *ARG0* and *ARG1* in SRL. Assume *KA* the set of key roles, such as *Agent*, *Org, Person,* and for a specific event mention em_i, we have

$$\sum_{a_{<i,j>} \in S_i} \sum_{k \in KA} y_{<i,j,k>} \geq 1 \qquad \exists AC_i \tag{11}$$

where template AC_i is the condition to apply C3 to event mention em_i and consists of two parts, the template body and the result. If the condition is satisfied, we apply the constraint to the model; otherwise, we ignore the constraint. For constraint C3 on event mention em_i, the template is as follows:

T1: *<event type of em_i, whether there are other event mentions with the same type (coreference event) occurring in this discourse (0/1), whether existing candidate entities can act as key roles(0/1)> <whether em_i contains at least one key role (0/1)>*

We first extract all T1's instances from the training set and then calculate the conditional probabilities $P(1|t1_j)$ where $t1_j$ is a sample in T1 and 1 indicates those event mentions in the training set which matched the sample $t1_j$ contains at least one key role. The condition AC_i of a special event mention em_i in the test set is true if there is a matched sample tl_j for event mention em_i and $P(1|t1_j)>\theta$ where θ is a threshold learned from the development set.

The second one is a constraint on the argument roles:

(C4) Maximal argument number constraints: normally, most roles just associate with only one argument. However, a role may contain more than one argument in a special context, such as role *Person* in a *Marry* or *divorce* event. For a special role R_k, we have

$$\begin{aligned} &\sum_{a_{<i,j>}\in S_i} y_{<i,j,k>} \le 1 \qquad \neg\exists AN_{<i,k>} \wedge \forall r_k \in R \\ &\sum_{a_{<i,j>}\in S_i} y_{<i,j,k>} \le n \qquad \exists AN_{<i,k>} \wedge \forall r_k \in R \end{aligned} \tag{12}$$

where $AN_{<i,k>}$ is the condition to apply C4 to role r_k and n is the maximal number of arguments with role r_k in an event mention em_i. Commonly, this number equals to entity number in a *coordinate* structure. The template for this constraint on role r_k in an event mention em_i is as follows:

T2: *<entity types of the role r_k, whether entities with coordinate structure existed in its candidates (0/1)> <whether the role r_k contains more than one argument (0/1)>*

The same learning method is also adopted to construct template T2 by first collecting corresponding samples and then creating condition $AN_{<i,k>}$ for role r_k in event mention em_i, i.e. $AN_{<i,k>}$ is true if there is a matched sample $t2_j$ for role r_k and conditional probability $P(1|t2_j)=1$ where $t2_j$ is a sample in T2 and 1 indicates that role r_k contains more than one argument.

The final one is a constraint on adjacent entities:

(C5) Modification structure constraints: In some syntactic structures which contain more than one entity, a specific entity in this structure may act as an argument with high probability. For two adjacent entity mentions $a_{<i,j>}$ and $a_{<i,j+1>}$ in event mention em_i, we have

$$\begin{aligned} &x_{<i,j>=0} \qquad \exists AS1_{<i,j>} \wedge a_{<i,j>} \in S_i \\ &x_{<i,j+1>=0} \qquad \exists AS2_{<i,j>} \wedge a_{<i,j+1>} \in S_i \\ &x_{<i,j>} + x_{<i,j+1>} \le 1 \qquad \exists AS3_{<i,j>} \wedge a_{<i,j>}, a_{<i,j+1>} \in S_i \end{aligned} \tag{13}$$

where $AS1_{<i,j>}$, $AS2_{<i,j>}$, $AS3_{<i,j>}$ are the conditions to apply one constraint of C5 to entity mentions $a_{<i,j>}$ or $a_{<i,j+1>}$.

Typically, two syntactic structures, *coordination* and *modification*, may contain more than one entity. For example, the structure of two adjacent entities in S2 is *coordination* while that of the other two adjacent entity mentions in S3 is *modification*.

S2: *父亲 (PER) 和我 (PER) (My father (PER) and I (PER))*

S3: *北京市 (GPE) 市长 (PER) (The mayor (PER) of Beijing (GPE))*

Since all entities in the *coordination* structure play the same role in an event mention and this structure will be used to learn the templates on constraint C4 and we only consider *modification* on constraint C5. Commonly, there are two forms to represent the syntactic relationship between two adjacent entity mentions:

1) *NP/NN+other phrase/POS tag+NP/NN*
2) *NP/NN+NP/NN*

Those phrases or POS tags in (1) may be QP, ADJP, NP, DEC, etc. We use following template to construct constraints C5 between two adjacent entity mentions $a_{<i,j>}$ and $a_{<i,j+1>}$:

T3: *<syntactic structure of adjacent entity mentions $a_{<i,j>}$ and $a_{<i,j+1>}$, palce of this adjacent entities pair in the NP (Begin / Middle / End), entity type of $a_{<i,j>}$, entity type of $a_{<i,j+1>}$> <Which one is selected as the argument ($a_{<i,j>}$ / $a_{<i,j+1>}$ / None / Both)>*

For each event mention, we extract all NPs from its syntactic tree and delete all NPs included in other NP firstly. Then for each NP, we mine all pairs of adjacent entities and construct the samples following template T3. Finally, for each pair of adjacent entities $a_{<i,j>}$ and $a_{<i,j+1>}$ in the test set which match a sample $t3_l$, we calculate conditional probabilities $P(a_{<i,j>}|t3_l)$ and $P(a_{<i,j+1>}|t3_l)$ to represent the probabilities of $a_{<i,j>}$ or $a_{<i,j+1>}$ being selected as argument respectively. The condition $AS1_{<i,j>}$ or $AS2_{<i,j>}$ will be set to true respectively if $P(a_{<i,j>}|t3_l)< \phi$ or $P(a_{<i,j+1>}| t3_l)< \phi$. Besides, condition $AS3_{<i,j>}$ will be set to true if $|P(a_{<i,j>}|t3_l)-P(a_{<i,j+1>}| t3_l)| \ge \delta$. Here, ϕ and δ are the thresholds learned from the development set.

5.2 Discourse-based Constraints

In discourse level, all event mentions are normally related with each other semantically, temporally, causally or conditionally. As we mentioned in Section 1, some arguments of an event mention can be inferred from its related event mentions in the same discourse since they describe different aspects of relevant topics.

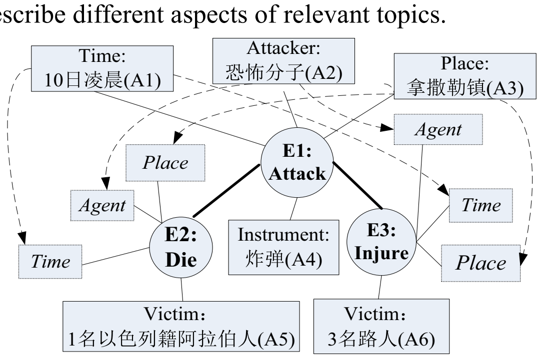

Figure 1 Sample S1 and its graphical representation

The relationship between arguments in relevant event mentions can be represented as an undirected graph and there are two types of nodes, trigger and argument, while the former is indicated by circle and the latter is indicated by rectangle. Figure 1 illustrates sample *S1* where the solid lines represent the relationship between relevant event mentions and their arguments and the dash lines represent the relationship inferred from relevant events. While those arguments in solid rectangles can be determined using the sentence-level joint model, other arguments in dash rectangles only can be inferred with the help of the discourse-level information from relevant events.

Same as corpus-based constraints, we also apply the form of template to construct discourse-based constraints. For two event mentions em_i and $em_{i'}$ in a discourse, the template has the following format:

T4: *<event type of em_i, event type of $em_{i'}$, role r_k in em_i, role $r_{k'}$ in $em_{i'}$> <whether role $r_{k'}$ in $em_{i'}$ can be filled by the argument with role r_k in em_i (0/1)>*

(C6) Event relation constraints: Similar to T1 and T2, we have following discourse-based constraint to enforce the consistency between related arguments in different event mentions which are related semantically, temporally, causally or conditionally.

$$y_{<i,j,k>} = y_{<i',j,k'>} \quad \exists AR_{<i,i',k,k'>} \wedge \forall em_i, em_{i'} \in D \wedge a_{<i,j>} \in S_i \cap S_{i'} \wedge r_k, r_{k'} \in R \tag{14}$$

where $AR_{<i,i',k,k'>}$ is true if probability $P(1|t4_{<i,i',k,k'>})>\partial$ and $t4_{<i,i',k,k'>}$ refers to template *<event type of em_i, event type of $em_{i'}$, role r_k in em_i, role $r_{k'}$ in $em_{i'}$>* and ∂ is a threshold learned from the development set.

Through imposing the constraints on the event relationship in a discourse, many more arguments may be inferred from related event mentions. Since it is reasonable to assume that an event mention should have at least one key role (not including *Time* and *Place*), we propose a simple rule to filter out those pseudo mentions and (as a result) corresponding pseudo arguments on the consistency between a trigger and its arguments.

(C7) Event mention constraints: if an event mention doesn't have an argument, besides *Time* and *Place*, this mention will be regarded as pseudo one and all its arguments will be deleted.

6 Experimentation and Analysis

In this section, we first describe the experimental setting and then evaluate our joint model of argument identification and role determination in Chinese event extraction.

6.1 Experimental Setting

For fair comparison, we adopt the same experimental setting as the state-of-the-art event extraction system (Li et al. 2012b) and all the experiments are done on the ACE 2005 Chinese corpus. Since only the training data of the ACE 2005 Chinese corpus are available, 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 follow the setting of ACE diagnostic tasks 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*)[2] with all entities annotated in the corpus kept. We use *Berkeley Parser*[3] and *Stanford Parser*[4] to create the constituent and dependency parse trees. Besides, the ME tool (*Maxent*)[5] is employed to train individual component classifiers and *lp_solver*[6] is used to construct the joint model.

6.2 Experimental Results

For fair comparison, all the experiments on argument extraction are done on the output of the trigger extraction system as described in the state-of-the-art event extraction system (Li et al. 2012b) and Table 2 shows its performance.

Trigger Identification			Event Type Determination		
P(%)	R(%)	F1	P(%)	R(%)	F1
74.4	71.9	73.1	71.4	68.9	70.2

Table 2 Performance of baseline trigger extraction on trigger identification and event type determination

Table 3 compares the performance of Li et al. (2012b), our baseline, our sentence-level joint model (S-JM) and our discourse-level joint model (D-JM). Here, the parameters λ, θ , δ, ϕ, ∂ are fine-tuned to 0.65, 0.10, 0.10, 0.50 and 0.80 respectively, using the development set. Besides, 18642 constrains have learned from the training set in D_JM while S_JM only contains 1693 constraints.

Performance / 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
S-JM	59.3	60.1	59.7	54.4	55.2	54.8
D-JM	60.9	65.6	63.2	55.4	59.6	57.4

Table 3 Performance comparison of argument extraction on argument identification and role determination

Table 3 shows that:

1) Our baseline slightly improves the performance over Li et al. (2012b) due to the incorporation of more refined features. This indicates the limitation of feature engineering with intra-event information.
2) Our sentence-level joint model 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 sentence-level constraints in the joint model is not effective to infer more arguments.
3) Compared to the sentence-level joint model, our discourse-level joint model improves the performance of argument identification and role determination by 3.5%

[2]http://ictclas.org/
[3] http://code.google.com/p/berkeleyparser/
[4] http://nlp.stanford.edu/software/lex-parser.shtml
[5] http://mallet.cs.umass.edu/
[6] http://lpsolve.sourceforge.net/5.5/

and 2.6% improvement in F1-measure respectively. This suggests the effectiveness of those corpus-based and discourse-based constraints and the ability of our discourse-level joint model to infer inter-sentence arguments. Table 4 shows the contributions of different corpus-based and discourse-based constraints while the discourse-based constraint C6 gains the highest improvement of argument identification and role determination in F1-measure respectively. Besides, it's obviously that these constraints interact with others. However, our experimental results show that such interaction is very limited since they address different issues.

Contribution / Constraint	Argument Identification			Argument Role Determination		
	P(%)	R(%)	F	P(%)	R(%)	F
S-JM	59.3	60.1	59.7	54.4	55.2	54.8
+C3	-0.7	+1.3	+0.3	-0.8	+1.3	+0.2
+C4	+1.0	-0.1	+0.5	+0.9	-0.1	+0.4
+C5	+1.6	-0.2	+0.7	+1.2	-0.2	+0.5
+C6	-1.5	+4.9	+1.5	-1.3	+3.8	+1.2
+C7	+1.2	-0.4	+0.5	+1.0	-0.4	+0.3

Table 4 Contributions of different corpus-based and discourse-based constraints on argument identification and role determination (incremental)

4) Table 2 shows 25.6% of pseudo trigger mentions are introduced into argument extraction and that is the main cause of the low performance of argument extraction. If we use the golden trigger extraction, our exploration shows that the precision and recall of argument identification can be up to 76.8% and 88.1% respectively. Table 5 shows the performance comparison of argument extraction on AI and RD given golden trigger extraction. Compared to the Baseline and S-JM, our D-JM improves the performance of argument identification by 5.3% and 4.5% improvement in F1-measure respectively, largely due to the dramatic increase in recall of 10.7% and 9.4%.

Performance / System	Argument Identification			Argument Role Determination		
	P(%)	R(%)	F	P(%)	R(%)	F
Baseline	76.2	77.4	76.8	70.4	72.0	71.2
S-JM	76.6	78.7	77.6	70.5	73.1	71.8
D-JM	76.8	88.1	82.1	70.8	82.1	76.0

Table 5 Performance comparison of argument extraction on argument identification and type determination (Golden trigger extraction)

6.3 Analysis

The experimental results show that the number of these corpus-based constraints are almost 11 times of that of S-JM, which can be learned from the training set, as long as the proper templates are used. These corpus-based constraints are effective in constraining probabilistic search, i.e., they may be directly applied for determining which entity can act as an argument in an NP, or enforcing the consistency on the arguments between relevant event mentions. This suggests that it is interesting to explore data-driven constraints learned from the training data.

Besides, the discourse-based constraint improves the performance significantly. Table 4 shows it enhances the F1-measure of argument identification and role determination by 1.5% and 1.2% respectively, largely due to a gain of 4.9% and 3.8% in recall. Unfortunately, it harms the precision due to mis-identifying pseudo arguments as true ones. Moreover, our exploration shows that our joint model can mine those arguments within a long distance which are un-annotated as arguments of a special event mention in the corpus. Actually, they are the true ones to our knowledge and are more than 35% of those pseudo arguments inferred by our discourse-based constraints. This ensures that our discourse-level joint model and discourse-based constraint is helpful to argument extraction.

We also propose a simple rule to impose the consistency between a trigger and its arguments. Our experimentation shows that it filters out almost 4.3% of pseudo arguments and as a result improves the precision by 1.2%. This filter rule also contributes to trigger extraction and helps remove 12.4% of pseudo trigger mentions and improves the precision from 74.4% to 77.2%. It suggests the close interaction of four components in event extraction and the potential of further joint modeling.

The contributions of this paper are:

1) In machine learning aspect, we propose a novel discourse-level joint model of argument identification and role determination, which can infer more arguments with the help of various kinds of sentence and discourse level information. Besides, we introduce some corpus-based constraints (learned from the training set) to enforce the consistency.
2) In linguistics aspect, we propose constraints based on the syntactic structure and the semantic relationship between event mentions in the discourse. The experimental results encourage the research towards such direction.

7 Conclusion

This paper studies discourse-level joint modeling of argument identification and role determination. In particular, various kinds of discourse-based constraints and corpus-based constraints, either automatically learned or linguistically motivated, are introduced to improve the performance. Evaluation on the ACE 2005 Chinese corpus shows the effectiveness of such joint model and our proposed constraints.

For the further work, we will explore better joint modeling and more effective discourse-based constraints in 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. 61070123, No. 61272260 and No. 61273320, the National 863 Project of China under Grant No. 2012AA011102.

References

[Barzilay and Lapata, 2006] Regina Barzilay and Miralla Lapata. Aggregation via set partitioning for natural language generation. In *Proceedings of NAACL 2006*, pages 359-366, 2006.

[Chambers and Jurafsky, 2011] Nathanael Chambers and Dan Jurafsky. Template-based information extraction without the templates. In *Proceedings of ACL 2011*, pages 976-986, 2011.

[Chen and Ji, 2009a] Zheng Chen and Heng Ji. Can one language bootstrap the other: a case study on event extraction. In *Proceedings of NAACL-HLT 2009 Workshop on Semi-supervised Learning for Natural Language Processing*, pages 66-74, 2009.

[Chen and Ji, 2009b] Zheng Chen and Heng Ji. 2009b. Language specific issue and feature exploration in Chinese event extraction. In *Proceedings of NAACL-HLT 2009*, pages 209-212, 2009.

[Denis and Baldridge, 2007] Pascal Denis and Jason Baldridge. Joint determination of anaphoricity and coreference resolution using integer programming. In *Proceedings of NAACL-HLT 2007,* pages 236–243, 2007.

[Dong and Dong, 2006] Zhengdong Dong and Qiang Dong. *HowNet and the computation of meaning*. World Scientific Pub Co. Inc., 2006.

[Do et al., 2012] Quang Xuan Do, Wei Lu and Dan Roth. Joint inference for event timeline construction. In *Proceedings of EMNLP 2012*, pages 677-687, 2012.

[Finkel and Manning, 2009] Jenny Rose Finkel and Christopher Manning. Nested named entity recognition. In *Proceedings of EMNLP 2009*, pages 141-150, 2009.

[Hong et al., 2011] Yu Hong, Jianfeng Zhang, Bin Ma, Jianmin Yao, Guodong Zhou and Qiaoming Zhu. Using cross-entity inference to improve event extraction. In *Proceedings of ACL 2011*, pages 1127-1136, 2011.

[Huang and Riloff, 2011] Ruihong Huang and Ellen Riloff. Peeling back the layers: detecting event role fillers in secondary contexts, In *Proceedings of ACL 2011*, pages 1137-1147, 2011.

[Kim 2000] Young-Joo Kim. Subject/object drop in the acquisition of Korean: A cross-linguistic comparison. *Journal of East Asian Linguistics*, 9(4): 325-351, 2000.

[Lee et al., 2012] Heeyoung Lee, Marta Recasens, Angel Chang, Mihai Surdeanu and Dan Jurafsky. Joint entity and event coreference resolution across documents. In *Proceedings of EMNLP 2012*, pages 489-500, 2012.

[Li et al., 2010] Junhui Li, Guodong Zhou and Hwee Tou Ng. Joint syntactic and semantic parsing of Chinese. In *Proceedings of ACL 2010*, pages 1108-1117, 2010.

[Li et al., 2012a] Peifeng Li, Guodong Zhou, Qiaoming Zhu and Libin Hou. Employing compositional semantics and discourse consistency in Chinese event extraction. In *Proceedings of EMNLP 2012*, pages 1006-1016, 2012.

[Li et al., 2012b] Peifeng Li, Qiaoming Zhu, Hongjun Diao and Guodong Zhou. Joint modeling of trigger identification and event type determination in Chinese event extraction. In *Proceedings of COLING 2012,* pages 1635-1653, 2012.

[Li and Zhou, 2012] Peifeng Li and Guodong Zhou. Employing morphological structures and sememes for Chinese event extraction. In *Proceedings of COLING 2012*, pages 1619-1634, 2012.

[Lu and Roth, 2012]Wei Lu and Dan Roth. Automatic event extraction with structured preference modeling. In *Proceedings of ACL 2012*, pages 835-844, 2012.

[Patwardhan and Riloff, 2007] Siddharth Patwardhan and Ellen Riloff. Effective information extraction with semantic affinity patterns and relevant regions. In *Proceedings of EMNLP-CoNLL 2007*, pages 717-727, 2007.

[Patwardhan and Riloff, 2009] Siddharth Patwardhan and Ellen Riloff. A unified model of phrasal and sentential evidence for information extraction. In *Proceedings of EMNLP 2009*, pages 151-160, 2009.

[Poon and Vanderwende, 2010] Hoifung Poon and Lucy Vanderwende. Joint inference for knowledge extraction from biomedical literature. In *Proceedings of NAACL-HLT 2010*, pages 813-821, 2010.

[Qin et al., 2011] Bing Qin, Yanyan Zhao, Xiao Ding, Ting Liu and Guofu Zhai. Event type recognition based on trigger expansion. *Tsinghua Science and Technology*. 15(3): 251-258, 2011.

[Riedel et al., 2009] Sebastian Riedel, Hong-Woo Chun, Toshihisa Takagi and Junichi Tsujii. A Markov logic approach to bio-molecular event extraction. In *Proceedings of BioNLP 2009*, pages 41-49, 2009.

[Riedel and McCallum, 2011] Sebastian Riedel and Andrew McCallum. Fast and robust joint models for biomedical event extraction. In *Proceedings of EMNLP 2011*, pages 1-12, 2011.

[Roth and Yih, 2004] Dan Roth and Wen-tau Yih. A linear programming formulation for global inference in natural language tasks. In *Proceedings of CoNLL 2004*, pages 1-8, 2004.

[Srikumar and Roth, 2012] Vivek Srikumar and Dan Roth. A joint model for extended semantic role labeling. In *Proceedings of EMNLP 2011*, pages 129-139, 2011.

[Song et al., 2012] Yang Song, Jing Jiang,Wayne Xin Zhao, Sujian Li and Houfeng Wang. Joint learning for coreference resolution with Markov Logic. In *Proceedings of EMNLP 2012*, pages 1245-1254, 2012.