Bilingual Semantic Role Labeling Inference via Dual Decomposition

HAITONG YANG, YU ZHOU, and CHENGQING ZONG, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences

This article focuses on bilingual Semantic Role Labeling (SRL); its goal is to annotate semantic roles on both sides of the parallel bilingual texts (bi-texts). Since rich bilingual information is encoded, bilingual SRL has been applied in many natural-language processing (NLP) tasks such as machine translation (MT), cross-lingual information retrieval (IR), and the like. A feasible way of performing bilingual SRL is using monolingual SRL systems to perform SRL on each side of bi-texts separately. However, it is difficult to obtain consistent SRL results on both sides of bi-texts in this way. Some works have tried to jointly infer bilingual SRL because there are many complementary language cues on both sides of bi-texts and they reported better performance than monolingual systems. However, there are two limits in the existing methods. First, the existing methods often require high inference costs due to the complex objective function. Second, the existing methods fully adopt the candidates generated by monolingual SRL systems, but many candidates are discarded in the argument pruning or identification stage of monolingual systems. In this article, we propose two strategies to overcome these limits. We utilize a simple but efficient technique: Dual Decomposition to search for consistent results for both sides of bi-texts. On the other hand, we propose a method called Bi-Directional Projection (BDP) to recover arguments discarded in monolingual SRL systems.

We evaluate our method on a standard parallel benchmark: the OntoNotes dataset. The experimental results show that our method yields significant improvements over the state-of-the-art monolingual systems. In addition, our approach is also better and faster than existing methods due to BDP and Dual Decomposition.

1. INTRODUCTION

Semantic Role Labeling (SRL) is a standard task in the natural-language processing (NLP) community. Its goal is to recognize related phrases and assign a joint structure (WHO did WHAT to WHOM, WHEN, WHERE, WHY, HOW) to each predicate of a sentence [Gildea and Jurafsky 2002]. Because rich semantic information is encoded,
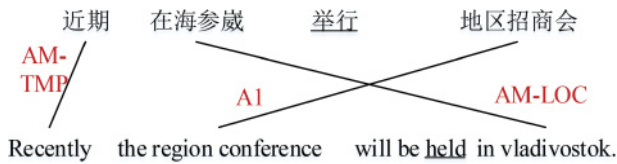
Fig. 1. In this parallel sentence pair, the predicate pair '举行 / hold' has the same argument structure between in Chinese and in English.

SRL has been widely applied in many fields of NLP, such as question and answering [Narayanan and Harabagir 2004], information extraction [Surdeanu et al. 2003], and machine translation [Wu and Fung 2009a; Liu and Gildea 2010; Xiong et al. 2012; Zhai et al. 2012, 2013].

The semantic equivalence of the parallel bi-texts means that they should have the same predicate semantic structure. In bilingual or multilingual text processing, such as MT and cross-lingual information retrieval, many researchers [Wu and Fung 2009b; Liu and Gildea 2010; Xiong et al. 2012; Zhai et al. 2012, 2013] have utilized the consistency of semantic representations of bi-texts to improve the performance of their systems. Semantic structures have two major advantages over syntactic structures. First, semantic structures tend to agree better between two languages than syntactic constituents [Fung et al. 2006]. This property motivates the approach of using the consistency of semantic roles to select MT outputs [Wu and Fung 2009b]. Second, the set of semantic roles of a predicate models the skeleton of a sentence, which is crucial to the readability of MT output [Liu and Gildea 2010].

Motivated by these observations, this article focuses on bilingual SRL. A conventional way to perform bilingual SRL is using monolingual SRL systems to perform SRL on each side of bi-texts separately, as has been done by Fung et al. [2007]. However, it is very difficult to obtain consistent SRL results on both sides of bi-texts in this way. For example, it may occur that '近期' is labeled as AM-LOC but 'Recently' is labeled as AM-TMP in monolingual SRL systems. In this case, we should abandon these obviously inconsistent results and search for better results for both sides. Some works have demonstrated that it is a better way to jointly infer bilingual SRL because both sides of bi-texts contain complementary language cues that can be utilized to correct the inconsistent cases. In fact, this observation has formed the basis for many cross-lingual tasks in NLP, such as syntax projection across languages [Yarowsky et al. 2001; Hwa et al. 2005; Ganchev et al. 2009], cross-lingual named entity recognition [Huang and Vogel 2002; Moore 2003] and IR [Si and Callan 2005]. In these cases, multilingual models yield increased accuracy because different languages present different ambiguities and therefore offer complementary constraints on the shared underlying labels.

Zhuang and Zong [2010a] first explored the problem of inferring bilingual SRL jointly. In their work, they constructed an Integer Linear Programming (ILP) framework to infer bilingual SRL in which some constraints are designed to enhance the consistency of bilingual SRL. They reported better performance over monolingual systems. However, there are two limits in their work. One is that, in the inference, they still adopted the argument candidates generated by monolingual systems, but in fact many arguments are discarded in the argument pruning or identification stage of monolingual systems. Instead, we believe that the complementary language cues in bi-texts could guide us to recover these missing arguments. The other limit is that the objective function of their bilingual model is very complex[1], but the inference technique of ILP is costly in

---

[1]There are many variables (more than 100) in the objective function, and relations between variables are constrained by many conditions.
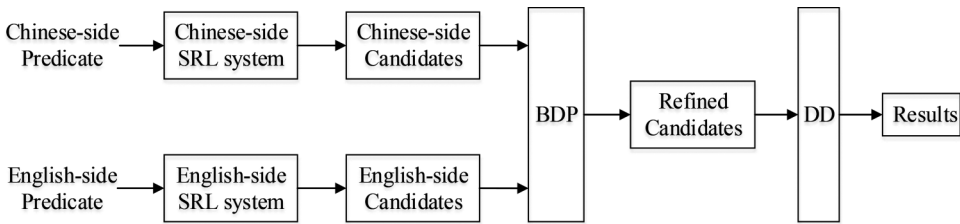
Fig. 2. The overview of our approach.

handling complex optimization problems. Moreover, if the data to be processed are very large, ILP often fails to meet the requirements of practical applications.

To overcome these limits, in this article, we propose two strategies for bilingual SRL. Specifically, to recover the missing arguments, we developed a greedy algorithm called Bi-Directional Projection (BDP) to finely enlarge the set of candidates. In BDP, we first use a monolingual system to generate the initial candidates for both sides; then, with the help of word alignment information, we greedily search for the candidate with greatest match in the other side for each candidate of one side. If the resulting candidate is not in the initial candidates set, it will be added. It is noted that some noisy candidates may be added incorrectly due to incorrect word alignment. Thus, we also design some simple, but effective, rules to eliminate these noisy candidates. In the inferring phase, in contrast to Zhuang and Zong [2010a], we employed a technique called Dual Decomposition (DD) to efficiently perform the inference. In recent years, DD has been widely accepted as a simple and efficient technique to resolve combination optimization problems, such as the combination of two lexicalized parsing models [Rush et al. 2010], syntax parsing [Koo et al. 2010; Wang et al. 2013b], and bilingual named entity recognition [Wang et al. 2013a]. In this work, we prove that the technique of DD could also efficiently infer bilingual SRL.

An overview of our approach to bilingual SRL is shown in Figure 2. Our approach works as follows: first, we use monolingual systems to generate the initial candidates; second, based on the initial candidates, the proposed BDP greedily search on both sides and obtain refined argument candidates; third, the technique of Lagrange Dual Decomposition is implemented to search for the consistent SRL results for both sides of bi-texts.

We have conducted experiments on a standard parallel benchmark—the OntoNotes dataset. Experimental results show that our method yields significant improvements over the state-of-the-art monolingual systems. Benefiting from reliable candidates generated by the BDP algorithm, our method achieved better performance than Zhuang and Zong [2010a], and our method is also much faster than theirs.

The rest of this article is organized as follows: Section 2 describes how we generate argument candidates on both sides of bi-texts using monolingual SRL systems. Section 3 presents the algorithm of BDP. Section 4 presents our joint inference model. Section 5 presents our experiments and results. Section 6 introduces related work. Our conclusions can be found in Section 7.

2. MONOLINGUAL SRL SYSTEMS

In this section, we introduce the monolingual SRL system, which is a baseline in the experiments and also generates initial argument candidates for BDP and the inference step.

An SRL system usually takes a parse tree (either handcrafted parse trees or auto parse trees generated by an automatic parser) as the input and assigns appropriate

semantic roles to the constituents in the parse tree, which are semantic arguments to the predicate in question. Most SRL systems work in stages, which consist of a pruning stage, an argument-identification stage, and an argument-classification stage. In the pruning stage, the goal is to eliminate the obvious nonargument candidates. The heuristic pruning method in Xue [2008] has been widely accepted by the SRL community; thus, we implemented his method in this article. The predominant approach to argument identification and argument classification is to formulate them as classification problems that can be solved with machine-learning techniques. In the stage of argument identification, all argument candidates are identified as an argument or a nonargument that is not related to the predicate in question; thus, argument identification is generally formulated as a binary classification task. Argument classification, which assigns appropriate semantic roles to candidates identified by the argument identification stage, is a natural multicategory classification problem.

2.1. Classifier

In the literature, many classification techniques, such as SVM [Pradhan et al. 2005], Perceptrons [Carreras et al. 2004], and Maximum Entropy [Xue and Palmer 2004; Xue 2008], have been successfully used to solve the tasks of argument identification and argument classification. In this work, we used a Maximum Entropy classifier with a tunable Gaussian prior. As a discriminative model, the Maximum Entropy classifier is scalable to handle a large set of features. Moreover, the Maximum Entropy classifier does multicategory classification naturally, thus can be straightforwardly applied in the problems here. Zhang Le's MaxEnt toolkit[2] is used for the specific implementation.

2.2. Features

Features are crucial to the performance of SRL systems. Many works [Xue and Palmer 2004; Xue 2008] have studied what features are discriminative for SRL. In Xue and Palmer [2004] and Xue [2008], the author experimentally proved that the stages of argument identification and classification require different features. Therefore, we follow his suggestions and utilize two groups of features: one for argument identification and the other for argument classification. The following lists describe the features in detail.

The features for argument identification:

—Predicate lemma
—Path from node to predicate
—Head word
—Head word's part-of-speech
—Verb class
—Predicate and head-word combination
—Predicate and phrase-type combination
—Verb class and head-word combination
—Verb class and phrase-type combination

All of these features are also used in argument classification. Here are some other features:

—Position: the relative position of the candidate argument compared with the predicate
—Subcat frame: the syntactic rule that expands the parent of the verb
—The first and the last word of the candidate
—Phrase type: the syntactic tag of the candidate argument
—Subcat frame+: the frame that consists of the NPs [Xue 2008].

---

[2]http://homepages.inf.ed.ac.uk/lzhang10/maxent_toolkit.html

Sent	Within three to five years ,		beihai	has constructed		the framework of a modernized city	
Args	[	AM-TMP]	[A0]	[Pred]	[	A1	]
loc		(0, 4)	(6, 6)			(9, 14)	
l		AM-TMP	A0			A1	
p		0.93	0.85			0.80	

Fig. 3. An argument output of monolingual SRL systems: location *loc*, label *l*, and classification probability *p*.

2.3. Output of Monolingual SRL System

One reason why we choose a maximum entropy classifier as our classifier is that it can easily incorporate arbitrary features. The other reason is that the maximum entropy classifier can directly output classification probabilities without any postprocessing that will be used in our joint inference stage. For clear illustration, we wrap the classification probabilities into the output of monolingual SRL systems. As shown in Figure 3, an argument output of monolingual SRL systems has three attributes: its location in the sentence *loc*, represented by the position of its first word and last word; its semantic role label *l*; and its classification probability *p*. Therefore, we annotate an argument output of a monolingual SRL system as a triple (loc, l, p). For example, the $A0$ argument in Figure 3 is $((6, 6), A0, 0.85)$.

2.4. Complementarity of Chinese and English SRL Systems

The semantic equivalence of the parallel bi-texts means that they should have the same predicate semantic structure, which implies that the aligned argument pairs in bi-texts (e.g., the pair of '近期' and 'Recently' in Figure 1) should have consistent labels. Nevertheless, the monolingual SRL systems of Chinese and English work independently. Thus, some inconsistent cases often occur on both sides of monolingual SRL systems. However, there are many complementarity clues that are beneficial to correct these mistakes. For example, in Figure 1, if '近期' is missed or annotated with a wrong label by a Chinese SRL system, but 'Recently' is labeled correctly by an English SRL system, intuitively, with the help of the English side, we can recognize the argument of '近期' on the Chinese side and give it a correct label.

From these observations, we can see that there are mainly two types of complementarity that can be utilized in the bilingual SRL system.

Location Complementarity means that, if one of an argument pair is found by the monolingual SRL system but the other is missed, we can track down the missed argument in the other language according to word alignment information. It is worth noting that there are also some exceptional cases, for example, the argument ellipsis on one side. These cases have little impact on the location complementarity, however, because, in these cases, the aligned points are null in the other language.

Label Complementarity means that an aligned argument pair in a bi-text should have consistent labels. For instance, in Figure 1, the labels for '近期' and 'Recently' are AM-TMP, which are consistent. However, if the label on one side is wrong but the label on the other side is right, we obtain inconsistent results for both sides. In these cases, we should abandon this result and search for a pair of consistent results. It is worth noting that the consistent labels for the aligned argument pair do not mean that they have to have the same label. For a brief illustration, an example is given in here. In
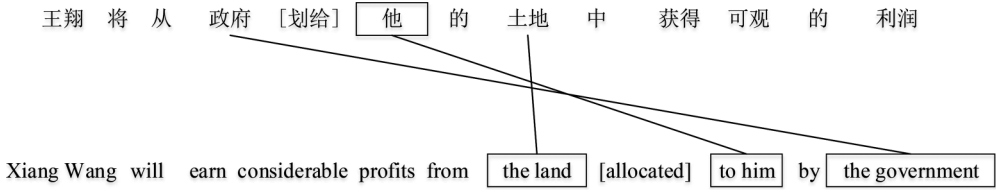
王翔　将　从　政府　[划给]　他　的　土地　中　获得　可观　的　利润

Xiang Wang   will   earn considerable profits from   the land   [allocated]   to him   by   the government

Fig. 4.   An example from a monolingual SRL system.

the following bi-text, the predicate pair is '下降' and 'decreased' and has three pairs of aligned arguments. Among them, the argument pair of '今年 一 至 二月份' and 'up to February of this year' have the same label 'AM-TMP', but the other two pairs have different labels because of the different usages of the predicates on both sides. Thus, it is not reasonable to force the aligned arguments to have the same label, and we need to consider the context of predicates to make a decision for bilingual SRL.

(a) [今年 一 至 二月份]$_{AM-TMP}$, [中国 一般 贸易 出口]$_{A0}$ [明显]$_{AM-ADV}$ [下降]$_{Pred}$.
(b) [up to February of this year]$_{AM-TMP}$, [China's general trade exports]$_{A1}$ have [decreased]$_{Pred}$ [noticeably]$_{A2}$.

Here, we summarized two categories of complementarity that can be beneficial to bilingual SRL. In Sections 3 and 4, we will describe how to utilize the two categories of complementarity to improve bilingual SRL.

3. BI-DIRECTIONAL PROJECTION ALGORITHM

As illustrated in Section 2, an SRL system works in a pipeline way and argument candidates are obtained through the pruning phase and argument-identification phase. However, the pipelined way often causes error propagation because the results of the argument pruning and identification stages are not always correct. For clearer illustration, an example is given in Figure 4. In the given bi-text, the predicate pair in question is '划给/ allocate'. The argument candidates generated by individual monolingual SRL systems are the ones embedded in a box. We note that, on the Chinese side, only '他' is identified as an argument of '划给', while '政府' and '土地' are discarded in the pruning due to parsing errors. In the SRL community, it is widely recognized that the overall performance of an SRL system is largely determined by the quality of syntactic parsers [Gildea and Palmer 2002], which is particularly notable in the pruning stage and identification stage. We note that the text related to '划给' on the Chinese side is a complex '的's phrase — '政府 划给 他 的 土地'. When the text is parsed, the problem of syntactic ambiguity occurs, which causes '政府' and '土地' to be discarded by the Chinese SRL system. However, we note that the translated text for '政府 划给 他 的 土地' is *the land allocated to him by the government*, which is a relatively simple text compared with the complex '的' phrase on the Chinese side and is easily parsed correctly. Therefore, on the English side, we obtain fully correct argument candidates from the English SRL system. In this case, the fully correct candidates of English side can be utilized to guide recovering the discarded candidates on the Chinese side because we can locate to the positions of the missed arguments on the Chinese side according to word alignment. We found that it is not always true that the English side obtains better candidates than the Chinese side and the opposite also occurs. Thus, an algorithm is expected to utilize the *Location Complementary* information of bi-texts to improve the performance of both sides simultaneously.

ALGORITHM 1: Bi-Directional Projection

Set $e\_Array = []; c\_Array = []$;
Generate initial candidates from individual monolingual SRL system, and put results
into e_Array and c_Array, respectively;
Input: e_Array and c_Array
Output: e_Array and c_Array
repeat
   for *each $e_k \in e\_Array$* **do**
      $e'_k = arg\,max_{arg \in c\_Array} sim(arg, e_k)$;
      if $e'_k \notin c\_Array$;
      then
         Put e'_k into c_Array;
      end
   end
   for *each $c_k \in c\_Array$* **do**
      $c'_k = arg\,max_{arg \in e\_Array} sim(arg, c_k)$;
      if $c'_k \notin e\_Array$;
      then
         Put c'_k into e_Array;
      end
   end
until *no new candidate is generated*;
return e_Array and c_Array

3.1. Algorithm Description

Motivated by the earlier observations, in this article, we propose an algorithm called
Bi-Directional Projection (BDP) to recover the argument candidates discarded in the
argument pruning and identification stages. In the algorithm, a crucial step is to find
the aligned candidate with equivalent semantics on the other side for a candidate on
one side. For example, in Figure 4, given 'the land' and 'the government' on the English
side, we need to find '土地' and '政府' on the Chinese side. Because word alignment
of bi-texts conveys semantic equivalence at the word level and provides the requisite
information for our goal, word alignment is our natural choice. If the word alignment
of the bi-text is known, we can find the missing candidates '土地' and '政府' of Figure 4
with little effort.

The pseudo-code of BDP is shown in Algorithm 1, in which we first generate the
initial candidates using monolingual SRL systems. Then, for each candidate on the
English side, we search all constituents of the parse tree on the Chinese side. If
the most similar candidate to the current candidate is not contained in the candidate
set on the Chinese side, this candidate is added into the candidate set on the Chinese
side. We call this process a projection. After the projection from English to Chinese is
done, we forward a projection from Chinese to English. Because in a projection there
may be new candidates generated, we iterated the BDP process until there were no
more new candidates produced in both directions' projections.

In BDP, a crucial step is to find the aligned candidate with semantic equivalence
on the other side for each candidate on one side. Therefore, we need a similarity
function to evaluate the similarity of a pair of candidates. Here, we adopt a simple and
effective function. Specifically, we approximate the similarity of a pair of candidates by
measuring the *word alignment-based word overlap* [Padó and Lapata 2009]. Assume
that we would like to evaluate the similarity of a Chinese candidate c and an English
candidate e. Let C and E denote the set of words in the yield of c and e, respectively,
and TC and TE the set of aligned words for c and e in the other language, respectively.

Then, the similarity of c and e is defined as:

$$sim(c, e) = \frac{|C \cap TE| + |E \cap TE|}{|C \cup E|}. \tag{1}$$

3.2. Candidates Filtration

In the experiments, we use the publicly available GIZA++ [Och and Ney 2003] to automatically induce the word alignment for bi-texts. However, the results generated by GIZA++ may contain some incorrect alignments, which can cause cascading errors in BDP. Therefore, we have to take measures to filter these noisy candidates. Here, we make the noise elimination using some simple, but effective, rules:

(1) *Length ratio*. The aligned candidates are translated text, and their length ratio should not be too large. If the length ratio is larger than a given threshold (μ), the newly added candidate will be filtered out.
(2) *Identification probability*. We also use the traditional module of argument identification to filter candidates. If the probability of argument identification is lower than a given threshold (τ), the newly added candidate will be filtered out.
(3) *Arguments cannot overlap with the predicate*. This is a common rule in the argument identification stage [Punyakanok et al. 2004]. If the candidate generated by BDP overlaps with the predicate, it will be filtered out.
(4) *Candidates far from the predicate*. This is a rule designed for the English side's projection. We found that, in many cases, the Chinese side has some arguments far from the predicate, for example, there may be one or several clauses between arguments and the predicate. However, these cases are seldom observed on the English side. In the experiments, we set that if there is one clause between the newborn candidate and the predicate, the candidate generated by BDP will be filtered out.

4. BILINGUAL SRL MODEL

In this section, we describe how our bilingual SRL model works. Our model can be conceptually decomposed into three components: the Chinese side, the English side, and the Argument Alignment. The objective function of our bilingual model can be formulated as the sum of two subobjectives with constraints:

$$\begin{aligned} &\max \ O_c \ + \ O_e \\ &\text{where } O_c \text{ and } O_e \text{ are constrained by } O_a. \end{aligned} \tag{2}$$

In this equation, O_c and O_e represent the quality of the SRL results on the Chinese side and English side, respectively. If there are not any constraints, we can find that the results of Equation (2) are the same as the monolingual systems. Therefore, we use the third module, Argument Alignment O_a, to enhance the consistency of bilingual SRL. It gives preference to the bilingual SRL results that have more consistent argument structures. In the following sections, we will introduce these three components and the inference technique optimizing Equation (2), respectively.

4.1. Components of Chinese and English

4.1.1. Chinese Component. For easy illustration, some notation is given here. We denote the entire semantic label set of the Chinese side as $\{l_1^c, l_2^c, \ldots l_{L_c}^c\}$[3]. Suppose that, on the Chinese side, there are N_c candidates denoted as $cand_1^c, \ldots, cand_{N_c}^c$. The probability of

---

[3]Note that there is not a 'null' label in the classification stage of our system.

assigning l_m^c to $cand_i^c$ is p_{im}^c that is generated by the monolingual system. Meanwhile, an indicator variable x_{im} is defined as:

$$x_{im} = \left[cand_i^c \text{ is assigned label } l_m^c\right]$$

to represent whether $cand_i^c$ is assigned a label l_m^c.

Therefore, the objective function of the Chinese side is

$$O_c = \sum_{i=1}^{N_c} \sum_{m=1}^{L_c} p_{im}^c x_{im}$$

$$\text{s.t. } \sum_{m=1}^{L_c} x_{im} = 1, \text{ for all } i$$

where the constraint $\sum_{m=1}^{L_c} x_{im} = 1$ guarantees that each candidate is assigned one and only one label.

4.1.2. English Component. Similar notations are defined on the English side. The entire semantic label set of the English side is denoted as $\{l_1^e, l_2^e, \ldots l_{L_e}^e\}$. Suppose that there are N_e candidates on the English side and are denoted as $cand_1^e, \ldots, cand_{N_e}^e$. The probability of assigning l_n^e to $cand_j^e$ is p_{jn}^e that is generated by the monolingual system. An indicator variable y_{jn} is defined as:

$$y_{jn} = \left[cand_j^e \text{ is assigned label } l_n^e\right]$$

to represent whether $cand_j^e$ is assigned a label l_n^e. Therefore, the objective function of the English side is

$$O_e = \sum_{j=1}^{N_e} \sum_{n=1}^{L_e} p_{jn}^c y_{jn}$$

$$\text{s.t. } \sum_{n=1}^{L_e} y_{jn} = 1, \text{ for all } j,$$

where the constraint $\sum_{n=1}^{L_e} y_{jn} = 1$ guarantees that each candidate is assigned one and only one label.

4.2. Argument Alignment

In Equation (2), we utilize the Argument Alignment module to enhance the consistency of bilingual SRL. Just as stated in Section 3, both sides of bi-texts have the same predicate argument structure, which indicates that each pair of aligned arguments has consistent labels, for example, the aligned pair of '近期' and 'Recently' in Figure 1 is consistent if labeled as AM-TMP. Moreover, the set of aligned argument pairs have been obtained through the BDP algorithm. Therefore, the goal of Argument Alignment is to determine whether a pair of aligned arguments is consistent if assigned a pair of labels.

4.2.1. Argument Alignment Model. The problem can be formulated as a binary classification task. Here, we also use a Maximum Entropy classifier to determine whether a pair of aligned candidates $cand_i^c$ and $cand_j^e$ are consistent when they were assigned a pair of labels l_m^c and l_n^e. An indicator variable $\phi(i, j, m, n)$ stands for the output of the Argument Alignment model. Its definition is as follows:

$$\phi(i, j, m, n) = \left[cand_i^c \text{ and } cand_j^e \text{ are consistent when assigned } l_m^c \text{ and } l_n^e\right].$$

Table I. The Argument Alignment Matrix

Ch\En	A0	A1	A2	A3	A4	AM*	NUL
A0	492	30	4	0	0	0	46
A1	98	853	43	2	0	0	8
A2	9	57	51	1	0	47	0
A3	1	0	2	6	0	0	0
A4	0	0	2	0	3	0	0
AM*	0	2	39	0	0	895	221
NUL	53	14	27	0	0	45	0

Note: NUL means the Argument ellipsis in one language.

Table II. The Argument Alignment Matrix
for the Predicate Pair ('增长', 'Grow')

Ch\En	A0	A1	A2	AM*	NUL
A0	0	16	0	0	0
A1	0	0	12	0	0
AM*	0	0	4	7	0
NUL	0	16	0	2	0

To train the model, we need a corpus with annotations of alignment relations between arguments. Zhuang and Zong [2010a] have manually aligned the arguments in 60 files (chtb_0121.fid to chtb_0180.fid) of the OntoNotes corpus. Therefore, we use these data as our training data. Statistics about the alignment relations between arguments are shown in Table I. Each entry in Table I is the number of times for which one type of Chinese argument aligns with one type of English argument. AM* stands for all adjunct types such as AM-TMP and AM-LOC, and NUL means that the argument on the other side cannot be aligned with any argument on this side. From Table I, we can get the following observations:

(1) The consistent argument pairs do not have to have the same labels. In general, an argument pair has the same label, but exceptions could occur in some cases. For example, a Chinese A0 argument aligns with an English A1 argument 30 times in the corpus.

(2) The consistent argument pairs have different labels from a limited scope. For example, a Chinese A0 argument aligns with an English A1 argument most of the time except for A0, but never aligns with an A4 or AM* argument.

4.2.2. Features for the Argument Alignment Model. We use the following features in the Argument Alignment Model.

Semantic role labels of argument pairs. From Table I, we can see that semantic role labels of argument pairs are a good indicator of whether they should align with each other. For example, a Chinese A0 argument aligns with an English A0 argument most of the time, but never aligns with an English AM* argument in Table I.

Predicate verb pair and label pair combination. Different predicate pairs have different argument alignment patterns. Let us take the Chinese predicate '增长' and the English predicate 'grow' as an example. The argument alignment matrix for all instances of the Chinese–English predicate pair ('增长', 'grow') in the manually aligned corpus is shown in Table II.

From Table II, we can see that all A0 arguments of '增长' align with A1 arguments of 'grow', which is in contrast to the results in Table I, in which a Chinese A0 argument tends to align with an English A0 argument. This phenomenon shows that a predicate pair can determine which types of arguments should align with each other. Therefore, we use the predicate pair combined with labels as a feature in the argument alignment model.

Syntactic tags and label pair combination. Syntactic tags are a good indicator of semantic roles. Therefore, we concatenated syntactic tags with label pairs as a feature in the Argument Alignment model.

4.3. Inferring Bilingual SRL

We formulate our bilingual SRL model in Equation (2). In this section, we illustrate how to optimize the objective function with constraints. The notation defined in Sections 4.1 and 4.2 are still used here. In addition, the set of aligned candidates pairs is denoted as A, which is obtained from the BDP algorithm. $\phi(i, j, m, n)$ is the output of the Argument Alignment module, representing whether an aligned pair $cand_i^c$ and $cand_j^e$ with the labels of l_m^c and l_n^e are consistent. Here, we set that $\phi(i, j, m, n) = 1$ if $x_{im} = 0$ or $y_{jn} = 0$. In other words, the constraint will be satisfied if $x_{im} = 0$ or $y_{jn} = 0$, and it only works when $(i, j) \in A$, $x_{im} = 1$ and $y_{jn} = 1$. Therefore, the entire objective function becomes:

$$O_c + O_e = \sum_{i=1}^{N_c} \sum_{m=1}^{L_c} p_{im}^c x_{im} + \sum_{j=1}^{N_e} \sum_{n=1}^{L_e} p_{jn}^e y_{jn}$$

$$\text{s.t. } \forall i, \sum_{m=1}^{L_c} x_{im} = 1; \; \forall j, \sum_{n=1}^{L_e} y_{jn} = 1 \text{ and } \forall (i, j) \in A, \phi(i, j, m, n) = 1 \text{ for all } m, n.$$

The Lagrange relaxation of this objective function is

$$L(X, Y, U) = O_c + O_e + \sum_{(i,j) \in A} \sum_{m,n} u_{ij}(m, n)(\phi(i, j, m, n) - 1),$$

where $u_{ij}(m, n)$ is the Lagrange multiplier for the aligned argument pair $cand_i^c$ and $cand_j^e$ with the labels of l_m^c and l_n^e, and X, Y, U denote the sets of the variables x, y, and u, respectively.

Then, the dual objective function is:

$$L(U) = \max_{X,Y} \left(O_c + O_e + \sum_{(i,j) \in A} \sum_{m,n} u_{ij}(m, n)(\phi(i, j, m, n) - 1) \right)$$

$$\text{s.t. } \forall i, \sum_{m=1}^{L_c} x_{im} = 1 \text{ and } \forall j, \sum_{n=1}^{L_e} y_{jn} = 1.$$

The dual problem is to find $\min L(U)$. We use the subgradient method [Boyd and Mutapcic 2003] to minimize the dual objective function. Following Rush et al. [2010], we define the subgradient of $L(U)$ as:

$$\gamma_{ij}(m, n) = \phi(i, j, m, n) - 1.$$

Then, $u_{ij}(m, n)$ is adjusted as follows:

$$u'_{ij}(m, n) = u_{ij}(m, n) - \delta * \gamma_{ij}(m, n),$$

where $\delta > 0$ is a small step size and is set to 0.1 in the experiments.

ALGORITHM 2: Subgradient Method

Set $k = 1$; % initialize the iteration num
Set $u_{ij}(m, n) = 0, \forall (i, j) \in A,$ for all m, n; % initialize the Lagrange multiplier
for $k = 1$ *to* K **do**
    $X^k, Y^k \leftarrow arg\,max_{X,Y}(O_c + O_e + \sum_{(i,j) \in A} \sum_{m,n} u_{ij}(m, n)(\phi(i, j, m, n) - 1));$
    if $\forall (i, j) \in A, \phi(i, j, m, n) = 1$ *for all* m, n
    then
        return (X^k, Y^k);
    else
        $u'_{ij}(m, n) = u_{ij}(m, n) - \delta * \gamma_{ij}(m, n)$
    end
end

Algorithm 2 presents the subgradient method to solve the dual objective. In the beginning, the algorithm initializes the Lagrange multiplier value with 0. In each iteration, the algorithm finds the best X^k and Y^k. If X^k and Y^k are consistent, then the solution is returned. Otherwise, adjust the Lagrange multipliers. It is worth noting that the max problem (Line 5) can be solved efficiently because the probability of assigning a label to each candidate has been obtained by the monolingual systems. From this optimization process, the Lagrange multiplier can be regarded as an intermediate component that generates punishments for the inconsistent results.

5. EXPERIMENTS

5.1. Experimental Setup

In our experiments, we use the Xinhua News portion of Chinese and English data in LDC OntoNotes Release 3.0 (LDC2009T24). This data contains parallel proposition annotations for 325 files (chtb_0001.fid to chtb_0325.fid) from a Chinese–English parallel treebank. The detailed description of this data can be found in Palmer et al. [2005]. The English part of this data contains proposition annotations only for verbal predicates. Therefore, we consider only verbal predicates in this work.

We use the GIZA++ toolkit [Och and Ney 2003] to perform automatic word alignment. In additon to the parallel PropBank data, we use an extra 4,500K Chinese–English sentence pairs to induce word alignments for both directions, with the default GIZA++ settings. The alignments are symmetrized using the intersection heuristic, which is known to produce high-precision alignments.

Our monolingual SRL systems are trained separately. The Chinese SRL system is trained on 640 files (chtb_0121.fid to chtb_0931.fid) in the Chinese Propbank 1.0; the English SRL system is trained on Sections 02-21 of the English Propbank. We use Berkeley parser and Charniak parser for auto parsing in Chinese and English, respectively. The parsers are also retrained on the training sets. The max iteration number in the inference is set to 30.

In the parallel corpus, we use 80 files (chtb_0001 to chtb_0080.fid) as the test data, 40 files (chtb_0081.fid to chtb_0120.fid) as the development data, and 60 files (chtb_0121.fid to chtb_0180.fid) as the training data. The data setting is the same as in Zhuang and Zong [2010a]. Table III shows the sentence and predicate distributions in each dataset. We note that there are a small number of one-to-many Chinese–English sentence pairs except for one-to-one sentence pairs. The quality of automatic word alignment on one-to-many Chinese–English sentence pairs is usually very poor, however. Therefore, we include only one-to-one Chinese–English sentence pairs in all data, and not all predicates in a sentence pair can be included. Only bilingual predicate pairs

Table III. Sentence and Predicate Counts

	Test	Development	Train
Articles	1–80	81–120	121–180
Chinese sentences	1,067	578	778
English sentences	1,182	620	828
Bilingual pairs	821	620	614
Chinese predicates	3,792	2,042	2,572
English predicates	2,864	1,647	1,860
Bilingual pairs	1476	790	982

Table IV. Results of Varying the Threshold μ
of Length Ratio on the Development Set

μ	P	R	F_1
1.0	90.2	84.9	87.5
1.5	90.2	85.0	87.5
2.0	90.1	85.2	87.6
2.5	90.1	85.2	87.7
3.0	90.1	85.5	**87.8**
3.5	89.9	85.7	87.7
4.0	89.7	85.9	87.7
4.5	89.5	86.0	87.7
5.0	89.4	86.1	87.7
5.5	89.2	86.1	87.6
6.0	89.2	86.1	87.6

Note: The length ratio is obtained by dividing the
English side's length by the Chinese side's length.
The score of P, R, and F_1 is the average of both sides
of bi-texts.

are included. A bilingual predicate pair is defined to be a pair of predicates in bi-texts, which align with each other in automatic word alignment. The setting for sentence pairs and predicate pairs is also the same as in Zhuang and Zong [2010a]. To train the Argument Alignment model, we need both gold samples and negative samples, but there are only gold samples in the manually annotated corpus. To produce sufficient negative samples, we fix one label on one side and randomly change the label on the other side. Finally, Zhang Le's MaxEnt toolkit is adopted to train the Argument Alignment model with 300 iterations and Gaussian prior 1.0.

5.2. Evaluation Metric and Parameter Estimation

For fair comparison of different SRL systems, we use the widely accepted criteria: *Precision* (**P**), *Recall* (**R**), and F_{score} (**F_1**). Following Zhang et al. [2004], we conducted statistical significance tests with 1,000 resampling size and 95% confidence interval. In the following tables, an asterisk is used to indicate that the difference between the proposed approach and the corresponding baseline system was statistically significant.

In BDP, we designed 4 heuristics rules to prevent the noisy candidates. Rules (3) and (4) can be directly applied in the argument-identification stage. But, for Rules (1) and (2), there are two hyperparameters, μ and τ; we tuned them on the development set. Tables IV and V show the results of independently varying μ and τ on the development data, respectively. They show that the best performance is achieved when $\mu = 3$ and $\tau = 0.4$. Therefore, we use these two thresholds in our following experiments.

Table V. Results of Varying the Threshold τ of
Identification Probability on the Development Set

τ	P	R	F_1
0.1	89.0	86.2	87.6
0.2	89.2	86.1	87.6
0.3	89.6	85.9	87.7
0.4	89.9	85.9	**87.8**
0.5	90.0	85.4	87.6
0.6	90.0	85.2	87.6
0.7	89.8	84.8	87.2
0.8	89.7	84.3	86.9
0.9	89.6	84.0	86.7

Note: The score of P, R, and F_1 is the average of both sides of bi-texts.

Table VI. Results of Different Bilingual SRL Systems

Systems	Chinese			English		
	P	R	F_1	P	R	F_1
Monolingual	80.4	74.1	77.2	77.6	74.2	75.8
Zhuang and Zong [2010a]	81.4	75.0	78.1*	78.7	75.3	77.0*
Ours	81.6	76.2	78.8*	78.7	76.3	77.5*

Note: The Symbol * indicates that the results are significantly better ($p < 0.05$) than for the baseline system.

5.3. Bilingual SRL Results

The main results of different approaches on the test set are shown in Table VI. We can see that both our system and Zhuang and Zong's [2010a] system significantly improve the F1 scores over monolingual SRL systems, which demonstrates that jointly inferring bilingual SRL is superior to the monolingual SRL systems. On the other hand, if we compare our system with Zhuang and Zong's system, our approach also outperforms theirs, which mainly benefits from the improvement of the recall score. The large improvement of the recall score indicates that BDP helps recover many candidates discarded in the argument-identification stage. Although the improvement of our method over Zhuang and Zong [2010a] is limited, our method is simple and efficient (see Section 5.5), which is particularly important if the data to be processed are very large.

5.4. Effects of BDP

BDP is designed to recover the missing candidates in the argument-pruning and identification stages. In this section, we investigate the effects of BDP on argument identification and SRL, respectively.

The results of argument identification on the test set are shown in Table VII. In the table, 'Bi' means the system without BDP, 'Bi-Ch' means the system with the projection from English to Chinese, 'Bi-En' means the system with the projection from Chinese to English, and 'Bi-All' means the system with the complete BDP. We can see that both 'Bi-Ch' and 'Bi-En' improve the performance of the Chinese side and the English side, respectively, while 'Bi-All' reaches the best performance on both sides, which indicates that Location Complementarity occurs widely on both sides of monolingual SRL systems, and BDP can recover a large number of missing candidates in the argument-identification stage.

Table VII. Effects of BDP on Argument Identification

Systems	Chinese			English		
	P	R	F_1	P	R	F_1
Bi	85.8	79.1	82.3	88.2	84.4	86.3
Bi-Ch	87.0	80.9	83.8	88.2	84.4	86.3
Bi-En	85.8	79.1	82.3	89.0	85.5	87.2
Bi-All	87.3	81.1	84.0*	89.4	85.7	87.5*

Note: The symbol * indicates that the results are significantly better ($p < 0.05$) than for the baseline system.

Table VIII. Effects of BDP on SRL in Different Systems

Systems	Chinese			English		
	P	R	F_1	P	R	F_1
Monolingual	80.4	74.1	77.2	77.6	74.2	75.8
Monolingual + BDP	80.4	75.0	77.6	77.5	74.8	76.1
Zhuang and Zong [2010a]	81.4	75.0	78.1	78.7	75.3	77.0
Zhuang and Zong [2010a] + BDP	81.4	76.0	78.6	78.5	76.3	77.4
Ours-	81.5	75.1	78.2	78.8	75.4	77.1
Ours	81.6	76.2	78.8	78.7	76.3	77.5

"Ours" means our whole system and "Ours-" means the system for which BDP is removed. "+ BDP" means that the systems are added extra candidates generated by BDP.

We further investigate the effects of BDP on SRL. The results of the test set are shown in the Table VIII. In the experiments, we measure the effects of BDP on three SRL systems: monolingual SRL systems, Zhuang and Zong's [2010a] system and our system. We can see that, after BDP is added, the performance of all SRL systems improved, which benefits from a large number of candidates recovered by BDP. The improvement of monolingual SRL systems is slightly lower than the others because monolingual SRL systems labeled both sides of bi-texts independently, while jointly inferring was implemented in Zhuang and Zong's [2010a] system and our system. Another interesting observation is that, whatever BDP is added, our system and Zhuang and Zong's [2010a] system have comparable performance. This indicates that, under the evaluation metric of F_1, the technique of DD does not prevail over ILP in inferring bilingual SRL. The advantage of DD over ILP is its simplicity and high efficiency (described in Section 5.5).

5.5. Efficiency of Dual Decomposition

Zhuang and Zong [2010a] utilized ILP to infer bilingual SRL, but the objective function in their method is very complex and the ILP technique requires costly inference. Nevertheless, from Algorithm 2, we can see that the Lagrange DD is very concise and fast. Here, we compare the time consumption of the two methods in inferring bilingual SRL; the results are shown in Figure 5. We can see that the time consumption of the DD is only one fourth of the ILP's consumption under the same hardware condition. In consideration of data being produced at a vast and unprecedented scale in modern society, we think our method is significantly superior to Zhuang and Zong [2010a] in practical applications.

5.6. Convergence

The technique of Lagrange DD is iterative. Thus, the convergence of the algorithm is an important indicator of a good algorithm. Figure 6 shows the percentage of cases in which the exact solutions are returned (we reach an agreement between the Chinese side and the English side) versus the number of iterations of the algorithm. The algorithm
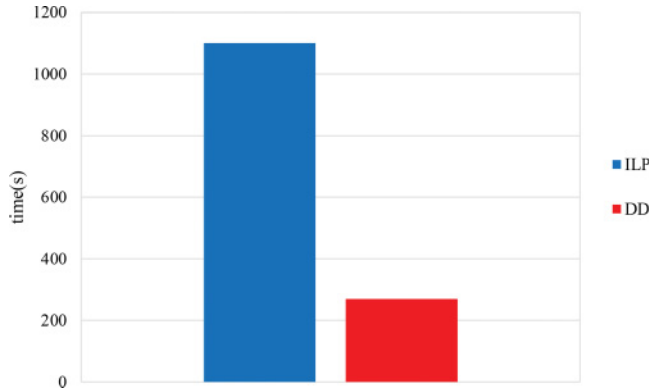
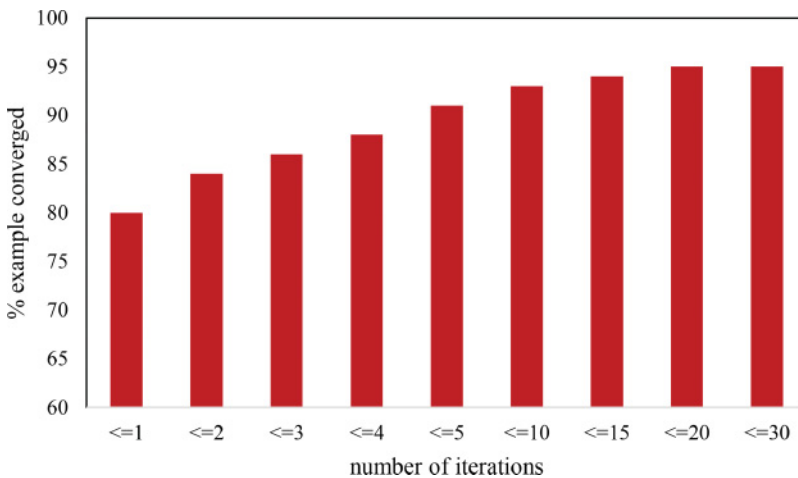Fig. 5.   Comparison of time consumption between DD and ILP.

Fig. 6.   Convergence results of Dual Decomposition. We show the percentage of examples in which an exact solution is returned by the algorithm versus the number of iterations of the algorithm.

Table IX. Statistics for the Errors in Monolingual SRL Systems and our System

	Chinese side's errors	English side's errors	All errors	Consistency
Monolingual	202	322	524	80%
Ours	160	270	430	94%

Note: The score of 'consistency' is computed by the Argument Alignment model.

produces exact solutions for over 94% of all examples if the maximum iteration number is set to 30. In addition, for over 90% of the examples, the algorithm returns an exact solution in 10 iterations or fewer. Although the dual decomposition algorithm is not guaranteed to give an exact solution, it is very successful at achieving this goal in our task.

5.7. Error Analysis

To further illustrate the effectiveness of our approach, we perform a manual error analysis on the results. Table IX shows the statistics on the errors in monolingual SRL systems and our system. We can see that our method obtained approximately 20% error reduction compared with monolingual SRL systems both on the Chinese side and

Table X. Some Examples in Our System

	Examples	Monolingual	Ours	Gold
1	**经 台 当局** 核准 的 台商 投资案	AM-MNR	A0	A0
	the investment proposals <u>approved</u> **by Taiwan authorities**	A0	A0	A0
2	**一九九五 和 一九九六 年**, 中韩双方 进行 了 会谈	AM-TMP	AM-TMP	AM-TMP
	in 1995 and 1996, China and North Korea <u>held</u> the conference	AM-LOC	AM-TMP	AM-TMP

Table XI. The Classification Errors for the
Arguments in Set *A* from Our System

Source	Number
Monolingual SRL systems	220
Argument Alignment module	100
All	320

the English side. The corresponding error reduction shows that our method is effective in jointly decoding bilingual SRL results. Moreover, the overall consistency of bilingual SRL results is also improved by 14%.

To better understand the advantages of jointly inferring bilingual SRL, some examples are given in Table X. In Example (1), although the argument '经 台 当局' is a prepositional phrase (PP), it is the agent (A0) of the predicate '核准'. In the training data of the Chinese side, the arguments with a PP tag are usually labeled as AM-MNR, AM-LOC, or AM-TMP, except for A0. We found that such a case (a PP as A0) occurs less than 10 times in the training data of the Chinese side. Thus, it is not surprising that the argument '经 台 当局' is incorrectly labeled as AM-MNR by the Chinese SRL system. Meanwhile, it can be seen that, although the argument '*by Taiwan authorities*' on the English side is also a PP, it is correctly labeled as A0 because a PP as A0, especially when the PP begins with 'by', is very common in the training data of the English side. Given the labels AM-MNR and A0, the module of Argument Alignment successfully recognizes the inconsistency of both sides. After jointly inferring of our approach, we obtain the correct labels for both sides.

In Example (2), the argument '*in 1995 and 1996*' is incorrectly labeled as AM-LOC by the English SRL system, but in fact this argument indicates the time of the predicate 'hold'. In an SRL system, AM-LOC and AM-TMP are easily confused because these two kinds of roles often occupy similar syntax contexts. However, the argument '一九九五 和 一九九六 年' on the Chinese side is labeled correctly as AM-TMP because the last word '年' (year) is a useful clue to clearly indicate the time. The argument pair with labels AM-TMP and AM-LOC is classified into the inconsistency by the module of Argument Alignment and, finally, our approach outputs the correct labels for both sides.

Although the proposed model has significantly enhanced the consistency of bi-texts, some challenges remain. We collect the classification errors for the arguments of the Argument Alignment set *A* in our system; the corresponding statistics are shown in Table XI. We found that the major sources of the remaining errors are monolingual SRL systems ($220/320 = 68\%$) and the Argument Alignment module ($100/320 = 32\%$).

— **Monolingual SRL systems.** The jointly inferring in our method is based on the candidates and the classification probability generated by monolingual SRL systems. According to the description of our method (Section 4.3), we find that the iteration process reaches the optimal through imposing punishment on the inconsistent results. However, in the case that the classification probability for the gold label in monolingual SRL systems is too low, it is hard for jointly inferring to correct this kind of error. In all errors, this kind of error accounts for 68%, and is the main source for the remaining errors.

—**The Argument Alignment module.** The Argument Alignment module is designed as the constraint to enhance the consistency of bi-texts and plays a core part in our approach. Thus, it is crucial to the performance of jointly inferring bilingual SRL. To evaluate the performance of the Argument Alignment module, we adopt a 10-fold cross-validation experiment on the training set. The average accuracy is about 85%. Although the Argument Alignment module performs well, it is far from perfect; it causes 32% of errors. To further analyze the source of these errors, we classify these errors into two categories: one is caused by the Argument Alignment module not recognizing the inconsistent results ($80/320 = 25\%$) and the other is caused by the Argument Alignment module mistaking the consistent results for the inconsistent ones ($20/320 = 7\%$).

6. RELATED WORK

Some previous work on monolingual SRL is related to our work. Gildea and Jurafsky [2002] first presented a system based on a statistical classifier, which is trained on a hand-annotated corpora, FrameNet. In their pioneering work, they used a gold or autoparsed parse tree as the input, then extracted various lexical and syntactic features to identify the semantic roles for a given predicate. After Gildea and Jurafsky [2002], there have been a large number of works on automatic semantic role labeling. Punyakanok et al. [2004] constructed an ILP architecture in which the dependency relationships among arguments are implied in the constraints. In addition, there have been many extensions in machine-learning models [Moschitti et al. 2008], feature engineering [Xue and Palmer 2004], and inference procedures [Toutanova et al. 2005; Punyakanok et al. 2004; Yang and Zong 2014; Yang et al. 2015; Zhuang and Zong 2010b].

Research on SRL utilizing parallel corpus is also related to our work. Fung et al. [2007] did pioneering work on studying argument alignment on Chinese–English parallel PropBank. They first identified and labeled the semantic structures using the Chinese and English shallow semantic parsers. Then, given the SRL results on both sides, they automatically induced the argument alignment between two sides. Padó and Lapata [2009] did research on cross-lingual annotation projection on an English–German parallel corpus. They performed SRL only on the English side, then mapped the English SRL result to the German side. The difference between our work and these works is that we fully utilize complementary language cues in parallel bi-texts to improve the performance of both sides.

The work most related to ours is Zhuang and Zong [2010a]. They constructed an ILP framework to incorporate not only linguistic constraints on source and target sides of bi-texts, but also the bilingual argument structure consistency requirement on bi-texts. They reported substantial improvements on both sides of bi-texts compared with the state-of-the-art monolingual SRL system. However, our work is different from theirs in two ways. First, they fully accepted the candidates from monolingual SRL systems as the candidates in the inferring phase while a number of arguments were discarded in the argument-pruning and identification stage. We developed an algorithm to recover these discarded arguments. On the other hand, they used the technique of ILP to infer the bilingual SRL, but the objective function of their model is very complex and ILP usually requires costly inference. In contrast to Zhuang and Zong [2010a], we utilized Lagrange DD to efficiently infer consistent results for both sides.

7. CONCLUSIONS

This article focuses on bilingual Semantic Role Labeling, whose goal is to annotate semantic roles on bi-texts. If we use monolingual SRL systems to make annotations on both sides of bi-texts, it is very difficult to obtain good and consistent SRL

results on both sides. Moreover, we note that there is a lot of complementarity information on both sides of bi-texts. We classified the complementarity information into two categories: Location Complementarity and Label Complementarity. In this article, we describe how to utilize the two categories of complementarity information to improve bilingual SRL. We mainly make efforts on the following two points: (1) in the argument-identification stage, we propose a Bi-Directional Projection algorithm to recover candidates discarded in the argument-pruning or identification phase; (2) in the inferring stage, we adopted the technique of Lagrange Dual Decomposition to infer consistent results for bi-texts. Compared with existing methods, we obtain better results and a faster processing speed. Finally, the proposed model does not utilize language-dependent information. Therefore, although our experiments are conducted on Chinese–English language pairs, it is expected that the proposed approach can be applied to other language pairs with little adaptation effort.

In our work, we still utilize monolingual SRL systems to generate candidates for joint inferring. In the future, we will consider how to perform joint learning in both sides of bi-texts without any monolingual SRL systems.

REFERENCES

Stephen Boyd and Almir Mutapcic. 2003. *Subgradient Methods*. Lecture notes of ee392o. Stanford University, Stanford, CA.

Xavier Carreras, Lluís Màrquez, and Grzegorz Chrupała. 2004. Hierarchical recognition of propositional arguments with perceptrons. In *HLT-NAACL 2004 Workshop: 8th Conference on Computational Natural Language Learning (CoNLL'04)*.

Pascale Fung, Zhaojun Wu, Yongsheng Yang, and Dekai Wu. 2006. Automatic learning of Chinese English semantic structure mapping. In *Proceedings of the 2006 ACL Spoken Language Technology Workshop*. 230–233. DOI:http://dx.doi.org/10.1109/SLT.2006.326797

Pascale Fung, Zhaojun Wu, Yongsheng Yang, and Dekai Wu. 2007. Learning bilingual semantic frames: Shallow semantic parsing vs. semantic role projection. In *Proceedings of the 11th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI'07)*.

Kuzman Ganchev, Jennifer Gillenwater, and Ben Taskar. 2009. Dependency grammar induction via bitext projection constraints. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL (ACL'09)*.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling for semantic roles. *Comput. Linguist.* 28, 3 (2002), 245–288.

Daniel Gildea and Martha Palmer. 2002. The necessity of parsing for predicate argument recognition. In *Proceedings of the Joint Conference of the 40th Annual Meeting of the ACL (ACL'02)*. DOI:http://dl.acm.org/citation.cfm?id=1073124

Fei Huang and Stephan Vogel. 2002. Improved named entity translation and bilingual named entity extraction. In *4th IEEE International Conference on Multimodal Interfaces (ICMI'02)*. DOI:http://dx.doi.org/10.1109/ICMI.2002.1167002

Rebecca Hwa, Philip Resnik, Amy Weinberg, Clara Cabezas, and Okan Kolak. 2005. Bootstrapping parsers via syntactic projection across parallel texts. *Natural Language Engineering* 11, 3, 311–325. DOI:http://dx.doi.org/10.1017/S1351324905003840

Terry Koo, Alexander M. Rush, Michael Collins, Tommi Jaakkola, and David Sontag. 2010. Dual decomposition for parsing with non-projective head automata. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP'10)*.

Ding Liu and Daniel Gildea. 2010. Semantic role features for machine translation. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING'10)*.

Robert C. Moore. 2003. Learning translations of named-entity phrases from parallel corpora. In *Proceedings of the 10th Conference on European Chapter of the Association for Computational Linguistics (EACL'03)*.

Alessandro Moschitti, Daniele Pighin, and Roberto Basili. 2008. Tree kernels for semantic role labeling. *Computational Linguistics* 34, 2, 193–224. DOI:http://dx.doi.org/10.1162/coli.2008.34.2.193

Srini Narayanan and Sanda Harabagiu. 2004. Question answering based on semantic structures. In *Proceedings of the 17th International Conference on Computational Linguistics (COLING'04)*.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics* 29, 1, 19–51. DOI:http://dx.doi.org/10.1162/089120103321337421

Sebastian Padó and Mirella Lapata. 2009. Cross-lingual annotation projection of semantic roles. *Journal of Artificial Intelligence Research* 36, 1, 307–340.

Martha Palmer, Nianwen Xue, Olga Babko-Malaya, Jinying Chen, and Benjamin Snyder. 2005. A parallel proposition bank II for Chinese and English. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky (CorpusAnno'05)*.

Sameer Pradhan, Kadri Hacioglu, Valerie Krugler, Wayne Ward, James H. Martin, and Daniel Jurafsky. 2005. Support vector learning for semantic argument classification. *Machine Learning* 60, 1–3, 11–39. DOI:http://dx.doi.org/10.1007/s10994-005-0912-2

Vasin Punyakanok, Dan Roth, Wen-tau Yih, and Dav Zimak. 2004. Semantic role labeling via integer linear programming inference. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING'04)*. DOI:http://dx.doi.org/10.3115/1220355.1220552

Alexander M. Rush, David Sontag, Michael Collins, and Tommi Jaakkola. 2010. On dual decomposition and linear programming relaxations for natural language processing. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP'10)*.

Luo Si and Jamie Callan. 2005. Multilingual retrieval by combining multiple multilingual ranked lists. In *Accessing Multilingual Information Repositories, 6th Workshop of the Cross-Language Evalution Forum (CLEF'05)*. DOI:http://dx.doi.org/10.1007/11878773_13

Philipp Sorg and Philipp Cimiano. 2012. Exploiting Wikipedia for cross-lingual and multilingual information retrieval. *Data and Knowledge Engineering* 74, 26–45. DOI:http://dx.doi.org/10.1016/j.datak.2012.02.003

Mihai Surdeanu, Sanda Harabagiu, John Williams, and Paul Aarseth. 2003. Using predicate-argument structures for information extraction. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics (ACL'03)*. DOI:http://dx.doi.org/10.3115/1075096.1075098

Kristina Toutanova, Aria Haghighi, and Christopher D. Manning. 2008. A global joint model for semantic role labeling. *Computational Linguistics* 34, 2, 161–191. DOI:http://dx.doi.org/10.1162/coli.2008.34.2.161

Mengqiu Wang, Wanxiang Che, and Christopher D. Manning. 2013a. Joint word alignment and bilingual named entity recognition using dual decomposition. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL'13)*.

Zhiguo Wang, Chengqing Zong, and Nianwen Xue. 2013b. A lattice-based framework for joint Chinese word segmentation, POS tagging and parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL'13)*.

Dekai Wu and Pascale Fung. 2009a. Can semantic role labeling improve SMT? In *Proceedings of the 13th Annual Conference of European Association for Machine Translation (EAMT'09)*.

Dekai Wu and Pascale Fung. 2009b. Semantic roles for SMT: A hybrid two-pass model. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics (NAACL'09)*.

Deyi Xiong, Min Zhang, and Haizhou Li. 2012. Modeling the translation of predicate-argument structure for SMT. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Jeju Island, Korea, 902–911.

Nianwen Xue. 2004. Calibrating features for semantic role labeling. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'04)*.

Nianwen Xue. 2008. Labeling Chinese predicates with semantic roles. *Computational Linguistics* 34, 2, 225–255. DOI:http://dx.doi.org/10.1162/coli.2008.34.2.225

Nianwen Xue and Martha Palmer. 2004. Calibrating features for semantic role labeling. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'04)*.

Haitong Yang, Tao Zhuang, and Chengqing Zong. 2015. Domain adaptation for syntactic and semantic dependency parsing using deep belief networks. *Transactions of the Association for Computational Linguistics* 3, 271–282.

Haitong Yang and Chengqing Zong. 2014. Multi-predicate semantic role labeling. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP'14)*.

David Yarowsky, Grace Ngai, and Richard Wicentowski. 2001. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of the 1st International Conference on Human Language Technology Research (HLT'01)*. DOI:http://dx.doi.org/10.3115/1072133.1072187

Feifei Zhai, Jiajun Zhang, Yu Zhou, and Chengqing Zong. 2012. Machine translation by modeling predicate-argument structure transformation. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING'12)*.

Feifei Zhai, Jiajun Zhang, Yu Zhou, and Chengqing Zong. 2013. Handling ambiguities of bilingual predicate-argument structures for statistical machine translation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL'13)*.

Ying Zhang, Stephan Vogel, and Alex Waibel. 2004. Interpreting Bleu/NIST scores: How much improvement do we need to have a better system?. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04)*. 2051–2054.

Tao Zhuang and Chengqing Zong. 2010a. Joint inference for bilingual semantic role labeling. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP'10)*.

Tao Zhuang and Chengqing Zong. 2010b. A minimum error weighting combination strategy for Chinese semantic role labeling. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING'10)*.