

Learning Event Expressions via Bilingual Structure Projection

Fangyuan Li¹, Ruihong Huang², Deyi Xiong^{1*}, Min Zhang¹

Soochow University, Suzhou, China¹

Texas A&M University, College Station, USA²

fyli@stu.suda.edu.cn, huangrh@cse.tamu.edu

{dyxiong, minzhang}@suda.edu.cn

Abstract

Identifying events of a specific type is a challenging task as events in texts are described in numerous and diverse ways. Aiming to resolve high complexities of event descriptions, previous work (Huang and Riloff, 2013) proposes multi-faceted event recognition and a bootstrapping method to automatically acquire both event facet phrases and event expressions from unannotated texts. However, to ensure high quality of learned phrases, this method is constrained to only learn phrases that match certain syntactic structures. In this paper, we propose a bilingual structure projection algorithm that explores linguistic divergences between two languages (Chinese and English) and mines new phrases with new syntactic structures, which have been ignored in the previous work. Experiments show that our approach can successfully find novel event phrases and structures, e.g., phrases headed by nouns. Furthermore, the newly mined phrases are capable of recognizing additional event descriptions and increasing the recall of event recognition.

1 Introduction

Event recognition aims to identify documents that describe a specific type of event. Accurate event recognition is challenging due to ambiguities of event keywords. In the previous work, Huang and Riloff (2013) (hereafter H&R) proposed multi-faceted event recognition method that uses event expressions as well as event defining characteristics (aka “event facets”, such as “agents” and “purpose”) to achieve high accuracy in identifying civil unrest events. They also presented a bootstrapping solution that can learn event expressions and event facet phrases from unannotated texts. However, to achieve high quality phrases, strict syntactic constraints have been enforced and their bootstrapping algorithm can only learn two particular types of V-O (Verb-Object) Structure for both event expressions and facet phrases. Obviously, diverse forms of other verb phrases and non-verb phrases exist to describe events and are ignored by the proposed algorithm. For instance, a verb phrase where two verbs are connected with a particular dependency relation “xcomp”¹, (e.g., “came out to demonstrate”) is one of these structures. Civil unrest events can also be invoked by some noun structure phrases, such as just a noun word phrase (e.g., “sit-ins”) or phrases starting with a noun (e.g., “disobedience of order”), even a passive form phrase structure like “rallies held (in)”.

In order to address this issue, we propose a simple yet effective bilingual structure projection method that explores syntactic divergences (Georgi et al., 2012) between two languages and mines new syntactic structures for event expressions and event facet phrases effectively using parallel corpora. This is inspired by many recent cross-lingual research that utilize the second language to provide a different view (Balcan and Blum, 2005; Burkett et al., 2010; Ganchev et al., 2012) and complementary cues (Che et al., 2013; Wang et al., 2013) in improving Natural language Processing (NLP) tasks for the target language, analogous to co-training (Chen and Ji, 2009; Wan, 2009; Hajmohammadi et al., 2015) but between two different languages. In order to learn new event phrases and their syntactic structures, we map phrases² back and

*Corresponding author

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>

¹“xcomp” is a dependency relation between a verb or an adjective and its open clausal complement in a dependency tree. In sentence “Workers came out to demonstrate”, the relation between verb “came” and verb “demonstrate” is “xcomp”.

²We start with initial phrases learned by H&R, thanks to the authors for sharing the learned phrases and evaluation data.

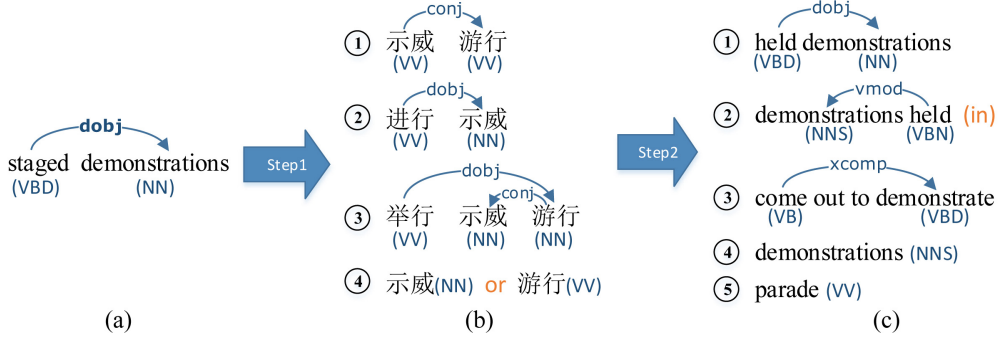


Figure 1: The bilingual structure mapping procedure

forth multiple times between two languages using parallel corpora to make full use of these divergence information. We choose Chinese as the pivot language to learn English event phrases and event facet phrases. On the one hand, both Chinese and English share a common sentence structure SVO (Subject-Verb-Object) and other similar sentence composition. On the other hand, these two languages have many significant differences, e.g., an uninflected language (Chinese) vs. an inflected language (English). The commonalities enable bilingual projection while the language divergences stimulate occurrences of new phrases and new phrase structures. The input to our bilingual structure projection system are two English verb phrase lists, event phrases and purpose facet phrases which are learned by H&R’s multi-faceted event recognition method. After each mapping step, new syntactic structures and new phrases are learned.

Figure 1 illustrates one iteration of bilingual structure mapping with examples. Given the English phrase “staged demonstrations” with the structure of “VBD<dobj>NN”, English sentences containing this phrase in the parallel corpora are identified. Then various Chinese phrases (Figure 1(b)) are generated when mapping the English phrase to its Chinese correspondence based on word alignments of the parallel sentences. Interestingly, a multi-word verb phrase in English can be expressed with only one noun (e.g., “示威”/demonstration) or a verb (e.g., “游行”/parade) in Chinese. Furthermore, we have observed that Chinese tends to use a conjunction of two verbs that have roughly the same meaning when referring to one single event, e.g., “示威 游行” in Figure 1(b). More examples are given in Figure 1. When we map these already diversified Chinese phrases back to English, new phrases with richer syntactic structures (Figure 1(c)) are generated, including some interesting noun structure phrases and one single-word phrase. To fully exploit language divergences, the bilingual structure projection run back and forth between the two languages and continues for several iterations.

Experiment results show that our approach can successfully find hundreds of new English phrasal structures, e.g., structures headed by nouns, and learn thousands of new event expression and event facet phrases. Furthermore, using the same evaluation data and evaluation method as in H&R, the newly mined phrases are capable of recognizing additional event descriptions, and significantly increasing the recall of event recognition by 8.2 points and the overall F_1 -score by 3.5 points.

This paper is structured as follows: Section 2 describes the H&R’s multi-faceted event recognition approach. Section 3 details our bilingual structure projection method and some heuristic rules used in our method. Section 4 describes our experimental design and evaluation results. Then section 5 discusses a variety of new phrases and structures generated by our approach. Section 6 introduces the related work of bilingual methods for various NLP tasks. Last, section 7 summarizes the bilingual structure method and expounds our future work.

2 Background

Accurately identifying documents that describe a specific type of event is a challenging task because events can be mentioned in various complex contexts. Using event keywords alone are barely reliable. For example, while the words “strike”, “rally” and “riot” are commonly used to describe civil unrest events, they frequently refer to other events that are dramatically different from civil unrest events including “bowling strike”, “rally car” and “imagination riot”. In the previous work, Huang and Riloff

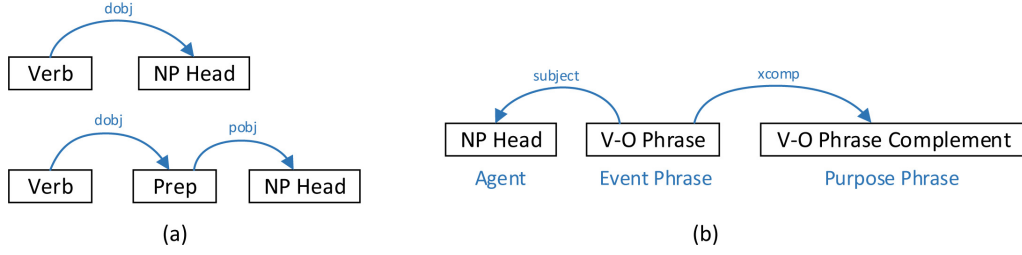


Figure 2: Syntactic constraints in H&R’s method. (a) shows two pre-defined phrase structures, (b) is the sentence pattern in H&R’s bootstrapping system

(2013) proposed the multi-faceted event recognition approach that uses both event expressions and event defining characteristics (called event facets) to accurately identify event occurrences. As described in the paper, agents and purpose are two types of event facets that are essential to distinguish many types of events. For instance, both natural disasters and military incidents can mention injuries and deaths as consequences, however, their agents are distinct. The agents of natural disasters have to be natural force while the agents of military incidents have to include military personnel. Similarly, purposes describe motivations of events and are extremely helpful to distinguish various types of events.

H&R also proposed a bootstrapping framework to learn event expressions and event facet dictionaries from the unannotated texts automatically requiring only minimal supervision with a few event keywords and a few seed phrases for each event facet. They observed that event facet phrases and event expressions tend to co-occur in event introductory sentences. Therefore, the bootstrapping system first learns event expressions from the sentences that contain both types of event facets and then learns more event facet phrases from the sentences that contain an event expression and a different type of event facet, in an iterative manner. Their multi-faceted event recognition with bootstrapped event dictionaries achieved high precision (88%) with a reasonable recall (71%) on identifying civil unrest events. However, to ensure high quality of learned phrases, strict syntactic constraints were enforced at both the phrasal and sentential level. Specifically, they only considered event expressions and purpose phrases as verb phrases in two types (Figure 2(a)), the first phrasal type is a verb followed by the head noun of its direct object and the second phrasal type is a verb with an attached prepositional phrase, while reasonably both event expressions and purpose phrases can exist in many other syntactic structures. Furthermore, within a sentence, specific dependency relations are required between both facet phrases and the main event expression (Figure 2(b)), the agent term has to be the syntactic subject of the event expression and the purpose phrase has to be a clausal complement of the event expression. Obviously, these harsh syntactic constraints pose limitations to the types of event phrases and event facet phrases that can be learned using this framework. Our research is committed to mine new phrases and phrase structures that go beyond these constraints leveraging divergences across two languages.

3 Learning Event Expressions via Bilingual Structure Projection

Our algorithm can iterate multiple times to learn new phrases in new syntactic structures automatically. In our experiments, we expand two types of phrases: event phrases (EP) and purpose phrases (PP) learned by H&R’s method.

3.1 One Iteration of Bilingual Structure Projection

In our bilingual projection, we use structured phrases for mapping. Structured phrases³ comprise both lexical and structural information. One iteration of the projection consists of two stages: mapping English event phrases to Chinese and projecting Chinese equivalents back to English. Next, we use an example as shown in Figure 2 to illustrate the projection process from English to Chinese.

³A structured phrase is defined as “start_node <relation1> in_node1 <relation2> in_node2 <relation3> ... <relationN> end_node”, where each node is a word, and the relation between two nodes is their dependency relation. Structured phrases capture both lexical and structural information for event expressions. Each structured phrase is essentially a path between two nodes in a dependency parse tree.

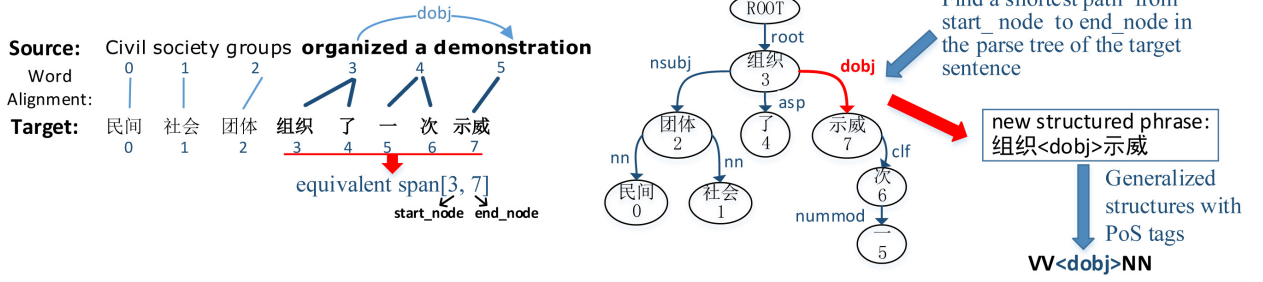


Figure 3: The illustration of projection from English to Chinese

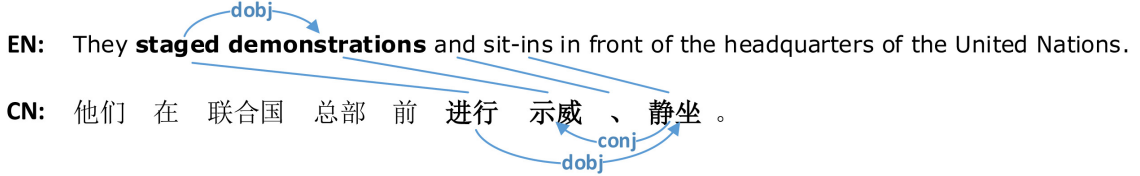


Figure 4: One verb with two coordinate objects

First, we identify an English event phrase on the source side of our parallel corpora, e.g., “organized a demonstration” in Figure 3, and extract the corresponding structured phrase “organized <dobj> demonstration” according to the dependency tree. Second, we detect phrase span of the translation equivalent for the extracted structured phrase by computing the aligned span on the target side via word alignments. In Figure 3, the equivalent span on the Chinese side is [3, 7]. In the third step, we take the leftmost and rightmost word of the translation equivalent span as the start_node and end_node. Then we further generate the structured phrase on the target side by finding the shortest path from the start_node to the end_node in the dependency tree, e.g., “组织 <dobj> 示威”. For the in-depth analysis in Section 5, we use part-of-speech (PoS) tags to replace words in the found structured phrase to obtain a generalized structure, e.g., VV<dobj>NN in Figure 3. Then, mapping the new generated Chinese structured phrases (“组织 <dobj> 示威” in Figure 3 for example) back to English in the second stage of the bilingual projection following the similar procedure as described above. After the two-stages mapping, we obtain diversified English event phrases.

3.2 Phrase Decomposition

From our training data, we have learned many phrases with a conjunction. We find that most of them follow two structure patterns. The first is that one verb has two coordinate objects that express two related events. For example, the Chinese equivalent of English phrase “staged demonstrations” in Figure 4 is “进行 示威”. However, there is not a direct dependency relation between “进行” and “示威” in the dependency tree. Instead, they are connected by a word “静坐” (sit-ins). Apparently, “staged demonstrations” (进行 示威) and “staged sit-ins” (进行 静坐) are two related events. The other pattern is an interesting phenomenon we have observed in our structure projection experiments. Chinese language tends to use a conjunction of two words that have roughly the same meaning when referring to an event while in English only one of the two coordinates is used to refer to the same event. For example, “捍卫 <dobj> 人权 <conj> 民主” (defend human rights and democracy) is a common expression in Chinese. However, in English, “defend human rights” is used to express the same meaning. To exploit this conjunction structure and linguistic divergence, we split such phrases into two separate phrases and keep both original phrases and decomposed phrases in the bilingual projection. For example, “进行 示威、静坐” is separated into two phrases “进行 示威” and “进行 静坐”.

3.3 Phrase Filtering

In order to alleviate error propagation from word alignments and dependency trees, we apply phrase filtering. Particularly, we adopt three strategies to filter out inappropriate phrases.

Filtering by phrase frequencies: We keep phrases that occur at least t times and discard phrases occurring less than t times to minimize the impact of word alignment and dependency parse errors. Parameter t is tuned on the development set as we harvest new phrases.

Structural filtering: We use syntactic structure information to rule out incomplete phrases. For instance, Chinese phrase “进行了” (carry on sth.) with a phrase structure “VV<asp>AS” is not a complete phrase since it does not have an object. Similarly, we filter out phrases ending with “AS”, “P”, “DEC”, “LC”, “PU”, “CD”, “MSP”⁴.

Filtering by phrase specificity: We keep phrases that are closely related to our topic. Some phrases occur many times during the learning procedure for two reasons. The first reason is that they are closely related to our topic. The second is because they are high-frequency phrases in our corpora. We have observed that some highly frequent and general phrases in our corpora often occur in the learning process, mainly due to word alignment errors or dependency parsing errors. Aiming to learn phrases that are specific event expressions, we define a metric called phrase specificity to avoid bringing in corpora-wide frequent phrases. The metric for phrase p is defined as follows:

$$phrase_specificity(p) = \frac{N_l}{N_c} * 100 \quad (1)$$

where N_l denotes the number of occurrences of phrase p in our projection procedure, N_c denotes the number of occurrences in our entire corpora. This metric measures how close a new phrase is related to the topic of events that we want to detect. If N_l s of two phrases are close to each other, but N_c of one phrase is bigger than that of the other, we deem the phrase with bigger N_c more likely to be a high-frequency phrase. In our experiments, we only consider phrases that have this metric over a certain threshold. We use a development set (section 4.1) to determine the threshold.

3.4 Iterative Projection

We further extend the projection process described in Section 3.1 with phrase decomposition (Section 3.2) and phrase filtering (Section 3.3) to an automatic iterative system. This allows us to use newly learned phrases to learn more new phrases. The most straightforward idea is executing the projection procedure in Section 3.1 many times. However in practice, the growth rate of the number of newly learned phrases is far beyond our imagination. During the iterative projection between the two languages, thousands of incomplete or incorrect phrases are generated. In order to control the growth rate of new phrases and to avoid generating bad phrases, we only keep new phrases that are found at least twice by the iterative system. We deem phrases learned repeatedly more reliable than those occasionally learned. For example, we can learn five different phrases: “举行<dobj>示威”(4)、“示威”(2)、“举行<dobj>游行”(2)、“举行<dobj>活动”(1)、“示威者”(1) when phrase “held<dobj>demonstrations” is mapped to Chinese. The numbers in brackets show the times of phrase learned. According to the strategy, the last two phrases are removed. This is different from filtering by phrase frequencies strategy in Section 3.3. The phrase frequency in section 3.3 means the total times of a new phrase learned by all original phrases. But in the iterative projection, we talk about the frequency of different new phrases learned by one particular original phrase.

4 Experiments

After each structure projection iteration, we appended the newly learned phrases to their corresponding phrase list (EP or PP) and ran the same event recognition evaluation procedure as in H&R’s but with the appended longer phrase lists.

4.1 Data

Our experiment bilingual data consists of 3.57M bilingual sentences from LDC corpora LDC2004E12, LDC2004T08, LDC2005T10, LDC2003E14, LDC2002E18, LDC2005T06, LDC2003E07,

⁴ AS: aspect markers, P: prepositions, DEC: Chinese “的” for relative clauses, LC: localizers, PU: punctuations, CD: cardinal numbers, MSP: some particles.

Method	Phrases	Recall	Precision	F ₁
H&R's Iter #4	EP:623	71	88	79
	PP:569			
Iteration 1	EP:1096	76.2	86.5	81.1
	PP:2219			
Iteration 2	EP:4273	79.2	86.0	82.5
	PP:4597			
Iteration 3	EP:8041	79.2	86.0	82.5
	PP:9169			
Iteration 4	EP:9868	79.2	86.0	82.5
	PP:11705			

Table 1: Results of the projection method using H&R's phrase lists as seed phrases for expansion and projection

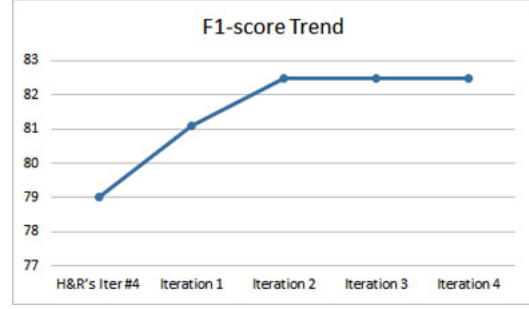


Figure 5: F1-score curve against the number of iterations

LDC2004T07. We ran Giza++ (Och, 2003) and Stanford dependency parser (De Marneffe et al., 2006; Chang et al., 2009) on the parallel sentence pairs to obtain word alignments and dependency trees. In addition, we used the same evaluation method and data as H&R's. The evaluation data contains 400 news articles that were randomly sampled from the English Gigaword Fifth Edition corpora (Parker et al., 2011). Each article contains one of six commonly used civil unrest keywords or their morphological variations. The development set contains 100 documents and the rest 300 documents are used as the test set.

4.2 Event Recognition with Expanded Phrases

We examine the effectiveness of our bilingual structure projection algorithm on the task of event recognition. We choose H&R's best result as our baseline. H&R's multi-faceted event recognition approach achieves the best result after four iterations of bootstrapping.

Our first experiment was designed to expand the EP and PP lists learned by H&R's method at the 4th iteration with our bilingual structure projection system. Our system ran for multiple iterations. According to the development data, the best F_1 -score was achieved after the first two iterations. Table 1 shows the event recognition performance of our bilingual structure projection method. The original multi-faceted event recognition approach at the 4th iteration has achieved a high accuracy (88%) with a relatively low recall (71%). After the first iteration of projection, we obtained an improvement of 5.2 points on recall and 2.1 points on F_1 -score over the baseline. With the newly learned phrases in the first iteration projection, the event recognition recall can be further improved by another 3 points after the second iteration. Overall, with a little loss in precision, the recall has increased by 8.2 points and the F_1 -score 3.5 points. We further observed that results cannot be elevated further after the 2nd iteration even with more phrases, as shown in Figure 5. We conjecture that the reasons are twofold. First, the limited original phrases may not supply more useful phrases after two iterations, which results in a saturated useful phrase list. Second, we do get more useful phrases. However the test data is not large enough so that all newly learned phrases can be found in the test data. Therefore, we cannot see further changes in performance. From the results, we can see the bilingual structure projection algorithm can mine thousands of new phrases. With the newly learned phrases, we can successfully identify additional civil unrest events in the test data.

Due to the noise in H&R's phrase lists (the precision is 88%, indicating 12% noisy phrases) and the features of bootstrap system, phrases learned in previous iterations often have a high precision, but the quality of phrases normally decrease in the succeeding iterations. We further conducted experiments with the phrase lists (EP and PP) learned from the first to the third iteration by H&R's method, which have a high quality. The results are shown in Table 2. In these three iterations, our bilingual structure projection algorithm can improve the recall with almost no loss in precision. This illustrates that our method can recall more phrases and patterns still with a high precision. Note that our method has already outperformed H&R's best result at the third iteration (73.3% in recall and 79.6% in F_1 -score) while H&R's method achieved this performance after the 4th iteration. Therefore, with bilingual structure projection, the number of iterations of the original bootstrapping learning process can be decreased.

phrase iteration	Method	EP numbers	PP numbers	Recall	Precision	F ₁
Iter #1	H&R's Method	145	124	50	88	63
	Bilingual Projection	279	888	53.5	90.0	67.1(+4.1)
Iter #2	H&R's Method	410	356	63	89	74
	Bilingual Projection	790	1387	68.3	88.5	77.1(+3.1)
Iter #3	H&R's Method	504	402	68	88	77
	Bilingual Projection	968	1501	73.3	87.1	79.6(+2.6)

Table 2: Results at the first three iterations

phrase iteration	Method	Recall	Precision	F ₁
TermLex	H&R's Method	66	85	74
	Bilingual Projection	61	81	70
PairLex	H&R's Method	10	91	18
	Bilingual Projection	12	100	21
TermSets	H&R's Method	59	83	69
	Bilingual Projection	57	73	64
PairSets	H&R's Method	68	84	75
	Bilingual Projection	79	78	79
ALLSets	H&R's Method	70	84	76
	Bilingual Projection	78	78	78
Average of Five	H&R's Method	54.6	85.4	62.4
	Bilingual Projection	57.4	82.0	62.4

Table 3: Results of SVM with the bilingual projection method

4.3 SVM Classifiers with Bilingual Structure Projection

H&R also experimented with a suite of supervised classifiers by engineering features based on their learned event dictionaries. In their presented results, supervised classifiers yielded worse event recognition performance than the multi-faceted approach that simply relies on exact match with learned event dictionaries. One guess for this inferior comparison is that their learned phrases are still not diverse and rich enough and their induced feature vectors are too sparse. We have learned many more phrase for both event expressions and the purpose facet through our bilingual structure projection method. We rebuilt the same set of supervised classifiers with the same features. But the features are induced based on the augmented EP lists and PP lists using our bilingual structure projection algorithm. Agent phrase lists (AP) keep the same as H&R's. We ran experiments on five SVM classifiers as shown in Table 3 and performed ten-fold cross validation on the test set, the same as H&R's. All features are binary. We use a vector of 0 and 1 to represent a document. TermLex encodes a binary feature for every phrase in all three phrase lists. PairLex encodes a binary feature for each pair combination from two different lists and requires them to occur in one same sentence. TermSets encodes three binary features for each list, a feature gets 1 when at least one phrase occurs in the document from the corresponding list. PairSets encodes three binary features and each feature represents a combination of two different lists (EP+PP, PP+AG, EP+AG). If any pair occurs in the same sentence, the value gets 1 otherwise 0. Last, the ALLSets encodes 7 binary features, the previous six features plus another binary feature of a sentence containing at least an entry combination from all three lists. Table 3 shows the comparison of our projection method and H&R's method. Although our expanded phrases do not work well on TermLex and TermSets, they still can improve other three classifiers in different degrees. The last row in Table 3 shows the average performance of two methods. Generally, compared to multi-faceted based phrases, our expanded phrases increase the recall, but lower the precision, overall F is the same. Our experiments reconfirm that multi-faceted event dictionary match based event recognition approach, while simple, is more effective than trained supervised classifiers that use dictionary matches as features.

5 Analysis: New Phrases and Structures

We further analyze syntactic structures of the newly learned phrases by bilingual structure projection. Due to linguistic divergences between English and Chinese, various novel new structures are observed in learned Chinese phrases and English phrases.

New Chinese Structures and Examples
NN: 静坐 (stage sit-ins), 怠工 (stop work), 罢工 (went on strike)
VV: 纵火 (set fire), 泄愤 (vent their anger)
VV<rcomp>VV: 纵火 焚烧 (set fire), 进行 绝食 (go on hunger strike)
VV<dobj>NN<conj>NN: 举行 游行 示威 (stage demonstrations), 加入 抗议 罢工 (join the strike and protest)
VV<dobj>NN<relcl>VV: 放火 焚烧 车辆 (set fire to vehicles), 表达 反对 呼声 (express opposition)

Table 4: Examples of new Chinese structures learned

New English Structures and Examples
NN: self-immolation, demonstrations, sit-ins
NN<prep>NN: overuse of force, boycott of elections, disobedience of order
NN<vmod>VBN: rallies held (in), objections expressed (by), rocks thrown (at), disturbances caused (by)
VV: demonstrated, parade
VB<xcomp>VB: cease (to) function, came (out to) demonstrate, pledged (to) support, urge (them to) resign
VB<dobj>NN<conj>NN: held rallies and demonstrations, staged sit-ins and hunger strikes
VB<dobj>NN<prep_of>NN: prevent acts of discrimination, condemned acts of terrorism

Table 5: Examples of new English structures learned

Table 4 shows examples of several new Chinese phrase structures. Interestingly, a multi-word verb phrase in English can be expressed with only one noun or verb word in Chinese, e.g., “went on strike” vs. “罢工” (a noun in Chinese), “vent their anger” vs. “泄愤” (a verb in Chinese). Even more interestingly, we have observed that Chinese tends to put together two coordinate words with roughly the same meaning when referring to one single event, e.g., “staged demonstrations” aligned to “举行 游行 示威” (“游行” and “示威” both mean demonstrations). The reason for putting two words with similar meanings together is to emphasize on the occurrence of the event. More examples are given in Table 4.

Table 5 shows a few examples of new English phrase structures. Dramatically different from the two pre-defined types of verb phrases as specified in H&R’s research, many new phrases are headed by nouns, including individual nouns “sit-ins”, nouns with a prepositional attachment “boycott of elections” and nouns modified by a passive voiced verb phrase “rallies held in”. In addition, we have seen some new verb structures in English phrases that consist of a single verb or a verb with complex objects as shown in Table 5.

6 Related Work

Recent years have witnessed increasing interests in leveraging bilingual corpora or resources to improve performance of monolingual NLP tasks. Generally, The introduction of bilingual corpora or resources serves two purposes. The first purpose is to alleviate the problem that we have few labeled instances in some resource-impovertished languages by a resource-rich language (Hwa et al., 2005; Ganchev et al., 2009; Das and Petrov, 2011; He et al., 2015). The second purpose is to leverage divergences found in different languages to obtain complementary cues (Li et al., 2012; Wang et al., 2013; Che et al., 2013) or extra information (Snyder et al., 2009; Burkett et al., 2010) from another language. Our projection method follows the latter.

In the first purpose, Das and Petrov (2011) explored existing abundant English labeled resources as features to assist building tools for eight European languages. Different to projecting labels as feature, Wang and Manning (2014) proposed a method that projected model expectations as feature for training. He et al. (2015) transferred the sentiment information of a resource-rich language to replenish the lost information of the target language.

In the second purpose, Chen and Ji (2009) proposed a bootstrap framework of co-training among two languages, which uses Chinese event extraction as a case study and bilingual texts as a new source of information. Burkett et al. (2010) attached a bilingual model as a second view (Balcan and Blum, 2005; Ganchev et al., 2012) onto original monolingual models, and used rich features from unannotated bitext to train parameters in bilingual models, which can help to reproduce training data of monolingual model. Che et al. (2013) exploited the complementary cues between two languages as bilingual constraints to help detect errors in a mono-lingual tagger task, which can improve the annotation quality of named entities. Zhu et al. (2013) translated English sentences into Chinese sentences (with the same topic) in ACE 2005 evaluation data with google machine translation system as a second text representation feature

so as to alleviate the data sparseness problem effectively.

Our method is also related to paraphrase learning (Bannard and Callison-Burch, 2005; Callison-Burch, 2008; Zhao et al., 2008; Snover et al., 2009; Ganitkevitch et al., 2013). However, there are two significant differences. First, paraphrase learning translates phrases strictly via word alignments while we use word alignments to find phrase spans on the target language. Second, our purpose is to obtain structured phrases (with syntactic constraints) rather than plain phrases as structured phrases can help us find new phrase structures as shown in Section 3.

7 Conclusion and Future Work

We have presented a bilingual structure projection algorithm that explores structural divergences between languages and can effectively dig up new phrase with various new structures by mapping phrases back and forth across two languages. We combine syntactic information with machine translation technology, not only can reduce the effect of word alignment errors, but also diversify the original two pre-defined event phrase structures. Our experiments show that the newly learned event phrases are capable of recognizing additional event descriptions and considerably increasing the recall of event recognition with minimal loss on precision. Bilingual structural divergences between human languages are common, the proposed bilingual structure projection algorithm is general and can be applied to any pair of languages, and easily extended to the scenario with multiple languages. In addition to event recognition, the proposed structure projection algorithm across languages is potentially useful to many other NLP tasks that utilize extraction patterns by automatically generating novel and diverse phrasal patterns. In our future work, we will attempt to explore the possibility and effect of expanding phrases among other language pairs and other NLP tasks using our bilingual structure projection method.

8 Acknowledgments

The authors were supported by National Natural Science Foundation of China (Grant Nos. 61403269, 61432013 and 61525205) and Natural Science Foundation of Jiangsu Province (Grant No. BK20140355). This research was also partially supported by Ruihong Huang’s startup funds in Texas A&M University. We also thank the anonymous reviewers for their insightful comments.

References

- Maria-Florina Balcan and Avrim Blum. 2005. A pac-style model for learning from labeled and unlabeled data. In *Learning Theory*, pages 111–126. Springer.
- Colin Bannard and Chris Callison-Burch. 2005. Paraphrasing with bilingual parallel corpora. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 597–604. Association for Computational Linguistics.
- David Burkett, Slav Petrov, John Blitzer, and Dan Klein. 2010. Learning better monolingual models with unannotated bilingual text. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, pages 46–54. Association for Computational Linguistics.
- Chris Callison-Burch. 2008. Syntactic constraints on paraphrases extracted from parallel corpora. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 196–205. Association for Computational Linguistics.
- Pi-Chuan Chang, Huihsin Tseng, Dan Jurafsky, and Christopher D Manning. 2009. Discriminative reordering with chinese grammatical relations features. In *Proceedings of the Third Workshop on Syntax and Structure in Statistical Translation*, pages 51–59. Association for Computational Linguistics.
- Wanxiang Che, Mengqiu Wang, Christopher D Manning, and Ting Liu. 2013. Named entity recognition with bilingual constraints. In *HLT-NAACL*, pages 52–62.
- Zheng Chen and Heng Ji. 2009. Can one language bootstrap the other: a case study on event extraction. In *Proceedings of the NAACL HLT 2009 Workshop on Semi-Supervised Learning for Natural Language Processing*, pages 66–74. Association for Computational Linguistics.

- Dipanjan Das and Slav Petrov. 2011. Unsupervised part-of-speech tagging with bilingual graph-based projections. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 600–609. Association for Computational Linguistics.
- Marie-Catherine De Marneffe, Bill MacCartney, Christopher D Manning, et al. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC*, volume 6, pages 449–454.
- 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 and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*, pages 369–377. Association for Computational Linguistics.
- Kuzman Ganchev, Joao Graca, John Blitzer, and Ben Taskar. 2012. Multi-view learning over structured and non-identical outputs. *arXiv preprint arXiv:1206.3256*.
- Juri Ganitkevitch, Benjamin Van Durme, and Chris Callison-Burch. 2013. Ppdb: The paraphrase database. In *HLT-NAACL*, pages 758–764.
- Ryan Georgi, Fei Xia, and William D Lewis. 2012. Measuring the divergence of dependency structures cross-linguistically to improve syntactic projection algorithms. In *LREC*, pages 771–778.
- Mohammad Sadegh Hajmohammadi, Roliana Ibrahim, Ali Selamat, and Hamido Fujita. 2015. Combination of active learning and self-training for cross-lingual sentiment classification with density analysis of unlabelled samples. *Information sciences*, 317:67–77.
- Xiaonan He, Hui Zhang, Wenhan Chao, and De Qing Wang. 2015. Semi-supervised learning on cross-lingual sentiment analysis with space transfer. In *2015 IEEE First International Conference on Big Data Computing Service and Applications (BigDataService)*, pages 371–377.
- Ruihong Huang and Ellen Riloff. 2013. Multi-faceted event recognition with bootstrapped dictionaries. In *HLT-NAACL*, pages 41–51.
- Rebecca Hwa, Philip Resnik, Amy Weinberg, Clara Cabezas, and Okan Kolak. 2005. Bootstrapping parsers via syntactic projection across parallel texts. *Natural language engineering*, 11(03):311–325.
- Qi Li, Haibo Li, Heng Ji, Wen Wang, Jing Zheng, and Fei Huang. 2012. Joint bilingual name tagging for parallel corpora. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 1727–1731. ACM.
- Franz Josef Och. 2003. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pages 160–167. Association for Computational Linguistics.
- Robert Parker, David Graff, Junbo Kong, Ke Chen, and Kazuaki Maeda. 2011. English gigaword fifth edition, june. *Linguistic Data Consortium, LDC2011T07*.
- Matthew G Snover, Nitin Madnani, Bonnie Dorr, and Richard Schwartz. 2009. Ter-plus: paraphrase, semantic, and alignment enhancements to translation edit rate. *Machine Translation*, 23(2-3):117–127.
- Benjamin Snyder, Tahira Naseem, and Regina Barzilay. 2009. Unsupervised multilingual grammar induction. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*, pages 73–81. Association for Computational Linguistics.
- Xiaojun Wan. 2009. Co-training for cross-lingual sentiment classification. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*, pages 235–243. Association for Computational Linguistics.
- Mengqiu Wang and Christopher D Manning. 2014. Cross-lingual pseudo-projected expectation regularization for weakly supervised learning. *Transactions of the Association for Computational Linguistics*, 2:55–66.
- Mengqiu Wang, Wanxiang Che, and Christopher D Manning. 2013. Effective bilingual constraints for semi-supervised learning of named entity recognizers. In *AAAI*. Citeseer.
- Shiqi Zhao, Haifeng Wang, Ting Liu, and Sheng Li. 2008. Pivot approach for extracting paraphrase patterns from bilingual corpora. In *ACL*, volume 8, pages 780–788.
- Zhu Zhu, Shoushan Li, Guodong Zhou, and Rui Xia. 2013. Bilingual event extraction: a case study on trigger type determination.