Bilingual Event Extraction: a Case Study on Trigger Type Determination

Zhu Zhu[†] Shoushan Li[†*] Guodong Zhou[†] Rui Xia[‡]

[†]Natural Language Processing Lab
Soochow University, China
{zhuzhu0020,
shoushan.li}@gmail.com,
gdzhou@suda.edu.cn

[‡]Department of Computer Science
Nanjing University of Science and
Technology
rxia@njust.edu.cn

Abstract

Event extraction generally suffers from the data sparseness problem. In this paper, we address this problem by utilizing the labeled data from two different languages. As a preliminary study, we mainly focus on the subtask of trigger type determination in event extraction. To make the training data in different languages help each other, we propose a uniform text representation with bilingual features to represent the samples and handle the difficulty of locating the triggers in the translated text from both monolingual and bilingual perspectives. Empirical studies demonstrate the effectiveness of the proposed approach to bilingual classification on trigger type determination.

1 Introduction

Event extraction is an increasingly hot and challenging research topic in the natural language processing (NLP) community (Ahn, 2006; Saun et al. 2006; Zhao et al. 2008). It aims to automatically extract certain types of events with the arguments to present the texts under a structured form. In event extraction, there are four primary subtasks, named trigger identification, trigger type determination, argument identification, and argument role determination (Chen and NG, 2012). As an important technology in information extraction, event extraction could be applied to many fields such as information retrieval, summarization, text mining, and question answering.

Recently, the dominative approach to event extraction is based on supervised learning where a set of labeled samples are exploited to train a model to extract the events. However, the availa-ble labeled data are rather sparse due to various kinds of event categories. For example, the event taxonomy in ACE 2005 [1] (Automatic Content Extraction) includes 8 types of events, with 33 subtypes, such as "*Marry/Life*" (subtype/type), and "*Transport/Movement*". Moreover, some subtypes such as "*Nominate/Personnel*" and "*Convict/Justice*" contain less than 10 labeled samples in the English and Chinese corpus respectively. Apparently, such a small scale of training data is difficult to yield a satisfying performance.

One possible way to alleviate the data sparseness problem in event extraction is to conduct bilingual event extraction with training data from two different languages. This is motivated by the fact that labeled data from a language is highly possible to convey similar information in another language. For example, **E1** is an event sample from the English corpus and **E2** is another one in the Chinese corpus. Apparently, **E1** and the English translation text of **E2**, share some important clues such as *meet* and *Iraq* which highly indicates the event type of "*Meet/Contact*".

E1: *Bush arrived in Saint Petersburg on Saturday, when he also briefly met German chancellor Gerhard Schroeder, whose opposition to the Iraq war had soured his relationship with Washington, at a dinner hosted by Putin.*

E2: *美国总统布什将于 2 月访问德国并与施罗德会谈，伊朗和伊拉克问题将是双方会谈的重点。(U.S. president George W. Bush will visit Germany in February and meet with Schroeder, Iran and Iraq will be the focus of the talks the two sides.)*

In this paper, we address the data sparseness problem in event extraction with a bilingual pro-

---

[1]http://www.nist.gov/speech/tests/ace/2005

cessing approach which aims to exploit bilingual training data to enhance the extraction performance in each language. As a preliminary work, we mainly focus on the subtask of trigger type determination. Accordingly, our goal is to design a classifier which is trained with labeled data from two different languages and is capable of classifying the test data from both languages. Generally, this task possesses two main challenges.

The first challenge is text representation, namely, how to eliminate the language gap between the two languages. To tackle this, we first employ Google Translate[2], a state-of-the-art machine translation system, to gain the translation of an event instance, similar to what has been widely done by previous studies in bilingual classification tasks e.g., Wan (2008); Then, we uniformly represent each text with bilingual word features. That is, we augment each original feature vector into a novel one which contains the translated features.

The second challenge is the translation for some specific features. It is well-known that some specific features, such as the triggers and their context features, are extremely important for determining the event types. For example, in **E3**, both trigger "*left*" and named entity "*Saddam*" are important features to tell the event type, i.e., "*Transport/Movement*". When it is translated to Chinese, it is also required to know trigger "*离开*"(*left*) and named entity "*萨达姆*" (*Saddam*) in **E4,** the Chinese translation of **E3.**

E3: *Saddam's clan is said to have **left** for a small village in the desert.*

E4: Chinese translation: *据 说* **萨达姆** *(Saddam) 家族 已经* **离开***(left) 沙漠 中 的 一个 小 村庄。*

However, it is normally difficult to know which words are the triggers and surrounding entities in the translated sentence. To tackle this issue, we propose to locate the trigger from both monolingual and bilingual perspectives in the translation text. Empirical studies demonstrate that adding the translation of these specific features substantially improves the classification performance.

The remainder of this paper is organized as follows. Section 2 overviews the related work on event extraction. Section 3 proposes our approach to bilingual event extraction. Section 4 gives the experimental studies. In Section 5, we conclude our work and give some future work.

2   Related Work

In the NLP community, event extraction has been mainly studied in both English and Chinese.

In English, various supervised learning approaches have been explored recently. Bethard and Martin (2006) formulate the event identification as a classification problem in a word-chunking paradigm, introducing a variety of linguistically motivated features. Ahn (2006) proposes a trigger-based method. It first identifies the trigger in an event, and then uses a multi-classifier to implement trigger type determination. Ji and Grishman (2008) employ an approach to propagate consistent event arguments across sentences and documents. Liao and Grishman (2010) apply document level information to improve the performance of event extraction. Hong et al. (2011) leverage cross-entity information to improve traditional event extraction, regarding entity type consistency as a key feature. More recently, Li et al. (2013) propose a joint framework based on structured prediction which extracts triggers and arguments together.

In Chinese, relevant studies in event extraction are in a relatively primary stage with focus on more special characteristics and challenges. Tan et al. (2008) employ local feature selection and explicit discrimination of positive and negative features to ensure the performance of trigger type determination. Chen and Ji (2009) apply lexical, syntactic and semantic features in trigger labeling and argument labeling to improve the performance. More recently, Li et al. (2012) and Li et al. (2013) introduce two inference mechanisms to infer unknown triggers and recover trigger mentions respectively with morphological structures.

In comparison with above studies, we focus on bilingual event extraction. Although bilingual classification has been paid lots of attention in other fields (Wan 2008; Haghighi et al., 2008; Ismail et al., 2010; Lu et al., 2011；Li et al., 2013), there is few related work in event extraction. The only one related work we find is Ji (2009) which proposes an inductive learning approach to exploit cross-lingual predicate clusters to improve the event extraction task with the main goal to get the event taggers from extra resources, i.e., an English and Chinese parallel corpus. Differently, our goal is to make the la-

---

beled data from two languages help each other without any other extra resources, which is original in the study of event extraction.

3 The Proposed Approach

Trigger type determination aims to determine the event type of a trigger given the trigger and its context (e.g., a sentence). Existing approaches to trigger type determination mainly focus on monolingual classification. Figure 1 illustrates the framework for Chinese and English.

In comparison, our approach exploits the corpora from two different languages. Figure 2 illustrates the framework. As shown in the figure, we first get the translated corpora of Chinese and English origin corpora through machine translation. Then, we represent each text with bilingual features, which enables us to merge the training data from both languages so as to make them help each other.
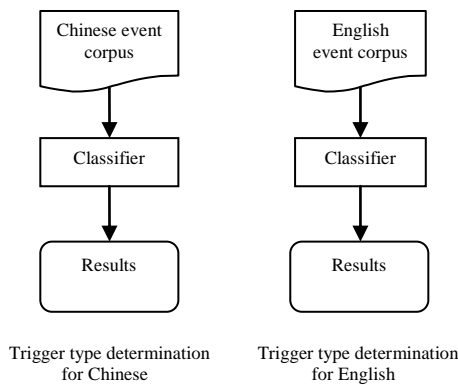

Figure 1: The framework of monolingual classification for trigger type determination
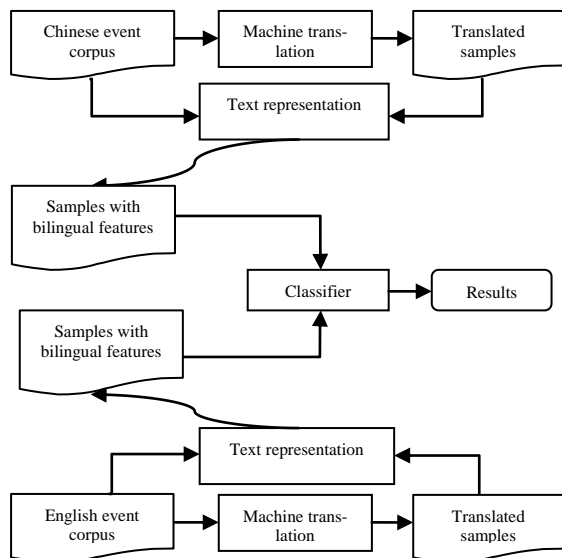

Figure 2: The framework of bilingual classification for trigger type determination

3.1 Text Representation

In a supervised learning approach, labeled data is trained to obtain a classifier. In this approach, the extracted features are the key components to make a successful classifier. Table 1 shows some typical kinds of features in a monolingual classification task for trigger type determination. To better understand these features, the real feature examples in **E3** are given in the table.

Given the feature definition, a monolingual sample x is represented as the combination of all the features, i.e.,

$$x = \begin{pmatrix} e_1, e_2 \ldots, e_n, Tri, POS\_Tri, Tri\_con, \\ POS\_con, Ent, Ent\_type, Ent\_subtype \end{pmatrix} \quad (1)$$

Features	Feature examples in **E3**
All words $(e_1, e_2 \ldots, e_n)$	*Saddam, clan, is, ... , desert*
Trigger (Tri)	*left*
POS of the trigger (POS_Tri)	*VBN*
Trigger's context words (Tri_con)	*...,have, for,...*
POS of trigger's context words (POS_con)	*...,VB,IN,...*
Entities around trigger (Ent)	*Saddam*
Entity type (Ent_type)	*PER*
Entity subtype ($Ent\_subtype$)	*individual*

Table 1: The features and some feature examples for trigger type determination

In bilingual classification, we represent a sample with bilingual features, which makes it possible to train with the data from two languages. To achieve this goal, we employ a single feature augmentation strategy to augment the monolingual features into bilingual features, i.e.,

$$x \Rightarrow x_{Chinese}, x_{English} \quad (2)$$

Specifically, a sample x is represented as follows:

$$x = \begin{pmatrix} \begin{pmatrix} c_1, c_2 \ldots, c_m, Tri_c, POS\_Tri_c, Tri_c\_con, \\ POS\_con, Ent_c, Ent_c\_type, Ent_c\_subtype \end{pmatrix} \\ \begin{pmatrix} e_1, e_2, \ldots, e_n, Tri_e, POS\_Tri_e, Tri_e\_con, \\ POS\_con, Ent_e, Ent_e\_type, Ent_e\_subtype \end{pmatrix} \end{pmatrix} \quad (3)$$

Where the tokens with the '*c*'/'*e*' subscript mean the features generated from the Chinese/English text. From the features, we can see that some

features, such as *Tri_con* and *Ent*, depend on the location of the trigger word. Therefore, locating the trigger in the translated text becomes crucial.

3.2 Locating Translated Trigger

Without loss of generality, we consider the case of translating a Chinese event sample into an English one. Formally, the word sequence of a Chinese event sample is denoted as $s_c = (c_1, c_2, \ldots, c_n)$, while the sequence of the translated one is denoted as $s_e = (e_1, e_2, \ldots, e_m)$. Then, the objective is to get the English trigger Tri_e in s_e, given the Chinese trigger word Tri_c in s_c. The objective function is given as follows:

$$\underset{1 \le k, l \le m}{\arg\max} P\left(e_{k\_l} = Tri_e\right) \qquad (4)$$

Where e_{k_l} denotes the substring $(e_k, e_{k+1} \ldots, e_l)$ in s_e and $1 \le k, l \le m$.

In this paper, the above function could be solved in two perspectives: monolingual and bilingual ones. The former uses the English training data alone to locate the trigger while the latter exploit the bilingual information to get the translated counterpart of the Chinese trigger.

The monolingual perspective: The objective is to locate the trigger with the monolingual information. That is,

$$\underset{1 \le k, l \le m}{\arg\max} P\left(e_{k\_l} = Tri_e \mid s_e, R_e\right) \qquad (5)$$

Where R_e denotes the training resource in English. In fact, this task is exactly the first subtask in event extraction named trigger identification, as mentioned in Introduction. For a simplified implementation, we first estimate the probabilities of $P\left(e_{k\_l} = Tri_e\right)$ in R_e with maximum likelihood estimation when $e_{k\_l} \in s_e$.

The bilingual perspective: The objective is to locate the trigger with the bilingual information. That is,

$$\underset{1 \le k, l \le m}{\arg\max} P\left(e_{k\_l} = Tri_e \mid s_e, s_c, Tri_c\right) \qquad (6)$$

Where Tri_c is the trigger word in Chinese and s_e is the translated text towards s_c. More generally, this can be solved from a standard word alignment model in machine translation (Och et al, 1999; Koehn et al, 2003). However, training a

word alignment requires a huge parallel corpus which is not available here.

For a simplified implementation, we first get the Tri_c's translation，denoted as $trans_{Tri_c}$，with Google Translate. Then, we estimate $P\left(e_{k\_l} = Tri_e\right)$ as follows:

$$P\left(e_{k\_l} = Tri_e\right) = \begin{cases} 0.9 & if \ e_{k\_l} = trans_{Tri_c} \\ \alpha & others \end{cases} \qquad (7)$$

Where 0.9 is an empirical value which makes the translation probability become a dominative factor when the translation of the trigger is found in the translated sentence. α is a small value which makes the sum of all probabilities equals 1.

The final decision is made according to both the monolingual and bilingual perspectives, i.e.,

$$\underset{1 \le k, l \le m}{\arg\max} \ P\left(e_{k\_l} = Tri_e \mid s_e, R_e\right) \\ \cdot P\left(e_{k\_l} = Tri_e \mid s_e, s_c, Tri_c\right) \qquad (8)$$

Note that we reduce the computational cost by make the word length of the trigger less than 3, i.e., $l - k \le 3$.

4 Experimentation

4.1 Experimental Setting

Data sets: The Chinese and English corpus for even extraction are from ACE2005, which involves 8 types and 33 subtypes. All our experiments are conducted on the subtype case. Due to the space limit, we only report the statistics for each type, as shown in Table 2. For each subtype, 80% samples are used as training data while the rest are as test data.

#	Chinese	English	*total*
Life	389	902	*1291*
Movement	593	679	*1272*
Transaction	147	379	*526*
Business	144	137	*281*
Conflict	514	1629	*2143*
Contact	263	373	*636*
Personnel	203	514	*717*
Justice	457	672	*1129*
total	*2710*	*5285*	*7995*

Table 2: Statistics in each event type in both Chinese and English data sets

Features: The features have been illustrated in Table 1 in Section 3.2.

Classification algorithm: The maximum entropy (ME) classifier is implemented with the public tool, Mallet Toolkits[3] .

Evaluation metric: The performance of event type recognition is evaluated with F-score.

4.2 Experimental Results

In this section, we evaluate the performance of our approach to bilingual classification on trigger type determination. For comparison, following approaches are implemented:

➢ **Monolingual**: perform monolingual classification on the Chinese and English corpus individually, as shown in Figure 1.

➢ **Bilingual**: perform bilingual classification with partial bilingual features, ignoring the context features (e.g., context words, context entities) under the assumption that the trigger location task is not done.

➢ **Bilingual_location**: perform bilingual classification by translating each sample into another language and using a uniform representation with all bilingual features as shown in Section 3.2. This is exactly our approach. The number of the context words and entities before or after the trigger words is set as 3.
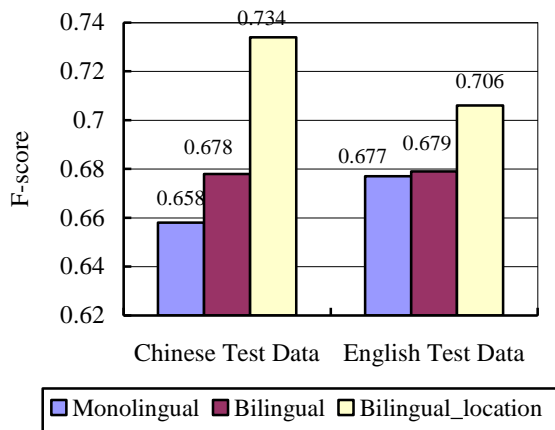

Figure 3: Performance comparison of the three approaches on the Chinese and English test data

Figure 3 shows the classification results of the three approaches on the Chinese and English test data. From this figure, we can see that **Bilingual_location** apparently outperform **Monolingual**, which verifies the effectiveness of using bilingual corpus. Specifically, the improvement by our approach in Chinese is impressive, reaching 7.6%. The results also demonstrate the importance of the operation of the trigger location,

without which, bilingual classification can only slightly improve the performance, as shown in the English test data.

The results demonstrate that our bilingual classification approaches are more effective for the Chinese data. This is understandable because the size of English data is much larger than that of Chinese data, 5285 vs. 2710, as shown in Table 2. Specifically, after checking the results in each subtype, we find that some subtypes in Chinese have very few samples while corresponding subtypes in English have a certain number samples. For example, the subtype of "*Elect/Personnel*" only contains 30 samples in the Chinese data while 161 samples can be found in the English data, which leads a very high improvement (15.4%) for the Chinese test data. In summary, our bilingual classification approach provides an effective way to handle the data sparseness problem in even extraction.

5 Conclusion and Future Work

This paper addresses the data sparseness problem in event extraction by proposing a bilingual classification approach. In this approach, we use a uniform text representation with bilingual features and merge the training samples from both languages to enlarge the size of the labeled data. Furthermore, we handle the difficulty of locating the trigger from both the monolingual and bilingual perspectives. Empirical studies show that our approach is effective in using bilingual corpus to improve monolingual classification in trigger type determination.

Bilingual event extraction is still in its early stage and many related research issues need to be investigated in the future work. For example, it is required to propose novel approaches to the bilingual processing tasks in other subtasks of event extraction. Moreover, it is rather challenging to consider a whole bilingual processing framework when all these subtasks are involved together.

---

[3] http://mallet.cs.umass.edu/

References

Ahn D. 2006. The Stages of Event Extraction. In *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*, pp.1~8.

Bethard S. and J. Martin. 2006. Identification of Event Mentions and Their Semantic Class. In *Proceedings of EMNLP-2006*, pp.146-154.

Chen C. and V. NG. 2012. Joint Modeling for Chinese Event Extraction with Rich Linguistic Features. In *Proceedings of COLING-2012*, pp. 529-544.

Chen Z. and H. Ji. 2009. Language Specific Issue and Feature Exploration in Chinese Event Extraction. In *Proceedings of NAACL-2009*, pp. 209-212.

Haghighi A., P. Liang, T. Berg-Kirkpatrick and D. Klein. 2008. Learning Bilingual Lexicons from Monolingual Corpora. In *Proceedings of ACL-2008*, pp. 771-779.

Hong Y., J. Zhang., B. Ma., J. Yao., and G. Zhou. 2011. Using Cross-Entity Inference to Improve Event Extraction. In *Proceedings of ACL-2011*, pp. 1127–1136.

Ismail A., and S. Manandhar. 2010. Bilingual Lexicon Extraction from Comparable Corpora Using In-domain Terms. In *Proceedings of COLING-2010*, pp.481-489.

Ji H. 2009. Cross-lingual Predicate Cluster Acquisition to Improve Bilingual Event Extraction by Inductive Learning. In *Proceedings of the Workshop on Unsupervised and Minimally Supervised Learning of Lexical Semantics*, pp. 27-35.

Ji H, and R. Grishman. 2008. Refining Event Extraction through Cross-Document Inference. In *Proceedings of ACL-2008*, pp. 254-262.

Koehn P., F. Och, and D. Marcu. 2003. Statistical Phrase-based Translation. In *Proceedings of HTL-NAACL-2003*, pp. 127-133.

Li P., and G. Zhou. 2012. Employing Morphological Structures and Sememes for Chinese Event Extraction. In *Proceedings of COLING-2012*, pp. 1619-1634.

Li P., Q. Zhu and G. Zhou. 2013. Using Compositional Semantics and Discourse Consistency to Improve Chinese Trigger Identification. In *Proceedings of COLING-2013*, pp. 399-415.

Li Q, H Ji, and H. Liang. 2013. Joint Event Extraction via Structured Prediction with Global Features. In *Proceedings of ACL-2013*, pp. 73-82.

Li S, R Wang, H Liu, and CR Huang. 2013. Active Learning for Cross-Lingual Sentiment Classification. In *Proceedings of Natural Language Processing and Chinese Computing*, pp. 236-246.

Liao S and R. Grishman. 2010. Using Document Level Cross-event Inference to Improve Event Extraction. In *Proceedings of ACL-2010*, pp. 789-797.

Lu B., C. Tan, C. Cardie and B. K. Tsou. 2011. Joint Bilingual Sentiment Classification with Unlabeled Parallel Corpora. In *Proceedings of ACL-2011*, pp. 320-330.

Och F., C. Tillmann, and H. Ney. 1999. Improved Alignment Models for Statistical Machine Translation. In *Proceedings of EMNLP-1999*, pp.20-28.

Tan H., T. Zhao, and J. Zheng. 2008. Identification of Chinese Event and Their Argument Roles. In *Proceedings of CITWORKSHOPS-2008*, pp. 14-19.

Wan X. 2008. Using Bilingual Knowledge and Ensemble Techniques for Unsupervised Chinese Sentiment Analysis. In *Proceedings of EMNLP-2008*, pp. 553-561.

Zhao Y., Y. Wang, B. Qin, et al. 2008. Research on Chinese Event Extraction. In *Proceedings of Journal of Chinese Information, 22(01)*, pp. 3-8.