Cross-lingual Predicate Cluster Acquisition to Improve Bilingual Event Extraction by Inductive Learning

Heng Ji
Computer Science Department
Queens College and The Graduate Center
The City University of New York
`hengji@cs.qc.cuny.edu`

Abstract

In this paper we present two approaches to automatically extract cross-lingual predicate clusters, based on bilingual parallel corpora and cross-lingual information extraction. We demonstrate how these clusters can be used to improve the NIST Automatic Content Extraction (ACE) event extraction task[1]. We propose a new *inductive learning* framework to automatically augment background data for low-confidence events and then conduct global inference. Without using any additional data or accessing the baseline algorithms this approach obtained significant improvement over a state-of-the-art bilingual (English and Chinese) event extraction system.

1 Introduction

Event extraction, the 'classical' information extraction (IE) task, has progressed from Message Understanding Conference (MUC)-style single template extraction to the more comprehensive multi-lingual Automatic Content Extraction (ACE) extraction including more fine-grained types. This extension has made event extraction more widely applicable in many NLP tasks including cross-lingual document retrieval (Hakkani-Tur et al., 2007) and question answering (Schiffman et al., 2007). Various supervised learning approaches

have been explored for ACE multi-lingual event extraction (e.g. Grishman et al., 2005; Ahn, 2006; Hardy et al., 2006; Tan et al., 2008; Chen and Ji, 2009). All of these previous literatures showed that one main bottleneck of event extraction lies in low recall. It's a challenging task to recognize the different forms in which an event may be expressed, given the limited amount of training data. The goal of this paper is to improve the performance of a bilingual (English and Chinese) state-of-the-art event extraction system without accessing its internal algorithms or annotating additional data.

As for a separate research theme, extensive techniques have been used to produce word clusters or paraphrases from large unlabeled corpora (Brown et al., 1990; Pereira et al., 1993; Lee and Pereira, 1999, Barzilay and McKeown, 2001; Lin and Pantel, 2001; Ibrahim et al., 2003; Pang et al., 2003). For example, (Bannard and Callison-Burch, 2005) and (Callison-Burch, 2008) described a method to extract paraphrases from largely available bilingual corpora. The resulting clusters contain words with similar semantic information and therefore can be useful to augment a small amount of annotated data. We will automatically extract cross-lingual predicate clusters using two different approaches based on bilingual parallel corpora and cross-lingual IE respectively; and then use the derived clusters to improve event extraction.

We propose a new learning method called *inductive learning* to exploit the derived predicate clusters. For each test document, a background document is constructed by gradually replacing the low-confidence events with the predicates in the same cluster. Then we conduct cross-document inference technique as described in (Ji and Grish-

---

27

man, 2008) to improve the performance of event extraction. This inductive learning approach matches the procedure of human knowledge acquisition and foreign language education: analyze information from specific examples and then discover a pattern or draw a conclusion; attempt synonyms to convey/learn the meaning of an intricate word.

The rest of this paper is structured as follows. Section 2 describes the terminology used in this paper. Section 3 presents the overall system architecture and the baseline system. Section 4 then describes in detail the approaches of extracting cross-lingual predicate clusters. Section 5 describes the motivations of using cross-lingual clusters to improve event extraction. Section 6 presents an overview of the inductive learning algorithm. Section 7 presents the experimental results. Section 8 compares our approach with related work and Section 9 then concludes the paper and sketches our future work.

2   Terminology

The event extraction task we are addressing is that of ACE evaluations. ACE defines the following terminology:

entity: an object or a set of objects in one of the semantic categories of interest
mention: a reference to an entity (typically, a noun phrase)
event trigger: the main word which most clearly expresses an event occurrence
event arguments: the mentions that are involved in an event (participants)
event mention: a phrase or sentence within which an event is described, including trigger and arguments

The 2005 ACE evaluation had 8 types of events, with 33 subtypes; for the purpose of this paper, we will treat these simply as 33 distinct event types. For example, for a sentence "*Barry Diller on Wednesday quit as chief of Vivendi Universal Entertainment*", the event extractor should detect all the following information: a "*Personnel_End-Position*" event mention, with "*quit*" as the trigger word, "*chief*" as an argument with a *role* of "position", "*Barry Diller*" as the person who quit the position, "*Vivendi Universal Entertainment*" as the

organization, and the time during which the event happened is "*Wednesday*".

3   Approach Overview

3.1   System Pipeline

Figure 1 depicts the general procedure of our approach. The set of test event mentions is improved by exploiting cross-lingual predicate clusters.
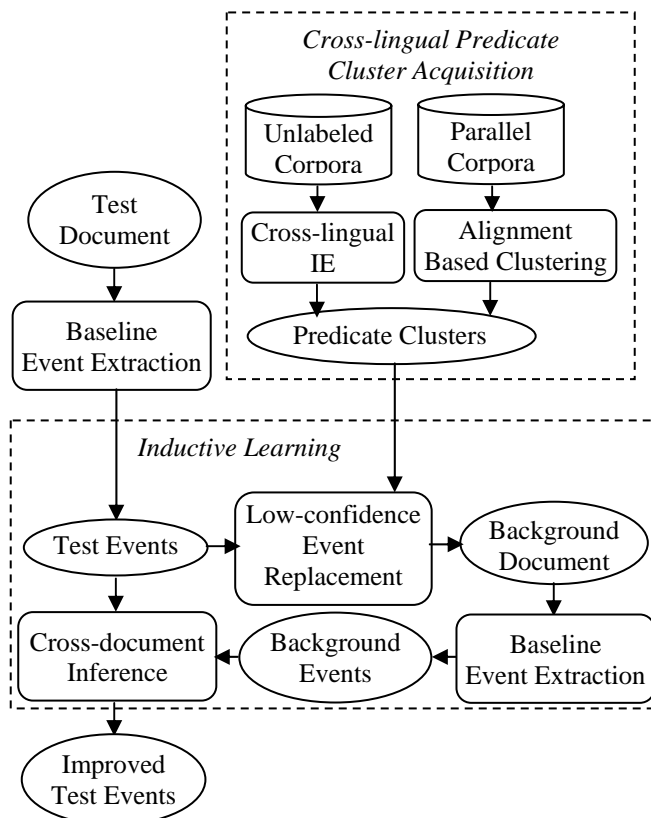

Figure 1. System Overview

The following section 3.2 will give more details about the baseline bilingual event tagger. Then we will present the predicate cluster acquisition algorithm in section 4 and the method of exploiting clusters for event extraction in section 6.

3.2   A Baseline Bilingual Event Extraction System

We use a state-of-the-art bi-lingual event extraction system (Grishman et al., 2005; Chen and Ji, 2009) as our baseline. The system combines pattern matching with a set of Maximum Entropy classifiers: to distinguish events from non-events;

to classify events by type and subtype; to distinguish arguments from non-arguments; to classify arguments by argument role; and given a trigger, an event type, and a set of arguments, to determine whether there is a reportable event mention. In addition, the Chinese system incorporates some language-specific features to address the problem of word segmentation (Chen and Ji, 2009).

4 Cross-lingual Predicate Cluster Acquisition

We start from two different approaches to extract cross-lingual predicate clusters, based on parallel corpora and cross-lingual IE techniques respectively.

4.1 Acquisition from Bilingual Parallel Corpora

In the first approach, we take use of the 852 Chinese event trigger words in ACE05 training corpora as our 'anchor set'. For each Chinese trigger, we search its automatically aligned English words from a Chinese-English parallel corpus including 50,000 sentence pairs (part of Global Autonomous Language Exploitation Y3 Machine Translation training corpora) to construct an English predicate cluster. The word alignment was obtained by running Giza++ (Och and Ney, 2003). In each cluster we record the frequency of each unique English word. Then we conduct the same procedure in the other direction to construct Chinese predicate clusters anchored by English triggers.

State-of-the-art Chinese-English word alignment error rate is about 40% (Deng and Byrne, 2005). Therefore the resulting cross-lingual clusters include a lot of word alignment errors. In order to address this problem, we filter the clusters by only keeping those predicates including the original predicate forms in ACE training data or English/Chinese Propbank (Palmer et al., 2005; Xue and Palmer, 2009).

4.2 Acquisition from Cross-lingual IE

Based on the intuition that Machine Translation (MT) may translate a Chinese trigger word into different English words in different contexts, we employ the second approach using cross-lingual IE techniques (Hakkani-Tur et al., 2007) on TDT5 Chinese corpus to generate more clusters. We apply the following two cross-lingual IE pipelines:

Chinese IE_MT: Apply Chinese IE on the Chinese texts to get a set of Chinese triggers *ch-trigger-set1*, and then use word alignments to translate (project) *ch-trigger-set1* into a set of English triggers *en-trigger-set1*;

MT_English IE: Translate Chinese texts into English, and then apply English IE on the translated texts to get a set of English triggers *en-trigger-set2*.

For any Chinese trigger *ch-trigger* in *ch-trigger-set1*, if its corresponding translation *en-trigger* in *en-trigger-set1* is the same as that in *en-trigger-set2*, then we add *en-trigger* into the cluster anchored by *ch-trigger*.

We apply the English and Chinese IE systems as described in (Grishman et al., 2005; Chen and Ji, 2009). Both cross-lingual IE pipelines need machine translation to translate Chinese documents (for English IE) or project the extraction results from Chinese IE into English. We use the RWTH Aachen Chinese-to-English statistical phrase-based machine translation system (Zens and Ney, 2004) for these purposes.

4.3 Derived Cross-lingual Predicate Clusters

Applying the above two approaches we obtained 438 English predicate clusters and 543 Chinese predicate clusters.

For example, for a trigger "伤(injure)", we can get the following two predicate clusters with their frequency in the parallel corpora:

伤→{injured:99 injuries:96 injury:76
 wounded:38 wounding:28 injuring:14 wounds:7
killed:4 died:2 mutilated:1 casualties:1 chop:1 killing:1 shot:1}.

injured → {受伤:1624 重伤:102 伤:99 轻伤:29 伤势:23 炸:12 打伤:10 爆炸:6 伤害:3 死亡:2 冲突:1 亡:1 烫伤:1 损失:1 出席:1 登陆:1 致残:1 自残:1 }

We can see that the predicates in the same cluster are not restrictedly synonyms, but they were generated as alternative translations for the same word and therefore represent similar meanings. More importantly, these triggers vary from very common ones such as 'injured' to rare words such as 'mutilate'. This indicates how these clusters can aid extracting low-confidence events: when deciding whether a word 'mutilate' indicates a "Life-

Injure" event in a certain context, we can replace it with other predicates in the same cluster and may provide us more reliable overall evidence.

Figure 2 presents the distribution of clusters which include more than one predicate.
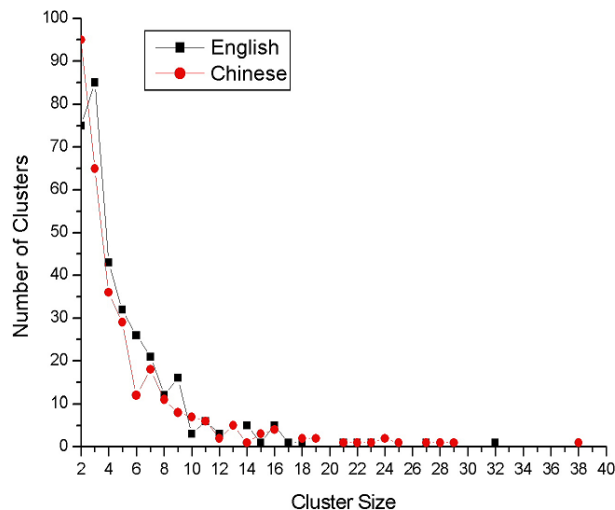

Figure 2. Cluster Size Distribution

We can see that most clusters include 2-9 predicates in both English and Chinese. However on average English clusters include more predicates. In addition, there are many more singletons in Chinese (232) than in English (101). This indicates that Chinese event triggers are more ambiguous.

5 Motivation of Using Cross-lingual Clusters for Event Extraction

After extracting cross-lingual predicate clusters, we can combine the evidence from all the predicates in each cluster to adjust the probabilities of event labeling. In the following we present some examples in both languages to demonstrate this motivation.

5.1 Improve Rare Trigger Labeling

Due to the limited training data, many trigger words only appear a few times as a particular type of event. This data sparse problem directly leads to the low recall of trigger labeling. But exploiting the evidence from other predicates in the same cluster may boost the confidence score of the candidate event. We present two examples as follows.

(1) English Example 1

For example, "*blown up*" doesn't appear in the training data as a "Conflict-Attack" event, and so it cannot be identified in the following test sentence. However, if we replace it with other predicates in the same cluster, the system can easily identify 'Conflict-Attack' events in the new sentences with high confidence values:

(a) Test Sentence:
Identified as "Conflict-Attack" Event with Confidence=0:

*He told AFP that Israeli intelligence had been dealing with at least 40 tip-offs of impending attacks when the Haifa bus was **blown up**.*

(b) Cross-lingual Cluster
炸毁 → *{ **blown up**:4 bombing:3 blew:2 destroying:1 destroyed:1 }*

(c) Replaced Sentences
Identified as "Conflict-Attack" Event with Confidence=0.799:

*He told AFP that Israeli intelligence had been dealing with at least 40 tip-offs of impending attacks when the Haifa bus was **destroyed**.*
    …

(2) Chinese Example 1

Chinese predicate clusters anchored by English words can also provide external evidence for event identification. For example, the trigger word "假释 (release/parole)" appears rarely in the Chinese training data but in most cases it can be replaced by a more frequent trigger "释放(release)" to represent the same meaning. Therefore by combining the evidence from "释放" we can enhance the confidence value of identifying "假释" as a "Justice-Release_Parole" event. For example,

(a) Test Sentence:
Identified as "Justice-Release_Parole" Event with Confidence=0:

这名嫌犯因为侵害案件**假释**出狱却又犯下了重罪。。 *(This suspect was released because of the violation case but committed a felony again.)*

(b) Cross-lingual Cluster
releasing → *{假释:4 释放:1 }*

(c) Replaced Sentences
Identified as "Justice-Release_Parole" Event with Confidence=0.964:

*这名嫌犯因为侵害案件**释放**出狱却又犯下了重罪.*
…

5.2 Improve Frequent Trigger Labeling

On the other hand, some common words are highly ambiguous in particular contexts. But the other less-ambiguous predicates in the clusters can help classify event types more accurately.

(1) English Example 2

For example, in the following sentence the "*Personnel-End_Position*" event is missing because "*step*" doesn't indicate any ACE events in the training data. However, after replacing "step" with other prediates such as "quit", the system can identify the event more easily:

(a) Test Sentence:
Identified as "Personnel-End_Position" Event with Confidence=0:

*Barry Diller on Wednesday **step** from chief of Vivendi Universal Entertainment, the entertainment unit of French giant Vivendi Universal.*

(b) Cross-lingual Cluster
下台 → *{ resign:6 **step**:5 quit:3}*

(c) Replaced Sentences
Classified as "Personnel-End_Position" Event with Confidence=0.564:

*Barry Diller on Wednesday **quit** from chief of Vivendi Universal Entertainment, the entertainment unit of French giant Vivendi Universal.*
…

(2) Chinese Example 2

Some single-character Chinese predicates can represent many different event types in different contexts. For example, the word "打" appears in 27 different predicate clusters, representing the meaning of hit/call/strike/form/take/draw etc. Therefore we can take use of other less ambiguous predicates in these clusters to adjust the likelihood of event classification.

For example, in the following test sentence, the word "打" indicates two different event types. If we replace these words with other predicates, we can classify them into different event types more accurately based on the evidence from replaced predicates and contexts.

(a) Test Sentence:
Event Classification for trigger word "打":

*就在几天前船长紧急**打** ("call", Phone-Write event with confidence 0) 电报求救，表示轮机长蔡明志已经在 10 天前被大陆渔工**打**("attacked/killed", Conflict-Attack event with confidence 0.528)死，自己也被殴**打**("attacked", Conflict-Attack event with confidence 0.946)，连人带船胁持到大陆。(Several days ago the Captain **called** urgent telegraphs to ask for help, expressing that the boat pilot Cai Mingzhi was already **killed** by mainland fishermen and he himself was **assaulted** and duressed to the mainland.)*

(b) Cross-lingual Cluster

call→*{ 打电话:6 电话:6 **打**:1 拨打:1 }*

attack→*{ 袭击:564 进攻:110 攻击:114 打击:24 反击:15 爆炸:15 突袭:15 击:8 偷:6 围攻:6 身亡:5 行凶:4 战争:3 死亡:3 丧生:2 谋杀:2 死:2 轰炸:2 侵略:2 入侵:2 设立:1 出兵:1 推翻:1 打死:1 劫持:1 **打**:1 遇害:1 咬:1 }*

(c) Replaced Sentences
Event Classification for trigger word "打" with higher confidence:

*就在几天前船长紧急**拨打** ("call", Phone-Write event with confidence 0.938) 电报求救，表示轮机长蔡明志已经在 10 天前被大陆渔工杀 ("attacked/killed", Conflict-Attack event with confidence 0.583)死，自己也被**袭击**("attacked", Conflict-Attack event with confidence 0.987)，连人带船胁持到大陆。*
…

Based on the above motivations we propose to incorporate cross-lingual predicate clusters to refine event identification and classification. In order

to exploit these clusters effectively, we shall generate additional background data and conduct global confidence. The sections below will present the detailed algorithms.

6 Inductive Learning

We design a framework of inductive learning to incorporate the derived predicate clusters. The general idea of inductive learning is to analyze information from all kinds of specific examples until we can draw a conclusion. Since the main goal of our approach is to improve the recall of event extraction, we shall focus on those events generated by the baseline tagger with low confidence. For those events we automatically generate background documents using the predicate clusters (details in section 6.1) and then conduct global inference between each test document and its background documents (section 6.2).

6.1 Background Document Generation

For each event mention in a test document, the baseline event tagger produces the following local confidence value:

- *LConf(trigger, etype)*: The probability of a string *trigger* indicating an event mention with type *etype* in a context sentence *S*;

If *LConf(trigger, etype)* is lower than a threshold, and it belongs to a predicate cluster *C*, we create an additional background document *BD* by:

- For each *predicate_i ∈ C*, we replace *trigger* with *predicate_i* in *S* to generate new sentence *S'*, and add *S'* into *BD*.

6.2 Global Inference

For each background document *BD*, we apply the baseline event extraction and get a set of background events. We then apply the cross-document inference techniques as described in (Ji and Grishman, 2008) to improve trigger and argument labeling performance by favoring interpretation consistency across the test events and background events.

This approach is based on the premise that many events will be reported multiple times from different sources in different forms. This naturally occurs in the test document and the background document because they include triggers from the same predicate cluster.

By aggregating events across each pair of test document *TD* and background document *BD,* we conduct the following statistical global inference:

- to remove triggers and arguments with low confidence in *TD* and *BD*;
- to adjust trigger and argument identification and classification to achieve consistency across *TD* and *BD*.

In this way we can propagate highly consistent and frequent triggers and arguments with high global confidence to override other, lower confidence, extraction results.

7 Experimental Results

7.1 Data and Scoring Metric

We used ACE2005 English and Chinese training corpora to evaluate our approach. Table 1 shows the number of documents used for training, development and blind testing.

Language	Training Set	Development Set	Test Set
English	525	33	66
Chinese	500	10	40

Table 1. Number of Documents

We define the following standards to determine the *correctness* of an event mention:

- *A trigger is correctly identified* if its position in the document matches a reference trigger.
- *A trigger is correctly identified and classified* if its event type and position in the document match a reference trigger.
- *An argument is correctly identified* if its event type and position in the document match any of the reference argument mentions.
- *An argument is correctly identified and classified* if its event type, position in the document, and role match any of the reference argument mentions.

Performance Language/System		Trigger Identification +Classification			Argument Identification			Argument Classification Accuracy	Argument Identification +Classification		
		P	R	F	P	R	F		P	R	F
English	Baseline	67.8	53.5	59.8	49.3	31.4	38.3	88.2	43.5	27.7	33.9
	After Using Cross-lingual Predicate Clusters	**69.2**	**59.4**	**63.9**	**51.7**	**32.7**	**40.1**	**89.6**	**46.3**	**29.3**	**35.9**
Chinese	Baseline	58.1	47.2	52.1	46.2	33.7	39.0	95.0	43.9	32.0	37.0
	After Using Cross-lingual Predicate Clusters	**60.2**	**52.6**	**56.1**	**46.8**	**36.7**	**41.1**	**95.6**	**44.7**	**35.1**	**39.3**

Table 2. Overall Performance on Blind Test Set (%)

7.2 Confidence Metric Thresholding

Before blind testing we select the thresholds for the trigger confidence *LConf(trigger, etype)* as defined in section 6.1 by optimizing the F-measure score of on the development set. Figure 3 shows the effect on precision and recall of varying the threshold for inductive learning using cross-lingual predicate clusters.
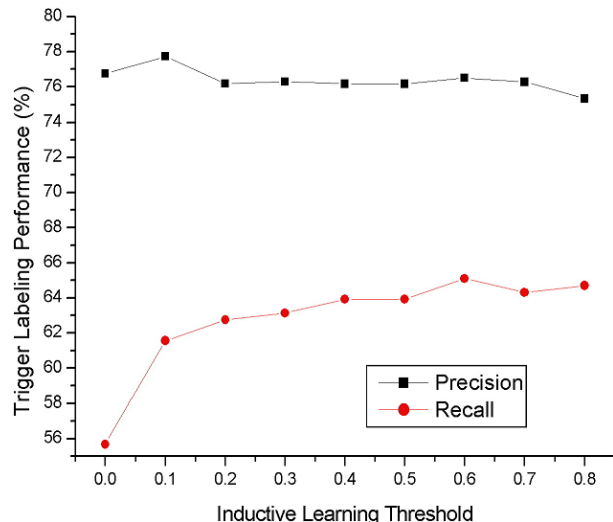

Figure 3. Trigger Labeling Performance with Inductive Learning Confidence Thresholding on English Development Set

We can see that the best performance on the development set can be obtained by selecting threshold 0.6, achieving 9.4% better recall with a little loss in precision (0.26%) compared to the baseline (with threshold=0) . Then we apply this threshold value directly for blind test. This optimizing procedure is repeated for Chinese as well.

7.3 Overall Performance

Table 2 shows the overall Precision (P), Recall (R) and F-Measure (F) scores for the blind test set.

For both English and Chinese, the inductive learning approach using cross-lingual predicate clusters provided significant improvement over the baseline event extraction system (about 4% absolute improvement on trigger labeling and 2%-2.3% on argument labeling). The most significant gain was provided for the recall of trigger labeling – 5.9% absolute improvement for English and 5.4% absolute improvement for Chinese.

Surprisingly this approach didn't cause any loss in precision. In fact small gains were obtained on precision for both languages. This indicates that cross-lingual predicate clusters are effective at adjusting the confidence values so that the events were not over-generated. The refined event trigger labeling also directly yields better performance in argument labeling.

We conducted the Wilcoxon Matched-Pairs Signed-Ranks Test on a document basis. The results show that for both languages the improvement using cross-lingual predicate clusters is significant at a 99.7% confidence level for trigger labeling and a 96.4% confidence level for argument labeling.

7.4 Discussion

For comparison we attempted a self-training approach: adding high-confidence events in the test set back as additional training data and re-train the event tagger. This produced 1.7% worse F-measure score for the English development set. It further

proves that using the test set itself is not enough, we need to explore new predicates to serve as background evidence.

In addition we also applied a bootstrapping approach using relevant unlabeled data and obtained limited improvement – about 1.6% F-measure gain for English. As Ji and Grishman (2006) pointed out, both self-training and bootstrapping methods require good data selection scheme. But not for any test set we can easily find relevant unlabeled data. Therefore the approach presented in this paper is less expensive – we can automatically generate background data while introducing new evidence.

An alternative way of incorporating the cross-lingual predicate clusters would follow (Miller et al., 2004), namely encoding the cluster membership as an additional feature in the supervised-learning procedure of the baseline event tagger. However in the situation where we cannot directly change the algorithms of the baseline system, our approach of inductive learning is more flexible.

8   Related Work

Our approach of extracting predicate clusters is related to some prior work on paraphrase or word cluster discovery, either from mono-lingual parallel corpora (e.g. Barzilay and McKeown, 2001; Lin and Pantel, 2001; Ibrahim et al., 2003; Pang et al., 2003) or cross-lingual parallel corpora (e.g. Bannard and Callison-Burch, 2005; Callison-Burch, 2008). Shinyama and Sekine (2003) presented an approach of extracting paraphrases using names, dates and numbers as anchors. Hasegawa et al. (2004) described a paraphrase discovery approach based on clustering concurrent name pairs.

Several recent studies have stressed the benefits of using paraphrases or word clusters to improve IE components. For example, (Miller et al., 2004) proved that word clusters can significantly improve English name tagging. The idea of using predicates in the same cluster for candidate trigger replacement is similar to Ge et al.(1998) who used local context replacement for pronoun resolution. To the best of our knowledge, our work presented the first experiment of using cross-lingual predicate paraphrases for the ACE event extraction task.

9   Conclusion and Future Work

In this paper we described two approaches to extract cross-lingual predicate clusters, and designed a new inductive learning framework to effectively incorporate these clusters for event extraction. Without using any additional data or changing the baseline algorithms, we demonstrated that this method can significantly enhance the performance of a state-of-the-art bilingual event tagger.

We have noticed that the current filtering scheme based on Propbank may be too restricted to keep enough informative predicates. In the future we will attempt incorporating POS tagging results and frequency information.

In addition we will extend this framework to extract cross-lingual relation and name clusters to improve other IE tasks such as name tagging, relation extraction, event coreference and event translation. We are also interested in automatically discovering new event types (non-ACE event types) or more fine-grained subtypes/attributes for existing ACE event types from the derived predicate clusters.

References

David Ahn. 2006. The stages of event extraction. *Proc. COLING/ACL 2006 Workshop on Annotating and Reasoning about Time and Events*. Sydney, Australia.

Colin Bannard and Chris Callison-Burch. 2005. Paraphrasing with Bilingual Parallel Corpora. *Proc. ACL 2005*.

Regina Barzilay and Kathleen McKeown. 2001. Extracting Paraphrases from a Parallel Corpus. *Proc. ACL 2001*.

Peter F. Brown, Vinvent J. Della pietra, Peter V. deSouza, Jenifer C. Lai, Robert L. Mercer. 1990. Class-based N-gram Models of Natural Language. *Computational Linguistics*.

Chris Callison-Burch. 2008. Syntactic Constraints on Paraphrases Extracted from Parallel Corpora. *Proc. EMNLP 2008*. Honolulu, USA.

Zheng Chen and Heng Ji. 2009. Language Specific Issue and Feature Exploration in Chinese Event Extraction. *Proc. HLT-NAACL 2009*. Boulder, Co.

Yonggang Deng and William Byrne. 2005. HMM Word and Phrase Alignment for Statistical Machine Translation. *Proc. HLT-EMNLP 2005*. Vancouver, Cananda.

Niyu Ge, John Hale and Eugene Charniak. 1998. A Statistical Approach to Anaphora Resolution. Proc. Sixth Workshop on Very Large Corpora

Ralph Grishman, David Westbrook and Adam Meyers. 2005. NYU's English ACE 2005 System Description. *Proc. ACE 2005 Evaluation Workshop*. Washington, US.

Dilek Hakkani-Tur, Heng Ji and Ralph Grishman. 2007. Using Information Extraction to Improve Cross-lingual Document Retrieval. *Proc. RANLP2007 workshop on Multi-source, Multilingual Information Extraction and Summarization*.

Hilda Hardy, Vika Kanchakouskaya and Tomek Strzalkowski. 2006. Automatic Event Classification Using Surface Text Features. *Proc. AAAI06 Workshop on Event Extraction and Synthesis*. Boston, Massachusetts. US.

Takaaki Hasegawa, Satoshi Sekine and Ralph Grishman. 2004. Discovering Relations among Named Entities from Large Corpora. *Proc. ACL 2004*. Barcelona, Spain.

Ali Ibrahim, Boris Katz and Jimmy Lin. 2003. Extracting Structural Paraphrases from Aligned Monolingual Corpora. *Proc. ACL 2003*.

Heng Ji and Ralph Grishman. 2006. Data Selection in Semi-supervised Learning for Name Tagging. *Proc. ACL 2006 Workshop on Information Extraction Beyond the Document*. Sydney, Australia.

Heng Ji and Ralph Grishman. 2008. Refining Event Extraction Through Cross-document Inference. *Proc. ACL 2008*. Ohio, USA

Lillian Lee and Fernando Pereira. 1999. Distributional Similarity Models: Clustering vs. Nearest Neighbors. *Proc. ACL1999*. pp. 33-40.

Dekang Lin and Patrick Pantel. 2001. DIRT-Discovery of Inference Rules from Text. *Proc. ACM SIGDD Conference on Knowledge Discovery and Data Mining*.

Scott Miller, Jethran Guinness and Alex Zamanian.2004. Name Tagging with Word Clusters and Discriminative Training. *Proc. HLT-NAACL2004*. pp. 337-342. Boston, USA.

Franz Josef Och and Hermann Ney. 2003. "A Systematic Comparison of Various Statistical Alignment Models", Computational Linguistics, volume 29, number 1, pp. 19-51.

Martha Palmer, Daniel Gildea and Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*. Volume 31, Issue 1. pp. 71-106.

Bo Pang, Kevin Knight and Daniel Marcu. 2003. Syntax-based Alignment of Multiple Translations: Extracting Paraphrases and Generating New Sentences. *Proc. HLT/NAACL 2003*.

Fernando Pereira, Naftali Tishby and Lillian Lee. 1993. Distributional Clustering of English Words. *Proc. ACL1993*. pp. 183-190.

Barry Schiffman, Kathleen R. McKeown, Ralph Grishman and James Allan. 2007. Question Answering using Integrated Information Retrieval and Information Extraction. *Proc. HLT-NAACL 2007*. Rochester, US.

Yusuke Shinyama and Satoshi Sekine. 2003. Paraphrase Acquisition for Information Extraction. *Proc. ACL 2003 workshop on Paraphrasing (IWP 2003)*.

Hongye Tan, Tiejun Zhao and Jiaheng Zheng. 2008. Identification of Chinese Event and Their Argument Roles. *Proc. Computer and Information Technology Workshops*.

Nianwen Xue and Martha Palmer. 2009. Adding semantic roles to the Chinese Treebank. *Natural Language Engineering*, 15(1):143-172.

Richard Zens and Hermann Ney. 2004. Improvements in Phrase-Based Statistical Machine Translation. *In HLT/NAACL 2004*. New York City, NY, US