Building Chinese Event Type Paradigm Based on Trigger Clustering

Xiao Ding, Bing Qin, Ting Liu*
Research Center for Social Computing and Information Retrieval
Harbin Institute of Technology, China
{xding, qinb, tliu}@ir.hit.edu.cn

Abstract

Traditional Event Extraction mainly focuses on event type identification and event participants extraction based on pre-specified event type annotations. However, different domains have different event type paradigms. When transferring to a new domain, we have to build a new event type paradigm. It is a costly task to discover and annotate event types manually. To address this problem, this paper proposes a novel approach of building an event type paradigm by clustering event triggers. Based on the trigger clusters, the event type paradigm can be built automatically. Experimental results on three different corpora – ACE (small, homogeneous, open corpus), Financial News and Musical News (large scale, specific domain, web corpus) indicate that our method can effectively build an event type paradigm and can be easily adapted to new domains.

1 Introduction

Event extraction techniques have been widely used in several different specific domains, such as musical reports (Ding et al., 2011), financial analysis (Lee et al., 2003), biomedical investigation (Yakushiji et al., 2001) and legal documents (Schilder et al., 2007). Traditional event extraction systems achieved excellent performance in some important information extraction benchmarks, such as MUC (Message Understanding Conference, Chinechor et al., 1994) and ACE (Automatic Content Extraction). However, most of these methods require pre-specified event types as their prior knowledge. For example, ACE defines an event as *a specific occurrence involving participants*, and it annotates 8 types and 33 subtypes of events (LDC, 2005). However, building an event type paradigm in this way not only requires massive human effort but also tends to be very data dependent. As a result, it may prevent the event extraction from being widely applicable. Since event types among domains are different, the event type paradigm of ACE, which does not define music related events, is useless for the music domain event extraction. So we have to build a totally different event type paradigm for the music domain from scratch.

Recently, some researchers have been aware of the limitations of only considering pre-defined paradigm as well. In the same vein, some studies work on the problem of relation extraction (Chambers and Jurasky, 2011 and 2009; Poon and Domingos, 2009 and 2008; Yates and Etzioni, 2009). Rosenfeld and Feldman (2006) built a high-performance unsupervised relation extraction system without target relations in advance. Hasegawa et al. (2004) discovered relations among named entities from large corpora by clustering pairs of named entities. However, most of the above work focuses on relation extraction rather than event extraction.

In contrast to the well-studied problem of relation extraction, only a few works focused on event extraction. For example, Li (2010) proposed a domain-independent novel event discovery approach. They exploited a cross-lingual clustering algorithm based on sentence-aligned bilingual parallel texts to discover event trigger clusters. Their motivation is to discover novel events for a new domain rather than build a new event type paradigm from scratch. So it takes domain specific event triggers as the input. However, it is also a costly task to annotate triggers for new domains.

To address above issues, this paper proposes a series of novel algorithms to automatically build event type paradigm. The proposed approach is based on the definition of event trigger: *the word that most clearly expresses an event's occurrence*, and our key observations: *triggers are the most important lexical units to represent events. A set of triggers with similar meaning or usage represents the same event type. Event types can*

* Email correspondence

International Joint Conference on Natural Language Processing, pages 311–319,
Nagoya, Japan, 14-18 October 2013.

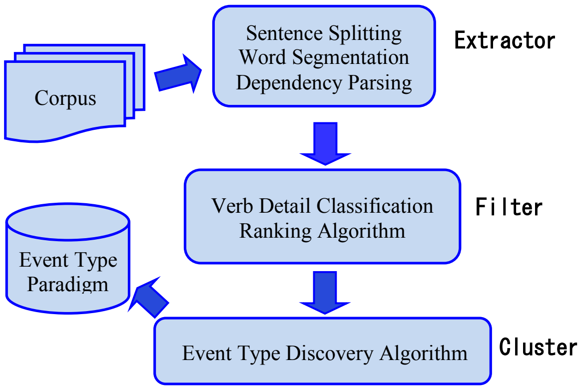

Figure 1. Architecture of the proposed system

be discovered based on trigger clustering. Our approach involves three steps: 1) we introduce a trigger extraction algorithm based on the dependency syntactic structure; 2) a trigger filter is then constructed to remove some noisy candidate triggers; 3) we develop an event type discovery algorithm based on our proposed trigger clustering methods. The clustered event types are used to construct an event type paradigm.

Experimental results show that our approach not only achieve significantly better performances than the baseline method, but also are more stable across different corpora. On the ACE, Financial News and Musical News corpus, the average accuracy is 73%. It shows that trigger clustering based method is effective on building an event type paradigm which is the premise of event extraction. We extract 33 event types for the ACE corpus, nine event types for the Financial News corpus and seven event types for the Musical News corpus.

Our contributions are as follows.

1. In this paper, we put forward the problem of event type paradigm building, and develop a novel framework as the solution.
2. This paper exploits a series of novel algorithms for automatically discovering and clustering domain independent event types.

The remainder of this paper is organized as follows. Section 2 presents our approach for event type paradigm building. Section 3 evaluates the proposed method. The related work on event extraction is discussed in Section 4, and we conclude the paper in Section 5.

2 Approach Overview

Since event trigger is the word that most clearly expresses an event's occurrence, the key idea of this paper is to automatically construct an event type paradigm by clustering event triggers. For example, in the ACE corpus, a set of event triggers {倒闭, 闭门, 关闭, 停业, 解散} ({bankrupt, shut down, close, close down, dismiss}) represents the sense of the event type "Business/End-Org". As shown in Figure 1, our system has three main components: trigger extractor, trigger filter and trigger cluster. The input of the system is a raw corpus, such as the ACE corpus, the Financial News corpus and the Musical News corpus, and the output is the event type paradigms which are shown in Table 2, Table 3 and Table 4.

```
Algorithm 1: TE algorithm
Input: Raw corpus D
Output: Candidate triggers
Foreach document d in raw corpus D Do
   d ← Paragraph Splitting
   d ← Sentence Splitting
   Foreach sentence s in document d Do
      s ← Word Segmentation
      s ← Chinese Dependency Parsing
      s ← Identify subject-predicate relation (SBV) pair
            (V_SBV, Sub) and verb-object relation (VOB)
            pair (V_VOB, Obj)
      If V_SBV = V_VOB = V_t, Then
         Extract V_t as candidate trigger
   End For
End For
```

Figure 2. The algorithm for trigger extraction

2.1 Trigger Extractor

An event trigger is the center of an event, which is an important feature for recognizing the event type. Kiyoshi Sudo (2003) summarized three classical models for representing events. All of these three models rely on the syntactic tree structure and the trigger is specified as a predicate in this structure. In order to accurately extract event triggers, we employ the predicate-argument model (Yangarder et al., 2000) which is based on a direct syntactic relation between a predicate and its arguments. We extract the syntactic relation for predicate-argument model by means of the HIT (Harbin Institute of Technology) Dependency Parser (Che et al., 2009). Based on the predicate-argument model, we propose a **t**rigger **e**xtraction algorithm (**TE**). The details are shown in Figure 2.

Take the following sentence as an example:

毛泽东 1893 年 出生 于 湖南湘潭。
1 2 3 4 5

→ *Mao Zedong was born in Xiangtan, Hunan Province in 1893.*
1 2 3 4 5 6 7

The HIT Chinese Dependency Parser dependencies are:

SBV (出生-3, 毛泽东-1)
→ (born-3, Mao Zedong-1)
VOB (出生-3, 湖南湘潭-5)
→ (born-3, Xiangtan, Hunan Province-5)

ADV (出生-3, 1893 年-2)
→ (born-3, 1893-7)
POB (湖南湘潭-5, 于-4)
→ (Hunan Province-5, in-4)

where each atomic formula represents a binary dependence from the governor (the first token) to the dependent (the second token). The SBV relation, which stands for the *subject-predicate structure*, means that the head is a predicate verb and the dependent is a subject of the predicate verb; the VOB dependency relation, which stands for the *verb-object structure*, means that the head is a verb and the dependent is an object of the verb; the ADV relation, which stands for the *adverbial structure*, means that the head is a verb and the dependent is an adverb of the verb; the POB relation, which stands for the *prep-object structure*, means that the head is an object and the dependent is a preposition of the object.

Since $V_{SBV} = V_{VOB} = V_t =$ 出生 (*born*) in this case, based on the predicate-argument model, the word "出生" should be extracted as a candidate event trigger.

2.2 Trigger Filter

Although we obtain some useful candidate triggers, certain meaningless candidate triggers come along in the results of the trigger extractor as well. Therefore, we introduce a trigger filter which uses heuristic rule and ranking algorithm to filter out these less informative candidates.

These rules are applied in order as follows:

Rule (1): Subdividing Verbs

Since event trigger words are extracted based on the predicate-argument model, most of these candidate trigger words are verb terms. However, not all of verb terms can be used as trigger words. For example, the copular verb (e.g. "is") rarely acts as the event trigger. To investigate which categories of verbs can serve as event triggers, we classify Chinese verbs into eight subclasses listed in Table 1. Such classification makes each subclass function as one grammatical role. For example, a modal verb will never be the predicate of a sentence and a nominal verb will always function as a noun.

We perform the verb sub-classification model based on the work by Liu et al. (2007). Statistically, about 94% of ACE Chinese event triggers are general verbs or nominal verbs and other types of verbs are rarely as trigger words. In order to ensure the accuracy of trigger clustering, we stress that the candidate trigger must be general verb or nominal verb.

Rule (2): Domain Relevance Ranking

Verb	Description	Examples
vx	copular verb	他 ***是*** 对 的 (He is right)
vz	modal verb	你 ***应该*** 努力 工作 (You should work hard)
vf	formal verb	他 要求 ***予以*** 澄清 (He'd demand an explanation)
vq	directional verb	他 认识 ***到*** 困难 (He has realized the difficulties)
vb	resultative verb	他 看 ***完*** 了 电影 (He has seen the movie)
vg	**general verb**	他 ***喜欢*** 踢 足球 (He likes playing football)
vn	**nominal verb**	参加 我们 的 ***讨论*** (Take part in our discussion)
vd	adverbial verb	产量 ***持续*** 增长 (Production increases steadily)

Table 1. The scheme of verb subclass

Domain relevancy degree is an important measure of the trigger's significance. According to the candidate trigger distribution in the domain corpus and the general corpus, we can compute its domain relevancy degree as follows:

$$DR(V_t) = Freq_D(V_t) / Freq_G(V_t) \qquad (1)$$

where $DR(V_t)$ is the domain relevancy degree of the candidate trigger V_t, $Freq_D(V_t)$ is the frequency count of the candidate trigger V_t in the domain corpus (financial and musical news), and $Freq_G(V_t)$ is the frequency count in the general corpus (People's Daily corpus). We will rank candidate triggers by their domain relevancy degrees and retain top N_t[1] candidate triggers.

2.3 Trigger Clustering and Event Type Paradigm Building

The trigger word is the most important lexical unit to represent events. A set of triggers with the same meaning and usage represents the same event type. Event type can be discovered based on trigger clustering. We propose the **e**vent **t**ype **d**iscovery (**ETD**) algorithm based on trigger clustering without giving the number of clusters in advance. The algorithm is shown in Figure 3.

For two triggers V_i and V_j in *ETD*, the similarity function $Sim(V_i, V_j)$ in clustering is calculated using semantic information provided by HowNet (Dong et al., 2006) as

$$Sim(V_i, V_j) = \frac{2N_s}{N_i + N_j} \qquad (2)$$

where N_s denotes the number of identical sememes in the DEFs (the concept definition in HowNet) of V_i and V_j; N_i and N_j denote the number of sememes in the DEFs of V_i and V_j, respect-

[1] We test different N_t on dev set; and N_t is 50% of candidate triggers achieved the best gains.

Algorithm 2: ETD algorithm

Input: Candidate triggers from Section 2.2 and Threshold θ (refer to Section 3.2)
Output: Event trigger clusters
Foreach trigger V_t in candidate triggers **Do**
    Compute the similarity (*Sim*) between V_t and the rest of other triggers, using function (2)
    If *Sim* >= θ **Then**
        add V_t to the related event type $ET_{re} \cup \{V_t\}$
    Else If *Sim* < θ **Then**
        set up a new event type ET_{new}
End For

Figure 3. The ETD algorithm

Algorithm 3: PAC model

Input: Verb-argument tuples <V_t, *Subj, Obj*>, where V_t is the trigger from Section 2.2 and Subj and Obj are the arguments of V_t; and Threshold θ (refer to Section 3.2)
Output: Event trigger and arguments clusters
Foreach tuple *p* in verb-argument tuples <V_t, *Subj, Obj*> **Do**
    Compute the similarity (*Sim*) between *p* and the rest of other tuples, using function (3) and (4)
    If *Sim* >= θ **Then**
        add V_t to the related event type $ET_{re} \cup \{V_t\}$
    Else If *Sim* < θ **Then**
        set up a new event type ET_{new}
End For

Figure 4. The PAC model

tively. Hownet uses sememes to interpret concepts. Sememes are regarded as the basic unit of the meaning. For example, "paper" can be viewed as a concept, and its sememes are "white", "thin", "soft", "flammable", etc.

As referred in Section 2.1, most of trigger words are verb terms. Polysemic verbs are a major issue in NLP, such as "to *fire* a gun" and "to *fire* a manager", where "*fire*" has two different meanings. The state-of-the-art verb sense disambiguation approach (Wagner et al., 2009) stresses that verbs which agree on their selectional preferences belong to a common semantic class. For example, "to *arrest* the suspect" and "to *capture* the suspect". Based on this approach, we propose a PAC (predicate-argument clustering) model which group the verbs based on their subcategorisation and selectional preferences. ETD considers only the verb subcategorisation, whereas PAC involves the verb argument tuple, such as <*bomb*, *US Army*, *weapon warehouse*>, where "US Army" and "weapon warehouse" are the subject word and the object word of the trigger word "bomb". The clustering process of PAC which is shown in Figure 4 is the same as ETD, except for the similarity measurement. PAC calculates the similarity between all the verb argument tuples by the following function:

$$Sim(Tuple_i, Tuple_j) = 2Sum_s / (Sum_i + Sum_j) \quad (3)$$

$$Sum_s = N_s + S_s + O_s,\ Sum_i = N_i + S_i + O_i,\ Sum_j = N_j + S_j + O_j \quad (4)$$

where S_s and O_s denotes the number of identical sememes in the DEFs of *Subj*$_i$ and *Subj*$_j$, *Obj*$_i$ and *Obj*$_j$; S_i and S_j denote the number of sememes in the DEFs of *Subj*$_i$ and *Subj*$_j$, respectively; O_i and O_j denote the number of sememes in the DEFs of *Obj*$_i$ and *Obj*$_j$, respectively.

A group of triggers are aggregated to a trigger cluster according to their semantic distance, and we view each trigger cluster as one kind of event type. Then all these event types are finally employed to construct an event type paradigm.

3 Experimental Results and Analysis

3.1 Experiment Settings

3.1.1 Data Description

In order to test how robust our approach is, we evaluate it using three different data sets: ACE 05, Financial News[2] and Musical News[3]. ACE 05 is a public corpus with a pre-defined event type paradigm. Financial News and Musical News are specific domain corpora collected by ourselves. To justify the effectiveness of our method, we carefully conducted user studies into two specific domain corpora. For each sentence in the data, two annotators were asked to label and cluster all potential triggers. The agreement between our two annotators, measured using Cohen's kappa coefficient, is substantial (kappa = 0.75). We asked the third annotator to adjudicate the trigger clusters on which the former tow annotators disagreed. Each trigger cluster is used to represent one type of event. All these events construct our final event type paradigm. In particular, we carry out experiment on 633 documents from the ACE 05 corpus, 6000 sentences from the Financial News corpus and 6000 sentences from the Musical News corpus, respectively. One third of these data is used as development set and the remaining data is used as test set.

The gold standard event type paradigm of ACE, Financial News and Musical News are shown in Table 2, Table 3 and Table 4.

3.1.2 Evaluation Measure

We adopt *F-Measure (F)* and *Purity* (Halkidi et al., 2001) to determine the correctness of an event cluster:

$$p(i,r) = n(i,r) / n_r \quad (5)$$

[2] http://www.10jqka.com.cn/
[3] http://yue.sina.com.cn/

Method	Corpus	F-Measure (%)	Purity (%)
Baseline	ACE	42.05	61.47
ETD	ACE	63.21	68.17
PAC	ACE	**69.57**	**70.24**
ETD	Financial News	71.52	74.81
PAC	Financial News	**74.42**	**76.18**
ETD	Musical News	72.23	78.35
PAC	Musical News	**75.08**	**80.28**

Table 5. F-Measure and Purity scores on the test set. All the improvements are significant ($p < 0.05$)

Types	Subtype
Life	Be-Born, Marry, Divorce, Injure, Die
Movement	Transport
Transaction	Transfer-Ownership, Transfer-Money
Business	Start-Org, Merge-Org, Declare-Bankruptcy, End-Org
Conflict	Attack, Demonstrate
Contact	Meet, Phone-Write
Personnel	Start-Position, End-Position, Nominate, Elect
Justice	Arrest-Jail, Release-Parole, Trial-Hearing, Charge-Indict, Sue, Convict, Sentence, Fine, Execute, Extradite, Acquit, Appeal, Pardon

Table 2. ACE event type paradigm

Event Type	Examples
Start-Org	MIUI is found in 2010 by Xiaomi Tech.
End-Org	Sears closed more stores as holiday sales slide.
Merge-Org	Two of Tucson's oldest and most respected landscape companies have decided to merge.
Declare Bankruptcy	American airlines are falling sharply for the second straight day on the fears that the company might be forced to file for bank-ruptcy.
Go-Public	Chinese video site Youku filed to go public on the New York Stock.
Raise-Price	Gold price rises higher in Hong Kong.
Cut-Price	Sony cuts Tablet S price by $100, 16GB version now $399.
Cooperation	Nokia and Microsoft announce plans for a broad strategic partnership to build a new global mobile ecosystem.
Investment	Tencent, one of the biggest web companies in China, is investing $300m in Digital Sky Technologies of Russia.

Table 3. Financial News event type paradigm

Event Type	Examples
Vocal Con-cert	Chinese rock singer Cui Jian is to hold his first concert in Beijing at the Capital Gymna-sium on Aug. 24.
Album	'Super Girls' release 1st album 'Terminal PK' on August 29, 2005.
Awards	Kanye West won best rap album at the 48th annual Grammy Awards in Los Angeles.
Sign-Org	Lady Gaga was signed with Streamline Rec-ords by the end of 2007.
Breakup-Org	Singer Chen Chusheng broke up with his agent E.E. Media after September.
Quit-Singing	Hong Kong pop queen and actress Faye Wong will soon quit her singing career.
Return-Stage	Faye Wong returned to the stage in 2010 amidst immense interest in the Sinosphere.

Table 4. Musical News event type paradigm

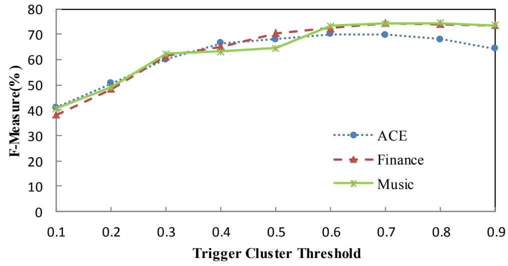

Figure 5. ETD algorithm with thresholding on the development set

$$r(i,r) = n(i,r) / n_i \quad (6)$$

$$f(i,r) = \frac{2 * p(i,r) * r(i,r)}{p(i,r) + r(i,r)} \quad (7)$$

$$F = \sum_i (n_i / n) \max\{f(i,r)\} \quad (8)$$

$$Purity = \sum_r (n_r / n) \max\{p(i,r)\} \quad (9)$$

where i is the gold standard event trigger cluster, and r is the event trigger cluster which has the most identical triggers with i. So n_i is the number of triggers in cluster i; n_r is the number of triggers in cluster r; n is the number of all triggers; and $n(i, r)$ is the number of identical triggers between i and r. For every cluster we first compute $p(i, r)$, $r(i, r)$ and $f(i, r)$, then we obtain *F-Measure* and *Purity* for the whole clustering result. Note that the evaluation is based on word instances rather than word types.

3.2 Selection of Trigger Cluster Threshold

During development, we tuned the trigger clustering threshold to find the best value. Figure 5 presents the effect on F-Measure of varying the threshold for trigger clustering. This figure shows that the best performance on the development set can be obtained by selecting threshold 0.6 for the ACE corpus, 0.7 for the Financial News corpus and 0.9 for the Musical News corpus. Figure 5 suggests that the performance did not dramatically change with the threshold when θ from 0.6 to 0.9. Hence, we can firstly set θ = 0.6 for new domains. We also test different threshold values for PAC, the result of which is the same as ETD. Then we directly use 0.6 as threshold value to the blind test.

3.3 Comparative Experiments

All the evaluation results are shown in Table 5. We *first* compare our approach with Li et al., 2010 (denoted as baseline) on the ACE 05 corpus. They exploit a cross-lingual clustering algorithm based on sentence-aligned bilingual parallel texts to discover event trigger clusters. The baseline approach can generate both English and Chinese event trigger clusters. We only compare with its Chinese result. Our approach achieves better performance than the baseline approach (about 8% significant improvement on Purity and more than 20% significant improvement on F-Measure). In addition, the baseline approach uses 1233 gold standard English event triggers and 852 gold standard Chinese event triggers in the ACE 05 as the input. However, we automatically extract event triggers based on our trigger extraction algorithm.

We carry out the *second* comparison experiment between ETD algorithm which is based on trigger clustering and PAC model which is based on predicate-argument clustering. The trigger and its corresponding arguments (selectional preferences) play an important role in our approach. We observe that the F-Measure score is boosted from 63.21% to 69.57% on the ACE corpus by using the PAC model. This can be explained by the reason that single trigger is not quite enough for representing event. Trigger's arguments can contribute to trigger disambiguation. The experiment results also confirm the assumption (Wagner et al., 2009) that verbs which agree on their selectional preferences belong to a common semantic class.

We also run the *third* comparison experiment using three different corpora (ACE 05, Financial News and Musical News) to evaluate the robustness and domain adaptiveness of our system. The performances on the specific domain corpora are better than that on the ACE corpus (about 5% absolute improvement on F-Measure and 6%-10% on Purity). The main reason is that the events in specific domain are more specific. In addition, the experiment results on both specific domain corpora can achieve good performance. This indicates that our system is domain independent.

In order to evaluate whether the filter rules used in Section 2.2 are effective, we introduce the *fourth* comparison experiment. We use the Purity score to evaluate the effectiveness of the two filter rules. The evaluation results are shown in Table 6. We find that the average improvement using only rule (1) is 6.93% absolute for three corpora compared with the performance without using rule; using only rule (2) is 5.84% absolute; and using both rules, the average improvement is 12.61% absolute. This indicates that our two filter rules can improve the experiment performance significantly.

Method \ Performance	Purity (%) ACE	Finance	Music
ETD	55.59	60.82	62.31
ETD+Rule(1)	65.13	66.51	69.37
ETD+Rule(2)	58.22	68.26	70.25
ETD+Rule(1)+Rule(2)	**68.17**	**74.81**	**78.35**
PAC	60.17	62.45	66.24
PAC+Rule(1)	68.04	69.24	72.38
PAC+Rule(2)	62.86	70.32	73.21
PAC+Rule(1)+Rule(2)	**70.24**	**76.18**	**80.28**

Table 6. The performance for filter rules

	Error types	Proportion
1	Trigger Extraction	33.0%
2	Trigger ambiguous	28.3%
3	Trigger Filter	19.5%
4	Others	19.2%

Table 7. Error types in the experiment

3.4 Discussion

Analysis of Experimental Errors

We first inspect the errors produced by our approach. The errors are mainly caused by the sparse event triggers in corpus. Table 7 shows the distribution of the errors in detail.

After error analysis, we found that the most number of errors are caused by trigger extraction. The main reasons are: firstly, not all of event triggers are verbs, such as "婚姻 (marriage)" for "Life/Marry" event, although it is reasonable to assume that event triggers are verbs because on average, there are more than 95% event triggers are verbs in our three different data sets. Secondly, since only verbs with subject and object are extracted, non-predicate verbs and the verbs without subject/object will not be extracted as candidate triggers. However, the coverage of possible triggers by our trigger extraction algorithm is reasonable good (more than 85%), because most of the trigger words appear repeatedly in the corpus, and their usages are varied. As long as one of their usages is fit for our extraction algorithm, they can be extracted as candidate triggers. Note that the goal of this paper is to build an event type paradigm for new domains. We concern more on the coverage of event type rather than event triggers. The event triggers extracted by us can cover all of event types. We will exploit more effective trigger extraction algorithm in future work.

Trigger ambiguity also accounts for a big proportion of the errors. As discussed in Section 2.3, we cannot judge the event type only by the trigger itself, such as "撤 (withdraw/dismiss)" for both "Personnel/End-Position" event and "Movement/ Transport" event. This kind of errors can be partially fixed by the PAC model. For example, we cluster "撤职务 (dismiss duties)" for "Personnel/End-Position" event and "撤军队 (withdraw troop)" for "Movement/Transport" event. These examples indicate that selectional preferences seem to be a reasonable feature even for highly ambiguous verbs like "撤" which encourages to improve argument extraction.

There are still some errors caused by trigger filter. This is mainly due to the fact that not all of triggers are general verb or nominal verb. Domain relevance ranking filter rule will ignore the common event types, which might also be very important for general event extraction, such as "Life/Die" event in the ACE corpus. More effective filter rules will be exploited in future.

Some other errors are caused by NLP tools, such as word segmentation, part-of-speech tagging and dependency parsing. We believe that our algorithms can be improved with the improvement of these NLP tools. In addition, there are about 10% of good event triggers extracted but put into the wrong cluster by trigger cluster.

Analysis of Different Corpus Sources

The third comparison experiment shows that the performance of our approach on three corpora is not very consistent (F-Measure 69.57%, 74.42% and 75.08% on the ACE, Financial and Musical corpus, respectively). The F-Measure on the ACE corpus is lower than that on the other two domain corpora. The performances on the other two domain corpora are comparable. The main reasons are as follows: firstly, the discrimination between some event types in the ACE paradigm is very small, such as the "Justice/Charge-Indict" event and the "Justice/Sue" event; the "Personnel/Nominate" event and the "Personnel/Start-Position" event; the "Life/Die" event and the "Conflict/Attack" event. Secondly, some events rarely occur in the ACE corpus, such as "Justice/Extradite" event occurs only three times in the ACE corpus. Thirdly, some events have a lot of triggers in the ACE corpus, but not all of these event triggers appear frequently. For example, the "Movement/Transport" event has 188 triggers and 64.89% of its triggers appear only once. As compared to ACE corpus, the similarity among event types in the other two corpora is low. Finally, we analyze that the quantity of event types also results in the different performance between the ACE corpus and the domain-specific corpus. There are 33 subtypes of events in the ACE corpus which are far more than the number of events in the Financial and Musical corpus.

Analysis of Different Filter Rules

The fourth comparison experiment indicates that both the filter rules are effective. As shown in Table 4, the improvement obtained using rule (1) is 7.87%, 6.79% and 6.14% on the ACE, Financial and Musical News corpus, respectively. The experiment result verifies that verb subdividing is helpful for the Chinese event extraction task. The improvement obtained using rule (2) is 2.69%, 7.87% and 6.97% on the ACE, Financial and Musical News corpus, respectively. The performances on all these three different corpora are improved by rule (2); however, it is obvious that rule (2) is not much effective on the ACE corpus (2.69%) compared with the other two domain-specific corpora (7.87% and 6.97%). The main reason is that the ACE corpus contains many common events and the domain-specific information is not very useful. For the other two domain-specific corpora, rule (2) has improved the performance more than rule (1) did. This is due to the fact that rule (2) is more effective on the domain-specific corpus.

4 Related Work

4.1 Word Cluster Discovery

Our approach of automatically building an event type paradigm is related to some prior work on word cluster discovery (e.g. Barzilay and McKeown, 2001; Ibrahim et al., 2003; Pang et al., 2003). Most of these works are based on machine translation techniques to solve paraphrase extraction problem. However, several recent researches have stressed the benefits of using word clusters to improve the performance of information extraction tasks. For example, Miller et al., (2004) proved that word clusters could significantly improve English name tagging performance. In the same vein, some studies work on the problem of relation extraction (Chambers and Jurasky, 2011 and 2009; Poon and Domingos, 2009 and 2008; Yates and Etzioni, 2009). In these work, "relation words" were extracted and clustered. In this paper, our work confirmed that trigger clusters are also effective for event type paradigm building. The problem of event trigger words extraction and clustering is also a challenge problem.

4.2 Traditional Event Extraction

The commonly used approaches for most event extraction systems are the knowledge engineering approach and the machine learning approach. Grishman et al., (2005) used a combination of pattern matching and statistical modeling techniques. They extract two kinds of patterns: 1) the sequence of constituent heads separating anchor and its arguments; and 2) a predicate argument sub-graph of the sentence connecting anchor to all the event arguments. In conjunction, they used a set of Maximum Entropy based classifiers for 1) Trigger labeling, 2) Argument classification and 3) Event classification. Ji and Grishman, (2008) further exploited a correlation between senses of verbs (that are the triggers for events) and topics of documents. They first proposed refining event extraction through unsupervised cross-document inference. Following Ji's work, Liao et al., (2010) used document level cross-event inference to improve event extraction. Chen and Ji, (2009) combined word-based classifier with character-based classifier; and explored effective features for the Chinese event extraction task. Liao and Grishman, (2010) ranked two semi-supervised learning methods for adapting the event extraction system to new event types. Hong et al, (2011) proposed a blind cross-entity inference method for event extraction, which well uses the consistency of entity mention to achieve sentence-level trigger and argument (role) classification. Lu and Roth, (2012) presented a novel model based on the semi-Markov conditional random fields for the challenging event extraction task. The model takes in coarse mention boundary and type information and predicts complete structures indicating the corresponding argument role for each mention.

However, for all the above approaches, it is necessary to specify the target event type in advance. Defining and identifying those types heavily rely on expert knowledge, and reaching an agreement among the experts or annotators requires a lot of human labor. Li et al., (2010) proposed a domain-independent novel event discovery approach. They exploited a cross-lingual clustering algorithm based on sentence-aligned bilingual parallel texts to discover event trigger clusters. Their motivation is to discover novel events for a new domain rather than build a new event type paradigm from scratch. Therefore, it takes domain specific event triggers as the input. However, it is also a costly task to annotate triggers for new domains. The motivation of this paper is to build event type paradigm from scratch rather than discover novel events based on the existing event type paradigm.

5 Conclusion and Future Work

Traditionally, in the topic of event detection, we have to categorize the events into various predefined event-types. In this paper, we aim to tackle the situation when the category of event-type is undefined, and we try to derive the event-types from the corpus. In particular, we automatically build an event type paradigm by using a trigger clustering algorithm: 1) we introduce a trigger extraction algorithm based on the dependency syntactic structure; 2) a trigger filter is then constructed to remove some noisy candidate triggers; 3) we develop an event type discovery algorithm based on our proposed trigger clustering methods. The clustered event types are used to construct an event type paradigm. Experimental results on three different corpora – ACE (small, homogeneous, open corpus), Financial News and Musical News (large scale, specific domain, web corpus) indicate that our method can effectively build an event type paradigm and that it is easy to adapt the proposed method to new domains.

In the future, more sophisticated algorithm will be exploited. Furthermore, a bottom-up event extraction system can be built based on our event type paradigm.

Acknowledgments

This work was supported by National Natural Science Foundation of China (NSFC) via grant 61273321, 61133012 and the Nation-al 863 Leading Technology Research Project via grant 2012AA011102.

References

R. Barzilay and K. McKeown. 2001. Extracting Paraphrases from a Parallel Corpus. In *Proceedings of ACL/EACL*.

N. Chambers and D. Jurafsky. 2009. Unsupervised Learning of Narrative Schemas and their Participants. In *Proceedings of ACL-IJCNLP*.

N. Chambers and D. Jurafsky. 2011. Template-Based Information Extraction without the Templates. In *Proceedings of ACL*.

W. Che, Z. Li, Y. Li, Y. Guo, B. Qin, T. Liu. 2009. Multilingual dependency-based syntactic and semantic parsing. In *Proceedings of CoNLL 2009*.

H. Chieu and H. Ng. 2002. A Maximum Entropy Approach to Information Extraction from Semi-structured and Free Text. In *Proceedings of AAAI*.

N. Chinchor, L. Hirschman and D. D. LewW. 1994. Evaluating Message Understanding Systems: An Analysis of the Third Message Uderstanding Conference (MUC-3). In *Computational Linguistics 3*, pages 409-449.

X. Ding, F. Song, B. Qin and T. Liu. 2011. Research on Typical Event Extraction Method in the Field of Music. *Journal of Chinese Information Processing Vol. 25, No. 2, 15-20*.

Z. Dong and Q. Dong. 2006. HowNet and the Computation of Meaning. World Scientific Publishing Co. Pte. Ltd. 2006.

R. Grishman, D. Westbrook and A. Meyers. 2005. Nyu's English ACE 2005 System Description. In *Proceedings of NIST 2005*.

M. Halkidi, Y. Batistakis, M. Vazirgiannis. On Clustering Validation Techniques[J]. *Intelligent Information Systems* 2001. 17(2-3): 107-145.

T. Hasegawa, S. Sekine and R. Grishman. 2006. Discovering Relations among Named Entities from Large Corpora. In *Proceedings of ACL*.

Y. Hong, J. Zhang, B. Ma, J. Yao, G. Zhou, Q. Zhu. 2011. Using Cross-Entity Inference to Improve Event Extraction. In *Proceedings of ACL*.

A. Ibrahim, B. Katz and J. Lin. 2003. Extracting Structural Paraphrases from Aligned Monolingual Corpora. In *Proceedings of the Second International Workshop on Paraphrasing (IWP 2003)*.

H. Ji and R. Grishman. 2008. Refining Event Extraction through Unsupervised Cross-document Inference. In *Proceedings of ACL*.

LDC. 2005. ACE (Automatic Content Extraction) English Annotation Guidelines for Events. http://projects.ldc.upenn.edu/ace/docs/English-Events-Guidelines_v5.4.3.pdf

C. S. Lee, Y. J. Chen, and Z. W. Jian, Ontology-based fuzzy event extraction agent for chinese e-news summarization. *Expert Systems With Applications*, vol. 25, no. 3, pp. 431–447, 2003.

H. Li, X. Li, H. Ji, and Y. Marton. Domain-Independent Novel Event Discovery and Semi-Automatic Event Annotation. In *Proceedings of PACLIC*.

S. Liao and R. Grishman. 2010. Filtered Ranking for Bootstrapping in Event Extraction. In *Proceedings of Coling*.

T. Liu, J. Ma, H. Zhang and S. Li. 2007. Subdividing Verbs to Improve Syntactic Pasing. *Journal of Electronics*, 24(3): 347-352.

W. Lu and D. Roth. 2012. Automatic Event Extraction with Structured Preference Modeling. In *Proceedings of ACL*.

S. Miller, J. Guinness and A. Zamanian. 2004. Name Tagging with Word Clusters and Discriminative Training. In *Proceedings of HLT/ACL*.

B. Pang, K. Knight and D. Marcu. 2003. Syntax-based Alignment of Multiple Translations: Extracting Paraphrases and Generating New Sentences. In *Proceedings of NAACL*.

H. Poon and P. Domingos. 2008. Joint Unsupervised Coreference Resolution with Markov Logic. In *Proceedings of EMNLP*.

H. Poon and P. Domingos. 2009. Unsupervised Semantic Parsing. In *Proceedings of EMNLP*.

B. Rosenfeld and R. Feldman. 2006. URES: an Unsupervised Web Relation Extraction System. In *Proceedings of the COLING/ACL*.

F. Schilder. 2007. Event Extraction and Temporal Reasoning in Legal Documents. In *Proceedings of Annotating, Extracting and Reasoning about Time*, LNAI-Volume 4795/2007, pp: 59-71.

K. Sudo, S. Sekine, and R. Grishman. 2001. Automatic pattern acquisition for Japanese information extraction. In *Proceedings of HLT*.

K. Sudo, S. Sekine, and R. Grishman. 2003. An Improved Extraction Pattern Representation Model for Automatic IE Pattern Acquisition. In *Proceedings of ACL*.

M. Thelen and E. Riloff. 2002. A Bootstrapping Method for Learning Semantic Lexicons using Extraction Pattern Contexts. In *Proc. of EMNLP*.

W. Wagner, H. Schmid, and S. Schulte im Walde. 2009. Verb Sense Disambiguation using a Predicate Argument-Clustering Model. In *Proceedings of the CogSci Workshop on Distributional Semantics beyond Concrete Concepts*.

A.Yakushiji, Y.Tateisi, Y.Miyao, and J.Tsujii, 2001. Event extraction from biomedical papers using a full parser. Pacif. Symp. Biocomp, 6, pp: 408-419.

M. Yankova 2003. Focusing on Scenario Recognition in Information Extraction. In *Proceedings of ECAL*.

A. Yates and O. Etzioni. 2009. Unsupervised Methods for Determining Object and Relation Synonyms on the Web. *Journal of Artificial Intelligence Research 34 (2009) 255-296*.

Q. Zhou and M. Sun. 1999. Build a Chinese treebank as the test suite for Chinese parsers. In *Proceedings of the workshop MAL'99 and NLPRS'99*.