Event Relation Recognition by Multi Part of Speech Association Distribution Characteristics

Chao Han, Feng Zhou *, Lanjiang Zhou and Xin Yan

Key Laboratory of Intelligent Information Processing, Kunming University of Science and Technology, Kunming, Yunnan, China

* Corresponding author: zf158@sina.com

Abstract. Event relation recognition, as one of natural language processing technologies, faces information stream of texts detecting event relation. By analyzing the influence of the words of different parts of speech on the relevance of events. And use the form of lexical chain to extract and store the relevant vocabulary between events, this paper propose an event relation recognition method based on lexical chain to detect latent semantic relation between events: whether events hold logical relation or not. Compared with the method based on dependency cue inference, the proposed method achieves 7.68% improvement.

Keywords. Multi part of speech distribution, lexical chain, event relation.

1 Introduction

Events refer to the objective facts that are involved in a number of roles and show some kind of behavior or state characteristics [1].

The occurrence of the event has the characteristics of objectivity, authenticity and so on. However, the occurrence of events is often not an isolated phenomenon. The occurrence of an event necessarily involves other events associated with it, such as the cause event, the result event, the concurrency event, and so on. The logical form of interdependence and association between events and their related events, called event relations. In the expression level of the text, the text information which expresses the event and the event relation contains a specific distribution rule. Therefore, it is great help to the realization of topic deduction and topic prediction for discrete events in large-scale information flows that the formation of a natural language analysis and information processing mechanism for automatic recognition and detection of event relations.

Event relation recognition is to achieve the shallow layer detection of the event logic relation, that is to judge whether there is a logical correlation between any events, and it is a kind of two element relation judgment. By parsing the text structure and semantic features, the results of "relevant" or "irrelevant" judgments are directly given to the text fragments (including phrases, clauses, sentences and paragraphs) describing the different natural events in the text.

The rest of the paper is organized as follows. Section 2 presents the related works. Section 3 analyzes the distribution characteristics of the multi part of speech association and construction of the lexical chain; Section 4 describes the method of Event Relationship Recognition Based on Lexical Chain in detail; Section 5 describes the experiment; Section 6 Conclusion.

2 Related work

Event detection is becoming a new research hotspot because of the increasing demand in automatic question answering, automatic summarization and event prediction. And the main mining methods in event logical relation recognition are divided into template matching and element analysis.

2.1 Template matching

One of the main methods of event relation detection is to use the pattern matching of event feature, according to the defined templates, event relation that matches with the template in the text is extracted. Chklovski [2] uses LSP (lexical syntactic pattern) to extract the resources with event relations, and the results are organized into a knowledge base called "VerbOcean". Chklovskii et al, extracted the event matching of six temporal relationships (similarity, strength, antonymy, enablement, happen, and before) using a manually collected LSP template. Manually defined event relation templates are often subject to a number of constraints, resulting in low recall rate of relational detection.

2.2 Element analysis

Most of the research on the event relation detection based on event elements inherited the distribution hypothesis of Harris [3]. The Harris hypothesis states that words in the same context have the same or similar meaning. Lin [4] proposed an unsupervised method, called DIRT algorithm, which combines the Harris distribution hypothesis with the idea of establishing dependency tree.

Ma Bin [5] through the analysis of event semantic dependency characteristics and events in the evolution of the semantic dependency of the law, proposed a semantic dependency clue based event correlation identification method. By locating the events in the text information flow, the method analyzes the syntactic dependency between events, and mining the reasoning clues (i.e. "dependent cues"), and then realizes the automatic identification of events.

3 Multiple part of speech semantic association distribution analysis and lexical chain

3.1 Multiple part of speech semantic association distribution analysis

The event can be thought of as consisting of a predicate and its accompanying six variables (who / whom, what, when, where, why, how), where predicates are usually verbs and the other six variables are nouns, Pronouns, adjectives and adverbs. In other words, the event is actually composed of a number of words with different parts of speech. It is found that even if the modified part (adjectives and adverbs) of the event is removed, The main information about the event can still be gotten. For example, a complete event: "Tom went to the big supermarket which is newly opened and close to his home early in the morning, and bought 3 pounds of fresh big crabs." After removing the modified part of the event: "Tom went to the supermarket to buy (3 pounds of) crabs." The main information of the event before and after processed is the same, the processed events are more concise in expression, which is mainly composed of verbs and nouns (and sometimes pronouns). It can be seen that in an event, verbs and nouns describe most of the information of the event, while adjectives and adverbs describe very little information about the event.

If there is a correlation between the main information of the two events, it is believed there is a correlation between the two events. This paper can convert the question of whether the two events are related to determine whether the words (verbs, nouns and pronouns) that describe the main information of the two events are related. Of course, a single word is not a complete expression of the main information of the event, so this paper determine the relevance of the two events is based on multiple words simultaneously related. So how many simultaneously related pair of words in the events are enough to determine that the two events are correlated? In order to explore this problem, this paper has carried on the

related statistical experiment. Four press releases, 967 news events are selected out In Sohu, Sina and other news sites, after manual labeling and statistics, the statistical results shown in Table 1 and Table 2 .

Table 1. The probabilities of event-related in the same type related words.

Quantity POS	1 pair	2 pair	3 pair
Nouns	26.77%	58.36%	83.54%
Verbs	61.34%	87.26%	
Pronouns	70.46%	83.85%	

Table 2. The probabilities of event-related in the multiple type related words.

Verbs & Nouns	Verbs & Pronouns	Pronouns & Nouns
93.24%	96.12%	89.67%

Table 1 shows that when there are one or more pairs of related words with the same part of speech in two events, the two events are the probability of the relevant event. And Table 2 shows that when there are multiple pairs of lexical correlations with different parts of speech in the two events, the two events are the probabilities of the relevant events. In the experiment, this paper find that the relevance of verbs and pronouns has a great impact on the relevance of events, while the relevance of nouns has little impact on the relevance of events. And the more the number of relevant vocabularies between events, the greater the probability of events being related events, while the probability of events being related events is generally high when there are multiple pairs of lexical correlations with different parts of speech between events. According to the feature of semantic relevance distribution, this paper proposes a method of event relation recognition based on the feature of multi part of speech association distribution. In terms of the processing of related words, this paper uses lexical chain to extract and store related vocabulary of event set.

3.2 Lexical chain

Halliday and Hasan [6] first proposed the concept of lexical chain in 1976. A lexical chain is a chain of words that are related to each other. There are many ways to construct a lexical chain. Morris and Hirst [7] first proposed a greedy algorithm to generate lexical chains, Barzilay and Elhadad [8] proposed a non-greedy algorithm model in 1997 to build a lexical chain. Silber and McCoy [9], Galleyz and McKeown [10] also proposed effective methods for constructing lexical chains. This paper argues that the lexical chain is a set of semantically related words which, using the method of semantic relevance is more suitable for the construction of lexical chain.

Semantic relevance of words reflects the degree of relevance between the semantic meaning of words. The computation method of semantic relatedness of words is divided into two categories, one is the machine learning method based on the corpus, the other is the method based on the semantic dictionary. This paper uses the computing method of the semantic relevance of Chinese words based on HowNet proposed by Li Gao [11] to calculate semantic relevance of words, this method takes into account various semantic relations among words.

3.3 Lexical chain construction algorithm

The construction of the lexical chain is based on the whole text. Firstly, the text is preprocessed, including text segmentation, low frequency word filtering, part of speech tagging and semantic annotation. Then after nouns which have been semantic annotated are processed, the semantic relevance between words is computed, the words whose semantic relevance value meets the specified criteria are taken as candidate words, and the synonyms are merged, then the words that contribute little significance to the meaning of literature are filtered out, Finally, semantically related words are

aggregated together as a lexical chain. In order to reduce the workload of the automatic indexing system, some automatic indexing system after word segmentation, will first filter out those words with low frequency words, and then semantic annotation and part of speech tagging. The method in this paper chooses the sentences as the unit, and then merging these candidates to build a new lexical chain. The concrete construction of the lexical chain construction algorithm is as follows:

- (1) Word segmentation, word frequency filtering, POS tagging and semantic tagging.
- (2) Set semantic correlation threshold s .
- (3) Based on sentence, calculate semantic correlation value between words in a sentence, adding the words whose correlation value is greater than s to the candidate words set H , the final result of H is $\{W_1, w_2, \dots, W_n\}$.
- (4) Select w_1 as the first element of the initial lexical chain L from H and remove w_1 from H .
- (5) From the remaining words of H , this paper selects the words whose semantic relevance is greater than s in L , add them to the lexical chain L , and delete these words from H at the same time. Repeat this step until no new words are added to the L , and L is a complete word chain.
- (6) Repeat steps (4) and (5) to build other lexical chains until there is no word in H .

4 Event relation recognition by multiple part of speech semantic association distribution characteristics

The language is the carrier of the event, and the word is the smallest particle of language, that is to say, the event is actually composed of words. The lexical chain is a collection of semantically related vocabularies, so there must be more or less connections between events in the same lexical chain. These connections are partial, weak, and this paper cannot rely solely on this association to determine whether there is a relationship between the events. Many times, however, the same event is often associated with a number of lexical chains, and a lexical chain often involves multiple events, so the event set and the lexical chain will form a staggered relationship network. So as long as these parts, weak but staggered connection enough, this paper can use them to infer and even quantify the correlation between the events. This is the core idea of the event relation recognition method based on the feature of the multi part of speech semantic association distribution.

The flow chart of the event relation recognition method based on the feature of the multi part of speech semantic association distribution is shown in Figure 1. The method is divided into three parts: event preprocessing, lexical chain construction and event correlation calculation.

4.1 Event preprocessing

The main work of the event preprocessing is to carry out some necessary processing of the event set, to prepare for the following work. The processing flow of event preprocessing is given below.

Fig.1. Flow chart of event relation recognition method.

- (1) The events in the event set are numbered one by one, and the processing operations such as word segmentation and part-of-speech tagging are carried out.
- (2) A new event expression is generated based on the result of the lexical chain construction in 5.2.. Such as e_1 (ChS_1), where e_1 represents the event with number 1 and ChS_1 represents the set of lexical chains associated with event 1.
- (3) Event pairs set ECS is formed a by pairing the events in the lexical chain and excluding duplicated event pairs, the expression is $ECS = \{ (e_1, e_2), (e_i, e_j), \dots, (e_n, e_m) \}$.

4.2 Lexical chain construction

In this section, this paper builds the lexical chain with the algorithm introduced in 4.3 and the words extracted from the event set. Specific work is as follows:

- (1) The lexical chain construction of the event set was carried out by the method of constructing the lexical chain introduced in 4.3.
- (2) The lexical chain is processed and add the decision parameters v and p , the expression is $ch_j (v_j, p_j)$. In which ch_j is the lexical chain whose number is j , the value of v_j is 0 or 1, which means that this lexical chain is (1) the verb chain or not (0); the value of p_j is also 0 or 1, which means that this lexical chain is (1) the pronoun chain or not (0),but v_j and p_j can not be 1 at the same time.
- (3) Construct a lexical chain set ChS , whose expression is $ChS = \{ ch_1, ch_2 \dots ch_j \dots ch_n \}$.

4.3 Event correlation calculation

According to the first two parts of the work, to quantify the degree of correlation between the events, and then infer whether the event pairs are relevant, and finally output the relevant event pair set as the judgment result. This specific work is as follows:

- (1) This paper gets the related lexical chain set of each event in the event pair from the event pair set ECS, and intersect them to get the common chain set $ChRS$. If there is an event pair who is composed of e_i and e_j , so $ChRS = ChS_i \cap ChS_j$.
- (2) The degree of correlation between events can be expressed as the correlation factor θ , denoted as $e_i e_j, 0 \leq \theta \leq 2$, the greater the value of θ , the greater the correlation between events. θ is calculated as follows.

The number of shared lexical chains of the event pair and the characteristics of the shared lexical chain (v and p) have an effect on the value of θ . The parameters are weighted according to the influence of these factors. The formula is defined as (1).

Where α and β are weighting coefficients.

- (3) Finally, the judgment of whether the events of the event pair are related is: If $\theta \geq X$, this paper believes that the event pair is the relevant event pair, otherwise the event pair is not relevant. Where X is a threshold set manually. The relevant event pair is then stored into the relevant event pair set ERS.
- (4) Repeat steps (1), (2), (3) until all pairs of events in the event pair set are judged and processed.

5 Experimental analysis

5.1 Experimental data

In this experiment, this paper collected 10 news reports from Xinhua, Phoenix, Sina, Universal and aother global news website, and each news report contains 25 news events. Each news event is annotated ("Relevant" or "irrelevant"). Eventually, 1248 events "relationship pairs" are obtained, where the "event pair" with logical relationship (ie, "Relevant") is 346 pairs, accounting for the total "event pairs" 27.7%.

5.2 Evaluation criteria

In the event relation recognition task, the performance of the system depends mainly on the number of "event pairs" of the correctly identified association relationship. The general evaluation indexes in the field of text retrieval are used: Precision, Recall and F value.

5.3 Experimental results and analysis

In order to test and compare the effect of the event relation recognition model proposed in this paper, a system based on dependency cue [2] is chosen as the baseline system. First this paper adjusts the weighting coefficients in formula (1) to obtain the best result, and then apply the best value to the experiment. The weighting coefficients are: $\alpha = 0.4$, $\beta = 0.6$. Then this paper sets the decision threshold X to 1, that is, when $\theta \geq 1$ that the event pair is related event pair. The experimental results are shown in Table 3. The method proposed in this paper compared to the Baseline in the F value obtained 7.68% increase, while the P value and R value were also obtained 11.18% and 5.11% increase.

Table 3. Experimental results.

Method	P/%	R/%	F/%
Baseline	59.41	52.03	55.48
Our-method	70.59	57.14	63.16

Compared with Baseline, it can be seen that the event relation recognition method based on the feature of the multi part of speech semantic association has improved the precision and the recall rate. Which improve the precision is more significant, while the recall rate is not very significant increase, is still at a low level. The main reason is that the method of event relation recognition based on the characteristics of multiple parts of speech semantic association distribution is mainly based on the number and the part of speech of the shared lexical chain to determine whether there is a relationship between the events, and there are some related events whose number of shared lexical chain is small, and even some related events have not shared vocabulary chain. For example, the sentence: "It is estimated that during the Vietnam War, the United States dropped 2 million bombs to Laos, and 30% of the shells did not explode, bringing a huge hidden danger for the post - war Lao people." It contains event e1: "It is estimated that during the Vietnam War, the United States dropped 2 million bombs to Laos", event e2 : "and 30% of the shells did not explode" and event e3: "bringing a huge hidden danger for the post - war Lao people". Event e2 and event e3 are related events, but because there is no the shared lexical chain between event e2 and event e3, the system can not pair them as an event pair, and it is impossible to make a correct judgment.

6 Conclusion

This paper proposes an event relation recognition method based on the feature of the multi part of speech semantic association distribution, which is based on the task of event relation identification in the same topic. By using the relevance of the lexical chain, and then the shared lexical chain between the event pairs is extracted. According to the number and characteristics of these shared lexical chains, the correlation degree of event pairs is calculated, the recognition of the correlation between events is achieved. Experimental results show that the proposed method has higher P value, R value and F value compared with the event relation method based on event dependency cues. However, there are still some related events that can not be matched and judged (omitted and misjudged) by the method of this paper, and more event features and lexical chain features need to be extracted, and according to these features to carry out event matching and judgment.

Acknowledgment

Foundation item: China National Natural Science Foundation (61562049),(61462055).

References

1. X Yang, Y Hong, Y Chen, J Yao, Q Zhu. An Overview of Event Relation Detection System. JCIP 29, 4 (2015): 25-32
2. Chklovski, P Pantel. Global path-based refinement of noisy graphs applied to verb semantics//proceedings of the Joint Conference on Natural Language Processing, Jeju Island, Korea, (2005): 792—803
3. zS Harris. Mathematical Structure of Language. New York,(1968)
4. D Lin, P Pantel. Discovery of Inference Rules from Text// Proceedings of the 7th ACM SIGKDD, San Francisco, California, USA, (2001): 323-328
5. B MA, Y Hong, X Yang, J Yao. Using Event Dependency Cue Inference to Recognize Event Relation. ASNUP 49,1 (2013): 109–116
6. MAK Halliday, R Hasan. Cohesion in English. London, UK: Longman, (1976)
7. J Morris, G Hirst. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. CL 17, 1 (1991): 21-48
8. R Barzilay, M Elhadad. Using Lexical Chains for Text Summarization. In Proceedings of the ACL Workshop on Intelligent Scalable Text Summarization. Madrid, Spain, (1997): 10-17
9. HG Silber, KF McCoy. Efficiently Computed Lexical Chains as an Intermediate Representation for Automatic Text Summarization. CL 28, 4 (2006): 487-496
10. M Galley, K Mckeown. Improving Word Sense Disambiguation in Lexical Chaining. International Joint Conference on Ijcai, (2003): 1486-1488
11. G Li, Q Dai, Q Wei. A new approach to compute semantic relevance of Chinese words //Proceedings - 2010 International Conference on Artificial Intelligence and Education, ICAIE (2010): 610-613