

CEC2.0-Based Detection Event Relation for Chinese Text

Xianchuan Wang^{1, 2, a *}, Zongtian Liu^{1, b}, Tao Liao^{1, c} and Qiang Li^{1, d}

¹ School of computer engineering and science, Shanghai University, Shanghai200444, China

² School of Computer and Information Engineering, Fuyang Normal College, Fuyang236037, China

^axch_wang@shu.edu.cn, ^bztliu@shu.edu.cn, ^ctliao@aust.edu.cn, ^d603265406@qq.com

Keywords: event; event relation; Chinese text; CEC2.0.

Abstract. Event is the basic unit of knowing, remembering and understanding the real world. Event is regarded as the basic unit of knowledge representation. We detect event relation between two events for some Chinese texts by matching event relation pair table, which are constructed based on the CEC2.0 (called the Chinese Event Corpus). The experimental result indicates that there is a good outcome for detecting event relation with this method, the value of the precision, recall and F-measure respective is 83.6%, 77% and 80.2%. The motivation is to know the dynamic process of event happening and developing in the real world by the event relation in Chinese texts.

Introduction

The problem that how to detect some needed semantic information from the massive unstructured texts is one of the hot issues. Events are not isolated in the real world, and there are some inherent relations among them[1]. Chinese text is an expression style that new agency reports event and event relation in real world with Chinese characters. Traditionally, text knowledge representation is mainly based on words/ concepts in text mining field. It causes the Tennis Ball Problem[2]. Event has some affiliated semantic information and is bigger knowledge representation unit than words/concepts. What's more, the methodology that event is used to represent knowledge can avoid the Tennis Ball Problem above.

Event relation in text is a semantic representation style with characters, to some content, which objectively indicates happening and dynamic developing of event in real world. we treat event as the basis unit of knowledge representation for text. And detect event relation between two events from massive Chinese texts, which supply some support for application of text mining, such as automatic abstraction and question and answering based on event. Meanwhile, we want offer some service for government in emergency field by the event relation detected.

Related works

There have been many research works in the field of event detection, event extraction and application since MUC-6 initiated the tasks about information extraction in the aspect of event. However, the research interests in event relation are in a fledging period all over the globe. LDC released TimeBank1.2 in 2003. The next year LDC released ACE 05 elevation corpus. There were many entity-relation and relation between event and state in ACE 05. Both of them is annotated relation with regard to specific tasks, and they are based on predicate. However, some predicates are not events indeed.

The research works mainly focus on some special event relation, such as, time relation, causal relation in the aspect of detecting event relation. Additionally, the main methods are pattern matching, element analyzing and rule reasoning. Chklovski[4] extract resource with some event relations by Lexical-Syntactic Pattern. Torisawa[5] mine verb pair with entailment relation by unrestricted distribution strategy, and filter the verb pair by Quasi-pattern proposed by Glickman. Lin[6] propose an unsupervised algorithm called DIRT to detect event relation. Szpektor[7] propose TE/ASE algorithm, which can detect event relation by two steps. Mani accomplish event relation classification

with semi-supervised method, which is based on rule and EM. Fu[8] do some research works on causal relation. MA[9] manually annotated related relation among events for Chinese texts collected, but he only aims at related relation among events.

As we expressed above, we can find that the research on detecting event relation is starting. Although, some researchers realize the importance that event relation is crucial for automatic abstraction, information retrieve, QA and so on, the event relation detected is simplification. There are complex semantic relation among events in real world, but the event relation detected above is far from describing real relation among these events.

Detect Event Relation

In this paper, the definition of event, event relation and some specific event relation, such as causal relation, following relation etc, are described in detail by Liu[3]. The Chinese text is regarded as the raw text data for detecting event relation in emergency domain. Firstly, we automatically crawl free Chinese texts from the Internet. Then, pre-process them, and automatically detect event trigger based on trigger table from CEC2.0. Finally, detect event relation based on event relation pair table from CEC2.0.

● Pre-processText

We crawl Chinese texts from the Internet with the Nutch by the key words that are earthquake, food poison, terrorist attack, traffic accident and fire. Each text crawled is segmented ,tagged lexical category and removed stop words etc with the language cloud supplied by the agency of HIT and Xunfei.

● Detect Event

Detecting event is a key step for detecting event relation. In this paper, detecting event is accomplished by detecting trigger. Trigger is the word in sentence, which can illustrate the occurrence of event. Trigger and the occurrence of event have a one to one relation. We detect trigger by matching the trigger-table, which is constructed from CEC2.0. The trigger table is called T1. Then we expand T1 with Chinese synonyms dictionary and obtain trigger table expanded called T2. Next, we extract verbs from the texts, and save these verbs to list L1. Finally, we traverse T2 with each word in L1, which is called matching algorithm by trigger table for detecting event.

● Detect event relation

In this paper, we detect event relation for Chinese text by matching event pair with the event relation pair table constructed from CEC2.0. The event relation detected include causal relation, following relation, accompanying relation, component relation and thought content relation. CEC2.0 is called the Chinese Event Corpus, which is semi-automatically constructed by Zongtian Liu and others, in which they annotated event, event semantic role ,event relation and so on. The data about event and event relation annotated in CEC2.0 is illustrated by Table 1.

Table 1. Event and Event Relation of CEC 2.0

Topic	#article	#event	#causal	#accompany	#follow	#component	#thought-content
Earthquake	63	1053	247	30	40	1	168
Fire	75	1216	87	72	131	7	109
Food-Poisoning	61	1109	97	77	117	0	92
Traffic-Accident	85	1790	351	276	323	1	155
Terrorist Attack	49	823	135	95	147	1	165
Total	333	5991	917	550	758	10	689

We take the transitivity of event relation into account, and expand the event relation annotated in CEC2.0. The number of expanded event relation is illustrated by Table 2.

Table 2. The number of event relation expanded

Topic	#causal	#accompany	#follow	#component	#thought-content
Earthquake	372	45	72	1	282
Fire	183	103	213	7	184
Food-Poisoning	227	125	194	0	148
Traffic-Accident	531	374	492	1	256
Terrorist Attack	319	169	236	1	272
Total	1632	816	1207	10	1142

There are two events e_i and e_j in CEC2.0. If there is a certain relation between them, then the two events composite an event relation pair $\langle e_i, e_j \rangle$. Therefore, event relation pair table called T3, which is composited by some event relation pairs like this, can express the event relation expanded in CEC2.0. when detecting event relation, firstly, each two triggers composite an event relation pair in each Chinese text. Then, traverse the event relation pair table T3 with each event relation pair above. If it exists in T3, then the two events of the event relation pair above have the corresponding event relation.

Experiments and Evaluation

● Datasets and Evaluating Method

We regard the Chinese texts of emergency domain as testing dataset to detect event relation, which have 5 topics including terrorist attack, traffic accident, food poison, fire and earthquake, and the Chinese texts number of each topic is 200. The training dataset is all the texts annotated in CEC2.0. according to Table1 and Table2, we know that the number of triggers and event relation pairs respective is 5991 and 4807, which is contained in trigger table and event relation pair table constructed from CEC2.0. The approach we use for evaluating the result of detecting event relation is the universal parameters, which are the precision, recall and F-measure.

● Experiment Result and Analysis

Table 3. Detect Event and Event Relation

Topic	Detect Event			Detect Event Relation		
	P	R	F	P	R	F
Earthquake	72.4	68.3	70.3	81.9	76.3	79.0
Fire	75.1	70.6	72.8	83.4	77.1	80.1
Food Poisoning	74.7	69.8	72.2	85.3	76.9	78.9
Traffic Accident	77.6	72.8	75.1	87.8	80.2	83.8
Terrorist Attack	70.5	68.1	69.3	79.8	74.7	77.2
Mean	74.1	69.9	71.9	83.6	77.0	80.2

Table 3 tell us some data about precision, recall and F-measure for detecting event and event relation in each topic of Chinese texts. Under the circumstance that the number of texts is the same in each topic. The precision, recall and F-measure for detecting event and event relation of traffic accident topic text is higher than that of other topic texts. The precision, recall and F-measure for detecting event and event relation of terrorist attack is minimum among these topic texts. The mean value of precision, recall and F-measure for detecting event and event relation respective is 74.1%, 69.9%, 71.9%, 83.6%, 77% and 80.2%. After analyzing the reason why there are the phenomenon on

the precision, recall and F-measure described above, we find that the trigger table and event relation pair table is constructed from CEC2.0. What's more, the number of texts in traffic accident topic is larger than the other topic texts. The number of texts in terrorist attack topic is the least among the five topics. Those are indicated in Table 1. Therefore, the number of trigger and event relation pair in traffic accident is larger than that in other topics. However, the number of texts in each topic is different in CEC2.0, which result in the number of event relation pair is different from each other. And, to some extent, this phenomenon effect the result of detecting event relation.

Summary

Event is the basic unit of knowing and understanding, which is the way to describe the real world for human. Event is a coarse-grained unit of knowledge representation compared with that of words/concepts, which is suitable for the cognitive laws for human. There are inherent relation among events. We treat Chinese texts in emergency domain as the testing datasets, and the texts in CEC2.0 is the training datasets. We detect event relation by matching event relation pair table which built from CEC2.0. The event relation detected includes causal relation, following relation, accompanying relation, component relation and thought content relation. The result of detecting event relation is well, and the value of the precision, recall and F-measure respective is 83.6%, 77% and 80.2%. However, this method has a drawback that is based on the CEC2.0 which can be treated as domain knowledge base. This is the limit of the method we used in this paper.

Acknowledgements

This paper is supported by the National Natural Science Foundation of China under Grant No.61273328 and No.61305053.

References

- [1] Zong-tian, Liu.,etc., Research on Event-oriented Ontology Model. Computer Science, 2009. 36(11): p. 189-192,199.
- [2] Merlini, D., R. Sprugnoli, and M.C. Verri, The tennis ball problem. Journal of Combinatorial Theory, Series A, 2002. 99(2): p. 307-344.
- [3] Liu, W., et al. Extending OWL for modeling event-oriented ontology. in Complex, Intelligent and Software Intensive Systems (CISIS), 2010 International Conference on. 2010. IEEE.
- [4] Chklovski, T. and P. Pantel. VerbOcean: Mining the Web for Fine-Grained Semantic Verb Relations. in EMNLP. 2004.
- [5] Kazama, J.i. and K. Torisawa. Exploiting Wikipedia as external knowledge for named entity recognition. in Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL). 2007.
- [6] Lin, D. and P. Pantel. DIRT@ SBT@ discovery of inference rules from text. in Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. 2001. ACM.
- [7] Szpektor, I. and I. Dagan. Learning entailment rules for unary templates. in Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1. 2008. Association for Computational Linguistics.
- [8] Fu, J., et al., Using dual-layer CRFs for event causal relation extraction. IEICE Electronics Express, 2011. 8(5): p. 306-310.
- [9] MA Bin, et al., Using Event Dependency Cue Inference to Recognize Event Relation. Journal of Peking University, 2013. 49(1): p. 109-116.