Mining Large-scale Event Knowledge from Web Text

Ya-nan Cao, Peng Zhang, Jing Guo, Li Guo

Institute of Information Engineering, Chinese Academy of Science, Beijing, China

{caoyanan, zhangpeng, guoli}@iie.ac.cn

Abstract

This paper addresses the problem of automatic acquisition of semantic relations between events. While previous works on semantic relation automatic acquisition relied on annotated text corpus, it is still unclear how to develop more generic methods to meet the needs of identifying related event pairs and extracting event-arguments (especially the predicate, subject and object). Motivated by this limitation, we develop a three-phased approach that acquires causality from the Web text. First, we use explicit connective markers (such as “because”) as linguistic cues to discover causal related events. Next, we extract the event-arguments based on local dependency parse trees of event expressions. At the last step, we propose a statistical model to measure the potential causal relations. The results of our empirical evaluations on a large-scale Web text corpus show that (a) the use of local dependency tree extensively improves both the accuracy and recall of event-arguments extraction task, and (b) our measure improves the traditional PMI method.

Keywords: event knowledge, web text mining, event-argument extraction, causality measurement

1 Introduction

As a source of intelligence, the general notion of causality has been a popular subject of study in many fields, particularly the artificial intelligence. Causality refers to the relation between two events when the occurrence of one event (the cause) leads to the occurrence of another one (the effect). It helps to predict the future; achieve goals on the basis of actions; diagnose problems and explain why something has happened.

Relating to this interest, building causality knowledge bases in realistic applications has been actively studied. For example, some researchers used the causality knowledge base to support the Question-Answering (QA) system in answering ‘why’ question (Chang, 2004; Girju, 2003; Pechsiri, 2007), which is among the most crucial forms of questions. And it is proved to be able to improve the QA performance. In general natural language understanding (NLU) tasks, the early work in (Heckerman, 1997) proposed a plan recognition method for discourse understanding using knowledge
about the cause and effect of an action. Furthermore, Saba pointed out that “NLU is, for the most part, a commonsense reasoning process at the pragmatic level” (Saba, 2006). And in his work, some reference must be resolved by recourse to causality knowledge. For instance, in the sentence “John shot a policeman, he immediately fell down”, we could infer that “he” refers to “a policeman” relied on the causal relation “\( \text{Shot}(x,y) \rightarrow \text{Fall\_down}(y) \)”. However, these knowledge-intensive systems result in a bottleneck due to the high cost of building and maintaining a huge knowledge base. To address this issue, many researchers have been concerned with automatic acquisition of causal relations between event expressions (typically verbs or verb phrases). On one hand, some approaches depended on predefined linguistic patterns employed in supervised learning models (Girju, 2003; Inui, 2005; Pechsiri, 2007). They identified causality knowledge from annotated closed corpora, and it is limited to scale up. On the other hand, unsupervised methods proposed heuristic statistical scores to evaluate potentially related events (Torisawa, 2003, 2006; Beamer, 2009; Riaz, 2010), but mostly have an unsatisfactory accuracy below 60%.

Although these corpus-based approaches for causality knowledge acquisition have considerable potential, most reviewed researches escaped from an important aspect in event relation acquisition that each event has arguments. They have just focused on identifying causality expressions without mentioning how to extract the arguments of the event-pairs and if so, they mainly relied on manually annotated corpus (Pechsiri, 2007; Inui, 2005) or extracted argument-shared structures (Torisawa, 2003, 2006). So, the major challenge in causality knowledge acquisition is the need for more generic methods that would acquire massive causal relations from unannotated corpus and learn the argument structures of event-pairs.

Motivated by this background, in this paper, we propose an approach to mine causal relations between events with argument structures from the Web. And our work involves three subtasks: (1) the identification of causality expressions; (2) the extraction of cause and effect pairs (3) the measure of extracted relations. We firstly use explicit causal connective markers as linguistic cues to discover causality relations. Then, event-pairs with the predicate-argument structure are extracted based on local dependency parse trees. And finally, we propose a statistical score \( S \) to measure the causal association between potential related events, and prune relations with low \( S \) value. Experimental results demonstrate the effectiveness of our approach, which had a precision around 80%.

The remainder of this paper is organized as follows. In Section 2, we present relevant previous work and point out the shortcomings of current researches. Section 3 describes the proposed method for mining and pruning causal relations from the Web. In Section 4, experiments are designed to evaluate our approaches. And finally, we conclude in Section 5.

## 2 Previous Work

The early works attempted to extract causal relations using knowledge-based inference technologies (Kaplan, 1991) These studies were based on hand-coded, domain-specific knowledge bases which are difficult to scale up for realistic applications. Recently there has been increasing interest in automatic causality extraction from texts, which can be classified into two approaches: the pattern-based approach and statistic-based approach.

Existing statistical methods for causality acquisition used one or more distribution characteristics of two events in the text. These major features are: (1) Co-Occurrence feature: the cause event and effect one may co-occur frequently; (2) Object-Sharing feature: the related two events may share a common participant; (3) Temporal feature: the cause event occur before (or simultaneously with) the effect event; (4) Distance feature: the two events may appear inside locally coherent text (in the same sentence particularly).
The pioneering work in (Torisawa, 2003) constructed a statistical model for extracting commonsense inference rules from coordinate verb phrases based on Co-Occurrence feature and Object-Sharing feature. Each rule was selected from the highest combined probability of two verbs and their shared participant. Torisawa further extended his work in (Torisawa, 2006) by emphasizing the occurrence frequency of a single verb, which indicated how generic the meaning of the verb is. The precision of the improved model achieved 60%.

Other unsupervised approaches focused on special data sets or special type of events. Beamer applied a statistical measure, Causal Potential, on a text corpus of screen plays where the verb events are already temporally ordered and annotated (Beamer, 2009). This measure combined Co-Occurrence feature and Temporal feature, which was calculated by point-wise mutual information (PMI) and directional bigram frequencies. Riaz identified causal relations between scenario-specific events in two phases (Riaz, 2010): they firstly mined event dependencies using the measure Effect-Control-Dependency, which is derived from PMI, and then identified the direction of the causal relationship (Cause and the Effect roles). The F-Measure of experiments on two sets of web news articles is respectively 52% and 60%.

Overall, aforementioned unsupervised approaches have achieved around 45-60% accuracy when determining whether or not two events are in a causal relationship. For there are various other types of relations between events including temporal, entailment, etc., it is difficult to embody a specific semantic relation just using statistical features. Instead, more researchers utilized generic lexico-syntactic co-occurrence patterns to mine causal knowledge.

Khoo manually constructed a set of graphical patterns that indicate the presence of a causal relation in sentences, and which part of sentence represents the cause or the effect (Khoo, 2000). These patterns are matched with the syntactic parse trees of sentences, and the parts of the parse tree that match with the slots in the patterns are extracted as the cause or the effect. Khoo applied 68 graphical patterns on 100 medical abstracts in the Medline database [MEDINE 2001], and reported the unsatisfactory result with an accuracy of around 50%. This low-precision problem requires an additional component for pruning extracted relations. A popular way is to incorporate a classifier trained with supervision (Zhang, 2014).

Girju automatically discovered the pattern <NP1 Verb NP2>, where the verb is a synonym of cause (such as product) reflecting the causal relationship between events expressed by noun phrases. To resolve potential ambiguity of the verb, this work trained a C4.5 decision tree to learn the semantic constraints of NP1 and NP2. The precision of this classifier is 65.6%. Chang also focused on the nominal event expressions, which used the lexical NP pairs as the pattern, and exploited the probability distribution of NP1 and NP2 to quantify the co-occurrence preferences of both phrases in a large corpus (Chang, 2004). The Naïve Bayesian (NB) model used for disambiguation had a precision of 81%.

More researches aimed to acquire relationships between verbs or verb phrases indicating the causal event and effect one. Inui used explicit connective marker tame, such as “because”, “since”, “as the result”, etc., to discover causal relation from two adjacent sentences (Inui, 2005). In this work, Inui further classified the causal relation into four subtypes mainly based on event-agents’ volitionality, which is learned by the Support Vector Machine (SVM) model. The result reported was satisfactory with the precision of about 85%. (Pechsiri, 2007) used verb-pair rules learnt by two different machine learning techniques (NB and SVM) to identify causality from multiple Elementary Discourse Units. The average precision of both models exceed 80%.

For the current supervised learning of causal relation classifier, causality-annotated corpus is required. During these learning procedures, the participants (subject and object) of an event and their semantic information are often adopted as important features. However, most reviewed researches have focused on identifying causality expressions without mentioning how to extract the arguments of
the events and if so, they mainly relied on manually annotated corpus. As we all know, the construction of such corpus would take much effort. So, a major challenge in causality knowledge acquisition is the need for more generic methods that would acquire massive causal relations from unannotated corpus and learn the relations between verbs with varied argument structures. This issue would be addressed in this paper.

3 Approach

![Three-phased causality acquisition approach](image)

The method we explore in this paper is illustrated in Figure 1. The overall process has three phrases: we firstly use lexico-syntactic patterns not only to recognize causal relations from the web, but also to identify pairs of event expressions; then, we extract the predicate-argument structure of each event expression based on its dependency parse tree in local scale; in the final phase, we propose a statistical score $S$ to measure the causal association between potential related events, and prune relations with low $S$ value.

3.1 Identifying Causal Relations from the Web

From a linguistic aspect, causality in natural language is expressed either implicitly or explicitly. The difference is that implicit causal relation is expressed without obvious cue phrases. In order to identify causal relations from the web by the search engine, we use explicit connective markers as linguistic cues.

In Kim’s work (Kim, 2007), the syntactic tags of the cue phrase are: (1) Causal verb, e.g. allow, cause, lead to, or contribute to; (2) Prepositional, e.g. due to, or as a consequence; (3) Subordinate, e.g. because or as; (4) Adverbial, e.g. subsequently or consequently; and (5) Noun, e.g. the cause of or the effect of. There is a similar causal expression classification performed on Chinese text. The following examples show Chinese causal relations categorized according to part of speech of the cue phrases.

a) Conjunction

- 因为队员受伤，所以主办方推迟了比赛。

  *Because* the player get hurt, *(so)* the organization postponed this play.
b) Causal Verb
   — 地震引起了海啸。
   The earthquake caused tidal waves.

c) Casual Noun
   — 船下沉的原因是超载。
   The cause of the boat accident is overloading.

According to the previous research experience, we focus on the most frequent and less ambiguous
cue phrase conjunction, and incorporate specific causal verbs to improve the recall of the acquired
knowledge. Note that a pair of causal conjunctions may co-occur in a Chinese sentence, such as the
example in (a), while it is not allowed in English. In order to concentrate on our goal, we use the
patterns to indicate the boundary of cause event and effect event, as well as to indicate the causal
semantic relations.

Our lexico-syntactic patterns contain a pair of connective markers being correlated with each other
and an end mark indicating the boundary of the causal expression, which are uniformly expressed as
follows.

\[
\text{C\_CON\_Marker} \quad [\star] \quad \text{E\_CON\_Marker} \quad [\star] \quad \langle \text{End\_Marker}\rangle
\]

In this pattern, \text{C\_CON\_Marker} is a causal conjunction, while \text{E\_CON\_Marker} is the corresponding
conjunction or a causal verb. Event expressions connected by these two markers are generally short
phrases or clauses. \text{End\_Marker} indicates a punctuation or an empty word such as interjection,
auxiliary word, etc. While \text{\star} is a wildcard, \text{[\star]} matches from one to \text{N} arbitrary words. \text{N} is used as
a window of event expressions, and it is assigned an empirical value indicating the maximum number
of words in a regular event mention. The parts of a sentence which match with these two slots are
extracted as a cause event and its effect event.

When we perform the extraction process, this pattern is automatically instantiated with given word
lists to generate query terms. Issuing these query terms, we take advantage of the search engine (such
as Google) to retrieve potential causal relations from the Web. And subsequently, we extract pairs of
event expressions from corresponding snippets matching with the event-slots.

### 3.2 Extracting Arguments of Causal Events

In the previous phase, co-occurrence pairs of event expressions are acquired from the web corpus.
Next, we extract the events’ predicate-argument structures (i.e., “[Subj] [Pre] [Obj]” instances) relied
on dependency syntactic structures of event expressions.

This idea is inspired by the event extraction work in (Zhao, 2008), which recognized event-
arguments mainly based on the dependency-path feature employed in a maximum entropy classifier.
Compared with the well-known phrase-structure, the dependency structure of a sentence is more likely
to reflect the semantic relations between contiguous or noncontiguous words. And we further find that,
the dependency structure can map to our event-argument structure in limited corpus, in spite of its
weakness in generic semantic role labeling tasks (Xue, 2008). So, dependencies are used as the
syntactic theory of choice.

Here, we use the dependency representations proposed in (NLP Toolkit 2011) to describe our
method. As shown in Figure 2, a dependency tree composes of some contiguous words in a sentence.
Every word in sentence can be viewed as a node of tree. Two nodes holding a dependency relation
constitute a dependency pair, in which one node is the head (e.g. “criticize”) while the other is a
dependent (e.g. “student”). A dependency relation is represented by a directed arc pointing from the
head to the dependent with a functional category label (e.g. “SBV”). So, a dependent can be viewed as
a child node of its head. The core word has a “HED” dependency to the virtual root, and we define it the level-1 node.

![Dependency tree of an event expression](image)

**Figure 2:** Dependency tree of an event expression

In the dependency tree of an event expression, which refers to a verb phrase or a clause here, the head word is a verb. It’s intuitive that the head verb is the predicate of an involved event. And the distribution of its arguments in dependency paths has some regularity. It’s suggested by the observation that we can always find the subject and object in direct dependents (level-2 nodes) of the head, or in the level-3 nodes. We summarize these strong regularities as unambiguous rules, examples of which are demonstrated in Table 1.

| Dep. Relation | Rule Instance |
|---------------|---------------|
| SBV           | Rule1. if node.pos=“n” or node.cont∈ “personal pronouns” then node is Subj |
| VOB           | Rule2. if node.pos=“n” or node.cont∈ “personal pronouns” then node is Obj |
| ADV           | Rule3. if node.pos=“p” and child_node.relation=“POB” then |
|               | if node.cont=“被(by)” then child_node is Subj |
|               | else child_node is Obj |
|               | Rule4. if node.pos= “v” and node.cont! = “被(by)” then node is Pre.Mod |
| ATT           | Rule5. if node.pos=“u” and child_node.relation=“DE” then |
|               | if child_node.pos=“n” or child_node.cont∈ “personal pronouns” then |
|               | child_node is Subj |
|               | else child_node is Pre.Mod |
| QUN           | Rule6. Pre.Mod |
| VV            | Rule7. if child_node.relation=“VOB” then node is Coo_Pre |
| CMP           | Rule8. if node.pos=“v” or node.pos=“adj” then node is Obj.CMP |
| COO           | Rule9. if child_node.relation=“LAD” or child_node.relation=“PUN” then node is Coo_Pre |
| MT            | Rule10. Pass over |

**Note:** A node in our rules has three attributes: “*.cont” indicates the corresponding word; “*.pos” indicates part-of-speech of the word; and “.relation” indicates the dependency relation from this word.

**Table 1:** Examples of rules for event-arguments extraction

In Table1, the first column shows possible categories of dependency relations from the head, and rules on corresponding dependents are listed in the second column. These rules mainly use features about single node (including the word’s part-of-speech) and relations between dependency pairs. These rules are also applied to some special phrase structures such as passive structure (e.g. rule3 and rule 4) and coordinate construction (e.g. rule 7 and rule 9).
During the procedure of recognizing event-arguments, we firstly extract the head verb as the predicate, traverse its child nodes and make corresponding rules effective according to the dependency category. From the instance sentence in Figure 1, we extract an event structure “[Subj teacher] [Pre criticize] [Obj student]” successively using rule 1, rule 3, and rule 10.

3.3 Measuring Causal Association

The lexico-syntactic patterns for causality acquisition define point-wise causal assertions. If pattern instances are found in texts, the extracted event pairs suggest but not confirm a causal relation. It may happen that these detected relations are accidental or are only valid in the given contexts. That is, they cannot be considered commonly agreed. In this phase, we perform a statistical analysis over event frequencies to assess and prune candidate causal relations.

To score the pairs for causal association, we propose a measure called Causal Strength which gauges how likely these two events are to be in a causal relationship without prior knowledge of any context. Causal Strength is calculated via the following formula:

$$S(e_1, e_2) = \log\left(\frac{P(e_2 | e_1)}{P(e_2)}\right) + \log\left(\frac{P(e_1 | e_2)}{\max P(e_1, e_2)}\right)$$

There are two main intuitions behind our causal strength $S$. The first term comes from the notion of probabilistic causation which defines it in terms of the causal event’s occurrence increasing the probability of the result event (Mellor, 1995). Thus $S(e_1, e_2)$ has high values when $P(e_2 | e_1) > P(e_2)$ and has low values when $P(e_2 | e_1) < P(e_2)$. It is consistent to the measure Point-Wise Mutual Information (PMI), which is used to capture dependencies between variables. However, PMI has the disadvantage of giving higher weights to strongly dependent but rare events.

This problem is addressed by the second term. The second term comes from the assumption that two events with higher co-occurrence frequency are more likely to be related. Hence, more frequent event pairs are given higher score. In this term, we emphasize the contribution of the cause event by using most frequent pair $(e_1, e)$ in the denominator.

Satisfying both of these intuitions results in high values of $S$, while lacking in one or both of them lowers the value of $S$. We get rid of event pairs with lower $S$ value than given minimum threshold.

4 Experiments

In this section, we evaluate the effectiveness of our causality acquisition approach and report the experimental results.

4.1 Experimental Settings

For our experiments, we manually selected 7 frequent causal conjunctions and 3 causal verbs listed in Table 2. These cues compose 16 pairs of markers in regular collocation, which are instantiated in extraction patterns. And we use 7 as the empirical value of the event window $N$.

Because the Google Search Engine returns at most 1000 items of users’ retrieval results, we employ a concrete verb in causal event expression to get more focused relations. That is, a wildcard between Cause_Marker and Effect_Marker is instantiated by a verb, and the patterns are used to identify its effect events. For this issue, we built a lexicon of over 10,000 common verbs with transitive labels, obtaining 8,387 transitive verbs and 4,732 intransitive. We send the pattern instances, as query items, to the Google Search Engine and download relevant texts returned from the Web.

As a preliminary filter, we use End_Marker to delete the useless suffix from these extracted causal expressions. And sentences in which the length of effect event is more than 7 are discarded. The final
The corpus consists of 1,960,000 sentences. We call it the Causal Corpus in the following experiments.

| Cause_Marker | Effect_Marker | End_Marker |
|--------------|---------------|------------|
| 因 (because of) | 而 (so) | 。 (period) |
| 因为 (since) | 因而 (thus) | ； (semicolon) |
| 由于 (due to) | 所以 (therefore) | 的 (auxiliary word) |
|              | 因此 (hence) | 了 (auxiliary word) |
|              | 引起 (cause) | 着 (auxiliary word) |
|              | 导致 (lead to) | 吧 (auxiliary word) |
|              | 使得 (make) |            |

Table 2: Wordlists for causality extraction patterns

4.2 Effectiveness of Event-Argument Extraction

We evaluate event-argument extraction approach by comparing with two other automatic methods used in previous works. One is relied on semantic role labeling (SRL) technique to identify the subject (Arg0) and the object (Arg1) referred in (Riaz, 2010). The other one used structure-mapping rules based on whole dependence parser, which is similar to (Khoo, 2000). The difference between the Whole Parser method and Local Parser method is whether parsing the causal expression or respectively parsing the two event expressions.

In order to reduce evaluation costs, we randomly select 5,000 sentences from the Causal Corpus as the test data-set. Using a Chinese NLP Toolkit (2011), we get the whole dependency trees, local dependency structures and semantic role labeling results of these sentences. In our approach, we use 36 predefined rules to extract event-argument structures. We note that, if an event expression contains multiple contiguous verbs, then its dependency tree is usually ambiguous. To improve the precision, we remove some obvious wrong dependency trees using heuristic rules.

| Method | Precision | Recall | F-Measure |
|--------|-----------|--------|-----------|
| Local Parser | 93.7% | 90.5% | 92.1% |
| Local SRL | 62.0% | 53.8% | 57.6% |
| Whole Parser | 84.0% | 65.0% | 73.3% |

Table 3: A comparison of performance of three event-arguments extraction approaches

The Comparison result is shown in Table 3. For SRL is a higher level natural language understanding task than syntactic parsing, it is more difficult to achieve satisfactory effectiveness. And current SRL technique is not robust enough to apply in realistic application. So, The Local SRL method just has the lowest F-Measure of 57.6%, although the SRL result directly reflects event-arguments of verbs. Compared with Whole Parser method, the locally dependency parsing extensively improves both the accuracy and recall of event arguments extraction task, and its F-Measure achieved 92.1%.

4.3 Effectiveness of the Causal Association Measure

In this experiment, we test the effectiveness of the measure for Causal Association. All the probabilities used in our method are estimated by maximum likelihood estimation from event frequencies in Causal Corpus. To address the data-sparse problem, when we calculate the frequency of an event, we consider those with synonymous arguments refer to the same event. For example, “[Subj teacher][Pre criticize][Obj student]” and “[Subj master][ Pre blame][Obj student]” are
synonymous events. Considering the evaluation cost, here we selected 20 groups of synonymous verbs (144 verbs in all) which act as the predicate in the cause event. And we get 6,307 event pairs.

To see the effect of our measure, we compare $S$ with another score $PMI$. And we calculate these scores respectively using the whole argument structure (-E) or the head verb (-V) as an event.

![Figure 3: Comparisons between $S$ with $PMI$](image)

Figure 3 shows the performance of both the $S$ measure and $PMI$ measure. As the results, all the causal relations are ranked according to each score. This graph plots the precision of the top $N$ relations in the ranked list. It’s obvious that the measure $S$ outperformed the $PMI$ scores. And it further proved that the importance of Co-Occurrence feature referred in Section 2. The best precision and recall of $S$-E method achieved 89.3% and 83.0%, respectively.

5 Conclusion and future work

Motivated by the needs of generic methods to acquire specific relations between events, we explore an automatic three-phased approach. We take the causal relation as example in this paper. We first use lexico-syntactic patterns to not only recognize causal relations from the web text, but also identify pairs of event expressions. Then, we extract the predicate-argument structure of each event expression based on its dependency parser tree in local scale. At the last step, we propose a statistical score $S$ to measure the causal association between potential related events, and prune relations with low $S$ value. The experimental results have shown that (a) the use of local dependency tree extensively improves both the accuracy and recall of event-arguments extraction task; (b) our measure which is an improvement of $PMI$ has better performance.

There are two interesting directions in the future. First, identifying causality boundary automatically rather than just using separators in a pattern. Second, the three arguments referred in our work are not enough in some special cases. For example, the effects of events “he works carelessly” and “he works carefully” are commonly different. To distinguish from these causes, we will introduce new arguments as the solution.
Acknowledgement

This work was supported by the NSFC (No.61370025), 863 projects (No.2011AA010703 and No.2012AA012502), 973 project (No.2013CB329606), and the Strategic Leading Science and Technology Projects of Chinese Academy of Sciences (No.XDA06030200).

Reference

Beamer, B. and Girju, R. (2009). Using a Bigram Event Model to Predict Causal Potential. Computational Linguistics and Intelligent Text Processing(CICLing) 2009, Lecture Notes in Computer Science(LNCS) 5449, pp.430-441.

Chang, D. S. and Choi, K. S. (2004). Causal Relation Extraction Using Cue Phrase and Lexical Pair Probabilities. IJCNLP 2004, Lecture Notes in Computer Science, 3248, pp. 61-70.

Girju, R. (2003). Automatic detection of causal relations for question answering. In Proceedings of the ACL 2003 workshop on Multilingual summarization and question answering, pp.76-83.

Heckerman, D., Meek, C. and Cooper, G. (1997). A Bayesian approach to causal discovery. Tech. rep., Microsoft research Advanced Technology Division, Microsoft Corporation, Technical Report MSR-TR-97-05.

Inui, T., Inui, K. and Matsumoto, Y. (2005). Acquiring causal knowledge from text using the connective marker tame. ACM Transactions on Asian Language Information Processing (TALIP), 4(4), pp. 435-474.

Kaplan, R. M. and Berry-Rogghe, G. (1991). Knowledge-based acquisition of causal relationships in text. In Knowledge Acquisition, 3(3), pp.317-337.

Khoo, C. S. G., Chan, S. and Niu, Y. (2000). Extracting causal knowledge from a medical database using graphical patterns. In Proceedings of The 38th Annual Meeting of The Association for Computational Linguistics (ACL), pp.336-343.

Kim, S., Bracewell, R. H. and Wallace, K. M. (2007). A framework for automatic causality extraction using semantic similarity. In ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, USA.

Pechsiri, C. and Kawtrakul, A. (2007). Mining Causality from Texts for Question Answering System. IEICE TRANSACTIONS on Information and Systems, 90(10), pp. 1523-1533.

Riaz, M. and Girju, R. (2010). Another Look at Causality: Discovering Scenario-Specific Contingency Relationships with No Supervision. In Proceedings of the Fourth International Conference on Semantic Computing(ICSC), pp. 361-368.

Saba, W. S. (2006). Language, logic and ontology: uncovering the structure of commonsense knowledge. arXiv: cs.AI/0610067 v3.

Torisawa, K. (2003). An Unsupervised Learning Method for Commonsensical Inference Rules on Events. In Proceedings of the Second CoLogNet-ELsNET Symposium.

Torisawa, K. (2006). Acquiring Inference Rules with Temporal Constraints by Using Japanese Coordinated Sentences and Noun-Verb Co-occurrences. In Proceedings of the HLT/NAACL 2006, pp.57–65.

Xue, N. W. (2008). Labeling Chinese Predicates with Semantic Roles. Association for Computational Linguistics(CL), Volume 34, Number 2, pp.225-255.

Zhang, P., Zhou, C., Wang, P., Guo, B., Zhu, X.Q. and Guo, L. (2004). E-tree: An efficient indexing structure for ensemble models on data streams. IEEE Transactions on Knowledge and Data Engineering.

Zhao, Y. Y., Qin, B., Che, W. X. and Liu, T. (2008). Research on Chinese Event Extraction. Journal of Chinese Information Processing, vol.22, No.1.