Recognizing Implicit Discourse Relations through Abductive Reasoning with Large-scale Lexical Knowledge

Jun Sugiura, Naoya Inoue, and Kentaro Inui

Tohoku University, 6-3-09 Aoba, Aramaki, Aoba-ku, Sendai, 980-8579, Japan
{jun-s,naoya-i,inui}@ecei.tohoku.ac.jp

Abstract. Discourse relation recognition is the task of identifying the semantic relationships between textual units. Conventional approaches to discourse relation recognition exploit surface information and syntactic information as machine learning features. However, the performance of these models is severely limited for implicit discourse relation recognition. In this paper, we propose an abductive theorem proving (ATP) approach for implicit discourse relation recognition. The contribution of this paper is that we give a detailed discussion of an ATP-based discourse relation recognition model with open-domain web texts.

Keywords: Discourse Relation, Abductive Reasoning, Lexical Knowledge, Association Information

1 Introduction

Discourse relation recognition is the task of identifying the semantic relationship between textual units. For example, given the sentence *John pushed Paul towards a hole.*(S1) *Paul didn't get hurt.*(S2), we identify a contrast relationship between textual units (S1) and (S2). Discourse relation recognition is useful for many NLP tasks such as summarization, question answering, and coreference resolution.

The traditional studies on discourse relation recognition divided discourse relations into two distinct types according to the presence of discourse connectives between textual units: (i) an *explicit* discourse relation, or discourse relation with discourse connectives (e.g. *John hit Tom because he is angry.*). (ii) an *implicit* discourse relation, or discourse relation without discourse connectives (e.g. *John hit Tom. He got angry.*). Identifying an implicit discourse relation is much more difficult than identifying an explicit discourse relation. In this paper, we focus on the task of implicit discourse relation recognition.

Conventional approaches to implicit discourse relation recognition exploit surface information (e.g. bag of words) and syntactic information (e.g. syntactic dependencies between words) to identify discourse relations [1, 2, etc.]. However, the performance of these models is severely limited as mentioned in Sec. 2.2. We believe that the problem of these approaches is two fold: (i) they do not capture causality between the events mentioned in each textual units, and (ii) they

do not capture the factuality of these events. We believe that this information plays a key role for implicit discourse relation recognition. Suppose we want to recognize a contrast relation between S_1 and S_2 in the first example. To recognize the discourse relation, we need to at least know the following information: (i) commonsense knowledge: *pushing into a hole* usually causes *getting hurt*; (ii) factuality: *Paul did not get hurt.* Finally, combining (i) and (ii), we need to recognize the unusualness of discourse: something against our commonsense knowledge happened in S_2. As described in Sec. 3, our investigation revealed that we have several patterns of reasoning and need to combine several kinds of reasoning to identify a discourse relation.

Motivated by this observation, we propose an abductive theorem proving (ATP) approach for implicit discourse relation recognition. The key advantage of using ATP is that the declarative nature of ATP abstracts the flow of information away in a modeling phase: we do not have to explicitly specify when and where to use a particular reasoning. Once we give several primitive inference rules to an ATP system, the system automatically returns the best answer to a problem, combining the inference rules.

In this paper, we attempt to answer the following open issues of ATP-based discourse relation recognition: (i) does it really work on real-life texts?; (ii) does it work with a large knowledge base which is not customized for solving target texts? The contribution of this paper is as follows. We give a detailed discussion of an ATP-based discourse relation recognition model with open-domain web texts. In addition, we show that our ATP-based model is computationally feasible with a large knowledge base.

The structure of this paper is as follows. In Sec. 2, we describe abduction and give an overview of previous efforts on discourse relation recognition. In Sec. 3, we describe our ATP-based discourse relation recognition model. In Sec. 4, we report the results of pilot evaluation of our model with large lexical knowledge.

2 Background

2.1 Weighted abduction

Abduction is inference to the best explanation. Formally, logical abduction is defined as follows:

- **Given:** Background knowledge B, and observations O, where both B and O are sets of first-order logical formulas.
- **Find:** A *hypothesis* (or *explanation*, *abductive proof*) H such that $H \cup B \models O, H \cup B \not\models \perp$, where H is a conjunction of literals. We say that p is *hypothesized* if $H \cup B \models p$, and that p is *explained* if $(\exists q)\ q \rightarrow p \in B$ and $H \cup B \models q$.

Typically, several hypotheses H explaining O exist. We call each of them a *candidate hypothesis* and each literal in a hypothesis an *elemental hypothesis.* The goal of abduction is to find the best hypothesis among candidate hypotheses

by a specific measure. In the literature, several kinds of evaluation measure have been proposed, including cost-based and probability-based [3, 4, etc.].

In this paper, we adopt the evaluation measure of *weighted abduction*, which is proposed by Hobbs et al. [4]. In principle, the evaluation measure gives a penalty for assuming specific and unreliable information but rewards for inferring the same information from different observations. We summarize the primary feature of this measure as follows (see [4] for more detail):

- Each elemental hypothesis has a positive real-valued cost;
- The cost of each candidate hypothesis is defined as the sum of costs of the elemental hypotheses;
- The best hypothesis is defined as the minimum-cost candidate hypothesis;
- If an elemental hypothesis is explained by other elemental hypothesis, the cost becomes zero.

2.2   Related work

Discourse relation recognition is a prominent research area in NLP. Most researchers have primarily focused on explicit discourse relation recognition, employing statistical machine learning-based models [5, 6, etc.] with superficial and syntactic information. The performance of explicit discourse relation recognition is comparatively high; for instance, Lin et al. [7] achieved an 80.6% F-score.

The performance of implicit discourse relation recognition is, however, relatively low (25.5% F-score). Most existing work on implicit discourse relation recognition [1, 2, etc.] extend the feature set of [5] with richer lexico-syntactic information. For example, Pitler et al. [2] exploit a syntactic parse tree and sentiment polarity information of words contained in textual units to generate a feature set. However, the performance is not as high as a practical level.

An abductive discourse relation recognition model is originally presented in Hobbs et al. [4]. However, Hobbs et al. [4] reported the results in a fairly closed setting: they tested their model on two test texts with manually encoded background knowledge which is required to solve the discourse relation recognition problems that appear in two texts. Therefore, it is an open question whether the abductive discourse relation recognition model works in an open setting where the wider range of real-life texts and large knowledge base are considered.

3   Abductive theorem proving for discourse relation recognition

In this section, we describe our discourse relation recognition model. We employ ATP to recognize a discourse relation. Given target discourse segments, we abductively prove that there exists a coherence relation (i.e. some discourse relation) between the discourse segments using background knowledge. We axiomatize (i) definition of discourse relations and (ii) lexical knowledge (e.g. causal knowledge of events) in the background knowledge, which serve as a proof of the existence of a coherence relation.

The motivation of using <u>abductive</u> theorem proving is that we can <u>assume</u> a proposition with the cost even if we fail to find a complete proof of a coherence relation between discourse segments, as mentioned in Sec. 2. By choosing the minimum-cost abductive proof, we can identify the most likely discourse relation.

We first show how to axiomatize the definition of discourse relations (Sec. 3.1). We then conduct an example-driven investigation of lexical knowledge which is required to solve a few real-life discourse relation recognition problems in order to identify a type of lexical knowledge needed for an ATP-based recognition model (Sec. 3.2). In Sec. 3.2, we make sure that our developed theory works on a general-purpose inference engine as we expected. We use the lifted first-order abductive inference engine *Henry*, which is developed by one of the authors.[8] To perform deeper analysis of the inference results, we also improved the existing visualization module provided by Henry. The inference engine and visualization tool are publicly available at `https://github.com/naoya-i/henry-n700/`.

3.1   Axiomatizing definitions of discourse relations

We follow the definitions of discourse relations provided by Penn Discourse Tree-Bank (PDTB) 2.0 [9], a widely used and large-scale corpus annotated with discourse relations.[1] The PDTB defines four coarse-grained discourse relations, but it is still rather difficult to identify all discourse relations. Therefore, we adopt two-way classification: whether it is adversative (*Comparison* in PDTB) or resultative (*Temporal*, *Contingency*, *Expansion* in PDTB). Because a resultative relation can be regarded as relations other than an adversative relation, we first axiomatize the definition of adversative and then consider the other relation.

According to the PDTB Annotation Manual [12], an adversative relation consists of two subtypes: *Concession* and *Contrast*. These subtypes are defined below, respectively.

Concession One of the arguments describes a situation A which causes C, while the other asserts (or implies) ¬ C. One argument denotes a fact that triggers a set of potential consequences, while the other denies one or more of them.
Contrast Arg1 and Arg2 share a predicate or a property and the difference is highlighted with respect to the values assigned to this property.

The condition of *Concession* can be described as the following axiom:

$$event(e_1, type1, F, x_1, x_2, x_3, s_1) \ \wedge \ event(e_2, type2, NF, y_1, y_2, y_3, s_2)$$
$$\wedge \ cause(e_1, F, e_2, F) \ \wedge \ event(e_u, type2, EF, y_1, y_2, y_3, s_u) \ \Rightarrow \ Adversative(s_1, s_2).$$

This axiom says that if event e_1 occurs in segment s_1 (roughly corresponding to Arg1 in PDTB) and that event is expected to cause an event of $type2$ while such an event of $type2$ does actually not occur in segment s_2 (roughly corresponding to Arg2 in PDTB), then the discourse relation between s_1 and s_2 is Adversative. The examples using this type of axiom to recognize discourse relations will

---

[1] For other corpora, see [10, 9, 11, etc.].

be mentioned later. On the other hand, a typical pattern where the *Contrast* relation holds can be described, for example, as follows:

$$value(e_1, Pos, s_1) \ \wedge \ value(e_2, Neg, s_2) \ \Rightarrow \ Adversative(s_1, s_2).$$

This axiom says that when the sentiment polarity of e_1 in segment s_1 is Positive and the sentiment polarity of e_2 in segment s_2 is Negative, the discourse relation between s_1 and s_2 is Adversative. The examples of axioms described here represent formation conditions of Adversative and take some variation due to their value of factuality or sentiment.

Furthermore, these axioms can represent conditions of Resultative. For instance, if the sentiment polarity of e_2 in segment s_2 is the same as that of segment s_1, then the discourse relation between s_1 and s_2 is Resultative as below:

$$value(e_1, Pos, s_1) \ \wedge \ value(e_2, Pos, s_2) \ \Rightarrow \ Resultative(s_1, s_2).$$

In total, we created 21 axioms for the definition of discourse relations.

Finally, we add the following axioms to connect the definition of discourse relations with the existence of a coherence relation between discourse segments:

$$Adversative(s_1, s_2) \Rightarrow CoRel(s_1, s_2) \tag{1}$$

$$Resultative(s_1, s_2) \Rightarrow CoRel(s_1, s_2), \tag{2}$$

where $CoRel(s_1, s_2)$ indicates that there exists a coherence relation between segments s_1 and s_2. Given target discourse segments S_1, S_2, we prove $CoRel(S_1, S_2)$ using the axioms described above and lexical knowledge which is described in the next section.

We formally describe our meaning representation. First, we use $cause(e_a, f_a, e_c, f_c)$ to represent that event e_a with factuality f_a causes event e_c with factuality f_c. Second, we represent an event by using $event(e, t, f, x_1, x_2, x_3, s)$, where e is the event variable, t is the event type of e, f is the factuality of event e, x_1, x_2, x_3 are arguments of event and s is the segment which event e belongs to. Factuality of event e can take one of the following four values: F (Fact; e occurred), NF (NonFact; e did <u>not</u> occur), EF (Expected-Fact; e is expected to occur), and ENF (Expected-NonFact; e is expected <u>not</u> to occur). In addition, the value (sentiment polarity) of event e is represented as $value(e, v, s)$. v is either Pos (Positive) or Neg (Negative).

3.2   Example-driven investigation of lexical knowledge

Next, we manually analyzed a small number of samples for each discourse relation to investigate what types of knowledge are required to explain those samples and how to axiomatize them. In this paper, we manually convert each sample text into the logical forms, extracting main verbs in its matrix clauses as predicates.[2]

---

[2] In future work, we will exploit the off-the-shelf semantic parser (e.g. Boxer [13]) to automatically get the logical forms. Using automatic semantic parsers brings some challenges to us, e.g. how to represent the verbs in embedded clauses. We do <u>not</u> address these issues in this work, because we want to focus on the investigation of types of world knowledge that are required to identify discourse relations.
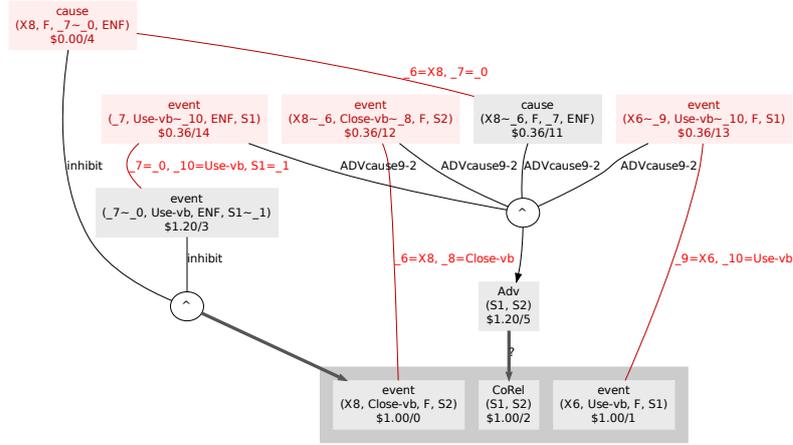
Fig. 1. Example of the abductive proof automatically produced by our system. The black directed edges: backward-chainings. The red undirected edges: unification. The labels of undirected red edges: unifier. The terms starting with a capital letter: constant; otherwise, variable. "$X \sim Y$": Y is unified with X. The grayed nodes: explained literals. The red nodes: hypothesized literals.

The dataset consists of text which we collect from the Web.[3] This website provides English texts for adult English learners as a second language. We collect 16 discourse segment pairs in which half of them can be regarded as Adversative and the others can be regarded as Resultative.

Let us take one of the simplest samples, example (2).

(2) S1: A lot of traffic once used Folsom Dam Road.
    S2: Right now, the road is closed.
    (Topic=Working, StoryID=174)

In this example, S_1 and S_2 are in the Adversative relation. While the Folsom Dam Road was once used by a lot of traffic, it is not usable now because it is closed. Something was once used but it is now unusable; therefore, the Adversative relation holds. This can be described as the following axiom:

Condition of Adversative

$$event(e_1, type1, F, x_1, x_2, x_3, s_1) \wedge event(e_2, type2, F, y_1, y_2, y_3, s_2)$$
$$\wedge\ cause(e_2, F, e_u, ENF) \wedge event(e_u, type1, ENF, x_1, x_2, x_3, s_1) \Rightarrow Adversative(s_1, s_2)$$

The causality relation between the "closed" event and the "unusable" event can be described as:

---

[3] http://www.cdlponline.org/

Relation between events

$$event(e_1, Use, ENF, x_1, x_2, x_3, s_1) \wedge cause(e_2, F, e_1, ENF)$$
$$\Rightarrow event(e_2, Close, F, y_1, x_2, y_3, s_2)$$

Note that we have $cause(e_2, F, e_1, NF)$ in the left-hand side of the axiom. We use this literal to accumulate the type of reasoning. In an abductive proof, we expect this literal to unify with an elemental hypothesis generated by the axioms of discourse relation.

Fig. 1 shows the result of applying the proposed model to example (2).[4] In Fig. 1, the observations consists of three literals; occurrence of event whose type is Use in segment S_1, occurrence of event whose type is Close in segment S_2, and $CoRel(S_1, S_2)$ which is a symbol of existence of some discourse relation between the segments.

To see how our model combines multiple pieces of knowledge, let us take another example.

(3) S1: Right now, the road is closed.
S2: Most of the people who used the road every day are angry.
(Topic=Working, StoryID=174)

The "closed" event causes the "unusable" event and the "unusable" event than further causes the "angry" event, which can be explained by combining the knowledge that being "unusable" is negative in sentiment polarity and the knowledge that a negative event may cause someone to be angry. These pieces of knowledge can be axiomatized as follows. The proof graph is shown in Fig. 2.

Condition of Resultative

$$event(e_1, type1, F, x_1, x_2, x_3, s_1) \wedge event(e_2, type2, F, y_1, y_2, y_3, s_2)$$
$$\wedge \ cause(e_1, F, e_u, EF) \wedge event(e_u, type2, EF, y_1, y_2, y_3, s_2) \Rightarrow Resultative(s_1, s_2)$$

Relation between events

$$event(e_1, Use, ENF, x_1, x_2, x_3, s_1) \ \wedge \ cause(e_2, F, e_1, ENF)$$
$$\Rightarrow event(e_2, Close, F, y_1, x_2, y_3, s_2)$$
$$event(e_1, Angry, EF, x_1, x_2, x_3, s_1) \ \wedge \ cause(e_2, f, e_1, EF)$$
$$\Rightarrow value(e_2, Neg, s_2) \ \wedge \ event(e_2, type, f, y_1, y_2, y_3, s_2)$$

Transitivity

$$cause(e_1, f_1, e_2, f_2) \ \wedge \ cause(e_2, f_2, e_3, f_3) \ \Rightarrow \ cause(e_1, f_1, e_3, f_3)$$

Polarity

$$value(e_1, Neg, s_1) \ \Rightarrow \ event(e_1, Use, ENF, x_1, x_2, x_3, s_1)$$

Through the investigation as illustrated above, we reached the conclusion that the axioms in Table 1 can recognize discourse relations for most examples.

---

[4] Throughout this paper, we omit the arguments of events from our representation in a proof graph for readability. Since we do not have a lexical knowledge between nouns and verbs, this simplification does not affect to the result of inference.
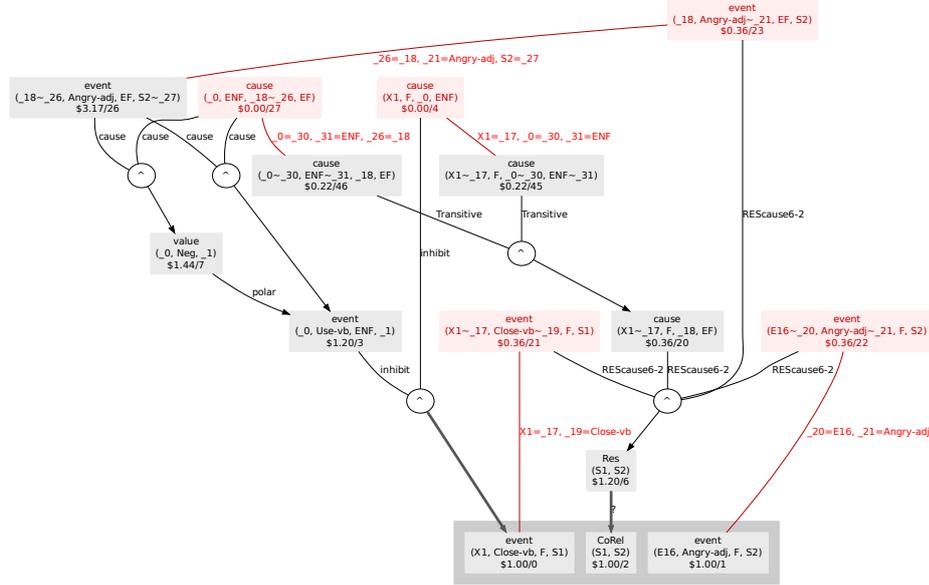
Fig. 2. Example of the abductive proof automatically produced by our system. See the description of Fig. 1.

4   Pilot large-scale evaluation

As mentioned in the previous section, our model assumes that axioms encoding lexical knowledge are automatically extracted from a large lexical resources (see "To be automatically acquired" axioms in Sec. 3.2.) In this section, we extract the axioms of causal relations and synonym/hyperonym relations from WordNet [14] and FrameNet [15], both popular and large lexical resources, and then apply our model to the example texts presented in Sec. 3. Regarding sentiment polarity, we plan to extract the axioms from a large-scale sentiment polarity lexicon such as [16] in future work.

We clarify that our primary focus here is the feasibility of our ATP-based discourse relation recognition model with a large knowledge base. The quantitative evaluation of our model (e.g. the predictive accuracy of discourse relations) is future work. Therefore, we first report how to incorporate WordNet and FrameNet axioms into our knowledge base (Sec. 4.1) and then preliminarily report the computational time of inference required to solve the example problems, showing some interesting output (Sec. 4.2).

Table 1. Set of axiom type.

Scale	Type of knowledge	Examples of axiom
Small-scale (Manually written)	Definitions of discourse relations	$event(e_1, type1, F, x_1, x_2, x_3s_1)$ $\wedge\ event(e_2, type2, F, y_1, y_2, y_3, s_2)$ $\wedge\ cause(e_2, F, e_u, ENF)$ $\wedge\ event(e_u, type1, ENF, x_1, x_2, x_3, s_1)$ $\Rightarrow\ Adversative(s_1, s_2);$ $value(e_1, Pos, s_1) \wedge value(e_2, Pos, s_2)$ $\Rightarrow\ Resultative(s_1, s_2)$
	Transitivity	$cause(e_1, f_1, e_2, f_2) \wedge cause(e_2, f_2, e_3, f_3)$ $\Rightarrow\ cause(e_1, f_1, e_3, f_3)$
Large-scale (Automatically acquired)	Causal relations	$event(e_1, Use, ENF, x_1, x_2, x_3, s_1) \wedge cause(e_2, F, e_1, ENF)$ $\Rightarrow\ event(e_2, Close, F, y_1, x_2, y_3, s_2);$ $event(e_1, Angry, EF, x_1, x_2, x_3, s_1) \wedge cause(e_2, f, e_1, EF)$ $\Rightarrow\ value(e_2, Neg, s_2) \wedge event(e_2, type, f, y_1, y_2, y_3, s_2)$
	Synonym/Hyponym	$event(e_1, Attack, F, x_1, x_2, x_3, s_1)$ $\Rightarrow\ event(e_1, Destroy, F, x_1, x_2, x_3, s_1)$
	Sentiment polarity	$value(e_1, Neg, s_1) \Rightarrow\ event(e_1, Die, F, x_1, x_2, x_3, s_1);$ $value(e_1, Pos, s_1) \Rightarrow\ event(e_1, Die, NF, x_1, x_2, x_3, s_1)$

4.1   Automatic axiom extraction from linguistic resources

We summarize the axioms extracted from WordNet and FrameNet in Table 2. For each resource, we extract two kinds of axioms. First, we generate axioms that map a word to the corresponding WordNet synset or FrameNet frame (Word-Synset, or Word-Frame types). The example WordNet axiom in Table 2 enables us to hypothesize that a "*die*"-typed event can be mapped to WordNet synset 200358431. Second, we also encode a semantic relation between synsets or frames. For example, the causal relation between *Getting* frame and *Giving* frame is encoded as the axiom in the Table. [5]

4.2   Results and discussion

We have tested our large-scale discourse relation recognition model on the randomly selected 7 texts as those presented in Sec. 3. We restricted the maximum

---

[5] Note that the mapping axioms have bi-directional implications. By using the bi-directional axioms, we can combine the knowledge from FrameNet and WordNet to perform a robust inference. For instance, we can do an inference like: *pass away* → *synsetA* → *die* → *FNDeath* if we do not have a direct mapping from *pass away* to *FNDeath*. Since the framework is declarative, we do not have to specify when and where to use a particular type of knowledge, which results in a robust reasoning.

Table 2. Axioms automatically extracted from WordNet and FrameNet.

Type	Resources	Example	# axioms
Word-Synset	WordNet	$event(e, Die, f, x_1, x_2, x_3, s)$ $\Leftrightarrow event(e, WNSynset200358431, f, x_1, x_2, x_3, s)$	169,362
Word-Frame	FrameNet	$event(e, Shoot, f, x_1, x_2, x_3, s)$ $\Leftrightarrow event(e, FNUseFireArm, f, x_1, x_2, x_3, s)$	10,358
Causal relations	WordNet relationsA[*1]	$event(e_1, WNSynset20036712, EF, x_1, x_2, x_3, s_1)$ $\wedge cause(e_1, EF, e_2, F)$ $\Rightarrow event(e_2, WNSynset200358431, F, y_1, y_2, y_3, s_2)$	35,440
	FrameNet relations[*2]	$event(e_1, FNGiving, EF, x_1, x_2, x_3, s_1)$ $\wedge cause(e_1, EF, e_2, F)$ $\Rightarrow event(e_2, FNGetting, F, y_1, y_2, y_3, s_2)$	6,584
Synonym/ Hyponym	WordNet relationsB[*3]	$event(e, WNSynset200060063, f, x_1, x_2, x_3, s)$ $\Rightarrow event(e, WNSynset200358431, f, x_1, x_2, x_3, s)$	177,837

*1: Causality, Entailment, Antonym. *2: IsCausativeOf, InheritsFrom, PerspectiveOn, Precedes, SeeAlso, SubFrameOf, Uses. *3: Meronym, Hyperonym.

number of backward-chaining steps to 2 due to the computational feasibility. For each problem, on average, the number of potential elemental hypotheses was 13,034 and the (typed) number of axioms that were used to generate candidate abductive proofs was 142. The time of inference required to solve each problem was 7.00 seconds on average.

Now let us show one of the proof graphs automatically produced by our system.[6] In Figure 3, we show the abductive proof graph for the following discourse:

S_1: Only 56 people died from the explosion,
S_2: but many other problems have been caused because of it.
　　(Topic=Activity, StoryID=241)

Although we suffer from the insufficiency of lexical knowledge, the abductive engine gave us the best proof where two segments are tied with a resultative relation. In the proof graph, *Die* and *CauseProblem* events are used to prove "event" literals hypothesized by the axiom of discourse relation. Note that the causal relation between these events is not proven but assumed with $0.36.

The overall results indicate that we now have a good environment to develop ATP-based discourse processing.

5   Conclusions

We have explored an abductive theorem proving (ATP)-based approach for implicit discourse relation recognition. We have investigated the type of axioms required for an ATP-based discourse relation recognition and identified five types of axioms. Our result is based on real-life Web texts, and no previous work has

---

[6] For the simplicity, we used only one axiom for the axiom of discourse relation.
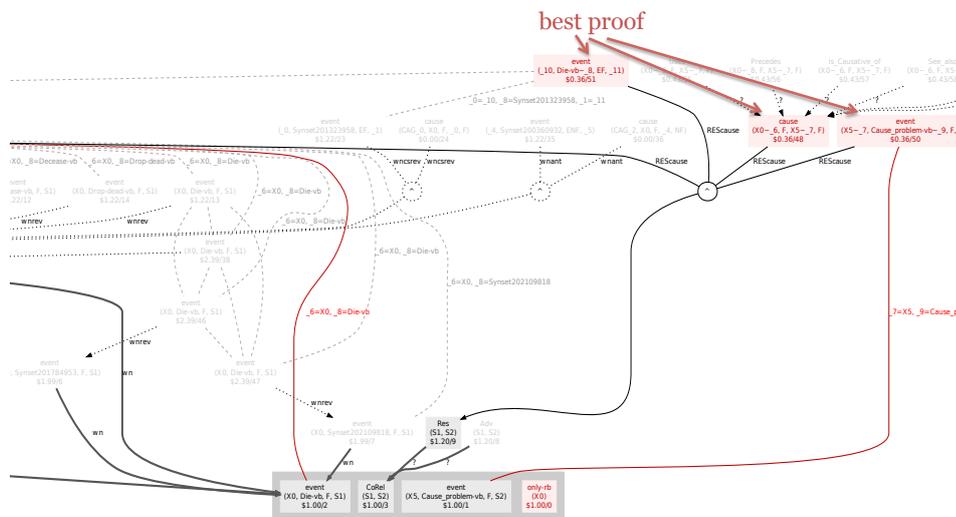
Fig. 3. Abductive proof with potential elemental hypotheses. The grayed-out nodes are those which are *potentially* included in the best proof, but not actually included in the best proof. Similarly, the dotted edges are *potential* explainer-explainee relationships between elemental hypotheses.

done an investigation in the same setting. Also, we have preliminarily evaluated our model with a large knowledge base. We have automatically constructed axioms of lexical knowledge from WordNet and FrameNet, which results in around four hundred thousand inference rules. Our experiments showed the great potential of our ATP-based model and that we are ready to develop ATP-based discourse processing in a real-life setting.

Our future work includes three directions. First, we will create a larger knowledge base, exploiting the linguistic resources which have recently become available. As a first step, we plan to axiomatize Narrative Chain [17], ConceptNet,[7] and Semantic Orientations of Words [16] to extend our axioms with the large knowledge resources. Second, we plan on applying the technique of automatic parameter tuning for weighted abduction [18] to our model. Third, we plan to create a dataset for abductive discourse processing, where we annotate simple English texts with some discourse phenomena including discourse relations and coreference etc. As the source text, we will use materials for ESL (English as Second Language) learners,[8] a set of syntactically and lexically simple texts, so that we can trace the detailed behavior of abductive reasoning process. About the semantic representation, we plan to use the Boxer semantic parser [13] to automatically get the event arguments.

---

[7] http://conceptnet5.media.mit.edu/

[8] http://www.cdlponline.org/

References

1. Lin, Z., Kan, M., Ng, H.: Recognizing implicit discourse relations in the penn discourse treebank. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. (2009) 343–351
2. Pitler, E., Nenkova, A.: Using syntax to disambiguate explicit discourse connectives in text. In: Proceedings of the ACL-IJCNLP 2009. (2009) 13–16
3. Charniak, E., Goldman, R.: Probabilistic abduction for plan recognition. Brown University, Department of Computer Science (1991)
4. Hobbs, J., Stickel, M., Appelt, D., Martin, P.: Interpretation as Abduction. Artificial Intelligence **63**(1-2) (1993) 69–142
5. Marcu, D., Echihabi, A.: An unsupervised approach to recognizing discourse relations. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. (2002) 368–375
6. Subba, R., Eugenio, B.D.: An effective discourse parser that uses rich linguistic information. In: NAACL. (2009) 566–574
7. Lin, Z., Ng, H., Kan, M.: A PDTB-Styled End-to-End Discourse Parser. Arxiv preprint arXiv:1011.0835 (2010)
8. Inoue, N., Inui, K.: Large-scale cost-based abduction in full-fledged first-order predicate logic with cutting plane inference. In: Proceedings of the 13th European Conference on Logics in Artificial Intelligence. (2012) 281–293
9. Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., Webber, B.: The Penn Discourse TreeBank 2.0. In: Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC). (2008) 2961–2968
10. Carlson, Lynn and Marcu, Daniel and Okurowski, Mary Ellen: RST Discourse Treebank, LDC2002T07. Number LDC2002T07, Linguistic Data Consortium (2002)
11. Wolf, F., Gibson, E., Fisher, A., Knight, M.: The Discourse GraphBank: A database of texts annotated with coherence relations. LDC (2005)
12. Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A., Robaldo, L., Webber, B.: The Penn Discourse Treebank 2.0 Annotation Manual. IRCS Technical Report (2007)
13. Bos, J.: Wide-Coverage Semantic Analysis with Boxer. In Bos, J., Delmonte, R., eds.: Proceedings of STEP. Research in Computational Semantics, College Publications (2008) 277–286
14. Fellbaum, C., ed.: WordNet: an electronic lexical database. MIT Press (1998)
15. Ruppenhofer, J., Ellsworth, M., Petruck, M., Johnson, C., Scheffczyk, J.: FrameNet II: Extended Theory and Practice. Technical report, Berkeley, USA (2010)
16. Takamura, H., Inui, T., Okumura, M.: Extracting semantic orientations of words using spin model. In: ACL. (2005) 133–140
17. Chambers, N., Jurafsky, D.: Unsupervised learning of narrative schemas and their participants. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. (2009) 602–610
18. Yamamoto, K., Inoue, N., Watanabe, Y., Okazaki, N., Inui, K.: Discriminative learning of first-order weighted abduction from partial discourse explanations. In: CICLing (1). (2013) 545–558