Acquiring and Generalizing Causal Inference Rules from Deverbal Noun Constructions

Shohei Tanaka[1]   *Naoaki Okazaki*[2,3]   *Mitsuru Ishizuka*[1]

(1) University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan
(2) Tohoku University, 6-3-09 Aramaki Aza Aoba, Aobaku, Sendai 980-8579, Japan
(3) PREST, Japan Science and Technology Agency (JST), Japan
`shh.tanaka at gmail.com, okazaki at ecei.tohoku.ac.jp,`
`ishizuka at i.u-tokyo.ac.jp`

ABSTRACT

This paper presents a novel approach for inducing causal rules by using deverbal nouns as a clue for finding causal relations. We collect verbs and their deverbal forms from FrameNet, and extract pairs of sentences in which event verbs and their corresponding deverbal forms co-occur in documents. The most challenging part of this work is to generalize an instance of causal relation into a rule. This paper proposes a method to generalize and constrain causal rules so that the obtained rules have the high chance of applicability and reusability. In order to find a suitable constraint for a causal rule, we utilize relation instances extracted by an open-information extractor, and build a classifier to choose the most suitable constraint. We demonstrate that deverbal nouns provide a good clue for causal relations and that the proposed method can induce causal rules from deverbal noun constructions.

KEYWORDS: causal relation, rules, pattern generalization, semantic inference, knowledge acquisition.

Proceedings of COLING 2012: Posters, pages 1209–1218,
COLING 2012, Mumbai, December 2012.

1209

1 Introduction

Performing semantic inference is important for natural language applications such as Question Answering (QA), Information Extraction, and Discourse Analysis. One of the missing links for semantic inference is the availability of commonsense knowledge in computers. In this paper, we focus on acquiring knowledge about causal relations between events.

Previous work on causal rule acquisition targeted at simple rules each of whose head is represented by a single literal or n-ary predicate: for example, Girju (2003) collected causal rules between nouns (e.g., *hunger* ⇒ *headache*); and Pantel et al. (2007) acquired causal rules between verbs (e.g., Y *announced the arrest of* X ⇒ X *is charged by* Y). However, humans perform more complicated inferences to predict outcomes of an event.

Let us consider the following example: *Google acquires Android Inc. The acquisition will enhance Google's competition in mobile phones.* The first sentence mentions an acquisition event with the verb *acquire*. Starting with its deverbal noun *acquisition*, the second sentence describes the possible outcome of the acquisition event. Referring to events explained in the preceding sentences, deverbal nouns often provide good clues for identifying cause-effect relations. However, acquiring the following causal rule from the above example is of no use:

$$\texttt{acquire}(\texttt{X}, \textit{Android}) \Rightarrow \texttt{compete-in}(\texttt{X}, \textit{mobile phones}) \qquad (1)$$

Even though we generalize the causal relation by replacing the company name *Google* with a variable X, it is unlikely to reuse the causal knowledge that *if a company acquires Android, the company will compete in mobile phones.* Having said that, the following rule may be too generic:

$$\texttt{acquire}(\texttt{X}, \texttt{Y}) \Rightarrow \texttt{compete-in}(\texttt{X}, \texttt{Z}) \qquad (2)$$

This rule only expresses that *if a company acquires something, the company will compete in some area.* This causality may be supported by a lot of activities in the real world, but it does not provide a good hint for predicting the value of Z. In contrast, inducing the following causal rule would be more preferable in terms of the reusability and predictability:

$$\texttt{acquire}(\texttt{X}, \texttt{Y}) \wedge \texttt{specialized-in}(\texttt{Y}, \texttt{Z}) \Rightarrow \texttt{compete-in}(\texttt{X}, \texttt{Z}) \qquad (3)$$

Here, we complemented a predicate `specialized-in(Y,Z)` as a constraint, i.e., as a premise in the head (left-hand side) of the rule, even though this is not explicit in the original text. Humans accept the causal relation mentioned in the above text because we have a prior knowledge about Rule 3 and the truth of the predicate `specialized-in`(*Android, mobile phone*).

This paper presents a novel approach for inducing causal rules like Rule 3 from the sentences with deverbal nouns (as in the above example). The contributions of this paper are twofold:

1. We focus on verbs and their deverbal nouns that co-refer to the same events. The use of deverbal nouns was not explored in the previous work on causal knowledge acquisition. We investigate the advantage of this approach empirically.

2. We present a method for generalizing and constraining causal relations by making use of relation instances acquired automatically from a large corpus. Previous work replaced the same mention (string) in a pattern with a variable to induce an inference rule (template). In contrast, this work unveils hidden predicates and variables that are not stated explicitly in text, but are crucial for explaining causal relations. This part is very challenging because we need to combine pieces of predicates obtained from different texts.

2 Proposed Method

The proposed system uses FrameNet[1] (Fillmore, 1976; Baker et al., 1998) for obtaining a list of verbs and their deverbal nouns (e.g., *acquisition* and *purchase* as deverbal nouns of verbs *buy* and *acquire*). Finding documents containing both verbs and their deverbal nouns in the corpus, the system extracts text fragments in which causal relations are expressed by pairs of sentences, "*A verb B ...*" and "*Deverbal-noun ...*". Here, *A* and *B* present named entities[2], and "*Deverbal-noun ...*" denotes a sentence starting with the deverbal noun of *verb*. We call the former sentence "*A verb B ...*" a *head sentence* and the latter sentence "*Deverbal-noun ...*" a *body sentence*. We apply several NLP analyses including part-of-speech tagging, dependency parsing, named entity recognition, and coreference resolution for obtaining causal relation instances as dependency trees with variables (Section 2.1). Independently of this process, the system extracts relation instances from the corpus by using ReVerb[3] (Fader et al., 2011) (Section 2.2). Searching for candidates of relation instances that can be inserted to a causal rule as a constraint, the system chooses the best relation instance as a constraint (Section 2.3).

2.1 Extracting causal relations using deverbal nouns

Using the list of verbs and their deverbal nouns extracted from FrameNet, the proposed system finds documents that contain both verbs and their corresponding deverbal nouns. For example, we extract a document when it contains a verb *buy* and its deverbal nouns (*purchase*, *acquisition*, *procurement*, etc). We employ Stanford Core NLP[4] for fundamental NLP analyses including sentence splitting, tokenization, part-of-speech tagging, named entity recognition, dependency parsing, and coreference resolution. The system extracts a causal relation from every sentence pair containing a verb and its deverbal noun. In this section, we use the following example to explain the process of generalizing causal relations into causal rules.

UnitedHealth buys Pacificare. The acquisition also gives UnitedHealth new operations in Nevada.

Firstly, the proposed method extracts a predicate and its arguments as the event referred to by the head (first) sentence. We define that: a predicate is a verb; arguments are a subject and object of the verb in the dependency tree; and arguments must mention named entities. For instance, the system extracts buy(*UnitedHealth, Pacificare*) from the example. In this study, we assume that a named entity presents either a person, location, or organization recognized by Stanford Core NLP. We replace mentions of each argument with a variable such as A and B, and generalize the predicate into a pattern buy(A, B). We call the pattern and variables extracted from the head sentence *head pattern* and *entity variables*, respectively.

2.1.1 Simplifying a pattern from the sentence with a deverbal noun

Sentences with deverbal nouns are often so specific that we cannot reuse corresponding patterns as bodies of causal rules. For example, the pattern from the example, *the acquisition also gives* A *new operations in Nevada*, is too specific. Therefore, we simplify a pattern from the body sentence (*body pattern* hereafter) by applying the following procedure.

---

[1] https://framenet.icsi.berkeley.edu/fndrupal/
[2] We use newswire text as a corpus, where the current NLP tools (e.g., POS tagger and NER) were designed to perform well. Because articles in the newswire domain mostly describe events occurring with named entities (e.g., companies, organizations, people), we do not think the requirement of variables *A* and *B* was strong.
[3] http://reverb.cs.washington.edu/
[4] http://nlp.stanford.edu/software/corenlp.shtml

1. Remove nodes whose depths (distances from the root node) are more than three in the dependency tree. We assume that these words are unnecessary for body patterns.
2. Replace every noun node with a variable (e.g., X) whose depth is no more than three. These variables will be used for generalizing the causal relation.
3. Keep nodes whose depths are one or two.
4. For each variable X, resolve it to a variable in the head pattern, A or B, if the variable X satisfies the following rules:
   - The variable X is a part of the named entity in the head pattern. For example, when X is *Google* and A is *Google Inc*, we replace X with A.
   - The variable is the initials of the named entity. For example, when X is *HP* and A is *Hewlett-Packard*, we replace X with A.
5. If a node is recognized as a numerical expression (tagged as either "Time", "Money", "Percent", "Date" and "Number" by Stanford Core NLP), replace the node with a special variable representing its semantic class. For example, we replace *$1,500,000* with MONEY.
6. Remove nodes that have certain syntactic relations (adverbial modifiers, appositional modifiers, adjectival modifiers and complementizers) with their parents. Nodes under these relations unnecessary for body patterns, describing specific/additional information.
7. Remove a body pattern if it ends with words other than nouns. This rule removes body patterns in passive voice, for example, *the acquisition of A was announced.*

We call variables that were unresolved to entity variables after this procedure *unconstrained variables*. The procedure yields a body pattern *the acquisition gives* A *operations in* X. Combining the head and body patterns, we obtain the following causal relation,

$$\mathtt{buy}(A, B) \Rightarrow \textit{the acquisition gives A operations in X} \tag{4}$$

Meanwhile, it would be better for the usability of causal rules if we could paraphrase the body pattern *the acquisition gives* A *operations in* X into a predicate representation `operate-in(A, X)` or a simpler textual pattern like A *will operate in* X. As the first attempt for using deverbal nouns, we leave the task as a future work; in this study we focus on generalizing causal rules.

2.2 Finding possible constraints for causal rules

So far, we obtained generalized causal rules with variables. However, these rules are too generic to represent a causal relation; for example, it is inadequate to fill any location name (e.g., *Tokyo* and *London*) in the unconstrained variable X of the rule, buy(A, B) ⇒ *The acquisition gives* A *operations in* X. Therefore, we would like to find constraints for unconstrained variables so that a rule is likely to instantiate a causal relation. The basic idea for inducing constraints is to associate unconstrained variables (e.g., X) with entity variables (e.g., A and/or B). In other words, if we found a relation associating either of the pairs (X, A) or (X, B), we could use the relation as the constraint for the variable X.

For example, if we were aware of a relation instance `headquartered-in`(*Pacificare*, *Nevada*), we could transform Rule 4 into:

$$\mathtt{buy}(A, B) \wedge \mathtt{headquartered\text{-}in}(B, X) \Rightarrow \textit{The acquisition gives A operations in X} \tag{5}$$

With the predicate `headquartered-in(B, X)` as a constraint (premise), Rule 5 has a higher chance of realizing the causality than Rule 4. In this way, we solve the problem of inducing
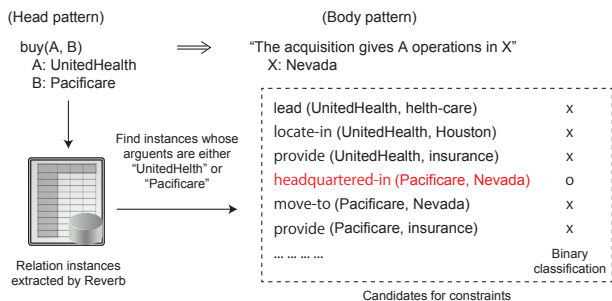
Figure 1: Choosing a constraint as a binary classification problem.

constraints by finding relation instances that associate unconstrained variables with entity variables. An easy and secure approach for the problem would be to extract a relation instance from the target document from which body and head patterns are extracted. However, there is no guarantee that a target document has a sentence associating unconstrained variables with entity variables. For example, the target document may not include a sentence like *Pacificare maintains headquarters in Nevada.* Therefore, we extract relation instances by applying ReVerb, an Open Information Extractor, to a large text corpus. We use the collection of relation instances as a knowledge base to explain unconstrained variables. In this study, we use the ClueWeb09 corpus[5] as a large text corpus.

2.3 Choosing a relation instance for inducing a constraint

A naive approach for associating unconstrained variables (e.g., X) with entity variables (e.g., A and B) would be to find relation instances that match to the query *(A, X) or *(B, X), where * denotes a wildcard. However, this query is inflexible in that it assumes an exact match for the value of X. In addition, if the query finds multiple relation instances (see "Candidates for constraints" in Figure 1), we need a mechanism to rank the relation instances. Therefore, we formalize the problem of choosing a relation instance for a constraint as a binary classification problem: choose a relation instance that yields the highest confidence score in the candidate relation instances. In order to allow flexible matching on the value of unconstrained variables (e.g., X), we relax the query such that it retrieves relation instances containing either the value of A or B,

$$*(A, *) \text{ or } *(B, *) \text{ or } *(*, A) \text{ or } *(*, A) \quad (6)$$

Figure 1 illustrates this process. Because we relaxed the query, the retrieved relation instances may not refer to the value of X (e.g., *Nevada*). At the same time, the retrieved instances may include multiple relations (e.g., headquartered-in and move-to) that refer to the value of X. Thus, we design several features to choose a relation instance that is suitable for the causal rule as a constraint. In the descriptions of the features, we denote: X as the value of the unconstrained variable; X' as the value of the argument other than A and B in the retrieved relation instance; R as the text representation of the retrieved relation (e.g., *has a headquarter in* for headquarter-in relation).

---

[5] http://lemurproject.org/clueweb09.php/

1. Word overlap between X and X'. Representing an argument X as a vector w_X whose elements present occurrences of words in X, and $w_{X'}$ similarly, this feature computes a cosine similarity between the vectors w_X and $w_{X'}$.
2. Word overlap between R and the target document. Representing the relation R as a vector w_R whose elements present occurrences of words in R and the vector of target document w_D similarly, this feature computes a cosine similarity between the vectors w_R and w_D.
3. Overlap of documents supporting the relation R and the body pattern. We define d_R as the set of documents containing the relation R, in other words, documents from which ReVerb yields the relation R. We also define d_b containing all the words in the body pattern. This feature measures the overlap of the two sets d_R and d_b by using the Jaccard coefficient.
4. Overlap of documents supporting the relation R and X'. This feature measures the overlap of two sets of documents that containing the relation R and the value of X', respectively, by using the Jaccard coefficient.
5. Overlap of documents supporting the relation R and X. This feature measures the overlap of two sets of documents that containing the relation R and the value of X, respectively, using the Jaccard coefficient.
6. Overlap of documents supporting X and X'. This feature measures the overlap of two sets of documents containing the values of X and X' by using the Jaccard coefficient.
7. Context similarity between X and X'. We represent an argument X as a vector c_X whose elements present frequencies of words that co-occur with X within sentences. We also define $c_{X'}$ similarly. This feature computes a cosine similarity between the vectors c_X and $c_{X'}$ as a distributional similarity between X and X'.

In order to build a classifier for ranking constraints, we manually prepared a training set. In this study, we used a verb *acquire* (belonging to the frame "Getting") as the target verb. Using its deverbal nouns, the system extracted ten causal relations from the corpus. The system found 100 relation instances for each causal relation. Then we asked a human annotator to label each relation instance as: positive if a relation is suitable as a constraint for the causal relation; and negative otherwise. In this way, we obtained 1,000 training instances for the classifier. Although the training set might look small in numbers, we think this is sufficient because the designed features do not include lexicalized features. We use liblinear[6] as an implementation of linear kernel SVMs for modeling the classifier. The system computes the dot product of the feature vector and the weight vector to compute the score of a relation instance.

3 Experiments

We conducted two experiments to evaluate the proposed method. The first experiment investigates the ability of deverbal nouns as clues for causal relations (without any generalization). The second experiment evaluates the correctness of causal rules. In these experiments, we used the portion of L.A. Times (about 300,000 articles) in English Gigaword Corpus Third Edition[7].

3.1 Deverbal nouns as clues for causal relations

Because no resource exists for evaluating causal relations between verbs, we built an evaluation set manually, selecting 10 verbs (frames) for this evaluation[8]. For each verb in the target verb

---

[6] http://www.csie.ntu.edu.tw/~cjlin/liblinear/
[7] http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2007T07
[8] We chose verbs that are frequent in the ClueWeb09 corpus, but excluded some verbs that do not have deverbal nouns (e.g., *be*), and that do not introduce causal relations (e.g., *like*).

Method	Precision	Recall	F_1
Baseline method (causal)	0.0445	0.1574	0.0694
Baseline method (causal + other)	0.1440	0.2165	0.1730
Proposed method (causal)	0.5357	0.2778	0.3659
Proposed method (causal + other)	0.6607	0.1457	0.2387

Table 1: Precision and recall on locating causal relations

set, we randomly sampled five documents in which both the verb and one of its deverbal nouns appear. This process obtained 50 documents (five for each verb) as an evaluation set. Then we asked a human annotator to mark pairs of verbs and other expressions (including verbs and nouns) that have causal relations in the documents. In addition, we also asked the annotator to mark pairs if they do not have causal relations but other relations (e.g., similar and associated). In this way, we obtained 108 pairs in causal relations and 146 pairs in other relations.

We prepared a baseline method that assumes a pair of relations sharing the same argument to have a causal relation. The baseline method uses ReVerb to extract relation instances in each document in the test set. For example, when ReVerb finds relation instances visit(*the prime minister*, *US*) and meet(*the prime minister*, *the president*) from the same document, the baseline method yields *visit* ⇒ *meet*.

Table 1 reports precision and recall of the proposed and baseline methods for locating causal relations. In the table, "causal" uses causal relations identified by the annotator as the gold standard, and "causal + other" uses causal and other relations as the gold standard. Our method performed much better than the baseline method in terms of precision and F_1 score. The baseline method did not work well for finding causal relations (0.0694 F_1 score), but found causal and other relations to some extent (0.1730 F_1 score). In contrast, the proposed method gained 0.3659 F_1 score in finding causal relations, but the F_1 score decreased to 0.2387 when we include other relations for the evaluation. This fact suggests that deverbal nouns can locate causal relations selectively, separating from other types of associations.

3.2 Extraction of causal rules

Using the same set of the 50 documents in Section 3.1, we evaluated the correctness of the rules extracted by a system. We asked the human subject to mark each rule extracted by a system into: *causal* if the rule presents a causal relation; *related* if the head and body of the rule does not present a causal relation but have some relation; and *incorrect* if the rule is incorrect.

We compare four methods including a baseline and the proposed method and their variants. "ReVerb+ReVerb" applies ReVerb to a target document, and finds causal rules such as verb1(A, B) ⇒ verb2(B, X), using the identical argument B as the bridge to connect *verb1* and *verb2*. In order to insert a constraint for the causal rule, it searches for relation instances verb3(A, X), verb3(B, X), verb3(X, A), or verb3(X, B) in the database constructed in Section 2.2. This method selects the relation instance with the highest score (computed by ReVerb) as a constraint. "ReVerb+SVM" extracts causal rules similarly to "ReVerb+ReVerb", but selects a relation for a constraint for a causal rule by using the SVM classifier described in Section 2.3. "Proposed method+ReVerb" extracts causality rules by using the proposed method. When this method selects a constraint for a causal rule, it selects the relation instance with the highest score computed by ReVerb. "Proposed method+SVM" is identical to the proposed method; this

Method	Causal	Causal + Related
ReVerb+ReVerb	0.1667	0.4902
ReVerb+SVM	0.1176	0.4804
Proposed method+ReVerb	0.2946	0.5982
Proposed method+SVM	0.3750	0.6339

Table 2: Accuracy of causal rules extracted by the systems

setting uses the SVM classifier to select a relation instance as a constraint for a causal rule.

Table 2 reports the average of accuracy values computed on the gold standard prepared by a human subject. The proposed method using SVM achieved the highest performance (0.3750 for causality). The SVM-based constraint selector boosted the correctness of causal rules for the proposed method (0.2946 → 0.3750). The baseline method could yield rules representing some association (0.4902 for causal and other relations), but failed to produce causal rules (0.1667). The SVM-based constraint selector did not contribute to the baseline method. This is probably because we trained the constraint selector for the proposed method. We observed that the half of rules extracted by the proposed method were judged incorrect. Analyzing these false cases, we found that these errors appeared in the phase of selecting constraints.

4 Related Work

The previous work on automatic acquisition of causal knowledge can be categorized into three groups in terms of types of inference rules: noun-noun causality (Girju, 2003; Chang and Choi, 2006; Saeger et al., 2011), verb-verb causality (Lin and Pantel, 2001; Chklovski and Pantel, 2004; Torisawa, 2006; Pantel et al., 2007; Abe et al., 2008; Beamer and Girju, 2009; Do et al., 2011; Hashimoto et al., 2012), and inference rules of other types (e.g., entailment) (Pekar, 2006; Szpektor and Dagan, 2008; Aharon et al., 2010; Schoenmackers et al., 2010; Berant et al., 2010, 2011; Gordon and Schubert, 2011; Berant et al., 2012). However, causal rules extracted by the previous work were limited to those without variables (e.g., *lean* ⇒ *kiss*) or those with the same set of variables (e.g., X *leaves for* Y ⇒ X *gets to* Y) in the head and body of a rule. In contrast, our work is the first approach that leverages deverbal nouns that directly express causal relations, and generalizes causal relations into causal rules with multiple variables.

5 Conclusion

In this paper, we presented a novel approach for inducing causal rules from the sentences with deverbal nouns. We conducted two experiments, and demonstrated that deverbal nouns present a good clue for causal relations and that the proposed method can generalize causal relations into causal rules. In this work, we did not address the problem of paraphrasing the body pattern (e.g., *the acquisition gives* A *operations in* X) into a predicate representation (e.g., operate-in(A,X)) or a simpler textual pattern (e.g., A *will operate in* B). This task would be an immediate future work of this study. In addition, we would like to extend the approach of rule generalization to causal relations identified by other clues (e.g., distributional similarity of verbs) and to other types of semantic relations, for example, entailment relations.

Acknowledgments

References

Abe, S., Inui, K., and Matsumoto, Y. (2008). Acquiring event relation knowledge by learning cooccurrence patterns and fertilizing cooccurrence samples with verbal nouns. In *Proceedings of the Third International Joint Conference on Natural Language Processing*, IJCNLP 2008, pages 497–504.

Aharon, R. B., Szpektor, I., and Dagan, I. (2010). Generating entailment rules from FrameNet. In *Proceedings of the ACL 2010 Conference Short Papers*, ACL 2010 (short), pages 241–246.

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The Berkeley FrameNet project. In *Proceedings of the 17th international conference on Computational linguistics (Coling 1998) and the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, ACL 1998, pages 86–90.

Beamer, B. and Girju, R. (2009). Using a bigram event model to predict causal potential. In *Proceedings of the 10th International Conference on Intelligent Text Processing and Computational Linguistics*, CICLing 2009, pages 430–441.

Berant, J., Dagan, I., Adler, M., and Goldberger, J. (2012). Efficient tree-based approximation for entailment graph learning. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, ACL 2012, pages 117–125.

Berant, J., Dagan, I., and Goldberger, J. (2010). Global learning of focused entailment graphs. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL 2010, pages 1220–1229.

Berant, J., Dagan, I., and Goldberger, J. (2011). Global learning of typed entailment rules. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, ACL HLT 2011, pages 610–619.

Chang, D.-S. and Choi, K.-S. (2006). Incremental cue phrase learning and bootstrapping method for causality extraction using cue phrase and word pair probabilities. *Information Processing and Management*, 42(3):662–678.

Chklovski, T. and Pantel, P. (2004). VerbOcean: Mining the web for fine-grained semantic verb relations. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, EMNLP 2004, pages 33–40.

Do, Q. X., Chan, Y. S., and Roth, D. (2011). Minimally supervised event causality identification. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, EMNLP 2011, pages 294–303.

Fader, A., Soderland, S., and Etzioni, O. (2011). Identifying relations for open information extraction. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, EMNLP 2011, pages 1535–1545.

Fillmore, C. J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, 280(1):20–32.

Girju, R. (2003). Automatic detection of causal relations for question answering. In *Proceedings of the ACL 2003 workshop on multilingual summarization and question answering*, pages 76–83.

Gordon, J. and Schubert, L. K. (2011). Discovering commonsense entailment rules implicit in sentences. In *Proceedings of the TextInfer 2011 Workshop on Textual Entailment*, TIWTE 2011, pages 59–63.

Hashimoto, C., Torisawa, K., De Saeger, S., Oh, J.-H., and Kazama, J. (2012). Excitatory or inhibitory: A new semantic orientation extracts contradiction and causality from the web. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, EMNLP-CoNLL 2012, pages 619–630.

Lin, D. and Pantel, P. (2001). DIRT – Discovery of inference rules from text. In *Proceedings of ACM Conference on Knowledge Discovery and Data Mining*, KDD 2001, pages 323–328.

Pantel, P., Bhagat, R., Chklovski, T., and Hovy, E. (2007). ISP: Learning inferential selectional preferences. In *In Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL HLT 2007, pages 564–571.

Pekar, V. (2006). Acquisition of verb entailment from text. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, HLT-NAACL '06, pages 49–56.

Saeger, S. D., Torisawa, K., Tsuchida, M., Kazama, J., Hashimoto, C., Yamada, I., Oh, J.-H., Varga, I., and Yan, Y. (2011). Relation acquisition using word classes and partial patterns. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, EMNLP 2011, pages 825–835.

Schoenmackers, S., Etzioni, O., Weld, D. S., and Davis, J. (2010). Learning first-order Horn clauses from web text. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP 2010, pages 1088–1098.

Szpektor, I. and Dagan, I. (2008). Learning entailment rules for unary templates. In *Proceedings of the 22nd International Conference on Computational Linguistics*, Coling 2008, pages 849–856.

Torisawa, K. (2006). Acquiring inference rules with temporal constraints by using Japanese coordinated sentences and noun-verb co-occurrences. In *Proceedings of the Human Language Technology Conference of North American chapter of the Association for Computational Linguistics*, HLT NAACL 2006, pages 57–64.