UIUC: A Knowledge-rich Approach to Identifying Semantic Relations between Nominals

Brandon Beamer,[1,4] Suma Bhat,[2,4] Brant Chee,[3,4] Andrew Fister,[1,4] Alla Rozovskaya,[1,4] Roxana Girju[1,4]
Department of Linguistics[1],
Department of Electrical and Computer Engineering[2],
Department of Library and Information Science[3],
Beckman Institute[4],
University of Illinois at Urbana-Champaign
{bbeamer, spbhat2, chee, afister2, rozovska, girju}@uiuc.edu

Abstract

This paper describes a supervised, knowledge-intensive approach to the automatic identification of semantic relations between nominals in English sentences. The system employs different sets of new and previously used lexical, syntactic, and semantic features extracted from various knowledge sources. At SemEval 2007 the system achieved an F-measure of 72.4% and an accuracy of 76.3%.

1 Introduction

The SemEval 2007 task on Semantic Relations between Nominals is to identify the underlying semantic relation between two nouns in the context of a sentence. The dataset provided consists of a definition file and 140 training and about 70 test sentences for each of the seven relations considered: *Cause-Effect*, *Instrument-Agency*, *Product-Producer*, *Origin-Entity*, *Theme-Tool*, *Part-Whole*, and *Content-Container*. The task is defined as a binary classification problem. Thus, given a pair of nouns and their sentential context, the classifier decides whether the nouns are linked by the target semantic relation. In each training and test example sentence, the nouns are identified and manually labeled with their corresponding WordNet 3.0 senses. Moreover, each example is accompanied by the heuristic pattern (query) the annotators used to extract the sentence from the web and the position of the arguments in the relation.

(1) 041 "He derives great joy and <e_1>happiness</e_1> from <e_2>cycling</e_2>." WordNet(e_1) = "happiness%1:12:00::", WordNet(e_2) = "cycling%1:04:00::", Cause-Effect(e_2,e_1) = "true", Query = "happiness from *"

Based on the information employed, systems can be classified in four types of classes: (A) systems that use neither the given WordNet synsets nor the queries, (B) systems that use only WordNet senses, (C) systems that use only the queries, and (D) systems that use both.

In this paper we present a type-B system that relies on various sets of new and previously used linguistic features employed in a supervised learning model.

2 Classification of Semantic Relations

Semantic relations between nominals can be encoded by different syntactic constructions. We extend here over previous work that has focused mainly on noun compounds and other noun phrases, and noun–verb–noun constructions.

We selected a list of 18 lexico-syntactic and semantic features split here into three sets: *feature set #1* (core features), *feature set #2* (context features), and the *feature set #3* (special features). Table 1 shows all three sets of features along with their definitions; a detailed description is presented next. For some features, we list previous works where they proved useful. While features F1 – F4 were selected from our previous experiments, all the other features are entirely the contribution of this research.

Feature set #1: Core features

This set contains six features that were employed in all seven relation classifiers. The features take into consideration only lexico-semantic information

Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007), pages 386–389,
Prague, June 2007.

No.	Feature	Definition
Feature Set #1: Core features		
F1	**Argument position** (Girju et al., 2005; Girju et al., 2006)	indicates the position of the arguments in the semantic relation (e.g., Part-Whole(e_1, e_2), where e_1 is the *part* and e_2 is the *whole*).
F2	**Semantic specialization** (Girju et al., 2005; Girju et al., 2006)	this is the prediction returned by the automatic WordNet IS-A semantic specialization procedure.
F3, F4	**Nominalization** (Girju et al., 2004)	indicates whether the nouns e_1 (F3) and e_2 (F4) are nominalizations or not. Specifically, we distinguish here between *agential nouns*, *other nominalizations*, and *neither*.
F5, F6	**Spatio-Temporal features**	indicate if e_1 (F5) or e_2 (F6) encode time or location.
Feature Set #2: Context features		
F7, F8	**Grammatical role**	describes the grammatical role of e_1 (F7) and e_2 (F8). There are three possible values: *subject*, *direct object*, or *neither*.
F9	**PP Attachment**	applies to NP PP constructions and indicates if the prepositional phrase containing e_2 attaches to the NP containing e_1.
F10, F11	**Semantic Role**	is concerned with the semantic role of the phrase containing either e_1 (F10) or e_2 (F11). In particular, we focused on three semantic roles: *Time*, *Location*, *Manner*. The feature is set to 1 if the target noun is part of a phrase of that type and to 0 otherwise.
F12, F13, F14	**Inter-noun context sequence**	is a set of three features. F12 captures the sequence of stemmed words between e_1 and e_2, while F13 lists the part of speech sequence in between the target nouns. F14 is a scoring weight (with possible values 1, 0.5, 0.25, and 0.125) which measures the similarity of an unseen sequence to the set of sequence patterns associated with a relation.
Feature Set #3: Special features		
F15, F16	**Psychological feature**	is used in the *Theme-Tool* classifier; indicates if e_1 (F15) or e_2 (F16) belong or not to a predefined set of psychological features.
F17	**Instrument semantic role**	is used for the *Instrument-Agency* relation and indicates whether the phrase containing e_1 is labeled as em Instrument or not.
F18	**Syntactic attachment**	is used for the *Instrument-Agent* relation and indicates whether the phrase containing the *Instrument* role attaches to a noun or a verb

Table 1: The three sets of features used for the automatic semantic relation classification.

about the two target nouns.

Argument position (F1) indicates the position of the semantic arguments in the relation. This information is very valuable, since some relations have a particular argument arrangement depending on the lexico-syntactic construction in which they occur. For example, most of the noun compounds encoding Stuff-Object / Part-Whole relations have e_1 as the part and e_2 as the whole (e.g., *silk dress*).

Semantic specialization (F2) is a binary feature representing the prediction of a semantic specialization learning model. The method consists of a set of iterative procedures of specialization of the training examples on the WordNet IS-A hierarchy. Thus, after all the initial noun–noun pairs are mapped through generalization to *entity* – *entity* pairs in WordNet, a set of necessary specialization iterations is applied until it finds a boundary that separates positive and negative examples. This boundary is tested on new examples for relation prediction.

The *nominalization* features (F3, F4) indicate if the target noun is a nominalization and, if yes, of what type. We distinguish here between *agential nouns*, *other nominalizations*, and *neither*. The features were identified based on WordNet and NomLex-Plus[1] and were introduced to filter some of negative examples, such as *car owner*/THEME.

Spatio–Temporal features (F5, F6) were also introduced to recognize some near miss examples, such as Temporal and Location relations. For instance, *activation by summer* (near-miss for *Cause-Effect*) and *mouse in the field* (near-miss for *Content-Container*). Similarly, for *Theme-Tool*, a word acting as a Theme should not indicate a period of time, as in $<e_1>$*the appointment*$</e_1>$ *was for more than one* $<e_2>$*year*$</e_2>$. For this we used the information provided by WordNet and special classes generated from the works of (Herskovits, 1987), (Linstromberg, 1997), and (Tyler and Evans, 2003).

[1] NomLex-Plus is a hand-coded database of 5,000 verb nominalizations, de-adjectival, and de-adverbial nouns. http://nlp.cs.nyu.edu/nomlex/index.html

Feature set #2: Context features
This set takes advantage of the sentence context to identify features at different linguistic levels.

The *grammatical role* features (F7, F8) determine if e_1 or e_2 is the *subject, direct object,* or *neither*. This feature helps filter out some instances with poor context, such as noun compounds and identify some near-miss examples. For example, a restriction imposed by the definition of *Theme-Tool* indicates that in constructions such as Y/Tool *is used for V-ing* X/Theme, neither X nor Y can be the subject of the sentence, and hence Theme-Tool(X, Y) would be false. This restriction is also captured by the nominalization feature in case X or Y is an agential noun.

PP attachment (F9) is defined for NP PP constructions, where the prepositional phrase containing the noun e_2 attaches or not to the NP (containing e_1). The rationale is to identify negative instances where the PP attaches to any other word before NP in the sentence. For example, *eat* $<e_1>$*pizza*$</e_1>$ *with* $<e_2>$*a fork*$</e_2>$, where *with a fork* attaches to the verb *to eat* (cf. (Charniak, 2000)).

Furthermore, we implemented and used two *semantic role* features which identify the semantic role of the phrase in a verb–argument structure, phrase containing either e_1 (F10) or e_2 (F11). In particular, we focus on three semantic roles: *Time, Location, Manner*. The feature is set to 1 if the target noun is part of a semantic role phrase and to 0 otherwise. The idea is to filter out near-miss examples, expecially for the *Instrument-Agency* relation. For this, we used ASSERT, a semantic role labeler developed at the University of Colorado at Boulder[2] which was queried through a web interface.

Inter-noun context sequence features (F12, F13) encode the sequence of lexical and part of speech information between the two target nouns. Feature F14 is a weight feature on the values of F12 and F13 and indicates how similar a new sequence is to the already observed inter-noun context associated with the relation. If there is a direct match, then the weight is set to 1. If the part-of-speech pattern of the new substring matches that of an already seen substring, then the weight is set to 0.5. Weights 0.25 and 0.125 are given to those sequences that overlap entirely or partially with patterns encoding other semantic relations in the same contingency set (e.g., semantic relations that share syntactic pattern sequences). The value of the feature is the summation of the weights thus obtained. The rationale is that the greater the weight, the more representative is the context sequence for that relation.

Feature set #3: Special features
This set includes features that help identify specific information about some semantic relations.

Psychological feature was defined for the *Theme-Tool* relation and indicates if the target noun (F15, F16) belongs to a list of special concepts. This feature was obtained from the restrictions listed in the definition of *Theme-Tool*. In the example *need for money*, the noun *need* is a psychological feature, and thus the instance cannot encode a *Theme-Tool* relation. A list of synsets from WordNet subhierarchy of *motivation* and *cognition* constituted the psychological factors. This was augmented with preconditions such as *foundation* and *requirement* since they would not be allowed as tools for the theme.

The *Instrument semantic role* is used for the *Instrument-Agency* relation as a boolean feature (F17) indicating whether the argument identified as Instrument in the relation (e.g., e_1 if Instrument-Agency(e_1, e_2)) belongs to an instrument phrase as identified by a semantic role tool, such as ASSERT.

The *syntactic attachment* feature (F18) is a feature that indicates whether the argument identified as Instrument in the relation attaches to a verb or to a noun in the syntactically parsed sentence.

3 Learning Model and Experimental Setting

For our experiments we chose libSVM, an open source SVM package[3]. Since some of our features are nominal, we followed the standard practice of representing a nominal feature with n discrete values as n binary features. We used the RBF kernel.

We built a binary classifier for each of the seven relations. Since the size of the task training data per relation is small, we expanded it with new examples from various sources. We added a new corpus of 3,000 sentences of news articles from the TREC-9 text collection (Girju, 2003) encoding *Cause-Effect* (1,320) and *Product-Producer* (721). Another col-

[2] http://oak.colorado.edu/assert/

[3] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

Relation	P	R	F	Acc	Total	Base-F	Base-Acc	Best features
Cause-Effect	69.5	100.0	82.0	77.5	80	67.8	51.2	F1, F2, F5, F6, F12–F14
Instrument-Agency	68.2	78.9	73.2	71.8	78	65.5	51.3	F7, F8, F10, F11, F15–F18
Product-Producer	84.5	79.0	81.7	76.3	93	80.0	66.7	F1–F4, F12–F14
Origin-Entity	86.4	52.8	65.5	75.3	81	61.5	55.6	F1, F2, F5, F6, F12–F14
Theme-Tool	85.7	41.4	55.8	73.2	71	58.0	59.2	F1–F6, F15, F16
Part-Whole	70.8	65.4	68.0	77.8	72	53.1	63.9	F1–F4
Content-Container	93.1	71.1	80.6	82.4	74	67.9	51.4	F1–F6, F12–F14
Average	79.7	69.8	72.4	76.3	78.4			

Table 2: Performance obtained per relation. Precision, Recall, F-measure, Accuracy, and Total (number of examples) are macro-averaged for system's performance on all 7 relations. Base-F shows the baseline F measure (all true), while Base-Acc shows the baseline accuracy score (majority).

lection of 3,129 sentences from Wall Street Journal (Moldovan et al., 2004; Girju et al., 2004) was considered for *Part-Whole* (1,003), *Origin-Entity* (167), *Product-Producer* (112), and *Theme-Tool* (91). We also extracted 552 *Product-Producer* instances from eXtended WordNet[4] (noun entries and their gloss definition). Moreover, for *Theme-Tool* and *Content-Container* we used special lists of constraints[5]. Besides the selectional restrictions imposed on the nouns by special features such as F15 and F16 (psychological feature), we created lists of containers from various thesauri[6] and identified selectional restrictions that differentiate between containers and locations relying on taxonomies of spatial entities discussed in detail in (Herskovits, 1987) and (Tyler and Evans, 2003).

Each instance in this text collection had the target nouns identified and annotated with WordNet senses. Since the annotations used different WordNet versions, senses were mapped to sense keys.

4 Experimental Results

Table 2 shows the performance of our system for each semantic relation. *Base-F* indicates the baseline F-measure (all true), while *Base-Acc* shows the baseline accuracy score (majority). The *Average* score of precision, recall, F-measure, and accuracy is macroaveraged over all seven relations. Overall, all features contributed to the performance, with a different contribution per relation (cf. Table 2).

5 Conclusions

This paper describes a method for the automatic identification of a set of seven semantic relations based on support vector machines (SVMs). The approach benefits from an extended dataset on which binary classifiers were trained for each relation. The feature sets fed into the SVMs produced very good results.

[4] http://xwn.hlt.utdallas.edu/

[5] The *Instrument-Agency* classifier was trained only on the task dataset.

[6] Thesauri such as TheFreeDictionary.com.

Acknowledgments

We would like to thank Brian Drexler for his valuable suggestions on the set of semantic relations.

References

E. Charniak. 2000. A Maximum-entropy-inspired Parser. In the Proceedings of *the 1st NAACL Conference*.

R. Girju, A. Giuglea, M. Olteanu, O. Fortu, O. Bolohan, and D. Moldovan. 2004. Support vector machines applied to the classification of semantic relations in nominalized noun phrases. In the Proceedings of *the HLT/NAACL Workshop on Computational Lexical Semantics.*

R. Girju, D. Moldovan, M. Tatu, and D. Antohe. 2005. On the semantics of noun compounds. *Computer Speech and Language*, 19(4):479–496.

R. Girju, A. Badulescu, and D. Moldovan. 2006. Automatic discovery of part-whole relations. *Computational Linguistics*, 32(1).

R. Girju. 2003. Automatic detection of causal relations for question answering. In the Proceedings of *the ACL Workshop on "Multilingual Summarization and Question Answering - Machine Learning and Beyond"*.

A. Herskovits. 1987. *Language and spatial cognition: An interdisciplinary study of the prepositions in English*. Cambridge University Press.

S. Linstromberg. 1997. *English Prepositions Explained*. John Benjamins Publishing Co., Amsterdam/Philaderphia.

D. Moldovan, A. Badulescu, M. Tatu, D. Antohe, and R. Girju. 2004. Models for the semantic classification of noun phrases. In the Proceedings of *the HLT/NAACL Workshop on Computational Lexical Semantics*.

A. Tyler and V. Evans. 2003. *The Semantics of English Prepositions: Spatial Sciences, Embodied Meaning, and Cognition*. Cambridge University Press.