

# ISI: Automatic Classification of Relations Between Nominals Using a Maximum Entropy Classifier

Stephen Tratz and Eduard Hovy

Information Sciences Institute  
University of Southern California  
Marina del Rey, CA 90292  
{stratz,hovy}@isi.edu

## Abstract

The automatic interpretation of semantic relations between nominals is an important subproblem within natural language understanding applications and is an area of increasing interest. In this paper, we present the system we used to participate in the SEMEVAL 2010 Task 8 Multi-Way Classification of Semantic Relations between Pairs of Nominals. Our system, based upon a Maximum Entropy classifier trained using a large number of boolean features, received the third highest score.

## 1 Introduction

Semantic interpretation of the relations between nominals in text is an area of growing interest within natural language processing (NLP). It has potential uses for a variety of tasks including machine translation (Baldwin and Tanaka, 2004) and question answering (Ahn et al., 2005). The related and more narrowly-focused problem of automatic interpretation of noun compounds is the focus of another SEMEVAL task (Butnariu et al., 2009).

In this paper, we discuss the overall setup of SEMEVAL 2010 Task 8 (Hendrickx et al., 2010), present the system we used to participate, and discuss our system's performance. Our system, which consists of a Maximum Entropy classifier trained using a large variety of boolean features, received the third highest official score of all the entries.

## 2 Related Work

The groundwork for SEMEVAL 2010 Task 8 was laid by an earlier SEMEVAL task (Girju et al., 2007). For SEMEVAL 2007 Task 4, participants provided *yes* or *no* answers as to whether a particular relation held for each test example. For SEMEVAL 2010, instead of providing a binary out-

put for a single class, participants were required to perform multi-way classification, that is, select the most appropriate relation from a set of 10 relations including the OTHER relation.

The selection of a semantic relation for a pair of nominals within a sentence is somewhat similar to the task of noun compound interpretation, which is a more restricted problem focused only upon the nouns within noun compounds. Some of the recent work on this problem includes that of Butnariu et al. (2009), Girju (2007), Girju et al. (2005), Kim and Baldwin (2005), Nakov (2008), Nastase et al. (2006), Turney (2006), and Ó Séaghdha and Copestake (2009).

## 3 Task Overview

The task is, given a pair of nominals within their sentence context, select the most appropriate semantic relation from the set of available relations and indicate the direction of the relation. Though the final score was based upon the output of the system trained using the whole training dataset, participants were also required to submit three additional label sets using the first 12.5%, 25%, and 50% of the training data.

### 3.1 Relation Scheme

The relations were taken from earlier work on noun compounds by Nastase and Szpakowicz (2003).

A total of 10 relations were used including CAUSE-EFFECT, COMPONENT-WHOLE, CONTENT-CONTAINER, ENTITY-ORIGIN, ENTITY-DESTINATION, INSTRUMENT-AGENCY, MEMBER-COLLECTION, MESSAGE-TOPIC, OTHER, and PRODUCT-PRODUCER. Since each relation except the OTHER relation must have its direction specified, there are a total of 19 possible labels.

## 3.2 Data

The training and testing datasets consist of 8000 and 2717 examples respectively. Each example consists of a single sentence with two of its nominals marked as being the nominals of interest. The training data also provides the correct relation for each example.

## 4 Method

### 4.1 Classifier

We use a Maximum Entropy (Berger et al., 1996) classifier trained using a large number of boolean features. Maximum Entropy classifiers have proven effective for a variety of NLP problems including word sense disambiguation (Tratz et al., 2007; Ye and Baldwin, 2007). We use the implementation provided in the MALLET machine learning toolkit (McCallum, 2002). We used the default Gaussian prior parameter value of 1.0.

### 4.2 Features Used

We generate features from individual words, including both the nominals and their context, and from combinations of the nominals.

To generate the features for individual words, we first use a set of word selection rules to select the words of interest and then run these words of interest through a variety of feature-generating functions. Some words may be selected by multiple word selection rules. For example, the word to the right of the first nominal will be identified by the *word 1 to the right of the 1st nominal* rule, the *words that are 3 or less to the right of the 1st nominal* rule, and the *all words between the nominals* rule. In these cases, the actual feature is the combination of an identifier for the word selection rule and the output from the feature-generating function. The 19 word-selection rules are listed below:

#### Word-Selection Rules

- The {1st, 2nd} nominal (2 rules)
- Word {1, 2, 3} to the {left, right} of the {1st, 2nd} nominal (12 rules)
- Words that are 3 or less to the {left, right} of the {1st, 2nd} nominal (4 rules)
- All words between the two nominals (1 rule)

The features generated from the individual words come from a variety of sources including word orthography, simple gazetteers, pattern

matching, WordNet (Fellbaum, 1998), and Roget's Thesaurus.

#### Orthographic Features

- Capitalization indicator
- The {first, last} {two, three} letters of each word
- Indicator if the first letter of the word is a/A.
- Indicator for the overall form of the word (e.g. jump -> a, Mr. -> Aa., SemEval2 -> AaAa0)
- Indicators for the suffix types (e.g., de-adjectival, de-nominal [non]agentive, de-verbal [non]agentive)
- Indicators for a wide variety of affixes including those related to degree, number, order, etc. (e.g., ultra-, poly-, post-)
- Indicators for whether or not a preposition occurs within either term (e.g., 'down' in 'breakdown')

#### Gazetteer and Pattern Features

- Indicators if the word is one of a number of closed classes (e.g. articles, prepositions)
- Indicator if the word is listed in the U.S. Census 2000's most common surnames list
- Indicator if the word is listed in the U.S. Census 2000's most common first names list
- Indicator if the word is a name or location based upon some simple regular expressions

#### WordNet-based Features

- Lemmatized version of the word
- Synonyms for all NN and VB entries for the word
- Hypernyms for all NN and VB entries for the word
- All terms in the definitions ('gloss') for the word
- Lexicographer file names for the word
- Lists of all link types (e.g., meronym links) associated with the word
- Part-of-speech indicators for the existence of NN/VB/JJ/RB entries for the word
- All sentence frames for the word
- All part, member, substance-of holonyms for the word

#### Roget's Thesaurus-based Features

- Roget's divisions for all noun (and verb) entries for the word

Some additional features were extracted using combinations of the nominals. These include features generated using The Web 1T corpus (Brants and Franz, 2006), and the output of a noun compound interpretation system.

### Web 1T N-gram Features

To provide information related to term usage to the classifier, we extracted trigram and 4-gram features from the Web 1T Corpus (Brants and Franz, 2006). Only n-grams containing lowercase words were used. The nominals were converted to lowercase if needed. Only n-grams containing both terms (including plural forms) were extracted. We included the n-gram, with the nominals replaced with N1 and N2 respectively, as individual boolean features. We also included versions of the n-gram features with the words replaced with wild cards. For example, if the nominals were ‘food’ and ‘basket’ and the extracted n-gram was ‘put\_N1\_in\_the\_N2’, we also included ‘\*\_N1\_in\_the\_N2’, ‘\*\_N1\*\_the\_N2’, etc. as features.

### Noun Compound System Features

We also ran the nominals through an in-house noun compound interpretation system and took its output as features. We will not be discussing the noun compound interpretation system in detail in this paper. It uses a similar approach to that described in this paper including a Maximum Entropy classifier trained with similar features that outputs a ranked list of a fixed set of semantic relations. The relations ranked within the top 5 and bottom 5 were included as features. For example, if “Topic of Communication” was the third highest relation, both “top:3:Topic of Communication” and “top\*:Topic of Communication” would be included as features.

### 4.3 Feature Filtering

The aforementioned feature generation process creates a very large number of features. To determine the final feature set, we first ranked the features according to the Chi-Squared metric. Then, by holding out one tenth of the training data and trying different thresholds, we concluded that 100,000 features was roughly optimal. For the cases where we used 12.5%, 25%, and 50%, we tested on the remaining training data and came up with different cutoffs: 25,000, 40,000, and 60,000, respectively.

## 5 Results

Each participating site was allowed to submit multiple runs based upon different systems or configurations thereof. The results for the best performing submissions from each team are presented in Table 1. The official metric for the task was F1 macroaveraged across the different relations. We are pleased to see that our system received the third highest score.

Our results by the different relation types are shown in Table 2. We note that the performance on the OTHER relation is relatively low.

Top Results				
System	Macroaveraged F1			
	12.5%	25%	50%	100%
UTD	73.08	77.02	79.93	82.19
FBK_IRST	63.61	70.20	73.40	77.62
<b>ISI</b>	<b>66.68</b>	<b>71.01</b>	<b>75.51</b>	<b>77.57</b>
ECNU	49.32	50.70	72.63	75.43
TUD	58.35	62.45	66.86	69.23
ISTI	50.49	55.80	61.14	68.42
FBK_NK	55.71	64.06	67.80	68.02
SEKA	51.81	56.34	61.10	66.33
JU	41.62	44.98	47.81	52.16
UNITN	16.57	18.56	22.45	26.67

Table 1: Final results (macroaveraged F1) for the highest ranking (based upon result for training with the complete training set) submissions for each site. 12.5%, 25%, 50%, and 100% indicate the amount of training data used.

Results by Relation			
Relation	P	R	F1
Cause-Effect	87.77	87.50	87.63
Component-Whole	73.21	75.32	74.25
Content-Container	82.74	84.90	83.80
Entity-Destination	81.51	81.51	81.51
Entity-Origin	81.86	75.19	78.38
Instrument-Agency	64.34	58.97	61.54
Member-Collection	84.62	84.98	84.80
Message-Topic	75.91	79.69	77.76
Product-Producer	70.83	66.23	68.46
Other	43.28	45.37	44.30

Table 2: Precision, recall, and F1 results for our system by semantic relation.

## 6 Conclusion

We explain the system we used to participate in the SEMEVAL 2010 Task 8: Multi-Way Classification of Semantic Relations Between Pairs of Nominals and present its results. The overall approach is straight forward, consisting of a single Maximum Entropy classifier using a large number of boolean features, and proves effective, with our system receiving the third highest score of all the submissions.

## 7 Future Work

In the future, we are interested in utilizing parsing and part-of-speech tagging to enrich the feature set. We also want to investigate the relatively low performance for the OTHER category and see if we could develop a method to improve this.

## Acknowledgements

Stephen Tratz is supported by a National Defense Science and Engineering Graduate Fellowship. We would like to thank the organizers of this task for their hard work in putting this task together.

## References

- Ahn, K., J. Bos, J. R. Curran, D. Kor, M. Nissim, and B. Webber. 2005. Question Answering with QED at TREC-2005. In *Proc. of TREC-2005*.
- Baldwin, T. & T. Tanaka 2004. Translation by machine of compound nominals: Getting it right. In *Proc. of the ACL 2004 Workshop on Multiword Expressions: Integrating Processing*.
- Berger, A., S. A. Della Pietra, and V. J. Della Pietra. 1996. A Maximum Entropy Approach to Natural Language Processing. *Computational Linguistics*, 22.
- Brants, T. and A. Franz. 2006. Web 1T 5-gram Corpus Version 1.1. Linguistic Data Consortium.
- Butnariu, C. and T. Veale. 2008. A concept-centered approach to noun-compound interpretation. In *Proc. of 22nd International Conference on Computational Linguistics (COLING 2008)*.
- Butnariu, C., S.N. Kim, P. Nakov, D. Ó Séaghdha, S. Szpakowicz, and T. Veale. 2009. SemEval Task 9: The Interpretation of Noun Compounds Using Paraphrasing Verbs and Prepositions. In *Proc. of the NAACL HLT Workshop on Semantic Evaluations: Recent Achievements and Future Directions*.
- Fellbaum, C., editor. 1998. WordNet: An Electronic Lexical Database. MIT Press, Cambridge, MA.
- Girju, R., D. Moldovan, M. Tatu and D. Antohe. 2005. On the semantics of noun compounds. *Computer Speech and Language*, 19.
- Girju, R., P. Nakov, V. Nastase, S. Szpakowicz, P. Turney, and D. Yuret. 2007. SemEval-2007 Task 04: Classification of Semantic Relations between Nominals In *Proc. of the 4th Semantic Evaluation Workshop (SemEval-2007)*.
- Hendrickx, I., S. N. Kim, Z. Kozareva, P. Nakov, D. Ó Séaghdha, Sebastian Padó, M. Pennacchiotti, L. Romano, and S. Szpakowicz. 2010. Improving the interpretation of noun phrases with cross-linguistic information. In *Proc. of the 5th SIGLEX Workshop on Semantic Evaluation*.
- Girju, R. 2007. Improving the interpretation of noun phrases with cross-linguistic information. In *Proc. of the 45th Annual Meeting of the Association of Computational Linguistics (ACL 2007)*.
- Kim, S.N. and T. Baldwin. 2005. Automatic Interpretation of Compound Nouns using WordNet::Similarity. In *Proc. of 2nd International Joint Conf. on Natural Language Processing*.
- McCallum, A. K. MALLET: A Machine Learning for Language Toolkit. <http://mallet.cs.umass.edu>. 2002.
- Nakov, P. 2008. Noun Compound Interpretation Using Paraphrasing Verbs: Feasibility Study. In *Proc. the 13th International Conference on Artificial Intelligence: Methodology, Systems, Applications (AIMSA'08)*.
- Nastase V. and S. Szpakowicz. 2003. Exploring noun-modifier semantic relations. In *Proc. the 5th International Workshop on Computational Semantics*.
- Nastase, V., J. S. Shirabad, M. Sokolova, and S. Szpakowicz 2006. Learning noun-modifier semantic relations with corpus-based and Wordnet-based features. In *Proc. of the 21st National Conference on Artificial Intelligence (AAAI-06)*.
- Ó Séaghdha, D. and A. Copestake. 2009. Using lexical and relational similarity to classify semantic relations. In *Proc. of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2009)*.
- Tratz, S., A. Sanfilippo, M. Gregory, A. Chappell, C. Posse, and P. Whitney. 2007. PNNL: A Supervised Maximum Entropy Approach to Word Sense Disambiguation In *Proc. of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*.
- Turney, P. D. 2006. Similarity of semantic relations. *Computation Linguistics*, 32(3):379-416
- Ye, P. and T. Baldwin. 2007. MELB-YB: Preposition Sense Disambiguation Using Rich Semantic Features. In *Proc. of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*.