

# ECNU: Effective Semantic Relations Classification without Complicated Features or Multiple External Corpora

Yuan Chen<sup>†</sup>, Man Lan<sup>†,§</sup>, Jian Su<sup>§</sup>, Zhi Min Zhou<sup>†</sup>, Yu Xu<sup>†</sup>

<sup>†</sup>East China Normal University, Shanghai, PRC.

<sup>§</sup>Institute for Infocomm Research, Singapore.

lanman.sg@gmail.com

## Abstract

This paper describes our approach to the automatic identification of semantic relations between nominals in English sentences. The basic idea of our strategy is to develop machine-learning classifiers which: (1) make use of class-independent features and classifier; (2) make use of a simple and effective feature set without high computational cost; (3) make no use of external annotated or unannotated corpus at all. At SemEval 2010 Task 8 our system achieved an F-measure of 75.43% and a accuracy of 70.22%.

## 1 Introduction

Knowledge extraction of semantic relations between pairs of nominals from English text is one important application both as an end in itself and as an intermediate step in various downstream NLP applications, such as information extraction, summarization, machine translation, QA etc. It is also useful for many auxiliary tasks such as word sense disambiguation, language modeling, paraphrasing and discourse relation processing.

In the past decade, semantic relation classification has attracted a lot of interest from researchers and a wide variety of relation classification schemes exist in the literature. However, most research work is quite different in definition of relations and granularities of various applications. That is, there is little agreement on relation inventories. SemEval 2010 Task 8 (Hendrickx et al., 2008) provides a new standard benchmark for semantic relation classification to a wider community, where it defines 9 relations including CAUSE-EFFECT, COMPONENT-WHOLE, CONTENT-CONTAINER, ENTITY-DESTINATION, ENTITY-ORIGIN, INSTRUMENT-AGENCY, MEMBER-COLLECTION, MESSAGE-TOPIC,

PRODUCT-PRODUCER, and a tenth pseudo-relation OTHER (where relation is not one of the 9 annotated relations).

Unlike the previous semantic relation task in SemEval 2007 Task 4, the current evaluation provides neither query pattern for each sentence nor manually annotated word sense (in WordNet semantic) for each nominals. Since its initiative is to provide a more realistic real-world application design that is practical, any classification system must be usable without too much effort. It needs to be easily computable. So we need to take into account the following special considerations.

1. The extracted features for relation are expected to be easily computable. That is, the steps in the feature extraction process are to be simple and direct for the purpose of reducing errors possibly introduced by many NLP tools. Furthermore, a unified (global) feature set is set up for all relations rather than for each relation.
2. Most previous work at SemEval 2007 Task 4 leveraged on external theauri or corpora (whether unannotated or annotated) (Davidov and Rappoport, 2008), (Costello, 2007), (Beamer et al., 2007) and (Nakov and Hearst, 2008) that make the task adaption to different domains and languages more difficult, since they would not have such manually classified or annotated corpus available. From a practical point of view, our system would make use of less resources.
3. Most previous work at Semeval 2007 Task 4 constructed several local classifiers on different algorithms or different feature subsets, one for each relation (Hendrickx et al., 2007) and (Davidov and Rappoport, 2008). Our approach is to build a global classifier for all relations in practical NLP settings.

Based on the above considerations, the idea of our system is to make use of external resources as less as possible. The purpose of this work is two-fold. First, it provides an overview of our simple and effective process for this task. Second, it compares different features and classification strategies for semantic relation.

Section 2 presents the system description. Section 3 describes the results and discussions. Section 4 concludes this work.

## 2 System Description

### 2.1 Features Extraction

For each training and test sentence, we reduce the annotated target entities  $e1$  and  $e2$  to single nouns *noun1* and *noun2*, by keeping their last nouns only, which we assume to be heads.

We create a global feature set for all relations. The features extracted are of three types, i.e., lexical, morpho-syntactic and semantic. The feature set consists of the following 6 types of features.

**Feature set 1: Lemma of target entities  $e1$  and  $e2$ .** The lemma of the entities annotated in the given sentence.

**Feature set 2: Stem and POS of words between  $e1$  and  $e2$ .** The stem and POS tag of the words between two nominals. First all the words between two nominals were extracted and then the Porter's stemming was performed to reduce words to their base forms (Porter, 1980). Meanwhile, OpenNLP postag tool was used to return part-of-speech tagging for each word.

**Feature set 3: syntactic pattern derived from syntactic parser between  $e1$  and  $e2$ .** Typically, the verb phrase or preposition phrase which contain the nominals are important for relation classification. Therefore, OpenNLP Parser was performed to do full syntactic parsing for each sentence. Then for each nominal, we look for its parent node in the syntactic tree until the parent node is a verb phrase or preposition phrase. Then the label of this phrase and the verb or preposition of this phrase were extracted as the syntactic features.

Besides, we also extracted other 3 feature types with the aid of WordNet.

**Feature set 4: WordNet semantic class of  $e1$  and  $e2$ .** The WordNet semantic class of each annotated entity in the relation. If the nominal has two and more words, then we examine the semantic class of " $w1\_w2$ " in WordNet. If no result returned from WordNet, we examine the semantic

class of head in the nominal. Since the cost of manually WSD is expensive, the system simply used the first (most frequent) noun senses for those words.

**Feature set 5: meronym-holonym relation between  $e1$  and  $e2$ .** The meronym-holonym relation between nominals. These information are quite important for COMPONENT-WHOLE and MEMBER-COLLECTION relations. WordNet3.0 provides meronym and holonym information for some nouns. The features are extracted in the following steps. First, for nominal  $e1$ , we extract its holonym from WN and for nominal  $e2$ , we extract its Synonyms/Hypernyms. Then, the system will check if there is same word between  $e1$ 's holonym and  $e2$ 's synonym & hypernym. The yes or no result will be a binary feature. If yes, we also examine the type of this match is "*part\_of*" or "*member\_of*" in holonym result. Then this type is also a binary feature. After that, we exchange the position of  $e1$  and  $e2$  and perform the same processing. By creating these features, the system can also take the direction of relations into account.

**Feature set 6: hyponym-hypernym relation between nominal and the word "container".** This feature is designed for CONTENT-CONTAINER relation. For each nominal, WordNet returns its hypernym set. Then the system examine if the hypernym set contains the word "container". The result leads to a binary feature.

### 2.2 Classifier Construction

Our system is to build up a global classifier based on global feature set for all 9 non-Other relations. Generally, for this multi-class task, there are two strategies for building classifier, which both construct classifier on a global feature set. The first scheme is to treat this multi-class task as an multi-way classification. Since each pair of nominals corresponds to one relation, i.e., single label classification, we build up a 10-way SVM classifier for all 10 relations. Here, we call it multi-way classification. That is, the system will construct one single global classifier which can classify 10 relations simultaneously in a run. The second scheme is to split this multi-class task into multiple binary classification tasks. Thus, we build 9 binary SVM classifiers, one for each non-Other relation. Noted that in both strategies the classifiers are built on global feature set for all relations. For the second multiple binary classification, we also exper-

imented on different prob. thresholds, i.e., 0.25 and 0.5. Furthermore, in order to reduce errors and boost performance, we also adopt the majority voting strategy to combine different classifiers.

### 3 Results and Discussion

#### 3.1 System Configurations and Results

The classifiers for all relations were optimized independently in a number of 10-fold cross-validation (CV) experiments on the provided training sets. The feature sets and learning algorithms which were found to obtain the highest accuracies for each relation were then used when applying the classifiers to the unseen test data.

Table 1 summaries the 7 system configurations we submitted and their performance on the test data.

Among the above 7 system, SR5 system shows the best macro-averaged F1 measure. Table 2 describes the statistics and performance obtained per relation on the SR5 system.

Table 3 shows the performance of these 7 systems on the test data as a function of training set size.

#### 3.2 Discussion

The first three systems are based on three feature sets, i.e., F1-F3, with different classification strategy. The next three systems are based on all six feature sets with different classification strategy. The last system adopts majority voting scheme on the results of four systems, i.e., SR1, SR2, SR4 and SR5. Based on the above series of experiments and results shown in the above 3 tables, some interesting observations can be found as follows.

Obviously, although we did not perform WSD on each nominal and only took the first noun sense as semantic class, WordNet significantly improved the performance. This result is consistent with many previous work on Semeval 2007 Task 4 and once again it shows that WordNet is important for semantic relation classification. Specifically, whether for multi-way classification or multiple binary classification, the systems involved features extracted from WordNet performed better than the others not involved WN, for example, SR4 better than SR1 (74.82% vs 60.08%), SR5 better than SR2 (75.43% vs 72.59%), SR6 better than SR3 (72.19% vs 68.50%).

Generally, the performance of multiple binary classifier is better than multi-way classifier. That means, given a global feature set for 9 relations, the performance of 9 binary classifiers is better than a 10-way classifier. Specifically, when F1-F3 are involved, SR2 (72.59%) and SR3 (68.50%) are both better than SR1 (60.08%). However, when F1-F6 feature sets are involved, the performance of SR4 is between that of SR5 and SR6 in terms of macro-averaged  $F_1$  measure. With respect to accuracy measure (Acc), SR4 system performs the best.

Moreover, for multiple binary classification, the threshold of probability has impact on the performance. Generally, the system with prob. threshold 0.25 is better than that with 0.5, for example, SR2 better than SR3 (72.59% vs 68.50%), SR5 better than SR6 (75.43% vs 72.19%).

As an ensemble system, SR7 combines the results of SR1, SR2, SR4 and SR5. However, this majority voting strategy has not shown significant improvements. The possible reason may be that these classifiers come from a family of SVM classifiers and thus the random errors are not significantly different.

Besides, one interesting observation is that SR4 system achieved the top 2 performance on TD1 data amongst all participating systems. This shows that, even with less training data, SR4 system achieves good performance.

#### Acknowledgments

This work is supported by grants from National Natural Science Foundation of China (No.60903093), Shanghai Pujiang Talent Program (No.09PJ1404500) and Doctoral Fund of Ministry of Education of China (No.20090076120029).

#### References

- I. Hendrickx, S. N. Kim, Z. Kozareva, P. Nakov, D. Ó Séaghdha, S. Padó, M. Pennacchiotti, L. Romano and S. Szpakowicz. SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations Between Pairs of Nominals. In *Proceedings of the 5th SIGLEX Workshop on Semantic Evaluation*, pp.94-99, 2010, Uppsala, Sweden.
- D. Davidov and A. Rappoport. Classification of Semantic Relationships between Nominals Using Pattern Clusters. *Proceedings of ACL-08: HLT*, pp.227-235, 2008.
- F. J. Costello. UCD-FC: Deducing semantic relations using WordNet senses that occur frequently

Run	Feature Set	Classifier	P (%)	R (%)	F <sub>1</sub> (%)	Acc (%)
SR1	F1-F3	multi-way classification	70.69	58.05	60.08	57.05
SR2	F1-F3	multiple binary (prob. threshold =0.25)	74.02	71.61	72.59	67.10
SR3	F1-F3	multiple binary (prob. threshold =0.5)	80.25	60.92	68.50	62.02
SR4	F1-F6	multi-way classification	75.72	74.16	74.82	70.52
SR5	F1-F6	multiple binary (prob. threshold =0.25)	75.88	75.29	<b>75.43</b>	70.22
SR6	F1-F6	multiple binary (prob. threshold =0.5)	83.08	64.72	72.19	65.81
SR7	F1-F6	majority voting based on SR1, SR2, SR4 and SR5	74.83	75.97	75.21	70.15

Table 1: Summary of 7 system configurations and performance on the test data. Precision, Recall, F1 are macro-averaged for system’s performance on 9 non-Other relations and evaluated with directionality taken into account.

Run	Total #	P(%)	R (%)	F <sub>1</sub> (%)	Acc (%)
Cause-Effect	328	83.33	86.89	85.07	86.89
Component-Whole	312	74.82	65.71	69.97	65.71
Content-Container	192	79.19	81.25	80.21	81.25
Entity-Destination	292	79.38	86.99	83.01	86.99
Entity-Origin	258	81.01	81.01	81.01	81.01
Instrument-Agency	156	63.19	58.33	60.67	58.33
Member-Collection	233	73.76	83.26	78.23	83.26
Message-Topic	261	75.2	73.18	74.17	73.18
Product-Producer	231	73.06	61.04	66.51	61.04
Other	454	38.56	40.09	39.31	40.09
Micro-Average		76.88	76.27	76.57	70.22
Macro-Average		75.88	75.29	75.43	70.22

Table 2: Performance obtained per relation on SR5 system. Precision, Recall, F1 are macro-averaged for system’s performance on 9 non-Other relations and evaluated with directionality taken into account.

Run	TD1		TD2		TD3		TD4	
	F <sub>1</sub> (%)	Acc (%)	F <sub>1</sub> (%)	Acc (%)	F <sub>1</sub> (%)	Acc (%)	F <sub>1</sub> (%)	Acc (%)
SR1	52.13	49.50	56.58	54.84	58.16	56.16	60.08	57.05
SR2	46.24	38.90	47.99	40.45	69.83	64.67	72.59	67.10
SR3	39.89	34.56	42.29	36.66	65.47	59.59	68.50	62.02
SR4	<b>67.95</b>	<b>63.45</b>	<b>70.58</b>	<b>66.14</b>	<b>72.99</b>	<b>68.94</b>	74.82	<b>70.52</b>
SR5	49.32	41.59	50.70	42.77	72.63	67.72	<b>75.43</b>	70.22
SR6	42.88	36.99	45.54	39.57	69.87	64.00	72.19	65.81
SR7	58.67	52.71	58.87	53.18	72.79	68.09	75.21	70.15

Table 3: Performance of these 7 systems on the test data as a function of training set size. The four training subsets, TD1, TD2, TD3 and TD4, have 1000, 2000, 4000 and 8000 (complete) training samples respectively. F1 is macro-averaged for system’s performance on 9 non-Other relations and evaluated with directionality taken into account.

in a database of noun-noun compounds. *ACL SemEval’07 Workshop*, pp.370C373, 2007.

B. Beamer, S. Bhat, B. Chee, A. Fister, A. Rozovskaya and R.Girju. UIUC: A knowledge-rich approach to identifying semantic relations between nominals. *ACL SemEval’07 Workshop*, pp.386-389, 2007.

I. Hendrickx, R. Morante, C. Sporleder and A. Bosch. ILK: machine learning of semantic relations with

shallow features and almost no data. *ACL SemEval’07 Workshop*, pp.187C190, 2007.

P. Nakov and M. A. Hearst. Solving Relational Similarity Problems Using the Web as a Corpus. In *Proceedings of ACL*, pp.452-460, 2008.

M. Porter. An algorithm for suffix stripping. In *Program*, vol. 14, no. 3, pp.130-137, 1980.