FBK_NK: a WordNet-based System for Multi-Way Classification of Semantic Relations

Matteo Negri and Milen Kouylekov
FBK-Irst
Trento, Italy
{negri,kouylekov}@fbk.eu

Abstract

We describe a WordNet-based system for the extraction of semantic relations between pairs of nominals appearing in English texts. The system adopts a lightweight approach, based on training a Bayesian Network classifier using large sets of binary features. Our features consider: *i)* the context surrounding the annotated nominals, and *ii)* different types of knowledge extracted from WordNet, including direct and explicit relations between the annotated nominals, and more general and implicit evidence (*e.g.* semantic boundary collocations). The system achieved a Macro-averaged F1 of 68.02% on the "Multi-Way Classification of Semantic Relations Between Pairs of Nominals" task (Task #8) at SemEval-2010.

1 Introduction

The "Multi-Way Classification of Semantic Relations Between Pairs of Nominals" task at SemEval-2010 (Hendrickx et al., 2010) consists in: *i)* selecting from an inventory of nine possible relations the one that most likely holds between two annotated nominals appearing in the input sentence, and *ii)* specifying the order of the nominals as the arguments of the relation. In contrast with the semantic relations classification task (Task #4) at SemEval-2007 (Girju et al., 2007), which treated each semantic relation separately as a single two-class (positive vs. negative) classification task, this year's edition of the challenge presented participating systems with a more difficult and realistic *multi-way* setup, where the relation *Other* can also be assigned if none of the nine relations is suitable for a given sentence. Examples of the possible markable relations are reported in Table 1[1].

The objective of our experiments with the proposed task is to develop a Relation Extraction system based on shallow linguistic processing, taking the most from available thesauri and ontologies. As a first step in this direction, our submitted runs have been obtained by processing the input sentences only to lemmatize their terms, and by using WordNet as the sole source of knowledge.

Similar to other approaches (Moldovan and Badulescu, 2009; Beamer et al., 2009), our system makes use of *semantic boundaries* extracted from the WordNet IS-A backbone. Such boundaries (*i.e.* divisions in the WordNet hierarchy that best generalize over the training examples) are used to define pairs of high-level synsets with high correlation with specific relations. For instance, <*microorganism#1, happening#1*> and <*writing#1, consequence#1*> are extracted from the training data as valid high-level collocations respectively for the relations *Cause-Effect* and *Message-Topic*. Besides exploiting the WordNet IS-A hierarchy, the system also uses the holo-/meronymy relations, and information derived from the WordNet glosses to capture specific relations such as *Member-Collection* and *Product-Producer*. In addition, the context surrounding the annotated nominals is represented as a *bag-of-words/synonyms* to enhance the relation extraction process. Several experiments have been carried out encoding all the information as large sets of binary features (up to ∼6200) to train a Bayesian Network classifier available in the Weka[2] toolkit. To capture both the *relations* and the *order* of

[1] In the first example the order of the nominals is (<e2>,<e1>), while in the others is (<e1>,<e2>)

[2] http://www.cs.waikato.ac.nz/ml/weka/

Proceedings of the 5th International Workshop on Semantic Evaluation, ACL 2010, pages 202–205,
Uppsala, Sweden, 15-16 July 2010. ©2010 Association for Computational Linguistics

1	Cause-Effect(e2,e1)	A person infected with a particular <e1>flu</e1> <e2>virus</e2> strain develops an antibody against that virus.
2	Instrument-Agency(e1,e2)	The <e1>river</e1> once powered a <e2>grist mill</e2>.
3	Product-Producer(e1,e2)	The <e1>honey</e1> <e2>bee</e2> is the third insect genome published by scientists, after a lab workhorse, the fruit fly, and a health menace, the mosquito.
4	Content-Container(e1,e2)	I emptied the <e1>wine</e1> <e2>bottle</e2> into my glass and toasted my friends.
5	Entity-Origin(e1,e2)	<e1>This book</e1>is from the 17th <e2>century</e2>.
6	Entity-Destination(e1,e2)	<e1>Suspects</e1> were handed over to the <e2>police station</e2>.
7	Component-Whole(e1,e2)	<e1>Headlights</e1> are considered as the eyes of the <e2>vehicle</e2>.
8	Member-Collection(e1,e2)	Mary looked back and whispered: 'I know every <e1>tree</e1> in this <e2>forest</e2>, every scent'.
9	Message-Topic(e1,e2)	Here we offer a selection of our favourite <e1>books</e1> on military <e2>history</e2>.

Table 1: SemEval-2010 Task #8 semantic relations.

their arguments, training sentences having opposite argument directions for the same relation have been handled separately, and assigned to different classes (thus obtaining 18 classes for the nine target relations, plus one for the *Other* relation).

The following sections overview our experiments, describing the features used by the system (Section 2), and the submitted runs with the achieved results (Section 3). A concluding discussion on the results is provided in Section 4.

2 Features used

The system uses two types of boolean features: WordNet features, and context features.

2.1 WordNet features

WordNet features consider different types of knowledge extracted from WordNet 3.0.

Semantic boundary collocations. Collocations of high-level synsets featuring a high correlation with specific relations are acquired from the training set using a bottom-up approach. Starting from the nominals annotated in the training sentences (<e1> and <e2>), the WordNet IS-A backbone is climbed to collect all their ancestors. Then, all the ancestors' collocations occurring at least *n* times for at most *m* relations are retained, and treated as boolean features (set to 1 for a given sentence if its annotated nominals appear among their hyponyms). The *n* and *m* parameters are optimized on the training set.

Holo-/meronymy relations. These boolean features are set to 1 every time a pair of annotated nominals in a sentence is *directly* connected by holo-/meronyny relations. They are particularly appropriate to capture the *Component-Whole* and *Member-Collection* relations, as in the 8th example in Table 1 (where *tree#1* is an *holonym* of *forest#1*). Due to time constraints, we did not explore the possibility to generalize these features considering transitive closures of the nominals' hypo-/hypernyms. This possibility could allow to handle sentences like *"A <e1>herd</e1> is a large group of <e2>animals</e2>."* Here, though *herd#1* and *animal#1* are not directly connected by the meronymy relation, all the *herd#1* meronyms have *animal#1* as a common ancestor.

Glosses. Given a pair of annotated nominals <e1>,<e2>, these features are set to 1 every time either <e1> appears in the gloss of <e2>, or vice-versa. They are intended to support the discovery of relations in the case of consecutive nominals (*e.g. honey#1* and *bee#1* in the 3rd example in Table 1), where contextual information does not provide sufficient clues to make a choice. In our experiments we extracted features from both tokenized and lemmatized words (both nominals, and gloss words). Also in this case, due to time constraints we did not explore the possibility to generalize the feature considering the nominals' hypo-/hypernyms. This possibility could allow to handle sentences like examples 1 and 4 in Table 1. For instance in example 4, the gloss of "*bottle*" contains two hypernyms of *wine#1*, namely *drink#3* and *liquid#1*, that could successfully trigger the *Content-Container* relation.

Synonyms. While the previous features operate with the annotated nominals, WordNet synonyms are used to generalize the other terms in the sentence, allowing to extract different types of contextual features (see the next Section).

2.2 Context features

Besides the annotated nominals, also specific words (and word combinations) appearing in the surrounding context often contribute to trigger the target relations. Distributional evidence is captured by considering word contexts *before*, *between*, and *after* the annotated nominals. To this aim, we experimented with windows of different size, containing words that occur in the training set a variable number of times. Both the parameters (*i.e.* the size of the windows, and the number of occurrences) are optimized on training data. In our experiments we extracted contextual features from lemmatized sentences.

3 Submitted runs and results

Our participation to the SemEval-2010 Task #8 consisted in four runs, with the best one (FBK_NK-RES1) achieving a Macro-averaged F1 of 68.02% on the test data. For this submission, the overall training and test running times are about 12'30" and 1'30" respectively, on an Intel Core2 Quad 2.66GHz with 4GB RAM.

FBK_NK-RES1. This run has been obtained adopting a conservative approach, trying to minimize the risk of overfitting the training data. The features used can be summarized as follows:

- Semantic boundary collocations: all the collocations of <e1> and <e2> ancestors occurring at least 10 times in the training set (*m* param.), for at most 3 relations (*n* param.);
- Holo-/meronymy relations between the annotated nominals;
- Glosses: handled at the level of *tokens*;
- Context features: *left*, *between*, and *right* context windows of size 3-ALL-3 words respectively. Number of occurrences: 25 (*left*), 10 (*between*), 25 (*right*).

On the **training set**, the Bayesian Network classifier (trained with 2239 features, and evaluated with 10-fold cross-validation) achieves an Accuracy of 65.62% (5249 correctly classified instances out of 8000), and a Macro F1 of 78.15%.

FBK_NK-RES2. Similar to the first run, but:

- Semantic boundary collocations: *m*=9, *n*=3;
- Glosses: handled at the level of *lemmas*;
- Context features: *left*, *between*, and *right* context windows of size 4-ALL-1 words respectively (occurrences: 25-10-25).

Run	1000	2000	4000	8000
FBK_NK-RES1	55.71	64.06	67.80	**68.02**
FBK_NK-RES2	54.27	63.68	67.08	67.48
FBK_NK-RES3	54.25	62.73	66.11	66.90
FBK_NK-RES4	44.11	58.85	63.06	65.84

Table 2: Test results (Macro-averaged F1) using different amounts of training sentences.

Based on the observation of system's behaviour on the training data, the objectives of this run were to: *i)* add more collocations as features, *ii)* increase the importance of terms appearing in the *left* context, *iii)* reduce the importance of terms appearing in the *right* context, and *iv)* increase the possibility of matching the nominals with gloss terms by considering their respective lemmas. On the **training set**, the classifier (trained with 2998 features) achieves 66.92% Accuracy (5353 correctly classified instances), and a Macro F1 of 79.56%.

FBK_NK-RES3. Similar to the second run, but considering the synonyms of the most frequent sense of the words *between* <e1> and <e2>.

The goal of this run was to generalize the context *between* nominals, by considering word lemmas. On the **training set**, the classifier (trained with 2998 features) achieves an Accuracy of 64.94% (5195 correctly classified instances), and a Macro F1 of 77.38%.

FBK_NK-RES4. Similar to the second run, but considering semantic boundary collocations occurring at least 7 times in the training set (*m* param.), for at most 3 relations (*n* param.).

The goal of this run was to further increase the number of collocations used as features. On the **training set**, the classifier (trained with 6233 features) achieves achieves 68.12% Accuracy (5449 correct classifications), and 82.24% Macro F1.

As regards the results on the test set, Table 2 reports the scores achieved by each run using different portions of the training set (1000, 2000, 4000, 8000 examples), while Figure 1 shows the learning curves for each relation of our best run.

4 Discussion and conclusion

As can be seen from Table 2, the results contradict our expectations about the effectiveness of our less conservative configurations and, in particular, about the utility of using larger amounts of semantic boundary collocations. The performance

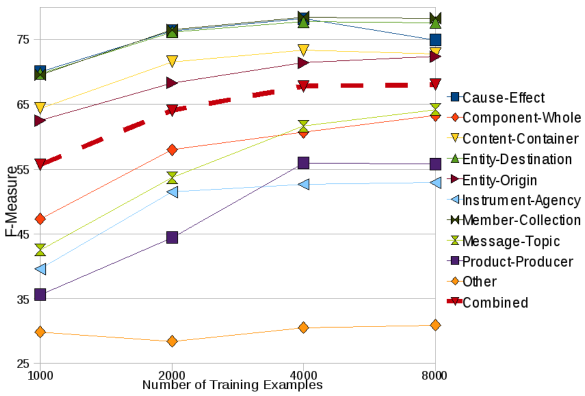

Figure 1: Learning curves on the test set (FBK_NK-RES1).

decrease from Run2 to Run4[3] clearly indicates an overfitting problem. Though suitable to model the training data, the additional collocations were not encountered in the test set. This caused a bias towards the *Other* relation, which reduced the overall performance of the system.

Regarding our best run, Figure 1 shows different system's behaviours with the different target relations. For some of them (*e.g. Entity-Destination*, *Cause-Effect*) better results are motivated by the fact that they are often triggered by frequent unambiguous word patterns (*e.g.* *"<e1>has been moved to a <e2>"*, *"<e1> causes <e2>"*). Such relations are effectively handled by the context features which, in contrast, are inadequate for those expressed with high lexical variability. This is particularly evident with the *Other* relation, for which the acquired context features poorly discriminate positive from negative examples even on the training set.

For some relations additional evidence is successfully brought by the WordNet features. For instance, the good results for *Member-Collection* demonstrate the usefulness of the holo-/meronymy features.

As regards semantic boundary collocations, to check their effectiveness we performed a *post-hoc* analysis of those used in our best run. Such analysis was done in two ways: *i)* by counting the number of collocations acquired on the training set for each relation R_i, and *ii)* by calculating the ambiguity of each R_i's collocation on the training set (*i.e.* the average number of other relations activated by the collocation). The analysis revealed that the top performing relations (*Member-Collection*, *Entity-Destination*, *Cause-Effect*, and *Content-Container*) are those for which we acquired lots of unambiguous collocations. These findings also explain the poor performance on the *Instrument-Agency* and the *Other* relation. For *Instrument-Agency* we extracted the lowest number of collocations, which were also the most ambiguous ones. For the *Other* relation the high ambiguity of the collocations extracted is not compensated by their huge number (around 50% of the total collocations acquired).

In conclusion, considering *i)* the level of processing required (only lemmatization), *ii)* the fact that WordNet is used as the sole source of knowledge, and *iii)* the many possible solutions left unexplored due to time constraints, our results demonstrate the validity of our approach, despite its simplicity. Future research will focus on a better use of semantic boundary collocations, on more refined ways to extract knowledge from WordNet, and on integrating other knowledge sources (*e.g.* SUMO, YAGO, Cyc).

Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement n. 248531 (CoSyne project).

References

B. Beamer, A. Rozovskaya, and R. Girju 2008. *Automatic Semantic Relation Extraction with Multiple Boundary Generation.* Proceedings of The National Conference on Artificial Intelligence (AAAI).

R. Girju, P. Nakov, V. Nastase, S. Szpakowicz, P. Turney, and D. Yuret 2007. *SemEval-2007 task 04: Classification of semantic relations between nominals.* Proceedings of the 4th Semantic Evaluation Workshop (SemEval-2007).

I. Hendrickx et al. 2010. *SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations Between Pairs of Nominals* Proceedings of the 5th SIGLEX Workshop on Semantic Evaluation.

D. Moldovan, A. Badulescu 2005. *A Semantic Scattering Model for the Automatic Interpretation of Genitives.* Proceedings of The Human Language Technology Conference (HLT).

[3]The only difference between Run2 and Run4 is the addition of around 4000 semantic boundary collocations, which lead to an overall 2.4% F1 performance decrease. The decrease mainly comes in terms of Recall (from 65.91% in Run2 to 63.35% in Run4).