Non-atomic Classification to Improve a Semantic Role Labeler for a Low-resource Language

Richard Johansson
Språkbanken, Department of Swedish, University of Gothenburg
Box 100, SE-40530 Gothenburg, Sweden
richard.johansson@gu.se

Abstract

Semantic role classification accuracy for most languages other than English is constrained by the small amount of annotated data. In this paper, we demonstrate how the frame-to-frame relations described in the FrameNet ontology can be used to improve the performance of a FrameNet-based semantic role classifier for Swedish, a low-resource language. In order to make use of the FrameNet relations, we cast the semantic role classification task as a *non-atomic label prediction task*. The experiments show that the cross-frame generalization methods lead to a 27% reduction in the number of errors made by the classifier. For previously unseen frames, the reduction is even more significant: 50%.

1 Introduction

The FrameNet lexical database and annotated corpora, based on the theory of semantic frames (Fillmore et al., 2003), have allowed the implementation of automatic systems to extract *semantic roles* (Gildea and Jurafsky, 2002; Johansson and Nugues, 2007; Màrquez et al., 2008; Das et al., 2010).

Since the original FrameNet is developed for the English language, most research on semantic role extraction has focused exclusively on English. However, the English FrameNet has inspired similar efforts for other languages. For instance, the ongoing development of a Swedish FrameNet (Borin et al., 2010) allows us to investigate the feasibility of using this resource in constructing an automatic role-semantic analyzer for Swedish. However, due to the fact that the Swedish FrameNet annotation process is in a fairly early stage, not much annotated material is available, and this limits the performance attainable by automatic classifiers trained on these data. In particular, the scarce amount of data makes it very hard for the machine learning methods to discern general linguistic facts concerning the syntactic–semantic linking patterns, such as the relation between the voice of a verb, the syntactic functions of its arguments, and the semantic roles of the arguments (Dowty, 1991).

In this paper, we show that the inter-frame relations described in the FrameNet ontology allow us to generalize across frames. This allows the classifier to learn general linguistic facts, and it also leads to more efficient use of the annotated data. To allow this kind of generalization, we formulate the semantic role selection problem as a classification task with non-atomic labels. This cross-frame generalization method reduces the number of errors made by the classifier by 27%, improving the accuracy from 54.4 to 66.5. When evaluating on frames for which the classifier has not been trained, the accuracy improves from 7.2 (random performance) to 53.4, a 50% error reduction.

2 The Swedish FrameNet

The Swedish FrameNet, SweFN, is a lexical resource under development (Friberg Heppin and Toporowska Gronostaj, 2012), based on the English version of FrameNet constructed by the Berkeley research group (Baker et al., 1998). It is found on the SweFN website[1], and is available as a free resource.

The SweFN frames and frame names correspond to the English ones, with some exceptions, as do the selection of frame elements including definitions and internal relations. The meta-information about the frames, such as semantic relations between frames, is also transferred from the Berkley FrameNet. Compared to the Berkeley FrameNet, SweFN is expanded with information about the domain of the frames, at present: general language, the medical and the art domain.

[1] http://spraakbanken.gu.se/eng/swefn

*First Joint Conference on Lexical and Computational Semantics (*SEM)*, pages 95–99,
Montréal, Canada, June 7-8, 2012. ©2012 Association for Computational Linguistics

At the time of writing this paper, SweFN covered 519 frames with around 18,000 lexical units. The lexical units are gathered from SALDO, a free Swedish electronic association lexicon (Borin and Forsberg, 2009). A lexical unit from SALDO cannot populate more than one frame. At present there are 31 frames in SweFN which do not match a frame in the Berkeley FrameNet. Of these, there are eight completely new frames while the others have been modified in some way.

Crucially for the work presented in this paper, each frame is exemplified with at least one sentence. The number of sentences is currently 2,974. The most well-annotated frames are EXPERIENCER_OBJ with 38 sentences, CAUSE_MOTION with 21, and CAUSE_HARM with 19. These sentences form the training material used in the following sections.

3 System Implementation

In this section, we describe the implementation of our semantic role labeling system. In order to be useful on its own, such a system needs to solve several tasks: (1) identification of predicate words; (2) assignment of frames to predicate words; (3) identification of role fillers; (4) assignment of semantic role labels to role fillers. In this paper, we focus exclusively on the semantic role classification task.

3.1 Baseline: A Classifier for Swedish Semantic Roles

Following most previous implementations, we used a syntactic parse tree as the basis of the semantic role extraction; we assumed that every semantic role span coincides with the projection of a subtree in the syntactic tree. The tasks of segmentation and labeling then reduce to a classification problem on syntactic tree nodes. Each sentence was parsed by the LTH dependency parser (Johansson and Nugues, 2008a), which we trained on a Swedish treebank (Nilsson et al., 2005). Figure 1 shows a sentence annotated with a dependency tree and semantic roles.

The semantic role labeling classifier was implemented as a linear multiclass classifier with a flexible output space depending on the frame of the given predicate; we trained this classifier using an online learning algorithm (Crammer et al., 2006). In addition, we imposed a uniqueness constraint on the labels output by the classifier, so that every role may appear only once for a given predicate.

We considered a large number of features for the classifier (Table 1). Most of these are commonly used features taken from the standard literature on semantic role labeling. We then applied a standard greedy forward feature selection procedure to determine which of them to use. The features containing SALDO ID refer to the entry identifiers in the SALDO lexicon. Note that the POS tags have coarse and fine variants, such as VERB and VERB-FINITE-PRESENT-ACTIVE respectively, and we used both of them.

Semantic role classifiers rely heavily on lexical features (Johansson and Nugues, 2008b), and this may lead to brittleness; in order to increase robustness, we added features based on hierarchical clusters constructed using the Brown algorithm (Brown et al., 1992). The Brown algorithm clusters word into hierarchies represented as bit strings. Based on tuning on a development set, we found that it was best not to use the full bit string, but only a prefix if the string was longer than 12 bits.

FRAME
DEPENDENCY RELATION PATH
FRAME ELEMENTS
POSITION
VOICE
ARGUMENT HEAD SALDO ID
ARGUMENT HEAD LEMMA
ARGUMENT HEAD POS (FINE)
PREDICATE POS (FINE)
ARGUMENT POS (COARSE)
ARGUMENT RIGHT CHILD POS (COARSE)
ARGUMENT WORD
PREDICATE WORD CLUSTER
ARGUMENT WORD CLUSTER

Table 1: List of classifier features.

3.2 A Classifier Using Non-atomic Semantic Role Labels

The classifier described above is a quite typical example of how semantic role classifiers are normally implemented: each frame is independent of all other frames. However, in our case, when the amount of training data is quite small, the limitations of this standard approach become apparent:

- Since there are many frames, the amount of training data for each frame is very limited.

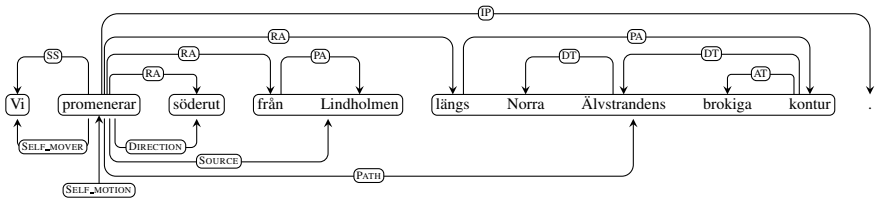

Figure 1: A sentence with dependency syntax (above) and semantic role structure (below).

- Basic linguistic facts, such as which roles are likely to appear in subject position, have to be relearned for each frame.

To remedy these problems, we developed a classifier using *non-atomic labels*: instead of just a simple label INGESTION:INGESTOR, the classifier can predict several labels, using some sort of decomposition into meaningful parts. In §3.3, we will describe several such decompositions.

As described above, our baseline classifier is a standard linear classifier. Assume that the frame F defines a set of semantic roles $r_1, \ldots, r_n$, then the classifier predicts a semantic role r^* for a given argument a using this model:

$$r^* = \arg\max_{r \in F} w \cdot \Phi(a, r)$$

Here Φ is a feature function describing features of the argument a taking the semantic role r, and w is a weight vector produced by some training algorithm.

This classifier model can easily be generalized to the non-atomic case. We then assume that each role r can be decomposed using a decomposition function D, which returns a set of labels. We now apply the feature function to each sub-label l instead of the main label r.

$$r^* = \arg\max_{r \in F} \sum_{l \in D(r)} w \cdot \Phi(a, l)$$

Non-atomic classification has been described in a number of publications. It is fairly common in text categorization, where *hierarchical* classification is probably the most common type. One of the most similar to ours is the action classifier by Roth and Tu (2009), which handled a large label set by decomposing the labels into meaningful parts.

3.3 Generalization Methods

We investigated several ways of analyzing the labels, and most of them were based on the properties of the frames, defined in the FrameNet ontology. The Swedish FrameNet currently does not define such properties, but since the frames and frame elements are for the most part based on their English counterparts, we used the English ontology. In case of mismatch, we just left the label in its original state.

The first method we tried was based on frame-to-frame relations. We used the following relations:

INHERITANCE: specific to general, e.g. COMMUNICATION_NOISE to COMMUNICATION.

SUBFRAME: from component to complex, e.g. SETTING_OUT to TRAVEL.

CAUSATIVE-OF: causative to inchoative, e.g. CAUSE_TEMPERATURE_CHANGE to INCH._CHANGE_OF_TEMP..

INCHOATIVE-OF: inchoative to stative, e.g. INCH._CHANGE_OF_TEMP. to TEMPERATURE.

USING: child to parent, e.g. COMMUNICATION_NOISE to MAKE_NOISE.

PERSPECTIVE-ON: perspectivized to neutral, e.g. RIDE_VEHICLE to USE_VEHICLE.

To analyze a label in terms of frame-to-frame relations, we applied the transitive closure of each relation and returned the resulting set. For instance, when applying the Inheritance relation to the INGESTION:INGESTOR label, we get the following set: { INGESTION:INGESTOR, INGEST_SUBSTANCE:INGESTOR, MANIPULATION:AGENT, INTENT._AFFECT:AGENT, INTENT._ACT:AGENT, TRANS._ACTION:AGENT }.

The second method was based on the *semantic type* of the semantic role. For instance, the INGESTION:INGESTOR role needs to be filled by an entity of the semantic type SENTIENT. The decomposition of this role then simply becomes { INGESTION:INGESTOR, SENTIENT }.

The third method was based on the simple notion *label generalization*: if two semantic roles

in two different frames have the same name, then we use the same label. For instance, we change the INGESTION:INGESTOR and INGEST_SUBSTANCE:INGESTOR to INGESTOR. We normalized the spelling, punctuation, and capitalization of the labels before generalizing.

4 Experiments

We evaluated the classifier on the example sentences in the Swedish FrameNet. The frame and the argument were given to the classifier, which then had to predict the semantic role. We evaluated in two different ways: *In-frame* evaluation, where a 5-fold cross-validation was carried out over the set of sentences, and *Out-frame* evaluation, where the cross-validation was done over the set of frames. The out-frame setting simulates the situation where a new frame has been defined, but no training data have been annotated. Without any sort of cross-frame generalization, the classification in the out-frame setting becomes a random baseline.

Table 2 shows the results of using the frame-to-frame relations for analyzing the semantic role labels. We see that decomposition based on Inheritance is by far the most effective of these, although the highest performance is obtained when combining all types of relation-based decompositions.

Classifier	In-frame	Out-frame
Baseline	54.4	7.2
Inheritance	58.7	28.1
Using	55.8	20.5
Subframe	54.8	11.5
Causative-of	54.5	9.7
Perspective-on	54.5	8.1
Inchoative-of	54.4	8.0
All except Inheritance	56.0	24.0
All relations	59.6	36.9

Table 2: Classification results with generalization based on frame-to-frame relations.

The effect of analyzing labels in terms of semantic type is similar. The in-frame performance is higher than that of relation-based decomposition, while the out-frame performance is a bit lower. The two generalization methods seem to complement each other, since we get a higher performance by combining them. Table 3 shows the results.

Classifier	In-frame	Out-frame
Semantic type	61.7	31.7
Semantic type + relations	63.5	42.6

Table 3: Adding semantic type generalization.

Finally, Table 4 shows the effect of using label generalization. This is by far the most effective method. However, we get even higher performance by combining it with the other two methods.

Classifier	In-frame	Out-frame
Label generalization	65.9	51.5
LG + ST + relations	66.5	53.4

Table 4: Results with label generalization.

5 Discussion

When developing NLP systems for a low-resource language, it is crucial to make effective use of the available data. In the case of FrameNet semantic role classification, one way to improve the use of the data is to generalize the roles across the frames. This also makes sense from a theoretical point of view, since predicting multiple labels allows the machine learner to learn general facts as well as specifics.

This work builds on previous work in multi-label classification. For the task of FrameNet semantic role classification, the work most closely related to ours is that by Matsubayashi et al. (2009), which defined a classifier making use of *role groups*; the effect of the role groups turns out to be similar to our non-atomic classification approach.

Our experiments showed very significant error reductions. This was especially notable in the case of out-frame evaluation, which is to be expected since the baseline in this case was a random selection. The best classifier used all three types of label decomposition, and achieved a 26% in-frame and a 50% out-frame error reduction.

Acknowledgements

The research presented here was supported by the Swedish Research Council (the project *Swedish Framenet++*, VR dnr 2010-6013) and by the University of Gothenburg through its support of the Centre for Language Technology and Språkbanken (the Swedish Language Bank).

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, pages 86–90, Montréal, Canada.

Lars Borin and Markus Forsberg. 2009. All in the family: A comparison of SALDO and WordNet. In *Proceedings of the Nodalida 2009 Workshop on WordNets and other Lexical Semantic Resources – between Lexical Semantics, Lexicography, Terminology and Formal Ontologies*, Odense, Denmark.

Lars Borin, Dana Dannélls, Markus Forsberg, Maria Toporowska Gronostaj, and Dimitrios Kokkinakis. 2010. The past meets the present in the Swedish FrameNet++. In *Proceedings of EURALEX*.

Peter F. Brown, Peter V. deSouza, Robert L. Mercer, Vincent J. Della Pietra, and Jenifer C. Lai. 1992. Class-based n-gram models of natural language. *Computational Linguistics*, 18(4):467–479.

Koby Crammer, Ofer Dekel, Joseph Keshet, Shai Shalev-Schwartz, and Yoram Singer. 2006. Online passive-aggressive algorithms. *Journal of Machine Learning Research*, 2006(7):551–585.

Dipanjan Das, Nathan Schneider, Desai Chen, and Noah A. Smith. 2010. Probabilistic frame-semantic parsing. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 948–956, Los Angeles, United States.

David R. Dowty. 1991. Thematic proto-roles and argument selections. *Language*, 67(3):574–619.

Charles Fillmore, Christopher Johnson, and Miriam Petruck. 2003. Background to FrameNet. *International Journal of Lexicography*, 16(3):235–250.

Karin Friberg Heppin and Maria Toporowska Gronostaj. 2012. The rocky road towards a Swedish FrameNet. In *Proceedings of LREC-2012 (to appear)*.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288.

Richard Johansson and Pierre Nugues. 2007. Semantic structure extraction using nonprojective dependency trees. In *Proceedings of SemEval-2007*, pages 227–230, Prague, Czech Republic, June 23-24.

Richard Johansson and Pierre Nugues. 2008a. Dependency-based syntactic–semantic analysis with PropBank and NomBank. In *Proceedings of the CoNLL Shared Task*, pages 183–187, Manchester, United Kingdom.

Richard Johansson and Pierre Nugues. 2008b. The effect of syntactic representation on semantic role labeling. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 393–400, Manchester, United Kingdom.

Lluís Màrquez, Xavier Carreras, Ken Litkowski, and Suzanne Stevenson. 2008. Semantic role labeling: An introduction to the special issue. *Computational Linguistics*, 34(2):145–159.

Yuichiroh Matsubayashi, Naoaki Okazaki, and Jun'ichi Tsujii. 2009. A comparative study on generalization of semantic roles in FrameNet. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 19–27, Suntec, Singapore.

Jens Nilsson, Johan Hall, and Joakim Nivre. 2005. MAMBA meets TIGER: Reconstructing a Swedish treebank from antiquity. In *Proceedings of NODALIDA Special Session on Treebanks*.

Dan Roth and Yuancheng Tu. 2009. Aspect guided text categorization with unobserved labels. In *Proceedings of the IEEE Conference on Data Mining*, Miami, United States.