Single Classifier Approach for Verb Sense Disambiguation based on Generalized Features

Daisuke Kawahara[†], **Martha Palmer**[‡]

[†]Graduate School of Informatics, Kyoto University, Kyoto, Kyoto 606-8501, Japan

[‡]Department of Linguistics, University of Colorado Boulder, Boulder, CO 80302, USA

dk@i.kyoto-u.ac.jp, martha.palmer@colorado.edu

Abstract

We present a supervised method for verb sense disambiguation based on VerbNet. Most previous supervised approaches to verb sense disambiguation create a classifier for each verb that reaches a frequency threshold. These methods, however, have a significant practical problem that they cannot be applied to rare or unseen verbs. In order to overcome this problem, we create a single classifier to be applied to rare or unseen verbs in a new text. This single classifier also exploits generalized semantic features of a verb and its modifiers in order to better deal with rare or unseen verbs. Our experimental results show that the proposed method achieves equivalent performance to per-verb classifiers, which cannot be applied to unseen verbs. Our classifier could be utilized to improve the classifications in lexical resources of verbs, such as VerbNet, in a semi-automatic manner and to possibly extend the coverage of these resources to new verbs.

Keywords: verb sense disambiguation, single classifier, word representations

1. Introduction

A verb plays a primary role in conveying the meaning of a sentence. Since capturing the sense of a verb is essential for natural language processing (NLP), lexical resources for verbs play an important role in NLP.

VerbNet is one of such lexical resources, in which verbs are organized into classes on the basis of their syntactic and semantic behavior (Kipper-Schuler, 2005). It has been used in many NLP applications that need to consider semantics in particular, such as word sense disambiguation (Dang, 2004), semantic parsing (Swier and Stevenson, 2005; Shi and Mihalcea, 2005) and discourse parsing (Subba and Di Eugenio, 2009). To make use of VerbNet in such practical applications, it is necessary to map each verb token in a text to a VerbNet class. This is a task of verb sense disambiguation, which has been resolved by supervised approaches in recent years (Girju et al., 2005; Abend et al., 2008; Chen and Eugenio, 2010; Brown et al., 2011; Croce et al., 2012).

Most previous supervised approaches to verb sense disambiguation create a classifier for each verb that reaches a frequency threshold (e.g., 10 times). These methods, however, have a significant practical problem that they cannot be applied to rare or unseen verbs. In order to overcome this problem, we propose a single supervised classifier for this task. This classifier exploits generalized features of a verb and its modifiers in order to better deal with rare or unseen verbs. Furthermore, the classifier could be utilized to improve the classifications in VerbNet and to possibly extend the coverage of VerbNet to new verbs.

2. Related Work

As mentioned in Section 1, there have been supervised approaches to verb sense disambiguation that classify verbs into a VerbNet class (Girju et al., 2005; Abend et al., 2008; Chen and Eugenio, 2010; Brown et al., 2011; Croce et al., 2012). These methods basically train a supervised classifier for each verb or use class membership constraints (Abend et al., 2008), which limit the class candidates of a verb to its seen classes in the training data. Therefore, it is difficult or impossible to deal with rare or unseen verbs when applying these models to new data.

Among them, Chen and Eugenio (2010) tried a single classifier model as well as per-verb classifiers. The single classifier achieved an accuracy of 90.8% and the per-verb classifier achieved 96.7% for polysemous verbs in the sentences in VerbNet. Although they mentioned that their single classifier can handle unseen verbs, they did not propose a method for improving the single classifier to the level of the per-verb classifiers.

3. Resources

3.1. SemLink

The Semlink project (Loper et al., 2007) is aimed at creating a mapping of PropBank (Palmer et al., 2005), FrameNet (Baker et al., 1998), WordNet and VerbNet to one another.[1] This project includes a corpus that annotates each verb token in the Wall Street Journal corpus (of the Penn Treebank) with a VerbNet class, a PropBank frame and a FrameNet frame. We employ the VerbNet class annotations of this corpus. The corpus is split into the standard division of syntactic parsing: sections 02-21 for training (60,450 tokens), section 00 for development (3,167 tokens) and section 23 for testing (3,508 tokens).[2] In this training set, the unique number of verb classes is 233.[3]

3.2. Word Representations

To provide a classifier for verb sense disambiguation with semantic or generalized features of a verb and its modifiers, we use the following three kinds of word representations.

[1]http://verbs.colorado.edu/semlink/

[2]Version 1.2.2c is used in this paper.

[3]Sub-classes are ignored.

Brown These are clusters induced by the Brown clustering algorithm (Brown et al., 1992). A word is represented as a bit string. We use the Brown clusters (the number of clusters: 3,200) created by Turian et al. (2010).[4] This data covers 247,339 words.

SENNA These are the distributed word representations trained via a neural network model (Collobert et al., 2011).[5] A word is represented as a 50-dimensional vector. This data covers 130,000 words.

RNN-{80, 640, 1600} These are the distributed word representations trained via a recurrent neural network language model (Mikolov et al., 2013).[6] A word is represented as 80-, 640- and 1600-dimensional vectors. This data covers 82,390 words.

4. Single Classifier for Verb Sense Disambiguation based on Generalized Features

We propose a single classifier for assigning a VerbNet class to a verb token in a text. The features of this classifier consist of basic features and generalized features. Generalized features are used to give the classifier generalization abilities across verbs.

4.1. Basic Features

We extract a verb and its modifiers from a dependency parse and use them as basic features on the basis of the work of Chen and Palmer (2009). These basic features are utilized in all the models of our experiments. An input sentence is converted to Stanford collapsed dependencies (de Marneffe et al., 2006)[7] and the following features are extracted from these dependencies:

- lemma and part-of-speech tag of the target verb
- lemma and part-of-speech tag of each word that depends on the verb, as distinguished by the dependency relation

For instance, from the following sentence, the features listed in Table 1 are extracted.

(1) Children may then observe birds at the feeder.

verb.lemma:observe, verb.pos:VB,
nsubj.lemma:children, nsubj.pos:NNP,
aux.lemma:may, aux.pos:MD,
advmod.lemma:then, advmod.pos:RB,
dobj.lemma:bird, dobj.pos:NNS,
prep_at.lemma:feeder, prep_at.pos:NN

Table 1: Basic features extracted from the sentence "Children may then observe birds at the feeder." *.lemma means a lemma and *.pos means a part-of-speech tag.

4.2. Generalized Features

For generalized features, we use each of the three types of word representations described in section 3.2. The following features are calculated for a verb and its direct object (if it exists) and used with the basic features.[8]

- For the word representations based on neural network models (SENNA and RNN-*), we first apply K-means clustering to each of the word representations (K = 100, 320, 1000, 3200, 10000).[9] Then, we use the cluster numbers of all five settings as features.
- A Brown cluster is represented as a bit string (e.g., the word "bird" belongs to `10111110010`). Following previous work (Turian et al., 2010), we use the first 4, 6, 10 and 20 bits as features.

5. Experiments and Discussions

5.1. Experimental Settings and Results

In the experiments, we use the SemLink corpus with the split described in Section 3.1. The basic features are extracted from gold-standard parses to examine the pure effects of generalized features.

We adopt Opal (Yoshinaga and Kitsuregawa, 2010)[10] as a machine learning implementation.[11] This tool enables online learning using a polynomial kernel. As the parameters of Opal, we used the passive-aggressive algorithm (PA-I) with the polynomial kernel of degree 2 as a learner and the extension to multi-class classification (Matsushima et al., 2010), and set the aggressiveness parameter C to 0.001, which achieved the best performance on the development set. Other parameters are set to the default values of Opal. The number of classes is 233, which is the number of unique VerbNet classes that appear in the training set.

We measure accuracy of classifications, which is calculated by the proportion of the number of correct classifications to the number of all verb tokens. Table 2 lists the accuracy of a baseline method based only on basic features (DEP) and the proposed methods based on three kinds of word representations (DEP+Brown, DEP+SENNA, DEP+RNN-{80, 640, 1600}). This table lists not only the overall accuracy but also the accuracy only for polysemous verbs, which have more than one class in the tokens of the training set. As a result, DEP+RNN-1600 outperformed the baseline method and also the other models based on generalized features. Also, while increasing the dimension of word representation vectors, the accuracy was slightly improved. Figure 1 shows the cumulative accuracy for infrequent verbs. From this figure, we can see that the accuracy of the verbs that

[4]http://metaoptimize.com/projects/wordreprs/

[5]http://ronan.collobert.com/senna/

[6]http://www.fit.vutbr.cz/~imikolov/rnnlm/

[7]http://nlp.stanford.edu/software/lex-parser.shtml

[8]Due to space limitation, ablation studies are not included in this abstract, but just using a verb and its direct object achieved the best performance on the development set.

[9]The original word vectors can be used as features, but the training with these vectors are very slow and the performance is slightly lower than that of the K-means features in our preliminary experiments.

[10]http://www.tkl.iis.u-tokyo.ac.jp/~ynaga/opal/

[11]Opal achieved a better performance than Support Vector Machines in our preliminary experiments.

	all	poly
DEP	0.9555	0.8965
DEP+Brown	0.9595	0.9076
DEP+SENNA	0.9618	0.9076
DEP+RNN-80	0.9629	0.9104
DEP+RNN-640	0.9632	0.9076
DEP+RNN-1600	**0.9655**	**0.9141**

Table 2: Classification accuracy for all verb tokens (all) and only polysemous verb tokens (poly).

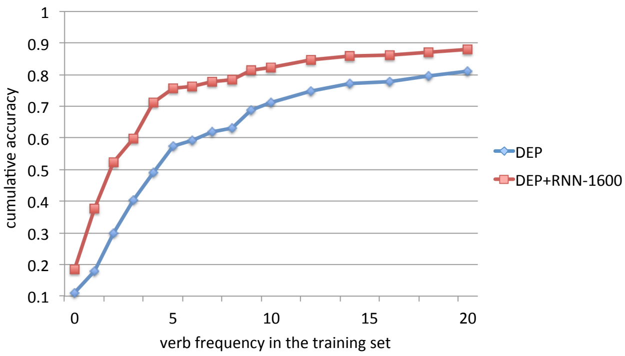

Figure 1: Cumulative accuracy for infrequent verbs.

occur less than 10 times in the training set was improved by 10-20%.

We also compared the performance of the single classifier with per-verb classifiers, which had been used in previous work. A per-verb classifier was trained for each of 594 verbs that appear 10 or more times in the training set. These per-verb classifiers also use the generalized features. However, the generalized features for verbs are meaningless for the per-verb classifiers, and only the generalized features for direct objects would work. For comparison, we evaluated only the tokens of these 594 verb from the test set (3,349 out of 3,508 tokens). Table 3 lists the accuracy of the per-verb classifiers and the single classifiers. The single classifier DEP+RNN-1600 slightly outperformed the per-verb classifiers (though the difference is not significant). Although Chen and Eugenio (2010) reported the accuracy of a single classifier was lower than that of per-verb classifiers, our single classifier achieved equivalent performance to the per-verb classifiers by using generalized features. It is difficult to compare our results with those in previous work due to the use of different data set. However, we anticipate that our results would match or even surpass the results of Croce et al. (2012), which achieved a state-of-the-art accuracy of 93.78% on obsolete SemLink data using per-verb classifiers.

	all	poly
per-verb	0.9684	0.9027
single (DEP)	0.9663	0.8962
single (DEP+RNN-1600)	**0.9716**	**0.9130**

Table 3: Comparison between single classifiers and per-verb classifiers. The column of "all" means the accuracy for all verb tokens and that of "poly" means the accuracy for polysemous verb tokens.

5.2. Discussions

We examined erroneous classifications by DEP+RNN-1600 in the development set. Major errors were still caused by unseen and rare verbs in the training set, such as "disgorge," "resubmit," "ration,“ "snap" and "encircle." To improve the accuracy of these verbs, it is necessary to consider more generalized or semantic features, such as dynamic dependency neighbors (Dligach and Palmer, 2008) and kernel-based structural similarity (Croce et al., 2012).

Some polysemous verbs achieved a low classification accuracy. For example, "carry" has three VerbNet classes, i.e., carry-11.4, cost-54.2 and fit-54.3, but several carry-11.4 tokens were misclassified into fit-54.3. One of such tokens is found in the following sentence:

(2) ... who has been willing to let Mr. Markey carry the legislation in recent months.

The major sense of carry-11.4 is to carry some concrete object somewhere, but it has also the sense of legislation or voting. This case may suggest dividing carry-11.4 into more fine-grained classes.

Another example is "fail," which has two VerbNet classes, i.e., succeed-74 and neglect-75. The following token is labeled as neglect-75 in the SemLink corpus, but it was misclassified into succeed-74.

(3) ... the government said that orders for manufactured goods and spending on construction failed to rise in September.

Since the sense difference between succeed-74 and neglect-75 is subtle, it was very difficult to distinguish it with our classifier. Actually, these two classes are not distinguished in the OntoNotes sense groupings.[12]

6. Conclusion

This paper described a method for verb sense disambiguation based on VerbNet and SemLink. This method consists of a single classifier that can handle any rare or unseen verbs by exploiting generalized features. Our experimental results show that the proposed method achieves equivalent performance to per-verb classifiers, which cannot be applied to unseen verbs.

As discussed in the previous section, errors of our classifier may suggest modifications of the classifications in VerbNet. If these suggestions were efficiently performed, this would help lexicographers find clues for modification. Furthermore, it is possible to extend the coverage of VerbNet by applying the classifier to a raw corpus and extracting out-of-vocabulary verbs with plausible VerbNet classes in a semi-automatic manner.

[12] http://verbs.colorado.edu/html_groupings/fail-v.html

7. Acknowledgments

This work was supported by Kyoto University John Mung Program and JSPS KAKENHI Grant Number 25540140. We also gratefully acknowledge the support of the National Science Foundation Grant NSF-IIS-1116782, A Bayesian Approach to Dynamic Lexical Resources for Flexible Language Processing. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We also appreciate the fruitful discussions with Dr. Jinying Chen.

8. References

Omri Abend, Roi Reichart, and Ari Rappoport. 2008. A supervised algorithm for verb disambiguation into VerbNet classes. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING2008)*, pages 9–16.

Collin Baker, Charles J. Fillmore, and John 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 (COLING-ACL98)*, pages 86–90.

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.

Susan Windisch Brown, Dmitriy Dligach, and Martha Palmer. 2011. VerbNet class assignment as a WSD task. In *Proceedings of the 9th International Conference on Computational Semantics (IWCS2011)*.

Lin Chen and Barbara Di Eugenio. 2010. A maximum entropy approach to disambiguating VerbNet classes. In *Proceedings of the 2nd Interdisciplinary Workshop on Verbs, The Identification and Representation of Verb Features*.

Jinying Chen and Martha Palmer. 2009. Improving English verb sense disambiguation performance with linguistically motivated features and clear sense distinction boundaries. *Language Resources and Evaluation*, 43(2):181–208.

Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12:2493–2537.

Danilo Croce, Alessandro Moschitti, Roberto Basili, and Martha Palmer. 2012. Verb classification using distributional similarity in syntactic and semantic structures. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL2012)*, pages 263–272.

Hoa Trang Dang. 2004. *Investigations into the role of lexical semantics in word sense disambiguation*. Ph.D. thesis, University of Pennsylvania.

Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC2006)*, pages 449–454.

Dmitriy Dligach and Martha Palmer. 2008. Novel semantic features for verb sense disambiguation. In *Proceedings of ACL-08: HLT, Short Papers*, pages 29–32.

Roxana Girju, Dan Roth, and Mark Sammons. 2005. Token-level disambiguation of VerbNet classes. In *Proceedings of the Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes*.

Karin Kipper-Schuler. 2005. *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. Ph.D. thesis, University of Pennsylvania.

Edward Loper, Szu-Ting Yi, and Martha Palmer. 2007. Combining lexical resources: mapping between PropBank and VerbNet. In *Proceedings of the 7th International Workshop on Computational Linguistics*.

Shin Matsushima, Nobuyuki Shimizu, Kazuhiro Yoshida, Takashi Ninomiya, and Hiroshi Nakagawa. 2010. Exact passive-aggressive algorithm for multiclass classification using support class. In *Proceedings of 2010 SIAM International Conference on Data Mining (SDM2010)*, pages 303–314.

Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. 2013. Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT2013)*, pages 746–751.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Lei Shi and Rada Mihalcea. 2005. Putting pieces together: Combining FrameNet, VerbNet and WordNet for robust semantic parsing. In *Computational Linguistics and Intelligent Text Processing*, pages 100–111. Springer.

Rajen Subba and Barbara Di Eugenio. 2009. An effective discourse parser that uses rich linguistic information. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT2009)*, pages 566–574.

Robert Swier and Suzanne Stevenson. 2005. Exploiting a verb lexicon in automatic semantic role labelling. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT-EMNLP2005)*, pages 883–890.

Joseph Turian, Lev-Arie Ratinov, and Yoshua Bengio. 2010. Word representations: A simple and general method for semi-supervised learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL2010)*, pages 384–394.

Naoki Yoshinaga and Masaru Kitsuregawa. 2010. Kernel slicing: Scalable online training with conjunctive features. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING2010)*, pages 1245–1253.