

1 Related Works

Memory based learning has been shown to be an effective choice for word sense disambiguation tasks. (Mooney, 1996) Veenstra et al. (2000) argues that, because memory-based learning does not require linguistic knowledge but only training examples, it is well suited for domains where large numbers of features and sparseness of data interact to make life difficult for many other machine-learning methods.¹ In a comparative study of various supervised learning algorithms, an unoptimized version of the k-nearest neighbors algorithm achieved results competitive with other, more computationally complex algorithms. (Mooney, 1996)²

On the other hand, memory based algorithms come with significant storage costs. To address this, the IB1, IB2, and IB3 algorithms were developed extend the k-nearest neighbor algorithm to reduce storage requirements, with a small or even zero loss of classification accuracy. (Aha et al., 1991)

The Tilburg Memory-Based Learner (TiMBL) provides an open-source implementation of IB1 and IB2 learning algorithms, as well as IGTREE, a decision tree based optimization and TRIBL and TRIBL2, hybrids of IB1 and IGTREE and IB2 and IGTREE respectively. (Daelemans et al., 2004) In addition to the algorithms, TiMBL allows for adjusting various parameters, including but not limited to distance metric, feature weighting, number of neighbors, and tree-order.³

As it has been demonstrated that both parameter optimization (Daelemans and Hoste, 2002) and feature optimization (Daelemans et al., 2003) can result in a higher variation in results than differences between algorithms, it is therefore necessary to optimize features and parameters for each task. For this reason, we evaluate the performance of TiMBL with various features sets, and optimize over various parameters using

¹This generalization is particularly true of the SENSEVAL-3 Word Sense Disambiguation Task, which involves disambiguating word-senses using a training set containing as few as 97 and at most 257 examples.

²As a memory based method, the k-nearest neighbor algorithm has a training time of 0, compared to the non-linear training times of Naive-Bayes, Perceptron and other types of classifiers. Additionally, the testing time is linear, which makes this algorithm very desirable with respect to time complexity. (Mooney, 1996)

³Tree-order is, of course, only a parameter for the IGTREE, TRIBL, and TRIBL2 algorithms.

brute-force method.⁴

Although initial results (Mihalcea et al., 2004) suggested that k-nearest neighbor algorithms underperformed others for SENSEVAL-3's Romanian Word Sense Disambiguation task, Dinu and Kübler (2007) demonstrated that a narrower selection of features provided greater accuracy in WSD than the methods used during the SENSEVAL-3 competition.

It has additionally been shown that TiMBL, with optimized features and parameters, can be used to extend the training corpus with unlabelled data. (Kübler and Zhekova, 2009) This semi-supervised approach is particularly useful for WSD, due to the relative scarcity and questionable quality of human annotated sense classification. Le et al. (2006) outlines the problems for using semi-supervised approaches to WSD, and presents solution and a bootstrapping algorithm which informs our approach.

References

- David W Aha, Dennis Kibler, and Marc K Albert. 1991. Instance-based learning algorithms. *Machine learning*, 6(1):37–66.
- Walter Daelemans and Véronique Hoste. 2002. Evaluation of machine learning methods for natural language processing tasks. In *3rd International conference on Language Resources and Evaluation (LREC 2002)*. European Language Resources Association (ELRA).
- Walter Daelemans, Véronique Hoste, Fien De Meulder, and Bart Naudts. 2003. Combined optimization of feature selection and algorithm parameters in machine learning of language. In *ECML*, pages 84–95. Springer.
- Walter Daelemans, Jakub Zavrel, Kurt van der Sloot, and Antal Van den Bosch. 2004. Timbl: Tilburg memory-based learner. *Tilburg University*.
- Georgiana Dinu and Sandra Kübler. 2007. Sometimes less is more: Romanian word sense disambiguation revisited. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing, RANLP*.
- Sandra Kübler and Desislava Zhekova. 2009. Semi-supervised learning for word sense disambiguation: Quality vs. quantity. In *RANLP*, pages 197–202.
- Anh-Cuong Le, Akira Shimazu, and Le-Minh Nguyen. 2006. Investigating problems of semi-supervised learning for word sense disambiguation.

⁴Full details are discussed in the methods section.

tion. In *Computer Processing of Oriental Languages. Beyond the Orient: The Research Challenges Ahead*, pages 482–489. Springer.

Rada Mihalcea, Vivi Nastase, Timothy Anatolievich Chklovski, Doina Tatar, Dan Tufis, and Florentina T Hristea. 2004. An evaluation exercise for romanian word sense disambiguation.

Raymond J Mooney. 1996. Comparative experiments on disambiguating word senses: An illustration of the role of bias in machine learning. *arXiv preprint cmp-lg/9612001*.

Jorn Veenstra, Antal Van den Bosch, Sabine Buchholz, Walter Daelemans, et al. 2000. Memory-based word sense disambiguation. *Computers and the Humanities*, 34(1-2):171–177.