

Word Sense Disambiguation: The Case for Combinations of Knowledge Sources

Mark Stevenson

(University of Sheffield)

Stanford, CA: CSLI Publications (CSLI studies in computational linguistics, edited by Ann Copestake) (distributed by the University of Chicago Press), 2003, xvi+175 pp; hardbound, ISBN 1-57586-389-8, \$67.50; paperbound, ISBN 1-57586-390-1, \$25.00

Reviewed by

Susan McRoy

University of Wisconsin–Milwaukee

In this book, Stevenson describes his work on applying and evaluating empirical methods for word sense disambiguation (WSD) in large texts. His approach combines several individually weak knowledge sources using a memory-based machine-learning algorithm. It contrasts with earlier methods that relied on hand-crafted rules to combine information from multiple knowledge sources, such as that of McRoy (1992), and with previous empirical approaches, such as that of Yarowsky (1992), that used small lexical samples. Previous rule-based approaches used many of the same knowledge sources and disambiguated about the same percentage of words but lacked a mechanism for evaluating disambiguation accuracy over large test sets. Large-scale empirical approaches to word sense disambiguation have become possible because of the availability of tagged text with categories from WordNet and a mapping, created by Knight and Luk (1994), from WordNet categories to senses from the *Longman Dictionary of Contemporary English (LDOCE)*.

According to the author, this book is based largely on his Ph.D. dissertation, with some extensions to address comments that were made by readers of the thesis and also some discussion of recent work. The book follows the traditional structure of a thesis: a discussion of the foundations of word sense disambiguation and a survey of earlier work by others, followed by a detailed presentation of the author's approach to the problem, his implementation of the approach, and its evaluation. It adds a foreword written by Yorick Wilks that succinctly describes the main contribution of the work and the methods that were used.

Stevenson is concerned with the problem of how to discriminate multiple meanings or senses of the same word. In particular, the book addresses the task of sense tagging, which Stevenson defines as the task of annotating all the words in a document with senses from a given lexicon. Sense tagging is useful for machine translation and may also be of some (but smaller) benefit to work on information retrieval. Stevenson distinguishes sense tagging from general semantic disambiguation and semantic tagging, which do not constrain the types of semantic annotations used, and also from sense disambiguation, which does not require that all words be disambiguated. (Alternatives to lexicon senses for annotation might be semantic features such as HUMAN or ANIMATE.) The author notes that because sense tagging is the most constrained word sense disambiguation task, "a solution to the sense tagging problem would also be a

solution to the other, more general, WSD tasks." Also, sense tagging could potentially replace other discriminators for ambiguities of pronunciation or part of speech.

The weak sources of knowledge that Stevenson's learning algorithm combines include selectional preferences, thesaurus or topic classes, and distributional information gleaned from a corpus. He quantitatively validates the importance of combining multiple sources by showing that while individually these sources could only provide accuracy at the 60% level, properly combined they could achieve accuracy as high as 92% (page 88). The 92% figure is based on disambiguating words only to the homograph level and counting all words. His algorithm achieves 70–90% accuracy when we consider only content words such as adverbs, adjectives, nouns, and verbs and disambiguation down to the level of individual senses.

The first half of the book provides an introduction to the task and concepts of word sense disambiguation, suitable for anyone who is beginning work in this area. For example, the first chapter characterizes different forms of word sense disambiguation, including sense tagging, which, as mentioned, is the focus of the book. The second chapter describes some earlier approaches to word sense disambiguation and the SENSEVAL evaluation framework. The third chapter describes the content and organization of lexical resources commonly used for WSD, including *LDOCE*, *Roget's Thesaurus*, and WordNet. Chapter 4 presents a general characterization of the application knowledge sources for WSD as either filters or partial taggers and discusses the information from *LDOCE* and WordNet that would be most useful for such algorithms. The framework itself seems to cover most, if not all, prior work. Thus, the author's goal seems to be to provide a way of unifying prior approaches, instead of choosing among them.

The remainder of the book describes Stevenson's approach to sense tagging and his evaluation of the results. Chapter 5 considers part-of-speech tagging using syntactic categories from *LDOCE* and the contribution that it can make to sense tagging. In Stevenson's experiments, 94% of words could be assigned the correct *LDOCE* sense (to the homograph level), using only information from the Brill tagging and a mapping from its tags to those of *LDOCE*. This result would seem to contradict Stevenson's claim that a combination of sources is needed to reach this same level of accuracy, although he did find that none of the other knowledge sources (simulated annealing, selectional preferences, or the broad context, as identified by subject codes) provides this accuracy individually.

Chapter 6 (which is about 25 pages long) discusses Stevenson's implementation of a sense tagger using senses from *LDOCE* and the architecture based on combining the results of several filters and partial taggers. Neither algorithms for computing weak knowledge sources nor algorithms for combining weak knowledge sources are entirely new; however, they are explained succinctly here, along with some adaptations that were made. What is novel is the application of a machine-learning algorithm to combine the outputs of its filters and taggers. The particular algorithm is Daelemans et al.'s (1999) TiMBL memory-based learning system, which is explained in some detail.

Chapters 7 and 8 discuss the evaluation of word sense disambiguation algorithms. Chapter 7 describes several previous evaluation strategies, as well as providing important background on the mapping between WordNet senses and the senses of *LDOCE*, and also Stevenson's mapping of an evaluation corpus that was annotated with WordNet senses to a corpus that is tagged with *LDOCE* senses. Chapter 8 describes Stevenson's experiments to evaluate his own system.

The book would be most appropriate to students, researchers, or practitioners who are learning about word sense disambiguation for the first time or to anyone who has not considered word sense disambiguation since the late 1980s, when nonempirical

approaches were still common. The introduction is very basic and would be readily understandable by a student just taking her first course in computational linguistics. The background is similarly intended for those unfamiliar with word sense disambiguation, although the level of detail in some of the discussions is somewhat uneven. A key issue for Stevenson is how systems acquire disambiguation information, such as whether they require tagged or untagged corpora for training and whether they make use of a machine-readable dictionary. There is a discussion of the 1998 and 2001 SENSEVAL competitions (Kilgarriff and Palmer 2000; Preiss and Yarowsky 2001), which standardize the comparison of different approaches to word sense disambiguation.

The presentation of the book would have benefited from a bit more careful editing. The book contains numerous minor punctuation and grammatical errors and errors in the bibliographic references. The background section is uneven in its level of detail and its citations to related work. There are detailed discussions of Wilks's preference semantics, Cowie, Guthrie, and Guthrie's (1992) work using simulated annealing,¹ and Yarowsky's algorithm for tagging words with categories from *Roget's Thesaurus*. Readers would also likely have appreciated having a more comprehensive index, including, for example, entries for authors whose work has been cited.

References

- Cowie, Jim, Joe Guthrie, and Louise Guthrie. 1992. Lexical disambiguation using simulated annealing. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92)*, Nantes, France, pages 359–365.
- Daelemans, Walter, Jakub Zavrel, Ko van der Sloot, and Antal van den Bosch. 1999. TiMBL: Tilburg Memory Based Learner, version 2.0, Reference Guide. Technical Report ILK 99-01, Induction of Linguistic Knowledge Group, Tilburg University, The Netherlands.
- Kilgarriff, Adam and Martha Palmer, editors. 2000. Special issue on SENSEVAL. *Computers and the Humanities*, 34(1–2).
- Knight, Kevin and Steve Luk. 1994. Building a large knowledge base for machine translation. In *Proceedings of the 12th National Conference of the American Association for Artificial Intelligence (AAAI-94)*, Seattle, WA, pages 773–778.
- McRoy, Susan. 1992. Using multiple knowledge sources for word sense discrimination. *Computational Linguistics*, 18(1):1–30.
- Preiss, Judita and David Yarowsky, editors. 2001. *Proceedings of the SENSEVAL-2 Workshop*, Toulouse, France.
- Selman, Bart and Graeme Hirst. 1985. A rule-based connectionist parsing system. In *Proceedings of the Seventh Annual Conference of the Cognitive Science Society*, Irvine, CA, pages 212–221.
- Selman, Bart and Graeme Hirst. 1987. Parsing as an energy minimization problem. In Lawrence Davis, editor, *Genetic Algorithms and Simulated Annealing (Research Notes in Artificial Intelligence)*, Pitman, pages 141–154.
- Yarowsky, David. 1992. Word-sense disambiguation using statistical models of Roget's categories trained on large corpora. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92)*, Nantes, France, pages 454–460.

Susan McRoy's research spans a number of areas, including designing and implementing systems for tutoring and training; achieving robust human-machine communication; detecting and repairing communication errors; developing tools for real-time intelligent dialog systems; and investigating interactions among speech, gaze, and gesture during communications among people. McRoy's address is Department of Electrical Engineering and Computer Science, P.O. Box 784, University of Wisconsin-Milwaukee, Milwaukee, WI 53201; e-mail: mcroy@uwm.edu.

¹ Readers interested in the history of the use of simulated annealing in NLP will also want to consider the work by Selman and Hirst (1985, 1987).