Appears in *Workshop Notes for the Workshop on Learning Language in Logic,* Bled, Slovenia, pp. 7-14, June 1999
Also appears in *Learning Language in Logic,* J. Cussens and S. Dzeroski (Eds.), pp. 57-66, Springer Verlag, Berli

Learning for Semantic Interpretation: Scaling Up Without Dumbing Down

Raymond J. Mooney

Department of Computer Sciences
University of Texas
Austin, TX 78712-1188, USA
mooney@cs.utexas.edu
http://www.cs.utexas.edu/users/mooney

Abstract. Most recent research in learning approaches to natural language have studied fairly "low-level" tasks such as morphology, part-of-speech tagging, and syntactic parsing. However, I believe that logical approaches may have the most relevance and impact at the level of semantic interpretation, where a logical representation of sentence meaning is important and useful. We have explored the use of inductive logic programming for learning parsers that map natural-language database queries into executable logical form. This work goes against the growing trend in computational linguistics of focusing on shallow but broad-coverage natural language tasks ("scaling up by dumbing down") and instead concerns using logic-based learning to develop narrower, domain-specific systems that perform relatively deep processing. I first present a historical view of the shifting emphasis of research on various tasks in natural language processing and then briefly review our own work on learning for semantic interpretation. I will then attempt to encourage others to study such problems and explain why I believe logical approaches have the most to offer at the level of producing semantic interpretations of complete sentences.

1 Introduction

The application of machine learning techniques to natural language processing (NLP) has increased dramatically in recent years under the name of "corpus-based," "statistical," or "empirical" methods. There has been a dramatic shift in computational linguistics from manually constructing grammars and knowledge bases to partially or totally automating this process by using statistical learning methods trained on large annotated or unannotated natural language corpora.

The success of statistical methods in speech recognition (Stolcke, 1997; Jelinek, 1998) has been particularly influential in motivating the application of similar methods to other aspects of natural language processing. There is now a variety of work on applying learning methods to almost all other aspects of language processing as well (Charniak, 1993; Brill & Mooney, 1997; Manning & Schütze, 1999), including syntactic analysis (Charniak, 1997), semantic disambiguation and interpretation (Ng & Zelle, 1997), discourse processing and information extraction (Cardie, 1997), and machine translation (Knight, 1997). Some concrete publication statistics clearly illustrate the extent of the revolution in natural language research. According to data recently collected by Hirschberg (1998), a full 63.5% of the papers in the Proceedings of the Annual Meeting of the Association for Computational Linguistics and 47.4% of the papers in the journal *Computational Linguistics* concerned corpus-based research in 1997. For comparison, 1983 was the last year in which there were no such papers and the percentages in 1990 were still only 12.8% and 15.4%.

Nevertheless, traditional machine learning research in artificial intelligence, particularly logic-based learning, has had limited influence on recent research in computational linguistics. Most current learning research in NLP employs statistical techniques inspired by research in speech recognition, such as hidden Markov models (HMMs) and probabilistic context-free grammars (PCFGs). There has been some recent research on logic-based language learning (Mooney & Califf, 1995; Cohen, 1996; Freitag, 1998), in particular, a recent body of European *inductive logic programming* (ILP) research on language (Cussens, 1997; Manandhar, Džeroski, & Erjavec, 1998; Kazakov & Manandhar, 1998; Eineborg & Lindberg, 1998; Lindberg & Eineborg, 1998; Cussens, Džeroski, & Erjavec, 1999; Lindberg & Eineborg, 1999). However, most of this research has focused on relatively "low level" tasks such as morphological analysis and part-of-speech tagging and has not conclusively demonstrated superior performance when compared to competing statistical methods for these tasks.

In contrast, most of our own recent research on applying ILP to NLP has focused on learning to parse natural-language database queries into a semantic logical form that produces an answer when executed in Prolog (Zelle & Mooney, 1993, 1994, 1996; Zelle, 1995; Mooney, 1997; Thompson & Mooney, 1999; Thompson, 1998; Thompson, Califf, & Mooney, 1999). There is a long tradition of representing the meaning of natural language statements and queries in first-order logic (Allen, 1995; Dowty, Wall, & Peters, 1981; Woods, 1978). However, we know of no other recent research specifically on learning to map language into logical form. Nevertheless, we believe this is the most suitable NLP task for ILP, since the desired output is a logical representation that is best processed using logic-based methods.

This paper first presents a brief historical view of the shifting emphasis of research on various tasks in natural language processing. Next, it briefly reviews our own work on learning for semantic interpretation. Finally, it summarizes the arguments in favor of semantic interpretation as the most promising natural-language application of logic-based learning.

2 A Brief Historical Review of NLP Research

From the very early days of NLP research, answering natural-language questions in a particular domain was a key task (Green, Wolf, Chomsky, & Laughery, 1963; Simmons, 1965, 1970). Although syntactic analysis was a major component of this task, the production of a semantic interpretation that could be used to retrieve answers was also very important. The semantic analysis of language was a particular focus of NLP research in the 1970's, with researchers exploring tasks ranging from responding to commands and answering questions in a micro-world (Winograd, 1972) to answering database queries (Woods, 1977; Waltz, 1978; Hendrix, Sacerdoti, Sagalowicz, & Slocum, 1978) and understanding short stories (Charniak, 1972; Schank, 1975; Charniak & Wilks, 1976; Schank & Abelson, 1977; Schank & Riesbeck, 1981).

Research in this era attempted to address complex issues in semantic interpretation, knowledge representation, and inference. The systems that were developed could perform interesting semantic interpretation and inference when understanding particular sentences or stories; however, they tended to require tedious amounts of application-specific knowledge-engineering and were therefore quite brittle and not easily extended to new texts or new applications and domains. The result was systems that could perform fairly in-depth understanding of narrative text; but were restricted to comprehending three or four specific stories (Dyer, 1983).

Disenchantment with the knowledge-engineering requirements and brittleness of such systems grew, and research on in-depth semantic interpretation began to wane in the early to mid 1980's. The author's own thesis research in the mid 1980's focused on attempting to relieve the knowledge-engineering bottleneck by using *explanation-based learning* (EBL) to automatically acquire the larger knowledge structures (*scripts* or *schemas*) needed for narrative understanding (DeJong, 1981; Mooney & DeJong, 1985; DeJong & Mooney, 1986). However, this approach still required a large amount of existing knowledge that could be used to construct detailed explanations for simpler stories.

In order to avoid the difficult problems of detailed semantic analysis, NLP research began to focus on building robust systems for simpler tasks. With the advent of statistical learning methods that could successfully acquire knowledge from large corpora for more tractable problems such as speech recognition, part-of-speech tagging, and syntactic parsing, significant progress has been made on these tasks over the past decade (Jelinek, 1998; Manning & Schütze, 1999). Also, much current NLP research is driven by applications to arbitrary documents on the Internet and World Wide Web (Mahesh, 1997), and therefore cannot exploit domain-specific knowledge. Consequently, much current NLP research has more the flavor of traditional *information retrieval* (Sparck Jones & Willett, 1997), rather than AI research on language *understanding*. This overall trend is succinctly captured by the recently coined clever phrase "scaling up by dumbing down."

Unfortunately, there is relatively little research on using learning methods to acquire knowledge for detailed semantic interpretation. Research on corpus-based word-sense disambiguation addresses semantic issues (Ng & Zelle, 1997; Ide & Véronis, 1998); however, only at the level of interpreting individual words rather than constructing representations for complete sentences. Research on learning for information extraction also touches on semantic interpretation; however, existing methods learn fairly low-level syntactic patterns for extracting specific target phrases (Cardie, 1997; Freitag, 1998; Bikel, Schwartz, & Weischedel, 1999; Soderland, 1999; Califf & Mooney, 1999). Nevertheless, there has been a limited amount of research on learning to interpret complete sentences for answering database queries (Zelle & Mooney, 1996; Miller, Stallard, Bobrow, & Schwartz, 1996; Kuhn & De Mori, 1995).

3 CHILL: ILP for Semantic Interpretation

Our own research on learning for semantic interpretation has involved the development of a system called CHILL (Zelle, 1995) which uses ILP to learn a deterministic shift-reduce parser written in Prolog. The input to CHILL is a corpus of sentences paired with semantic representations. The parser learned from this data is able to transform these training sentences into their correct representations, as well as generalizing to correctly interpret many novel sentences.

CHILL is currently able to handle two kinds of semantic representations: a case-role form based on *conceptual dependency* (Schank, 1975) and a Prolog-based logical query language. As examples of the latter, consider two sample queries for a database on U.S. geography, paired with their corresponding logical form:

What is the capital of the state with the highest population?

```
answer(C, (capital(S,C), largest(P, (state(S),
                                     population(S,P))))).
```

What state is Texarkana located in?

```
answer(S, (state(S), eq(C,cityid(texarkana,_)), loc(C,S))).
```

CHILL treats parser induction as a problem of learning rules to control the actions of a shift-reduce parser. During parsing, the current context is maintained in a stack of previously interpreted constituents and a buffer containing the remaining input. When parsing is complete, the buffer is empty and the stack contains the final representation of the input. There are three types of operators used to construct logical queries. First is the introduction onto the stack of a predicate needed in the sentence representation due to the appearance of a word or phrase at the front of the input buffer. A second type of operator unifies two variables appearing in the current items in the stack. Finally, a stack item may be embedded as an argument of another stack item.

A generic parsing shell is provided to the system, and the initial parsing operators are produced through an automated analysis of the training data using general templates for each of the operator types described above. During learning, these initial overly-general operators are specialized so that the resulting parser deterministically produces only the correct semantic interpretation of each the training examples. The introduction operators require a semantic lexicon as background knowledge that provides the possible logical representations of specific words and phrases. CHILL initially required the user to provide this lexicon; however, we have recently developed a system called WOLFIE that learns this lexicon automatically from the same training corpus (Thompson & Mooney, 1999; Thompson, 1998).

CHILL has been used successfully to learn natural-language interfaces for three separate databases: 1) a small database on U.S. geography, 2) a database of thousands of restaurants in northern California, and 3) a database of computer jobs automatically extracted from the Usenet newsgroup `austin.jobs` (Califf & Mooney, 1999). After training on corpora of a few hundred queries, the system learns parsers that are reasonably accurate at interpreting novel queries for each of these applications. For the geography domain, the system has learned semantic parsers for Spanish, Japanese, and Turkish, as well as English. Below are some of the interesting novel English queries that the geography system can answer although it was never explicitly trained on queries of this complexity:

- What states border states through which the Mississippi runs?
- What states border states that border Texas?
- What states border states that border states that border states that border Texas?
- What states border the state that borders the most states?
- What rivers flow through states that border the state with the largest population?
- What is the largest state through which the Mississippi runs?
- What is the longest river that flows through a state that borders Indiana?
- What is the length of the river that flows through the most states?

CHILL is described in a bit more detail in the article in this volume by Thompson and Califf (2000), which focuses on the automatic selection of good training sentences.

4 Semantic Interpretation and Learning in Logic

Our research on CHILL demonstrates that ILP can help automate the construction of useful systems for semantic interpretation of natural language. Since the desired output of semantic interpretation is a logical form, ILP methods are particularly well suited for manipulating and constructing such representations. In addition, ILP methods allow for the easy specification and exploitation of background knowledge that is useful in parsing and disambiguation. The current version of CHILL makes significant use of semantic typing knowledge and background predicates for finding items in the stack and buffer that satisfy particular constraints.

There has been significant work on using statistical methods for performing tasks such as part-of-speech tagging and syntactic parsing, and recent results demonstrate fairly impressive performance on these tasks. Logic based methods have currently been unable to demonstrate superior performance on these tasks, and due to the limited context that apparently usually suffices for these problems, are unlikely to easily overtake statistical methods. However, there has been very little research demonstrating successful application of statistical methods to semantic interpretation. Consequently, this problem presents a promising opportunity for demonstrating the advantages of logic-based methods.

Although corpora consisting of tens of thousands of annotated sentences exist for tasks such as part-of-speech tagging and syntactic parsing (Marcus, Santorini, & Marcinkiewicz, 1993), very little data exists for semantic analysis. [1] Consequently, the identification of important and representative tasks and the construction of significant corpora of semantically interpreted sentences are leading requirements for furthering research in this area. Although developing a good semantic representation and annotating sentences with logical form can be a fairly time-consuming and difficult job, a dedicated effort similar to that already undertaken to produce large treebanks for syntactic analysis could produce very sizable and useful semantic corpora.

Part of the resistance to exploring semantic analysis is that, given the current state of the art, it almost inevitably leads to domain dependence. However, many useful and important applications require NLP systems that can exploit specific knowledge of the domain to interpret and disambiguate queries, commands, or statements. The goal of developing general learning methods for this task is exactly to reduce the burden of developing such systems, in the same way that machine learning is used to overcome the knowledge-acquisition bottleneck in developing expert systems. It is largely the difficulty of engineering specific applications that has prevented natural-language interface technology from becoming a wide-spread method for improving the user-friendliness of computing systems. Learning technology for automating the development of such domain-specific systems could help overcome this barrier, eventually resulting in the wide-spread use of NL interfaces.

Consequently, I strongly encourage others to consider investigating learning for semantic interpretation using either statistical, logic-based, or other methods. A larger community of researchers investigating this problem is critical for making important and significant progress. In particular, logic-based approaches are the only ones to have demonstrated significant success on this problem to date, and it is the most promising and natural application of ILP to NLP.

Acknowledgements

This research was supported by the National Science Foundation under grant IRI-9704943.

[1] Some of the corpora we have developed are available from `http://www.cs.utexas.edu/users/ml`.

Bibliography

Allen, J. F. (1995). *Natural Language Understanding (2nd Ed.)*. Benjamin/Cummings, Menlo Park, CA.

Bikel, D. M., Schwartz, R., & Weischedel, R. M. (1999). An algorithm that learns what's in a name. *Machine Learning*, *34*, 211–232.

Brill, E., & Mooney, R. J. (1997). An overview of empirical natural language processing. *AI Magazine*, *18*, 13–24.

Califf, M. E., & Mooney, R. J. (1999). Relational learning of pattern-match rules for information extraction. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence*, pp. 328–334 Menlo Park, CA. AAAI Press.

Cardie, C. (1997). Empirical methods in information extraction. *AI Magazine*, *18*, 65–79.

Charniak, E. (1972). Toward a model of children's story comprehension. Tech. rep. TR266, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Charniak, E. (1997). Statistical techniques for natural language parsing. *AI Magazine*, *18*, 33–43.

Charniak, E., & Wilks, Y. (Eds.). (1976). *Computational Semantics*. North-Holland, Amsterdam.

Charniak, E. (1993). *Statistical Language Learning*. MIT Press.

Cohen, W. W. (1996). Learning to classify English text with ILP methods. In De Raedt, L. (Ed.), *Advances in Inductive Logic Programming*, pp. 124–143. IOS Press, Amsterdam.

Cussens, J. (1997). Part-of-speech tagging using Progol. In *Proceedings of the Seventh International Workshop on Inductive Logic Programming*, pp. 93–108 Berlin. Springer.

Cussens, J., Džeroski, S., & Erjavec, T. (1999). Morphosyntactic tagging of Slovene using Progol. In Džeroski, S., & Flach, P. (Eds.), *Proceedings of the Ninth International Workshop on Inductive Logic Programming* Berlin. Springer-Verlag.

DeJong, G. (1981). Generalizations based on explanations. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, pp. 67–70 San Francisco. Morgan Kaufman.

DeJong, G. F., & Mooney, R. J. (1986). Explanation-based learning: An alternative view. *Machine Learning*, *1*, 145–176. Reprinted in *Readings in Machine Learning*, J. W. Shavlik and T. G. Dietterich (eds.), Morgan Kaufman, San Mateo, CA, 1990.

Dowty, D. R., Wall, R. E., & Peters, S. (1981). *Introduction to Montague Semantics*. D. Reidel, Dordrecht, Holland.

Dyer, M. (1983). *In Depth Understanding*. MIT Press, Cambridge, MA.

Eineborg, M., & Lindberg, N. (1998). Induction of constraint grammar-rules using Progol. In *Proceedings of the Eighth International Workshop on Inductive Logic Programming*, pp. 116–124 Berlin. Springer.

Freitag, D. (1998). Toward general-purpose learning for information extraction. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and COLING-98*, pp. 404–408 New Bunswick, New Jersey. Association for Computational Linguistics.

Green, B. F., Wolf, A. K., Chomsky, C., & Laughery, K. (1963). Baseball: An automatic question answerer. In Feigenbaum, E. A., & Feldman, J. (Eds.), *Computers and Thought*, pp. 207–216. McGraw Hill, New York. Reprinted in *Readings in Natural Language Processing*, B. Grosz, K. Spark Jones, and B. Lynn Webber (eds.), Morgan Kaufman, Los Altos, CA, 1986.

Hendrix, G. G., Sacerdoti, E., Sagalowicz, D., & Slocum, J. (1978). Developing a natural language interface to complex data. *ACM Transactions on Database Systems*, *3*, 105–147.

Hirschberg, J. (1998). Every time I fire a linguist, my performance goes up, and other myths of the statistical natural language processing revolution. Invited talk, Fifteenth National Conference on Artificial Intelligence (AAAI-98).

Ide, N., & Véronis, J. (1998). Introduction to the special issue on word sense disambiguation: The state of the art. *Computational Linguistics*, *24*, 1–40.

Jelinek, F. (1998). *Statistical Methods for Speech Recognition*. MIT Press, Cambridge, MA.

Kazakov, D., & Manandhar, S. (1998). A hybrid approach to word segmentation. In *Proceedings of the Eighth International Workshop on Inductive Logic Programming*, pp. 125–134 Berlin. Springer.

Knight, K. (1997). Automating knowledge acquisition for machine translation. *AI Magazine*, *18*, 81–96.

Kuhn, R., & De Mori, R. (1995). The application of semantic classification trees to natural language understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17*, 449–460.

Lindberg, N., & Eineborg, M. (1998). Learning constraint grammar-style disambiguation rules using inductive logic programming. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and COLING-98*, pp. 775–779 New Bunswick, New Jersey. Association for Computational Linguistics.

Lindberg, N., & Eineborg, M. (1999). Improving part of speech disambiguation rules by adding linguistic knowledge. In Džeroski, S., & Flach, P. (Eds.), *Proceedings of the Ninth International Workshop on Inductive Logic Programming* Berlin. Springer-Verlag.

Mahesh, K. (Ed.). (1997). *Papers from the AAAI Spring Symposium on Natural Language Processing for the World Wide Web*, Menlo Park, CA. AAAI Press.

Manandhar, S., Džeroski, S., & Erjavec, T. (1998). Learning multilingual morphology with CLOG. In *Proceedings of the Eighth International Workshop on Inductive Logic Programming*, pp. 135–144 Berlin. Springer.

Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing.* MIT Press, Cambridge, MA.

Marcus, M., Santorini, B., & Marcinkiewicz, M. (1993). Building a large annotated corpus of English: The Penn treebank. *Computational Linguistics*, *19*, 313–330.

Miller, S., Stallard, D., Bobrow, R., & Schwartz, R. (1996). A fully statistical approach to natural language interfaces. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, pp. 55–61 New Bunswick, New Jersey. Association for Computational Linguistics.

Mooney, R. J. (1997). Inductive logic programming for natural language processing. In Muggleton, S. (Ed.), *Inductive Logic Programming: Selected papers from the 6th International Workshop*, pp. 3–22. Springer-Verlag, Berlin.

Mooney, R. J., & Califf, M. E. (1995). Induction of first-order decision lists: Results on learning the past tense of English verbs. *Journal of Artificial Intelligence Research*, *3*, 1–24.

Mooney, R. J., & DeJong, G. F. (1985). Learning schemata for natural language processing. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, pp. 681–687 San Francisco. Morgan Kaufman.

Ng, H. T., & Zelle, J. (1997). Corpus-based approaches to semantic interpretation in natural language processing. *AI Magazine*, *18*, 45–64.

Schank, R. C. (1975). *Conceptual Information Processing.* North-Holland, Oxford.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures.* Lawrence Erlbaum and Associates, Hillsdale, NJ.

Schank, R. C., & Riesbeck, C. K. (1981). *Inside Computer Understanding: Five Programs plus Miniatures.* Lawrence Erlbaum and Associates, Hillsdale, NJ.

Simmons, R. F. (1965). Answering English questions by computer: A survey. *Communications of the Association for Computing Machinery*, *8*, 53–70.

Simmons, R. F. (1970). Natural language question-answering systems: 1969. *Communications of the Association for Computing Machinery*, *13*, 15–30.

Soderland, S. (1999). Learning information extraction rules for semi-structured and free text. *Machine Learning*, *34*, 233–272.

Sparck Jones, K., & Willett, P. (Eds.). (1997). *Readings in Information Retrieval.* Morgan Kaufmann, San Francisco, CA.

Stolcke, A. (1997). Linguistic knowledge and empirical methods in speech recognition. *AI Magazine*, *18*, 25–31.

Thompson, C. A., & Califf, M. E. (2000). Improving learning in two natural language tasks: Choosing examples intelligently. In *This volume.*

Thompson, C. A., Califf, M. E., & Mooney, R. J. (1999). Active learning for natural language parsing and information extraction. In *Proceedings of the Sixteenth International Conference on Machine Learning*, pp. 406–414 San Francisco. Morgan Kaufman.

Thompson, C. A., & Mooney, R. J. (1999). Automatic construction of semantic lexicons for learning natural language interfaces. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence*, pp. 487–493.

Thompson, C. A. (1998). *Semantic Lexicon Acquisition for Learning Natural Language Interfaces.* Ph.D. thesis, Department of Computer Sciences, University of Texas, Austin, TX. Also appears as Artificial Intelligence Laboratory Technical Report AI 99-278 (see http://www.cs.utexas.edu/users/ai-lab).

Waltz, D. L. (1978). An English language question answering system for a large relational database. *Communications of the Association for Computing Machinery*, *21*, 526–539.

Winograd, T. (1972). *Understanding Natural Language.* Academic Press, Orlando, FL.

Woods, W. A. (1977). Lunar rocks in natural English: Explorations in natural language question answering. In Zampoli, A. (Ed.), *Linguistic Structures Processing.* Elsevier North-Holland, New York.

Woods, W. A. (1978). Semantics and quantification in natural language question answering. In Yovits, M. (Ed.), *Advances in Computers, vol. 17*, pp. 2–64. Academic Press, New York. Reprinted in *Readings in Natural Language Processing*, B. Grosz, K. Spark Jones, and B. Lynn Webber (eds.), Morgan Kaufman, Los Altos, CA, 1986.

Zelle, J. M. (1995). *Using Inductive Logic Programming to Automate the Construction of Natural Language Parsers.* Ph.D. thesis, Department of Computer Sciences, University of Texas, Austin, TX. Also appears as Artificial Intelligence Laboratory Technical Report AI 96-249.

Zelle, J. M., & Mooney, R. J. (1993). Learning semantic grammars with constructive inductive logic programming. In *Proceedings of the Eleventh National Conference on Artificial Intelligence*, pp. 817–822 Menlo Park, CA. AAAI Press.

Zelle, J. M., & Mooney, R. J. (1994). Inducing deterministic Prolog parsers from treebanks: A machine learning approach. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, pp. 748–753 Menlo Park, CA. AAAI Press.

Zelle, J. M., & Mooney, R. J. (1996). Learning to parse database queries using inductive logic programming. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pp. 1050–1055 Menlo Park, CA. AAAI Press.