Book Reviews

The Balancing Act: Combining Symbolic and Statistical Approaches to Language

Judith L. Klavans and Philip Resnik (editors)
(Columbia University and University of Maryland)

Cambridge, MA: The MIT Press
(Language, speech, and communication
series), 1996, xiv+186 pp; hardbound,
ISBN 0-262-11218-3, $35.00;
paperbound, ISBN 0-262-61122-8, $17.50

Reviewed by
Takenobu Tokunaga
Tokyo Institute of Technology

The Balancing Act is intended as an exploration into research combining rule-based and corpus-based approaches, that is, qualitative and quantitative methods. The book consists of eight chapters, six of which represent papers selected and revised from those originally presented at the workshop of the same name held following the 1994 annual meeting of the Association for Computational Linguistics, with two papers newly added. The papers contained in the book can be classified as follows: Chapters 1 (Steven Abney) and 5 (Shyam Kapur and Robin Clark) are concerned with the statistical modeling of linguistic theories, and in particular Chomskyan theory. Chapters 3 (Béatrice Daille) and 4 (Vasileios Hatzivassiloglou) deal with corpus-based lexical acquisition. Chapters 2 (Hiyan Alshawi) and 8 (Carolyn Penstein Rosé and Alex H. Waibel) discuss the task of balancing the roles of statistics and rules in the context of a working system. Chapter 6 (Patti Price) provides a good survey of the history of the marriage of speech recognition and language processing research. Finally, Chapter 7 (Lance A. Ramshaw and Mitchell P. Marcus) describes a part-of-speech tagger.

The first paper in the book is Abney's, entitled "Statistical methods and linguistics." Abney's main claim is that statistical approaches contribute not only to language processing, but also to the understanding of language itself. He argues that stochastic grammars play an important role in the handling of a number of theoretical linguistic issues. The principal emphasis of the paper is on syntactic issues, with little description given of the incorporation of semantic and pragmatic factors into a statistical system, or of their interaction with syntax.

In contrast to Abney's general discussion of the contributions of statistics to theoretical linguistics, the work presented by Kapur and Clark in Chapter 5 gives a specific example of this type of statistical application. In this chapter, Kapur and Clark introduce the notion of a Parameter-Setting Machine (PSM), and propose a means of using corpus data to learn the parameters required to facilitate Chomskyan theory. Unfortunately, the title of the paper, "The automatic construction of a symbolic parser via statistical techniques," is rather misleading given that the paper deals only with word-order parameters; but despite this, the paper gives an interesting insight into the extent to which such an approach can be applied to learning other parameters.

Chapters 3 and 4 deal with the acquisition of lexical knowledge from corpora.

In Chapter 3, "Study and implementation of combined techniques for automatic extraction of terminology," Daille discusses the automatic extraction of terminology for specific domains. As an example domain, she considers French telecommunications, and attempts to extract compound terms as component word pairs, reasoning that terms composed of three or more words can be reduced to a set of pairwise relations. The proposed extraction process is made up of two phases, namely the linguistic filtering phase and the statistical filtering phase. In the first phase, each candidate term is checked against predefined, linguistically motivated, patterns such as `N adj` or `N1 prep N2`. Collocation statistics are then calculated from the candidate terms derived during the first stage. This statistical filtering phase involves more than ten kinds of statistical measures, including simple frequency and mutual information. The results obtained from the different types of statistical filters indicate that, surprisingly, simple frequency returns the best result. Daille attributes this to the use of linguistic filters prior to statistical calculation.

Chapter 4, "Do we need linguistics when we have statistics? A comparative analysis of the contributions of linguistic cues to a statistical word grouping system," discusses a similar acquisition task based on linguistic knowledge and focuses on grouping adjectives with regard to their patterns of usage. Following his previous work (Hatzivassiloglou and McKeown 1993), Hatzivassiloglou thoroughly investigates the effectiveness of various kinds of linguistic knowledge, and also the cost of acquiring linguistic knowledge. The cost of knowledge acquisition represents a unique viewpoint in this type of research and constitutes yet another important facet of the overall balancing act.

There is no doubt that the success of statistical techniques in speech recognition has had a considerable influence on the language processing community. In Chapter 6, "Combining linguistics with statistical methods in automatic speech understanding," Price discusses the evolution of speech understanding from speech recognition and language processing research. She summarizes the cultural gaps between these research areas (her Table 6.1), and goes on to propose ways in which they can contribute to each other.

Both Chapters 2 and 8 deal with speech translation, but with different emphases. In Chapter 2, "Qualitative and quantitative models of speech translation," Alshawi discusses the question of which components of an overall system statistical approaches should be introduced into. In Chapter 8, "Recovering from parser failures: A hybrid statistical and symbolic approach," Rosé and Waibel propose a method of using parse fragments to recover from a parse error. Their system tries to construct a feature-structure-style intermediate representation from the set of parse fragments, relying on collocational preferences for slot types and fillers.

In Chapter 7, "Exploring the nature of transformation-based learning," Ramshaw and Marcus discuss the advantages of a Brill-like approach (Brill 1992) over pure stochastic (hidden Markov model) and symbolic (decision tree) methods, using both qualitative and quantitative evaluation. They claim that Brill's approach suffers less from over-training than is the case for HMMs or decision trees because the rules extracted in Brill's approach are more loosely interdependent. This is due to HMMs being overly monolithic and the structure of decision trees being too rigidly hierarchical. In their evaluation, Ramshaw and Marcus analyzed the dependencies between learned rules and found that more general rules tend to apply at an early stage, but that more specific rules may override this effect at a later stage if necessary. While the paper leaves much room for further exploration of Brill's approach, particularly of its more theoretical aspects, it does point to this being one promising approach to combining symbolic and statistical techniques.

From the experiences of corpus-based NLP research over the last decade, it is obvious that there is an inherent limitation to knowledge obtained from unannotated text. As a result, corpora have been annotated, and new attempts have been made to extract linguistic knowledge from these more information-rich sources. There has been a widespread belief that describing linguistic knowledge from scratch is expensive, but annotating corpora is not free either. Thus, further exploration of the cost-effectiveness of corpus-based research is required. That is, we are at the stage of being able to discuss where we should invest human labor in order to obtain linguistic knowledge of a certain level of quality.

References

Brill, Eric. 1992. A simple rule-based part of speech tagger. In *Proceedings of the Third Conference on Applied Natural Language Processing, Association for Computational Linguistics*, Trento, Italy, pages 152–155.

Hatzivassiloglou, Vasileios and Kathleen McKeown. 1993. Towards the automatic identification of adjectival scales: Clustering adjectives according to meaning. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, Columbus, Ohio, pages 172–182.

Takenobu Tokunaga is an Associate Professor of the Department of Computer Science, Tokyo Institute of Technology. He received his Ph.D. from the Tokyo Institute of Technology in 1991. His research interests include the application of NLP techniques to information retrieval tasks. Tokunaga's address is: Department of Computer Science, Tokyo Institute of Technology, 2-12-1 Ookayama Meguro, Tokyo 152, Japan; e-mail: take@cs.titech.ac.jp