

Briefly Noted

Opinion Mining and Sentiment Analysis

Bo Pang and Lillian Lee
(Yahoo! Research and Cornell University)

Boston: Now Publishers, 2008, x+137 pp; paperbound, ISBN 978-1-60198-150-9, \$99.00, €99.00, £79.00 [Originally published as *Foundations and Trends in Information Retrieval*, 2(1–2), 2008, 1–135.]

Over the last decade or so there has been growing interest in research on computationally analyzing opinions, feelings, and subjective evaluation in text. This burgeoning body of work, variously called “sentiment analysis,” “opinion mining,” and “subjectivity analysis,” addresses such problems as distinguishing objective from subjective propositions, characterizing positive and negative evaluations, determining the sources of different opinions expressed in a document, and summarizing writers’ judgments over a large corpus of texts. Potential applications include Web mining for consumer and political opinion summarization, business and government intelligence analysis, and improving text analysis applications such as information retrieval, question answering, and text summarization.

In this well-written book, Pang and Lee survey the current state of the art in opinion mining and sentiment analysis, broadly construed, with the goal of fitting this diverse research area into a unified framework. After a brief introduction to the area (Chapter 1) and survey of application areas (Chapter 2), the authors present their view of the central challenges that unify this research area in Chapter 3, largely by contrasting it with “traditional,” “fact-based” text analysis. The book then surveys the full range of extant approaches, dividing them into sentiment classification and extraction (Chapter 4), and opinion summarization (Chapter 5). This survey is quite thorough as regards computational work in the area, though it lacks detailed reference to relevant linguistics research such as in the study of modality (Nuyts 2001; Kärkkäinen 2003), cognitive linguistics (Stein and Wright 1995; Langacker 2002), and appraisal theory (Martin and White 2005). This lacuna is justified, however, by the (perhaps unfortunate)

fact that little computational work to date relates to this literature.¹

A distinctive and valuable feature of the book is the inclusion of material on the relationship between subjective language and its social and economic impact (Chapter 6). This discussion helps to place the technical work in its larger context, pointing towards opportunities and risks in its application in various situations. Also particularly valuable is Chapter 7, on publicly available resources, which includes much useful information about available data sets, relevant competitive evaluations, and tutorials/bibliographies in the area. Although much of this information is likely to become outdated, the authors also maintain a companion Web site² which presumably will feature updates to this resource list.

The book provides a useful resource for application developers as well as for researchers, though some readers might have benefited from a more extensive discussion of real-world applications and how various techniques can be used as components of larger systems.

Overall, this slim and entertaining volume is an excellent and timely survey of an exciting and growing research area within computational linguistics.—*Shlomo Argamon, Illinois Institute of Technology*

References

- Banfield, Ann. 1982. *Unspeakable Sentences: Narration and Representation in the Language of Fiction*. Routledge & Kegan Paul Books.
- Kärkkäinen, Elise. 2003. *Epistemic Stance in English Conversation: A Description of Its Interactional Functions, With a Focus on I Think*. John Benjamins Publishing Company.

1 To be fair, there is a brief discussion in Chapter 1 of Banfield’s work (1982) and Quirk et al.’s notion of *private states* (1985) and their influence on the development of the notion of *subjectivity*, but it would have been nice to see a broader discussion relating the computational state of the art to linguistic theory.

2 www.cs.cornell.edu/home/lllee/opinion-mining-sentiment-analysis-survey.html.

- Langacker, Ronald W. 2002. Deixis and subjectivity. *Grounding: The Epistemic Footing of Deixis and Reference*, pages 1–28.
- Martin, James R. and Peter R. R. White. 2005. *The Language of Evaluation: Appraisal in English*. Palgrave Macmillan.
- Nuyts, Jan. 2001. Subjectivity as an evidential dimension in epistemic modal expressions. *Journal of Pragmatics*, 33(3):383–400.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman, London and New York.
- Stein, Dieter and Susan Wright, editors. 1995. *Subjectivity and Subjectivisation: Linguistic Perspectives*. Cambridge University Press, Cambridge, UK.