Book Reviews

Sentiment Analysis and Opinion Mining

Bing Liu
(University of Illinois at Chicago)

Reviewed by
Claire Cardie
Cornell University

This 2012 book is written as a comprehensive introductory and survey text for sentiment analysis and opinion mining, a field of study that investigates computational techniques for analyzing text to uncover the opinions, sentiment, emotions, and evaluations expressed therein. As such, it aims to be accessible to a broad audience that includes students, researchers, and practitioners, as well as to cover all important topics in the field.

With regard to the first aim, the book is very much a success: The writing is clear and concise, informative examples motivate each new topic, terminology is clearly defined, and descriptions of key algorithms are provided in the running text along with short (usually one-line) descriptions of each piece of relevant related work. The latter, in particular, makes the book an excellent platform from which to dive into the quickly expanding body of literature on sentiment and opinion analysis. In addition, I believe that the book should be easily accessible to anyone with a computer science background.

With regard to Liu's second aim of covering all important topics in the field, the degree to which the book succeeds is a matter of, well, opinion. Let me explain. Liu's early research was in data mining and Web mining; not surprisingly then, the book is written from this perspective. It is very much centered around the analysis of user-generated opinions in social media. Liu's particular expertise is in the area of product reviews; hence, the bulk of the book's examples are from this domain. Furthermore, the book focuses on techniques that are first and foremost applicable to **aspect-based sentiment analysis**—fine-grained analysis of opinions regarding specific aspects of products and services. For the most part, investigations in this area have been restricted to reviews of electronics products (e.g., cameras), hotels, and restaurants with their associated entity-specific aspects (e.g., weight, photo quality, and ease of use for cameras; rooms, front desk service, and cleanliness for hotels; and food, service, ambience, and cost for restaurants).

In contrast, the similarly named survey of Pang and Lee (2008)—*Opinion Mining and Sentiment Analysis*—is more even-handed in its selection of topics and techniques and is written from the point of view of natural language processing (NLP) and computational linguistics. Pang and Lee, for example, are aware of prior work in the field on fact and event-based text analysis and, within that context, focus consciously on the description of "*new challenges*[1] raised by sentiment-aware applications" as well as the methods

---

1 Emphasis added.

proposed to address them. As a result, the survey proves to be an easy, comfortable (and entertaining) read for those with an NLP-centric ancestry.

Not so with the Liu survey, because the goals and assumptions that underlie aspect-based sentiment analysis can be at odds with many of those at the core of computational linguistics and NLP. But do not despair! This is a good thing! Precisely because of Liu's different tack on opinion and sentiment analysis, for many readers the book will be a wonderful source of ideas for new problems to work on in the field. In particular, the language of product reviews is quite different from that of other opinion-oriented text (e.g., editorials, blogs, position papers, political arguments, and even movie reviews). Product reviews tend to be quite short; they describe a single, known product; the opinion expressions themselves tend to be product-specific. Other genres of opinion-oriented text are generally longer; they can discuss virtually any topic or set of topics and, hence, are likely to exhibit a greater variety of sentence and discourse structure, including a virtually unlimited (and, out of context, ambiguous) opinion expression vocabulary, the presence of opinion holders that are not the author, and implicit opinion targets.

With this contrast in writing genres in mind, reading the book becomes a thought experiment in determining whether the techniques it covers will perform well on opinion-oriented texts beyond product reviews; if not, how and when will they fail; and in what circumstances might more complex language understanding components like parsing, semantic interpretation, or discourse analysis be helpful in analyzing product reviews?

The book contains eleven chapters (with a short summary of the book in a final twelfth chapter). Chapter 1 introduces the problem of sentiment analysis. It discusses the differences in terminology that exist in industry vs. academia and briefly describes approximately 20 recent[2] applications-oriented sentiment analysis research efforts published largely at venues outside of NLP. The latter is a nice entree into the applied sentiment analysis literature beyond the standard NLP conferences.

Chapter 2 provides an abstraction of the opinion mining problem and formally defines an *opinion* in terms of its components—the opinion holder, the entity and aspect of that entity that is the target of the opinion, the sentiment expressed, and the time that the opinion was expressed. The remaining chapters are organized around this definition and the sentiment-based applications that it enables. Thus, there are chapters on document-level sentiment classification and rating prediction (Chapter 3), sentence-level subjectivity and sentiment classification (Chapter 4), and aspect-based sentiment analysis (Chapter 5, roughly one-quarter of the book). Following these are shorter chapters on sentiment lexicon generation, opinion summarization, the analysis of comparative opinions, opinion retrieval (vs. Web search), and determining review quality (Chapters 6, 7, 8, 9, and 11, respectively).

Chapter 10 is a somewhat longer chapter devoted to opinion spam detection. Here Liu describes techniques to identify fake product reviews—some that rely primarily on the review content and available meta-data, and others based on identifying atypical behaviors of the reviewer(s).

The chapter on aspect-based sentiment analysis is, by far, the most detailed, but there are some nice surprises in other chapters including sections on handling sarcastic sentences, learning a priori objective terms that imply an opinion, and analyzing

---

2 Most papers are from 2010 and 2011.

opinions in contexts, as well as multiple sections that address cross-language and cross-domain issues.

Overall, the book is a very valuable resource for those interested in understanding the quickly expanding literature on sentiment and opinion analysis, especially techniques for aspect-based sentiment analysis. For NLP researchers, it can also serve as a source of new problems to tackle in the analysis of opinion-oriented text.

Reference

Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2:1–135.

This book review was edited by Pierre Isabelle.

Claire Cardie is a Professor in the Computer Science and Information Science departments at Cornell University. She has served as an action editor for the *Journal of Machine Learning Research (JMLR)*, an associate editor for the *Journal of Artificial Intelligence Research (JAIR)*, and an editorial board member for *Machine Learning* and *Computational Linguistics*. She was Program Chair (with Pierre Isabelle) of the joint ACL/COLING conference in 2006. E-mail: `cardie@cs.cornell.edu`.