

Book Review

Sentiment Analysis: Mining Opinions, Sentiments, and Emotions

Bing Liu

(University of Illinois at Chicago)

Cambridge University Press, 2015, 381 pp.; hardcover, ISBN 9781107017894, \$80

Reviewed by

Jun Zhao, Kang Liu, and Liheng Xu

Institute of Automation, Chinese Academy of Sciences

With the increasing development of Web 2.0, such as social media and online businesses, the need for perception of opinions, attitudes, and emotions grows rapidly. Sentiment analysis, the topic studying such subjective feelings expressed in text, has attracted significant attention from both the research community and industry. Although we have known sentiment analysis as a task of mining opinions expressed in text and analyzing the entailed sentiments and emotions, so far the task is still vaguely defined in the research literature because it involves many overlapping concepts and sub-tasks. Because this is an important area of scientific research, the field needs to clear this vagueness and define various directions and aspects in detail, especially for students, scholars, and developers new to the field. In fact, the field includes numerous natural language processing tasks with different aims (such as sentiment classification, opinion information extraction, opinion summarization, sentiment retrieval, etc.) and these have multiple solution paths. Bing Liu has done a great job in this book in providing a thorough exploration and an anatomy of the sentiment analysis problem and conveyed a wealth of knowledge about different aspects of the field.

Liu is a leading figure in this research area. Not only has he made important contributions to the understanding of opinions and sentiments expressed in text, but he also has significantly influenced the design of real-life sentiment analysis algorithms and the building of practical sentiment analysis systems. This book has at least three significant merits and meets the needs of different types of readers.

1. First, it is praiseworthy that the book gives detailed definitions of opinions and probably all important pertinent concepts, including sentiment, opinion target, time of opinion, opinion holder, opinion reason, opinion qualifier, and the often-neglected opinion types such as comparative opinions, fact-implied opinions, and so forth. The book explains and illustrates these concepts clearly, which facilitates a comprehensive and principled understanding of the sentiment analysis problem. In our opinion, the definitions in this book for sentiment analysis are more complete and profound than any other publication that we have seen.
2. This book not only presents the main sub-tasks of sentiment analysis, such as sentiment classification at different discourse levels, opinion summarization, opinion search, and emotion identification, but also covers

many emerging sentiment-related topics, such as sentiment analysis of debates and discussions, mining of intentions, and deceptive opinion detection. This shows that Liu has extensive research experience as well as knowledge of industry applications and knows the importance of these emerging tasks, which are in demand from the industry. Furthermore, this book includes the latest methods and technologies for each related task published in the mainstream conference proceedings and journals. What is especially valuable is that it describes first-hand experience in building a sentiment analysis system. Liu thus brings to readers a global view of the whole area from research to practice, and meets the needs of different readerships, from experienced scholars in natural language processing to graduate students and industry practitioners interested in sentiment analysis and social media analysis. This book thus serves as a comprehensive yet in-depth survey of references in sentiment analysis.

3. It is also commendable that the book gives a balanced treatment of both linguistic approaches and machine learning approaches to solving the sentiment analysis problem. In recent years, the mainstream methods in the research literature have been based on supervised learning with elaborate feature engineering. State-of-the-art techniques also use deep learning models to learn effective features directly from raw data as an alternative to manual feature engineering. However, as mentioned in this book, supervised learning methods provide no linguistic interpretations and do not generate knowledge for linguists or industry developers to gain insights of the problem. Practically, when errors occur in an application, it is hard to know what is wrong and how to fix it. Fortunately, this book provides a comprehensive list of linguistic constructs and perspectives that are instrumental for sentiment analysis, which make up for the deficiency of black-box approaches using pure machine learning. Moreover, it also lists and elaborates on many specific linguistic phenomena that are critical for effective classification of sentiment such as negation (Chapter 5), modality (Chapter 5), and comparison (Chapter 8). We believe that this book will enable the reader to gain not just a comprehensive understanding of the computation methods but also deep linguistic insights of the sentiment analysis problem and its possible solutions.

The book is organized into 13 chapters. The first two chapters introduce the basics and define the sentiment analysis problem. Chapters 3–9 discuss the core sentiment analysis tasks (e.g., sentiment classification, aspect analysis, and opinion summarization) and their current solution methods. Chapters 10–13 investigate the emerging themes from recent research and applications (e.g., analysis of debates, intentions, fake opinions, and review quality).

Specifically, Chapter 1 motivates and gives an overview of the whole book. It describes the expression of sentiment as one of the most important and complicated phenomena of human language. The goal of sentiment analysis is to computationally extract sentiments, opinions, and emotions expressed in text, which is different from the goal of traditional linguistic studies aimed at understanding the human language. Technically, analysis of sentiment can be divided into several levels according to different

discourse granularity, such as document, sentence, and aspect or sentiment-and-target levels. Liu took a structured approach to write this book.

Chapter 2 gives the definition of sentiment analysis, along with discussions of many key concepts such as subjectivity, affect, emotion, and mood. An opinion can either be defined as a quadruple or quintuple, and sentiment involves the type, orientation, and intensity about an opinion. Many succinct examples are also given to make the concepts easily understandable.

Chapter 3 discusses document-level sentiment classification, which is probably the most studied problem in sentiment analysis. It usually involves direct applications of (supervised or unsupervised) machine learning algorithms. Although the problem is called sentiment classification, it could also be taken as a regression problem. Most sentiment classification approaches categorize documents into only positive, negative, and objective types, whereas emotion classification involves more categories with overlapping meanings and is more difficult to perform accurately.

Document-level sentiment classification is, however, too coarse for practical applications. More elaborate analysis is applied to sentence and aspect levels, as introduced in Chapters 4 and 5. These two problems are the most practically useful research topics of sentiment analysis. In particular, aspect-level analysis forms the core of applications of sentiment analysis as it aims to identify the atomic unit of information contained in sentiment, opinion, and emotion expressions, which is the pair of sentiment and its target. For sentiment identification, the book shows that beyond positive and negative orientations of sentiment expressions, many sophisticated language phenomena also need to be considered in analysis (e.g., conditional expressions, sarcasm, sentiment composition, and negation). To deal with such sophisticated language phenomena, term-level features are no longer sufficient. One has to incorporate deeper knowledge of syntax, semantics, and discourse. Although significant efforts have been made, the problem is still far from being solved.

Chapters 6 and 7 describe aspect extraction and sentiment lexicon generation, respectively. The two topics are inter-related because sentiment lexicons are often used to express opinions on aspects or targets. Therefore, many researchers have studied the two tasks together. Furthermore, aspects and sentiment expressions also have common characteristics, such as, for example, the fact that they are both domain-specific. Although the two problems are challenging, these chapters provide valuable resources and practical algorithms for their solutions.

Chapter 8 studies comparative expressions related to sentiment. The task is also challenging because of the flexible usage of comparatives and the difficulty of identifying the preferred entity set. Limited research attention has been paid to this problem, but it is very useful in applications.

Chapter 9 provides several approaches to generating opinion summaries. Opinion summarization is quite different from conventional single-document or multi-document summarization, because it is centered on opinion targets and produces quantitative sentiment ratings for targets. The output of opinion summarization can be interactively visualized, which gives really interesting and easily understandable results that can be seen by users in a single glance.

Chapters 10–13 introduce several related topics to sentiment analysis, namely, the study of debates, intentions, deceptive opinions, and quality of reviews. These are emerging themes from recent research. We believe there will be significant research activities and applications on these topics in the years to come.

As a whole, this book serves as a useful introduction to sentiment analysis along with in-depth discussions of linguistic phenomena related to sentiments, opinions, and

emotions. Although many sentiment analysis methods are based on machine learning as in other NLP tasks, sentiment analysis is much more than just a classification or regression problem, because the natural language constructs used to express opinions, sentiments, and emotions are highly sophisticated, including sentiment shift, implicated expression, sarcasm, and so on. Liu has described these issues and problems very clearly. Readers will find this book to be inspiring and it will arouse their interests in sentiment analysis.

Jun Zhao is a professor in the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. His primary research focus is information extraction and question answering. Zhao's e-mail address is jzhao@nlpr.ia.ac.cn. *Kang Liu* is an associate professor in the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. His primary research focus is opinion mining, information extraction, and machine learning. Liu's e-mail address is kliu@nlpr.ia.ac.cn. *Liheng Xu* received a Ph.D. degree from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. His primary research focus is opinion mining and deep learning. Xu's e-mail address is lhxu@nlpr.ia.ac.cn.