

BOOK REVIEW

Lei Lei and Dilin Liu, *Conducting Sentiment Analysis*. Cambridge University Press, 2021. pp. 104. ISBN 978-1-108-82921-2 (paperback), 978-1-108-90967-9 (E-book)

Conducting Sentiment Analysis, coauthored by Lei Lei and Dilin Liu, is the fourth Element of the *Cambridge Elements in Corpus Linguistics Series*. As its name suggests, it is a book dedicated to sentiment analysis. This technique, also known as opinion mining, makes use of computer technologies to identify and extract human subjective sentiments and opinions toward individuals, events, topics, services, or products that are being evaluated (Liu, 2015). The primary objective of sentiment analysis is to automatically determine the sentiment polarity of a given text, that is, whether the text is positive, neutral, or negative in attitudes, emotions, or opinions (Mäntylä et al., 2018; Taboada, 2016). It is noteworthy that although sentiment analysis is sometimes associated with emotion analysis, the latter is essentially a “more specialized subcategory of sentiment analysis” (p. 1) because it involves more detailed dimensions of emotion (e.g., anger, anticipation, disgust, fear, joy, sadness, surprise, and trust).

This book, targeting its readership of students and researchers in corpus linguistics, covers basic concepts, application scenarios, practical methods, and implementations of sentiment analysis. Specifically, it comprises six chapters, consisting of the background of sentiment analysis, a review of methods, a step-by-step tutorial, two chapters of detailed case studies, and a conclusion. In addition, the datasets and codes used in the book are publicly available at www.cambridge.org/sentimentanalysis.

The authors begin Chapter 1 with a definition and introduction of sentiment analysis. The introduction then shifts to previous works of sentiment analysis. As the Internet has become one of the major venues of communication, millions of people express their opinions and sentiments via blogs, Twitter, Facebook, forum, or other online media. These opinions and sentiments are relevant to people’s daily lives which in turn may provide useful insights into the understanding of public opinions and hence are invaluable assets for stakeholders (e.g., governments, manufacturers, and service providers) to make informed decisions and respond accordingly (Birjali et al., 2021). Given the volume and value of these opinionated texts, efficient computational tools are needed to facilitate the automatic extraction of opinions and sentiments. Sentiment analysis is such a powerful tool capable of processing a large volume of digital text data and quickly capturing an overview of the opinions and sentiments expressed in the text. For this reason, it has attracted growing research interests over the past decade or so and its reach has been expanded from computer science to areas such as marketing, finance, health care, communications, political science, and social studies (Zhang et al., 2018). Likewise, sentiment analysis also has the potential to benefit linguistic research, particularly the study of corpus linguistics, from a methodological perspective.

In Chapter 2, the authors critically review and compare two major methods for sentiment analysis: the unsupervised (or lexicon-based) and supervised machine-learning methods. On the one hand, the unsupervised/lexicon-based method does

not require training datasets but makes use of sentiment lexicons to gauge the polarity of a given text. Sentiment lexicon is one of the key ingredients of this approach. For end users, many off-the-shelf lexicons are publicly available in R. Nonetheless, readers are advised to be cautious in choosing lexicons for texts of different domains because some lexicons are domain-independent but others are not. For example, while the Jockers lexicon (Jockers, 2017) is supposed to work well on general texts such as news and novels, the Loughran–McDonald lexicon (Loughran & McDonald, 2016) is made for analyzing texts of business and finance and may not be suitable for analyzing texts in other fields. On the other hand, the supervised machine-learning method is more sophisticated. This approach first requires all the texts in a target corpus to be manually tagged for sentiment polarity. Then linguistic features (e.g., words, phrases, and syntactic patterns) that are distinctive of a text and its corresponding sentiment polarity are chosen according to the purpose of a task and the experience of the researcher. When the selection of features is done, the tagged texts are split into two subsets, that is, the training set and the test set. Next, the computer algorithms automatically “learn” the association between linguistic features and the sentiment polarity from texts in the training set. Finally, the test set is used to evaluate whether the models have achieved an acceptable level of performance (e.g., accuracy and precision). If the results are considered satisfactory, the training of the model is completed and the trained models are ready for sentiment analysis tasks with a larger corpus in a similar domain or field. Compared with the lexicon-based method, the supervised machine-learning method usually features better accuracy. However, this approach requires human annotations for the training and the test datasets, a process which is often time-consuming; and for a different task, this work may need to be redone. It should be noted that regardless of the methodological differences, sentiment analysis can be applied at document level, sentence level, and aspect level. Document-level analysis is concerned with the overall sentiment of a given text or document, sentence-level analysis computes the sentiment of each sentence in a text, and aspect-level analysis evaluates different aspects of a product (e.g., quality, price, and service).

Chapter 3 is the only chapter that offers a step-by-step tutorial that focuses exclusively on the technical details of how to conduct sentiment analysis in the R language. Readers are expected to have a basic knowledge of programming with R before entering into this section. For demonstration and comparison purposes, this chapter uses an open collection of tweets about services provided by U.S. airlines as the dataset for all the sentiment and emotion analysis. The authors first illustrate how the supervised machine-learning sentiment analysis is implemented with R. The complete workflow such as data preparation, model training, prediction, and classification is presented with codes and explanations in plain language. So even novice readers will have no difficulty in understanding. It then moves on to the section on unsupervised/lexicon-based method, in which the authors provide two examples to show how lexicon-based sentiment analysis and emotion analysis can be done respectively with existing R packages. In short, the hands-on tutorials provided in this chapter help readers to understand the procedures of different methods for sentiment analysis and to conduct the analysis from scratch on their own.

Chapter 4 offers an interesting case study of sentiments and emotions in political speeches. Using a corpus of the *State of the Union addresses* (SOTUs) between 1790 and 2020, the authors conduct a lexicon-based sentiment analysis and a lexicon-based emotion analysis to trace the diachronic changes of sentiments and emotions in

SOTUs. Results of the study show that the overall sentiments and emotions in the SOTUs are largely positive and remain relatively stable across the past 230 years. On the other hand, the results also reveal substantial fluctuations in terms of the mean sentiment scores among some of the SOTUs, suggesting that some presidents are seemingly more positive than others in their speeches. A qualitative analysis of the text of the most positive/negative SOTUs shows that social and historical events such as war and economic crisis are closely related to the sentiment polarity. More importantly, this chapter is also enlightening because it expands the application of sentiment analysis to the corpus-based study of political texts. To be specific, it demonstrates how the expressions of sentiments and emotions in the SOTUs can be captured, analyzed, visualized, reported, and discussed with reference to major historical events (e.g., the American Civil War) and differences in language use (e.g., style of speech) among the U.S. presidents.

In Chapter 5, the second case study investigates sentiments and emotions in online movie reviews. The corpus used in this study is a large sample of movie review texts extracted from the Internet Movies Database (IMDb). The 50,000 reviews in the corpus are manually labeled for sentiment polarity, with half of the reviews tagged as positive and the other half tagged as negative (neutral reviews were not included in the dataset). The study first tests the power of supervised machine-learning sentiment analysis at document level. The model trained for this task yields a satisfactory result in terms of overall accuracy. However, a close reading of the misclassified samples suggests that the excessive use of negative/positive words in the reviews may influence the accuracy of the model. To better understand the movie reviews, the authors then conduct an extra lexicon-based emotion analysis at the aspect level. It is found that the mean emotion scores of “trust,” “anticipation,” and “joy” are higher than those of “fear,” “sadness,” “anger,” “surprise,” and “disgust.” A further qualitative analysis of the review texts provides new insights into the emotion scores. That is, if a review text is mainly concerned with the stories and plots of the movie itself rather than the reviewer’s personal evaluation of the movie, the results of emotion analysis may be unreliable.

The authors draw a short conclusion with only one page in Chapter 6. In the conclusive part, they reiterate the achievements and limitations elaborated in previous chapters, and briefly mention the future directions of research in sentiment analysis. For example, although sentiment analysis techniques have improved over the past decades, the development of lexicons and annotated training datasets is seemingly lagged behind. Moreover, most of the current studies have focused on languages such as “English, Chinese, Arabic, and Indian” (p. 95). Pertinent research on other languages is far from sufficient.

Conducting Sentiment Analysis is a timely book for and a noteworthy contribution to the study of corpus linguistics. As noted earlier, although sentiment analysis has been successfully used to help gauge sentiments and emotions in various domains and fields (Birjali et al., 2021; Liu & Lei, 2018), its application in corpus linguistics, or linguistic research as a whole, is rather scarce. Traditionally, the study of corpus linguistics usually employs a number of techniques such as word lists, concordance lines, collocates, cluster, and keyness lists to investigate the patterns of language use (McEnery & Hardie, 2011). In other words, traditional methods in corpus linguistics describe how language is used in context. Sentiment analysis, on the other hand, makes use of natural language processing (NLP) and machine learning (ML) techniques to gauge sentiments and emotions in language. That is, it offers

an innovative approach to explore the affective dimension of language use. Since the linguistic expression of sentiments and emotions is a universal human trait (Taboada, 2016), it is of great significance for students and professionals of corpus linguistics to master techniques such as sentiment analysis to address issues concerning the affective dimensions of language. The book fills this gap by providing timely additions to the methodological arsenal for corpus linguistic studies. Readers will surely benefit from reading this book because it keeps them updated with the state-of-the-art text analytical techniques for corpus linguistic research.

To sum up, this book defines sentiment analysis as an interdisciplinary research field deeply rooted in both linguistics and computer science. It offers novel techniques and perspectives for readers to examine the affective dimension of human language. The detailed hands-on tutorials are easy to follow and the case studies are interesting and inspiring. Without doubt, *Conducting Sentiment Analysis* will definitely be a valuable addition to the reading list and a must-read for novice researchers of corpus linguistics.

Acknowledgments. Work on this review is supported by Sichuan Foreign Languages & Literature Research Center (No. SCWYH18-06) and the Social Science Fund of Sichuan Province (No. SC21WY002).

Lan Yi

*School of Foreign Languages, Chengdu Jincheng College,
Sichuan, China*

Ju Wen

*School of Liberal Education, Chengdu Jincheng College,
Sichuan, China*

doi:10.1017/langcog.2022.27

References

- Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226(16), 107134. <https://doi.org/10.1016/j.knosys.2021.107134>
- Jockers, M. L. (2017). *Syuzhet: Extract sentiment and plot arcs from text*. <https://github.com/mjockers/syuzhet>
- Liu, B. (2015). *Sentiment analysis*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139084789>
- Liu, D., & Lei, L. (2018). The appeal to political sentiment: An analysis of Donald Trump's and Hillary Clinton's speech themes and discourse strategies in the 2016 US presidential election. *Discourse, Context & Media*, 25(5), 143–152. <https://doi.org/10.1016/j.dcm.2018.05.001>
- Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), 1187–1230. <https://doi.org/10.2139/ssrn.2504147>
- Mäntylä, M. V., Graziotin, D., & Kuuttila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, 27(1), 16–32. <https://doi.org/10.1016/j.cosrev.2017.10.002>
- McEnery, T., & Hardie, A. (2011). What is corpus linguistics? In T. McEnery & A. Hardie (Eds.), *Corpus linguistics* (pp. 1–24). Cambridge University Press. <https://doi.org/10.1017/CBO9780511981395.002>
- Taboada, M. (2016). Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics*, 2(1), 325–347. <https://doi.org/10.1146/annurev-linguistics-011415-040518>
- Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *WIREs Data Mining and Knowledge Discovery*, 8(4), e1253. <https://doi.org/10.1002/widm.1253>