

Cambridge University Press

978-1-107-01789-4 - Sentiment Analysis: Mining Opinions, Sentiments, and Emotions

Bingliu

Frontmatter

[More information](#)

---

## Sentiment Analysis

### *Mining Opinions, Sentiments, and Emotions*

Sentiment analysis is the computational study of people's opinions, sentiments, emotions, and attitudes. This fascinating problem is increasingly important in business and society. It offers numerous research challenges but promises insight useful to anyone interested in opinion analysis and social media analysis.

This book gives a comprehensive introduction to the topic from a primarily natural language processing point of view to help readers understand the underlying structure of the problem and the language constructs that are commonly used to express opinions and sentiments. It covers all core areas of sentiment analysis; includes many emerging themes, such as debate analysis, intention mining, and fake-opinion detection; and presents computational methods to analyze and summarize opinions. It will be a valuable resource for researchers and practitioners in natural language processing, computer science, management sciences, and the social sciences.

Bing Liu is a professor of computer science at the University of Illinois at Chicago. His current research interests include sentiment analysis and opinion mining, data mining, machine learning, and natural language processing. He has published extensively in top conferences and journals, and his research has been cited on the front page of the *New York Times*. He is also the author of two books: *Sentiment Analysis and Opinion Mining* (2012) and *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data* (first edition, 2007; second edition, 2011). He currently serves as the Chair of ACM SIGKDD and is an IEEE Fellow.

Cambridge University Press  
978-1-107-01789-4 - Sentiment Analysis: Mining Opinions, Sentiments, and Emotions  
Bingliu  
Frontmatter  
[More information](#)

---

Cambridge University Press

978-1-107-01789-4 - Sentiment Analysis: Mining Opinions, Sentiments, and Emotions

Bingliu

Frontmatter

[More information](#)

---

# Sentiment Analysis

*Mining Opinions, Sentiments, and Emotions*

BING LIU

*University of Illinois at Chicago*



Cambridge University Press  
978-1-107-01789-4 - Sentiment Analysis: Mining Opinions, Sentiments, and Emotions  
Bingliu  
Frontmatter  
[More information](#)

CAMBRIDGE  
UNIVERSITY PRESS

32 Avenue of the Americas, New York, NY 10013-2473, USA

Cambridge University Press is part of the University of Cambridge.  
It furthers the University's mission by disseminating knowledge in the pursuit of  
education, learning, and research at the highest international levels of excellence.

[www.cambridge.org](http://www.cambridge.org)  
Information on this title: [www.cambridge.org/9781107017894](http://www.cambridge.org/9781107017894)

© Bing Liu 2015

This publication is in copyright. Subject to statutory exception  
and to the provisions of relevant collective licensing agreements,  
no reproduction of any part may take place without the written  
permission of Cambridge University Press.

First published 2015

Printed in the United States of America

*A catalog record for this publication is available from the British Library.*

*Library of Congress Cataloging in Publication Data*

Liu, Bing, 1963–

Sentiment analysis : mining opinions, sentiments, and emotions / Bing Liu.  
pages      cm

Includes bibliographical references and index.

ISBN 978-1-107-01789-4 (hardback)

1. Natural language processing (Computer science)    2. Computational linguistics.    3. Public  
opinion – Data processing.    4. Data mining.    I. Title.

QA76.9.N38L58    2015

006.3'12–dc23      2014036113

ISBN 978-1-107-01789-4 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for  
external or third-party Internet websites referred to in this publication and does not guarantee that  
any content on such websites is, or will remain, accurate or appropriate.

Contents

Preface	<i>page</i> xi
Acknowledgments	xv
<b>1 Introduction</b>	<b>1</b>
1.1 Sentiment Analysis Applications	4
1.2 Sentiment Analysis Research	8
1.2.1 Different Levels of Analysis	9
1.2.2 Sentiment Lexicon and Its Issues	10
1.2.3 Analyzing Debates and Comments	11
1.2.4 Mining Intentions	12
1.2.5 Opinion Spam Detection and Quality of Reviews	12
1.3 Sentiment Analysis as Mini NLP	14
1.4 My Approach to Writing This Book	14
<b>2 The Problem of Sentiment Analysis</b>	<b>16</b>
2.1 Definition of Opinion	17
2.1.1 Opinion Definition	17
2.1.2 Sentiment Target	19
2.1.3 Sentiment of Opinion	20
2.1.4 Opinion Definition Simplified	22
2.1.5 Reason and Qualifier for Opinion	24
2.1.6 Objective and Tasks of Sentiment Analysis	25
2.2 Definition of Opinion Summary	29
2.3 Affect, Emotion, and Mood	31
2.3.1 Affect, Emotion, and Mood in Psychology	31
2.3.2 Affect, Emotion, and Mood in Sentiment Analysis	36
2.4 Different Types of Opinions	39
2.4.1 Regular and Comparative Opinions	39
2.4.2 Subjective and Fact-Implied Opinions	40
2.4.3 First-Person and Non-First-Person Opinions	44

vi	Contents
2.4.4	Meta-Opinions 44
2.5	Author and Reader Standpoint 45
2.6	Summary 45
3	<b>Document Sentiment Classification 47</b>
3.1	Supervised Sentiment Classification 49
3.1.1	Classification Using Machine Learning Algorithms 49
3.1.2	Classification Using a Custom Score Function 56
3.2	Unsupervised Sentiment Classification 57
3.2.1	Classification Using Syntactic Patterns and Web Search 57
3.2.2	Classification Using Sentiment Lexicons 59
3.3	Sentiment Rating Prediction 61
3.4	Cross-Domain Sentiment Classification 63
3.5	Cross-Language Sentiment Classification 65
3.6	Emotion Classification of Documents 67
3.7	Summary 68
4	<b>Sentence Subjectivity and Sentiment Classification 70</b>
4.1	Subjectivity 72
4.2	Sentence Subjectivity Classification 73
4.3	Sentence Sentiment Classification 76
4.3.1	Assumption of Sentence Sentiment Classification 77
4.3.2	Classification Methods 78
4.4	Dealing with Conditional Sentences 80
4.5	Dealing with Sarcastic Sentences 82
4.6	Cross-Language Subjectivity and Sentiment Classification 84
4.7	Using Discourse Information for Sentiment Classification 86
4.8	Emotion Classification of Sentences 87
4.9	Discussion 88
5	<b>Aspect Sentiment Classification 90</b>
5.1	Aspect Sentiment Classification 91
5.1.1	Supervised Learning 92
5.1.2	Lexicon-Based Approach 93
5.1.3	Pros and Cons of the Two Approaches 96
5.2	Rules of Sentiment Composition 98
5.2.1	Sentiment Composition Rules 99
5.2.2	DECREASE and INCREASE Expressions 106
5.2.3	SMALL_OR_LESS and LARGE_OR_MORE Expressions 109
5.2.4	Emotion and Sentiment Intensity 112
5.2.5	Senses of Sentiment Words 112
5.2.6	Survey of Other Approaches 114
5.3	Negation and Sentiment 116
5.3.1	Negation Words 116
5.3.2	Never 119

Contents	vii
5.3.3 Some Other Common Sentiment Shifters	121
5.3.4 Shifted or Transferred Negations	122
5.3.5 Scope of Negations	122
5.4 Modality and Sentiment	123
5.5 Coordinating Conjunction <i>But</i>	127
5.6 Sentiment Words in Non-opinion Contexts	129
5.7 Rule Representation	131
5.8 Word Sense Disambiguation and Coreference Resolution	133
5.9 Summary	135
<b>6 Aspect and Entity Extraction</b>	<b>137</b>
6.1 Frequency-Based Aspect Extraction	138
6.2 Exploiting Syntactic Relations	140
6.2.1 Using Opinion and Target Relations	141
6.2.2 Using Part-of and Attribute-of Relations	147
6.3 Using Supervised Learning	149
6.3.1 Hidden Markov Models	150
6.3.2 Conditional Random Fields	151
6.4 Mapping Implicit Aspects	153
6.4.1 Corpus-Based Approach	153
6.4.2 Dictionary-Based Approach	154
6.5 Grouping Aspects into Categories	157
6.6 Exploiting Topic Models	159
6.6.1 Latent Dirichlet Allocation	160
6.6.2 Using Unsupervised Topic Models	163
6.6.3 Using Prior Domain Knowledge in Modeling	168
6.6.4 Lifelong Topic Models: Learn as Humans Do	171
6.6.5 Using Phrases as Topical Terms	174
6.7 Entity Extraction and Resolution	179
6.7.1 Problem of Entity Extraction and Resolution	179
6.7.2 Entity Extraction	183
6.7.3 Entity Linking	184
6.7.4 Entity Search and Linking	185
6.8 Opinion Holder and Time Extraction	186
6.9 Summary	187
<b>7 Sentiment Lexicon Generation</b>	<b>189</b>
7.1 Dictionary-Based Approach	190
7.2 Corpus-Based Approach	193
7.2.1 Identifying Sentiment Words from a Corpus	194
7.2.2 Dealing with Context-Dependent Sentiment Words	195
7.2.3 Lexicon Adaptation	197
7.2.4 Some Other Related Work	198
7.3 Desirable and Undesirable Facts	199
7.4 Summary	200

viii	Contents
<b>8</b>	<b>Analysis of Comparative Opinions</b> 202
8.1	Problem Definition 202
8.2	Identify Comparative Sentences 206
8.3	Identifying the Preferred Entity Set 207
8.4	Special Types of Comparison 209
8.4.1	Nonstandard Comparison 209
8.4.2	Cross-Type Comparison 211
8.4.3	Single-Entity Comparison 212
8.4.4	Sentences Involving <i>Compare</i> and <i>Comparison</i> 214
8.5	Entity and Aspect Extraction 215
8.6	Summary 216
<b>9</b>	<b>Opinion Summarization and Search</b> 218
9.1	Aspect-Based Opinion Summarization 219
9.2	Enhancements to Aspect-Based Summary 221
9.3	Contrastive View Summarization 224
9.4	Traditional Summarization 225
9.5	Summarization of Comparative Opinions 225
9.6	Opinion Search 226
9.7	Existing Opinion Retrieval Techniques 227
9.8	Summary 229
<b>10</b>	<b>Analysis of Debates and Comments</b> 231
10.1	Recognizing Stances in Debates 232
10.2	Modeling Debates/Discussions 235
10.2.1	JTE Model 236
10.2.2	JTE-R Model: Encoding Reply Relations 240
10.2.3	JTE-P Model: Encoding Pair Structures 243
10.2.4	Analysis of Tolerance in Online Discussions 245
10.3	Modeling Comments 246
10.4	Summary 248
<b>11</b>	<b>Mining Intentions</b> 250
11.1	Problem of Intention Mining 250
11.2	Intention Classification 254
11.3	Fine-Grained Mining of Intentions 256
11.4	Summary 258
<b>12</b>	<b>Detecting Fake or Deceptive Opinions</b> 259
12.1	Different Types of Spam 262
12.1.1	Harmful Fake Reviews 262
12.1.2	Types of Spammers and Spamming 263
12.1.3	Types of Data, Features, and Detection 265
12.1.4	Fake Reviews versus Conventional Lies 267
12.2	Supervised Fake Review Detection 269
12.3	Supervised Yelp Data Experiment 272



Contents	ix
12.3.1 Supervised Learning Using Linguistic Features	273
12.3.2 Supervised Learning Using Behavioral Features	274
12.4 Automated Discovery of Abnormal Patterns	275
12.4.1 Class Association Rules	276
12.4.2 Unexpectedness of One-Condition Rules	277
12.4.3 Unexpectedness of Two-Condition Rules	280
12.5 Model-Based Behavioral Analysis	282
12.5.1 Spam Detection Based on Atypical Behaviors	282
12.5.2 Spam Detection Using Review Graph	283
12.5.3 Spam Detection Using Bayesian Models	284
12.6 Group Spam Detection	285
12.6.1 Group Behavior Features	288
12.6.2 Individual Member Behavior Features	290
12.7 Identifying Reviewers with Multiple Userids	291
12.7.1 Learning in a Similarity Space	292
12.7.2 Training Data Preparation	293
12.7.3 d-Features and s-Features	294
12.7.4 Identifying Userids of the Same Author	295
12.8 Exploiting Burstiness in Reviews	298
12.9 Some Future Research Directions	300
12.10 Summary	301
<b>13 Quality of Reviews</b>	<b>303</b>
13.1 Quality Prediction as a Regression Problem	303
13.2 Other Methods	305
13.3 Some New Frontiers	306
13.4 Summary	307
<b>14 Conclusions</b>	<b>309</b>
Appendix	315
Bibliography	327
Index	363

Cambridge University Press  
978-1-107-01789-4 - Sentiment Analysis: Mining Opinions, Sentiments, and Emotions  
Bingliu  
Frontmatter  
[More information](#)

---

# Preface

---

Opinion and sentiment and their related concepts, such as evaluation, appraisal, attitude, affect, emotion, and mood, are about our subjective feelings and beliefs. They are central to human psychology and are key influencers of our behaviors. Our beliefs and perceptions of reality, as well as the choices we make, are to a considerable degree conditioned on how others see and perceive the world. For this reason, our views of the world are very much influenced by others’ views, and whenever we need to make a decision, we often seek out others’ opinions. This is true not only for individuals but also for organizations. From an application point of view, we naturally want to mine people’s opinions and feelings toward any subject matter of interest, which is the task of *sentiment analysis*. More precisely, sentiment analysis, which is also called *opinion mining*, is a field of study that aims to extract opinions and sentiments from natural language text using computational methods.

The inception and rapid growth of sentiment analysis coincide with those of social media on the web, such as reviews, forum discussions, blogs, and microblogs, because for the first time in human history, we now have a huge volume of opinion data recorded in digital forms. These data, also called *user-generated content*, prompted researchers to mine them to discover useful knowledge. This naturally led to the problem of sentiment analysis or opinion mining because these data are full of opinions. That these data are full of opinions is not surprising, because the primary reason why people post messages on social media platforms is to express their views and opinions, and therefore sentiment analysis is at the very core of social media analysis. Since early 2000, sentiment analysis has grown to be one of the most active research areas in natural language processing. It is also widely studied in data mining, web mining, and information retrieval. In fact, the research has spread from computer science to management science and social science because of its importance to business and society as a whole. In recent years, industrial activities surrounding sentiment analysis have also thrived. Numerous start-ups have emerged. Many large corporations, for example, Microsoft, Google, Hewlett-Packard, and Adobe, have also built their

own in-house systems. Sentiment analysis systems have found applications in almost every business, health, government, and social domain.

Although no silver bullet algorithm can solve the sentiment analysis problem, many deployed systems are able to provide useful information to support real-life applications. I believe it is now a good time to document the knowledge that we have gained in research, and, to some extent, in practice, in a book. Obviously, I don't claim that I know everything that is happening in the industry, as businesses do not publish or disclose their algorithms. However, I have built a sentiment analysis system myself in a start-up company and served clients on projects involving social media data sets in a large variety of domains. Over the years, many developers of sentiment analysis systems in the industry have also told me roughly what algorithms they were using. Thus, I can claim that I have a reasonable knowledge of practical systems and their capabilities and firsthand experience in solving real-life problems. I try to pass along those nonconfidential pieces of information and knowledge in this book.

In writing this book, I aimed to take a balanced approach, analyzing the sentiment analysis problem from a linguistic angle to help readers understand the underlying structure of the problem and the language constructs commonly used to express opinions and sentiments and presenting computational methods to analyze and summarize opinions. Like many natural language processing tasks, most published computational techniques use machine learning or data mining algorithms with the help of text-specific clues or features. However, if we only focus on such computational algorithms, we will miss the deep insights of the problem, which in turn will hinder our progress on the computational front. Most existing machine learning algorithms are black boxes. They do not produce human-interpretable models. When something goes wrong, it is hard to know the cause and how to fix it.

In presenting linguistic constructs and perspectives, I do not follow the linguistic tradition in writing because the knowledge and the way that the knowledge is presented in the traditional linguistics literature are mainly for people to understand rather than for computers to operationalize to solve real-life problems. Although the knowledge of human beings and instructions for computers can largely intersect, they also have major differences. As a case in point, when I was working on the problem of mining opinions from conditional sentences, I read several linguistics books about conditionals. However, to my surprise, I found almost no linguistic knowledge that can be operationalized computationally to help solve the problem. I believe this is partially because the current computation technologies are not mature enough to have the same understanding capability as people and partially because much of the linguistic knowledge is not meant for computers to use. Another feature of this book is that it is not just about studying the language for human understanding per se, as much of the traditional linguistic literature does; it is also about practical applications of mining sentiment and opinion expressed in natural language, for which we not only want to recognize sentiment or opinion expressions and their polarities (or orientations) but also to

extract several other pieces of important information associated with sentiment or opinion. For example, we want to identify the real-world entities or topics that a sentiment or opinion is about. These entities or topics are called *opinion* (or *sentiment*) *targets*. Extracting opinion targets is extremely important in practice. For example, in the sentence “*I am disgusted by tax increase for the poor,*” if we only find that the sentence expresses a negative sentiment and/or an emotion of *disgust* from the sentence author, it is not that useful in practice. But if we also find that the negative sentiment is toward ‘*tax increase for the poor,*’ which is the target of the negative sentiment or emotion, the information becomes much more valuable. I hope this book can serve to encourage linguists to develop a comprehensive theory about sentiment and opinion and their associated concepts.

I write this book as an introductory text to the field of sentiment analysis and as a research survey. In many places, it is one or the other, and in some other places, it is a mixture of both. The reason for this mixed or somewhat unusual presentational style is that there are few mature techniques or algorithms for sentiment analysis, although numerous researchers have attempted to solve each subproblem. In many cases, we can see from the accuracy of the results of the published papers that they are not yet ready for prime time. Another reason for the mixed presentational style of this book is that most existing research methods are direct applications of machine learning and data mining algorithms employing text features. Because many books on machine learning and data mining cover these algorithms extensively, these algorithms are thus not detailed in this book. This book also does not detail the basics of linguistics or natural language processing, such as part-of-speech tagging, syntactic parsing, shallow parsing, and grammar. Although these topics are very important to sentiment analysis too, again, they have been covered in numerous books on natural language processing. This book thus assumes that readers know the basics of machine learning and natural language processing.

I tried to cover all major developments of the field in this book. It is thus quite comprehensive. Evidence of this is that the book cites more than six hundred publications from all major conferences and journals. I organize the book as follows. Chapter 1 introduces the book and gives the motivations for the study of sentiment analysis. We see that sentiment analysis is a fascinating and yet challenging problem with almost unlimited practical applications. Chapter 2 defines the sentiment analysis problem and discusses many of its related issues. Here we see that although sentiment analysis is a natural language processing problem, it can be defined structurally. Through the definition, we can transform unstructured text to structured data. This facilitates subsequent qualitative and quantitative analyses, which are critical for real-life applications. We also see that sentiment analysis is a multifaceted problem with many challenging and interrelated subproblems.

Chapter 3 studies the topic of document-level sentiment classification, which classifies an opinion document (e.g., a product review) as expressing a positive or negative sentiment. Chapter 4 studies the same classification problem but focuses on each individual sentence. Related problems of sentiment rating

Cambridge University Press

978-1-107-01789-4 - Sentiment Analysis: Mining Opinions, Sentiments, and Emotions

Bingliu

Frontmatter

[More information](#)

prediction, transfer learning, and multilingual sentiment classification are also discussed in these two chapters.

Chapters 5 and 6 go to the fine-grained level to study the most important topic of aspect-based sentiment analysis, which not only classifies sentiment but also identifies the target of sentiment or opinion. Most practical sentiment analysis or opinion mining systems in industry are based on this fine-grained level of analysis. Chapter 5 focuses on aspect sentiment classification, and Chapter 6 focuses on aspect or target extraction.

Chapter 7 describes research that compiles sentiment lexicons. A sentiment lexicon is a list of words and phrases (e.g., *good*, *amazing*, *bad*, *horrible*) that people often use to express positive or negative opinions. Chapter 8 studies opinions expressed in comparative sentences. Chapter 9 focuses on opinion summarization and opinion search. Chapter 10 looks into a different type of sentiment (agreement and disagreement) expressed in online debates and discussions, which involve extensive interactive exchanges among participants. Chapter 11 investigates intention mining, which aims to discover intentions expressed in language.

Chapter 12 switches to a very different topic: detecting fake or deceptive online opinions. Chapter 13 studies the problem of ranking online reviews based on their usefulness so that users can view the most useful reviews first. Chapter 14 concludes the book and discusses some future research.

The book is suitable for students, researchers, and practitioners who are interested in social media analysis and natural language processing in general and sentiment analysis or opinion mining in particular. It is written not only for the computer science audience but also for researchers and practitioners in management sciences and social sciences. Consumer sentiments and public opinions are central to many management and social science areas such as marketing, economics, communication, and political science. Lecturers can readily use the book in class for courses on natural language processing, social media analysis, social computing, and text and data mining. Lecture slides are available online.

# Acknowledgments

---

Many researchers have assisted me technically in writing this book. Without their help, the book might never have become reality. First of all, I would like to thank my former and current students – Junsheng Chen, Zhiyuan Chen, Xiaowen Ding, Geli Fei, Murthy Ganapathibhotla, Mingqing Hu, Nitin Jindal, Abhinav Kumar, Huayi Li, Arjun Mukherjee, Ramanathan Narayanan (Northwestern University), Federico Alberto Pozzi (visiting student from University of Milano-Bicocca), Guang Qiu (visiting student from Zhejiang University), Sathish Ramadoss, Jianfeng Si (visiting student from City University of Hong Kong), William Underwood, Andrea Vaccari, Vivek Venkataraman, Zhongwu Zhai (visiting student from Tsinghua University), and Lei Zhang – for contributing numerous research ideas over the years. Discussions with many researchers also helped shape the book: Shuanhu Bai, Jim Blomo, Erik Cambria, Malu G. Castellanos, Dennis Chong, Umesh Dayal, Eduard Dragut, Boi Faltings, Ronen Feldman, Christiane D. Fellbaum, Zhiqiang Gao, Alexander Gelbukh, Riddhiman Ghosh, Natalie Glance, Meichun Hsu, Joshua Huang, Minglie Huang, Jing Jiang, Birgit König, Xiao-li Li, Qian Liu, Boia Marina, Sharon Meraz, Enza Messina, Tieyun Qian, Jidong Shao, Mehrbod Sharifi, Hao Wang, Jan Wiebe, Qiang Yang, Clement Yu, Philip S. Yu, ChengXiang Zhai, Fangwei Zhang, Yuanlin Zhang, Jun Zhao, Xiaoyan Zhu, and Chengqing Zong. I am also in debt to the two anonymous reviewers; despite their busy schedules, they read the book very carefully and gave me numerous excellent suggestions. I have taken each and every one of them into consideration while improving the book. The web and the social media have made writing this book so much easier. I have found much valuable information from them that has helped in writing the book tremendously.

I would also like to express my gratitude to the National Science Foundation, Google, HP Labs, and the Microsoft Corporation for their generous support of my research over the years. The Department of Computer Science at the University of Illinois at Chicago provided computing resources and a supportive environment for this project.

On the publication side, it was a pleasure working with the helpful staff of Cambridge University Press. I thank my editor, Lauren Cowles, who asked me in August 2011 whether I was interested in writing a book about sentiment analysis for Cambridge University Press. It has been a wonderful experience working with her since. I also thank my copy editor, Holly T. Monteith, for helping me improve the presentation and my production editor, Sonika Rai, for guiding me through the final production process.

Finally, I thank my parents, brother, and sister for their constant support and encouragement. My deepest gratitude goes to my own family: Yue, Shelley, and Kate. They have helped me in so many ways. My wife has taken care of almost everything at home and put up with me and the long hours that I have spent on this book. I dedicate this book to them.

Bing Liu  
Chicago, USA  
July 10, 2014