

Contents

Preface	<i>page</i> xi
Acknowledgments	xv
1 Introduction	1
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
2 The Problem of Sentiment Analysis	16
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

2.4.4	Meta-Opinions	44
2.5	Author and Reader Standpoint	45
2.6	Summary	45
3	Document Sentiment Classification	47
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	Sentence Subjectivity and Sentiment Classification	70
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	Aspect Sentiment Classification	90
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

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
6	Aspect and Entity Extraction	137
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
7	Sentiment Lexicon Generation	189
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

8	Analysis of Comparative Opinions	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 Comparisons	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
9	Opinion Summarization and Search	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
10	Analysis of Debates and Comments	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
11	Mining Intentions	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
12	Detecting Fake or Deceptive Opinions	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

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
13	Quality of Reviews	303
13.1	Quality Prediction as a Regression Problem	303
13.2	Other Methods	305
13.3	Some New Frontiers	306
13.4	Summary	307
14	Conclusions	309
	Appendix	315
	Bibliography	327
	Index	363