

# Sentiment Analysis and Opinion Mining



# Synthesis Lectures on Human Language Technologies

## **Editor**

**Graeme Hirst, University of Toronto**

Synthesis Lectures on Human Language Technologies publishes monographs on topics relating to natural language processing, computational linguistics, information retrieval, and spoken language understanding. Emphasis is placed on important new techniques, on new applications, and on topics that combine two or more HLT subfields.

<http://www.morganclaypool.com/toc/hlt/1/1>

## **Sentiment Analysis and Opinion Mining**

Bing Liu

2012

## **Discourse Processing**

Manfred Stede

2011

## **Bitext Alignment**

Jörg Tiedemann

2011

## **Linguistic Structure Prediction**

Noah A. Smith

2011

## **Learning to Rank for Information Retrieval and Natural Language Processing**

Hang Li

2011

## **Computational Modeling of Human Language Acquisition**

Afra Alishahi

2010

**Introduction to Arabic Natural Language Processing**

Nizar Y. Habash  
2010

**Cross-Language Information Retrieval**

Jian-Yun Nie  
2010

**Automated Grammatical Error Detection for Language Learners**

Claudia Leacock, Martin Chodorow, Michael Gamon, Joel Tetreault  
2010

**Data-Intensive Text Processing with MapReduce**

Jimmy Lin, Chris Dyer  
2010

**Semantic Role Labeling**

Martha Palmer, Daniel Gildea, Nianwen Xue  
2010

**Spoken Dialogue Systems**

Kristiina Jokinen, Michael McTear  
2009

**Introduction to Chinese Natural Language Processing**

Kam-Fai Wong, Wenjie Li, Ruifeng Xu, Zheng-sheng Zhang  
2009

**Introduction to Linguistic Annotation and Text Analytics**

Graham Wilcock  
2009

**Dependency Parsing**

Sandra Kübler, Ryan McDonald, Joakim Nivre  
2009

**Statistical Language Models for Information Retrieval**

ChengXiang Zhai  
2008

© Springer Nature Switzerland AG 2022  
Reprint of original edition © Morgan & Claypool 2012

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

Sentiment Analysis and Opinion Mining  
Bing Liu

ISBN: 978-3-031-01017-0    paperback

ISBN: 978-3-031-02145-9    ebook

DOI 10.1007/978-3-031-02145-9

A Publication in the Springer series

*SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES*

Lecture #16

Series Editor: Graeme Hirst, University of Toronto

**Series ISSN**

ISSN 1947-4040    print

ISSN 1947-4059    electronic

# Sentiment Analysis and Opinion Mining

**Bing Liu**

University of Illinois at Chicago

*SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES #16*

## ABSTRACT

Sentiment analysis and opinion mining is the field of study that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from written language. It is one of the most active research areas in natural language processing and is also widely studied in data mining, Web mining, and text mining. In fact, this research has spread outside of computer science to the management sciences and social sciences due to its importance to business and society as a whole. The growing importance of sentiment analysis coincides with the growth of social media such as reviews, forum discussions, blogs, micro-blogs, Twitter, and social networks. For the first time in human history, we now have a huge volume of opinionated data recorded in digital form for analysis.

Sentiment analysis systems are being applied in almost every business and social domain because opinions are central to almost all human activities and are key influencers of our behaviors. Our beliefs and perceptions of reality, and the choices we make, are largely conditioned on how others see and evaluate the world. For this reason, when we need to make a decision we often seek out the opinions of others. This is true not only for individuals but also for organizations.

This book is a comprehensive introductory and survey text. It covers all important topics and the latest developments in the field with over 400 references. It is suitable for students, researchers and practitioners who are interested in social media analysis in general and sentiment analysis in particular. Lecturers can readily use it in class for courses on natural language processing, social media analysis, text mining, and data mining. Lecture slides are also available online.

## KEYWORDS

sentiment analysis, opinion mining, emotion, affect, evaluation, attitude, mood, social media, natural language processing, text mining.

# Acknowledgments

I would like to thank my former and current students—Zhiyuan Chen, Xiaowen Ding, Geli Fei, Murthy Ganapathibhotla, Minqing Hu, Nitin Jindal, Huayi Li, Arjun Mukherjee, Quang Qiu (visiting student from Zhejiang University), William Underwood, Andrea Vaccari, 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: Malu G. Castellanos, Dennis Chong, Umesh Dayal, Eduard Dragut, Riddhiman Ghosh, Natalie Glance, Meichun Hsu, Jing Jiang, Birgit König, Xiaoli Li, Tiejun Qian, Gang Xu, Philip S. Yu, Clement Yu, and ChengXiang Zhai. I am also very grateful to two anonymous reviewers. Despite their busy schedules, they read the book very carefully and gave me many excellent suggestions. I have taken each and every one of them into consideration while improving the book. On the publication side, I thank the Editor, Dr. Graeme Hirst, and the President and CEO of Morgan & Claypool Publishers, Mr. Michael Morgan, who have managed to get everything done on time and provided me with many pieces of valuable advice. Finally, my greatest gratitude goes to my own family: Yue, Shelley, and Kate, who have helped in so many ways.

# Contents

Preface.....	xiii
<b>1. Sentiment Analysis: A Fascinating Problem.....</b>	<b>1</b>
1.1 Sentiment Analysis Applications.....	2
1.2 Sentiment Analysis Research.....	3
1.2.1 Different Levels of Analysis .....	4
1.2.2 Sentiment Lexicon and Its Issues .....	5
1.2.3 Natural Language Processing Issues .....	6
1.3 Opinion Spam Detection .....	7
1.4 What's Ahead.....	7
<b>2. The Problem of Sentiment Analysis .....</b>	<b>9</b>
2.1 Problem Definitions .....	10
2.1.1 Opinion Defintion.....	10
2.1.2 Sentiment Analysis Tasks .....	14
2.2 Opinion Summarization.....	17
2.3 Different Types of Opinions.....	18
2.3.1 Regular and Comparative Opinions .....	18
2.3.2 Explicit and Implicit Opinions.....	19
2.4 Subjectivity and Emotion .....	19
2.5 Author and Reader Standpoint .....	21
2.6 Summary .....	21
<b>3. Document Sentiment Classification.....</b>	<b>23</b>
3.1 Sentiment Classification Using Supervised Learning.....	24
3.2 Sentiment Classification Using Unsupervised Learning .....	28
3.3 Sentiment Rating Prediction .....	30
3.4 Cross-Domain Sentiment Classification .....	31

3.5	Cross-Language Sentiment Classification.....	34
3.6	Summary .....	36
<b>4.</b>	<b>Sentence Subjectivity and Sentiment Classification .....</b>	<b>37</b>
4.1	Subjectivity Classification.....	38
4.2	Sentence Sentiment Classification.....	41
4.3	Dealing with Conditional Sentences .....	43
4.4	Dealing with Sarcastic Sentences .....	44
4.5	Cross-Language Subjectivity and Sentiment Classification .....	45
4.6	Using Discourse Information for Sentiment Classification .....	47
4.7	Summary .....	47
<b>5.</b>	<b>Aspect-Based Sentiment Analysis .....</b>	<b>49</b>
5.1	Aspect Sentiment Classification.....	50
5.2	Basic Rules of Opinions and Compositional Semantics.....	53
5.3	Aspect Extraction .....	58
5.3.1	Finding Frequent Nouns and Noun Phrases .....	59
5.3.2	Using Opinion and Target Relations .....	61
5.3.3	Using Supervised Learning .....	62
5.3.4	Using Topic Models .....	62
5.3.5	Mapping Implicit Aspects .....	66
5.4	Identifying Resource Usage Aspect .....	67
5.5	Simultaneous Opinion Lexicon Expansion and Aspect Extraction.....	68
5.6	Grouping Aspects into Categories.....	71
5.7	Entity, Opinion Holder, and Time Extraction .....	73
5.8	Coreference Resolution and Word Sense Disambiguation.....	75
5.9	Summary .....	76
<b>6.</b>	<b>Sentiment Lexicon Generation .....</b>	<b>79</b>
6.1	Dictionary-Based Approach.....	80
6.2	Corpus-Based Approach .....	83
6.3	Desirable and Undesirable Facts.....	87
6.4	Summary .....	88
<b>7.</b>	<b>Opinion Summarization.....</b>	<b>91</b>
7.1	Aspect-Based Opinion Summarization.....	91
7.2	Improvements to Aspect-Based Opinion Summarization.....	94

7.3	Contrastive View Summarization.....	95
7.4	Traditional Summarization.....	96
7.5	Summary .....	97
<b>8.</b>	<b>Analysis of Comparative Opinions.....</b>	<b>99</b>
8.1	Problem Definitions .....	99
8.2	Identify Comparative Sentences.....	102
8.3	Identifying Preferred Entities.....	103
8.4	Summary .....	105
<b>9.</b>	<b>Opinion Search and Retrieval.....</b>	<b>107</b>
9.1	Web Search vs. Opinion Search .....	107
9.2	Existing Opinion Retrieval Techniques.....	108
9.3	Summary .....	111
<b>10.</b>	<b>Opinion Spam Detection.....</b>	<b>113</b>
10.1	Types of Spam and Spamming.....	114
10.1.1	Harmful Fake Reviews .....	115
10.1.2	Individual and Group Spamming.....	115
10.1.3	Types of Data, Features, and Detection.....	116
10.2	Supervised Spam Detection.....	117
10.3	Unsupervised Spam Detection .....	120
10.3.1	Spam Detection Based on Atypical Behaviors .....	120
10.3.2	Spam Detection Using Review Graph .....	123
10.4	Group Spam Detection .....	124
10.5	Summary .....	125
<b>11.</b>	<b>Quality of Reviews .....</b>	<b>127</b>
11.1	Quality as Regression Problem.....	127
11.2	Other Methods.....	129
11.3	Summary .....	130
<b>12.</b>	<b>Concluding Remarks.....</b>	<b>133</b>
	<b>Bibliography.....</b>	<b>135</b>
	<b>Author Biography .....</b>	<b>167</b>

# Preface

Opinions are central to almost all human activities and are key influencers of our behaviors. Our beliefs and perceptions of reality, and the choices we make, are, to a considerable degree, conditioned upon how others see and evaluate the world. For this reason, when we need to make a decision we often seek out the opinions of others. This is not only true for individuals but also true for organizations.

Opinions and related concepts such as sentiments, evaluations, attitudes, and emotions are the subjects of study of *sentiment analysis and opinion mining*. The inception and rapid growth of the field coincide with those of the social media on the Web, e.g., reviews, forum discussions, blogs, micro-blogs, Twitter, and social networks, because for the first time in human history, we have a huge volume of opinionated data recorded in digital forms. 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 text mining. In fact, it has spread from computer science to management sciences and social sciences due to its importance to business and society as a whole. In recent years, industrial activities surrounding sentiment analysis have also thrived. Numerous startups have emerged. Many large corporations have built their own in-house capabilities. Sentiment analysis systems have found their applications in almost every business and social domain.

The goal of this book is to give an in-depth introduction to this fascinating problem and to present a comprehensive survey of all important research topics and the latest developments in the field. As evidence of that, this book covers more than 400 references from all major conferences and journals. Although the field deals with the natural language text, which is often considered the unstructured data, this book takes a structured approach in introducing the problem with the aim of bridging the unstructured and structured worlds and facilitating qualitative and quantitative analysis of opinions. This is crucial for practical applications. In this book, I first define the problem in order to provide an abstraction or structure to the problem. From the abstraction,

we will naturally see its key sub-problems. The subsequent chapters discuss the existing techniques for solving these sub-problems.

This book is suitable for students, researchers, and practitioners who are interested in social media analysis in general and sentiment analysis in particular. Lecturers can readily use it in class for courses on natural language processing, social media analysis, text mining, and data mining. Lecture slides are also available online.