

Mining Product Features from Online Reviews

by

Weishu Hu



Master of Software Engineering

2010



**Faculty of Science and Technology
University of Macau**

Mining Product Features from Online Reviews

by

Weishu Hu

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Software Engineering

Faculty of Science and Technology
University of Macau

2010



Approved by _____
Supervisor

Date _____

In presenting this thesis in partial fulfillment of the requirements for a Master's degree at the University of Macau, I agree that the Library and the Faculty of Science and Technology shall make its copies freely available for inspection. However, reproduction of this thesis for any purposes or by any means shall not be allowed without my written permission. Authorization is sought by contacting the author at

Address:
Telephone:
Fax:
E-mail:



University of Macau

Abstract

MINING PRODUCT FEATURES FROM ONLINE
REVIEWS

by Weishu Hu

Thesis Supervisor:
Associate Professor, Zhiguo Gong
Master of Software Engineering

With the advance of the internet, e-commerce systems have become extremely important and convenient to human being. More and more products are sold on the Web, and more and more people are purchasing products online. As a result, an increasing number of customers post product reviews at merchant websites and express their opinions and experiences in any network space such as internet forums, discussion groups, and blogs. So there is a large amount of data records related to products on the Web, which are useful for both manufacturers and customers. Mining product reviews becomes a hot research topic, and existing researches mostly base on product features to analyze the opinions. So mining product features is the number one step to further reviews processing. In this thesis, we present how to mine product features efficiently and accurately. The proposed extraction approach is different from the previous methods because we only mine the features of the product from opinion sentences in which the customers have expressed their positive or negative sentiment. In order to find opinion sentences, a SentiWordNet-based algorithm is proposed. There are three steps to perform our task: (1) Identifying opinion sentences in each review which is positive or negative via SentiWordNet; (2) Mining product features that have been commented on by customers from opinion sentences; (3) Pruning feature to remove these incorrect features. Compared to previous work, our experimental result achieves higher precision and recall. It executes fast enough for practical use.

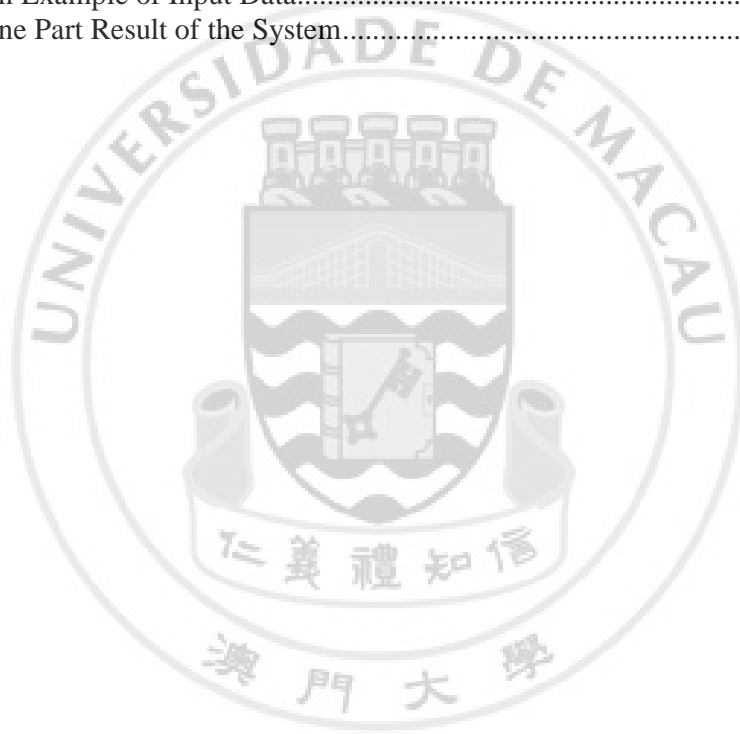
Table of Contents

LIST OF FIGURES	III
LIST OF TABLES.....	IV
LIST OF ABBREVIATIONS.....	V
ACKNOWLEDGMENTS	VI
CHAPTER 1: INTRODUCTION.....	1
1.1 BACKGROUND.....	1
1.2 MOTIVATION.....	3
1.3 OBJECTIVE.....	5
1.4 SOLUTION: SENTIWORDNET-BASED APPROACH.....	6
1.5 CONTRIBUTIONS OF THE THESIS.....	7
1.6 ORGANIZATION OF THIS THESIS.....	8
CHAPTER 2: RELATED WORK	10
2.1 EXISTING RESEARCH ON PRODUCT FEATURES MINING.....	11
2.1.1 MINING PRODUCT FEATURES BY ASSOCIATION RULE.....	12
2.1.2 UNSUPERVISED INFORMATION-EXTRACTION SYSTEM.....	13
2.1.3 APPLY A PROBABILITY-BASED ALGORITHM TO EXTRACT PRODUCT FEATURES.....	14
2.2 OPINION SENTENCE IDENTIFICATION.....	15
2.2.1 EXTRACTING OPINION SENTENCES RELEVANT TO AN OPEN-DOMAIN BLOG.....	16
2.2.2 INTRODUCTION OF OPINIONFINDER.....	17
2.3 CLASSIFYING THE SENTIMENT OF OPINIONS.....	18
2.3.1 DETERMINING THE SENTIMENT OF OPINIONS.....	19
2.3.2 INTRODUCTION OF SENTIWORDNET.....	21
2.4 CONCLUSION OF RELATED WORK.....	23
CHAPTER 3: THE PROPOSED METHOD.....	24
3.1 SYSTEM OVERVIEW.....	25
3.2 PART-OF-SPEECH TAGGING (POS).....	26
3.3 OPINION SENTENCES IDENTIFICATION.....	28
3.4 FEATURES GENERATION.....	32
3.4.1 EXPLICIT FEATURE.....	33
3.4.2 IMPLICIT FEATURE.....	35
3.4.3 FEATURE PRUNING.....	36
CHAPTER 4: EXPERIMENT.....	37
4.1 DATA.....	38
4.2 EVALUATION APPROACH.....	39
4.2.1 PRECISION AND RECALL.....	39
4.2.2 F-MEASURE.....	39
4.3 RESULT AND DISCUSSION.....	40
CHAPTER 5: CONCLUSION AND FUTURE WORK.....	43
BIBLIOGRAPHY	45



LIST OF FIGURES

<i>Number</i>	<i>Page</i>
Figure 1: An Example of Summary	2
Figure 2: A System Architecture	19
Figure 3: The Architect of Product Features Extraction System	25
Figure 4: An Example of Output	25
Figure 5: Head Part of SentiWordNet.....	29
Figure 6: An Example of Input Data.....	38
Figure 7: One Part Result of the System.....	42

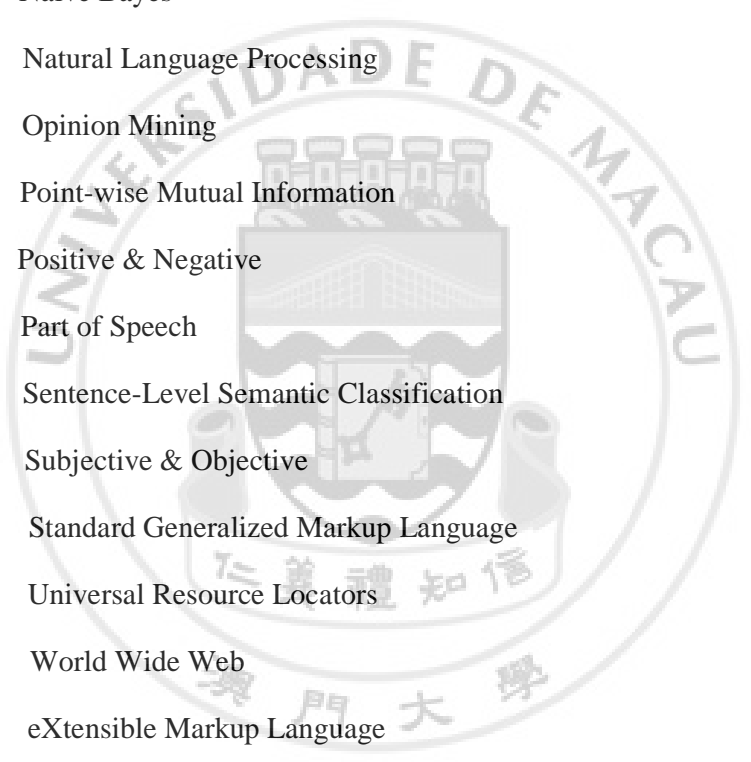


LIST OF TABLES

<i>Number</i>	<i>Page</i>
Table 1: Illustration of the terms for Precision and Recall.....	39
Table 2: Precision and Recall at Opinion Sentences Identification.....	40
Table 3: Comparison of Hu's, Opine's and our results	42



LIST OF ABBREVIATIONS



AMA -	Association Mining Algorithm
CBA -	Classification Based on Associations
JWI -	Java Wordnet Interface
NB -	Naïve Bayes
NLP -	Natural Language Processing
OM -	Opinion Mining
PMI -	Point-wise Mutual Information
PN -	Positive & Negative
POS -	Part of Speech
SLSC -	Sentence-Level Semantic Classification
SO -	Subjective & Objective
SGML -	Standard Generalized Markup Language
URL -	Universal Resource Locators
WWW -	World Wide Web
XML -	eXtensible Markup Language

ACKNOWLEDGMENTS

I would like to convey my sincere appreciation to my supervisor Prof. Zhiguo Gong for his professional guidance, great patience and valuable suggestions throughout the whole process of thesis writing. Prof. Gong always encouraged me when the research could not move forward. He provided much useful idea to me, and also gave instructive tutorial to improve my research capability. Prof. Gong was also an important mentor like a friend. He provided a critical but supportive perspective on my research. He would gently remind me of the importance of conciseness, clarity, readability and proper use of technical terms in writing both papers and thesis.

I acknowledge with gratitude all my research group members, Guangyi Xiao, Yiyang Yang, Haitao Zou, whose cooperation helped me to improve my research. We worked in a cooperative and collegial environment.

I am grateful to other professors in University of Macau, Prof. Enhua Wu, Dr. Jingzhi Guo, Dr. Qiwen Xu, Dr. Si Yain Whar, Dr. Tham Yiu Kwok, Dr. Vong Chi Man, Dr. Xiaoshan LI, Dr. Yan Zhuang, and Dr. Fong Chi Chiu. Their impressed teaching inspired me to make further research during my graduate study.

Last but not least, I wish to give my thanks to my family and my girl friend Ziwei Yang for their support and encouragement all these years.