香港浸會大學
HONG KONG BAPTIST UNIVERSITY

HKBU Scholars

MASTER'S THESIS

Semantic analysis for extracting fine-grained opinion aspects

Zhan, Tianjie

Date of Award:
2010

Link to publication

General rights
Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Download date: 18 Mar, 2025

Semantic Analysis for Extracting Fine-grained Opinion Aspects

ZHAN Tianjie

A thesis submitted in partial fulfillment of the requirements

for the degree of

Master of Philosophy

Principle Supervisor: Dr. LI Chun Hung

Hong Kong Baptist University

November 2010

Abstract

Web documents of user-generated content such as forum discussions and product reviews attracted research interests in opinion mining. Previous works on extraction of opinion aspects and words obtain lexical features based on grammatical rules and WordNet-like lexical resources together with machine learning analyses. Rule based approaches involve in using syntactical and semantic rules to learn the co-occurrence patterns of relevant lexicons for identifying opinion and its aspect. To enhance identification of words' semantic similarity, the WordNet-like lexical database provides semantic relations, e.g. synonyms and hypernyms, for distinguishing distinct opinion aspects and sentiments. But such approaches often represents sentence as "bag of word" or "sequence of word" without fully utilizing semantic dependency relations in a sentence which is essential for extracting semantic relevant words to represent opinion aspects or opinions. To address above problems, this dissertation considers two issues on the topic of extracting fine-grained opinion aspects in sentence level.

The first issue is how to take full advantage of semantic information within a sentence. We introduce a syntactic dependency parser as baseline to develop our sentential semantic parser for learning the composition structure of a sentence. Starting with the atomic sentence which contains only one verb predicate comprising of a subject, main verb and object, compound sentences and complex sentences are parsed into several atomic sentences within a hierarchical layered structure. For example, a compound sentence could be parsed as an atomic sentence with its subject or object as an atomic sentence. Based on the assumption that parts of a atomic sentence can be divided into core parts (subject-verb-object or subject-verb) and other parts for modifying the components of core parts, atomic sentence can be parsed into representation comprising of a <subject, verb, object> triple and a set of head-modifier dependency relations.

The second issue is how to identify distinct finer-grain opinion aspects in discriminative semantic orientation. To address the problem, we focus on identifying discriminative clusters of semantic relevant lexicons to indicate distinct opinion aspects,

among which noun is most representative and essential for describing specific opinion aspect. To develop related noun features and their semantic dependent context features, we extract noun fragments and their semantic neighbor features which are determined by their adjacency relationship with noun fragment within the sentence structure. Given the obtained noun fragments and their semantic neighbors, we propose two approaches to extract finer-grained opinion aspects in an unsupervised manner. The first approach is to employ non-negative matrix factorization (NMF) method on the co-occurrence frequency matrix between noun fragments and their semantic neighbor features. To combine the two views of noun-sentence co-occurrence and association relationship between noun fragment and its semantic neighbors, the second approach takes adjectives as the representative of semantic neighbor features, and a tri-factorization generative model is developed to factorize the corpus along three views of noun fragments, adjectives and sentences to generate semantic relevant clusters for opinion aspect representation.

Finally, the experimental results demonstrate the advantage of our NMF-based approach over the conventional approaches based on noun-sentence co-occurrence frequency matrix. In the second approach, experimental results indicate that our tri-factorization model outperforms the baseline models based on latent Dirichlet allocation. Effectiveness of the two approaches is shown by improved numerical measurement as well as recognition of semantically cohesive fine-grained opinion aspects.

Contents

DECLERATION i

Abstract ii

Acknowledgements iv

1. Introduction 1

1.1 Introduction to Basic Concept, Methodology and Techniques 1

1.1.1 Opinion 1

1.1.2 Opinion Aspect 2

1.1.3 Feature-based Opinion Mining 3

1.2 Research Objective 4

1.3 Organization of this Dissertation 7

2. Background of Opinion Mining 8

2.1 Opinion Mining 8

2.1.1 Tasks, Applications and Challenges of Opinion Mining 8

2.1.2 General Framework in Opinion Mining 9

2.1.3 Approaches to Opinion Mining 11

2.2 Deep Semantic Parsing 12

2.3 Opinion Aspect Extraction 14

2.4 Chapter Summary 14

3. Natural Language Analysis 16

3.1 Natural Language Analysis 16

3.1.1 Part-of-speech Tagging 16
3.1.2 Semantic Dependency Parsing 17
3.2 Opinion Mining with NLP 19
3.2.1 Syntactical Pattern based Approaches 20
3.2.2 Semantic Pattern based Approaches 21
3.3 Chapter Summary 23
4. Sentential Semantic Parsing 24
4.1 Introduction 24
4.2 Sentence Modeling 27
4.3 Sentential Semantic Parsing via Part-of-sentence Recognition 33
4.4 Generative Process 41
4.5 Implementation 45
4.6 Empirical Evaluation 48
4.7 Visualization Graph 52
4.8 Chapter Summary 54
5. Opinion Aspect Mining with Nominal Semantic Structure 55
5.1 Introduction 55
5.2 Problem Definition 56
5.3 Our Nominal Semantic Structure Model 56
5.3.1 Noun Fragment 57
5.3.2 Nominal Semantic Structure 57
5.3.3 Nominal Semantic Neighborhood 58
5.4 Nominal Semantic Structure Parsing 59

5.4.1 Adjective 59
5.4.2 Verb Predicate 59
5.4.3 Preposition 60
5.5 Co-clustering Approach 61
5.5.1 Non-negative Matrix Factorization 61
5.5.2 Clusters Evaluation 62
5.6 Experiments 63
5.7 Related Works 69
5.8 Conclusions 69
6. A Tri-factorization Model for Opinion Aspect Mining 70
6.1 Introduction 70
6.2 Feature Extraction 71
6.3 Our Tri-factorization Generative Model 73
6.4 Inference and Parameter Estimation 75
6.4.1 Latent Variable Inference 76
6.4.2 Parameter Estimation 77
6.4.3 Hyper-parameter Estimation 78
6.5 Evaluation 78
6.5.1 Perplexity 78
6.5.2 Average Cluster Entropy 79
6.5.3 Normalized Mutual Information Index 80
6.6 Experiments 80
6.7 Conclusion 88

7. Conclusions and Future works 89
7.1 Summary 89
7.2 Feature Works 90
Reference 92
CURRICULUM VITAE 98