

MASTER'S THESIS

Conditional random fields based method for feature-level opinion mining and results visualization

Qi, Luole

Date of Award:
2012

[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

Conditional Random Fields based Method for Feature-level Opinion Mining and Results Visualization

Qi Luole

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

Principal Supervisor: Dr. Li Chen

Hong Kong Baptist University

February 2012

Abstract

The tasks of feature-level opinion mining usually include the extraction of product entities from product reviews, the identification of opinion words that are associated with the entities, and the determining of these opinions' polarities (e.g., positive, negative, or neutral). In recent years, two major approaches have been proposed to determining opinions at the feature level: supervised methods such as the one based on lexicalized hidden Markov model (L-HMMs), and unsupervised methods such as rule-based and association rule mining based techniques. However, few attention has been paid to applying more discriminative learning models to achieve the goal. On the other hand, little work has evaluated their algorithms' performance for identifying *intensifiers*, *entity phrases* and *infrequent entities*. In this thesis, we in particular adopt the Conditional Random Fields (CRFs) model to perform opinion mining tasks. Relative to related approaches, we have not only highlighted the algorithm's ability in mining different types of review elements, but also integrated the self-tagging process and the optimization of learning functions so as to further enhance the performance. Our method was compared to both supervised L-HMMs based method and association rule mining based unsupervised methods in the experiments. The results confirm the accuracy of the CRF-based method in accomplishing the opinion mining task from several aspects.

How to present the mining results to the users in an effective way is another issue in this thesis. As we do not know whether users prefer the mining results or the original reviews, we conducted our first user study to find out the answer. Based on the analysis of the collected data, we know that users prefer the extracted

information and they think they can help them make decision when they face several choices. Thus we designed three different interfaces within a comparison matrix to display the extracted information. The three interfaces are opinion cloud interface, barchart interface and table format interface. We also conduct another user study to find out which interface is the best for comparing products. Through the analysis of the experiment results, we conclude that users like the barchart interface most. These findings help us to improve the current interface design and also give us a direction of our future work.

Keywords: User Reviews, Feature-Level Opinion Mining, Conditional Random Fields (CRFs), E-Commerce, Interface, Visualization, Comparison Matrix, User study

Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Tables	viii
List of Figures	x
Chapter 1 Introduction	1
1.1 Background	1
1.2 Our Contributions	5
1.3 Main Findings	6
Chapter 2 Related Work	7
2.1 Document Level Opinion Mining	7
2.2 Feature Level Opinion Mining	8
2.2.1 Unsupervised Approaches	9
2.2.2 Supervised Approaches	11
2.2.3 Compared Approaches	13
2.2.4 Association Rule Based Approach	13
2.2.5 L-HMM Based Approach	15

2.3	Visualization of Opinion Mining Results	16
2.3.1	Document-Level Visualization	16
2.3.2	Feature-Level Visualization	23
Chapter 3	Proposed Approach	27
3.1	CRF based Approach for Opinion Mining	27
3.1.1	Our Problem Statement	27
3.1.2	Pre-Processing	29
3.1.3	Defining Tags and Preparing Training Set	29
3.1.4	Word Expansion and Self-Tagging Process	32
3.1.5	Defining Learning Functions for the CRFs Model	33
3.1.6	Training CRF Model	36
3.1.7	Decoding and Determining Orientations for Opinions and In- tensifiers	37
3.1.8	Determining Opinion Sentence and Sentence Orientation . . .	38
3.2	Experiment Setup and Results	39
3.3	Experiment 1: Testing on Learning Functions and Self-Tagging Process	40
3.3.1	Testing on CRFs Learning Functions	40
3.3.2	Testing on Self-Tagging Process	41
3.4	Experiment 2: Comparing CRFs based Methods with Related Ones .	43
3.4.1	Compared Algorithms in the Experiment	43
3.4.2	Results - Feature-Level	46
3.4.3	Results - Sentence level	50
Chapter 4	Visualization of Opinion Mining Results	59
4.1	Introduction	59
4.1.1	Opinion Cloud Interface	60
4.1.2	Barchart Interface	62
4.1.3	Table format Interface	65

Chapter 5	User Studies	67
5.1	The First User Study	68
5.1.1	Experiment Design and Procedure	68
5.1.2	Participants	71
5.1.3	Result Analysis	73
5.2	The Second User Study	77
5.2.1	Experiment Design and Procedure	77
5.2.2	Participants	79
5.2.3	Result Analysis	80
Chapter 6	Conclusions	84
	Bibliography	86
	Curriculum Vitae	92