

FUZZY BASED IMPLICIT SENTIMENT ANALYSIS ON QUANTITATIVE
SENTENCES

AMIR HOSSEIN YAZDAVAR

UNIVERSITI TEKNOLOGI MALAYSIA

FUZZY BASED IMPLICIT SENTIMENT ANALYSIS ON QUANTITATIVE
SENTENCES

AMIR HOSSEIN YAZDAVAR

A dissertation submitted in partial fulfillment of the
requirement of the award of the degree of
Master of Science (Computer Science)

Faculty of Computing
Universiti Teknologi Malaysia

SEPTEMBER 2013

This dissertation is dedicated to my beloved mother and father for their endless support and encouragement

ACKNOWLEDGEMENT

Praise be to Allah, the Most Gracious and Most Merciful Who has created the mankind with knowledge and wisdom. First and foremost i would like to express my thanks to Almighty ALLAH on the successful completion of this research work and thesis.

Apart from the efforts of myself, the success of any project depends largely on the encouragement and guidelines of many others. I would like to thank my dear wife Monireh for her personal support, kindness and great patience at all times. My parents have given me their unequivocal support throughout, as always, for which my mere expression of thanks likewise does not suffice.

I would like to show my greatest gratitude to my supervisor Professor Dr. Naomie Salim for her advice and guidance provided throughout my study. Her trust, knowledge and friendly personality have always been an inspiration for me and will deeply influence my career and future life. My gratitude and special thanks is extended to my examiners Dr. Alex Sim Tze Xiang and Dr. Roliana Ibrahim for their helpful suggestions and comments.

ABSTRACT

With the rapid growth of social media on the web, emotional polarity computation has become a flourishing frontier in the text mining community. However, it is challenging to understand the latest trends and summarise the state or general opinions about products due to the big diversity and size of social media data and this creates the need of automated and real time opinion extraction and mining. On the other hand, the bulk of currently research has been devoted to study the subjective sentences which contain opinion keyword and limited work has been reported for objective statement that implies sentiment. In this regard, fuzzy based knowledge engineering model has been developed for sentiment classification of special group of such sentences including the change or deviate from desired range or value. Drug reviews are the rich source of such statements. Therefore, in this research, 210 reviews were collected from patient's review for building corpus. These reviews have been selected from different cholesterol lowering drugs. Medical experts cooperated in this research for building Gold standard corpus. Pre-processing operations including extracting medical terms and their corresponding values have been done on this corpus. An appropriate technique has been developed to map each of these medical terms to their corresponding values. Resulted documents were stored into XML file. Determining sentiment polarity of each sentence employing fuzzy knowledge based system is the next step of this research. The main conclusion through this study is, in order to increase the accuracy level of drug opinion mining system, objective sentences which imply opinion should be taken into consideration.

ABSTRAK

Dengan pertumbuhan pesat media sosial di web, terdapat se jumlah besar dokumen mengandungi seperti pendapat ulasan, forum, perbincangan, blog dan twitter yang telah menarik ramai penyelidik bukan sahaja di dalam Pemprosesan Bahasa Tabiee, tetapi juga, sains politik, sains, dan perubatan. Walau bagaimanapun, sebahagian besar masa penyelidikan bertumpu kepada mengkaji ayat-ayat yang subjektif yang mengandungi kata kunci bersifat pendapat. Terlalu sedikit penyelidikan yang telah dilaporkan bagi pernyataan objektif yang menunjukkan sentimen. Dalam hal ini, pengetahuan berasaskan kejuruteraan model kabur telah dibangunkan untuk klasifikasi mengelaskan kepada sentimen kumpulan khas termasuk pertukaran atau perbezaan dari julat nilai yang dikehendaki. Ulasan ubatan adalah sumber yang kaya dengan kenyataan seperti itu. Oleh itu, dalam kajian ini 210 ulasan telah dikumpulkan daripada kajian pesakit untuk membina korpus. Ulasan ini telah dipilih melibatkan ubat untuk merendahkan kolesterol. Pakar-pakar perubatan bekerjasama dalam kajian ini untuk membina korpus piawai emas. Operasi pra-pemprosesan yang telah dilakukan ke atas korpus ini termasuk, mengekstrak terma perubatan dan nilai-nilai yang berkaitan. Satu pendekatan yang sesuai telah direka untuk memetakan setiap terma perubatan kepada nilai nilai yang sepadan. Dokumen yang terhasil disimpan ke dalam fail XML. Sentimen setiap ayat yang menggunakan sistem berasaskan pengetahuan kabur adalah langkah seterusnya dalam kajian ini . Rumusan kajian menunjukkan pemprosesan untuk meningkatkan ketepatan system perlombongan pendapat bagi ubat-ubatan ayat objektif yang membayangkan pendapat perlu diambil kira.

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
	DECLARATION	ii
	DEDICATION	iii
	ACKNOWLEDGEMENT	iv
	ABSTRACT	v
	ABSTRAK	vi
	TABLE OF CONTENTS	viii
	LIST OF TABLES	xii
	LIST OF FIGURES	xiii
	LIST OF ABBREVIATIONS	xv
	LIST OF APPENDICES	xvi
1	INTRODUCTION	1
	1.1 Introduction	1
	1.2 Problem Background	2
	1.3 Problem Statement	4
	1.4 Project Aim	5
	1.5 Project Objectives	5
	1.6 Project Scope	5
	1.7 Significance of Project	6
	1.8 Organization of the Report	6
	1.9 Summary	7
2	LITERATURE REVIEW	8
	2.1 Introduction	8

2.2	Sentiment Analysis and Opinion Mining	8
2.2.1	Document Level Analysis	9
2.2.1.1	Unsupervised Learning Approach	9
2.2.1.2	Supervised Method	10
2.2.2	Sentence Level Analysis	11
2.2.2.1	Subjectivity Classification	11
2.2.2.2	Sentence Sentiment Classification	13
2.2.3	Aspect Based Sentiment Analysis	14
2.2.3.1	Aspect Extraction	14
2.2.3.2	Grouping Aspect into Categories	17
2.2.3.3	Aspect Sentiment Classification	18
2.2.3.4	Entity, Time, Opinion holder	19
2.3	Biomedical text Mining	19
2.3.1	Extracting Information in Medical	20
2.3.1.1	Name Entity Recognition	20
2.3.1.2	Relation Extraction	23
2.3.1.3	Event Extraction	23
2.3.1.4	Knowledge Resources and Tools	24
2.3.2	Fuzzy Application System in Medical Domain	25
2.4	Sentiment Analysis in Medical Domain	26
2.4.1	Previous Work and Challenges	26
2.5	Summary	30
3	METHODOLOGY	31
3.1	Introduction	31
3.2	Operational Framework	31
3.2.1	Planning Phase	35
3.2.2	Collecting Drug Reviews	35
3.2.3	Preparing Performance Measurements	36
3.2.4	Analyzing and Categorizing Numerated Sentences	36

3.2.5	Technique Identification	37
3.2.5.1	GATE Preprocessing Framework	38
3.2.6	Removing Non-numerated Sentences	39
3.2.7	Extracting Quantitative Medical Terms and Values	40
3.2.8	Mapping Values to Terms	40
3.2.9	Storing Opinionated Sentences with their Features	41
3.2.10	Preparing Input Data for Fuzzy System	41
3.2.11	Development of Fuzzy Logic Technique	41
3.2.11.1	Knowledge Acquisition by Interview and Literature	42
3.2.11.2	Structure of Proposed Fuzzy Expert System	45
3.2.11.3	Implementation of Fuzzy Expert System	50
3.2.12	Preparing Data For evaluation	50
3.2.13	Programming of techniques	50
3.2.14	Evaluation	50
3.3	Instrumentation	52
3.4	Summary	53
4	EXPERIMENTAL RESULTS AND DISCUSSION	54
4.1	Introduction	54
4.2	Preprocessing and Intermediate Results and Analysis	54
4.2.1	Building Corpus Results	55
4.2.1.1	Corpus Data Distribution	57
4.2.1.2	Gold Standard Corpus Characteristics	58
4.2.2	Corpus Text Segmentation Results	60
4.2.2.1	English Tokeniser Results	61
4.2.2.2	Sentence Splitter Results	62
4.2.2.3	ANNIE Gazetteer Results	62
4.2.2.2	JAPE Transducer Results	64

4.2.3	Storing Result of Extraction Phase into XML	66
4.2.4	Result of Parsing XML File	67
4.3	Results for Fuzzy Expert System	67
4.3.1	Results of Fuzzification	68
4.3.2	Results of Rule Evaluation	69
4.3.3	Results of Rule Aggregation and Defuzzification	70
4.4	Results of Preparation XML File for Evaluation	70
4.5	Overall Evaluation using Precision & Recall and Analysis	71
4.6	Discussion	76
4.7	Summary	77
5	CONCLUSION	78
5.1	Introduction	78
5.2	Findings	78
5.3	Contribution of Study	79
5.4	Future Work	79
	REFERENCES	81
	Appendices A-G	84-89

LIST OF TABLES

TABLE NO.	TITLE	PAGE
2.1	Accepted ranges of some quantitative medical terms	22
2.2	Different types of numerated sentences in medical reviews	29
3.1	Various types of Statin drugs effectiveness for lowering LDL	42
3.2	Various types of Statin drugs effectiveness for increasing HDL	43
3.3	Various types of Statin drugs effectiveness for lowering Triglyceride	43
3.4	Various types of Statin drugs effectiveness for lowering Cholesterol	44
3.5	Niacin's effect on various cholesterol items	45
3.6	List of used Fuzzy sets	47
3.7	Hardware requirement	52
3.8	List of used software	53
4.1	Corpus characteristics	57
4.2	Polarity distributions over the Gold standard corpus	59
4.3	Various forms of quantitative medical terms employed by patients	63
4.4	Extracted Increasing & decreasing verbs	63
4.5	Confusion matrix for proposed algorithm	72
4.6	Overall precision and recall	72
4.7	A sample of false positive & false negative result	75

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
3.1	Operational Framework	33
3.2	Diagrammatic summary of research methodology	34
3.3	Gate preprocessing framework	39
3.4	Sentiment fuzzy set	49
4.1	A Sample of collected reviews	55
4.2	A Sample of stored reviews in XML file	56
4.3	A Sample of constructed corpus in GATE	56
4.4	Medical terms distribution over the constructed corpus	57
4.5	Drug type distributions over the constructed corpus	58
4.6	A sample of Gold standard preparation process	59
4.7	Polarity Distributions over Different Types of Sentences	60
4.8	GATE processing resources	61
4.9	GATE Tokenization result	61
4.10	A sample of GATE sentence splitter result	62
4.11	A sample of GATE Flexible Gazetteer result	64
4.12	A Sample of regular expression used for mapping each medical term to its corresponding values	65
4.13	A sample of mapping quantitative medical terms to their values result	65

4.14	Distribution of Different Opinionated Sentences Extracted from Corpus	66
4.15	A sample result of extraction phase stored into XML file	67
4.16	A sample of fuzzification step	68
4.17	Result of fuzzification process	69
4.18	A sample of rule evaluation process	69
4.19	A sample of defuzzification process	70
4.20	A sample of annotated sentence with its sentiment stored into XML file	71
4.21	Result of GATE Annotation Diff Tool	72

LIST OF ABBREVIATIONS

NLP	-	Natural Language Processing
SVM	-	Support Vector Machines
POS	-	Part Of Speech
NER	-	Named Entity Recognition
PMI	-	Pointwise Mutual Information
KA	-	Knowledge Acquisition
UMLS	-	Unified Medical Language System
OOP	-	Object Oriented Program
GATE	-	General Architecture for Text Engineering
JAPE	-	Java Annotation Patterns Engine
HMM	-	Hidden Markov Model
CRF	-	Conditional Random Field
API	-	Application Programming Interface
ANNIE	-	A Nearly-New Information Extraction
XML	-	Extensible Markup Language
HMM	-	Hidden Markov Model
COG	-	Center Of Gravity
PSO	-	Particle Swarm Optimization

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A	Sample of Corpus Collection Stored in XML file	84
B	Annotation Guideline	85
C	Sample of annotated drug reviews by medical experts	86
D	List of medical terms mapped to appropriate entity	87
E	Sample of regular expression for mapping medical terms to their corresponding values	88
F	Sample of generated fuzzy rules	89

CHAPTER 1

INTRODUCTION

1.1 Introduction

Opinion mining or in the other word sentiment analysis, is the study of “what the others think”. Long time before web used as a media for transferring information and opinion, people usually ask each other or their friends to make decision through the various issues such as buying product or planning to vote. However nowadays, there is no limitation to ask others opinion since internet and Web provide a vast pool of reviews from millions of people that we even did not know them.

According to one survey cited by Pang et al. (2008) more than 80 percent of Web users have done at least one online search while purchasing a product. However, monitoring and finding the other’s idea might be confusing and overwhelming since finding relevant sites and reliable opinion through a huge volume of opinionated text in each sites, seems to be impossible. Thus, there is a clear need to build an automatic system to help finding and extracting opinions among different entities. We need to take this into consideration that, sentiment analysis is a Natural Language Processing (NLP) problem. Thus, most of the NLP aspects like negation word, coreference resolution problem and word sense disambiguation are involved in here too. As a result, dealing with this problem, require overcoming a number of challenges. In this regard, one of the basic problems involved in sentiment analysis is considering only sentiment word such as good, wonderful, awful and etc is far from sufficient since one word might have different

orientation in various applications. Furthermore, there are sentences that do not contain any sentiment word whereas they imply opinion. Indeed, these sentences state desirable or undesirable factual information. We believe in order to sentiment analysis system achieve next level of accuracy these sentences should be taken into account.

In the present thesis, we investigate special kind of factual sentences which contain quantitative measurement term that implicitly express sentiment.

1.2 Problem Background

According to (Liu, 2010) opinion can be defined as a quintuple $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ where e_i is the name of entity, a_{ij} denotes an aspect of entity e_i , s_{ijkl} shows the sentiment orientation on aspect a_{ij} of entity e_i , h_k is opinion holder (the one who express this idea) and t_l is the time when this idea expressed by its opinion holder. By this definition, sentiment analysis is the task to find all quintuples in the given opinionated document.

Opinions fall within two categories based on the way they are expressed; explicit opinion and implicit opinion. An explicit opinion includes subjective sentences which have opinion keyword. These kinds of statement are easy to detect. The bulk of currently research is devoted to this category.

On the other hand, implicit opinion is an objective statement that implies sentiment. Indeed, they state one desirable or undesirable fact. Limited work has been done in this category. According to Zhang et al., (2011), one special kind of these sentences which is noun and noun phrase that imply opinion has been taken into consideration. However, there exist other approaches which remained unexplored. One group of such approaches is related to change or deviate from desired range or value. Regarding this issue in some application domains, the values of an item have specific properties which denote the sentiment. Indeed, change of

these values to the normal and optimal interval or deviation from the norm range might express positive or negative opinion respectively. For example let the optimal value for total Cholesterol have been defined below than 200 mg/dl, then the sentence “this drug lowered my Cholesterol form 300 to 190” bear positive sentiment since the Cholesterol decreased into optimal value. And in sentence “don’t take this drug, it puts my Blood pressure into 18” implicit opinion by changing from normal range has been expressed.

Furthermore, based on our observation, significant changes might also denote sentiment even the new value would not place in a normal range. For instance in the sentence “it dropped my Cholesterol level from 580 into 250”, although the second value of Cholesterol would not place in the optimal value (below than 200mg/dl) but the sentence express positive sentiment. On the other hand, the sentence “it increased my Cholesterol level from 250 into 580” denotes negative sentiment polarity, thus it is important to consider change direction.

It is worth mentioning to recall that, among these kinds of sentences, there exist some kinds of sentences that do not express any sentiment while containing changes in numeric values. As an example “my doctor changes the normal dosage of Welchol from 624mg to 300mg” is a factual sentence that should be grouped into non-opinionated sentence. In addition, there are many numeric fields which do not show any sentiment and need to be filtered e.g. “Have never felt so bad, like a 100 year old woman (I am 63)”.

In the present thesis, we investigate numerated sentences, not only to categorize them into opinionated and non-opinionated but also to determine whether they contain positive or negative sentiment polarity.

1.3 Problem Statement

Determining sentiment orientation from sentences have been studied by many researchers, considering different method to handle this problem although a large portion of their effort have been devoted to subjective statement that contain sentiment word, objective sentences which contain desirable or undesirable fact can imply sentiment too. One group of these statements that remains unexplored to date, contains numeric values denote change and deviation from normal range. Drug reviews are the rich source of such statement. In this regard, one special characteristic of such sentences is related to certain degree of uncertainty and imprecision involved in them. In the light of our observation, this uncertainty can be regarded from different point of view. First, a large portion of quantitative medical term associated with predefined ranges which try to identify a patient's status e.g. for Total Cholesterol, below than 200mg/dl considered as desirable, between 200 and 239 is borderline high and greater than 240 defined as high. Thus, these terms can be accepted as fuzzy.

Second, changes in the value of these quantitative terms, might demonstrate improvement, stable condition, and exacerbation of a patient, regardless of where the second value placed. For example, the sentence "my total Cholesterol dropped from 580 to 240" shows improvement in total Cholesterol while the second value of it (240) placed into high range. Intuitively, changes can be grouped into slight, medium and high increased or decreased which can be denoted by fuzzy set theory.

Furthermore, patient's sentiment might be regarded as positive, neutral and negative based on the factors that have been stated, second numeric value and changes, side effect, opinion words and etc.

Consequently, fuzzy logic is an ideal choice to deal with this problem. Therefore, the present study tries to determine sentiment polarity of numerated sentence by employing fuzzy set theory. There are four main issues which can achieve the goal of this study; what is the method to extract numeric variable in a sentence? How this numbers related to their specific entities? Which numerated

sentence bear a sentiment? And what is the sentiment orientation of opinionated sentence?

1.4 Project Aim

The work aims to develop fuzzy algorithm to classify numerated factual statement contain implicit opinion.

1.5 Project Objectives

This study aim to accomplish the following objectives;

- I. To extract quantitative aspects and their corresponding values.
- II. To determine opinionated sentences through objective sentences that contained numerated facts.
- III. To develop fuzzy rule based decision system for the purpose of identifying sentiment orientation of numerated opinionated objective sentences that contain implicit opinion.
- IV. To evaluate suggested algorithm using precision and recall measures.

1.6 Project Scope

- I. The corpus collection will consist from 210 drug reviews collected from website: www.askapatient.com
- II. 110 reviews will be used for generating fuzzy rules by expert.

- III. 100 reviews will be utilized for building test set for determining sentiment orientation.
- IV. The corpus collection will be stored in XML file.
- V. GATE will be employed to annotate numbers and extract medical terms.
- VI. This thesis will apply GATE JAPE Grammar for mapping extracted quantitative medical terms to their associated values.
- VII. Fuzzy rules will be formulated from doctors for inference step.
- VIII. Fuzzy expert system approach will be taken to accomplish sentiment classification.

1.7 Significance of Project

This study represent a positive step toward implicit opinion mining to achieve higher accuracy system by exploiting fuzzy set theory and natural language processing techniques.

1.8 Organization of the Report

This study is organized into five chapters. Chapter1 represents the introduction of the study, problem background, project aim, scope, objectives and the significance of the project. Chapter 2 discusses the previous work and the literature review. Chapter 3 explains the methodology employed in this thesis. Chapter 4 represent the experimental results and finally conclusion and findings are shown in Chapter 5.

1.9 Summary

In conclusion, by rapid growth of social media on the web, we confront a huge volume of opinionated documents such as reviews, forum, discussion, blogs and twitters which have been attracted many researchers due to its integral role in not only Natural Language Processing, but its effect on medical, political and social science. However, there are many challenges which have not been solved and require more research. In this study, we deal with one of this problem to increase the accuracy level of existing system.

REFERENCES

- Abbod, M. F., D. G. von Keyserlingk, D. A. Linkens and M. Mahfouf (2001). Survey of utilisation of fuzzy technology in medicine and healthcare. *Fuzzy Sets and Systems* 120(2): 331-349.
- Aronson, A. R. (2001). Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proceedings of the AMIA Symposium*, American Medical Informatics Association.
- Bhatia, R. S., A. Graystone, R. A. Davies, S. McClinton, J. Morin and R. F. Davies (2010). Extracting information for generating a diabetes report card from free text in physicians notes. *Proceedings of the NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents*, Association for Computational Linguistics.
- Capuzzi, D. M., J. R. Guyton, J. M. Morgan, A. C. Goldberg, R. A. Kreisberg, O. Brusco and J. Brody (1998). Efficacy and safety of an extended-release niacin (Niaspan): a long-term study. *The American journal of cardiology* 82(12): 74U-81U.
- Cui, H., V. Mittal and M. Datar (2006). Comparative experiments on sentiment classification for online product reviews. *Proceedings of the National Conference on Artificial Intelligence*, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.
- Cunningham, H. (2002). GATE, a general architecture for text engineering. *Computers and the Humanities* 36(2): 223-254.
- Ding, X., B. Liu and P. S. Yu (2008). A holistic lexicon-based approach to opinion mining. *Proceedings of the international conference on Web search and web data mining*, ACM.

- Hai, Z., K. Chang and J. Kim (2011). Implicit feature identification via co-occurrence association rule mining. *Computational Linguistics and Intelligent Text Processing*: 393-404.
- Hu, M. and B. Liu (2004). Mining and summarizing customer reviews. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM.
- Jakob, N. and I. Gurevych (2010). Extracting opinion targets in a single-and cross-domain setting with conditional random fields. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics.
- Jin, W., H. H. Ho and R. K. Srihari (2009). A novel lexicalized HMM-based learning framework for web opinion mining. *Proceedings of the 26th Annual International Conference on Machine Learning*, Citeseer.
- Jones, P. H., M. H. Davidson, E. A. Stein, H. E. Bays, J. M. McKenney, E. Miller, V. A. Cain and J. W. Blasetto (2003). Comparison of the efficacy and safety of rosuvastatin versus atorvastatin, simvastatin, and pravastatin across doses (STELLAR Trial). *American Journal of Cardiology* 92(2): 152-160.
- Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of Natural Language Processing*: 627-666.
- Liu, B. and L. Zhang (2012). A survey of opinion mining and sentiment analysis. *Mining Text Data*: 415-463.
- Pang, B. and L. Lee (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, Association for Computational Linguistics.
- Pang, B. and L. Lee (2008). *Opinion mining and sentiment analysis*. Now Pub.
- Rosier, A., A. Burgun and P. Mabo (2008). Using regular expressions to extract information on pacemaker implantation procedures from clinical reports. *AMIA Annual Symposium Proceedings*, American Medical Informatics Association.
- Sarawagi, S. (2008). Information extraction. *Foundations and trends in databases* 1(3): 261-377.
- Simpson, M. S. and D. Demner-Fushman (2012). Biomedical text mining: A survey of recent progress. *Mining Text Data*: 465-517.

- Su, Q., X. Xu, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen and Z. Su (2008). Hidden sentiment association in chinese web opinion mining. *Proceedings of the 17th international conference on World Wide Web*, ACM.
- Swaminathan, R., A. Sharma and H. Yang (2010). Opinion mining for biomedical text data: Feature space design and feature selection. *the Ninth International Workshop on Data Mining in Bioinformatics (BIOKDD 2010)*.
- Turney, P. D. (2002). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, Association for Computational Linguistics.
- Yalamanchi, D. (2011). *Sidedefective-system to mine patient reviews: sentiment analysis*. Rutgers University-Graduate School-New Brunswick
- Yu, H. and V. Hatzivassiloglou (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. *Proceedings of the 2003 conference on Empirical methods in natural language processing*, Association for Computational Linguistics.
- Yu, J., Z. J. Zha, M. Wang and T. S. Chua (2011). Aspect ranking: identifying important product aspects from online consumer reviews. *Computational Linguistics*: 1496-1505.
- Zhang, L. and B. Liu (2011a). extracting resource terms for Sentiment analysis. *Proceedings of iJcnLP-2011*.
- Zhang, L. and B. Liu (2011b). Identifying noun product features that imply opinions. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers*.
- Zhou, X., H. Han, I. Chankai, A. Prestrud and A. Brooks (2006). Approaches to text mining for clinical medical records. *Proceedings of the 2006 ACM symposium on Applied computing*, ACM.
- Zhou, X., H. Han, I. Chankai, A. A. Prestrud and A. D. Brooks (2005). Converting semi-structured clinical medical records into information and knowledge. *Data Engineering Workshops, 2005. 21st International Conference on*, IEEE.
- Zhuang, L., F. Jing, X. Y. Zhu and L. Zhang (2006). Movie review mining and summarization. *Conference on Information and Knowledge Management: Proceedings of the 15 th ACM international conference on Information and knowledge management*.