

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.