

Fuzzy Based Implicit Sentiment Analysis on Quantitative Sentences

Amir Hossein Yazdavar^{a,*}, Monireh Ebrahimi^a, Naomie Salim^a

^aFaculty of Computing, Universiti Teknologi Malaysia, Johor, Malaysia

* Corresponding author email address: yazdavar@gmail.com

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 summarize 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 current research has been devoted to study the subjective sentences which contain opinion keywords and limited work has been reported for objective statements that imply sentiment. In this paper, fuzzy based knowledge engineering model has been developed for sentiment classification of special group of such sentences including the change or deviation from desired range or value. Drug reviews are the rich source of such statements. Therefore, in this research, some experiments were carried out on patient's reviews on several different cholesterol lowering drugs to determine their sentiment polarity. The main conclusion through this study is, in order to increase the accuracy level of existing drug opinion mining systems, objective sentences which imply opinion should be taken into account. Our experimental results demonstrate that our proposed model obtains over 72% F1 value.

Keywords: Text mining, Natural language processing, Sentiment analysis, Fuzzy set theory

1. Introduction

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 and Lee, 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 about different topics.

According to (Liu, 2015) 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 keywords. These kinds of statement are easy to detect. The bulk of current research is devoted to this category. On the other hand, implicit opinion is implied in objective statements. Indeed, they state one desirable or undesirable fact (Kim and Hovy, 2004). Limited work has been done in this category. One special kind of these sentences which is noun and noun phrase that imply opinion has been taken into consideration by (Zhang and Liu, 2011). However, they give a general approach for all domains whereas opinion mining (especially implicit opinion mining) is a highly domain and context dependent task (Rushdi et al., 2011) and to reach good and applicable results it would be beneficial to exploit some domain knowledge and features. One group of such domain dependent problems is related to change or deviation from desired range or value. In some application domains, the values of an item have specific properties which denote the sentiment. Indeed, change of these values to the 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" implies 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 deviating from normal range has been expressed. Furthermore, based on our observation, significant changes might also denote