Original

Patient opinion mining to analyze drugs satisfaction using supervised learning

Vinodhini Gopalakrishnan *, Chandrasekaran Ramaswamy

Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar 608002, India

Abstract

Opinion mining is a very challenging problem, since user generated content is described in various complex ways using natural language. In opinion mining, most of the researchers have worked on general domains such as electronic products, movies, and restaurants reviews but not much on health and medical domains. Patients using drugs are often looking for stories from patients like them on the internet which they cannot always find among their friends and family. Few studies investigating the impact of social media on patients have shown that for some health problems, online community support results in a positive effect. The opinion mining method employed in this work focuses on predicting the drug satisfaction level among the other patients who already experienced the effect of a drug. This work aims to apply neural network based methods for opinion mining from social web in health care domain. We have extracted the reviews of two different drugs. Experimental analysis is done to analyze the performance of classification methods on reviews of two different drugs. The results demonstrate that neural network based opinion mining approach outperforms the support vector machine method in terms of precision, recall and *f*-score. It is also shown that the performance of radial basis function neural network method is superior than probabilistic neural network method in terms of the performance measures used.
© 2017 Universidad Nacional Autónoma de México, Centro de Ciencias Aplicadas y Desarrollo Tecnológico. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Sentiment; Opinion; Health; Drugs; Classification

1. Introduction

Natural language processing (NLP) is a set of computational techniques for analyzing natural language texts that allows computers to understand human language. Opinion mining discipline places itself at the crossroads of information retrieval and computational linguistics. These are the two fields from which opinion mining gathers and combines many concepts, ideas and methods (Yu & Hatzivassiloglou, 2003; Wilson, Wiebe, & Hoffmann, 2005; Xia, Zong, & Li, 2011; Ye, Zhang, & Law, 2009). The primary unit of NLP is the language term. Each language term has various linguistic features, such as grammatical category, meaning, sense, co-occurrence similarity, and contextual relationships that are employed for term classification and subjectivity analysis. The NLP tasks require a

knowledge base for information extraction and analysis. While some techniques are necessary for building a knowledge base, other techniques use existing knowledge bases to analyze documents. Researchers have presented great contributions in this area and diverse approaches have been employed to accomplish the opinion mining (Miao, Li, & Dai, 2009; Tang, Tan, & Cheng, 2009).

Opinion mining has led to development of a significant number of tools that are used to analyze consumer opinion for most major business environments including travel, housing, consumables, materials, products, services, and many others. But not much work was carried out on health care domain (Tang et al., 2009; Vinodhini & Chandrasekaran, 2014; Xu, Liao, Li, & Song, 2011). Drug surveillance is a major factor of drug safety once a drug has been released to the public use. Drug trials are often done in limited number of test subjects where the probability to detect uncommon adverse effects is minimal. Also the volunteers or patients participating in drug trials are also different from those receiving licensed medications, differing in age, co-morbidity and poly-pharmacy. Thus it is necessary to study

* Corresponding author.
   E-mail address: g.t.vino@gmail.com (V. Gopalakrishnan).

the safety of marketed drugs on an epidemiological scale. It is also important to understand how the general population uses a particular drug, perceive its safety, reactions and efficiency. The objective of this work is to use neural network methods for opinion classification in health care domain.

This paper is outlined as follows. Section 2 narrates the related work. Section 3 discusses the methodology used to develop the models. The data source used is reported in Section 4. The various classification methods used are introduced in Section 5. Sections 6 and 7 discuss the results. Section 8 concludes the work.

2. Literature review

Many studies on opinion classification have used machine learning algorithms with support vector machine (SVM) being the most commonly used. SVM has been used extensively for opinion miming on online reviews (Baccianella, Esuli, & Sebastiani, 2009; Jian, Chen, & Wang, 2010; Pang, Lee, & Vaithyanathan, 2002; Tan & Zhang, 2008; Tan & Zhang, 2008; Xia et al., 2011; Ye, Lin, & Li, 2005; Ye et al., 2009). Some research works exists for opinion mining in medical domain using machine learning techniques. Greaves, Ramirez-Cano, Millett, Darzi, and Donaldson (2012) analyzed patient's opinions about different performance aspects of hospitals in the United Kingdom. Ali, Sokolova, Schramm, and Inkpen (2013) used a subjectivity lexicon and machine learning algorithms to sentiment analysis of messages posted on forums dedicated to hearing loss. Sokolova and Bobicev (2011) used supervised machine learning methods to analyze sentiments and opinions expressed in health related user written texts. They analyzed opinions posted on a general medical forum. Sokolova and Bobicev (2011) presented a mining method for personal health information in Twitter. Wang, Chen, Tan, Wang, and Sheth (2012) proposed a combination of machine learning and rule based classifiers for opinion classification in suicide notes.

There are also very few studies specifically on opinion mining in the drug domain. Na et al. (2012) proposed a rule-based system for polarity classification of drug reviews. Yalamanchi (2011) developed a system called SidEffective to analyze patient's sentiments about a particular drug. Goeuriot et al. (2012) built a health-related sentiment lexicon and used it for polarity classification of drug reviews. Wiley, Jin, Hristidis, and Esterling (2014) used SentiWordNet and word-emotion lexicons for opinion classification of drug reviews of online social networks. Sharif, Zaffar, Abbasi, and Zimbra (2014) developed a framework that extracts important semantic and sentiment, that are better able to analyze the experiences of people when they discuss adverse drug reactions as well as the severity and the emotional impact of their experiences. Xia, Gentile, Munro, and Iria (2009) proposed a multi-step approach, where in the initial

step a standard topic classifier is learned from the data and the topic labels, and in the ensuing step several polarity classifiers, one per topic, are learned from the data and the polarity labels.

2.1. Motivation and contributions

In recent years, we witnessed the advance in neural network methodology, like fast training algorithm for deep multilayer neural networks (Chen, Liu, & Chiu, 2011; Ghiassi, Skinner, & Zimbra, 2013; Jian et al., 2010; Luong, Socher, & Manning, 2013; Moraes, Valiati, & Neto, 2013; Noferesti & Shamsfard, 2015; Sharma & Dey, 2012). Neural network techniques have found success in several NLP tasks recently such as opinion mining (Bobicev, Sokolova, Jafer, & Schramm, 2012; Socher, Lin, Manning, & Ng, 2011). Socher et al. (2011) also introduced prediction architecture based on recursive neural networks that can be used to provide a competitive syntactic parser for natural language sentences from the Penn Treebank.

However, the literature does not contribute much work in opinion classification using neural networks, especially probabilistic neural network (PNN) and radial basis function neural network (RBFN). But many researchers have proved that PNN and RBFN model is more effective than other models for data classification in various other domains (Adeli & Panakkat, 2009; Hajmeer & Basheer, 2002; Savchenko, 2013).

This motivates us to investigate the effectiveness of neural networks in multi class opinion classification. Based on the literature survey done, the major contributions of our work are as follows, PNN and RBFN based on function approximation are used for opinion classification which is not considered so far in opinion mining literature. The results obtained for neural network based method are compared with SVM as baseline method. We have applied opinion mining to health care domain, particularly focusing on public opinions on drugs.

3. Problem design

The following is the summary of our methodology for developing and validating the prediction models (Fig. 1).

  i. Perform data preprocessing.
 ii. Construct the vector space representation for the drug reviews selected for drug I and drug II.
iii. Apply the following classification methods using the respective training data set
    a. SVM.
    b. The neural network model using PNN.
    c. The neural network model using RBFN.
 iv. Predict the class (positive or negative) of each review in the test data set.
  v. Compare the prediction results with actual values.

Fig. 1. Problem design.