An Unsupervised Aspect Detection Model for Sentiment Analysis of Reviews

Ayoub Bagheri¹, Mohamad Saraee², and Franciska de Jong³

¹ Intelligent Database, Data Mining and Bioinformatics Lab, Electrical and Computer Engineering Department, Isfahan University of Technology, Isfahan, Iran
a.bagheri@ec.iut.ac.ir
² School of Computing, Science and Engineering, University of Salford, Manchester, UK
m.saraee@salford.ac.uk
³ University of Twente, Human Media Interaction, P.O. Box 217, 7500 AE Enschede, The Netherlands
f.m.g.dejong@utwente.nl

Abstract. With the rapid growth of user-generated content on the internet, sentiment analysis of online reviews has become a hot research topic recently, but due to variety and wide range of products and services, the supervised and domain-specific models are often not practical. As the number of reviews expands, it is essential to develop an efficient sentiment analysis model that is capable of extracting product aspects and determining the sentiments for aspects. In this paper, we propose an unsupervised model for detecting aspects in reviews. In this model, first a generalized method is proposed to learn multi-word aspects. Second, a set of heuristic rules is employed to take into account the influence of an opinion word on detecting the aspect. Third a new metric based on mutual information and aspect frequency is proposed to score aspects with a new bootstrapping iterative algorithm. The presented bootstrapping algorithm works with an unsupervised seed set. Finally two pruning methods based on the relations between aspects in reviews are presented to remove incorrect aspects. The proposed model does not require labeled training data and can be applicable to other languages or domains. We demonstrate the effectiveness of our model on a collection of product reviews dataset, where it outperforms other techniques.

Keywords: sentiment analysis, opinion mining, aspect detection, review mining.

1 Introduction

In the past few years, with the rapid growth of user-generated content on the internet, sentiment analysis (or opinion mining) has attracted a great deal of attention from researchers of data mining and natural language processing. Sentiment analysis is a type of text analysis under the broad area of text mining and computational intelligence. Three fundamental problems in sentiment analysis are: aspect detection, opinion word detection and sentiment orientation identification [1-2].

Aspects are topics on which opinion are expressed. In the field of sentiment analysis, other names for aspect are: features, product features or opinion targets [1-5].
Aspects are important because without knowing them, the opinions expressed in a sentence or a review are of limited use. For example, in the review sentence “after using it, I found the size to be perfect for carrying in a pocket”, “size” is the aspect for which an opinion is expressed. Likewise aspect detection is critical to sentiment analysis, because its effectiveness dramatically affects the performance of opinion word detection and sentiment orientation identification. Therefore, in this study we concentrate on aspect detection for sentiment analysis.

Existing aspect detection methods can broadly be classified into two major approaches: supervised and unsupervised. Supervised aspect detection approaches require a set of pre-labeled training data. Although the supervised approaches can achieve reasonable effectiveness, building sufficient labeled data is often expensive and needs much human labor. Since unlabeled data are generally publicly available, it is desirable to develop a model that works with unlabeled data. Additionally due to variety and wide range of products and services being reviewed on the internet, supervised, domain-specific or language-dependent models are often not practical. Therefore the framework for the aspect detection must be robust and easily transferable between domains or languages.

In this paper, we present an unsupervised model which addresses the core tasks necessary to detect aspects from review sentences in a sentiment analysis system. In the proposed model we use a novel bootstrapping algorithm which needs an initial seed set of aspects. Our model requires no labeled training data or additional information, not even for the seed set. The model can easily be transform between domains or languages. In the remainder of this paper, detailed discussions of existing works on aspect detection will be given in section 2. Section 3 describes the proposed aspect detection model for sentiment analysis, including the overall process and specific designs. Subsequently we describe our empirical evaluation and discuss important experimental results in section 4. Finally we conclude with a summary and some future research directions in section 5.

2 Related Work

Several methods have been proposed, mainly in the context of product review mining [1-14]. The earliest attempt on aspect detection was based on the classic information extraction approach of using frequently occurring noun phrases presented by Hu and Liu [3]. Their work can be considered as the initiator work on aspect extraction from reviews. They use association rule mining (ARM) based on the Apriori algorithm to extract frequent itemsets as explicit product features, only in the form of noun phrases. Their approach works well in detecting aspects that are strongly associated with a single noun, but are less useful when aspects encompass many low-frequency terms. The proposed model in our study works well with low-frequency terms and uses more POS patterns to extract the candidates for aspect. Wei et al. [4] proposed a semantic-based product aspect extraction (SPE) method. Their approach begins with preprocessing task, and then employs the association rule mining to identify candidate product aspects. Afterward, on the basis of the list of positive and negative opinion