7th International conference on Intelligent Human Computer Interaction, IHCI 2015

Semi-supervised Aspect Based Sentiment Analysis for Movies using Review Filtering

Deepa Anand[a]*, Deepan Naorem[a]

[a]*CMR Institute of Technology,AECS Layout, Bangalore 560048, India*

Abstract

Aspect based Sentiment Analysis (ABSA) is a subarea of opinion mining which enables one to gain deeper insights into the features of items which interest the users by mining reviews. In this paper we attempt to perform ABSA on movie review data. Unlike other domains such as camera, laptops restaurants etc, a major chunk of movie reviews is devoted to describing the plot and contains no information about user interests. The presence of such narrative content may potentially mislead the review analysis. The contribution of this paper is two-fold: a two class classification scheme for plots and reviews without the need for labeled data is proposed. The overhead of constructing manually labeled data to build the classifier is avoided and the resulting classifier is shown to be effective using a small manually built test set. Secondly we propose a scheme to detect aspects and the corresponding opinions using a set of hand crafted rules and aspect clue words. Three schemes for selection of aspect clue words are explored - manual labeling (M), clustering(C) and review guided clustering (RC). The aspect and sentiment detection using all the three schemes is empirically evaluated against a manually constructed test set. The experiments establish the effectiveness of manual labeling over cluster based approaches but among the cluster based approaches, the ones utilizing the review guided clue words performed better.

*Keywords:*Aspect Based Sentiment Analysis; Feature Based Opinion Mining; Naive Bayes; Natural Language Processing

1. Introduction

The web provides an excellent platform for users to express their views and opinions on a gamut of topics varying from products in e-commerce, political reforms they desire to their feelings on various day to day happenings. Free flow reviews allow the users to be much more expressive than plain rating systems by providing the freedom to articulate the aspects about the topic which is most important to them. The various forms of such user expression in the form of free flowing text manifesting in various forms such as blogs, reviews, and tweets provide a goldmine of

information which can be exploited in various ways, for product Recommendations, text summarization etc and has given rise to the broad area of research known as Opinion Mining [1].

Sentiment Analysis is a sub area of Opinion Mining which aims at inferring the emotion of a user towards a particular item expressed through the reviews. Though sentiments could fall under several classes such as anger, misery, happiness etc, the primary emotions of interest in most cases is the positivity or negativity of the review. Plenty of research has gone into accurate classification of reviews[1][2] under these categories but a variation of sentiment analysis called aspect based or feature based sentiment analysis[1][3][4] delves deeper to understand the reason for positive or negative orientation and investigates the item features for which the user has expressed an opinion. For instance, in the movie review snippet given below one can infer that the user is happy with choice of cast but not the story.

... the might of Tom Hanks and Julia Roberts could not salvage this horrible joke of a story ...

Many techniques have been proposed for the problem using a combination of common sense knowledge, based on Natural Language Techniques normally encoded as rules, Supervised and unsupervised techniques[9][12]. The common sense knowledge encoded as rules may be domain dependent or domain independent. ABSA has been attempted in various domains such as laptops [3], restaurants [9], movies [6][7] etc. Though there exists labeled dataset for applying supervised techniques in the laptop and restaurant domains such a dataset is lacking in the movie domain and thus we attempt an approach in this direction. We use a combination of domain knowledge along with semi/un supervised techniques to build an aspect and sentiment detector. An important step in achieving this is to tackle the problem which is unique to the movie domain i.e. reviews containing narrative content which may mar the accurate computation of the actual polarity of the review and its aspects. However creation of manually labeled data for filtering out plot sentences is time consuming and we outline an approach to overcome this hurdle.

To detect aspects and sentiments we employ hand crafted rules based on the work presented in [6][9] to detect potential sentiment aspect pairs. The ABSA task is then two-fold: detect the aspects (map the potential aspect words detected to the corresponding aspect categories) occurring in the sentence and estimate the polarity of the user towards these aspects. To detect the aspect category we make use of words which would be indicators of the presence of the aspect - we call such words aspect clue words. Three approaches for selecting the clue words are attempted here: manual, semi-supervised clustering and review word guided semi-supervised clustering. Whereas the former requires manual effort and needs to be tailored for a domain the clustering approaches require minimal intervention. We thus justify the effort expended in manual selection by evaluating the performance improvement it offers over un/semi supervised approaches. We also demonstrate that a review word based supervised clustering is able to offer better performance than the method without.

The paper is structured as follows: the next section discusses the relevant background for ABSA; Section 3 presents the proposed approach whereas section 4 presents the experimental results. Section 5 concludes the work and describes the future enhancements planned.

2. Literature Survey

The ease of generating plain text content in different forms varying from blogs and tweets to reviews, discussion forum posts etc have opened endless possibilities in the ways in which these can be utilized. Ecommerce companies tap into the goldmine of user generated content available aplenty, in order to serve customers better and boost their sales. Sentiment Analysis [2] and particularly ABSA [1], which poses several challenges inherent to processing natural language, is an active area of research in this direction. Several challenges which still have not been adequately addressed are implicit aspect detection, mapping aspect words to categories, resolving anaphora references etc [1]. Researchers combine techniques from common sense rules, unsupervised and supervised techniques to perform these tasks.

The lack of labeled data has lead to several researchers to explore unsupervised learning techniques to learn both aspects and their sentiments expressed in plain text. Particularly the fact that aspects are normally described by opinion words and opinion words in turn will have a target aspect can be used to iteratively expand the sentiment and aspect lexicon. The expansion is done with the help of rules to associate aspects and sentiments [1]. Jo and Oh