

Final Project Report

Xiao Liu, Wenxiang Zheng

December 13, 2014

1 Abstract

This paper is the final report of our project for Sentiment Analysis on movie reviews. In the introduction section, we review our proposal and literature reviews, and we also presents the layout of this final report. We provide some basic definitions and concepts that are related to such research topic in this section as well. In addition, the paper provides a description of the dataset that will be tested in our research project. In the Baseline Research Methods section, we address some baseline methods for Aspect-based Sentiment analysis and ranking prediction in our project, and we provide some advantages and disadvantages for each method along with relavant examples. In the Experiment section, we present the testing resutls for each of the methods that has been tested in the final experiment. Finally, this paper concludes with a discussion about some of the problems that we faced during final experiment and the future work.

2 Introduction

In the first few weeks of this semester, we had been very interested in Sentiment Analysis. After our discussions and literature reviews, we decided to focus on Sentiment Analysis on movie reviews for our term project.

In our research proposal, we proposed the importance of Sentiment Analysis on movie reviews to both movie producers and customers, and we discussed how an automatic Sentiment Analysis application can help the movie review system. We also described the dataset that will be examined for this

data mining project, and presented some candidate research questions for our term project.

After discussing our research questions and the direction of this research with professor, we have spotted the research question that we will work for throughout the semester for our term project. In the literature review, we present our research question and discuss some potential Sentiment Analysis methods from previous researches that can be implemented for our term project. In general, there are two major Sentiment Analysis methods: Sentiment Classification and Aspect-based Sentiment Analysis. After further discussion with the professor about our research topic and research method, we decided to work on Aspect-based Sentiment Analysis on movie reviews.

In the Project Midterm Report, we addressed the well-defined research question, and demonstrated the baseline methods for this project in Sentiment Analysis on movie reviews. The base line methods are divided into three categories: Review Ranking Predictions, Aspect Extraction, and Aspect-Based Sentiment Analysis.

In this Final Project Report, we present the results from final experiment for ranking predictions, in which we center on Naive Bayes as baseline method and use Bag of Words, Bi-gram, Uni-Bi-gram, and Stop Words as features to evaluate the performance of the experiment. Following the results for ranking prediction, we show the results for aspect extraction and explain the methods that have been implemented to improve aspect extraction. We also address five major errors, which we considered to have higher possibilities in affecting the accuracies in ranking predictions. We conclude the final report with some of the challenges in final experiment, especially the challenge for us to finish the test on Sentiment Analysis for each aspect, some preliminary ideas for future work, and a summary for what we have learned from this project.

3 Definitions

3.1 Basic Definitions

- Sentiment Analysis: are computational studies of opinions, sentiments, subjectivity, evaluations, attitudes, appraisal, affects, views, and emotions.
- Sentiment Classification: are deployed both on the document level and

the sentence level. As indicated by the term classification, it is basically a text classification problem that classifies a whole opinion document/sentence based on the overall sentiment of the opinion holder.

- Aspect-based Sentiment Analysis: are referred to determining the opinions and sentiments expressed on different features or aspects of entities.

3.2 Data Set Description

We grab the data set from Stanford SNAP group and it is a data set of Amazon movie reviews crawled from the web. For each instance, it contains the information as shown in Table ???. The data span a period of more than 10 years, including all 8 million reviews up to October 2012. Reviews include product and user information, ratings, and a plaintext review. We also have reviews from all other Amazon categories. There are in total 7911684 instances in this data set. The reviews are collected from 253059 users for 253059 movies and among these reviewers, 16341 of them have reviews more than 50 pieces. The reviews' quality are pretty high as the median number of words per review is 101 [mcauley2013amateurs].

Table 1: Dataset

Name	Example
product/productId	B00006HAXW
review/userId	A1RSDE90N6RSZF
review/profileName	Joseph M. Kotow
review/helpfulness	9/9
review/score	5.0
review/time	1042502400
review/summary	Pittsburgh - Home of the OLDIES
review/text	I have all of the doo ...

4 Research Questions

After observing the dataset together with some basic literature review, in this project, we are more interested in making predictions and analysing the

common features based on this dataset. We have the following well-defined candidate question:

- How to generate aspect-based summaries for each movie.

5 Related Works

Sentiment Analysis and Opinion Mining are computational studies of opinions, sentiments, subjectivity, evaluations, attitudes, appraisal, affects, views, emotions, etc., expressed in text [pang2008opinion]. Since the sentiment analysis has been a hot topic over years, there are a number of publications in this research area. Meanwhile, a number of opinion mining applications in the market, such as the opinion observe [liu2005opinion] that conducts analysis on the cellphone reviews; Aspect-based opinion summary for both the Bing search engine and Google product search [blair2008building]; tools like OpinionEQ¹ that integrates a few sentiment analysis functions; and live track of movies that predicts user ratings from the Twitter posts [thet2010aspect]. In this report of literature review, we report the related works in two perspectives: (1) sentiment classification, and (2) aspect-based sentiment analysis.

Before stepping into the first branch, we formalize the definitions of some related terms. *Opinions* are those words expressed one's feeling to an object. In one piece of opinion, there are opinion targets, features, sentimental positive or negative, opinion holder as well as the time [dale2000handbook]. For example, in the sentence "Alex bought a Cannon camera two weeks ago, and he loves it because the pictures are beautiful and high quality."

- (Target: Cannon camera; Features: picture quality; Sentimental pos or neg: positive"love"; Opinion holder: Alex; Time: two weeks ago)

Usually, we use the quintuples to describe an opinion to make the unstructured data into structured data [liu2007web].

5.1 Sentiment Classification

Sentiment classifications are deployed both on the document level and the sentence level. As indicated by the term classification, it is basically a text classification problem that classifies a whole opinion document/sentence,

¹<http://www.opinioneq.com>.

based on the overall sentiment of the opinion holder [turney2002thumbs; pang2002thumbs]. Obviously, for a classification problem, both the **unsupervised** and **supervised** learning are adopted.

- **Unsupervised:** Unsupervised methods derive a sentiment metric for text without training corpus, and it has been widely used, since the early time when this topic was first introduced. It is a fascinating problem for researchers to study; however, the sentiment classification is hard to deploy in the real research and experiment as there are many potential challenges in this method. Turney [turney2002thumbs] predicates the sentiment orientation of a review by the average semantic orientation of the phrases in the review that contain adjectives or adverbs, which is denoted as the semantic oriented method. They use three steps in this unsupervised classification: POS tags, Sentiment orientation(SO) estimation of the extracted phrases, and Average SO computing. Kim and Hovy [kim2004determining] build three models to assign a sentiment category to a given sentence by combining the individual sentiments of sentiment-bearing words. Hiroshi [hiroshi2004deeper] use the technique of deep language analysis for machine translation to extract sentiment units in text documents. Devitt and Ahmad [devitt2007sentiment] explore a computable metric of positive or negative polarity in financial news text.
- **Supervised:** Supervised methods consider the sentiment analysis task as a classification task and use labeled corpus to train the classifier. In majority, three classification techniques are tried: Naive Bayes, Maximum entropy, and Support vector machine. A few features cater to the researchers are *term frequency*, *POS tag*, *opinion words and phrases*, *negations*, *syntactic dependency*, etc. Since the work of Pang et al [pang2002thumbs], various classification models and linguistic features have been proposed to improve the classification performance. Mullen and Collier [mullen2004sentiment]; Wilson et al. [wilson2005recognizing]. Most recently, McDonald et al. [titov2008joint] investigate a structured model for jointly classifying the sentiment of text at varying levels of granularity. Blitzer et al. [blitzer2007biographies] investigate domain adaptation for sentiment classifiers, focusing on online reviews for different types of products. Andreevskaia and Bergler [andreevskaia2007clac] present a new system consisting of the ensemble of a corpusbased classifier and a lexicon-based classifier with precision-based vote weighting.

This Sentiment Classification method has been well-studied and implemented in many research projects; however, the potential challenges to deploy such method are also obvious. After reviewing related literatures in this topic, the Sentiment Classification method has the following limitations:

- The Sentiment Classification work for only one object in the document or sentence
- This method cannot extract different opinions
- This method cannot correctly extract indirect/unobvious opinions
- this method does not work for comparison reviews

For example, in the sentence "We bought the car last month and the windshield wiper has fallen off." There are two targets mentioned in this sentence, car and windshield wiper, and the opinion identification is unobvious. The Sentiment Classification method cannot detect whether an opinion towards the car or the windshield wiper in this sentence.

5.2 Aspect-based Sentiment Analysis

Sentiment classification method at both the document and sentence levels is quite useful; however, it does not find out what people like or dislike. In this case, another branch of Sentiment Analysis called Aspect-based Sentiment Analysis emerges. This method extracts entities and aspects (target, feature, opinion, and time) from documents.

To extract the entities, some methods are considered by researchers: Distributional similarity [jo2011aspect](which compares the surrounding text of candidates using cosine or PMI), PU learning [liu2002partially](which learns from positive and unlabeled examples), and bayesian sets [heller2005bayesian].

To extract the aspects, Liu first introduces a frequency-based method in 2004 [hu2004mining] because he considers the reviews from different people are irrelevant. When aspects/features are discussed, the words used converge. Later on, various improved methods are applied based on the first one. Zhuang et al [zhuang2006movie] improve the recall due to loss of infrequent aspects by using opinion words to extract the aspects; Popescu and Etzioni [popescu2005opine] improve the precision by removing the

frequent noun phrases that may not be aspects using part-of its relationship; Qiu [qiu2011opinion] applies the double propagation (DP) approach, which uses dependency of opinions and aspects to extract both aspects and opinion words.

5.3 Discussion

According to the data set that we obtain, the meta-data of the data set is well-structured and intuitive so that we know the name of review commenter, movie name, time of the review, and the review itself from the database. After considering the advantages and disadvantages of the two sentiment analysis methods, we believe that the Aspect-based Sentiment Analysis method is the most suitable method to deploy in our research topic. This method can help us extracting different opinions in the review as reviewers usually have different opinions for different parts of a movie. In addition, this method can help us identifying comparison opinions and unobvious/indirect opinions in the review. Although the Aspect-based Sentiment Analysis method seems to match the database that we use in the following research, there are several potential problems that we need to consider:

- Identify different comparative and implicit opinions
- Identify reviewer's emotions
- Measurement of the level of opinions that matches related ratings

6 Baseline Research Methods

In order to answer the research question, there are three steps in Aspect-based Sentiment Analysis:

- Ranking Prediction using Naive Bayes
- Aspect Extraction using Double Propagation
- Aspect Synonym Deduction using WordNet
- Aspect Sentiment Classification using Naive Bayes

6.1 Ranking Prediction by Naive Bayes

We use Naive Bayes for movie view ranking prediction. In order to obtain different results for performance evaluation and receive the most accurate prediction, we test the method by five different features. The features been tested are:

- Bag of Words
- Bi-gram
- Bag of Words and Bi-gram
- Bag of Words with Stop Words Out
- Bag of Words on JJ, JJS, VB (Stop Words Out)

6.2 Aspect Extraction by Double Propagation

According to literature reviews about this research topic, there are three major methods for Aspect Extraction: Frequent, Infrequent, and Double Propagation. After considering the advantages and disadvantages of all three methods for our research project on movie reviews, which contains massive words regarding aspects, entities, and opinions, we decided to use Double Propagation method as our baseline method for Aspect Extraction.

Frequent method is the easiest way to apply on Aspect-based Sentiment Analysis; however, there can be so many infrequent words in movie reviews toward the movies casts, content, and features, it will be difficult to extract those infrequent aspects when using Frequent method.

The Infrequent method provides better precision with smaller drop in call. The major parts of the Infrequent methods are part-of-relationship and the Web with PMI score. The mathematical expression of the Infrequent is given as $PMI(word1, word2) = \log((p(word1word2)/[(p(word1)*p(word2))])$. However, the Infrequent method is not ideal when considering movie reviews, since it is generally domain dependent, where a lot of times the aspects will approximate with the nearest noun to the opinion word. For example, in the example The iPhone 6 is very good, but the I dont like the iPhone6s camera, the aspect extraction can be less accurate because in the second part of the sentence, the opinion word is nearest to iPhone6 instead of camera.

The Double Propagation method, which is a bootstrapping method, could be the ideal method for us to mining the movie reviews, since this method uses dependency of opinions and aspects to extract both opinion and aspect words with domain independent method. It inputs a set of seed opinion words, while it doesn't need aspect seeds.

6.3 Identify Aspect Synonym by WordNet

We use the WordNet method for this step. Generally, the WordNet sees every two synonyms as the same term, which means word A and word B will be regarded as synonyms only if there is a synset containing A and B that appear in the top two senses of both words. There are some clustering features in the WordNet, including lexical similarity, distributional information, syntactical constrains.

6.4 Aspect Sentiment Classification by Naive Bayes

We use Naive Bayes for Aspect Sentiment Classification. There are three steps in sentence classification: Part-of-speech (POS) tagging, Sentiment Orientation of extracted phrases, and computation of the average SO of all phrases. The mathematical expression for the second step is:

- $SO(\text{phrase}) = PMI(\text{phrase}, \text{excellent}) - PMI(\text{phrase}, \text{poor})$

The bigger SO means that the opinion of the aspect is positive. After computing the average SO for all phrase, the method returns the overall SO, which means the overall opinion for a certain aspect is positive or negative. $PMI(\text{top-notch plot}, \text{excellent}) = 2.780$ $PMI(\text{top-notch plot}, \text{poor}) = 0.110$ $SO(\text{top-notch plot}) = 2.670$ $PMI(\text{really unforgettable}, \text{excellent}) = 0.629$ $PMI(\text{really unforgettable}, \text{poor}) = 0.237$ $SO(\text{really unforgettable}) = 0.392$ $\text{mean}(SO) = (2.670 + 0.392) / 2 = 3.062$ $\therefore \text{threshold} = 3.062$ \therefore positive

7 Final Experiment

7.1 Preprocessing

8 Ranking Prediction

8.1 Bag of Words

8.2 Bi-gram

8.3 Bag of Words and Bi-Gram

8.4 Bag of Words with Stop Words Out

8.5 Bad of Words on JJ,JJS,VB (Stop Words Out

8.6 Result Analysis

9 Aspect Extraction

10 Discussion

11 Future Plan

We are going to step-by-step realize the functions of the system we proposed. Since we would finally make an aspected-based summary for each movie in the dataset, the schedule of our plan is as follows:

Table 2: Schedule

Step	Function
1	Aspect Extraction using Double Propagation
2	Apsect Synonym Deduction using WordNet
3	Aspect Sentiment Classification using Naive Bayes
4	Some Visualizations if time permits

12 Contribution

Xiao contributed on research topic and data set selections, and she contributed to most of the related work part of this report. Wenxiang contributed to most of the writeup of this paper, and he also contributed on proofreading and editing the paper.