

Product Feature Discovery and Ranking for Sentiment Analysis from Online Reviews

Nitish Gupta¹, Shashwat Chandra²
Advisor: Dr. Amitabha Mukerjee²
{gnitish,chandras,amit}@iitk.ac.in

¹ Dept. Of Electrical Engineering

² Dept. Of Computer Science and Engineering

November 11, 2013

1 Introduction

An important task of review mining is to extract people's opinions and sentiments on features of a product. Eg. The sentence '*The phone has a good battery life.*' expresses positive review about the '*battery life*' of the '*phone*'. Here '*battery life*' is a feature of the product '*phone*' and the user expresses positive review about it. In an unsupervised environment, extraction of features is the most important and difficult aspect of review mining. Feature Ranking on the extracted feature-candidates is important for a twofold reason, one to increase the precision of the top-ranked candidates and the second to find those features which the users consider the most important while reviewing a particular product. Sentiment Analysis will help in judging how good or bad the users consider a product and due to what reasons.

2 Related Work

A lot of work on extraction and ranking of product features in opinion documents has been done in [1], [2], [3] and [4]. *Double Propagation*[1] is a state-of-the-art unsupervised algorithm to find the features of a product. It makes use of the fact that most of the reviews have a particular dependancy grammar in them which can be exploited to find the features in an unsupervised manner. For ranking of features [2] propose to use the well-known webpage ranking algorithm HITS.

3 Our Approach

To achieve the first and foremost task of extracting the product for which reviews are given, we propose to use the part-whole relation. Eg. In the sentence '*engine of the car*', '*engine*' is a 'part' of the 'whole' product '*car*'. Also in '*screen of the phone*', the '*whole*' aspect '*phone*' is the product for which the review is given.

Due to the fact that the features of a product are nouns, noun phrases and the opinion words modifying them are adjectives we propose to use a very naive frequentist approach along with the *Double-Propagation* algorithm to extract features. Feature Ranking will be achieved by comparing the frequency with which the different features occur in the reviews and also the frequency of the opinion words which are used along with them.

4 Dataset

We will use the dataset given on Prof. Bing Liu's page. We will also create our own corpus on online product reviews from '*Flipkart*'.

References

- [1] Qiu, Guang, et al. "Opinion word expansion and target extraction through double propagation." *Computational linguistics* 37.1 (2011): 9-27.
- [2] Zhang, Lei, et al. "Extracting and ranking product features in opinion documents." *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. Association for Computational Linguistics, 2010.
- [3] Liu, Bing. "Sentiment analysis and opinion mining." *Synthesis Lectures on Human Language Technologies* 5.1 (2012): 1-167.
- [4] Zhai, Zhongwu, et al. "Clustering product features for opinion mining." *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 2011.