Feature Extraction from Micro-blogs for Comparison of Products and Services

Peng Zhao, Xue Li, and Ke Wang
Nanjing University, 210093 China,
The University of Queensland, QLD 4072 Australia,
Simon Fraser University, Vancouver, Canada
zp10@software.nju.edu.cn,
xueli@itee.uq.edu.au,
wangk@cs.sfu.ca

Abstract. Social networks are a popular place for people to express their opinions about products and services. One question would be that for two similar products (e.g., two different brands of mobile phones), can we make them comparable to each other? In this paper, we show our system namely OpinionAnalyzer, a novel social network analyser designed to collect opinions from Twitter micro-blogs about two given similar products for an effective comparison between them. The system outcome is a structure of features for the given products that people have expressed opinions about. Then the corresponding sentiment analysis on those features is performed. Our system can be used to understand user’s preference to a certain product and show the reasons why users prefer this product. The experiments are evaluated based on accuracy, precision/recall, and F-score. Our experimental results show that the system is effective and efficient.

Keywords: social network, feature extraction, opinion mining, sentiment analysis.

1 Introduction

In recent years, social networks are becoming a popular place for people to express their opinions about social-economical issues or on products and services. Twitter, as a representative of social networks, is a microblogging service [16]. A tweet is a post or status update on Twitter. There has been a significant increase of daily posted tweets in recent years: 50 million in 2010, 200 million in 2011, and 400 million in 2012.

The direct implication of this trend is that it is difficult for consumers to compare two similar products or services that offer similar functions but with different properties. Many studies have been done on analyzing consumer reviews from forums [6] [5] [10] [13]. Bing Liu et al [9] explored opinion features and they analyzed certain types of customers’ reviews, but their paper didn’t go further to conduct sentiment analysis for the features of the products and services. Popescu and Etzioni in [13] conducted the product feature extraction and the sentiment
analysis. However, their approach is based on the known consumer reports that are written by experts and this kind of second-hand information can be biased or misleading. They did not use the features discussed by the first-hand consumers in social networks. Both [6] and [13] worked on consumer reviews, their work would provide only qualitative assessment rather than quantitative analysis on consumer opinions toward those product features. To the best of our knowledge, our work is the first of this kind that directly compare and contrast two similar products or services based on the social network analysis.

In this paper, we approach the feature extraction problem based on the text in tweets for the given products or services. There are three questions to be answered, 1) what features can we extract from tweets about the given products? 2) what are people’s opinions about the products based on the features extracted? and 3) how do we make recommendations based on the sentiment analysis of those extracted features? For the first question, the products are made up with some features that people would like to comment, so the first question turns out to be a feature extraction problem. For the second question, people would like, love, or even hate the products because of some features of the products. People’s emotion regarding the features of products can be divided into three kinds: positive, negative, or neutral. Thus the second question becomes a sentiment analysis problem. As for the third question, recommendations can be made if the overall statistical information can be made available for all of those features identified over the given products. So we conduct comparisons on the given products not only by offering recommendations but also by explaining why people like them, based on the sentiment analysis on the extracted features.

It should be pointed out that the physical features of a product can be easily obtained from product specifications where a product is advertised. Here, we do not want to get product features in this way because there are many features that consumers do not care and there are features that might not be listed in product specifications, for which consumers may like to comment. Therefore for a particular product, features should be chosen by consumers in anyway they prefer. We call the user-preferred features as hot features.

There are many feature extraction algorithms such as those given in Weka [17] like Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) [7], and Latent Dirichlet Allocation (LDA) [1]. An insightful review on feature extraction from text can be found in Hu and Liu [6] and an overview on effective feature extraction approaches can be found in Guyon et al [4]. Sentiment analysis is an approach used to categorize the overall attitude of a sentence or a paragraph towards a certain subject [13] [20].

In this paper, a new algorithm based on the formal concept analysis [3] is proposed to solve the feature extraction problem in our OpinionAnalyzer system. The difference between our feature extraction algorithm and the other feature extraction algorithms is that the feature space of a product or service is often hierarchically structured and we need an approach that will extract features in a lattice structure with a partial order, so to make features comparable to each other. As far as we know, this is the first work that extracts features and