Chapter 6
Mining Product Reviews in Web Forums

S. Hariharan
Pavendar Bharathidasan College of Engineering and Technology, India

T. Ramkumar
A.V.C. College of Engineering, India

ABSTRACT
Internet has brought a major drift in user community. Apart from its well-known usage, it also promotes social networking. Research on such social networking has advanced significantly in recent years which have been highly influenced by the online social websites. People perceive the web as a social medium that allows larger interaction among people, sharing of knowledge, or experiences. Internet or social web forums act as an agent to reproduce some general information that would benefit the users. A product review by the user is a more accurate representation of its real-world performance and web-forums are generally used to post such reviews. Though commercial review websites allow users to express their opinions in whatever way they feel, the number of reviews that a product receives could be very high. Hence, opinion mining techniques can be used to analyze the user-reviews, classify the content as positive or negative, and thereby find out how the product fares. This paper focuses its attention on providing a recommendation to the products available on the web by analyzing the context to score the sentences for each review by identifying the opinion and feature words using a novel algorithm.

1. INTRODUCTION

Internet has been considered as a medium to exchange information or knowledge. Due to the rapid growth of information online, almost every activity in human life has become automated. For instance consider a user, who wishes to purchase a music system or digital camera. The user can sit comfortably in his place and browse for the best products through online websites which provides about each products. Such commercial sites allow people to express their opinions on the products that they have purchased. Hearst (1992) and Wiebe (1994) originally proposed the idea of mining direction-based text, namely, text containing opinions, sentiments, affects, and biases.

The current scenario is that, social web-forums allows each individual to post review which may be a discussion groups or blogs which enables the new user to buy product based on the views.
of expressed by different users. These reviews are useful for both customers and manufacturers. Opinions or sentiments about the products can be analyzed about these products and their performances can be analyzed effectively. Opinions are hidden in long forum posts and blogs, which allows human to find relevant sources, summarize them and organize them into usable forms. Sun et al. (2009) proposed the system for sentimental classification and comparison of products from both subjective and objective perspective on various feature levels.

To make decisions about online reviews, opinions are so important and we should take some special interest in mine such positive and negative opinions. Review comments from the sites are considered as valuable because they cover a lot more products than those formal review sites. Facts and opinions are the categories of textual information. All entities and events in the world have objective statements that are called facts. On contrary subjective statements that reflect people’s sentiments or perceptions are called opinion. Recent research on such opinion mining and sentimental analysis have been performed based on unsupervised learning approach (Popescu & Etzioni, 2005), analyzing lexical features and their relationships (Riloff et al., 2006), sentimental word based scheme (Cai et al., 2008) and frequency of occurrence of product features (Wong & Lam, 2005).

Opinion or sentiment bearing words (e.g., great, amazing, wonderful, bad, and poor) are mainly used in the research on opinion mining. As per the review, work on mining such opinion bearing words and to identify their semantic orientations (i.e., positive or negative) helps both merchants and buyers (Su et al., 2008). Since a product may have hundreds or even thousands of reviews it is difficult for buyers to choose right decision and also for manufacturer to keep track of change if products of different kinds are produced. Hu and Liu (2004) aimed to mine and to summarize all the customer reviews of a product in a system. Only those opinions that are given by customers as positive or negative for each product is summarized; this makes the task different from traditional text summarization.

User generated contents are created and published by the end users who surf the web on a daily basis. Examples of such online documents are blogs, newsgroup postings and discussion forums. There are two types of textual information available on the web namely facts and opinions. Facts are objective expressions about entities, events and their properties. Facts can be expressed with topic keywords. Currently available search engines search only for facts and they are not appropriate for opinion retrieval or search. But the user generated content on the web such as personal experiences and opinions about a product or a movie in the form of reviews play a very important role in business, education, e-commerce, etc. Online review websites allow users to express their opinions for the information they are interested in. So there is a huge amount of information available online, however they fail to provide the knowledge about the products.

Customers could assess a product by reading opinions of other customers, which will help them to decide whether to purchase the product or not. On the other hand, the manufacturers can improve their products by tracking the feedback from the customers. With the emergence of e-commerce, the number of reviews that a product receives is quite high and many reviews are lengthy with only few sentences containing the actual opinions on the product. It is not easy for a customer to read all the reviews in order to decide on whether to buy the product or not. If the customer reads only a few reviews, then the opinion might be biased. Manufacturers also find it difficult to analyze all the reviews and to interpret the customer opinions. This paper attempts to provide a methodology to mine the reviews available in web forums and to express the opinions from them.

The rest of this paper is organized as follows: In Section 1 we have exposed the web as a medium of information gathering. Section 2 focuses on