Chapter 10
Mechanism for Crawling, Filtering, and Presenting Opinionated Content on Online Products to the Customers

Rosy Madaan
CSE Department, FET, Manav Rachna International Institute of Research and Studies, India

ABSTRACT

There is a large amount of data available on the web in form of opinions, which need to be accessed for mining opinions. This is an ever-growing field that brings together the reviews, blogs, discussions on forums, Twitter, microblogs, and social networks. A user may be looking for opinions on some commodity or product for making decision regarding purchase for which there is the need of a system based on question answering. This gives rise to a question answering (QA) system. This system works on all the aspects of question answering along with the mining of opinions. The chapter discusses all the modules of the question answering system along with how the opinions are mined. The details of implementation along with the performance analysis of the proposed system are given in the chapter. On performance evaluation, a high value of opinion accuracy has been found that shows that the system performs well.

DOI: 10.4018/978-1-6684-5255-4.ch010

INTRODUCTION

The information present in form of text on the WWW can be classified broadly into two major categories namely facts and opinions (Sasikal and Mary, 2020). Facts are statements written about events and the entities existing in the world in an objective form. Opinions are the statements written in an objective form that reflect sentiments (Hu and Liu, 2004) or perceptions of people about the real world events and entities. Much of the existing research on processing the textual information present on the WWW has been focussed on how factual information can be mined and retrieved. The related areas are information retrieval, Web search, natural language processing and text mining (Boiy and Moens, 2008). This is the era of huge digital content on the web in form of opinions and a little work has been done in the direction of processing the opinions. However, opinions are so important that whenever a person needs to make a decision about purchase of any item, he/she discusses with other customers (Popescu and Etzioni, 2007) and collect their opinions. This is applicable for both the individuals and the organizations.

In order to gather opinions about its products, a company needs to conduct surveys, organize focused groups and to employ external consultants. Also, the companies need to collect opinions of its competitors. The task of finding opinion sources and monitoring them on the Web is actually a formidable task because there exist a big number of diverse sources existing on the WWW and a huge volume of information is present on each source. Classifying an opinion is a very important aspect of opinion mining which has been studied widely in the NLP community. Opinion classification is defined as "determining whether an opinion on an object is positive, negative or neutral". For example, the system classifies the reviews gathered on movies into three type's i.e. positive, negative or neutral reviews. This type of problem is clearly a classification learning problem. In opinion classification, the words which are related to the topic are not important. Instead, the words that indicate polarity of an opinion (Bhatia, Sarma, and Bhatia, 2015) as positive, negative or neutral are important. The words excellent, great, amazing, worst, horrible, bad, worst, etc. can be used to decide the polarity. Most of the techniques for opinion classification apply some forms of machine learning techniques. The whole task of presenting opinions (Bhatia, Sarma, and Bhatia, 2018) to the user involves finding relevant sources, extract sentences comprising of opinions, summarizing them and organizing them into usable forms. This calls for the need of an automated system that takes a user's question as input and presents opinion(s) as answer(s). Such a system is termed as Question Answering (QA) system.

The paper has been organized in the following manner: section 2 discusses recent research carried out in this area; section 3 presents the proposed system for question