Chapter 13
Open Issues in Opinion Mining

Vishal Vyas
Pondicherry University, India

V. Uma
Pondicherry University, India

ABSTRACT

Opinions are found everywhere. In web forums like social networking websites, e-commerce sites, etc., rich user-generated content is available in large volume. Web 2.0 has made rich information easily accessible. Manual insight extraction of information from these platforms is a cumbersome task. Deriving insight from such available information is known as opinion mining (OM). Opinion mining is not a single-stage process. Text mining and natural language processing (NLP) is used to obtain information from such data. In NLP, content from the text corpus is pre-processed, opinion word is extracted, and analysis of those words is done to get the opinion. The volume of web content is increasing every day. There is a demand for more ingenious techniques, which remains a challenge in opinion mining. The efficiency of opinion mining systems has not reached the satisfactory level because of the issues in various stages of opinion mining. This chapter will explain the various research issues and challenges present in each stage of opinion mining.

INTRODUCTION

Opinion mining (OM)/Sentiment Analysis (SA) is related to deriving insight through analysis of user’s thoughts (reviews, posts, blogs etc.) about entities such as products, movies, people etc. Evaluation of reviews posted by users on e-commerce and social networking is of much use as it contains highly rated information.

DOI: 10.4018/978-1-5225-6117-0.ch013
Calculation of average inclination of opinion towards any entity not only helps businesses organizations to gain profits but also helps an individual in getting the right opinion about something unfamiliar.

Natural Language Processing (NLP) deals with actual text element processing. The text element is transformed into machine format by NLP. Artificial Intelligence (AI) techniques are applied on information provided by NLP to determine whether text sentence is positive or negative. Text mining can also be used in extracting the opinion. The difference is that, in text mining, data mining techniques are used to identify the opinion.

In both the techniques, content from the text corpus is pre-processed and opinion word extraction is performed before the opinion is derived. The raw text contains unwanted words which have no contribution to the opinion and such words are removed in preprocessing. The clean data is the output of the preprocessing stage. Various issues such as noise removal, missing values etc. are to be dealt in preprocessing stage. Many methods are available for pre-processing of text in opinion mining. The time complexity involved in pre-processing is high as compared to proceeding stages. High usage of abbreviations is normal when it comes to publishing more information using fewer characters. For instance in Twitter, only 280 characters are allowed in single post. Though various acronym dictionaries such as netlingo, urban dictionary are available, to deal with emerging slangs is a difficult task. Real-time updation of such dictionaries is the need of the hour.

Considering the next stage in opinion mining, the objective to create a predictive model for opinion mining can be fulfilled effectively by proper feature selection. Better feature selection not only produces accurate results but it also reduces the time complexity. In-depth knowledge of the problem domain is a prerequisite for feature selection. Filter, wrapper and embedded methods are used for feature selection in text mining. Presently, selection of a feature is a big issue, as the orientation of opinion changes with respect to the domain. Opinion mining is not limited to textual data but extending it to data in different formats such as real-time video, audio etc. is a real challenge.

Analysis being the next stage in OM, with what sentiment the author of the text is giving the opinion is identified through classification of the text. The chapter discusses the following challenges involved in classification of text for opinion mining. Researchers mainly use online reviews on movies and products for opinion mining. It is hard to identify whether the content is authentic or fake. Singh (2018) discussed a model to classify whether the review is authentic or fake. For better opinion mining, the elimination of spam content is necessary. The chapter discusses issues in the application of opinion mining for spam detection. Sentiment detection