Chapter 88
Machine Learning Approaches for Sentiment Analysis

Basant Agarwal
Malaviya National Institute of Technology, India

Namita Mittal
Malaviya National Institute of Technology, India

ABSTRACT

Opinion Mining or Sentiment Analysis is the study that analyzes people's opinions or sentiments from the text towards entities such as products and services. It has always been important to know what other people think. With the rapid growth of availability and popularity of online review sites, blogs', forums', and social networking sites' necessity of analysing and understanding these reviews has arisen. The main approaches for sentiment analysis can be categorized into semantic orientation-based approaches, knowledge-based, and machine-learning algorithms. This chapter surveys the machine learning approaches applied to sentiment analysis-based applications. The main emphasis of this chapter is to discuss the research involved in applying machine learning methods mostly for sentiment classification at document level. Machine learning-based approaches work in the following phases, which are discussed in detail in this chapter for sentiment classification: (1) feature extraction, (2) feature weighting schemes, (3) feature selection, and (4) machine-learning methods. This chapter also discusses the standard free benchmark datasets and evaluation methods for sentiment analysis. The authors conclude the chapter with a comparative study of some state-of-the-art methods for sentiment analysis and some possible future research directions in opinion mining and sentiment analysis.

1. INTRODUCTION

It has always been important to know what other people think. With the rapid growth of popularity and availability of online review sites, blogs, forums, and social networking sites necessity of analysing and understanding these reviews has arisen. Companies and people can use the opinion given in these reviews for better decision making for example a user can know about pros and cons of various features of the products that can help in taking decision of purchasing items. E-commerce companies can use

the users' opinion for improving their product quality and to know the current trends. Opinion Mining or Sentiment Analysis is the study that analyse people's opinion, sentiment towards entities such as products, services etc. in the text (Liu, 2012). Sentiment analysis research can be categorized among Document level, Sentence level and Aspect/Feature level sentiment analysis. Document level sentiment analysis classifies a review document as positive or negative sentiment polar document. It considers a document as a single unit. Sentence level sentiment analysis takes a sentence to extract the opinion or sentiment expressed in that sentence. Aspect based sentiment analysis deals with the methods that identify the entities in the text about which an opinion is expressed (Liu, 2012). Further, the sentiments expressed about these entities are identified. Other important tasks in sentiment analysis and opinion mining research are opinion summarisation, opinion retrieval, spam review detection, etc. Solutions for the challenges incurred in these problems come from NLP, cognitive science, information retrieval, machine learning etc. Sentiment analysis research challenges and existing solutions are nicely presented in detail (Liu, 2012). Detailed survey of sentiment analysis research through various techniques is presented by Liu, (2012) and Cambria et al. (2012). However, these surveys do not discuss in detail about the machine learning approaches for sentiment analysis specifically that is the objective of this chapter. Machine learning approaches have been widely applied to sentiment classification mostly for document level sentiment classification.

There are several challenges being faced in the sentiment analysis research. Firstly, words used for expressing sentiment are domain specific. For example, word "*unpredictable*" has positive orientation in movie review domain but may be negative oriented for car review domain. Secondly, to identify the subjective portion of text from the overall review, because same words can be used in subjective and objective sentences. For example, "*author used very crude language*" and "*crude oil is extracted from sea beds*". In this example, same word crude is used for expressing sentiment in first sentence, however second sentence is purely objective (Verma et. al 2009). Thirdly, thwarted expectations are difficult to handle. In certain cases most of the text represent positive or negative polarity, out of sudden polarity of overall text is reversed. For example, "*This film has a great cast. It has excellent storyline and nice cinematography. However it can't hold up the audiences*". Most of the review analysis research is based on the movie reviews and product review. Movie review sentiment classification face the challenge of handling the real facts which is generally mixed with actual review data. People generally discuss about the general traits of actors, plot of movie and relate the movie to their normal life. It is very difficult to extract the opinion from the reviews when there is a discussion of characteristics of artist and in the end overall movie is disliked. One of the biggest challenges of movie review analysis is to handle the negated opinion. Product review domain significantly differs from movie review dataset. In product reviews, reviewer generally writes both positives and negative opinion, because some features of the product are liked and some are disliked. These types of reviews are difficult to classify into positive or negative class. Generally, product review dataset contains more comparative sentences than movie review dataset, which is difficult to classify.

Main approaches for sentiment analysis can be categorized into Semantic Orientation based approaches (Turney 2002, Agarwal et al. 2013b), Knowledge based (Cambria et al. 2013) and Machine learning algorithm(Pang et al. 2002, Agarwal et al. 2013a). Sentiment orientation based approaches for sentiment analysis works in three steps. First of all sentiment-rich features are extracted. Further, semantic orientations of these sentiment-rich features are determined and finally overall semantic orientation of the document is computed by aggregating the semantic orientations of all the features in the document. Knowledge-based approaches initially use parser to extract the features, and further these features are

Related Content

Finding Associations in Composite Data Sets: The CFARM Algorithm
M. Sulaiman Khan, Maybin Muyeba, Frans Coenen, David Reidand Hissam Tawfik (2013). *Developments in Data Extraction, Management, and Analysis (pp. 162-186).*
www.irma-international.org/chapter/finding-associations-composite-data-sets/70797

The Process and Application of XML Data Mining
Richi Nayak (2009). *Handbook of Research on Text and Web Mining Technologies (pp. 249-272).*
www.irma-international.org/chapter/process-application-xml-data-mining/21728

Hybridization of Biogeography-Based Optimization and Gravitational Search Algorithm for Efficient Face Recognition
Lavika Goel, Lavanya B.and Pallavi Panchal (2019). *Advanced Metaheuristic Methods in Big Data Retrieval and Analytics (pp. 258-279).*
www.irma-international.org/chapter/hybridization-of-biogeography-based-optimization-and-gravitational-search-algorithm-for-efficient-face-recognition/216102

Efficient Open Domain Question Answering With Delayed Attention in Transformer-Based Models
Wissam Siblini, Mohamed Challaland Charlotte Pasqual (2022). *International Journal of Data Warehousing and Mining (pp. 1-16).*
www.irma-international.org/article/efficient-open-domain-question-answering-with-delayed-attention-in-transformer-based-models/298005

Elasticity in Cloud Databases and Their Query Processing
Goetz Graefe, Anisoara Nica, Knut Stolze, Thomas Neumann, Todd Eavis, Ilia Petrov, Elaheh Pourabbasand David Fekete (2013). *International Journal of Data Warehousing and Mining (pp. 1-20).*
www.irma-international.org/article/elasticity-cloud-databases-their-query/78284