

Chapter 58

A Novel Algorithm for Sentiment Analysis of Online Movie Reviews

Bisma Shah

Jamia Hamdard, India

Farheen Siddiqui

Jamia Hamdard, India

ABSTRACT

Others' opinions can be decisive while choosing among various options, especially when those choices involve worthy resources like spending time and money buying products or services. Customers relying on their peers' past reviews on e-commerce websites or social media have drawn a considerable interest to sentiment analysis due to realization of its commercial and business benefits. Sentiment analysis can be exercised on movie reviews, blogs, customer feedback, etc. This chapter presents a novel approach to perform sentiment analysis of movie reviews given by users on different websites. Also, challenges like presence of thwarted words, world knowledge, and subjectivity detection in sentiments are addressed in this chapter. The results are validated by using two supervised machine learning approaches, k-nearest neighbor and naive Bayes, both on method of sentiment analysis without addressing aforementioned challenges and on proposed method of sentiment analysis with all challenges addressed. Empirical results show that proposed method outperformed the one that left challenges unaddressed.

INTRODUCTION

Sentiment Analysis (SA) is an on-going field of research in text mining field and is used for the computational treatment of opinions, sentiments and subjectivity of text. It is used to understand the mind-sets, opinions and sentiments of people in general in regards to a specific product or a movie or an occasion. It helps in identifying and extracting subjective information in source materials and categorizes them as positive, negative, or neutral. Decision making process has always been influenced by the huge data

DOI: 10.4018/978-1-6684-6303-1.ch058

available and the human nature to always rely on opinion of other people regarding a particular product or an event. This unique feature has a significant impact on deciding matters that have financial, medical, social, business or other ramifications. Keeping in view the large amount of comments or reviews for a particular product and the colossal advancement in web users, there arises a dire need to build up a framework that gathers, constructs, analyzes and characterizes the remarks or reviews present online. A review is usually written by a person who has used a particular product or a service. The nature of review is immensely affected by an individual's interests, opinions and viewpoints. People who give biased reviews consequently impact the reputation of a forum or an organization to which the review is being contributed. Increase in the number of such people has posed a huge challenge to characterize and sort out the real issues and prospects of the product by virtue of which a user questions the authenticity of the content. In view of improving the scope of a particular product, huge organizations depend on individual reviews of clients and consider it to be of incredible significance in putting content construct advertisements on websites that effectively help a forthcoming buyer. Also movie enthusiasts and voters employ the same approach for analysis of certain information as an ever increasing number of individuals are utilizing social networking sites, online shopping and trend analysts who in the wake of scrutinizing the reviews available settle on different issues. For instance putting the promotion of a Kitchen Aid Mixer on a food blog impacts purchase choices as well as goes far in altering the advertising technique. Reviewers are being enthusiastically promoted by the advertising division of an organization by sending samples of product to be assessed or supporting discounts in blogs or in social networking locales like Facebook and Twitter. This has prompted the expansion in the volume of information accessible and the need to characterize the accessible data effectively as these have a larger impact on the reputation of a particular organization.

Large amount of data generated online by reviews can be processed so as to extract useful information from them, using suitable methods and approaches of Sentiment analysis, thereby supporting operational, managerial, and strategic decision making (Liu, 2010). But it is not sufficient to consider just the subjective nature of opinion for decision making (Buche, Chandak, & Zadgaonkar, 2013). Additionally, the written work aptitudes and selection of words by benefactors to a great extent rely on upon the proficiency of language and the demeanor of the author. There exist different kinds of opinions of users like regular, implicit, direct, indirect and comparative. The flexibility of expression and anonymity also accompanies a cost. Individuals with concealed plans effortlessly game the framework to give people the impression that they are independent members from the general population and post fake sentiments to promote or defame some targeted products, services, affiliations, or individuals without revealing their actual intentions, or the individual or affiliation that they are secretly working for. People with these intentions are popularly known as *opinion spammers* and their exercises are called *opinion spamming* (Abbasi, Chen, Thoms, & Fu, 2008). Moreover, online reviews are composed by the customers from their edge of interests and inclinations, they can be a mix of a positive and negative opinion and that may not necessarily help in categorizing reviews as positive or negative. For instance, consider the sentence "The Chinese dishes of this restaurant are not as good as their Thai dishes". Relative opinions like these are different in natural language processing. When a positive word "good" is negated like "not as good as", readers find it challenging to assimilate even how good the Thai dishes were, because this decides the taste of the Chinese dishes too. So while handling negation, it must be properly figured out that the presence of negation updates which part of the meaning expressed (Liu, 2012; Abbasi, France, Zhang, & Chen, 2011).

32 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/a-novel-algorithm-for-sentiment-analysis-of-online-movie-reviews/308536

Related Content

Elasticity in Cloud Databases and Their Query Processing

Goetz Graefe, Anisoara Nica, Knut Stolze, Thomas Neumann, Todd Eavis, Iliia 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

The Sentiment Revealed in Social Networks during the Games of the Brazilian Team in the 2014 World Cup: A Conceptual Approach of Actor-Network Theory

Rita Paulino (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 1705-1716).
www.irma-international.org/chapter/the-sentiment-revealed-in-social-networks-during-the-games-of-the-brazilian-team-in-the-2014-world-cup/150237

The IT Readiness for the Digital Universe

Pethuru Raj (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 757-777).
www.irma-international.org/chapter/the-it-readiness-for-the-digital-universe/150192

Web Mining for the Integration of Data Mining with Business Intelligence in Web-Based Decision Support Systems

Marcos Aurélio Domingues, Alípio Mário Jorge, Carlos Soaresand Solange Oliveira Rezende (2015). *Integration of Data Mining in Business Intelligence Systems* (pp. 120-142).
www.irma-international.org/chapter/web-mining-for-the-integration-of-data-mining-with-business-intelligence-in-web-based-decision-support-systems/116812

Approximating Proximity to Fast and Robust Distance-Based Clustering

Vladimir Estivill-Castroand Michael Houle (2002). *Data Mining: A Heuristic Approach* (pp. 1-21).
www.irma-international.org/chapter/approximating-proximity-fast-robust-distance/7582