A Comparative Study of Different Classification Techniques for Sentiment Analysis

Soumadip Ghosh, Academy of Technology, Kolkata, India

https://orcid.org/0000-0003-4817-5363

Arnab Hazra, Academy of Technology, Kolkata, India

Abhishek Raj, Academy of Technology, Kolkata, India

https://orcid.org/0000-0002-8121-3440

ABSTRACT

Sentiment analysis denotes the analysis of emotions and opinions from text. The authors also refer to sentiment analysis as opinion mining. It finds and justifies the sentiment of the person with respect to a given source of content. Social media contain vast amounts of the sentiment data in the form of product reviews, tweets, blogs, and updates on the statuses, posts, etc. Sentiment analysis of this largely generated data is very useful to express the opinion of the mass in terms of product reviews. This work is proposing a highly accurate model of sentiment analysis for reviews of products, movies, and restaurants from Amazon, IMDB, and Yelp, respectively. With the help of classifiers such as logistic regression, support vector machine, and decision tree, the authors can classify these reviews as positive or negative with higher accuracy values.

KEYWORDS

Decision Tree, Logistic Regression, Machine Learning, Sentiment Analysis, Support Vector Machine

1. INTRODUCTION

Sentiment analysis or opinion mining (Neethu M. et al., 2013) refers to emotions and opinions by analysis of texts, processing of natural languages to methodically identify, extract, count, and study some interesting information. Sentiment analysis has gained popularity in the recent past. The idea of performing analysis on texts is important for marketing research, where analysts wish to find out some useful information from customer feedback. It is vastly applied to various forms of customer feedback such as reviews and survey responses found on the web and social media. Commercial websites such as Amazon, eBay, Yelp and IMDb provide users the platform required to express their opinions towards any specific product or subject. Individuals post reviews of movies they have watched on websites like IMDb.

Performing analysis of sentiments from various data sources found on the web is valuable for any organization to maintain quality control of their products. For instance, getting user feedback means requesting people with surveys on every aspect the organization is interested in. One of the sources of doing this is web blogs and another one is electronic discussion boards, where individuals can talk about different types of topics or can request other people's views. This approach is beneficial for numerous reasons. Primarily, the people who share their views usually have more noticeable opinions than the average, which are furthermore convincing others to read them. Secondarily, product and service reviews obtained from commercial web sites also help us to choose which products to buy and which services to use. Furthermore, the individual reviews obtained from personal blogging sites are mostly unbiased and have individual experience towards a specific product or service. Mining these opinions is thus carrying valuable information for the improvement of the business.

Opinion mining is a technique of categorizing opinions articulated in the text sentences (Manning et al., 2008) obtained from several data sources. Basically, text sentences carry personal review or attitude concerning any specific product or subject. Opinion mining of small texts is thought-provoking because they are contextually limited. Decisions are to be made based on the inadequate texts provided by the user. We refer to this method as a supervised learning technique as it can categorize each user review correctly (Pang, Lee, & Vaithyanathan, 2002).

Machine learning (ML) (Witten I. H. et al., 2011) based classification models are trained with data sets containing text sentences and their performances are evaluated as well. Classification techniques such as Logistic Regression (LR) (Cramer J. S., 2002), Support Vector Machine (SVM) (Cortes C., & Vapnik V., 1995) and Decision Tree (DT) (Quinlan J. R.,1987) from ML domain can be applied to text data for performing sentiment analysis. These research studies (Kamal S. et al, 2016, 2017, 2018) contributed some methods which we have applied in our work.

The different sections of the research paper are as follows. In the first section, we have introduced about sentiment analysis and described its importance in business. Section 2 provides literature reviews that worth mentioning in this domain. Section 3 presents the data set description which is followed by the proposed methodology in section 4. Section 5 describes and analyzes results with explanations. Finally, section 6 is attributed to the conclusion and future works.

2. LITERATURE REVIEW

There are several ML-based types of research available to classify sentiments from the text. Some of them are listed below.

ML consists of several classification models such as Artificial Neural Network (ANN), SVM, decision tree, Logistic Regression, etc. These techniques are employed to categorize reviews of products. The research study (Mejova Y. et al., 2009) showed that using the presence of every character, frequency of occurrences of every character, text sentence containing negation, etc. as the features to build feature vector. He also showed that using unigram and bigram approaches one could create feature vectors efficiently in Sentiment analysis.

The research work (Domingos P., 1997) proposed that the Naive Bayes classifier could do well using dependent features for a certain problem. This work (Niu Z. et al., 2012) developed a new classifier based on the Bayesian algorithm. The model employed some effective approaches for the selection of a feature, computation of weight and classification. The research study (Barbosa L., & Feng J., 2010) designed a two-step analysis method which was an automatic sentiment analysis for classifying tweets. In the first step, tweets were classified into subjective and objective tweets. Then, in the second step, subjective tweets were classified as positive and negative tweets.

The research work (Celikyilmaz A. et al., 2010) developed a word clustering method based on the pronunciation of words. This method is applicable for normalizing noisy tweets. There are some words with the similar pronunciation but dissimilar meanings. So, to eliminate this kind of conflict, methods were developed. In the stated method, words having the same pronunciation were clustered

Related Content

Revolutionizing Healthcare: Synergizing Cloud Robotics and Artificial Intelligence for Enhanced Patient Care
Jaspreet Kaur (2024). *Shaping the Future of Automation With Cloud-Enhanced Robotics (pp. 272-287).*
www.irma-international.org/chapter/revolutionizing-healthcare/345546

Gait Rhythm of Parkinson's Disease Patients and an Interpersonal Synchrony Emulation System based on Cooperative Gait
Hirotaka Uchitomi, Kazuki Suzuki, Tatsunori Nishi, Michael J. Hove, Yoshihiro Miyake, Satoshi Orimoand Yoshiaki Wada (2014). *Robotics: Concepts, Methodologies, Tools, and Applications (pp. 1197-1211).*
www.irma-international.org/chapter/gait-rhythm-of-parkinsons-disease-patients-and-an-interpersonal-synchrony-emulation-system-based-on-cooperative-gait/84945

The Compilation and Validation of a Collection of Emotional Expression Images Communicated by Synthetic and Human Faces
Louise Lawrenceand Deborah Abdel Nabi (2013). *International Journal of Synthetic Emotions (pp. 34-62).*
www.irma-international.org/article/the-compilation-and-validation-of-a-collection-of-emotional-expression-images-communicated-by-synthetic-and-human-faces/97677

Physics and Cognitive-Emotional-Metacognitive Variables: Learning Performance in the Environment of CTAT
Sarantos I. Psycharis (2009). *Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and Artificial Intelligence (pp. 379-390).*
www.irma-international.org/chapter/physics-cognitive-emotional-metacognitive-variables/21517

Design and Validation of Force Control Loops for a Parallel Manipulator
Giuseppe Carbone, Enrique Villegasand Marco Ceccarelli (2013). *Advanced Engineering and Computational Methodologies for Intelligent Mechatronics and Robotics (pp. 206-224).*
www.irma-international.org/chapter/design-validation-force-control-loops/76450