Chapter 36
Investigating Machine Learning Techniques for User Sentiment Analysis

Nimesh V Patel

C.U Shah University, Wadhawan (Surendranagar-Gujarat), India

Hitesh Chhinkaniwala

Adani Institute of Infrastructure Engineering, Gujarat, India

ABSTRACT

Sentiment analysis identifies users in the textual reviews available in social networking sites, tweets, blog posts, forums, status updates to share their emotions or reviews and these reviews are to be used by market researchers to do know the product reviews and current trends in the market. The sentiment analysis is performed by two methods. Machine learning approaches and lexicon methods which are also known as the knowledge base approach. These. In this article, the authors evaluate the performance of some machine learning techniques: Maximum Entropy, Naïve Bayes and Support Vector Machines on two benchmark datasets: the positive-negative dataset and a Movie Review dataset by measuring parameters like accuracy, precision, recall and F-score. In this article, the authors present the performance of various sentiment analysis and classification methods by classifying the reviews in binary classes as positive, negative opinion about reviews on different domains of dataset. It is also justified that sentiment analysis using the Support Vector Machine outperforms other machine learning techniques.

INTRODUCTION

Sentiment Analysis finds outs the tests of the user from their reviews provided on social networking sites, forums, blogs, e-commerce company websites. This sentiment analysis on product or item reviews can be wrapped with other web recommender systems. Like search engines, recommenders take into account the personality characteristics, past behavior of each user, a product's feature-based reviews and

sentiment analysis about the product for certain aspects about the product, other geographical information, and market conditions.

Sentiment analysis and classification can be performed by basically two methods: (1) Machine learning based and (2) Lexicon based techniques. The machine learning techniques can be further divided into two types (i) supervised and (ii) unsupervised learning. There are certain differences between these techniques. Supervised learning technique requires more domain knowledge and information seeds can be provided from outside to train model, but it can be applied on the multi-aspect dataset so that can reflect more user needs for specific aspects of a product. Whereas unsupervised learning does not require domain Knowledge, no rating calculations to be performed to find out sentiment scores and word ordering is not required do process sentiment analysis. Unsupervised learning diagrams and trigrams also polarized with more number of granularity levels. It can be applied to multiple domains. Another type of sentiment analysis methods is lexicon based sentiment analysis. Lexicon based technique is further divided into various categories. (i) The Dictionary Based and (ii) Corpus-based methods. The dictionary-based methods require large domain-specific knowledge, it provides aspect-sentiments extraction and Evaluation, capable of treating of negation words like NOT, also able to analyze Semantic relationships like the synonym, antonym, adjective. Whereas Corpus-Based Methods are able to deal with detection of spam and fake opinion, Segment wise score calculation, but requires deep knowledge of explicit-implicit, regular-irregular, syntactical-semantic rules. (Nimesh et al., 2015) provided more details regarding issues, challenges, working characteristics, and behaviour of sentiment analysis methods.

The work presented in this article contains five parts. The first part gives the primary sense about the significance of sentiment analysis, recommending process using sentiment analysis and the introduction of some well-known methods to perform sentiment analysis. The second part includes the related work carried out in an area by various researchers. The third part explains the process steps performed in sentiment analysis using machine learning techniques along with pre-processing tasks and other subsequent tasks to be carried out to evaluate the performance of machine learning based sentiment analysis. Next part includes discussion about the various evaluation criteria. The fifth section includes information about the used datasets, experimental setup, and the results achieved by execution. The final part includes the justification of the proposed targeted work.

RELATED WORK

In movie review domain where the application of sentiment analysis has the greatest potential to generate value, it is an interesting and challenging testing ground for different sentiment classification approaches. It was hypothesized that this was due to the tendency of reviewers to rate the individual elements of items differently from the item as a whole within the same review. (Bo Pang et al., 2002) suggest, the machine learning methods and features used when classifying movie reviews do not have to be specific to that domain but its benefits may easily transfer to other areas where sentiment classification can be applied. For binary (positive and negative) sentiment classification, (Turney et al., 2002) proposed the positive and negative terms can be counted and expressions in a review used to determine its polarity. (Kennedy et al., 2006) who had considered negation words, intensifiers, and diminishes taken into account and managed for improving the accuracy of the system. In this article Support Vector Machines (SVMs)

10 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/investigating-machine-learning-techniques-for-user-sentiment-analysis/307478

Related Content

Improve Imbalanced Multiclass Classification Based on Modified SMOTE and Feature Selection for Student Grade Prediction
Siti Dianah, Ali Selamatand Ondrej Krejcar (2022). *Handbook of Research on New Investigations in Artificial Life, AI, and Machine Learning (pp. 371-389).*
www.irma-international.org/chapter/improve-imbalanced-multiclass-classification-based-on-modified-smote-and-feature-selection-for-student-grade-prediction/296811

ACO_NB-Based Hybrid Prediction Model for Medical Disease Diagnosis
Amit Kumar, Manish Kumarand Nidhya R. (2021). *Handbook of Research on Disease Prediction Through Data Analytics and Machine Learning (pp. 526-536).*
www.irma-international.org/chapter/aconb-based-hybrid-prediction-model-for-medical-disease-diagnosis/263336

Autonomous Navigation Using Deep Reinforcement Learning in ROS
Ganesh Khekareand Shahrukh Sheikh (2021). *International Journal of Artificial Intelligence and Machine Learning (pp. 63-70).*
www.irma-international.org/article/autonomous-navigation-using-deep-reinforcement-learning-in-ros/277434

A Study on Evaluation and Analysis of Edge Detection Operators
Pinaki Pratim Acharjya, Subhankar Joardar, Santanu Koleyand Subhabrata Barman (2023). *Meta-Learning Frameworks for Imaging Applications (pp. 65-102).*
www.irma-international.org/chapter/a-study-on-evaluation-and-analysis-of-edge-detection-operators/331478

A Review on Time Series Motif Discovery Techniques an Application to ECG Signal Classification: ECG Signal Classification Using Time Series Motif Discovery Techniques
Ramanujam Elangovanand Padmavathi S. (2019). *International Journal of Artificial Intelligence and Machine Learning (pp. 39-56).*
www.irma-international.org/article/a-review-on-time-series-motif-discovery-techniques-an-application-to-ecg-signal-classification/238127