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.

KEYWORDS

Feature Extraction, Machine Learning, Naïve Bayes, Opinion Mining, Maximum Entropy, Recommender System, Sentiment Analysis, Support Vector Machine

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

DOI: 10.4018/IJDSST.2019070101

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
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) method was proposed, and it was inferred that this machine learning algorithm performs significantly better than the term-counting method by considering negation words. (Pang et al., 2002) and (Pang et al., 2004) have compared the performance of various classification algorithms when determining the sentiment of a document, and also proved that SVMs were generally the best approach. Unigrams, bigrams, part-of-speech (POS) tags and term positions were considered as features, and using unigrams gives the best results. When using very simple features with multi-aspect sentiment analysis the SVM classification algorithm can be effective for performing sentiment analysis.

Arjun Mukherjee, Bing Liu, (2012) done aspect extraction by providing some seed words then the model will extract and clusters aspect terms into categories simultaneously. It needs to be considered that many discovered aspects can be meaningless/non-useful and totally unsupervised methods(Suin Kim, et al. 2013; Yulan et al., 2009) can’t fulfill all user aspects. This model reflects more user’s needs to be discovered for specific aspects. A hierarchical technique suggested by (Suin Kim et al., 2013) by two-level tree formation for aspect-sentiments. Seeded as the model to discover aspect based sentiments for given seeded word. In this model Aspect or Sentiment polarity is modelled as a distribution of words and infer aspect-sentiment tree from review text by including the likelihood- how the review corpus is generated, prior-how the tree is generated in prior, the posterior-review observation for analysis. This model is more accurate and flexibility to discover aspect with more than two sentiments. The similar algorithm for content-based recommender system based on sentiment analysis was suggested by (Keshav R, Arvind Swaminathan et al., 2014). It extracts opinion and classifies the reviews and
A Best Practise Business Intelligence Framework for the Telecommunications Industry: An Empirical Study
[www.igi-global.com/article/a-best-practise-business-intelligence-framework-for-the-telecommunications-industry/203636?camid=4v1a](www.igi-global.com/article/a-best-practise-business-intelligence-framework-for-the-telecommunications-industry/203636?camid=4v1a)

Selection of Cloud Delivery and Deployment Models: An Expert System Approach
[www.igi-global.com/article/selection-of-cloud-delivery-and-deployment-models/211181?camid=4v1a](www.igi-global.com/article/selection-of-cloud-delivery-and-deployment-models/211181?camid=4v1a)