Research Paper E-ISSN NO : 2455-295X | VOLUME : 9 | ISSUE : 6 | SPECIAL ISSUE JUNE-2023

NOVEL DEEP LEARNING APPROACHES FOR OPINION MINING WITH MACHINE LEARNING

KANIMOZHI.J	PH.D RESEARCH SCHOLAR, DEPARTMENT OF COMPUTER SCIENCE, NALLAMUTHU GOUNDER MAHALINGAM COLLEGE, POLLACHI, TAMILNADU, INDIA.
DR.R.MANICKA CHEZIAN	ASSOCIATE PROFESSOR, DEPARTMENT OF COMPUTER SCIENCE, NALLAMUTHU GOUNDER MAHALINGAM COLLEGE, POLLACHI, TAMILNADU, INDIA.

ABSTRACT:

The popularity of online marketplaces over the past several decades, online vendors and merchants now request feedback from their customers on the goods they have purchased. As a consequence, millions of evaluations are produced every day, which makes it challenging for a potential customer to decide whether to buy the goods or not. For product makers, it is challenging and time-consuming to analyse this massive volume of comments. The challenge of categorizing reviews according to their general semantic content (positive, negative, neutral) is examined. SVM and Nave Bayes, two different supervised machine learning approaches, have been tested on Amazon beauty goods to perform the study. Then, their accuracy levels were contrasted. The outcomes demonstrated that the SVM technique performs better than the Nave Bayes.

KEYWORDS:

INTRODUCTION

online marketplaces over the past several decades, online vendors and merchants now request feedback from their customers on the goods they have purchased. Every day, millions of evaluations are written about various goods, services, and locations online. Due of this, the Internet is now considered to be the most significant source for information about products and services. People frequently shop online at Amazon, one of the biggest e-commerce sites, where they may read hundreds of evaluations left by other consumers about the things they want to buy. These evaluations provide insightful thoughts regarding a product, such as its features, quality, and suggestions, which aids buyers in understanding nearly all of its details. Customers gain from this, but it also benefits merchants who produce their own goods since it makes it easier for them to comprehend customers' wants.

2. SENTIMENT CLASSIFICATION AND ANALYSIS

Sentiment analysis, commonly referred to as opinion mining, is a challenge in natural language processing (NLP), which entails finding and extracting subjective data from text sources. The goal of sentiment classification is to categorise user evaluations that have been published as either positive or negative, so the system does not have to fully comprehend the semantics of each word or document.

Yet, it's not enough to just assign words a good or bad connotation. There are several difficulties present. It is not always effective to categorise words and sentences based on their preceding positive or negative polarity. For instance, the word "amazing" has a positive connotation by itself, but if it is used with a word of negation, such as "not," the context might radically shift

2.1 SENTIMENT CLASSIFICATION USING MACHINE LEARNING METHODS

By employing example data, machine learning seeks to create an algorithm that will improve the system's performance. There are two primary phases to the sentiment analysis solution that machine learning offers. The model must first "learn" from training data before it can be used to categorise unobserved data in the subsequent stage.

Different categories may be used to classify machine learning algorithms:

Learning that is either supervised, semi-supervised, or unsupervised.

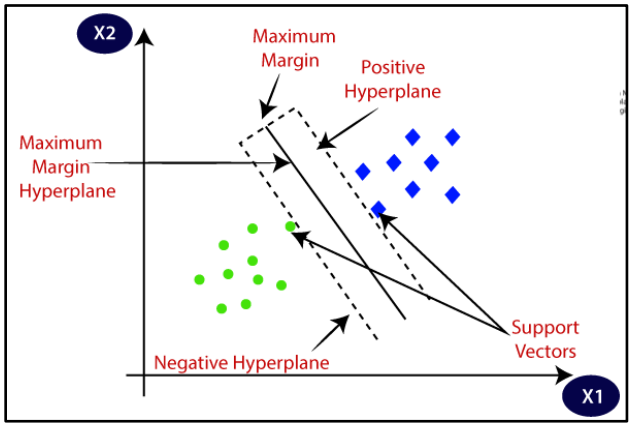
In supervised learning, the algorithm learns from training data in a manner similar to a teacher monitoring their pupils' learning (Brownlee 2016). The supervisor is somehow instructing the algorithm as to what results it should produce. Data for both the required input and output are therefore given. Additionally, the training data must already be labelled.

Unsupervised learning is taught on unlabeled data with no matching output, as opposed to supervised learning. The underlying structure of the data collection should be discovered by the algorithm on its own. This implies that in order to decide the output without having the correct answers, it must find comparable patterns in the data. Clustering is one of the most crucial techniques for unsupervised learning issues. Semi-supervised learning, which combines the advantages of supervised and

unsupervised learning, refers to issues when just a portion of the training data set is tagged. This is helpful when gathering data is inexpensive but identifying it takes time and money. Due to the fact that this method tends to increase the accuracy of the final model while requiring significantly less time and money to develop, it is very advantageous both in theory and in practice.

SUPPORT VECTOR MACHINE

support vector machines (SVM) are a supervised learning technique that may be utilised to address sentiment categorization issues. This method uses a decision plane where labelled training data is inserted. An algorithm then generates an optimum hyperplane which divides the data into several groups or classes. The optimal hyperplane is the one that divides the classes by the greatest margin, as shown in figure 1. This is accomplished by selecting a hyperplane with a maximum distance from the closest data for each class.

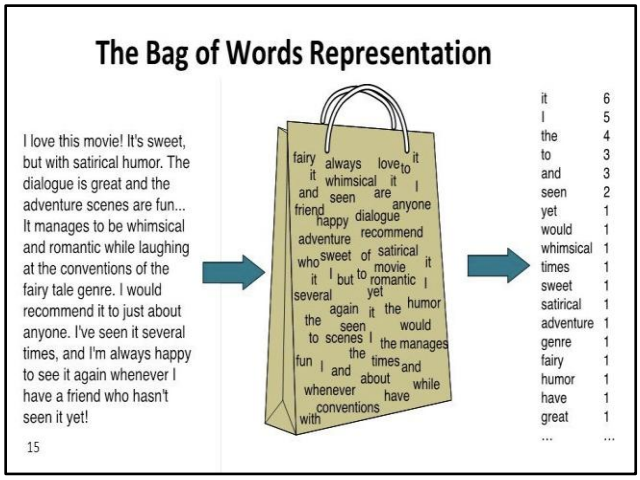


NAÏVE BAYES TECHNIQUES

Another machine learning method noted for its strength despite its simplicity is naive Bayes. This classifier, which is based on the Bayes theorem, makes the assumption that the features—which in text classification are often words—are independent of one another. Although Naive Bayes classifiers have demonstrated to work very well, this assumption is untrue. Prior to using the Nave Bayes model to solve text classification issues, feature extraction should be performed.

FEATURE EXTRACTION

The input (in this example, text data) must be processed since machine learning algorithms can only deal with fixed-length vectors of integers rather than raw text. The Bag of Words Model of Text, a widely used feature extraction technique, is the process for turning the texts into features. The method involves assembling several bags of words from the training data set, each of which is assigned a different number. The frequency of each word in the text is indicated by this number. Figure 2 provides a straightforward illustration of the Bag of Words paradigm. Because of the words' location in the document being ignored, the model is known as a bag of words.



RELATED WORK

Three supervised machine learning algorithms, Naive Bayes, SVM, and N-gram model, have been tried on internet evaluations about various tourist locations throughout the world, according to a recent study by Ye et al. In this study, it was discovered that properly-trained machine learning algorithms perform very well at categorizing reviews of tourism destinations.

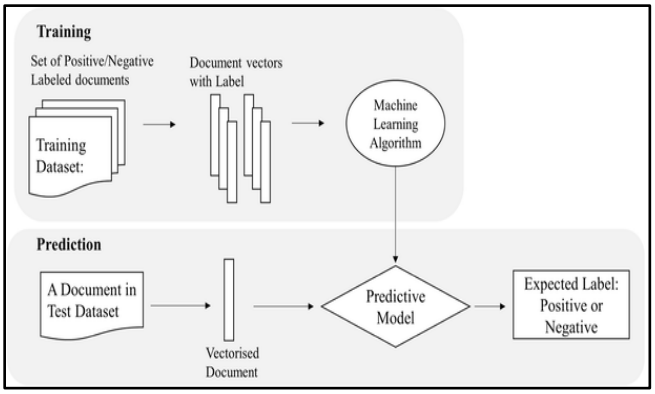
Additionally, they showed that the SVM and N-gram model outperformed the Naive Bayes approach in terms of outcomes. However, by adding more training data sets, the disparity between the algorithms was dramatically decreased.

Naive Bayes and SVM are reportedly the two methods most frequently employed to solve sentiment classification issues (Joachims 1998; Pang et al. 2002; Ye et al. 2009). Therefore, this thesis attempts to apply the supervised machine learning techniques of Nave Bayes and SVM to the Amazon website's reviews.

MACHINE LEARNING CLASSIFIERS

Each classifier algorithm must first be taught before being tested in order to conduct the experiments. The data was split into two data sets—training data sets and testing data sets—in order to train and apply the classifiers. As was already indicated, this research has involved two experiments. The classifiers were trained and evaluated twice on the reviews themselves and twice on the review summaries in each trial.

Step to transform the review texts into numerical features before being fed to the algorithms. The Bag of Words approach was employed for this. The Nave Bayes and SVM classifiers were trained in the third stage. The next stage was to apply the trained classifiers to the test data in order to evaluate their performance by contrasting the real labels that were not provided to the algorithms with the anticipated labels.



RESULT

The study's findings are presented in this section. The accuracy number displays the proportion of the testing data set that the model successfully categorised. Two different machine learning algorithms' performance on two sets of experiments.

ProductId	Na¨ıve Bayes	SVM
B004O	89.98%	89.43%
B0043	89.98%	89.15%
B0069	89.98%	84.13%
B000Z	89.98%	84.12%
B0015	89.98%	89.41%

CONCLUSION

In this work, SVM and Nave Bayes, two different machine learning algorithms, were applied to reviews on Amazon.com. The study's findings demonstrated that, when the entire data set was utilised as the training and testing data set, the SVM strategy outperformed the Naive Bayes approach in terms of accuracy. The Naive Bayes approach outperformed the SVM algorithm as the number of reviews fell.

REFERENCES

1. Richard A Berk. Statistical learning from a regression perspective. Springer, 2016.

2. Jason Brownlee. Supervised and unsupervised machine learning algorithms, Mar 2016.

3. Xing Fang and Justin Zhan. Sentiment analysis using product review data. Journal of Big Data, 2(1):5, 2015.

4. PimwadeeChaovalit and Lina Zhou. Movie review mining: A comparison between supervised and unsupervised classification approaches. In System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International

5. P´adraig Cunningham, Matthieu Cord, and Sarah Jane Delany. Supervised learning. In Machine learning techniques for multimedia, pages 21–49. Springer, 2008.

6. SajibDasgupta and Vincent Ng. Topic-wise, sentiment-wise, or otherwise?: Identifying the hidden dimension for unsupervised text classification. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2, pages 580–589. Association for Computational Linguistics, 2009.

7. Emma Haddi, Xiaohui Liu, and Yong Shi. The role of text pre-processing in sentiment analysis. Procedia Computer Science, 17:26–32, 2013.

8. Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 168–177. ACM, 2004.

9. SAURAV KAUSHIK. An introduction to clustering and di↵erent methods of clustering, Nov 2016.

10. JayashriKhairnar and MayuraKinikar. Machine learning algorithms for opinion mining and sentiment classification. International Journal of Scientific and Research Publications, 3(6):1–6, 2013.

11. Jure Leskovec and RokSosiˇc. Snap: A general-purpose network analysis and graph-mining library. ACM Transactions on Intelligent Systems and Technology (TIST), 8(1):1, 2016.

12. Bing Liu. Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1):1–167, 2012. Bing Liu. Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press, 2015.

13. Bingwei Liu, Erik Blasch, Yu Chen, Dan Shen, and Genshe Chen. Scalable sentiment classification for big data analysis using naive bayes classifier. In Big Data, 2013 IEEE International Conference on, pages 99–104. IEEE, 2013.

14. Jingjing Liu, Yunbo Cao, Chin-Yew Lin, Yalou Huang, and Ming Zhou. Low-quality product review detection in opinion summarization. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), 2007.

15. Zeenia Singla, Sukhchandan Randhawa, and Sushma Jain. Statistical and sentiment analysis of consumer product reviews. In Computing, Communication and Networking Technologies (ICCCNT), 2017 8th International Conference on, pages 1–6. IEEE, 2017.