COMPARING SENTIMENT ANALYSIS APPROACHES ON AMAZON REVIEWS

Ch.Devi[#1], K. Chaitrika[#2], P.V. sowmya[#3], N. Pooja Nagavalli [#4], K.Harini[#5]
Department of Computer Science & Engineering, Vignan's Institute of Engineering for Women, Vizag

ABSTRACT
The world we live in today is constantly becoming digital. As more and more of the world becomes digital, E-commerce is booming because it makes products accessible to customers without requiring them to leave their homes. Consumers these days rely heavily on online products, so a review's significance is growing. Thousands of opinions must be checked by the consumer before choosing a product to fully comprehend it. But we can make this process simpler by building a machine learning model which can go through thousands of reviews polarize them into positive, negative, and neutral and learn from them. To polarize the reviews, we use sentiment analysis. There are various methods for sentiment analysis for Amazon product reviews. The goal of this paper is to compare the various sentiment analysis methods like pre-trained models such as Vader, and traditional models such as Logistic Regression (LR) and Support vector machine (SVM), and deep learning model such as simple long short-term memory (LSTM) recurrent neural network and identify the best model by using all the methods on a large scale amazon dataset. The top-performing model can be applied to future product reviews' algorithmic sentiment classification.
Keywords: Sentiment analysis, Vader, Linear Regression, SVM, Deep learning, long-short term memory

INTRODUCTION

Sentiment analysis is a procedure which looks at written language to determine whether an expression is neutral, good, or negative. sentiment analysis tells the customers opinion on the products. it mainly defines the emotion of the particular product, and it divides the opinion into positive, negative, neutral. The main task of this paper is to label neutral, negative, and positive feedback on amazon products and different brands. The items which are having more positive reviews in amazon, the customer will automatically prefer to buy that particular product. So, the main motive of the amazon customer for buying the item is mainly depends upon the reviews/ratings. This paper's objective is to gather data regarding Amazon product reviews, pre-process them, and assess the data in order to gain useful insights. We are going to perform Sentiment Analysis u sing pre-trained models, various Deep Learning, and conventional Machine Learning methods. The effectiveness of each algorithm can then be compared to determine which one does sentiment analysis and prediction the best. The top-performing model can be applied to future product reviews' algorithmic sentiment classification. Here, the pre-trained model refers to the Vader approach (Valence Aware Dictionary for Sentiment Reasoning) and t The term "conventional methods" in this study refers to techniques that use a manually created feature extractor and a linear classifier like the Logistic Regression, SVM (Support vector machine) and LSTM is the deep learning algorithm employed. All the models are compared, and the results back up the notion that deep learning models are "data hungry" and perform better when trained on more examples.

RELATED WORKS

[1] The research papers corresponding to sentiment analysis that has been done newly. It works on finding emotions from reviews. The main frequently used programming language was R programming and Python programming, Elli, Maria and Yi-F an gathered sentiment of the product from the reviews and used the probe to create a commerce model. [2]The author practiced current supervised learning algorithms to conclude the review score using the text. They used the input as 70% training input and 30% testing input. In this paper the author had used various types of classifiers for finding out the precision and recall values. [3] In this paper the author practiced and increased the existing project in the area of NLP and with the help amazon

review datasets finding the opinion of the product. They used decision list classifier and naive bayes to name the probe as favourable, unfavourable, and neutral. [4] the author planned to create a model that displays the opinion in the form of graphs, they have utilized the data scratching from the amazon URL for collecting the data and pre-processed it. The paper finally summarizes the product review to be main point but there is not much more perfection, they indicate the final output into statistical charts. [5]they are predicting the ratings/reviews based on the customers giving the rating on the particular product using a pack of words, these design tested apply unigrams and bigrams, they pre-owned a division of amazon video game user probe from UCSD Time based design did not performed well and unigram had a accuracy 15.89% better than bigrams.[6]in this paper they used simple algorithms so it is safe to recognize, the system gives high validity on the SVM and it cannot effort well on the big/large datasets.

So far, in many research papers related to by-product probe of sentiment analysis we mainly deal with people's emotions, evaluations in the context. In paper [7] we mainly deal with Vader approach in different sentiment lexicons such as semantic orientation lexicons, sentiment intensity lexicons and lexicons through context awareness.

In this paper [9] The author mainly mentioned about amazon mobile phone probe that are classified into favourable, unfavourable and ordinary according to star rating since 1-star & 2-star ratings are considered as unfavourable, three star rating are ordinary whereas 4&5 are favourable, Accordingly all proposed models are practiced on feature origin models. The logistic regression model applied with bag of words (BOW), TF-IDF. The BOW achieved an accuracy of 70%, Again the result evaluation is on random forest model (RFM) with different factor of origin approaches. The RF with glove displays an validity of 80% the reason behind using glove is a word embedding method which aims in building low structural vector.

[10] This Paper focuses on gathering Amazon reviews and categorizing them as positively or negatively using various machine learning methods. Precision, Recall, and F-measure are used to gauge production. Yet machine learning is unable to keep up with the performance as the volume of diverse data to be handled increases. Deep Learning was introduced as a result. The most widely used Deep Learning models for Natural Language Processing (NLP) are the Convolution Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short Term Memory (LSTM). RNNs exhibit what is known as short-term memory. They can only recall and interpret what was in the prior neuron. One of the downsides of RNN is the vanishing gradient problem. S. Hochreiter suggested a solution to this known as LSTM[11][12] can manage long-term dependencies for a longer period of time.

[14] In this paper the authors analyzed reviews in order to anticipate the sentiment, whether it is favourable or unfavourable. This strategy is suggested to use the LSTM model. One consideration must be regarded is the pack of words which is clear to distinguish the document and design the languages. The overfitting issue that resulted from the increase of LSTM units is a restriction.

When compared to other models, LSTM demonstrated the best performance. But so far, [15] employed the same classification method as [14] LSTM here. Yet, the main difference is that here, multiple datasets were analysed rather than just one.

PROPOSED SYSTEM

Sentiment analysis can be examined as a opinion categorization process. Gathering customer reviews for products is the first stage. The alternative procedure is data generation, which involves stripping the review data down and transforming it into a format that may be applied for an effective grouping. Also, the coming step is sentiment reflection used to make opinions hidden in a review visible to machines, you need to manually assign sentiment markers (positive, neutral or negative) to words and expressions. The coming step is word It is an approach of word encoding that enables words with analogous meaning to have a analogous representation. In our design we use common word embedding approach called Word2Vec model.

Proposed System Architecture

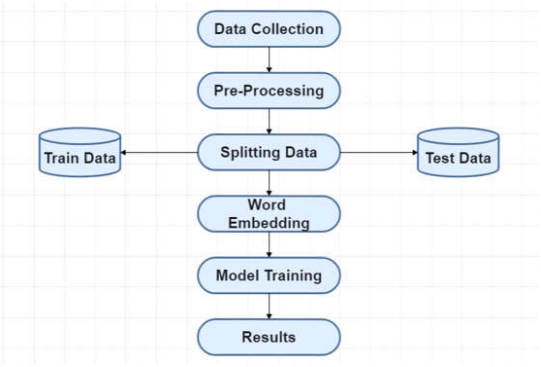

Fig 1: Proposed system architecture

Data Collection

The reviews and product details from Amazon were collected for the electronics dataset. This dataset contains attributes like (Product name, Brand name, Price, Rating, Review, helpfulness votes).

Splitting Data

The process entails splitting the dataset into two parts. One is the Training dataset which comprises 70% of the dataset, and the other is the training dataset which comprises 30% of the dataset. The first subset which is the training dataset is used to fit the model. The input is supplied to the model instead of training it is using the second subset, and its outputs are then checked to the accepted value.

Train Dataset: It is used for fitting the ml model.

Test Dataset: It is used to evaluate the trained ml model.

The goal is to evaluate how well the ml model performs on unseen data. We anticipate applying the model in this way. Where we do not have the predicted output or goal values, the forecast must first be fitted to the data that is currently available with known inputs and outcomes.

Pre-processing

The Dataset includes null characters, repetitive letters, emoticons, preceding or trailing whitespace, etc. All of this increases the problem's complexity and complicates the classification procedure.

First, some simple text cleaning was done, including removing tuples with null values, lowercase text conversion and the removal of URLs and emails. The second procedure was to eliminate vowels that appeared in a sequence for more than two times, as doing so normalizes words by making, for instance, two words that are written differently (such as gooood and good) equal. After then, the term "not" was used in place of negative constructions like "can't,""don't,""isn't,""never," etc. By using tags instead of emoticons, there are now only two types of emoticons: either positive or negative. (":)", ":-)", ":-]", etc. for "positive" and ":(", ":", ":[", ":-(", etc. for "negative"). After then, punctuation, whitespace, and English stop words like "a,""the," and "an" were eliminated.

Specialized Python packages were used for the pre-processing methods indicated above, such as the pyenchant library, which verifies the spelling of the words and changes any that are misspelt.

A label names sentiment is added. For reviews with rating above or equal to three the sentiment is labelled to 1 and for reviews below 3 the sentiment label is assigned 0.

Word embedding

Sentiment analysis in neural networks can be accomplished through supervised learning if documents are suitably represented. However, traditional techniques like bag-of-words have limitations as they do not consider word order or the context in which they are used. Moreover, such techniques produce sparse and high-dimensional vectors, making it difficult to capture semantic relationships. Word embedding technique overcomes these limitations by creating vectors of continuous real values that represent linguistic patterns and regularities. These vectors encode semantic links and context, resulting in similar words having similar vector representations. In contrast, bag-of-words does not distinguish between the different meanings of the same word. In this study, pre-trained word embeddings were used to compile single feature vectors from the reviews' phrases and words.

Word2Vec

Word2Vec is a computationally efficient predictive model it is not a single algorithm group of model architectures and optimizations are used to understand word embeddings from larger datasets. It was developed by Tomas Mikolov and colleagues in 2013 at google as an effort to improve the effectiveness of neural-network-based word embedding training, and it has now come to be accepted as the industry norm. It decodes the relationship between words into a vector after learning the link between various terms using the meaning and sense drawn from nearby words. The key advantage of the method is the ability to learn high-quality word embeddings quickly (low space and time complexity), enabling the learning of larger embeddings (more dimensions) (more features) from significantly larger text corpus (millions to billions of words).

Model comparisons

In this paper we compared the pre-trained model such as Vader, traditional models such as Logistic Regression and SVM, and deep learning model such as simple long short-term memory (LSTM) recurrent neural network.

VADER Lexicon and Rule-Based Sentiment Analysis Tool

VADER is an approach for sentiment analysis that uses a pre-built lexicon of words and their corresponding sentiment scores to determine the overall sentiment of a given text. Unlike other machine learning-based approaches that require training on large datasets, VADER is a rule-based approach that can analyse text in real-time.

The VADER approach works by first tokenizing the text into individual words and then assigning a sentiment score to each word based on its polarity (positive, negative, or neutral) and intensity (how strongly it conveys that sentiment). These scores are then combined to calculate an overall sentiment score for the entire text, ranging from -1 (most negative) to +1 (most positive), with 0 represents a neutral sentiment.

One of the key advantages of the VADER approach is its ability to handle sentiment in context. For example, the word "killer" has a negative sentiment score, but in the context of a movie review that describes a character as a "killer actor," the sentiment is likely to be positive. VADER can recognize these nuances and adjust its sentiment scores accordingly.

Overall, the VADER approach for sentiment analysis is a simple yet effective way to quickly analyse the sentiment of text data. Its rule-based approach and pre-built lexicon make it easy to implement and use, while its ability to handle sentiment in context allows it to provide more accurate results than other approaches.

Traditional Models

1. Logistic Regression

Logistic regression is a widely used statistical approach for sentiment analysis that is based on the binary classification of text data into positive and negative sentiment categories. The approach involves training a logistic regression model on a labelled dataset of text samples and their corresponding sentiment labels, using features such as word frequency, part-of-speech tags, and sentiment lexicons to represent the text data. The model then predicts the sentiment label for new, unlabelled text samples by calculating the chances that the sample comes under the good or bad sentiment category. Logistic regression is a simple and interpretable approach that can provide good accuracy for sentiment analysis tasks, especially when combined with feature engineering techniques and regularization methods to avoid overfitting.

2. **Support Vector Machine**

Support Vector Machines (SVMs) are a popular machine learning algorithm for sentiment analysis that aim to separate positive and negative text samples with a hyperplane that maximizes the margin between the two classes. SVMs are trained on a labelled dataset of text samples and their corresponding sentiment labels, using features such as bag-of-words, n-grams, and word embeddings to represent the text data. The trained model can then classify new, unlabelled text samples by determining which side of the hyperplane they fall on. SVMs are effective for sentiment analysis tasks with high-dimensional feature spaces and non-linear decision boundaries and can achieve high accuracy by properly tuning their kernel functions and regularization parameters.

Deep-learning Model

Long Short-Term Memory (LSTM) Model:

Recurrent neural networks (RNN) are advantageous for processing sequential data because, they are not like feed forward neural networks, they may use their internal "memory" to handle a succession of inputs. The same work is completed by RNN for each member of a sequence, and each output is dependent upon all previous calculations. A specific kind of RNN that can learn long-term dependencies is the LSTM network. On top of a word embedding matrix that was randomly initialised, we applied word-based LSTM. The final output in the LSTM output sequence is coupled to a single neuron with sigmoid activation function.

Results

Evaluation metrics are crucial for gauging classification performance. The most frequently used tool for this is an accuracy measure. The percentage of test datasets that the classifier correctly classifies is what is known as a classifier's correctness; however, for the text analysis the accuracy measure alone is never sufficient to provide accurate decision-making. Therefore, we used some other metrics along with accuracy measure to assess the performance of the classifier. The metrics that are used along accuracy are precision, recall, F1-score.

A) Accuracy: The amount of accurate predictions done by the model is called accuracy.

B) Precision: It is defined as the proportion of times the model classifies a sentiment as positive.

C) Recall: It determined how many actual positives the model actually determine by classifying them as positive.

D) F1-score: It is a metric for evaluating the accuracy of a binary classifier, which combines the precision and recall of the system into a single score, representing the harmonic mean of the two.

Following are the evaluation results of the various models used for sentiment analysis.

Vader Approach

	precision	recall	f1-score	support
neg	0.86	0.52	0.64	5097
pos	0.64	0.91	0.75	4903
accuracy			0.71	10000
macro avg	0.75	0.71	0.70	10000
weighted avg	0.75	0.71	0.70	10000

Fig 2: Summary of Vader model

The above figure displays the classification report of Vader model correctness of 71%.

- **Logistic Regression**

	precision	recall	f1-score	support
neg	0.84	0.88	0.86	1649.00
pos	0.87	0.83	0.85	1651.00
accuracy	0.85	0.85	0.85	0.85
macro avg	0.85	0.85	0.85	3300.00
weighted avg	0.85	0.85	0.85	3300.00

Fig 3: Summary of Logistic Regression model

The above figure displays the classification report of Logistic regression model which has the correctness of 85%.

- **Support vector machine**

	precision	recall	f1-score	support
neg	0.86	0.89	0.87	1649.00
pos	0.89	0.85	0.87	1651.00
accuracy	0.87	0.87	0.87	0.87
macro avg	0.87	0.87	0.87	3300.00
weighted avg	0.87	0.87	0.87	3300.00

Fig 4: Summary of Support Vector model

The above figure shows the classification report of SVM model which has the correctness of 87%.

- **LSTM**

```
Model: "sequential_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding_1 (Embedding)     (None, None, 128)         2560000

 lstm_1 (LSTM)               (None, 128)               131584

 dense_1 (Dense)             (None, 2)                 258

 activation_1 (Activation)   (None, 2)                 0

=================================================================
```

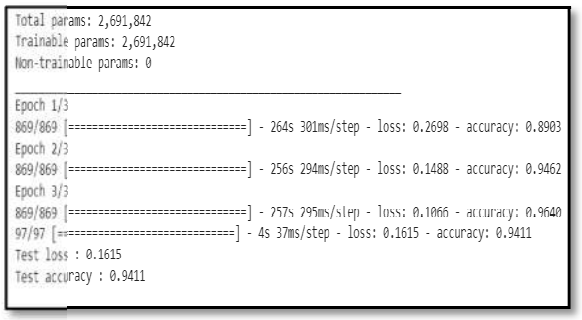

Fig 5: LSTM model Accuracy

The above figure shows the classification report of LSTM model which has the correctness of 94%.

CONCLUSION

This study compares many models in order to determine which ones work better. Using a sizable dataset, we examined numerous conventional and deep learning models.

Results Comparison

Document Representation	Model	Accuracy	F1-Score
NA	Vader	0.71	0.71
TF-IDF	Logistic Regression	0.85	0.85
TF-IDF	Support Vector Machine	0.87	0.87
Word2Vec	LSTM	0.93	0.9320

Fig 5: Comparing all models.

The analysis demonstrates that deep learning models outperform classical designs on huge datasets, but deep learning models require more training data and time to generalize effectively. The most effective design is the deep learning model the LSTM with word2vec , while the most efficient traditional model is Linear SVM classifier with TF-IDF. These skilled models can be used to automatically label sentiment in upcoming product reviews.

REFERENCES

[1] Elli, Maria Soledad, and Yi-Fan Wang. "Amazon Reviews, business analytics with sentiment analysis." 2016.

[2] Xu, Yun, Xinhui Wu, and Qinxia Wang. "Sentiment Analysis of Yelp‘s Ratings Based on Text Reviews." (2015).

[3] Rain, Callen. "Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning. "Swarthmore College.

[4] Bhatt, Aashutosh, et al. "Amazon Review Classification and Sentiment Analysis." International Journal of Computer Science and Information Technologies 6.6 (2015): 5107-5110.

[5] Chen, Weikang, Chihhung Lin, and Yi-Shu Tai."Text-Based Rating Predictions on Amazon Health & Personal Care Product Review." (2015)

[6] Tanjim Ul Haque, Nudrat Nawal Saber, Faisal Muhammad Sha. “Sentiment Analysis on Large Scale Amazon Product Reviews”. 2018.

[7] C.J. Hutto, Eric Gilbert. ” VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text”. 2015.

[8] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. In Proc. WLSM-11s.

[9] Meenakshi, Arkav Banerjee, Nishi Intwala & Vidya Sawant. “Sentiment Analysis of Amazon Mobile Reviews”. 2018.
[10] Desai, J., Majumdar, J., and Kumar K. (2016). sentiment analysis and opinion mining of online consumer reviews. A paper presented at the 2016 IEEE International Conference on Computational Intelligence and Computing (ICCIC). IEEE.
[11] Extended Short-Term Memory, S. Hochreiter and J. Schmidhuber, Journal of Neural Computation, Vol. 9 No. 8, pp. 1735-1780, 1997.
[12] A tutorial on long short-term memory recurrent neural networks called "understanding LSTM" Staudemeyer, Ralf C.
[13] Andrew Pulver, Siwei Lyu. “LSTM with Working Memory”. 2017.
[14] J. Bodapati, N. Veeranjaneyulu and S. Shaik, "Sentiment Analysis from Movie Reviews Using LSTMs", Ingénierie des systèmes d information, vol. 24, no. 1, pp. 125-129, 2019.
[15] “Text-based Sentiment Analysis Using LSTM”, Dr.G.S.N.Murthy, Shanmukha Rao Allu, Bhargavi Andhavarapu, Mounika Bagadi, and Mounika Belusonti, International Journal of Engineering Research and, vol. 9, no. 05,2020.Available: 10.17577/ijertv9is050290.