

Sentiment analysis applied on Amazon reviews

Jiaqi Li^{1,4,7}, Qi Pan^{2,5}, Yihao Wang^{3,6}

¹Faculty of Science, The University of Hong Kong, Hong Kong, 999077, Hong Kong
China

²Faculty of Science and Technology, Beijing Normal University-Hong Kong Baptist
University United International College, Zhuhai, 519087, China

³Faculty of Science and Technology, Beijing Normal University-Hong Kong Baptist
University United International College, Zhuhai, 519087, China

⁴u3597326@connect.hku.hk

⁵q030026113@mail.uic.edu.cn

⁶q030026154@mail.uic.edu.cn

⁷corresponding author

Abstract. With the rapid growth of e-commerce, accurately capturing buyers' sentiments through their reviews is increasingly vital for online marketplaces. In this paper, we aim to deal with sentiment analysis in these reviews by exploring effective methods to analyze them. We use a review dataset containing user ratings and comments on Amazon products. Applying the two-step methodology of data preprocessing and model building, we intend to employ models like LSTM and SVM to analyze Amazon customer reviews and gain insights into their performance. The findings of this study may also allow e-commerce platforms to provide better service to sellers and buyers.

Keywords: Sentiment Analysis, Amazon, e-commerce

1. Introduction

The rise of e-commerce since 2010 has dramatically reshaped global retailing, placing online marketplaces at the center of contemporary trade practices. These platforms play a pivotal role in shaping customers' shopping experiences. However, two critical challenges need addressing: the precise calculation of buyer-generated product ratings and an effective system to evaluate product review sentiment.

Accurate product ratings ensure fair product evaluation and trustworthy customer feedback. They offer invaluable insights into customer satisfaction and product quality. However, misleading ratings can distort potential buyers' perceptions, leading to customer dissatisfaction, revenue losses for sellers, and even tarnishing the platform's reputation.

Similarly, the importance of correctly ranking reviews cannot be overstated. Reviews offer detailed customer feedback and product experiences. But improperly ordered studies can lead to financial losses for consumers and manufacturers and limit the platform's utility for future buyers. Given these challenges, our research team employs the 'Amazon reviews' dataset from Kaggle for sentiment analysis.

This dataset, rich with user ratings and comments on electronics products, offers a fertile ground for our investigation.

The primary objective of this article is to tackle issues about sentiment analysis of e-commerce reviews. We harness methodologies such as SVM and LSTM to analyze user reviews and extract underlying sentiments. The insights gained from this research have the potential to significantly enhance shopping experiences, catalyze seller sales, and drive improvements in e-commerce platforms. By deepening our understanding of factors influencing customer behavior, we can expect e-commerce platforms to better cater to the needs of both consumers and producers, in line with findings from previous research [1].

2. Literature Review

Over the past few years, sentiment analysis has emerged as a highly discussed research subject in e-commerce. Many studies have been conducted to apply sentiment analysis to customer reviews.

User reviews are proven to impact customer behaviors greatly. Products with higher ratings and more positive reviews were likely to be purchased in dramatically larger quantities by customers [2]. On top of that, the study also discovered that the number and the length of reviews were also essential indicators that influences customer behaviors to a great extent [3]. The great significance buyers attach to thoughts is the main reason to use models and do sentiment analysis on online shopping reviews in e-commerce.

One frequently used method for sentiment analysis is the SVM classifier. Being a supervised machine learning algorithm, SVM categorizes data based on predefined features. In specific sentiment analysis, SVM can classify the sentiments into positive, negative, or neutral type based on certain words and phrases appearing in the text [4]. For example, they conducted research that used SVM to classify movie reviews into positive or negative sentiments.

Another highly recognized method for sentiment analysis is the LSTM neural network. LSTM can process sequential data and model temporal dependencies in texts [5]. Used in sentiment analysis, LSTM can categorize the sentiment of a sentence or paragraph by analyzing the context and relationships between words. For example, a study by Tang et al. used LSTM to classify restaurant reviews into positive or negative sentiments [6].

3. Data

In our research, we employed two distinct datasets to train and test models for sentiment analysis of Amazon reviews. The methodology was chosen by systematic design to optimize model performance.

3.1. Training Dataset

For model training, we selected a large training dataset consisting of one million reviews [7]. This dataset plays a vital role in the following aspects:

1. Maximizing the advantages of big data: Training on a large set enables us to capture the underlying patterns and structures, enhancing the generalization capability of the model.
2. Satisfying the need of deep learning: Being a deep learning model, the LSTM model requires a sufficient amount of data for effective learning. Training on a larger dataset facilitates the acquisition of more complicated features, consequently improving the model's overall performance.
3. Mitigating overfitting: Training on a large-scale dataset allows for a deeper understanding of the distribution of the data and reduces the risk of overfitting to data noise or specific sample features.

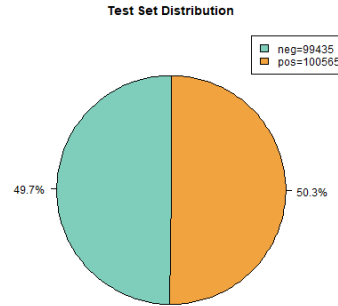


Figure 1. Test Set Distribution

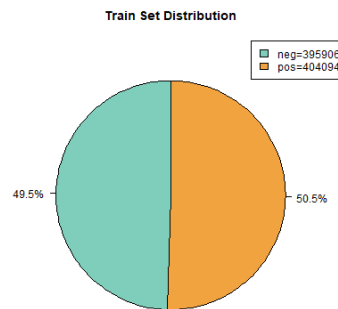


Figure 2. Train Set Distribution

3.2. Testing Dataset

A test dataset with 5000 Amazon reviews was used to evaluate the model's performance [8]. Employing this supplementary smaller dataset brings the following advantages:

1. Emulation of Real-world Scenarios: Testing on a smaller dataset better emulates real-world scenarios, providing deeper insights into the model's performance in real-life situations.
2. Flexibility and Scalability: This approach not only allows experimentation on training and test sets of various sizes but also exhibits considerable scalability, enabling adjustments based on the availability of data.

This set of 5000 data points includes textual data (the actual textual content of buyer reviews) and numerical data (the ratings given to each review, ranging from 1 star to 5 stars). In this research, we categorized the ratings from 1 to 5 stars as follows: Ratings of 1-3 stars were considered negative evaluations, while ratings of 4-5 stars were deemed positive evaluations.

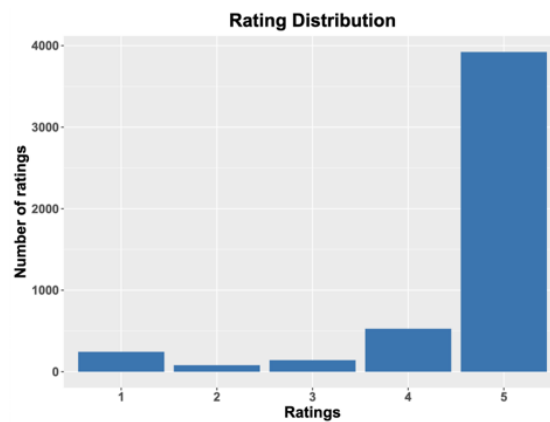


Figure 3. Rating Distribution

In conclusion, in this study, the data was partitioned into a larger training dataset and a smaller testing dataset. This approach offers advantages of capturing intricate patterns, enhancing generalization capabilities, and emulating real-world scenarios. It does not only align with the requirements of deep learning models such as LSTM but also enables a more authentic and reliable evaluation of model performance on unseen data.

4. Methodology

Our methodology follows two steps. One is preprocessing and cleaning the data; the next step is to build the model frameworks and test them with our data.



Figure 4. Flowchart of Research

4.1. Preprocessing:

In the process of data preprocessing, we did the following steps to do data cleaning:

1. Converting all numbers to 0 reduces noise and improves computational efficiency.
2. Removing all website data isn't helpful for sentiment analysis.
3. Remove special characters and spaces and lowercase all letters.

4.2. Model Structure (LSTM):

LSTM networks, which are adept at managing long-term dependencies [5], were used. LSTMs can extract critical information from reviews and adapt to varying text lengths, thus being suitable for sentiment analysis [9]. We trained and tested this model using Amazon reviews, PyTorch's nn for word encoding, and a two-layer LSTM network for sentiment prediction.

In this neural network, we employed the binary cross-entropy loss function

$$L(y, \hat{y}) = -\frac{1}{n} \sum_{i=1}^n [y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)]$$

, where

- y represents the actual labels, taking values of either 0 or 1.
- $\hat{y}$ denotes the model's predicted output, typically a value between 0 and 1, representing the probability of being classified as the positive class.
- n stands for the number of samples.

4.3. Model Structure (SVM):

The SVM algorithm was used for classifying reviews. SVM can handle linear and non-linear classifications [9], and is flexible for various applications. We implemented SVM using Python's LinearSVC and the TF-IDF method for text-to-vector conversion. The model's performance was evaluated using accuracy, recall, and F1 score metrics.

In this classifier, we employed the Hinge Loss Function $L(y) = \max(0, 1 - y \cdot f(x))$

, where

- y represents the actual label of the sample (usually -1 or 1).
- $f(x)$ stands for the model's prediction for the sample x . Essentially, it measures the degree of alignment between the predicted output and the actual label, penalizing instances of misclassification.

5. Results

Firstly, we employed a confusion matrix to showcase the results.

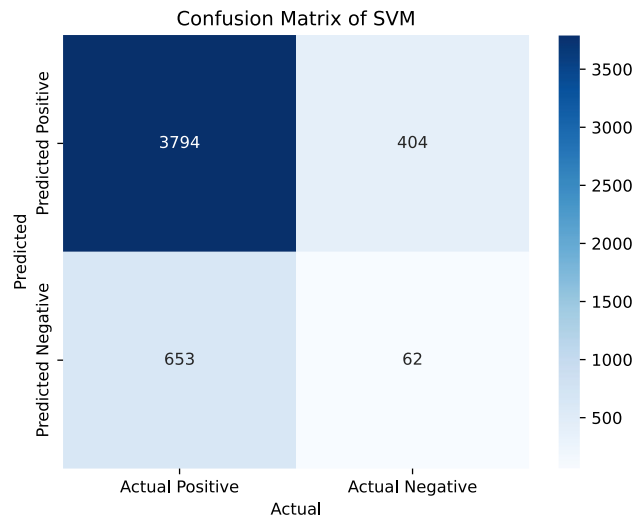


Figure 6. Confusion Matrix of SVM

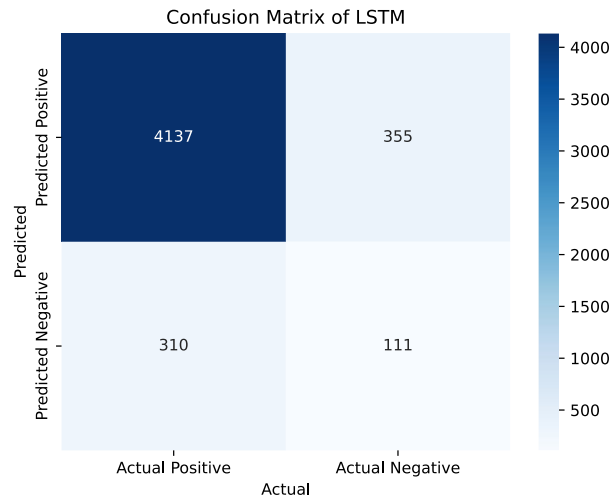


Figure 7. Confusion Matrix of LSTM

The SVM model demonstrated an F1-score of 0.9139, a precision of 0.9739, a recall of 0.8532, and an accuracy of 0.8544. These results show that the SVM model performs well regarding precision and overall classification. However, the recall value suggests that the model could still be improved in identifying positive instances more effectively.

The LSTM model achieved an F1-score of 0.9516, a precision of 0.9839, a recall of 0.9303, and an accuracy of 0.9143. The LSTM model displayed superior performance in all evaluation metrics compared to the SVM model, indicating its higher effectiveness in sentiment classification. The LSTM model's advantage is its ability to learn long-term dependencies in textual data, which is particularly valuable in sentiment analysis tasks, as it captures the intricate nuances in customer reviews [5]. On the other hand, the SVM model, although delivering reasonably accurate results, falls short in capturing complex and context-dependent patterns within text.

Table 1. Comparison of two models' indicators

Model	Accuracy	F1 Score	Precision	Recall
LSTM	91.03%	95.16%	97.39%	93.03%
SVM	85.44%	91.39%	98.39%	85.32%

After visualizing the data, we can more clearly compare the effects of the two models:

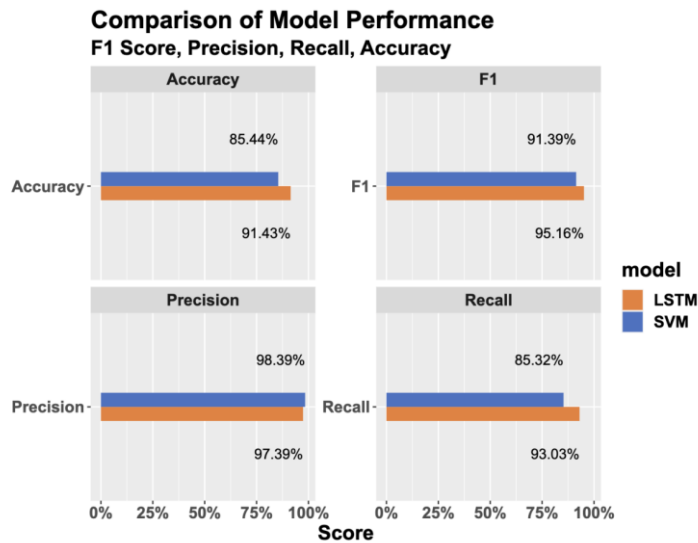


Figure 8. Visual model comparison

Regarding the figures, we can conclude that the LSTM model outperforms the SVM model in sentiment analysis of Amazon reviews. These results demonstrate the suitability of LSTM models for sentiment analysis tasks of online reviews. They also highlight the potential of the LSTM model on sentiment analysis and call for further research and development in this area.

6. Conclusion

In conclusion, our research first demonstrated the importance of sentiment analysis in e-commerce and then used proper methodology to prove the potential of LSTM models in providing valuable analysis on review sentiment. We employed a two-step methodology of data preprocessing and model building, utilizing a dataset of Amazon reviews containing user ratings and comments on various products. WE ANALYZED CUSTOMER REVIEWS using LSTM and SVM models and gained insights into their performances, respectively.

Our findings revealed that the LSTM model is generally more capable of processing sequence-based textual data and accurately capturing sentiment patterns than the SVM model. The result highlights the suitability of LSTM for sentiment analysis tasks of online reviews.

By implementing the LSTM model for sentiment analysis, e-commerce platforms can identify and prioritize trustworthy and positive reviews. This algorithm thus enables better customer purchasing experiences and higher seller revenues. On top of that, these findings may contribute to the ongoing development of algorithms and techniques for sentiment analysis in various fields, which may foster a deeper understanding of customer sentiment and improve the overall service quality on e-commerce platforms.

Acknowledge

Professor Patrick Houlihan and Teaching Assistant Beth strongly supported this work. The authors would like to thank them for both academic help and manuscript reviewing. Also, Jiaqi Li, Qi Pan, and Yihao Wang contributed equally and collaborated closely in this project and should be considered co-first authors.

References

- [1] Racherla, P., Mandviwalla, M., & Connolly, D. J. (2012). Factors affecting consumers' trust in online product reviews. *Journal of Consumer Behaviour*, 11(2), 94-104.
- [2] Hu, N., Pavlou, P. A., & Zhang, J. (2017). On self-selection biases in online product reviews. *MIS Quarterly*, 41(2), 449-471
- [3] Mudambi, S. M., & Schuff, D. (2010). What makes a helpful review? A study of customer reviews on Amazon.com. *MIS Quarterly*, 34(1), 185-200.
- [4] Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing* (pp. 79-86).
- [5] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
- [6] Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., & Qin, B. (2016). Learning sentiment-specific word embedding for twitter sentiment classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 155-160).
- [7] <https://www.kaggle.com/datasets/bittlingmayer/amazonreviews>
- [8] <https://www.kaggle.com/datasets/tarkkaanko/amazon>
- [9] Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. In *Advances in neural information processing systems* (pp. 649-657).
- [10] Schölkopf, B., & Smola, A. J. (2002). *I am learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.