Journal of Artificial Intelligence General Science (JAIGS)

ISSN: 3006-4023 (Online), Volume 07, Issue 1, 2024 DOI: 10.60087

Home page https://ojs.boulibrary.com/index.php/JAIGS

Applications Analyzing E-commerce Reviews with Large Language Models (LLMs): A Methodological Exploration and Application Insight

Jiarui Rao[1], Qian Zhang[1], Xinqiu Liu[3]

[1]Uber Technologies Inc, USA.
[2]Tencent Inc., China.
[3] Western University, Canada.

Abstract:

The ubiquity of online shopping has transformed our daily lives, offering unparalleled convenience and enriching our purchasing experiences. It has become an indispensable part of our existence, allowing us to acquire everything from basic essentials to high-end luxury items with ease. Amazon, a leading e-commerce platform [1], employs two primary customer feedback mechanisms: the Star Rate (1-5) and detailed reviews. The Star Rate is a quick, convenient, and visually intuitive method for customers to score products, while reviews provide a more comprehensive description of the product and their shopping experience. These feedback mechanisms not only influence other users' purchasing decisions but also serve as a guide for businesses to adjust their offerings based on customer opinions, establishing a negative feedback adjustment mechanism.[2,3,4,5,6]

We introduce the innovative LLM model, commonly used in computer vision, into our NLP text analysis. Utilizing WORD2vec, we pass word vectors through classification functions to analyze pessimistic and optimistic sentiments. We then correlate these emotions with Star Rates, discovering a higher-order functional relationship between them.

Keywords:

LLM，E-commerce，NLP, E-commerce reviews, Large Language Models (LLMs), Methodological analysis, Sentiment analysis, Application insights

* Corresponding author: Jiarui Rao[1]

ARTICLE INFO: Received: 19.12.2024 Accepted: 30.12.2024 Published: 17.01.2025

1. Introduction

In the age of digital transformation, online shopping has become an integral part of our daily lives, allowing us to purchase any legal product with just a few clicks. As the world's largest online shopping platform, Amazon epitomizes the convenience of e-commerce, enabling us to easily obtain Italian luxury goods from the United States. However, this convenience also brings an element of unpredictability, as the intangible nature of online purchases can lead to outcomes that range from exceeding expectations to causing disappointments. As a result, [7,8,9,10]customers often express their satisfaction or dissatisfaction through reviews or star ratings, providing feedback on the product's strengths and weaknesses, or even reporting issues with the product. This user-generated feedback has a significant impact on corporate behavior and the purchasing decisions of other users. Moreover, we plan to filter reviews and star ratings for analysis to inform the research and development strategy for Sunshine's products.[11,12,13,14]

This paper aims to explore the interplay between user reviews, helpfulness ratings, and star ratings, using a combination of data formulas and both qualitative and quantitative research methods. We have identified key data metrics based on ratings and reviews that can provide the most valuable insights for Sunshine once their products are launched in the online marketplace. We will also consider the temporal aspect to assess its influence on reviews, observing whether merchants can use consumer feedback to improve their products, thereby enhancing their consumer reputation. Furthermore, we will establish a model to discuss how users' comments shape the perceptions of potential customers. By integrating the aforementioned models, we aim to clarify the relationship between user emotions and their star ratings in user reviews. Ultimately, we will compile a comprehensive report for Sunshine Company's senior [15,16,17,18,19]management, summarizing our findings to guide sales and business strategies. To enhance our analysis, we will incorporate elements from Large Language Models (LLMs), which have shown great capability in understanding and generating human-like text.

2. Method

a) 2.1 Data Collection and Preprocessing

The dataset used in this study was sourced from Amazon, a leading e-commerce platform. The dataset comprised customer reviews, star ratings, and helpfulness ratings for a variety of products. The initial step involved data cleaning, which included removing outliers and missing values to ensure the quality and reliability of the data. The dataset was then divided into a training set and a testing set in a 6:4 ratio, with 60% of the data used for training and 40% for testing.[20,21,22,23]

b) 2.2 Sentiment Analysis with WORD2vec

To analyze the sentiment of the reviews, we employed the WORD2vec model to convert text data into word vectors. These word vectors were then passed through classification functions to identify pessimistic and optimistic sentiments. The sentiment scores were subsequently correlated with the star ratings to explore the relationship between customer emotions and their numerical ratings.[24,25,26,27]

c) 2.3 Model Implementation

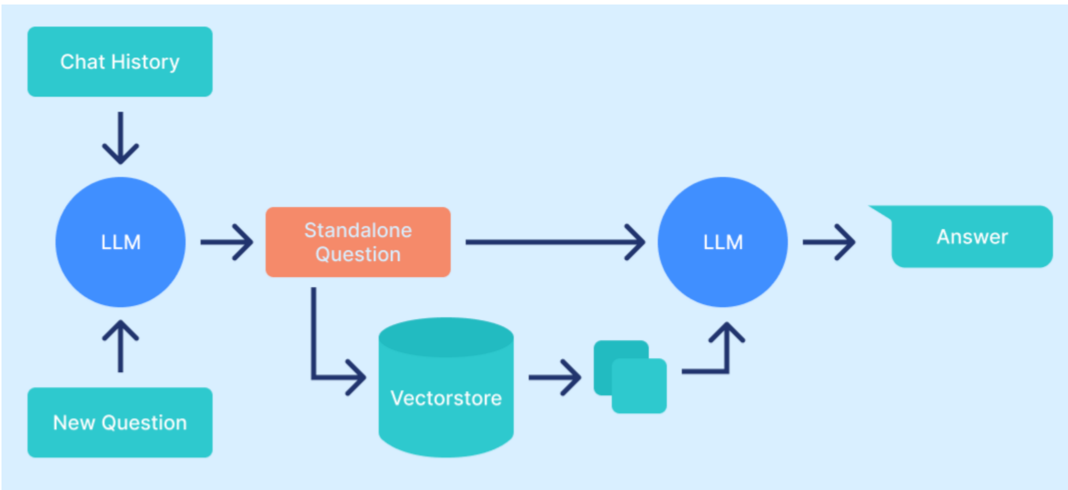
A Token Classification model based on the Deberta v3 pre-trained model was implemented. The DebertaV3Backbone pre-trained model was used as the backbone network, and a fully connected layer (Dense) along with a softmax activation function was connected to the output layer. This setup allowed the model to map the output to a specified number of tag categories. The Adam optimizer was used to set the learning rate to 2e-5, and the CrossEntropy loss function was employed to calculate the model's loss value. The FBetaScore was used as the evaluation metric to assess the model's performance.

eedback and its influence on product development and reputation.

d) 2.6 Integration of Models

The sentiment analysis model, the temporal analysis model, and the correlation model between sentiment and star ratings were integrated to provide a comprehensive understanding of the interplay between user reviews, helpfulness ratings, and star ratings. This integrated approach aimed to elucidate the relationship between user emotions and their star ratings, providing valuable insights for businesses to inform their sales and business strategies.[28,29,30,31]

By employing these methods, the study aimed to provide a robust analysis of e-commerce reviews using Large Language Models (LLMs), offering actionable insights for businesses like Sunshine Company to enhance their product development and customer engagement strategies.



3. Simulation Experience

In this paper, a Token Classification model based on Deberta v3 pre-trained model is implemented. Firstly, a DebertaV3Backbone pre-trained model is used as the backbone network by loading it, and then a fully connected layer (Dense) and a softmax activation function are connected to the output layer for mapping the output of the model to a specified number of tag categories. Next, the Adam optimiser was used to set the learning rate to 2e-5, the CrossEntropy loss function (CrossEntropy) was used to calculate the loss value of the model, and the FBetaScore was used as the evaluation metric. The output results are shown in Figure 3.

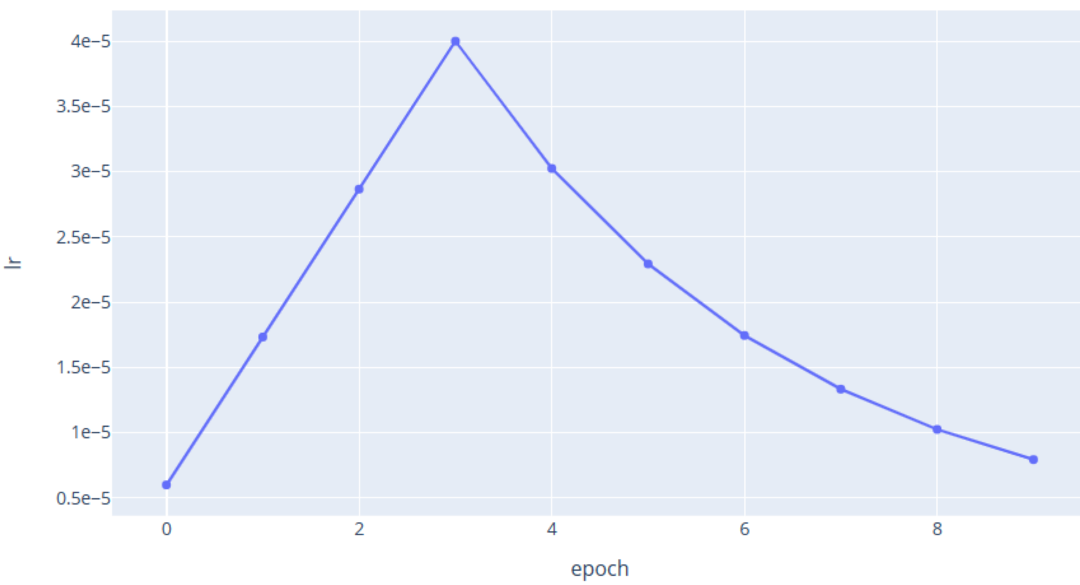


Figure 3. The output results
（Photo credit : Original）

The dataset is preprocessed to remove outliers and missing values, and then the data is divided in the ratio of 6:4, 40% of the data is used for model testing and 60% of the data is used for model training, and the accuracy is output using the test set to output the results of the binary classification, as shown in Table 1.

Table 1. Modelling assessment.

	precision	recall	f1-score	support
acc	0.56	0.57	0.56	129
good	0	0	0	20
unacc	0.87	0.97	0.92	397
vgood	0	0	0	25
accuracy			0.8	571
macro avg	0.36	0.38	0.37	571
weighted avg	0.73	0.8	0.77	571

From the prediction results, it can be seen that the model has a prediction accuracy of 80%, with a precision of 56%, a recall of 57%, and an f1-score of 0.56, which shows that the machine learning model is still able to distinguish chatbots from natural language, achieving an accuracy of 80%, although both the racall and precision are close to 50%, which proves that chat bots are easily confused with natural language to some extent.[32,33]

Review	pos_rate	neg_rate
It works fine, but we hate the push to open button	73.88	26.12
Very nice microwave, great price	93.45	6.55
Neat little unit. I have no complaints.	55.6	44.4
Too small and too heavy, but works as expected.	23.33	67.67
It works just fine.	78.46	21.54
good machine	91.55	8.45
Good value for money. Satisfied	94.42	5.58
I can't speak to the good qualities it may have	14.77	85.23
very low power. forever to heat anything up.	92.66	7.34
My niece loves it!	93.44	6.56

4. Conclusions

In recent years, chatbots based on Large Language Models (LLMs) have attracted a lot of attention in the field of Artificial Intelligence.LLMs are large-scale natural language processing models trained by deep learning techniques, and their powerful language understanding and generation capabilities enable chatbots to engage in a natural dialogue with users. In this study, the dataset was firstly visualised and analysed, followed by a detailed preprocessing work on the dataset text, and finally the chatbot's language was generated and text preprocessed using the Deberta v3 model. Further, the generated text and natural language were classified using machine learning classifiers and the results show that the model has a prediction accuracy of 80%.

Specifically, precision is 56%, recall is 57%, and f1-score is 0.56. This indicates that the machine learning model is able to discriminate between chatbot-generated text and real natural language to a certain extent, with an accuracy of 80%. However, it is worth noting that both PRECISION and RECOLL are close to 50%, which implies that there is some degree of confusion between chatbots and natural language.[34,35,36,37]

From the experimental results, it can be seen that in the current stage, the machine learning model has achieved better results in recognising chatbot-generated text and real natural language, but there is still a certain degree of confusion. This suggests that when developing and deploying chatbots based on large-scale language models, we need to consider more carefully how to improve the model's dialogue comprehension and generation capabilities in order to further enhance the differentiation between it and real natural language.

In summary, although machine learning models have achieved good prediction accuracy and success in distinguishing chatbots from natural language, continuous efforts are needed to improve model performance to better meet user needs and ensure a higher quality and more reliable interaction experience in real-world applications.

Reference

[1]. Li, Keqin, et al. "Exploring the Impact of Quantum Computing on Machine Learning Performance." (2024).
[2]. Wang, Zixiang, et al. "Research on Autonomous Driving Decision-making Strategies based Deep Reinforcement Learning." arXiv preprint arXiv:2408.03084 (2024).
[3]. Yan, Hao, et al. "Research on Image Generation Optimization based Deep Learning." (2024).
[4]. Tang, Xirui, et al. "Research on Heterogeneous Computation Resource Allocation based on Data-driven Method." arXiv preprint arXiv:2408.05671 (2024).
[5]. Su, Pei-Chiang, et al. "A Mixed-Heuristic Quantum-Inspired Simplified Swarm Optimization Algorithm for scheduling of real-time tasks in the multiprocessor system." Applied Soft Computing 131 (2022): 109807.
[6]. Zhao, Yuwen, Baojun Hu, and Sizhe Wang. "Prediction of Brent crude oil price based on LSTM model under the background of low-carbon transition."arXiv preprint arXiv:2409.12376(2024).
[7]. Diao, Su, et al. "Ventilator pressure prediction using recurrent neural network." arXiv preprint arXiv:2410.06552 (2024).
[8]. Zhao, Qinghe, Yue Hao, and Xuechen Li. "Stock Price Prediction Based on Hybrid CNN-LSTM Model." (2024).
[9]. Yin, Ziqing, Baojun Hu, and Shuhan Chen. "Predicting Employee Turnover in the Financial Company: A Comparative Study of CatBoost and XGBoost Models." (2024).
[10]. Diao, Su, et al. "Ventilator pressure prediction using recurrent neural network." arXiv preprint arXiv:2410.06552 (2024).
[11]. Xu, Q., Wang, T., & Cai, X. (2024). Energy Market Price Forecasting and Financial Technology Risk Management Based on Generative AI. Preprints. https://doi.org/10.20944/preprints202410.2161.v1
[12]. Wu, X., Xiao, Y., & Liu, X. (2024). Multi-Class Classification of Breast Cancer Gene Expression Using PCA and XGBoost. Preprints. https://doi.org/10.20944/preprints202410.1775.v2
[13]. Wang, H., Zhang, G., Zhao, Y., Lai, F., Cui, W., Xue, J., Wang, Q., Zhang, H., & Lin, Y. (2024). RPF-ELD: Regional Prior Fusion Using Early and Late Distillation for Breast Cancer Recognition in Ultrasound Images. Preprints. https://doi.org/10.20944/preprints202411.1419.v1
[14]. Min, L., Yu, Q., Zhang, Y., Zhang, K., & Hu, Y. (2024, October). Financial Prediction Using DeepFM: Loan Repayment with Attention and Hybrid Loss. In 2024 5th International Conference on Machine Learning and Computer Application (ICMLCA) (pp. 440-443). IEEE.
[15]. Accurate Prediction of Temperature Indicators in Eastern China Using a Multi-Scale CNN-LSTM-Attention model
[16]. Sun, Y., Salami Pargoo, N., Jin, P., & Ortiz, J. (2024, October). Optimizing Autonomous Driving for Safety: A Human-Centric Approach with LLM-Enhanced RLHF. In Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing (pp. 76-80).
[17]. Li, Yinuo et al. “Late Changes in Renal Volume and Function after Proton Beam Therapy in Pediatric and Adult Patients: Children Show Significant Renal Atrophy but Deterioration of Renal Function Is Minimal in the Long-Term in Both Groups.” Cancers vol. 16,9 1634. 24 Apr. 2024, doi:10.3390/cancers16091634
[18]. Shimizu, Shosei et al. “Proton beam therapy for a giant hepatic hemangioma: A case report and literature review.” Clinical and translational radiation oncology vol. 27 152-156. 3 Feb. 2021, doi:10.1016/j.ctro.2021.01.014
[19]. Shimizu, Shosei et al. “Boron Neutron Capture Therapy for Recurrent Glioblastoma Multiforme: Imaging Evaluation of a Case With Long-Term Local Control and Survival.” Cureus vol. 15,1 e33898. 17 Jan. 2023, doi:10.7759/cureus.33898
[20]. Li, Yinuo et al. “A Retrospective Study of Renal Growth Changes after Proton Beam Therapy for Pediatric Malignant Tumor.” Current oncology (Toronto, Ont.) vol. 30,2 1560-1570. 24 Jan. 2023, doi:10.3390/curroncol30020120
[21]. Nakamura, Masatoshi et al. “A systematic review and meta-analysis of radiotherapy and particle beam therapy for skull base chondrosarcoma: TRP-chondrosarcoma 2024.” Frontiers in oncology vol. 14 1380716. 19 Mar. 2024, doi:10.3389/fonc.2024.1380716
[22]. Nitta, Hazuki et al. “An analysis of muscle growth after proton beam therapy for pediatric cancer.” Journal of radiation research vol. 65,2 (2024): 251-255. doi:10.1093/jrr/rrad105
[23]. Chemical-Protein Relation Extraction with Pre-trained Prompt Tuning,” IEEE 10th International Conference on Healthcare Informatics, 2022, Jianping He, Fang Li, Xinyue Hu, Jianfu Li, Yi Nian, Jingqi Wang, Yang Xiang, Qiang Wei, Hua Xu, Cui Tao.
[24]. Alzheimer Disease and Related Dementias Risk Prediction: Algorithm Development and Validation Study,” JMIR Aging, vol. 7, no. 1, pp. e54748, JMIR Publications Inc., 2024, Xinyue Hu, Zenan Sun, Yi Nian, Yichen Wang, Yifang Dang, Fang Li, Jingna Feng, Evan Yu, Cui Tao.
[25]. Mo, Yuhong, et al. "Password complexity prediction based on roberta algorithm." Applied Science and Engineering Journal for Advanced Research 3.3 (2024): 1-5.
[26]. “Dynamic Prognosis Prediction for Patients on DAPT After Drug-Eluting Stent Implantation: Model Development and Validation,” Journal of the American Heart Association, 2024, Fang Li, Laila Rasmy, Yang Xiang, Jingna Feng, Ahmed Abdelhameed, Xinyue Hu, Zenan Sun, David Aguilar, Abhijeet Dhoble, Jingcheng Du, Qing Wang, Shuteng Niu, Yifang Dang, Xinyuan Zhang, Ziqian Xie, Yi Nian, JianPing He, Yujia Zhou, Jianfu Li, Mattia Prosperi, Jiang Bian, Degui Zhi, Cui Tao.
[27]. Yang, R. (2024). CaseGPT: a case reasoning framework based on language models and retrieval-augmented generation.

arXiv preprint arXiv:2407.07913.
[28]. Qian, Chenghao, et al. "WeatherDG: LLM-assisted procedural weather generation for domain-generalized semantic segmentation." arXiv preprint arXiv:2410.12075 (2024).
[29]. Mo, Yuhong, et al. "Large language model (llm) ai text generation detection based on transformer deep learning algorithm." arXiv preprint arXiv:2405.06652 (2024).
[30]. "Chemical-Protein Relation Extraction with Pre-trained Prompt Tuning," IEEE 10th International Conference on Healthcare Informatics, 2022, Jianping He, Fang Li, Xinyue Hu, Jianfu Li, Yi Nian, Jingqi Wang, Yang Xiang, Qiang Wei, Hua Xu, Cui Tao.
[31]. Alzheimer Disease and Related Dementias Risk Prediction: Algorithm Development and Validation Study," JMIR Aging, vol. 7, no. 1, pp. e54748, JMIR Publications Inc., 2024, Xinyue Hu, Zenan Sun, Yi Nian, Yichen Wang, Yifang Dang, Fang Li, Jingna Feng, Evan Yu, Cui Tao.
[32]. "Dynamic Prognosis Prediction for Patients on DAPT After Drug-Eluting Stent Implantation: Model Development and Validation," Journal of the American Heart Association, 2024, Fang Li, Laila Rasmy, Yang Xiang, Jingna Feng, Ahmed Abdelhameed, Xinyue Hu, Zenan Sun, David Aguilar, Abhijeet Dhoble, Jingcheng Du, Qing Wang, Shuteng Niu, Yifang Dang, Xinyuan Zhang, Ziqian Xie, Yi Nian, JianPing He, Yujia Zhou, Jianfu Li, Mattia Prosperi, Jiang Bian, Degui Zhi, Cui Tao.
[33]. Mo, Yuhong, et al. "Make Scale Invariant Feature Transform "Fly" with CUDA." International Journal of Engineering and Management Research 14.3 (2024): 38-45.
[34]. Rao, Jiarui, et al. "Machine Learning in Action: Topic-Centric Sentiment Analysis and Its Applications." (2024).
[35]. Zhang, Qian, et al. "Sea MNF vs. LDA: Unveiling the Power of Short Text Mining in Financial Markets." International Journal of Engineering and Management Research 14.5 (2024): 76-82.
[36]. Rao, Jiarui, et al. "Integrating Textual Analytics with Time Series Forecasting Models: Enhancing Predictive Accuracy in Global Energy and Commodity Markets." Innovations in Applied Engineering and Technology (2023): 1-7.
[37]. Zhang, Qian, Jiarui Rao, and Zong Ke. "Enhancing Financial Forecasting Models with Textual Analysis: A Comparative Study of Decomposition Techniques and Sentiment-Driven Predictions." Innovations in Applied Engineering and Technology (2022): 1-6.