International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 10 | Oct – 2025 DOI: 10.55041/ISJEM05103
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Sentiment Analysis of Customer Feedback using Deep Reinforcement Learning

Prathmesh Chavan … (Department of Data Science, Dr. D. Y. Patil Arts, Commerce and Science College, Pimpri)
Rajguru Bhosale ... (Department of Data Science ,Dr. D. Y. Patil Arts, Commerce and Science College, Pimpri)

ABSTRACT

Sentiment analysis of large-scale textual data is a challenging task, particularly when dealing with imbalanced class distributions commonly found in real-world review datasets. In this work, we propose a robust pipeline that integrates SBERT-based sentence embeddings with a reinforcement learning framework for multi-class sentiment classification. Specifically, we employ a Proximal Policy Optimization (PPO) Actor-Critic model, enhanced with an imbalance-aware reward mechanism and Monte Carlo dropout, to improve prediction stability and address minority-class recognition. The pipeline first transforms customer reviews into high-dimensional contextual embeddings, which are then fed into the reinforcement learning agent to make sequential classification decisions. Experiments conducted on a sizable customer review dataset demonstrate that the proposed method achieves a weighted F1 score of 0.77 and an overall accuracy of 0.8183, indicating strong performance on majority classes while partially mitigating the effects of class imbalance. The study highlights the effectiveness of combining pre-trained embeddings with policy-gradient reinforcement learning, offering a practical approach for real-world sentiment analysis tasks. Further, we discuss the implications of reward shaping and dropout for stability, providing insights into designing scalable and reliable sentiment classification systems.

Keywords:

Sentiment Analysis, SBERT Embeddings, Deep Reinforcement Learning, Proximal Policy Optimization (PPO), Imbalanced Data

1. INTRODUCTION

sentiment analysis remains a fundamental task in natural language processing (NLP), enabling organizations to extract actionable insights from large volumes of customer feedback generated on e-commerce platforms, social media, and review portals. Traditional machine learning approaches often struggle to capture deep contextual relationships in text and are particularly sensitive to class imbalance, leading to poor performance on minority classes such as neutral or negative sentiments.

Recent advances in transformer-based sentence embeddings, such as SBERT (all- mpnet-base-v2), offer rich contextual representations that encode semantic nuances effectively, providing a robust foundation for downstream sentiment classification. However, even with high- quality embeddings, conventional classifiers may not optimally align with real-world objectives, particularly when evaluation metrics like macro-F1 or minority-class recall are critical.

To address these challenges, this work integrates SBERT embeddings with a deep reinforcement learning (DRL) framework using Proximal Policy Optimization (PPO). We introduce an imbalance-aware reward mechanism that explicitly accounts for skewed class distributions, combined with Monte Carlo dropout to improve prediction stability. Further, we leverage FP16 mixed- precision training and early stopping to enhance computational efficiency and prevent overfitting. Our modular pipeline precomputes embeddings, applies optional dimensionality reduction, and trains an RL-based policy for sentiment prediction, effectively balancing performance across all classes.
This approach enables the model to prioritize minority-class recognition without sacrificing majority-class accuracy, providing a practical and scalable solution for large- scale, imbalanced sentiment analysis tasks. The proposed design also emphasizes reproducibility and efficiency, making it suitable for deployment in real-

world production systems.

2. OBJECTIVE

The primary objectives of this study are:

1. Develop a reproducible and scalable pipeline that integrates transformer-based embeddings with deep reinforcement learning (DRL) for multi-class sentiment classification of customer reviews (Negative, Neutral, Positive).
2. Mitigate the effects of severe class imbalance by employing reward shaping and class-aware optimization strategies.
3. Enhance computational efficiency in RL training by precomputing and compressing sentence embeddings. Assess model performance using macro-F1 and per-class F1 metrics to ensure improved recognition of minority classes without significant degradation on majority classes.
4. Establish a flexible methodology that can be extended to other text classification and NLP tasks.

3. Problem Statement:

Real-world customer review datasets suffer from a strong **class imbalance**, where positive reviews greatly outnumber neutral and negative ones. Traditional supervised models often overfit to the majority class, resulting in poor recognition of minority sentiments. This research addresses the imbalance factor by integrating **Deep Reinforcement Learning (DRL)** into sentiment analysis, enabling the model to learn reward-driven strategies that promote balanced performance and improve sentiment prediction across all classes in large-scale review data.4

4. METHODOLOGY

4.1 Data collection and preprocessing

We used a dataset of 205,052 Flipkart product reviews, containing text summaries and corresponding ratings. Ratings (1–5) were mapped to three sentiment classes: Negative (1–2), Neutral (3), and Positive (4– 5). The resulting class distribution was highly imbalanced: Positive – 142,607 (≈69.3%), Negative – 23,745 (≈11.6%), Neutral – 14,024 (≈6.8%). Stratified sampling was applied to create train, validation, and test splits, ensuring proportional representation of each class.

Text preprocessing included: Lowercasing all text. Removing HTML tags and punctuation. Stripping extra spaces. Removing English stopwords using NLTK.

4.2 Feature Extraction

We employed SBERT (all-mpnet-base-v2) to compute sentence embeddings for each review. These embeddings capture rich contextual information and semantic meaning of the reviews. All embeddings were precomputed and stored to reduce repeated transformer computations during reinforcement learning training.

4.3 Reinforcement Learning (RL) Formulation:

Sentiment classification is framed as a reinforcement learning task, where each embedding is an observation and the agent selects one of three actions: Negative, Neutral, or Positive. Correct predictions receive positive rewards, while incorrect ones are penalized, with class-specific weighting to handle imbalance. The agent is trained using a PPO Actor-Critic network, with Monte Carlo dropout for stability and early stopping based on validation performance. This setup enables effective optimization of class-aware rewards while maintaining robust performance across classes.

5. MODEL EVALUATION AND PERFORMANCE

The proposed sentiment classification framework, combining SBERT embeddings with a PPO Actor-Critic reinforcement learning model, was rigorously evaluated using weighted F1, Macro-F1, per-class F1 scores, and confusion matrices.

The model achieved a weighted F1 of 0.77 and an overall accuracy of 0.8183, reflecting strong performance on the predominant Positive class while moderately capturing the minority Neutral and Negative classes.

The integration of an imbalance-aware reward mechanism with Monte Carlo dropout enhanced prediction stability and improved representation for underrepresented classes. These results demonstrate that the combination of precomputed contextual embeddings and reinforcement learning provides a robust and scalable solution for imbalanced multi-class sentiment analysis.

Precomputed embeddings reduced training time, and confusion matrix analysis showed most misclassifications occurred between Neutral and Negative classes, highlighting areas for potential improvement in reward shaping or sampling strategies.

Method	Approach	Weighted F1
PPO Actor-Critic	Imbalance-Aware Reward + MC Dropout	0.77
PPO Actor-Critic	Imbalance-Aware Reward + FP16 + MC Dropout + Early Stopping	0.71
DRL Actor-Critic + MC Dropout	Weighted Focal Reward + MC Dropout	0.78
DRL Actor-Critic	Class-weighted rewards	0.78
DRL REINFORCE	Class-weighted rewards	0.78

Per-class F1 (Proposed): Negative 0.78, Neutral 0.68, Positive 0.80

5.RESULT ANALYSIS

The PPO-based sentiment classifier with SBERT embeddings achieved an overall accuracy of 0.8183 and weighted F1 of 0.77, performing best on Positive sentiments while Neutral and Negative classes had lower scores due to class imbalance

Monte Carlo dropout improved prediction stability by reducing variance, and the imbalance-aware reward guided policy optimization to better recognize minority classes.

7. CONCLUSION

This study presents a robust framework for multi-class sentiment analysis by integrating SBERT embeddings with a Proximal Policy Optimization (PPO) Actor-Critic model. The imbalance-aware reward and Monte Carlo dropout enhanced prediction stability and addressed class imbalance, enabling more reliable recognition of minority sentiment classes. Experimental results on large-scale customer review data demonstrate that the proposed method achieves high accuracy and weighted F1 scores while maintaining balanced performance across classes. The findings highlight the effectiveness of combining pre-trained contextual embeddings with reinforcement learning for real-world sentiment classification. Future work may explore integrating more advanced transformer models, dynamic reward shaping, and scaling to multilingual or multi-domain datasets.

REFERENCES

- Devlin, J., et al. (2019). BERT: Pre- training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT.
- He, P., Gao, J., & Chen, W. (2021). DeBERTa: Decoding-enhanced BERT with Disentangled Attention. ICLR.
- Schulman, J., et al. (2017). Proximal Policy Optimization Algorithms. arXiv preprint.
- Sanh, V., et al. (2019). DistilBERT: a distilled version of BERT. NeurIPS Workshop.