Comparative Study of Deep Learning Models for News Text Classification Using CNN, LSTM, and BiGRU

Team Members: Mohammad Varaliya (Leader), Abhishek Rane, Ayush Bhalgat, Shreyash Patekar, Dhruv Panchal

Mentor: Prof. Mayura Nagar

Institute: K. J. Somaiya Institute of Management, Somaiya Vidyavihar University, Mumbai

Abstract

This study compares the performance of different deep learning architectures, including Convolutional Neural Networks (CNN), Long Short-Term Memory Networks (LSTM), and Bidirectional Gated Recurrent Units (BiGRU), for the task of news text classification. Traditional machine learning models such as Logistic Regression and Support Vector Machines (SVM) achieved near 78% accuracy, while CNN provided improved results through local pattern recognition. LSTM underperformed due to training instability and convergence issues. BiGRU demonstrated the highest accuracy of 80.94% and a weighted F1-score of 80.42%, underscoring the value of incorporating bidirectional contextual understanding for text classification tasks.

1. Introduction

The rapid expansion of digital news platforms has resulted in a continuous influx of textual information, making manual classification inefficient and inconsistent. Automated text classification systems are essential for organizing information, enabling targeted content delivery, and improving search relevance. While traditional machine learning approaches rely on manual feature engineering, deep learning methods automatically learn semantic and contextual representations. This transition has significantly enhanced performance in natural language processing applications.

2. Literature Review

Kim (2014) demonstrated that CNNs effectively capture local n-gram patterns useful for sentence classification. Cho et al. (2014) introduced the GRU architecture to address the limitations of standard RNNs, particularly vanishing gradients. Liu et al. (2016) highlighted that bidirectional models, which process text sequences in both forward and backward directions, lead to improved contextual understanding and classification accuracy.

3. Dataset & Preprocessing

The dataset used in this study consists of 209,527 news articles sourced from an online news publication. Each sample includes fields such as headline, short description, authors, publication date, and category. To achieve balanced classification, the top 10 most frequently occurring categories were selected, resulting in 124,787 samples for training and evaluation. The preprocessing pipeline included text normalization by converting text to lowercase, removing punctuation, URLs, and special characters. Stopwords were filtered out to remove common but contextually irrelevant terms. Lemmatization was applied to reduce words to their base forms. For machine learning models, TF-IDF vectorization generated sparse text representations. For deep learning models, tokenization mapped each word to an index, followed by padding sequences to a uniform length of 100 tokens. This ensured consistent input dimensions across the neural architectures.

4. Model Architecture

CNN: The Convolutional Neural Network used 1D convolution filters to identify meaningful local patterns across text. It effectively captured phrase-level context. LSTM: Designed to capture long-range dependencies, LSTM networks contain gating mechanisms to manage memory flow. However, in this study, the model experienced training instability and failed to converge well. BiGRU: The Bidirectional GRU processes information in both forward and backward directions, enabling richer context representation. This model achieved the best results, indicating the importance of bidirectional context in language modeling.

5. Model Performance Comparison

Model	Accuracy	F1-Score
Logistic Regression	77.91%	76.90%
SVM	77.79%	76.83%
CNN	79.76%	79.16%
LSTM	28.53%	12.66%
BiGRU	80.94%	80.42%

6. Applications

• News categorization and organization • Personalized content recommendation systems • Media trend tracking and analytics • Detection and monitoring of misinformation

7. Challenges

• Headlines may lack sufficient context • Vocabulary overlap leads to category confusion • Deep models require high computational resources and memory

8. Future Scope

Future enhancements include the use of pretrained embeddings such as GloVe and fastText to provide richer semantic vector representations. Additionally, Transformer-based models such as BERT and GPT can be fine-tuned for improved contextual encoding and deeper semantic understanding.

9. Conclusion

The results clearly demonstrate that the BiGRU model outperforms CNN and LSTM due to its superior ability to capture bidirectional contextual dependencies. This makes BiGRU a strong candidate for real-world text classification applications.

References

[1] Kim, Y. (2014). CNN for Sentence Classification.
[2] Cho et al. (2014). GRU Networks.
[3] Liu et al. (2016). Bidirectional RNN for Text Classification.