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

Key Highlights of the Study:

- Dataset refined to top 10 categories (124,787 records)
- Logistic Regression and SVM gave ~78% baseline performance
- CNN improved performance by capturing local semantic features
- LSTM underfit due to training/convergence limitations
- BiGRU achieved highest accuracy (80.94%) and F1-score (80.42%)
- Demonstrated importance of bidirectional contextual learning
- Suggests next-step improvements with GloVe and Transformer models

The rapid expansion of online news sources has led to an overwhelming amount of textual information published on a daily basis. Categorizing news articles manually is not only time-consuming but also inconsistent due to human subjectivity. Therefore, automated text classification systems have become essential for organizing and processing news content effectively. This project examines the performance of multiple machine learning and deep learning architectures to determine the most effective model for news text classification. The dataset used in this study contains approximately 210,000 news articles, each associated with metadata such as headline, short description, publication date, and category label. To maintain uniformity and reduce imbalance, only the ten most frequent categories were selected for model training and evaluation, resulting in 124,787 samples. These samples underwent a systematic preprocessing pipeline that included text normalization, stopword removal, tokenization, and lemmatization. The cleaned dataset was then split into training, validation, and test sets, ensuring reliable performance comparison across models. Traditional machine learning models were implemented initially to establish baseline performance. Logistic Regression and Support Vector Machines (SVM) were trained using TF-IDF feature vectors. Both models performed similarly, reaching an accuracy of around 78%. While this level of performance indicates that linear classifiers can capture basic text relevance, they are limited in their ability to understand deeper semantic and contextual relationships within language. To address this limitation, advanced deep learning models were introduced. These models work by learning dense vector representations of text and capturing both local and sequential dependencies. The first neural model tested was a Convolutional Neural Network (CNN). CNNs are known for their ability to detect meaningful word patterns and phrase structures by applying convolution operations across text sequences. This model achieved an accuracy of 79.76%, outperforming the baseline models and showing that local context contributes significantly to classification quality. The next architecture tested was the Long Short-Term Memory (LSTM) network.

LSTM networks are designed to remember longer sequences by mitigating the vanishing gradient problem commonly observed in standard recurrent neural networks (RNNs). However, in this study, the LSTM model struggled to converge effectively, resulting in an accuracy of just 28.53%. The underperformance suggests training instability or insufficient tuning, indicating that LSTM alone was not suitable for the dataset under the given configuration. The final architecture evaluated was the Bidirectional Gated Recurrent Unit (BiGRU). GRUs are simplified variants of LSTM that retain essential gating mechanisms while reducing computational overhead. The bidirectional configuration allows the model to read sequences both forwards and backwards, capturing context more thoroughly than standard RNNs. This model demonstrated superior performance, achieving an accuracy of 80.94% and a weighted F1-score of 80.42%. The success of BiGRU highlights the importance of incorporating bidirectional context for improved semantic understanding in news text classification. Overall, this study highlights the significance of contextual modeling in text classification and establishes BiGRU as a strong candidate for applications involving news categorization.