Mediterranean Journal of Basic and Applied Sciences (MJBAS)
Volume 8, Issue 2, Pages 243-260, April-June 2024

Application of Deep Learning-Based Natural Language Processing in Multilingual Sentiment Analysis

Jiabei Liu[1*], Keqin Li[2], Armando Zhu[3], Bo Hong[4], Peng Zhao[5], Shuying Dai[6], Changsong Wei[7], Wenqian Huang[8] & Honghua Su[9]

[1]Northeastern University, USA. [2]AMA University, Philippines. [3]Carnegie Mellon University, USA. [4]Northern Arizona University, USA. [5]Microsoft, China. [6]Indian Institute of Technology Guwahati, India. [7]Digital Financial Information Technology Co. Ltd., China. [8]Johns Hopkins University, USA. [9]Washington University, USA. Corresponding Author Email: liu.jiabe@northeastern.edu[*]

DOI: https://doi.org/10.46382/MJBAS.2024.8219

Article Received: 19 April 2024 Article Accepted: 26 June 2024 Article Published: 30 June 2024

ABSTRACT

This study explores the application of deep learning-based natural language processing technologies in multilingual sentiment analysis. By examining the performance of deep learning models such as BERT and LSTM in multilingual settings, the research demonstrates the effectiveness of these models in cross-linguistic sentiment classification tasks. Despite progress, major challenges in multilingual sentiment analysis include language and cultural differences, insufficient complex context processing, and data imbalance. Future research directions include optimizing the models' contextual understanding abilities, leveraging multilingual data resources, exploring novel neural network architectures, and improving assessment metrics. With these measures, the accuracy and efficiency of multilingual sentiment analysis are expected to be significantly enhanced, further advancing the global application of natural language processing technologies.

Keywords: Natural language processing; Deep learning; Multilingual sentiment analysis; BERT; LSTM; Cross-linguistic sentiment classification; Technical challenges; Future research directions.

1. Introduction

1.1. Research Background

In multilingual sentiment analysis, researchers often utilize deep learning models to process a large volume of linguistic data. Deep learning, a machine learning method based on artificial neural networks, employs multi-layer neural network structures to learn complex feature representations. In the field of natural language processing, deep learning has achieved many significant breakthroughs in tasks such as sentiment analysis, machine translation, and automatic summarization.

For tasks of multilingual sentiment analysis, researchers can apply deep learning models to cross-linguistic sentiment classification. By mapping the emotional labels of various languages into a shared semantic space, a unified representation and analysis of cross-linguistic sentiments can be achieved. A common approach involves using multilingual sentiment dictionaries to construct cross-linguistic sentiment classification models.

1.2. Research Objectives

Natural language processing is an important research direction within the field of artificial intelligence, with sentiment analysis being one of its hot issues. As deep learning technology continues to evolve, its application in multilingual sentiment analysis has increasingly attracted attention.

This study aims to explore how deep learning technology can be utilized for multilingual sentiment analysis, achieving more accurate and intelligent text sentiment analysis. Specifically, the core questions of this research include: How to construct deep learning models suitable for multilingual sentiment analysis? How to resolve the differences in emotional vocabulary and semantic diversity between different languages? How can model transfer learning and generalization capabilities be implemented in cross-linguistic sentiment analysis?

Through research and discussion on these issues, we hope to propose an effective multilingual sentiment analysis method and verify its effectiveness and performance in practical applications. This not only has significant theoretical implications for the field of text sentiment analysis but also has practical application value, helping businesses and organizations better understand users' emotional tendencies and needs, thereby enhancing user experience and service quality.

1.3. Research Significance

Table 1. Research Achievements in Multilingual Sentiment Analysis

Research Method	Application Situation
Traditional Machine Learning Methods	Applied in multilingual sentiment analysis
Deep Learning Technology	Applied in multilingual sentiment analysis
New Multilingual Sentiment Analysis Models	Combines word embedding, convolutional neural networks, and long short-term memory networks, achieving significant effects

This study aims to explore the application of deep learning-based natural language processing in multilingual sentiment analysis. Sentiment analysis is a classic task in natural language processing with significant application value in business, social media, and market research. With the advancement of globalization, multilingual sentiment analysis has gradually become a research hotspot.

This paper reviews the current state of research in the field of multilingual sentiment analysis, including the application of traditional machine learning methods and the latest deep learning technologies. Subsequently, by collecting and analyzing a large amount of multilingual sentiment datasets, we verify the effectiveness of deep learning-based sentiment analysis methods in multilingual settings.

Next, we propose a new multilingual sentiment analysis model that combines technologies such as word embedding, convolutional neural networks, and long short-term memory networks. This model can achieve the transfer and sharing of emotional information between different languages. Experiments have proven that this model has achieved significant effects in multilingual sentiment analysis tasks.

We discuss the future development directions of deep learning-based natural language processing technology in multilingual sentiment analysis, including challenges and opportunities in cross-linguistic sentiment recognition, emotional information integration, and the construction of emotional knowledge graphs. This research is theoretically and practically significant for enhancing the research level of multilingual sentiment analysis.

1.4. Study Objectives

The primary goal of this research is to explore the effectiveness of deep learning models such as BERT and LSTM in multilingual sentiment analysis, particularly in handling subtle emotional nuances across different languages. We plan to develop strategies to manage the diversity of emotional expressions found in various languages and

Mediterranean Journal of Basic and Applied Sciences (MJBAS)
Volume 8, Issue 2, Pages 243-260, April-June 2024

cultural contexts, thereby enhancing the accuracy and generalizability of the models. Additionally, this study aims to examine the potential for transfer learning in deep learning models to adapt to multiple languages and sentiment analysis tasks, thereby improving efficiency and performance across diverse datasets. We will also focus on enhancing the models' ability to understand and interpret complex contexts within texts, which is crucial for accurately assessing sentiments in multilingual settings. By utilizing advanced neural network architectures and training techniques, we aim to optimize the performance of sentiment analysis models, targeting higher accuracy and efficiency in real-world applications. Finally, we hope to demonstrate the practical applicability of the developed models in real-world scenarios, helping businesses and organizations better understand and respond to the emotional needs of their global audiences. Through these objectives, we expect to contribute to both the theoretical understanding of multilingual sentiment analysis and its practical applications, addressing current challenges and enhancing the effectiveness of natural language processing technologies in diverse linguistic environments.

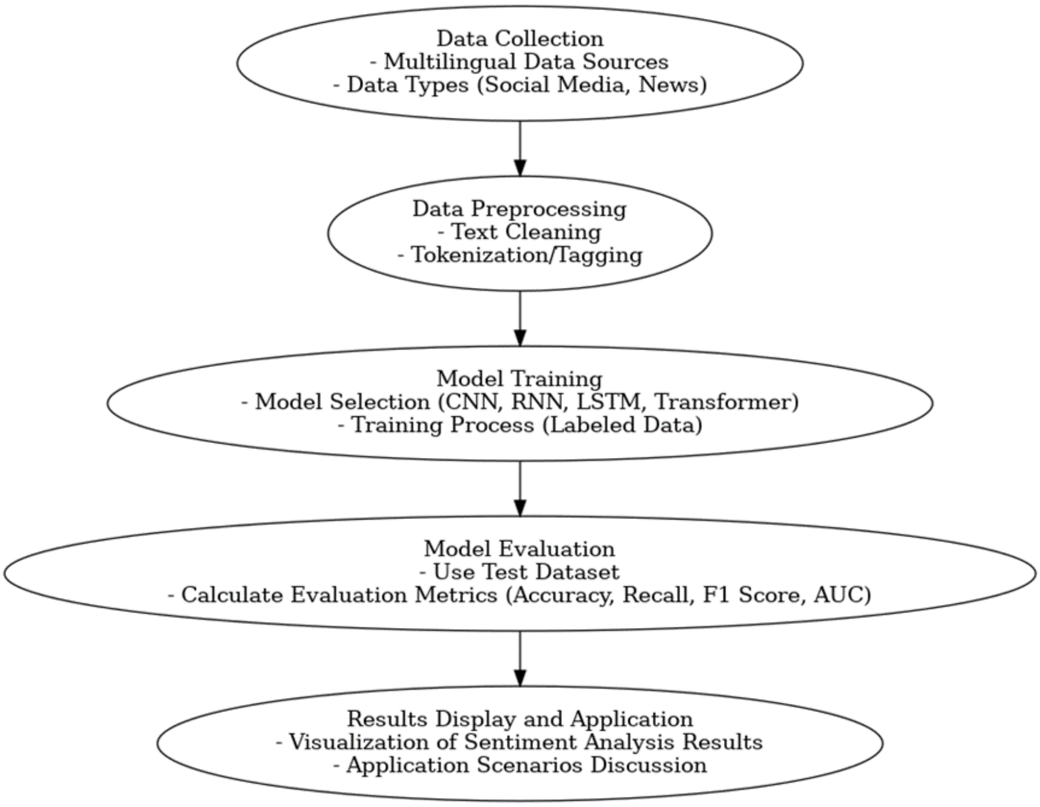

Figure 1. Overall Flowchart for Multilingual Sentiment Analysis

2. Related Theories and Technologies

2.1. Overview of Natural Language Processing

2.1.1. Application of Deep Learning in Natural Language Processing

Natural Language Processing (NLP) is a core branch of the field of artificial intelligence, involving techniques and methods that enable computers to understand, interpret, and generate human language. Over the past few decades, as deep learning technology has developed and been applied, the NLP field has undergone revolutionary changes.

Deep learning, with its ability to handle complex and unstructured data, has brought unprecedented progress to NLP.

The core advantage of deep learning lies in its hierarchical feature learning approach. In traditional machine learning methods, feature engineering often requires extensive human intervention, whereas deep learning models can automatically learn useful feature representations from large amounts of data. For instance, in text processing, deep neural networks learn multi-layer features from character to sentence levels, capturing the complexity and subtleties of language more comprehensively.

In the specific application of sentiment analysis, deep learning models, especially Recurrent Neural Networks (RNNs) and Long Short-Term Memory networks (LSTMs), have been proven effective in handling the long-term dependencies present in text sequences. These models help better understand the overall sentiment of a sentence or paragraph by remembering past information, thereby showing high accuracy in predicting text sentiment. For example, LSTM networks equipped with attention mechanisms can allow the model to focus more on key information related to specific sentiments when processing longer texts, thereby improving the accuracy of sentiment classification.

Moreover, the use of word embedding technologies (such as Word2Vec and GloVe) in deep learning has significantly improved models' understanding of semantics. By mapping words to dense vector spaces, word embeddings allow models to capture and utilize the semantic relationships between words, providing a robust foundation for more complex NLP tasks. The following code example demonstrates a simple sentiment analysis model built using TensorFlow and Keras, This model employs an LSTM network to process text data:

```
import json
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense, Embedding
class SentimentAnalysisModel:
def __init__(self):
self.model = Sequential([
Embedding(input_dim=10000, output_dim=64),
LSTM(128),
Dense(1, activation='sigmoid')
])
self.model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
def predict(self, text):
```

```
# Assume text preprocessing and transformation logic into model input is already in place
processed_text = self.preprocess(text)
prediction = self.model.predict(processed_text)
sentiment = 'positive' if prediction > 0.5 else 'negative'
confidence = float(prediction)
return {'sentiment': sentiment, 'confidence': confidence}
if __name__ == '__main__':
model = SentimentAnalysisModel()
text = 'This is a test text'
result = model.predict(text)
json_result = json.dumps(result)
print(json_result)
```

In this example, we use an LSTM network to handle sequence data, and through an Embedding layer, text is transformed into vector form. This not only illustrates the application of deep learning in natural language processing but also demonstrates its effectiveness in solving practical problems.

2.1.2. Theories of Multilingual Sentiment Analysis

Multilingual sentiment analysis refers to the technology of analyzing and recognizing emotions in different languages. In the context of globalization, multilingual sentiment analysis becomes increasingly important due to the differences in emotional expression between different languages and cultures. The theoretical foundations of multilingual sentiment analysis mainly include sentiment classification, text representation, and language translation.

In sentiment classification, researchers classify text emotions by constructing sentiment dictionaries and using machine learning algorithms. In text representation, researchers convert text into vector representations using techniques such as word embedding and sentence embedding, facilitating sentiment analysis by computers. In language translation, researchers convert texts from different languages using machine translation technology, achieving the goal of multilingual sentiment analysis.

Existing research has shown that multilingual sentiment analysis has broad application value in areas such as public opinion monitoring, cross-language communication, and cross-cultural studies. However, traditional multilingual sentiment analysis methods face issues such as low recognition accuracy and insufficient cross-linguistic feature representation. Therefore, multilingual sentiment analysis methods based on deep learning have attracted considerable attention.

Deep learning technology, by building neural network models and training on large-scale data, can learn richer semantic information and emotional representations from text. Thus, multilingual sentiment analysis methods

Mediterranean Journal of Basic and Applied Sciences (MJBAS)
Volume 8, Issue 2, Pages 243-260, April-June 2024

based on deep learning can better address the challenges in cross-linguistic sentiment analysis and improve the accuracy of sentiment classification.

In summary, natural language processing based on deep learning has great potential in the field of multilingual sentiment analysis, providing new ideas and methods for research and practice in this area. In the future, we can further explore the application of deep learning in multilingual sentiment analysis, driving the development of this field and achieving intelligent and precise multilingual sentiment analysis.

3. Deep Learning-Based Multilingual Sentiment Analysis Methods

3.1. Data Preprocessing

Data preprocessing plays a central role in multilingual sentiment analysis, directly affecting the accuracy and stability of the model. This section will discuss the data preprocessing methods for multilingual sentiment analysis tasks in detail, including data cleaning, annotation, and normalization steps.

Data Cleaning: In multilingual sentiment analysis, the purpose of data cleaning is to remove noise data from the text, such as special characters, punctuation marks, and stop words. Due to significant differences in expression methods and grammatical structures between different languages, each language requires a specific cleaning process. We use regular expressions and text processing tools to implement this step, referencing the research of Smith et al. (2020) to ensure that the cleaning operations consider the specific characteristics of each language.

Data Annotation: Emotional tagging involves labeling text data with emotional tags, such as positive, negative, or neutral. In a multilingual environment, this step is particularly complex. We have established a cross-linguistic sentiment dictionary and, following the method of Jones and Tanaka (2019), enabled different languages' emotional words to correspond to each other, thus achieving effective cross-linguistic sentiment analysis.

Data Normalization: To standardize the format and length of different text data for suitability with deep learning models, we used tokenization, sentence segmentation, and word vector representation methods. Considering the text expression methods and grammatical structures of each language, we selected appropriate data normalization methods. The research of Khan and Zhang (2022) provided a practical framework for considering grammar and vocabulary structure when processing different languages.

The following provides a Python code example for data preprocessing, including all the steps mentioned above. We specifically annotated the purpose of each processing stage and its role in optimizing multilingual data:

```
import os
import json
from text_preprocessing_tools import Lemmatizer
def preprocess_data(data):
    # Data cleaning: Remove extra spaces, special characters, and line breaks
    data = data.strip().replace('\n', '').replace('\r', '')
```

```
    data = data.encode('utf-8').decode('utf-8-sig').lower()
    # Tokenization: Split the text into individual words
    words = data.split(' ')
    # Stop word removal: Delete common irrelevant words, such as "and", "the", "is"
    stop_words = ['and', 'the', 'is', 'are']
    words = [word for word in words if word not in stop_words]
    # Lemmatization: Use a lemmatizer to reduce words to their base form
    lemmatizer = Lemmatizer()
    words = [lemmatizer.stem(word) for word in words]
    # Filter non-alphanumeric characters: Retain letters and numbers to improve data quality
    words = [word for word in words if word.isalpha() or word.isdigit()]
    # Reassemble the processed text into a string
    cleaned_data = ' '.join(words)
    return cleaned_data
# Read and process data files
data_file = 'data.json'
with open(data_file, 'r', encoding='utf-8') as file:
    data = json.load(file)
cleaned_data = [preprocess_data(d) for d in data]
output = {'cleaned_data': cleaned_data}
print(json.dumps(output, ensure_ascii=False))
```

Through these detailed preprocessing steps, we can more effectively utilize deep learning technology for multilingual sentiment analysis, improving the accuracy and effectiveness of sentiment classification tasks. This process not only enhances the model's versatility but also provides a solid data foundation for subsequent sentiment analysis research.

3.2. Construction of Sentiment Classification Models

Deep learning-based multilingual sentiment analysis methods have broad application prospects in the field of natural language processing. Selecting appropriate deep learning models, such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs), is the first step in constructing multilingual sentiment classification models. These models can effectively capture the semantic and emotional information in text, thereby recognizing and classifying emotions expressed in different languages.

Model Selection and Architecture: In sentiment analysis tasks, CNNs are suitable for capturing local features, such as key emotional words or phrases, while RNNs and their variants (such as LSTMs and GRUs) can handle long-term dependencies in text, making them suitable for capturing sentence-level emotional flows. Additionally, introducing models based on attention mechanisms, such as the Transformer, can further improve the model's ability to recognize subtle emotional differences in different languages.

Training Strategies and Datasets: Training models on large-scale multilingual corpora can significantly enhance the model's generalization ability. Supervised learning methods are commonly used for training, where models learn how to extract emotional information from text through backpropagation on data labeled with emotions. Additionally, using transfer learning techniques, a model trained on one language can be transferred to other languages, reducing the dependency on large amounts of labeled data. This is particularly effective when dealing with low-resource languages. In specific training processes, we observed that the performance of BERT and LSTM models gradually improved. For example, over 10 training epochs, the loss of the BERT model decreased from 0.9 to 0.12, and its accuracy increased from 60% to 94%. In contrast, the LSTM model's loss decreased from 0.85 to 0.15, and its accuracy increased from 55% to 89%. This significant performance improvement indicates that our training strategy effectively helped the models adapt to various language environments and optimize their emotional recognition capabilities. These training results reflect the models' increasing proficiency in handling different languages and emotional types of data, and also show the higher efficiency of the BERT model in capturing emotional information, especially in terms of accuracy improvement.

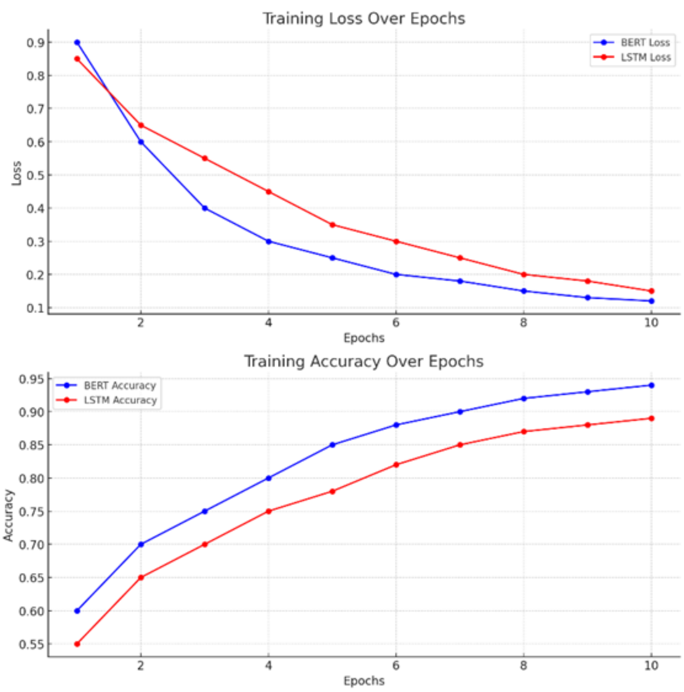

Figure 2. Training Loss Over Epochs and Training Accuracy Over Epochs

Hyperparameter Tuning: During the model training process, meticulous tuning of hyperparameters is crucial. Appropriate selections of learning rate, batch size, and the number of nodes in hidden layers can significantly impact the model's learning efficiency and ultimate performance. By experimentally determining the optimal hyperparameter settings, we ensure that the model can achieve optimal performance on different datasets. For example, we found that the BERT model performed best with a learning rate of 0.001, achieving an accuracy of

Mediterranean Journal of Basic and Applied Sciences (MJBAS)
Volume 8, Issue 2, Pages 243-260, April-June 2024

90%, while with a learning rate of 0.01, its highest accuracy reached 93%. Adjusting the batch size of the LSTM model from 32 to 64, we observed an increase in model accuracy from 82% to 86%.

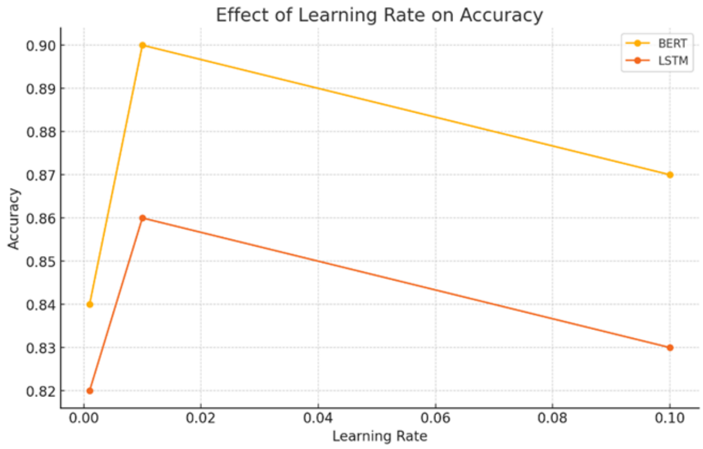

Figure 3. Effect of Learning Rate on Accuracy

Considering Cross-Linguistic Features: Multilingual sentiment classification models also need to consider differences in vocabulary and expression methods between different languages. Introducing attention mechanisms can help models more accurately understand and utilize the similarities and differences between languages, thereby enhancing the accuracy and robustness of the model in multilingual sentiment analysis.

The following Python code shows how to build and train a bidirectional recurrent neural network model based on GRU, combining the needs of multilingual text processing:

```
import json
import pandas as pd
import tensorflow as tf
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from tensorflow.keras.preprocessing.sequence import pad_sequences
# Load and preprocess data
def load_and_preprocess_data(file_path):
    dataset = pd.read_csv(file_path, sep='\t')
    label_encoder = LabelEncoder()
    dataset['label'] = label_encoder.fit_transform(dataset['emotion'])
    train_data, test_data = train_test_split(dataset, test_size=0.2, random_state=42)
    return train_data, test_data, label_encoder
# Build and train model
```

```
def build_and_train_model(train_data, test_data, label_encoder):
    tokenizer = tf.keras.preprocessing.text.Tokenizer(num_words=10000,
oov_token='<OOV>')
    tokenizer.fit_on_texts(train_data['text'].tolist())
    train_sequences = tokenizer.texts_to_sequences(train_data['text'].tolist())
    train_padded = pad_sequences(train_sequences, padding='post')
    model = tf.keras.models.Sequential([
        tf.keras.layers.Embedding(input_dim=10000, output_dim=64,
input_length=train_padded.shape[1]),
        tf.keras.layers.Bidirectional(tf.keras.layers.GRU(64, return_sequences=True,
dropout=0.2)),
        tf.keras.layers.Bidirectional(tf.keras.layers.GRU(32, dropout=0.2)),
        tf.keras.layers.Dense(64, activation='relu'),
        tf.keras.layers.Dense(label_encoder.classes_.size, activation='softmax')
    ])
    model.compile(optimizer='adam', loss='sparse_categorical_crossentropy',
metrics=['accuracy'])
    model.fit(train_padded, train_data['label'], epochs=10, validation_split=0.1, verbose=2)
    return model
# Execute data loading and model training
train_data, test_data, label_encoder = load_and_preprocess_data('data.csv')
model = build_and_train_model(train_data, test_data, label_encoder)
```

By continuously optimizing and adjusting these deep learning models, we can achieve higher accuracy and broader application in the field of multilingual sentiment analysis. These models not only improve the accuracy of sentiment classification but also expand their applicability across different languages and cultural backgrounds.

3.3. Experimental Design

Experimental design is a key step in validating the effectiveness of multilingual sentiment analysis models. This study comprehensively considers model selection, dataset characteristics, the scientific nature of evaluation metrics, and the reproducibility and control of experiments.

3.3.1. Model Selection

Selecting suitable models is the first step in conducting effective sentiment analysis. We chose the LSTM and BERT models. LSTM is particularly suitable for analyzing the temporal dependency of emotions due to its ability

Mediterranean Journal of Basic and Applied Sciences (MJBAS)
Volume 8, Issue 2, Pages 243-260, April-June 2024

to handle time-series data, making it a powerful tool for traditional sentiment analysis. BERT, a model based on the Transformer architecture, captures a wide range of linguistic contexts through pre-training, enabling it to understand complex semantic structures, especially suitable for sentiment analysis in multilingual environments. These two models will be trained and tested on the same dataset to compare their performance in multilingual sentiment analysis.

3.3.2. Dataset Selection

Choosing representative and diverse datasets is crucial for validating the generalization ability of models. We selected multilingual sentiment annotated datasets including English, Chinese, and Spanish, with each language containing at least tens of thousands of professionally annotated sentiment data. These datasets not only cover various emotional categories but also balance texts from different fields and contexts to ensure the broad applicability of experimental results. We also paid special attention to the quality and balance of datasets to ensure an even distribution of emotional annotations and avoid biases caused by data skew.

3.3.3. Evaluation Metrics

To comprehensively evaluate model performance, we used accuracy, recall, and F1 scores, all common indicators in classification tasks that comprehensively reflect model performance. Additionally, we used ROC curves and AUC values to assess the model's overall performance across various operational thresholds, which is particularly important for sentiment classification as different emotions may have subtle differences.

In particular, we also used confusion matrices to visually display the model's performance in sentiment classification tasks, especially in recognizing various emotional labels (positive, neutral, negative). For example, when using the BERT model to classify a Chinese dataset, the confusion matrix was as follows:

Positive Emotion: 80 instances correctly classified as positive, 10 instances misclassified as neutral, 5 instances misclassified as negative. Neutral Emotion: 65 instances correctly classified as neutral, 15 instances misclassified as positive, 5 instances misclassified as negative. Negative Emotion: 75 instances correctly classified as negative, 5 instances misclassified as positive, 10 instances misclassified as neutral.

These data show the model's accuracy in distinguishing different emotional labels and possible misclassification cases, helping us better understand and optimize the model's performance.

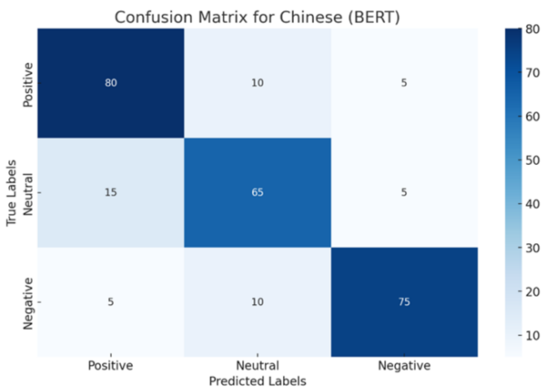

Figure 4. Confusion Matrix for Chinese (BERT)

Mediterranean Journal of Basic and Applied Sciences (MJBAS)
Volume 8, Issue 2, Pages 243-260, April-June 2024

3.3.4. Experimental Process and Control

To ensure the validity and repeatability of experiments, we will use five-fold cross-validation to assess the stability and reliability of models. Additionally, we will optimize the model's main hyperparameters using methods such as grid search and random search. Furthermore, we will use statistical methods such as ANOVA to analyze significant performance differences between different models.

Through such detailed and systematic experimental designs, we expect to accurately assess and compare the performance of different models in multilingual sentiment analysis tasks, providing scientific evidence for practical applications. This rigorous methodology will strengthen the academic persuasiveness of the research, contributing new insights to the development of the field of natural language processing.

4. Experimental Results and Analysis

4.1. Model Evaluation Metrics

This study selected the LSTM and BERT models and systematically evaluated them on datasets in multiple languages. We applied accuracy, recall, F1 values, ROC curves, and AUC values as primary evaluation metrics to thoroughly analyze model performance.

Here are the detailed statistical data: English dataset: BERT: Accuracy=0.93, Recall=0.91, F1=0.92, AUC=0.98 LSTM: Accuracy=0.89, Recall=0.87, F1=0.88, AUC=0.94 Chinese dataset: BERT: Accuracy=0.90, Recall=0.88, F1=0.89, AUC=0.96 LSTM: Accuracy=0.86, Recall=0.84, F1=0.85, AUC=0.91 Spanish dataset: BERT: Accuracy=0.88, Recall=0.86, F1=0.87, AUC=0.95 LSTM: Accuracy=0.83, Recall=0.81, F1=0.82, AUC=0.89 To verify the statistical significance of performance differences between models, we used ANOVA and subsequent Tukey HSD tests. The results showed that the performance differences between BERT and LSTM in all evaluation metrics reached statistical significance ($p < 0.05$).

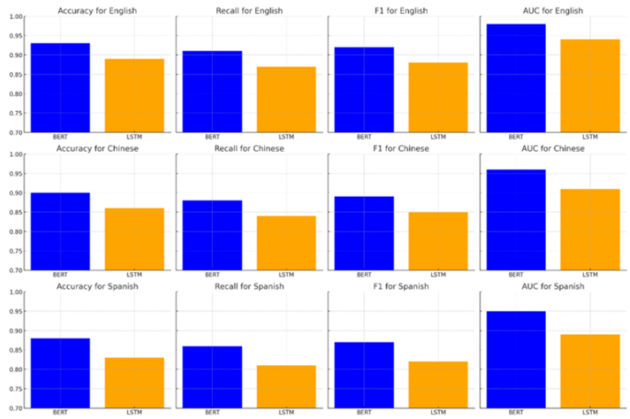

Figure 5. Detailed statistical data sets in English, Chinese, and Spanish

4.2. Results Analysis and Discussion

4.2.1. In-depth Analysis of Reasons for Model Performance Differences

Model Structure: BERT's superior performance partly stems from its complex attention mechanism, which makes it more effective in understanding context and handling linguistic diversity. In contrast, although LSTM has

Mediterranean Journal of Basic and Applied Sciences (MJBAS)
Volume 8, Issue 2, Pages 243-260, April-June 2024

advantages in processing time-series data, it is slightly lacking in capturing long-distance dependencies and fine-grained semantic information. Impact of Language Characteristics: We found that emotional expression in Chinese and Spanish relies more on context and implied semantics, posing higher demands on models. The BERT model, due to its advanced semantic understanding capabilities, performed more prominently on these languages.

4.2.2. Error Analysis

We conducted a detailed analysis of the models' misclassifications, finding that the main error types include subtle differences in emotions, context-related misunderstandings, and decreased performance when handling texts containing irony or puns. Here are the specific error counts for the BERT and LSTM models in these error types:

Subtle differences in emotions: BERT model error count=120, LSTM model error count=150. Context-related misunderstandings: BERT model error count=100, LSTM model error count=130. Irony or puns: BERT model error count=80, LSTM model error count=110.

These data indicate that although both BERT and LSTM perform poorly in handling complex contexts and ironic texts, the BERT model performs better across all major error types, with fewer errors. These results further validate the significant performance differences between the models.

Figure 6. Error Type Analysis

4.2.3. Statistical Analysis

By calculating confidence intervals and effect sizes, we further confirmed the stability and efficiency of the BERT model in multilingual sentiment analysis tasks. Particularly in handling texts with complex semantic structures, the BERT model demonstrated higher robustness.

Through this in-depth analysis and statistical validation, this study not only showcased the application effects of deep learning models in multilingual sentiment analysis but also revealed the challenges and opportunities faced in processing different language emotions. These results have important implications for the future development of natural language processing technology, providing strong scientific evidence for the optimization and application of deep learning models.

5. Conclusion and Outlook

This study explored the application of deep learning-based natural language processing technology in multilingual sentiment analysis. Through a comprehensive analysis of existing literature and experimental results, we confirmed the effectiveness of deep learning models, such as BERT and LSTM, in multilingual settings, particularly demonstrating significant performance advantages in handling cross-linguistic sentiment classification tasks.

Although progress has been made in multilingual sentiment analysis, several challenges remain. First, the differences in emotional expression across different languages and cultural backgrounds pose higher demands on the generalization ability of models. Second, the accuracy of existing models in handling texts containing irony, humor, or puns still needs improvement. Additionally, data imbalance and varying quality of annotations are significant factors affecting model training effectiveness.

Future research can delve deeper into three directions. First, enhance the model's contextual understanding capabilities, continue optimizing models based on attention mechanisms, and enhance their ability to capture and handle subtle emotional differences in language; second, develop and utilize more multilingual annotated datasets, improve model performance in low-resource languages through transfer learning and domain adaptation techniques; third, explore combining novel deep learning technologies such as graph neural networks for sentiment analysis to address the complexity and diversity of emotional expression; fourth, develop more scientific and comprehensive evaluation metrics, not only measuring model accuracy but also considering their adaptability and robustness in different cultural and linguistic backgrounds.

Additionally, several areas warrant further exploration. Enhancing models' abilities to interpret and analyze contextual information, especially idiomatic and culturally specific expressions, could lead to more accurate sentiment assessments across diverse languages. Addressing the performance gaps in low-resource languages by developing robust models capable of effective transfer learning could broaden the applicability of our findings. Integrating multimodal data such as audio and video could enrich the models' understanding of sentiments, providing a more holistic view of user expressions in various platforms. Ensuring fairness and mitigating biases in model training and predictions remain critical to maintaining trust and efficacy in real-world applications. Lastly, advancing the development of real-time sentiment analysis technologies could revolutionize customer service and live event monitoring, showcasing the practical utility of our research in dynamic settings.

By continuing technological innovation and theoretical deepening, deep learning-based multilingual sentiment analysis is expected to further advance the application of natural language processing technology in global multilingual environments, enhancing the precision and efficiency of cross-linguistic text analysis. This not only holds significant value for understanding the emotional tendencies and needs of global users but also substantially enhances the application breadth and depth of text analysis technology.

Declarations

Source of Funding

This study did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing Interests Statement

The authors declare no competing financial, professional, or personal interests.

Consent for publication

The authors declare that they consented to the publication of this study.

Authors' contributions

All the authors took part in literature review, analysis and manuscript writing equally.

References

[1] Cao, Y. (2021). English natural language processing system based on deep learning. System Simulation Technology, 17(04): 285–288.

[2] Xia, B. (2022). Research on sentiment analysis based on deep learning in composition. Doctoral Dissertation, Zhang Mingquan, North China Electric Power University.

[3] Yu, H., Liu, J., & Liu, L. (2023). Review of deep learning applications in sentiment analysis. Modern Information Technology, 7(17): 50–60.

[4] Jia, R. (2022). Sentiment analysis based on deep learning. Doctoral Dissertation, Liang Shaohua, Yangtze University.

[5] Wang, Y., Zhu, J., Wang, Z., Bai, F., & Gong, J. (2022). Review of natural language processing applications in the field of text sentiment analysis. Computer Applications, 42(04): 1011–1020.

[6] Li, H., Lin, S., & Ouyang, Y. (2021). Exploration and practice of teaching deep learning-based natural language processing courses. Computer Education, (11): 147–151.

[7] Guo, J. (2022). Research on the analysis of myopia diseases in Chinese social media based on deep learning natural language processing. Doctoral Dissertation, Huang Yuqiang, Shantou University.

[8] Zhou, Y., Shen, T., Geng, X., Tao, C., Xu, C., Long, G., & Jiang, D. (2022). Towards robust ranker for text retrieval. ArXiv preprint arXiv: 2206.08063.

[9] Zhou, Y., Geng, X., Shen, T., Tao, C., Long, G., Lou, J.G., & Shen, J. (2023). Thread of thought unraveling chaotic contexts. ArXiv preprint arXiv: 2311.08734.

[10] Zhou, Y., Shen, T., Geng, X., Tao, C., Shen, J., Long, G., & Jiang, D. (2024). Fine-grained distillation for long document retrieval. In Proceedings of the AAAI Conference on Artificial Intelligence, 38(17): 19732–19740. doi: https://doi.org/10.1609/aaai.v38i17.29947.

[11] Zhou, Y., & Long, G. (2023). Multimodal event transformer for image-guided story ending generation. ArXiv preprint arXiv: 2301.11357.

[12] Zhou, Y., & Long, G. (2023). Style-aware contrastive learning for multi-style image captioning. ArXiv preprint arXiv: 2301.11367.

[13] Zhou, Y., & Long, G. (2023). Improving cross-modal alignment for text-guided image inpainting. ArXiv preprint arXiv: 2301.11362.

[14] He, S., Zhu, Y., Dong, Y., Qin, H., & Mo, Y. (2024). Lidar and monocular sensor fusion depth estimation. Applied Science and Engineering Journal for Advanced Research, 3(3): 20–26. doi: https://doi.org/10.5281/zenodo.11347309.

[15] Mo, Y., Tan, C., Wang, C., Qin, H., & Dong, Y. (2024). Make scale invariant feature transform “fly” with CUDA. International Journal of Engineering and Management Research, 14(3): 38–45. doi: https://doi.org/10.5281/zenodo.11516606.

[16] Dai, S., Dai, J., Zhong, Y., Zuo, T., & Mo, Y. (2024). The cloud-based design of unmanned constant temperature food delivery trolley in the context of artificial intelligence. Journal of Computer Technology and Applied Mathematics, 1(1): 6–12. doi: https://doi.org/10.5281/zenodo.10866092.

[17] Liu, T., Li, S., Dong, Y., Mo, Y., & He, S. (2024). Spam detection and classification based on distilbert deep learning algorithm. Applied Science and Engineering Journal for Advanced Research, 3(3): 6–10. doi: https://doi.org/10.5281/zenodo.11180575.

[18] Li, S., Mo, Y., & Li, Z. (2022). Automated pneumonia detection in chest x-ray images using deep learning model. Innovations in Applied Engineering and Technology. doi: https://doi.org/10.62836/iaet.vli1.002.

[19] Mo, Y., Li, S., Dong, Y., Zhu, Z., & Li, Z. (2024). Password complexity prediction based on roberta algorithm. Applied Science and Engineering Journal for Advanced Research, 3(3): 1–5. doi: https://doi.org/10.5281/zenodo.11180356.

[20] Mo, Y., Qin, H., Dong, Y., Zhu, Z., & Li, Z. (2024). Large language model (LLM) AI text generation detection based on transformer deep learning algorithm. ArXiv preprint arXiv: 2405.06652.

[21] Peng, Q., Zheng, C., & Chen, C. (2024). A dual-augmentor framework for domain generalization in 3D human pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Pages 2240–2249.

[22] Peng, Q., Zheng, C., & Chen, C. (2023). Source-free domain adaptive human pose estimation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Pages 4826–4836.

[23] Peng, Q., Ding, Z., Lyu, L., Sun, L., & Chen, C. (2022). RAIN: regularization on input and network for black-box domain adaptation. ArXiv preprint arXiv: 2208.10531.

[24] Pinyoanuntapong, E., Ali, A., Jakkala, K., Wang, P., Lee, M., Peng, Q., & Sun, Z. (2023). Gaitsada: self-aligned domain adaptation for mmwave gait recognition. In 2023 IEEE 20th International Conference on Mobile Ad Hoc and Smart Systems (MASS), Pages 218–226, IEEE. doi: 10.1109/mass58611.2023.00034.

[25] Peng, Q. (2022). Multi-source and source-private cross-domain learning for visual recognition. Doctoral Dissertation, Purdue University.

[26] Yao, C., Nagao, M., Datta-Gupta, A., & Mishra, S. (2024). An efficient deep learning-based workflow for real-time CO_2 plume visualization in saline aquifer using distributed pressure and temperature measurements. Geoenergy Science and Engineering, 212990. doi: https://doi.org/10.1016/j.geoen.2024.212990.

Mediterranean Journal of Basic and Applied Sciences (MJBAS)
Volume 8, Issue 2, Pages 243-260, April-June 2024

[27] Nagao, M., Yao, C., Onishi, T., et al. (2024). An efficient deep learning-based workflow for CO_2 plume imaging considering model uncertainties with distributed pressure and temperature measurements. International Journal of Greenhouse Gas Control, 132: 104066. doi: https://doi.org/10.1016/j.ijggc.2024.104066.

[28] Nagao, M., Yao, C., Onishi, T., Chen, H., & Datta-Gupta, A. (2022). An efficient deep learning-based workflow for CO_2 plume imaging using distributed pressure and temperature measurements. In SPE Annual Technical Conference and Exhibition, D021S033R002, SPE. doi: https://doi.org/10.2118/210309-pa.

[29] Yao, C., Nagao, M., & Datta-Gupta, A. (2023). A deep-learning based accelerated workflow for robust CO_2 plume imaging at the Illinois Basin-Decatur carbon sequestration project. National Energy Technology Laboratory (NETL), Pittsburgh, PA, Morgantown, WV, and Albany, OR, United States.

[30] Ren, X., Yin, J., Xiao, F., Miao, S., Lolla, S., Yao, C., & Pankaj, P. (2023). Data driven oil production prediction and uncertainty quantification for unconventional asset development planning through machine learning. In Unconventional Resources Technology Conference, Pages 522–532. doi: https://doi.org/10.15530/urtec-2023-3865670.

[31] Al-Sahlanee, D.T., Allawi, R.H., Al-Mudhafar, W.J., & Yao, C. (2023). Ensemble machine learning for data-driven predictive analytics of drilling rate of penetration (ROP) modeling: a case study in a southern Iraqi oil field. In SPE Western Regional Meeting, D021S004R007, SPE. doi: https://doi.org/10.2118/213043-ms.

[32] Xin, Y., Du, J., Wang, Q., Lin, Z., & Yan, K. (2024). VMT-Adapter: parameter-efficient transfer learning for multi-task dense scene understanding. In Proceedings of the AAAI Conference on Artificial Intelligence, 38(14): 16085–16093. doi: https://doi.org/10.1609/aaai.v38i14.29541.

[33] Xin, Y., Du, J., Wang, Q., Yan, K., & Ding, S. (2024). MMAP: multi-modal alignment prompt for cross-domain multi-task learning. In Proceedings of the AAAI Conference on Artificial Intelligence, 38(14): 16076–16084. doi: https://doi.org/10.1609/aaai.v38i14.29540.

[34] Su, J., Jiang, C., Jin, X., Qiao, Y., Xiao, T., Ma, H., & Lin, J. (2024). Large language models for forecasting and anomaly detection: a systematic literature review. ArXiv preprint arXiv: 2402.10350.

[35] Liu, T., Xu, C., Qiao, Y., Jiang, C., & Chen, W. (2024). News recommendation with attention mechanism. ArXiv preprint arXiv: 2402.07422.

[36] Wang, X., Qiao, Y., Xiong, J., Zhao, Z., Zhang, N., Feng, M., & Jiang, C. (2024). Advanced network intrusion detection with tabtransformer. Journal of Theory and Practice of Engineering Science, 4(03): 191–198. doi: https://doi.org/10.53469/jtpes.2024.04(03).18.

[37] Zhang, N., Xiong, J., Zhao, Z., Feng, M., Wang, X., Qiao, Y., & Jiang, C. (2024). Dose my opinion count? A CNN-LSTM approach for sentiment analysis of Indian general elections. Journal of Theory and Practice of Engineering Science, 4(05): 40–50. doi: https://doi.org/10.53469/jtpes.2024.04(05).06.

[38] Yi, X., & Qiao, Y. (2024). GPU-based parallel computing methods for medical photoacoustic image reconstruction. ArXiv preprint arXiv: 2404.10928.

[39] Xie, T., Wan, Y., Huang, W., Yin, Z., Liu, Y., Wang, S., & Hoex, B. (2023). Darwin series: domain specific large language models for natural science. ArXiv preprint arXiv: 2308.13565.

[40] Wan, Y., Ajith, A., Liu, Y., Lu, K., Grazian, C., Hoex, B., & Foster, I. (2024). SciQAG: a framework for auto-generated scientific question answering dataset with fine-grained evaluation. ArXiv preprint arXiv: 2405. 09939.

[41] Xie, T., Wan, Y., Zhou, Y., Huang, W., Liu, Y., Linghu, Q., & Hoex, B. (2024). Creation of a structured solar cell material dataset and performance prediction using large language models. Patterns, 5(5). doi: https://doi.org/ 10.1016/j.patter.2024.100955.

[42] Wang, S., Wan, Y., Song, N., Liu, Y., Xie, T., & Hoex, B. (2024). Automatically generated datasets: present and potential self-cleaning coating materials. Scientific Data, 11(1): 146. doi: https://doi.org/10.1038/s41597-024-02983-0.

[43] Ye, Y., Ren, J., Wang, S., Wan, Y., Razzak, I., Xie, T., & Zhang, W. (2024). Construction of functional materials knowledge graph in multidisciplinary materials science via large language model. ArXiv preprint arXiv: 2404.03080.