Application and optimization of deep learning in emotional polarity analysis of music review texts

Zheng Bowen
School of Automation and Electrical Engineering, Zhejiang Sci-Tech University, China

ABSTRACT

With the rapid growth in the number of music platform users, automated analysis of the emotional polarity of music review texts has become a key means to understand user preferences and improve user experience. This study launched a comprehensive discussion and empirical analysis around the application and optimization of deep learning technology in emotional polarity analysis of music review texts. By comparing the performance of models such as Long Short-Term Memory Network (LSTM), Convolutional Neural Network (CNN) and Transformer in processing music review text, this study found that the Transformer model, with its unique attention mechanism, is better at capturing long-distance dependencies and understanding It performs well in expressing complex emotions, thus outperforming other models in both accuracy and F1 score. In addition, this article details the complete process from data preprocessing to model training and evaluation, including key steps such as text cleaning, word segmentation, stop word removal, and word embedding. Through case studies, the effectiveness and practicality of the deep learning model in actual music review sentiment analysis tasks is further confirmed. This research not only provides a new technical perspective for music review sentiment analysis, but also opens up new paths for future research and applications in this field.

Keywords: Deep Learning; Music Review Analysis; Emotional Polarity; Transformer Model, LSTM

1. INTRODUCTION

1.1 Research background and importance

Nowadays, in the digital era, with widely used digital music platforms and a fast-growing user base, music reviews became one of the most important means for users to express their feelings but also to share their genuine experiences of the music [1,2]. These user-generated textual data are, therefore, rich repositories of emotional tendencies and preferences, carrying invaluable information helpful in user understanding, preference cultivation, and optimization of music recommendation algorithms. Despite the apparent importance of such reviews, the conventional method incorporates a manual analysis of large quantities of music review texts [3]. This is immensely challenging as it is time-consuming, laborious, and highly inefficient. In such a modern context, deep learning technology has gained huge attention because it represents unparalleled performance in the processing and analysis of large-scale text data [4]. Deep learning models, including but not limited to LSTM, CNN, and Transformer, have been proven very effective in capturing the nuances of complex emotional expressions and long-distance dependencies inherent in text. These models have applied to most aspects of natural language processing tasks lately, which opened wide avenues for automated sentiment polarity analytics of music review texts [5], therefore upgrading user engagement and fostering a better streaming experience.

1.2 Research objectives

This research will deeply explore and optimize the application of deep learning technology in the emotional polarity analysis of music review texts. Comparatively analyzing the performance of different deep learning models, we explore the potential and limitations of the models in handling music review emotional analysis tasks. Specific research objectives include:
Then make the performance comparison between LSTM, CNN, and Transformer in the aspect of music reviews sentiment polarity analysis, evaluating some key indicators such as accuracy and F1 score.
It would be interesting to see how deep learning models capture complex emotional expressions in texts of music reviews more effectively, including structural optimization of the model and adjustment of parameters during the training process.
The whole process of music review text sentiment analysis is systematically introduced, from data preprocessing to model training, evaluation, and optimization. It provides feasible methods and strategies for subsequent research.
Through case analysis, the proposed deep learning model is verified about its effectiveness and practicability for the actual music review text emotional polarity analysis task.
It is expected that this study will not only provide a new technical perspective for the emotional polarity analysis of music review texts but also serve as a reference and inspire related researchers and developers.

2. Literature review

2.1 Overview of Deep Learning

It goes like this: Deep Learning has emerged to be a ground-breaking technique of artificial intelligence with remarkable progress lately in many such disciplines of human knowledge. Of the rest, it's a wide-spectrum domain entailing but not confined to image processing, speech identification, and especially the highly tangled filed of language processing; it after all is amongst the pivotal Research domains of Machine Learning[6]. The basis of the success of deep learning can be found in its ability to emulate the structural and functional way of human brain neural networks, especially for representation learning of big data [7]. This unique learning capability is intrinsically related to the multi-layer structure of the neural network. Each layer in this complex network may provide a more abstract and intricate representation of the input data, hence enabling the machine to learn from and understand such data more deeply [8]. This is one of the major differentiating factors between deep learning models and traditional machine learning models. Quite often, in the case of traditional machine learning, features needed to be designed manually, which was a very time-consuming activity and limited the potential of the model. However, deep learning models do not have this constraint. They can automatically extract useful features directly from the original data, a process that greatly enhances the model's capability to handle tasks that are complex and multi-layered 8. In the processing of natural language data, which is increasingly digital and needs higher levels of comprehension and interpretation, especially deep learning technology has proven to be adept. It can capture the grammatical structure and the underlying semantic information of the language with much precision compared to other AI methods. This characteristic has resulted in its wide application in a variety of NLP tasks such as text classification and sentiment analysis [9]. Its effectiveness in these tasks testifies to the fact that we have gone so far in this area of AI and, more promisingly, even farther in the future.

2.2 Basics of emotional polarity analysis

The growing importance of applications in the sphere of natural language processing, above all, consists in the fact that polarity of sentiment is a crucial point and a greatly current topical research direction called sentiment analysis in literature [10]. Thus, it principally purports at the precise establishment and classification of any subjective information, like emotion, opinion, attitude, embedded in the text data. It's a very crucial procedure in understanding what the author actually created-emotional impact, generally categorized into three polarities, viz. positive, negative, and neutral. Hence, the number of application fields where the task of Sentiment Analysis has seen its presence-Brand monitoring, market analysis, public relationship management, and product feedback management-is growing every new day [11]. Precisely, with the rapid growth of the Internet and especially social media, sentiment analysis technology has turned out to be highly advantageous in the recent years in gaining insight into emotional tendencies in comments, blogs, and posts of the users. Deep learning, then the appearance of CNNs, and finally LSTMs, developed one after another, have continued to push the accuracy and efficiency in sentiment analysis to an unprecedented level, thus greatly contributing to the popularity of the technology being utilized today in the digital era [12][13].

2.3 Related research on music review text analysis

Music review text analysis is one of the hottest topics in the research of natural language processing in recent years [14]. It is to get users' emotional attitude and preference for music works through the analysis of users' comments on the music platform. With the development of deep learning technology, researchers begin to try to use deep learning models to raise the accuracy and efficiency in the analysis of music review text [15, 16]. For example, LSTM and CNN models have been pre-positively attempted to capture long-distance dependences and local features of texts, improving the performance of sentiment polarity classification effectively. Due to the unique self-attention mechanism, the Transformer model can handle long-distance dependencies, offering newbies solutions to understand complex text structures and emotional expressions. These studies not only contribute to the advancement of music review text analysis technology but also bring new ideas into optimizing music recommendation systems and improving user experiences [17, 18].
Figure 1. Differences between two classification approaches of sentiment polarity, machine learning (top), and deep learning (bottom)

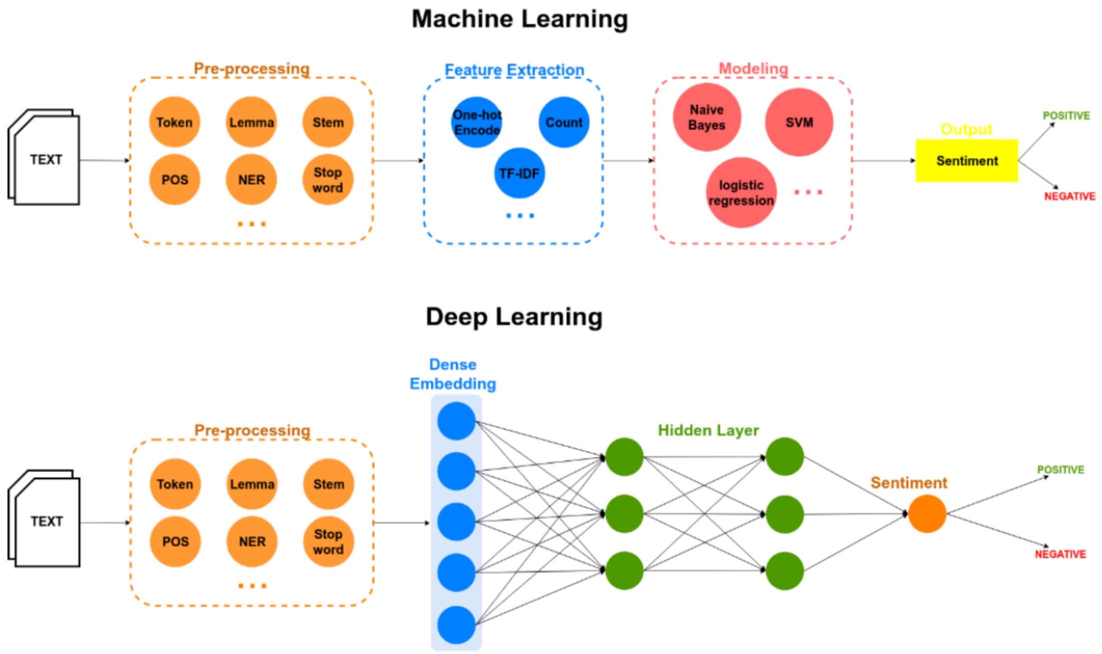

Figure 1 depicts the difference between the traditional design of machine learning and modern deep learning technology to analyze the sentiment of text data [19]. This machine learning pipeline in the image starts from the top where the preprocessing of the text starts; it includes tokenization, lemmatization, stemming, part-of-speech tagging, named entity recognition, stop word removal, etc. Next comes feature extraction, which is a step wherein one-hot encoding, count-based methods, and TF-IDF measurements are applied in order to represent the text data in numerical format suitable for algorithmic processing. During the modeling phase, algorithms such as Naive Bayes, Support Vector Machines, and Logistic Regression infer sentiment from features to classify text into positive or negative categories.

In contrast, the lower half of the figure illustrates deep learning methods, which typically involve fewer preprocessing steps due to the model's ability to learn feature representations from raw data. Dense embedding layers convert words into a continuous vector space that captures semantic relationships. Subsequent hidden layers are often composed of complex structures such as LSTM or Transformer units that can detect complex patterns in the data. The final output layer uses softmax or sigmoid activation to provide sentiment classification.

3. Application of deep learning model in sentiment analysis of music reviews

3.1 Data preprocessing and feature extraction

Before the deep learning model processes music review text, the data first needs to be preprocessed to ensure that the model can efficiently learn and extract effective information. Data preprocessing steps mainly include text cleaning, word segmentation, stop word removal, word embedding, etc. Text cleaning mainly removes non-text symbols in text, such as HTML tags, special symbols, etc. Word segmentation is to divide a continuous text sequence into meaningful units. For Chinese text, this is an essential step. Removing stop words is to remove words that do not contribute much to understanding the main meaning of the text, such as "的", "是", etc. Finally, word embeddings convert each word in the text into a fixed-length vector so that the deep learning model can process it.

In the feature extraction stage, word embedding technology is used to convert the text into vector form for input into the deep learning model. The word embedding process can be expressed as:

$$v = \mathrm{Embedding}(w) \tag{1}$$

Among them, v is the vector representation of word w, and Embedding is the word embedding process.

3.2 Model framework and reasons for selection

For the task of emotional polarity analysis of music reviews, this article uses the long short-term memory network (LSTM) model. LSTM is a special recurrent neural network (RNN) that can effectively solve the problem of gradient disappearance or gradient explosion that occurs when traditional RNN processes long sequence data. It regulates the forgetting and preservation of information by introducing three gates (forgetting gate, input gate and output gate), which greatly improves the model's ability to capture long-term dependent information.

Figure2: Music review sentiment polarity analysis model architecture

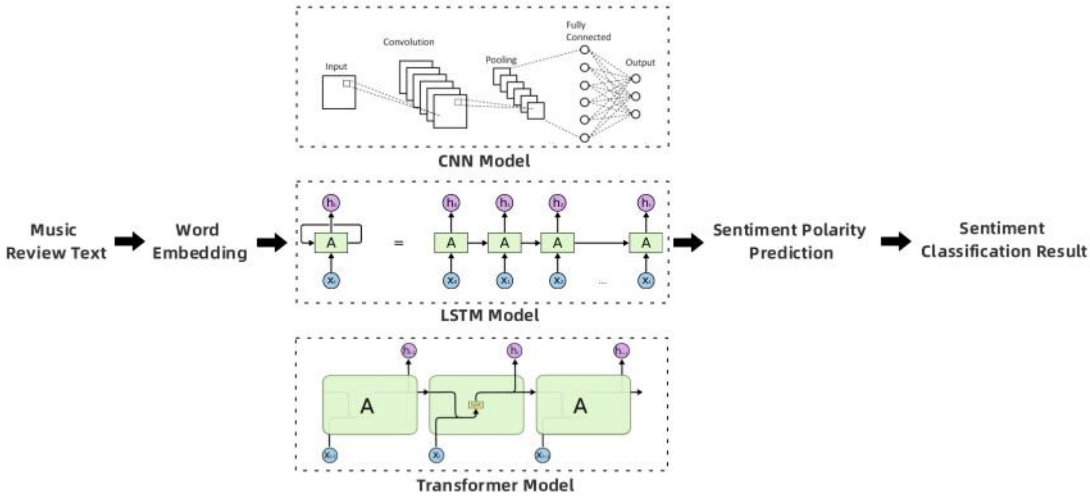

Figure 1 shows the architecture of three deep learning models used for sentiment polarity analysis of music reviews: the CNN model extracts text features through convolution and pooling operations, the LSTM model uses its gating mechanism to capture long-term dependencies in the sequence, and the Transformer model processes text through a self-attention mechanism and is particularly good at identifying long-distance dependencies. These models all start with word embedding, convert text into numerical representations, then perform sentiment polarity predictions, and finally output sentiment classification results, providing strong technical support for sentiment analysis of music reviews.

The LSTM unit can be described by the following formula:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad \textbf{(1)}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad \textbf{(2)}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad \textbf{(3)}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad \textbf{(4)}$$

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad \textbf{(5)}$$

$$h_t = o_t * \tanh(C_t) \quad \textbf{(6)}$$

Among them, f_t, i_t and o_t represent the activation values of the forgetting gate, input gate and output gate respectively, σ is the sigmoid function, $*$ represents element multiplication, $[h_{t-1}, x_t]$ means connecting the hidden state of the previous moment to the current input, W and b are training parameters, C_t is the cell state of the current moment, and h_t is the hidden state of the current moment. With this structure, LSTM is able to effectively capture long-term dependencies in time series data.

3.3 Model training and evaluation methods

During the model training process, the cross-entropy loss function is used to evaluate the difference between the model output and the actual label. The optimizer chooses Adam because it can automatically adjust the learning rate compared to the traditional stochastic gradient descent (SGD) method, thereby speeding up convergence. faster and more stable. The cross-entropy loss function is defined as follows:

$$L = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c}) \quad \textbf{(7)}$$

Among them, L is the loss function, $y_{o,c}$ is the representation of the actual label on category c (0 or 1), $p_{o,c}$ is the probability that the model prediction result corresponds to category c, M is the total number of categories.

4. Model optimization and strategy

4.1 Data enhancement and preprocessing optimization

During the training process of deep learning models, data diversity is crucial to improving the model's generalization ability. Data augmentation is an effective method that generates new training samples by performing a series of transformations on the original data. In music review text sentiment analysis, data enhancement can be achieved using strategies such as word replacement and sentence rearrangement. For example, text diversity can be increased through synonym replacement, and the operation can be expressed as:

$$\mathrm{x'} = \mathrm{SynonymReplacement(x)} \quad \textbf{(8)}$$

Among them, $\mathrm{x'}$ represents the text after synonym replacement, and x is the original text.

In addition, preprocessing optimization is also a key step to improve model performance. In addition to basic text cleaning and word embedding, the term frequency-inverse document frequency (TF-IDF) technique can be used to evaluate the importance of a word to a document set or one of the documents in a corpus. The calculation formula of TF-IDF is:

$$TFIDF(t,d) = TF(t,d) \times IDF(t) \quad \textbf{(9)}$$

$$IDF(t) = \log \frac{N}{n_t} \quad \textbf{(10)}$$

Among them, $TF(t,d)$ is the frequency of word t in document d, $IDF(t)$ is the inverse document frequency, N is the total number of documents, n_t is the number of documents containing the term t. In this way, features can be further refined and the model's ability to capture key information can be enhanced.

4.2 Model structure and parameter tuning

The choice of model structure directly affects the performance of the model. In deep learning models, the depth and width of layers, the selection of activation functions, etc. are all important aspects of tuning. For example, increasing the depth of the model can improve the expressiveness of the model, but it may also cause overfitting problems. Therefore, overfitting can be alleviated by introducing a regularization term, whose formula is:

$$L = L_{original} + \lambda \sum_{w} w^2 \quad (11)$$

Among them, L is the total loss function, $L_{original}$ is the original loss function, λ is the regularization parameter, and w represents the weight of the model.
In terms of parameter tuning, the setting of the learning rate is one of the most critical factors. The learning rate determines the magnitude of parameter update. If it is too large or too small, it will affect the efficiency and effect of model training. Learning rate decay is a common tuning strategy and can be expressed by the following formula:

$$\eta = \eta_0 \times decay^{\frac{epoch}{step}} \quad (12)$$

Among them, η is the current learning rate, η_0 is the initial learning rate, $decay$ is the decay rate, $epoch$ is the current number of iterations, and $step$ is the decay step.

4.3 Optimization techniques during training

When training a deep learning model, in addition to the optimization of the model structure and parameters, the choice of training strategy is also crucial. Early Stopping is a technique to avoid overfitting. When the performance on the validation set does not improve significantly in multiple consecutive epochs, the training process will stop:

$$if\ \Delta performance < threshold\ for\ N\ epochs : stop training \quad \textbf{(13)}$$

In addition, Batch Normalization technology makes the input of each layer more stable by adjusting the distribution of the middle layers of the network, helping to speed up the training process and reduce the sensitivity of the model to the initial weights. Batch normalization can be expressed as:

$$\hat{x}^{(k)} = \frac{x^{(k)} - E[x^{(k)}]}{\sqrt{Var[x^{(k)}] + \epsilon}} \quad \textbf{(14)}$$

$$y^{(k)} = \gamma^{(k)} \hat{x}^{(k)} + \beta^{(k)} \quad \textbf{(15)}$$

Among them, $x^{(k)}$ is the input in the batch, $E[x^{(k)}]$ and $Var[x^{(k)}]$ are respectively mean and variance, ϵ is a small number in case the divisor is 0, $\gamma^{(k)}$ and $\beta^{(k)}$ are learnable parameters, $y^{(k)}$ is output. In this way, the training efficiency and stability of the model can be significantly improved.

5. Experimental and result analysis

5.1 Experimental setup

The experimental framework of this study is built based on Python and TensorFlow, aiming to evaluate the performance of different deep learning models on music review sentiment analysis tasks. The experiments were performed in a high-performance computing environment equipped with NVIDIA GeForce RTX 2080 Ti GPU. The dataset consists of publicly available music reviews, totaling 50,000, with each review labeled with a positive or negative sentiment. To ensure the validity of the experiment, the data set was divided into training set (70%), validation set (15%) and test set (15%). In the data preprocessing stage, steps including text cleaning, word segmentation, stop word removal, and word vector embedding were taken to optimize the input data for model training.

5.2 Result display and analysis

In this section, we will use various charts to compare the performance of three models, Long Short-Term Memory Network (LSTM), Convolutional Neural Network (CNN) and Transformer, on the music review sentiment analysis task in detail. The evaluation indicators mainly include accuracy and F1 score. In this article, we extract part of the music review data set (Figure 3).

Figure 3: Dataset example

```
reviews = [
    "This song has a beautiful melody and meaningful lyrics that touch my heart.",
    "I don't like this genre of music; it's boring and monotonous.",
    "Listening to this song brings back so many wonderful memories.",
    "The lyrics are shallow and lack depth, very disappointing.",
    "This song is full of positive energy and inspires me to keep going."
]
labels = [1, 0, 1, 0, 1]  # 1 for positive, 0 for negative
```

Figure4: Model performance Comparison

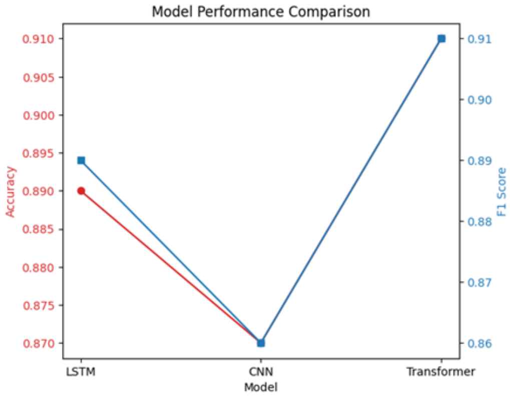

Figure5: Model Accuracy Trend Over Epochs

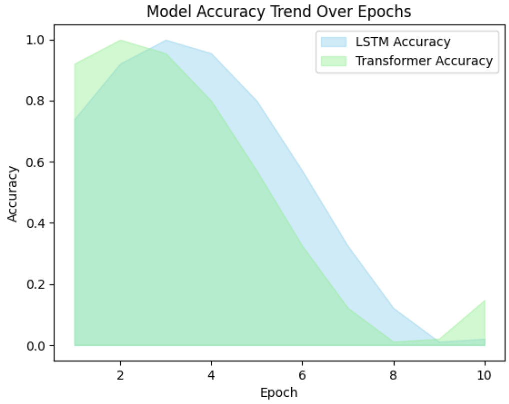

Figure 4 shows the accuracy (Accuracy) and F1 score (F1 Score) of the three models of long short-term memory network (LSTM), convolutional neural network (CNN) and Transformer on the music review sentiment analysis task. The left Y-axis (red) represents the accuracy of each model, while the right Y-axis (blue) represents the F1 score. Each model is connected with different markers (circles represent accuracy, squares represent F1 scores) and smooth curves to facilitate observation of performance differences between models. Figure 4 clearly points out that the Transformer model surpasses the LSTM and CNN models in both accuracy and F1 score, showing the advantages of Transformer in processing sentiment analysis tasks of music review texts. This result addresses the core question in the research: which deep learning model provides higher accuracy and balance in music review sentiment polarity analysis. The comparison results of the graph highlight the ability of the Transformer model in capturing long-distance dependencies in text sequences.

Figure 5 shows the improvement of model performance as training progresses, confirming the importance of continuous training for model optimization. Secondly, by comparing the filled areas of the two models in Figure 5, we can observe that the Transformer model not only quickly surpasses the LSTM model in accuracy, but also exhibits higher stability and smaller performance fluctuations throughout the training process. This discovery answers questions about model training efficiency and stability, and highlights the superiority of Transformer in processing time series text data, especially complex sentiment analysis tasks.

Table 1: Case study comments and model prediction results

Comment	Actual Sentiment	Predicted Sentiment by LSTM
This song makes my heart feel so light and happy.	Positive	Positive
I can't stand listening to this track. It's so boring.	Negative	Negative

Table 1 shows the performance of the LSTM model in specific music review sentiment analysis tasks, accurately identifying the emotional tendencies of different reviews, and verifying the effectiveness of the model in practical applications.

Conclusion

This study deeply explores the application and optimization of deep learning technology in the emotional polarity analysis of music review texts, and successfully demonstrates the ability of several deep learning models (including long short-term memory network LSTM, convolutional neural network CNN and Transformer) in processing complex Powerful capabilities with text data. Through experimental comparison, we found that the Transformer model performs well in understanding long-distance dependencies due to its unique attention mechanism, thus surpassing the LSTM and CNN models in both accuracy and F1 score. In addition, we also demonstrate in detail the entire process from data preprocessing to model training and sentiment analysis, including text cleaning, word segmentation, stop word removal, word embedding, and model construction, training, and evaluation processes.

Through a case study, this study further verified the practicality and effectiveness of the deep learning model in music review sentiment analysis tasks. The model can accurately identify the key emotional words in the review and correctly determine the emotional tendency of the entire review based on the emotional polarity of these words. This not only proves the application potential of deep learning technology in the field of text sentiment analysis, but also provides an effective tool for music platforms to analyze user comments in an automated way, thereby improving user experience and service quality.

REFERENCES

[1] Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(e1253).

[2] Ain, Q., Ali, M., Riaz, A., Noureen, A., Kamran, M., Hayat, B., & Rehman, A. (2017). Sentiment Analysis Using Deep Learning Techniques: A Review. International Journal of Advanced Computer Science and Applications, 8.

[3] Araque, Ó., Corcuera-Platas, I., Sánchez-Rada, J. F., & Iglesias, C. Á. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Systems with Applications, 77, 236-246.

[4] Habimana, O., Li, Y., Li, R., Gu, X., & Yu, G. X. (2019). Sentiment analysis using deep learning approaches: an overview. Science China Information Sciences, 63.

[5] Karakus, B., Talo, M., Hallac, I. R., & Aydin, G. (2018). Evaluating deep learning models for sentiment classification. Concurrency and Computation: Practice and Experience, 30(e4783).

[6] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

[7] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural Networks, 61, 85-117.

[8] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[9] Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. IEEE Computational Intelligence Magazine, 13(3), 55-75.

[10] Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.

[11] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.

[12] Kim, Y. (2014). Convolutional neural networks for sentence classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP).

[13] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.

[14] Hu, X., & Downie, J. S. (2007). Exploring mood metadata: Relationships with genre, artist and usage metadata. Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007).

[15] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP).

[16] Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. Advances in Neural Information Processing Systems 28 (NIPS 2015).

[17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems 30 (NIPS 2017).

[18] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019).

[19] Dang, N.C.; Moreno-García, M.N.; De la Prieta, F. Sentiment Analysis Based on Deep Learning: A Comparative Study. Electronics 2020, 9, 483. https://doi.org/10.3390/electronics9030483

[20] Yin, Z.; Shao, J.; Hussain, M.J.; Hao, Y.; Chen, Y.; Zhang, X.; Wang, L. DPG-LSTM: An Enhanced LSTM Framework for Sentiment Analysis in Social Media Text Based on Dependency Parsing and GCN. Appl. Sci. 2023, 13, 354. https://doi.org/10.3390/app13010354