ISSN: 2181-3906 2025

International scientific journal
«MODERN SCIENCE AND RESEARCH»
VOLUME 4 / ISSUE 6 / UIF:8.2 / MODERNSCIENCE.UZ

SELECTING AND JUSTIFYING DEEP LEARNING MODELS FOR EMOTION CLASSIFICATION

Nuriddinova Nasibaxon Umidjon qizi
Master's student of Applied Mathematics and Informatics, FSU.
https://doi.org/10.5281/zenodo.15650572

Abstract. *This paper focuses on the selection and justification of deep learning models for emotion classification tasks. It provides a comprehensive analysis of various neural network architectures, including Convolutional Neural Networks, Recurrent Neural Networks, Long Short-Term Memory networks, and Transformer models, assessing their performance in recognizing and classifying human emotions from multimodal data sources. The study examines the strengths and limitations of each model with respect to data type, training efficiency, computational complexity, and generalization capabilities. Furthermore, criteria for optimal model selection tailored to real-world emotion recognition applications are discussed. The findings contribute to enhancing the accuracy and robustness of emotion classification systems and offer valuable guidelines for researchers and practitioners developing advanced affective computing solutions.*

Keywords: *Deep Learning, Emotion Classification, Convolutional Neural Networks, Recurrent Neural Networks, Long Short-Term Memory, Transformer Models, Model Selection, Neural Network Performance.*

ВЫБОР И ОБОСНОВАНИЕ МОДЕЛЕЙ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ КЛАССИФИКАЦИИ ЭМОЦИЙ

Аннотация. *В данной работе рассматривается выбор и обоснование моделей глубокого обучения для задачи классификации эмоций. Представлен всесторонний анализ различных архитектур нейронных сетей, включая сверточные нейронные сети, рекуррентные нейронные сети, сети с долговременной кратковременной памятью и модели Transformer, с оценкой их эффективности в распознавании и классификации человеческих эмоций на основе многомодальных данных. Исследование анализирует сильные и слабые стороны каждой модели с учетом типа данных, эффективности обучения, вычислительной сложности и способности к обобщению. Кроме того, обсуждаются критерии оптимального выбора моделей, адаптированных к практическим приложениям в области распознавания эмоций. Полученные результаты способствуют повышению точности и надежности систем классификации эмоций и предоставляют ценные рекомендации для исследователей и разработчиков в области аффективных вычислений.*

Ключевые слова: *Глубокое Обучение, Классификация Эмоций, Сверточные Нейронные Сети, Рекуррентные Нейронные Сети, Долговременная Кратковременная Память, Модели Transformer, Выбор Модели, Производительность Нейронных Сетей.*

Introduction

In recent years, deep learning techniques have revolutionized the field of artificial intelligence, becoming pivotal tools for the automatic recognition and classification of human emotions.

Emotion classification plays a critical role in enhancing human-computer interaction, psychological assessment, healthcare applications, and marketing strategies. However, selecting an appropriate deep learning model for emotion classification presents significant challenges due to the complexity and variability of emotional data, diverse neural network architectures, computational requirements, and the efficiency of training processes. This paper aims to provide a comprehensive analysis of various deep neural network architecturesincluding Convolutional Neural Networks, Recurrent Neural Networks, Long Short-Term Memory networks, and Transformer models focusing on their effectiveness and suitability for emotion classification tasks.

Additionally, the study discusses essential criteria and considerations for justifying the selection of specific models in real-world applications. The insights gained from this research are expected to guide future developments in affective computing, improving the accuracy and robustness of emotion recognition systems and advancing human-machine interfaces.

Main part

In this study, we will consider in detail the methodological foundations of selecting and applying deep learning models for effective classification of emotions in the communication process. At the current stage of development of modern artificial intelligence, deep learning technologies are recognized as one of the most promising approaches to identifying emotions in communication. The issue of identifying and classifying emotions is complex, and in this process it is necessary to take into account various modalities, such as text, voice, facial expressions.

Therefore, an important methodological issue is to evaluate all existing models in this area and select the most optimal options from among them. When choosing deep learning models for the task of classifying emotions, the main attention should be paid to the following criteria: the level of accuracy of the model, the consumption of computational resources, the ability to work in real time, the amount of data required for training, and the degree of adaptability of the model to various communication contexts. This set of parameters allows us to choose the most optimal solution for our study. It is worth noting that each criterion has its own importance, and their relative importance depends on the specific task and application situation.

The accuracy of the model is undoubtedly one of the main indicators. High accuracy in emotion classification provides a good result, but this indicator alone is not enough. When assessing the accuracy, it is important to consider not only the percentage of overall correct predictions, but also individual indicators for each emotion class. For example, some emotions (for example, neutral or happiness) are often well identified, while others (for example, irony or mixed emotions) are often incorrectly classified. Therefore, it is advisable to evaluate each emotion class using criteria such as F1-score, precision, recall.

The consumption of computational resources is an important factor determining the real-world applicability of the model. Models with large transformer architectures can provide high accuracy, but they require a large amount of memory and computing power. This makes them difficult to use on mobile devices or in resource-constrained environments. Therefore, when choosing a model, it is necessary to take into account parameters such as its size, the time required for training and prediction, and energy consumption. The ability to work in real time is especially important in interactive systems.

To quickly identify emotions in the communication process, the model must work with minimal latency. Otherwise, the system may be inconvenient for the user.

The ability to work in real time is closely related to the architecture of the model, the number of parameters, and how it is optimized.

The amount of data required for training is also an important criterion. Some complex models require very large data sets to show high results. This can be a serious obstacle, especially for resource-constrained languages such as Uzbek. Therefore, it is preferable to choose models that can perform well even with a small amount of data or use transfer learning methods. The degree to which a model can adapt to different communication contexts is one of the factors that determine its practical value. Communication can occur in different domains (e.g., social media, customer support, educational platforms) and in different formats (formal, informal, professional).

Therefore, it is important that the model chosen can adapt to different contexts and accurately classify emotions.

There are several main deep learning architectures used for emotion classification today.

Among them, recurrent neural networks (RNNs), in particular, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models, show high results in detecting temporal relationships in text data. The advantage of these models is that they are able to take into account the contextual features of the dialogue and preserve the meaning of sentences. The advantage of LSTM models is their ability to sort out important information through mechanisms for storing and forgetting long-term relationships. This is especially important in cases where emotional expressions are often related to information said at the beginning of the dialogue.

GRU models, on the other hand, are simpler than LSTM and can show good results even with a small number of parameters. Their advantage is that they are trained faster and require less memory. In our study, it may be appropriate to use GRU-based models, especially in cases where real-time operation is important. Convolutional neural networks (CNN) are also showing high efficiency in processing text data. They are especially useful in detecting local features in the communication process and analyzing relationships between words. The advantage of CNN models is the ability to perform parallel computing, which allows them to run faster than LSTM and GRU models. CNN models also allow for the effective detection of important word combinations and phrases related to emotions in texts.

Some studies also use hybrid architectures based on the combination of CNN and RNN models. In this approach, the CNN part is used to detect local features, and the RNN part is used to take into account long-term relationships. Our study also aims to test such hybrid architectures and evaluate their effectiveness. In recent years, transformer architectures, in particular BERT (Bidirectional Encoder Representations from Transformers) and its various modifications, have brought revolutionary changes in language and speech processing. The advantage of these models is the ability to effectively take into account contextual dependencies and deeply understand the semantic meaning of the text. BERT and its variants provide a high level of accuracy in emotion classification, but they require a large amount of computational resources, which complicates their application in real-time systems.

The BERT model is famous for its pre-training/fine-tuning paradigm, which allows it to be adapted for special tasks such as emotion classification. Our research aims to study effective methods for adapting the BERT model to the task of classifying emotions in the Uzbek language.

At the same time, it is also important to analyze lightweight and optimized variants of BERT, such as RoBERTa, DistilBERT, ALBERT, since they allow achieving high results even with less computational resources. Other models belonging to the transformer family, such as GPT (Generative Pre-trained Transformer) and its variants, can also be used for the task of classifying emotions. A distinctive feature of GPT models is their generative nature, which allows them not only to classify emotions, but also to generate emotional content. This can be useful in creating artificial datasets for testing emotion classification systems.

Another important aspect of our research is the study of multimodal approaches.

Emotions are expressed not only through text, but also through voice timbre, speech rate, pauses, and other paralinguistic features. Therefore, in our study, it is also appropriate to consider multimodal deep learning architectures that combine text and voice data. Such an approach can significantly improve the accuracy of emotion classification. Among multimodal approaches, architectures that combine text and voice deserve special attention. Such models use, for example, models such as BERT or RoBERTa for text data, and wav2vec2 or HuBERT for voice data. Then, the features obtained for individual modalities are combined and a general prediction is made for emotion classification. Our research aims to test different variants of such models and evaluate their effectiveness.

When selecting models, their ability to adapt to dialogues in Uzbek and other Turkic languages is also important. Many modern models are optimized mainly for English and other common languages, and do not take into account the linguistic features of Turkic languages.

Therefore, in our study, it is necessary to consider optimal methods for adapting models to Uzbek data. In this regard, multilingual models, such as XLM-RoBERTa, mBERT (multilingual BERT) or mT5, are of particular interest. These models are trained on multilingual datasets and allow them to work with data in different languages, including Uzbek. High accuracy can be achieved by fine-tuning such models with emotional data in Uzbek.

Also, when choosing any model, it is necessary to address the issue of the dataset needed for its training. The emotional dataset in Uzbek may be limited. In such cases, data augmentation methods can be used, such as translating existing data, creating new samples through paraphrasing, and artificial data generation. Based on the preliminary analysis, it was found that the most promising approach is to use transformer-based models to classify emotions in communication. In particular, multilingual models such as XLM-RoBERTa also show good results on Uzbek texts. However, in order to effectively use such models, they need to be retrained on the basis of an emotional dataset in Uzbek.

At the same time, in situations where resource consumption is required to be optimized in real-time systems, more lightweight models can be used. In such cases, the possibility of transferring the knowledge of large transformer models to smaller and faster neural networks using the knowledge distillation approach is considered. In addition to choosing a model architecture, special attention should be paid to optimizing the training process.

For this, it is important to correctly select hyperparameters such as various optimization algorithms (Adam, AdamW, RAdam, etc.), the number of training epochs (epochs), and batch size. It is also envisaged to use regularization methods such as dropout, weight decay, and early stopping to solve the overfitting problem.

During the model evaluation process, it is important to check how it works on different data sets using the cross-validation method. This allows you to assess the generalization ability of the model and identify overfitting problems. For evaluation, along with standard criteria such as accuracy, precision, recall, and F1-score, confusion matrix analysis is also performed, which allows you to determine which emotion classes are confused with each other.

Conclusion

The selection of appropriate deep learning models for emotion classification is a multifaceted task that requires careful consideration of data characteristics, model architecture, computational complexity, and real-world applicability. This study has critically examined prominent neural network architectures Convolutional Neural Networks, Recurrent Neural Networks, Long Short-Term Memory networks, and Transformer-based models highlighting their respective advantages and limitations in processing multimodal emotional data. The analysis demonstrated that no single model universally outperforms others across all scenarios; instead, the optimal choice depends on the specific application context, available data modalities, and performance requirements.

Furthermore, the justification of model selection must incorporate both quantitative performance metrics and qualitative factors such as interpretability and scalability. By integrating these criteria, researchers and practitioners can develop more robust, efficient, and accurate emotion classification systems. The ongoing advancements in deep learning architectures and training methodologies promise continued improvements in affective computing, ultimately enhancing human-computer interaction and broadening the scope of emotion-aware technologies.

Future research should focus on hybrid models and adaptive frameworks capable of dynamically adjusting to diverse emotional contexts and data types. Additionally, ethical considerations related to privacy and bias in emotion recognition warrant increased attention to ensure responsible deployment of these technologies.

References

1. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
3. Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. *IEEE Transactions on Affective Computing*, 10(1), 18–31.
4. Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2018). From Facial Expression Recognition to Interpersonal Relation Prediction. *International Journal of Computer Vision*, 126, 863–877.

5. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
6. Karimov, O. R., & Tursunov, B. T. (2020). Application of deep learning methods in emotion recognition from speech signals. *Journal of Computer Science and Engineering*, 14(2), 45–52. (in Uzbek)
7. Karimova, D. M., & Usmonov, F. K. (2019). Neural networks in affective computing: An overview. *Uzbek Journal of Information Technologies*, 4(1), 34–42. (in Uzbek)
8. Ekman, P. (1992). An Argument for Basic Emotions. *Cognition and Emotion*, 6(3-4), 169–200.
9. Li, S., Deng, W., & Du, J. (2017). Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. *IEEE Conference on Computer Vision and Pattern Recognition*, 2852–2861.
10. Wang, S., & Wang, Y. (2021). Emotion Recognition Using Multimodal Deep Learning Approaches: A Review. *IEEE Transactions on Neural Networks and Learning Systems*.