The Application of Machine Learning and Deep Learning-Based Algorithms in Facial Expression Recognition

Xi Liu
Qingdao Cornerstone Bilingual School, QingDao, China

Keywords: Deep Learning, Machine Learning, Facial Expression Recognition.

Abstract: Facial expression recognition is essential for enhancing human-computer interaction and empathy in artificial intelligence, aiding in many domains such as mental health assessments, improving customer service, and enabling more intuitive educational technologies. This paper gives a comprehensive review of the applications of machine learning and deep learning in facial expression recognition. The research includes face emotion recognition framework based on machine learning, traditional machine learning algorithm and deep learning algorithm. The facial emotion recognition framework based on machine learning includes several steps, such as data set collection, pre-processing, model building and training. Traditional machine learning algorithms mainly include two methods for facial expression recognition based on decision tree and support vector machine. Deep learning is implemented using artificial neural networks. However, there are still some challenges and limitations, such as model complexity, data quality reliability issues, and balance of interpretation accuracy, and user understanding, and trust are also factors to consider.

1 INTRODUCTION

Facial expression recognition refers to the use of predictive models to learn facial features to predict different expressions, including crying, anger. The importance of facial expression recognition, the emotional monitoring of medical patients, the status monitoring of drivers, and the status monitoring of students in class. Facial expression recognition is difficult, and artificial intelligence can accurately predict it. Accuracy is the hardest, because facial expression features are hard to separate. Although different facial expressions have different facial expression features, they are too similar to separate. For example, opening mouth can mean smile, cry, surprise, or other emotions (He, 2005).

Artificial intelligence is a big improvement from “data processing” to “knowledge processing”. Artificial intelligence is one of the branches from computer science. It has already been called one of the three cutting- edge technologies in the 21 centuries. It develops rapidly in the last 30 years. It has been applied widely in many scientific fields, and it got numerous achievements. Artificial intelligence applies in computer science, psychology, philosophy, linguistics, and other subjects (Zhi, 2018). It has been used in medical diagnosis, logistics warehousing, equipment manufacturing, online learning, tourism transportation, and other fields (Li, 2024; Qiu, 2024; Sun, 2020; Wang, 2024; Wu, 2024; Zhou, 2023). Because identifying potential information from massive data set is very hard, artificial intelligence import a lot of advanced machine learning technologies. Machine learning technologies include neural networks, support vector machine, genetic algorithm. The most important applications of artificial intelligence in tourism transportation are intelligent driving and intelligent recommendation. Intelligent driving can get the information such as the position of people, position of cars, and position of obstacle by sensor around the car. Controller can use this information to plan the safest route. Artificial intelligence also is used in space exploration. If some areas that people can’t get in, the robot can replace people to get in (Zhao, 2017). The remaining part of this paper also includes methods, discussions, as well as the conclusion section. The method will provide a detailed description of some of the methods related to facial expression recognition implemented by others, and discuss the current progress, shortcomings, and future prospects of this field. The conclusion will summarize the entire article.

Liu, X.
The Application of Machine Learning and Deep Learning-Based Algorithms in Facial Expression Recognition.
DOI: 10.5220/0012972900004508

In *Proceedings of the 1st International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2024)*, pages 786-789
ISBN: 978-989-758-713-9

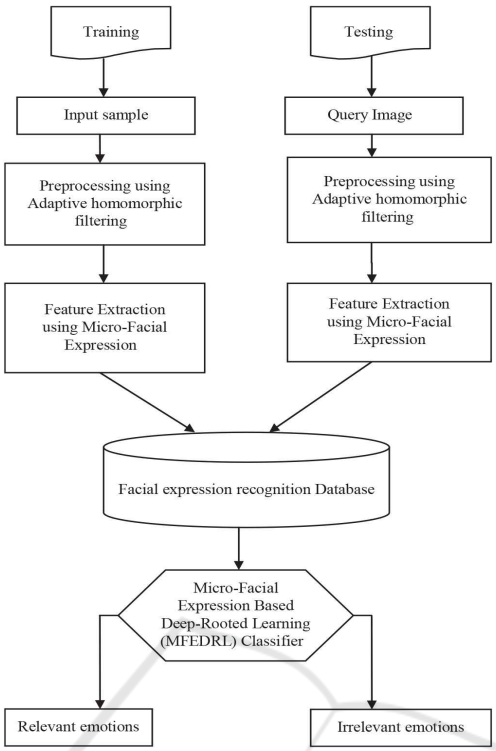

Figure 1: The framework of machine learning-based facial emotion recognition (Lalitha, 2019).

2 METHOD

2.1 Framework of Machine Learning-Based Facial Emotion Recognition

Framework of machine learning-based facial emotion recognition typically shown in Figure 1 includes many steps. First step is data set collection. Participants are required to try their best to maintain a neutral expression, when they watch a video that stimulates emotions in a laboratory environment. Then record a video that is about participants' changes in facial expression. Next annotate the video manually and collect micro emoji video data (Lv, 2024). The second step is data set preprocessing. There are several ways to preprocess data set such as normalization and data enhancement. Normalization simplifies the process by mapping data to a 0 to 1 range, making calculations more straightforward. This method transforms dimensional expressions into dimensionless ones, facilitating easier data handling. Normalization method mainly includes four type such as linear function transformation, logarithmic function transformation, arc tangent function transformation, and normalize the mean of 0 1 (Li, 2011). Data enhancement is a technology that generate new training samples from existing training samples. It also is a way which is low cost and effectiveness in improving the performance and accuracy of machine learning models at a type of data constrained environment (Meng, 2021). Another step is model construction. The most commonly used method for building models is decision tree and random forest. Decision tree is like a classification process. Defining the various branches of the decision tree by subdividing the remote sensing data set from one set to another. The decision tree is composed of a root node, a series of internal nodes, and terminal nodes. Every node has a parent node and two or more child nodes (Li, 2003). Random forests are a combining classifiers base on statistical learning theory. It combines decision tree algorithm and bootstrap resampling method. The core of this algorithm is to construct a tree classifier, which then uses voting for classification and prediction. This approach has found application in a wide range of classification, screening, and prediction tasks (Cao, 2024). Training models is a crucial phase, beginning with the setting of experimental parameters as its foundational operation. This is followed by testing the model, and the final step involves deploying the models.

2.2 Traditional Machine Learning Algorithms

2.2.1 Decision Tree Based Facial Expression Recognition Algorithm

The basic idea of decision tree algorithm is to construct a decision tree by training given data. Every path from the roots to the leaves of a tree is a rule embedded in the data. The working principle of the C5.0 pentagonal model is based on providing maximum. Each sub sample is defined by the first segmentation, and then segmented again based on different fields. Repeat this process until sub samples cannot be further segmented. Last, the lowest level segmentation will be rechecked and any segmentation that does not significantly contribute to the model values will be removed or trimmed.

2.2.2 Facial Expression Recognition Algorithm Based on the Support Vector Machine

Vector computation is the new generation of mathematical algorithms based on statistical learning theory. Classification includes two types. One is linear. Another one is nonlinear. Nonlinear is more complicated than linear. Support vector machine can solve nonlinear problems easily. The feature dimension increase dramatically can be solved by using kernel function. First, confirm distribution situation of enter data within the space. Because the data of images has randomness, unable to accurately define data. The core of a high- performance Support Vector Machine (SVM) is composed of kernel functions. The relationship between two samples has been determined in kernel function. There is no direct relation between mapping itself and kernel function. Using kernel function can skip calculating mapping process. The purpose is to integrate and statistically analyze the data of the image set to establish a boundary.

First step is using corner features and texture features of the face as landmark features for different expressions. Then extract the above features into an array form using effective algorithms. Then using simple and efficient support vector machines as classification learning tools. Next using sample labels to establish mathematical models of expression features for different expressions. Last Calculate the recognition accuracy and validate the reliability of the model through samples (Long, 2021).

2.3 Deep Learning Algorithms

2.3.1 Facial Expression Recognition Algorithm Based on Artificial Neural Network

BP neural network learns from samples. Adjust the connection weights in the network. Then achieve non logical induction. Neural network models have many features. First, there is no feedback link between neurons in each layer. Second, there is nothing between the mountains and rivers of each dynasty. Last, only adjacent layers of neurons have connections. Propagate forward to the shadow node before input signal. After passing through the action function. Then propagate the output information of the lead node to the output node. Last, provide the output result.

3 DISCUSSION

Although significant progresses have been achieved, there are many limitations and challenges. Interpretability refers to the degree to which humans can understand the reasons and processes behind decision-making. These complex models, such as deep neural networks, have achieved significant performance improvements, but their internal working mechanisms are often difficult to explain like black boxes. Understanding the decision-making process of these models poses a huge challenge. The accuracy of interpretability highly depends on the quality and reliability of input data. In the real world, data often contains noise values or biases, which directly affect the interpretability of the model. Ensuring data quality and understanding how data affects model decisions are key to achieving effective interpretation. When increasing the interpretability of a model, some performance and accuracy may be sacrificed. For example, simpler models such as decision trees are usually easier to interpret, but they are not able to handle complex data relationships. Finding the best balance between interpretability and performance is a key challenge. Even if explanations can be provided technically, the understanding and trust of users in these explanations is also an important consideration. Different user groups, such as professionals and ordinary users may have completely different needs and understanding of explanations. Designing an easy to understand and trustworthy way of explanation is crucial for achieving the correct goal of XAI. Applicability, also known as functionality, refers to the various

performance of a product or project that meets its intended use. In the future there will be many opportunities, and this field will develop rapidly. There are several possible solutions. Expert system is one of them.

An expert system is an intelligent computer program system that contains a large amount of knowledge and experience at the level of experts in a certain field. It can apply artificial intelligence technology and computer technology to reason and make judgments based on the knowledge and experience in the system, simulate the decision-making process of human experts, and solve complex problems that require human experts to handle. In short, an expert system is a computer program system that simulates human experts solving domain problems.

Another one is SHapley Additive exPlanations (SHAP). SHAP interprets the output of any machine learning model in a unified way. SHAP links game theory with local interpretation, combining previous methods, and attributing unique, consistent, and locally accurate additive features based on expected representation. Transfer learning, as the name suggests, is the process of transferring the trained model parameters to a new model to assist in its training. Considering that most data or tasks are correlated, transfer learning allows for the sharing of already learned model parameters with new models. This method accelerates and optimizes the model's learning efficiency by leveraging existing knowledge, rather than starting from zero like most networks.

4 CONCLUSION

This work completed a comprehensive review of machine learning and deep learning in facial expression recognition has been completed. Many approaches such as machine learning, deep learning, and ANN were investigated. Until now, there also are many challenges and limitations. The first one is Model complexity. The second one is Data quality and reliability. Next one is Data quality and reliability. Then another one is the balance between interpretation and accuracy. The last one is User understanding and trust. However, this paper has relatively little discussion on application scenarios such as how facial recognition is used in medical settings, which can be considered in the future.

REFERENCES

Cao, Z. F. 2024. Research on optimization of random forest algorithm [Dissertation]. Capital University of Economics and Business.

Feng, X. D., Li, S. C., Li, S. C., et al. 2011. Normalization methods for multi-field information of a mine roof water inrush model test. Journal of Coal Science and Engineering, 36(3), 5.

He, L., Zou, C., Bao, Y., & Zhao, L. 2005. Research progress on facial expression recognition of faces. Journal of Circuits and Systems, 10(1), 6.

Lalitha, S. D., & Thyagharajan, K. K. 2019. Micro-facial expression recognition based on deep-rooted learning algorithm. International Journal of Computational Intelligence Systems, 12(2), 903-913.

Li, M., He, J., Jiang, G., & Wang, H. 2024. DDN-SLAM: Real-time Dense Dynamic Neural Implicit SLAM with Joint Semantic Encoding. arXiv preprint arXiv:2401.01545.

Li, S., & Zhang, E. X. 2003. Research on remote sensing image classification method based on decision trees. Regional Research and Development, 22(1), 17-21.

Long, Y., Chang, L., & Ren, X. G. 2021. Research on facial expression recognition method based on support vector machine. Value Engineering, 40(11), 2.

Lü, L., Han, T., Fang, Y. L., et al. 2024. A crowdsourcing-based method for collecting micro-expression datasets. Patent No. 202110636486. Issued March 30, 2024.

Meng, W. 2021. What is data augmentation? Computer and Network.

Qiu, Y., Hui, Y., Zhao, P., Cai, C. H., Dai, B., Dou, J., ... & Yu, J. 2024. A novel image expression-driven modeling strategy for coke quality prediction in the smart cokemaking process. Energy, 130866.

Sun, G., Zhan, T., Owusu, B.G., Daniel, A.M., Liu, G., & Jiang, W. 2020. Revised reinforcement learning based on anchor graph hashing for autonomous cell activation in cloud-RANs. Future Generation Computer Systems, 104, 60-73.

Wang, H., Zhou, Y., Perez, E., & Roemer, F. 2024. Jointly Learning Selection Matrices For Transmitters, Receivers And Fourier Coefficients In Multichannel Imaging. arXiv preprint arXiv:2402.19023.

Wu, Y., Jin, Z., Shi, C., Liang, P., & Zhan, T. 2024. Research on the Application of Deep Learning-based BERT Model in Sentiment Analysis. arXiv preprint arXiv:2403.08217.

Zhao, N., & Xian, S. S. 2017. Current status and key technologies of artificial intelligence applications. Journal of the China Academy of Electronics and Information Technology, 12(6), 3.

Zhi, G. 2018. What is artificial intelligence? Popular Science, 000(1), 44-45.

Zhou, Y., Osman, A., Willms, M., Kunz, A., Philipp, S., Blatt, J., & Eul, S. 2023. Semantic Wireframe Detection. publica.fraunhofer.de.