Deep Learning and Machine Learning Based Facial Expression Recognition Employed in Mental Health

Hanyu Lin[a]
Computer Science and Technology, Beijing Information Science & Technology University, Beijing, China

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

Abstract: Facial expressions play an important role in human communication, conveying a wide range of emotions without the need for verbal communication. In recent years, Facial Expressions Recognition (FER) has found applications in various domains, particularly in the medical field. The technology was originally developed, using mostly machine-learning model algorithms, including Support Vector Machine (SVM) etc. However, as the dimension of the characteristics increases, it is difficult to obtain more feature samples. To solve problem, a framework for Principal Components Analysis (PCA)+Latent Dirichlet Allocation (LDA) is proposed. After a few years, the development of deep learning gave the FER experiments many large models to use, such as Convolutional Neural Network (CNN). However, the complexity of facial expressions, compounded by factors such as illumination, posture, and occlusion, poses challenges for accurate recognition using traditional methods, the Long-Short-Term-Memory (LSTM) layer structure is added to the CNN, and it developed into a new model called LSTM-CNN. Deep learning excels in handling complex data and large-scale datasets, making it the preferred choice for FER due to its adaptability and end-to-end learning capability. However, its lack of interpretability and challenges with complex data can hinder accuracy and trust in results, especially in medical applications like mental health diagnosis. Preprocessing data and refining identification algorithms are crucial steps to improve accuracy in FER projects.

1 INTRODUCTION

The expression on the face is really complex that through the control of eye muscles, face muscles and oral muscles, people can make more than 20 species expressions, such as ‘happy’, ‘sad’, ‘fear’. Those expressions are usually used in people’s daily life to correctly express their feelings at the time without saying any words. Due to the complexity of facial expressions and the effects of illumination, posture and occlusion, traditional methods can’t accurately identify the features of expressions to judge which expression is it. Identifying emotions based solely on the eyes is particularly prone to inaccuracies. Deep learning, on the other hand, offers a robust solution by using large datasets for model training in many domains (Liu, 2021; Liu, 2023, Qiu, 2020). It can pick up more elaborate characteristic and learn the complexity relationship between different expressions. After the training, it can give a higher accuracy in identification, and cut down time required, which can be considered as an effective solution.

In recent years, Facial Expressions Recognition (FER), as a new technique, has been employed in many areas, especially in medical treatment field. It can sequence specific features of expressions from a given live videos or images. For instance, Hearst et al. proposed Support Vector Machine (SVM) and soon applicable in FER to solve the data classification problems by finding an optimal decision hyperplane (Hearst, 1998). In addition, many researchers proposed a series of more efficient feature extraction algorithms, like Principal Components Analysis (PCA), Latent Dirichlet Allocation (LDA), Local Binary Patterns (LBP), combined with SVM, Artificial Neural Network (ANN) to improve the accuracy of recognition. As time progresses, fast-paced life brings kinds of mental disease to many people, and those who are living with mental disease are getting younger. Mental health has actually become an important globalization social issue. This

[a] https://orcid.org/0009-0004-0303-3163

Lin, H.
Deep Learning and Machine Learning Based Facial Expression Recognition Employed in Mental Health.
DOI: 10.5220/0012936800004508
In *Proceedings of the 1st International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2024)*, pages 285-288
ISBN: 978-989-758-713-9

issue has allowed for the rapid development of technical means of mental treatment in recent years. The identification of facial expressions is an important technique used in treatment to identify the inconspicuous expressions changes. By analysing the features of expressions which are captured by facial expressions recognition technology, it can be understood the emotions state of patients more accurately. With deep learning came into a new step that had been get a great development. The application of CNN in FER is considered. Breuer et al. use CNN to calculate the relationship between the facial action coding system and action units. Then, they use the Extended Cohn-Kanade (CK+), NovaEmotions and FER2013 datasets to prove their findings, identified the micro-expressions based CNN models. And they introduce a basic LSTM Recurrent Neural Network (RNN) to achieve most state-of-the-art accuracy (Breuer and Kimmel, 2017). Peng also proposed to introduce the LSTM model into CNN to improve CNN models, and made a comparison in SVM, CNN and novel CNN (Peng, 2021).

This paper focuses on facial expressions identification which is used in the treatment of mental diseases. The content of several other parts of the paper is as follows. First of all, the details of some existing methods will be introduced in section 2. Then, this paper will discuss the shortcoming and development direction in the existing methods in section 3. And the last section is to summarize the first three sections, conclude and reorganize the result from the paper.

2 METHODS

2.1 Traditional Machine-Learning Based Facial Expressions Recognition Models

2.1.1 SVM

As a traditional machine learning model, SVM is a kind of classification algorithm which has been used in FER to classify different expressions from images (Bellamkonda, 2015; Tsai and Chang, 2018). Because the data is usually images data, each facial expression is treated as a data point and the features of those data points were extracted from images. To classify the expressions, this algorithm can find a hyperplane in the vector space of features. The hyperplane can put the expression into two sides, one is healthy emotion, the other is unhealthy emotion (Peng, 2021). The advantage is the ability to handle the high-dimensional space of feature and to do well with the nonlinear classification problem with suitable kernel functions.

2.1.2 PCA

PCA has no explicit model structure, it's a kind of linear transformation method, can put the data into a new coordinate like a projector (Abdulrahman, 2014). When PCA using in FER, it can compress the expression features data to solves the problem from identifies the similarities and differences in those data. By calculating the expressions eigenvalues and the expressions eigenvectors of the covariance matrix, then the eigenvector corresponding to the largest eigenvalues was selected as the principal components to project the raw data onto the principal components. It's a kind of technique that can achieve dimension reduction from 3D to 2D even through there are N expression images (Deng, 2005). It can reduce the amount of expression features while retaining most of the information, reducing redundant information between data at the same time.

2.1.3 LBP

LBP has no explicit model structure, it's a kind of method that can extract the expression features from images, it can transform the facial image into a grayscale image. The original LBP operator is defined in the 3×3 traditional neighbourhood model, with the centre of the neighbourhood as a threshold. Compared the grey value of the 8 pixels with the grey value of centre. If the surrounding pixel is greater than the central pixel value, the position of the pixel is marked as 1, otherwise 0 (Bellamkonda, 2015). By describing the local texture of the image, it can capture the characteristics of texture, which is benefit for texture recognition.

2.1.4 FER Based on SVM with LBP and PCA and PCA+LDA

Muzammil Abdulrahman proposed a model using with LBP and PCA with SVM for FER (Abdulrahman, 2017). Using PCA and LBP algorithms to extract expression features, and SVM are used for classification.

PCA is a supervised learning algorithm that is mainly used for feature dimension reduction and classification. The goal is to find an optimal projection, so that the data of different categories can be separated in the new feature space, while the data within the same category are clustered as closely as possible (Deng, 2005). LDA differs from PCA. PCA

aims to maximize the variance of the data, while LDA aims to maximize differences between classes and minimize differences within classes.

Deng proposed that as the dimension of the characteristics increases, it is difficult to obtain more feature samples. To solve this significant problem, the author proposed a new framework PCA+LDA. This framework can describe that PCA maps the original t-dimensional feature acquisition to the f-dimensional feature as an intermediate space, and then LDA projects the PCA output onto the new g-dimensional feature vector (Deng, 2005).

2.2 Deep-Learning Based Facial Expressions Recognition Models

2.2.1 CNN

The Neural Network is a mathematical model or computational model that mimics the structure and function of the biological neural network. The structure of CNN can be divided into three large layers: input layer, hidden layer and output layer (Li, 2021; Qiu, 2024). In the hidden layer, there are three layers: Convolutional Layer for extracting the expression features, Pooling Layer for down sampling without damage identification results, Fully Connected Layer for classification.

When the convolution step is 1, the convolution core scans the elements of the feature map one by one. As an image of 16×16 input into the input layer, after a convolution kernel with a unit step size and no full of 5×5, it will output a feature map image which size of 12×12. After the data is convolved by the convolution kernel in the convolutional layer to extract the expression features. The pooling layer had a size of 2×2 polling box means that the length of the moving step was 2, which just like to the compression of the image. This process is performed repeatedly until the number of times the operational requirements are met (Peng, 2021).

2.2.2 LSTM-CNN

X. L. Peng then proposed to introduce LSTM into CNN. The LSTM as a layer can reduce the impact of the background colour, because the algorithm can make features more prominent which are continuous changes when people change the expression. This layer was configured with 64 hidden neurons, weights initialized using glorot_normal initialization, the bias initialized to 0 (Peng, 2021). The structure of LSTM can be put into three parts: input gate, forget gate and output gate. The input gate, calculated by the tanh function, is utilized to extract expression features from the candidate state. Then in the next gate by forgetting the old information and adding newly information to updates the old cell. In the last, the output gate will output the new cell (Yang, 2023). This combination takes advantage of the ability of LSTM to process sequence data and the advantages of the CNN in extracting spatial features, allowing the model to perform better in tasks such as face emotion recognition.

3 DISCUSSIONS

To compare the deep learning models with machine learning, deep learning has better performance on processing complex data and solving different problems. The main reason why FER chooses deep learning is that it can better handle complex features and large-scale data sets of face data, and can realize end-to-end learning process. Also, deep learning models are more adaptable and generalized, they can be better able to handle large-scale data and complex patterns, and also have good predictive power for the data which has never been seen before.

However, there are still some disadvantages for deep learning models. The interpretability of AI is that although these models can run completely, but there hardly have any explanation for why specific features over other specific features are selected during training, or how the correlation in the trained data is represented in the selection of features (Chakraborty, 2017). For example, in the medical scenario, due to the interpretability of AI, the reasons for the formation of facial expressions provided by patients are not explained, which may lead to doctors and patients to distrust the results from deep learning model, that will affect the doctor's judgment of the condition, and then affect the application and development of mental health.

Another reason for the low accuracy of deep learning models is because the original data is too complex and unique, so it is hindered when using some feature extraction algorithms. S. Li et al. proposed that different illumination, background and head posture can have a large impact on the data images, thus affecting the performance of FER (Li and Deng, 2020). When conducting the FER project, if the accuracy of the result is low, experimental group can choose to preprocess the data and adjust the identification algorithm. Reduce the number of expressions first, after ensuring a steady accuracy rate, then add to the type quantity. Here are some ways to preprocess the data by adjusting the image to a uniform size, scaling the pixel value of the image to a fixed range (usually [0,1] or [-1,1]), and standardizing the pixels and pose or angle values.

In the future, FER is expected to become an important tool in the field of mental health, helping to identify people's emotional changes. Combining deep learning and emotional intelligence, more accurate expression recognition can be realized to assist the diagnosis of mental health problems such as depression and anxiety. This technology may be combined with advanced biosensing hardware technology to provide a more comprehensive assessment of emotion (Deng, 2019; Sugaya, 2019). In the future, the development of these technologies may lead to more mental health AIDS and improve the mental health status of individuals.

4 CONCLUSIONS

In this paper, a review of machine learning and deep learning in FER was provided. This paper discussed models in methods of machine learning and deep learning. In machine learning, there are several models like SVM, LBP, PCA and PCA+LDA. In deep learning, there are several models like CNN and CNN-LSTM. Overall, machine learning is less accurate than deep learning in FER. But deep learning also has problems such as the inability to handle complex data and interpretability. This paper there are only limited models and algorithms about machine learning and deep learning. In the future the further study plans to increase the exploration of the usage scenarios and the method exploration of how the data are processed.

REFERENCES

Abdulrahman, M., Gwadabe, T. R., Abdu, F. J., & Eleyan, A. 2014, April. Gabor wavelet transform based facial expression recognition using PCA and LBP. In 2014 22nd signal processing and communications applications conference (SIU) (pp. 2265-2268). IEEE.

Abdulrahman, M., & Eleyan, A. 2015, May. Facial expression recognition using support vector machines. In 2015 23nd signal processing and communications applications conference (SIU) (pp. 276-279). IEEE.

Breuer, R., & Kimmel, R. 2017. A deep learning perspective on the origin of facial expressions. arXiv preprint arXiv:1705.01842.

Bellamkonda, S., & Gopalan, N. P. 2018. A facial expression recognition model using support vector machines. IJ Mathematical Sciences and Computing, 4, 56-65.

Chakraborty, S., Tomsett, R., Raghavendra, R., Harborne, D., Alzantot, M., Cerutti, F., ... & Gurram, P. 2017, August. Interpretability of deep learning models: A survey of results. In 2017 IEEE smartworld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, Internet of people and smart city innovation (pp. 1-6). IEEE.

Deng, H. B., Jin, L. W., Zhen, L. X., & Huang, J. C. 2005. A new facial expression recognition method based on local Gabor filter bank and PCA plus LDA. International Journal of Information Technology, 11(11), 86-96.

Deng, X., Li, L., Enomoto, M., Kawano, Y., 2019. Continuously frequency-tuneable plasmonic structures for terahertz bio-sensing and spectroscopy. Scientific reports, 9(1), p.3498.

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. 1998. Support vector machines. IEEE Intelligent Systems and their applications, 13(4), 18-28.

Li, S., & Deng, W. 2020. Deep facial expression recognition: A survey. IEEE transactions on affective computing, 13(3), 1195-1215.

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. 2021. A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems, 33(12), 6999-7019.

Liu, Y., Liu, L., Yang, L., Hao, L. and Bao, Y., 2021. Measuring distance using ultra-wideband radio technology enhanced by extreme gradient boosting decision tree (XGBoost). Automation in Construction, 126, p.103678.

Liu, Y. and Bao, Y., 2023. Intelligent monitoring of spatially-distributed cracks using distributed fiber optic sensors assisted by deep learning. Measurement, 220, p.113418.

Peng, X. 2021. Research on Emotion Recognition Based on Deep Learning for Mental Health. Informatica, 45(1).

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.

Qiu, Y., Yang, Y., Lin, Z., Chen, P., Luo, Y., & Huang, W. 2020. Improved denoising autoencoder for maritime image denoising and semantic segmentation of USV. China Communications, 17(3), 46-57.

Sugaya, T., Deng, X., 2019. Resonant frequency tuning of terahertz plasmonic structures based on solid immersion method. 2019 44th International Conference on Infrared, Millimeter, and Terahertz Waves, p.1-2.

Tsai, H. H., & Chang, Y. C. 2018. Facial expression recognition using a combination of multiple facial features and support vector machine. Soft Computing, 22, 4389-4405.

Yang, W. 2023. Extraction and analysis of factors influencing college students' mental health based on deep learning model. Applied Mathematics and Nonlinear Sciences.