Two Methods to Realize Facial Expression Recognition and Classification in Tough Conditions

Shuyi Wang,
Research School of Computer Science,
Australian National University,
Canberra Australia,
Email: u7094877@anu.edu.au

Abstract. In recent years, more and more applications of human-computer interaction appear, and an important basis for human-computer interaction is facial expression, so the main problem of this paper is to implement two different methods for facial expression recognition and classification. The two methods are decision tree and convolutional neural network. This study implements the two technologies using python and pytorch, inputs the same data set obtained in the tough environment, and compares the accuracy of its output. This study also discusses the improvement of CNN model accuracy through pruning. Through experiments, it is concluded that the accuracy of the decision tree classifier is lower than that of the CNN model, and pruning can improve the accuracy of the CNN model.

Keywords: Facial expression recognition, Decision tree, Neuron network, Pruning.

1 Introduction

Face recognition is the most intuitive and effective method of human biometric recognition technology. It involves computer vision, artificial intelligence and many other disciplines. It can play a significant role in human-computer interaction, medical treatment and other fields and has a good application prospect.

The traditional facial expression recognition mainly extracts facial features, so the selection of features directly affects the recognition rate, so the robustness is poor. At present, the most mainstream classification methods include decision tree classification and neural network classification.

Therefore, this research aims to implement these two technologies to solve the problem of facial expression recognition, and to find an algorithm that can both improve the recognition rate and robustness by comparing these two popular facial recognition technologies. In addition, this study also discusses the improvement of the accuracy of the CNN model through pruning.

2 Method

2.1 Data set

This study uses a new static facial expression database named Static Facial Expressions in the Wild (SFEW) which extracted from a temporal facial expressions database Acted Facial Expressions in the Wild (AFEW) proposed by Abhinav Dhall, Roland Goecke, Simon Lucey and Tom Gedeon[1].

By looking at the data set, it can be obtained that the data set has a total of 675 rows and 12 columns, each row represents an entity, and each column represents an attribute. The first column represents the name of the image, which is of type Object. The second column represents labels, which are of integer type, contains seven labels and represents facial expressions with numbers. 1 represents angry, 2 represents disgust, 3 represents fear, 4 represents happy, 5 represents neutral, 6 represents sad, and 7 represents surprise. The remaining ten columns represent the two ways of first five principal components of the image, all of float type.

2.2 Data preparation

It can be noticed that first five principle components of PHOG features only has 674 rows and there are infinite decimals in it, so use the command *np.nan_to_num()* to make it finite. Because the data set is already in the form of converting pixel values into principal components, there is no need to do this step and can use the data set directly.

In order to train and test the model, we divide the data set into a training set and a test set, and randomly select 80% of the data set as the training set and 20% as the test set. In addition, for the CNN model, change the data type of the data set to tensor type.

2.3 Decision tree classification

Decision tree is a kind of supervised learning, so some samples need to be given, and the samples need to include input data and output. Therefore, we first divide the training set obtained during preparation into training data and training labels. The same is true for the test set.

By reading the material [2], can understand three methods and basic principles of decision tree classification. The attribute selection method used in the code of this study is information entropy. We hope to find an attribute R with such characteristics: the information gain before and after splitting with attribute R is the largest than other attributes. The definition of information is as formulae (1), m represents the number of a certain category in the data set D, Pi represents the probability that any record in D belongs to the category, and Info(D) represents the amount of information required to separate the different categories of the data set D.

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2 p_i \,. \tag{1}$$

Info() in formulae (1) is actually the entropy, if the uncertainty of the category of a data set is higher, its entropy is greater. The amount of information needed to separate the different classes of the data set D is shown in formulae (2).

$$Info_R(D) = \sum_{j=1}^{k} \frac{|D_j|}{|D|} \times Info(D_j) \,. \tag{2}$$

Information gain represents the amount of information that attribute R brings to classification. The definition is shown in formulae (3). The largest attribute of gain is the attribute we expect.

$$Gain(R) = Info(D) - Info_R(D) \,. \tag{3}$$

The decision tree classifier is obtained by fitting the training set, and then the classifier is used to predict the output of the test set, and finally the decision tree of the prediction result is visualized as Figure 1 and Figure 2.

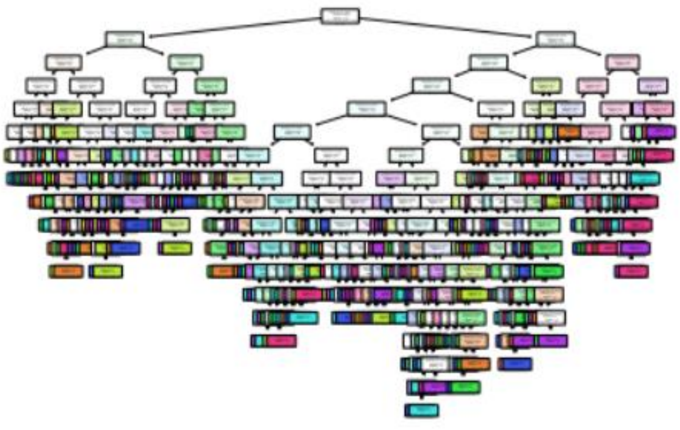

Fig. 1. Decision tree visualization, using first five principal components of LPQ features.

Fig. 2. Decision tree visualization, using first five principal components of PHOG features.

2.4 Convolutional neural network

Traditional convolutional neural networks generally include convolutional layers, pooling layers and fully connected layers. In the convolutional layer, the convolution kernel parameter values are optimized through backpropagation, and the parameters are shared. The neural network can reduce the weight parameters through shared parameters and sparse connections, and at the same time prevent the occurrence of overfitting. After the convolutional layer, the connection with pooling layer can perform feature dimensionality reduction and compress the data volume. In the fully connected layer, each neuron is connected to the neuron of the previous layer to obtain the recognition rate. [4]

In the CNN model constructed in this study, there are two layers of convolution, uses ReLU activation function, uses cross entropy loss, and stochastic gradient descent optimizer. The input image undergoes a convolution operation of a 3X3X64 convolution kernel, and then performs a 1X1 pooling to obtain an output. Then this output is subjected to a convolution operation of a 3X3X256 convolution kernel and perform a 1X1 pooling to obtain a new output. After the convolution is completed, flatten the data and then enter the fully connected layer. The data passes through the hidden layers. Finally, it passes through the output layer containing 7 neurons. Set the epoches to 500 and the learning rate to 0.01.

3 Results

The results of the decision tree classifier for LPQ features are shown in Table 1. The results of the decision tree classifier for PHOG features are shown in Table 2. The accuracy of the former is 0.15, while the accuracy of the latter is higher, but overall, the accuracy is very low. It can be obtained that by converting the images to first 5 principal components of pyramid of histogram of gradients features, then using the data to train the model, the result will be more accurate.

Table 1. Decision tree output, using LPQ features.

Facial expressions	Precision	Recall	F1-score
1	0.28	0.29	0.29
2	0.00	0.00	0.00
3	0.07	0.11	0.09
4	0.29	0.22	0.25
5	0.06	0.04	0.05
6	0.13	0.08	0.10
7	0.17	0.21	0.19
accuracy			0.15

Table 2. Decision tree output, using PHOG features.

Facial expressions	Precision	Recall	F1-score
1	0.11	0.12	0.12
2	0.00	0.00	0.00
3	0.24	0.28	0.26
4	0.40	0.22	0.29

5	0.17	0.13	0.15
6	0.14	0.12	0.13
7	0.14	0.16	0.15
accuracy			0.16

The advantages and disadvantages of the decision tree classifier are as follows. The calculation of decision tree classification is very simple and can handle irrelevant feature data, but it is not sensitive to missing intermediate values, and when there are many features, it will cause overfitting.

The test results of the convolutional neural network are shown in Table 3. The accuracy of using the CNN model is higher than that of the decision tree classifier, and similarly, the accuracy of using PHOG features is higher than that of LPQ features. After pruning the CNN model, the model is compressed, and the accuracy is improved slightly.

CNN can share convolution kernels, automatically perform feature extraction, and process high-dimensional data. However, the gradient descent algorithm may converge to a local minimum instead of a global minimum, and the pooling layer may cause some information to be lost.

Each method has its advantages and disadvantages. Generally speaking, it is better to use CNN for facial recognition because of its higher accuracy. Choosing a suitable optimizer can also avoid the problem of convergence to a local minimum. Because facial expressions have many features, the decision tree classification may be over-fitting, so it is also not the most suitable.

Table 3. CNN model output.

Feature type	Test loss	Test accuracy	Test accuracy after pruning
LPQ	4.984	0.134	0.147
PHOG	3.548	0.167	0.172

4 Discussion

Compared with the results of L.K. Milne, T.D. Gedeon and A.K. Skidmore [3], the conclusions are basically the same, not very different. It can be concluded that these two methods can handle classification problems on different data sets, and their performance is basically the same.

Two researches both found that the neural network model is more suitable for facial expression recognition, and decision tree classification may lead to overfitting. Both have obtained the result that decision tree has high false positive probability. However, this study does not come to the conclusion that the use of neural network models will also overfit, which may be related to the different sizes of the two data sets and the neural network structure and parameter settings used.

5 Conclusion

In conclusion, both neural network model and decision tree classification can realize facial expression recognition and classification, and the results are similar. The accuracy of the CNN model is slightly higher. Pruning the neural network model can improve its accuracy. When using facial expression recognition and classification, in order to avoid overfitting, it is better to use a neural network model instead of a decision tree classifier. When the accuracy of the neural network model is not as expected, it can be tried to improve its performance through pruning.

6 Future work

Future work mainly includes two aspects. On the one hand, it is to explore whether there is a way to avoid the overfitting problem of the decision tree classifier. The other is to explore methods to improve the accuracy of convolutional neural network models. For example, it can be tried to increase the number of data sets through data enhancement.

References

1. Abhinav Dhall, Roland Goecke, Simon Lucey, Tom Gedeon," Static Facial Expression Analysis in Tough Conditions: Data, Evaluation Protocol and Benchmark".
2. Three ways of decision tree classification algorithm, https://blog.csdn.net/qq_41587243/article/details/88314403.
3. L.K. Milne, T.D. Gedeon, A.K. Skidmore," CLASSIFYING DRY SCLEROPHYLL FOREST FROM AUGMENTED SATELLITE DATA: COMPARING NEURAL NETWORK, DECISION TREE & MAXIMUM LIKELIHOOD".
4. "Research and Implementation of Facial Expression Recognition Based on Data Augmentation and Convolutional Neural Network", 1009-3044(2020).
5. Hao Liu, "Facial Expression Recognition Method Based on Eigenface and Neural Network", Journal of Liaocheng University, 2014.
6. Yu Chenglin, Pei Hailong, "Face recognition framework based on effective computing and adversarial neural network and its implementation in machine vision for social robots", Computers & Electrical EngineeringVolume 92, 2021.
7. Shaohuai Wu; Bo Wang, "Facial expression recognition based on computer deep learning algorithm: taking cognitive acceptance of college students as an example", Journal of Ambient Intelligence and Humanized Computing2021. PP 1-12.