Face Emotions Recognization by Deep Learning Method and Evaluate through Decision Tree

Wenzheng Yang
Research School of Computer Science,
Australian National University
u7102294@anu.edu.au

Abstract: At present, many researches use deep learning to recognize and classify images, but do not use another model to evaluate the results. In this paper, I use convolutional neural network to recognize different emotional images, then classify them, and compare the results with different neural networks. The results show that the performance of convolutional neural network is better than that of fully connected neural network and LSTM neural network. Then the decision tree is used to evaluate the convolutional neural network and simulate the classification. At the same time, the convolutional neural network and the decision tree have achieved high accuracy.

Keywords: face emotions, neural networks , deep learning, decision tree

Introduction

Data Set

In this article, I used an extended version of the dataset. The first five principal components of local phase quantification (LPQ) feature is different from the previous dataset. LPQ is based on the calculation of STFT for local image windows, then first five components of pyramid of histogram of gradients (Phog) features, Phog is the pyramid of histogram of oriented gradients[1], so if we use the previous dataset, the problem becomes to use eigenvalues as input, Then a multi classification network with five input dimensions and seven output dimensions is constructed.
In the extended version of the dataset, there are corresponding pictures in the seven emotions, and the pictures are used as the original data. Convolution neural network is needed to extract image features. There are many facial expression images in the dataset, and the information of the image is stored in every pixel of the image. We can extract image features by defining convolution kernel and other operations, so as to realize the classification of facial emotions.

Deep Learning Model

The purpose of this paper is to recognize and classify facial emotions according to image features. The convolutional neural network is trained by using the image as the input. The results of convolution neural network are compared with those of fully connected neural network and LSTM neural network. In previous studies, they use the back propagation neural network to recognize set of faces under different noise environment conditions.

When we use convolution network for image recognition and classification, the main steps include: using convolution layer to extract features, using pooling layer to extract main features, using full connection layer to summarize the features of each part, and finally generating a classifier for image prediction and recognition. Compared with the traditional artificial neural network model, convolution neural network has more hidden layers. Its unique convolution and pooling operations have higher efficiency in image processing. It has incomparable advantages in image recognition and location and other forms of two-dimensional graphics tasks.[2] In the previous research, I mainly used the full connection neural network and LSTM neural network, and did not get a high accuracy, but using convolution neural network got better results, the prediction accuracy can reach more than 90%.

Decision Tree and Characteristic Input

After the construction of neural network, the internal classification mechanism of neural network is still difficult to understand intuitively, so I use the technology of literature [3] to explain the classification of neural network, and the input pattern is classified according to its influence on each specific output. Open a set of output patterns to generate features on the pattern. This can be achieved by many statistical methods.
In the previous research, I calculated the average value of each feature, took the average value of each feature as the feature input, I got the input that can activate each feature, then extracted the feature rules, and used the feature rules to classify the data. In this article, I extract features from images, and then classify images by feature rules. Image is used to train decision tree, and decision tree is used to simulate and evaluate neural network classification.

Methodology

Data preprocessing

Different from the data set I used before, the previous data set is characteristic data. I need to read the data and then eliminate the invalid data. Because the previous data set is arranged orderly according to the classification, the original data needs to be unordered, and sometimes the value range of the characteristic value may be biased, which is not conducive to the processing of neural network, So it is necessary to normalize the data, and then eliminate the correlation between different features.
In the extended data set, the content of the data set becomes facial emotion image. First of all, we need to divide the data set into training set and test set to prevent over fitting. Then we use the functions in Python to do a series of operations on the image. We use transforms. Randomresizedcrop() to randomly crop the given image to different sizes and aspect ratios, and then scale the cropped image to a specific size. Through this step, we can get that even if the image is only a part of the object, we also think it is this kind of object. Transformation. Randomhorizontalflip() through this step, we rotate the image randomly with a given probability. Transforms. Totensor() transforms the given image into tensor. Finally, the image is normalized by transforms. Normalize(). When obtaining training data, these operations are carried out. Because some operations have random attributes, each epoch has different ways of image processing, so data enhancement is realized.

Fig.1. part of the dataset

Convolutional neural network

For the image in the data set, each pixel stores the information of the image. For convolution layer, we can generate convolution kernel randomly by function to extract certain features from image. The convolution kernel is multiplied by the digital matrix of the image, and then added to get the convolution layer output. The output value is obtained by interacting with different convolution kernels, and the most suitable convolution kernel is obtained by mutual judgment. The larger the output value of the convolution layer, the more the convolution kernel can represent the characteristics of the image. Then, we use batchnorm2d() to normalize the data, so that the network performance will not be unstable because of the large data before relu. The input of pooling layer is the output matrix of convolution layer. The purpose is to reduce the number of training parameters, reduce the dimension of eigenvector, reduce the over fitting phenomenon and reduce the transmission of noise. There are two common pooling layer forms. In this paper, the maximum pooling is used, and the maximum value in the specified area is selected to represent the whole area. In this paper, a total of four convolution pooling. The work of convolution layer and pooling layer is to extract features and reduce the parameters brought by the original image. For the final output, we need a full connection layer to generate a classifier. The function of full join layer is to map feature representation to sample space. Finally, forward propagation.

In this paper, the cross entropy loss function is used. For the last layer, the gradient of the weight is no longer related to the derivative of the activation function, but only proportional to the difference between the output value and the real value. At this time, the convergence is faster, and because of the back propagation, the update of the whole weight will be faster.

I used the Adam optimizer in previous studies, but in this article I use the SGD optimizer. The training speed of this optimizer is very fast for large data sets. However, it is possible to introduce noise and can not completely overcome the local optimal solution.

The steps to use the loss function and optimizer are to calculate the loss, clear the gradient, error back propagation, and update the parameter values.

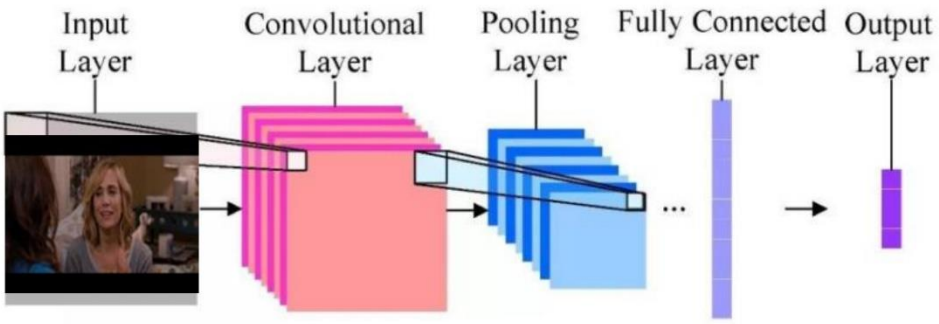

Fig.2. cnn model for face emotions

Fully connected neural network and LSTM neural network

In the previous study, I used Adam optimizer, which considers the first moment estimation and second-order moment estimation of gradient to calculate the updated step length. The optimizer is simple to implement, high efficiency and less memory. Its parameters are descriptive, usually not need to adjust or just a few fine-tuning, and can also achieve automatic adjustment of learning rate[4] and as for LSTM neural network, In one approach, classification, segmentation, and context integration are all carried out by 2D LSTM networks, allowing texture and spatial model parameters to be learned within a single model. The networks efficiently capture local and global contextual information over raw RGB values and adapt well for complex scene images. [5]

Decision Tree and Characteristic Input

Decision tree learning includes three steps: feature selection, decision tree generation and decision tree pruning. The essence of decision tree is the same as that of neural network classification, which is to summarize the classification rules from the training data set, and then classify the data. The criteria of feature selection in decision tree are information gain and information gain ratio. We will choose the segmentation with the maximum information gain, that is, recursive selection of optimal features. If the decision tree over considers all the data and causes over fitting, we need to prune the decision tree and minimize the loss function.

Researchers from various disciplines such as statistics, machine learning, pattern recognition, and Data Mining have dealt with the issue of growing a decision tree from available data.[6]

In this paper, before the decision tree classification of the original image, the original image is first transformed into gray image, and then the label of each facial emotion is obtained, that is, feature extraction. Information entropy is used to train decision tree.

Results and Discussion

Parameter Analysis

Convolutional neural network has four layers. The first layer has 3 inputs and 32 outputs, and the convolution kernel size is 3. The second layer has 32 inputs and 64 outputs, and so on. The fourth layer has 256 outputs. The kernel size of convolution layer is 2, and the stride is 2

By setting the dropout function in the full connectivity layer, the neurons are randomly inactivated to play the role of regularization. The parameters are 0.2 and 0.5 respectively. When the epoch is 246, we can get the optimal accuracy.

Results of classification:

When the learning rate is 0.1, the loss and accuracy curve of convolutional neural network is obtained

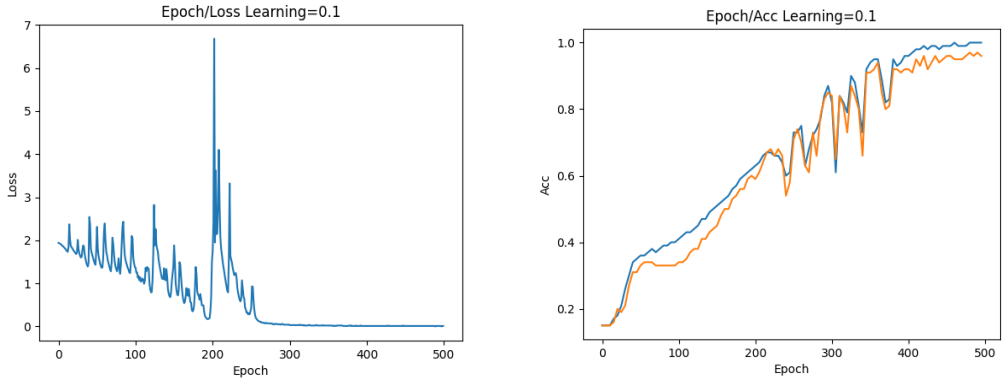

Fig.3.the loss and accuracy of 0.1

When the learning rate is 0.01, the loss and accuracy curve of convolutional neural network is obtained

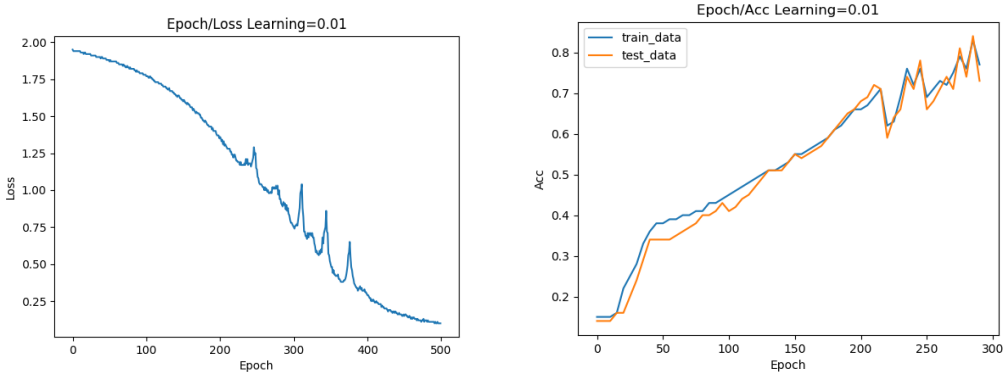

Fig.4.the loss and accuracy for 0.01

The pridiction result of decision tree:

```
The accuracy is 100.00%
```

Fig.5. the accuracy of decision tree

Fully connected neural network:

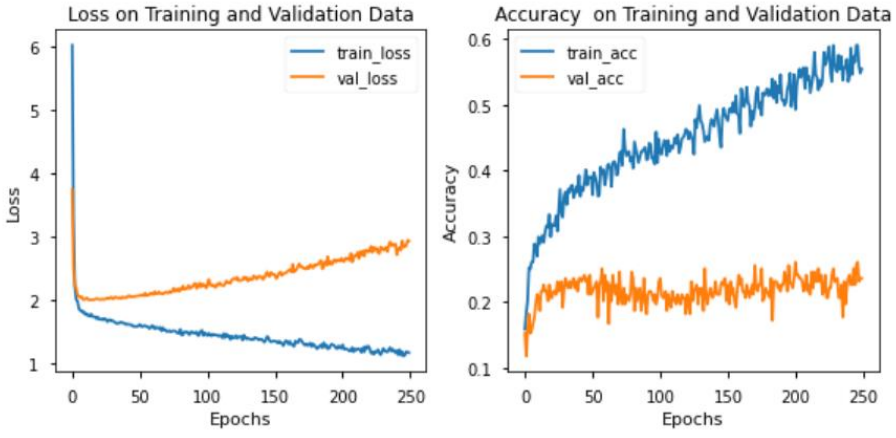

Fig.6. The result of fully connected data

LSTMneural network:

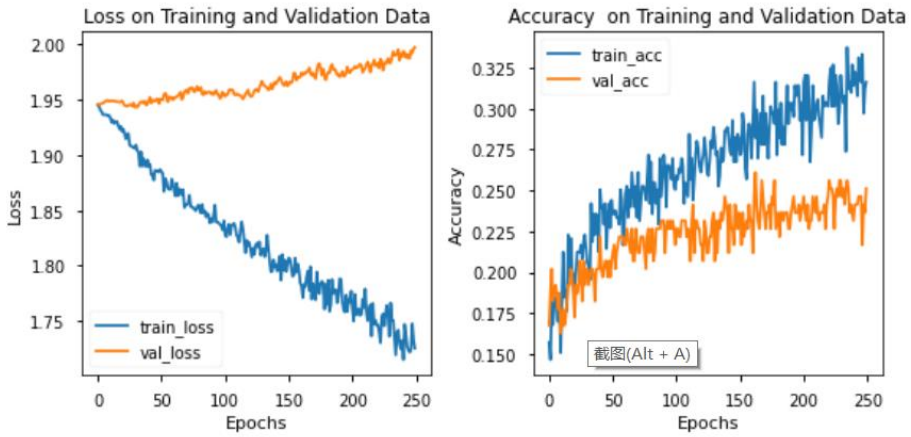

Fig.7. The result of LSTM nn

Discussion

Through our images, we can find that, within a certain range, with the increase of learning rate, the loss will gradually decrease, and then increase. We can find the best learning rate in the minimum loss interval. At the same time, through the analysis, I found that in convolution neural network, the convolution layer has a small proportion of parameters and a large proportion of calculation, while in the full connection layer, the convolution layer has a large proportion of parameters and a small proportion of calculation. Therefore, when we reduce the network parameters or weight clipping, we mainly focus on the full connection layer, while when we optimize the calculation, we focus on the convolution layer. When convolutional neural network underfitting, we can try to reduce the learning rate and increase the complexity of neural network. When it overfitting, we can try regularization and add dropout layer. When the convolutional neural network converges completely, we can try to adjust the learning rate and the number of epochs.

In the comparison of fully connected neural network, LSTM neural network and convolution neural network, we can find that the accuracy of convolution neural network is much higher than the other two neural networks. The parameters of fully connected neural network are too many, which leads to over fitting, while the pooling layer of convolutional neural network can prevent over fitting. And convolution neural network can recognize image transformation, which is equivalent to a prior condition. LSTM neural network is suitable for text generation based on time series, such as machine translation and speech recognition.

In the part of decision tree, we simulate the classification process of neural network through the classification mechanism of decision tree. For example, in softmax, the upper neural nodes are input to the softmax layer after calculation, and the parameters of the trained softmax layer are compared to select the most appropriate node output, and then the classification is completed. We use decision tree to explain neural network classification. By extracting feature inputs, we find the corresponding output for each input, and then predict the output.

In paper [7], we get the influence of each assignment on the final score by generating rules.Similarly, we can extract rules to find the image features that have the greatest influence on the prediction results.

Conclusion and Future Work

Conclusion

Convolution neural network can play a good role in facial emotion recognition and classification. In this paper, I use a lot of methods to improve the performance of convolutional neural network, such as increasing the number of neural network layers, adjusting the parameters, and finally get the convolutional neural network with an accuracy of more than 90%. Through the comparison of convolution neural network, fully connected neural network and LSTM neural network. Convolution neural network is more suitable for image processing, while LSTM neural network is more suitable for text generation. Before forecasting, it is necessary to determine what kind of neural network the data set is suitable for.
For decision tree, I use the classification of decision tree to compare with the classifier in convolutional neural network, which is to extract image features and classify images. I can also limit the height of the tree to find the features that have the greatest impact on image classification.

Future work

For image classification, transfer learning is efficient and powerful. It can complete the training in a short time without the help of GPU. In the next step, I want to combine transfer learning with deep learning, and select the pre trained Imagenet to initialize the model. And some study use DCGAN to generate samples and training in image recognition model, which based by CNN. This method can enhance the classification model and effectively improve the accuracy of image recognition.[8] . By reading this, I want to try more neural networks and to find the most effient model. For decision tree, I want to combine neural network model with decision tree, not only use decision tree to simulate neural network classification, but also use decision tree to evaluate neural network classification. At the same time, I want to use random forest to solve the problem of weak generalization ability of decision tree. For random forest, we should pay attention to the number and maximum depth of decision tree.

References:

1.Dhall, A., Goecke, R., Lucey, S. and Gedeon, T., 2011, November. Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. In 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops) (pp. 2106-2112). IEEE.
2.Zhao, K., He, T., Wu, S. *et al.* Application research of image recognition technology based on CNN in image location of environmental monitoring UAV. *J Image Video Proc.* **2018,** 150 (2018). https://doi.org/10.1186/s13640-018-0391-6
3.Gedeon T. D. and Turner H. S.:Explaining student grades predicted by a neural network. In Proceedings of 1993 International Joint Conference on Neural Networks(1993)
4 Zhu, Q. (2020) "Predicted the authenticity of anger through LSTMs and three-layer neural network and explain result by causal index and characteristic input pattern," 3rd ANU Bio-inspired Computing conference (ABCs 2020), paper 83, 7 pages, Canberra

5. Wonmin B, Thomas M. B, Federico R, Marcus L; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3547-3555

6. .Rokach L., Maimon O. (2005) Decision Trees. In: Maimon O., Rokach L. (eds) Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA. https://doi.org/10.1007/0-387-25465-X_9

7.Kawakami K. Supervised sequence labeling with recurrent neural networks. Ph. D. dissertation, PhD thesis. Ph.D. thesis. (2008).

8.Fang, W, Zhang, F, Sheng, VS & Ding, Y 2018, 'A method for improving CNN-based image recognition using DCGAN', *Computers, Materials and Continua*, vol. 57, no. 1, pp. 167-178. https://doi.org/10.32604/cmc.2018.02356