Comparison on machine learning based algorithms for face expression recognition

Jiayi Bai[1,4,†], Jiawei He[2,†], Fengyuan Liu[3,†]

[1]Legacy Christian Academy, 334 Giotto, United States
[2]Legacy Christian Academy, 2813 Martel Dr. Dayton OH, United State
[3]The Shiyan school Attached To Shijazhuang No.2 Middle School, Shijia Zhuang, Hebei, China

[4]23baijiayi@iusd.org
[†]These authors are equally contributed.

Abstract. With the development of machine learning methods, most of tasks has benefited from it. Numerous works have been done to improve the accuracy and robustness of these methods. In this paper, the authors aim to exploit the effectiveness of machine learning algorithms in face expression recognition (FER) task. Three methods, including support vector machine, random forest and convolutional neural network are introduced to test the performance in FER task. Three methods are tested in the public dataset, FER2013. The conclusion is drawn as follow. All these methods can handle the image recognition, which is a 7-class classification task. Furthermore, nerual network achieves the best performance, which achieves 65% accuracy. The paper exploit the potential of machine learning methods and give a brief attempt on applying the popular algorithms into real-world image recognition tasks.

Key Words: Face expression recognition, support vector machine, neural network, random forest.

1. Introduction

With the emerging development of deep learning based methods [1,2,3,4], many method with novel backbones are proposed, such as AlexNet [5], VGG [6], ResNet [7], DenseNet [8] and inception [9] network. Facial expressions play an important role in human communication [10]. They can express a wide range of emotions, from joy and happiness to disgust and fear. Computers can be trained to recognize these expressions, and the classification of emotions using machine learning is a process of using algorithms to automatically detect and assign labels to instances of data. This can be done using a variety of techniques, but the most common approach is to use supervised learning, in which a training dataset is used to train the machine learning algorithm to identify patterns in the data that correspond to different emotions. Once the algorithm has been trained, it can then be applied to new data in order to classify emotions in that data. This would be a valuable tool for applications such as law enforcement, security, and customer service. Some previous studies have come in to result that Deep learning (DL) based emotion detection has been shown to outperform traditional methods that rely on image processing. This allows for a more accurate understanding of the emotions conveyed in the data, leading to improved

performance. In this research, we will be looking at the various strengths and weaknesses of each method and seeing how well they perform on this particular type of task. By doing this, we came to the same conclusion that neuro networking is the best approach for image classification tasks.

2. Methods

2.1. Support Vector Machine

SVM, stand by Support Vector Machine, are supervised learning models with associated learning algorithms that analyze data for classification. It constructs a hyperplane or set of hyperplanes in a high or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers' detection. The primary capability of SVM is to separate two data and identify a hyperplane between each of them. After completing the learning, the program can be used in the test to match the test data with training data. Finally, the classification result of SVM is obtained.

The advantage of the SVM is:

1. SVM works relatively well when there is a clear margin of separation between classes.
2. SVM is effective in cases where the number of dimensions is greater than the number of samples.
3. SVM is relatively memory efficient.
4. SVM is more effective in high dimensional spaces.

2.2. Random Forest

RF, stand by Random Forest, it is a model that make output predictions by combining outcomes from a sequence of regression decision trees. Each tree is constructed independently and depends on a random vector sampled from the input data, with all the trees in the forest having the same distribution.

Decision trees are the building blocks of a random forest algorithm. It builds by decision nodes, leaf nodes, and a root node, three parts, so when a dataset is put in the random forest model, it will create a lot of random subsets and a spontaneous decision tree first, then select features from the data becoming the blocks, after lots of training and repeating the previous steps, the random forest is well ready for testing.

The advantage of the RF is:

1. It can perform both regression and classification tasks.
2. It can handle large datasets efficiently.
3. The random forest algorithm provides a higher level of accuracy in predicting outcomes over the decision tree algorithm.
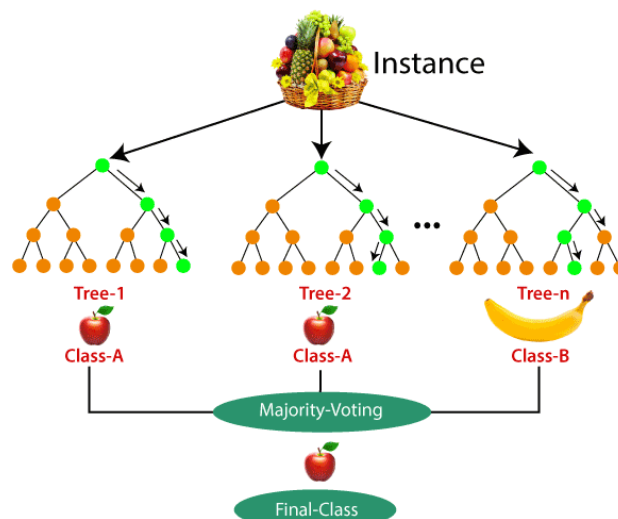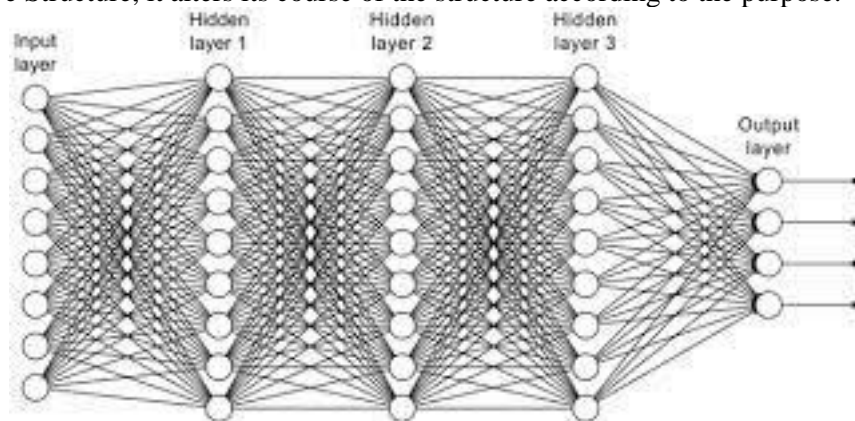

Figure 1. Framework of random forest.

2.3. Neural Network

NN, stand by Neural Network reflect the behavior of the human brain, allowing computer programs to recognize patterns and solve common problems. Neural networks are comprised of a node layers, containing an input layer, one or more hidden layers, and an output layer. Each node, or artificial neuron, connects to another and has an associated weight and threshold. There are multiple layers in Hidden Layer, including Fully Connected Layer, Convolutional Connected Layer, Max Pooling Layer, Dropout Layer, etc. These layers can be combined to build a complex network, eventually get to get a kind high success rate.

The advantage of the NN is:

1. NN is very suitable for visual analysis, it has good image classification ability.
2. Adaptive Structure, it alters its course of the structure according to the purpose.

Figure 2. Framework of neural network.

3. Results

The first set of this research is to use FER2013 dataset to make a Face expression recognition program by using different ways like SVM(Support Vector Machine), RF(Random Forest), NN(Neural Network) and make a comparison of the Accuracy. FER2013 humankind face expression dataset is concerned by 35886 photos of face in different expressions saved by csv file. There are 28708 training pictures,3589 Public test pictures and 3589 private test pictures. Every single picture is black and white at the size of 48×48and.There are totally seven kinds of expressions which match to 0anger,1 disgusrt,2fear,3happy,4sad,5surprised,6normal. Figure 4 shows the data samples.

Figure 4. Sample of FER2013.

Table 1. Results of three machine learning methods.

Method	Accuracy
Neural Network	65%
Random forest	38%
SVM	30%

As shown in Table 1, 65% accuracy is pretty good number compare to what SVM model, and random forest got 38%. So, our conclusion is SVM does not have a better ability to extract image features. And CNN can get higher accuracy by getting more dataset and more complex model.

4. Conclusion

In this paper, the authors aim to exploit the potential on how to apply the machine learning methods into real-world image classification task. Face expression recognition (FER) is widely used in intelligent monitoring, human interaction and so on. To this end, the authors is to introduce machine learning methods into FER task. Three algorithms including neural network, support vector machine and random forest are used for comparison. These methods are tested on the public dataset, FER2013, which contains 7 types of facial expression. The experiments show that the neural network is the best method, which obtains 65% accuracy. It is a satisfying result with respective to a 7-class classification task. The paper gives a brief attempt on bring the machine learning method to real image recognition application. In the future, the researchers can apply more efficient modules, such as attention method and transformer to the basic neural network to achieve better performance.

References

[1]   Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.
[2]   LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.
[3]   Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., ... & Iyengar, S. S. (2018). A survey on deep learning: Algorithms, techniques, and applications. ACM Computing Surveys (CSUR), 51(5), 1-36.
[4]   Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. IEEE Access, 7, 53040-53065.
[5]   Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105.
[6]   Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
[7]   He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
[8]   Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).
[9]   Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).
[10] Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., ... & Bengio, Y. (2013). Challenges in representation learning: A report on three machine learning contests. In International conference on neural information processing (pp. 117-124).