Cascade Network with Recurrent Structure to Detect Face Emotion under Limit Input

Kaiyuan Xing[1]

Research School of Computer Science
Australian National University, Acton 2601, ACT, Australia
u6994560@anu.edu.au

Abstract. Cascade network is a classic but widely used method in classification, especially for the smaller datasets and limited computing resource. In this paper, I propose a reformed cascade network to detect whether the people are genuinely angry or only make a pose. As an update version of my previous research, this cascade network I proposed targeting the processed video frame series. Comparing to the cascade network with adding layers, this new model will feature in recurrent structure so that it will meet the need to process time-related features. The new model shows a better performance compare to the previous version. Generally, it recorded a 5% improvement with the second version of the dataset, ending at 85%. Even though it still doesn't meet the benchmark provided by the dataset provider(95% claimed)[1], the proposed network can be trained with single-GPU computer and managed to have a decent result.

Keywords: Cascade Network · Emotion Detection · Cascade Tower · Recurrent Neural Network.

1 Introduction

1.1 Cascade Network

Cascade Correlation algorithm is designed for training artificial neural networks. The structure starts with directly-linked input-output fully-connection units, which is the minimum structure. After training the minimum structure, one single neuron will be inserted into the previous network one by one and connected with all input and hidden neurons. The new structure will be trained again with all previous neuron weights been detached from gradient calculation. Initially, we use Fahlman's "quickprop" algorithm[2] to calculate both output weights and hidden unit weights.

Compared to standard multilayer perceptron architecture, the cascade correlation structure has several advantages. The cascade network with fewer neurons can solve more complex problems[3]. Usually, cascade networks contain fewer weights and recalculated quicker because most of the weights are locked or 'frozen'. Besides, the cascade network will not face the back-propagation precision error problem cause only one layer of weights is changed each time. As a trade-off, the cascade network cannot fully take advantage of the modern parallel calculation processor. Currently, there are several new inspired cascade network, including fixed-size training[4], multilayer group training or chained cascade network[8].

1.2 Emotion Detection

When people communicate with each other face-to-face, the facial expression is an important indicator to show their emotion. However, facial expressions can't always reflect in the genuine feeling, and finding out whether the emotion is genuine or fake is a challenging task. Recent studies[1, 5] suggest the pupillary response may be an essential part to classify real or posed expression, which means that we can learn genuine emotion from processed images focus on pupils rather than the whole picture. It will significantly reduce the data size of the emotion accuracy detection but bring other challenges.

1.3 Recurrent Network

Sequential data prediction is widely recognized as a significant problem in machine learning and artificial intelligence. Comparing with traditional multilayer neural network, it accepts a list of input without a predetermined limit on size. The networks remember the past hidden results and combine the inputs with the past results as the new input. They can retain the result while generating outputs. It will provide a feature that the same input may not result in the same output due to the previous inputs.

1.4 Paper Introduction

My research in this paper proposed several different kinds of cascade networks dealing with emotion veracity detection, which is a subclass of classification. The dataset I use contained pupil information when the participants were watching several videos. The pupil information was captured for each frame, which means it is a sequential data. However, these data are pre-processed data, which means that the linear-calculation(not convolution-calculation, which targets pictures) combined with recurrent structure is suitable for this task. Unlike traditional multilayer linear deep neural network, it contains less parameters and requires less computing resources and storage space.

2 Method

In this section, I will discuss about the traditional cascade network and how I try to modify it due to the uniqueness of the dataset and the complexity of cascade network.

2.1 Cascade Network Structure

Originally, the cascade network is a progressively expanding structure. It only has a direct connection between input and output units. After the initial training of the weights between input and output units, the weights become frozen temporarily. A new cascade unit will be inserted into the inputs. The output will be retrained, producing a new output weight based on extended input. The steps above can be repeated iteratively. The module is shown in Fig. 1(a).

Initially, the input should directly connect to the output. However, because the type of input is the sequential data, I can't feed all the input at the same time. However, feed one piece of data in the sequence each time is too slow. Then the network will receive several pieces(2, 4 or 8) at the same time. To enable the recurrent part, the model will treat the input and its previous result as the new input[9-11]. Due to the lack of dimension of inputs (only six dimensions), which are processed pictures only abstract the information from pupils, not whole face, I will add a linear hidden layer to expand the dimensions directly. Khoo[6] used an identical method to maintain the computability, even though he is trying to decrease the dimensions. The width of this extra layer is 16, 32, or 64.

When it comes to the added unit, the traditional unit is one single neuron. According to Treadgold and Gorden[7], they set up a maximum depth of the cascade nodes by using the tower structure. Inspired by this, I expand the unit to multiple neurons shown in Fig. 1(b). The additional neurons will help the structure to fit the dataset and acquire a higher accuracy faster than adding neurons one by one. The number of neurons is set as a quarter of the extra layer width. For example, if the width of the extra layer is 16, then 4 neurons will be implemented at each time.

In the previous research, I implement two kinds of the non-linear structure to the cascade unit. However, both of these structures are not providing enough gain to accuracy. In this time, I implement a recurrent structure into the model. The result of the past input will be recorded in each layer, and it will combine with the next input as the total input. Fig. 1(c) shows the total structure with both recurrent. The general model with all technology mentioned above will be used in the experiment.

2.2 Dataset

The dataset used in the experiment is the Anger[1] dataset, promoted in 2017 by Chen. Twenty-two observers were asked to watch 20 movie clips. After watching, they should answer several answers, including whether the figure in the movie clip is emotionally angry or posed an angry face. During their watching, their reaction, including their pupil information, is recorded and processed. Each watching data contains left and right pupils dilation information in every second. The length of the data is not all the same. The dataset only provides the data from 20 observers. Besides, some of the information is missing. After preprocessing, I got 391 pieces of data.

2.3 Training Details

Generally, the cascade correlation is using the "quickprop" algorithm as the back-propagation method. However, I use the Adam algorithm to achieve a fluid and faster back-propagation process. The initial weight and bias set randomly, and all the activations between neurons are using ReLU algorithm, except the neurons to the output, which is used as sigmoid activation. The input layer has 2, 4 or 8 units, each for one parameter of left or right pupil dilation information. The output layer has two units, each for the possibility of Genuine or Posed. Because it is a 2-way classification task, the cross-entropy loss is an appropriate way to do the measurement.

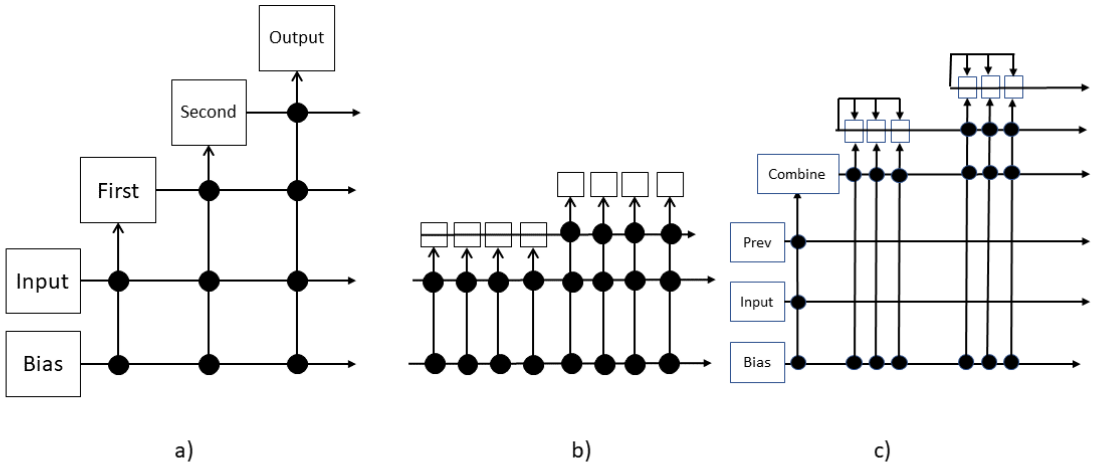

Fig. 1. The general model of the cascade network and several variant for the specific units

All the modules will be initially training for 4000 epochs, with learning rate setting at 2e-4, there will be a 5% decay after every 200 epochs. For each adding unit, the module will be training for 2000 epochs. For most of the cases, I will add 5 units to the module gradually, which make the total training epoch to 14000.

The dataset will be split into the training set and the testing set. One posed and one genuine video, which contains around 40 data will be located in the testing set. All the remaining data will be used as training and validate data. This leave-one-out method is also used in Hossain[5] training, which is similar to this one.

3 Experiment and Result

As a result of previous experiments, I have the conclusion that a hidden layer and multi-length cascade unit will significantly increase the accuracy. In all of these experiments, I was always setting the hidden layer width as four times of the neuron width, which is the best setting in the previous research. However, the size of a frame-group has not decided yet. From the result shown in the table above, a larger frame-size will always generally have higher accuracy. When the hidden layer width is 16, and the frame-group size at 1, the accuracy is 63.5%, 5 percent lower than the previous experiment with the same setting. With the increase of the frame-group size, the accuracy is fluently raising, and have a better result than previous finally when frame-group size is 8, standing at 71.4 percent, which is 3 percent higher than the previous experiment. The parameter of 8-long frame-group have around 83 percent more parameters comparing with the 1-long one.

The model with 32-width hidden layer shows a similar trend with the 16-width one. 1-long frame group(70.5) worse than the previous result(74.5) and 8-long catch up(77.9). However, a 64-width hidden layer model shows something different: the length of the frame group seems not relevant to the result. All of these results are around 80 or 81, even same with the previous result. It could be explained that a hidden layer wide enough has enough parameters to fit the dataset so that the length of the frame group is not relevant anymore. The good news is that training a wider frame group is faster than a narrow one when using GPU due to the parallel computation.

Due to the conclusion from the previous experiment, I only test the multi-width direct cascade unit or its recurrent version. Implementing the recurrent network to the cascade units shows a mixed result. On the one hand, when the frame-width is 4, the recurrent cascade unit models indicate degenerate results, around 1 or 2 percent. On the other hand, the recurrent cascade versions with frame-width at 8 have an impression mark. The accuracy raises about 4 percent compared to the normal version. Especially when the hidden-width length is 64, the accuracy stands at 84.3, a 4 percent higher than the best result in the previous experiment. The reason for the different performance between different frame-width is complicated. It may relate to local optimum and could be improved by better training technique.

Table 1. Accuracy between the different width of hidden layer, neuron and frame-group size

Archeticture			Accuracy	Previous Accuracy
Frame-group size	Hidden Layer Width	Neuron Width		
1	16	4	63.5	68.5
2			66.2	
4			68.6	
8			71.4	
1	32	8	70.5	74.5
4			76.2	
8			77.9	
1	64	16	80.0	80.25
4			80.4	
8			80.8	

Overall, adding recurrent structures improve the performance of the final result. With proper training and parameters setting, the model has a 4 percent compared to the previous experiment. However, it still falls short compared with the benchmark. More traditional way, including kNN, SVM, or multiple-layer neural network are more suitable for this task.[5]

Table 2. Accuracy between the different unit structure

Archeticture			Accuracy
frame-group size	Hidden Layer Width	Cascade Type	
4	16	Direct	68.6
		Recurrent	67.5
8	16	Direct	71.4
		Recurrent	74.8
4	32	Direct	76.2
		Recurrent	75.2
8	32	Direct	77.9
		Recurrent	81.2
4	64	Direct	80.4
		Recurrent	79.2
8	64	Direct	80.8
		Recurrent	84.3
Previous Best Record			80.25

4 Conclusion and Possible Future Work

In summary, my study shows that the cascade network with appropriate reform may have a affordable result generally, however it have no way to match the baseline. The cascade network I designed having bottleneck when try to dealing with the classification task with low-dimension input.

The future work will be mostly focus on better back-propagation methods seeking more efficient way to solve the low-dimension classification problem, including adding the depth of the hidden neuron or other non-linear methods.

References

1. Chen, L., Gedeon, T., Hossain, M. Z., & Caldwell, S. (2017, November). Are you really angry?: detecting emotion veracity as a proposed tool for interaction. In Proceedings of the 29th Australian Conference on Computer-Human Interaction (pp. 412-416). ACM.
2. Fahlman, S. E.; Lebiere, C. The Cascade-Correlation Learning Architecture. In NIPS 2; Touretzky, D. S., Ed.; Morgan-Kaufmann (pp. 524-532)
3. Wilamowski, B.(2010, Jul) Challenges in applications of computational intelligence in industrial electronics, in Proc. IEEE Int. Symp. Ind. Electron. (pp. 15–22).
4. Huang, G., Song, S., & Wu, C., (2012, November). Orthogonal Least Squares Algorithm for Training Cascade Neural Networks, in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 59, no. 11, (pp. 2629-2637), doi: 10.1109/TCSI.2012.2189060.
5. Hossain, Z., & Gedeon, T. (2017, May). Classifying Posed and Real Smiles from Observers' Peripheral Physiology. 11th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '17), EAI
6. Khoo S. & Gedeon T. (2009) Generalisation Performance vs. Architecture Variations in Constructive Cascade Networks. In: Köppen M., Kasabov N., Coghill G. (eds) Advances in Neuro-Information Processing. ICONIP 2008. Lecture Notes in Computer Science, vol 5507. Springer, Berlin, Heidelberg

7. Treadgold, N., & Gedeon, D.(1998)Exploring architecture variations in constructive cascade networks, 1998 IEEE International Joint Conference on Neural Networks Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98CH36227), Anchorage, AK, (pp. 343-348 vol.1), doi: 10.1109/IJCNN.1998.682289.
8. Ouyang, W., Wang K., Zhu X., & Wang X.(2017)Chained Cascade Network for Object Detection, 2017 The IEEE International Conference on Computer Vision (ICCV) (pp. 1938-1946)