From Constructive Cascade Neural Networks to LSTM for Deceive Detection System

Yuehan Cao

School of Computer Science
College of Engineering and Computer Science
Australian National University

Abstract. Detecting deceives through facial thermal images could be complex due to the high dimensionality of multivariate time series data. Therefore, efficiently processing the time series data could be essential in this binary classification problem. This paper applied constructive cascade neural network and long short-term memory recurrent neural network on data with different preprocessing methods. Constructive cascade neural network is applied to achieve a balance between the network complexity and the problem complexity, while LSTM can make efficient use of temporal information in dynamical systems. The result shows that trained LSTM provides better performance than cascade neural network on deceive detection.

Keywords: Cascade neural network; LSTM; Multitasking learning; Generalization

1 Introduction

Facial thermal image processing can be considered as the most ideal method for detecting indicators of physiological emotions and has been widely used in the commercial, government, military, and social life [1]. It has been seen in movies that people will use lie detector based on the change of heart rate to detect deceive which is not that efficient [2]. Thus, I am interested in how reliable thermal image could be on deceive detection.

Although it is still uncertain what the effect of the fight or flight response is on the temperature distribution and superficial blood flow changes on the face [3], many research findings have observed the temperature changes on periorbital, cheek, and carotid during emotional fluctuation [4,5].

We collected our experimental data using a mock crime protocol, in which the participants were divided into two groups: deceptive and truthful [3]. Some of the subjects were asked to deceive in questions relevant to deception. A thermal camera was used to capture images with a frame rate set to 10 fps. Assume simultaneous vasoconstriction and vasodilation in the cutaneous superficial blood vessels of the face for 'fight or flight' responses through facial thermal imaging, we manually selected five ROIs (regions of interest) on the face that was most likely to correlate with the emotional states in which we were interested. Two methods were introduced in this paper to preserve the interrelations in the input time sequence.

1.1 Constructive cascade neural network and Granger causality

Constructive cascade algorithm was designed as a self-built feedforward neural network, as the network architectures are dynamically created during the learning process. Constructive cascade neural network has been proved to perform well on a standard face image database which improves all the frontal pose data and an acceptable cost to only 0.9% drop in performance on multiple different poses [7].

However, cascade neural are memoryless for its current output only depends on the current input. It may lose dynamic information in time sequence. Thus, we introduced extended Granger causality in our recorded thermal time series to assess the blood flow changes, which can preserve relevance interrelations in the input sequence. Multivariate GC has been widely employed to assess directional influence between time series in neuroscientific studies [6]. It is a powerful tool for estimating the causality among time series [8]. 20 features related to 5 ROIs were feed to constructive cascade neural network with each of them is an extracted extended Grange causality value which represents the instantaneous interaction between two effective connected ROIs related to deceptive behavior [1].

1.2 LSTM recurrent neural network

Apart from calculating causality in multivariate time series problems, Sepp Hochreiter and Jürgen Schmidhuber proposed long short-term memory recurrent neural networks in 1997 [9], which can make efficient use of temporal information in the input sequence without vanishing gradient problem. Different from the feedforward neural network, interrelations between the current input and relevant past information are processed to produce the output [10]. With the ability to explicitly modeling time-series data, the temperature recorded in time sequence can be directly fed to LSTM recurrent neural network to determine deceive or truth through dynamic thermal information.

2 Method

2.1 Constructive Cascade Neural Network

2.1.1 Network Topology

Most real-world problems like image processing have extremely large input which makes the slowness of the backpropagation algorithm. Cascade architecture is designed to reduce wasted motion in the training process without the loss of performance.

The constructive cascade neural network used in this paper is modified from typical cascade structures like cascade correlation. CasCor shown in figure 1 has two main ideas. The cascade structure ensures the hidden units are added to the neural network one at a time and frozen since added. To lower the loss, new hidden units were created by maximizing the magnitude of the correlation between the updated output and the residual error signal we are trying to eliminate [11]. Without the knowledge of the size depth and connective pattern of the original network, a small cascade layer is built automatically. Since new hidden units are added with fixed input weights, the cascade algorithm learns fast with only the output connections are trained repeatedly (figure 1).

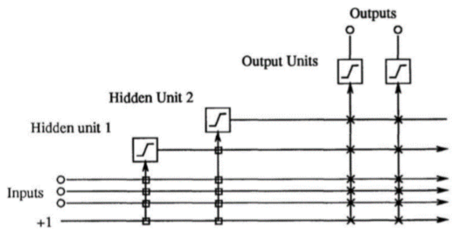


Figure 1: CasCor

Different from adding a single neuron one at a time, our constructive cascade neural network adds cascade layers with a fixed small size in order to reduce the computational complexity and limit the effect of the cascade [4]. One cascade layer is partially connected to the input and hidden layers then fully-connected to the output layer. More cascade layers will fetch information from preceding layers including the former cascade layers.

In this specific problem, each cascade layer is made up of 4 neurons. Each is designed to take its inputs from the input layer in 7 neighborhood neurons and hidden layers situated in 3 adjacent neurons. Neurons in cascade layers also receive inputs from all preceding cascade layers.

2.1.2 Training Methodology

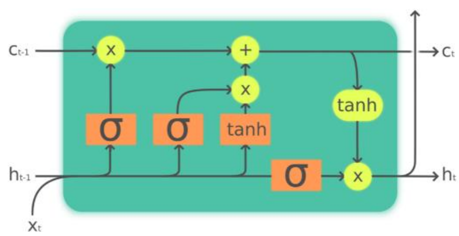
After using eGC to pre-process, our dataset has 31 rows and 21 columns with 20 features and 1 target. Weights and biases of the network were randomly initialized. Sigmoid activation function was applied to the hidden layer. Since the asymptotes of the function are symmetric, it is understandably used while performing binary classification and converge faster than non-symmetric ones. Mini-batch gradient descent, as a mixture of Stochastic gradient descent and batch gradient descent, is implemented for network learning. Calculate the mean gradient of each mini-batch and update the weights in the neural network to realize vectorized implementation as well as the parameter update frequency.

30 epochs are chosen to avoid overfitting. A cascade layer was supposed to add to the network when a certain number of epochs is reached (the training limitation of the current structure) or when the loss is less than a certain small value. However, with 31 data in the experiment, a 20% testing set only contains 6 data which may lead to unstable testing accuracy. Thus, cascade layers were added manually up to two and a comparison of the results will be included in the next section. The cross-entropy loss allows us to find a better local optimum for a comparable environment and with randomly initialized weights in binary classification problems [6].

2.2 Long Short-term Memory Recurrent Neural Network Topology

2.2.1 Network Topology

Feedforward neural network can only be applied to problems whose inputs and targets can be encoded with vectors of fixed dimensionality, and is memoryless. RNN is designed to model sequences of arbitrary length with a feedback connection from its output to its input. The hidden state holds the memory of the network as well as a representation of the current input. However, RNNs are unstable and



the backpropagating gradients can explode or vanish through a long time.

LSTM RNN is an improvement over the general recurrent neural networks, where the nonlinear units in the hidden layer are replaced by memory blocks. A block is made up of memory cells, input, output, and forget gates (figure 2). The input gate determines the portion of the inputs in the new time step goes into the hidden unit. The output date controls the extent to which the value in the cell is used to compute the output of the LSTM unit. The forget gate tells the time of the contents in the memory cell to be forgotten. The input modulation gate modulates the information that goes into the memory cell allowing for faster convergency. The cells are responsible for keeping track of the dependencies follows the input sequence.

2.2.2 Training Methodology

The original data set is composed of 31 tables for the 31 participants. The minimum and maximum temperatures of 5 face regions were extracted for each question. We only focused on a particular question which was most relevant to deceive or truth classification, and specifically include samples that have been tracked over the entire period to create thermal time series feed to LSTM (20s).

We apply min-max normalization on the values of each sample to achieve data uniformity. A feature vector containing 5 minimum and 5 maximum temperatures, which sums up to 10 as the input size to LSTM. 27 subjects are found to valid for this RNN model, which represents the batch size. The optimal number of LSTM cells were found by hyperparameter search over 20, 50, 100 cells. With an initial learning rate of 1e-1, and find to be optimized at 5e-1 in a set of learning rate range from 1e-3 to 1. The number of training epochs kept constant at 1000. Cross-entropy loss was used to this model as well to improve the performance in binary classification problems [6]. LSTM were trained via the SGD optimizer, because it uses the cost gradient of 1 example at each iteration to converge faster.

3 Experimental Results and Discussion

To evaluate the performance of the above two models, we calculated the percentage of correct labelled samples on both training and testing sets. We do k fold cross validation to show the accuracy result is reliable.

3.1 eGC and Cascade Neural Network

Number of layers	No cascade layer				One cascade layer				Two cascade layers			
	1	2	3		1	2	3		1	2	3	
Training Accuracy	86.76	85.66	86.73	86.38	92.08	93.92	93.56	93.43	97.14	96.42	95.70	96.42
Testing Accuracy	57.5	55.83	60.83	58.05	62.5	55.83	67.5	61.94	65.00	57.5	65.83	62.79

Figure 3. Training and Testing Accuracy

Figure 2 shows the average performance on different number of cascade layers. The training accuracy is used to assess how the performance is. The higher the correct percentage on test set the model gets, the better the performance is. Use the basic structure of backpropagation neural network with a hidden layer has a one-on-one connection to the input layer, a network with a total 220 weights has been set up with the generalized result around 65%. Once some weights from input to hidden have been pruned and reduced the total number of weights to 70, the accuracy decreased to 58%. In this case, we are doing cascade to having a better correct ratio without much increment in time complexity (number of weights). Network with one cascade layer has 108 weights with a 4% increase in performance compared to partially connected no cascade network. Two cascade layer has the best result with an overall 63% accuracy on testing data by 142 weights net. As a result of the trade-off between complexity (number of weights) and generalization performance (testing accuracy), it can be observed that with a 40% cut down of the number of weights, constructive cascade neural network cost less than 2 percent drops in average performance.

3.2 LSTM

For many to one structured LSTM, we compared the output of the last hidden cell with the target to assess the performance. The experiment was performed 5 times with different training and testing set to make most of the examples in the dataset are eventually used for both training and testing. The performance of the proposed model on facial thermal dataset is summarized in figure 4. Table a gives the training and testing accuracy after 1000 epoch training. Plot b and c represent the smoothed cross entropy loss of training and testing set against iteration cycle respectively. The training loss was 5.29 at iteration 0 and eventually became 2 at the 1000th epoch, while testing loss decreased to 5.2. It has been shown that the loss decreased slightly faster in the first 500 iterations and getting more and more closer to 0 at the end of the training process. In general, both training and testing loss decreasingly decreased during the training period. Above all, we can assume that the model is not overfitting, although the training accuracy is 100%.

Experiment	1	2	3	Average
Training accuracy (%)	100.0	100.0	100.0	100
Testing accuracy (%)	83.33	78.67	86.67	82.89

Figure 4.a Accuracy

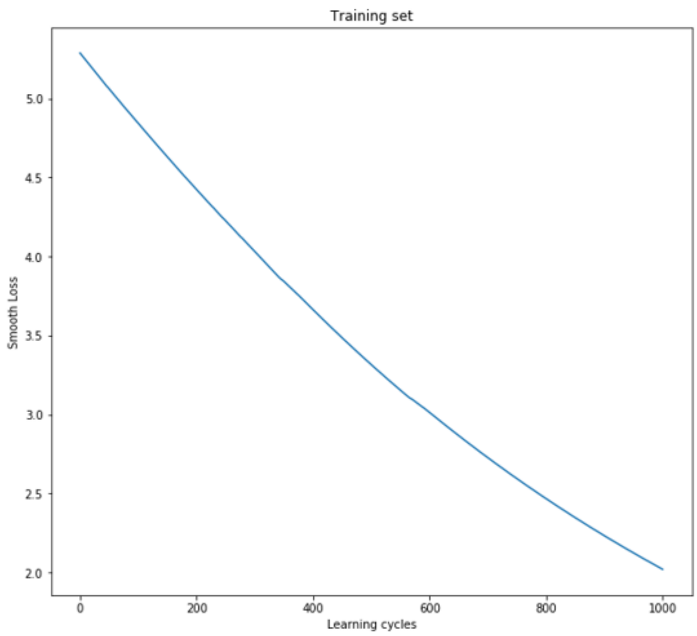


Figure 4.b Training set loss

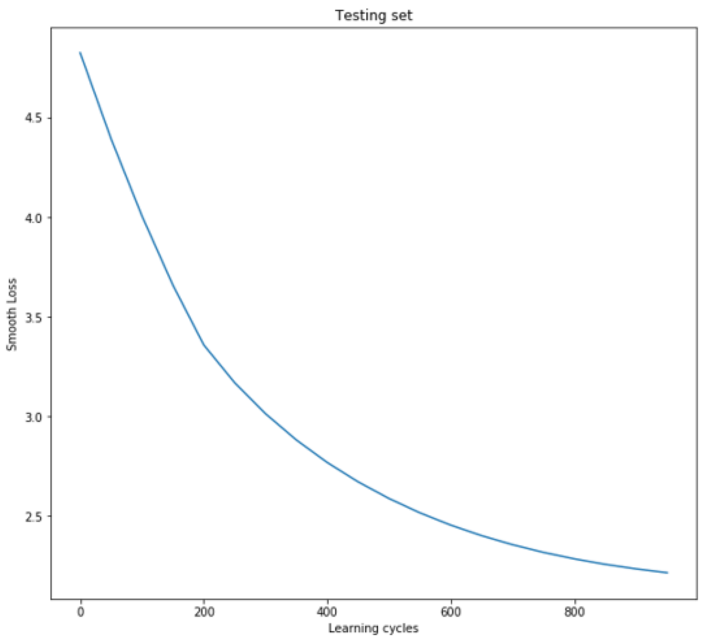


Figure 4.c Testing set loss

Compare the maximum performance on data applied eGC of the cascade neural network classifier with the four classifiers implemented in [3], decision tree, KNN, LDA, SVM have classification accuracy 53.8,58.9,58.9,58.9 respectively for training all features, the cascade technique has a slightly better result as 62.3%. Meanwhile, the LSTM model has achieved a significantly better result around of 83% correctness on the testing set. Thus, LSTM has been the most capable model for doing this multivariate time series classification. A significant drawback of LSTM is that it requires more training time due to the computational complexity of connections between hidden cells.

4 Conclusion and Future Work

This paper has compared two neural network models on multivariate time series classification tasks. To avoid the loss of interrelations in the input sequence, the cascade neural network chooses to pre-process the data with eGC. Five ROIs on the face were selected and the GC indexes between each pair of ROIs were calculated respectively. The LSTM recurrent neural network modified from a typical feedforward neural network preserved the internal dynamics by establishing connections between hidden cells. After training, the LSTM provides a higher classification rate than those techniques using GC indexes. Constructive cascade neural network is comparatively better performed than decision tree, KNN, LDA, and SVM [3]. Techniques like dynamic data mining (feature ranking) [13] should be extended to improve the performance of neural networks.

Due to the generality of the input to LSTM, it can be widely applied to sequence modeling tasks, not only in classification but also prediction, regression like music recognition, and voice detection. We could explore in the future ways to have LSTM deal with higher-level presentations like fully convolutional neural networks, which have been shown to achieve state-of-the-art performance on the tasks of classifying time series sequences [13].

Reference

1. Nguyen H., Kotani K., Chen F., Le B. (2014) A Thermal Facial Emotion Database and Its Analysis. In: Klette R., Rivera M., Satoh S. (eds) Image and Video Technology. PSIVT 2013. Lecture Notes in Computer Science, vol 8333. Springer, Berlin, Heidelberg
2. Heinz WernerGödert, Hans-GeorgRill, GerhardVosse. (2000) Psychophysiological differentiation of deception: the effects of electrodermal lability and mode of responding on skin conductance and heart rate. International Journal of Psychophysiology Volume 40, Issue 1, February 2001, Pages 61-75
3. Derakhshan, A., Mikaeili, M., Nasrabadi, A. M., & Gedeon, T. (2018). Network physiology of "fight or flight" response in facial superficial blood vessels. Physiological measurement, vol. 40, no. 1, p. 014002.
4. Pollina A, Stuart M and Robert G 2015 Hemifacial skin temperature changes related to deception Int. J. Glob. Res. Comput. Sci. 6
5. K. Hong, P. Yuen, T. Chen, A. Tsitiridis, F. Kam, M. Richardson, D. James, W. Oxford, J. Piper, F. Thomas, S. Lightman, ‘Detection and classification of stress using thermal imaging technique,’ Proceedings of the SPIE, 7486, pp. 0I01-0I09, 2009.
6. Gabriele Bellucci, Sergey Chernyak, Morris Hoffman, Gopikrishna Deshpande, Olga Dal Monte, Kristine M. Knutson, Jordan Grafman & Frank Krueger. (2015) Effective connectivity of brain regions underlying third-party punishment: Functional MRI and Granger causality evidence. Journal Social Neuroscience Volume 12, 2017, p. 124-134.
7. 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, HeidelbergScott
8. C. W. J. Granger. (1969) Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica, Vol. 37, No. 3. (Aug., 1969), pp. 424-438.
9. Hochreiter, Sepp & Schmidhuber, Jürgen. (1997). Long Short-term Memory. Neural computation. 9. 1735-80. 10.1162/neco.1997.9.8.1735.
10. Michael Hüsken, Peter Stagge. (2001). Recurrent neural networks for time series classification. Neurocomputing. Volume 50, January 2003, Pages 223-235.
11. E. Fahlman and Christian Lebiere, the Cascade Correlation Learning Architecture. D. S. Touretzky (ed.), Advances in Neural Information Processing Systems 2, pages 524–532, 1990.
12. Tamás D. Gedeon. (1997). Data Mining of Inputs: Analysing Magnitude and Functional Measures. International Journal of Neural SystemsVol. 08, No. 02, pp. 209-218.
13. Fazle Karim, Somshubra Majumdar, Houshang Darabi, (Senior Member, IEEE), Shun Chen. (2017). LSTM Fully Convolutional Networks for Time Series Classification. Digital Object Identifier 10.1109/ACCESS.2017.2779939