AI as Mendacity Indicators: A Comparative Study on Neural Networks/ Deep Learning Models for Thermal Imaging-Based Deceive Classification

Tong Cai

Research School of Computer Science,
Australian National University, Canberra, Australia
{u6493011@anu.edu.au}

Abstract. Neural Network (NN) has been intensively applied on classifying tasks over decades, while its application on the continuously updating data of huge sizes remains a difficulty. Meanwhile, the LSTM (Long Short-Term Memory) RNN (Recurrent Neural Network), a Deep Learning (DL) approach, has demonstrated a decent capability on processing sequential data. Performance comparison of LSTM and the Constructive Cascade Neural Network (ConstrCasc) proposed in [1] to identify the genuineness of narrations is presented in this paper. These models are trained using a dataset that contains the narrator's thermal time series of their minimum /maximum temperature values in five facial ROIs (region of interests), which are captured by a thermal camera (thermal-deceive) [2]. In addition to the original format with raw temperature values, the dataset was further extracted by computing the effective connectivity between these thermal time series, quantified by an extended version of the multivariate Granger causality (eGC). Two results are presented during the investigation. First, the LSTM model reaches an overall accuracy rate of 66.47% on the prediction with around 6% better than the ConstrCasc architectures, performing no worse than the preliminarily results without feature selection published in the thermal-deceive paper [2]. While the networks applying the ConstrCasc technique [1] have a minor degradation on prediction accuracy, scoring around 58% to 60%. On the other hand, there exists a gigantic discrepancy in the training time consumption between these two models, which can be majorly due to the different natures of two formats of the dataset, as well as the drawbacks of LSTM.

Keywords: LSTM, cascade neural network, deception detection, classification, facial thermal imaging, training efficiency, performance improvement

1 Introduction

Spotting deceit relies solely on humans' ability has been proven critically deficient in existing literature since our authenticity judgments can be easily twisted by various factors such as cognitive biases and stereotypes [3], which imposes a significant challenge on crime interrogations. Thus, evaluating the veracity of message contents with state-of-art technology is a study of promising exploration values in many areas. Frequently used measurements involve traditional means like stress measurement – sweating of skin, increased pulse and breathe rates; less invasive behavioural cues like oculomotor patterns and acoustic features [4]; and even physiological reactions including cardiovascular and electrodermal activities in recent years [5].

There is a multitude of approaches to handle the analysis for deception detection data. Among which, Neural Network with no doubt manifests excelling at classification tasks. However, even with an unprecedented development over the past few decades, simple feedforward NN is still confronted with a great challenge when dealing with high-dimensional, continuously evolving data of enormous size [6]. In this case, DL techniques are generally considered to be effective solutions for extracting relevant features from a huge dataset and a long short-term memory RNN can be a good way to manage sequence learning or sequence translation, like speech to text comprehension and real-time language translation. Moreover, it also has correspondingly great application values on tasks like video captioning, image to text, etc.

2 Method

2.1 Neural Network Topology

LSTM

The many to one LSTM architecture applied in this paper consists of two recurrent layers (two LSTMs stacked together with the second layer taking the output of the first one as its input) and followed by fully-connected linear layer as shown in Fig. 1. The hidden state within each LSTM layer captures 50 features of its input with a drop-out probability of 0.2 to regularise the network, therefore, prevent over-fitting. Two output neurons in the final output layer accommodate two labels (deceptive or truthful) respectively. The most probable label is computed through log-softmax as the prediction, which will punish bigger mistakes in likelihood space more.

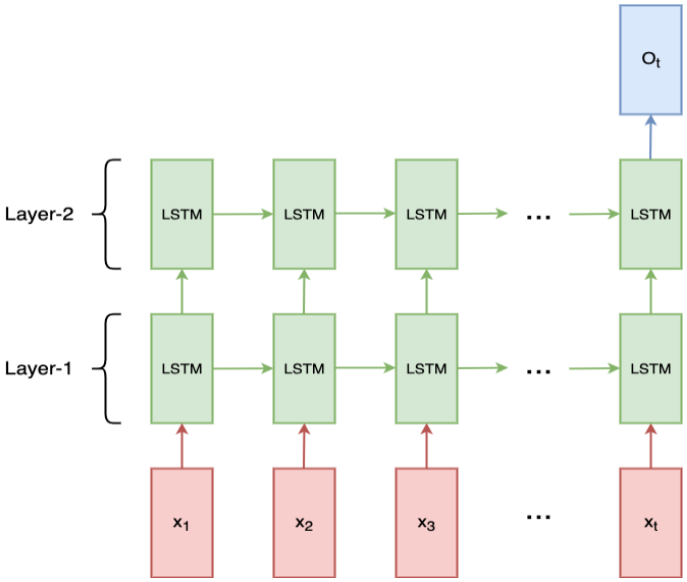

Fig. 1 Two-layer-stacked LSTM taking in input vector **x** (sequence_length, batch, input_features).

ConstrCasc (introduced in Assignment 1)

The source of information about this technique comes from a previous paper investigating the cascade neural network's generalization performance and architecture variations on face recognition tasks [1]. With this technique, we intend to reduce the complexity of the network while retaining the performance. Some modifications are made when applying the proposed network structure to the thermal-deceive dataset of the eGC format [2]. Instead of constructing a convolutional neural network (CNN), a simple neural network (NN) conforms better. The main superiority of CNN is that it captures significant features (e.g. edges) in an image without human supervision. However, the eGC format of the thermal-deceive dataset is made of 20 features for each subject but not image vectors [2], thus all two-dimensional layers mentioned in the ConstrCasc paper [1] are transformed to one-dimensional ones in this case.

The initial network contains one hidden layer to guarantee some computations are performed in the first place. Instead of a 16 by 16 hidden layer with a total of 256 hidden neurons as suggested in ConstrCasc paper [1], there are only 4 hidden neurons in our hidden layer, along with 20 neurons in the input layer decided by the size of the input features and 2 output neurons representing two indicating labels respectively. Preserving the characteristics of constructive cascade algorithm, the hidden neurons are partially connected to the input neurons and fully connected to the output neurons.

For each cascade we add, 2 neurons are present to receive inputs from both input and hidden layers, then transfer information to the output layer. The smaller number of cascade neurons than hidden neurons restricts the functionality of cascade layers and at the meantime trims the total number of hidden neurons. Latter cascade layer has a one-to-one connection with all its preceding cascade layers. Since the dataset is rather simple, two cascade layers at most (in addition to the existing hidden layer) will be more than enough to learn its pattern. Note that meanwhile all cascade layers are partially connected to both hidden and input layers, which to a great extent reduces the complexity of the network when we have large inputs and a consequently large number of weights. Adjacent hidden neurons extract information from two five-neuron neighbourhood on the input layer; each cascade neuron takes 10-neuron neighbourhood on the input layer and 2-neuron neighbourhood on the hidden layer.

Three different variants of the same architecture discussed above are presented in this report: network with a hidden layer but no added cascade layer, one cascade layer added and eventually two added.

2.2 Dataset

The thermal-deceive dataset was originally proposed by Derakhshan, Mikaeili, Nasrabadi and Gedeon [2] containing the minimal and maximal temperature values of five facial ROIs (i.e. periorbital, forehead, perinasal, cheek and chin), collected from 31 subjects under a mock crime scenario. Each participant was directed either to perform a 'criminal' act and meanwhile to conceal his 'crime' during the questioning or not. Their psycho-physiological responses were monitored by a thermal camera at 10 Hz for up to 20 seconds as they answered each question.

For the time series format of the dataset, the following pre-processing measurements are taken before inputting the data into LSTM:

1) 27 of these participants have a complete 20-second recording while the rest 4 were recorded for shorter periods. To unify the size of inputs, the information of the 4 participants were removed.

2) MinMaxScaler has been adopted as a measure to normalise the range of our statistics. Even though the skin temperatures of the human body are stable in a relatively small interval overall (between 33.5 and 36.9 C.), protruding parts generally have lower temperatures [7]. Because of this, nuances could result in a wide divergence on the results and the data of different ROIs should be normalised separately. MinMaxScaler beats many other normalization techniques because it preserves the shape of the original distribution without changing the information embedded in the original data. For this classification task, we intend to preserve as many outliers as possible to detect any abnormal fluctuations on facial temperatures [8].

Due to the limited size of the dataset (27 samples only after removing incomplete information), k-fold cross-validation is carried out. The 27 participants are split into 5 groups so that 20% of them are selected as the test set for each round. 5-split is chosen instead of conventional 10-split out of stabilisation consideration.

The eGC format is a refinement based on the time series version, whose principal is the causal interactions between brain regions [2]. As previous literature suggested, temperature redistribution on the facial cutaneous vasculature caused by a redirection of blood flows will be invoked because of the trigger of stress responses, especially for periorbital areas [9]. We can thus assess the blood flow changes as an indicator of the physiological responses. Ranking and feature selection are not implemented due to the limited size of our dataset. Also, the current format of the dataset has been properly normalised so no more pre-processing is needed.

For each variant of the ConstrCasc networks, these 31 participants are split into 10 groups and one of them is selected as the test set for each round.

2.3 Training Methodology

LSTM

For each k-fold split, 1000 epochs are run (in the experiment, the prediction accuracy of 500 epochs is more volatile). Tests are carried out after every 50 epochs of training to keep track of the progress. The learning algorithm applied is Stochastic Gradient Descent (SGD) for all learning process, with the loss computed by negative log-likelihood loss (NLLloss) every epoch. Log-SoftMax as an activation function is used with NLLloss as this combination encourages high confidence on correct predictions and imposes higher loss when the confidence on the correct class is low. Different learning rates were tested and it is eventually settled to e^{-1}.

ConstrCasc

The weight-sharing strategy, as proposed in [1], is not applied in this dataset as the absence of symmetrical features. The activation function applied is the hyperbolic tangent function, with Resilient propagation (RPROP) as learning algorithm during all learning process following the reference paper [1]. The RPROP beats conventional Backpropagation in two aspects: faster speed of convergence and adaptive step size of each weight dynamically [10].

Starting from one hidden layer but no cascade layer between input and output, 10-fold cross-validation is conducted with each split repeating for 200 epochs at most, and the loss is computed by cross-entropy loss every epoch. The training process terminates in advance and a new layer is added if the loss gets lower than 0.05 or the accuracy rate has remained the same for 10 rounds (the accuracy rate has reached a plateau). This prevents dissipation of time and computational resources since our model does not usually take long to fit a simple dataset. Compared to the standard mean squared error (MSE) used in ConstrCasc paper [1], the cross-entropy loss is more suitable to this dataset as we are outputting the probability of labels and the data are pre-normalised prior to the training. The above training procedure is repeated for each variant.

3 Results and Discussion

3.1 Results

We evaluated the dataset discussed in Section 2.2 using the LSTM model proposed in Section 2.1. The problem to be resolved is to classify the credibility of words based on the temporal variation of temperatures of a person's facial regions. The overall variations of the testing accuracies of five different split sets are given in Fig. 2. The experiment is manually run for 10 rounds. Table 1 contains the average performance of these petitions, along with the results of ConstrCasc networks with increasingly adding numbers of cascade layers for comparison purpose.

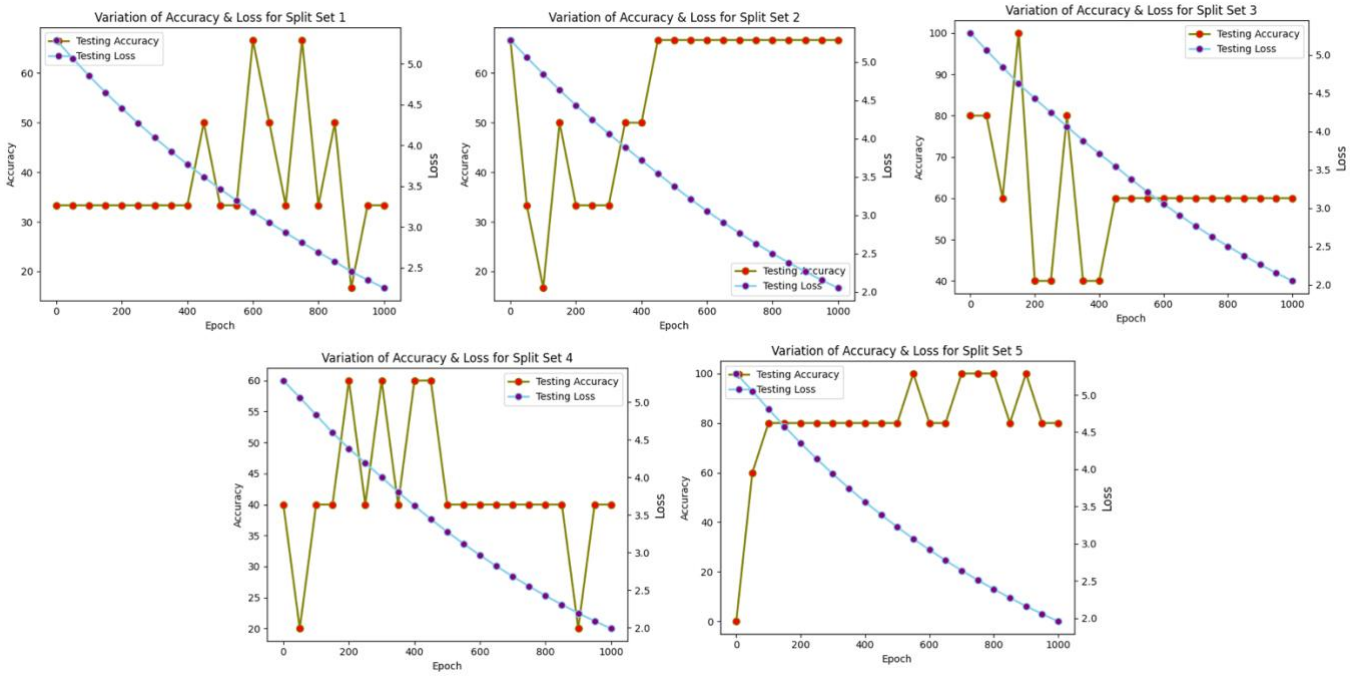

Fig. 2 The overall variation of testing accuracies and loss using 5 different split sets on LSTM

Table 1. The average performance of each network over 10 runs

Architecture	LSTM	No cascade layers	1 cascade layer	2 cascade layers
Training & Testing time s	736.62	2.08	4.27	6.90
Testing accuracy %	66.47	58.83	59.58	60
Standard deviation of testing accuracy %	6.25	6.96	5.13	5.13

3.2 Discussion

Drastic rise and fall of the testing accuracies can be observed from the graphs shown in Fig. 2. This is a consequence of small test sets. They often lead to unstable testing results – one misclassified sample introduces a drastic drop in the accuracy rate.

Based on the average performance of all four networks as exhibited in Table 1, a satisfying improvement by approximately 6% is presented on LSTM compared to the ConstrCasc models. This can be partially caused by a loss of information when using the pairwise time series of temperature values to estimate the effective connectivity with eGC. Some interactions may not be sensitive to identify and thus aren't fully preserved during the estimation. Besides, the limited size of the eGC format of the dataset restricts the generalization of the ConstrCasc models, as well as the functionality of adding cascade layers – they do not preserve an acceptable accuracy on the premise of simplifying a fully-connected network structure.

However, this moderate improvement is accompanied by a sacrifice of the training /testing efficiency - the ConstrCasc networks are far less time-consuming than LSTM is. One of the reasons why is that the dataset used for ConstrCasc (eGC format) is more concentrated and have a tiny size by contrast, while the time series version is of higher dimensionality and thus require more space during training.

In terms of the overall prediction stability, all four models are at the equivalently tolerable level with only about 1% difference. Comparing to the mean accuracy rate of 67.7 in the thermal-deceive paper when taking all features into account [2], we can conclude that we obtain approximately good performance with LSTM (66.47% on average).

4 Conclusion and Future Work

The main objective of this study is to compare the effectiveness of LSTM and a constructive cascade NN in determining the authenticity of words by analysing the information perceived by a thermal camera. Temperature value sequences of 5 ROIs are measured and used to compute the eGC indexes between each pair of them, in which case the GC indexes epitomise the vasoconstriction and vasodilation on the facial cutaneous vasculature – a primary cause of the temperature variations. Two different formats are each the input to an investigating model. According to the test result, LSTM outperforms the ConstrCasc NN by around 6% while at the expense of spending a hundred times more on training and testing. A significant part of the reasons resulting in this is the different constitution of the dataset. Overall, in this resource-constrained environment, even with an acceptable prediction accuracy, training with LSTM may not be the most ideal choice to tackle tasks with time series data due to its enormous time consumption, memory occupancy, high chance of overfitting, etc.

A future study can focus on exploring the applications of more powerful models for sequential data like Temporal convolutional network (TCN), ResNet, etc., given the current growing demand for sequence modelling. Apart from that, methods of extracting relevant information from time series data to reduce the dimensionality are as well of remarkable valuation to delve into.

References

1. Khoo, S., & Gedeon, T. (2008, November). Generalisation Performance vs. Architecture Variations in Constructive Cascade Networks. In International Conference on Neural Information Processing (pp. 236-243). Springer, Berlin, Heidelberg.
2. 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.
3. Thomas H. Feeley & Melissa J. Young (1998) Humans as lie detectors: Some more second thoughts, Communication Quarterly, 46:2, 109-126, DOI: 10.1080/01463379809370090.
4. Gonzalez Billandon, J., Aroyo, A., Pasquali, D., Tonelli, A., Gori, M., Sciutti, A., ... & Rea, F. (2019). Can a robot catch you lying? A machine learning system to detect lies during interactions. Frontiers in Robotics and AI, 6, 64.
5. Pollina, D. A., Dollins, A. B., Senter, S. M., Brown, T. E., Pavlidis, I., Levine, J. A., & Ryan, A. H. (2006). Facial skin surface temperature changes during a "concealed information" test. Annals of Biomedical Engineering, 34(7), 1182-1189.
6. Sadouk, L. (2018). CNN Approaches for Time Series Classification. In Time Series Analysis-Data, Methods, and Applications. IntechOpen.
7. Bierman, W. (1936). The temperature of the skin surface. Journal of the American Medical Association, 106(14), 1158-1162.
8. Hale, J. (2019). Scale, Standardize, or Normalize with Scikit-Learn. Towardsdatascience. Retrieved from https://towardsdatascience.com/scale-standardize-or-normalize-with-scikit-learn-6ccc7d176a02.
9. Derakhshan, A., Mikaeili, M., Khalilzadeh, M. A., & Mohammadian, A. (2014, July). Preliminary study on facial thermal imaging for stress recognition. In Intelligent Environments (Workshops) (pp. 66-73).
10. Riedmiller, M., & Braun, H. (1993, March). A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In IEEE international conference on neural networks (pp. 586-591). IEEE.