International Journal of Engineering Trends and Applications (IJETA) – Volume 8 Issue 5, Sep-Oct 2021

RESEARCH ARTICLE OPEN ACCESS

Comparison on Time Delay Neural Network and Long Short Term Memory networks

ANUPAMA USHA
Bhavans Vidya Mandir, Elamakkara, Kochi
Kerala

ABSTRACT
This paper makes an effort to compare the two popular AI neural network architectures namely Time delay neural network(TDNN) and Long Short Term Memory (LSTM) networks.A TDNN is like a feedforward network because time aspect is only inserted through its inputs.It just maps the past and present values.LSTM is based on RNN(Recurrent neural network).LSTM overcomes the problems in RNN as it can learn long term dependencies and are suitable for long term data.This paper attempts to systematically compare the pros and cons of both the algorithms by analysing the findings in other articles on this topic.
Keywords —RNN,dropout dependency learning,,neural network.

I. INTRODUCTION

Artificial neural network(ANN) is a technology based on the study of brain and neural system in animals. It simulates the electrical activity of the brain and nervous system. Various nodes arranged in multiple layers communicate based on weighted data signals .

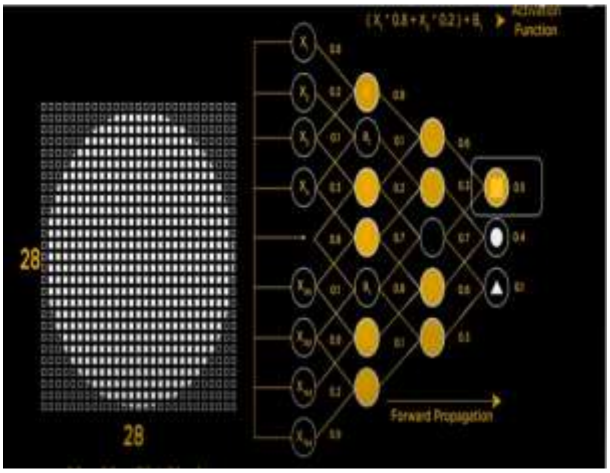

Figure 1: Working of a neural network

Figure 1 is an example neural network wherein a system can correctly recognize mathematical shapes using the concept of forward propagation and back propagation. ANNs consist of a layer of input nodes and a layer of output nodes, connected by one or more layers of hidden nodes. Input layer nodes pass on information t hidden nodes based on activation functions. ANN performance depends on the degree to which the system has been trained with data.

TDNN is a simple way to map past and present values. The delays remain constant throughout the training procedure and are calculated before itself using trial and error method. On the other hand, the memory feature of RNN can capture this feature by learning the time dependencies. LSTM can learn long term dependencies and are suited for learning sequential data like rhythm learning, music composition, grammar learning etc.

II. TDNN

Theory

Time delay neural network is a multilayer ANN architecture whose purpose is to 1) classify patterns with shift-invariance, and 2) model context at each layer of the network.
Shift-invariant classification means that the classifier does not require explicit segmentation prior to classification. For the classification of a temporal pattern (such as speech), the TDNN thus avoids having to determine the beginning and end points of sounds before classifying them. For contextual modelling in a TDNN, each neural unit at each layer receives input not only from activations/features at the layer below, but from a pattern of unit output and its context[1]. Time-delay neural networks work on sequential data, e.g., time series, by augmenting the input with time-delayed copies of previous inputs:

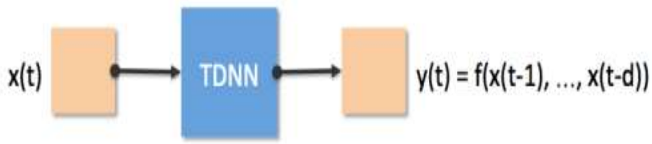

Figure 2: TDNN mathematical representation (David, nd)

Applications

TDNN can recognize phonemes. Each TDNN consists of an input layer, hidden layer and the phone state layer. Back-propagation was applied to train the networks in bootstrapping phase, to fit phoneme targets. Above the two phone state layers, the DTW algorithm is applied to find the optimal path of phone hypotheses for the word models. In the word layer the activation of the phone state units along the optimal paths are accumulated. The highest score of the word units represents the recognized word. In the second learning process the networks are trained to fit word targets. The error derivatives are back-propagated from the word units through the best path in the DTW layer down to the front-end TDNNs, ensuring that the network is optimized for the actual evaluation task, which is word and not phoneme recognition [2].

RNNs can be used for speech enhancement. Spoken sounds are not of uniform length and precise segmentation is difficult. By scanning a sound over past and future, the TDNN is able to construct a model for the key elements of that sound in a time-shift invariant manner (It means that when the input shifts the output also shifts but stays otherwise unchanged.).

It can be used for handwriting recognition where shift invariance was used. Two-dimensional TDNNs were later applied to other image-recognition tasks under the name of "Convolutional Neural Networks", where shift-invariant training is applied to the x/y axes of an image[1].

III. LSTM

Theory

It is a type of RNN capable of learning order dependence on domains like machine translation, speech recognition etc. It is a complex area of deep learning. standard RNNs fail to learn in the presence of time lags greater than 5 – 10 discrete time steps between relevant input events and target signals. The vanishing error problem casts doubt on whether standard RNNs can indeed exhibit significant practical advantages over time window-based feedforward networks. A recent model, “Long Short-Term Memory” (LSTM), is not affected by this problem. LSTM can learn to bridge minimal time lags in excess of 1000 discrete time steps by enforcing constant error flow through “constant error carrousels” (CECs) within special units, called cells [5]. An LSTM layer consists of a set of recurrently connected blocks, known as memory blocks. These blocks can be thought of as a differentiable version of the memory chips in a digital computer. Each one contains one or more recurrently connected memory cells and three multiplicative units – the input, output and forget gates – that provide continuous analogues of write, read and reset operations for the cells. The net can only interact with the cells via the gates [5].

LSTM can capture long-range dependencies. It can have memory about previous inputs for extended time durations. There are 3 gates in an LSTM cell. Memory manipulations in LSTM are done using these gates. Long short-term memory (LSTM) utilizes gates to control the gradient propagation in the recurrent network’s memory.

Forget Gate: Forget gate removes the information that is no longer useful in the cell state

Input Gate: Additional useful information to the cell state is added by input gate

Output Gate: Additional useful information to the cell state is added by output gate

This gating mechanism of LSTM has allowed the network to learn the conditions for when to forget, ignore, or keep information in the memory cell.[14]

Applications

Recognizing digits in an optimal way is a challenging problem. Recent deep learning based approaches have achieved great success on handwriting recognition. English characters are among the most widely adopted writing systems in the world [7].

It can be used for sign language processing though it’s difficult to extract text from visual signs. Many improved sign language translations models have been proposed.

All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

IV. COMPARISON

	Momentum	Default decay rate	Accuracy	Optimum Dropout value	Dependency learning
LSTM	0.5 - 0.9.	0.97	Best	20%	Can learn long term dependencies
TDNN	0.5(default)	Not applicable	Average	50%	Extremely low

Momentum- Momentum is a unique hyper parameter which allows the accumulation of the gradients of the past steps to determine the direction to go with, instead of using the gradient of only the current step to guide the search.

Decay rate-The rate at which weights decay to zero exponentially, if no other weight update is scheduled.

Dropout- Such a layer helps avoid overfitting in training by bypassing randomly selected neurons, thereby reducing the sensitivity to specific weights of the individual neurons. While dropout layers can be used with input layers, they shouldn’t be used with output layers as that may mess up the output from the model and the calculation of error. While adding more complexity may risk overfitting (by increasing nodes in dense layers or adding more number of dense layers and have poor validation accuracy), this can be addressed by adding dropout.

V. CONCLUSIONS

When we move to LSTM, we are introducing more controlling factors and thus bring in more flexibility in controlling the outputs. LSTM has proved that it can be very useful in time series forecasting problems. A possible limitation is that it requires large memory bandwidth to be computed. The properties of TDNN makes it suitable for limited applications like phoneme recognition etc.

ACKNOWLEDGMENT

This comparison was brought out as a analysis of other detailed research papers which threw insight into the various RNN models and on the importance of learning long term dependencies. The author wishes to thank all the authors of the articles referred.

REFERENCES

[1] Time delay neural network. https://en.wikipedia.org/wiki/Time_delay_neural_network

[2] Irwin King, Jun Wang & Laiwan Chan (2006) Neural Information processing

[3] David Hasenfratz https://dhasenfratz.com/2017/01/09/time-delay-neural-networks/

[4] Felix A. Gers, et al., Learning to Forget: Continual Prediction with LSTM, 2000

[5] Alex Graves, et al., Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures, 2005.

[6] Sarita Yadav, Ankur Pandey, Pulkit Aggarwal, Rachit Garg, Vishal Aggarwal(2018) Handwriting Recognition using LSTM Networks International Journal of New Technology and Research (IJNTR), ISSN:2454-4116, Volume-4, Issue-3, Pages .116-119 https://www.ijntr.org/download_data/IJNTR04030056.pdf

[7] J. Martens and I. Sutskever, "Training deep and recurrent neural networks with Hessian-Free optimization." Neural Networks: Tricks of the Trade, Springer, 2012, pp. pp 479-535.

[8] F. A. Gers, J. Schmidhuber, and F. Cummins. "Learning to forget: Continual prediction with LSTM." in Proceedings of the 9th International Conference on Artificial Neural Networks, IEEE, 1999, vol. 2, pp. 850- 855.

[9] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber. "Learning precise timing with LSTM recurrent networks," Journal of Machine Learning Research, vol. 3, pp. 115-143, 2003.

[10] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling." ISCA, pp. 338-342, 2014.

[11] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition," ArXiv e-prints, 2014.

[12] Chung, C. Gulcehre, K. Cho, and Y. Bengio. "Empirical evaluation of gated recurrent neural networks on sequence modeling," arXiv:1412.3555, 2014.

[13] Santhoopa Jayawardhana, https://towardsdatascience.com/sequence-models-and-recurrent-neural-networks-rnns-62cadeb4f1e1,2020