OPEN ACCESS

Middle East Research Journal of Engineering and Technology
ISSN: **2789-7737** (Print) **2958-2059** (Online)
Frequency: Bi-Monthly
DOI: 10.36348/merjet.2021.v01i01.009

KUWAIT SCHOLARS PUBLISHER
Website: http://www.kspublisher.com/
Email: office.kspublisher@gmail.com

Analysis on Artificial Recurrent Neural Network for Sequence Classification of Audio Data

Er. Sukhmanpreet Singh[*1], **Er. Nardeep Singh**[1], **Er. Sandeep Kaur**[1], **Er. Mandeep Kaur**[1]
[1]Assistant Professors, CSE Dept. BGIET Sangrur, Punjab

ABSTRACT: Recurrent neural networks (RNNs) are capable of learning features and long-term dependencies from sequential and time-series data. The RNNs have a stack of non-linear units where at least one connection between units forms a directed cycle. A well-trained RNN can model any dynamical system; however, training RNNs is mostly plagued by issues in learning long-term dependencies. In this paper, we present a survey on RNNs and several new advances for newcomers and professionals in the field. In this paper literature survey on Artificial Recurrent Neural Network has been presented.

Keywords: Recurrent neural networks (RNN), Bacterial Foraging Optimisation (BFO).

REVIEW PAPER

***Corresponding Author:**
Er. Sukhmanpreet Singh
Assistant Professors, CSE Dept.
BGIET Sangrur, Punjab.

How to cite this paper:
Sukhmanpreet Singh *et al.;* "Analysis on Artificial Recurrent Neural Network for Sequence Classification of Audio Data". Middle East Res J. Eng. Technol, 2021 Nov-Dec 1(1): 58-60.

Article History:
| Submit: 09.11.2021 |
| Accepted: 24.12.2021 |
| Published: 30.12.2021 |

1. INTRODUCTION

Neural networks are a data driven technique, and in recent years have become popular, in part due to an increase in available data, as well as optimisations in computer hardware. These networks have been used to solve a plethora of problems, such as image classification, text generation and object detection. The fundamental building block of a deep network is a single perceptron, which is explained in the next section.

2. Simple Neural Networks

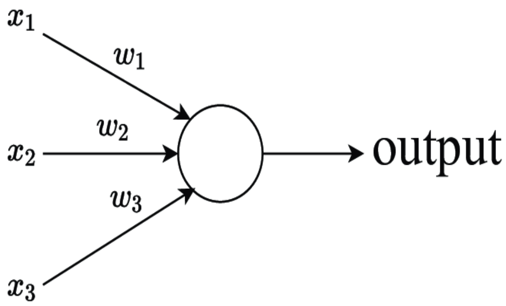

Figure 1 A single perceptron, with three inputs.

A neural network consists of several artificial neurons (aka perceptrons) connected as nodes in a graph. Each of these nodes will have an output activation a according to the function:

$$\alpha = f\left(\sum_{i=1}^{N} W_i x_i + b\right) \qquad (1)$$

Where N is the number of inputs, W is a vector of weights, and b of biases, which are learned parameters of the model.

3. RECURRENT NEURAL NETWORK

Standard neural networks will have one input, and one output. Instead, a recurrent neural network can have multiple inputs, and multiple outputs by re-running the network on the new data, while maintaining some internal data between the two network runs. The weights of the neurons remain the same, but the activations differ because the input data differs, and there is state information that is retained.

This can be thought of as having a 'flip-flop' inside a neural network these networks can store and act on data that happened an arbitrary length of time ago. In this way, a single RNN layer has two inputs and two outputs, with one set in data and the other in 'time'.The use of the word 'time' and 'time-steps' is loose, in this context it refers to a single recurrence of the network, and describes traversing any sequence: be those words in a sentence, instructions in a program, or frames in a video, for example. An ‘unrolling’ of an RNN in time can be seen in Figure 1.3. Here the network has one input xt and one output ht per time-step. The weights of the recurrent network A remain the same across time steps, but the activations differ as the internal memory changes over time, and the input differs over time.

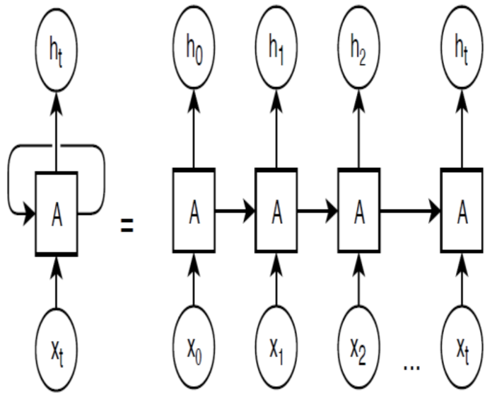

Figure 2: A basic recurrent neural network

There are many layouts an RNN can have, which can be seen in Figure 1.4. The first one is the most basic, and it represents a typical DNN, with one input, and one output. The next type depicted is a one-to-many network, where the output is a sequence, and the input is a single data sample. An example of this is an image caption generator where a single image is served as input to the network, and the network produces a sentence — a sequence of words. This sentence has no fixed length, the network runs until the ‘end of line’ character is produced by the network.

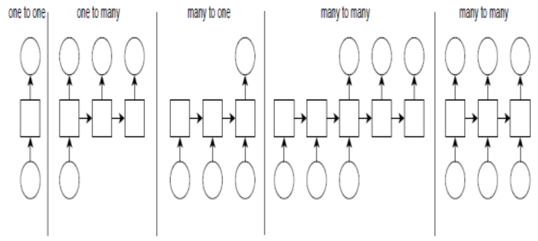

Figure 3: Several different layouts of RNNs

The third network shown is a many-to-one configuration. The next two types of RNN are both many-to-many. The first is where a whole sequence is input before a sequence is output, which is the format of the language translator presented in [1]. The final example is where there is an output for every input during the sequence, an example of which is video classification that may change as the video progresses.

4. LITERATURE REVIEW

Mishra et al. (2012) have proposed a closely related end-to-end method based on HOG features and a SVM classifier. Methods based in sliding window yield good text localization results. Their main drawback compared to connected component-based methods is their high computational cost, as sliding window approaches are confronted with a huge search space. Moreover, these methods are limited to the detection of a single language and orientation for which they have been trained on. Connected component-based methods, on the other hand, are based on a typical bottom-up pipeline: first, they apply a segmentation algorithm to extract regions (connected components); then, they classify the resulting regions into character or background; and finally, the identified characters are grouped into longer sequences (i.e. words or text lines).

Said Kassim Katungunya et. al. (2016) presented a method for handwritten digits recognition using multi-layer sigmoid neural network trained by thousands of images of handwritten digits from training set. They showed that the used approach is fast in computation and highly accurate compared to other methods. Based on this method further research can be done on this method to extend the capability of neural network to tackle most complex problems of machine learning.

Kourosh Kiani and Elmira Mohsenzadeh Korayem (2016) shows the use of SNN Model, in order to have robust learning and classification of handwritten digits, i.e., to have a learning process which is persistent against changes and high noise levels. Due to the similarities among handwritten digits, the classifications have been erratic but the Deep Belief Network the first layer, composed of 225 neurons (15*15 pixels for each image), works as the receptor of input images. The middle layer is used for processes, encoding and network learning, while the last layer, which is composed of 10 neurons (as we have 10 distinct classes), does the job of prediction and classification of images. The model was implemented using MATLAB and we have used Hoda Persian handwritten digits dataset as our input images. They obtained results show good accuracy (95%), the learning and classification of images of handwritten digits with high levels of noise.

H. Garud et al. (2014) presented a fast and effective CC scheme specifically suited for automobile video cameras. The proposed CC scheme uses combination of a static method for illuminant estimation called White Patch Retinex (WPR) and a computationally efficient linear transformation model for WB correction of the images. The proposed scheme also introduces novel temporal filtering of the WB parameters to avoid the field flicker noise. Exhaustive testing of the scheme in laboratory and real life test conditions have shown the it to be effective under various lighting conditions. The presented work ashows details of its implementation on an embedded processing platform to achieve HD video processing at the frame rate of 60 frames per second.

C. E. Portugal-Zambrano et al. (2016) focused on the implementation of a computer vision system combining a hardware prototype and a software module. The hardware was developed to capture the images of coffee beans, the software uses a White-Patch algorithm as a image enhancement procedure, color histograms as feature extractor and SVM for the classification task, a database of 1930 images was collected, they used 13 categories of defects described in the SCAA standard of evaluation. Results of classification achieved a 98.8% of overall detection accuracy, therefore the proposed. system proved to be effective in classifying physical defects of green coffee beans.

R. Sethi et al. (2.015) proposed a new image enhancement approach that modifies the gray world algorithm by finding the color cast using fuzzy logic and then removing the color cast by optimizing the correction method using Bacterial Foraging Optimisation (BFO). The proposed approach is adaptive in nature as it finds the intensity of color cast instead of assuming it which improves the quality of underwater images. Computed results have enhanced visual details, contrast and color performance.

Y. Zhang and X. Lu (2018) proposed a LSTM-CTC model by combining CTC training with LSTM model based on the principle of Connectionist Temporal Classification (CTC). By inserting the Softmax vector output from the top of LSTM into the CTC model and using the CTC decoding method, the presented model reduces the loss of the entire sequence and predicts the sequence label correctly in the prediction probability of the LSTM output.

Bhati S. et al (2019) proposed a two-step strategy to use machine learning methods for PD detection. In the first step, they use Long Short-Term Memory (LSTM)-based siamese networks to learn feature representations that highlight the information related to speech articulation and prosody relevant for PD detection. Siamese networks are trained on data pairs employing a Spanish corpus containing 52 patients and 56 control subjects. In the second step, they trained a classifier to make decisions about the presence or absence of PD employing the features provided by the LSTM networks.

C. Li et al. (2020) proposed to use artificial intelligence speech recognition to assist SSN on-site and in real time. The speech recognition algorithm proposed work is based on improved parameter adaptive spectral subtraction (IPASS) and long short-term memory (LSTM) network. The algorithm first uses IPASS to preprocess the speech data of reading out the safety notification collected on the operation site, and then uses LSTM for speech recognition. The presented results show that the proposed speech recognition algorithm has good anti-noise performance and could accurately recognize the on-site speech, and afterward, the speech will be matched with the safety notification to check if it has been read properly.

5. CONCLUSION

In this paper, we systematically review major and recent advancements of RNNs in the literature and introduce the challenging problems in training RNNs. A RNN refers to a network of artificial neurons with recurrent connections among them. The recurrent connections learn the dependencies among input sequential or time-series data. The ability to learn sequential dependencies has allowed RNNs to gain popularity

REFERENCES

- Sutskever, I., Vinyals, O., and Le, Q. V., “Sequence to sequence learning with neural networks”, In Advances in neural information processing systems, pages 3104–3112, 2014.
- Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8):1735–1780.
- Mishra, K. Alahari, C. Jawahar, Top-down and bottom-up cues for scene text recognition, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, pp. 2687–2694, 2012.
- Said Kassim Katungunya , Xuewen Ding and Juma Joram Mashenene, (2016) Automatic Recognition of Handwritten Digits Using Multi-Layer Sigmoid Neural Network, International Journal of Science and Research (IJSR), ISSN (Online): 2319-7064, Paper ID: NOV162037, Volume 5 Issue 3, March 2016.
- Kourosh Kiani and Elmira Mohsenzadeh Korayem, “Classification of Persian handwritten digits using spiking neural networks”, 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI), Electronic ISBN**:** 978-1-4673-6506-2, 2016.
- H. Garud, U. K. Pudipeddi, K. Desappan and S. Nagori, "A fast color constancy scheme for automobile video cameras," 2014 International Conference on Signal Processing and Communications (SPCOM), pp. 1-6, 2014.
- E. Portugal-Zambrano, J. C. Gutiérrez-Cáceres, J. Ramirez-Ticona and C. A. Beltran-Castañón, "Computer vision grading system for physical quality evaluation of green coffee beans, 2016 XLII Latin American Computing Conference (CLEI), pp. 1-11, 2016.