Recurrent Neural Network and its Various Architecture Types

Trupti Katte

Assistant Professor, Computer Engineering Department, Army Institute of Technology, Pune, Maharashtra, India

***Abstract*----Recurrent neural network are network with dynamic capabilities to generate and process temporal information. Recurrent neural network are network can deep learn the input with its various architecture and identify outputs. LSTM network model was the first RNN with greatest achievement in pattern recognition contest in 2014. RNN can be used in its various architecture forms depending on the needs. Here we provide brief summary of various RNN networks up to now with Why? How? When? to use this network. Different architecture showed that recurrent neural network is mostly network with feed -forward and if want to store some information fed-back.**

***Keywords*----Recurrent neural Network, RNN, LSTM, NTM, Neural.**

I. INTRODUCTION

A ***Neuron/ Nerve cell*** is a cell that carries electrical impulses. **Neurons** are most important part in brain and are the basic units of nervous system. ***Neural Network*** is the computer system which modeled based on concept of human brain and nervous systems neuron. ***Artificial Neural Network*** are the computing systems which are based on concept of biological neural networks. It consists of connections and units/Node (input, output, hidden).Nodes represent artificial neuron and connections represent direction of flow of information from one node to other.

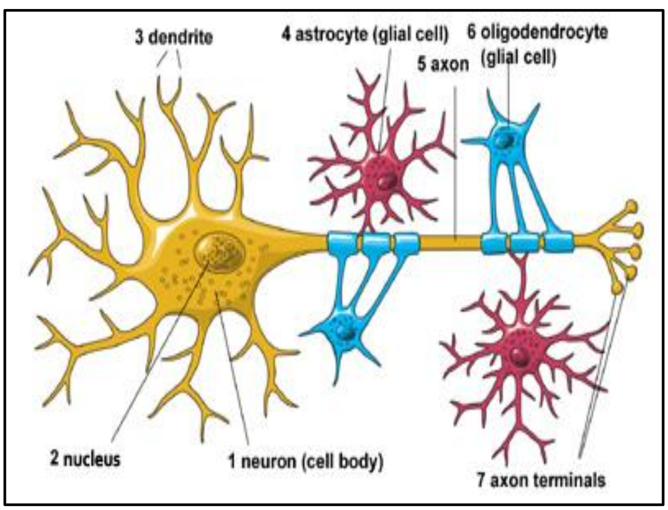


Fig 1.1 Neuron (Biological basic unit of nervous system)

The ***Recurrent Neural Network (RNN)*** is the network with loops, which allows information to persist in network. RNN has feed-back connection to the network itself, which allows activations to flow back in a loop, learn sequences and information to persist. RNN are extremely powerful in modeling sequential data, speech or text and applied on non-sequential data to train in a non-sequential manner. RNN can be used for image, video captioning, word prediction, word translation, image processing, speech recognition, speech processing [1], natural language processing, music processing applications, etc.

II. RECURRENT NEURAL NETWORK ARCHITECTURE TYPES

Different types of RNN architectures up to now are listed below:

1. *Fully Recurrent Neural Network*

Fully recurrent neural network (FRNN) developed in the 1980s, which can learn temporal sequences, either in batch mode or online. FRNN consists of two layers, input and output layer of linear and non-linear units, resp. The units in input layer is fully connected to every units of output layer by adjustable weights. Each unit has a real-valued time-varying activation function. The output units have some knowledge of their prior activations, which feedback the activations to the input layer units. Learning in FRNNs is by mapping input sequences and activations, to another set of output sequences. This continue to feedback to input sequences and finding output sequences over multiple time steps, and over time discover abstract representations.

2. *Recursive Neural network*

Network created in differentiable graph like structure by recursively applying same set of weights to network in topological order. Such networks are also trained by automatic differentiation [2] in reverse mode. It corresponds to linear chain structure and are used in natural language processing, processing distributed representation of structure. Variation of recursive neural network is Recursive Neural Tensor Network which uses tensor-based composition function on every network nodes.

No	RNN Architectures
1	Fully Recurrent Neural Network
2	Recursive Neural Network
3	Hopfield Network
4	Elman Networks And Jordan Networks or Simple Recurrent Network (SRN)
5	Echo State Network
6	Neural History Compressor
7	Long Short-Term Memory (LSTM)
8	Gated Recurrent Unit
9	Bi-Directional Recurrent Neural Network
10	Continuous-Time Recurrent Neural Network (CTRNN)
11	Hierarchical Recurrent Neural Network
12	Recurrent Multilayer Perceptron Network
13	Multiple Timescales Model
14	Neural Turing Machines (NTM)
15	Differentiable Neural Computer (DNC)
16	Neural Network Pushdown Automata (NNPDA)

Fig 2.1 RNN Architecture Types

3. Hopfield Network

Network in which all connections are symmetric, developed by John Hopfield (1982).Hopfield network consists of a set of N interconnected neurons which update their activation values asynchronously and independently of other neurons [3].On applying new input, output is calculated and given as feedback to input layers to adjust input, this process continues till output becomes constant. We can learn network by using Hebbian Learning rule or by using Storkey Learning rule. A variation on the Hopfield network is the Bidirectional Associative Memory (BAM). This NN is used for Content Addressable Memory (CAM).

4. Elman networks and Jordan networks or Simple Recurrent Network (SRN)

An Elman network is 3-layer network, with additional context units. It consists of input layer, output layer and middle-hidden layer which is connected with context unit with weight w=1 [4]. At each time input unit applied with learning unit and fed-forward. Context units as back-connections to network coming from hidden layer, saves previous values of hidden units. Jordan Network is nearly same as that of Elman network, only difference is that instead of fed-back context unit from hidden layer; here context units (state layer) are fed-back from Output layer.

5. Echo State Network

Echo state network is network with rare (typically 1%) connectivity with hidden layer [5]. The hidden neurons having fixed connectivity and its weights are randomly assigned. output neurons weights can be learned so that the network can (re)produce specific time-based patterns. The weights which are modified during training are connected from hidden-to-output neurons. The efficient C++ library for different Echo state network: a reservoir, and a similar concept is: Liquid state machine.

6. Neural history compressor

It is an unsupervised stack of RNN [6]. Input to next level is predicted based on learning previous layer input. If inputs are not predicted then they are given as input to next higher level with adding more hidden units. So each higher level consists of inputs which are unpredicted and compressed information representation of RNN. Network can be considered as network with two levels: higher level (conscious) chunker, lower level (subconscious) automatizer. It can partially solve vanishing gradient problem (1992) of automatic differentiation or back propagation in neural Networks.

7. Long short-term memory (LSTM)

LSTM is a system which can learn task by using deep learning and avoids vanishing gradient problem [7]. LSTM is normally improved by recurrent gates called "forget" gates and to learn tasks it requires memory of event happened in history. LSTM can be learned by Connectionist Temporal Classification (CTC) which achieves both alignment and recognition for weights.

8. Gated Recurrent Units

Gating mechanism in RNN introduced by Kyunghyun Cho [8] (2014). This mechanism lacks an output gate and has fewer parameters than LSTM. Its performance is similar to that of LSTM on polyphonic music and speech signal modeling.

9. Bi-directional RNN

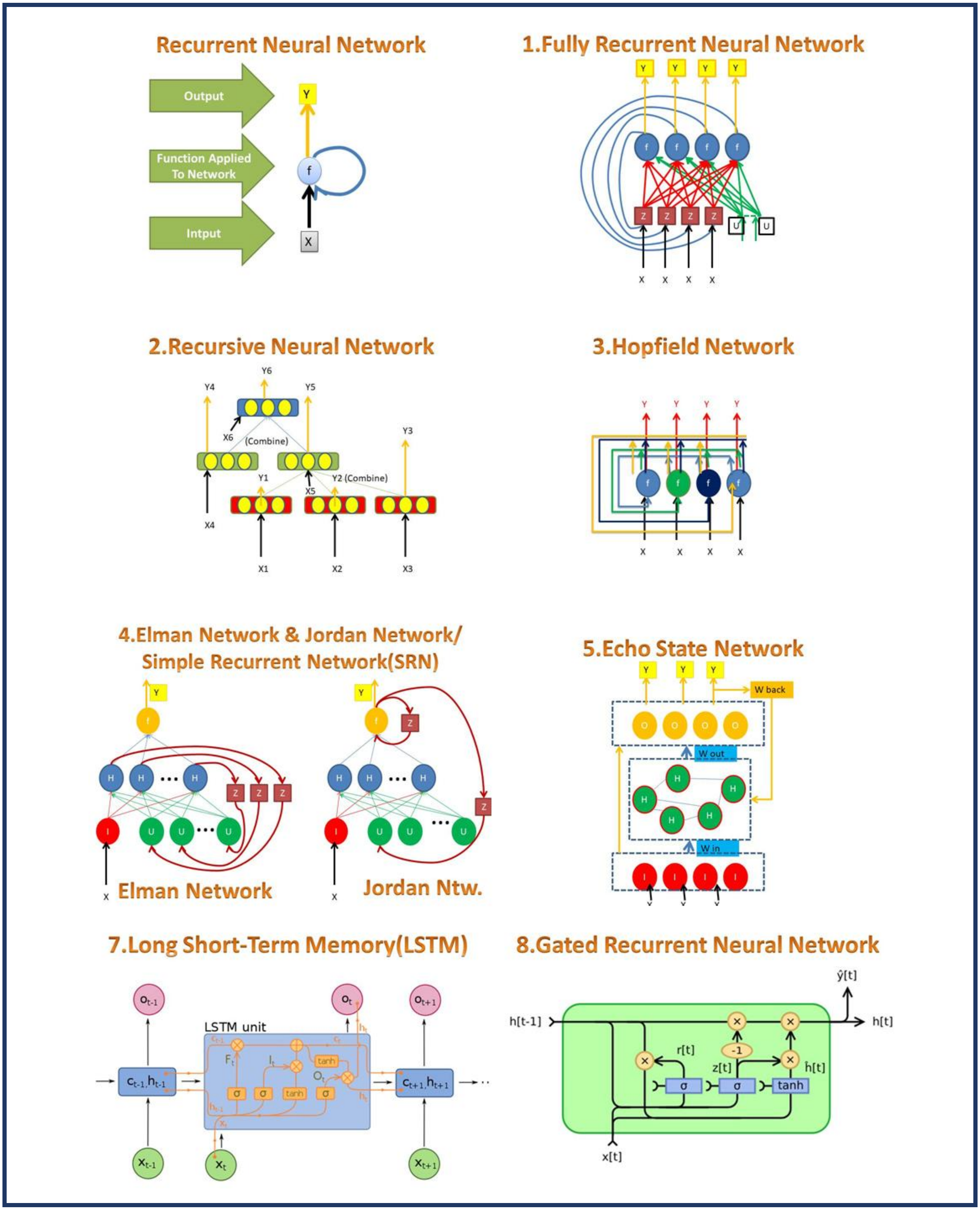
Bi-directional Recurrent Neural Network predicts each element of a finite sequence based on its past/previous and future/next situation. It works in both direction for processing sequence from left-to-right and right-to-left and concatenating their output. This technique is useful when combined with LSTM [9].

10. Continuous time RNN

It uses ordinary differential equations on system to model the effects on a neuron of the incoming spike train. CTRNNs effectively used in evolutionary robotics to address co-operation, vision, and minimal cognitive behavior [10]. Shannon sampling theorem, can be viewed as Continuous time Recurrent Neural Network.

11. Hierarchical RNN

This network work in generating response for a network in top-down hierarchy of layers. One part generates lower network primitive’stop-to-down and other part maintains global goals of top-down sequencing.


Recurrent Neural Network
Output
Function Applied To Network
Intput
Y
f
X
1.Fully Recurrent Neural Network
Y Y Y Y
f f f f
Z Z Z Z
U U
X X X X
2.Recursive Neural Network
Y6
Y4
Y5
X6
(Combine)
Y3
Y1
Y2 (Combine)
X5
X4
X1
X2
X3
3.Hopfield Network
Y Y Y Y
f f f f
X X X X
4.Elman Network & Jordan Network/
Simple Recurrent Network(SRN)
Y
f
H H ••• H
Z Z Z
I
U U ••• U
X
Elman Network
Y
f
Z
H H ••• H
Z
I
U U ••• U
X
Jordan Ntw.
5.Echo State Network
Y Y Y
W back
O O O O
W out
H H H H H
W in
I I I I
7.Long Short-Term Memory(LSTM)
LSTM unit
tanh
σ σ tanh σ
8.Gated Recurrent Neural Network
ŷ[t]
h[t-1]
h[t]
r[t]
-1
z[t]
ĥ[t]
σ σ tanh
x[t]

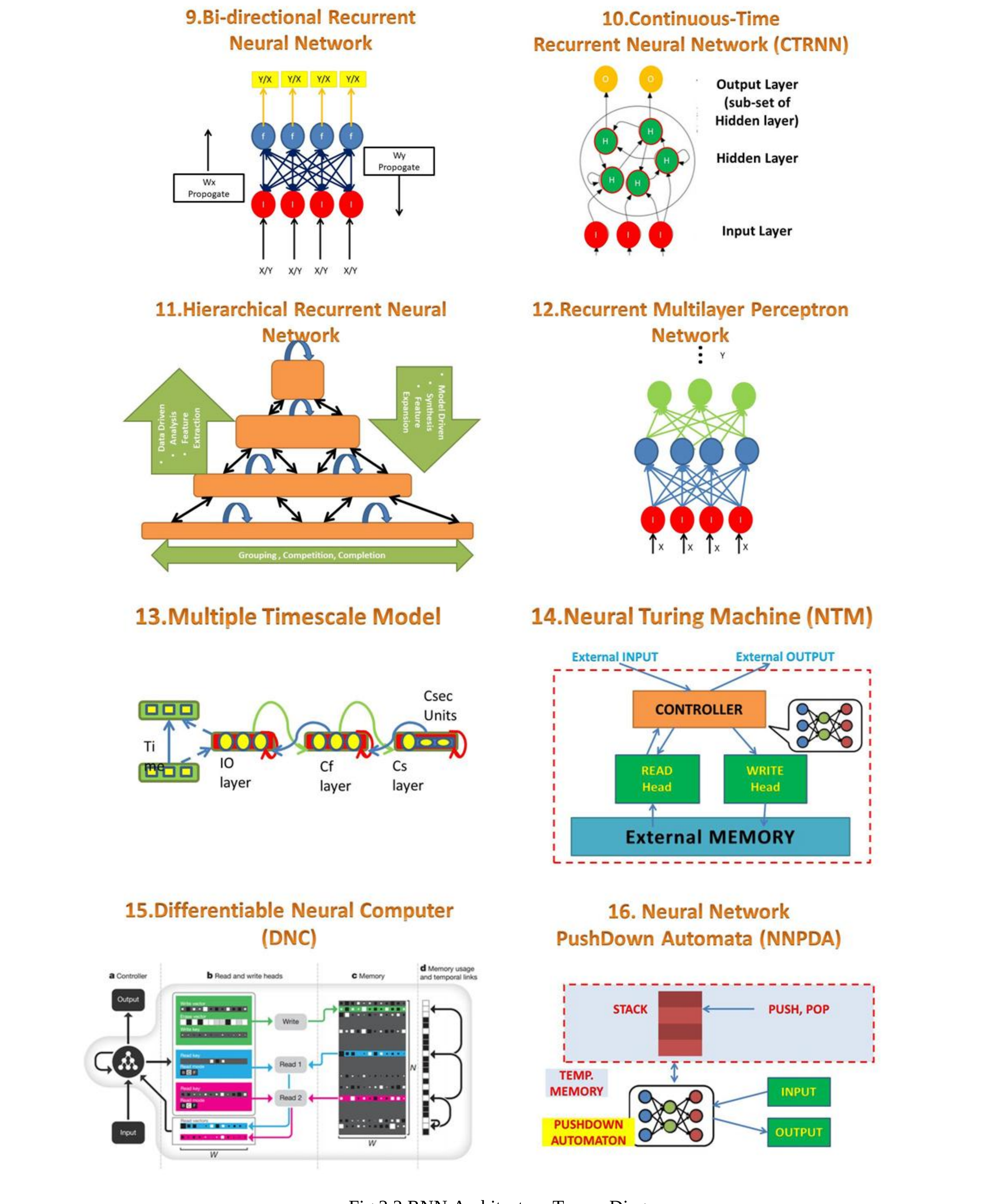


Fig 2.2 RNN Architecture Types: Diagram

Topological limits on information flow decreases nosiness between different parts of the network and improves system performance. Bottom-level neuron and Top-level neurons automatically generates neuron response dynamics in fast and slow speed, respectively. A hierarchical neural network used in controlling a mobile robot [11].

12. Recurrent Multilayer Perceptron

A general RMLP contains a sequence of cascaded, feed-forward, fully connected sub networks [12]. Each subnet/sub network consists of layers of nodes. In this network all the layers are connected in feed-forward; no feed-back connection, except only the last layer is allowed to have feedback connections amongst its nodes. Recurrent Multi-Layer Perceptron (RMLP) is used in identification and control of non-linear dynamic processes. This network architecture has the famous properties of multi-layer perceptron's and it integrates temporal behavior.

13. Multiple Timescales Model

Multiple timescale RNN model is used to simulate functional hierarchy in brain: neural system. Functional hierarchy can be thought as hierarchy in which complex elements can be broken down in to simple elements and simple elements can be combined into complex element. In simple way functional hierarchy can be categorized into two hierarchy parts: hierarchy in space (e.g. Visual information processing), hierarchy in time (e.g. Auditory information processing). Using Multiple timescale in neural activity this model functional hierarchy can self-organizes functional hierarchy [13]. Other research in auditory information processing in Hindustani classical music was done by author [17,18, 19].

14. Neural Turing Machines (NTM)

NTM was introduced by Alex Graves in 2014[14]. This model is combination of fuzzy pattern matching and power of programmable computers. NTM system is extended neural network with controller communicating with external memory resources. It is analogous to a Turing Machine or Von Neumann architecture but difference is that it can be efficiently trained with gradient descent. NTM with LSTM canconclude simple algorithms such as sorting, copying, and associative recall.

15. Differentiable Neural Computer

DNC was published by same person who published NTM, Alex Graves in 2016 [15] with model havingauto associative memory. This model is extension of NTM, with additional memory and temporal attention, and abstract, robust than NTM. DNC model is with differentiable subcomponent and similar to that of Von-Neuman architecture with training provided by gradient descent.

16. Neural Network Push Down Automata (NNPDA)

Neural network pushdown automata are neural network with context free grammars or pushdown automata.NNPDA is like NTM with tape replaced by equivalent stacks. This model attempt to give exact mathematical description, theoretical analysis of related issues.

III. CONCLUSION

Recurrent neural network has greater capability to deep learn. We now know the different architecture used for recurrent neural network with focus on techniques of learning the network. Changes in needs can change the architecture of RNN. With enhancement and addition of new features new architecture of RNN can be developed. Most recent network architecture up to now is differential neural computer with additional feature in NTM. The basic LSTM model can be used in combination to form the new strong network. Because boom of deep learning in market, advancement in each RNN can be effectively used to solve the given problem and RNN architecture are not limited to these architectures only. Further enhancement and additional component new RNN network with stronger capability to solve the problem can be formed.

REFERENCES

[1] Sak, Hasim; Senior, Andrew; Beaufays, Francoise (2014). "Long Short-Term Memory recurrent neural network architectures for large scale acoustic modeling" (PDF).

[2] Griewank, Andreas; Walther, Andrea (2008). Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation (Second ed.). SIAM.

[3] Ramya, C; Kavitha, G., and Shreedhara, K.S., Recalling of Images using Hopfield Neural Network Model, National Conference on Computers, Communication and Controls-2011.

[4] Cruse, Holk; Neural Networks as Cybernetic Systems, 2nd and revised edition.

[5] Jaeger, Herbert; Haas, Harald (2004-04-02). "Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication". Science.

[6] Schmidhuber, Jürgen (1992). "Learning complex, extended sequences using the principle of history compression". Neural Computation. 4 (2): 234–242.

[7] Gers, Felix; Schraudolph, Nicol N.; Schmidhuber, Jürgen. "Learning Precise Timing with LSTM Recurrent Networks ".ResearchGate. pp. 115–143.

[8] Cho, Kyunghyun; van Merrienboer, Bart; Gulcehre, Caglar; Bahdanau, Dzmitry; Bougares, Fethi; Schwenk, Holger; Bengio, Yoshua (2014). "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation". arXiv:1406.1078 .

[9] Graves, Alex; Schmidhuber, Jürgen (2005-07-01). "Framewise phoneme classification with bidirectional LSTM and other neural network architectures". Neural Networks. IJCNN 2005. 18 (5): 602–610.

[10] Harvey, Inman; Husbands, P.; Cliff, D. (1994), "Seeing the light: Artificial evolution, real vision", 3rd international conference on Simulation of adaptive behavior: from animals to animats 3.

[11] Paine, Rainer W.; Tani, Jun (2005-09-01). "How Hierarchical Control Self-organizes in Artificial Adaptive Systems". Adaptive Behavior. 13 (3): 211–225.

[12] K. Tutschku"Recurrent Multilayer Perceptrons for Identification and Control: The Road to Applications". CiteSeerX

[13] Yamashita, Yuichi; Tani, Jun (2008-11-07). "Emergence of Functional Hierarchy in a Multiple Timescale Neural Network

Model: A Humanoid Robot Experiment". PLOS Computational Biology. 4

[14] Graves, Alex; Wayne, Greg; Danihelka, Ivo (2014). "Neural Turing Machines". arXiv:1410.5401

[15] Graves, Alex; Wayne, Greg; Reynolds, Malcolm; Harley, Tim; Danihelka, Ivo; Grabska-Barwińska, Agnieszka; Colmenarejo, Sergio Gómez; Grefenstette, Edward; Ramalho, Tiago (2016-10-12). "Hybrid computing using a neural network with dynamic external memory". Nature. 538: 471–476.

[16] Trupti Katte , "Multiple Techniques for Raga Identification in Indian Classical Music", In proc. Of Vol.4, Issue 6, NCRTCST-2013.

[17] Trupti Katte, BhavanaTiple, "Raga Identification By Pakad Matching In North Indian Classical Music", In proc (IJCET), Vol.4, Issue 6, pp161-166, Nov. Dec. 2013.

[18] Trupti Katte, BhavanaTiple, "Techniques for Indian classical Raga identification- A Survey" In Proc. IEEE Pune,Dec. 2014. India.

AUTHORS BIOGRAPHY

Trupti Katte received ME degree in computer engineering from University of Pune (SPPU). Currently, she is an assistant professor at Army Institute of Technology.