An Unsupervised learning based LSTM model: A new architecture

DR. SAJJAD MOHSIN, FATIMA ZAKA
Department of Computer Science
COMSATS Institute of Information Technology
Park Road, Chak Shahzad, Islamabad
PAKISTAN
smohsin@comsats.edu.pk, fatima.zaka@gmail.com

Abstract: - Recurrent Neural Networks (RNNs) have shown good results with real-world temporal contextual data, but for input sequences with long time lags they fail. Long Short Term Memory (LSTM) model was built to address the issue of large time lags in input data successfully. However, LSTM found lacking for the tasks pertaining to lower level cognitive processing, specifically, information processing, storage & recall, and also whether they could learn in an unsupervised manner. Sustained Temporal Order Recurrent (STORE) networks are designed to encode the order of temporal data, and then could recall the encoded data in veridical as well as non-veridical order employing unsupervised learning. In this research we have propose the Fusion of supervised learning base LSTM propose by Jurgen and unsupervised learning based STORE proposed by Grossberg. To alternate between two approaches as well as mimicking brain in information processing during sleep time (internal input) we proposed CCS (Consolidation Control Unit), built on an in-depth cognitive foundation, to overcome the inability of LSTM to learn in unsupervised manner and to work with lower level cognitive processing. We conclude by providing experimental proof of the efficiency of proposed model by comparing it with original LSTM model.

Key-Words: - LSTM, STORE, fusion, CCS, unsupervised learning, RNNs

1 Introduction

Few examples exist of unsupervised learning with respect to temporal data and employing recurrent nets to model lower level cognitive processes. One example is a hybrid of recurrent neural network (employing the extended kalman filters for training the recurrent net) with ART2. Vieira and Lee proposed with the ability to adaptively learn in response to varying input patterns and then it further transfers this learning to dynamically growing group of simple recurrent nets[1]. Similarly Adaptive Resonance Theory (ART) neural networks have been used to cater to temporal [2] as well as spatial-temporal input sequences [3]. A breakthrough occurred later on when Bradski et al proposed a variant of basic ART architecture that could store as well as recall various temporal input sequences[4]. Sustained Temporal Order Recurrent (STORE) model although could not process the stored values to perform any tasks, but their key objective was to serve as working memory to other networks[4].

Long Short Term Memory (LSTM) performs well where recurrent neural networks fail specially with respect to time lag problems, and performs better than other approaches to solve various bench mark problems[5]. But the gradient learning algorithm of LSTM cannot support unsupervised learning (error-based learning algorithm). LSTM proposed by Klapper.et.al was trained using two unsupervised learning algorithm to discriminate groups of temporal sequences[6]. But no comparison with other benchmark solutions was provided in this work and application scope was limited. Furthermore, a preliminary critical analysis determined that the main focus of the application of LSTM for solving cognitive reasoning problems remains on high level reasoning problems. Problems like learning languages, speech recognition, and hand-writing recognition were quite successfully handled by LSTM architecture, and exceeded in terms of performance and robustness while in comparison with other approaches[7-9]. Therefore LSTM architecture needs to be modified to cater to lower-level reasoning tasks with ease as well. Also although LSTM can extract information conveyed by the temporal order of widely separated inputs but the basic architecture does not facilitate the non-veridical recall of any previously given temporal sequences.

In this paper a new approach is proposed to model the cognitive information processing and storage employing LSTM model. The proposed solution involves making certain additions to the basic LSTM architecture by fusing it with a working memory and the most suitable choice for working memory was STORE, as it caters to temporal data, supports unsupervised learning, and

ISBN: 978-960-474-270-7

allows veridical as well and non-veridical recall of information.
The rest of the paper is organized as follows; section II reviews the biological plausibility giving the foundation for the proposed architecture. Section III gives computational model as well as the working of the new architecture; the experimental setups and results are given in section IV, and section V concludes the paper with discussion on the proposed model and comparative results.

2 Cognitive Foundation

All the widely accepted theories of cognitive information processing like the “stage theory” [10] and the “level of processing theory” [11] explain that humans while learning and processing information make use of various distinct levels of elaboration. Then we have theories like parallel-distributed processing and connectionist [12] [13]: as their names indicate these theories give the idea that information is processed simultaneously but by several different parts located in our memory system. How the brain stores the data, and learns it in what order is still not known, the known part is that it stores and learns this data in an order that makes it’s working efficient and robust. The brain learns new information and classifies input data based on target data provided by external teacher, this is supervised learning for brain. The brain by employing unsupervised learning can form perceptions based on previously learned concepts and information; and use the new data from sensory organs and these pre-formed perceptions to take decisions and perform tasks in an optimized manner.

According to above given arguments in order to suggest a system modeling cognitive information processing skills we need to benchmark certain qualities like: it can process in parallel, where tasks are divided in different units of processing; the system should definitely be able to cater to real world data (which is temporal in nature mostly); and this system should be able to encode the information as well as the order of input, and recall information even in non-veridical order. Similarly one cannot ignore the required element of unsupervised learning in the proposed system.

3 LSTM-STORE: Architecture and Computational Model

Based on the cognitive foundation provided in the section II, a new architecture is proposed by fusion of LSTM and STORE for modeling the lower lever cognitive information processing. LSTM-STORE gives a solution with levels of processing and also where processing is being performed in parallel, in distinctive areas of network. The distinct areas of parallel processing (as shown in figure 1) are the LSTM and STORE, and the Consolidation Control System (CCS) unit enables the proposed model to alternates between supervised and unsupervised processing details about CCS is given ahead. Thus we get a new robust solution with the underlying information processing approach applied is connectionist. The temporal input (real world data with time lags) is simultaneously provided to the STORE and LSTM units; while LSTM learns to classify the sequence the STORE unit encodes the order of the sequence. During external input presentation LSTM-STORE learns a sequence from external teacher in a supervised manner. In absence of external input STORE unit trains the LSTM network for encoded information and the information recalled by STORE to train LSTM could be in veridical as well as non-veridical order.

Consolidation Control System: Figure 1 shows a control unit CCS (Consolidation Control System) that is being used to adjust the connection from the STORE unit to the LSTM unit, this connection is set to zero initially, when input is coming from external input unit. This is the supervised learning phase and in it the external input and target pair train the network. When the external input sequence ends, the CCS activates the connection from STORE to LSTM, via this connection the previous input pattern is fed back to LSTM in non-veridical order of either recency or bowing and even primacy if the need arises. This is the unsupervised learning phase, as now the teacher is STORE an internal component of model. The CCS also controls the representation of input to the STORE unit.

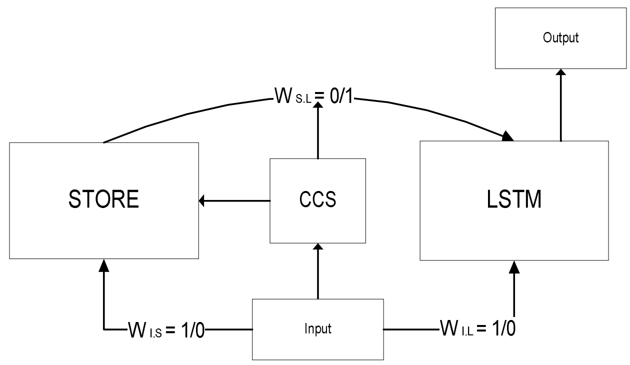

Figure 1: LSTM-STORE Architecture

An input sequence is presented to the LSTM-STORE model through the time interval [α-t_i, t_i]. The input sequence is fed simultaneously to both the LSTM and STORE unit, LSTM while learns the sequence through the truncated Back-Propagation Through Time (BPTT) algorithm, STORE encodes the temporal order of the sequence[14,15]. The item representation to the STORE is in format I= I_1, I_2… I_n.

The affect of I is stored by the activity pattern across the STORE network; this activity pattern at any time t_j in given as: $z_1(t_j), z_2(t_i), \ldots z_k(t_j) \ldots$ [4] and the weights of the connection between Input unit, LSTM and STORE unit (shown in figure 1) during the supervised learning phase are set to:

$$W_{I.S} = 1$$
$$W_{I.L} = 1$$
$$W_{S.L} = 0$$

At time (ti+1) when external input pattern ends, the unsupervised learning phase begins, at this phase Consolidation Control System switches the connection weights to:

$$WS.L = 1$$

And in the absence of external input the following connection weights become:

$$WI.S = 0$$
$$WI.L = 0$$

At this instance the activity pattern across STORE will contain the complete affect of previously presented input sequence and the activity pattern will be of the form: $z_1(t_i), z_2(t_i), \ldots z_k(t_i) \ldots$ Now the previous input pattern will be recalled and will be provided to the LSTM unit for rehearsal. The temporal order of this input sequence will now change; and now it will either present primacy order if:

$$z_{k-1}(t_i) > z_k(t_i)$$

Or it will present recency if:

$$z_k(t_i) > z_{k-1}(t_i)$$

Or bowing which combines the affect of both primacy and bowing.

4 Comparative Experimental Results

Task: The task selected to test the above described architecture was simple but sufficient to test the suitability of our proposed solution with respect to lower cognitive reasoning task. These tasks are input storage, recall (veridical as well as non-veridical order) and processing. The LSTM and LSTM-STORE were trained for a given input sequence in specific order. For testing, the same sequence was presented to both the LSTM and LSTM-STORE network out of order to check if the networks could classify the sequence, and how efficient the classification process will be. The input sequences had the long time lags that only LSTM architecture could handle adequately.

Comparison: The training and testing sequences used for both the LSTM and LSTM-STORE architectures were similar, and results were obtained for both the models for same parameters and input target sequences.

Architecture: A simplest possible setup was implemented here to test the model; STORE1 was chosen as working memory to encode order. It has two layered architecture, STORE1 can not only successfully encode the invariance principal [4] but it is also robust as the input durations do not affect the stored activity pattern. This is suitable for the type of input patterns for which LSTM is termed effective. The number of input and output unit was set as 1, and output layer was biased. The number of memory cell blocks for LSTM was set at 2, with each block having size 2; learning rate was set at 0.1 for our experiment.

Results: One important thing to note before moving to the result is that the initial weights for LSTM unit's connection are randomly chosen between the range [-0.1, 0.1]. And for the connections between CCS, LSTM and STORE units the weights alternates between 0 and 1 according to learning phase. Also the training set and testing set both were unique; with each set ten trials were conducted with different initial weights (all weights are between the above mentioned range of [-0.1,0.1]). Similar situations were used to test both the LSTM and LSTM-STORE architectures. Interestingly where LSTM cannot perform well LSTM-STORE models learn to solve the task with improved results.

The results of the ten trials for LSTM and LSTM-STORE architecture are given in the Table 1:

Table 1

Trial No.	Total Epochs to reach the set value of MSE* for LSTM	Success at Epoch (the epoch at which the sequence is successfully classified for LSTM)	Total Epochs to reach the set value of MSE for LSTM-STORE	Success at Epoch (the epoch at which the sequence is successfully classified for LSTM-STORE)
1	101	101	60	49
2	128	127	74	64
3	121	115	68	57
4	111	106	63	52
5	125	116	70	59
6	124	119	70	59
7	109	105	63	52
8	111	108	64	53
9	122	116	68	57
10	115	111	65	54

MSE: Mean Square Error, value set same for LSTM and LSTM-STORE

The above table shows the reduced number of epochs in case of LSTM-STORE architecture while in comparison with the original LSTM architecture. Also the LSTM-STORE achieves success in classifying sequences earlier then the LSTM for the test set.

5 Conclusion

This research work is an attempt to model the lower level cognitive tasks. The focus remains on modeling ability of brain to learn new information with the aid of external teachers and to learn derived knowledge in an unsupervised manner based on the existing learned knowledge. The approach of modeling adopted was of neural networks, as neural networks are simulators of the neural behavior of brain due to their neural plausibility they can imitate several cognitive behaviors with ease as compared to symbolic solutions[16].
By augmenting Long Short Term Memory with a working memory like Sustained Temporal Order Recurrent, which is designed to handle temporal data, we can improve the performance as shown in Figure 2. The average increase in performance is 57%, if we compare the results obtained from LSTM and LSTM-STORE architecture.
The implications of this are quite significant both with respect to cognitive modeling of lower level reasoning tasks but also with respect to language processing, hand writing recognition and speech and vision processing which are categorized as higher level cognitive reasoning tasks. We can improve and optimize the neural network architectures being employed to model these tasks by instilling in these networks the underlying processes like perception and storage.

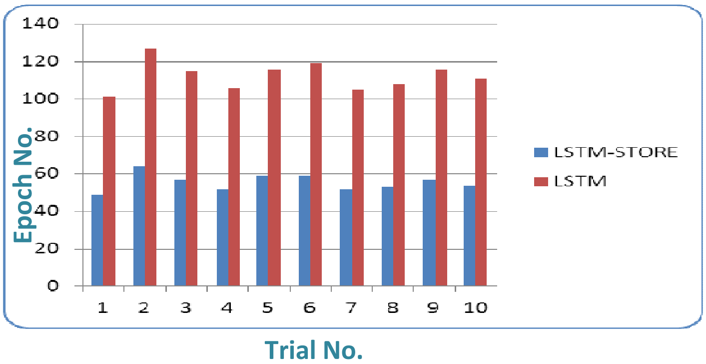

Figure 2: Chart comparing the performance of LSTM and LSTM-STORE architecture with the criteria: at which epoch network learned to classify all input sequences with out any misclassification. At X-axis we have the Trial No. and at Y-axis we have the epoch number.

For future work the architecture could be improved and tested with different type of input data especially real world input tasks like speech and visual processing. Further experiments could be performed to verify several ideas proposed by cognitive psychology researchers. Certain improvements in the architecture could also be proposed to enable it to suit better to higher level cognitive tasks and compare performance with LSTM model.

References:

[1] Vieira, F. H. T. and Lee, L. L. A Neural Architecture Based on the Adaptive Resonant Theory and Recurrent Neural Networks. In: International Journal of Computer Science and Applications. Vol. 4 Issue 3, pp 45-56. 2007

[2] Wang, D. L. & Yuwono, B. Incremental learning of complex temporal patterns. IEEE Transactions on Neural Networks. 1465-1481, 1996

[3] Silva S. and Caetano. Using Artificial Recurrent Neural Nets to Identify Spectral and Spatial Patterns for Satellite Imagery Classification of Urban Areas. 1997

[4] Bradski, G., Carpenter G. A., Grossberg S. STORE working memory networks for storage and recall of arbitrary temporal sequences. Biological Cybernetics, 71(6), October, pp 469-480, 1994.

[5] Hochreiter S. & Schmidhuber. J. Long Short-Term Memory. Neural Computation, 9(8):1735-1780, 1997

[6] Klapper-Rybicka M., Schraudolph N. N., Schmidhuber J. Unsupervised Learning in LSTM Recurrent Neural Networks. Proceedings of the International Conference on Artificial Neural Networks, pg. 684-691. 2001

[7] Gers F. A., Schmidhuber J. LSTM Recurrent Networks Learn Simple Context Free and Context Sensitive Languages. IEEE Transactions on Neural Networks 12(6) 1333-1340. 2001

[8] Graves A., Schmidhuber J., Frame-wise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks 18:5-6, pp. 602-610. 2005

[9] Graves A., Liwicki M., Fernandez S., Bertolami R., Bunke H., Schmidhuber J. A Novel Connectionist System for Improved Unconstrained Handwriting Recognition. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, pg. no. 5. 2009

[10] Atkinson, R., and Shiffrin, R. Human memory: A proposed system and its control processes. In: K Spence & J Spence (Eds.). The psychology of learning and motivation: Advances in research and theory (Vol. 2). New York: Academic Press. 1968.

[11] Craik, F. & Lockhart, R. Levels of processing: A framework for memory research. Journal of Verbal Thinking and Verbal Behavior, 11, 671-684. 1972

[12] Rumelhart, D. Schemata: The building blocks of cognition, In R. Spiro, B. Bruce, & W. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, NJ: Erlbaum. 1980

[13] McClelland, J. & Rumelhart, D. Parallel distributed processing. Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models. Cambridge, MA: MIT Press. 1986

[14] Williams R. J. and Zipser, D. Gradient based learning algorithms for recurrent networks and their computational complexity. In Backpropagation: Theory, Architectures and Applications. Hillsdale, NJ: Erlbaum.1992

[15] Werbose P. J. Generalization of Backpropagation with application to a recurrent gas market model. Neural Networks. 1, 1988

[16] Stillings, N. A., Weisler, S. W., Chase, C. H., Feinstein, M. H., Garfield, J. L., Rissland, E. L. Cognitive Science: An Introduction. MIT Press. 1995