EXPERIMENTS ON SEQUENCE PROCESSING WITH NEURAL NETWORKS

Luyuan Fang and William H. Wilson
(Discipline of Computer Science, Flinders University of South Australia,
Bedford Park, S.A. 5042, AUSTRALIA)
Tao Li
(Department of Computer Science, Monash University, Clayton, VIC 3168,

ABSTRACT

Neural networks for storing and for classifying sequences are studied in this paper. A model of sequential neural networks is proposed here. This model can be used as a classifier for sequences and as a memory for storage and retrieval of sequences. Some experiments have been conducted.

INTRODUCTION

The recent resurgence of neural network research has brought together many disciplines in cognitive science such as cognitive psychology, artificial intelligence, computer science, neurophysiology, and cybernetics. This is not without any special reason. Neural network research is the ground for all these disciplines and ties together many issues about cognition. Many models of neural networks have been proposed. Neural networks have also been applied to solving combinatorial problems.

Most of the neural networks take a finite size input (viewed as a window). These networks have difficulties handling variable size or conceptually infinite input sequence. Some researchers have considered the sequential presentation problem. But there remains much research work to be accomplished in this area. There seem to have three main aspects in the study of sequential neural networks: 1) grammar learning and sequence identification, 2) sequence neural memory, and 3) partial match retrieval.

Grammar Learning. It is often needed to classify sequences of patterns that obey some grammar. Sequential patterns can be of variable, and sometimes infinite, length. An example of sequential pattern classification is the parity problem in which a network is required to output a 1 if the number of 1's in the input sequence is even and to output a 0 otherwise. One would like to teach a neural network to learn to classify odd-even input patterns. Here the emphasis is placed on learning rather than the traditional way of designing finite state automata. The neural net is expected to generalize from the given examples in learning.

Sequence Memory. Kohonen proposed a sequential memory model. The main purpose of his sequential memory is to store sequences of patterns and to recall these sequences at a later time. The difference between this sequence memory and other associative memories is that the order of input/output patterns for the sequence memory must be preserved. Again the emphasis is placed on the learning aspect of the neural network. For example, the music notes represents the sequence to be stored and the name of the song can be viewed as the background (as in Kohonen's model).

Partial Match Retrieval. This involves the storage of some sequences and the retrieval of the sequences at a later time when only a small part of a sequence is given. This happens very often in our life. For example, humans learn to sing songs. The music and the words of a song often contain repetitions of some segments. When one hears a segment of music, he/she can immediately recognize the song. This is partial match retrieval of sequences. It deserves special attention due to the commonality of such phenomenon.

A MODEL OF SEQUENTIAL NEURAL MACHINE

Teaching a neural network to learn to recognize the important features of sequences is very important to neural network research. Unfortunately this area of research has not been well-explored yet. A model of sequential neural network was recently proposed by Gallant.

Gallant's Sequential Associative Memory. Stephen Gallant proposed an interesting model of sequential neural network which consists of highly interconnected subnetworks called Sequential Associative Memories (SAM's) and some trainable cells having modifiable connection weights. His model consists of a number of
SAMs and some three states perceptron units. A SAM consists of a number of perceptrons, the weights of which are generated at random and remain fixed. SAM cells are used to determine a state for recalling previous inputs and outputs and to enable the modifiable cells in the network to learn difficult tasks. The structure of this model is shown in Figure 1.

![Diagram of Sequential Network Model](image)

**Figure 1. The sequential network model of Gallant**

An easy learning algorithm, the pocket learning algorithm, was used to the trainable units for changing the connection weights. The network was able to learn to solve the parity problem, to do arithmetics of variable

**A Sequential Neural Network for Classification.** The reason for using random networks of SAMs was not explained in ref. However, one finds that the SAM model has difficulty learning sequences with repeated segments such as in music. This is because the connections and weights of SAMs are fixed. Hence the internal states generated by the SAMs are also fixed. So the network will have trouble memorizing if two distinct segments both follow the same segment. For example, a music is composed of three segment types, seg1, seg2, and seg3. If the music is ordered as seg1 seg2 seg1 seg3, the SAM model will have difficulty learning this sequence. This is because seg2 and seg3 both follow seg1. After the network has seen seg1, it will not always be able to correctly decide which of seg2 and seg3 will follow due to the fixed connections in the network of SAMs.

The ability to learn sequences with arbitrarily ordered and repeated segments, it is necessary to make modifiable the subnetwork for generating the next state. Although it is difficult to automatically recognize important segments of reasonable size, such problem can be overcome with a little human assistance. If easy learning algorithms are employed, the sequence of state transitions can be readily specified and the network can learn to recognize sequences with arbitrarily ordered and repeated segments.

A model of sequential neural network is proposed here. This model consists of two subnetworks, a state transition network and an output network. Each of the subnetworks contains three layers of perceptrons with modifiable connections and weights. The outputs of the state subnetwork are connected to the delay elements D1, D2, ..., Dk. The state subnetwork also takes some of its inputs from the delay elements. Figure 2 illustrates the new model.

This model certainly looks like a typical sequential circuit. It should also behave like a sequential circuit. However, there is an important difference between this model and the traditional sequential circuits. That is, this neural network learns to recognize (or to memorize) sequences in contrast to a traditional finite state machine which is designed to perform certain tasks. The network produces desired sequential behaviour according to what it was taught but not what it was designed for.
Figure 2. A Sequential Neural Network Model.

To use the sequential neural network for classification, the connection weights are adjusted according to the given input sequences and corresponding output sequences during the training process. The network learns the characteristics about the sequences by adjusting the weights. During the classification process, a sequence is presented to the network as input. The network produces desired outputs by extracting the important features of the input sequence. Different outputs are generated for sequences with different features. Hence a trained network can classify input sequences by important features.

Computing Activations for Cells. Let $C_i$ denote the activation of the $i$th perceptron cell in the network, where $C_i$ can assume the values $\{+1, -1, 0\}$ corresponding to \{true, false, unknown\}. If the cell $C_i$ has $m$ inputs, then $C_i$ is computed as

** $C_i(t+\delta t)=+1$ if $\sum_{j=1}^{m} C_j(t)W_{ij} > 0,$

** $C_i(t+\delta t)=-1$ if $\sum_{j=1}^{m} C_j(t)W_{ij} < 0,$

** $C_i(t+\delta t)=0$ if $\sum_{j=1}^{m} C_j(t)W_{ij} = 0,$

where $W_{ij}$ is the connection weight from the $j$th cell to the $i$th cell. The threshold is assumed to be 0 for every perceptron unit in the network. For computational reasons, we assume, without loss of generality, an indexing scheme for cells in which each cell can be connected to only those cells with lower index numbers.

Computing Next States and Outputs for the Network. To produce deterministic sequential behaviour, the network must use synchronous update or iterative update with fixed order. Using the cell indexing scheme
given in the previous subsection, synchronous and iterative updates are equivalent. Assuming that a network consists of \( n \) cells \( C_1, \ldots, C_n \) and \( p \) inputs \( I_1, \ldots, I_p \), the iterative update algorithm is described below.

FOR \( i = 1 \) TO \( n \) DO
BEGIN
\[ C_i \leftarrow 0; \]
FOR \( j = 0 \) TO \( p - 1 \) DO \[ C_i \leftarrow C_i + I_{j+1}W_{ij}; \]
FOR \( j = p \) TO \( i - 1 \) DO \[ C_i \leftarrow C_i + C_{j-p+1}W_{ij}; \]
END.

In the above, a cell is assumed to be connected to all inputs and all cells with lower indices.

This iterative update algorithm is used to compute the next state and output for the network. When an input vector \( \mathbf{N} \) is given, the activations of the cells are updated and the network state and output are changed.

1) Form the new input vector \( \mathbf{I} \) for the state subnetwork by combining the current network input vector \( \mathbf{N} \) with the contents of the delay units;
2) Compute a new state from the new state input vector \( \mathbf{I} \) by applying the iterative update algorithm to the state subnetwork and store the new state in the delay units;
3) Form the new input vector \( \mathbf{O} \) for the output subnetwork by combining the current network input vector \( \mathbf{N} \) with the current contents of the delay units;
4) Compute the new output from the vector \( \mathbf{O} \) by applying the iterative update algorithm to the output subnetwork.

To learn and to recognize sequences, the network must be presented with sequential input patterns. During the recognition process, the procedure described above will be used to generate next states to compute outputs for the network.

**EASY LEARNING**

The learning algorithm presented here is an "easy" one. For an easy learning algorithm, the desired activation of every cell, whether an input/output cell or an internal cell, must be specified.

**Learning of Connection Weights.** A slightly modified perceptron learning procedure is adopted here. Integer weights are assumed for connection weights. Let \( <E, C> \) be a training pattern for a cell \( C \) with \( m \) inputs, where \( E = (E_1, E_2, \ldots, E_m) \) is the given input vector and \( C \) is the desired activation. Also let \( W(t) = (W_{11}, W_{12}, \ldots, W_{m1}) \) be the current weight vector for the cell \( C \). Then the weight vector \( W \) is updated during learning according to the following rules,

\[
\begin{align*}
\text{if } EW(t) & \text{ and } C \text{ have the same sign, } W(t+\delta t) = W(t); \\
\text{otherwise, } W(t+\delta t) = W(t) + CE. & \quad (1)
\end{align*}
\]

\( EW(t) \) is an inner product. When \( EW(t) \) and \( C \) have the same sign, the network correctly classifies the pattern. Thus there is no need to modify the weights. Otherwise, the weights are adjusted accordingly. If the set of sample patterns is known to be linearly separable, rule (1) above can be repeatedly applied until a sample pattern is correctly classified. In other words, the rule \( W = W + CE \) is repeatedly applied until \( EW \) and \( C \) have the same sign. Then \( W(t+\delta t) \) is set to \( W \). This modified rule is used more often in practice.

**A Simple Learning Algorithm.** This algorithm is a modified version of the pocket learning algorithm.\(^2\) Let the set \( S \) of sample sequences be \( S_1, S_2, \ldots, S_l \) and the corresponding lengths be \( k_1, k_2, \ldots, k_l \). The set of sample sequences can be listed in the following form.

<table>
<thead>
<tr>
<th>( R_1 )</th>
<th>( R_2 )</th>
<th>( R_{max} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_1 )</td>
<td>( S_{11} )</td>
<td>( S_{12} \ldots S_{1k_1} )</td>
</tr>
<tr>
<td>( S_2 )</td>
<td>( S_{21} )</td>
<td>( S_{22} \ldots S_{2k_2} )</td>
</tr>
<tr>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( S_l )</td>
<td>( S_{l1} )</td>
<td>( S_{l2} \ldots S_{lk_l} )</td>
</tr>
</tbody>
</table>
where $R_{max} = \max(k1, k2, \ldots, k_l)$ and each $S_{ij}$ is an input vector. This has the form of a matrix. The network can learn the patterns row-wise or column-wise.

In row-wise learning, a complete sequence $S_t$ is presented to the network each time. For classification purpose, the initial state $D_0$ is the same for every input sequence. Let $D_{ij}$ be the desired network state when input pattern $S_{ij}$ is presented and $Y_{ij}$ be the corresponding output pattern. Row-wise learning algorithm is described as follows:

1) Randomly select a sequence $S_t$ from the set $S$ of sample sequences and delete $S_t$ from $S$;  
2) FOR $j = 1$ TO $k_l$ DO
   a) Present $S_{ij}$ to the network;
   b) Form an input vector $I$ for the state subnetwork, where $I = S_{ij}^*D_{ij}$;
   c) Specify the desired activation for each cell in the state subnetwork (by combining $D_{ij}$ and the internal cell activations);
   d) Apply cell learning procedure to adjust connection weights in the state subnetwork; test run (using iterative update) with input vector $I$ to see if the output is $D_{ij}$;
   e) Form an input vector $O$ for the output subnetwork, where $O = S_{ij}^*D_{ij}$;
   f) Specify the desired activation for each cell in the output subnetwork (by combining $Y_{ij}$ and the internal cell activations);
   g) Apply cell learning procedure to adjust connection weights in the output subnetwork; test run (using iterative update) with $O$ to see if the output is $Y_{ij}$;
3) If either of the tests in 2e and 2g fails, then record the number of sequences which have been learned so far as the pocket length; Otherwise, go to step 1.

This algorithm is repeatedly applied until all the sample sequences are correctly classified or time runs out. A modification to the above algorithm would be such that steps 2d and 2g test-run on all the previously tried sample sequences instead of just the current sample sequence. The column-wise learning algorithm is similar to the row-wise learning algorithm.

**STORAGE AND RETRIEVAL OF SEQUENCES**

In this section, Kohonen's sequence memory is first examined. The sequential neural network is then considered for the same application.

**Kohonen's Sequence Memory.** Kohonen's sequence memory consists of an auto associative network and a number of delay elements as shown in Figure 3. Some inputs of the auto associative network come from the delay elements and some outputs are connected to the delay elements. This forms a feedback loop and generates internal states.

![Figure 3. Kohonen's sequence memory.](image)

Among all the network inputs, some serve the purpose of sequence inputs and some serve the purpose of backgrounds. If the input sequence represents a music, the background pattern represents the name of the
The sequences are stored with the names. When a name is given, the sequence can be recalled. In Figure 3, \(P_1, \ldots, P_n\) are the sequence inputs and \(B_1, \ldots, B_k\) are the background inputs.

**Neural Sequential Machine as Sequential Memory.** The sequential neural network can be used to store and to retrieve sequences. To store several sequences, the network can be initialized to several distinct starting states. Therefore, the background inputs are not needed.

The network configuration is the same as discussed in section 2. Background inputs are eliminated. To enable the network to learn a sequence, a distinct starting state \(D_0\) is given. The learning algorithm discussed in section 3 is then applied to the network. To retrieve the sequence, the network is initialized to the state \(D_0\). Then a null input is given to the network. This forms a partial pattern. The network can continue with this partial pattern from state to state and eventually retrieve the stored sequence.

**PARTIAL MATCH RETRIEVAL**

It is often desired to store some sequences in a network and to match a specific sequence at a later time when a segment of the sequence is given. This seems similar to sequence classification. But they are not exactly the same. Partial match retrieval occurs so frequently in practice that it deserves special attention.

For an application of this type, a special output port is set for each sequence to be stored. If a total of \(n\) sequences are to be stored, the network should have \(n\) output ports. For all the sequences, a common starting state \(D_0\) is used. During the learning process, the output is set to \(+1\) for every pattern in the sequence. The network learns the \(n\) sequences as described in section 3.

For the purpose of partial match retrieval, a competitive network is added to the output end of the sequential neural network. This competitive network, as shown in Figure 4, consists of \(n\) cells, each of which is connected to all \(n\) output ports of the sequential neural network.

![Figure 4. Partial match classification with competition network.](image)

The connection from the \(i\)th output port to the \(i\)th cell is assigned a weight \(+1\) for \(1 \leq i \leq n\). The other connections of a cell are assigned the same weight \(-\Delta\), where \(0 < \Delta < 1/n\). These cells remember their previous states (activations). The short-term memory trace for these cells is defined as,

\[
\frac{dC_i(t)}{dt} = O_i(t) \cdot \sum_{j \neq i} O_j(t) \Delta_i
\]

where \(C_i\) is the activation of the \(i\)th competitive cell and \(O_j\) is the value of the \(j\)th output port. During the partial match retrieval process, the network is initialized to starting state \(D_0\). The partial input sequence is then presented to the network. At the end of the partial sequence, select the cell which has the highest activation value. If cell \(C_j\) is selected, the sequence \(S_j\) matches the partial sequence.

**LEARNING SEQUENTIAL PATTERNS: SOME EXPERIMENTS**

Some simple simulation experiments have been performed to study the behaviour of the sequential neural network. Although simple, these experiments prove the viability of the model proposed in this paper.
The first and the simplest experiment was conducted to solve the parity problem for sequences of arbitrary length. The network was implemented as a Moore machine rather than the Mealy machine in Figure 1. The state network has a single delay element and has two inputs, one from the delay element and one from the network input. There is no cell in the output network. The output of the state network serves as the network output.

Another experiment was conducted to solve the problem of deciding whether the total number of 1's in a sequences can be divided by 3. The network should output the remainder. This is similar in nature to the parity problem but the complexity is higher. The network also learned to solve this problem.

Another experiment is also being conducted to solve the problem of repeated segments. This problem neither can be solved by Kohonen's model nor by Gallant's model. The approach we plan to take is to assign different internal states to repeated segments if they are followed by different segments. This problem should also be solvable by our neural network.

SUMMARY

A new model of sequential neural network is proposed here. The applications of such a model is discussed. Three important application areas are also studied. These areas include the classification of sequences, the storage and retrieval of full sequences, and the partial match classification of sequences. The sequential neural network model presented in this paper can be applied to all the three areas of applications. Some experiments have been conducted to prove the feasibility of the model. In the near future, we will perform more experiments to show the potential of the sequential neural network. Theoretically, the sequential neural network model can solve some problems which are not solvable in previous research.

REFERENCES