Structure and basic functions of cognitive neural network machine

Vasiliy Osipov*

SPIIRAS, 199178, Saint-Petersburg, 14-th line of V.I., 39, Russia

Abstract. The structure of a cognitive machine based on a recurrent neural network with controlled elements is considered. The basic functions of this machine are defined and analyzed. The expanded capabilities of the cognitive machine to solve creative tasks are discussed. The results of modelling are given.

1 Introduction

In the coming years, small-scale cognitive neural network machines (CNMs) can form the basis for control systems of autonomous robots operating in difficult-to-forecast and difficult-to-reach conditions. These machines are understood as those that are capable of perceiving and connecting events (signals). Such CNMs should form the corresponding space-time models and develop creative solutions. These machines must quickly solve a wide range of difficult-formalized creative tasks.

There are several approaches to the creation of CNMs. These include digital, analog-digital and analog approaches. Examples of the digital approach are CNMs based on TrueNorth chips developed under the DARPA SyNAPSE program [1, 2]. This chip emulates 1 million neurons and 256 million synaptic connections between them. Among the analog-digital CNMs, it is possible to single out the systems developed within the framework of the BrainScaleS project: Neuromorphic processors. To create them, wafer with 384 HICANN chips is used. Each wafer allows emulating 200,000 neurons and 49 million synaptic connections [3]. As an analog cognitive neural network, there is a resistive processor of IBM [4], and others [5].

Now there is an opportunity to implement large neural networks and cognitive machines on chips with memristors. However, the level of development of artificial neural networks and ways of cognitive processing of information in them largely does not satisfy the needs of practice [6 - 13]. The possibilities of classical models of neural networks and the cognitive machines created on their basis do not go far beyond the recognition and classification of patterns [14].

The most promising models include artificial recurrent neural networks (RNNs) of real time. Unfortunately, the depth of information processing in traditional RNNs of real time is not high.

In recent years RNNs with controlled elements have been developed [15-19], on the basis of which perspective cognitive neural network machines can be created. These networks can be endowed with transparent logical structures and provide not only fast, but also deep processing of heterogeneous information.

Despite the results obtained, many issues related to the structural properties and capabilities of CNMs remain undeveloped. It is necessary to clarify the structural aspects of such machines and to determine their basic functions, allowing to provide expanded opportunities for solving creative tasks.

The article specifies the structure of a cognitive neural network machine based on a recurrent neural network with controlled elements. The system of extended basic functions of such a machine, which allows solving many creative tasks, is formed and discussed.

2 Structure of CNM

The basis of such a machine can be a two-layer pulse recurrent neural network with controlled elements, as well as blocks of forward and reverse transformation of signals (Fig. 1).

![Fig. 1. Generalized structure of CNM.](image-url)

The direct transformation of signals can involve first decomposing them into space-frequency components. Then each component of the signal is converted into a sequence of single pulses (SP) with a frequency and a phase, as functions of the amplitude and phase of the component. The frequencies of the components are

* Corresponding author: osipov_vasiliy@mail.ru
assigned to the numbers of the generated pulse sequences. As a result, the input of the neural network is given by successive single pulse populations (SPPs), carrying all the information about the input signals. At the output of the neural network, the resulting signals are formed from output SPPs. As part of the cognitive machine, it is proposed to use a recurrent neural network with a refined structure, reflected in Figure 2.

Fig. 2. Recurrent neural network with controlled elements.

The peculiarity of the RNN is that many functions for providing cognitive processing of information are assigned to the control unit. In this case, control actions can affect not only the characteristics of the synapses, but the capabilities of neurons. In this network, each neuron of one layer is in general connected with all the neurons of the interacting layer. Neurons of the same layer of bonds do not. This limitation can be justified by the need to minimize possible conflicts between neurons. The pulse delay time in the formed two-layer network circuits is less than the refractory time of neurons after excitation.

Fig. 3. Detailed structure of the recurrent neural network: neurons are indicated by circles, single delays - by squares, memristors - by fat dots, and control actions - by input small arrows.

At the level of individual neurons, single delays and synapses, such a neural network can be represented in the form of two controlled identical circuits, Fig. 3. The basis of these schemes are the controlled memristor crosbars, columns of neurons and single delays. In such a network, neuron layers are considered as two-dimensional structures. Moreover, due to the spatial shifts of SPPs, when transferring them from layer to layer, the network can be endowed with various linear, spiral, and other logical structures. The unambiguity between the input and output in the RNN is ensured by the priority of short bonds with the lowest attenuation. In addition, in such a network it is possible to rotate the processed SPPs and change the parameters of the divergence of the single pulses themselves. Changing these and other parameters depending on the current states of the layers can significantly affect on signal processing.

Fig. 4 shows one example of the space-time structure of RNN with three signaling systems, the basis of the CNM. Strengthening and weakening of interactions between signal systems is feasible by controlling the spatial parameters of single pulses diverging in the network. The same way it is feasible to form and interrupt the processing cycles of signals in the network, switching from one solved task to another.

Fig. 4. Space-time structure of the RNN with three signaling systems: 1, 3, 4 — directions of forwarding SPPs along the layers and between them; 2 — lines, breaking down the layers into logical fields due to SPP shifts along the layers; 5 — neurons.

The CNMs created on the basis of such RNN can have wide functional capabilities. These machines are endowed with additional basic functions associated with controlling the spatial and temporal parameters of the signals processed in the network. Let us define the expanded system of basic functions of the cognitive machine below.

3 Basic functions

When forming a system of such functions, we will proceed from possible actions on signals in CNM. If we exclude from them the actions for direct and inverse transformation of signals, it all comes down to the functions realized by a recurrent neural network with controlled elements. Analyzing the capabilities of the
The realization of F1 can be carried out with significant changes in the space-time relations between single pulses both within a SPP and their sequences. The greatest changes in such relationships are characteristic of spiral structures, and the smallest ones are linear RNNs. The current level of memorizing the results of signal binding depends on many factors. It is affected by the spatial and temporal distance between signals, their surroundings, the weights of the involved synapses, the total network load, and other factors.

Table 1. Basic functions of signal processing

<table>
<thead>
<tr>
<th>Function Symbol</th>
<th>Function Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>Promotion of SPPs over the network and the binding of these pulses.</td>
</tr>
<tr>
<td>F2</td>
<td>Storage of binding results.</td>
</tr>
<tr>
<td>F3</td>
<td>Acyclic and cyclic associative extraction of signals from the RNN memory with current signals.</td>
</tr>
<tr>
<td>F4</td>
<td>Suppression of false single pulses in signals propagated through the network, taking into account the previously stored information.</td>
</tr>
<tr>
<td>F5</td>
<td>Amplification and attenuation of signals.</td>
</tr>
<tr>
<td>F6</td>
<td>Spatial compression and expansion of SPPs when transmitting them from layer to layer.</td>
</tr>
<tr>
<td>F7</td>
<td>Rotations of such SPPs in the space of layers.</td>
</tr>
<tr>
<td>F8</td>
<td>Shifts of SPPs along the layers of the network.</td>
</tr>
<tr>
<td>F9</td>
<td>The change in the divergence parameters (shapes and directionality of the cross sections) of the single pulses themselves when they are transmitted from layer to layer.</td>
</tr>
<tr>
<td>F10</td>
<td>Compression and stretching of signals in time.</td>
</tr>
<tr>
<td>F11</td>
<td>Signal shifts in time.</td>
</tr>
<tr>
<td>F12</td>
<td>Space convolutions of SPPs.</td>
</tr>
<tr>
<td>F14</td>
<td>Time convolutions of complex signal constructs to simpler ones.</td>
</tr>
</tbody>
</table>

According to F3, acyclic and cyclic associative extractions of signals from different memory areas can be simultaneously performed in the network. The type and power of recalling information from memory largely depend on its current direction, the number of pulses in RNN. Suppression of false single pulses according to F4 is feasible at low levels of signal amplification in the network. Rotations (F6) of the SPPs in the RNN along with other functions make it possible to expand the space of possible states of signal processing. An important place in cognitive processing is occupied by the function F9, which provides for an operative change in the spatial orientation of the associative interaction of signals in the RNN.

Based on the F14, complex sequences of actions on signals at different space-time levels of their representation can be formed. Some of the behavioral programs can be hard-coded into the memory of a cognitive machine when it is developed. At the same time, it is provided opportunities for self-learning based on these programs in the process of accumulating experience.

The main creative tasks that can potentially be solved when implementing the functions F1 - F14 in CNMs based on large RNN with controlled elements include the following:
1. The tasks of filtering and the reconstruction of distorted signals.
2. Problems of classification and recognition of dynamic one-, two- and three-dimensional signals.
3. Tasks of adaptive forecasting of various events in the interest of autonomous robots.
4. The tasks of looking to the past and analyzing the history of events.
5. The tasks of transforming some signals into others, including associative translation of speech from one language to another.
6. Creative synthesis and realization of the behavior programs of the cognitive machine itself and autonomous robot.

The list of such tasks can be significantly expanded when correlated with specific areas of application of CNMs.

4 Modelling results

By modeling, we studied empowering CNMs with advanced functions. Two-layer recurrent neural networks with linear and spiral space-time structures were implemented programmatically. The size of the layers did not exceed seven thousand neurons, and the number of synapses is ranged from one 1 to nine million. The possibilities of the RNN with various space-time structures and functions were compared. In particular, the two-layer RNN with 2016 neurons in each layer was studied. At the expense of spatial shifts of SPPs, each layer of the network was divided into six lines of 8 logical fields, 6 × 7 neurons each. Subsequent SPPs input to RNN, moved along the layers a spiral way during the signal processing. When transferring SPPs from layer to layer the parameters of the spatial divergence of single pulses changed on the current states of layers. Through this, the RNN generated controlled signal processing cycles.

Examples of smoothed spatial distribution of relative charges carried by currents in a diverging beam without and with taking into account the individual difference in potentials at the synapses are provided in Figure 5. Analysis of Figure 5 shows that the current states of neurons in the receiving layer may substantially alter the charge distribution in the divergent signal beams. Depending on the current states of RNN layers the same input signal retrieves different associated signals from the network memory.

Successful formation of cycles of associative call of information from the RNN memory for this associative and spatial addressing is realized when there are regions of the unexcited neurons after the inducing signals. Otherwise, input signals retrieve information from the memory which must precede them. To deeper process
input signals, pauses are necessary between them, or deep processing should be temporarily separated from shallow one.

Figure 6 shows how, depending on the current states of layers, which change the spatial parameters of divergent signal beams, transmitted from layer to layer, the mode of functioning of the network is changed. In Fig. 6a, the RNN cyclically retrieves stored signals. The dips on the curve in Figure 6a reflect the fact that SPPs leave the RNN.

When new SPPs changing the RNN state come to the RNN input, the cyclic information retrieval is interrupted (Fig. 6b). Note that RNN capabilities on associative information retrieval depend not only on the number of training samples, but on the threshold of neuron excitation as well. The experiments showed that under certain conditions even with a single training sequence of SSPs the neural network is capable to associative retrieval of related signals.

It was confirmed that the more possibilities for changing the space-time and energy relationships between the signals in the RNN, the wider the range of creative tasks that CNM can potentially solve. For example, controlling only the directionality of the cross-sectional shapes of divergent pulses, depending on the current states of the layers, it was possible to organize and interrupt the processing cycles of the RNN signals in a wide range. By controlling only internal time, the RNN can adaptively predict possible events.

5 Conclusion

As a result of the study it can be argued that implementation of the CNMs with the proposed functionality can significantly increase their space of states and the capabilities of cognitive information processing. Due to additional space-time processing functions and depending on the current state of layers, more selective binding and storing of signals becomes feasible. More elaborated models of processed events are formed and stored in the network due to this mechanism. Opportunities for associative information retrieval from memory are improved as well. In contrast to the known solutions, the extended space-time processing makes available mechanisms to manage interactions of RNN subsystems, the memory load, the processing level, switching from one task to another.

Associative and spatial addressing to memory is proposed to be realized through rapid changes of spatial parameters of the signals being processed. Note that technically this addressing in RNN can also be
accomplished by changing the attenuation functions of the diverging signals. Spatial shifts of the processed signals along the RNN layers can be justified by slow changes of the spatial parameters of the elements of the network itself because of minimizing potential conflicts among neurons.

The considered new solutions are focused at analog implementation of RNNs with memristors. Analog implementation allows overcoming the complexity problem of creating large RNNs with the number of artificial neurons and synapses, comparable to characteristics of biological systems.

The potential possibilities of the studied cognitive machine allow solving not only the problems of classification and recognition of various signals. This machine can, on the basis of experience, carry out an operative search for information of interest, adaptively analyze current and predict future events, synthesize and implement complex programs of various actions. Undoubtedly, when creating such machines in hardware, work will be required to provide the necessary experience. Later, this experience can be easily transferred from one machine to another without significant costs.

In the coming years, when creating small-size cognitive machines based on large RNNs with controlled elements, it is possible to go far beyond the existing limitations of cognitive processing of information.

The proposed approach for space-time processing of signals may be used in the development of promising small cognitive analog neural machines.

References

15. V. Osipov, M. Osipova, Method and device of intellectual processing of information in neural network, RU Patent 2413304 (2011)
18. V. Osipov, Advances in Neural Networks, LNCS 9719, 177-184 (2016)