Recognition of stimulus received by recurrent neural network

S I Bartsev1,2,3 and G M Markova2

1 Institute of Biophysics SB RAS, Federal Research Center, Krasnoyarsk Scientific Center SB RAS, 50, Akademgorodok, Krasnoyarsk, 660036, Russian Federation
2 Siberian Federal University, 79, Svobodny pr., Krasnoyarsk, 660041, Russian Federation
3 E-mail: bartsev@yandex.ru

Abstract. The study is concerned with the comparison of two methods for identification of stimulus received by artificial neural network using neural activity pattern that corresponds to the period of storing information about this stimulus in the working memory. We used simple recurrent neural networks learned to pass the delayed matching-to-sample test. Neural activity was detected at the period of pause between receiving stimuli. The analysis of neural excitation patterns showed that neural networks encoded variables that were relevant for the task during the delayed matching-to-sample test, and their activity patterns were dynamic. The method of centroids allowed identifying the type of the received stimuli with efficiency up to 75% while the method of neural network-based decoder showed 100% efficiency. In addition, this method was applied to determine the minimal set of neurons whose activity was the most significant for stimulus recognition.

1. Introduction

It is known from neurophysiological studies that the information relevant to the current problem solving stores in the working memory in a dynamic form [1-4]. Patterns of neural activity that match certain content are highly variable. Decoding information processed by the brain using corresponding neural activity patterns is an urgent task of cognitive neuroscience. The neural correlates concept [5] is a suitable framework for this task since the possible solution is the identification of the minimal set of neurons or minimal neural activity that corresponds to the certain content. It is required to assess whether patterns related to the same content are reproducible.

The question of whether relevant information about the stimulus received by neural network stores in a static or dynamic form had been investigated on artificial recurrent neural networks (RNN) [6]. Studies showed that encoding information about an input stimulus received during the delayed matching-to-sample (DMTS) test is dynamic. Therefore features identified using biological neural networks were reproduced on artificial neural networks, so RNN can be used for modeling of neural activity.

It is important to note that the task assigned to RNN in the study [6] is simplified. The essence of this simplification is the following: since the duration of pause between first and second stimulus is fixed (200 ms), the way of how RNN stores the representation of first stimulus doesn’t matter. If, under these conditions, the patterns of excitation of RNN neurons are represented as movement along non-intersecting trajectories in the space of neural activity [1], then these trajectories can be located in any...
way. The only requirement for RNN is to achieve the necessary point in the neural activity space by the time the second stimulus arrives.

The situation becomes far more complicated and close to biology if the duration of pause before the receipt of the second stimulus is chosen randomly from a given interval. In this case, the question of how RNN stores the information about the stimulus that must be available for use during the entire pause becomes more acute. Some possible answers to this question are presented in the study [1]: the information stores in static excitation pattern, that makes its use easier, or in dynamic pattern, which is more plausible in terms of neurophysiology. In the second case, the question arises: if the patterns of neuronal excitation are very changeable, how should an internal decoder be arranged for adequate responding?

The aim of this study is to compare the effectiveness of two methods for identification of stimulus received by simple RNN using its neural activity pattern during the time period, when RNN stores the information about this stimulus in working memory. We provide a comparison of centroid and neural network-based decoding methods.

2. Materials and methods

We used simple RNN with two inputs, two output neurons and varied number of internal neurons (25, 30). RNN’s response $y_o^{(t)}$ at time $t$ is calculated as:

$$
\begin{align*}
y_h^{(t)} &= f_h\left(W_h \cdot y_h^{(t-1)} + W_i \cdot x^{(t)}\right) \\
y_o^{(t)} &= f_o\left(W_o \cdot y_h^{(t)}\right)
\end{align*}
$$

(1)

where $W_h$, $W_i$, $W_o$ are matrices of weight coefficients of internal neurons, inputs and output neurons, respectively; $x^{(t)}$ – vector of input signals at time $t$; vectors $y_h^{(t)}$ and $y_h^{(t-1)}$ are excitation levels of internal neurons at times $t$ and $t-1$. Activation functions of internal and output neurons are presented as:

$$
\begin{align*}
f_h(x) &= \frac{1}{2} \left( x + \frac{x}{a + |x|} + 1 \right), \\
f_o(x) &= \begin{cases} 
0, & \text{if } x \leq 0 \\
0 \cdot x, & \text{if } x > 0 \text{ & } x < 1 \\
1, & \text{if } x \geq 1
\end{cases}
\end{align*}
$$

(2, 3)

Training of RNN was carried out using the back propagation algorithm with propagation depth equal to 6 clock cycles (taks). In this case, the specific type of learning algorithm does not matter, since it is supposed to consider neural activity of the trained RNNs, the structure of which is formed regardless of the training method [7].

A task that RNNs were trained to solve was DMTS test. Firstly, one of three randomly selected stimuli (input vectors) entered the input: $A = (01)$, $B = (10)$, $C = (11)$. After a pause, duration of which varied randomly from 3 to 6 taks, a second stimulus, also randomly selected, entered the input. At the third step after the presentation of the second stimulus RNN produced a signal (10) on the output neurons if these stimuli coincided and a signal (01) otherwise.

To recognize received stimuli using neuronal excitation patterns we used the centroid method [8]. We computed the average level of excitation of every RNN neuron corresponding to each type of stimulus (A, B, C) during the storing period. Then we measured the total distance between this level and every neuron’s values of excitation at different time steps while RNN stored information about certain stimulus. A type of the stimulus received by RNN was determined by minimal value of this distance.
For neural network-based decoding method we applied single-layer neural networks (DN) consisting of three neurons with a linear activation function (3) connected with each of the inputs. The number of inputs was equal to the number of RNN’s neurons, the neural activity of which was used for decoding. The input data for DN were the excitation levels of RNN’s neurons taken from the pause period between the presentations of stimuli (3-6 takts inclusive). For DN’s training we used the back propagation algorithm.

3. Results and discussion
RNNs of the described configuration learned to pass DMTS test. Since the duration of the pause between receiving of first and second stimulus varied, RNNs stored information about the first received stimulus, being ready to use it at any possible moment when the second stimulus appears. Thus the representation of the identity of the first stimulus during the pause was formed in the RNN’s neural activity patterns in form of a dynamic invariant. Neuronal excitation patterns of RNN that consists of 25 neurons are shown in figure 1. Visual acquaintance with neural activity dynamics let us to identify high variability in excitation patterns during the interval between stimulus arrivals and lack of obvious signs of static patterns.

![Figure 1](image1.png)

**Figure 1.** Examples of dynamic patterns of RNN that consists of 25 internal neurons. To demonstrate the dynamics, duration of pause between stimulus arrivals was fixed at 6 takts (clock cycles). Vertical lines indicate (from left to right): receipt of the first stimulus, receipt of the second stimulus, RNN response. Letters in the upper left corner are pairs of received stimuli.

In previous studies we demonstrated on RNNs passing DMTS test that there is no simple correlation between the RNN excitation pattern and the type of stimulus, information about which is stored in neural activity [9]. This result was obtained by calculating mean squared error between the level of each neuron excitation at successive times for a certain stimuli and between the level of excitation of each neuron at same moments of time, but with different stimuli.

We consider the centroid method [8] as a possible way for identifying dynamic invariants of neural activity. The result of decoding stimuli received by three RNNs with 25 neurons and three RNNs with 30 neurons is presented in the table 1. Here and below, when recording neural activity, we considered the levels of neuronal excitation of RNNs trained to pass DMTS test with a random duration of pause between the stimuli which correspond to 4 takts before the presentation of the second stimulus.

| Table 1. Application of the centroid method for identification of the stimulus received by RNNs with 25 and 30 neurons. |
As shown in the table 1, on average the efficiency of the centroid method for identification of the stimulus received by RNN using its neural activity is higher for RNN with 25 neurons (75%) than with 30 neurons (50%).

Application of this method to randomly generated data instead of RNN’s neural activity showed 25-50% “coincidences”. So, the efficiency of the centroid method when considering real data on neural activity of RNNs with 30 neurons is comparable to that for random set of numbers.

Low accuracy of identification can be explained by high variability of signal that encodes stimuli and also by the fact that the centroid method does not take into account the individual dynamics of neurons. For a more accurate recognition, we need a method which allows to determine the dynamic invariant of the received stimulus representation and to consider individual neuronal dynamics. The neural network-based decoding method meets these requirements.

This method showed 100% efficiency: DNs learned to recognize stimuli, the information about which was stored in RNN’s neural activity patterns, unmistakably for both RNNs with 25 and 30 neurons. Also, the method allowed identifying a set of neurons, whose excitation levels were the most significant for the stimulus recognition. To achieve this, the structure of DN was reduced as follows: synapses with the minimal absolute values that connected DN’s neurons to one of the outputs were equated to zero. It was followed by additional training of DN to achieve the initial quality of functioning. We repeated this procedure until the minimal possible amount of inputs with nonzero synapses was determined, at which it was still possible for DN to operate with the required efficiency.

We use the neural activity data of three RNNs with 25 neurons and three RNNs with 30 neurons; three DNs were trained for each of RNNs. Using the reduction procedure, we identified $7 \pm 1$ significant neurons for RNNs with 25 neurons and $9 \pm 2$ significant neurons for RNNs with 30 neurons.

To illustrate the task that DN had to solve, we provide a picture of excitation levels of the most significant neurons of RNN with 30 neurons, corresponding to information storing about stimulus type A and B. This neural activity is shown in figure 2.

| Number of takt | Input stimulus | Result of decoding     |
|----------------|----------------|------------------------|
|                |                | **25 neurons** | **30 neurons** |
|                |                | RNN 1   | RNN 2   | RNN 3   | RNN 1   | RNN 2   | RNN 3   |
| 3              | A              | A       | A       | A       | B       | B       | A       |
| 4              | A              | B       | B       | A       | A       | B       | A       |
| 5              | A              | A       | A       | A       | A       | C       | A       |
| 6              | A              | A       | A       | A       | A       | C       | B       |
| 3              | B              | B       | B       | B       | B       | B       | A       |
| 4              | B              | C       | C       | B       | C       | B       | A       |
| 5              | B              | B       | B       | B       | A       | C       | B       |
| 6              | B              | B       | B       | B       | A       | C       | B       |
|                | C              | B       | B       | C       | B       | B       | A       |
| 4              | C              | C       | C       | A       | C       | C       | C       |
| 5              | C              | C       | C       | C       | C       | C       | C       |
| 6              | C              | B       | B       | C       | A       | C       | B       |

Accuracy       67%  67%  92%  50%  42%  58%

As shown in the table 1, on average the efficiency of the centroid method for identification of the stimulus received by RNN using its neural activity is higher for RNN with 25 neurons (75%) than with 30 neurons (50%).
Figure 2. Activity of the neurons of RNN with 30 neurons, which were selected by DN as the most significant for decoding stimuli A and B. The type of the received stimulus is shown in parentheses. Captions in the legend are ordinal numbers of RNN neurons.

4. Conclusion
The method of neural network-based decoder, suggested in the current study, allows identifying the certain type of the stimulus received by recurrent neural network and stored in the dynamic pattern of its neural activity. Using the method of centroids we managed to recognize, on average, 75% of stimuli received by neural networks with 25 neurons and 50% – by networks with 30 neurons. At the same time, the efficiency of the described neural network-based method was 100% for networks of both sizes. The advantage of this method is the possibility to account for individual excitation patterns of every neuron of the original neural network. Moreover, this method allows determining the minimum required subset of neurons that provide encoding of the received stimulus. Focusing on identifying the minimal amount of neurons, whose activity underlies a certain phenomenon, is typical for the concept of neural correlates [5], so the suggested method can be referred to this concept.

Acknowledgements
The research was funded by RFBR, Krasnoyarsk Territory and Krasnoyarsk Regional Fund of Science, project number 20-41-240003.

References
[1] Meyers E M 2018 J. Neurophysiol. 120(5) 2260-8
[2] Meyers E M, Freedman D J, Kreiman G, Miller E K and Poggio T 2008 J. Neurophysiol. 100(3) 1407-19
[3] Barak O, Tsodyks M and Romo R 2010 J. Neurosci. 30 9424-30
[4] Stokes M G, Kusunoki M, Sigala N, Nili H, Gaffan D and Duncan J 2013 Neuron 78(2) 364-75
[5] Crick F and Koch C 2003 Nat. Neurosci. 6(2) 119-26
[6] Miconi T 2017 Biologically plausible learning in recurrent neural networks reproduces neural dynamics observed during cognitive tasks Elife 6 e20899
[7] Bartsev S I and Bartseva O D 2002 Dokl. Biochem. Biophys. 386 235-8
[8] Crowe D A, Averbeck B B and Chafee M V 2010 J. Neurosci. 30(35) 11640-53
[9] Bartsev S, Baturina P, Markova G 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1047 012078