Nonlinear Dynamics and Neuroscience

Izvestiya Vysshikh Uchebnykh Zavedeniy. Applied Nonlinear Dynamics. 2024;32(5)

Article DOI: 10.18500/0869-6632-003121

A spiking binary neuron — detector of causal links*

M. V. Kiselev[1]✉, *D. A. Larionov*[1,2], *A. M. Urusov*[1,3]

[1]I. N. Ulianov Chuvash State University, Cheboksary, Russia
[2]Cifrum, subsidiary of Rosatom, Moscow, Russia
[3]Gazprombank (JSC), Moscow, Russia
E-mail: ✉mkiselev1@yandex.ru, denis.larionov@gmail.com, imurusov@mail.ru

Received 31.10.2023, accepted 11.04.2024, available online 10.09.2024, published 30.09.2024

Abstract. *Purpose.* Causal relationship recognition is a fundamental operation in neural networks aimed at learning behavior, action planning, and inferring external world dynamics. This operation is particularly crucial for reinforcement learning (RL). In the context of spiking neural networks (SNNs), events are represented as spikes emitted by network neurons or input nodes. Detecting causal relationships within these events is essential for effective RL implementation. *Methods.* This research paper presents a novel approach to realize causal relationship recognition using a simple spiking binary neuron. The proposed method leverages specially designed synaptic plasticity rules, which are both straightforward and efficient. Notably, our approach accounts for the temporal aspects of detected causal links and accommodates the representation of spiking signals as single spikes or tight spike sequences (bursts), as observed in biological brains. Furthermore, this study places a strong emphasis on the hardware-friendliness of the proposed models, ensuring their efficient implementation on modern and future neuroprocessors. *Results.* Being compared with precise machine learning techniques, such as decision tree algorithms and convolutional neural networks, our neuron demonstrates satisfactory accuracy despite its simplicity. *Conclusion.* We introduce a multi-neuron structure capable of operating in more complex environments with enhanced accuracy, making it a promising candidate for the advancement of RL applications in SNNs.

Keywords: spiking neural network, binary neuron, spike timing dependent plasticity, dopamine-modulated plasticity, anti-Hebbian plasticity, reinforcement learning, neuromorphic hardware.

For citation: Kiselev MV, Larionov DA, Urusov AM. A spiking binary neuron — detector of causal links. Izvestiya VUZ. Applied Nonlinear Dynamics. 2024;32(5):589–605. DOI: 10.18500/0869-6632-003121

Introduction

If we aim to develop an intelligent system based on neural networks capable of learning adaptive behavior towards specific objectives, it is imperative to equip it with the capacity to identify and incorporate causal relationships between events occurring both within the network and in the external environment. These relationships may encompass consistent sequences of patterns forming a unified spatiotemporal pattern, directives to actuators generated by the

*The paper presents materials of a talk given at the conference “Neuroinformatics — 2023”.

network, and subsequent responses from the external environment, as well as events preceding a reward and the reward itself. Therefore, the capability to discern causes and effects should constitute a foundational functionality within network structures or individual neurons. In most scenarios, the learning network lacks access to a priori knowledge outlining causal relationships in its environment. Instead, it must deduce them from observed temporal alignments of different events, under the assumption that if event B is frequently witnessed within a specific time frame following event A, then A serves as the cause and B as the consequence.

In this research, we investigate the integration of this particular functionality within spiking neural networks (SNNs), specifically within an individual neuron embedded in the SNN. In the context of SNN, information is encoded through sequences of spikes, making it necessary to frame this problem within the spike domain. We shall formally define the task to be addressed by this neuron, referred to as the causal relationship detector. The neuron receives spikes from a group of other neurons. It is postulated that the concurrent activation of an unidentified subset of these presynaptic neurons signifies event A (the cause), while the activation of a distinct neuron indicates event B (the consequence). Given the assumption that event B consistently or frequently occurs within a time window of T_p following event A, the neuron's objective is to identify the specific presynaptic neurons belonging to this subset and fire whenever event A occurs. Similar to other learning challenges encountered in SNNs, this problem is resolved by introducing appropriate adjustments to the synaptic weights of the neuron-detector. Importantly, and notably complicating the matter, both the presynaptic neurons and the neuron-detector itself may produce not only individual isolated spikes but also tightly clustered and prolonged spike sequences. This aspect introduces a significant challenge, as it limits the applicability of the majority of existing synaptic plasticity models that rely on the relative timing of individual pre- and post-synaptic spikes.

Numerous studies have demonstrated how SNN can express causal relationships between observations. However, the present research offers a unique combination of three distinctive attributes:

1. Causal relationships are identified by a single neuron, rather than relying on a network-wide approach.
2. The temporal dimension of causal relationships is emphasized, with events-causes occurring certain time before their corresponding consequences.
3. Specially designed local synaptic plasticity rules are used for learning.

The majority of works consider this problem without a temporal aspect. In this context, the problem is often called Bayesian inference or causal inference. It has a close relation to supervised learning, where the network should determine which factors, either independently or in combination, most reliably indicate that the given object belongs to the target class. Recent examples of such research taking various approaches to tackle causal/Bayesian inference problems can be found in [1–3]. It is noteworthy that one of the primary objectives of these works is supervised learning, which does not explicitly incorporate time. Even when dealing with time-series data, the mentioned works treat the entire time series as a single entity, assigning a label to the entire sequence. In contrast, our research aligns more closely with reinforcement learning where all the signals (input signals, network commands, and reward/punishment) exist within the continuous time where time delays carry significance.

Furthermore, our approach resonates with Friston's concepts of free energy [4], as understanding the "expected events" is essential for quantifying the "surprise" or unexpected events in terms of free energy. In essence, this research delves into the intricacies of recognizing causal relationships within the temporal context, which holds particular relevance for dynamic, time-dependent systems.

It's worth noting that our research problem bears resemblance to another significant area of interest among machine learning and neural network researchers, namely, time-series forecasting. Time-series forecasting techniques aim to predict the future values of specific variables (whether discrete or continuous) based on their current and recent values, and possibly those of other related variables. Naturally, if we can identify causal relationships between the values of certain variables or the states of an object and the values of specific parameters in the future, it equips us with a tool for predicting these future values. However, our primary objective differs from traditional time series forecasting, as we are not focused on predicting the exact value of a particular variable at an exact point in time. Instead, our aim is to infer causal rules that indicate that after event A, it is expected that event B will occur within a time interval of length T_p. Consequently, our problem can be more appropriately characterized as future event prediction.

Remarkably, there are relatively few applications of SNNs to this problem. One of such approaches is described in [5]. It utilizes the NeuCube system [6], which is centered around the Liquid State Machine (LSM) [7]. The LSM is a large, chaotic, non-plastic SNN designed to transform a combination of time-series data and static (or slowly changing) parameters into a high-dimensional representation, represented by the current firing frequencies of neurons within the LSM. Due to the great number of neurons in the LSM, representations of various spatiotemporal patterns in the form of LSM neuronal activity tend to be linearly separable. Consequently, classification problems related to such representations can be efficiently solved using simple linear classifiers. As described in [5], several instances of NeuCube's application to rare event forecasting have been demonstrated. One specific example, the prediction of strokes, is examined in greater depth in [8]. While the LSM-based approach has demonstrated success across a wide range of tasks, it has a notable drawback in that it requires a large, computationally intensive LSM to achieve efficiency. In contrast, our approach effectively addresses a similar problem using just a single neuron, offering a more resource-efficient solution.

Paper [9] shows how special SNN structures can be used to infer the graph of causal relationships, but again without a temporal aspect as mentioned before.

At last, there exists another branch of SNN research closely aligned with our study. As described below, we employ a combination of two synaptic plasticity models commonly referred to, albeit in a very approximate sense, as Hebbian plasticity and dopamine plasticity to address the problem of identifying causal links. Hebbian plasticity, when applied to the plasticity of spiking neurons, is often denoted as the STDP plasticity model (Spike Timing Dependent Plasticity) [10], while dopamine plasticity is typically associated with reward-related effects. The collective terminology for these merged plasticity models is R-STDP (Reward Spike Timing Dependent Plasticity). Numerous works have explored and proposed various R-STDP models [11–14], with some of them having already undergone testing in real-world applications [15]. Presently, there is no universally accepted consensus on how to best combine these two types of synaptic plasticity, resulting in a wide array of proposed models. Moreover, unlike our approach where reward takes the form of a spike signal, these models typically represent reward as a global real-valued variable. To the best of our knowledge, no prior work has employed similar synaptic plasticity rules for the specific purpose of detecting causal links.

Furthermore, our specific objective was to craft a simplified plasticity model to facilitate efficient implementation on contemporary and forthcoming neuroprocessors. According to work [16], there is a clear trend in the emergence and development of software and hardware systems that allow not only the inference of convolutional neural networks converted into SNN form, but also the use of arbitrary models of spiking neurons and plasticity rules, opening up the possibility of continuous learning.

In the subsequent section, we will provide a detailed description of our innovative synaptic plasticity model, which combines Hebbian (effectively, anti-Hebbian) and dopamine plasticity.

Following this, we will consider the application of this model to the problem of reward prediction in reinforcement learning, using as an example the task of reward prediction in the "ping-pong" RL task. In conclusion, we will outline our vision of how neurons of this kind can form network structures capable of inferring complex graphs of causal relationships essential for constructing external world models in RL. Subsequently, we will assess the advantages and limitations of our approach and outline future research plans in this direction.

1. Methods and materials

In our research, we investigate the learning process of a single spiking neuron in the context of identifying causal relationships among events. This neuron (Fig. 1) is connected to a group of presynaptic neurons denoted as set C, whose activity is presumed to represent various events in a general sense. In our study, we interpret these events as potential triggers for another event, which we will refer to as the "target event". This target event is signaled by spikes from a separate presynaptic neuron denoted as S, which is not a part of set C. The time of the j-th spike generation by the i-th presynaptic neuron from the set C is denoted by t_{ij}. The moments when the neuron S fires will be denoted as T_i. We say that some event is a cause of the target event if the target event is frequently observed not later than the time T_p after that event. As we have mentioned, these possible causes of the target event are indicated by the specific activity of the presynaptic neurons, which the learning neuron should try to recognize. T_p is the temporal constant fixing time scale of the concrete domain or task. It is assumed that the target events are rare — it means that T_p is much less than the minimum value of interspike intervals $T_i - T_{i-1}$. It is the only important assumption — without it, our problem of finding causal relationships seems to become senseless.

We introduce the concept of a "target period", which encompasses the time interval of length T_p preceding each T_i . The learning neuron should mark the target periods by its activity (spikes emitted by it at the times T_i^*). If it learns to do it with sufficient accuracy, it signifies that it has successfully recognized the causal link between a specific event (corresponding to the distinct activity of a presynaptic neuron that triggers the neuron to fire) and the target event.

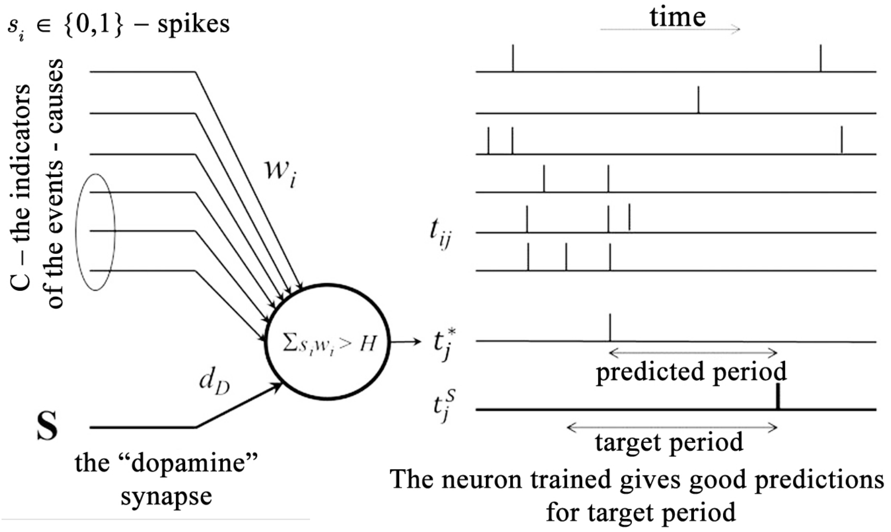
$s_i \in \{0,1\}$ – spikes
C – the indicators of the events - causes
w_i
$\Sigma s_i w_i > H$
t_j^*
d_D
S
the "dopamine" synapse
time
t_{ij}
predicted period
t_j^S
target period
The neuron trained gives good predictions for target period

Fig 1. The schematic model of binary spiking neuron and the temporal diagram of its pre- and postsynaptic spikes (for the trained state)

To assess the accuracy of this recognition, we introduce the concept of a "prediction period". Each prediction period begins at some moment T_i^* and ends either after time T_p from that moment or at some moment T_i — whichever occurs first. The total duration of time t_{err}, during which the target period and prediction period do not overlap, serves as a natural metric for measuring the inaccuracy in predicting the target event. The learning neuron's objective is to maximize the metric represented by the equation:

$$R = 1 - \frac{t_{\text{err}}}{t_{\text{tar}}}, \tag{1}$$

where t_{tar} denotes the total duration of the target periods.

In our study, we use the simplest neuron model called "binary neuron". This neuron operates in discrete time. In each time quantum, it receives spikes through its plastic synapses with weights w_i. It is assumed that the size of these quanta is of the order of the possible mismatch time of the input spikes from the set C indicating the causal event, so that a significant part of these spikes arrive within one quantum. If the sum of the weights of the synapses that received spikes in a given quantum is greater than the threshold value H, the neuron fires. This choice makes our result general — in fact, it does not depend on the concrete neuron model. After the respective time discretization, any neuron model can be approximated by the binary neuron which retains the main spiking neuron property — neuron fires when several strong excitatory synapses obtain spikes during short time period. In order to make the weights w_i dimensionless, we set $H = 1$.

1.1. General idea of the method and rules of synaptic plasticity used in the study. We assume that information about potential causal events leading to the target event is encoded within spikes originating from the presynaptic neurons in set C. The synapses responsible for transmitting these spikes are plastic, and their synaptic weights should be modified in such a way that would force the neuron to fire during the target period.

Activity of the learning neuron and modification of its synaptic weights should be interrelated in the following way:

a. Untrained neuron should be inactive — the plasticity mechanism should strengthen those synapses which would force the neurons to fire in the correct time. For this reason, we set weights of all plastic synapses to zero at start.
b. If the neuron fires in the wrong time (outside of target periods) then the synapses which helped it to fire should be depressed.
c. If the neuron fires in the correct time then nothing should be done with its synapses — otherwise modifying its synaptic weights can move it from a trained state.

These behaviors are achieved through the specific design of synaptic plasticity rules. It is important that properties of the plastic synapses are totally different from the single synapse through which the presynaptic neuron S is connected. We will call it the "dopamine" synapse" because spikes coming to it control the plasticity of all other synapses.

The principles A, B, and C specified above are fulfilled due to the following combination of the two plasticity rules:

1. Dopamine plasticity. Every time the neuron receives a spike from the S neuron all the plastic synapses having obtained a spike during the time T_P before this "dopamine" spike are potentiated.
2. Anti-Hebbian plasticity. All synapses contributing to neuron firing are depressed.

It is evident that the equilibrium between dopamine plasticity and anti-Hebbian plasticity, when properly balanced, satisfies the conditions A, B and C, ensuring the successful training and functioning of the neuron.

1.2. Model of synaptic plasticity in detail. Similar to our previous studies [17, 18], the rules of synaptic plasticity used in this paper are additive and apply to a variable called "synaptic resource", denoted by W, rather than directly to the synaptic weight, denoted by w. The functional relationship between W and w is expressed by the formula

$$w = w_{\min} + \frac{(w_{\max} - w_{\min}) \cdot \max(W, 0)}{w_{\max} - w_{\min} + \max(W, 0)}, \tag{2}$$

where $w_{\min}$ and $w_{\max}$ are constants. It is obvious that w values lay inside the range $[w_{\min}, w_{\max})$ — while W runs from $-\infty$ to $+\infty$. In this study, $w_{\min} < 0$, while $w_{\max} > 0$ so that synaptic plasticity can make excitatory synapse inhibitory and vice versa.

As mentioned earlier, the synaptic plasticity model consists of two separate and independent components. They are described in the following subsections.

1.2.1. Anti-Hebbian plasticity. The standard STDP (spike timing dependent plasticity) model [10] states that spikes coming short time before postsynaptic spike emission potentiate the synapses receiving them. This concept aligns with Donald Hebb's principle, which asserts that synaptic plasticity should encode causal relationships within neuron firings; in essence, synapses responsible for inducing neuron firing should be strengthened. This principle has been proven by plenty of neurophysiological observations. However, in-depth investigations into plasticity within biological neurons have revealed multiple instances of entirely distinct synaptic plasticity models existing in nature [19, 20]. Furthermore, examples of plasticity rules acting in the direction opposite to Hebbian principle (anti-Hebbian plasticity) have been observed in different organisms [21]. It makes us conclude that different kinds of synaptic plasticity are suitable for the solution of different problems. Moreover, the standard STDP model becomes senseless or self-contradictory in the case (which is quite common in the biological brain) when presynaptic and postsynaptic spikes are not stand alone in time but are packed in tight sequences (spike trains). In this case, it is senseless to say that presynaptic spike comes before or after postsynaptic spike because there are many postsynaptic spikes in the close neighborhood before and after the given presynaptic spike.

For these reasons, we have devised our own variant of the anti-Hebbian plasticity model and applied it to the problem at hand. Let us elaborate on this model.

As previously mentioned, weight modifications in the standard STDP are bound to single pre- and post-synaptic spikes. However, in the presence of spike trains, these rules lose their applicability. Consequently, in our model, the synaptic plasticity acts are bound to postsynaptic spike trains instead of single spikes. We refer to these spike trains as "tight spike sequences" (TSS). Specifically, taking the constant $ISI_{\max}$ as a measure of "tightness" of TSS, we define TSS as a sequence of spikes adhering to the following criteria:

1. There were no spikes during time $ISI_{\max}$ before the first spike in TSS.
2. Interspike intervals for all neighboring spikes in TSS are not greater than $ISI_{\max}$.
3. There are no spikes during time $ISI_{\max}$ after the last spike in TSS.

In this work, $\mathrm{ISI}_{\max}$ is set to the value of T_p.

Our anti-Hebbian plasticity model adheres to the following rules:

1. Resource of any synapse can change at most once during a single TSS. Here and below, TSS refers to postsynaptic spike train.
2. Resources of only those synapses are changed which receive at least one spike during TSS.

All synaptic resources change (decrease) by the same value d_H, regardless of the exact time of presynaptic impulses.

1.2.2. Dopamine plasticity. The described neuron has a plasticity-inducing synapse (from S). When it obtains a spike, synaptic resources of all plastic synapses of the neuron having received at least one presynaptic spike during the time interval of the length T_P in the past are modified (increased) by the same value d_D.

We emphasize that dopamine plasticity and anti-Hebbian plasticity are two completely logically independent mechanisms. Their combination gives the desired effect of neuron learning.

1.2.3. Neuron stability. In our model, the synaptic plasticity values d_H and d_D are not constant. Instead, they are dynamic and exhibit characteristics that adapt over learning. Initially, during the early stages of learning, these values need to be sufficiently large. However, for a well-trained neuron that consistently makes accurate predictions, they should approach zero. This adaptation is crucial to prevent further modifications to the neuron's synaptic weights, which could potentially disrupt its established trained state. To account for this adaptive behavior, we introduce an additional component s into the neuron's state, denoted as "stability". The synaptic plasticity decrease exponentially as the stability value s increases, according to the equations

$$d_H = \overline{d_H} \cdot \min(2^{-s}, 1), \quad d_D = \overline{d_D} \cdot \min(2^{-s}, 1). \tag{3}$$

Here $\overline{d_H}$ and $\overline{d_D}$ are constants of the neuron model. To balance anti-Hebbian and dopamine plasticity (which is necessary to satisfy condition C from subsection 1.1), we set $\overline{d_H} = \overline{d_D}$.

The stability value of a neuron changes in two cases:

1. It is decreased by the constant d_s every TSS.
2. It is adjusted by the value $d_s \cdot \max\left(2 - \frac{|t_{\mathrm{TSS}} - \mathrm{ISI}_{\max}|}{\mathrm{ISI}_{\max}}, -1\right)$ whenever a presynaptic dopamine spike is received. Here t_{TSS} is the time interval between the most recent TSS onset and the dopamine spike.

We see that if TSS began exactly the time $\mathrm{ISI}_{\max}(= T_p)$ ago before dopamine spike (i.e. target event) the neuron stability increase is maximum and is equal to d_s — if to take into account its decrease by d_s in accordance with rule 1. This corresponds to the most accurate prediction of the target event and serves as evidence that the neuron is well trained. Conversely, if a dopamine spike occurs when the neuron has remained inactive for an extended period, it suggests inadequate training, and as a result, its stability is decreased by d_s to facilitate further training.

1.3. Test task — the ping-pong game. The described technique holds great promise in the realm of reinforcement learning (RL). While supervised learning can be conceptualized as identifying causal relationships between predictors as causes and the target value as a consequence, RL tasks encompass a broader spectrum of causal link determination, explicitly incorporating the element of time. Reward signals come rarely and, possibly, with significant delay from the world states or the agent actions they evaluate. In order to overcome problems with insufficient evaluation signal frequency, the mechanism of intermediate goals should be utilized, but it is also based on inferring causal links.

Futhermore, the key point of the most comprehensive realization of RL, known as model-based RL, is creation by the agent of the internal model of world dynamics and world reactions to agent actions. The model creation mechanism unavoidably includes inference of a network of causal relationships between world state changes and agent actions. Thus, it is entirely reasonable to assert that causal link inference is the most basic operation in RL.

Keeping this in mind, we selected the first task to test capability of the neuron model proposed from the RL domain. Namely, we chose one of the RL tasks drawn from the well-established ATARI games benchmark set [22]. This task centers around the computerized ping-pong game, where a ball traverses within a square area, rebounding off its walls. The area has only three walls. Instead of the left wall, the racket moves in the vertical direction on the left border of this square area. The racket is controlled by the agent which can move it up and down. When the ball hits the racket it bounces back and the agent obtains a reward signal. If the ball crosses the left border without hitting the racket the agent gets punishment and the ball is returned to a random point of the middle vertical line of the area, gets random movement direction and speed and the game continues. Using the reward/punishment signals received, the agent should understand that its aim is to reflect the ball and learn how to do it.

In our example, the network (in fact, a single neuron) should solve the very first problem — to "understand" what conditions cause obtaining reward in the near future.

Let us describe the input information coming to the neuron's plastic synapses. This information includes the current positions of the ball and the racket and the ball velocity. While the ultimate formulation of this problem would involve primary raster information (i.e., the screen image), our focus here extends beyond computer vision and delves into the realm of causal relationships. Consequently, we assume that preceding layers have already processed the primary raster data and converted it into the spike-based description of the world state, which forms the basis of our neuron's input.

The input nodes that are sources of spikes sent to the learning neuron are subdivided into the following sections:

1. The ball X coordinate. Consists of *30 nodes* capturing the ball's horizontal position. The horizontal dimension is broken to 30 bins. When the ball is in the bin i, the i-th node emits spikes with frequency 300 Hz. To establish spatial and temporal scales we assume that the size of the square area is 10×10 cm (so that the boundary coordinates are ±5 cm) and the discrete emulation time step is 1 ms.
2. The ball Y coordinate. Consists of *30 nodes* capturing the ball's vertical position. Similar to X but for the vertical axis.
3. The ball velocity X component. Consists of *9 nodes* capturing the ball's horizontal velocity. When the ball is reset in the middle of the square area, its velocity is set to the random value from the range [10, 33.3] cm/s. Its original movement direction is also random but it is selected so that its X component would not be less than 10 cm/s. The whole range of possible ball velocity X component values is broken to 9 bins such that the probabilities to find the ball at a random time moment in each of these bins are approximately equal. While the ball X velocity is in some bin, the respective input node emits spikes with 300 Hz frequency.
4. Y component of the ball velocity. Consists of *9 nodes* capturing the vertical velocity of the ball. The same logic as for the X velocity component.
5. Y coordinate of the paddle. Consists of *30 nodes* capturing the vertical position of the paddle. Similar to the Y coordinate of the ball. The racket is 1.8 cm in size, so it occupies slightly more than 5 vertical intervals.
6. Relative position of the ball and racket in the near zone. Consists of *25 nodes* capturing the positions of the ball close to the racket. The square field of view of 3×3 cm moves with the racket, so that the center of the racket is always in the center of the left border of this field. The field of view is divided into 5×5 square zones. When the ball is in a zone, the corresponding input node generates spikes with a frequency of 300 Hz.

In total, there are 133 input nodes transmitting their spikes to the learning neuron. The neuron's objective is to discern the conditions that lead to obtaining a reward signal within a 100 ms time frame, and accordingly, we set $T_p = 100$ ms.

1.4. Selection of neuron parameters using a genetic algorithm. While our neuron model appears to be relatively simple, it incorporates several parameters that require appropriate calibration. There are four key parameters in our model:

- The maximum synaptic resource change $\overline{d_H}$. This parameter controls learning speed. Its low values make learning slow, high values may make it unstable.
- Minimum value of synapse weight $w_{\min}$. It is negative.
- Maximum value of synapse weight $w_{\max}$.
- The stability change speed d_S.

The optimum values of $\overline{d_H}$ and d_s are determined by strength of causal links — in case of weak determinism or high noise, great values of these parameters will lead to learning instability. $w_{\min}$ and $w_{\max}$ should be selected on the basis of mean input spike flow — the input node count and the mean spike frequency per node.

Although the general principles for setting these parameters are sufficiently clear, we decided to find their optimum values using the genetic algorithm and the very wide search ranges: [0.03, 1] for $\overline{d_H}$; [0.003, 1] for $w_{\min}$; [0.03, 1] for $w_{\max}$; [0.003, 3] for d_s. For setting their random values, the log-uniform distribution was used. The population size was 300; the elitism level was 0.1; the mutation probability per chromosome was 0.5. The optimization criterion was R (1). It was measured for the last 600 seconds of 2000-second record of ping-pong game where the racket moved chaotically. The total number of rewards was 951. The genetic algorithm terminated when 3 successive generations showed no R increase.

2. Results and discussion

The best result achieved in our experiments was obtained in the 17th generation of the genetic algorithm. The R value was 0.553. The optimal parameter values were: $\overline{d_H} = 0.056$; $w_{\min} = -0.017$ (zero value slightly reduces R); $w_{\max} = 0.48$ (i.e., at least three input spikes are required to trigger); $d_s = 0.23$.

Considering the inherent fuzziness in the relationship between the current world state and soon obtaining reward due to the discrete world description and chaotic racket movement, this R value appears satisfactory. To objectively evaluate the performance, we decided to compare our learning neuron with traditional machine learning methods. To ensure a fair comparison, we selected two machine learning algorithms of completely different nature: decision tree and convolutional neural network. All the algorithms were trained on the same binary signal data from the input nodes, with each emulation step serving as a learning example. The target value was binary, labeling target periods. The machine learning algorithms were applied on the same data as our learning neuron (the first 1 400 000 steps) and created scoring models that returned the probability that the current step belongs to the target period. The steps when the score returned exceeds a certain threshold (determined through optimization) were considered analogously to neuron firing. The prediction periods in the last 600 000 simulation steps were determined using the rule described at the beginning of Section 1.1 and the R value was calculated. The optimum threshold score value was found from the requirement that the R value for the first 1 400 000 steps should be maximum. The decision tree algorithm used the information gain split criterion. The network included 2 convolutional ReLU layers.

The maximum R value obtained by decision tree is 0.742, convolutional network gives R equal to 0.731. The observed proximity of results shown by very different methods proves validity of our approach to determination of theoretical limit for R value in our problem. Thus, the estimation of this limit equal to 0.75 seems to be realistic.

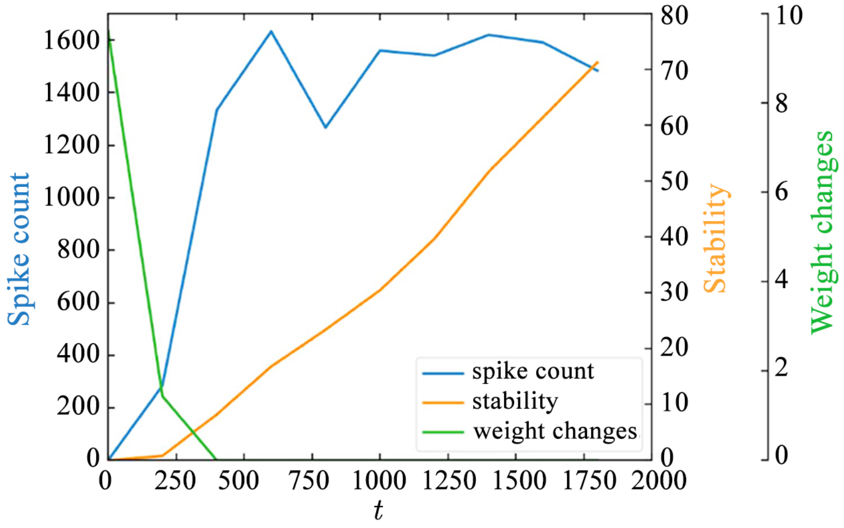

Fig 2. Dynamics of firing frequency, stability and weight changes of the learning neuron

Although the result achieved by our neuron (74% of the theoretical maximum) might be considered modest, we regard our model as successful. Indeed, our neuron is very simple. Being considered as a predictive model, it contains only 133 degrees of freedom. In contrast, the decision tree model includes 51 levels and 403 non-terminal nodes. In essence, the function of our neuron is similar to the conjunction of logical values corresponding to a few of its strongest synapses (see below). Coincidence of these factors is treated as a cause of the target event. However, it is evident that obtaining reward in our example may be preceded by several significantly different conditions. Therefore, it is rather a disjunction of several conjunctions. We propose that a network of these described neurons could yield much more precise reward predictions, and we will explore the potential structure of such a network below.

Let's examine the learning process and its outcomes. The dynamics of the neuron activity, its stability and the total weight change (the sum of absolute values of the individual synaptic weight changes) are represented on Fig. 2. Since the original weights of all plastic synapses were 0, only the dopamine plasticity worked at first. During the first 100 seconds the neuron did not fire. After 250 seconds, its firing frequency stabilized. Due to the weight stabilization mechanism (as can be seen from Fig. 2, neuron's stability grows almost linearly), the synaptic weights did not change after 400 seconds. In reality, the learning process took 400 seconds instead of the planned 1400 seconds.

The learning results are presented on Fig. 3 depicting values of synaptic resources of the learning neuron at the 2000th second. The leftmost plot corresponds to 30 input nodes coding the ball X coordinate. The vertical axis of all plots except the rightmost one displays synaptic resource value. The second plot corresponds to 30 input nodes coding the Y coordinate of the ball (the blue line) and the racket (the orange line). The next two plots represent 9+9 input nodes coding the horizontal and vertical components of the ball velocity. The rightmost plot shows the color-coded values of synaptic resources of the 25 input nodes indicating the location of the ball within a 5×5 grid that moves with the racket. The distribution of synaptic resource values in these plots appears reasonable and in line with expectations.

In conclusion, we can assert that our neuron is capable of detecting causal relationships between observed events. However, as previously discussed, a single neuron, even when adequately

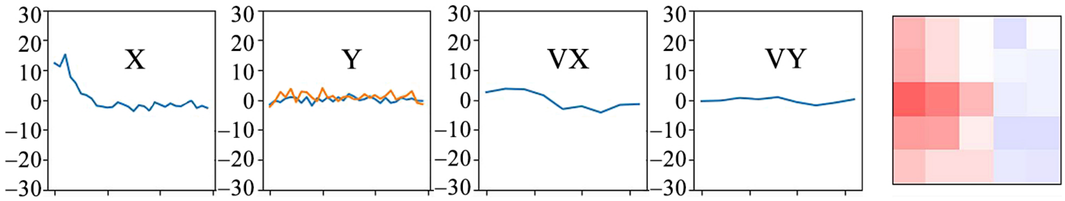

Fig 3. The synaptic resource of the neuron trained (color online)

trained, is not an extremely precise predictor of the imminent occurrence of the target event. This is because the target event may be triggered by several significantly different precursors. To achieve more accurate predictions, a network of neurons of this kind is necessary.

We can propose a hypothesis regarding the potential architecture of such a network (see Fig. 4). In this network, the neurons recognizing different causes of the target event (the blue L rings) enter the columnar structure where each column corresponds to one separate cause. These columns engage in competition to recognize events-causes using lateral inhibitory connections between their "winner-takes-all" (WTA) neurons. The firing ("winning") WTA neuron blocks not only the other WTA neurons but also the GATE neurons in the other columns. If one L neuron fires then the other L neurons should not fire because L neurons should recognize different causal links. If some L neuron fires after the winner it will not receive a reward signal because this signal will not pass through the GATE blocked by the winner. The synapses which caused its firing will be suppressed by anti-Hebbian plasticity and this situation will not repeat next time. If the winning neuron fired correctly it will be rewarded since its GATE is not blocked. We believe that this architectural approach has the potential to recognize complex networks of causal relationships, and we plan to explore and test it in our future research.

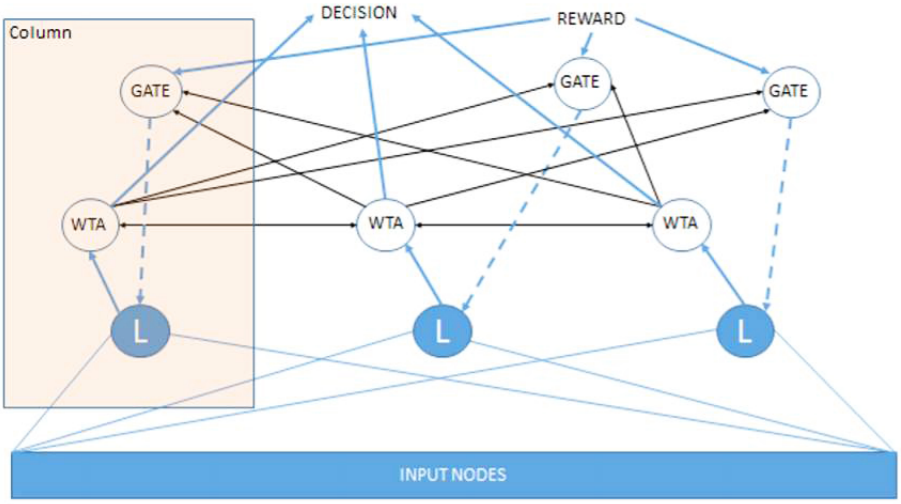

Fig 4. The possible structure including several neurons recognizing causal links for prediction of the target event which may be a consequence of several significantly different causes. The blue arrows are excitatory connections, the black arrows are blocking, the dashed arrows are reward

Conclusion

In conclusion, the ability to discern causal relationships within data streams is a fundamental attribute for any learning system engaging with the real world. In this research, we have showcased the realization of this critical function at the level of an individual neuron, facilitated by a novel combination of anti-Hebbian and dopamine plasticity mechanisms. Given the pivotal role of such mechanisms in the context of implementing reinforcement learning within Spiking Neural Networks (SNNs), we tested our model on a simplified RL problem, namely, the ATARI ping-pong computer game. Our findings, including an estimation of the theoretical upper limit for prediction accuracy in this task, underscore the efficiency of the proposed neuron model in identifying causal links.

Along with this, it is obvious that the neuron described in the work is not complex enough to describe cause-and-effect relationships between many different factors. In this sense, there is a direct analogy with ANNs. It is known that even a single neuron has a limited ability to approximate a narrow class of multidimensional nonlinear functions, but a multilayer network of such neurons is required to approximate a wide class of functions.

Furthermore, we introduced an SNN architecture featuring neurons of the aforementioned type, with the capacity to infer networks of causal relationships directly from raw data. In the forthcoming research, we aim to rigorously test and fine-tune this SNN framework. Additionally, our ongoing investigations will focus on augmenting the capabilities of this SNN to capture temporal aspects of causal relationships, shifting from the question "What are the potential consequences of given events?" to the more nuanced query of "When are these consequences likely to materialize?". We believe that this study is a significant step toward enhancing the capacity of artificial neural networks to model and understand complex causal dynamics in real-world environments.

References

1. Moreno-Bote R, Drugowitsch J. Causal inference and explaining away in a spiking network. Scientific Reports. 2015;5:17531. DOI: 10.1038/srep17531.
2. Lansdell BJ, Kording KP. Neural spiking for causal inference and learning. PLoS Computational Biology. 2023;19(4):e1011005. DOI: 10.1371/journal.pcbi.1011005.
3. Skatchkovsky N, Jang O, Simeone O. Bayesian continual learning via spiking neural networks. Frontiers in Computational Neuroscience. 2022;16:1037976. DOI: 10.3389/fncom.2022.1037976.
4. Friston K. The history of the future of the bayesian brain. Neuroimage. 2012;62(2):1230–1233. DOI: 10.1016/j.neuroimage.2011.10.004.
5. Kasabov N, Scott NM, Tu E, Marks S, Sengupta N, Capecci E, Othman M, Doborjeh MG, Murli N, Hartono R, Espinosa-Ramos JI, Zhou L, Alvi FB, Wang G, Taylor D, Feigin V, Gulyaev S, Mahmoud M, Hou Z-G, Yang J. Evolving spatio-temporal data machines based on the NeuCube neuromorphic framework: Design methodology and selected applications. Neural Networks. 2016;78:1–14. DOI: 10.1016/j.neunet.2015.09.011.
6. Kasabov NK. NeuCube: a spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data. Neural Networks. 2014;52:62–76. DOI: 10.1016/j.neunet.2014.01.006.
7. Schliebs S, Fiasché M, Kasabov N. Constructing robust liquid state machines to process highly variable data streams. International Conference on Artificial Neural Networks ICANN 2012: Artificial Neural Networks and Machine Learning — ICANN. 2012;604–611. DOI: 10.1007/978-3-642-33269-2_76.

8. Doborjeh M, Doborjeh Z, Merkin A, Krishnamurthi R, Enayatollahi R, Feigin V, Kasabov N. Personalized spiking neural network models of clinical and environmental factors to predict stroke. Cognitive Computation. 2022;14:2187–2202. DOI: 10.1007/s12559-021-09975-x.
9. Fernando C. From blickets to synapses: Inferring temporal causal networks by observation. Cognitive Science. 2013;37(8):1426–1470. DOI: 10.1111/cogs.12073.
10. Markram H, Gerstner W, Sjöström PJ. A history of spike-timing-dependent plasticity. Frontiers in Synaptic Neuroscience. 2011;3:4. DOI: 10.3389/fnsyn.2011.00004.
11. Kerr RR, Grayden DB, Thomas DA, Gilson M, Burkitt AN. Coexistence of reward and unsupervised learning during the operant conditioning of neural firing rates. PLoS ONE. 2014;9(1):e87123. DOI: 10.1371/journal.pone.0087123.
12. Yuan M, Wu X, Yan R, Tang H. Reinforcement learning in spiking neural networks with stochastic and deterministic synapses. Neural Computation. 2019;31(12):2368–2389. DOI: 10.1162/neco_a_01238.
13. Mozafari M, Ganjtabesh M, Nowzari-Dalini A, Thorpe SJ, Masquelier T. Bio-inspired digit recognition using reward-modulated Spike-Timing-Dependent Plasticity in deep convolutional networks. Pattern Recognition. 2019;94:87–95. DOI: 10.1016/j.patcog.2019.05.015.
14. Frémaux N, Sprekeler H, Gerstner W. Functional requirements for reward-modulated Spike-Timing-Dependent Plasticity. The Journal of Neuroscience. 2010;30(40):13326–13337. DOI: 10.1523/JNEUROSCI.6249-09.2010.
15. Juarez-Lora A, Ponce-Ponce VH, Sossa H, Rubio-Espino E. R-STDP spiking neural network architecture for motion control on a changing friction joint robotic arm. Frontiers in Neurorobotics. 2022;16:904017. DOI: 10.3389/fnbot.2022.904017.
16. Ivanov D, Chezhegov A, Kiselev M, Grunin A, Larionov D. Neuromorphic artificial intelligence systems. Frontiers in Neuroscience. 2022;16:959626. DOI: 10.3389/fnins.2022.959626.
17. Kiselev M, Ivanov A, Ivanov D. Approximating conductance-based synapses by current-based synapses. Advances in Neural Computation, Machine Learning, and Cognitive Research IV. Neuroinformatics 2020. Studies in Computational Intelligence. 2020;925:394–402. DOI: 10.1007/978-3-030-60577-3_47.
18. Kiselev MV. A Synaptic Plasticity rule providing a unified approach to supervised and unsupervised learning. Proceedings of International Joint Conference on Neural Networks. 2017:3806–3813. DOI: 10.1109/IJCNN.2017.7966336.
19. Ho VM, Lee JA, Martin KC. The cell biology of synaptic plasticity. Science. 2011;334(6056): 623–628. DOI: 10.1126/science.1209236.
20. Citri A, Malenka RC. Synaptic plasticity: Multiple forms, functions, and mechanisms. Neuropsychopharmacology Reviews. 2008;33:18–41. DOI: 10.1038/sj.npp.1301559.
21. Roberts PD, Leen TK. Anti-hebbian spike-timing-dependent plasticity and adaptive sensory processing. Frontiers in Computational Neuroscience. 2010;4:156. DOI: 10.3389/fncom.2010.00156.
22. Jiajun F. A review for deep reinforcement learning in atari: Benchmarks, challenges, and solutions. ArXiv:abs/2112.04145. 2022.