

УДК 004.852

Исследование метода обучения с подкреплением спайковой нейронной сети

Козов А. В.¹, Чернышев А. С.^{1,*}

[*alexey.chemushev@gmail.com](mailto:alexey.chemushev@gmail.com)

¹МГТУ им. Н.Э. Баумана, Москва, Россия

В статье рассмотрен метод обучения с подкреплением спайковой нейронной сети, построенной при применении определенного типа связи между нейронами, называемого гедонистическим синапсом. Приведено описание метода, алгоритма получения тестовых данных и способов оценки качества обучения, использованных при исследовании. Представлены выводы по результатам работы и возможные направления дальнейших исследований.

Ключевые слова: спайковая нейронная сеть, обучение с подкреплением, оценка качества классификации

Введение

Искусственные нейронные сети в настоящее время имеют широкую область применения. Большой интерес представляет такой класс нейронных сетей как спайковые, также называемые импульсными, нейронные сети. Спайковые нейросети – это одно из направлений исследования резервуарных вычислений в области рекуррентных сетей неустойчивых состояний (Liquid State Machine) [1]. В основе таких сетей лежит представление взаимодействия между отдельными нейронами в виде спайков — сигналов одинаковой амплитуды и длительности. Передаваемая информация при этом закодирована посредством времени возникновения и частоты отдельных импульсов (спайков). Это отличает спайковые нейросети от классических искусственных нейронных сетей, в которых эта информация представлена некоторым действительным числом. В работе [2] предложено условное разбиение нейронных сетей по принципу их работы на поколения, и спайковые нейросети выделены как отдельное, третье, поколение искусственных нейронных сетей.

Отличия в принципе функционирования сетей третьего и предыдущих поколений требуют пересмотра, приспособления или разработки новых правил и методов работы с ними [3]. В работе доктора Сеунга (Sebastian H. Seung) [4] предложен метод обучения с подкреплением, применимый для определенного типа связи между нейронами, названного гедонистическим синапсом. Такая связь представляет собой модель стохастического си-

напса, частично отражающего свойства своего биологического аналога. Гедонистический синапс изменяет собственную активность в зависимости от своей активности в предыдущие моменты времени и подкрепления. Положительное («награда») или отрицательное («наказание») подкрепление показывает, насколько активность данного синапса соответствует ожидаемой. Функции параметров синапса заданы таким образом, что синапс стремится реже получать «наказание» и чаще — «награду». По этой причине синапс был назван гедонистическим.

1. Модель гедонистического синапса

Гедонистический синапс реализует механизм стохастического градиентного подъема по поверхности ожидаемой награды. Предложенная модель является теоретической по своей сути и не претендует на серьезное нейробиологическое основание. Тем не менее, она отражает некоторые известные биологические феномены, например, такой как выработка оперантного условного рефлекса [5].

Пусть ситуация, когда синапс передает возникающий в пресинаптической области импульс в постсинаптическую, будет называться проведением входного спайка синапсом. Модель гедонистического синапса устроена так, что синапс может провести входной импульс (в терминах нейробиологи – высвободить нейромедиаторы) с некоторой вероятностью s . Эта вероятность зависит от состояния, которое имеет синапс в момент возникновения передаваемого спайка: доступности (A) или невосприимчивости (R). В состоянии невосприимчивости синапс не может производить высвобождение нейромедиаторов, и, соответственно, $s = 0$ при любых входных спайках (любой пресинаптической активности). В состоянии невосприимчивости синапс переходит сразу после успешного проведения спайка, и через некоторое время, заданное константой τ_r , меняет свое состояние на A . В состоянии доступности вероятность того, что он высвободит нейромедиаторы задана зависимостью

$$s = \frac{1}{1 + e^{-q-c}},$$

которая является функцией от показателя натренированности синапса q и показателя c . Роль последнего в том, чтобы производить кратковременную потенциацию активных синапсов (Short Term Potentiation). Чем больше значение этих двух переменных, тем выше вероятность проведения спайка.

Переменные q и c изменяют свои значения во времени и имеют различную динамику. Изменение переменной $q(t)$ происходит на большом промежутке времени, обучение нейрона охарактеризовано именно этим показателем. Его можно описать как вероятностный вес синапса. Переменную $c(t)$ можно охарактеризовать как катализатор спайков синапса — при большом числе проведенных импульсов синапс увеличивает вероятность своего спайка.

Каждый входной импульс оставляет так называемый след $e(t)$ на синапсе, который угасает за заданный промежуток времени. Этот след выступает в качестве ключевого по-

казателя в обучении синапса следовать награде. Угасание следа описывает динамика, заданная уравнением

$$\frac{de}{dt} = -\frac{e}{\tau_e},$$

где τ_e — временная константа длительности затухания следа.

Величина переменной c возрастает при каждом входном спайке на небольшую константу Δc , после чего c угасает с динамикой, аналогичной динамике переменной e , но с временной константой τ_c .

При включении системы награды и наказания в данную модель, где уровень награды задан переменной $h(t)$, динамику изменения вероятностного веса синапса $q(t)$ можно описать уравнением

$$\frac{dq}{dt} = \eta h(t) e(t),$$

где η — малый коэффициент скорости обучения.

След активности синапса $e(t)$ играет роль градиента стратегии системы (policy gradient [6]). Изменение вероятностного веса в соответствии с данной политикой приведет к повышению правдоподобности модели и стабилизации ответа сети на стимул. При входном спайке переменная e возрастает на значение Δe : если спайк был успешно проведен («release»), тогда $\Delta e = 1 - s$, или же $\Delta e = -s$ в случае неудачи («failure»). В состоянии невосприимчивости синапса изменения e не происходит ($\Delta e = 0$). Корреляция $e(t)$ с подкреплением $h(t)$ — это центральный механизм в данной модели, который позволяет изменять вероятностные веса нейронов в соответствии с ожидаемой наградой, максимизируя ее.

Когда на синапс в состоянии доступности поступает входной спайк, изменения описанных выше параметров происходят в следующем порядке: попытка высвобождения нейромедиатора с вероятностью s , изменения e на Δe , изменение c на Δc . Если же синапс находится в состоянии невосприимчивости, то изменения этих параметров не происходит.

2. Оценка качества обучения нейронной сети

В настоящее время нет устоявшихся критериев по оценке качества обучения спайковых нейронных сетей, а некоторые стандарты оценок, например, скорость реагирования сети на изменение класса входного сигнала, для спайковых сетей не разработаны. Без проведения отдельных исследований на эту тему, целесообразным является применение наиболее простых и практически значимых критериев. При этом использование сразу нескольких критериев позволяет более объективно оценить работу классификатора. В данной работе для каждой эпохи обучения нейросети выполнена оценка обученности по двум критериям: показатель числа ошибок и разделимость плотности распределения вероятности классификации.

В качестве самого простого способа оценки успешности работы сети может быть выбран показатель числа ошибок. Он представляет собой отношение ошибочных класси-

фикаций к общему числу поданных примеров. Если ассоциировать активность нейронов выходного слоя с распознаваемыми классами согласно их номерам, то ошибками будут следующие ситуации:

- номер наиболее активного нейрона в выходном слое не совпадает с номером класса, поданного на вход;
- имеется несколько одинаково активных нейронов при условии, что активность остальных нейронов выходного слоя меньше;
- активных нейронов нет.

Такая простая оценка позволяет сразу получить представление о текущем уровне обученности сети.

Вторая оценка демонстрирует, насколько хорошо могут быть разделимы выходные сигналы сети. Графически эта оценка представляет собой набор диаграмм плотности распределения вероятности классификации примеров из текущей эпохи, а также построенная по этим данным кривая ошибок (ROC-кривая) [7]. В качестве вероятности классификации p_i за некоторое время использовано отношение числа импульсов n_i , которые произвел соответствующий классу i нейрон, к числу всех импульсов нейронов выходного слоя n_{Σ} за это время:

$$p_i = \frac{n_i}{n_{\Sigma}}, i = 1, 2, \dots, N.$$

Здесь N — число распознаваемых классов. При моделировании для каждой пары распознаваемых классов строятся соответствующие диаграммы оценки плотности (рис. 1).

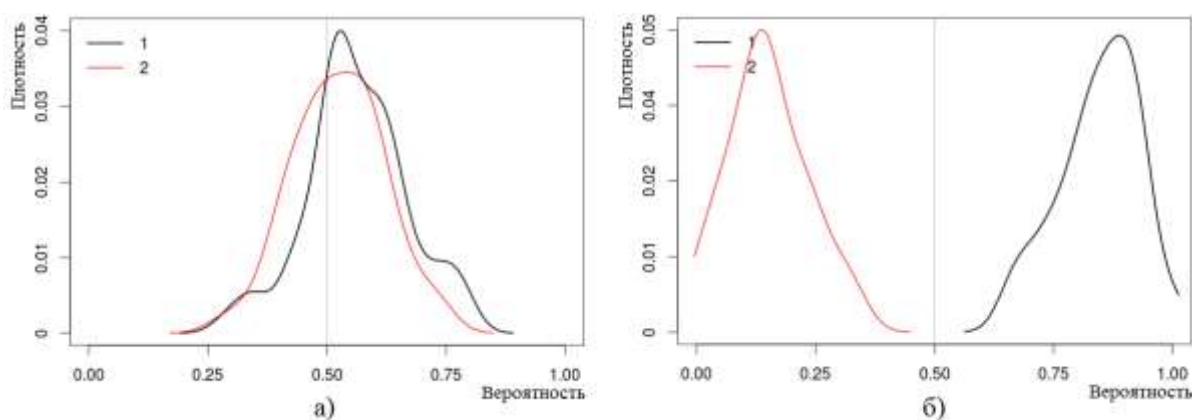


Рис. 1. Классификация примеров двух классов. Выходные данные сети, оценка плотности вероятности классификации образца как класс «1» при предъявлении примеров классов «1» (черные линии) и «2» (красные линии). Изображены результаты моделирования: а) плохо разделяемые сигналы, будет получена низкая оценка обученности; б) хорошо разделяемые сигналы, будет получена высокая оценка обученности

ROC-кривая строится по последовательности полученных вероятностей и показывает зависимость доли верных положительных классификаций от доли ложных положительных классификаций при различных значениях порога решающего правила. Количественную интерпретацию этой оценки можно получить путем подсчета площади, ограниченной кривой ошибок и осью долей ложных положительных классификаций (AUC). Расчет AUC

происходит для каждой пары распознаваемых классов, итоговое значение получено как среднее арифметическое. Как и показатель количества ошибок, AUC-показатель изменяется в пределах от 0 до 1, но, в противоположность первой оценке, чем выше значение AUC, тем более качественный ответ был выдан нейронной сетью. В качестве обобщенной оценки обученности может быть использована взвешенная сумма показателя количества ошибок и величины $(1 - \text{AUC})$.

3. Тестовые данные

Первоначальная настройка и отладка сети, построенной с использованием модели гедонистического синапса, была проведена на простейших тестовых данных. Они представляют собой набор из двух чередующихся спайковых последовательностей. В первой последовательности нейроны входного слоя производят импульсы поочередно в порядке своих номеров по возрастанию, во второй – по убыванию (рис. 2, а).

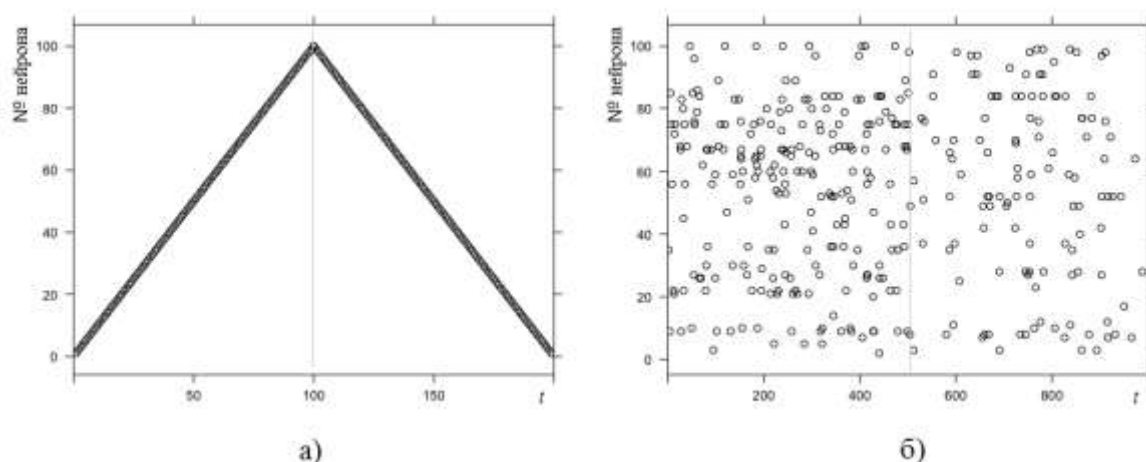


Рис. 2. Карты спайков нейронов входного слоя сети. По оси абсцисс отложено время симуляции t , по оси ординат – номера нейронов входного слоя; каждая окружность означает спайк нейрона в момент времени t , тонкая линия посередине – граница классов. Примеры: а) два простых класса для отладки; б) «сложные» классы

После того как была достигнута корректная работа сети, тестовые примеры были заменены более сложными. Каждый класс представляет собой уникальный паттерн последовательности спайков. Он получен при соотношении с каждым нейроном входного слоя некоторой интенсивности порождения спайков (рис. 2, б). В обучающем примере классы остаются неизменными, чередуясь друг с другом. Такой способ получения обучающих примеров приведен в работе [8]. Для испытания обученной сети использованы дополнительные тестовые данные, полученные добавлением некоторого шума между исходными тестовыми классами. В каждой обучающей выборке задано по 50 примеров данных для каждого класса продолжительностью по 500 миллисекунд. Соответственно, за одну эпоху происходит обучение нейронной сети на 100 образцах данных. По результатам их обработки выполняется оценка качества обучения за эпоху. Итоговая оценка для нескольких

последовательно проведенных эпох может быть рассчитана как среднее арифметическое оценок по эпохам.

4. Исследование метода обучения

Для исследования обучаемости нейросети с гедонистическими синапсами была использована библиотека динамических нейронных сетей dnn [9]. Была построена нейронная сеть из трех слоев нейронов: входного, скрытого и выходного. Модель нейронов — “Leaky integrate-and-fire” [1, 3]. Связи между слоями — стохастические возбуждающие, внутри скрытого слоя присутствуют также тормозные синапсы. Генерация подкрепления в системе задана таким образом, что в случае, если спайк был произведен нейроном, номер которого совпадал с номером класса текущего примера, система получает положительное подкрепление («награда»), в противоположной ситуации — отрицательное («наказание»). Общая награда в системе представлена в двух формах — как текущая (или мгновенная) величина награды r и ее среднее значение r_m за большой промежуток времени. Расчет текущего значения происходит по формуле

$$\frac{dr}{dt} = \frac{r - \Sigma}{\tau_{trace}},$$

где Σ — сумма всех «наград» и «наказаний» за текущий такт времени симуляции; τ_{trace} — временная константа мгновенной награды. Среднее значение награды задано временной константой $\tau_{m t}$ и зависимостью

$$\frac{dr_m}{dt} = \frac{r_m - r}{\tau_{m t}}.$$

В качестве величины уровня награды для гедонистического синапса использована разность между этими двумя значениями: $h = r - r_m$, как это рекомендовано в оригинальной работе [4] для уменьшения шума при обучении.

Нейронная сеть имеет большое число параметров, от которых зависят скорость и качество обучения. Эти зависимости, а также оптимальные значения параметров могут быть определены эмпирическим путем (для предварительной оценки) и при помощи байесовского метода оптимизации, который позволяет находить экстремумы сложных функций. Моделирование с различными параметрами выявило границы областей допустимых значений для параметров гедонистических синапсов и некоторые зависимости для параметров нейронной сети в целом. Например, высокая связность сети (каждый нейрон с вероятностью 0,9 имеет возбуждающую связь с нейронами другого слоя) увеличивает время, необходимое для достижения заданного уровня оценки качества обучения, но при этом обеспечивает его воспроизводимость, позволяя с большей вероятностью получить одинаковую оценку за одинаковое время при нескольких симуляциях.

Эффективность байесовской оптимизации [10] обеспечена благодаря использованию предположений о характере и поведении функции для определения точек испытаний в области поиска. В основе метода лежит теорема Байеса, связывающая апостериорную и априорную вероятности события, наступившего при некотором условии. Вначале необходи-

мо выбрать наиболее достоверное предположение о характере исследуемой функции, в данном случае – о зависимости качества обучения нейронной сети от ее параметров. Для этого была проведена серия последовательных симуляций для получения значений качества обучения при различных параметрах из определенной ранее области их допустимых значений. По полученному набору значений параметров и соответствующих им значений функции происходит расчет и оценка возможных предположений. В качестве основного предположения должно быть выбрано то, которое наиболее точно предсказывает значения функции-оригинала при заданных параметрах. Далее можно провести непосредственно Байесовскую оптимизацию с целью определить такой набор изменяемых параметров, чтобы функция оценки качества обучения сети имела наилучшее, с точки зрения используемого критерия, значение.

Результаты байесовской оптимизации подтвердили первоначальные предположения о прямом влиянии корреляции временной константы синапса τ_e и константы текущей награды τ_{trace} на качество обучения. Для достижения устойчивости при обучении сети эти параметры должны иметь примерно равные значения. Для исследуемой системы они имеют порядок 100 миллисекунд при $\tau_{m, t}$ равном продолжительности одного класса. Активность нейронов выходного слоя и сети в целом определена соотношением констант τ_c , Δc , τ_r . По результатам оптимизации эти константы получили достаточно большие значения: $\tau_c = 205$, $\Delta c = 39$ и $\tau_r = 12$. Высокая активность синапсов не переходит в гиперактивность сети благодаря продолжительному периоду невосприимчивости отдельного синапса (τ_r). Стоит отметить, что найденные значения констант обеспечивают высокую скорость обучения нейронной сети. Уже после второй эпохи активность сети во входном и скрытом слоях приобретает устойчивый характер для каждого из классов и практически не меняется при дальнейшем обучении. Процесс стабилизации активности в этих слоях не зависит от коэффициента скорости обучения синапса η , поскольку влияние параметра натренированности q на вероятность провести спайк значительно меньше влияния параметра c . Из этого можно сделать вывод, что при данных параметрах процесс обучения сети сосредоточен в выходном слое и незначителен в других слоях. Благодаря тому, что выходной слой имеет существенно меньший размер по сравнению с другими слоями, обеспечена высокая скорость обучения и устойчивость работы нейронной сети.

Заключение

Предложенный в работе [4] метод обучения с подкреплением для спайковых нейронных сетей на основе модели гедонистического синапса показал свою работоспособность. Проведено исследование по определению параметров синапса, обеспечивающих высокие значения скорости, качества и устойчивости обучения таких сетей. Гедонистический синапс может применяться в системах распознавания образов, в том числе, для классификации многомерных временных рядов. Определение оптимальной структуры сети с гедонистическими синапсами и возможность их практического применения могут быть направлениями дальнейших исследований.

Список литературы

1. Lukoševičius M., Jaeger H. Reservoir computing approaches to recurrent neural network training // Computer Science Review. 2009. Vol. 3. №. 3. Pp. 127-149.
DOI: [10.1016/j.cosrev.2009.03.005](https://doi.org/10.1016/j.cosrev.2009.03.005)
2. Maass W. Networks of spiking neurons: The third generation of neural network models // Neural Networks. 1997. Vol. 10. №. 9. Pp. 1659-1671. DOI: [10.1016/S0893-6080\(97\)00011-7](https://doi.org/10.1016/S0893-6080(97)00011-7)
3. Paugam-Moisy H., Bohte. S.M. Computing with Spiking Neuron Networks // Handbook of Natural Computing. Springer. 2009. pp 335-376. DOI: [10.1007/978-3-540-92910-9_10](https://doi.org/10.1007/978-3-540-92910-9_10)
4. Seung H. S. Learning in Spiking Neural Networks by Reinforcement of Stochastic Synaptic Transmission // Neuron. 2003. Vol. 40. №. 6. Pp. 1063–1073.
DOI: [10.1016/S0896-6273\(03\)00761-X](https://doi.org/10.1016/S0896-6273(03)00761-X)
5. Skinner, B.F. "The Behavior of Organisms: An Experimental Analysis". New York: Appleton-Century-Crofts. 1938
6. Peters J., Bagnell J.A. Policy Gradient Methods // Encyclopedia of Machine Learning. Springer US. 2011. Pp. 774-776
7. Bradley A.P. The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms // Pattern Recognition. 1997. Vol. 30. № 7. Pp. 1145–1159.
DOI: [10.1007/978-0-387-30164-8_640](https://doi.org/10.1007/978-0-387-30164-8_640)
8. Kappel D., Nessler B., Maass W. STDP Installs in Winner-Take-All Circuits an Online Approximation to Hidden Markov Model Learning // PLoS Comput. Biol. 2014. Vol. 10. № 3.
DOI: [10.1371/journal.pcbi.1003511](https://doi.org/10.1371/journal.pcbi.1003511)
9. Library of Dynamic Neural Networks // <https://github.com> : веб-сайт. Режим доступа: <https://github.com/alexeyche/dnn> (дата обращения 01.05.2016).
10. Brochu E., Cora V.M., De Freitas N. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. CoRR abs. 2010

Studying a Reinforcement Learning Technique for the Spiking Neural Network

A.V. Kozov¹, A.S. Chernyshev^{1,*}

[*alexey.chemushev@gmail.com](mailto:alexey.chemushev@gmail.com)

¹Bauman Moscow State Technical University, Moscow, Russia

Keywords: spiking neural network, reinforcement learning, evaluation of classification

The article describes a reinforcement learning technique for the spiking neural network. Spiking artificial neural networks, also called neural networks of the third-generation are a special class of the artificial neural networks, in which the signal is a set of impulses (spikes) of the same amplitude and duration. There is a need to use special techniques to learn these networks. Development and research of such techniques is a relevant task now. The article discusses a reinforcement learning technique for the neural network using a particular type of connections between neurons, called a hedonistic synapse. The hedonistic synapse partly reflects the properties of its real biological counterpart. It alters its activity depending on the intrinsic activity in previous times and on the reinforcement in the system. The study was conducted to define an impact of such synapse parameters on the learning efficiency of the spiking neural network. The article shows a mathematical model of the hedonic synapse and gives a description of its parameters and variables. It describes the role of these variables and parameters in reinforcement learning the neural network. To assess the learning quality are used such criteria as an error index and AUC. The article describes their application to have a generalized assessment of learning.

It also describes a generation technique and gives examples of test data used in the study. A technique to calculate the reinforcement for the network is described depending on the output signals. The impact of the variables and parameters of the hedonic synapse model on the learning quality for the neural network is analysed. The article reviews the features of the Bayesian optimization technique and describes its using to optimize quality of classification. It provides an in-depth analysis of the optimization results and comes to the conclusion, as a result of the study, that such a learning technique can be used to classify the multivariate time series. The article defines potential trends for further research in the field concerned that is an optimization of the neural network structure and its applications.

References

1. Lukoševičius M., Jaeger H. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 2009, vol. 3, no. 3, pp. 127-149. DOI: [10.1016/j.cosrev.2009.03.005](https://doi.org/10.1016/j.cosrev.2009.03.005)
2. Maass W. Networks of spiking neurons: The third generation of neural network models // *Neural Networks*, 1997, vol. 10, no. 9, pp. 1659-1671. DOI: [10.1016/S0893-6080\(97\)00011-7](https://doi.org/10.1016/S0893-6080(97)00011-7)
3. Paugam-Moisy H., Bohte. S.M. *Computing with Spiking Neuron Networks. Handbook of Natural Computing*. Springer. 2009. pp. 335-376. DOI: [10.1007/978-3-540-92910-9_10](https://doi.org/10.1007/978-3-540-92910-9_10)
4. Seung H.S. Learning in Spiking Neural Networks by Reinforcement of Stochastic Synaptic Transmission // *Neuron*, 2003, vol. 40, no. 6, pp. 1063–1073. DOI: [10.1016/S0896-6273\(03\)00761-X](https://doi.org/10.1016/S0896-6273(03)00761-X)
5. Skinner, B.F. *The Behavior of Organisms: An Experimental Analysis*. New York: Appleton-Century-Crofts, 1938.
6. Peters J., Bagnell J.A. *Policy Gradient Methods. Encyclopedia of Machine Learning*. Springer US, 2011, pp. 774-776. DOI: [10.1007/978-0-387-30164-8_640](https://doi.org/10.1007/978-0-387-30164-8_640)
7. Bradley A.P. The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms. *Pattern Recognition*, 1997, vol. 30, no. 7. Pp. 1145–1159. DOI: [10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2)
8. Kappel D., Nessler B., Maass W. STDP Installs in Winner-Take-All Circuits an Online Approximation to Hidden Markov Model Learning. *PLoS Comput. Biol.*, 2014, vol. 10, no. 3. DOI: [10.1371/journal.pcbi.1003511](https://doi.org/10.1371/journal.pcbi.1003511)
9. Library of Dynamic Neural Networks. <https://github.com> : web-site. Available at: <https://github.com/alexeyche/dnn> (accessed 04.03.16).
10. Brochu E., Cora V.M., De Freitas N. *A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning*. CoRR abs., 2010