

УДК 004.81+004.853+004.855+159.942

## РАЗРАБОТКА СИСТЕМЫ ЭМОЦИОНАЛЬНОЙ ОЦЕНКИ НА ОСНОВЕ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ И НЕЙРОБИОЛОГИЧЕСКИ ИНСПИРИРОВАННЫХ МЕТОДОВ

Е.Ю. Майорова<sup>1</sup>, М.О. Таланов<sup>2</sup>, Р. Лоуи<sup>3</sup>

<sup>1,2</sup> Казанский (Приволжский) федеральный университет

<sup>3</sup> University of Gothenburg

<sup>1</sup>eugeniamaiorova@gmail.com; <sup>2</sup>max.talanov@gmail.com; <sup>3</sup>robert.lowe@his.se

### **Аннотация**

Объектом проведенного исследования является эмоциональная оценка искусственного интеллекта. В качестве системы реализации эмоциональной оценки выбрана система обучения с подкреплением. В результате симуляции построенной модели получены графики, показывающие активность структур мозга, участвующих в процессе их воздействия друг на друга. В ходе настройки системы удалось добиться четырех вспышек активности на таламусе вместо ожидаемых пяти.

**Ключевые слова:** NEST, NeuCogAR, куб Лёвхейма, эмоциональная оценка

### **ВВЕДЕНИЕ**

Эмоции – это филогенетический непрерывный механизм, гибко адаптирующийся к любым стимулам и ответам [5]. Гибкость эмоций получается путем отделения поведенческой реакции от события-стимула. Эмоциональный процесс – это оценка события по его значимости для индивида и предположение (прогнозирование) возможных действий, направленных на изменение отношений между человеком и окружающей средой [6]. Основная проблема состоит в том, что современные вычислительные системы не могут испытывать эмоции, что в итоге приводит к их нежизнеспособности в условиях реального мира.

Актуальность темы проведенного исследования подтверждена американским ученым Марвином Мински (1927–2016), одним из основателей Лаборатории искусственного интеллекта в Массачусетском технологическом университете: в своей книге [7] он говорит о значимости эмоционального процесса для создания

мыслящего искусственного интеллекта.

## **ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ КАК СПОСОБ ЭМОЦИОНАЛЬНОЙ ОЦЕНКИ**

### **Принятие решений и обучение с подкреплением**

Поведенческая нейробиология решает различные вопросы, в числе которых – вопрос принятия решений. Выбор оптимального действия основан на прогнозах долгосрочных последствий [29]. Например, 1) часто вкусу предшествует запах: почувствовав приятный аромат с кухни, мы предполагаем, что вскоре сможем вкусно поесть; 2) мы не будем переходить дорогу даже на зеленый сигнал светофора, если на нас с высокой скоростью летит автомобиль.

Вопрос принятия решений был подробно исследован в поведенческой психологии на основе парадигмы Павлова. Также появились и другие доказательства, что существуют области, контролирующие различные аспекты усвоенного поведения. Так как вопрос принятия решений связан с прогнозированием, очевидно, что он связан и с обучением.

Современная интерпретация обучения с подкреплением исторически возникла из двух отдельных исследований. Первое исследование провели Ричард Саттон и Эндрю Барто в направлении изучения создания искусственного интеллекта: они разработали основные алгоритмы и концепции обучения с подкреплением [33]. Второе исследование выполнили Димитрий Бертсекас и Джон Цициклис, специалисты в области исследования операций и оптимального управления. Их разработки заключаются в стохастических приближениях к методам динамического программирования, которые были названы нейродинамическим программированием [34]. Слияние этих двух направлений исследований сформулировало поведенчески вдохновленные алгоритмы эвристического обучения с подкреплением и предоставило инструменты для анализа свойств этих алгоритмов [29].

Обучение с подкреплением – это обучение какого-либо агента действиям, чтобы максимизировать сигнала поощрения (вознаграждения), который принимает числовые значения. Обучение с подкреплением включает в себя следующие элементы: стратегия, функция вознаграждения, функция ценности и модель среды (в системах, имитирующих обучение с подкреплением) [35]. Стратегия

определяет поведение агента в каждый момент времени, функция поощрения является характеристикой вознаграждения, функция ценности характеризует состояние с учетом долгосрочных перспектив, а модель среды имитирует действия рассматриваемой среды.

В процессе обучения агент взаимодействует со средой, получает от нее ответ, в зависимости от которого осуществляет дальнейшие действия. Каждое новое состояние характеризуется новым вознаграждением. Максимизация награды является целью обучения с подкреплением, для достижения которой могут использоваться различные алгоритмы: динамическое программирование, прогнозирование и управление Монте-Карло, SARSA, TD-обучение и Q-обучение [36–41]. Все они были рассмотрены Саттоном и Барто [35].

### **Значение оценочной обратной связи**

Как уже было сказано, цель обучения с подкреплением заключается в максимизации выгоды для интеллектуального агента. Здесь мы можем рассматривать две задачи: прогнозирование и управление.

В задаче прогнозирования обучение с подкреплением используется только для определения некой политики (стратегии), которая описывает для каждого посещаемого состояния ожидаемую в будущем награду [8].

Задача управления ставит целью поиск набора стратегий, который максимизирует награду в режиме реального времени, т. е. во время переходов от одного состояния к другому.

Для поиска оптимального маршрута используются алгоритмы, упомянутые выше. При этом существует два вида нейронных архитектур для прогнозирования и контроля: с обратной связью и без обратной связи.

Классический пример обучения с подкреплением без обратной связи – собака Павлова [42]. В данном случае собака является интеллектуальным агентом, пища – безусловным раздражителем, звонок – условным. Условный раздражитель предсказывает безусловный, после обучения у агента появляется реакция на условный раздражитель. В данном случае нет обратной связи, т. к. реакция на условный раздражитель не влияет на представление стимулов.

В случае существования обратной связи архитектура обучения представляет

собой замкнутую схему. Замкнутость необходима для решения задачи управления, т. к. действия агента влияют на свои собственные входные значения. Эта схема изображена на рис. 1. Точка  $X_0$  является сигналом обратной связи [8, 35].



Рис. 1. Архитектура обучения с подкреплением с обратной связью

На рис. 2 показана эта же архитектура, расширенная до так называемой модели Актёр–Критик (Actor–Critic). Эта модель получила такое название благодаря аналогии с актёром, который выбирает действие, и критиком, который эти действия критикует. Критик представляет собой функцию оценочной стоимости [8, 35].

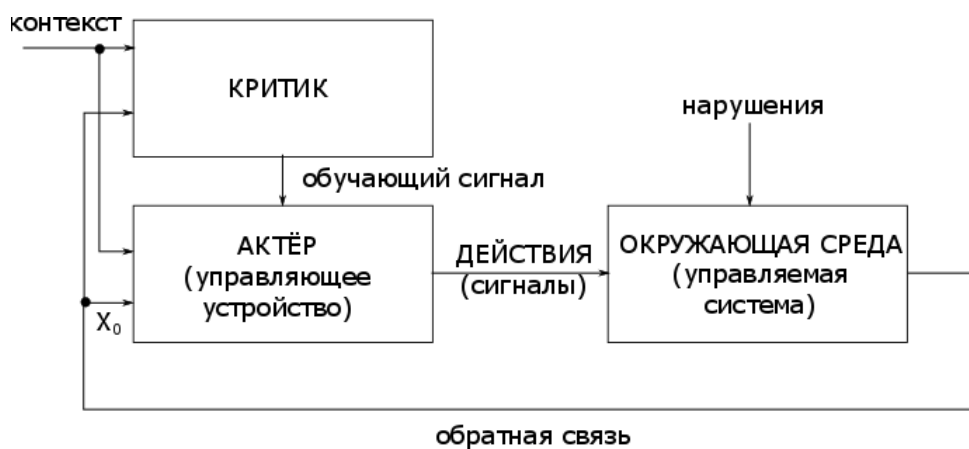


Рис. 2. Архитектура Актёр–Критик

Таким образом, обучение с подкреплением предполагает взаимодействие между некими средой и обучающейся системой, т. е. наличие обратной связи.

### Постановка задачи

Задачу обучения с подкреплением можно привести к математической модели. Рассмотрим, как она будет выглядеть в общем виде.

Пусть  $S$  – множество всех принимаемых состояний, и оно конечно. В процессе обучения с подкреплением агент взаимодействует со средой. Другими словами, происходит игра между ними. Ниже приведен порядок действий в этой игре. Инициализация стратегии и состояния среды: для каждого момента времени агент выбирает действие; среда генерирует награду; среда генерирует новое состояние; агент корректирует стратегию [43].

Эта игра будет называться марковским процессом принятия решений, если вероятность получения нового состояния и награды в нем зависит от текущего состояния и действия, совершенного агентом [44]. Задача состоит в том, чтобы найти такую стратегию, при которой суммарная награда будет максимальной.

### **Обзор методов решения задачи**

Саттон и Барто выделили три фундаментальных метода решения задачи обучения с подкреплением: динамическое программирование, метод Монте-Карло и метод временной разницы (Temporal Difference, TD-метод) [35].

Методы динамического программирования предполагают, что модель окружающей среды идеальна и выполняется марковское свойство. Это предположение является следствием того, что применяются такие методы очень редко. Основная идея таких методов состоит в том, что функция ценности используется для создания и поиска оптимальных стратегий. Алгоритм можно описать следующим образом: 1) Выбирается стратегия; 2) До тех пор, пока стратегия не устойчива: стратегия оценивается – функция оценки стратегии; стратегия улучшается – улучшенная стратегия. Методы динамического программирования могут быть эффективными для определенного круга задач.

Методы Монте-Карло основаны на осредненной выгоде для некой выборки. Их используют для заданий, которые можно разбить на эпизоды. Здесь оценивание происходит не на каждом шаге, а после завершения очередного эпизода. Методы Монте-Карло имеют преимущества перед методами динамического программирования, например, они менее чувствительны к выполнению марковского свойства.

Методы временной разницы в настоящее время являются наиболее важными среди всех методов обучения с подкреплением. Они объединяют в себе идеи методов динамического программирования и Монте-Карло.

### Метод временной разницы (TD-метод)

Большинство алгоритмов, названных выше, опирается на так называемый TD-метод – метод временной разницы [8]. Он был предложен Саттоном и Барто в 1990 году в качестве альтернативы модели Рескорла–Вагнера, которая учитывает торможение и блокирование, а также предполагает, что обучение происходит только тогда, когда награда не равна ожиданию [45]. У этой модели отмечаются некоторые недостатки. Например, она не может описывать события второго порядка: если стимул А предсказывает стимул В, а стимул В предсказывает состояние S, то очевидна связь между А и S, что никак не отражается в этой модели. Вместе с тем, TD-метод устраняет недостатки модели Рескорла–Вагнера [29].

Целью TD-метода является оценка ситуации с точки зрения будущих наград. Происходит это следующим образом: состояние и подкрепительный сигнал (награда) передаются из окружающей среды критику. Он оценивает состояние функцией и высчитывает ошибку предсказания, которая используется для обучения и у критика, и у актёра. Актёру передается значение, на основании полученных данных он генерирует стратегию. На рис. 3 изображена схема TD-обучения.



Рис. 3. TD-метод

Согласованность между последовательными состояниями лежит в основе TD-метода, с его помощью можно прогнозировать события, даже если среда не известна [29].

## Выбор нейрональной архитектуры обучения с подкреплением

Как же обучение с подкреплением происходит на нейрональном уровне? Многие нейрональные архитектуры модели Актёр–Критик, описанные в литературе, сосредоточены на взаимодействии базальных ганглиев и коры. В этих архитектурах Критик описывается более подробно, чем Актёр [8]. В настоящее время описаны следующие архитектуры: параллельные взаимные архитектуры, модель Хоука; идеализированные параллельные взаимные архитектуры; расходящиеся взаимные архитектуры; параллельные невзаимные архитектуры; расходящиеся невзаимные архитектуры.

Особое внимание стоит уделить схеме, предложенной Вёргёттенем и Порром и описывающей связь между TD-обучением и базальными ганглиями [8]. На рис. 4 показаны эти связи с основными входами и выходами, а также отмечены возбуждающие и подавляющие связи между нейронами. Обозначения: ЧС(r) – черная субстанция (pars reticulata), БШ(in) – бледный шар (pars interna), БШ(ex) – бледный шар (pars externa), ЧС(c) – черная субстанция (pars compacta), ВОП – вентральная область покрышки, RRA – ретрорубральная область.

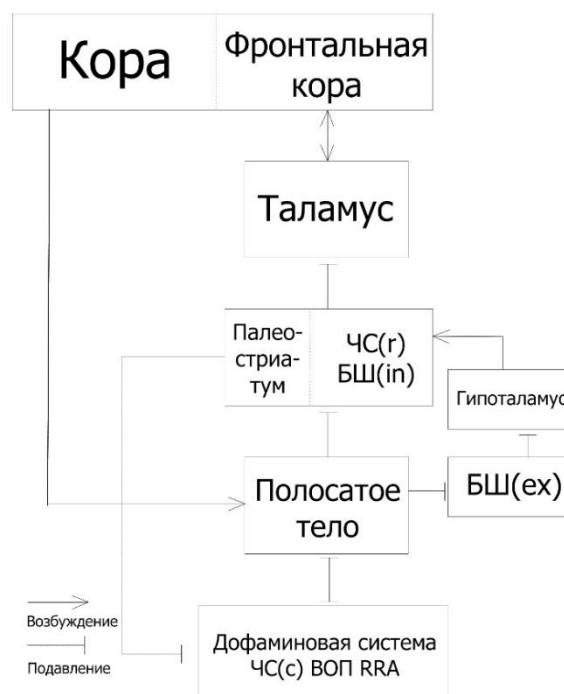


Рис. 4. Связь между TD-обучением и базальными ганглиями

## Составление нейрональной модели

Для построения высокоуровневой модели системы эмоциональной оценки на основе обучения с подкреплением были выбраны метод временной разницы, а также нейрональная архитектура для его реализации.

Преобразуем схему так, чтобы ее использование в NEST не вызывало затруднений. Окончательный вариант представлен на рис. 6.

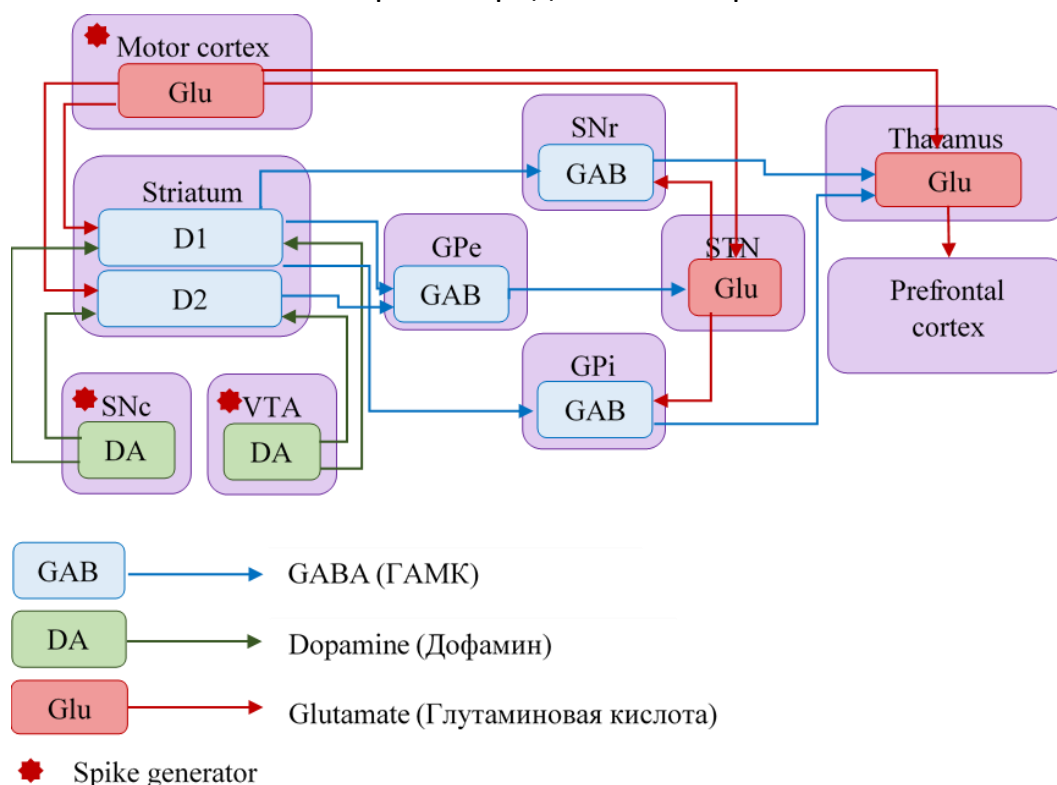


Рис. 6. Нейрональная архитектура

Спайковые генераторы (Spike generator) устанавливаются на двигательную кору, черную субстанцию (compacta), вентральную область покрышки.

Двигательная кора передает возбуждающие (Glutamate) сигналы в таламус, гипоталамус и полосатое тело. Под воздействием дофамина черная субстанция и вентральная область покрышки передают сигнал полосатому телу. В свою очередь, полосатое тело подавляюще (GABA) воздействует на черную субстанцию (reticulata) и бледный шар. Бледный шар (externa) передает подавляющие сигналы гипоталамусу. Гипоталамус возбуждает черную субстанцию (reticulata) и бледный шар (interna), которые подавляюще воздействуют на таламус. В итоге таламус получает сигналы от двигательной коры, черной субстанции (reticulata) и бледного шара (interna) и передает возбуждающий сигнал префронтальной коре.



## Реализация

Представим модель, описанную выше, на языке Python. За связи на рис. 6 отвечает следующий код:

```
connect(motor[Cortex], thalamus[thalamus_Glu], syn_type=Glu)
connect(motor[Cortex], striatum[D1], syn_type=Glu)
connect(motor[Cortex], striatum[D2], syn_type=Glu)
connect(motor[Cortex], stn[stn_Glu], syn_type=Glu)
connect(striatum[D1], snr[snr_GABA])
connect(striatum[D1], gpi[gpi_GABA])
connect(striatum[D1], gpe[gpe_GABA])
connect(striatum[D2], gpe[gpe_GABA])
connect(gpe[gpe_GABA], stn[stn_Glu])
connect(stn[stn_Glu], snr[snr_GABA], syn_type=Glu)
connect(stn[stn_Glu], gpi[gpi_GABA], syn_type=Glu)
connect(gpi[gpi_GABA], thalamus[thalamus_Glu])
connect(snr[snr_GABA], thalamus[thalamus_Glu])
connect(thalamus[thalamus_Glu], motor[FrontalCortex], syn_type=Glu)
if dopa_flag:
    <...>
    connect(snc[snc_DA], striatum[D1], syn_type=DA_ex)
    connect(snc[snc_DA], striatum[D2], syn_type=DA_in)
    connect(vta[vta_DA0], striatum[D1], syn_type=DA_ex)
    connect(vta[vta_DA0], striatum[D2], syn_type=DA_in)
```

Выше был описан порядок воздействия разных структур мозга друг на друга. Здесь мы видим, что, действительно, сначала воздействует двигательная кора (1 – 4), затем полосатое тело воздействует на черную субстанцию и бледный шар (5 – 8), бледный шар (externa) – на гипоталамус (9). Гипоталамус воздействует на черную субстанцию (reticulata) и бледный шар (interna) (10 – 11), которые воздействуют на таламус (12 – 13). В конце таламус воздействует на префронтальную кору (14). Дофаминовое воздействие вынесено в отдельный блок (15): черная субстанция (compacta) и вентральная область покрышки воздействуют на полосатое тело (b – e). Типы связей (возбуждающая, подавляющая, воздействие дофамина)

---

также указаны, за них отвечает параметр `syn_type`.

Значения используются при запуске генератора. Он генерирует спайки по пуассоновскому распределению и действует с определенной силой, измеряемой в mV, т. е., по сути, генератор заряжает нейрон и повышает мембранный потенциал. Также у генератора есть частота. На языке Python для него используется соответствующее условие `generator_flag`. Рассмотрим код подробнее:

```
if generator_flag:
    connect_generator(motor[Cortex], 9.9, T, rate=300, coef_part=1, weight=2)
    delta = [1.0, 1.5, 0.38, 0.8, 0.33]
    k = 9.9
    iter = 0
    for i in range(5) :
        connect_generator(snc[snc_DA], k, k + 3.3, rate = 300, weight=delta[iter],
            coef_part=1)
        connect_generator(vta[vta_DA0], k, k + 3.3, rate = 300, weight=delta[iter],
            coef_part=1)
        iter += 1
        k += 3.3
```

Подключаем генератор к двигательной коре (2), время действия – от 9,9 мс до T (задается, в нашем случае T = 40 мс), частота – 300, параметр `coef_part` отвечает за процент нейронов, которые будут подцеплены детектором (используется для отображения результатов в виде графиков), сила генератора `weigh` равна 2. Затем нужно подключить еще два генератора, объявляем их (3). Чтобы дать генераторам время на «разгон», время действия поставим от 9,9 мс (4). Далее запускаем генераторы в цикле, будем изменять силу на каждом шаге. Один шаг длится 3,3 мс.

#### *Тестирование модели с различными данными*

До сих пор мы говорили лишь о воздействиях различных структур мозга друг на друга. Однако зачастую наличия воздействия недостаточно, например, воздействие может быть слабым и совершенно незначительным. Но у этой проблемы есть решение: можно определить синаптическую силу этого воздействия для любой связи. Такая сила называется весом, обозначается в программе с помощью параметра `weight_coef`.

---

Нам нужно настроить веса таким образом, чтобы была видна активность структур мозга, участвующих в обучении с подкреплением. Ожидается активность таламуса под воздействием дофамина, 5 вспышек. На рисунках 7 и 8 показана активность таламуса до настройки системы.

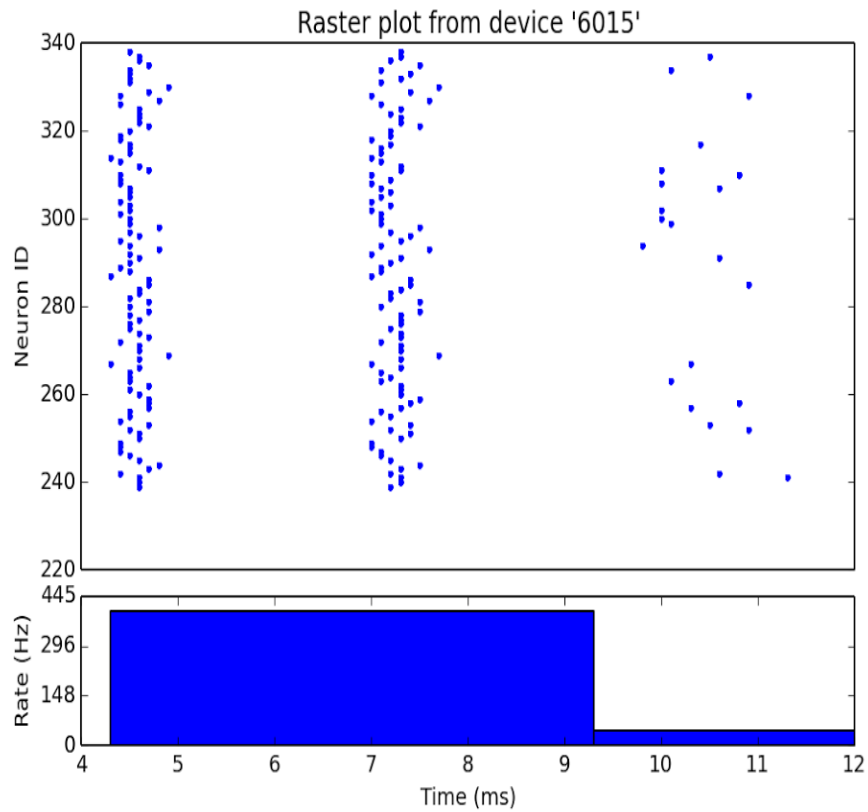


Рис. 7. Активность на таламусе до настройки системы. Спайки

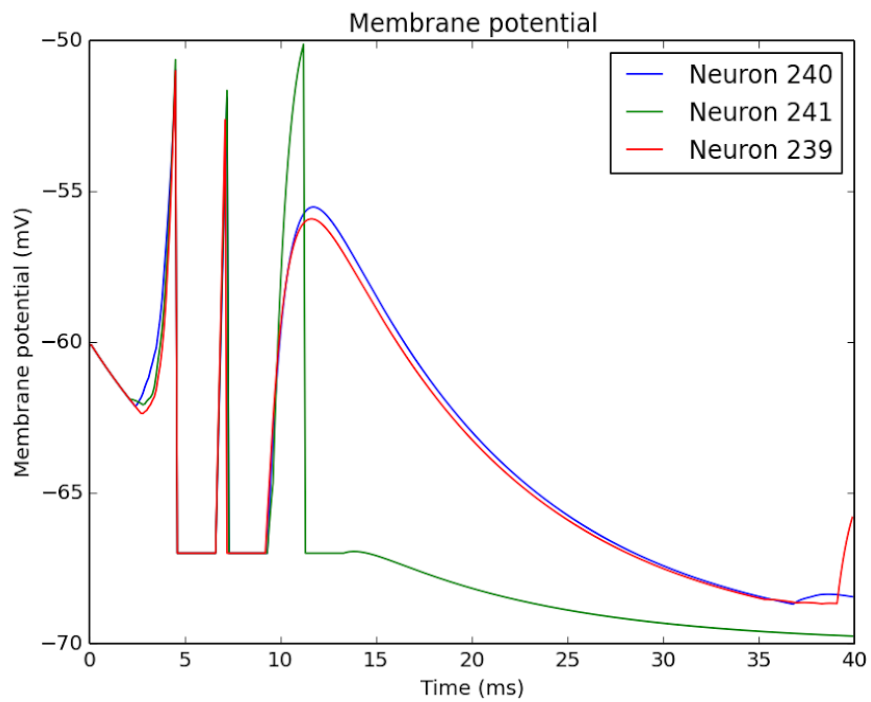


Рис. 8. Активность на таламусе до настройки системы. Мембранный потенциал  
Тестирование системы с различными параметрами привело к результатам, показанным на рисунках 9 и 10.

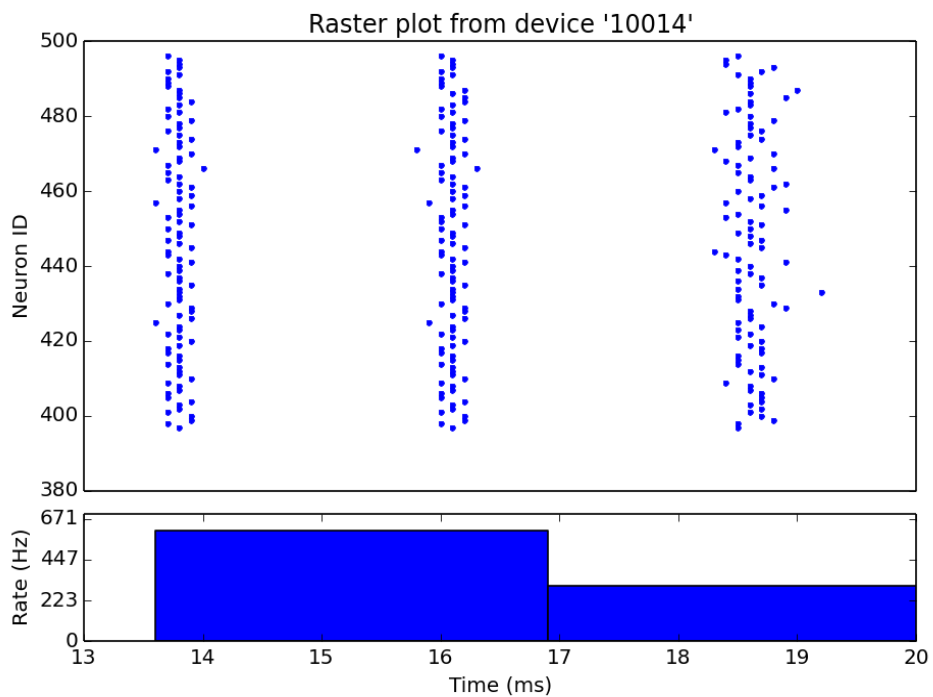


Рис. 9. Активность на таламусе после настройки системы. Спайки

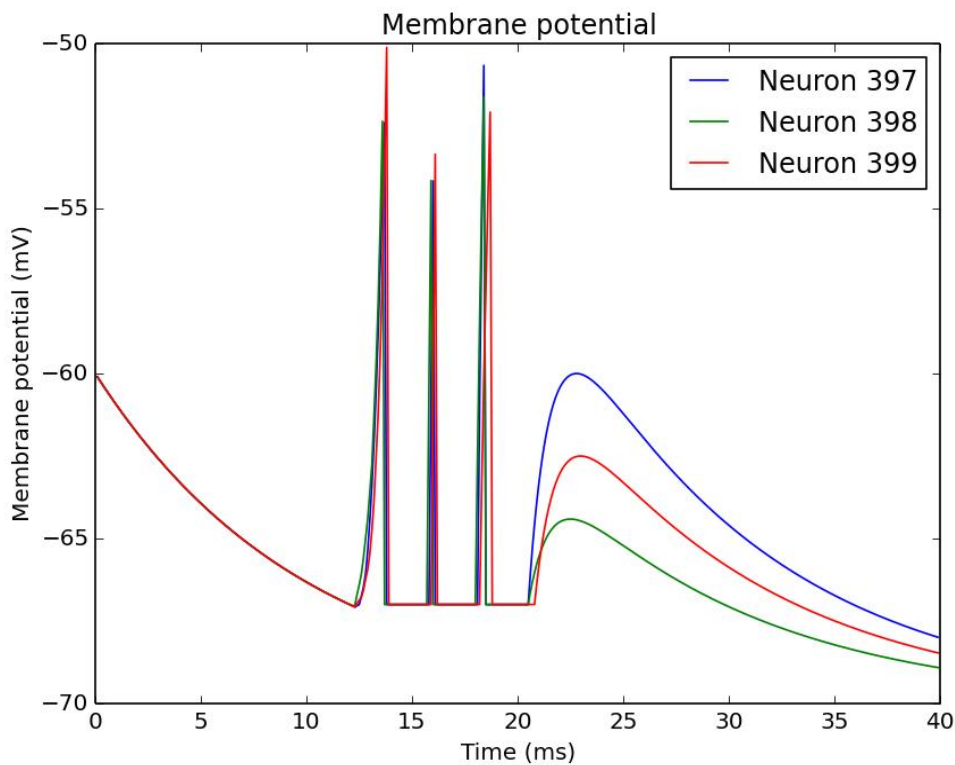


Рис. 10. Активность на таламусе после настройки системы.

#### Мембранный потенциал

На рис. 10 мы видим четыре вспышки вместо ожидаемых 5. Это произошло из-за подавления бледного шара (*interna*) и черной субстанции (*reticulata*) и восстановления потенциала таламуса; как видно на рис. 9, первые 3,3 мс таламус не активен, а первая вспышка отсутствует. Далее система ведет себя так, как и ожидалось. Спайки для бледного шара (*interna*) и черной субстанции (*reticulate*) отсутствуют, так как выделяется дофамин.

Данных результатов мы добились, используя следующие веса (если вес не указан, по умолчанию используется значение, равное единице):

```
connect(motor[Cortex], thalamus[thalamus_Glu], syn_type=Glu, weight_coef=0.5)
```

```
connect(motor[Cortex], striatum[D1], syn_type=Glu)
```

```
connect(motor[Cortex], striatum[D2], syn_type=Glu)
```

```
connect(motor[Cortex], stn[stn_Glu], syn_type=Glu)
```

```
connect(striatum[D1], snr[snr_GABA], weight_coef=0.1)
```

```
connect(striatum[D1], gpi[gpi_GABA], weight_coef=0.1)
```

```
connect(striatum[D1], gpe[gpe_GABA], weight_coef=0.1)
```

```
connect(striatum[D2], gpe[gpe_GABA], weight_coef=0.1)
connect(gpe[gpe_GABA], stn[stn_Glu], weight_coef=0.5)
connect(stn[stn_Glu], snr[snr_GABA], syn_type=Glu, weight_coef=2)
connect(stn[stn_Glu], gpi[gpi_GABA], syn_type=Glu, weight_coef=2)
connect(gpi[gpi_GABA], thalamus[thalamus_Glu], weight_coef=3)
connect(snr[snr_GABA], thalamus[thalamus_Glu], weight_coef=3)
connect(thalamus[thalamus_Glu], motor[FrontalCortex], syn_type=Glu)
```

## ЗАКЛЮЧЕНИЕ

Создание системы эмоциональной оценки на основе обучения с подкреплением – этап в разработке эмоционального искусственного интеллекта. Эмоциональная оценка неразрывно связана с прогнозированием, которое, в свою очередь, связано с обучением. В ходе работы была составлена и реализована нейронная модель. В результате моделирования получены графики, показывающие активность структур мозга, участвующих в процессе их воздействия друг на друга. Так как воздействие не может быть однозначным, а каждая связь обладает синаптической силой, то такую силу, называемую весом, нужно настроить. В ходе настройки системы мы добились четырех всплесков активности на таламусе вместо ожидаемых пяти. Это произошло из-за подавления бледного шара (*interna*) и черной субстанции (*reticulata*) и восстановления потенциала таламуса.

## СПИСОК ЛИТЕРАТУРЫ

1. *Максим Таланов*. Эмоциональные вычисления. URL: <http://postnauka.ru/video/45297>.
2. *Lowenstein G., Lerner J.S.* The role of affect in decision-making // In R. Davidson, K. Scherer, H. Goldsmith (Eds.) *Handbook of Affective Science*. New York: Oxford University Press, 2003. P. 619–642.
3. *Максим Таланов*. Эмоциональный искусственный интеллект. URL: <http://postnauka.ru/video/45296>.
4. *Tom Ziemke, Robert Lowe*. On the Role of Emotion in Embodied Cognitive Architectures: From Organisms to Robots. Springer Science+Business Media, LLC 2009. P. 71–73.
5. *David Sander, Didier Grandjean, Klaus R. Scherer*. A systems approach to appraisal mechanisms in emotion. Geneva Emotion Research Group, Department of

Psychology, University of Geneva, 2005. P. 140–148.

6. *Petta P.* The role of emotion in a tractable architecture for situated cognizers // In: Trapp R., Petta P., Payr S. Eds. Emotions in Humans and Artifacts. Cambridge, MA: MIT Press, 2003. P. 87–88.

7. *Minsky Marvin.* The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind. Simon and Schuster, 2007. P. 256–258.

8. *Wörgötter F., Porr B.* Temporal Sequence Learning, Prediction, and Control – a Review of different models and their relation to biological mechanisms. Department of Psychology, University of Stirling, 2005. P. 45.

9. *Ortony A., Norman D., Revelle W.* Affect and proto-affect in effective functioning // In: Fellous J-M, Arbib M.A., Eds. Who need emotions? New York: Oxford University Press, 2005.

10. *Damasio A.R.* The feeling of what happens: body, emotion and the making of consciousness. Heinemann: London, 1999. 400 p.

11. *Rolls E.* Emotion explained. Oxford: Oxford University Press, 2005.

12. *Phelps E.* Emotion and cognition: Insights from studies of the human amygdala // Annu. Rev. Psychol. 2006. V. 57. P. 27–53.

13. *Scherer K.R., Ekman P.* On the nature and function of emotion: a component process approach // In: Approaches to Emotion. Hillsdale, N.J.: Lawrence Erlbaum, 1984. P. 293–317.

14. *Paulus Martin P., Angela J.Yu.* Emotion and decision-making: affect-driven belief systems in anxiety and depression // Trends in Cognitive Sciences. September 2012. V. 16, No 9. P. 476–483.

15. *Kahneman D., Tversky A.* Prospect theory: an analysis of decision under risk // Econometrica. 1979. V. 47. P. 263–291.

16. *Mukherjee K.* A dual system model of preferences under risk // Psychol. Rev. 2010. V. 117. P. 243–255.

17. *Hsee C.K., Rottenstreich Y.* Music, pandas, and muggers: on the affective psychology of value // J. Exp. Psychol. Gen. 2004. V. 133. P. 23–30.

18. *Kusev P., van Schaik P.* Preferences under risk: content-dependent behavior and psychological processing // Front. Psychol. 2011. V. 2. P. 269–271.

19. *Breazeal C.* Designing sociable robots. Cambridge, MA: MIT Press, 2002. 244 p.
20. *Kelley A.E.* Neurochemical networks encoding emotion and motivation: An evolutionary perspective // In: Fellous J-M., Arbib M.A., Eds. Who needs emotions? The brain meets the robot. New York: Oxford University Press, 2005.
21. *Max Talanov, Jordi Vallverdu, Salvatore Distefano, Manuel Mazzara, Radhakrishnan Delhibabu.* neuromodulating cognitive architecture: towards biomimetic emotional AI // Advanced Information Networking and Applications (AINA), 2015 IEEE 29th International Conference. P. 587–592.
22. *Аллахвердов В.М., Богданова С.И. и др.* Психология: учеб. / отв. ред. А.А. Крылов. М.: Проспект, 2005. С. 214–217.
23. *Vernon David.* Artificial Cognitive Systems: a Primer. The MIT Press Cambridge, Massachusetts London, England, 2014. 288 p.
24. *McCarthy J., Hayes P.J.* Some philosophical problems from the standpoint of artificial intelligence at the Wayback Machine // In: Meltzer B., Michie D., Eds. Machine Intelligence. Edinburgh: Edinburgh University Press, 1969. No 4. P. 463–502 (archived August 25, 2013).
25. *Таланов Максим.* Марвин Минский и эмоциональные машины. URL: <https://postnauka.ru/faq/58727>.
26. *Lövheim H.* A new three-dimensional model for emotions and monoamine neurotransmitters // Med Hypotheses. 2012. V. 8. P. 341–348.
27. *Tomkins S.* Affect theory // In: P. Ekman, W. Friesen, P. Ellsworth, Eds. Emotions in the Human Face. Cambridge: Cambridge University Press, 1982. P. 355–395.
28. *Smith Craig A., Lazarus Richard S.* Emotion and Adaptation // In: L.A. Pervin, Ed. Handbook of Personality: Theory and Research. New York: Guilford, 1990. P. 609–637.
29. *Lazarus Richard S.* Progress on a cognitive-motivational-relational theory of emotion // American Psychologist. 1991. V. 46, No 8. P. 819–834.
30. *Niv Yael.* Reinforcement learning in the brain // Psychology Department & Princeton Neuroscience Institute, Princeton University, 2009.



31. *Barto A.G.* Adaptive critic and the basal ganglia // In J.C. Houk, J.L. Davis, D.G. Beiser, Eds. Models of information processing in the basal ganglia. Cambridge: MIT Press, 1995. P. 215–232.
  32. *Schultz W., Dayan P., Montague P.R.* A neural substrate of prediction and reward // Science. 1997. No 275. P. 1593–1599.
  33. *Wickens J.R., Kotter R.* Cellular models of reinforcement // In: J.C. Houk, J.L. Davis, D.G. Beiser, Eds. Models of information processing in the basal ganglia. MIT Press, 1995. P. 187–214.
  34. *Barto A.G., Sutton R.S., Watkins C.J.C.H.* Learning and sequential decision making // In: M. Gabriel, J. Moore, Eds. Learning and computational neuroscience: Foundations of adaptive networks. Cambridge, MA: MIT Press, 1990. P. 593–602.
  35. *Bertsekas D.P., Tsitsiklis J.N.* Neuro-dynamic programming. Athena Sc., Scientific, 1996. 512 p.
  36. *Sutton R.S., Barto A.G.* Reinforcement Learning. An Introduction. Bradford Books, MIT Press, Cambridge, MA, 2002 edition, 1998. 320 p.
  37. *Bellman R.E.* Dynamic Programming. Princeton: Princeton University Press, 1957. 392 p.
  38. *Sutton R.S.* Learning to predict by the methods of temporal differences // Machine Learning. August 1988. V. 3, Issue 1. P. 9–44.
  39. *Sutton R.S.* Generalization in reinforcement learning: successful examples using sparse coarse coding // In: D.S. Touretzky, M.C. Mozer, M.E. Hasselmo, Eds. Advances in Neural Information Processing Systems: Proceedings of the 1995 Conference. Cambridge, MA, 1996. P. 1038–1044.
  40. *Rummery G.A.* Problem solving with reinforcement learning. PhD thesis. Cambridge University, Cambridge, 1995. 52 p.
  41. *Watkins C.J.C.H.* Learning from delayed rewards. PhD thesis. University of Cambridge, Cambridge, England, 1989. 234 p. URL: [https://www.cs.rhul.ac.uk/home/chrisw/new\\_thesis.pdf](https://www.cs.rhul.ac.uk/home/chrisw/new_thesis.pdf).
  42. *Watkins C.J.C.H., Dayan P.* Technical note: Q-Learning // Machine Learning. 1992. V. 7, Issue 8. P. 279–292. URL: <http://www.gatsby.ucl.ac.uk/~dayan/papers/cjch.pdf>.
  43. *Pavlov I.P.* Conditioned reflexes. London: Oxford University Press, 1927. URL: <http://s-f-walker.org.uk/pubsebooks/pdfs/Conditioned-Reflexes-Pavlov.pdf>.
-

44. *Воронцов К.В.* Обучение с подкреплением (Reinforcement Learning) URL: <http://www.machinelearning.ru/wiki/images/archive/3/35/20140621071329!Voron-ML-RL-slides.pdf>.
45. *Bellman R.* A Markovian decision process // Journal of Mathematics and Mechanics. 1957. No 6. P. 716–719.
46. *Rescorla R.A., Wagner A.R.* A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement // In: A.H. Black, W.F. Prokasy, Eds. Classical conditioning II: Current research and theory. New York, NY: Appleton-Century-Crofts, 1972. P. 64–99.
47. *Gewaltig Marc-Oliver, Diesmann Markus.* NEST (NEural Simulation Tool) // Scholarpedia. 2007. V. 2, No 4. P. 1430. URL: [http://www.scholarpedia.org/article/NEST\\_\(NEural\\_Simulation\\_Tool\)](http://www.scholarpedia.org/article/NEST_(NEural_Simulation_Tool)).
48. Supercomputers Ready for Use as Discovery Machines for Neuroscience // Frontiers in Neuroinformatics. November 2012. V. 6. P. 1–12.
49. *Diesmann M., Gewaltig M.* NEST: an environment for neural systems simulations // Forschung und wissenschaftliches Rechnen, Beiträge zum Heinz-Billing-Preis. 2001. Bd. 58. S. 43–70.
50. *Picard R.W.* Affective Computing. MIT Press, 1997.
- 

## **THE SYSTEM OF EMOTIONAL APPRAISAL BASED ON REINFORCEMENT LEARNING AND BIO-INSPIRED METHODS**

**E. Mayorova<sup>1</sup>, M. Talanov<sup>2</sup>, R. Lowe<sup>3</sup>**

<sup>1,2</sup> *Kazan (Volga Region) Federal University*

<sup>3</sup> *University of Gothenburg*

<sup>1</sup>[eugeniamaiorova@gmail.com](mailto:eugeniamaiorova@gmail.com); <sup>2</sup>[max.talanov@gmail.com](mailto:max.talanov@gmail.com); <sup>3</sup>[robert.lowe@his.se](mailto:robert.lowe@his.se)

### **Abstract**

I research and lecture in Cognitive Science where my particular interest is in emotions – neural networks modeling and applications – and animal and human learning.

**Keywords:** *appraisal, emotional appraisal, reinforcement learning*

---

## REFERENCES

1. *Talanov M.* Emotional computing. URL: <http://postnauka.ru/video/45297>.
2. *Lowenstein G., Lerner J.S.* The role of affect in decision-making // In R. Davidson, K. Scherer, H. Goldsmith (Eds.) *Handbook of Affective Science*. New York: Oxford University Press, 2003. P. 619–642.
3. *Talanov M.* Emotional artificial intelligence. URL: <http://postnauka.ru/video/45296>
4. *Tom Ziemke, Robert Lowe.* On the Role of Emotion in Embodied Cognitive Architectures: From Organisms to Robots. Springer Science+Business Media, LLC 2009. P. 71–73.
5. *David Sander, Didier Grandjean, Klaus R. Scherer.* A systems approach to appraisal mechanisms in emotion. Geneva Emotion Research Group, Department of Psychology, University of Geneva, 2005. P. 140–148.
6. *Petta P.* The role of emotion in a tractable architecture for situated cognizers // In: Trapp R., Petta P., Payr S. Eds. *Emotions in Humans and Artifacts*. Cambridge, MA: MIT Press, 2003. P. 87–88.
7. *Minsky Marvin.* The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind. Simon and Schuster, 2007. P. 256–258.
8. *Wörgötter F., Porr B.* Temporal Sequence Learning, Prediction, and Control – a Review of different models and their relation to biological mechanisms. Department of Psychology, University of Stirling, 2005. P. 45.
9. *Ortony A., Norman D., Revelle W.* Affect and proto-affect in effective functioning // In: Fellous J-M, Arbib M.A., Eds. *Who need emotions?* New York: Oxford University Press, 2005.
10. *Damasio A.R.* The feeling of what happens: body, emotion and the making of consciousness. Heinemann: London, 1999. 400 p.
11. *Rolls E.* Emotion explained. Oxford: Oxford University Press, 2005.
12. *Phelps E.* Emotion and cognition: Insights from studies of the human amygdala // *Annu. Rev. Psychol.* 2006. V. 57. P. 27–53.

13. *Scherer K.R., Ekman P.* On the nature and function of emotion: a component process approach // In: *Approaches to Emotion*. Hillsdale, N.J.: Lawrence Erlbaum, 1984. P. 293–317.

14. *Paulus Martin P., Angela J.Yu.* Emotion and decision-making: affect-driven belief systems in anxiety and depression // *Trends in Cognitive Sciences*. September 2012. V. 16, No 9. P. 476–483.

15. *Kahneman D., Tversky A.* Prospect theory: an analysis of decision under risk // *Econometrica*. 1979. V. 47. P. 263–291.

16. *Mukherjee K.* A dual system model of preferences under risk // *Psychol. Rev.* 2010. V. 117. P. 243–255.

17. *Hsee C.K., Rottenstreich Y.* Music, pandas, and muggers: on the affective psychology of value // *J. Exp. Psychol. Gen.* 2004. V. 133. P. 23–30.

18. *Kusev P., van Schaik P.* Preferences under risk: content-dependent behavior and psychological processing // *Front. Psychol.* 2011. V. 2. P. 269–271.

19. *Breazeal C.* *Designing sociable robots*. Cambridge, MA: MIT Press, 2002. 244 p.

20. *Kelley A.E.* Neurochemical networks encoding emotion and motivation: An evolutionary perspective // In: *Fellous J-M., Arbib M.A., Eds. Who needs emotions? The brain meets the robot*. New York: Oxford University Press, 2005.

21. *Max Talanov, Jordi Vallverdu, Salvatore Distefano, Manuel Mazzara, Radhakrishnan Delhibabu.* neuromodulating cognitive architecture: towards biomimetic emotional AI // *Advanced Information Networking and Applications (AINA), 2015 IEEE 29th International Conference*. P. 587–592.

22. *Allahverdov V.M., Bogdanova S.I. et al.* *Psychology: textbook*. Prospect Publishing, 2005. P. 214–217.

23. *Vernon David.* *Artificial Cognitive Systems: a Primer*. The MIT Press Cambridge, Massachusetts London, England, 2014. 288 p.

24. *McCarthy J., Hayes P.J.* Some philosophical problems from the standpoint of artificial intelligence at the Wayback Machine // In: *Meltzer B., Michie D., Eds. Machine Intelligence*. Edinburgh: Edinburgh University Press, 1969. No 4. P. 463–502 (archived August 25, 2013).

25. *Talanov M.* Marvin Minsky and the emotional machines. URL: <https://postnauka.ru/faq/58727>.

26. *Lövheim H.* A new three-dimensional model for emotions and monoamine neurotransmitters // *Med Hypotheses*. 2012. V. 8. P. 341–348.
27. *Tomkins S.* Affect theory // In: P. Ekman, W. Friesen, P. Ellsworth, Eds. *Emotions in the Human Face*. Cambridge: Cambridge University Press, 1982. P. 355–395.
28. *Smith Craig A., Lazarus Richard S.* Emotion and Adaptation // In: L.A. Pervin, Ed. *Handbook of Personality: Theory and Research*. New York: Guilford, 1990. P. 609–637.
29. *Lazarus Richard S.* Progress on a cognitive-motivational-relational theory of emotion // *American Psychologist*. 1991. V. 46, No 8. P. 819–834.
30. *Niv Yael.* Reinforcement learning in the brain // Psychology Department & Princeton Neuroscience Institute, Princeton University, 2009.
31. *Barto A.G.* Adaptive critic and the basal ganglia // In J.C. Houk, J.L. Davis, D.G. Beiser, Eds. *Models of information processing in the basal ganglia*. Cambridge: MIT Press, 1995. P. 215–232.
32. *Schultz W., Dayan P., Montague P.R.* A neural substrate of prediction and reward // *Science*. 1997. No 275. P. 1593–1599.
33. *Wickens J.R., Kotter R.* Cellular models of reinforcement // In: J.C. Houk, J.L. Davis, D.G. Beiser, Eds. *Models of information processing in the basal ganglia*. MIT Press, 1995. P. 187–214.
34. *Barto A.G., Sutton R.S., Watkins C.J.C.H.* Learning and sequential decision making // In: M. Gabriel, J. Moore, Eds. *Learning and computational neuroscience: Foundations of adaptive networks*. Cambridge, MA: MIT Press, 1990. P. 593–602.
35. *Bertsekas D.P., Tsitsiklis J.N.* *Neuro-dynamic programming*. Athena Sc., Scientific, 1996. 512 p.
36. *Sutton R.S., Barto A.G.* *Reinforcement Learning. An Introduction*. Bradford Books, MIT Press, Cambridge, MA, 2002 edition, 1998. 320 p.
37. *Bellman R.E.* *Dynamic Programming*. Princeton: Princeton University Press, 1957. 392 p.
38. *Sutton R.S.* Learning to predict by the methods of temporal differences // *Machine Learning*. August 1988. V. 3, Issue 1. P. 9–44.
39. *Sutton R.S.* Generalization in reinforcement learning: successful examples

using sparse coarse coding // In: D.S. Touretzky, M.C. Mozer, M.E. Hasselmo, Eds. Advances in Neural Information Processing Systems: Proceedings of the 1995 Conference. Cambridge, MA, 1996. P. 1038–1044.

40. *Rummery G.A.* Problem solving with reinforcement learning. PhD thesis. Cambridge University, Cambridge, 1995. 52 p.

41. *Watkins C.J.C.H.* Learning from delayed rewards. PhD thesis. University of Cambridge, Cambridge, England, 1989. 234 p. URL: [https://www.cs.rhul.ac.uk/home/chrisw/new\\_thesis.pdf/](https://www.cs.rhul.ac.uk/home/chrisw/new_thesis.pdf/).

42. *Watkins C.J.C.H., Dayan P.* Technical note: Q-Learning // Machine Learning. 1992. V. 7, Issue 8. P. 279–292. URL: <http://www.gatsby.ucl.ac.uk/~dayan/papers/cjch.pdf>.

43. *Pavlov I.P.* Conditioned reflexes. London: Oxford University Press, 1927. URL: <http://s-f-walker.org.uk/pubsebooks/pdfs/Conditioned-Reflexes-Pavlov.pdf>.

44. *Vorontsov K.V.* Reinforcement Learning. URL: <http://www.machinelearning.ru/wiki/images/archive/3/35/20140621071329!Voron-ML-RL-slides.pdf>.

45. *Bellman R.* A Markovian decision process // Journal of Mathematics and Mechanics. 1957. No 6. P. 716–719.

46. *Rescorla R.A., Wagner A.R.* A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement // In: A.H. Black, W.F. Prokasy, Eds. Classical conditioning II: Current research and theory. New York, NY: Appleton-Century-Crofts, 1972. P. 64–99.

47. *Gewaltig Marc-Oliver, Diesmann Markus.* NEST (NEural Simulation Tool) // Scholarpedia. 2007. V. 2, No 4. P. 1430. URL: [http://www.scholarpedia.org/article/NEST\\_\(NEural\\_Simulation\\_Tool\)](http://www.scholarpedia.org/article/NEST_(NEural_Simulation_Tool)).

48. Supercomputers Ready for Use as Discovery Machines for Neuroscience // Frontiers in Neuroinformatics. November 2012. V. 6. P. 1–12.

49. *Diesmann M., Gewaltig M.* NEST: an environment for neural systems simulations // Forschung und wissenschaftliches Rechnen, Beiträge zum Heinz-Billing-Preis. 2001. Bd. 58. S. 43–70.

50. *Picard R.W.* Affective Computing. MIT Press, 1997.

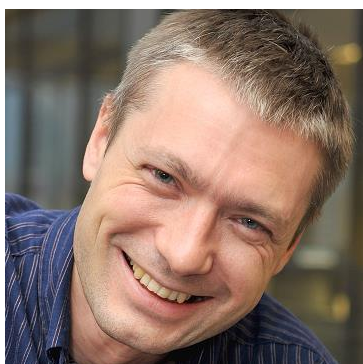
## СВЕДЕНИЯ ОБ АВТОРАХ



**МАЙОРОВА Евгения Юрьевна** – выпускница 2016 года Высшей школы информационных технологий и информационных систем Казанского (Приволжского) федерального университета.

**Eugeniya Yurievna MAYOROVA**, bachelor of Higher School Information technologies and systems of Kazan Federal University.

email: eugeniamaiorova@gmail.com



**ТАЛАНОВ Максим Олегович** – кандидат технических наук, руководитель Лаборатории машинного понимания Высшей школы информационных технологий и информационных систем Казанского (Приволжского) федерального университета.

**Maxim Olegovich TALANOV**, PhD, head of Machine Cognition Lab in Higher Institute of Information Technology in Kazan (Volga Region) Federal University.

email: max.talanov@gmail.com



**ЛОУ Роберт** – доктор, доцент и глава исследовательской лаборатории ICE Lab в университете Гётеборга.

**Robert LOWE** is the docent and research leader at the ICE Lab and at the Interaction Lab at the University of Gothenburg.

email: robert.lowe@his.se

*Материал поступил в редакцию 15 мая 2016 года*