

УДК 004.81

DOI: 10.25559/SITITO.16.202002.331-337

Естественный способ преодоления катастрофической забывчивости нейронных сетей

А. А. Куталев

АО «ИнфоВотч», г. Москва, Россия

121357, Россия, г. Москва, ул. Верейская, д. 29, стр. 134

kutalev@yandex.ru

Аннотация

Проблема катастрофической забывчивости проявилась в моделях нейронных сетей на базе коннекционистского подхода, которые активно исследуются начиная со второй половины 20-го века. Предпринимались многочисленные попытки и были предложены различные способы решения этой проблемы, но до самого последнего времени значимых успехов достичь не удавалось. В 2016 году случился значительный прорыв – группа ученых из DeepMind предложила метод эластичного закрепления весов (EWC), который позволяет успешно бороться с проблемой катастрофической забывчивости. К сожалению, хотя нам известны случаи использования этого метода в реальных задачах, он пока не получил повсеместного распространения. В этой работе мы хотим предложить альтернативные подходы к преодолению катастрофической забывчивости, основанные на суммарном абсолютном сигнале, прошедшем через связь в нейронной сети, которые демонстрируют схожую с EWC эффективность и, при этом, имеют существенно меньшую вычислительную стоимость. Эти подходы имеют более простую реализацию и представляются нам по своей сути более близкими к процессам, происходящим в мозге животных для сохранения выученных ранее навыков при последующем обучении. Мы надеемся, что простота реализации этих методов послужит их более широкому применению.

Ключевые слова: нейронные сети, катастрофическая забывчивость, эластичное закрепление весов, алгоритм обратного распространения, суммарный абсолютный сигнал.

Для цитирования: Куталев, А. А. Естественный способ преодоления катастрофической забывчивости нейронных сетей / А. А. Куталев. – DOI 10.25559/SITITO.16.202002.331-337 // Современные информационные технологии и ИТ-образование. – 2020. – Т. 16, № 2. – С. 331-337.

© Куталев А. А., 2020



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.



Natural Way to Overcome Catastrophic Forgetting in Neural Networks

A. A. Kutalev

JSC InfoWatch, Moscow, Russia
29 Vereiskaya St. b. 134, Moscow 121357, Russia
kutalev@yandex.ru

Abstract

The problem of catastrophic forgetting manifested itself in models of neural networks based on the connectionist approach, which have been actively studied since the second half of the 20th century. Numerous attempts have been made and various ways to solve this problem have been proposed, but until very recently substantial successes have not been achieved. In 2016, a significant breakthrough occurred – a group of scientists from DeepMind proposed the method of elastic weight consolidation (EWC), which allows us to successfully overcome the problem of catastrophic forgetting. Unfortunately, although we were aware about the cases of using this method in real tasks, it has not yet obtained widespread distribution. In this paper, we want to propose alternative approaches for overcoming catastrophic forgetting, based on the total absolute signal passed through the connection. These approaches demonstrate similar efficiency as EWC and, at the same time, have less computational complexity. These approaches have a simpler implementation and seem to us to be essentially closer to the processes occurring in the brain of animals to preserve previously learned skills during subsequent training. We hope that the ease of implementation of these methods will serve their wider application.

Keywords: neural network, catastrophic forgetting, elastic weight consolidation, back propagation, total absolute signal.

For citation: Kutalev A.A. Natural Way to Overcome Catastrophic Forgetting in Neural Networks. *Sovremennye informacionnye tehnologii i IT-obrazovanie* = Modern Information Technologies and IT-Education. 2020; 16(2):331-337. DOI: <https://doi.org/10.25559/SITITO.16.202002.331-337>



Введение

Проблема катастрофической забывчивости ([1], [2]) оказывает влияние на множество задач в современном машинном обучении, мешая сохранять навыки предобученных нейронных сетей (далее НС) при последующем обучении. Например, при адаптации продуктов, являющихся передовыми в индустрии ML (таких, как BERT, GPT и др.), к прикладным задачам конкретного разработчика. При этом происходит дообучение нейросети на специализированных обучающих наборах данных без доступа к оригинальным наборам данных, на которых проводилось первичное обучение НС.

Суть проблемы катастрофической забывчивости состоит в том, что нейронная сеть, обученная на некотором обучающем наборе **A**, при дальнейшем обучении на другом обучающем наборе **B** в отсутствие набора **A** быстро теряет навык, полученный при обучении на **A**. Такое поведение совсем не похоже на поведение высших живых организмов, которые способны длительное время сохранять выученные навыки при обучении другим задачам. Это несоответствие побуждает нас искать способы преодоления указанной проблемы.

Из проведенных нейрофизиологических исследований и исследований в машинном обучении ([1] – [5]) можно сделать вывод, что корень проблемы в том, что обучение на наборе **B** изменяет связи нейросети, важные для сохранения навыка, полученного при обучении на наборе **A**. Логично было бы попробовать сохранить важные для выученного навыка связи каким-либо образом при дальнейшем обучении на другом обучающем наборе данных.

Относительно недавно, в 2016 году, группой ученых из DeepMind был предложен следующий этой логике метод преодоления проблемы катастрофической забывчивости [5], основанный на эластичном закреплении весов (EWC). Однако, до сих пор этот метод не получил повсеместного распространения.

В этой работе мы хотим предложить метод преодоления проблемы катастрофической забывчивости, который является развитием метода эластичного закреплении весов, представленного в [5]. В отличие от EWC «важность» веса связи в нашем алгоритме базируется не на диагональных элементах информационной матрицы Фишера, а на суммарном абсолютном сигнале, прошедшем через связь в процессе обработки сетью обучающих примеров после завершения цикла обучения. Как и EWC, наш подход имеет линейную вычислительную стоимость по количеству параметров сети и количеству примеров в обучающем наборе. Однако, он позволяет избежать построения дополнительных вычислительных графов (для вычисления диагонали матрицы Фишера) и может быть интегрирован непосредственно в процесс обучения нейронной сети.

Также мы рассмотрим случай, для которого метод эластичного закреплении весов терпит фиаско.

Результаты

Алгоритм, предложенный в [5], показывает впечатляющую способность сохранять навыки, полученный на одном обучающем наборе данных, при обучении на других наборах данных. В то же время механизм решения проблемы катастрофической забывчивости, используемый живыми организмами, наверняка имеет другую природу: сложно представить, что каждый отдельный дендрит или синапс нервной системы животного

способен рассчитывать соответствующий элемент матрицы Фишера. Это заставляет искать другие методы сохранения навыка в нейронной сети.

Следуя логике метода эластичного закреплении весов EWC, каждому весу связи w_i в НС сопоставляется его «важность» для навыка, полученного при обучении на первом обучающем наборе **A**. «Важность» веса w_i показывает насколько сильно будет штрафовать изменение веса w_i при обучении на следующем обучающем наборе **B**. То есть, изменение веса w_i штрафуются тем сильнее, чем больше была его «важность».

Штраф реализуется добавлением регуляризатора в функцию потерь:

$$L = L_B + \frac{\lambda}{2} \sum_i F_i (w_i - w_{A,i}^*)^2,$$

где L_B – функция потерь при обучении на наборе **B** без закреплении весов, $w_{A,i}^*$ – веса нейронной сети после обучения на наборе **A**, а в качестве «важности» i -го веса в EWC выступает F_i – соответствующий i -й диагональный элемент информационной матрицы Фишера. Теоретическое обоснование и подробности вычисления F_i можно найти в [5].

В качестве альтернативы F_i мы предлагаем использовать другую величину, характеризующую «важность» связи в нейронной сети. А именно, суммарный абсолютный сигнал, прошедший через связь при обработке обучающей сетью всех примеров из обучающего набора **A**:

$$S_{A,i}^w = \frac{1}{n} \sum_k |x_{k,i} w_{A,i}|,$$

где $x_{k,i}$ – сигнал, поданный на вход i -й связи при обработке нейронной сетью k -го примера, n – количество примеров в наборе **A**, а $w_{A,i}$ – вес i -й связи в обученной на **A** нейронной сети. В качестве «важности» j -го смещения нейронной сети возьмем величину:

$$S_{A,j}^b = \frac{1}{n} \sum_k |y_{k,j}|,$$

где $y_{k,j}$ – выходной сигнал (активация) j -го нейрона, к сумматору которого относится смещение b_j .

Таким образом, функция потерь при обучении на последующем наборе данных **B** приобретает вид:

$$L = L_B + \frac{\lambda}{2} \sum_i S_{A,i}^w (w_i - w_{A,i}^*)^2 + \frac{\lambda}{2} \sum_j S_{A,j}^b (b_j - b_{A,j}^*)^2,$$

где $S_{A,i}^w$ и $S_{A,i}^b$ «важности» весов и смещений соответственно, полученные после обучения на наборе **A**.

При дальнейшем обучении на наборе данных **C** мы должны добавить соответствующие члены регуляризации. Тогда функция потерь при обучении на **C** примет вид:

$$L = L_C + \frac{\lambda}{2} \sum_i (S_{A,i}^w + S_{B,i}^w) (w_i - w_{B,i}^*)^2 + \frac{\lambda}{2} \sum_j (S_{A,j}^b + S_{B,j}^b) (b_j - b_{B,j}^*)^2.$$

То есть, при последовательном обучении «важность» весов и смещений накапливается:

$$S_{AB,i} = S_{A,i}^w + S_{B,i}^w,$$

и так далее по всем последующим обучающим наборам.

Необходимо отметить, что в оригинальной статье [5], описывающей алгоритм эластичного закреплении весов, при последовательном обучении на нескольких обучающих наборах применяется добавление к функции потерь регуляризатора по каждому завершеному обучающему набору. Что приводит к



дополнительному расходу памяти и вычислительных ресурсов по мере увеличения количества последовательных обучений. В [6] была показана математическая некорректность такого подхода, и указана корректность накопления путем простого суммирования «важностей» весов, рассчитанных как диагональные элементы матрицы Фишера, для применения с единственным регуляризатором, использующим коэффициенты $w_{N_i}^*$, полученные по завершении обучения на последнем обучающем наборе перед текущим. В случае же использования суммарного абсолютного сигнала в качестве «важности» веса суммирование накопленных «важностей» при последовательном обучении появляется естественным образом из самого определения суммарного абсолютного сигнала.

При обучении методом градиентного спуска изменения весов сети на шаге обучения имеют вид:

$$w_i' = w_i - \alpha \nabla_i,$$

где α – заданная константа скорости обучения,

$\nabla_i = \frac{\partial L}{\partial w_i}$ – градиент функции потерь по весам связей. Вместо

добавления регуляризатора в функцию потерь L мы можем на каждом шаге обучения ослаблять градиент функции потерь пропорционально величине «важности» связи S_i^w , накопленной при обучении на предыдущих наборах данных. В случае нулевой «важности» весов, то есть при обучении на первом наборе данных, градиент должен иметь коэффициент 1 и убывать по мере роста «важности» связи. Для выполнения этих условий используем множитель

$$\frac{1}{1 + \lambda S_i^w}:$$

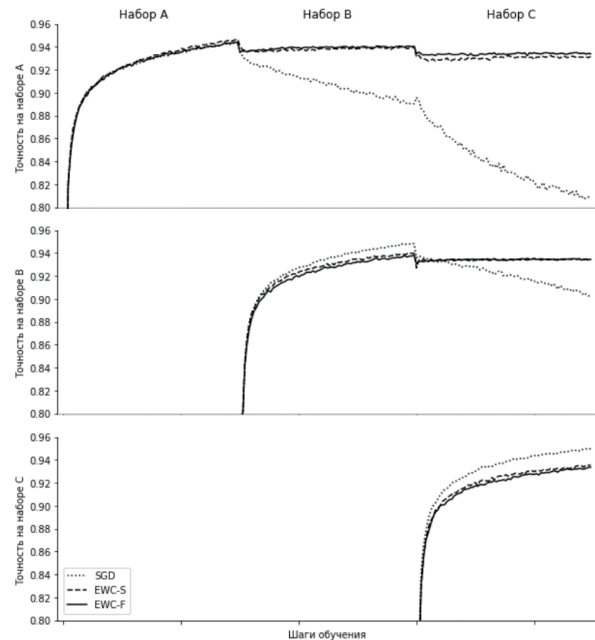
$$w_i' = w_i - \frac{\alpha}{1 + \lambda S_i^w} \nabla_i.$$

Для обучения смещений аналогично:

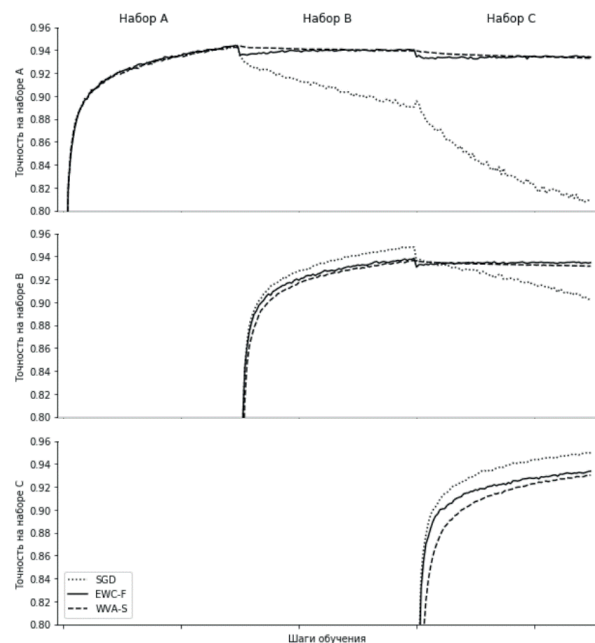
$$b_j' = b_j - \frac{\alpha}{1 + \lambda S_j^b} \nabla_j.$$

Мы назвали такой подход методом ослабления скорости весов (WVA – Weight Velocity Attenuation). В методе WVA в качестве «важности» веса или смещения вместо суммарного абсолютного сигнала S могут также выступать и F – соответствующие весу или смещению диагональные элементы матрицы Фишера. Таким образом, мы получим четыре возможных сочетания методов закрепления весов и замедления градиента с используемыми «важностями» весов на основе матрицы Фишера и суммарном абсолютном сигнале.

Для проверки этих методов мы провели эксперименты по последовательному обучению глубоких нейронных сетей с различным числом полносвязных слоев на нескольких обучающих наборах данных. Каждый из этих наборов был получен из набора данных MNIST случайным перемешиванием входов одинаково для всех примеров в наборе аналогично тому, как это сделано в [5].



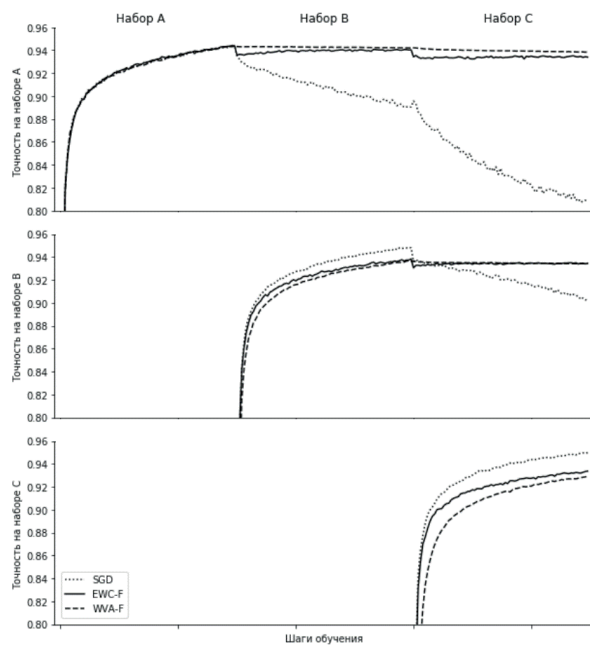
Р и с. 1. Сравнение EWC-S с EWC-F и SGD
F i g. 1. Comparison of EWC-S with EWC-F and SGD



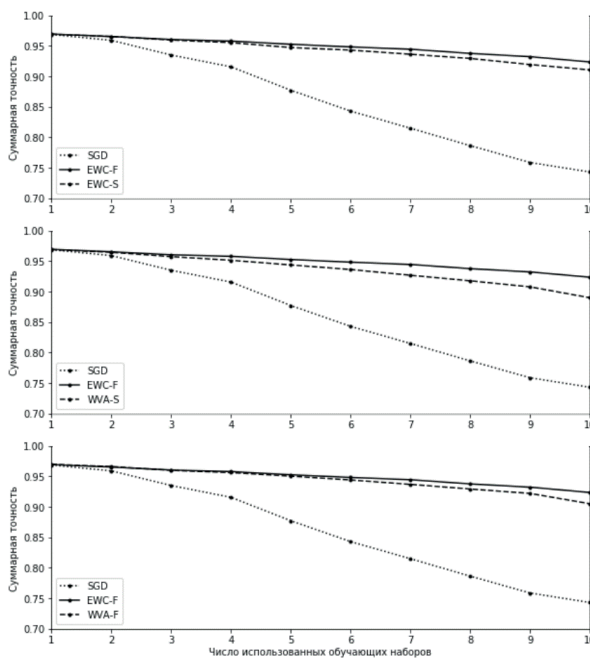
Р и с. 2. Сравнение WVA-S с EWC-F и SGD
F i g. 2. Comparison of WVA-S with EWC-F and SGD

Наши эксперименты показали, что все четыре вышеописанных метода демонстрируют почти одинаковую способность НС сохранять навыки при последовательном обучении нескольким обучающим наборам. Рисунки 1, 2 и 3 демонстрируют изменение точности на трех обучающих наборах А, В и С при последовательном обучении на этих наборах с использованием различных методов преодоления катастрофической забывчивости НС. Рисунок 1 позволяет сравнить метод эластичного за-





Р и с. 3. Сравнение WVA-F с EWC-F и SGD
 Fig. 3. Comparison of WVA-F with EWC-F and SGD



Р и с. 4. Сравнение EWC-S, WVA-S и WVA-F с EWC-F при последовательном обучении 10 наборам
 Fig. 4. Comparison of EWC-S, WVA-S, and WVA-F with EWC-F and SGD in 10-Set Sequential Training

крепления весов на основе суммарного абсолютного сигнала (EWC-S) с методом эластичного закрепления весов на основе диагональных элементов информационной матрицы Фишера (EWC-F) и с простым стохастическим градиентным спуском (SGD). Аналогично, рисунок 2 служит сравнению метода замедления скорости весов на основе суммарного абсолютного сигнала (WVA-S) с методами EWC-F и SGD, а рисунок 3 – сравнению метода замедления скорости весов на основе диагональ-

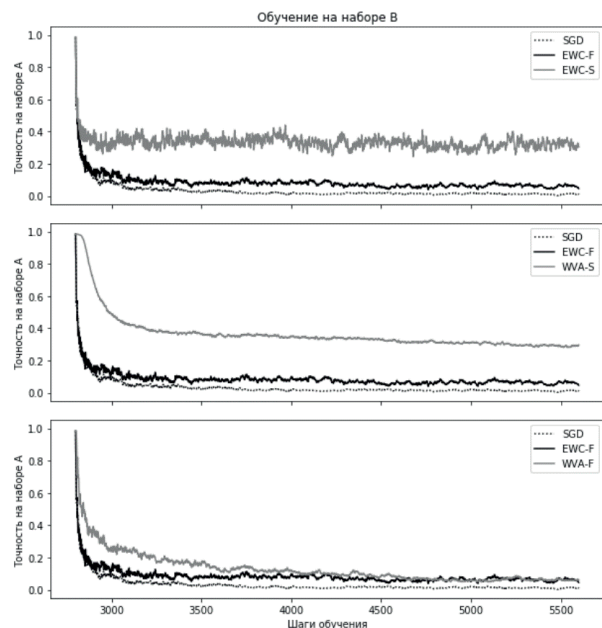
ных элементов информационной матрицы Фишера (WVA-F) с методами EWC-F и SGD. Графики на этих рисунках усреднены по 10 проходам. Видно, что при использовании простого градиентного спуска (SGD) точность быстро деградирует, в то время как остальные методы хорошо сохраняют навык. Рисунок 4 иллюстрирует деградацию суммарной точности на всех использованных наборах при последовательном обучении на десяти различных наборах. График на рисунке также усреднен по 10 проходам.

Эмпирическое наблюдение: для всякого метода с параметром λ_1 и элементами матрицы Фишера в качестве «важности» весов, нам всегда удавалось подобрать такой коэффициент λ_2 , что этот же метод на основе суммарного абсолютного сигнала с параметром λ_2 показывал такое же сохранение навыка при последовательном обучении на нескольких обучающих наборах.

Также, по нашим наблюдениям, при прочих равных условиях, методы ослабления скорости весов лишь слегка уступают методам закрепления весов. Что интуитивно предсказуемо, так как закрепление весов не дает важным весам сети уходить далеко от «правильных» значений. В то время как при замедлении скорости более «важные» веса могут удаляться от «правильных» значений сколь угодно далеко, просто это происходит существенно медленнее, чем для менее «важных» весов.

Наблюдения

Все вышеописанные методы решают проблему катастрофической забывчивости нейронных сетей лишь в случае, когда каждый из наборов данных в последовательном обучении включает примеры с активацией каждого из выходов нейронной сети. Если же в наборе содержатся примеры лишь части классов, распознаваемых нейронной сетью, то катастрофическая забывчивость быстро уничтожает предыдущие навыки при последовательном обучении даже при использовании методов закрепления весов или ослабления скорости весов.



Р и с. 5. Деградация точности на наборе А при обучении на наборе В
 Fig. 5. Degradation of precision on set A when trained on set B



Иллюстрацию проблемы можно увидеть на рисунке 3 (график усреднен по 10 проходам), где показана деградация точности при обучении на втором наборе **B** с использованием различных методов сохранения навыка, следующим за обучением на первом наборе **A**. Обучающие наборы **A** и **B** получены из MNIST: **A** содержит только примеры с цифрами 0-4, **B** примеры с цифрами 5-9. Видно, что лишь при использовании методов EWC-S и WVA-S, основанных на суммарном абсолютном сигнале, деградация навыка, полученного при обучении на **A**, идет существенно медленнее при обучении на **B** чем при использовании EWC-F и WVA-F. А при использовании метода замедления скорости весов на суммарном абсолютном сигнале WVA-S можно получить суммарную точность около 75% с использованием ранней остановки (что, однако, требует использования тестовой части набора **A** при обучении на **B**). В любом случае точность на наборе **A** деградирует существенно сильнее, чем при использовании «полных» обучающих наборов, включающих примеры каждого из классов, распознаваемых нейронной сетью.

Такое поведение при обучении НС сильно отличается от поведения обучения животных. Это можно показать на простом примере: кошка может учиться ходить и шевелить ушами последовательно и независимо. Тогда при обучении ходьбе не будут активироваться выходы ее мозга, управляющие ушами. А при обучении шевелению ушами не активируются выходы, управляющие лапами. Тем не менее, при обучении шевелению ушами навык ходьбы не деградирует существенно.

Частичное решение проблемы «неполных» обучающих наборов было предложено в [11], где при обучении на «неполном» обучающем наборе (split MNIST) кросс-энтропийная функция потерь применялась только для подмножества выходов сети, активируемых в этом наборе. Но такое решение, к сожалению, нельзя назвать универсальным.

Аналогично нашей работе в работах [11] и [12] также предложены альтернативные способы вычисления «важности» весов (SI в [11] и MAS в [12]), которые по результатам экспериментов в этих работах превосходят классический EWC из [5]. Все они, однако, требуют хранения как минимум 2^*N дополнительных параметров (где N – количество параметров нейронной сети), что при современных размерах нейронных сетей может ограничить область их применения, а также создания дополнительного вычислительного графа и расчета по нему «важностей» весов. В то же время метод WVA, представленный в этой работе, требует хранения лишь N дополнительных параметров («важностей» весов) и не требует дополнительного вычислительного графа с производными.

Рассуждения и выводы

То, что предложенные методы, основанные на суммарном абсолютном сигнале, прошедшем через связь, помогают сохранять навыки сети при последовательном обучении, позволяет выдвинуть гипотезу, что обучение в нервной системе человека и животных происходит по аналогичной схеме. Сначала сеть обучается на предоставленных примерах. Затем, когда на нескольких предоставленных сети примерах получен более чем удовлетворительный результат, происходит закрепление наиболее важных связей между нейронами. Это закрепление может происходить, например, с помощью выброса подкрепляющего гормона, который делает менее пластичными

связи, через которые прошло наибольшее число активаций. Судя по исследованиям [7] и [8], аналогичным образом ведет себя миелин – гормон, вырабатываемое количество которого связано с электрической активностью нейрона. То есть, чем больше активаций произойдет с нейроном, тем более толстым слоем покроет его миелиновая оболочка. Известно, что миелинизация нейронов служит ускорению прохождения сигнала по ним, но влияние миелиновой оболочки на пластичность связей нейрона пока не исследована.

Основываясь на проведенных экспериментах с замедлением обучения связей можно сделать предположение, что обучение мозга животного конкретной задаче (обучающему набору) приводит к образованию в мозге целой выделенной подсети, решающей эту задачу. Таким образом, в результате последовательного обучения нескольким задачам, мозг становится комплексом из подсетей, каждая из которых обучена решать отдельную задачу. Но эти подсети не разделены, то есть могут иметь общие нейроны и связи, и, таким образом, использовать навыки друг друга.

References

- [1] French R.M. Catastrophic forgetting in connectionist networks. *Trends in Cognitive Science*. 1999; 3(4):128-135. (In Eng.) DOI: [https://doi.org/10.1016/S1364-6613\(99\)01294-2](https://doi.org/10.1016/S1364-6613(99)01294-2)
- [2] McCloskey M., Cohen N.J. Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. *Psychology of Learning and Motivation*. 1989; 24:109-165. (In Eng.) DOI: [https://doi.org/10.1016/S0079-7421\(08\)60536-8](https://doi.org/10.1016/S0079-7421(08)60536-8)
- [3] McClelland J.L., McNaughton B.L., O'Reilly R.C. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*. 1995; 102(3):419-457. (In Eng.) DOI: <https://doi.org/10.1037/0033-295X.102.3.419>
- [4] Goodfellow I.J., Mirza M., Xiao D., Courville A.C., Bengio Y. An Empirical Investigation of Catastrophic Forgetting in Gradient-Based Neural Networks. *arXiv:1312.6211*. 2013. Available at: <https://arxiv.org/abs/1312.6211> (accessed 04.09.2020). (In Eng.)
- [5] Kirkpatrick J., Pascanu R., Rabinowitz N., Veness J., Desjardins G., Rusu A.A., Milan K., Quan J., Ramalho T., Grabska-Barwinska A., Hassabis D., Clopath C., Kumaran D., Hadsell R. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*. 2017; 114(13):3521-3526. (In Eng.) DOI: <https://doi.org/10.1073/pnas.1611835114>
- [6] Huszár F. Note on the quadratic penalties in elastic weight consolidation. *Proceedings of the National Academy of Sciences*. 2018; 115(11):E2496-E2497. (In Eng.) DOI: <https://doi.org/10.1073/pnas.1717042115>
- [7] Wake H., Lee P.R., Fields R. D. Control of Local Protein Synthesis and Initial Events in Myelination by Action Potentials. *Science*. 2011; 333(6049):1647-1651. (In Eng.) DOI: <https://doi.org/10.1126/science.1206998>
- [8] Miller D.J., Duka T., Stimpson C.D., Schapiro S.J., Baze W.B., McArthur M.J., Fobbs A.J., Sousa A.M., Šestan N., Wildman D.E., Lipovich L., Kuzawa C.W., Hof P.R., Sherwood



- C.C. Prolonged myelination in human neocortical evolution. *Proceeding of the National Academy of Sciences*. 2012; 109(41):16480-16485. (In Eng.) DOI: <https://doi.org/10.1073/pnas.1117943109>
- [9] Zacarias A., Alexandre L.A. SeNA-CNN: Overcoming Catastrophic Forgetting in Convolutional Neural Networks by Selective Network Augmentation. In: Pancioni L., Schwenger F., Trentin E. (ed.) *Artificial Neural Networks in Pattern Recognition. ANNPR 2018. Lecture Notes in Computer Science*. 2018; 11081:102-112. Springer, Cham. (In Eng.) DOI: https://doi.org/10.1007/978-3-319-99978-4_8
- [10] Li H., Barnaghi P., Enshaeifar S., Ganz F. Continual Learning Using Task Conditional Neural Networks. *arXiv:2005.05080*. 2020. Available at: <https://arxiv.org/abs/2005.05080> (accessed 04.09.2020). (In Eng.)
- [11] Zenke F., Poole B., Ganguli S. Continual Learning Through Synaptic Intelligence. *Proceedings of the 34th International Conference on Machine Learning. PMLR*. 2017; 70:3987-3995. International Convention Centre, Sydney, Australia. Available at: <http://proceedings.mlr.press/v70/zenke17a.html> (accessed 04.09.2020). (In Eng.)
- [12] Aljundi R., Babiloni F., Elhoseiny M., Rohrbach M., Tuytelaars T. Memory Aware Synapses: Learning What (not) to Forget. In: Ferrari V., Hebert M., Sminchisescu C., Weiss Y. (ed.) *Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science*. 2018; 11207:144-161. Springer, Cham. (In Eng.) DOI: https://doi.org/10.1007/978-3-030-01219-9_9
- [13] Thangarasa V., Miconi T., Taylor G.W. Enabling Continual Learning with Differentiable Hebbian Plasticity. In: *2020 International Joint Conference on Neural Networks (IJCNN)*. Glasgow, United Kingdom; 2020, p. 1-8. (In Eng.) DOI: <https://doi.org/10.1109/IJCNN48605.2020.9206764>
- [14] Kumaran D., Hassabis D., McClelland J.L. What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. *Trends in Cognitive Sciences*. 2016; 20(7):512-534. (In Eng.) DOI: <https://doi.org/10.1016/j.tics.2016.05.004>
- [15] Miconi T., Stanley K.O., Clune J. Differentiable plasticity: training plastic neural networks with backpropagation. *Proceedings of the 35th International Conference on Machine Learning. PMLR*. 2018; 80:3559-3568. Available at: <http://proceedings.mlr.press/v80/miconi18a.html> (accessed 04.09.2020). (In Eng.)
- [16] Zenke F., Gerstner W., Ganguli S. The temporal paradox of Hebbian learning and homeostatic plasticity. *Current Opinion in Neurobiology*. 2017; 43:166-176. (In Eng.) DOI: <https://doi.org/10.1016/j.conb.2017.03.015>
- [17] Li Z., Hoiem D. Learning without Forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2018; 40(12):2935-2947. (In Eng.) DOI: <https://doi.org/10.1109/TPAMI.2017.2773081>
- [18] Parisi G.I., Kemker R., Part J.L., Kanan C., Wermter S. Continual lifelong learning with neural networks: A review. *Neural Networks*. 2019; 113:54-71. (In Eng.) DOI: <https://doi.org/10.1016/j.neunet.2019.01.012>
- [19] Masse N.Y., Grant G.D., Freedman D.J. Alleviating catastrophic forgetting using context-dependent gating and synaptic stabilization. *Proceedings of the National Academy of Sciences*. 2018; 115(44):E10467-E10475. (In Eng.) DOI: <https://doi.org/10.1073/pnas.1117943109>
- [20] Mirzadeh S.I., Farajtabar M., Ghasezadeh H. Dropout as an Implicit Gating Mechanism For Continual Learning. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA; 2020. p. 945-951. (In Eng.) DOI: <https://doi.org/10.1109/CVPRW50498.2020.00124>
- [21] Soltoggio A., Stanley K.O., Risi S. Born to learn: The inspiration, progress, and future of evolved plastic artificial neural networks. *Neural Networks*. 2018; 108:48-67. (In Eng.) DOI: <https://doi.org/10.1016/j.neunet.2018.07.013>
- [22] Song S., Miller K.D., Abbott L.F. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*. 2000; 3:919-926. (In Eng.) DOI: <https://doi.org/10.1038/78829>
- [23] Lee K., Lee K., Shin J., Lee H. Overcoming Catastrophic Forgetting With Unlabeled Data in the Wild. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea (South); 2019. p. 312-321. (In Eng.) DOI: <https://doi.org/10.1109/ICCV.2019.00040>
- [24] Rostami M., Kolouri S., Pilly P.K. Complementary Learning for Overcoming Catastrophic Forgetting Using Experience Replay. In: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*. IJCAI; 2019. p. 3339-3345. (In Eng.) DOI: <https://doi.org/10.24963/ijcai.2019/463>
- [25] Schak M., Gepperth A. A Study on Catastrophic Forgetting in Deep LSTM Networks. In: Tetko I., Kůrková V., Karpov P., Theis F. (ed.) *Artificial Neural Networks and Machine Learning – ICANN 2019: Deep Learning*. ICANN 2019. *Lecture Notes in Computer Science*. 2019; 11728:714-728. Springer, Cham. (In Eng.) DOI: https://doi.org/10.1007/978-3-030-30484-3_56
- [26] Ribeiro J., Melo F.S., Dias J. Multi-task Learning and Catastrophic Forgetting in Continual Reinforcement Learning. In: Calvanese D., Iocchi L. (ed.) *GCAI 2019. Proceedings of the 5th Global Conference on Artificial Intelligence*. 2019; 65:163-175. (In Eng.) DOI: <https://doi.org/10.29007/g7bg>

Поступила 04.09.2020; принята к публикации 25.09.2020;
опубликована онлайн 30.09.2020.
Submitted 04.09.2020; revised 25.09.2020;
published online 30.09.2020.

Об авторе:

Куталев Алексей Анатольевич, специалист, программист-исследователь отдела прогнозирования, АО «ИнфоВотч» (121357, Россия, г. Москва, ул. Верейская, д. 29, стр. 134), ORCID: <http://orcid.org/0000-0003-2695-792X>, kutalev@yandex.ru

Автор прочитал и одобрил окончательный вариант рукописи.

About the author:

Alexey A. Kutalev, Specialist, Software Developer of the Prediction Department, JSC InfoWatch (29 Vereiskaya St. b. 134, Moscow 121357, Russia), ORCID: <http://orcid.org/0000-0003-2695-792X>, kutalev@yandex.ru

The author has read and approved the final manuscript.

