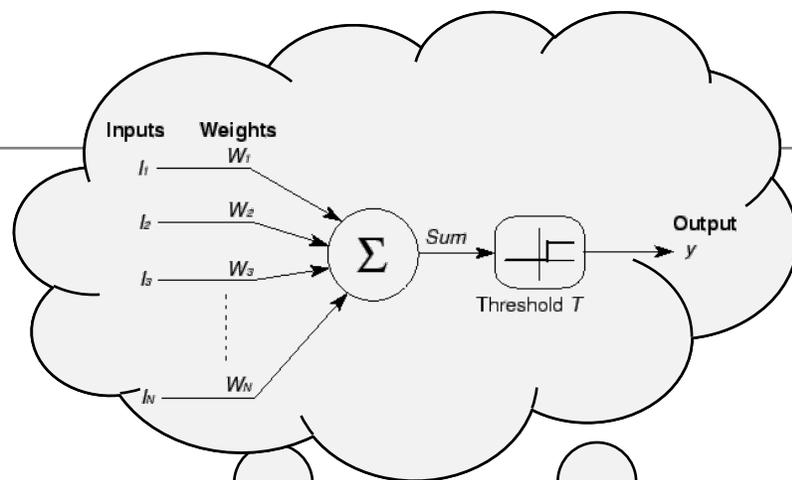


Нейронные сети

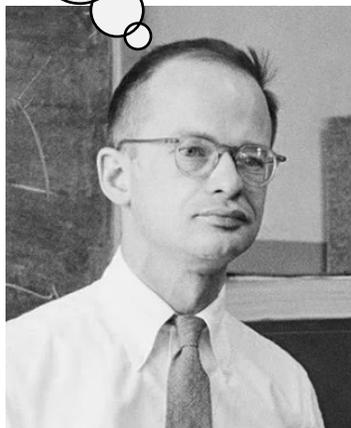
История нейронных сетей и глубокого обучения

Лекция 13

Математическая модель нейрона Мак-Каллока и Питтса (1943)



Уоррен Мак-Каллок
(Warren McCulloch)
1898—1969
американский
нейропсихолог,
нейрофизиолог, один из
основателей кибернетики



Уолтер Питтс (Walter
Pitts)
1923—1969 (46 лет)
американский
нейролингвист, логик и
математик

BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 5, 1943

A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY

WARREN S. MCCULLOCH AND WALTER PITTS

FROM THE UNIVERSITY OF ILLINOIS, COLLEGE OF MEDICINE,
DEPARTMENT OF PSYCHIATRY AT THE ILLINOIS NEUROPSYCHIATRIC INSTITUTE,
AND THE UNIVERSITY OF CHICAGO

Because of the "all-or-none" character of nervous activity, neural events and the relations among them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes. It is shown that many particular choices among possible neurophysiological assumptions are equivalent, in the sense that for every net behaving under one assumption, there exists another net which behaves under the other and gives the same results, although perhaps not in the same time. Various applications of the calculus are discussed.

I. Introduction

Theoretical neurophysiology rests on certain cardinal assumptions. The nervous system is a net of neurons, each having a soma and an axon. Their adjunctions, or synapses, are always between the axon of one neuron and the soma of another. At any instant a neuron has some threshold, which excitation must exceed to initiate an impulse. This, except for the fact and the time of its occurrence, is determined by the neuron, not by the excitation. From the point of excitation the impulse is propagated to all parts of the neuron. The velocity along the axon varies directly with its diameter, from less than one meter per second in thin axons, which are usually short, to more than 150 meters per second in thick axons, which are usually long. The time for axonal conduction is consequently of little importance in determining the time of arrival of impulses at points unequally remote from the same source. Excitation across synapses occurs predominantly from axonal terminations to somata. It is still a moot point whether this depends upon irreciprocity of individual synapses or merely upon prevalent anatomical configurations. To suppose the latter requires no hypothesis *ad hoc* and explains known exceptions, but any assumption as to cause is compatible with the calculus to come. No case is known in which excitation through a single synapse has elicited a nervous impulse in any neuron, whereas any

115

130 LOGICAL CALCULUS FOR NERVOUS ACTIVITY

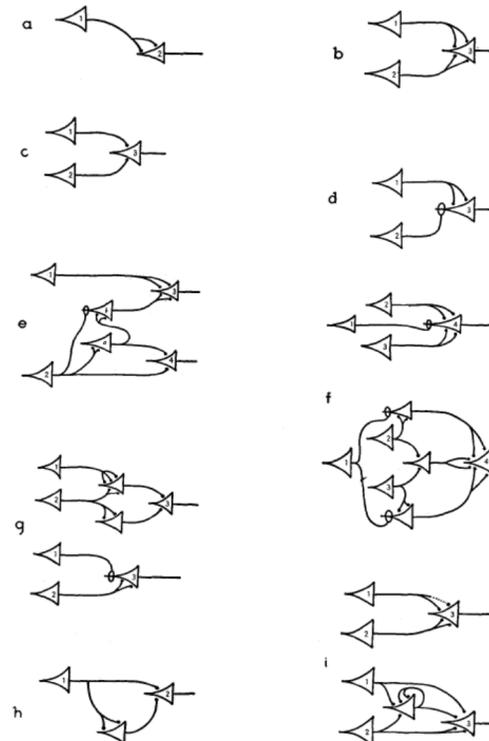


FIGURE 1

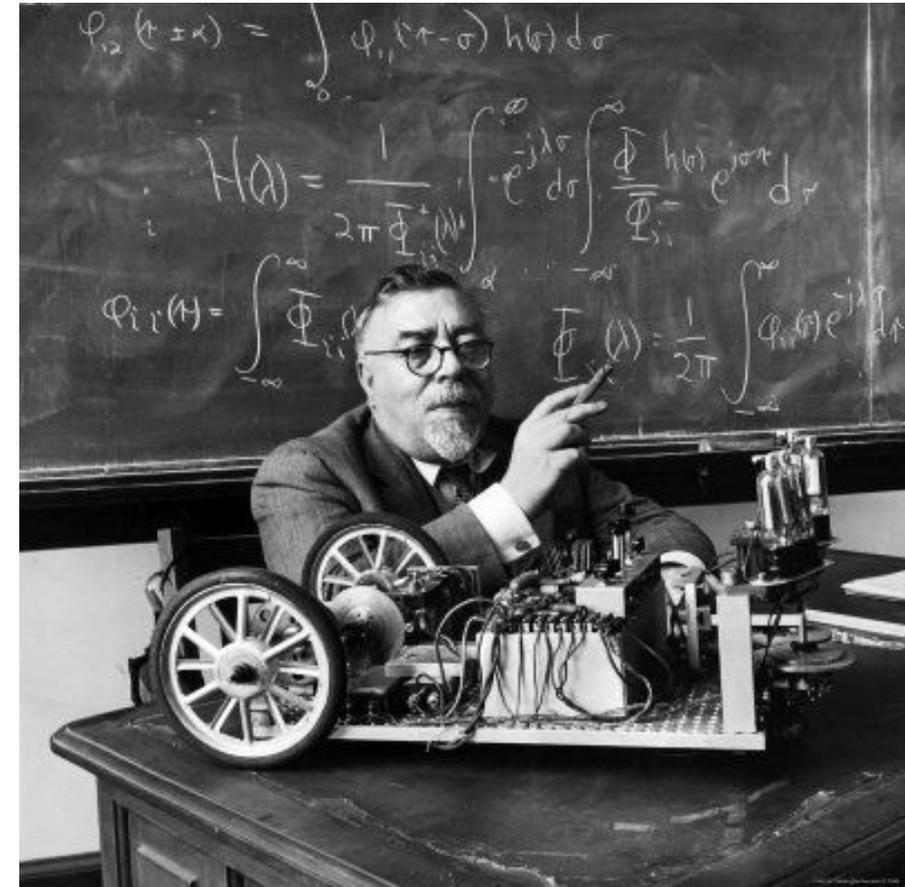
McCulloch W.S., Pitts W. A logical calculus of the ideas immanent in nervous activity // The Bulletin of Mathematical Biophysics. Kluwer Academic Publishers, 1943. Vol. 5, no. 4. P. 115–133.

[DOI:10.1007/BF02478259](https://doi.org/10.1007/BF02478259).

(Логическое исчисление мыслей, присущих нервной деятельности)

Норберт Винер (Norbert Wiener) 1894-1964

Американский математик, один из основоположников кибернетики и теории искусственного интеллекта



Закончив среднюю школу в 11 лет, Винер поступил в Колледж Тафтса и всего через три года стал бакалавром математики. Еще до наступления совершеннолетия Гарвард удостоил Винера докторской степени за его диссертацию по математической логике.

Норберт Винер с женой Маргарет, дочерьми Пегги и Барбарой и зятем Гордоном Райсбеком



Барбара

Гордон Райсбек

Маргарет

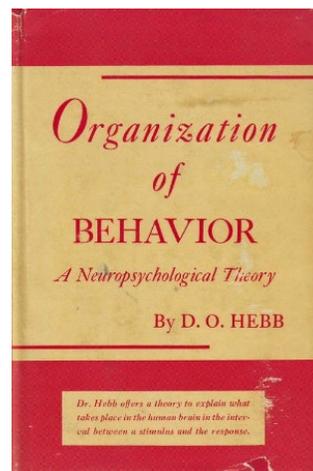
Пегги

Если связь между двумя нейронами часто используется, она от этого упражняется и становится сильнее.

Обучение по Хеббу (1949)



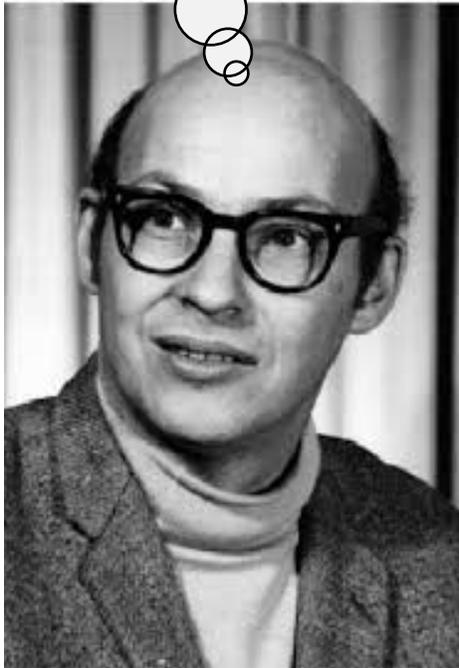
Дональд Хебб (Donald Hebb) 1904—1985
канадский физиолог и нейропсихолог



Когда аксон клетки А находится достаточно близко, чтобы возбудить клетку В, и многократно или постоянно участвует в том, чтобы ее активировать, в одной или обеих клетках происходит некий процесс роста или изменение метаболизма, в результате которого эффективность А как клетки, возбуждающей В, увеличивается.

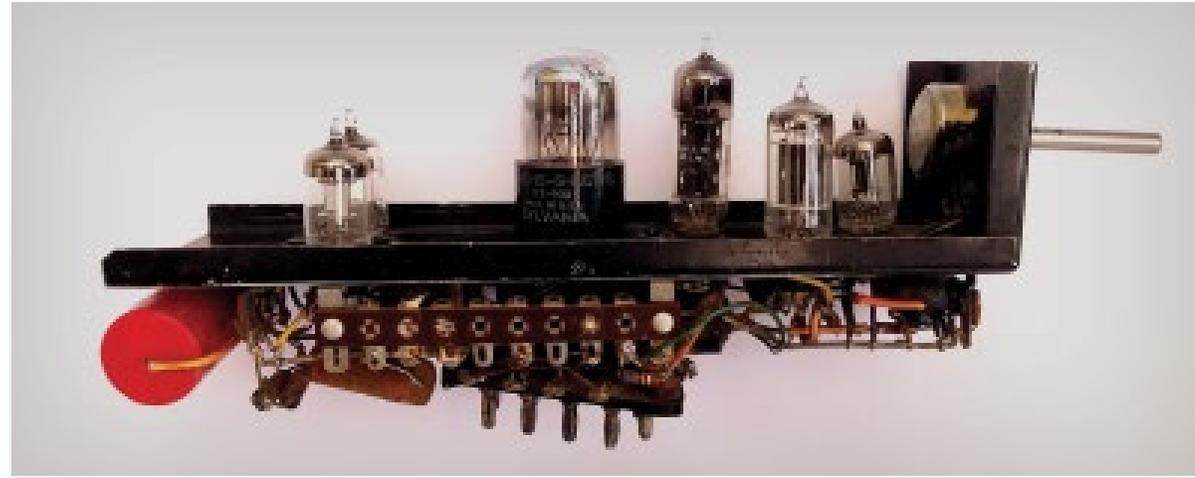
Правило Хебба
(1949)

SNARC (1951)



Марвин Минский (Marvin Minsky)
1927—2016

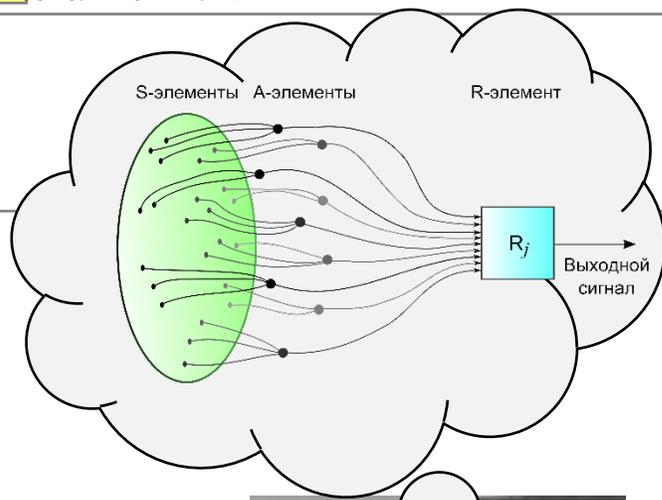
американский ученый – информатик, сооснователь
Лаборатории искусственного интеллекта в МТИ



SNARC
Stochastic Neural Analog
Reinforcement Calculator
1951

В 1951 году Марвин Минский и его аспирант Дин Эдмунд построили сеть из сорока синапсов. Синапсы были случайным образом соединены друг с другом и обучались по правилу Хебба на основе вознаграждений, которые им давали исследователи.

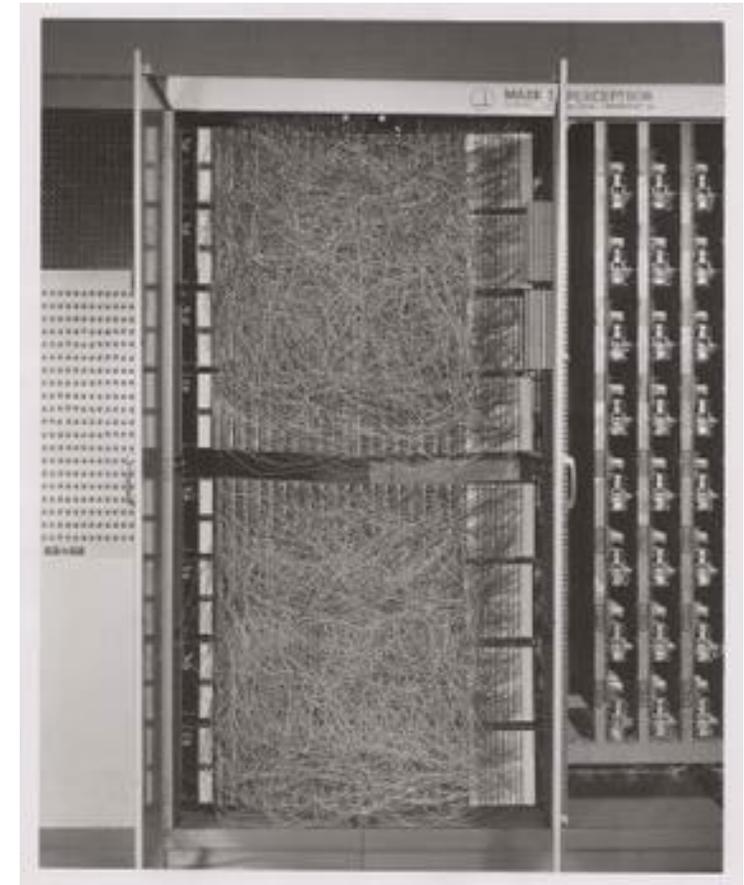
Персептрон Розенблатта (1962)



Фрэнк Розенблатт (Frank Rosenblatt)

1928—1971

американский учёный в области психологии, нейрофизиологии и искусственного интеллекта



**Mark I - первый
нейрокомпьютер
1962**

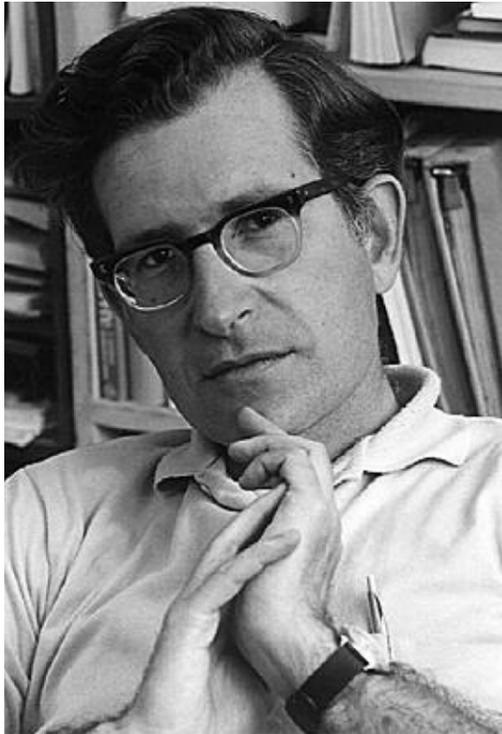
The New York Times

8 июля 1958 года



Психолог показывает эмбрион компьютера, разработанного, чтобы читать и становиться мудрее. Разработанный ВМФ... стоивший 2 миллиона долларов компьютер «IBM 704», обучился различать левое и правое после пятидесяти попыток... По утверждению ВМФ, они используют этот принцип, чтобы **построить** первую мыслящую машину класса «**Персептрон**», которая сможет читать и писать; разработку планируется завершить через год, **с общей стоимостью \$100000**... Ученые предсказывают, что позже **Персептроны** **смогут распознавать людей и называть их по имени, мгновенно переводить устную и письменную речь с одного языка на другой**. Мистер Розенблатт сказал, что в принципе возможно построить «мозги», которые смогут воспроизводить самих себя на конвейере и которые будут осознавать свое собственное существование.

Тонкости машинного перевода (1954-1964)



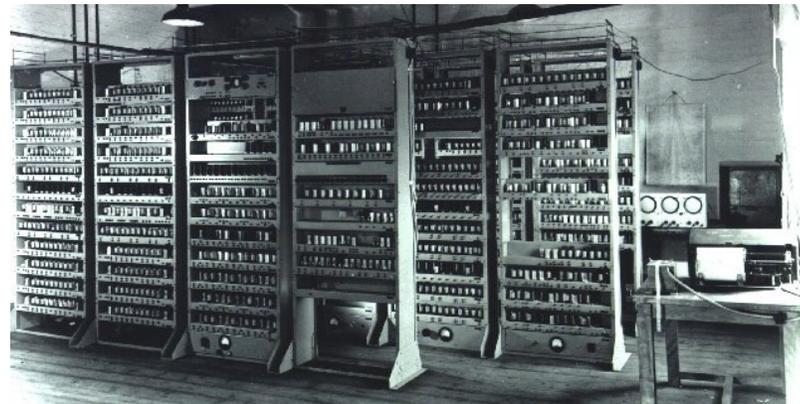
Ноам Хомский
(Noam Chomsky)

1928

американский учёный в области
математической лингвистики и
формальных грамматик

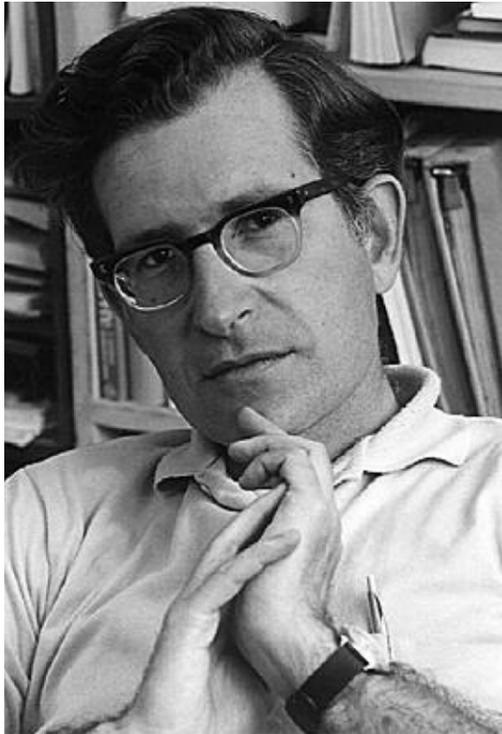
**The spirit is strong, but
the flesh is weak**
(Дух силен, но плоть
слаба)

**The vodka is good, but
the meat is rotten**
(Водка хорошая, но
мясо гнилое)



<Перевод на русский язык>

Тонкости машинного перевода (1954-1964)



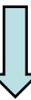
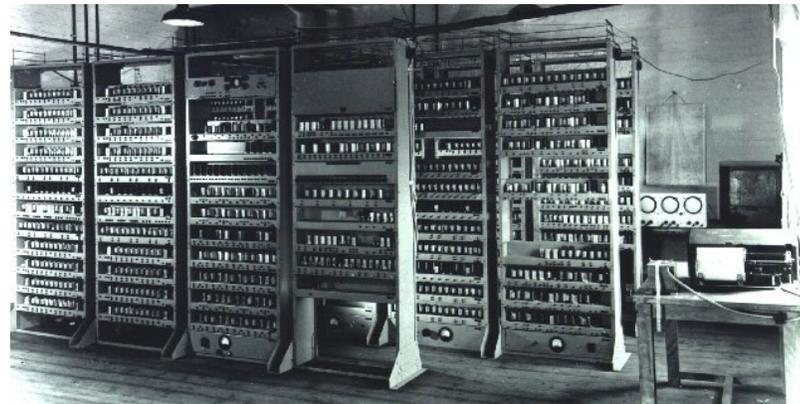
Ноам Хомский
(Noam Chomsky)

1928

американский учёный в области
математической лингвистики и
формальных грамматик

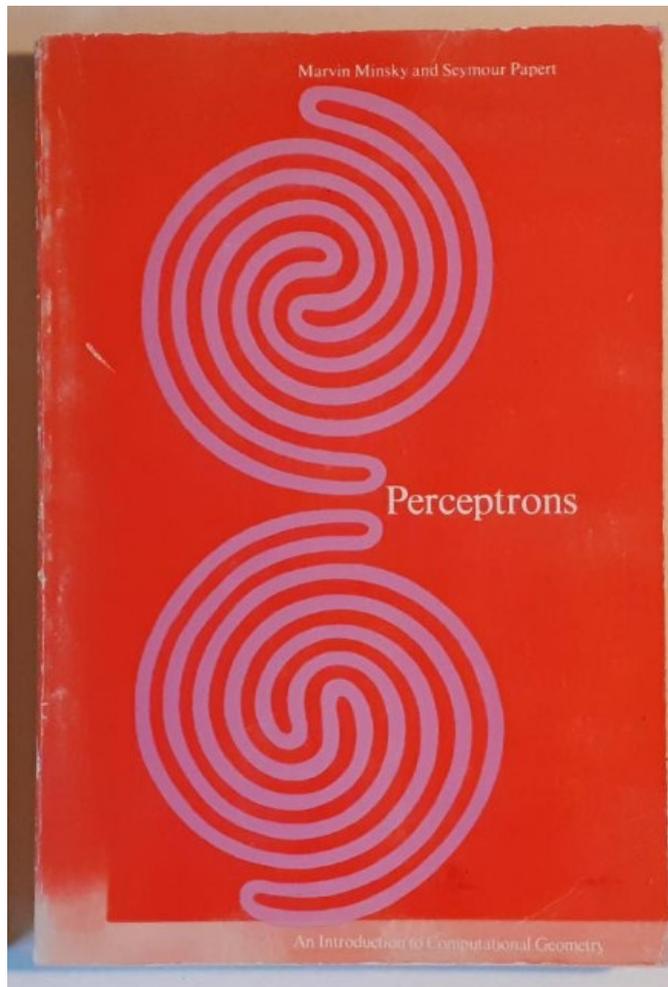
**Out of sight, out of
mind**
(С глаз долой, из
сердца вон)

Blind idiot
(Слепой идиот)



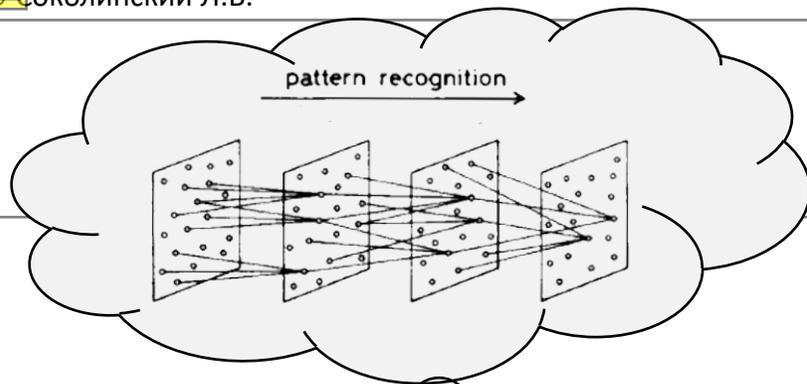
<Перевод на русский язык>

Критика персептрона (1969)



*Minsky M., Papert S. Perceptrons.
Cambridge, MA: MIT Press, 1969*

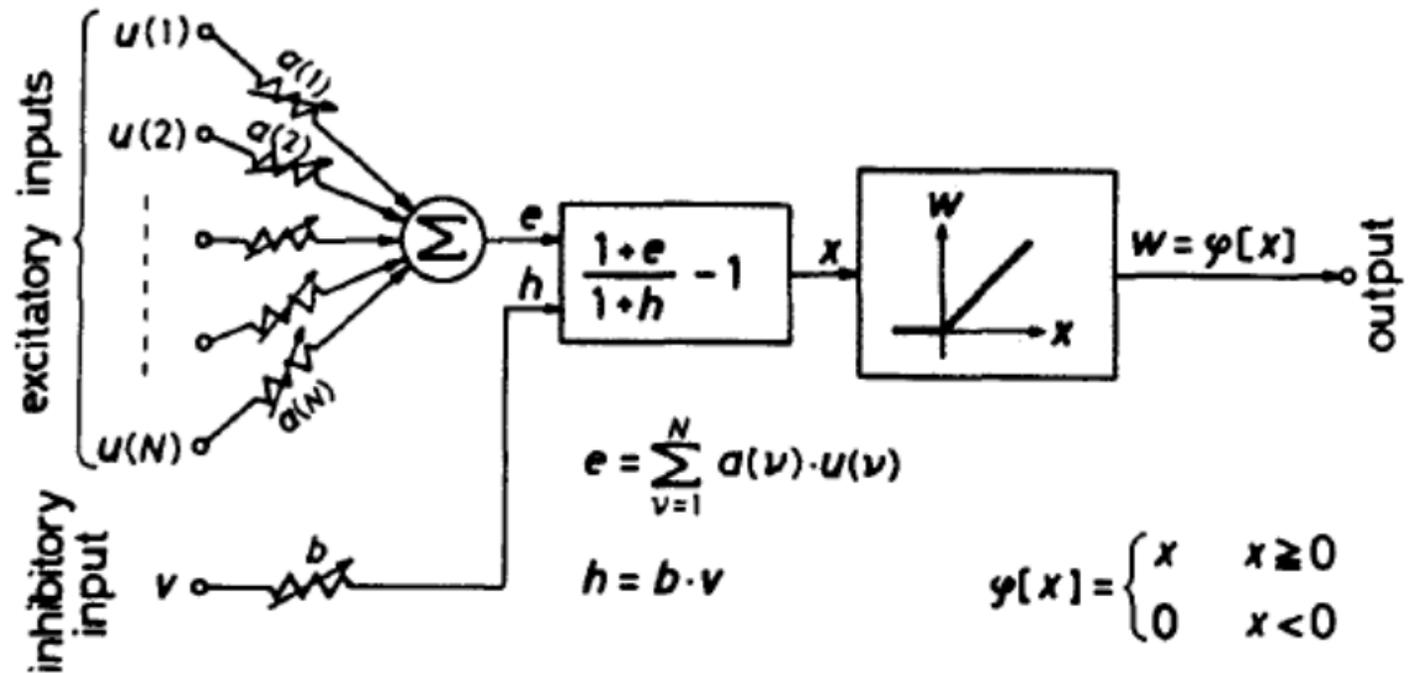
- Персептрон не может реализовать функцию XOR (сложение по модулю 2)
- Персептрон может разделять только те множества точек, между которыми можно провести гиперплоскость
- Сети с одним скрытым слоем не могут вычислять некоторые функции, если персептроны скрытого слоя не связаны со всеми входами



Неокогнитрон (1980)



Кунихико Фукусима (Kuniyiko Fukushima) 1936
японский ученый в области искусственных нейронных сетей и
глубокого обучения



Впервые предложен линейный выпрямитель ReLU

Fukushima K., Miyake S. Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Visual Pattern Recognition // Competition and Cooperation in Neural Nets. Lecture Notes in Biomathematics, vol. 45 / ed. Amari S., Arbib M.A. Berlin, Heidelberg: Springer, 1982. P. 267–285. https://doi.org/10.1007/978-3-642-46466-9_18.

Метод обратного распространения ошибки (1986)



Джеффри Хинтон
(Geoffrey Hinton)

1947

британский ученый в области информатики, «отец» глубокого обучения



Дэвид Румельхарт
(David Rumelhart)

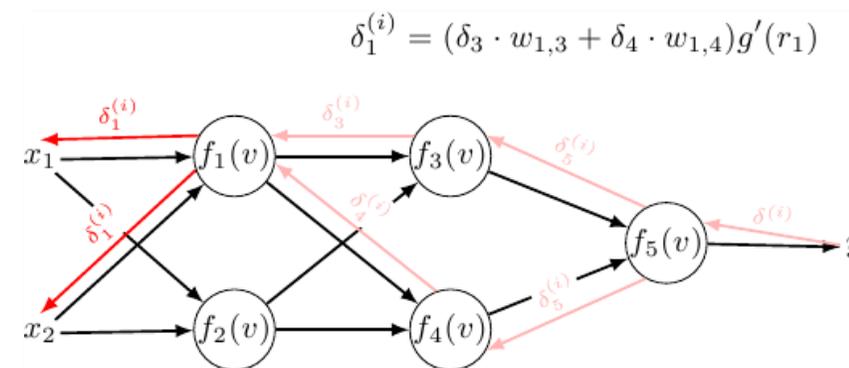
1942-2011

американский ученый в области математической психологии



Рональд Вильямс
(Ronald Williams)

американский ученый в области информатики, пионер глубокого обучения



Learning Internal Representations by Error Propagation

D. E. RUMELHART, G. E. HINTON, and R. J. WILLIAMS

THE PROBLEM

We now have a rather good understanding of simple two-layer associative networks in which a set of input patterns arriving at an input layer are mapped directly to a set of output patterns at an output layer. Such networks have no *hidden* units. They involve only *input* and *output* units. In these cases there is no *internal representation*. The coding provided by the external world must suffice. These networks have proved useful in a wide variety of applications (cf. Chapters 2, 17, and 18). Perhaps the essential character of such networks is that they map similar input patterns to similar output patterns. This is what allows these networks to make reasonable generalizations and perform reasonably on patterns that have never before been presented. The similarity of patterns in a PDP system is determined by their overlap. The overlap in such networks is determined outside the learning system itself—by whatever produces the patterns.

The constraint that similar input patterns lead to similar outputs can lead to an inability of the system to learn certain mappings from input to output. Whenever the representation provided by the outside world is such that the similarity structure of the input and output patterns are very different, a network without internal representations (i.e., a

Rumelhart D. E., Hinton G. E., Williams R. J. Learning Internal Representations by Error Propagation // Parallel Distributed Processing. Vol 1: Foundations / Cambridge, MA, USA: MIT Press, 1986.

Автокодировщики (1986)

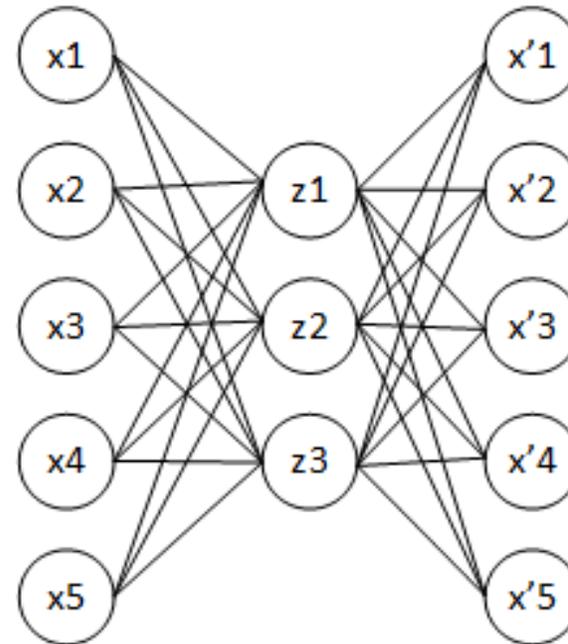
Learning Internal Representations by Error Propagation

D. E. RUMELHART, G. E. HINTON, and R. J. WILLIAMS

THE PROBLEM

We now have a rather good understanding of simple two-layer associative networks in which a set of input patterns arriving at an input layer are mapped directly to a set of output patterns at an output layer. Such networks have no *hidden* units. They involve only *input* and *output* units. In these cases there is no *internal representation*. The coding provided by the external world must suffice. These networks have proved useful in a wide variety of applications (cf. Chapters 2, 17, and 18). Perhaps the essential character of such networks is that they map similar input patterns to similar output patterns. This is what allows these networks to make reasonable generalizations and perform reasonably on patterns that have never before been presented. The similarity of patterns in a PDP system is determined by their overlap. The overlap in such networks is determined outside the learning system itself—by whatever produces the patterns.

The constraint that similar input patterns lead to similar outputs can lead to an inability of the system to learn certain mappings from input to output. Whenever the representation provided by the outside world is such that the similarity structure of the input and output patterns are very different, a network without internal representations (i.e., a



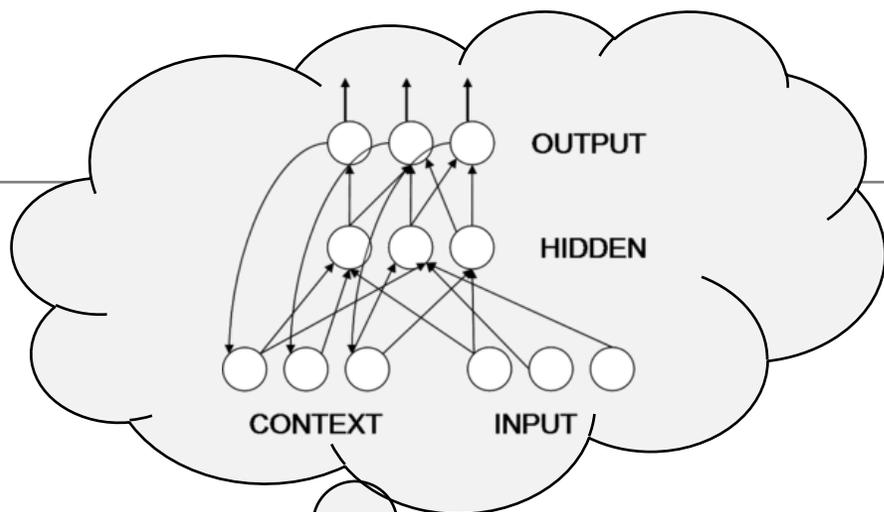
$$f(x) = x'$$

Основные применения:

- Уменьшение шума в данных
- Уменьшение размерности многомерных данных для визуализации

Rumelhart D. E., Hinton G. E., Williams R. J. Learning Internal Representations by Error Propagation // Parallel Distributed Processing. Vol 1: Foundations / Cambridge, MA, USA: MIT Press, 1986.

Рекуррентные сети (1986)



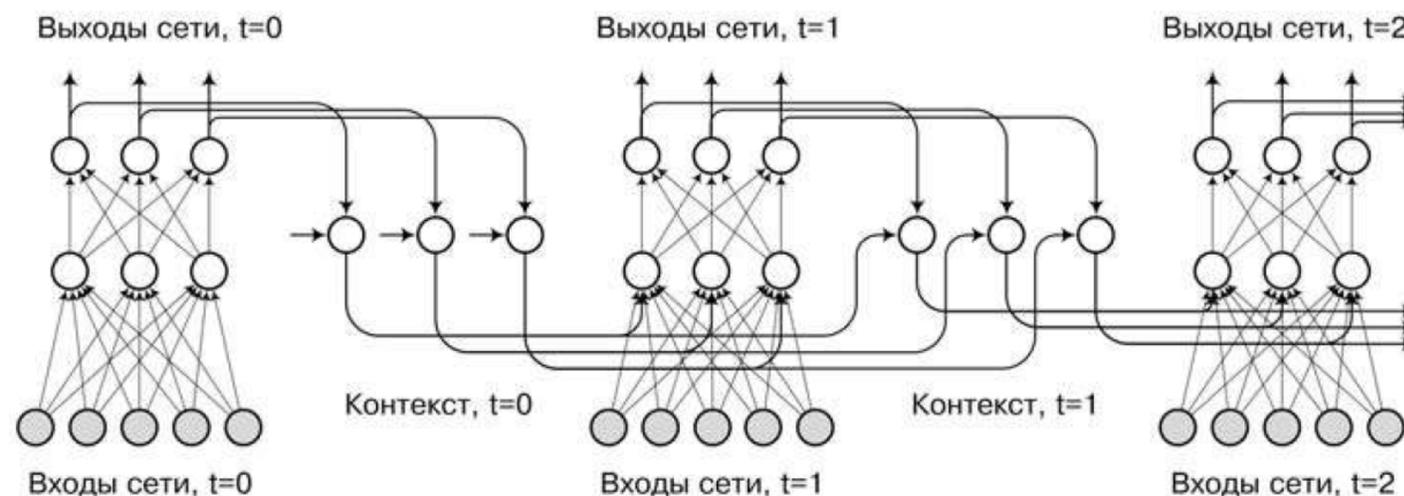
Майкл Джордан (Michael Jordan)

1956

американский ученый в области машинного обучения

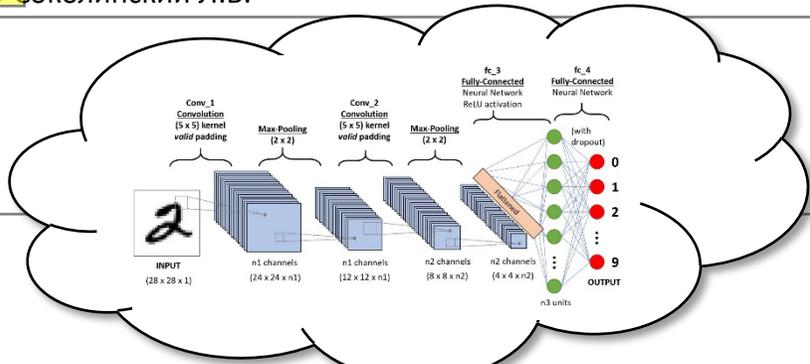
Области применения:

- Распознавание рукописного текста
- Распознавание речи
- Обработка видео
- ...



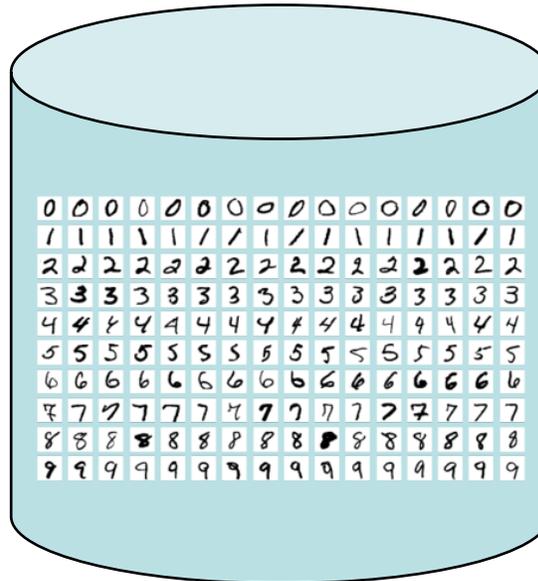
Jordan M. I. Serial Order: A Parallel Distributed Processing Approach: Tech. Rep. ICS Report 8604: Institute for Cognitive Science, University of California, San Diego, 1986.

Сверточные сети (1989)



Ян Лекун (Yann LeCun) 1960

французский ученый в области информатики, «отец» современных сверточных нейронных сетей



The MNIST database
<http://yann.lecun.com/exdb/mnist/>
 (1998)

NEURAL COMPUTATION

Communicated by Dana Ballard

Backpropagation Applied to Handwritten Zip Code Recognition

Y. LeCun
 B. Boser
 J. S. Denker
 D. Henderson
 R. E. Howard
 W. Hubbard
 L. D. Jackel

AT&T Bell Laboratories, Holmdel, NJ 07733 USA

The ability of learning networks to generalize can be greatly enhanced by providing constraints from the task domain. This paper demonstrates how such constraints can be integrated into a backpropagation network through the architecture of the network. This approach has been successfully applied to the recognition of handwritten zip code digits provided by the U.S. Postal Service. A single network learns the entire recognition operation, going from the normalized image of the character to the final classification.

1 Introduction

Previous work performed on recognizing simple digit images (LeCun 1989) showed that good generalization on complex tasks can be obtained by designing a network architecture that contains a certain amount of a priori knowledge about the task. The basic design principle is to reduce the number of free parameters in the network as much as possible without overly reducing its computational power. Application of this principle increases the probability of correct generalization because it results in a specialized network architecture that has a reduced entropy (Denker *et al.* 1987; Patarnello and Carnevali 1987; Tishby *et al.* 1989; LeCun 1989), and a reduced Vapnik-Chervonenkis dimensionality (Baum and Haussler 1989).

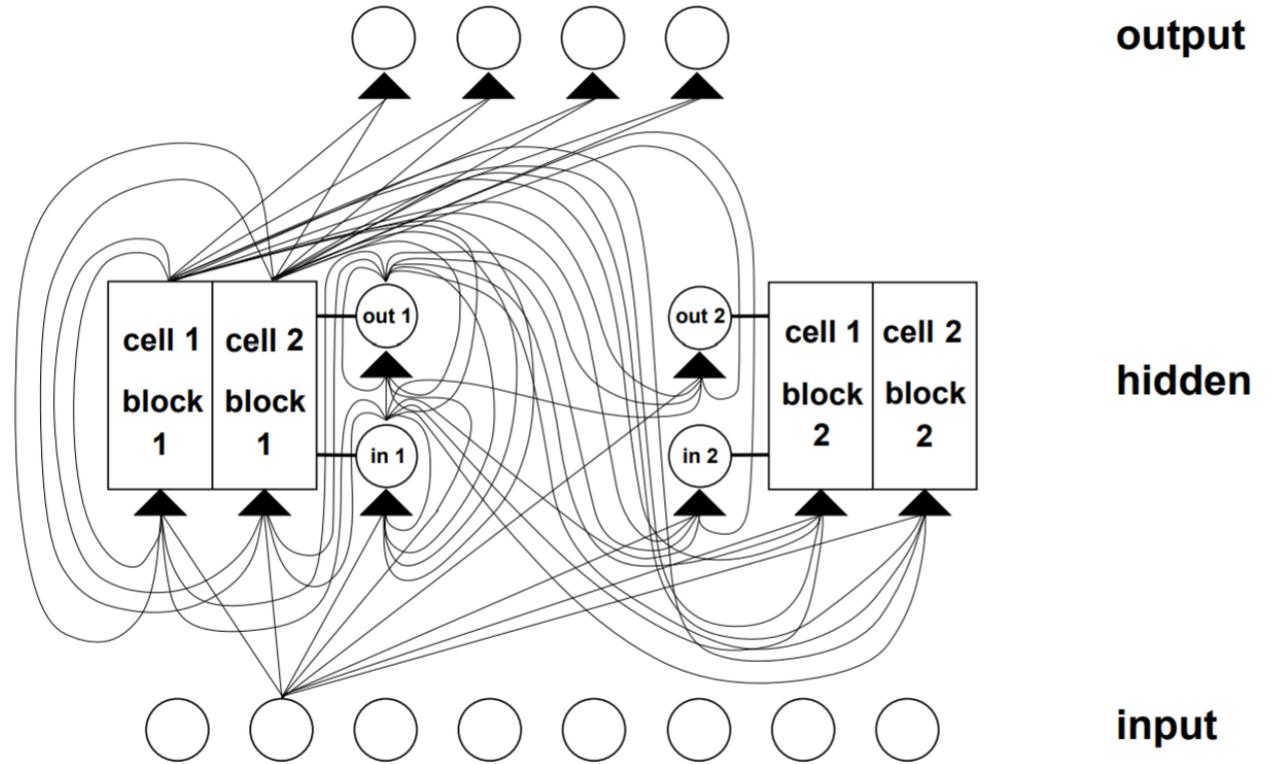
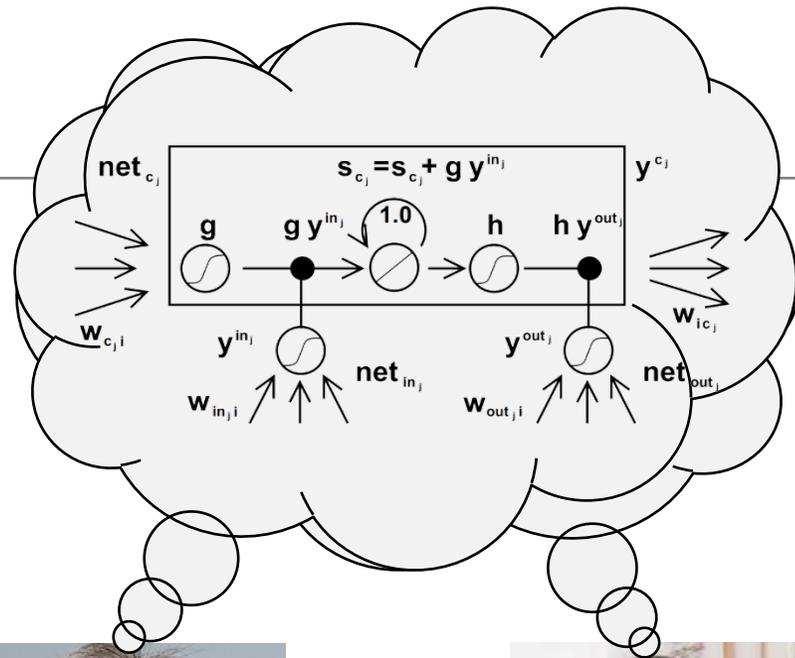
In this paper, we apply the backpropagation algorithm (Rumelhart *et al.* 1986) to a real-world problem in recognizing handwritten digits taken from the U.S. Mail. Unlike previous results reported by our group on this problem (Denker *et al.* 1989), the learning network is directly fed with images, rather than feature vectors, thus demonstrating the ability of backpropagation networks to deal with large amounts of low-level information.

Neural Computation 1, 541-551 (1989) © 1989 Massachusetts Institute of Technology

LeCun Y. et al. Backpropagation Applied to Handwritten Zip Code Recognition // Neural Computation. 1989. Vol. 1, no. 4. P. 541–551.

<https://doi.org/10.1162/neco.1989.1.4.541>.

Модель долгой краткосрочной памяти LSTM (1995)



Зепп Хохрайтер (Sepp Hochreiter) 1967

немецкий ученый в области машинного обучения



Юрген Шмидхубер (Jürgen Schmidhuber) 1963

немецкий ученый в области машинного обучения

Hochreiter S., Schmidhuber J. Long Short-Term Memory // Neural Computation. MIT Press Journals, 1997. Vol. 9, no. 8. P. 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.

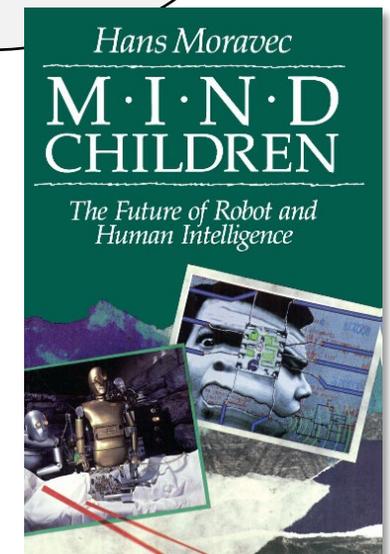
Парадокс Моравека (1988)



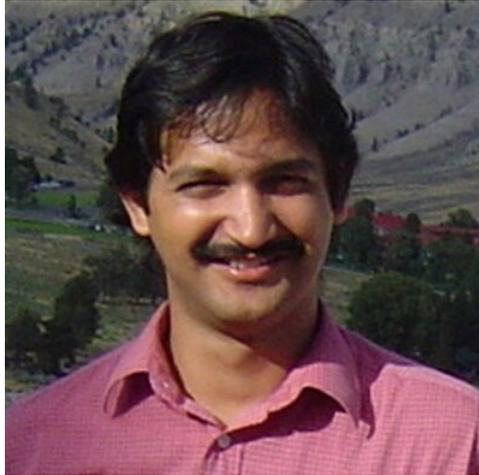
Ханс Моравек (Hans Moravec) 1948

австрийский ученый в области робототехники и
искусственного интеллекта

Почему сложные для человека задачи,
вроде решения дифференциальных
уравнений, просты для компьютера, а
простые, вроде узнавания лиц
годовалым ребенком, сложны или
невозможны?



Использование GPU (2009)



Rajat Raina



Anand Madhavan



Andrew Ng

Raina R., Madhavan A., Ng A. Large-Scale Deep Unsupervised Learning Using Graphics Processors // Proc. 26th ICML, 2009. — P. 873–880.

Large-scale Deep Unsupervised Learning using Graphics Processors

Rajat Raina
Anand Madhavan
Andrew Y. Ng

RAJAT@CS.STANFORD.EDU
ANAND@STANFORD.EDU
ANG@CS.STANFORD.EDU

Computer Science Department, Stanford University, Stanford CA 94305 USA

Abstract

The promise of unsupervised learning methods lies in their potential to use vast amounts of unlabeled data to learn complex, highly nonlinear models with millions of free parameters. We consider two well-known unsupervised learning models, deep belief networks (DBNs) and sparse coding, that have recently been applied to a flurry of machine learning applications (Hinton & Salakhutdinov, 2006; Raina et al., 2007). Unfortunately, current learning algorithms for both models are too slow for large-scale applications, forcing researchers to focus on smaller-scale models, or to use fewer training examples.

In this paper, we suggest massively parallel methods to help resolve these problems. We argue that modern graphics processors far surpass the computational capabilities of multicore CPUs, and have the potential to revolutionize the applicability of deep unsupervised learning methods. We develop general principles for massively parallelizing unsupervised learning tasks using graphics processors. We show that these principles can be applied to successfully scaling up learning algorithms for both DBNs and sparse coding. Our implementation of DBN learning is up to 70 times faster than a dual-core CPU implementation for large models. For example, we are able to reduce the time required to learn a four-layer DBN with 100 million free parameters from several weeks to around a single day. For sparse coding, we develop a simple, inherently parallel algorithm, that leads to a 5 to 15-fold speedup over previous methods.

Appearing in *Proceedings of the 26th International Conference on Machine Learning*, Montreal, Canada, 2009. Copyright 2009 by the author(s)/owner(s).

1. Introduction

We consider two well-known unsupervised learning models, deep belief networks (DBNs) and sparse coding, that can learn hierarchical representations of their input (Olahausen & Field, 1996; Hinton & Salakhutdinov, 2006). With the invention of increasingly efficient learning algorithms over the past decade, these models have been applied to a number of machine learning applications, including computer vision, text modeling and collaborative filtering, among others. These models are especially well-suited to problems with high-dimensional inputs, over which they can learn rich models with many latent variables or layers. When applied to images, these models can easily have tens of millions of free parameters, and ideally, we would want to use millions of unlabeled training examples to richly cover the input space. Unfortunately, with current algorithms, parameter learning can take weeks using a conventional implementation on a single CPU. Partly due to such daunting computational requirements, typical applications of DBNs and sparse coding considered in the literature generally contain many fewer free parameters (e.g., see Table 1), or are trained on a fraction of the available input examples.

In our view, if the goal is to deploy better machine learning applications, the difficulty of learning large models is a severe limitation. To take a specific case study, for two widely-studied statistical learning tasks in natural language processing—language modeling and spelling correction—it has been shown that simple, classical models can outperform newer, more complex models, just because the simple models can be tractably learnt using orders of magnitude more input data (Banks & Brill, 2001; Brants et al., 2007).

Analogously, in our view, scaling up existing DBN and sparse coding models to use more parameters, or more training data, might produce very significant performance benefits. For example, it has been shown that sparse coding exhibits a qualitatively different and highly selective behavior called “self-stopping” when

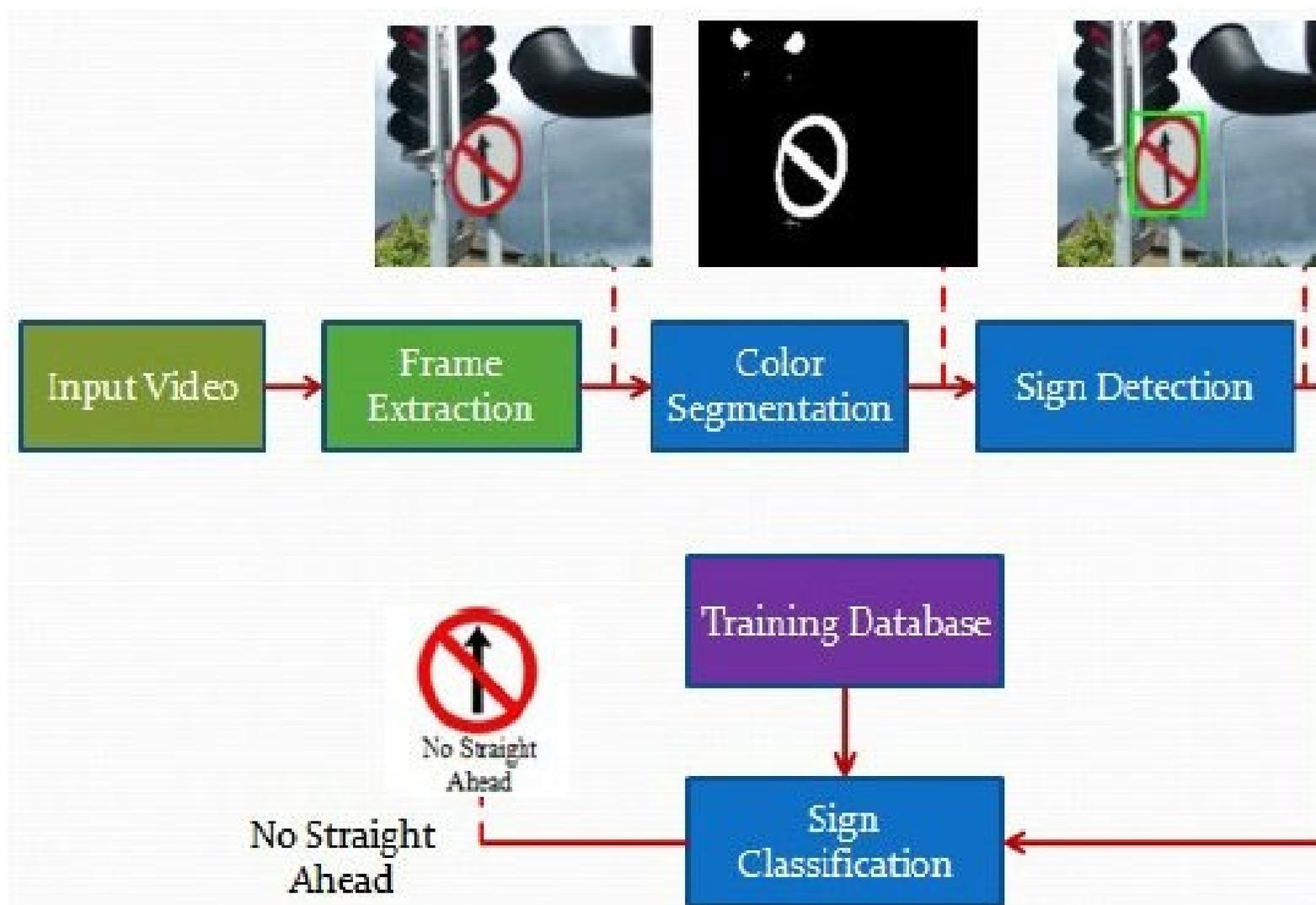
873



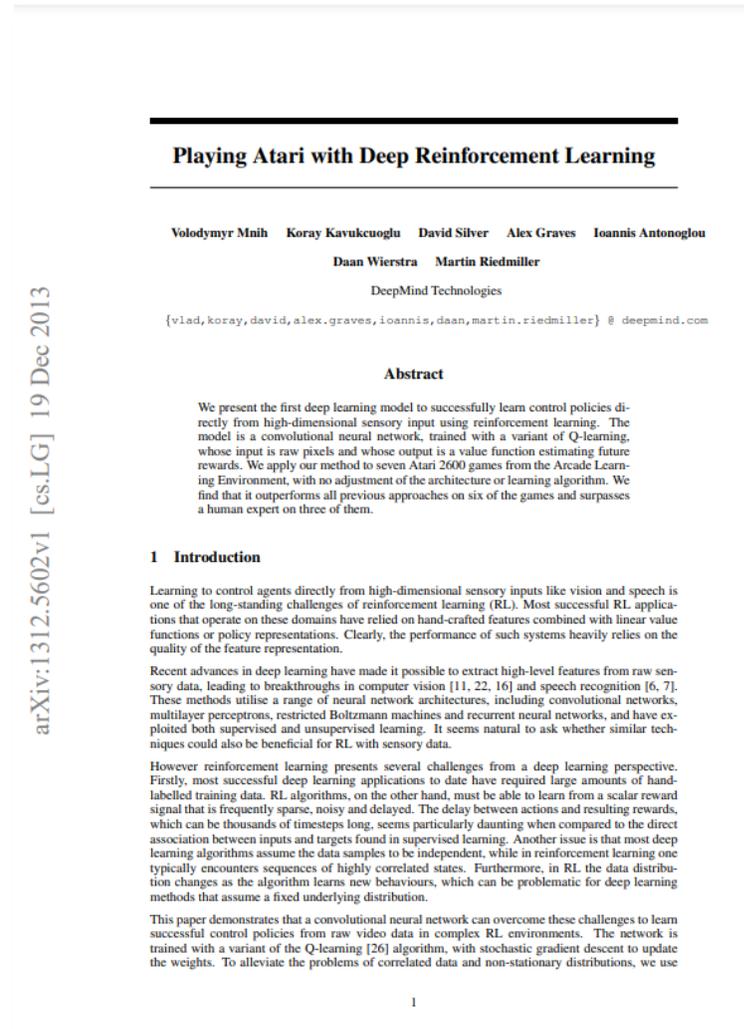
GPU NVIDIA H100 (2022)

- 989 Терафлопс (TF32)
- 456 тензорных ядер
- 14 592 ядер CUDA
- 80 ГБ HBM2 (3.35 ТБ/с)
- 3 779 210 руб.

Глубокая сверточная сеть распознает дорожные знаки на фотографиях лучше, чем человек (2011)

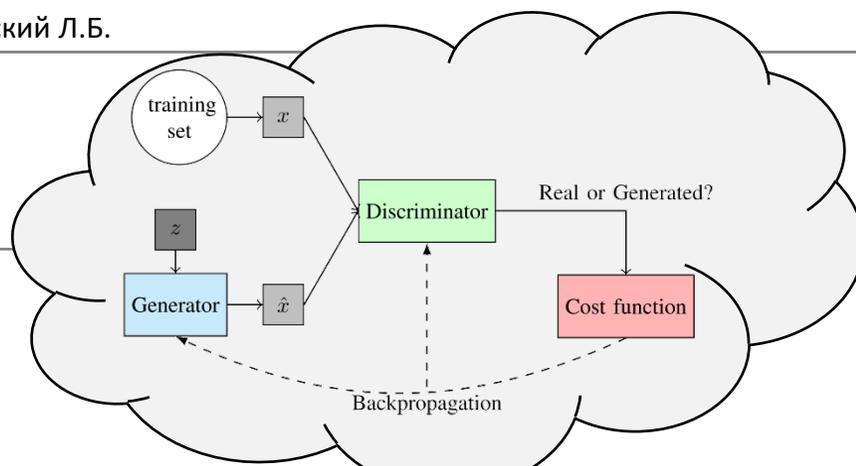


Обучение с подкреплением (2013)



<https://arxiv.org/abs/1312.5602>

Порождающие состязательные сети (2014)

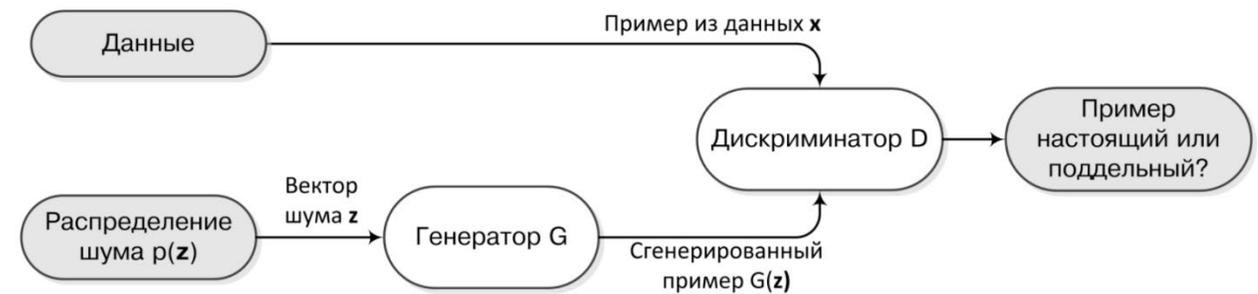


Ян Гудфеллоу (Ian Goodfellow)
1986

американский ученый в области машинного обучения, «отец» генеративно-состязательных сетей

Область применения:

- Создание искусственных изображений
- Реклама
- Кинематограф
- ...



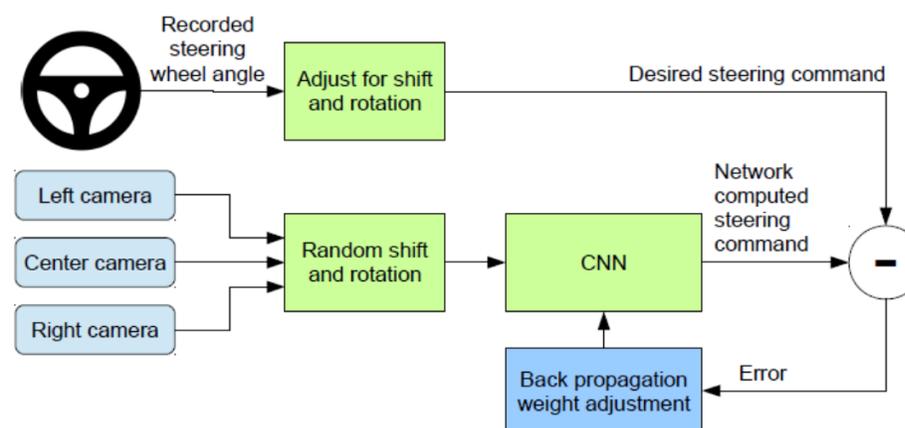
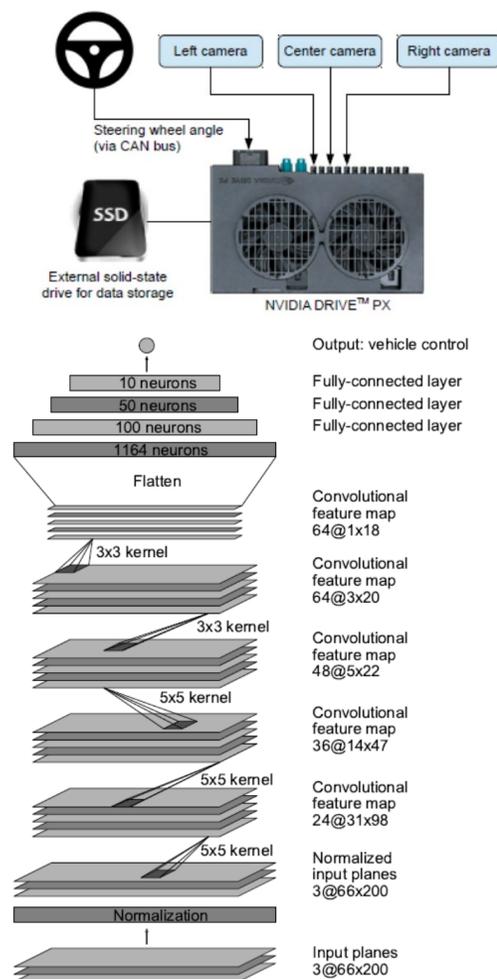
**Сгенерированные
лица людей**

Generative Adversarial Networks / I. J. Goodfellow et al. // ArXiv e-prints, 2014.

Нейронная сеть побеждает чемпиона мира по игре в го (2016)

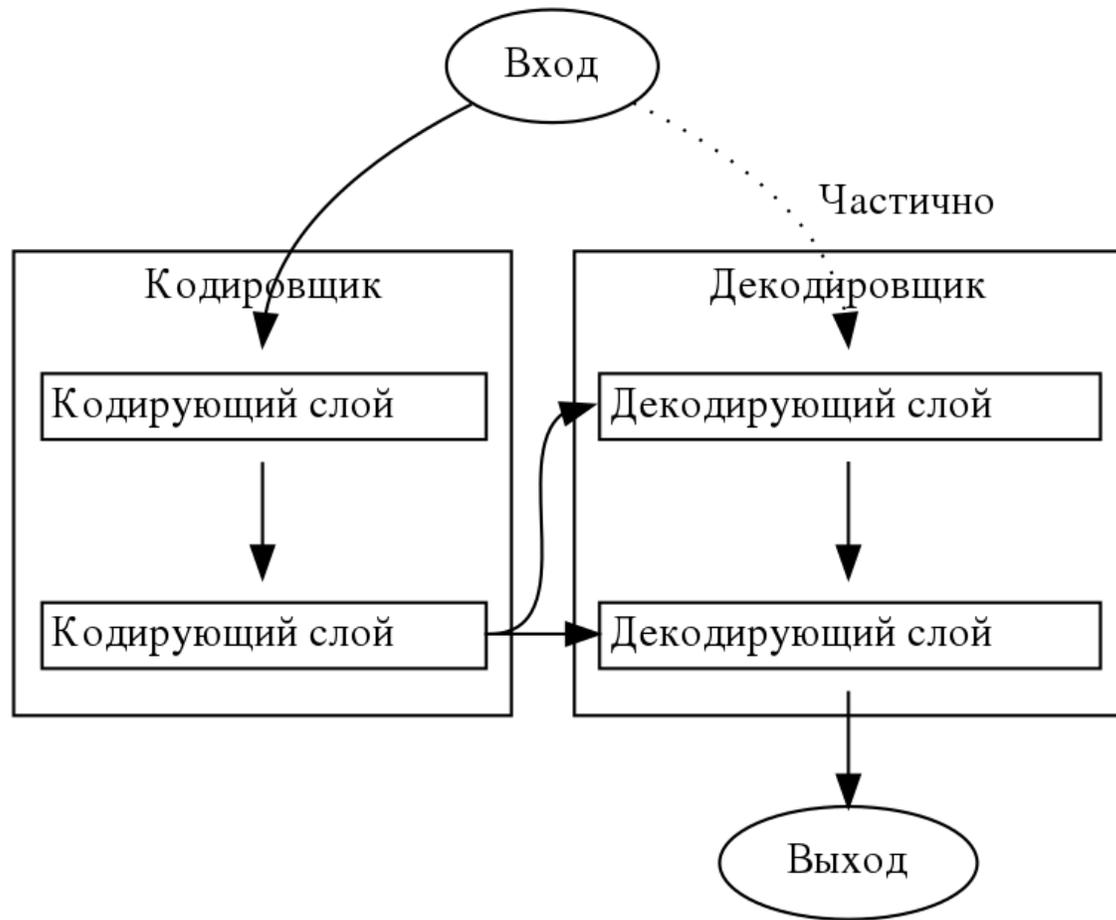


Нейронная сеть управляет автомобилем (2016)



End to End Learning for Self-Driving Cars / M. Bojarski et al. // arXiv, 2016. <http://arxiv.org/abs/1604.07316>. <https://arxiv.org/pdf/1604.07316.pdf>

Трансформеры (2017)



- Трансформер (Transformer) — архитектура глубоких нейронных сетей, разработанная Google для обработки текстов на естественном языке
- При обучении нейронная сеть определяет корреляции между различными словами в текстах (взятых, например, из «Википедии»), что позволяет ей генерировать собственные тексты
- В отличие от рекуррентных нейронных архитектур, которые обрабатывают текст последовательно, трансформеры делают это параллельно

Предсказание пространственной структуры белков (2021)



Article

Highly accurate protein structure prediction with AlphaFold

<https://doi.org/10.1038/s41586-021-03819-2>

Received: 11 May 2021

Accepted: 12 July 2021

Published online: 16 July 2021

Open access

Check for updates

John Jumper^{1,2*}, Richard Evans^{1*}, Alexander Pritzel^{1*}, Tim Green^{1*}, Michael Figurnov^{1*}, Oriol Ronneberger^{1*}, Kathryn Tunyasubunrond^{1*}, Shoaib Bates^{1*}, Augustin Zidek^{1*}, Anna Potapenko^{1*}, Alex Bridgland^{1*}, Clemens Meyer^{1*}, Simon A. Kohl^{1*}, Andrew J. Ballard^{1*}, Andrew Cowie^{1*}, Bernardino Romera-Paredes^{1*}, Stanislaw Mikolajczyk^{1*}, Rishabh Jaini^{1*}, Jonas Adler^{1*}, Trevor Back^{1*}, Sripriya Patil^{1*}, David Reineer^{1*}, Ellen Clancy^{1*}, Michal Zelinka^{1*}, Martin Steinegger^{1*}, Michalina Pacholika^{1*}, Tamas Berghammer^{1*}, Sebastian Bodenstein^{1*}, David Silver^{1*}, Cyril Vinyals^{1*}, Andrew W. Senior^{1*}, Koray Kavukcuoglu^{1*}, Pushmeet Kohli^{1*} & Demis Hassabis^{1,2*}

Proteins are essential to life, and understanding their structure can facilitate a mechanistic understanding of their function. Through an enormous experimental effort¹, the structures of around 100,000 unique proteins have been determined², but this represents a small fraction of the billions of known protein sequences³. Structural coverage is bottlenecked by the months to years of painstaking effort required to determine a single protein structure. Accurate computational approaches are needed to address this gap and to enable large-scale structural bioinformatics. Predicting the three-dimensional structure that a protein will adopt based solely on its amino acid sequence—the structure prediction component of the ‘protein folding problem’—has been an important open research problem for more than 50 years⁴. Despite recent progress^{5–11}, existing methods fall far short of atomic accuracy, especially when no homologous structure is available. Here we provide the first computational method that can regularly predict protein structures with atomic accuracy even in cases in which no similar structure is known. We validated an entirely redesigned version of our neural network-based model, AlphaFold, in the challenging 14th Critical Assessment of Protein Structure Prediction (CASP14)¹², demonstrating accuracy competitive with experimental structures in a majority of cases and greatly outperforming other methods. Underpinning the latest version of AlphaFold is a novel machine learning approach that incorporates physical and biological knowledge about protein structure, leveraging multi-sequence alignments, into the design of the deep learning algorithm.

The development of computational methods to predict three-dimensional (3D) protein structures from the protein sequence has proceeded along two complementary paths that focus on either the physical interactions or the evolutionary history. The physical interaction programs have been designed to integrate our understanding of molecular driving forces into either thermodynamic or kinetic simulation of protein folding^{13–15} or statistical approximation thereof^{16–18}. Although these are very appealing, this approach has proved highly challenging for even moderate-sized proteins due to the computational intractability of molecular simulation, the context dependence of protein stability and the difficulty of producing sufficiently accurate models of protein physics. The evolutionary programme has provided an alternative in recent years, in which the constraints on protein structure are derived from bioinformatics analysis of the evolutionary history of proteins, carried out biennially using recently solved structures that have not been deposited in the PDB or publicly disclosed so that it is a blind test

the steady growth of experimental protein structures deposited in the Protein Data Bank (PDB)¹⁹, the explosion of genomic sequencing and the rapid development of deep learning techniques to interpret these correlations. Despite these advances, contemporary physical and evolutionary history-based approaches produce predictions that are far short of experimental accuracy in the majority of cases in which a close homologue has not been solved experimentally and this has limited their utility for many biological applications.

In this study, we develop the first, to our knowledge, computational approach of predicting protein structure to near experimental accuracy in a majority of cases. The neural network AlphaFold that we developed was entered into the CASP14 assessment (May–June 2020) under the team name AlphaFold2 and a completely different model from our CASP14 AlphaFold system¹². The CASP14 assessment is carried out biennially using recently solved structures that have not been deposited in the PDB or publicly disclosed so that it is a blind test

¹DeepMind, London, UK; ²School of Biological Sciences, Seoul National University, Seoul, South Korea; ³Vertical intelligence institute, Seoul National University, Seoul, South Korea; ⁴These authors contributed equally: John Jumper, Richard Evans, Oriol Ronneberger, Kathryn Tunyasubunrond, Shoaib Bates, Augustin Zidek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislaw Mikolajczyk, Rishabh Jaini, Jonas Adler, Sripriya Patil, David Reineer, Ellen Clancy, Michal Zelinka, Martin Steinegger, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. *e-mail: john.jumper@deepmind.com; richard.evans@deepmind.com

Nature | Vol 596 | 16 August 2021 | 583

DOI:10.1038/s41586-021-03819-2

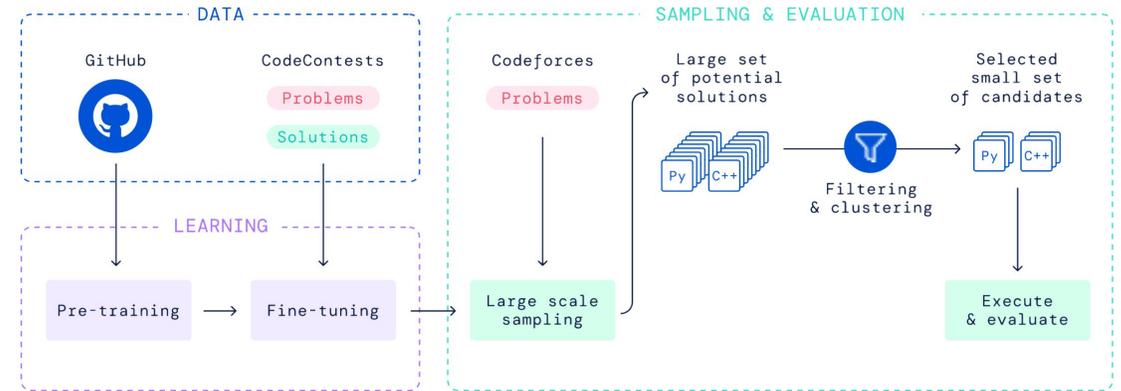
Not used to make prediction

Used to make prediction



AlphaFold

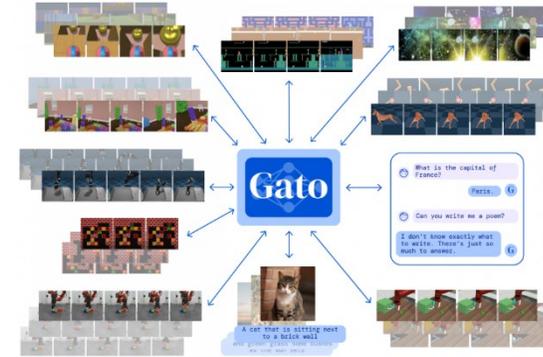
AlphaCode: нейросетевая модель, генерирующая код для решения задач спортивного программирования (2022)



- В соревнованиях по программированию на платформе Codeforces AlphaCode показала средний результат среди 5000 участников
- Используются большие и эффективные архитектуры на основе трансформеров
- Впервые система искусственного интеллекта успешно выступила на соревнованиях по программированию

- При предварительном обучении исходные тексты программ из GitHub случайным образом разделяются на две части: первая часть поступает в кодер в качестве входных данных, а декодер обучается генерировать вторую часть
- При точной настройке кодеру передаются описания задач, а декодер обучается генерировать решения
- В ходе соревнований AlphaCode генерирует множество вариантов решений для каждой задачи и выполняет их, чтобы отфильтровать плохие образцы и сгруппировать оставшиеся
- На проверку отправляется не более 10 различных решений для одной задачи

Gato: нейросетевая модель, способная выполнять 604 задачи (2022)



- Используется подход на основе трансформеров
- Одна и та же сеть с одинаковыми весами
 - Успешно играет на приставке Atari
 - Описывает изображения
 - Общается в чате
 - Строит конструкции из блогов с помощью роботизированной руки
 - ...
- Сеть имеет 1.2 миллиардов параметров (весов и смещений)

Яндекс Переводчик (2017)

Яндекс Переводчик Текст Сайты Документы Картинки Для бизнеса

АНГЛИЙСКИЙ ↔ РУССКИЙ

Yandex Translate is a web service provided by Yandex, intended for the translation of text or web pages into another language.

The service uses a self-learning statistical machine translation, developed by Yandex. The system constructs the dictionary of single-word translations based on the analysis of millions of translated texts. In order to translate the text, the computer first compares it to a database of words. The computer then compares the text to the base language models, trying to determine the meaning of an expression in the context of the text.

In September 2017, Yandex.Translate switched to a hybrid approach incorporating both statistical machine translation and neural machine translation models.]

Яндекс Переводчик - это веб-сервис, предоставляемый Яндексом, предназначенный для перевода текста или веб-страниц на другой язык.

Сервис использует самообучающийся статистический машинный перевод, разработанный Яндексом. Система создает словарь однословных переводов на основе анализа миллионов переведенных текстов. Чтобы перевести текст, компьютер сначала сравнивает его с базой данных слов. Затем компьютер сравнивает текст с базовыми языковыми моделями, пытаясь определить значение выражения в контексте текста.

В сентябре 2017 года Яндекс.Переводчик перешел на гибридный подход, включающий как статистический машинный перевод, так и модели нейронного машинного перевода.

- Яндекс Переводчик использует гибридную систему (сочетает в себе статистический и нейронный машинный перевод)
- Доступен перевод для 98 языков
- Направление перевода определяется автоматически
- Перевод с микрофона
- В мобильной версии реализован функция перевода при наведении камеры на текст

Kandinsky – нейросеть Сбербанка для генерации изображений и видео по текстовому описанию (2022)



Фото: Нейросеть «Кандинский»

- Поддерживает более 100 языков
- Генерация изображения по текстовому описанию
- Микширование двух изображений
- Микширование изображения и текстового описания дополнения
- Стилистическая переработка изображения

<https://www.sberbank.com/promo/kandinsky/>

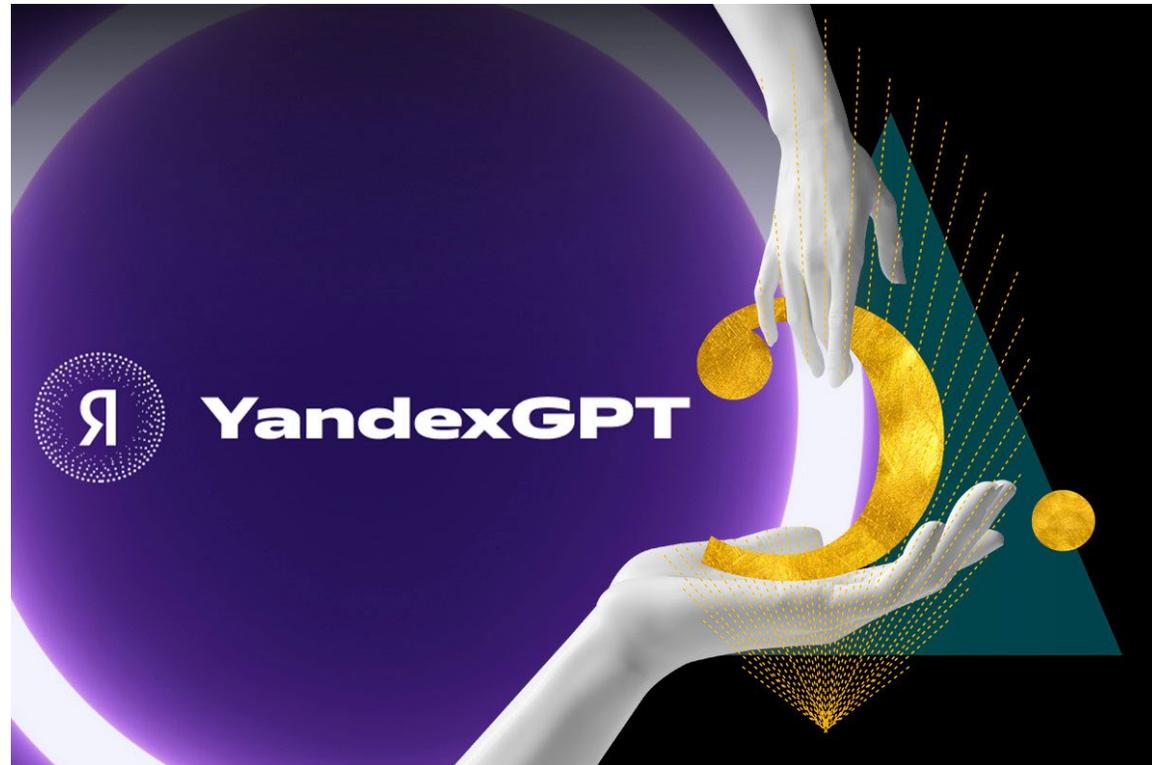
ChatGPT (2022)



<https://chat.openai.com/>

- ChatGPT (Generative Pre-trained Transformer) — чат-бот с генеративным искусственным интеллектом, разработанный компанией OpenAI
- Отвечает на вопросы в различных областях, включая историю, науку, технологии и многое другое
- Генерация текстов на разных языках, относящиеся к различным предметным областям
- Генерация программ на различных языках программирования

YandexGPT (2023)



https://ya.ru/alisa_davay_pridumaem

- YandexGPT— нейросеть семейства GPT от компании «Яндекс», которая генерирует тексты на основе данных из интернета
- Используется в виртуальном голосовом помощнике «Алиса»
- Может написать письмо или статью, объяснить непонятное слово или тему из учебника, придумать идею, дать совет и помочь с другими задачами
- YandexGPT 2 доступна на ya.ru в навике «Алиса, давай придумаем»
- Работает в приложении Яндекс, Яндекс Браузере, на Яндекс Станции и телевизорах с Алисой

Сбер представляет новую памятную серебряную монету к наступающему 2024 Новому году



- Нейросеть Сбера Kandinsky сгенерировала изображение Зеленого дракона, которое усилиями дизайнеров Московского монетного двора реализовались в финальном дизайне, воплотив символ 2024 года в серебре высшей пробы
- С Новым 2024 Годом!!!

Конец лекции

13

