WARSAW UNIVERSITY OF TECHNOLOGY
FACULTY OF MATHEMATICS
AND INFORMATION SCIENCE

Neural Networks

Lecture 1

Course objectives

Relay students a knowledge of artificial neural networks using information from biological structures. After completing the course (lecture and project) students should:

- have theoretical knowledge about principles of construction and operation of basic models,
- be able to select proper structure to execute the assumed functions,
- be able to select proper programming tools (languages, packages etc.) to carry out tasks,
- being the part of a team be able to carry out the tasks for team members,
- prepare and test computer program,
- prepare the final report.

Learning outcomes

knowledge

- a student knows theoretical background of operation and modelling of neuronlike elements and the rules of construction of neuronal multi layer structures

skills

- is able to analyse given net, prepare its functional description, carry out the proof of its correct work
- is able to analyse given net, prepare its functional description, carry out the proof of its correct work

Learning outcomes

skills (cont)

- can evaluate the usefulness of programming tools to model the network based on given parameters
- can obtain information from literature, databases and other selected sources appropriate for problems solved

soft competences

- can cooperate individually and in a work team, accepting various role in it

Learning outcomes realisation and verification

Assumed learning outcomes – student	course form	verification citeria	verification methods
knows theoretical background of operation and modelling of neuronlike elements and the rules of construction of neuronal multi layer structures	lecture (examples) exercises before-exam	discussion of various structures and modela	exam – written and/or oral part
is able to analyse given net, prepare its functional description, carry out the proof of its correct work	lecture (examples) project (exercises) project	completion of proper analysis and description	exam written part, project
can design a complex device related to solve a practical problem (i.e from the area of finanses or data classification)	lecture (examples) project (exercises)	design of a project of device, analysis of correctness	exam written part, project
can evaluate the usefulness of programming tools to model the network based on given parameters	exercises before-exam project exercises + consultations	selecdftion of a proper programming language with justification	project's course and pass
can obtain information from literature, databases and other selected sources appropriate for problems solved	project	bibliography selectios, justification	project's course and pass
can cooperate individually and in a work team, accepting various role in it	project	split of work within a team members, completion of entrusted tasks	teachers' observation

ECTS credits

- contact hours 75h:
  - lectures – 30h,
  - laboratory work – 30h
- preparation for laboratory work – 20h
- familiarize with basic literature – 15h
- computer program preparation, debugging, verification (out of lab) – 30h
- final report preparation – 10h
- preparation for the exam and written exam – 20h

Total students' workload 155h = 5 ECTS credits

Course Contents

Introduction

What cybernetics and biocybernetics are

Modeling

Neurocomputers and Neurocomputing

Comparison of humans and computers

Methods of learning

The nervous system

Course Contents

The brief overview of the brain

Biological neuron

Signal processing in the biological nervous system

The Artificial Neuron

McCulloch & Pitts Model

Single-layer Artificial Neural Network

Course Contents

Multi-layer Artificial Neural Network

Mathematical Model of a Single Neuron and a Network

The Rosenblatt's Perceptron

Method of Learning

Perceptron Representation

Perceptron limitations (XOR Problem)

Linear Separability

Course Contents

The Rosenblatt's Perceptron cont.

Overcoming the limitations

Existence Theorem

The Delta Rule

ADALINE model

The Backpropagation Algorithm

Course Contents

Associative Memories

3 - Layer Model

Kohonen Self-Organizing Model

Learning Method

Winner Takes All Rule

Neighborhood definition

Course Contents

Adaptive Resonance Theorem

- *ART Architecture*
- *Learning Method*

Hamming Model

Network for Logic Operations

Neural Networks for Compression

Optimization Problems

Neural Networks for Matrix Algebra Problems

Bibliography

- **T. Kohonen** *Associative Memory*, Springer, 1978
- **P. D. Wasserman** *Neural Computing, theory and practice,* Van Nostrand Reinhold 1989
- **R. Beale, T. Jackson** *Neural Computing, An Introduction*, A.Hilger IOP Publ. Co. Bristol 1990.
- **A. Cichocki, R. Unbehauen,** *Neural Networks for Optimization and Signal Processing*, J.Wiley 1993.

Bibliography

- **J. J. Hopfield** *Neural Networks and physical systems with emergent collective computational abilities*, Proc. Natl. Acad,. Sci. USA, 79, 1982
- **J. J. Hopfield** *Neurons with Graded Response have collective computational properties like those of two-state neurons*, Proc. Natl. Acad,. Sci. USA, 81, 1982

Bibliography

- **J. J. Hopfield, D. W. Tank** *„Neural" Computation and Decisions in Optimization Problems*, Biol. Cyber. 52, 141-152, 1985.
- **R. P. Lippman** *An introduction to Computing with Neural Networks,* IEEE ASSP Mag. April 1987
- **J. Kinoshita, N. G. Palevsky** *Computing with Neural Networks,* High Technology, May 1987
- **R. Hecht-Nielsen** *Neurocomputing, Picking the Human Brain*, IEEE Spectrum, March 1988

Bibliography

- **D. L. Alkon** *Memory Storage and Neural Systems*, Sci.Amer. July 1989
- **D. R. Hush, B. H. Horne** *Progress in Supervised Neural Networks,* IEEE Sign Proc.Mag. Jan. 1993
- **L.Rutkowski** *New Soft Computing Techniques for System Modelling, Pattern Classification and Image Processing,* Springer-Verlag, 2004

Bibliography

- **L.Rutkowski** *Flexible Neuro-Fuzzy Systems,* Kluwer Acad, Publ., 2004
- **L.Rutkowski** *Computational Intelligence,* Springer Verlag, 2008
- **Conf. Materials:** *Neural Networks and Soft Computing* 2000-2015

Bibliography - Polish

- **S. Osowski,** *Sieci neuronowe*, Ofic. Wyd. Pol. Warszawskiej, Warszawa 1994.
- **J. Korbicz, A. Obuchowicz, D. Uciński,** *Sztuczne sieci neuronowe, podstawy i zastosowania*, Akademicka Oficyna Wydawnicza PLJ, Warszawa 1994.
- **T. Kacprzak, K. Ślot,** *Sieci neuronowe komórkowe*, PWN 1995

Bibliography - Polish

- **T. Masters,** *Sieci neuronowe w praktyce*, WNT 1996
- **J. Zurada, M. Barski, W. Jędruch** *Sztuczne sieci neuronowe*, PWN 1996
- **S. Osowski** *Sieci neuronowe w ujęciu algorytmicznym*, WNT 1996.
- **L. Rutkowski** (ed) *Sieci neuronowe i neurokomputery* Wyd. Pol.Czest. 1996

Bibliography - Polish

- **D. Rutkowska, M. Piliński, L. Rutkowski** *Sieci neuronowe, algorytmy genetyczne i systemy rozmyte,* PWN 1997
- **R. Tadeusiewicz** *Elementarne wprowadzenie do technik sieci neuronowych z przykł. progr.,* Akad. Ofic.Wyd. PLJ 1998
- **S. Osowski** *Sieci neuronowe do przetwarzania informacji*, Ofic. Wyd. PW, 2000

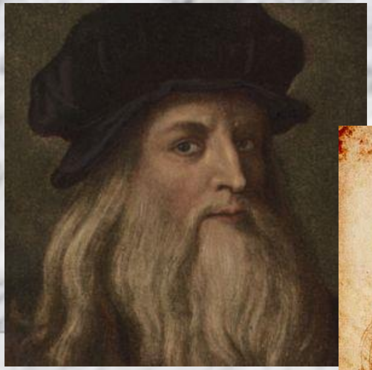
Bibliography - Polish

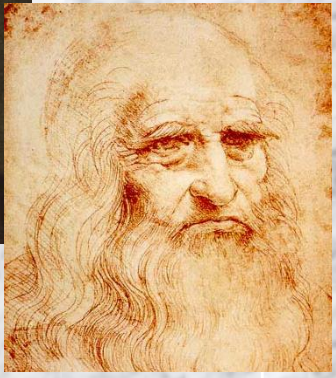
- **J. Mańdziuk** *Sieci neuronowe typu Hopfielda*, Akad. Ofic. Wyd. EXIT 2000
- **L. Rutkowski** *Biocybernetyka i inżynieria biomedyczna, t.6 Sieci Neuronowe, EXIT, 2000*
- **B. Borowik** *Pamięci asocjacyjne*, Mikom 2002
- **R. A. Kosiński** *Sztuczne sieci neuronowe*, WNT 2002
- **L. Rutkowski** *Metody i techniki sztucznej inteligencji*, PWN 2005

Bibliography - Journals

Neural Networks
IEEE Transaction on Neural Networks
Proceedings of the IEEE
IEEE Transaction on System, Man and Cybernetics
Artificial Intelligence
Computer IEEE
Neurocomputing
Network, Computation in Neural Systems

Introduction





History

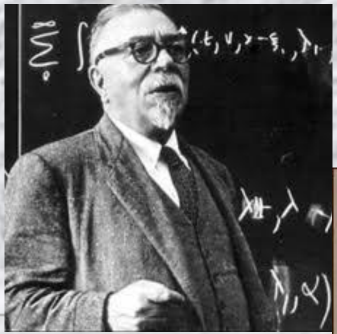
Born on April 15, 1452, in Vinci, Italy, Leonardo da Vinci was the epitome of a “Renaissance man.” Man of a curious mind and keen intellect, da Vinci studied the laws of science and nature, which greatly informed his work as a painter, sculptor, architect, inventor, military engineer and draftsman.

Specialization means to focus on a specific aspect of a larger topic.

is necessary, but ...

Synthesis is the act of combining elements to form something new.

Introduction



Cybernetics

Norbert Wiener, with Artur Rosenbluth, 1940th, analogy between humans and technical systems

Book:

***Cybernetics or Control and Communication in the Animal and the Machine* – 1948**

(*Cybernetyka – czyli sterowanie i komunikacja w zwierzęciu i maszynie* – 1971**)**

word from greek **–** **κύβερνετεσ** **-** helmsman

Introduction

Cybernetics

data transmission, on the base of mathematical logic, electronics, theory of probability, computer sciences

and

on the analogy between machines and living organisms

Introduction

Modeling

mathematical

physical

simulation

Model

formal description of a system or process allowing precise and logical analysis; background for technical realization, can be a prototype

Introduction

Modeling can be controversial because

object description is impossible

description is extremely complicated

description is general.

Some simplifications and limitations have to be used, next verified by the results

Introduction

We will model the nervous system, or precisely – the elements of the nervous system.
We do not intend to build the copy of any real nervous system.

Introduction

We are not attempting to build computer brains, not to mimic parts of real brains – we are aiming rather to discover the properties of models that take their behavior from extremely simplified versions of neural systems, usually on massively reduced scale.

Introduction

Stages of modeling

1. collection, analysis and evaluation of existing biological data, defining the useful properties
2. defining the possibilities for exact mathematical description

Introduction

Stages of modeling (cont.)

3. model of a process or structure
4. comparison of the results biological experiments
5. computer model
6. technical device

Introduction

Why neural modeling ???

1. Realization of important functions
2. The vast amount of information received from the environment and appropriate selection of this information,
3. Adaptability to varying conditions
4. The great reliability of a system - comprised of a huge number of elements –
   minor or major damage, do not lead to an interruption in the work of the system

Introduction

System reliability:

assuming

10^{10} elements

probability of correct functioning = 0,9999999999

theoretical probability of correctness of the system

< 0,367

but, it works !!!

Introduction

Nervous system

- system of data transmission, multilayer, hierarchical, and optimal
- mostly parallel processing
- perfect selection of important information

History

XVIII - XIX century

tissue excitation together with electrical processes

XX century

nervous system is composed from many cells

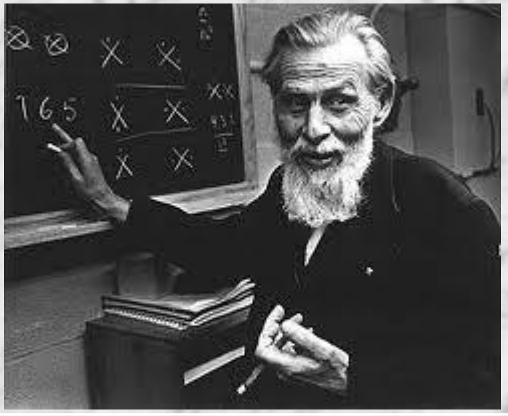
electrochemical processes inside cells

History

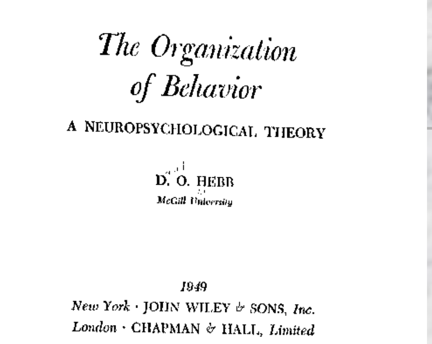
1943 McCulloch & Pitts model

The logical calculus of the ideas immanent in nervous activity

Formal neuron, on – off switch and can be combined to compute logical functions



History



The Organization of Behavior

A NEUROPSYCHOLOGICAL THEORY

D. O. HEBB

McGill University

1949

New York · JOHN WILEY & SONS, *Inc.*

London · CHAPMAN & HALL, *Limited*

1949 r. Hebb’s theory

The organization of Behavior

Concept of cell assemblies, behavior is coded by collections of neurons,

Hebb’s (or Hebbian) learning rule : ***„When an axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A’s efficiency, as one of the cells firing B, is increased.”***

The use of existing or active pathway strengthens the connections between the neurons

History

1962 Frank Rosenblatt's (an American sychologist) book

The Principles of Neurodynamics

model of the perceptron

1969 Marvin Minsky & Seymour Papert book

Perceptrons: An introduction to Computational Geometry

Perceptron are impractical and/or inadequate to solve problems - **death of the perceptron**

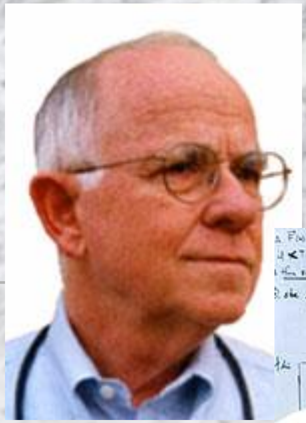
History

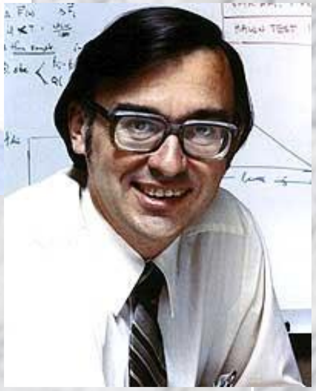
- Quote from Minsky and Papert's book, *Perceptrons*

"[The perceptron] has many features to attract attention: its linearity; its intriguing learning theorem; its clear paradigmatic simplicity as a kind of parallel computation.

There is no reason to suppose that any of these virtues carry over to the many-layered version. Nevertheless, we consider it to be an important research problem to elucidate (or reject) our intuitive judgment that the extension is sterile."

History





1960 Widrow & Hoff

Adaptive switching circuits

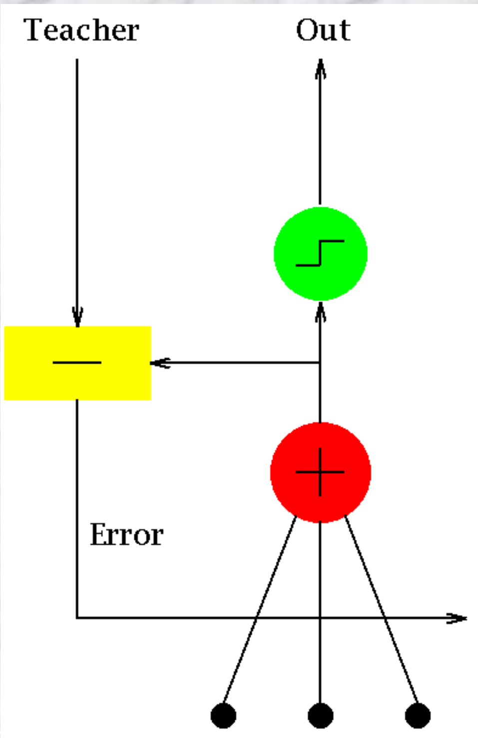
ADAptive LInear NEuron = ADALINE

rule:

difference between actual output and desired output is the background for error correction

History

- **ADALINE** is a single-layer artificial neural network and the name of the physical device that implemented this network. It is based on the McCulloch-Pitts neuron. It consists of a weight, a bias and a summation function.
- The difference between Adaline and the standard perceptron is that in the learning phase the weights are adjusted according to the weighted sum of the inputs (the net). In the standard perceptron, the net is passed to the activation (transfer) function and the function's output is used for adjusting the weights. There also exists an extension known as Madaline.



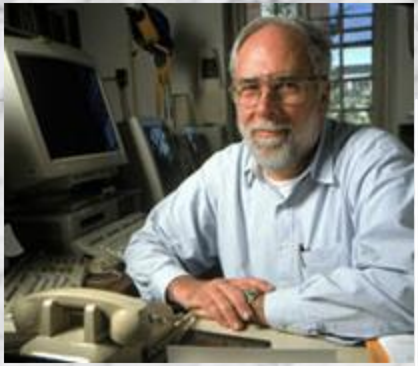
- 8 cells, 128 connections, 10^4/sec.

History

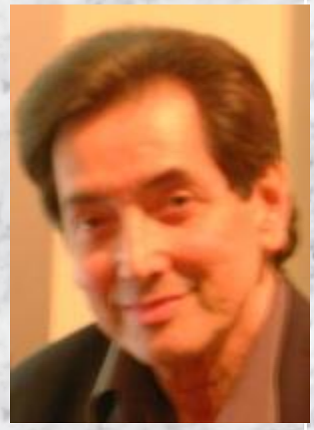
Teuvo Kohonen from Helsinki University of Technology has made many contributions to the field of artificial neuron networks, including the Learning Vector Quantization algorithm, fundamental theories of distributed associative memory and optimal associative mappings. His most famous contribution is the Self-Organizing Map (also known as the *Kohonen map* or *Kohonen artificial neural networks*, although Kohonen himself prefers *SOM*).

History

James Anderson from Brown University studied how brains and computers are different in the way they compute



Stephen Grossberg introduced in 1976 Adaptive Resonance Theory and Self-Organizing Maps for the learning. Outstar and Instar learning were combined by Grossberg in 1976 in a three-layer network for the learning of multi-dimensional maps.



History

In 1985-1990 Adaptive resonance theory (ART) was a theory developed by **Stephen Grossberg** and **Gail Carpenter** on aspects of how the brain processes information. It describes a number of neural network models which use supervised and unsupervised learning methods, and address problems such as pattern recognition and prediction

History

Kunihiko Fukushima from NHK Science and Technical Research Laboratories invented an artificial neural network, "Neocognitron ", which has a hierarchical multi-layered architecture and acquires the ability to recognize visual patterns through learning. He described a "Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position"

History

1982 John Joseph Hopfield

Neural Networks and Physical Systems with Emergent Collective Computational Abilities

New impulse for research !!!

Hopfield's Model

Hopfield found similarities between the neural networks and some physical, magnetic systems – the spin glass. Hopfield exploited an analogy to energy states in physics and introduced the *computational energy function*. Like a physical system, the network seeks its lowest energy state and with the iteration procedure converges to the stable state.

*

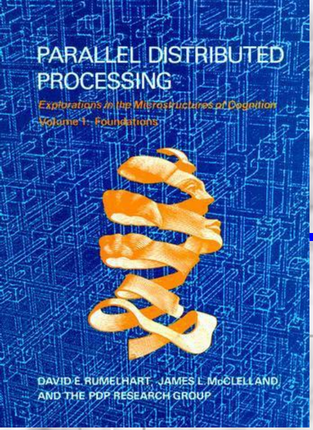
Hopfield's Model

System matches unknown input signal to one of previously stored signals.

Why Hopfield's works are so important ??

„stimulated" the interest in neural networks, gave the new way in the development in computers, united together the theory of neural networks with physics (particularly – optics, or optical information processing).

*



History

Backpropagation, an abbreviation for "backward propagation of errors", a method of training artificial neural networks used in conjunction with an optimisation method such as gradient descent. The method calculates the gradient of a loss function with respects to all the weights in the network. The gradient is fed to the optimization method which in turn uses it to update the weights, in an attempt to minimize the loss function.

The backpropagation algorithm was originally introduced in the 1970s, by Paul Werbos, wasn't fully appreciated until a famous 1986 book by David Rumelhart and James McCleeland *„Parallel Distributed Processing"* .

49

History

Boltzmann machine is a type of stochastic recurrent neural network invented by Geoffrey Hinton and Terry Seynowski in 1983. Boltzmann machines can be seen as the stochastic generative counterpart of Hopfield nets. The networks use well known ideas like simulated annealing.

Hardware implementation

From middle 80th the competition between laboratories and business from the electronic elements. The important parameters are:

- **number of neuronlike element in the network ,**
- **number of connections,**
- **the speed,**

Hardware implementation of neural networks in 1985-1988

Neurocomputer's name	Year	Number of elements	Number of connections	Speed	Creator
Mark III	1985	$8 \cdot 10^3$	$4 \cdot 10^5$	$3 \cdot 10^5$	R. Hecht-Nielsen, TRW
Neural Emulator Processor	1985	$4 \cdot 10^3$	$1.6 \cdot 10^4$	$4.9 \cdot 10^5$	C. Cruz, IBM
Mark IV	1986	$2.5 \cdot 10^5$	$5 \cdot 10^6$	$5 \cdot 10^6$	R. Hecht-Nielsen, TRW
Odyssey	1986	$8 \cdot 10^3$	$2.5 \cdot 10^5$	$2 \cdot 10^6$	A. Penz, Tex. Inst. CRL
Crossbar Chip	1986	256	$6.4 \cdot 10^4$	$6 \cdot 10^9$	L. Jackel, AT&T Bell Labs
Anza	1987	$3 \cdot 10^4$	$5 \cdot 10^5$	$1.4 \cdot 10^5$	R. Hecht-Nielsen, Neurocomp. Corp.
Parallon	1987	$9.1 \cdot 10^4$	$3 \cdot 10^5$	$3 \cdot 10^4$	S. Bogoch, Human Dev.
Anza plus	1988	10^6	$1.5 \cdot 10^6$	$6 \cdot 10^6$	R. Hecht-Nielsen, Neurocomp. Corp.

Neurocomputers

Neurocomputers
are computers, computer programs, or both, whose computational structure is very similar to the biological structure of the human brain.

*

Neurocomputers

Neurocomputers have been described as:

- neural computers
- neural networks machines
- artificial neural systems
- electronics neural systems
- parallel associative networks,
- parallel distributed processors
- sixth generation computers.

*

Neurocomputing

The field of **neurocomputing,** especially in the are of psychology, is often called connectionism.

*

Neurocomputers vs conventional computers

different tasks, different structure, so ... why expect similarities ???

Neurocomputers „exist" in the traditional computers, are simulated.

Neurocomputers should solve problems at which the brain seems very good and at which conventional computers and artificial intelligence seem poor.

*

Neurocomputers

Neurocomputers are both fast and excellent at recognizing patterns and thus they can also operate as expert systems. Like the brain they are self-organizing and essentially self-programming.

*

Comparison

Different structure and different rules, difficult to find the area of comparison.

Speed:

neuron sends approximately 1000 imp/sec

electronic chip – billion or more

Structure:

neural networks – parallel, many connections, (10 000)

electronic chip – serial (< 100)

*

Comparison

Computers are designed to carry out one instruction after another, extremely rapidly, whereas our brain works with many more slow units. Whereas computer can carry out a millions of operations every second - the brain respond about ten times per second. The computer is a high-speed, serial machine, and is used as such, compared to a slow, highly parallel nature of the brain.

*

Comparison

Computer usually has a long and complicated program, which gives it specific instructions as to what to do at every stage in its operation.
In such a computer its processing power is located, is concentrated in a single processing unit - central processing unit (CPU). The information on which computations or operations have to be performed are stored in the computer memory.

*

Comparison

As a result of a single processor - only one processing step can be executed in time. Moreover, when executing a processing step, the CPU has access only to a very small fraction of the memory. It means that in practice, only an insignificant portion of a system and systems' knowledge participates in the processing.

*

Comparison

It seem appropriate to distribute the processing capability across the computer's memory - each memory cell become an active processing element interacting with other such elements. This results in a massively parallel computer made up of an extremely large number of simple processing units - as many as these are memory cells.

*

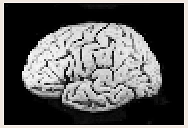
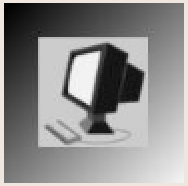
Comparison

Using such a massively parallel architecture would increase the computational power of a computer. This computer would be capable to execute many billions of operations per second.

The understanding of a neural architecture is very important for the development of massively parallel models of computation.

*

Comparison

	processing elements	element size	energy use	processing speed	style of computation	fault tolerant	learns	intelligent, conscious
	10^{14} synapses	10^{-6} m	30 W	100 Hz	parallel, distributed	yes	yes	usually
	10^{8} transistors	10^{-6} m	30 W (CPU)	10^{9} Hz	serial, centralized	no	a little	not (yet)

- • Volume: 1400 cm^3
- • Surface: 2000 cm^2
- • Weight: 1,5 kg
- • Cerebral cortex covering hemispheres contains 10^{10} nerve cells
- • Number of connections between cells: 10^{15}
- • Speed of sending/receiving information's = 10^{15}operations/sec

*

Software and Functional Comparisons

	Neurocomputers	Conventional Computers
Feedback Sensitivity	Excellent	None
Memory	High density Distributed, Associative	Low Density Localized, Specific
Database Search	Fast Close Match	Slow Exact Match
Mathematical and Algorithmic Ability	Poor	Excellent
Heuristic Ability	Excellent	Poor
Pattern Recognition Ability	Fast	Slow
Incomplete Pattern Recognition	Excellent	Poor

*

Hardware and Structural Comparisons

	Neurocomputers	Conventional Computers
Data Signal	Quasi-analog	Digital
Connectivity of Processing Elements	About 10 dynamically Changeable by Self-Programming	About 3 Not Changeable
Processing Sequence	Parallel, Simultaneous	Serial Independent
Site of Memory, Logic and Control	Nonlocal, Distributed in Connections	Localized to Processing Elements
Processing elements	Nonlinear. May be Nonthreshold. Arranged in Parallel	Linear, Threshold. Arranged in Series

*

Comparison of Fifth- and Sixth Generation Computers

	5th Generation	6th Generation
Main Usage	Artificial Intelligence	Pattern Recognition
Processing elements	VLSI	Artificial Neural Networks
Technologies	Silicon	Silicon, Optics, Molecular electronics
Architecture	Parallel Modules	Parallel Processing Elements
Connections	Externally Programmable	Dynamically Self-Programmable
Self-Learning	Limited	Good
Software Development	Major Role in Success	Minor Role in Success
Use of Neurobiology in Design	None	Moderate

*

Summary

Neurocomputer – it is information processing machine, composed from elements mimicking neural elements (neurons). These elements are of very simple construction:

- ***many inputs but one output only***
- ***incoming signals are summarized***
- ***the magnitude of the output signal depends from the input and so called threshold***

*

Summary

To distinguish the importance of the inputs signals are multiplied by *weights*.

So, the signal from out input can be different than identical signal from the another input.

*

Summary

Elements are connected forming the *net*.
Part of a net receive the input signals, the other part is connected to the net input, but the majority are interconnected to each other

structure of connections + weights

decides what neurocomputer will do

*

Summary

Main advantage:

ability for parallel processing

„Normal" computer perform operations in serial, while a neurocomputer perform many operations in parallel.

Even computer specially design for parallel processing – thousands processors – but neural networks – billions of processing elements.

*

Summary

Computer usually has a long and complicated program, which gives it specific instructions as to what to do at every stage in its operation.

Summary

The program for neurokomputer is in the structure of connections and the values of weights are its parameters. Moreover it has the learning capability.

Learning

Learning system is simple. The system has to solve the task with known answer
and we correct parameters in such a way – the system answer to be consistent with this answer.
Because about the elements' operation depends from its structure and weights

Learning =
change of weights

*

Learning

Two main rules:

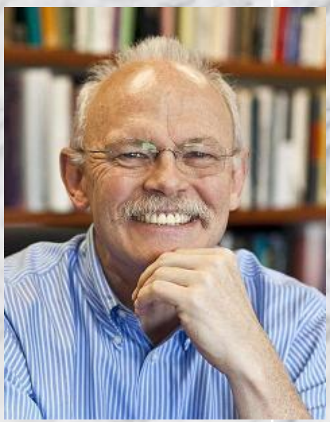
- **only neurons with wrong output signal are subject of the weights change**
- **the value of correction is proportional to the signal at the element input**

*

Learning

For the simple nets (1-2 layers) learning is simple. For the multilayer nets the special learning methods are used, more popular to ***the backpropagation method***

(*Parallel distributed processing..,* 1986, D.E.Rumelhart & J.L.McClelland, MIT)



*