Machine Learning and Data Mining: 2014

Volker Tresp

- Professor, LMU
- Principal Research Scientist, Siemens AG

I. Introduction

II. Non-Technical Perspectives

- »Logic and Philosophy
- »Psychology
- »Neurobiology

III. Technical Learning

IV. Details on the Lecture

I. Introduction

Why is machine learning of interest?

Thesis: Learning is one of the three fundamental mechanisms for the design and the improvement of autonomous (intelligent) systems

1: Intelligent Design

» Almost all practical solutions are based on intelligent design

» Engineer: the knowledgeable „Watchmaker“

» Programmer

Advantages:

» Explicit knowledge: the system is well understood and can be analysed and can be improved via analytic thinking

» Time constant: years

Disadvantage:

Need for an (expensive) designer (human)

2: Evolution

- Improvement via trial and error
- Biological evolution
  - the blind „watchmaker“
- Technical evolution
  - Evolutionary improvement of technical solutions

Advantages:

- Simple (blind)
- Self-Optimizing

Disadvantages:

- Time constant: years, decades, centuries, …
- Wasteful

3: Learning

Biological Learning:

» Lifelong optimization of the behavior of an individual via interaction with the environment

» Even primitive animals can learn

» Basic properties of animals (“natural law”)

» Feedback of the learning success (*reinforcement*)

» Time constants: days

Human

» Learning with a teacher

» Long time inheritance of the learned knowledge via culture and books

» Learning via processing of information

3: Learning (cont‘d)

Machine Learning

» Broadest sense: attempt to mimic biological learning for technical purposes

» Autonomous optimization of a technical system via interaction with reality or by analyzing acquired data

» „Learning instead of programming“

Characterization of Learning

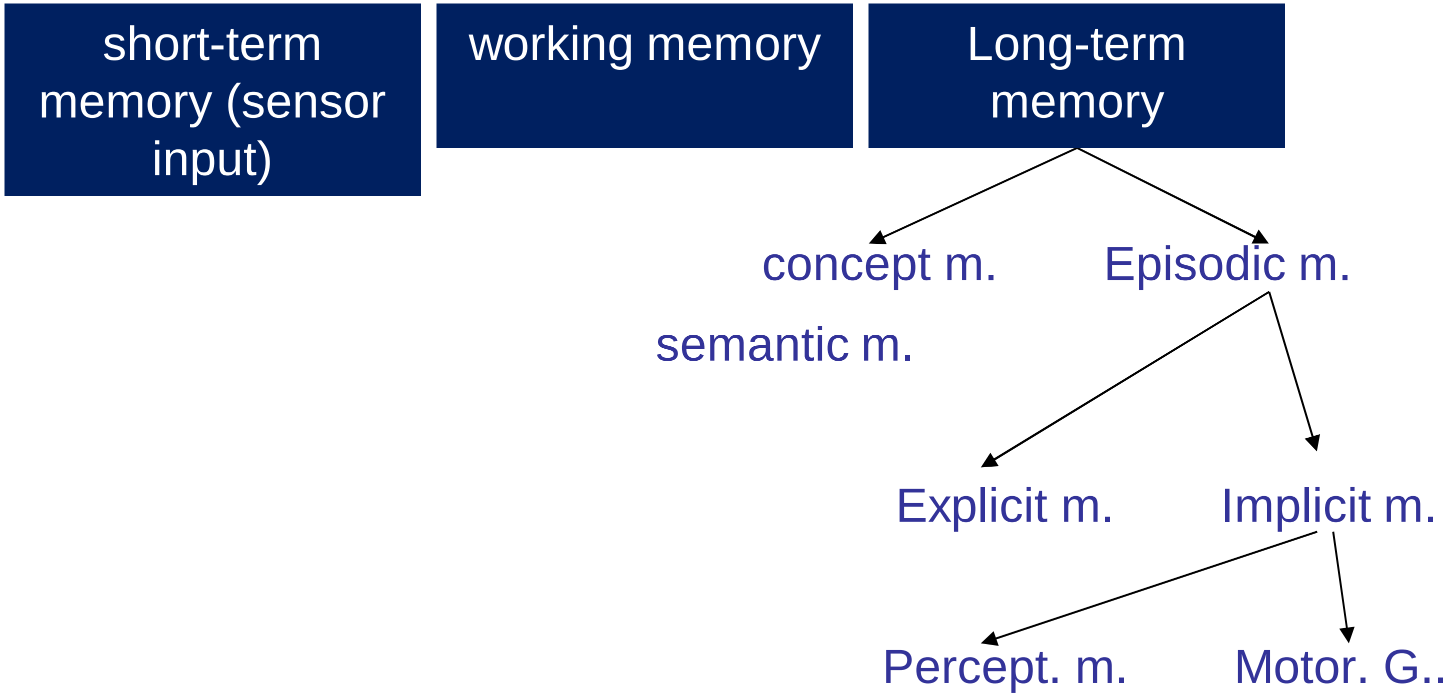
- **Learning is an exclusive property of living beings** (and computers?)
  - Even quite primitive animals can (adaption)
- Biological Learning:
  - (beneficial? permanent?) Modifications in the (central?) nervous system (based on interactions with the environment?)
- Machine Learning:
  - Beneficial changes in a technical system based on the analysis of data or based on the interaction with the environment by using learning algorithms

Etymological Origin

- Etymologically, “learning” belongs to the word group of "leisten", which originally means “to trace something”

- Even etymologically, “learning” has something to do with the idea of “leaving traces”

Leaving a trace: having a memory



- The human brain is incredible
  - Somehow my brain can store all movies, books, comics, music, paintings, lectures, which I have ever consumed. What else? Why?
  - It also becomes technically possible!

II. Non-technical Perspectives

1. *Philosophy*
2. *Psychology*
3. *Neurobiology*

Remark

Common threads in Philosophy, Psychology, Cognition, and even Biology:

Dominance of

- Internal mechanisms; inside-outside view
  - Logic: Deduction
  - Philosophy: Rationalism, Idealism
  - Biological intelligence: Psycho Analytics, Cognition
  - Inheritance
- External influences dominate; outside-inside view
  - Logic: Induction
  - Philosophy: Empiricism; Positivism
  - Biological Intelligence: Psychology; Behaviorism
  - External influence

II. Non-technical Perspectives

1. *Philosophy*
2. *Psychology*
3. *Neurobiology*

Basic Concepts: Deduction and Induction

Deduction:

» From the general to the specific (*top-down*)

» Axioms are given and theorems are derived via the machinery of deductive reasoning

» Axioms:

- » Can be simple facts ("Jack's height is 180 cm")
- » Complex Axioms ("If something is a dog, it is also a mammal")

» Basis for the classical Artificial Intelligence

Induction:

» Generalizes observations (*bottom-up*), to generalize and to justfy theories

» Inferring the validity of a hypothesis via observations and experiences

- » Simple Facts ("Jack's height is 180 cm") as in deduction
- » **Learned dependencies instead of assumes axioms!**

» Basis for Machine Learning

Rationalism (from latin: ratio = „reason“)

- Priority of rational reasoning
  - In knowledge acquisition; in contrast to other forms such as the senses or religious convention

- Representatives: Socrates (ca 470–399B.C.E.), René Descartes (1596–1650), Baruch Spinoza (1632–1677), Gottfried Leibniz (1646–1716), Immanuel Kant (1724–1804)
  - Since the Enlightenment, rationalism is usually associated with the introduction of mathematical methods into philosophy, as in Descartes, Leibniz, and Spinoza (Bourke 263). This is commonly called **continental rationalism**, because it was predominant in the continental schools of Europe, whereas in Britain empiricism dominated
  - Proponents of some varieties of rationalism argue that, starting with foundational basic principles, like the axioms of geometry, one could deductively derive the rest of all possible knowledge

Empiricism

- » More of a British tradition
- » All knowledge originates in the senses, the observations or the experiment
- » In contrast to Rationalism
- » „ There is nothing in the mind that was not first in the senses“ (Hobbes?)
- » Representatives: Francis Bacon (1562-1626), John Locke (1632-1704), George Berkeley (1685-1753), David Hume (1711-1776)

Idealism (dominant Philosophy of the 19th century) and Materialism

- **Idealism:** each form of matter, including human behavior, is a reflection of ideas
- **Materialism** (Ludwig Feuerbach, Karl Marx): history is not driven by ideas but by laws (historic-dialectic materialism)

- In philosophy, idealism is the group of philosophies which assert that reality, or reality as we can know it, <u>is fundamentally mental, mentally constructed, or otherwise immaterial.</u> Epistemologically, idealism manifests as a skepticism about the possibility of knowing any mind-independent thing. In a sociological sense, idealism emphasizes how human ideas — especially beliefs and values — shape society. As an ontological doctrine, idealism goes further, asserting that all entities are composed of mind or spirit. Idealism thus rejects physicalist and dualist theories that fail to ascribe priority to the mind
- Beginning with Immanuel Kant, German idealists such as G. W. F. Hegel, Johann Gottlieb Fichte, Friedrich Wilhelm Joseph Schelling, and Arthur Schopenhauer dominated 19th-century philosophy

Positivism (after 19th century)

- Knowledge is derived from positive findings
- Data derived from sensory experience, and logical and mathematical treatments of such data, are together the exclusive source of all authentic knowledge
  - Modern scientific thinking

- This view holds that society operates according to laws like the physical world. Introspective and intuitional attempts to gain knowledge are rejected
- The concept was developed in the early 19th century by the philosopher and founding sociologist, Auguste Comte. Comte argued that society operates according to its own laws, much as the physical world operates according to gravity and other laws of nature. Also: Ernst Mach.

- Stephen Hawking is a recent high profile advocate of positivism, at least in the physical sciences. In The Universe in a Nutshell he writes:
  - Any sound scientific theory, whether of time or of any other concept, should in my opinion be based on the most workable philosophy of science: the positivist approach put forward by Karl Popper and others. According to this way of thinking, a scientific theory is a mathematical model that describes and codifies the observations we make. A good theory will describe a large range of phenomena on the basis of a few simple postulates and will make definite predictions that can be tested… If one takes the positivist position, as I do, one cannot say what time actually is. All one can do is describe what has been found to be a very good mathematical model for time and say what predictions it makes
  - (Popper does not see himself as positivist)

Karl Popper: Scientific Discovery

» Is induction sound?

» Karl Popper (Sir, 1902-1994). *The Logic of Scientific Discovery*

» **If no finite set of observation can ever prove a theory, how can we ever accept a scientific theory as being true?**

» Popper accepts Empiricism as a valid means to increase knowledge, if one accepts that theories can only be tested but never be proven

Karl Popper

II. Non-technical Perspectives

1. *Philosophy*
2. *Psychology*
3. *Neurobiology*

Psychoanalytics

- Psychoanalysis was founded by Sigmund Freud (1856-1939)
- Hypothesis: people can be cured by making conscious their unconscious thoughts and motivations, thus gaining “insight”

Psychology as Empirical Science

Begin of empirical psychology:

» Herrmann von Helmholtz (1821-1894)

» Wilhelm Wundt (1832-1920) (Assistant to Helmholtz)

  » Wundt is considered to be the founder of psychology as a separate scientific field

  » From 1858 to 1863, he was assistant to Hermann von Helmholtz. “Theorie der Sinneswahrnehmungen”

» Gustav Theodor Fechner (1801–1887): Founder of Psychophysics

  » "the scientific study of the relation between stimulus and sensation"

Behaviorismus (1920-1960)

- *„Belief in the existence of consciousness goes back to the ancient days of superstition and magic“*
- Also as reaction to Sigmund Freud
- Rejection of theories that need to assume mental states
  - The inner structure (of the brain) is irrelevant
- The functioning can only be deduced from input (stimulus) and output (Reaction)
- Representatives: Iwan Pawlow (1849-1936), John Watson (1878-1958), B. F. Skinner

Learning in Psychology

- **Definition:** permanent changes in behavior trough training
- **Habituation:** Learning to ignore a stimulus
- **Classical Conditioning**: After training the saliva flow (conditional response) starts as a reaction to the bell (conditional stimulus) even without food (unconditional stimulus) being present
- **Operative Conditioning**: Learning that after ans action a certain consequence follows
  - A rat is trained to open the cage which leads to a reward (food)
- **Complex Learning**: Learning which goes beyond simple association. Application of a strategy for problem solving or for forming a complex spatial map

Classical Conditioning (Pawlow)

Unconditional stimulus (food)	⇨	Unconditional response (salivation)		
Unconditional stimulus (food) With conditional stimulus (bell)	⇨	Unconditional response (salivation)		
Conditional stimulus (bell)	⇨	Unconditional response (salivation)		

Cognitive Psychologie and Cognition

- » Attempt to understand the inner working of the „Black Box“
- » Reaction to Behaviorism
- » Human behavior is more than stimulus-response
- » Development is an active process of a subject
- » Acting is dominated not only by a stimulus but by active reasoning (children can develop)
- » The link between stimulus and behavior is the cognitive representation
- » Williams James (1842-1910), Herrmann von Helmholtz (1821-1894), Frederik Bartkett (1886-1969), George Miller (*The magic number seven*, 1956), Noam Chomsky (*Three Models of Language*, 1956)

Cognitive Psychology

» Perceptron

» Memory

» Language

» Learning

» Emotions

» *Reasoning*

» Planning

» Attention

» Construction of scenes and memories (construal)

» Consciousnes: Why did we develop consciousness?

Cognitive Psychology and Learning

Learning is defined relatively narrowly

» As inductive reasoning and not memorization

» Learning in problem solving

Although learning is ubiquitous

» The world “constructs” the brain

  » Of all possible languages, we learn one

    » The basis to Chomsky's linguistic theory is that the principles underlying the structure of language are biologically determined in the human mind and hence genetically transmitted (Chomsky)

» Memory and learning

  » Classification of new Objects

  » Learning of Prototypes

» …

II. Non-technical Perspectives

1. *Philosophy*
2. *Psychology*
3. *Neurobiology*

Neurobiology

- Learning: „Neuronal Plasticity“
- There must be a physical change if something is learned or memorized
- *Central mechanism: Synapses change their efficiency (synaptic plasticity)*
  - *Short-term plasticity:* the change lasts milliseconds to minutes
  - *Long-term plasticity:* the synaptic efficiency changes for hours to life-long

Neuron

- Resting potential: -70 mV.
- Depolarisation: > -50mV
  - -> Opening of the sodium channels; action potential
- Refractory period: during this time no new action potential can begenerated, independent of activation strength (appr. 2ms)
- Systems theory: *leaky integrator*

Synapse

Presynaptic: Presynaptic discharge of the action potential leads to the release of neuro transmitters

Postsynaptic: opening of ion channels and thus change of the postsynaptic membrane potential

Example: Aplysia

- Eric Richard Kandel (* 7. November 1929 in Vienna): US-American Neuroscientist with Austrian origin
- Study object: Californian see slug (*Aplysia californica*)
- Nobel price 2000
- Gill-withdrawal reflex with 24 sensory-neurons and 6 motor-neurons
- Habituation
  - Reduction of neurotransmitters with repeated stimuli
- Sensitization:
  - Increase of neurotransmitters with repeated (damaging) stimuli
- Association:
  - Light/electric chock

The Synapse (left) influences a second synapse (bottom)

Short-term memory (minutes):

» A weak stimulus results in the phosphorylation of proteins of the ion channels (weak arrow on the left) which results in an increase of neurotransmitters

Long-time memory (weeks):

» A strong stimulus (thick arrow, left) results in a increased level of the messenger cAMP (Cyclic adenosine monophosphate), which results in an amplification of the protein kinase

» This influences the cell's DNA and new proteins are generated

» This results in an increases efficiency of the synapse: its efficiency is increased and mode neurotransmitters are generated

Hebb Learning in Psychology und Neurophysiology

» Kandel's results supplied new evidence for the Hebb's law

- » "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."
- » "Neurons that fire together wire together"

» Hebb learning has been proven, i.e., in the neurons of the hippocampus

» Hebb formulates learning much more abstractly than Kandel

- » Open question: how much can one ignore biological details without loosing the essence (e.g. spike timing?)

Conclusion: Nontechnical Learning and Machine Learning

- Philosophy: Empiricism, Positivism; formalization of the empirical viewpoint (Popper)

» Psychology
  - » Behaviorists; Psychophysics
  - » Cognitive Psychology

» Biology
  - » Learning in Neurobiology is little understood
  - » Neuroinformatics: Inspiration from Biology
  - » Machine Learning produces learning algorithms and learning models, which are of interest biologically

III. Machine Learning

1. *Before the computer age: Statistics*
2. *Neuroinformatics*
3. *AI and Machine Learning*
4. *Neuroinformatics: Revival*
5. *Modern Machine Learning*
6. *Data mining; Big Data*

III. Machine Learning

1. *Before the computer age: Statistics*
2. *Neuroinformatics*
3. *AI and Machine Learning*
4. *Neuroinformatics: Revival*
5. *Modern Machine Learning*
6. *Data mining; Big Data*

Thomas Bayes (Rev., 1701 -1761)

» Updating the degree of belief in hypothesis based on observations

» *P(H=1):* degree of belief in the truthfulness of a hypothesis *H* (a priori assumption)

» *P(D|H=1):* Plausibility of the data *D*, if the hypothesis *H* is true (Likelihood)

» *P(D|H=0):* Plausibility of the data *D*, if the hypothesis *H* is false

» Bayes' theorem:

» $$P(H=1|D) = P(D|H=1)\, P(H=1) / P(D)$$

(a posteriori probability of the hypothesis)

» Pierre-Simon Laplace (1749–1827) then developed the theory

Bayesian Statistics is based on the concept of Subjektive Probability

- Subjective probability:
  - Before I throw a coin, what is the probability that it is a fair coin
  - I believe that the probability that party X wins the election is 45%

- Cox (1946): Cox's theorem implies that any plausibility model that meets the postulates is equivalent to the subjective probability model, i.e., can be converted to the probability model by rescaling

  - If a 1 corresponds to the belief that an event happens with certainty and if a 0 corresponds to the belief that an event does not happen, and numbers in between corresponds to degrees of certainty, then these numbers exactly behave as probabilities

Critique on Bayesian Statistics

Karl Pearson (1857 – 1936) now considered the <u>founder of modern statistics (frequentist statistics)</u>

» *"I felt like a buccaneer of Drake's days -... I interpreted that sentence of Galton to mean that there was a category broader than causation, namely correlation, of which causation was only the limit, and that this new conception of correlation brought psychology, anthropology, medicine, and sociology in large parts into the field of mathematical treatment."*

Sir Ronald Aylmer Fisher (Sir, 1890-1962)

» Criticism on the role of subjective probabilities: frequentists only make statements on repeatable experiments

» One evaluates if the data contradict a hypothesis but one does not make statements about the probability of a hypothesis

Egon Pearson (1895-1980) Son of Karl P., **Jerzy Neyman** (1894-1981)

III. Machine Learning

1. *Before the computer age: Statistics*
2. *Neuroinformatics*
3. *AI and Machine Learning*
4. *Neuroinformatics: Revival*
5. *Modern Machine Learning*
6. *Data mining; Big Data*

Remark

The two poles we found in Philosophy, Psychology and Biology are also present In the history of intelligent systems

Dominance of

- Internal mechanisms; inside-outside view
  - Classical Artificial Intelligence (AI)
- External influences dominate; outside-inside view
  - Statistics
  - Neural Networks
  - Machine Learning

Neuroinformatics

- First: focus on expressiveness of neural networks (and not learnability)
- *McCulloch and Pitts (1943):* first attempt to formalize brain functions via simple computational nodes (network of simple logical units)

Expressivenes of Neural Structures

- *John v. Neumann (1956):* investigated the error tolerance of Neural Networks ("*reliable computing with unreliable elements*")
- *John v. Neumann (1958):* Computer and the Brain
- *John von Neumann concludes that the brain operates in part digitally, in part analogically, but uses a peculiar statistical language unlike that employed in the operation of man-made computers*

Learning in Neuronal Structures

» *Hebb (1949):* Repeated activation of one neuron by another, across a particular synapse, increases its conductance (Hebb‘s theorem); "*Neurons that fire together wire together*"

» Hebb tried to explain classical conditioning via neual mechanisms

- *Wiener (1949):* Cybernetics, or control and communications in the animal and the machine *The whole world -- even the universe -- could be seen as* ***one big feedback system subject to the relentless advance of entropy,*** *which subverts the exchange of messages that is essential to continued existence (Wiener, 1954).* Book: *Cybernetics or Control and Communication in the Animal and the Machine* (1948)

Learning in Neuronal Structures: Assoziative Memory

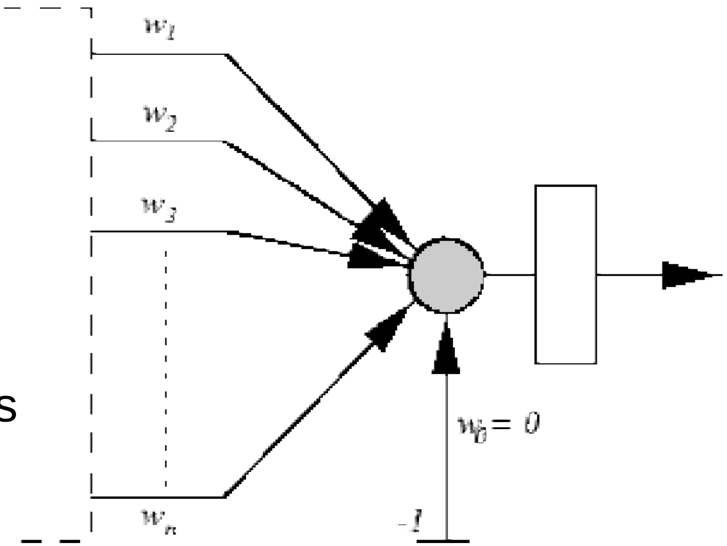
- W. K. Taylor (1950er), Karl Steinbuch (1961)
- Associative memory, “Lernmatrix”
- Relationship to Hebb Learning

Perceptron and ADALINE

» Minsky developed 1954 in his dissertation a neural computer he called SNARC (Stochastic Neural Analog Reinforcement Calculator)

» Rosenblatt developed 1958 the Perceptron learning rule and formulated a convergence proof; Mark I Perceptron

» Widrow and Hoff developed 1960 the ADALINE (ADaptives LINeares Element) (used in modems)



» Minsky and Papert published 1969 the book „Perceptrons“ and demonstrated the limitations of the Perceptrons and of the ADALINE (Exclusive-Or Problem)

III. Machine Learning

1. *Before the computer age: Statistics*
2. *Neuroinformatics*
3. *AI and Machine Learning*
4. *Neuroinformatics: Revival*
5. *Modern Machine Learning*
6. *Data mining; Big Data*

The Classical Area of Artificial Intelligence (AI): an Ice Age for Neuroinformatics

» After the book of Minsky and Papert funding almost exclusively went into the emerging field of AI

» No more funding for the study of learning systems

» A brief history of classical AI (1960s to 1980s)

AI and Induction

- Maschinelles Lernen is „Intelligence via Learning“
  - Induction
  - Empiricism
- (Classical) AI: Axioms permit the derivation of theorems form axioms via deducion

Four AI Goals

- » Understanding human thinking: Cognition
- » Indistinguishable from human acting (Turing Test)
  - » Language, knowledge representation, reasoning, learning, vision, robotics
- » Rational (optimal) (not necessarily human) Reasoning: Logic
- » Rational (optimal) (not necessarily human) Acting: Agents

AI: Roots

Philosophical Basis:

» Vienna School (Rudolf Carnap (1891-1970))

» **Logical Positivism:** All knowledge can be characterized by logical theories …

» Ludwig Wittgenstein (1889-1951), Bertrand Russel (1872-1970)

Birth of AI: Dartmouth Workshop (1956)

John McCarthy (Dartmouth, later Stanford) (1927-2011)

» Naming: AI (To distinguish from Cybernetics); Inventor of LISP

Marvin Minsky (1927-) (MIT)

» SAINT (calculus integration); ANALOGY (geometric analogy); STUDENT (algebra); Blocks World; The Society of Mind (1985); Anti-Logic and Anti-Neuro

Claude Shannon (1916-2001) (Bell Labs) Inventor of Information Theoriy

Arthur Samuel (1901-1990) IBM; checkers program

Ray Solomonoff (1926-2009) (MIT) Founder of Algorithmic Probability

John von Neumann Institute for Advanced Study; Inventor of Game Theory

Allen Newell (1927-1992)(CMU), **Herbert Simon** (1916-2001) (CMU) (Nobel P.):

» General Problem Solver (GPS): a program to solve general problems (terminated after 10 years)

» Representative of strong AI: Intelligence is independent of substrate

Nathaniel Rochester (IBM), Trenchard More (Princeton), Oliver Selfridge (MIT), Cliff Shaw

Further Development

Early Enthusiasm (1952-1969)

- In the first AI phase there was an unlimited expectation with respect to the capabilities of computers to „solve tasks for which intelligence is required, if they would be executed by humans " (Minsky).
- Herbert Simon (1957)
  - Within the next 10 years a computer will be world champion in chess and will derive an important mathematical theorem
  - In don't want to chock you …There are now in the world machines that think … in a visible future the range of problems they can handle will be coextensive with the range to which the human mind has been applied…
- In 1958 McCarthy proposed to formalize the complete human knowledge in form of a homogeneous formal representation, first order predicate logic.

First Reality-Dose (1966-1973)

- Translation of Russian into English was stopped: "the spirit is willing but the flesh is weak" became "the vodka is good but the meat is rotten"
- Reasoning did not scale up

Knowledge-based Systems

Knowledge-based Systems(1969-1979)

» Expert systems: In an expert system. There is a formal knowledge representation, for example as a set of rules, and is applied to facts to infer new facts

» Bruce Buchanan: *DENDRAL* (1969); inferring molecular structure from mass spectroscopy data; first knowledge intensive system

» Ed Feigenbaum (Stanford): *Heuristic Programming Project* (HPP)

» Feigenbaum, Buchanan, Edward Shortliffe; *MYCIN:* Diagnose blood infections; extensive interviewing of experts; uncertainty factors

» Progress in NLP: Eugene Charniak, Roger Shank

» PROLOG

AI becomes an Industry (1980- and a few years later)

» McDermott: **R1** (DEC, 1982); Configuration of new computer systems; each major company has an AI group

» Japan (1981) Fifth Generation Project; 10-year project for the realization of intelligent computers based on PROLOG

Collapse (1984) of many Silicon Valley Start-Ups (<u>Beginning of the AI winter</u>)

Maschinelle Learning in Classical AI

» Machine Learning was not in focus in classical AI (*only deductive inference is sound*)

» The field wanted to distinguish itself from statistics and probability

» Focus on symbolic Machine Learning

» Out off this tradition

   » Case-based reasoning (*case-based reasoning*, CBR) (Schank, 1977)

   » Learning of decision trees (Ross Quinlan's ID3, 1979, Rivest)

   » *Inductive Logic Programming* (Stephen Muggleton, 1991)

      » Intuitively attractive: The goal is to extract simple logical rules

      » Powerful: One can learn (*first-order*) Prolog Rules (Turing-equivalent)

III. Machine Learning

1. *Before the computer age: Statistics*
2. *Neuroinformatics*
3. *AI and Machine Learning*
4. *Neuroinformatics: Revival*
5. *Modern Machine Learning*
6. *Data mining; Big Data*

Revival of Neuroinformatics

- There was increasing interest in Neuroinformatics around the mid 80s; end of the neural winter (1969-1982)
- End of the AI hype: maybe the substrate is relevant after all
- Learning in focus; opposition to rule-based approaches
- Fascination Brain: despite the biological complexity there should be a simple organizational principal, which leads to intelligence via learning. Maybe intelligence can only be reached via learning?
- Technical high performance solutions could be achieved relatively easily

Revival of the Assoziative Memory

- **John Hopfield (1982, 1984):** *Neural networks and physical systems with emergent collective computational abilities*
- Achievements:
  - Associative memory(Hebb learning)
  - Combinatorial optimization
- Contributions from statistical physics (Spin-glasses)
- Interesting features: Nonlinear, parallel, error tolerant, feedback
- Implementation as optical computer?
- Relationship to brain functioning
- At the end: solutions were not technically competitive
- Independent and earlier: Stephen Grossberg, Teuvo Kohonen

Multi-layer Perzeptron

Ackley, Hinton, Sejnowsky (1985): *Boltzmann Machine*
» Discriminative Learning
» Theoretically very interesting but not as practical as the MLP

Rumelhart, Hinton, Williams (1986): Multi-layer Perceptron (MLP)
» MLP: a robust powerful tool for modeling high-dimensional nonlinear dependencies
» Solution to the *exclusive-or-problems,* Nettalk
» MLP: superior modeling tool for high-dimensional problems
» *Neuroinformatics breakthrough*
» Interest in Statistics

» Since 1988 a certain hype („Learning instead of programming“)

III. Machine Learning

1. *Before the computer age: Statistics*
2. *Neuroinformatics*
3. *AI and Machine Learning*
4. *Neuroinformatics: Revival*
5. *Modern Machine Learning*
6. *Data mining; Big Data*

Modern Machine Learning

- » In a similar way as Logic dominates classical AI, there is a great influence of statistical thinking in Machine Learning from the early 1990s onward
- » In the second half of the 1990s 90er modern machine learning became Statistical Machine Learning
- » AI became highly influenced by Machine Learning

Main Phases

- Late 1980s until around 2000
  - MLP, Neural Computation, strong influence from physics (mean field theory), model diversity (Hopfield network, Kohonen networks, Boltzmann machine, …)
- Since second half of 1990s
  - Statistical Learning Theory, Support Vector Machines, margin approaches
  - Bayesian networks and Bayesian learning, Gaussian Processes
  - Committee machines, ensemble methods
- Since early 2000s
  - Nonparametric Statistics
- Since second half of 2000s
  - Factorization approaches; Web and search
- Since around 2010
  - Deep Learning, MLP revival

III. Machine Learning

1. *Before the computer age: Statistics*
2. *Neuroinformatics*
3. *AI and Machine Learning*
4. *Neuroinformatics: Revival*
5. *Modern Machine Learning*
6. *Data mining; Big Data*

Fields by Applications Foci

- **Statistics**
  - Focus on the significance of a dependency (does the medication work or not) and interpretability
  - Few data points, few variables
- **Machine Learning**
  - Vision of a intelligent learning machine
- **Data Mining**
  - Analysis of large data bases; discovery of patterns in data
  - Today: also WWW focus
  - Large number of variables, large data sets
- **Big Data**
  - Google, Facebook and others make huge profits in analysing huge data sets (Map Reduce, Hadoop, …)
    - Is this also relevant to other industries? Spy on your customer?
  - Very Large Databases meets Machine Learning, Data Mining, Statistics
    - Log-linear models, Random Forests, Deep Learning

Data Mining

- Data Mining as part of the KDD - Process (**Knowledge Discovery in Databases** (KDD))
- History:
  - 1989 IJCAI Workshop on KDD
  - 1995 KDD Conference
  - 1998 SIG KDD Conference
- Based on available statistical and machine learning approaches. But also development of novel approaches
  - Frequent Item Sets, pattern discovery, Association Rules
  - DBSCAN (Ester, Kriegel, Sander, XU) ...

Today

- ML has increasing impact on other areas such as Vision, Speech Recognition, Information Retrieval, Information Extraction, Bioinformatics, …
- The big player in the age of information, as Microsoft, Google, Yahoo, Amazon, Facebook, are hiring huge amounts of machine learners
- Big Data Hype: most large companies have activities and are looking for “data scientists”

IV. Details on the Lecture

The Lecture

- Technical foundation of approaches which are in focus today

- Mathematics
  - Linear Algebra (Vectors, Matrices, …)
  - Probability
  - Statistics

Literature

Lecture

- *The Elements of Statistical Learning: Data mining, Inference and Prediction*. Hastie, Tibshirani, Friedman: Springer (2nd Ed.). [Moderne Statistik; frequentistisch] Download at http://www-stat.stanford.edu/~tibs/ElemStatLearn/
- *Machine Learning: a Probabilistic Perspective. Kevin Murphy: MIT Press [very popular; Bayesian oriention]*
- *Bayesian Reasoning and Machine Learning. David Barber*. Download at *http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage* *[Bayesiani orientation]*
- *Pattern Classification. Duda, Hart, Storck: Wiley [Pattern recognition]*
- *Pattern Recognition and Machine Learning*. Bishop: Springer *[Bayesian touch]*
- *Data Mining: Concepts and Techniques*. Han and Kamber: Morgan Kaufmann [Datamining]
- *Artificial Intelligence-a Modern Approach*. Russel and Norvig, Prentice Hall [All of AI]
- *Kernel Methods for Pattern Analysis.* John Shawe-Taylor and Nello Cristianini: Cambridge University Press [Kernel approaches]
- *Machine Learning*. Tom Mitchel: McGraw-Hill [Some excellent Chapters; some outdated]
- Andrew Ng’s coursera course: https://www.coursera.org/course/ml

Literature (cont`d)

Time Series

- *Time Series Analysis*. Hamilton

Reinforcement Lernen and Game Theorie

- *Reinforcement Learning: an Introduction*. Sutton and Barto: MIT Press
- *Fun and Games: A Text on game Theory*. Binmore and Linster, Houghton Mifflin

Statistics

- *Bayesian Data Analysis*. Gelman, Carlin, Stern, Rubin: Chapman
- *Heckerman's Tutorial:* http://research.microsoft.com/research/pubs/view.aspx?msr_tr_id=MSR-TR-95-06
- *Statistik.* Fahrmeir, Kuenstler, Pigeot, Tutz: Springer
- *Introduction to Mathematical Statistics.* Hogg, Craig: Prentice Hall
- *Probability, Random Variables and Stochastic Processes.* Papoulis, McGraw, Hill

Movie

- Youtube Video: The Machine That Changed The World Part 3
  - https://www.youtube.com/watch?v=-NRHAaOJAVE

- 10:20: Selfridge, Shannon, Minsky
- 22:20: Failures
- 40:10: Perceptron
- 45:00: Nettalk

http://www.dbs.informatik.uni-muenchen.de/cms/Maschinelles_Lernen_und_Data_Mining

Zeit und Ort

Veranstaltung	Zeit	Ort	Beginn
Vorlesung	Mi, 9.00 s.t. - 12.00 Uhr	Raum 061 (Oettingenstr. 67)	09.04.2014
Übung	Do, 14.00 - 16.00 Uhr	Raum 169 (Oettingenstr. 67)	17.04.2014
	Do, 16.00 - 18.00 Uhr	Raum 169 (Oettingenstr. 67)	17.04.2014

Planung

Vorlesung		Übung	
Datum	**Thema**	**Datum**	**Blatt**
09.04.14	Introduction	10.04.14	*selbstständig*: R - Einführung (Sourcen)
16.04.14	Perceptron; Linear Algebra (review); Linear Regression	17.04.14	Blatt 1
23.04.14	Basis Functions; Neural Networks	24.04.14	
30.04.14	Deep Learning; Kernels	entfällt (Feiertag)	
07.05.14	Probability (review)	08.05.14	
14.05.14	Frequentist and Bayesian Statistics; Linear Classifiers	15.05.14	
21.05.14	Support Vector Machine	22.05.14	
28.05.14	Model Comparison	entfällt (Feiertag)	
04.06.14	Principal Component Analysis	05.06.14	
11.06.14	Recurrent Neuronal Networks and Time Series (Guest Lecture)	12.06.14	
18.06.14	Graphical Models	entfällt (Feiertag)	
25.06.14	Reinforcement Learning	26.06.14	
02.07.14	Overview over other relevant topics: Clustering, Random Forest, etc.	03.07.14	

Übungsbetrieb

Zur Vertiefung der Vorlesung werden 2-stündige Übungen angeboten, in denen die vorgestellten Verfahren weiter erläutert und an praktischen Beispielen veranschaulicht werden. Da es sich mitunter um Programmieraufgaben handelt, ist eine vorherige Vorbereitung des aktuellen Übungsblattes erwünscht um Fragen diesbezüglich besser beantworten zu können.