A historical perspective on Machine Learning (on the occasion of the 25th Benelearn) Luc De Raedt

A historical perspective on Machine Learning (on the occasion of the 25th Benelearn)

Luc De Raedt



A historical perspective on Machine Learning (on the occasion of the 25th Benelearn)

Luc De Raedt





Based on a true story of Machine Learning

A historical perspective on Machine Learning (on the occasion of the 25th Benelearn)

Luc De Raedt





Based on a true story of Machine Learning



PERSONAL PERSPECTIVE

Ryszard Michalski ...

Ryszard Michalski, Tom Mitchell, Jaime Carbonell

Ryszard Michalski, Tom Mitchell, Jaime Carbonell



1983



Preface

The ability to learn is one of the most fundamental attributes of intelligent behavior. Consequently, progress in the theory and computer modeling of learning processes is of great significance to fields concerned with understanding intelligence. Such fields include cognitive science, artificial intelligence, information science, pattern recognition, psychology, education, epistemology, philosophy, and related disciplines.

The recent observance of the silver anniversary of artificial intelligence has been heralded by a surge of interest in machine learning—both in building models of human learning and in understanding how machines might be endowed with the ability to learn. This renewed interest has spawned many new research projects and resulted in an increase in related scientific activities. In the summer of 1980, the First Machine Learning Workshop was held at Carnegie-Mellon University in Pittsburgh. In the same year, three consecutive issues of the *International Journal of Policy Analysis and Information Systems* were specially devoted to machine learning (No. 2, 3 and 4, 1980). In the spring of 1981, a special issue of the *SIGART Newsletter* No. 76 reviewed current research projects in the field.

Before 1980 — Handbook of AI 1981 overview

THE HANDBOOK OF ARTIFICIAL INTELLIGENCE

Volumes I and II by Avron Barr and Edward A. Feigenbaum Volume III by Paul R. Cohen and Edward A. Feigenbaum

VOLUME I

- I. Introduction A. Artificial Intelligence
- B. The Al Handbook
- C. The Al literature

II. Search

- A. Overview
- B. Problem representation
- 1. State-space representation 2. Problem-reduction representation
- 3. Game trees
- C. Search methods
- 1. Blind state-space search
- 2. Rind AND/OR graph search
- 3. Heuristic state-space search
- a. Basic concepts in heuristic search
- b. A*-Optimal search for an optimal solution
- c. Relaxing the optimality requirement
- d. Bidirectional search
- 4. Heuristic search of an AND/OR graph
- 5. Game tree search
- a. Minimax procedure
- b. Alpha-beta pruning
- c. Heuristics in game tree search
- D. Sample search programs I. Logic Theorist
- 2. General Problem Solver
- 3. Gelernter's geometry theorem-proving machine
- 4. Symbolic integration programs
- 5. STRIPS
- 6. ABSTRIPS

III. Knowledge Representation A. Overview

- R. Survey of representation techniques.
- C. Representation schemes
- L. Logie
- 2. Procedural representations
- 3. Semantic networks
- 4. Production systems
- S. Direct (analogical) representations
- 6. Semantic primitives 7. Frames and scripts

IV. Understanding Natural Language

- A. Overview
- R Machine translation
- C. Grammars 1. Formal grammars
- 2. Transformational grammars
- 3. Systemic grammar
- 4. Case grammars
- D. Parsing
- 1. Overview of parsing techniques 2. Augmented transition networks

- The General Syntactic Processor
 Text generation
 Natural language processing systems
 Early natural language systems
 Wilks's machine translation system
 LUNAR
 SHRDLU
 MARGIE
 SAM and PAM
 LIFER
-
- V. Understanding Spoken Language A. Overview B. Systems architecture C. The ARPA SUR projects 1. HEARSAY 2. HARPY 3. HWIM 4. The SRI/SDC speech systems

VOLUME II

- VI. Programming Languages for AI Research
- A. Overview
- B. LISP
 C. AI programming-language features
 1. Overview
 2. Data structures
- 3. Control structures 4. Pattern matching
- 5. Programming environment D. Dependencies and assumptions
- VII. Applications-oriented AI Research: Science
- A. Overview
- B. TEIRESIAS C. Applications in chemistry
- 1. Chemical analysis 2. The DENDRAL programs
- a. Heuristic DENDRAL b. CONGEN and its extensions
- c. Meta-DENDRAL
- 3. CRYSALIS 4. Applications in organic synthesis
- D. Other scientific applications 1. MACSYMA
- 2. The SRI Computer-based Consultant
- 3. PROSPECTOR
- 4. Artificial Intelligence in database management
- VIII. Applications-oriented AI Research: Medicine
 - A. Overview B. Medical systems L. MYCIN
- 2. CASNET 3. INTERNIST

- 4. Preseat Illness Program 5. Digitalis Therspy Advisor 6. IRIS 7. EXPERT
- IX. Applications-oriented AI Research: Education

5. Walts

6. Shirai

8. Kanade

2. Color

6. Texture

2. Motion

7. Markworth

1. Visual input

3. Preprocessing

5. Region analysis

1. Intrinsic images

3. Stereo vision

4. Range finders

E. Algorithms for vision

F. Devices and systems

I. Robetic vision

3. ACRONYM

A. Overview

B. Rote learning

L. Issues

2. Template matching

4. Relaxation algorithms

C. Learning by taking advice

D. Learning from examples

a. Version space

a. AQ11

+ HACKER

B. STRIPS and ABSTRIPS

C. Nonhierarchical planning

E. Refinement of skeletal plans

D. Hierarchical planners

d LEX

A. Overview

L NOAH

2. MOLGEN

2. Mostow's operationalizer

3. Learning single concepts

4. Learning multiple concepts

b. Meta-DENDRAL

and pattern-recognition sy

b. Data-driven knowledge-space oper

5. Learning to perform multiple-step tasks

a. Samuel's Checkers Player

b. Waterman's Poker Player

s. Grammatical inference

XV. Planning and Problem Solving

c. Generating and testing plausible hypotl

XIV. Learning and Inductive Inference

5. Shape-from methods

and recognition

1. Pyramids and quad trees

C. Early processing of visual data

4. Edge detection and line finding

D. Representation of scene characteristics

6. Three-dimensional shape description

3. Linguistic methods for computer vision

2. Organization and control of vision systems

2. Rote learning in Samuel's Checkers Planar

- A. Overview B. ICAI systems design C. Intelligent CAI systems 1. SCHOLAR 2. WHY 3. SOPHEE
 - 4. WEST 5. WUMPUS
 - 6. GUIDON
 - 7. BUGGY
 - 8. EXCHECK
 - D. Other applications of AI to education
- X. Automatic Programming
- A. Overview
- B. Methods of program specification
- C. Basie approaches
- D. Automatic programming systems 1. PSI and CHI
- 2. SAFE
- 3. The Programmer's Apprentice
- 4. PECOS
- 5. DEDALUS
- 6. Protosystem I 7. NLPO
- 7. NLPQ 8. LIERA

VOLUME III

- XI. Models of Cognition
- A. Overview B. General Problem Solver
- E. General Protoem S
- C. Opportunistic problem solving D. EPAM
- E. Semantic network models of memory 1. Quillian's semantic memory system

B. The resolution rule of inference

C. Nonresolution theorem proving

B. Blocks-world understanding

D. The Bayer-Moore Theorem Prover

2. HAM

XII. Automatic Deduction

E. Nonmonotonie logies.

F. Logic programming

3. ACT 4. MEMOD F. Belief systems

A. Overview

XIII. Vision

https://archive.org/details/handbookofartific02barr

A. Ourview

I. Roberts

2. Gupman

4. Huffman-Clowes

3. Falk

7. BUGGY 8. EXCHECK D. Other applications of AI to education B. Before applications of AI to education A. Overview B. Methods of program specification C. Basic approaches D. Automatic programming systems 1. PSI and CHI 2. SAFE 3. The Programmer's Apprentice 4. PECOS 5. DEDALUS 6. Protosystem I 7. NLPQ 8. LIBRA

VOLUME III

- XI. Models of Cognition
 - A. Overview

h

nce

ment

licine

- B. General Problem Solver
- C. Opportunistic problem solving
- D. EPAM
- E. Semantic network models of memory
 - 1. Quillian's semantic memory system 2. HAM
 - 3. ACT
 - 4. MEMOD
- F. Belief systems
- XII. Automatic Deduction
 - A. Overview
 - B. The resolution rule of inference
 - C. Nonresolution theorem proving
 - D. The Boyer-Moore Theorem Prover
 - E. Nonmonotonic logics
 - F. Logic programming

XIII. Vision

- A. Overview
- B. Blocks-world understanding
 - I. Roberts
 - 2. Guzman
 - 3. Falk
 - 4. Huffman-Clowes

a. F.	 Shape-from methods Three-dimensional shape description Chorock of Al Pyramids and quad trees Template matching Linguistic methods for computer vision Relaxation algorithms Devices and systems Robotic vision Organization and control of vision systems ACRONYM 	981	overview
αv.	Learning and Inductive Inference		
Α.	Overview		
B	Rote learning		
	I. Issues		
	2. Rote learning in Samuel's Checkers Player		
C.	Learning by taking advice		
	1. Issues		
D	2. Mustow's operationaliser		
D.	Learning from examples		
	2 Learning in control		
	and nattern-recognition systems		
	3. Learning single concents		
	a. Version space		
	b. Data-driven knowledge-space operators		
	c. Generating and testing plausible hypothes	es.	
	d. Schema instantiation		
	4. Learning multiple concepts		
	a. AQ11		
	b. Meta-DENDRAL		
	c. AM		
	5. Learning to perform multiple-step tasks		
	a. Samaer's Checkers Player		
	6. Waterman's Foker Player		
	d LEX		
	s. Grammatical inference		
	e onteriori incontre		
CV.	Planning and Problem Solving		
A	Overview		
В,	STRIPS and ABSTRIPS		
C.	Nonhierarchical planning		
D.	Hierarchical planners		
	1. NOAH		
	2 MOLCEN		

4. Range finders

E. Refinement of skeletal plans

https://archive.org/details/handbookofartific02barr



Donald Michie



Donald Michie was a British researcher in artificial intelligence. During World War II, Michie worked for the Government Code and Cypher School at Bletchley Park, contributing to the effort to solve "Tunny," a German teleprinter cipher. Wikipedia

Born: November 11, 1923, Yangon, Myanmar (Burma)

Died: July 7, 2007, England, United Kingdom

Children: Jonathan Michie, Susan Michie

Books: Donald Michie: Machine Intelligence, Biology and More, more

Education: Rugby School, Balliol College

Awards: IJCAI Award for Research Excellence

Menace (Michie 63)





Menace (Michie 1963)

Learns Tic-Tac-Toe **Hardware:**

287 Boxes

(1 for each state)

Pearls in 9 colors

(1 color per position)

Play principle:

Choose box corresponding to current state Choose pearl at random from box Play corresponding move

Learning algorithm:

Game lost -> retain all pearls used (*negative reword - reinforcement*)

Game won -> for each select pearl, add a pearl of the same color to box (*positive reward - reinforcement*)



BOXES (1968)

https://www.youtube.com/watch?v=qF2fFMrNUCQ

• basis of reinforcement learning

	PART ONE	GENERAL ISSUES IN MACHINE LEARNING	1
	Chapter 1	An Overview of Machine Learning Jaime G. Carbonell, Ryszard S. Michalski, and Tom M. Mitchell	3
STHBOLEC COMPUTATION	1.1 1.2 1.3 1.4 1.5	Introduction The Objectives of Machine Learning A Taxonomy of Machine Learning Research An Historical Sketch of Machine Learning A Brief Reader's Guide	3 3 7 14 16
Artificial Intelligence Approach An Artificial Intelligence Approach Edited by Ryazard S. Michalaki Jaime G. Carbonell Tom M. Mitchell	Chapter 2 2.1 2.2 2.3 2.4 2.5 2.6 2.7	Why Should Machines Learn? Herbert A. Simon Introduction Human Learning and Machine Learning What is Learning? Some Learning Programs Growth of Knowledge in Large Systems A Role for Learning Concluding Remarks	25 25 28 30 32 34 35
	PART TWO	LEARNING FROM EXAMPLES	39
	Chapter 3	A Comparative Review of Selected Methods for Learning from Examples Thomas G. Dietterich and Ryszard S. Michalski	41
	3.1 3.2	Introduction Comparative Review of Selected Methods	41 49



Why should machines learn ?

Herbert Simon (1916-2001)

Turing Award 1975, Nobel prize Economics 1978

The only partially satisfactory definition I've been able to find is that learning is any change in a system that allows it to perform better the second time on repetition of the same task or on another task drawn from the same population. The change should be more or less irreversible—not irreversible in the sense that you can't unlearn (although that sometimes is hard, especially unlearning bad habits) but irreversible in that the learning doesn't go away rapidly and autonomously. Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more efficiently and more effectively the next time.



"Learning is any change in a system that produces a more or less permanent change in its capacity for adapting to its environment."

Herbert Simon

- I. I would give a very high priority to research aimed at simulating, and thereby understanding, human learning. It may be objected that such research is not AI but cognitive psychology or cognitive science or something else. I don't really care what it is called; it is of the greatest importance that we deepen our understanding of human learning, and the Al community possesses a large share of the talent that can advance us toward this goal.
- 2. I would give a high priority, also, to basic research aimed at understanding why human learning is so slow and inefficient, and correspondingly, at examining the possibility that machine learning schemes can be devised that will avoid, for machines as well as people, some of the tediousness of learning.
- 3. I would give a high priority to research on the natural language interface between computer systems and human users. Again, it does not matter whether you call it research on learning or research on understanding. We do want systems, particularly in the knowledge engineering area, in which we don't have to know the internal language or representation in order to interact with them. This is especially true, as I have just argued, if the : systems are to be cumulative over many years.
- 4. I think there is an important place for research on programming from incomplete instructions (automatic programming), which is not unrelated to 12.1 the preceding item. Giving instructions to a skilled programmer is different from writing the program yourself-else why hire the programmer? It is a very important research question to ask whether we can get the computer to be the skilled programmer.

٠

5. My final priority is research on discovery programs-programs that discover new things. We may regard discovery itself as a form of learning, but in addition we will want to give a discovery system learning capabilities because we will want it to preserve and to be able to use all the new things it finds.

Chapter 4	A Theory and Methodology of Inductive Learning Ryszard S. Michalski	83
41	Introduction	83
4.2	Types of Inductive Learning	.87
4.3	Description Language	94
4.4	Problem Background Knowledge	96
4.5	Generalization Rules	103
4.6	The Star Methodology	112
4.7	An Example	116
4.8	Conclusion	123
4.A	Annotated Predicate Calculus (APC)	130
PART THREE	LEARNING IN PROBLEM-SOLVING AND PLANNING	135
Chapter 5	Learning by Analogy: Formulating and Generalizing Plans from Past Experience Jaime G. Carbonell	137
5.1	Introduction	137
5.2	Problem-Solving by Analogy	139
5.3	Evaluating the Analogical Reasoning Process	149
5.4	Learning Generalized Plans	151
5.5	Concluding Remark	159
Chapter 6	Learning by Experimentation: Acquiring and Refining Problem-Solving Heuristics Tom M. Mitchell, Paul E. Utgoff, and Ranan Banerji	163
6.1	Introduction	163
6.2	The Problem	164
6.3	Design of LEX	167
6.4	New Directions: Adding Knowledge to Augment	180
6.5	Summary	189
Chapter 7	Acquisition of Proof Skills in Geometry John R. Anderson	191
7.1	Introduction	191
7.2	A Model of the Skill Underlying Proof Generation	193
7.3	Learning	201
7.4	Knowledge Compilation	202

7.5	Summary of Geometry Learning	217
Chapter 8	Using Proofs and Refutations to Learn from Experience Frederick Hayes-Roth	221
8.1 8.2 8.3 8.4	Introduction The Learning Cycle Five Heuristics for Rectifying Refuted Theories Computational Problems and Implementation Techniques	221 222 225 234
8.5	Conclusions	238
PART FOUR	LEARNING FROM OBSERVATION AND	241
	DISCOVERT	
Chapter 9	The Role of Heuristics in Learning by Discovery: Three Case Studies Douglas B. Lenat	243
9.1	Motivation	243
9.2	Overview	245
9.3	Case Study 1: The AM Program; Heuristics Used to Develop New Knowledge	249
9.4	A Theory of Heuristics	263
9.5	Case Study 2: The Eurisko Program; Heuristics Used to Develop New Heuristics	276
9.6	Heuristics Used to Develop New Representations	282
9.7	Case Study 3: Biological Evolution; Heuristics Used to Generate Plausible Mutations	286
9.8	Conclusions	302
Chapter 10	Rediscovering Chemistry With the BACON System Pat Langley, Gary L. Bradshaw, and Herbert A. Simon	307
10.1	Introduction	307
10.2	An Overview of BACON.4	309
10.3	The Discoveries of BACON.4	312
10.4	Rediscovering Nineteenth Century Chemistry	319
10.5	Conclusions	326

Chapter 11	Learning From Observation: Conceptual Clustering Ryszard S. Michalski and Robert E. Stepp	331
11.1	Introduction	332
11.2	Conceptual Cohesiveness	333
11.3	Terminology and Basic Operations of the Algorithm	336
11.4	A Criterion of Clustering Quality	344
11.5	Method and Implementation	345
11.6	An Example of a Practical Problem: Constructing a Classification Hierarchy of Spanish Folk Songs	358
11.7	Summary and Some Suggested Extensions of the Method	360
PART FIVE	LEARNING FROM INSTRUCTION	365
Chapter 12	Machine Transformation of Advice into a Heuristic Search Procedure	367
10.1	Introduction	367
12.1	Kinds of Knowledge Lised	370
12.2	A Slightly Non-Standard Definition of Heuristic	374
12.0	Search	••••
12.4	Instantiating the HSM Schema for a Given Problem	378
12.5	Refining HSM by Moving Constraints Between	384
12.6	Evaluation of Generality	398
12.7	Conclusion	399
12.A	Index of Rules	403
Chapter 13	Learning by Being Told: Acquiring Knowledge for Information Management Norm Haas and Gary G. Hendrix	405
13.1	Overview	405
13.2	Technical Approach: Experiments with the KLAUS Concept	408
13.3	More Technical Details	413
13.4	Conclusions and Directions for Future Work	418
13.A	Training NANOKLAUS About Aircraft Carriers	422

331	Chapter 14	The Instructible Production System: A Retrospective Analysis Michael D. Rychener	429
	14.1	The Instructible Production System Project	430
332	14.2	Essential Functional Components of Instructible	436
333		Systems	
336	14.3	Survey of Approaches	443
	14.4	Discussion	453
344	DADT CIV		461
340	FART SIA	AFFLIED LEARNING STSTEWS	401
360	Chapter 15	Learning Efficient Classification Procedures and their Application to Chess End Games J. Ross Quinlan	463
	15.1	Introduction	463
365	15.2	The Inductive Inference Machinery	465
0.07	15.3	The Lost N-ply Experiments	470
367	15.4	Approximate Classification Rules	474
	15.5	Some Thoughts on Discovering Attributes	477
	15.6	Conclusion	481
007			
367	Chapter 16	Inferring Student Models for Intelligent	483
367 370 374	Chapter 16	Inferring Student Models for Intelligent Computer-Aided Instruction Derek H. Sleeman	483
367 370 374 378	Chapter 16	Inferring Student Models for Intelligent Computer-Aided Instruction Derek H. Sleeman Introduction	483 483
367 370 374 378	Chapter 16 16.1 16.2	Inferring Student Models for Intelligent Computer-Aided Instruction Derek H. Sleeman Introduction Generating a Complete and Non-redundant Set	483 483 488
367 370 374 378 384	Chapter 16 16.1 16.2	Inferring Student Models for Intelligent Computer-Aided Instruction Derek H. Sleeman Introduction Generating a Complete and Non-redundant Set of Models	483 483 488
367 370 374 378 384	Chapter 16 16.1 16.2 16.3	Inferring Student Models for Intelligent Computer-Aided Instruction Derek H. Sleeman Introduction Generating a Complete and Non-redundant Set of Models Processing Domain Knowledge	483 483 488 503
367 370 374 378 384 398	Chapter 16 16.1 16.2 16.3 16.4	Inferring Student Models for Intelligent Computer-Aided Instruction Derek H. Sleeman Introduction Generating a Complete and Non-redundant Set of Models Processing Domain Knowledge Summary	483 483 488 503 507
367 370 374 378 384 398 399 403	Chapter 16 16.1 16.2 16.3 16.4 16.A	Inferring Student Models for Intelligent Computer-Aided Instruction Derek H. Sleeman Introduction Generating a Complete and Non-redundant Set of Models Processing Domain Knowledge Summary An Example of the SELECTIVE Algorithm: LMS-I's Model Generation Algorithm	483 483 488 503 507 510
367 370 374 378 384 398 399 403 405	Chapter 16 16.1 16.2 16.3 16.4 16.4 16.A Comprehensiv	Inferring Student Models for Intelligent Computer-Aided Instruction Derek H. Sleeman Introduction Generating a Complete and Non-redundant Set of Models Processing Domain Knowledge Summary An Example of the SELECTIVE Algorithm: LMS-I's Model Generation Algorithm re Bibliography of Machine Learning Paul E. Utgoff and Bernard Nudel	483 483 488 503 507 510 511
367 370 374 378 384 398 399 403 405	Chapter 16 16.1 16.2 16.3 16.4 16.4 16.A Comprehensiv Glossary of Se	Inferring Student Models for Intelligent Computer-Aided Instruction Derek H. Sleeman Introduction Generating a Complete and Non-redundant Set of Models Processing Domain Knowledge Summary An Example of the SELECTIVE Algorithm: LMS-I's Model Generation Algorithm re Bibliography of Machine Learning Paul E. Utgoff and Bernard Nudel	483 483 488 503 507 510 511 551
367 370 374 378 384 398 399 403 405 405 408	Chapter 16 16.1 16.2 16.3 16.4 16.A 16.A Comprehensiv Glossary of Se About the Auth	Inferring Student Models for Intelligent Computer-Aided Instruction Derek H. Sleeman Introduction Generating a Complete and Non-redundant Set of Models Processing Domain Knowledge Summary An Example of the SELECTIVE Algorithm: LMS-I's Model Generation Algorithm re Bibliography of Machine Learning Paul E. Utgoff and Bernard Nudel Elected Terms In Machine Learning	483 483 488 503 507 510 511 551 557
367 370 374 378 384 398 399 403 405 405 408 413	Chapter 16 16.1 16.2 16.3 16.4 16.4 16.A Comprehensiv Glossary of Se About the Author	Inferring Student Models for Intelligent Computer-Aided Instruction Derek H. Sleeman Introduction Generating a Complete and Non-redundant Set of Models Processing Domain Knowledge Summary An Example of the SELECTIVE Algorithm: LMS-I's Model Generation Algorithm Performance State Paul E. Utgoff and Bernard Nudel Elected Terms In Machine Learning	483 483 488 503 507 510 511 551 557 563
367 370 374 378 384 398 399 403 405 405 405 408 413 418	Chapter 16 16.1 16.2 16.3 16.4 16.4 16.A Comprehensive Glossary of Se About the Author Author Index	Inferring Student Models for Intelligent Computer-Aided Instruction Derek H. Sleeman Introduction Generating a Complete and Non-redundant Set of Models Processing Domain Knowledge Summary An Example of the SELECTIVE Algorithm: LMS-I's Model Generation Algorithm re Bibliography of Machine Learning Paul E. Utgoff and Bernard Nudel Elected Terms In Machine Learning	483 483 503 507 510 511 551 557 563

1980 ... 1986

- First workshops on Machine Learning (first conference in 1993)
- Focus on AI and Cognitive Science paradigm
- Focus on SYMBOLIC Methods, on HUMAN like learning, on AUTOMATED DISCOVERY
- IJCAI 85 in LA had 3000 academic participants (10 000 with industry included?) These were the days of expert systems
- No role for SUBSYMBOLIC methods / NEURAL NETS
- NIPS would start in 1986, with the revival of Neural Networks (Parallel Distributed Processing / Connectionism Rumelhart and McClelland)
- <u>https://www.youtube.com/watch?v=iIP4aPDTBPE</u> (1989)

From the Dartmouth 1956 proposal

The following are some aspects of the artificial intelligence problem:

- 1. **Automatic Computers** : If a machine can do a job, then an automatic calculator can be programmed to simulate the machine. The speeds and memory capacities of present computers may be insufficient to simulate many of the higher functions of the human brain, but the major obstacle is not lack of machine capacity, but our inability to write programs taking full advantage of what we have.
- 2. How Can a Computer be Programmed to Use a Language : It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning and rules of conjecture. From this point of view, forming a generalization consists of admitting a new word and some rules whereby sentences containing it imply and are implied by others. This idea has never been very precisely formulated nor have examples been worked out.
- 3. **Neuron Nets :** How can a set of (hypothetical) neurons be arranged so as to form concepts. Considerable theoretical and experimental work has been done on this problem by Uttley, Rashevsky and his group, Farley and Clark, Pitts and McCulloch, Minsky, Rochester and Holland, and others. Partial results have been obtained but the problem needs more theoretical work.

4. **Theory of the Size of a Calculation :** If we are given a well-defined problem (one for which it is possible to test mechanically whether or not a proposed answer is a valid answer) one way of solving it is to try all possible answers in order. This method is inefficient, and to exclude it one must have some criterion for efficiency of calculation. Some consideration will show that to get a measure of the efficiency of a calculation it is necessary to have on hand a method of measuring the complexity of calculating devices which in turn can be done if one has a theory of the complexity of functions. Some partial results on this problem have been obtained by Shannon, and also by McCarthy.

5. **Self-Improvement** Probably a truly intelligent machine will carry out activities which may best be described as self-improvement. Some schemes for doing this have been proposed and are worth further study. It seems likely that this question can be studied abstractly as well.

6. **Abstractions** A number of types of ``abstraction" can be distinctly defined and several others less distinctly. A direct attempt to classify these and to describe machine methods of forming abstractions from sensory and other data would seem worthwhile.

7. **Randomness and Creativity** A fairly attractive and yet clearly incomplete conjecture is that the difference between creative thinking and unimaginative competent thinking lies in the injection of a some randomness. The randomness must be guided by intuition to be efficient. In other words, the educated guess or the hunch include controlled randomness in otherwise orderly thinking.

1986

- The Machine Learning Journal was founded by Pat Langley, Ryszard Michalski, Jaime Carbonnell and Tom M. Mitchell
 - find an own venue for ML research ...
 - same focus / bias initially (cf. Langley, MLJ 2011)
- The 1st European Working Session on ML was organised in Orsay by Yves Kodratoff





Ryszard S. Michalski was a Polish-American computer scientist. Michalski was Professor at George Mason University and a pioneer in the field of machine learning. Wikipedia

Born: May 7, 1937, Kalush, Ukraine

Died: September 20, 2007, Fairfax, Virginia, United States

Fields: Machine learning, Artificial intelligence

Video 0;58-3:08 + 22:25-25:43 ? + 44:05-45

1989

- The 1st KDD workshop was organised at IJCAI 1989 in Detroit, attended by 67 participants (among which most of the key players in ML and KDD ...)
- Panel with Ross Quinlan, Pat Langley, and Larry Kerschberg
- Donald Michie predicts that ``The next area that is going to explode is the use of machine learning tools as a component of large scale data analysis'' (AI Week, March 15, 1990)
- First KDD conference 1995

Call for Participation: IJCAI-89 Workshop on Knowledge Discovery in Databases Sunday, August 20 (tentative), Detroit MI, USA

The growth in the amount of available databases far outstrips the growth of corresponding knowledge. This creates both a need and an opportunity for extracting knowledge from databases. Many recent results have been reported on extracting different kinds of knowledge from databases, including diagnostic rules, drug side effects, classes of stars, rules for expert systems, and rules for semantic query optimization.

Knowledge discovery in databases poses many interesting problems, especially when databases are large. Such databases are usually accompanied by substantial domain knowledge which can significantly facilitate discovery. Access to large databases is expensive - hence the need for sampling and other statistical methods. Finally, knowledge discovery in databases can benefit from many available tools and techniques from several different fields including expert systems, machine learning, intelligent databases, knowledge acquisition, and statistics.

Topics of interest include:

- o Discovery and use of approximate rules
- o Knowledge-based discovery methods
- o Integration of knowledge-based and statistical methods
- o Efficient heuristic algorithms for discovery
- o Automatic knowledge acquisition
- o Construction of expert systems from data
- o Discovery in medical and scientific data
- o Bias for human understandability of discovered knowledge
- o Learning query optimization rules and integrity constraints
- o Knowledge discovery as a threat to database security and privacy

"Donald Michie"

12:20 / 21:20 / 31:54

Benelearn 1991

- 1st Benelearn in Leuven, 70 participants, 10 talks, lunches paid by FWO, Invited talks by Yves Kodratoff and Katharina Morik
- 1992, Amsterdam (Van Someren)
- 1993, Brussels (Van de Velde)
- 1994, Rotterdam (Bioch) 27 presentations !
- 1995, Brussels (ULB)
- 1996, Maastricht
- 1997, Tilburg
- 1998, Wageningen ...

Learning and Knowledge

- Explicit goal to learn new "knowledge", focus on results that are "understandable"
- Explicit goal to reason with that knowledge (eg. in problem solving)
- Explicit goal to learn rich representations, to learn for use in expert systems
- Cf. e.g. Dietterich, MLJ 86, Michalski's trains etc.

Machine learning

- Was initially broadening its scope from purely symbolic and knowledge based to
 - probabilistic methods
 - reinforcement learning
 - case-based reasoning and instance based learning
 - problem solving ...
 - was getting a diverse and open-minded field !



cf Langley, MLJ 2011

ML as an experimental science

- The title of an MLJ paper by Kibler and Langley in 1988; see also Langley MLJ 2011
- Presented also as keynote as EWSL 88
- Point of view that ML systems should not be just systems that do something, but should be evaluated according to scientific principles, through setting up experiments in a systematic way
- UCI Database
- Side-effect : focus on tasks that are easy to evaluate, on particular datasets, on classification and regression ...
- Another side-effect: you have to beat the competition ...



cf Langley, MLJ 2011

The Experimental Study of Machine Learning

Pat Langley

(LANGLEY@PTOLEMY.ARC.NASA.GOV)

AI Research Branch, Mail Stop 244–17, NASA Ames Research Center, Moffett Field, CA 94035 USA

Dennis Kibler

(KIBLER@ICS.UCI.EDU)

Department of Information & Computer Science, University of California, Irvine, CA 92717 USA

icarining ocnaviors.

As normally defined, an *experiment* involves systematically varying one or more *independent* variables and examining their effect on some *dependent* variables. Thus, a machine learning experiment requires more than a single observation of a system's behavior; it requires a number of observations made under different conditions. In each case, one must measure some aspect of the system's behavior for comparison across the different conditions.

6 phases :

- 1. Formulating Hypotheses
- 2. Design experiments and select Samples
- 3. Running experiments and compile results
- 4. Test hypotheses
- 5. Explain unexpected results
- 6. Report

Introduction of SVMs

- Around 1992-95 by Vapnik, Cortes et al.
- Enormous boost in performance
- Principled theory, interesting mathematics coming from a new community (physics, optimisation...)
- But also had a profound influence on the nature of machine learning
- Side-effect shift of focus of ML, towards optimisation, math and Linear Algebra …
- Side-effect change of the field ...

ICML 2005



ICML 2005



Observations

- The social aspects of science (of ML?)
- Fields evolve, have biases, communities are dynamic, split, merge ...
- Quite important to retain identity, to remain broad enough, yet coherent enough, ...

• cf. The structure of scientific revolutions, Thomas Kuhn

Evolution ...





ML today

- Enormous progress made, impressive applications, used in many other fields as the enabling technology
- A healthy field
- Attracting loads of attention
- Unreasonable expectations ? a bit like AI ?
- Big data / data science ... splitting off ?

ML today

- Diversity could be better ? are not we converging too fast ? More exploration would be useful ?
 - in terms of methodology
 - in terms of tasks
- Let's play more with the problem set up?
- Links to AI, to human learning, to reasoning ?

What is next?

- Get AI more into the picture ...
- Machine learning : an AI approach ?

Can we automate Data Science / Machine Learning ?

• The robot scientist (Ross King et al. Nature 2004)



- Can we apply that idea to data science/machine learning itself?
- One possible solution to the lack of data scientists today





Advanced ERC Grant

Synthesising Inductive Data Models

Data Model

Inductive Model

- The synthesis system "learns the learning task". It identifies the right learning tasks and learns appropriate IMs
- 2. The system may need to restructure the data set before IM synthesis can start
- A unifying IDM language for a set of core patterns and models will be developed

018	- -	fx 12573		
4	В	c	D	
3	Time	Ins. sendent	Dependent	
4	Period	Va. ble	Variable	
5	1	4,8	6,809	
6	2	4,703	6,465	
7	3	4,748	6,569	
8	4	5,844	8,266	
9	5	5,192	7,257	
10	6	5,086	7,064	
11	7	5,511	7,784	
12	8	6,107	724	
13	9	5,052	992	
14	10	4,985	6.#22	
15	11	5,576	7.94	
16	12	6,647	9,650	
17	13	7,011	10,221	
18	14	8,452	12,573	
19	15			
20	16			
4 10	Linear Regr	ession +		

KU LEUVEN

Thanks

BTW: we are hiring PhD students and post-docs !