Artificial general intelligence

Symbolic, connectionist, hybrid

Antal Péter

ComBine Lab

Artificial Intelligence group

Department of Measurement and Information Systems

Agenda

- Homeworks
- Cognitive science
  - Cognitive architectures
  - Criticisms (symbol???)
- The connectionist approach
  - Connectionism/artificial neural networks
    - ANN „winter” 1.0
  - Parallel distributed processing
    - ANN „winter” 2.0
  - Deep learning
  - Neuroinspired-AI
- **Next: the neurobiological substrate**

Requirements

- Grading:
  - Two homeworks 50-50%.
    - Review, essay, programming..
  - MI Almanach
    - http://project.mit.bme.hu/mi_almanach/
    - http://project.mit.bme.hu/mi_almanach/books/kieg/aima/ch01s01
  - AGI topic lists
    - Suggestions are welcome
      - Email
        - Subject: [AGI:hw] NAME NEPTUNCODE keyword(s)

Introduction

AGI podcasts

- [?] Josh Tenenbaum Computational Cognitive Science
- 47. Peter Norvig. Artificial Intelligence: A Modern Approach

- **48. Gary Marcus. Hybrid of Deep Learning and Symbolic AI**
  - Gary Marcus is a professor emeritus at NYU, founder of Robust.AI and Geometric Intelligence, the latter is a machine learning company acquired by Uber in 2016. He is the author of several books on natural and artificial intelligence, including his new book Rebooting AI: Building Machines We Can Trust. Gary has been a critical voice highlighting the limits of deep learning and discussing the challenges before the AI community that must be solved in order to achieve artificial general intelligence.

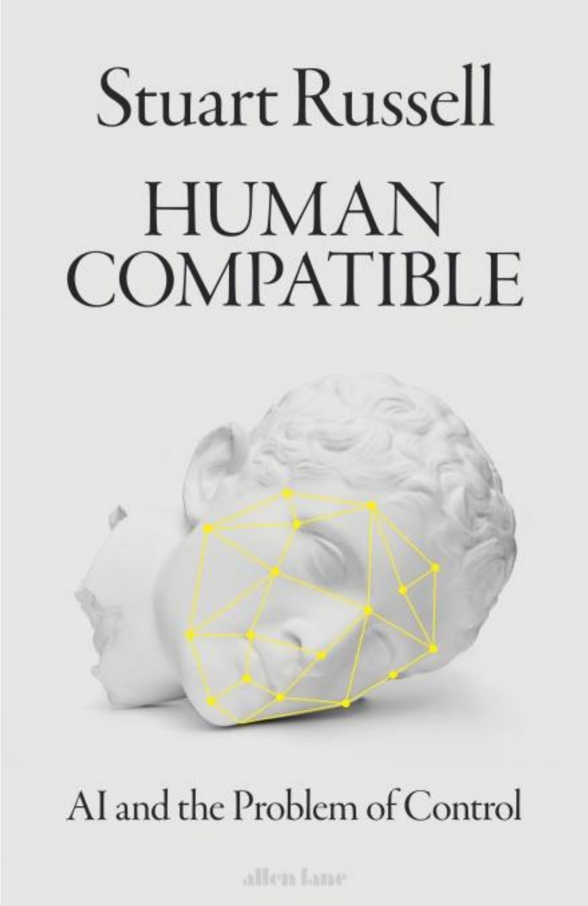
Marcus, G., 2018. Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*.

Davis, Ernest, and Gary Marcus. "Commonsense reasoning and commonsense knowledge in artificial intelligence." *Commun. ACM* 58.9 (2015): 92-103.

Marcus, G., 2017. Am I human?. *Scientific American*, *316*(3), pp.58-63.

Marcus, G., Rossi, F. and Veloso, M., 2016. Beyond the turing test. *Ai Magazine*, *37*(1), pp.3-4.

Human-compatible AI



https://spectrum.ieee.org/computing/software/many-experts-say-we-shouldnt-worry-about-superintelligent-ai-theyre-wrong

- The „gorilla", „paper clip",… problems
- The value-alignment problem
- Counterarguments
  - Switch off
  - Too far
  - Human(-level!?) intelligence is multifaceted („human"?)
  - Not possible
    - Artificial intelligence and life in 2030. *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*
  - Standard rules for safety and goal specification
  - Expected utility is +++..

Bostrom, N., 2016. Superintelligence: Paths, Dangers, Strategies, Reprint ed.
Davis, E., 2015. Ethical guidelines for a superintelligence. *Artificial Intelligence*, *220*, pp.121-124.

Computer models of mind/cognition

Books

- Boden, M., 1980. Artificial intelligence and natural man.
- Hofstadter, Douglas R. Gödel, Escher, Bach. Penguin Books, 1980.
- Boden, Margaret A. Computer models of mind. Cambridge University Press, 1988.
- Mérő, László. *Észjárások: a racionális gondolkodás korlátai és a mesterséges intelligencia*. Akadémiai Kiadó, 1989.
- Tibor, Vámos. Computer epistemology. Vol. 25. World Scientific, 1991.
- **Pléh, Csaba. *A megismeréstudomány alapjai: az embertől a gépig és vissza*. Typotex, 2013.**

Pléh Csaba: *A megismeréstudomány alapjai*

1. A MEGISMERÉSTUDOMÁNY (KOGNITÍV TUDOMÁNY) HELYE
2. A KOGNITÍV KUTATÁS KLASSZIKUS SZEMLÉLETE
3. A SZIMBÓLUMFELDOLGOZÓ GONDOLKODÁS NÉHÁNY RÉSZLETE
4. A SZIMBÓLUMFELDOLGOZÓ FELFOGÁS INHERENS BÍRÁLATA
5. A REPREZENTÁCIÓ FOGALMA A KOGNITÍV TUDOMÁNYBAN
6. A REPREZENTÁCIÓ „SZIGORÚBB" FOGALMA
7. GONDOLKODNAK-E A GÉPEK?
8. A KONNEKCIONISTA ALTERNATÍVA
9. A MODULOK PARLAMENTJE
10. BIOLÓGIAI ALTERNATÍVÁK
11. A TUDAT KÉRDÉSE A KOGNITÍV TUDOMÁNYBAN

A(G)I as “symbol manipulation”

- The Logic Theorist, 1955
  - ➔ see lectures on logic
- The Dartmouth conference ("birth of AI”, 1956)
- List processing (Information Processing Language, IPL)
- Means-ends analysis ("reasoning as search")
  - ➔ see lectures on planning
- The General Problem Solver
- Heuristics to limit the search space
  - ➔ see lecture on informed search
- The physical symbol systems hypothesis
  - intelligent behavior can be reduced to/emulated by symbol manipulation
- The unified theory of cognition (1990, cognitive architectures: Soar, ACT-R)

- Newel&Simon: Computer science as empirical inquiry: symbols and search, 1975

Constraints on mind

1. Behave as an (almost) arbitrary function of the environment (universality).
2. Operate in real time.
3. Exhibit rational, i.e., effective adaptive behavior.
4. Use vast amounts of knowledge about the environment.
5. Behave robustly in the face of error, the unexpected, and the unknown.
6. Use symbols (and abstractions).
7. Use (natural) language.
8. Exhibit self-awareness and a sense of self.
9. Learn from its environment.
10. Acquire its capabilities through development.
11. Arise through evolution.
12. Be realizable within the brain as a physical system.
13. Be realizable as a physical system.

Newell, A., 1980. Physical symbol systems. *Cognitive science*, *4*(2), pp.135-183.

A physical symbol system

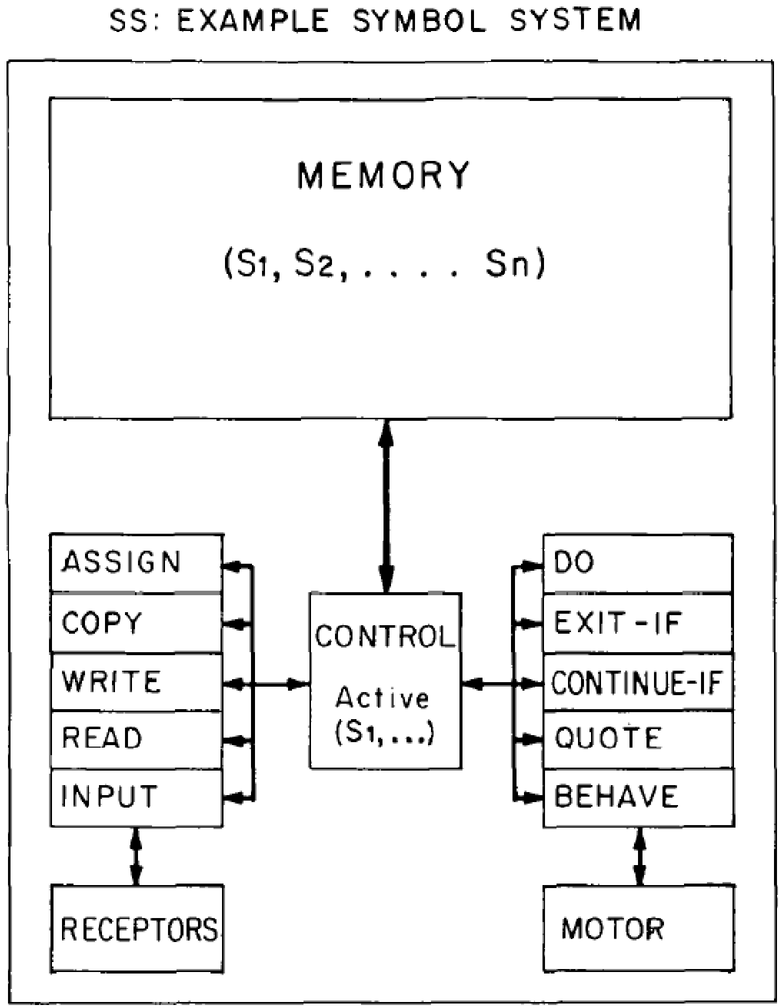


Figure 2. Structure of SS, a Paradigmatic Symbol System.

Newell, A., 1980. Physical symbol systems. *Cognitive science*, *4*(2), pp.135-183.

The physical symbol system hypothesis

Physical Symbol System Hypothesis: The necessary and sufficient condition for a physical system to exhibit general intelligent action is that it be a physical symbol system.

Necessary means that any physical system that exhibits general intelligence will be an instance of a physical symbol system.

Sufficient means that any physical symbol system can be organized further to exhibit general intelligent action.

General intelligent action means the same scope of intelligence seen in human action: that in real situations behavior appropriate to the ends of the system and adaptive to the demands of the environment can occur, within some physical limits.

Newell, A., 1980. Physical symbol systems. *Cognitive science*, *4*(2), pp.135-183.

Architectures: cognition

- SOAR
  - Newell, A., 1980. Physical symbol systems. *Cognitive science*, *4*(2), pp.135-183.
  - Laird, J.E., Newell, A. and Rosenbloom, P.S., 1987. Soar: An architecture for general intelligence. *Artificial intelligence*, *33*(1), pp.1-64.
  - Rosenbloom, P.S., Laird, J. and Newell, A. eds., 1993. The SOAR papers: Research on integrated intelligence.
- ACT-R (Adaptive Character of Thought, ACT-R)
  - Anderson, J.R. and Bellezza, F.S., 1993. Rules of the mind. Hillsdale, NJ: L.
  - Anderson, J.R., 2014. *Rules of the mind*. Psychology Press.
  - Anderson, J.R., 1996. ACT: A simple theory of complex cognition. *American psychologist*, *51*(4), p.355.
  - Lebiere, C. and Anderson, J.R., 1993, June. A connectionist implementation of the ACT-R production system. In *Proceedings of the fifteenth annual conference of the Cognitive Science Society* (pp. 635-640).
  - Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C. and Qin, Y., 2004. An integrated theory of the mind. *Psychological review*, *111*(4), p.1036.
  - Anderson, J.R., 2009. *How can the human mind occur in the physical universe?* (Vol. 3). Oxford University Press.

http://act-r.psy.cmu.edu/

https://en.wikipedia.org/wiki/ACT-R

ACT-R

- a cognitive architecture
- a theory for simulating and understanding human cognition

ACT-R Theory
- Architecture

Language Processing
- Analogy and Metaphor
- Language Learning
- Lexical and General Language Processing
- Parsing
- Sentence Memory

Perception and Attention
- Attention
- Driving and Flying Behavior
- Eye Movements
- Graphical User Interfaces
- Multi-Tasking
- Psychophysical Judgements
- Situational Awareness and Embedded Cognition
- Stroop
- Subitizing
- Task Switching
- Time Perception
- Visual Search

Problem Solving and Decision Making
- Choice and Strategy Selection
- Dynamic Systems
- Errors
- Game Playing
- Insight and Scientific Discovery
- Mathematical Problem Solving
- Programming
- Reasoning
- Spatial Reasoning and Navigation
- Tower of Hanoi
- Use and Design of Artifacts

Learning and Memory
- Category Learning
- Causal Learning
- Cognitive Arithmetic
- Declarative Memory
- Implicit Learning
- Interference
- Learning by Exploration and Demonstration
- List Memory
- Practice and Retention
- Reinforcement Learning
- Representation
- Skill Acquisition
- Updating Memory and Prospective Memory
- Working Memory

Other
- Cognitive Development
- Cognitive Workload
- Communication, Negotiation, and Group Decision Making
- Comparative (Architectures)
- Comparative (Inter-species)
- Computer Generated Forces, Video Games, and Agents
- fMRI
- Individual Differences
- Information Search
- Instructional Materials
- Intelligent Tutoring Systems
- Motivation, Emotion, Cognitive Moderators, & Performance
- Neuropsychology
- Tools
- Unrelated to ACT-R
- User Modeling

Uncategorized

http://act-r.psy.cmu.edu/publication/

Architectures: vision

- David Marr’s Tri-Level Hypothesis:
  - computational level:
    - what does the system do (e.g.: what problems does it solve or overcome) and similarly, why does it do these things
  - algorithmic level (sometimes representational level):
    - how does the system do what it does, specifically, what representations does it use and what processes does it employ to build and manipulate the representations
  - implementational/physical level:
    - how is the system physically realised (in the case of biological vision, what neural structures and neuronal activities implement the visual system)

Architectures: maps

- **Self-organizing map (SOM)**
  - https://en.wikipedia.org/wiki/Self-organizing_map
- Thousand Brains Theory of Intelligence
  - https://numenta.com/blog/2019/01/16/the-thousand-brains-theory-of-intelligence/
  - https://www.youtube.com/watch?v=-EVqrDlAqYo
  - Hawkins, J., Lewis, M., Klukas, M., Purdy, S. and Ahmad, S., 2018. A framework for intelligence and cortical function based on grid cells in the neocortex. *Frontiers in neural circuits*, *12*, p.121.
- Adaptive Resonance Theory (ART)
  - Grossberg, S., 2006. Adaptive resonance theory. *Encyclopedia of cognitive science*.

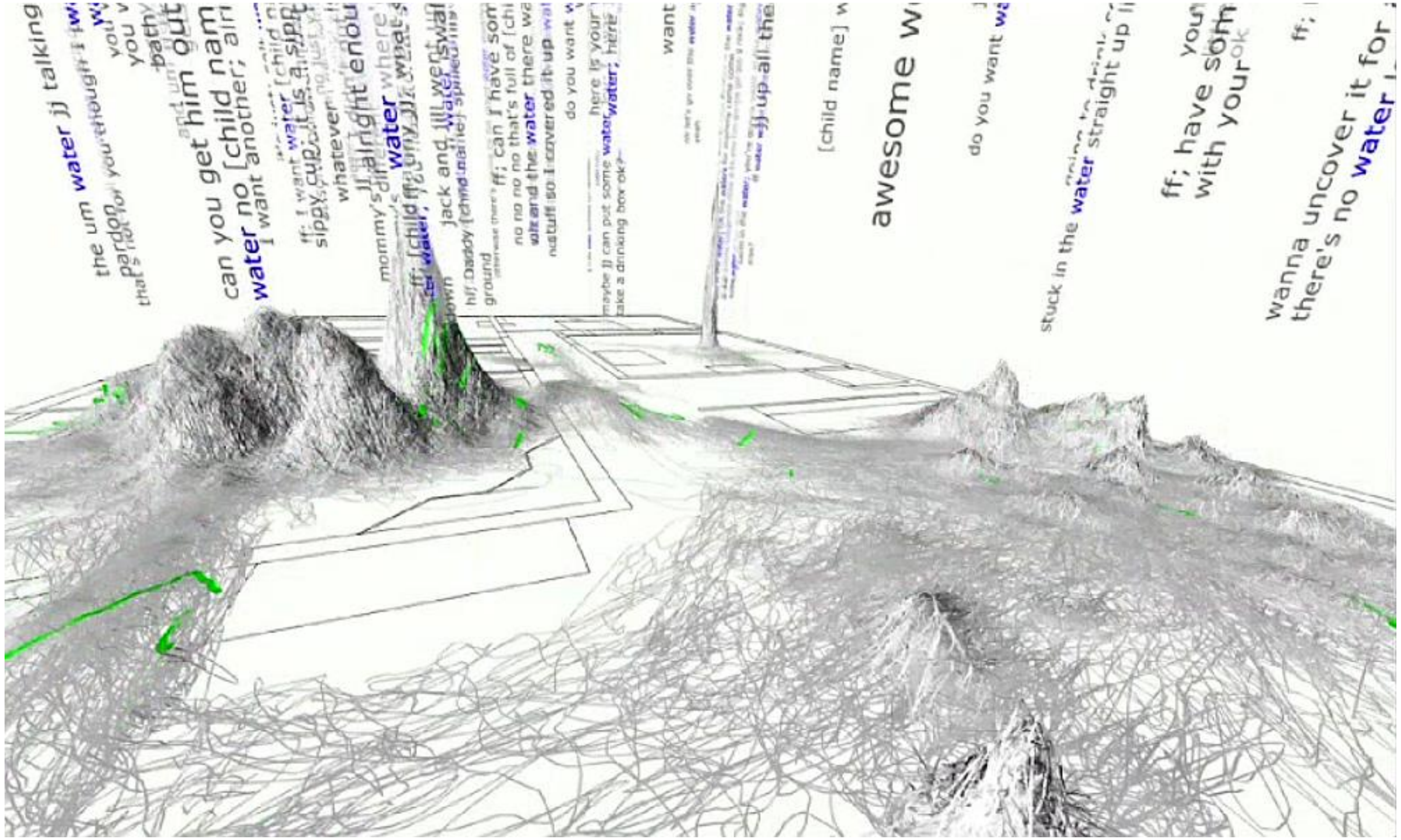
Architectures: language (of thought)

- Famous experiments about an inherent (~oldest) language
- N. Chomsky:
  - Universal grammar
  - Inherent language theory
- J. Fodor
  - Innate language module
  - Fodor, J.A., 1983. The modularity of mind. MIT press.
  - **Fodor, J.A. and Pylyshyn, Z.W., 1988. Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1-2), pp.3-71.**
- S. Pinker
  - innate capacity for language
  - Pinker, S., 2003. The language instinct: How the mind creates language. Penguin UK.

Language of thought

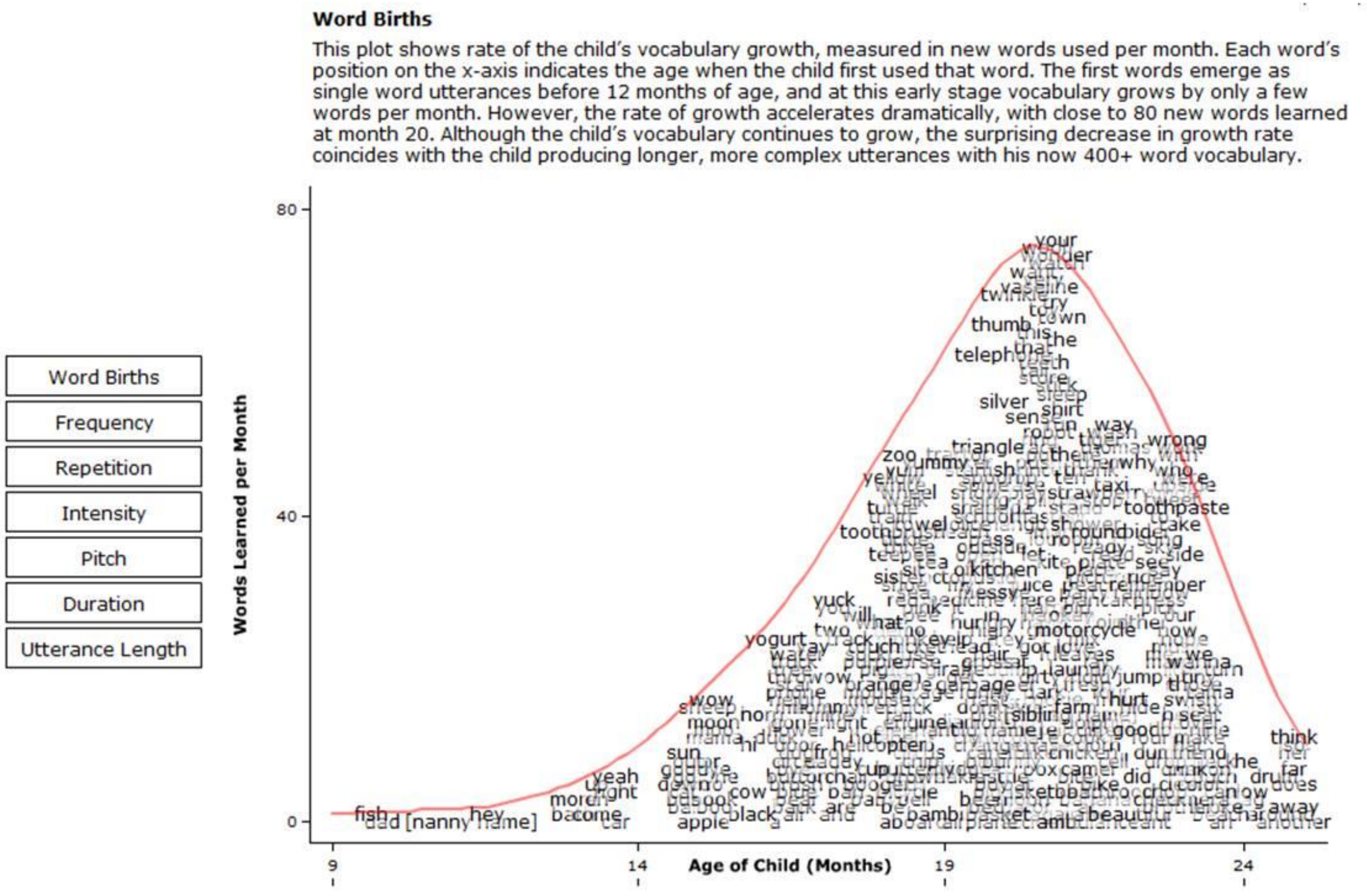
- Properties
  - Combinatorial syntax and semantics for mental representations
  - Productivity
  - Systematicity
  - Compositionality
  - Inferential coherence
- Intensional logic
  - Possible worlds semantics

Birth of a word



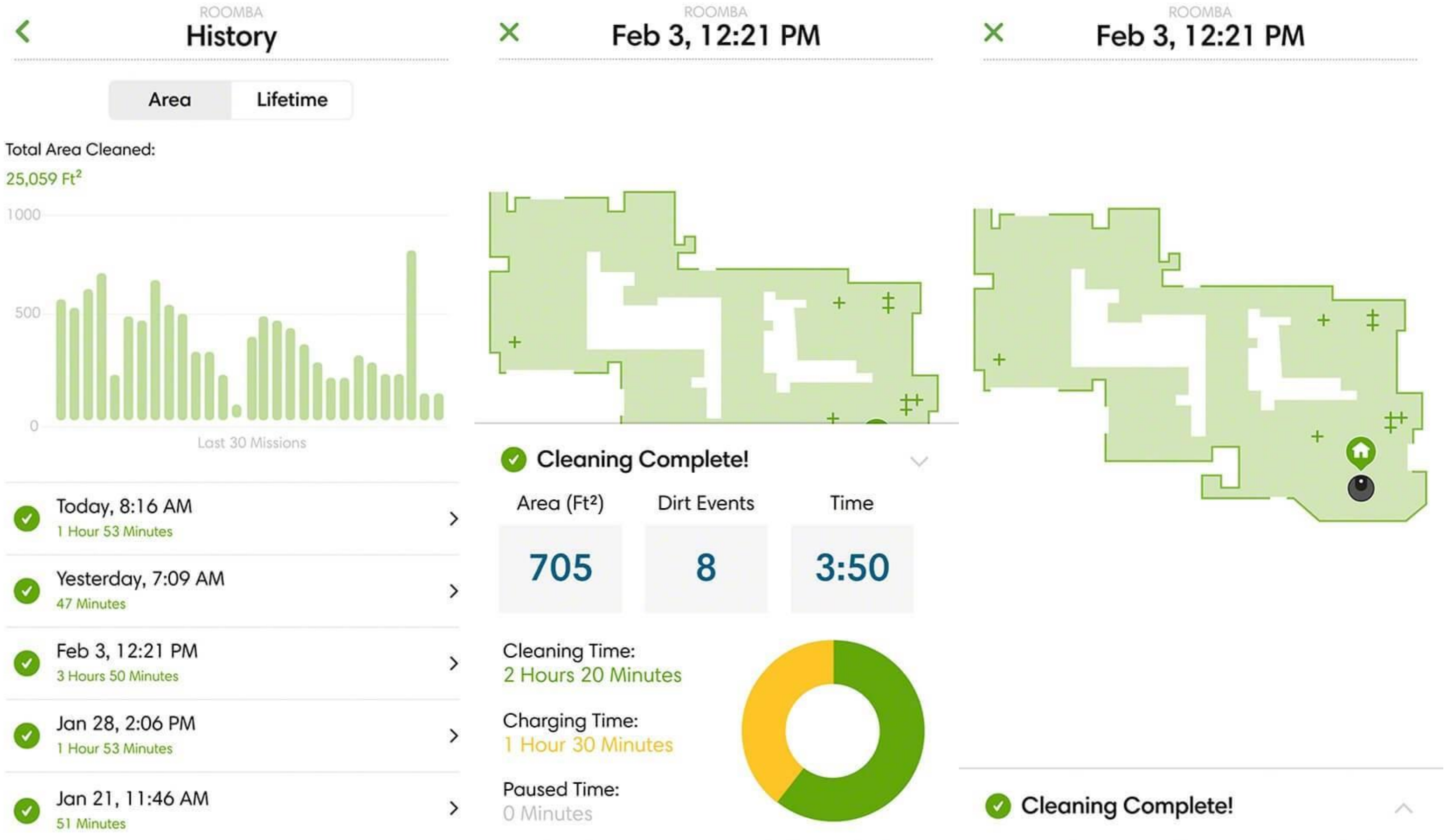
Roy, B.C., Frank, M.C., DeCamp, P., Miller, M. and Roy, D., 2015. Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, *112*(41), pp.12663-12668.

Learning rate of words



Roy, B.C., Frank, M.C., DeCamp, P., Miller, M. and Roy, D., 2015. Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, *112*(41), pp.12663-12668.

Intelligence without (symbolic) representation



Brooks, R.A., 1991. Intelligence without representation. *Artificial intelligence*, *47*(1-3), pp.139-159.

Beyond symbolic cognition (After GOFAI*)

*: Good Old Fashioned AI

Challenges for symbolic systems

- **Trivialities: sensation, perception, motoric/sensory-motoric capabilities**
- Symbols (as atomic concepts) + UTM (as centralized, sequential computation)
  ➔artificial neural networks, connectionism, parallel distributed computing
- Non-symbolic heuristics
  - Connectionist approaches to heuristics
- Uncertainties
  - Probabilistic Graphical Models (PGMs)
- Utilities(~values)
  - Utility theory, Decision theory
- Causality
  - Causality research

Beyond symbolic cognition: non-symbolic heuristics

Nature of expertise
(in rule-based production systems)

- Complex symbols (schemata, gestalt, patterns)
  - Flexible, multi-aspect, „grounded” concepts (~symbols)
    - ➔Sub-symbolic learning
- Complex rules
  - Meta-learning
- **Efficient inference: heuristics** (sub-thinking the right thing ;-)
  - [dictionary]*“A rule of thumb, simplification, or educated guess that reduces or limits the search for solutions in domains that are difficult and poorly understood.”*
  - *Prioritization of rules: timing, scoring,..*

Reminder: main properties of uninformed search

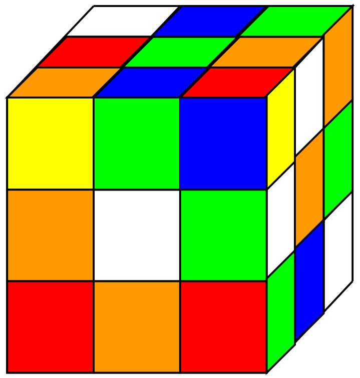
Criterion	Breadth-First	Uniform-cost	Depth-First	Depth-limited	Iterative deepening	Bidirectional search
Complete?	YES*	YES*	NO	YES, if $l \geq d$	YES	YES*
Time	b^{d+1}	$b^{C^*/e}$	b^m	b^l	b^d	$b^{d/2}$
Space	b^{d+1}	$b^{C^*/e}$	bm	bl	bd	$b^{d/2}$
Optimal?	YES*	YES*	NO	NO	YES	YES

A heuristic function

- [dictionary]*"A rule of thumb, simplification, or educated guess that reduces or limits the search for solutions in domains that are difficult and poorly understood."*
  - $h(n)$ = estimated cost of the cheapest path from node *n* to goal node.
  - If *n* is goal then $h(n)=0$

➔ for definition, derivation, effect, etc., see Appendix

Rubik's Cube



- The cardinality: 10^{19}
- Any position can be solved in 20 or fewer moves (where a half-twist is counted as a single move)
- average branching factor is ~13.3
- Invented in 1974 by Ernő Rubik.
- Rubik's cube current world records
  - http://www.youtube.com/watch?v=oC0B4b4J9Ys
- How can we guide the search process???

Agostinelli, F., McAleer, S., Shmakov, A. and Baldi, P., 2019. Solving the Rubik's cube with deep reinforcement learning and search. *Nature Machine Intelligence*, *1*(8), pp.356-363.

Solving the Rubik’s cube with deep reinforcement learning and search

The Rubik’s cube is a prototypical combinatorial puzzle that has a large state space with a single goal state. The goal state is unlikely to be accessed using sequences of randomly generated moves, posing unique challenges for machine learning. We solve the Rubik’s cube with DeepCubeA, a deep reinforcement learning approach that learns how to solve increasingly difficult states in reverse from the goal state without any specific domain knowledge. DeepCubeA solves 100% of all test configurations, finding a shortest path to the goal state 60.3% of the time. DeepCubeA generalizes to other combinatorial puzzles and is able to solve the 15 puzzle, 24 puzzle, 35 puzzle, 48 puzzle, Lights Out and Sokoban, finding a shortest path in the majority of verifiable cases.

Agostinelli, F., McAleer, S., Shmakov, A. and Baldi, P., 2019. Solving the Rubik’s cube with deep reinforcement learning and search. *Nature Machine Intelligence*, *1*(8), pp.356-363.

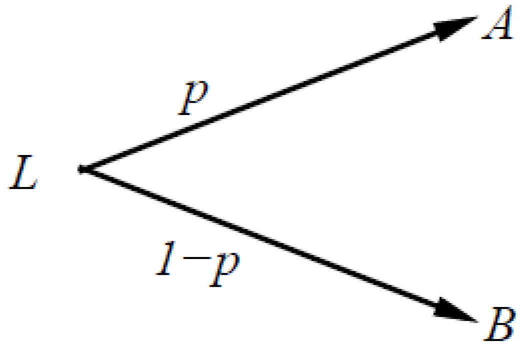
Long-term planning, meta-heuristics

- Dota 2
  - https://openai.com/blog/dota-2/
- Starcraft
  - Vinyals, O., Ewalds, T., Bartunov, S., Georgiev, P., Vezhnevets, A.S., Yeo, M., Makhzani, A., Küttler, H., Agapiou, J., Schrittwieser, J. and Quan, J., 2017. Starcraft ii: A new challenge for reinforcement learning. *arXiv preprint arXiv:1708.04782*.
- Hierarchical planning
  - Wu, B., 2019, July. Hierarchical macro strategy model for moba game ai. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, pp. 1206-1213).

Beyond symbolic cognition: utilities

Preferences

An agent chooses among prizes (A, B, etc.) and lotteries, i.e., situations with uncertain prizes



Lottery $L = [p, A;\ (1-p), B]$

Notation:

$A \succ B$	A preferred to B
$A \sim B$	indifference between A and B
$A \succsim B$	B not preferred to A

Rational preferences

Idea: preferences of a rational agent must obey constraints.
Rational preferences $\Rightarrow$
behavior describable as maximization of expected utility

Constraints:

Orderability
$$(A \succ B) \vee (B \succ A) \vee (A \sim B)$$
Transitivity
$$(A \succ B) \wedge (B \succ C) \Rightarrow (A \succ C)$$
Continuity
$$A \succ B \succ C \Rightarrow \exists\, p \ \ [p, A;\ 1-p, C] \sim B$$
Substitutability
$$A \sim B \Rightarrow [p, A;\ 1-p, C] \sim [p, B; 1-p, C]$$
Monotonicity
$$A \succ B \Rightarrow (p \geq q \Leftrightarrow [p, A;\ 1-p, B] \succsim [q, A;\ 1-q, B])$$

An irrational preference

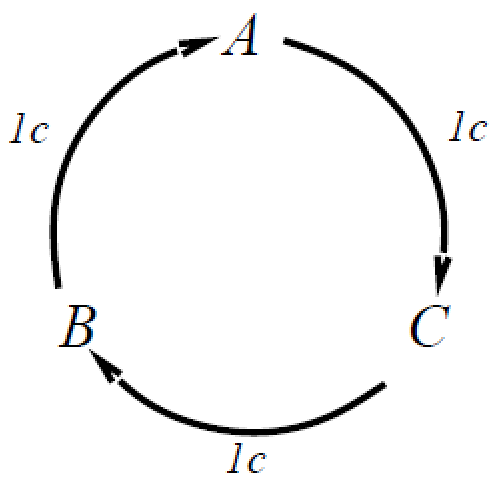
Violating the constraints leads to self-evident irrationality

For example: an agent with intransitive preferences can be induced to give away all its money

If $B \succ C$, then an agent who has C would pay (say) 1 cent to get B

If $A \succ B$, then an agent who has B would pay (say) 1 cent to get A

If $C \succ A$, then an agent who has A would pay (say) 1 cent to get C



Maximizing expected utility

Theorem (Ramsey, 1931; von Neumann and Morgenstern, 1944):
Given preferences satisfying the constraints
there exists a real-valued function U such that

$$U(A) \geq U(B) \quad \Leftrightarrow \quad A \succsim B$$
$$U([p_1, S_1; \ \ldots \ ; \ p_n, S_n]) = \Sigma_i \, p_i U(S_i)$$

MEU principle:

Choose the action that maximizes expected utility

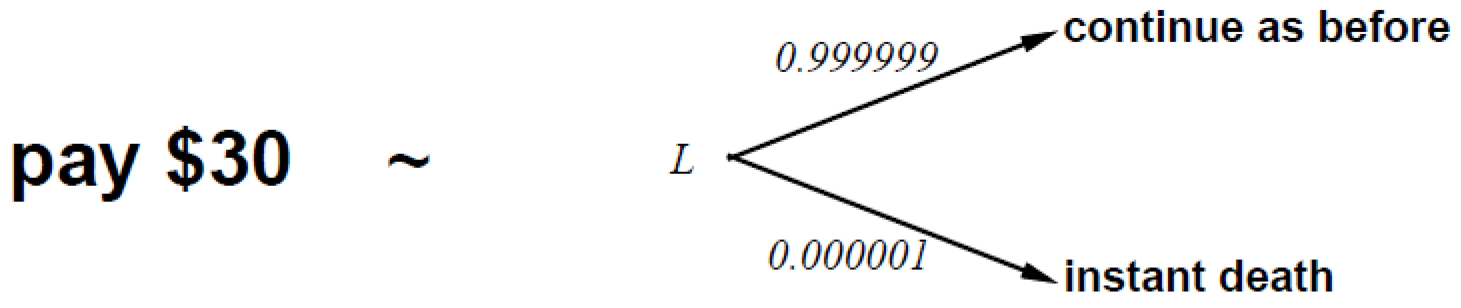
Note: an agent can be entirely rational (consistent with MEU)
without ever representing or manipulating utilities and probabilities

E.g., a lookup table for perfect tictactoe

Utilities

Utilities map states to real numbers. Which numbers?

Standard approach to assessment of human utilities:
  compare a given state A to a standard lottery L_p that has
    "best possible prize" $u_\top$ with probability p
    "worst possible catastrophe" $u_\bot$ with probability $(1-p)$
  adjust lottery probability p until $A \sim L_p$



Utility scales

Normalized utilities: $u_\top = 1.0$, $u_\perp = 0.0$

Micromorts: one-millionth chance of death
useful for Russian roulette, paying to reduce product risks, etc.

QALYs: quality-adjusted life years
useful for medical decisions involving substantial risk

Note: behavior is **invariant** w.r.t. +ve linear transformation

$$U'(x) = k_1 U(x) + k_2 \quad \text{where } k_1 > 0$$

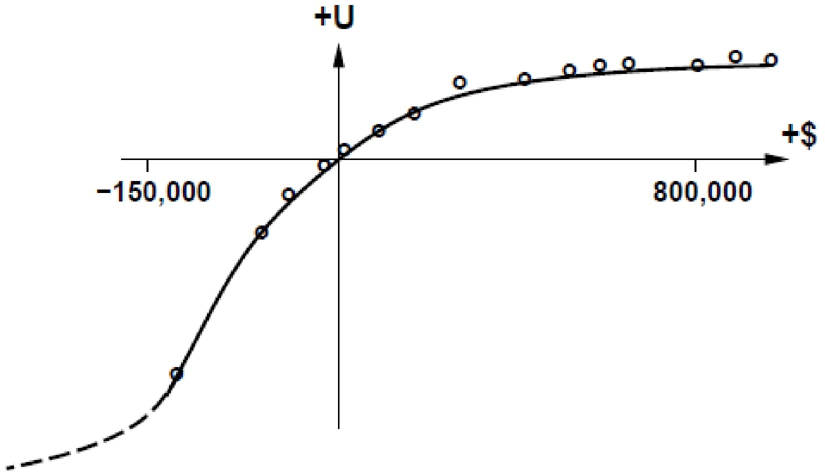
With deterministic prizes only (no lottery choices), only
ordinal utility can be determined, i.e., total order on prizes

Money

Money does **not** behave as a utility function. Given a lottery L with expected monetary value $EMV(L)$,
usually $U(L) < U(EMV(L))$, i.e., people are risk-averse.

Utility curve: for what probability p am I indifferent between a prize x and a lottery $[p, \$M;\ (1-p), \$0]$ for large M?

Typical empirical data, extrapolated with risk-prone behavior:



Beyond symbolic cognition: causality

Principles of causality

- Principles for a causal relation between X➔Y:
  - **Probabilistic association**,
  - **Temporal asymmetry**: X precedes temporally Y,
  - ~~(Physical locality)~~
  - Quantitative effect of interventions: **dose-effect relation**
    - necessity (i.e., if the cause is removed, effect is decreased)
    - sufficiency (if exposure to cause is increased, effect is increased)
  - **Counterfactuals**:
    - Y would not have been occurred with that much probability if Y hadn't been present
    - Y would have been occurred with larger probability if X had been present
  - **Bounded context-sensitivity** (~context-free): relevant on average
  - Plausible **explanation** (no alternative based on confounding).

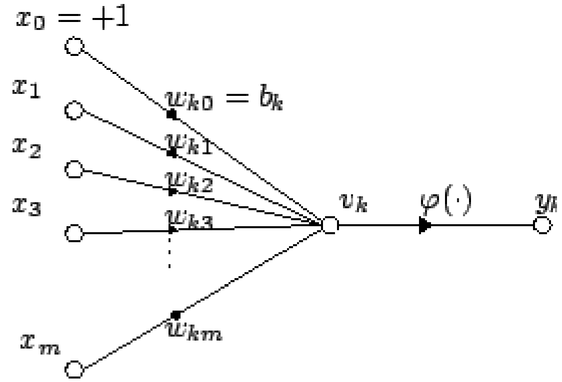
- Duality principle: rules/mechanisms vs. observations/interventions.

Beyond symbolic cognition: Artificial neural networks (ANN 1.0)

Birth of artificial neural networks (ANNs)

- 1943 McCulloch & Pitts: Boolean circuit model of brain
- 1949 Hebb: Organization of Behaviour: Hebbian learning fire together-wire together
- 1958 Frank Rosenblatt: Perceptron (Mark I Perceptron)
- 1959 Bernard Widrow and Marcian Hoff: (Multiple) ADAptive LINear Elements
- **1960-79 The physical symbol system hypothesis: search**
- 1969 Marvin Minsky, Seymour Papert: "Perceptrons"

  Credit assignment problem in multilayer perceptrons ➔ **1st ANN winter**
- (1973 Lighthill report "Artificial Intelligence: A General Survey (➔?) AI winter
  - https://en.wikipedia.org/wiki/Lighthill_report)

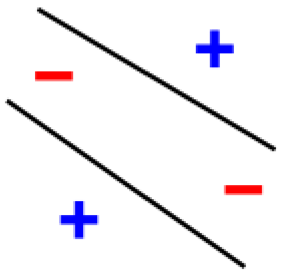


https://en.wikipedia.org/wiki/Artificial_neuron

The Credit Assignment Problem

The credit assignment problem concerns determining how the success of a system's overall performance is due to the various contributions of the system's components (Minsky, 1963).

Minsky, M. L. (1963). Steps toward artificial intelligence. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers And Thought* (pp. 406-450). New York, NY: McGraw-Hill.



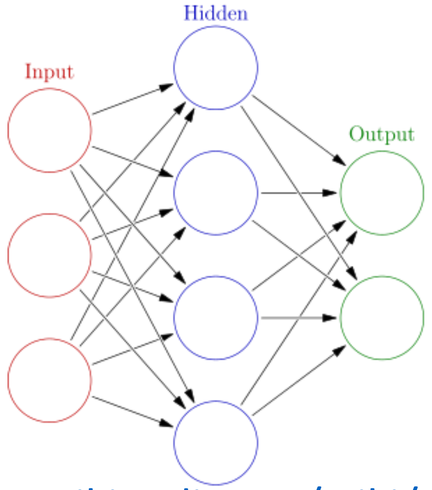
https://en.wikipedia.org/wiki/Linear_separability

http://www.bcp.psych.ualberta.ca/~mike/Pearl_Street/Dictionary/contents/C/creditassign.html

Connectionism, parallel distributed processing
ANN 2.0

Development of connectionism: the message passing paradigm

- 1969 Marvin Minsky, Seymour Papert: “Perceptrons” (➔?) NN winter
- 1974 Paul Werbos solved the backward flow of credit assignment (Freud’s work)
- 1979 Kunihiko Fukushima neocognitron precursor for convolutional neural networks
- 1982 Jon Hopfield recurrent network (Hopfield Net)
- 1985 Ackley, David H; Hinton Geoffrey E; Sejnowski, Terrence J Boltzmann machines
- 1986 Rumelhart, D.E., Hinton, G.E. and McClelland, J.L., 1986. A general framework for parallel distributed processing. *Parallel distributed processing: Explorations in the microstructure of cognition*
- Pearl, J., 1988. 88, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.
- 1988 Rumelhart, D.E., Hinton, G.E. and Williams, R.J., 1988. Learning representations by back-propagating errors. *Cognitive modeling*
- 1989 George Cybenko: universal approximation using sigmoid activation functions



https://en.wikipedia.org/wiki/Artificial_neural_network

The message passing paradigm
(the parallel distributed processing paradigm)

Precursors:
Forward/backward chaining in (Horn) logic
Propagating uncertainty factors

1980-90: PDP, inference in PGMs,…

Now (~2020): network-based propagation paradigm

Connectionism and cognitive architecture: pros-cons

Fodor, J.A. and Pylyshyn, Z.W., 1988. Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1-2), pp.3-71.

KLASSZIKUS SZEMLÉLET	*MEGKÉRDŐJELEZÉS ÉS FINOMÍTÁS*
egységes	moduláris
szimbolikus	szubszimbolikus
propozicionális	hálózatelvű
szekvenciális	párhuzamos
atomisztikus	procedurális
explicit	implicit
logikus, deduktív	intuitív, élményelvű
egyéni	szociális
testetlen	testre vonatkozó
önmagában tekinthető	evolúciós
modellálható	kimeríthetetlen
gépies, automatikus	emberi, jelentésorientált
igazságorientált	vágy irányította
tudásfüggetlen	tudás áthatotta

4.2 táblázat A klasszikus kognitivizmus és az alternatív irányok jellegzetes szembenállásai

Pléh Csaba, 2013. *A megismeréstudomány alapjai: az embertől a gépig és vissza*. Typotex.

Statistical complexity: 2nd ANN winter (1988-2006)

- 1974 Vapnik, V. and Chervonenkis, A., Theory of pattern recognition. (in Russian)
- 1989 Blumer, A., Ehrenfeucht, A., Haussler, D. and Warmuth, M.K., Learnability and the Vapnik-Chervonenkis dimension. Journal of the ACM (JACM), 36(4), pp.929-965.
- 1990 Haussler, D., *Probably approximately correct learning*. University of California, Santa Cruz, Computer Research Laboratory.
- 1992 Boser, B.E., Guyon, I.M. and Vapnik, V.N., A training algorithm for optimal margin classifiers. In Proceedings of the fifth annual workshop on Computational learning theory (pp. 144-152). ACM.
- 1996 Devroye, L., Györfi, L. and Lugosi, G., A probabilistic theory of pattern recognition springer. New York.
- **1998 Vapnik, V. and Vapnik, V., Statistical learning theory.**

+patents at Bell Labs:Y.LeCun

Convergence bounds for finite data size (ε accuracy,δ confidence)

sample complexity: $N_{\varepsilon,\delta}$

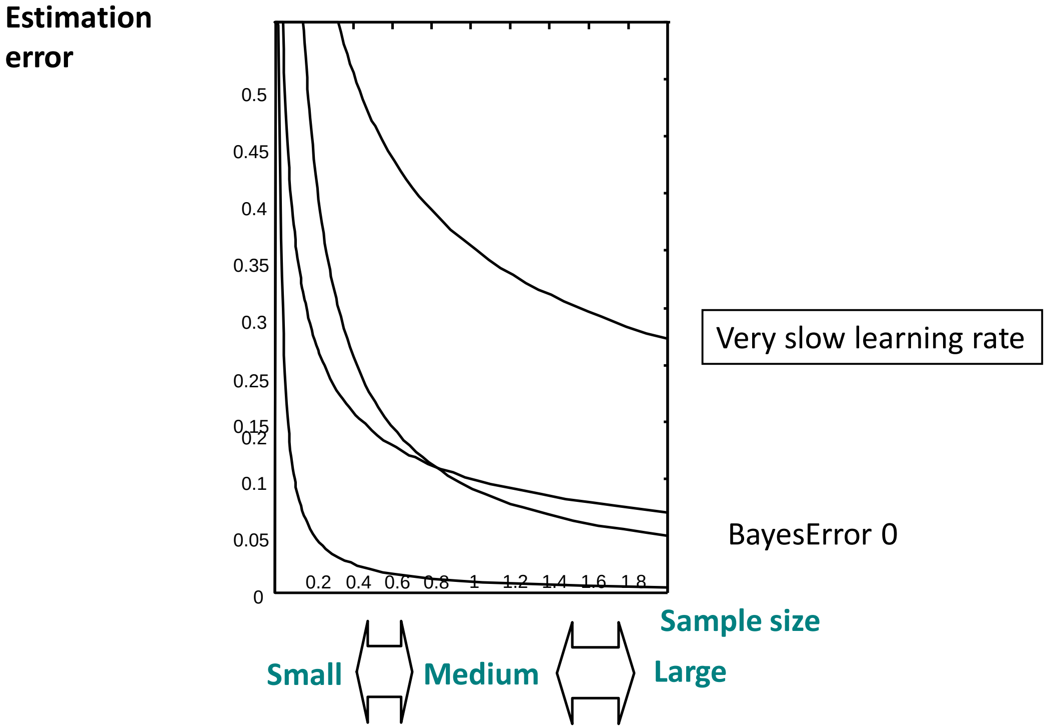
$$p(D_N : \varepsilon < | Error(\mathrm{Model}(D_N)) |) < \delta$$

Statistical complexity of learning

Sample complexity N(ε,δ):

$$\sup_{(X,Y)} P\left(L\left(\phi_n^*\right) - L_C > \varepsilon\right) \le \delta$$

Estimation error: e.g. $\boxed{O\left(\frac{\text{Class Complexity}}{\text{Data size}}\right)}$

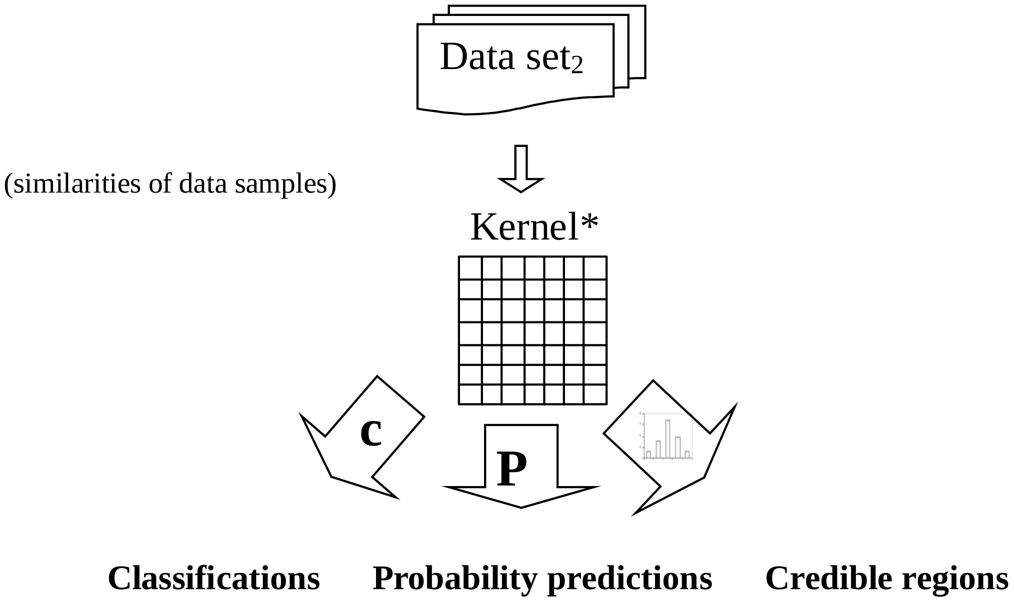


Kernel technologies

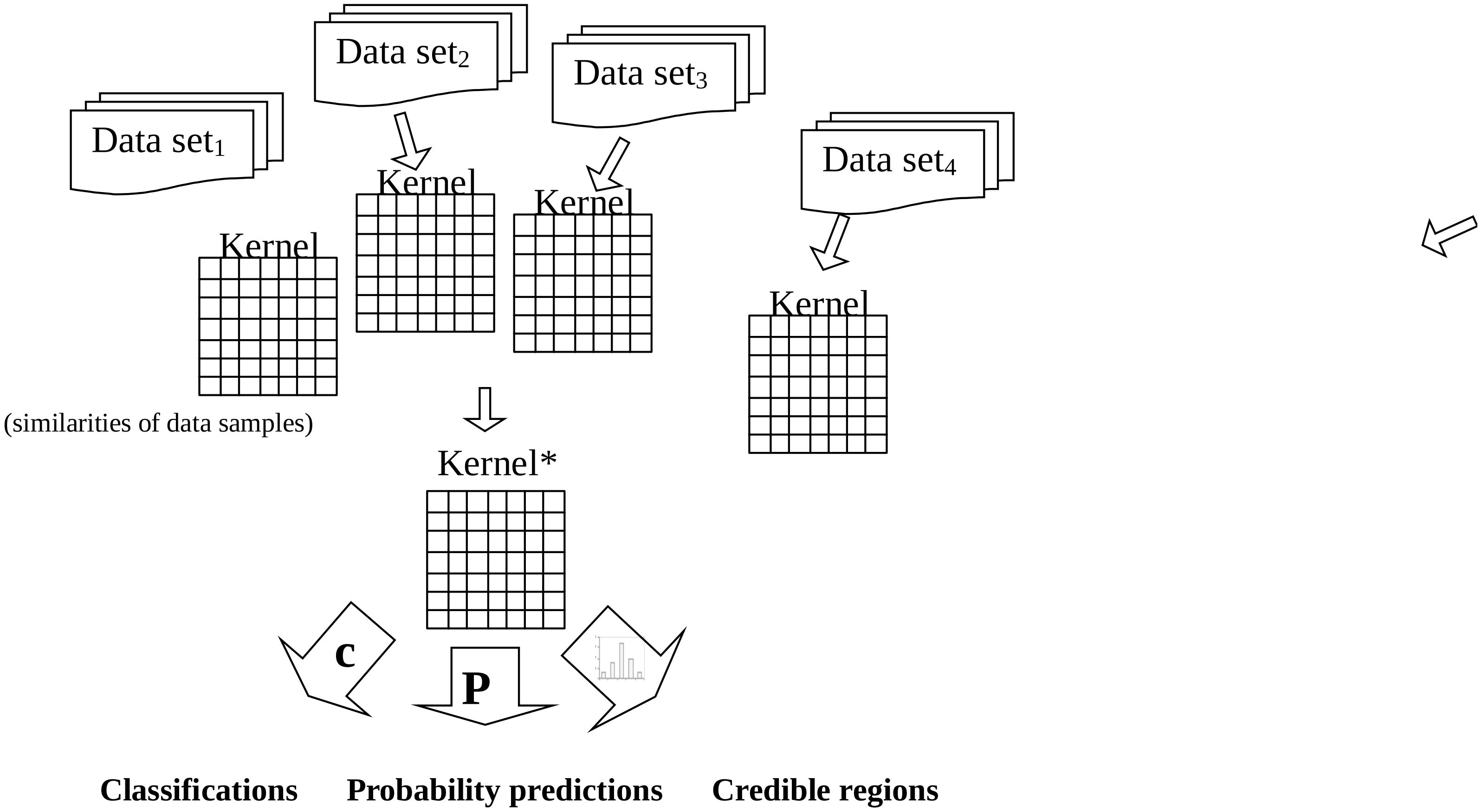
Kernel technologies

Precursors:
Reasoning with similarities, case-based resoning,..
universal consistency of 1/k-Nearest Neighbourhood

1, Statistical guarantees for inductive performance
2, Efficient computational complexities



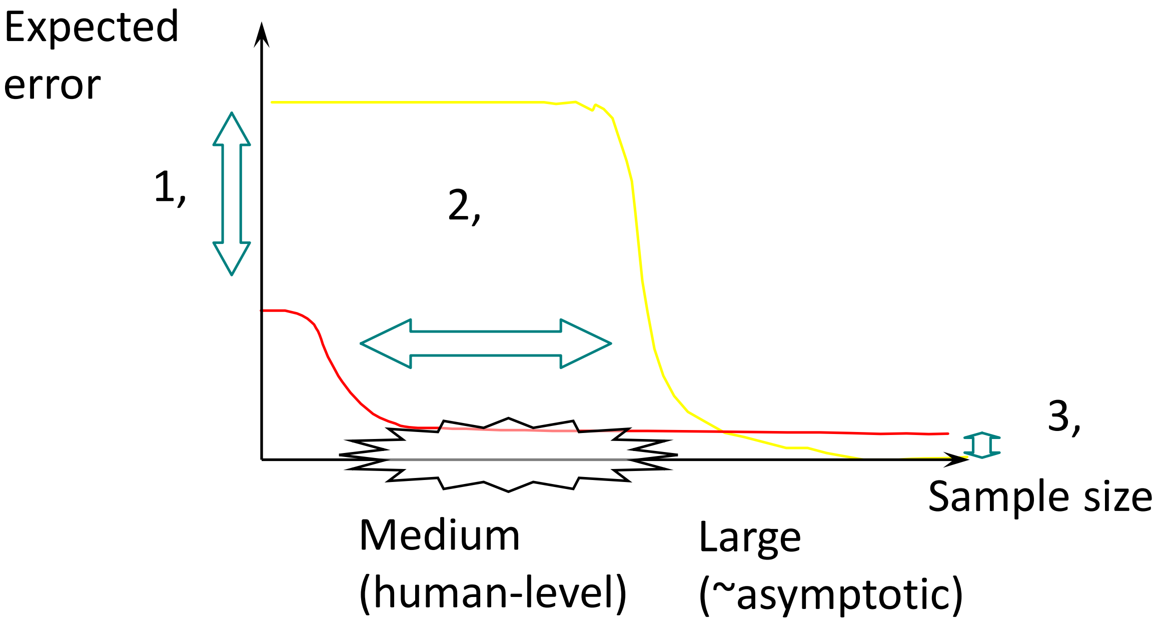
Multiple kernel learning: Multi-aspect intelligence



Bach, F.R., Lanckriet, G.R. and Jordan, M.I., 2004, July. Multiple kernel learning, conic duality, and the SMO algorithm. In *Proceedings of the twenty-first international conference on Machine learning* (p. 6). ACM.

Knowledge-based artificial neural networks (precursor for ANN 3.0)

Learning with prior knowledge



1, Initial error

2, Learning rate

3, Asymptotic error

Knowledge-based artificial neural networks

- 1993 Abu-Mostafa,Y.S.: Hints and the VC Dimension, Neural Computation, 5, 278-288
- 1994 Towell, G.G. and Shavlik, J.W., 1994. Knowledge-based artificial neural networks. *Artificial intelligence*, *70*(1-2), pp.119-165.
- 1995 Opitz,D.W.&J.W.Shavlik:Dynamically Adding Symbolically Meaningful Nodes to Knowledge-Based Neural Networks, Knowledge-Based Systems, 8(6):301-311
- 1995 P.Myllymaki. Mapping Bayesian Networks to Stochastic Neural Networks: A Foundation for Hybrid Bayesian-Neural systems. Ph.D. dissertation, University of Helsinki, No. A-1995-1, 1995
- 1998 P. Niyogi, T. Poggio, and F. Girosi. Incorporating prior information in machine learning by creating virtual examples. Proceedings of the IEEE, 86(11):2196–2209
- 1998 P. Antal. Applicability of prior domain knowledge formalised as Bayesian network in the process of construction of a classifier. In Proc. of the 24th Annual Conf. of the IEEE Industrial Electronic Society (IECON ’98), pages 2527–2531
- 2000 P. Antal, G. Fannes, H. Verrelst, B. De Moor, and J. Vandewalle. Incorporation of prior knowledge in black-box models: Comparison of transformation methods from Bayesian network to multilayer perceptrons. In Workshop on Fusion of Domain Knowledge with Data for Decision Support, 16th Uncertainty in Artificial Intelligence Conference, pages 42–48,
- **2000 I. Cloete and J. M. Zurada. Knowledge-Based Neurocomputing. MIT Press, Cambridge, MA, 2000**
- 2003 P. Antal, G. Fannes, Y. Moreau, and B. De Moor. Bayesian applications of belief networks and multilayer perceptrons for ovarian tumor classification with rejection. Artificial Intelligence in Medicine, 29:39–60

Architectures and Techniques for Knowledge-Based Neurocomputing

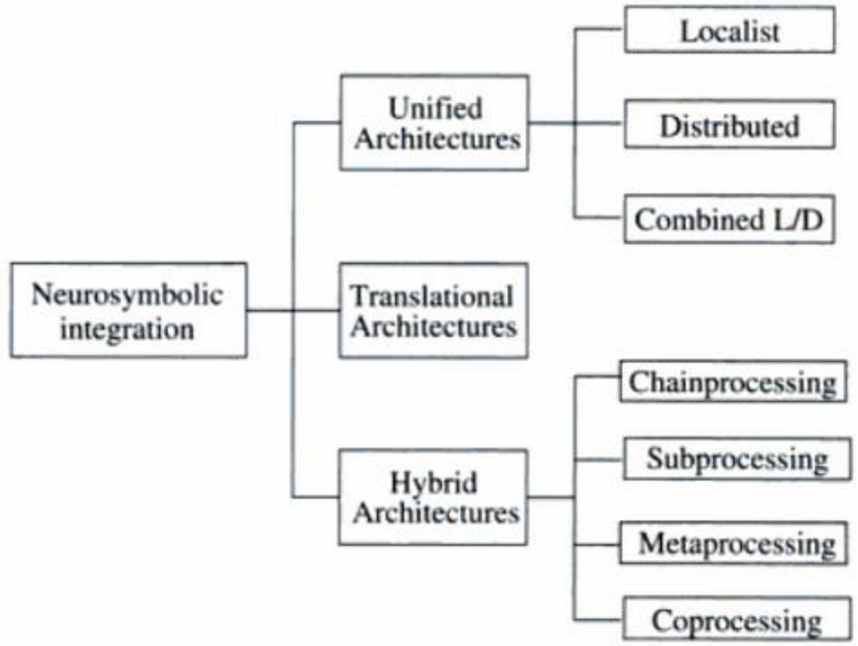
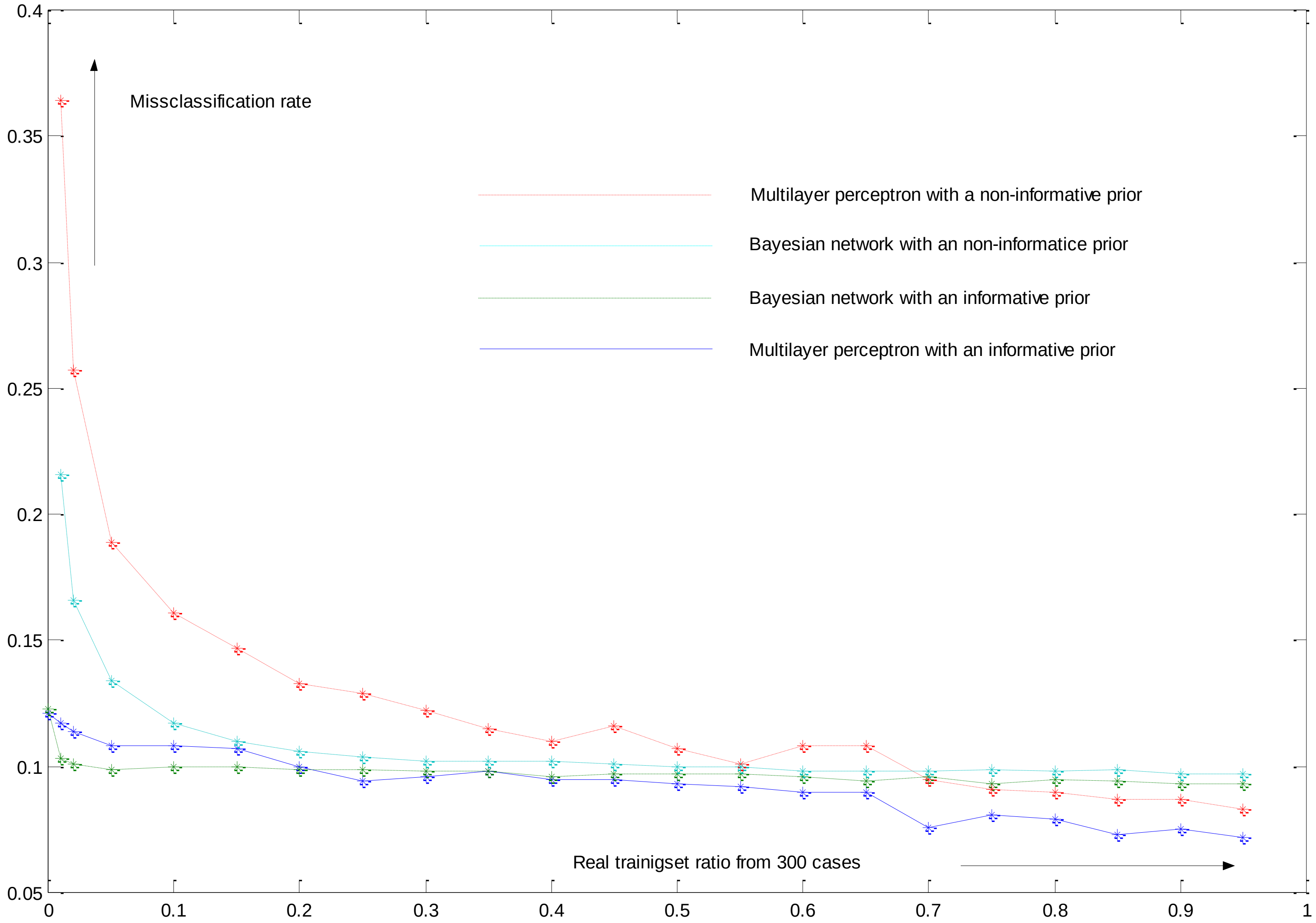


Figure 2.1 Classification of integrated neurosymbolic systems.

Informed neural networks



P. Antal, G. Fannes, D. Timmerman, Y. Moreau, B. De Moor: Bayesian Applications of Belief Networks and Multilayer Perceptrons for Ovarian Tumor Classification with Rejection, *Artificial Intelligence in Medicine*, vol. 29, pp 39-60, 2003

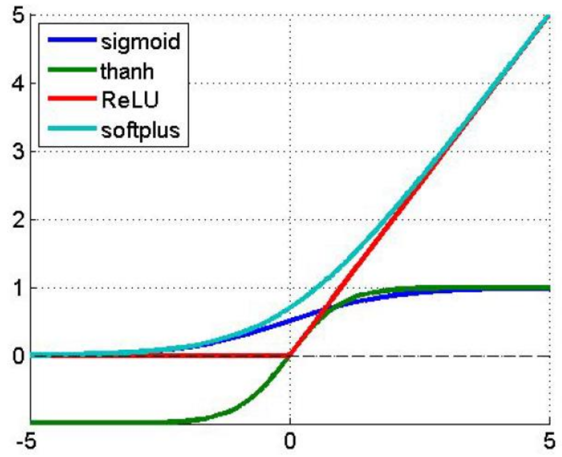
Deep learning
(Artificial neural networks 2.0++)

Return of ANNs as deep learning: 2006<

- 2006 - Hinton, Osindero, Yee-Whye Teh: A fast learning algorithm for deep belief nets
  - Learning by layers then global refinement
- 2010, Glorot, Bengio: Understanding the difficulty of training deep feedforward neural networks
  - Novel transfer function: ReLU,
  - Novel parameter initialization
- 2011 - Mohamed, A. R., Sainath, T. N., Dahl, G., Ramabhadran, B., Hinton, G. E., & Picheny, M. (2011, May). Deep belief networks using discriminative features for phone recognition.
  - GPUs (x10-x100 speed-up)
- 2011 - Google Brain
  - The work resulted in unsupervised neural net learning of an unprecedented scale - 16,000 CPU cores powering the learning of a whopping 1 billion weights
- 2012/14/15: AlexNet (8),GoogLeNet(22), ResNet(152 hidden layer)

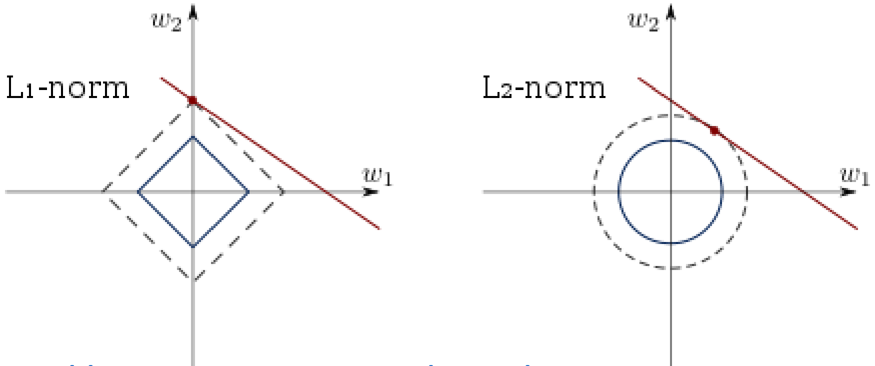
ANN learning in 1986 vs. in 2006

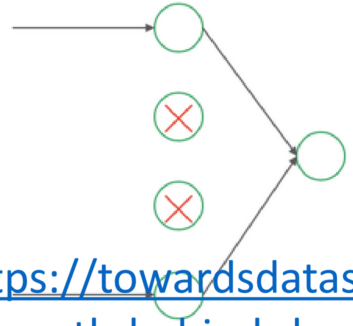
1. Our labeled datasets were thousands of times too small.
2. Our computers were millions of times too slow.
3. We initialized the weights in a stupid way.
4. We used the wrong type of non-linearity.



LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *Nature*, *521*(7553), p.436.

[L1/Lasso, parameter sharing (convolutional nets, long short-term memory), dropout, SGD/momentum/…/ADAM,..]





https://imgur.com/a/Hqolp#2dKCQHh

http://ruder.io/optimizing-gradient-descent/

https://towardsdatascience.com/simplified-math-behind-dropout-in-deep-learning-6d50f3f47275

https://bl.ocks.org/EmilienDupont/aaf429be5705b219aaaf8d691e27ca87

https://en.wikipedia.org/wiki/Lasso_(statistics)

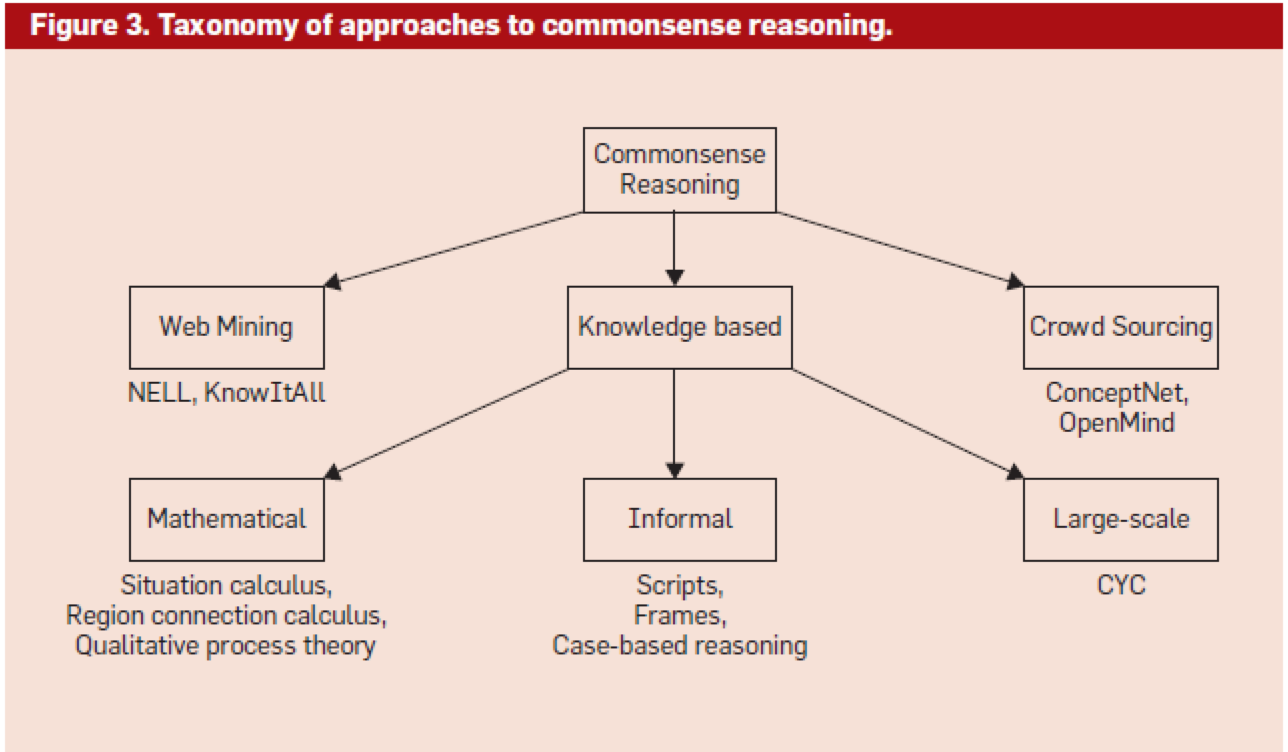
Deep learning problems (➔3rd ANN winter?)

1. Deep learning thus far is data hungry
2. Deep learning thus far is shallow and has limited capacity for transfer
3. Deep learning thus far has no natural way to deal with hierarchical structure
4. Deep learning thus far has struggled with open-ended inference
5. Deep learning thus far is not sufficiently transparent
6. Deep learning thus far has not been well integrated with prior knowledge
7. Deep learning thus far cannot inherently distinguish causation from correlation
8. Deep learning presumes a largely stable world, in ways that may be problematic
9. Deep learning thus far works well as an approximation, but its answers often cannot be fully trusted
10. Deep learning thus far is difficult to engineer with

Marcus, G., 2018. Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*.

AGI-inspired ANNs
(ANNs 3.0)

Large-scale data for inferring-learning commonsense knowledge?



https://mosaic.allenai.org/projects/mosaic-commonsense-benchmarks

SWAG: A Large-Scale Adversarial Dataset for Grounded Commonsense Inference

Rowan Zellers, Yonatan Bisk, Roy Schwartz, and Yejin Choi • EMNLP • 2018

Sources for commonsense knowledge

+Federated learning

- Emails („free” service providers)
- Chats („free” service providers)
- Everyday conversations (Siri, Alexa)
- Wearable electronics, basic physiological data (quantified self)…
- Decoding the inner speech/Thoughts-to-speech
  - Moses, D.A., Leonard, M.K., Makin, J.G. and Chang, E.F., 2019. Real-time decoding of question-and-answer speech dialogue using human cortical activity. *Nature communications*, *10*(1), pp.1-14.
  - Anumanchipalli, G.K., Chartier, J. and Chang, E.F., 2019. Speech synthesis from neural decoding of spoken sentences. *Nature*, *568*(7753), p.493.

Beyond standard learning

- Learning with prior knowledge
- Sequential/online learning
- Reinforcement learning
- Multitask learning
- Transfer learning
- Budgeted learning
- Active learning
- One-shot learning
- Federated learning
- …
- (Machine teaching)

Hilbert's twenty-three problems

Problem	Brief explanation	Status	Year Solved
1st	The continuum hypothesis (that is, there is no set whose cardinality is strictly between that of the integers and that of the real numbers)	Proven to be impossible to prove or disprove within Zermelo–Fraenkel set theory with or without the Axiom of Choice (provided Zermelo–Fraenkel set theory is consistent, i.e., it does not contain a contradiction). There is no consensus on whether this is a solution to the problem.	1940, 1963
2nd	Prove that the axioms of arithmetic are consistent.	There is no consensus on whether results of Gödel and Gentzen give a solution to the problem as stated by Hilbert. Gödel's second incompleteness theorem, proved in 1931, shows that no proof of its consistency can be carried out within arithmetic itself. Gentzen proved in 1936 that the consistency of arithmetic follows from the well-foundedness of the ordinal ε_0.	1931, 1936
3rd	Given any two polyhedra of equal volume, is it always possible to cut the first into finitely many polyhedral pieces that can be reassembled to yield the second?	Resolved. Result: No, proved using Dehn invariants.	1900
4th	Construct all metrics where lines are geodesics.	Too vague to be stated resolved or not.[h]	—
5th	Are continuous groups automatically differential groups?	Resolved by Andrew Gleason, depending on how the original statement is interpreted. If, however, it is understood as an equivalent of the Hilbert–Smith conjecture, it is still unsolved.	1953?
6th	Mathematical treatment of the axioms of physics (a) axiomatic treatment of probability with limit theorems for foundation of statistical physics (b) the rigorous theory of limiting processes "which lead from the atomistic view to the laws of motion of continua"	Partially resolved depending on how the original statement is interpreted.[9] Items (a) and (b) were two specific problems given by Hilbert in a later explanation.[1] Kolmogorov's axiomatics (1933) is now accepted as standard. There is some success on the way from the "atomistic view to the laws of motion of continua."[10]	1933–2002?
7th	Is a^b transcendental, for algebraic $a \neq 0,1$ and irrational algebraic b ?	Resolved. Result: Yes, illustrated by Gelfond's theorem or the Gelfond–Schneider theorem.	1934
8th	The Riemann hypothesis ("the real part of any non-trivial zero of the Riemann zeta function is ½") and other prime number problems, among them Goldbach's conjecture and the twin prime conjecture	Unresolved.	—
9th	Find the most general law of the reciprocity theorem in any algebraic number field.	Partially resolved.[i]	—
10th	Find an algorithm to determine whether a given polynomial Diophantine equation with integer coefficients has an integer solution.	Resolved. Result: Impossible; Matiyasevich's theorem implies that there is no such algorithm.	1970
11th	Solving quadratic forms with algebraic numerical coefficients.	Partially resolved.[11]	—
12th	Extend the Kronecker–Weber theorem on Abelian extensions of the rational numbers to any base number field.	Unresolved.	—
13th	Solve 7th degree equation using algebraic (variant: continuous) functions of two parameters.	The problem was partially solved by Vladimir Arnold based on work by Andrei Kolmogorov.[j]	1957
14th	Is the ring of invariants of an algebraic group acting on a polynomial ring always finitely generated?	Resolved. Result: No, a counterexample was constructed by Masayoshi Nagata.	1959
15th	Rigorous foundation of Schubert's enumerative calculus.	Partially resolved.	—
16th	Describe relative positions of ovals originating from a real algebraic curve and as limit cycles of a polynomial vector field on the plane.	Unresolved, even for algebraic curves of degree 8.	—
17th	Express a nonnegative rational function as quotient of sums of squares.	Resolved. Result: Yes, due to Emil Artin. Moreover, an upper limit was established for the number of square terms necessary.	1927
18th	(a) Is there a polyhedron that admits only an anisohedral tiling in three dimensions? (b) What is the densest sphere packing?	(a) Resolved. Result: Yes (by Karl Reinhardt). (b) Widely believed to be resolved, by computer-assisted proof (by Thomas Callister Hales). Result: Highest density achieved by close packings, each with density approximately 74%, such as face-centered cubic close packing and hexagonal close packing.[k]	(a) 1928 (b) 1998
19th	Are the solutions of regular problems in the calculus of variations always necessarily analytic?	Resolved. Result: Yes, proven by Ennio de Giorgi and, independently and using different methods, by John Forbes Nash.	1957
20th	Do all variational problems with certain boundary conditions have solutions?	Resolved. A significant topic of research throughout the 20th century, culminating in solutions for the non-linear case.	?
21st	Proof of the existence of linear differential equations having a prescribed monodromic group	Partially resolved. Result: Yes/No/Open depending on more exact formulations of the problem.	?
22nd	Uniformization of analytic relations by means of automorphic functions	Unresolved.	?
23rd	Further development of the calculus of variations	Too vague to be stated resolved or not.	—

https://en.wikipedia.org/wiki/Hilbert’s_problems

Smale's problems

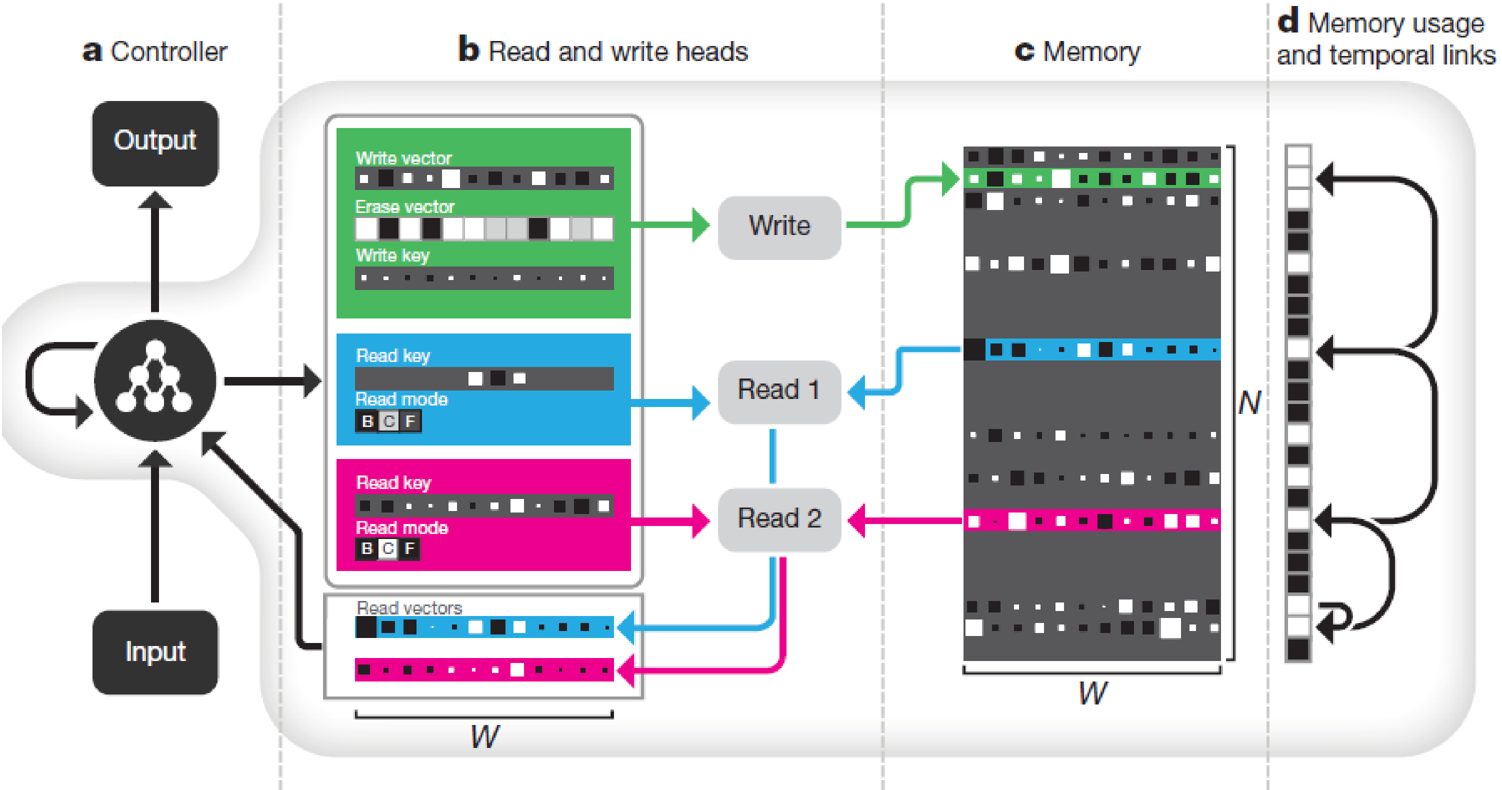
Problem	Brief explanation	Status	Year Solved
1st	Riemann hypothesis: The real part of every non-trivial zero of the Riemann zeta function is 1/2. (see also Hilbert's eighth problem)	Unresolved.	–
2nd	Poincaré conjecture: Every simply connected, closed 3-manifold is homeomorphic to the 3-sphere.	Resolved. Result: Yes, Proved by Grigori Perelman using Ricci flow.[3][4][5]	2003
3rd	P versus NP problem: For all problems for which an algorithm can *verify* a given solution quickly (that is, in polynomial time), can an algorithm also *find* that solution quickly?	Unresolved.	–
4th	Shub–Smale tau-conjecture on the integer zeros of a polynomial of one variable[6][7]	Unresolved.	–
5th	Can one decide if a Diophantine equation $f(x,y) = 0$ (input $f \in \mathbb{Z}[u,v]$) has an integer solution, (x,y), in time $(2^s)^c$ for some universal constant c? That is, can the problem be decided in exponential time?	Unresolved.	–
6th	Is the number of relative equilibria finite, in the n-body problem of celestial mechanics, for any choice of positive real numbers $m_1, \ldots, m_n$ as the masses?	Partially resolved. Proved for five bodies by A. Albouy and V. Kaloshin in 2012.[8]	2012
7th	Distribution[clarification needed] of points on the 2-sphere	Partially resolved. A noteworthy form of this problem is the Thomson Problem of equal point charges on a unit sphere governed by the electrostatic Coulomb's law. Very few exact N-point solutions are known while most solutions are numerical. Numerical solutions to this problem have been shown to correspond well with features of electron shell-filling in Atomic structure found throughout the periodic table.[9] A well-defined, intermediate step to this problem involving a point charge at the origin has been reported.[10]	–
8th	Extend the mathematical model of general equilibrium theory to include price adjustments	Gjerstad (2013)[11] extends the deterministic model of price adjustment to a stochastic model and shows that when the stochastic model is linearized around the equilibrium the result is the autoregressive price adjustment model used in applied econometrics. He then tests the model with price adjustment data from a general equilibrium experiment. The model performs well in a general equilibrium experiment with two commodities.	2013
9th	The linear programming problem: Find a strongly-polynomial time algorithm which for given matrix $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ decides whether there exists $x \in \mathbb{R}^n$ with $Ax \geq b$.	Unresolved.	–
10th	Pugh's closing lemma (higher order of smoothness)	Partially Resolved. Proved for Hamiltonian diffeomorphisms of closed surfaces by M. Asaoka and K. Irie in 2016.[12]	2016
11th	Is one-dimensional dynamics generally hyperbolic? (a) Can a complex polynomial T be approximated by one of the same degree with the property that every critical point tends to a periodic sink under iteration? (b) Can a smooth map $T : [0,1] \to [0,1]$ be C^r approximated by one which is hyperbolic, for all $r > 1$?	(a) Unresolved, even in the simplest parameter space of polynomials, the Mandelbrot set. (b) Resolved. Proved by Kozlovski, Shen and van Strien.[13]	2007
12th	Can a diffeomorphism of a compact manifold M onto itself be C approximated, all $r \geq 1$, by one $T : M \to M$ which commutes with only its iterates? In other words, what are the centralizers of a diffeomorphism?	Partially Resolved. Solved in the C^1 topology by Christian Bonatti, Sylvain Crovisier and Amie Wilkinson[14] in 2009. Still open in the C^r topology for $r > 1$.	2009
13th	Hilbert's 16th problem: Describe relative positions of ovals originating from a real algebraic curve and as limit cycles of a polynomial vector field on the plane.	Unresolved, even for algebraic curves of degree 8.	–
14th	Do the properties of the Lorenz attractor exhibit that of a strange attractor?	Resolved. Result: Yes, solved by Warwick Tucker using interval arithmetic.[15]	2002
15th	Do the Navier–Stokes equations in $\mathbb{R}^3$ always have a unique smooth solution that extends for all time?	Unresolved.	–
16th	Jacobian conjecture: If The Jacobian determinant of F is a non-zero constant and k has characteristic 0, then F has an inverse function $G : k^N \to k^N$, and G is regular (in the sense that its components are polynomials).	Unresolved.	–
17th	Solving polynomial equations in polynomial time in the average case	Resolved. C. Beltrán and L. M. Pardo found a uniform probabilistic algorithm (average Las Vegas algorithm) for Smale's 17th problem[16][17] F. Cucker and P. Bürgisser made the smoothed analysis of a probabilistic algorithm *à la Beltrán-Pardo* and then exhibited a deterministic algorithm running in time $N^{O(\log \log N)}$.[18] Finally, P. Lairez found an alternative method to de-randomize the algorithm and thus found a deterministic algorithm which runs in average polynomial time.[19] All these works follow Shub and Smale's foundational work (the "Bezout series") started in[20]	2008-2016
18th	Limits of intelligence (it talks about the fundamental problems of intelligence and learning, both from the human and machine side)[21]	Unresolved.	–

https://en.wikipedia.org/wiki/Smale’s_problems

Smale’s 18th problem: Limits of intelligence

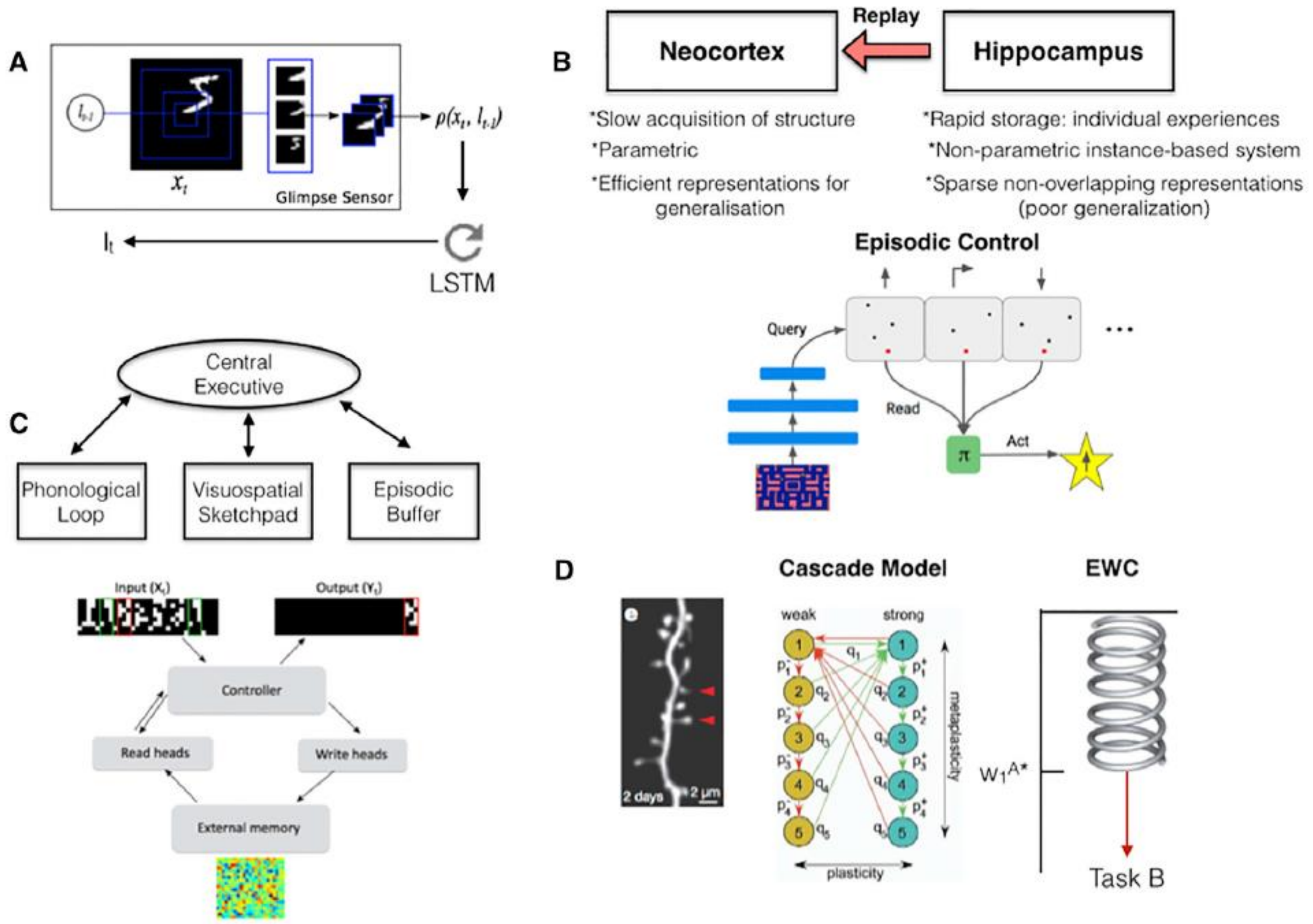
- Penrose (1991) attempts to show some limitations of artificial intelligence. Involved in his argumentation is the interesting question, "is the Mandelbrot set decidable?" (see problem 14) and implications of the Gödel incompleteness theorem.
- However a broader study is called for, one which involves deeper models of the brain, and of the computer, in a search of what artificial and human intelligence have in common,and how they differ.
- This project requires the development of a mathematical model of intelligence, with variations to take into account the differences between kinds of intelligence.
- It is useful to realize that there can be no unique model. Even in physics which is more clearly defined, one has classical mechanics, quantum mechanics, and relativity theory, each yielding its own insights and understandings and each with its own limitations. Models are idealizations with drastic simplifications which capture main truths.
- An important part of intelligent activity is problem solving. For this one has a traditional model, the Turing machine, as well as a newer machine which processes real numbers (see BCSS), referred to previously in problem 3. The Turing machine has been accepted as a reasonable model for the digital computer. We have argued for the alternative real number machine as a more appropriate model for the digital computer's use in scientic computation and in situations where arithmetic operations dominate (the Manifesto as reprinted as Chapter 1 of BCSS). Such mathematical models for human intelligence are less developed.
- There is one example of a general problem that comes to the forefront; that is the problem of equation solving for polynomial systems, over some field of numbers. The real numbers with inequalities are an important special case of this problem. Artificial intelligence hasencountered it in its study of robotics. Moreover, over any field, equation solving possesses a universality in a formal mathematical sense in the theory of NP completeness.
- One might ask, is there a form of intelligence that can solve general systems of polynomial equations. This problem is anticipated by the previous problems 3 and 17.
- **The use of the Turing machine versus its real counterpart is a manifestation of the age old conflict between the discrete and the continuous. I believe that the real number machine is the more important of the two for understanding the problem solving limitations of humans. ….**

Towards hybrid systems: Neural Turing Machines



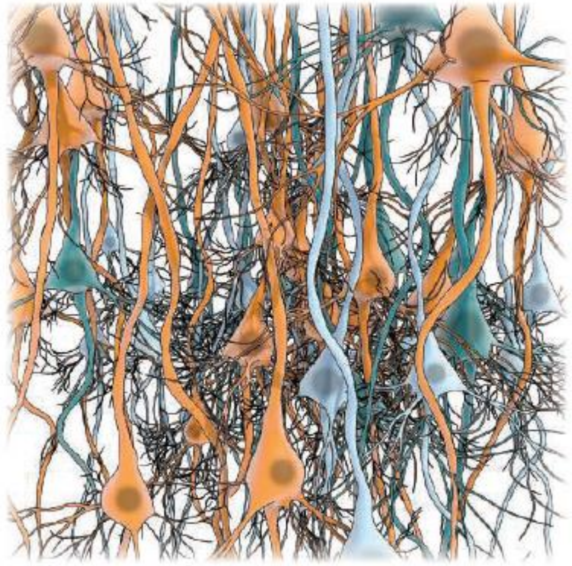
Graves, A., Wayne, G. and Danihelka, I., 2014. Neural turing machines. *arXiv preprint arXiv:1410.5401*.
Graves, A., Wayne, G., Reynolds, M., Harley, T., Danihelka, I., Grabska-Barwińska, A., Colmenarejo, S.G., Grefenstette, E., Ramalho, T., Agapiou, J. and Badia, A.P., 2016. Hybrid computing using a neural network with dynamic external memory. *Nature*, *538*(7626), p.471.

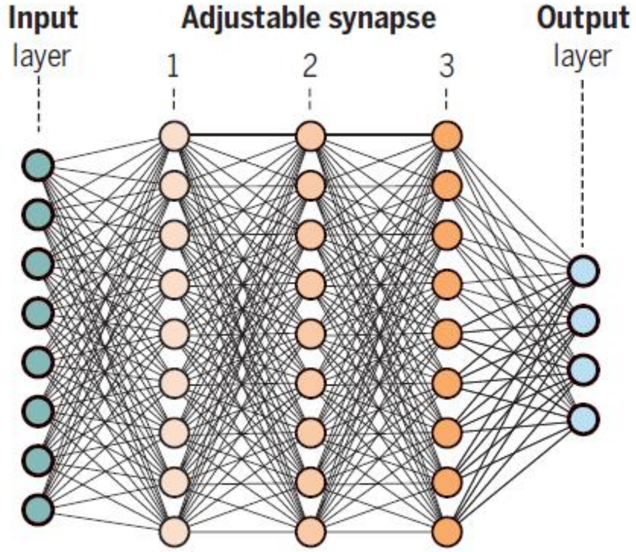
Neuroscience-Inspired Artificial Intelligence



Hassabis, D., Kumaran, D., Summerfield, C. and Botvinick, M., 2017. Neuroscience-inspired artificial intelligence. *Neuron*, *95*(2), pp.245-258.

Using neuroscience to develop artificial intelligence





Complex neural network
Connectivity in cortical networks includes rich sets of connections, including local and long-range lateral connectivity, and top-down connections from high to low levels of the hierarchy.

Informed AI network
Biological innate connectivity patterns provide mechanisms that guide human cognitive learning. Discovering similar mechanisms, by machine learning or by mimicking the human brain, may prove crucial for future artificial systems with human-like cognitive abilities.

Complexity of structure

• **Evolution**	**x100m years**	**10years**
• **Development**	**9 months (~25 years)**	**1sec**
• **Learning(&Development)**	**experimenting~25 years (+9m)**	**days (<weeks)**
• **Information content**	**DNA+epigenetics+nature(!!!)**	**xGB**
• **Computational model**	**nature(!!!)**	**UTM**

Ullman, S., 2019. Using neuroscience to develop artificial intelligence. *Science*, *363*(6428), pp.692-693.

Summary

- Cognitive science
  - Cognitive architectures
  - Criticisms
    - Symbols, values, causality
- The connectionist approach
  - Connectionism/artificial neural networks (ANN 1.0): 1943-1969
    - ANN „winter” 1.0
  - Parallel distributed processing (ANN 2.0): 1974-1988
    - ANN „winter” 2.0
    - Knowledge-based neurocomputing
  - Deep learning (ANN 2.0++): 2006-
    - Limits of deep ANNs
  - Neuroinspired-AI (ANN 3.0)
- **Next: the neurobiological substrate**

Heuristics

Problem solving with search

- A problem is defined by:
  - An initial state, e.g. *Arad*
  - Successor function *S(X)*= set of action-state pairs
    - e.g. *S(Arad)={<Arad $\rightarrow$ Zerind, Zerind>,…}*

  intial state + successor function = state space
  - Goal test, can be
    - Explicit, e.g. *x='at bucharest'*
    - Implicit, e.g. *checkmate(x)*
  - Path cost (additive)
    - e.g. sum of distances, number of actions executed, …
    - *c(x,a,y)* is the step cost, assumed to be >= 0

A solution is a sequence of actions from initial to goal state.

Optimal solution has the lowest path cost.

Best-first search

- General approach of informed search:
  - Best-first search: node is selected for expansion based on an *evaluation function* $f(n)$ in TREE-SEARCH().
- Idea: evaluation function measures distance to the goal.
  - Choose node which *appears* best
- Implementation:
  - $fringe$ is queue sorted in decreasing order of desirability.
  - Special cases: greedy search, A* search

A heuristic function

- [dictionary]*"A rule of thumb, simplification, or educated guess that reduces or limits the search for solutions in domains that are difficult and poorly understood."*
  - $h(n)$ = estimated cost of the cheapest path from node *n* to goal node.
  - If *n* is goal then $h(n)=0$

A* search

- Best-known form of best-first search.
- Idea: avoid expanding paths that are already expensive.
- Evaluation function $f(n)=g(n) + h(n)$
  - $g(n)$ the cost (so far) to reach the node.
  - $h(n)$ estimated cost to get from the node to the closest goal.
  - $f(n)$ estimated total cost of path through n to goal.

A* search

- A* search uses an admissible heuristic
  - A heuristic is *admissible* if it *never overestimates* the cost to reach the goal (~optimistic).

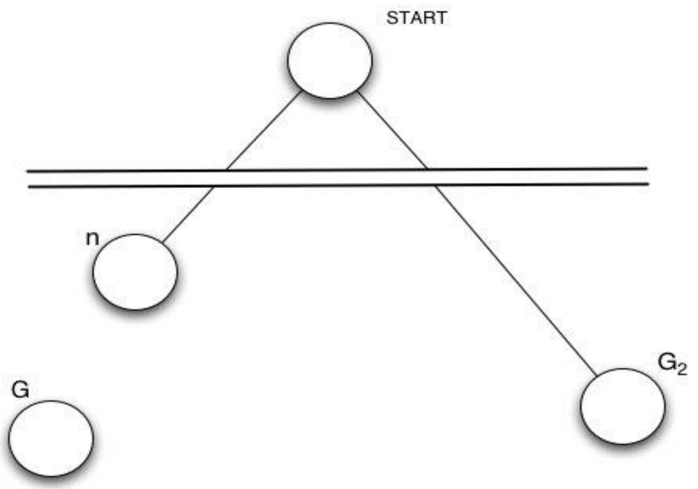
Formally:

1. $h(n) <= h^*(n)$ where $h^*(n)$ is the true cost from n
2. $h(n) >= 0$ so $h(G)=0$ for any goal G.

e.g. $h_{SLD}(n)$ never overestimates the actual road distance

Theorem: If $h(n)$ is admissible, A^* using `BEST-FIRST-SEARCH()` with selector function f(n)=h(n) is optimal.

Optimality of A*(standard proof)



- Suppose a suboptimal goal G_2 in the queue.
- Let n be an unexpanded node on a shortest to optimal goal G.

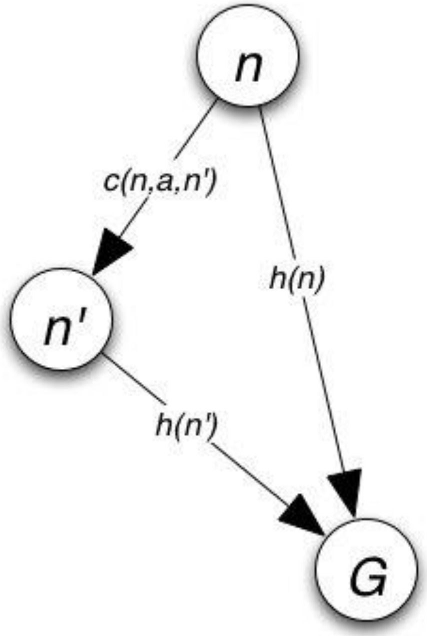
$f(G_2)$	$= g(G_2)$	since $h(G_2)=0$
	$> g(G)$	since G_2 is suboptimal
	$>= f(n)$	since h is admissible

Since $f(G_2) > f(n)$, A* will never select G_2 for expansion (i.e. for checking, but note that G_2 can be inside the queue).

Consistency

- A heuristic is consistent if

$$h(n) \le c(n,a,n') + h(n')$$



- If h is consistent, we have

$$\begin{aligned} f(n') &= g(n') + h(n') \\ &= g(n) + c(n,a,n') + h(n') \\ &\ge g(n) + h(n) \\ &\ge f(n) \end{aligned}$$

i.e. f(n) is non-decreasing along any path.

Theorem: If *h(n)* is consistent, A* using `GRAPH−SEARCH` is optimal

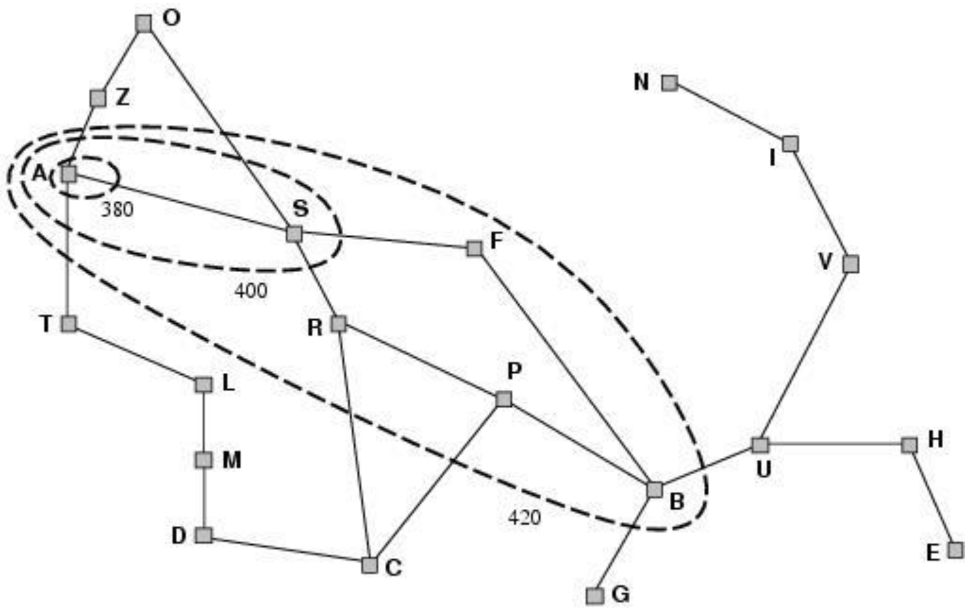
Optimality of A*(more usefull)

- A* expands nodes in order of increasing *f* value
- Contours can be drawn in state space
  - Uniform-cost search adds circles.

  - F-contours are gradually

  Added:
  1) nodes with *f(n)<C**
  2) Some nodes on the goal
  Contour (*f(n)=C**).

  Contour *i* has all nodes
  with $f=f_i$, where $f_i < f_{i+1}$.



A* search, evaluation

- Completeness: YES
  - Since bands of increasing *f* are added
  - Unless there are infinitly many nodes with $f<f(G)$

A* search, evaluation

- Completeness: YES
- Time complexity:
  - Number of nodes expanded is still exponential in the length of the solution.

A* search, evaluation

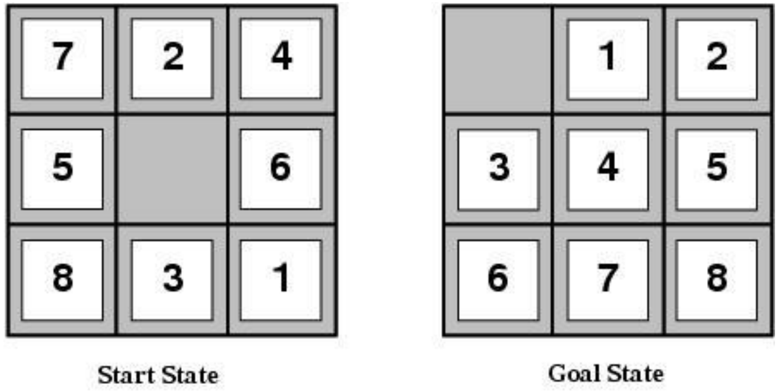
- Completeness: YES
- Time complexity: (exponential with path length)
- Space complexity:
  - It keeps all generated nodes in memory
  - Hence space is the major problem not time

A* search, evaluation

- Completeness: YES
- Time complexity: (exponential with path length)
- Space complexity:(all nodes are stored)
- Optimality: YES
    - Cannot expand f_{i+1} until f_i is finished.
    - A* expands all nodes with $f(n) < C^*$
    - A* expands some nodes with $f(n)=C^*$
    - A* expands no nodes with $f(n)>C^*$

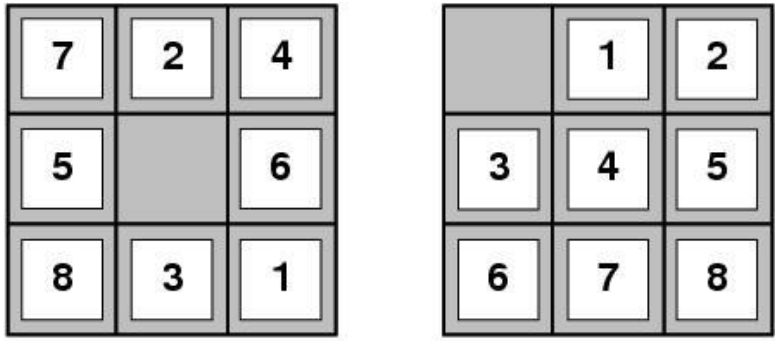
Also *optimally efficient* (not including ties)

Heuristic functions



- E.g for the 8-puzzle
  - Avg. solution cost is about 22 steps (branching factor +/- 3)
  - Exhaustive search to depth 22: 3.1 x 10^{10} states.
  - A good heuristic function can reduce the search process.

Heuristic functions



Start State Goal State

- E.g for the 8-puzzle knows two commonly used heuristics
- h_1 = the number of misplaced tiles
    - $h_1(s)=8$
- h_2 = the sum of the distances of the tiles from their goal positions (manhattan distance).
    - $h_2(s)=3+1+2+2+2+3+3+2=18$

Heuristic quality

- Effective branching factor b*
  - Is the branching factor that a uniform tree of depth *d* would have in order to contain *N+1* nodes. $N+1=1+b^*+(b^*)^2+\ldots+(b^*)^d$
  - Measure is fairly constant for sufficiently hard problems.
    - Can thus provide a good guide to the heuristic's overall usefulness.
    - A good value of b* is 1.

Inventing admissible heuristics

- Admissible heuristics can be derived from the exact solution cost of a relaxed version of the problem:
  - Relaxed 8-puzzle for h_1 : a tile can move anywhere

    As a result, $h_1(n)$ gives the shortest solution
  - Relaxed 8-puzzle for h_2 : a tile can move to any adjacent square.

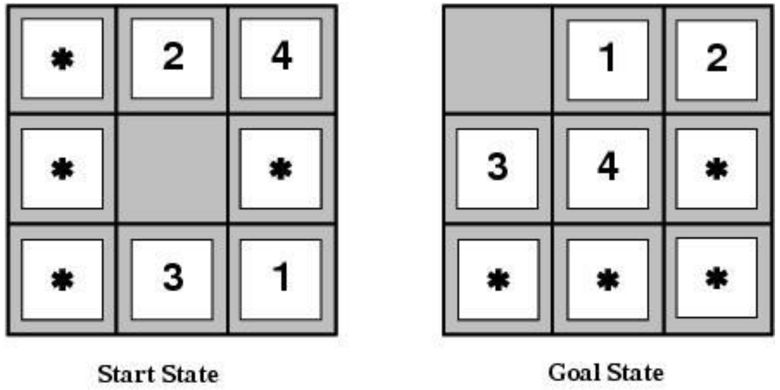
    As a result, $h_2(n)$ gives the shortest solution.

The optimal solution cost of a relaxed problem is no greater than the optimal solution cost of the real problem.

`ABSolver` found a useful heuristic for the *Rubic cube*.

Inventing admissible heuristics

- Admissible heuristics can also be derived from the solution cost of a subproblem of a given problem.
- This cost is a lower bound on the cost of the real problem.
- Pattern databases store the exact solution for every possible subproblem instance.
  - The complete heuristic is constructed using the patterns in the DB



Inventing admissible heuristics

- Another way to find an admissible heuristic is through learning from experience:
  - Experience = solving lots of 8-puzzles
  - An inductive learning algorithm can be used to predict costs for other states that arise during search.

Prieditis: Machine Discovery of Effective Admissible Heuristics, 1993

Heuristic quality and dominance

- 1200 random problems with solution lengths from 2 to 24.

	Search Cost			Effective Branching Factor		
d	IDS	$A^*(h_1)$	$A^*(h_2)$	IDS	$A^*(h_1)$	$A^*(h_2)$
2	10	6	6	2.45	1.79	1.79
4	112	13	12	2.87	1.48	1.45
6	680	20	18	2.73	1.34	1.30
8	6384	39	25	2.80	1.33	1.24
10	47127	93	39	2.79	1.38	1.22
12	3644035	227	73	2.78	1.42	1.24
14	–	539	113	–	1.44	1.23
16	–	1301	211	–	1.45	1.25
18	–	3056	363	–	1.46	1.26
20	–	7276	676	–	1.47	1.27
22	–	18094	1219	–	1.48	1.28
24	–	39135	1641	–	1.48	1.26

- If $h_2(n) >= h_1(n)$ for all n (both admissible)

  then h_2 *dominates* h_1 and is better for search