Jakovác Antal
(Adat- és Számításintenzív Tudományok kutatócsoport)

Understanding
Science
(Artificial) Intelligence

September 25, 2020

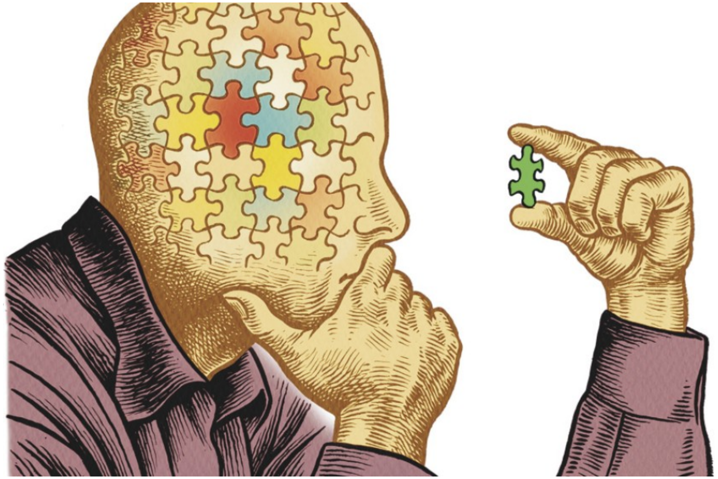
Outlines

- Challenges of understanding
- What is understanding?
- Understanding in science
- Number of relevant coordinates
- Learning
- Conclusions

Outlines

- Challenges of understanding
- What is understanding?
- Understanding in science
- Number of relevant coordinates
- Learning
- Conclusions

Challenges of understanding



“Deep in the human unconscious is a pervasive need for a logical universe that makes sense. But the real universe is always one step beyond logic.”

— Frank Herbert, Dune

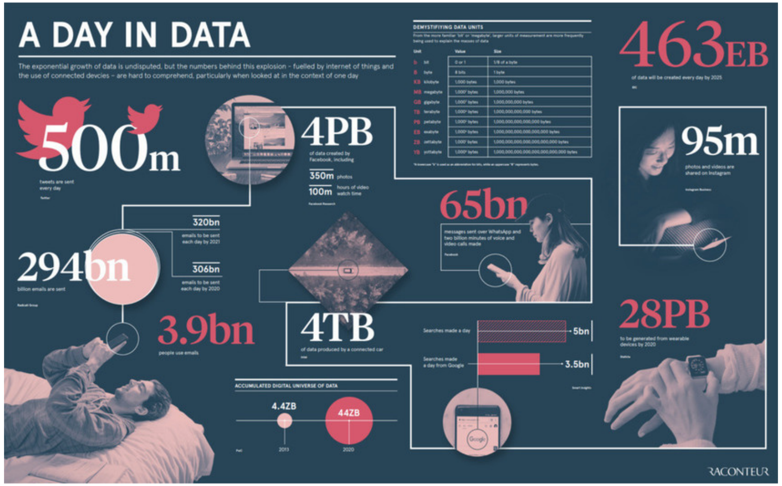
- Technical evolution, new tools, new developments makes necessary to rethink old concepts: e.g. “understanding”
- artificial intelligence understanding
- scientific understanding
- philosophy (knowledge vs. understanding, epistemology)

Big data / AI general view

Gigantic amount of data daily:

daily: $\sim 5\times10^{20}$ byte (CERN: $\sim 3\times10^{14}$ byte)
stored: $\sim 5\times10^{22}$ byte (~50 ZB)

We need organizing principles!



Known processes: statistics, models (calibration)

Unknown processes: learn systems with the help of artificial intelligence (AI)

AI successes and challenges

this person doesn’t exist!

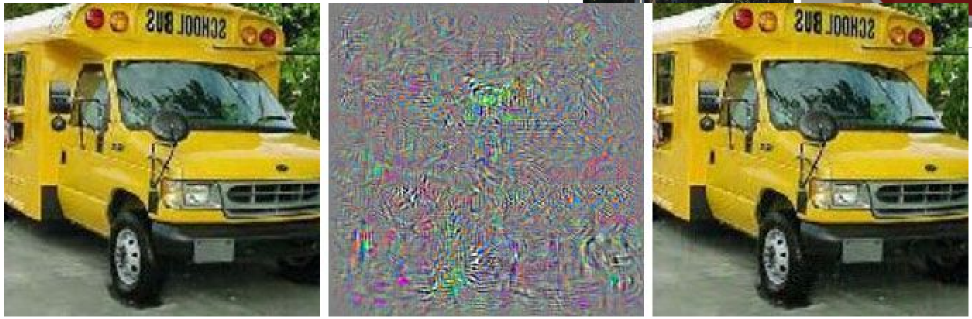
Successes

- table/computer games (chess, go, poker: AlphaZero, AlphaGo, Dota: OpenAI; Starcraft: DeepMind)
- image recognition
- image generation (face/animals/landscape Nvidia, StyleGAN2)
- NLP, translation (google translate)
- self driving cars

Challenges

- unexpected errors
- “adversarial attacks”
- long range correlation

does AI understand what it does?



AI is the scientific tool of XXI. century

(artificial intelligence, machine learning, etc. = AI)

Scientific approaches

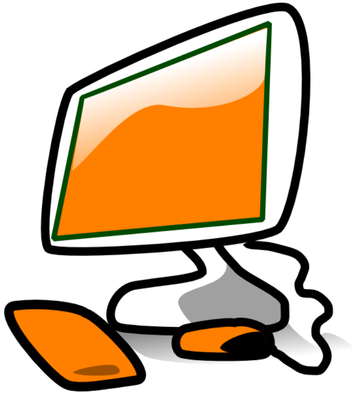
- **until mid XX. century:** scientists revealed the theory, using experiments, tried to solve the (differential) equations on paper with pencil
  - only linear systems are solvable (transformable to linear)
  - perturbation theory
  - coherent worldview: a generic problem is solvable at all scales (spectrum)

AI is the scientific tool of XXI. century

(artificial intelligence, machine learning, etc. = AI)

Scientific approaches

- **since mid XX. century**: new paradigm, scientists reveal the theory, using experiments, but the solutions are done with the help of computers
  - nonlinear systems are also solvable
  - lots of new phenomena (chaotic systems, phase transitions, etc.)
  - where is understanding? we see numbers, but what do they mean? (e.g. we can calculate proton mass from QCD, but we still do not have model for nuclear physics)
  - meaning of terms in a multiparameter model? (c.f. event generators)

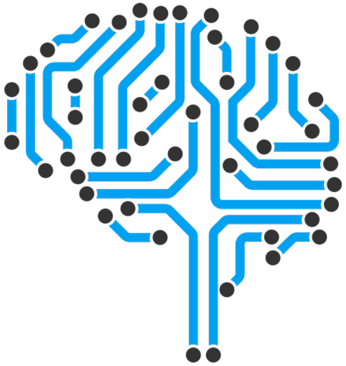


AI is the scientific tool of XXI. century

(artificial intelligence, machine learning, etc. = AI)

Scientific approaches

- **XXI. century**: both the theory and the solutions shall be done with the help of computers?
  - hope: synthesis/compression of big data,
    model generation for complicated systems
  - does anything remain from understanding?
    are the models understandable at all?
  - **clarifying the meaning of understanding is still more urgent...**



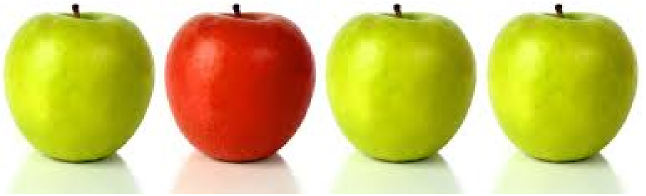
Outlines

- Challenges of understanding
- What is understanding?
- Understanding in science
- Number of relevant coordinates
- Learning
- Conclusions

What is understanding?

- What is the common of image recognition and QCD Lagrangian?
- What can be understood? (formula, book, joke, image)
- **is it uniquely defined?** e.g. what is in the picture?
  - colored pixels
  - paintbrushes, painting techniques
  - woman with a tabby cat in a room, …
  - artwork: Renoir: Woman with a cat
- *all are legitimate understanding of the same image*

Understanding: finding the common features



Loose definition of understanding:

- for a single image any descriptions are equally good
- **for a collection of images we should reveal what is common in them**
- then it is easy to
  - tell whether an image is an element of our collection (classification)
  - remember to any image in our collection (compression or decoding)
- ***understanding a collection: all the AI tasks are easy to perform***

Our goal is to make this loose definition mathematically precise.

Coordinatization of a finite set

- *how can we characterize an image up to a finite resolution?*
  - digitize, list the color of each pixel: N_{pixel} pixels and N_c colors
  - paint it with a thin brush, and record the process
  - start to list the content in human notions
    (breed of cat, color of the woman's skin, hair, wall, curtain, etc)
- in principle all can be represented by numbers
- all descriptions can be refined to give the same details!
  ⇨ **all are equivalent representations of individual images**
- X finite set, all possible images $|X|=N_c^{N_{pixel}}$
- all $\xi : X \rightarrow I^N$ **bijective maps (coordinatizations) are equivalent**
- *we mostly will use* $I=\{0,1\}$, $\xi_i(x)$ *are the coordinates,* $X \rightarrow I$ *"measurements".*

Statistics in subsets

- *joint distribution* of $i_1, i_2, \ldots, i_a$ coordinates over $C \subset X$ (δ are Kronecker-deltas)

$$p_\xi^C(\sigma_{i_1}, \sigma_{i_2}, \ldots \sigma_{i_a}) = \frac{1}{|C|} \sum_{x \in C} \delta(\xi_{i_1}(x) = \sigma_{i_1})\, \delta(\xi_{i_2}(x) = \sigma_{i_2}) \ldots \delta(\xi_{i_a}(x) = \sigma_{i_a})$$

- *expected value* of $q : I^N \rightarrow R$ coordinate function

$$E_\xi^C(q) = \frac{1}{|C|} \sum_{x \in C} q \circ \xi(x) = \sum_{\sigma \in I^N} q(\sigma)\, p_\xi^C(\sigma)$$

- ith coordinate is *independent*: $p_\xi^C(\sigma_i, \sigma_{i_2}, \ldots \sigma_{i_a}) = p_\xi^C(\sigma_i)\, p_\xi^C(\sigma_{i_2}, \ldots \sigma_{i_a}) \quad \forall a$

  or, equivalently, $E_\xi^C(\sigma_i \sigma_{i_2} \ldots \sigma_{i_a}) = E_\xi^C(\sigma_i)\, E_\xi^C(\sigma_{i_2} \ldots \sigma_{i_a}) \quad \forall a$
- a coordinate is *deterministic* in $C \subset X$ if $\forall x \in C,\ \xi_i(x) =$ constant (always independent)
- *uniform distribution* $p_\xi^C(\sigma) = \frac{1}{|I|} \sum_{v \in I} \delta(\sigma = v)$

Statistics in subsets

Examples:

- for $C=\{x\}$ containing a single element all coordinates are deterministic
- for $C=X$ complete set all coordinates are independent and uniformly distributed
- 3 bits example: $C=\{001, 010, 100, 111\}$
  - for all 3 bits: $p_\xi^C(\sigma)=\frac{1}{2}\left(\delta(\sigma=0)+\delta(\sigma=1)\right)$ uniform distribution
  - expected value of 3-bit correlation: $E_\xi^C(\sigma_1\sigma_2\sigma_3)=\frac{1}{4}\neq E_\xi^C(\sigma_1)E_\xi^C(\sigma_2)E_\xi^C(\sigma_3)=\frac{1}{8}$

pixels are not independent within C !

Lesson: *hidden, higher order correlations can appear in subsets!*

Corollary: elements of C are not easy to identify, pixel-wise information is not enough

Understanding subsets

Understanding ≡ finding a coordinatization where the coordinates are not correlated!

Technical definition of understanding of a single subset:

We understand a subset, if we give a coordinatization (*complete model*) where all coordinates are independent, and either deterministic (relevant), or uniformly distributed (irrelevant).

Technical definition of understanding of several disjunct subsets:

We understand a system of disjuncts subsets C_a (and their union $C = \cup_a C_a$), if we give a coordinatization (*common complete model*) where all coordinates are independent for all C_a and C, and are either deterministic, or uniformly distributed for any of the subsets (overall relevant, partially relevant, irrelevant).

Proof of existence

Does this coordinatization exist at all? Constructive proof:

- **special case:** assume that $|C_i| = 2^{N_i}$, $|C| = 2^{N_C}$ and $|X| = 2^N$ are all powers of 2. Order them to have $N \geqslant N_C \geqslant N_1 \geqslant \ldots \geqslant N_a$.
- Make an indexed list from elements of the subsets as

  (elements of C_1*, elements of* C_2*, … , elements of* C_a*, elements of* $X - C$*)*

- The index of list in binary numeral system gives the desired coordinatization, because the the C_i subset is placed from $2^{N_i} K$ to $2^{N_i}(K+1)$, and thus
  - the first $N - N_i$ bits are deterministic for C_i *(relevant bits)*
  - the remaining N_i bits are independent and uniformly distributed *(irrelevant bits)*.
- **for other cases:** we extend all the sets with dummy elements to achieve cardinality of power of 2 (practically affects only some bits)

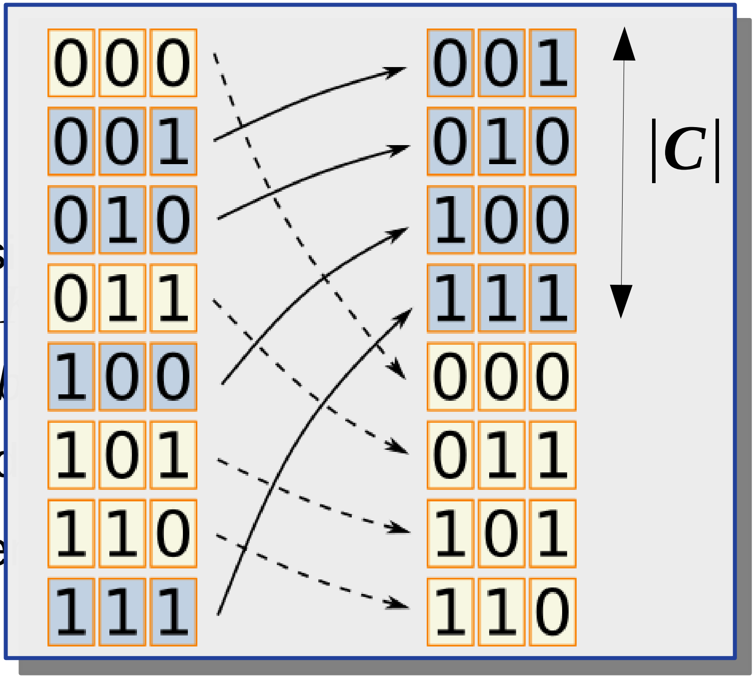
Proof of existence

Does this coordinatization exist at all? Constructive proof:

- **special case:** assume that $|C_i|=2^{N_i}$, $|C|=2^{N_C}$ and $|X|=2^{N}$ are all powers of 2. Order them to have $N \geqslant N_C \geqslant N_1 \geqslant ... \geqslant N_a$.
- Make an indexed list from elements of the subsets as

  (elements of C_1, elements of C_2, … , elements of

- The index of list in binary numeral system gives the des because the the C_i subset is placed from $2^{N_i}K$ to $2^{N_i}($
  - the first $N-N_i$ bits are deterministic for C_i *(relevant*
  - the remaining N_i bits are independent and uniformly
- **for other cases:** we extend all the sets with dummy ele of power of 2 (practically affects only some bits)



Proof of existence

Example: 4 bits, $C_1=\{0010, 0101, 1000, 1100\}$, $C_2=\{0110, 1111\}$, $C_3=\{0111, 1001\}$

index and ordered set: $set=\{0010, 0101, 1000, 1100, 0110, 1111, 0111, 1001, ...\}$

$index=\{0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111, ...\}$

corresponds to the permutation $0010 \rightarrow 0000, 0101 \rightarrow 0001, 1000 \rightarrow 0010, \ldots$

- for C_1: relevant bits first two (00), irrelevant the last two (uniform and independent)
- for C_2: relevant bits first three (010), irrelevant the last one (uniform)
- for C_3: relevant bits first three (011), irrelevant the last one (uniform)
- for C : relevant bit is the first (0), irrelevant the last three (uniform and independent)

How does it help understanding?

Let ξ be the common complete model for $C_{1,} C_{2,} \ldots , C_a , C$

- **classification:** $x \in C_i$ iff relevant bits of $\xi(x) =$ relevant bits of C_i
- **decoding:** to produce $x \in C_i$ we have to chose the relevant bits characteristic to C_i and the irrelevant bits independently, uniform randomly

$$\xi^{-1}(\sigma_{relevant} = C_{i,relevant} , \sigma_{irrelevant} = \mathrm{random}) \in C_i$$

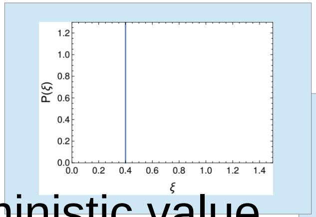
- **lossless data compression:** if we know that $x \in C_i$, the relevant bits can be built into the static part of the code, and we have to store the *irrelevant bits*.

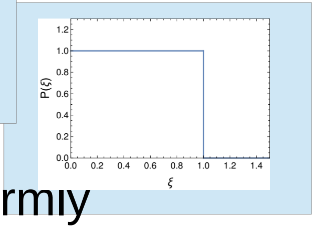
All the AI tasks can be solved by inspecting certain bits.

Continuous approximation

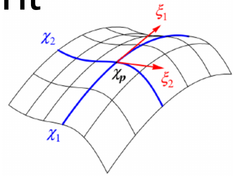
We can map n_b binary numbers to a floating point number with n_b significant binary figures.

$$x=\sum_{i=1}^{n_b}\frac{\sigma_i}{2^i}, \qquad x\in[0,1]$$





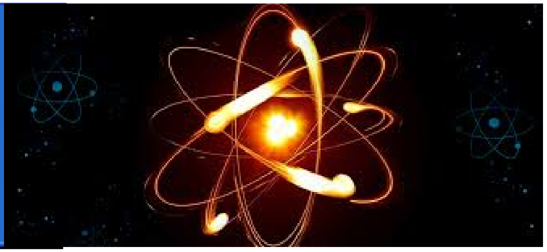
- *relevant coordinates*: floating point number will have a deterministic value
- *irrrelevant coordinates*: floating point coordinates will be independent, uniformly distributed in $[0,1]$
- we can start with real number representation, and look for a curvilinear coordinate system where the coordinates are either deterministic (relevant) or independent and uniformly distributed in $[0,1]$



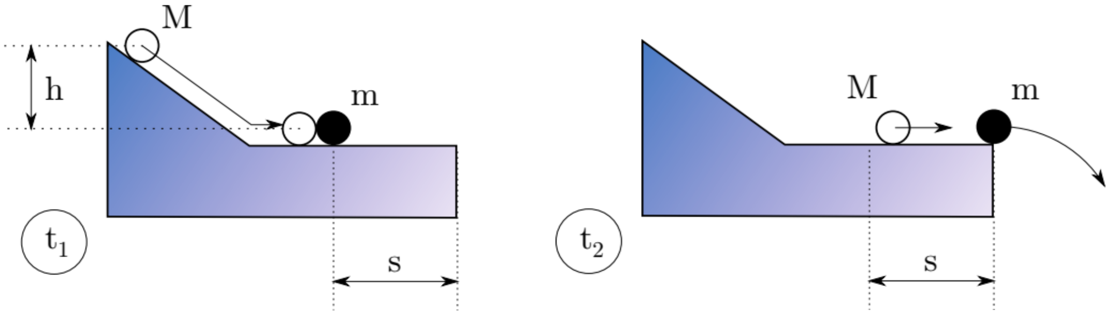
Outlines

- Challenges of understanding
- What is understanding?
- Understanding in science
- Number of relevant coordinates
- Learning
- Conclusions

Understanding in science



Example: measure the time difference between collision and falling from table

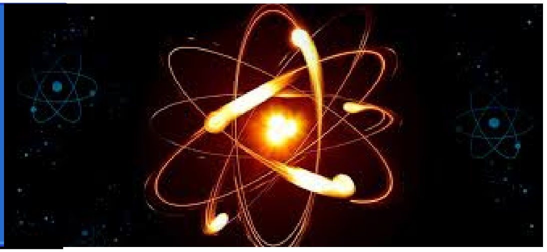


- subsets: experiments with the same measured value $C_i = C_{\Delta t}$
- *overall relevant coordinates*: experimental setup (solid balls, slopes, …)
- *overall irrelevant coordinates:* do not influence the result (color, experimenter, ...)
- *partially relevant quantities for* $C_{\Delta t}$: changing from experiment to experiment, and influence the result

$$\Delta t = \frac{s}{2\sqrt{2gh}}\left(1 + \frac{m}{M}\right)$$

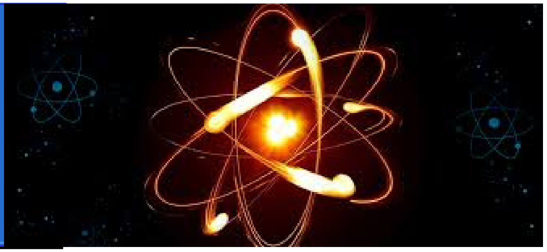
these yield the usual concepts of mechanics (position, acceleration, mass, ...)

Understanding in science



- *To be understood*: natural processes / states
- *original characterization:* collect "facts", all possible measurements (giving the measurement operator)
- *single out a measurement* M (can consist of several elementary measurements), this chooses the "phenomenon" which we want to understand
- C_i^M sets are **indexed by the measured value** of M
- *find a common complete model …*
  - overall relevant quantities: define the stable environment / discipline
  - irrelevant quantities: do not influence any measurement
  - *partially relevant quantities*: these are the **physical quantities**, these influence the measurements

Understanding in science



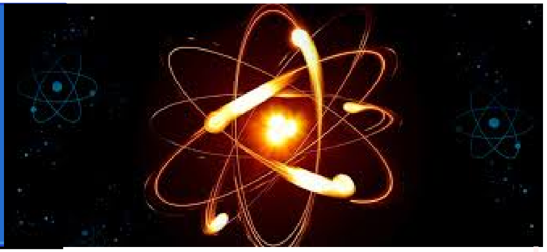
Natural law: we may realize that we have an overcomplete basis

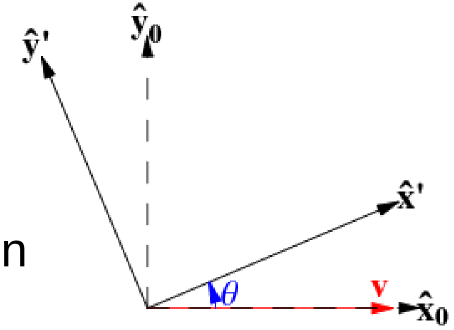
there are relations between the relevant quantities, e.g. $F = ma$

Units of measure: certain quantities appear together in measurable quantities, we can associate a unit to them.

Most important step is the distillation of **the relevant coordinates**, once we found them, their relation can be studied.

Symmetries and bases

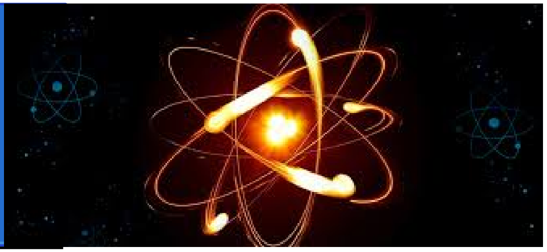




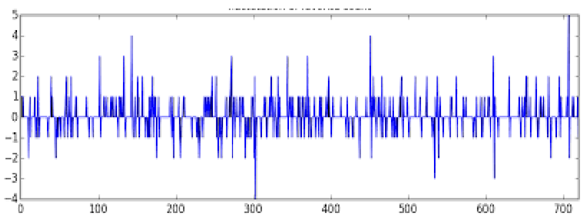
Relevant coordinates need basis, but it is not unique!

- space coordinates have no meaning, because we can make rotation
- in general global symmetries redefine the representation
- local (gauge) symmetries: “gauge fields at a given point” has no absolute meaning
- not the parameters of the Lagrangian matter, they are not directly measurable
- in general: **form of Lagrangian does not matter**, the same reality can have different representations
- examples: event generators, trained DNNs.

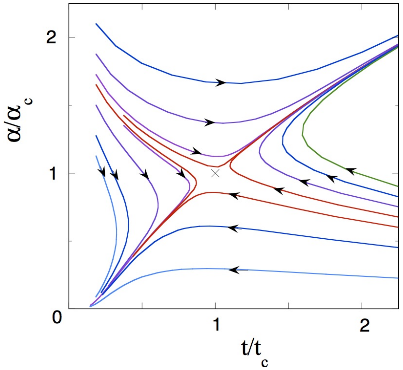
Precision and irrelevance



Subsets to understand: C_i^M indexed by the measurments of M.



- ***discrete representation***
  - *high precision*: all C_i^M consist of a single element: all coordinates relevant
  - *lower resolution* (i.e. change M): more and more irrelevant coordinates
  - *measure of relevance*: when a coordinate starts to be irrelevant
- ***continuous representation****:* resolution can be set by averaging on scale k
  - $\frac{\partial M}{\partial x_i} = g_i$ "couplings" depend on the scale (renormalization group)
  - *measure of relevance*: decrease for irrelevants, increase for relevants

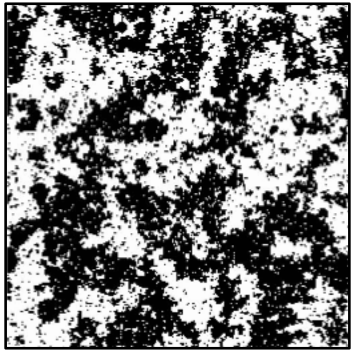


Outlines

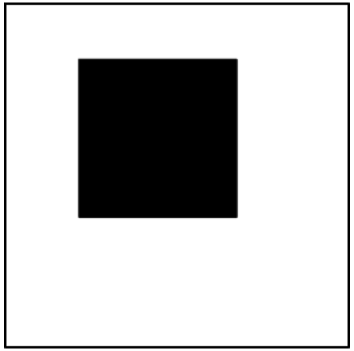
- Challenges of understanding
- What is understanding?
- Understanding in science
- Number of relevant coordinates
- Learning
- Conclusions

Number of relevant coordinates

Consider black-and-white images, $X=\{0,1\}^N$, and two subsets:

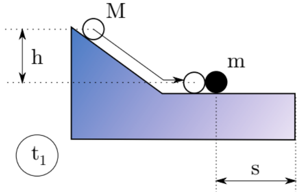


- **Ising-configurations** (with measured energy and IR correlators)
  - in IR limit the correlators are not independent
  - only **three relevant** quantities (β, J, h)
  - all other are irrelevant, irrespective of the image size
- **black square on white background**
  - very ordered image
  - compressible with top left corner coordinates and edge length
  - thus only **three irrelevant** quantities (x, y, a)
  - all other are relevant, irrespective of the image size

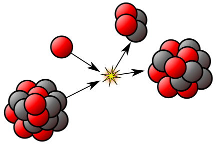


Number of relevant coordinates

Spectrum in number of relevant coordinates



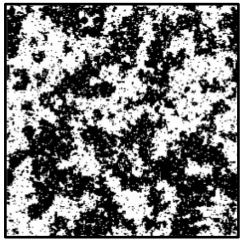
point mechanics
~ 5 relevant



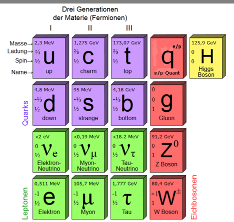
nuclear physics
20-? relevant

chemistry, biology
~ 100-? relevant

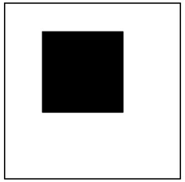
natural environment
? relevant ? irrelevant



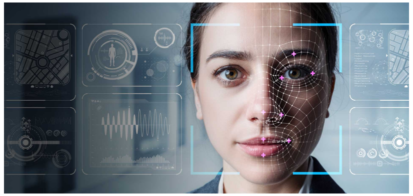
Ising model
3 relevant



Standard Model
21 relevant
(symmetries!)



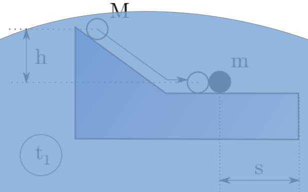
geometric images
~ 10-100 irrelevant



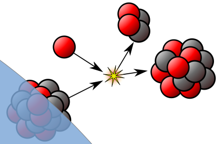
face recognitions
~ 30000 irrelevant

Number of relevant coordinates

Spectrum in number of relevant coordinates



point mechanics
~ 5 relevant

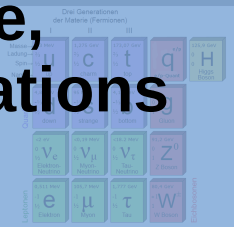


nuclear physics

chemistry, biology
~ 100-? relevant

natural environment
? relevant ? irrelevant

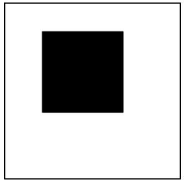
Few relevant quantities, identifyable, exact laws, relations



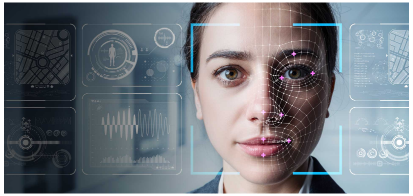
SCIENCE

Ising model
3 relevant

Standard Model
21 relevant
(symmetries!)



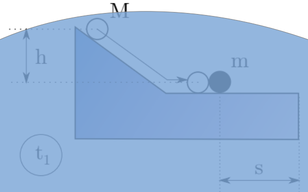
geometric images
~ 10-100 irrelevant



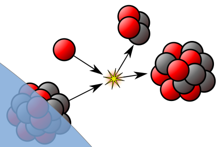
face recognitions
~ 30000 irrelevant

Number of relevant coordinates

Spectrum in number of relevant coordinates



point mechanics
~ 5 relevant

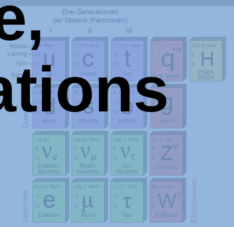


nuclear physics
~ 20-? relevant

chemistry, biology
~ 100-? relevant

natural environment
? relevant ? irrelevant

Few relevant quantities,
identifyable,
exact laws, relations



Ising model
3 relevant

SCIENCE

Standard Model
21 relevant
(symmetries!)

Lot of relevant quant.
unidentifyable,
similar role,
no exact description

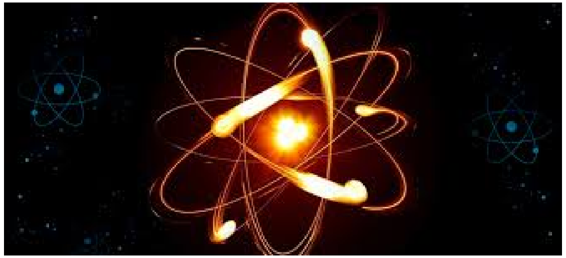
(ARTIFICIAL)
INTELLIGENCE

geometric images
~ 10-100 irrelevant

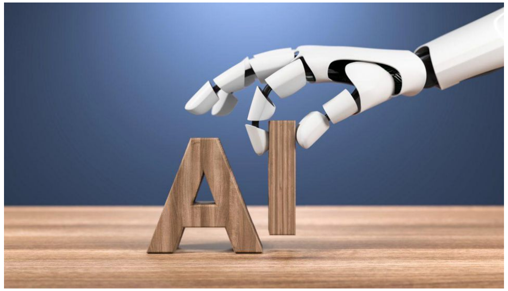
face recognitions
~ 30000 irrelevant

Number of relevant coordinates

According to the number of relevant quantities the nature of exploration changes



- **few relevant coordinates**
  - all terms are important, different (we even can name them: mass, position, gauge coupling, …)
  - recognition strategy: know and study them one-by-one
  - exact laws, understandable: **SCIENCE**



- **lot of relevant coordinates**
  - exact definition is not important, unidentifyable one-by-one (e.g. can be replaced by similar)
  - recognition strategy: train them together
  - laws are heuristic: **(ARTIFICIAL) INTELLIGENCE**

Science using AI

What to do with intermediate systems
(nuclear physics, biology, finance, meteorology, …)?

- we can try and approach it scientifically: find most important, identifyable parts
- **trust the rest to AI:** *(one of our present main projects)*
  - time evolution in coarsely discretized systems
  - Hamiltonian of complicated systems (hadronization, meteorology, finance, noise filtering)
  - train general DNN with examples from data (or exact simulation)
  - will not know the exact law, but better than ad hoc “scientific” guesses

Outlines

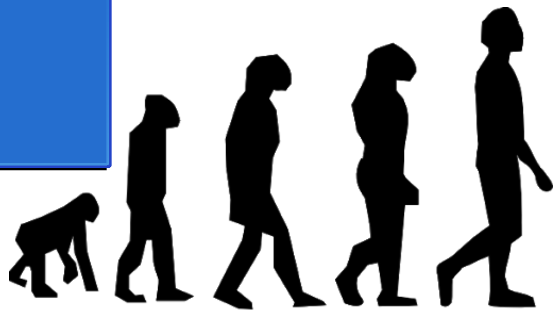
- Challenges of understanding
- What is understanding?
- Understanding in science
- Number of relevant coordinates
- Learning
- Conclusions

Learning from an expert

How can we train a multiparameter system?

- **show annotated examples:** supervised learning, reinforced learning
  - needs a huge number of input-output pairs
  - better to show them in random order
  - all neurons develop towards the common goal (backpropagation, gradient)
- **unsupervised techniques:** generative adversarial networks (GANs), variational autoencoders (VAEs)
  - VAE: goal is to reproduce the input after a bottleneck
  - GANS: generator tries to generate inputs, discriminator tries to tell apart the generated and real inputs

Gradual learning, measure of relevance, lossy compression



In reality inputs change continuously

- three types of relevance after long time; completely relevant: never change, irrelevants: change slowly, varies fast
- ***characteristic changing time: another measure of relevance***. Relevant never change, weakly irrelevant slowly change, strongly irrelevant fast change
- importance from the point of view of recognition:
  - relevant: not important, since they are constant
  - strongly irrelevant: not important, too fast and small changes in environment
  - **weakly irrelevant**: these are the most important coordinates (lossy compression)
- **learning strategy**: omit too steady or too volatile parts of the input; can be done on neuron level

Outlines

- Challenges of understanding
- What is understanding?
- Understanding in science
- Number of relevant coordinates
- Learning
- Conclusions

Conclusions

Understanding ≡ finding the relevant / irrelevant coordinates

- independent, either deterministic (relevant) or uniform (irrelevant)
- all AI tasks can be easily performed with looking at some coordinates classification: relevants, regression, compression: irrelevants
- *number of relevant coordinates* can vary vastly: from few (Ising model) to almost all (geometric images); inherent property, not a bug
- **scientific approach** is good for few relevant coordinates: all are important, have meaning, can be studied one-by-one, exact laws
- **intelligent approach** is good for plenty of relevant coordinates: not important one-by-one (symmetric), only cumulative effect important (training), heuristic laws
- **mixed strategy to approach intermediate systems**

Vége