Is Artificial Neural Network Intelligent?

Akira Imada

Brest State Technical University (Belarus)

Is AI Intelligent?

A history of arguments between pro- & anti-AI parties.

Harsh criticism to AI by Dreyfus

Alchemy and Artificial Intelligence (1965)

$\Downarrow$

What Computers Can't Do? (1972)

$\Downarrow$

What Computers Still Can't Do. (1992)

Alchemy and Artificial Intelligence (1965)

The field of AI exhibits a recurrent pattern:
early dramatic success followed by sudden unexpected difficulties.

"The first man to climb a tree could claim tangible progress toward reaching the moon."

Rebuff by Papert

"The Artificial Intelligence of Hubert L. Dreyfus: A Budget of Fallacies."(1965).

Rebuffs by the other big names in AI

Edward Feigenbaum:

"What does he offer us? Phenomenology! That ball of fluff. That cotton candy!"

Marvin Minsky:

"They misunderstand, and should be ignored."

(Actually no response after the 3rd edition.)

"Machine who thinks" by Pamela McCorduck (1992)

Papert:

"For Dreyfus, all social sciences are as wrong-headed as AI. This is not an attitude widely held in universities."

McCorduck:

"If Dreyfus is so wrong-headed, why haven't AI people made more effort to contradict him?"

Rodney Brooks (1991)

"Artificial intelligence started as a field whose goal was to replicate human level intelligence in a machine.

$\Downarrow$

Early hopes diminished as the magnitude and difficulty of that goal was appreciated.

$\Downarrow$

No one talks about replicating the full gamut of human intelligence any more."

Can computer play chess?

Dreyfus (1960):

"Program written by the Newell, Shaw, and Simon played legal chess, but was beaten in 35 moves by a ten-yea-old novice. "

'The New Yorker' (11 January 1966)

The Talk of the Town:

Computer was beaten in chess by a ten-yea-old novice.

... We don't care what the machine is going to do.

The New Yorker, Jun 11, 1966 Стр. 1 из 1

(ICN

THE NEW YORKER

THE TALK OF THE TOWN

Notes and Comment

FOR months, we've been uneasily eying a subway advertisement that pictures a complicated-looking electrical unit and inquires, rather belligerently, "When this circuit learns your job, what are you going to do?" Now, we know that the advocates of skill-retraining aren't trying to scare us, nor have any strange new machines moved into our office lately, but we *have* been wondering just how rampant automation is likely to become. For one thing, philosophical-minded friends have taken to spoiling parties for us by arguing over when the activity of a machine may legitimately be called "thinking." (They all assume, of course, that sooner or later science will build robots whose behavior will be indistinguishable from human behavior.) We've taken only small comfort from humanistic debaters who hold that, however smart a mechanical brain may become, it will never be capable of man's most sophisticated acts of mind, such as creative abstraction, moral choice, and falling in love. For even if machines can't achieve "consciousness" in the human sense, it's discomfiting to think of them grinding away at the very gates of our highest capacities. Wrestling with these lofty conceptual problems, our friends have apparently not thought to ask how well a digital computer (the only high-speed all-purpose information-processing device in existence) can manifest the simplest kinds of intelligent behavior. It stands to reason that before any mélange of wires and tubes occupies our desk and sends us scurrying off to "retrain for tomorrow's jobs" (whatever they may be) it will have mastered at least the rudimentary intellectual activity characteristic of children and, in some cases, animals—playing games, recognizing patterns, solving easy problems, reading sentences. With the help of the press, a few noisy researchers in the field of artificial intelligence have fostered the impression that such modest feats can indeed be performed by machines today. A wellspring of this scientific mythology seems to be a historic talk delivered in 1957 by H. A. Simon, one of the grandfathers of artificial intelligence. "It is not my aim to surprise or shock you—if, indeed, that were possible in an age of nuclear fission and prospective interplanetary travel," Mr. Simon said. "But the simplest way I can summarize it is to say that there are now in the world machines that think, that learn, and that create. Moreover, their ability to do these things is going to increase rapidly until—in a visible future—the range of problems they can handle will be coextensive with the range to which the human mind has been applied." Mr. Simon went on to predict that within ten years a digital computer would (a) win the chess championship of the world (unless barred, by rule, from competition), (b) discover and prove an important new mathematical theorem, (c) write music praised by critics, and (d) programmatically express most theories in psychology.

Well, we've just come across a lovely paper by Hubert L. Dreyfus, a professor at the Massachusetts Institute of Technology, which says computers can't, and won't. With Mr. Simon's decade almost up, we learn, a recurrent pattern has plagued artificial intelligence in every field it tackles: dramatic early success followed by unforeseen problems and then by disenchantment. In game playing, for instance, researchers developed a checkers program, about ten years ago, that was able to beat an ex-champion from Connecticut. In chess, however, where the number of possible moves and responses is so much greater, computer programs bogged down in the problem of exponential growth. A computer's attention cannot be attracted by areas on the board that look interesting. It cannot zero in on possibilities that appeal to a sort of "fringe consciousness." It can only count out alternative moves on an ever-branching tree of possibilities. At about the time of Mr. Simon's grand prognostication, a group at Los Alamos devised a chess program that played an inferior, though legal, game on a reduced board. Ever since that program beat one weak opponent, the forecasts of impending master play have grown increasingly emphatic, but no computer developed in the intervening years has failed to play a stupid game. A highly publicized program, in its latest recorded bout, was defeated in thirty-five moves by a ten-year-old novice. Yet the projected world championship is only a year off. The quest for a new mathematical theorem has a similar history. An ingenuous reader of W. R. Ashby, a leading authority in the field of computers and thought, might assume that Mr. Simon's prophecy about the important new theorem has already been realized. Professor Ashby recently hailed a particular program's powers: "Gelernter's theorem-proving program that discovered a new proof of the *pons asinorum* that demands no construction"—a proof which "the greatest mathematicians of 2000 years have failed to notice" and which "would have evoked the highest praise had it occurred." A glance at D. E. Smith's "History of Mathematics" (published

"Look, I've told you! We won't take the helicopter back to the Pan Am Building if it scares you. We'll take a taxi."

in 1925) reveals that the *pons asinorum*, or asses' bridge, is none other than the first theorem to be proved in Euclidean geometry—that opposite angles of an isosceles triangle are equal—and that the machine's "new" proof was originally introduced by Pappus in 300 A.D. Computers have still made no great impression as composers, and hopes for high-quality mechanical language translation have been frustrated, too, despite government-research investments of about sixteen million dollars, for computer programs have foundered in unsuspected ambiguities of syntax and semantics. (A short while back, one of the best in the field was asked to translate "Time flies like an arrow" from Russian into English and came up with "Time flies enjoy eating arrows.")

In the face of these and parallel stymies, workers in cognitive simulation refuse to be discouraged. Mr. Dreyfus compares their pertinacity to that of the alchemists, who, after initial triumphs in distilling quicksilver from what seemed to be dirt, labored fruitlessly for several hundred years to transmute lead into gold, feeling themselves to be ever on the verge of a breakthrough. By defining "progress" as "displacement toward the ultimate goal," today's alchemists, the cognitive-simulation workers, obscure the prospects for artificial intelligence. According to this definition, Dreyfus points out, the first man to climb a tree could claim progress toward flight to the moon. He suggests that the workers' unrelenting optimism is founded on the unfortunate Cartesian assumption that human information-processing must be analyzable into "simple determinate operations," or discrete steps, much like those achieved by a digital computer. If, in reality, "thinking" entails more mysterious kinds of mental activity as well, there are a great many offices that machines won't ever be moving into. Some weeks ago, in a pleasant turnabout, a collection of scientists at a Symposium on Computer Usage agreed that machines will forever be confused by complexity. "Computers can add and subtract one million numbers in one second," Dr. Warren S. McCulloch, of M.I.T., explained. "No man can do this, but man can slip into any of fourteen different modes of action—from sleeping to fighting—in about three-tenths of a second. He has about one trillion computing neurons that bring together two million separate biological components all at once."

So while cybernetics enthusiasts burble happily over the success of Compatibility Research, the computerized matchmaking service that is now pairing off college students on an international basis, we take some quiet satisfaction in the Lincoln Center Repertory Theatre's most recent stride toward modernization. An I.B.M. 1401 computer made such a botch of handling ticket requests that extra human help took forty-five straight days, working twelve hours a day, to straighten it out. In consequence, Robert Schlosser, subscription supervisor, dispatched to subscribers a letter that may be a harbinger. In it he announced the installation of "an improved method" for handling subscriptions—"substituting a manual process for the automated process in use until now." Now that the computer has been fired and some competent people have learned the machine's job, we don't care what the machine is going to do.

THE truffles at the Griststede's on University Place are kept in the cash register.

City Fragments

WE went over to the Brooklyn Museum the other day to see a new outdoor sculpture garden containing fragments of various nineteenth-century and early-twentieth-century buildings that have lately been torn down in the five boroughs of New York. The garden, which is the first of its kind, owes its existence in large part to the Anonymous Arts Recovery Society, a group of dedicated scavengers who for the past seven years have managed to save bits and pieces of architectural history by bribing wrecking crews, sifting mountains of debris, carting away heavy stones, and, when necessary, raiding demolition sites at night. The Museum has now rescued the rescuers, so to speak, with its garden, which occupies three-quarters of an acre in what was formerly its back yard. Built with funds set aside by the late Mr. Walter N. Rothschild for a memorial to his

http://archives.newyorker.com/global/print.asp?path=/djvu/Conde Nast/New Yorker/1... 2010/05/08

Yet another topic on machine translation

'Time flies like an arrow,' in Russian

$\Downarrow$

'Time flies enjoy eating arrows,' in English.

A chess match: Dreyfus vs. Computer

Papert:

"I organized the famous chess match. That was beautiful."

In a bulletin board in SIG AI

Papert wrote:

'A ten-year-old can beat the machine' – Dreyfus:

But the machine can beat Dreyfus.

Can computer play chess? [II]

Garry Kasparov vs. IBM's Deep Blue.

⇓

(in 6 game match)

1996:	Deep Blue	Kasparov	even	even	Kasparov	Kasparov
1997:	Kasparov	Deep-Blue	even	even	even	Deep Blue

Is Deep Blue intelligent?

Deep Blue's strategy

$\Downarrow$

A brute force to evaluate billions of future positions.

What should intelligence look like?

Dreyfus (1965)

"A little intelligence is not intelligence at all but stupidity.

Any program that does just one thing well is

at best more like an idiot savant than like an intelligent man."

Should intelligence be perfect?

Brooks (1991):

"It is unfair to claim that an elephant has no intelligence worth studying just because it does not play chess."

Contradiction & Intelligence?

Frosini (2009):

Contradiction can be seen as a virtue rather than as a defect.

$\Downarrow$

Constant presence of inconsistencies in our thoughts.

$\Downarrow$

Is contradiction accidental or is it the necessary companion of intelligence?

What about intelligence by NN?

McClelland (2009)

The author of 'Parallel Distributed Processing' (1986)

"Is a machine realization of truly human-like intelligence achievable?"

What is human-like intelligence?

Human-intelligence is spontaneous, flexible , and/or unpredictable, more or less.

Or even erroneous sometimes.

“I beg your pardon?”

Intelligent people try a different explanation for an easier understanding

while

others just repeat the same expression, maybe louder.

Let's make an agent be spontaneous!

Behaviors might differ even in a similar situations.

To see if it's possible or not

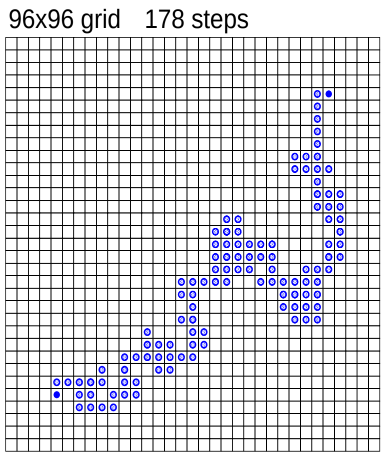
let's experiment

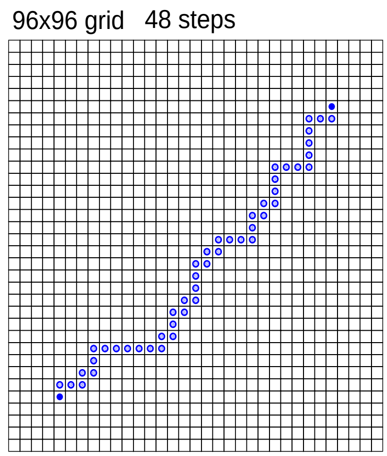
in a situation as simple as possible for an essential observation.

Path-planning as a benchmark

An agent plans a shortest route from start to goal
in a grid-world.

Examples of path from start to goal





Which route to be chosen?

The number of shortest paths from (0,0) to (m,n)

⇓

infinitely large.

⇓

$$\sum_{i=0}^{m+n} {}_mC_i \times {}_nC_{m+n-i}.$$

Can an agent take a different path from run to run?

What about from (0,0) to (m,0)?

In this case only one unique solution.

$\Downarrow$

Can an agent still be flexible from run to run?

Goal is

to make an agent behave differently even when it encounters the same situation as before.

NN with fixed weights

$\Downarrow$

never be intelligent

but only repeats exactly the same action in the same situation.

Floreano's approach (2000)

With McCulloch & Pitts neurons

$\Downarrow$

Modification of w_{ij} during actions of agent

with either one of the four Hebbian and Hebbian-like rules.

Let's repeat his experiment

with spiking neurons expecting to be more biologically plausible.

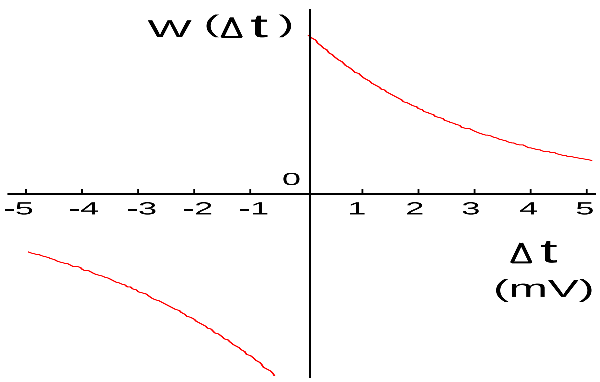
Spike Timing Dependent Plasticity (STDP)

A counter part of Hebbian rule for spiking neurons

What is STDP?

$$W(\Delta t) = \begin{cases} A_+ \exp(-\Delta t/\tau_+) & \text{if} \quad \Delta t \geq 0 \\ -A_- \exp(-\Delta t/\tau_-) & \text{if} \quad \Delta t < 0 \end{cases}$$

where $\Delta t = t_{post} - t_{pre}$



In short

potentiation occurs when a pre-synaptic neuron fires
shortly before a post-synaptic neuron

and

depression occurs when the post-synaptic neuron fires shortly after.

Meunier et al. (2005)

$\Downarrow$

"Up to now, nobody has been able to show how it is possible to learn with STDP..."

Farries et al. (2007)

$\Downarrow$

"Although synaptic plasticity is widely believed to be a major component of learning, it is unclear how STDP itself could serve as a mechanism for general purpose learning."

Two implicit implementations of STDP

- Di Paolo (2002)
  - Recurrent neural network with evolved STDP
  - Conductance-based integrate-and-fire (I&F) model.
- Florian (2005)
  - Feedforward neural network with reward-modulated STDP
  - Stochastic leaky I&F neurons.

Conductance-based I&F model

Membrane voltage $v(t)$ is:

$$\tau_m \frac{dv(t)}{dt} = V_{rest} - v + g_{ex}(t)(E_{ex} - v) + g_{in}(t)(E_{in} - v).$$

Di Paolo's implicit STDP

Using two functions $P^+(t)$ and $P^-(t)$

- When no firing occurs, they decay exponentially: $\tau \frac{d}{dt}P(t) = -P(t)$
- Every time a spike arrives at the synapse $P^+(t)$ is incremented by A^+

$$w_{ij}(t) \rightarrow w_{ij}(t) - w_{max}P^+(t)$$

- Every time the post-synaptic neuron fires $P^-(t)$ is decremented by A^-

$$w_{ij}(t) \rightarrow w_{ij}(t) - w_{max}P^-(t)$$

Stochastic Leaky Integrate & Fire Model

Membrane voltage of neuron i is

$$u_i(t) = u_r + (u_i(t - \delta t) - u_r) \exp(-\delta t / \tau) + \sum_j w_{ij} f_j(t - \delta t)$$

Reward-modulated STDP Learning

(Florian 2007)

$$w_{ij}(t+\delta t) = w_{ij}(t) + \gamma r(t+\delta t)\zeta_{ij}(t)$$

where

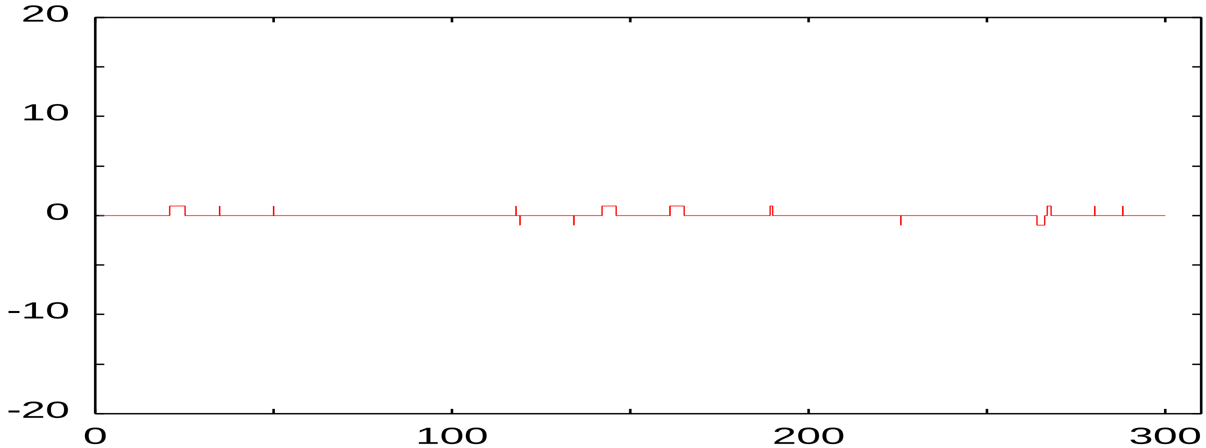
$$\zeta_{ij}(t) = P_{ij}^{+}(t) f_i(t) + P_{ij}^{-}(t) f_i(t)$$

$$P_{ij}^{+}(t) = P_{ij}^{+}(t-\delta t)\exp(-\delta t/\tau_{+}) + A_{+} f_j(t)$$

$$P_{ij}^{-}(t) = P_{ij}^{-}(t-\delta t)\exp(-\delta t/\tau_{-}) + A_{-} f_j(t)$$

Are we happy?

Simple heuristics can do it using a dice



Papert (1965)

"A very simple algorithm can sometimes obtain the same results as the holistic, intuitive human mind,"

Not sufficient to be intelligent

A different-action-even-in-an-identical-situation

$\Downarrow$

just a necessary condition at the best.

What else we need to be intelligent?

A spontaneous, flexible or unpredictable behavior

should be done consciously

Izhikevich (2006)

defined consciousness as *attention to memory.*

'Science of the conscious mind'

Ascoli et. al (2008)

Cognitive maps are made up of contexts,
such as spatial location mammals employ for their path-finding
using hippocampus.

Navigation by hippocampus

Muller's simulation (1996)

- Assume mapping from

  points in 2-D field where rat explores

  to

  pyramidal cells in a recurrent network of CA3.

- Mapping is one-to-one but randomly assigned points in 2-D space.
- Distance relation in 2-D space is stored as weight configuration.
- The shortest path in neural space is also shortest in 2-D space?

Is navigation by hippocampus intelligent?

- Merriam's kangaroo rats can learn the distribution of food patches around its nest in three evenings of foraging;
- Marmoset monkeys reliably relocate food sites and do not revisit a place where food was already eaten on that foraging trip;
- Black-capped chickadees hide insects and seeds in numerous, widely spread caches in trees over its home range.

Elephant cannot play chess, but...

It might not sound like an intelligent behavior, but as already quoted Brooks (1990), an elephant could be intelligent even if it cannot play chess.

Belief, desire, intention, and emotion

Bratman (1967)

The belief-desire-intention (BDI) model to construct multi agent system

or

Pereira et al. (2005)

a model of emotional BDI agents.

$\Downarrow$

These might help us design a NN closer to human intelligence.

Why spontaneous?

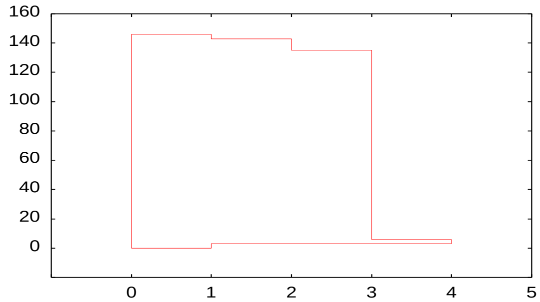
We sometimes need spontaneous and unpredictable intelligence rather than efficiency or effectiveness like in case of SONY's AIBO.

It learns excellently and acts differently in different situation but repeats the same action in a same situation.

McClelland (2009) again

"Over the next decade, the butterfly will finally emerge from the chrysalis, and truly parallel computing will take flight."

Mars Landrover Problem



Can we design a robot such that it navigates flexibly enough to take a different route from run to run, using a memory with some conscious intention, hopefully with belief and some sort of desire?

Thanks!

- Assume you like Chopin then you also like Chopin from CD?

- Assume you like Chopin then you also like Chopin from CD?
    - Different action even in a similar situation.

- Assume you like Chopin then you also like Chopin from CD?
  - Different action even in a similar situation.
- Is elephant intelligent or idiot?
  - Should intelligence be such perfect?

- Assume you like Chopin then you also like Chopin from CD?
  - Different action even in a similar situation.
- Is elephant intelligent or idiot?
  - Should intelligence be such perfect?
- Expert of just one thing is not intelligent but an idiot savant?