

IV130 Přínosy a rizika inteligentních systémů

Možnosti a limity strojů

Jiří Zlatuška

8. března 2024

Podstata výpočtů/výpočetních procesů; pochození mezi možností výpočtů

- Výpočty / výpočetní procesy motivovány kroky při formálních matematických/logických vývodech
- Z Gödelových výsledků známe nemožnost formulovat mechanickou/logickou bázi pro celou matematiku
- Výpočty lze popsat pomocí logických (nikoli matematických) vyvozovacích pravidel – jako zjednodušování struktury důkazů
- Meze možností výpočtů: nerozhodnutelné problémy
- Entropie a informace: od mezí účinnosti partních strojů po meze komunikací a výpočtů
- Fyzikálně dané meze možností výpočetních procesů

Vyčíslitelnost a nerozhodnutelné problémy

- 1936: Alonzo Church dokázal pomocí λ -kalkulu a Alan Turing o několik měsíců později pomocí Turingových strojů, že **odpověď na Hilbertův problém rozhodování (Entscheidungsproblem – odpověď na otázku, zda je formule dokazatelná v daném kalkulu) je negativní**
- Existence *nerozhodnutelných problémů* v informatice
- Churchova-Turingova teze (Kleene, 1952) – formalizace vyčíslitelných funkcí / vyčíslitelnosti jsou rovnocenné, se stejnou vyjadřovací silou

Termodynamika a (fyzikální) šipka času

- Zákony newtonovské fyziky/kvantové mechaniky jsou reverzibilní vzhledem k času; změny fyzikálních stavů v nich nemají orientaci vzhledem k roku času
- Tok tepla je ale časově orientován; jeho chování není symetrické vůči času, ale obsahuje směrovost
- Rozdíl teplot lze využít pro konání práce: parní stroj se stal základem průmyslové revoluce
- Vynaložením práce lze tok tepla obrátit (princip chladničky nebo tepelné pumpy)
- Teplo jako energie chaotického pohybu molekul
- Entropie jako míra neuspořádanosti v systému
- Druhý termodynamický zákon: celkové entropie v izolovaném systému nemůže v čase nikdy klesnout
- Překážka spontánního uspořádání rychlých a pomalých molekul v plynu s rovnoměrným rozložením teploty (a tedy zákaz vytvoření tepelného gradientu, který by umožňoval z neuspořádaného pohybu vytěžit práci)
- Tok času je dán „šipkou času“ podle Druhého termodynamického zákona, v důsledku čehož dochází k nevratným procesům

Informace a entropie - fyzika

- James Clerk Maxwell ve své práci z roku 1871 “Theory of Heat” navrhuje démona třídícího pomalé a rychlé molekuly a tedy přenášejího teplo z chladného na teplejší – ukazuje se, že zpracování informací o rychlosti molekul vede k vykonávání práce a nemožnost takového démona souvisí s mezemi pro nakládání s energií v počítačích
- Leo Szilard v roce 1922 (publikováno 1929), “On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings”, vyvodil vztah třídění molekul a práci s pamětí a měřeními a navrhl „Szilardův stroj“ převádějící teplo z jeho okolí na práci s poklesem entropie odpovídajícím informacím o dané molekule (předešel Shannona, ale v podstatě bez povšimnutí).

Informace a entropie - komunikace

- Claude Shannon: “fundamentální problém komunikace” jako schopnost příjemce reprodukovat zdrojovou zprávu poslanou odesilatelem přes komunikační kanál (zatížený šumem)
- Práce z roku 1948 “The Mathematical Theory of Communication”
- Informace jako míra odlišnosti
- Entropie jako míra nejistoty o predikci stavu ($\log_2(n)$ bitů pro n hodnot, s entropií předpokládající stejnou pravděpodobnost pro každý z nich)

Termodynamika a informace

- Druhý termodynamický zákon (celková entropie nemůže v izolovaném systému nikdy v čase klesnout)
- Rolf Landauer (1961): „jakákoli logicky nevratná manipulace s informacemi musí být doprovázena odpovídajícím vzrůstem entropie příslušného zařízení na zpracování informací nebo prostředí, v němž pracuje“
- V polovině 19. století vytvořili fyzikové zkoumající fundamentální meze účinnosti *parních strojů* termodynamiku jako vědní disciplínu.
- Shannon rozšířil analýzu založenou na entropii na *komunikace*.
- Landauer udělal podobné rozšíření pro *výpočty*.
- Výsledkem toho je: *Pouze reverzibilní (vratné) operace lze provádět bez disipace tepla.*
- Landauer: “Informace jsou fyzikální!” (“Irreversibility and heat generation in the computing process”, 1961)

Informační fyzika

- John Archibald Wheeler: “It from Bit” („Něco [pochází] z bitů“) (Information, physics, quantum: the search for links”, 1989)
- Wheeler: Informace se nacházejí v jádru fyziky a každé „něco“, ať už částice nebo pole, svou existenci odvozuje z pozorování.
- Fyzikální svět jsou proto primárně informace!
- **Informace → Fyzikální zákony → Hmota**

Entropie rozšířená na gravitaci

- Entropie rozšířená na gravitaci: Jacob Bekenstein přiřadil v roce 1974 entropii černým dírám na základě plochy jejich horizontu událostí vyjádřené v Planckových plochách $4\pi G\hbar/c^3$, tj. zhruba 10^{-70} m². Stephen Hawking v roce 1975 odvodil záření odpovídající teplotě, jež má původ v Hawkingově-Bekensteinově entropii černých děr. (Pro elementární odvození viz Leonard Susskind: “The Black Hole War: My Battle with Stephen Hawking to Make the World Safe for Quantum Mechanics”, 2008; česky Argo 2013)
- Předpokládáme-li vesmír složený z hmoty/polí v kvantových stavech, tvoří stavy hmoty spočetnou množinu. Seth Lloyd v roce 2002 ukázal, že pro vesmír existuje maximální množství bitů rovnající se 10^{122} (“Computational capacity of the universe”, 2002).
- Lloyd 2006: Komputační vesmír jako opice náhodně klepající do klávesnic psacích strojů, kde tyto opice odpovídají kvantovým fluktuacím programujícím vesmír (“The computational universe”, 2006)

Meze možností výpočtů

- Bekensteinova mez: horní mez entropie, nebo informací v ohraničené oblasti prostoru (nebo maximální množství informací potřebných pro dokonalý popis systému až na kvantovou úroveň)
- Lidský mozek s objemem 1260 cm^3 má Bekensteinovu mez rovnu 2.6×10^{42} bitů
- Margolusův–Levitinův teorém (1998): výpočty nemohou probíhat rychleji než 6×10^{33} operací/sec/joule (<https://arxiv.org/pdf/quant-ph/9710043>)
- Seth Lloyd (2010): Ve vesmíru nemohlo proběhnout více než 10^{120} operací na 10^{90} bitech (<https://arxiv.org/abs/quant-ph/0110141>)

Meze výpočtů

- Exponenciální složitost některých problémů z nich může udělat problémy nejen obtížné, ale i ve skutečnosti *neřešitelné*: např. rozhodnutí, zda lze mapu s jedním milionem oblastí obarvit třemi barvami, může vést k situacím vyžadujícím 2^{1000} kroků výpočtu; což celkově i na nejrychlejší počítači podléhajícím Lloydovým omezením znamená nejméně 10^{242} let výpočtů (ve vesmíru, který je pouze 10^{10} let starý).
- Obdobné meze plynou z vlastností výpočtů prohledávajících celý stavový prostor: např. hra Go má na plné desce 19×19 více než 10^{170} možných pozic, prohledávat je pro jistotu výhry nelze.
- Obdobně „Shannonovo číslo“ jako odhad počtu možných her v šachu rovný 10^{120} z článku “Programming a Computer for Playing Chess” z roku 1950, opět mimo možnosti prohledávání.

- Sama výkonnost počítačů/výpočetní techniky/komunikací není dostatečnou odpovědí na řešení reálných problémů
- Přesnost a rychlost nestačí k vyvážení exponenciální složitosti problémů vedoucích k prohledávání všech možností (stavového prostoru všech možných cest k řešení)
- „Intelligence“ odpovídá hledání aproximativních řešení s počítačnou racionalitou jako základem pro implementaci pomocí strojů: AI – umělá inteligence

Stručná historie AI

- Letní projekt workshopu (letní školy) v Dartmouth v roce 1956: John McCarthy a Claude Shannon iniciovali setkání s cílem
- „Vycházíme z myšlenky, že každý aspekt učení nebo jakýkoli jiný rys inteligence lze principiálně popsat natolik přesně, že bude možné sestavit stroj, který ji bude simulovat. Pokusíme se zjistit, jak naučit stroje používat jazyk, vytvářet abstrakce a pojmy, řešit druhy problémů, které jsou nyní výhradně lidskou doménou, a zlepšovat se. Domníváme se, že v jednom či několika z těchto problémů může být učiněn značný pokrok, stačí, když se pečlivě vybraná skupina vědců sejde a budou spolu přes léto pracovat.“
- Obecně považováno za „zrození“ umělé inteligence (AI), nicméně během následujících let následovala rezoluční metoda dokazování Alana Robinsona či program hrající dámu Arthura Samuela, nikoli však strojové učení či automatický překlad.
- Zpráva pro britskou vládu 1973 uvádí: „V žádné, ani dílčí, oblasti zde [v oblasti AI] objevy dosud dosažené nedošly ke slibovaným výrazným dopadům.“

Stručná historie AI

- 1940-1950 počáteční práce
 - 1943 McCulloch a Pitts formulují model mozku na bázi Boolských obvodů
 - 1950 Turingův článek *Computing Machinery and Intelligence*
 - 1958 Rosenblatt navrhuje *perceptron*
- 1950-1970 úvodní nadšení
 - 50. léta rané programy AI (a jazyk LISP), hraní dámy, dokazování teorémů
 - 1956 Dartmouthské setkání a termín „umělá inteligence“ (artificial intelligence, AI)
 - 1965 Robinsonův algoritmus pro logické dokazování (rezoluční metoda)
 - 1969 monografie Minsky&Papert: *Perceptrons* zastavila v oblasti neuronových sítí výzkum do konce 80. let
- 1970-90 „znalostní přístupy“
 - 1969-1979 raný vývoj znalostních systémů
 - 1980-1988 tvorba „expertních systémů“ pro průmyslové aplikace
 - 1988-1993 oblast expertních systémů považována za neperspektivní (tzv. „zima AI“)
- 1990-2012 statistické přístupy
 - Začlenění pravděpodobnosti a nejistoty do formulace modelů
 - Prohloubení technické úrovně práce, Bayesovské sítě,
 - Aktéři a učící se systémy jako nový přístup (tzv. „jaro AI“)

Stručná historie AI

- 2012- renesance neuronových sítí
 - Big data, neuronové sítě, kombinace velkých cloudových systémů a velkých dat
 - Návrat k některým podoblastem AI, „hluboké učení“ od cca 2011
 - AI nasazována k použití v mnoha oblastech průmysl
 - 2012 Alexnet Alex Krizhevsky pod vedením Geoffreya Hintonu trénuje model neuronové sítě a využíví architekturu grafických procesorů (GPU) tro trénování
 - Ilya Sutskevar přidává myšlenku trénování na bázi obrázků ImageNet
 - 2015-18 velký rozmach klasifikačních aplikací na baázi této architektury
- 2018- velké jazykové modely
 - GPT-n (generativní předtrénované transformátory) v OpenAI
 - LLM – velké jazykové modely generující predikce pokračování textu/dialogu na bázi transformátorů a hlubokých sítí
 - Alec Radford, Karthik Narasimhan, Tim Salimans & Ilya Sutskever: *Improving Language Understanding by Generative Pre-Training*, Open AI, červen 2018 – GPT-1
 - 2022. ChatGPT umožňující široké testování GPT-3 širokou veřejností
 - 2023: GPT-4
 - Chatboty některých dalších firem (Gemini od Google resp- DeepMind, Anthropic, Meta, ...)

Meze AI?

I. J. Good a koncept inteligenční exploze

- Kolega Alana Turing from Bletchley Parku (kryptografie a Collosus)
- Články *Speculations Concerning the First Ultraintelligent Machine* a *Logic of Man and Machine* (oba z roku 1965)
- Úspěch v budování obecné AI může být největší událostí v lidské historii ...
- ... a je důležité zajistit, aby to také nebyl úspěch poslední.
- „První ultrainteligenční stroj může být posledním vynálezem, který člověk kdy udělá.“
- Ale: Vše, co přinesla inteligence je výsledkem inteligence, ...
- ... a pokud naši iteligenici dokážeme zesílit, nenexistují meze toho, kam se lidstvo může dostat.

I. J. Good (1965)

- “Ultrainteligentní stroj by mohl navrhovat ještě lepší stroje; bezpochyby by poté došlo k **‘inteligenční explozi’** a lidská inteligence by zůstala daleko pozadu. ... Je zvláštní, že tento aspekt se mimo science fiction zmiňuje tak zřídka.“
- I. J. Good také v roce 1968 spolupracoval jako konzultant se Stanleyem Kubrickem při realizaci filmu *2001: Vesmírná odysea*.

AI jako stroje mající IQ?

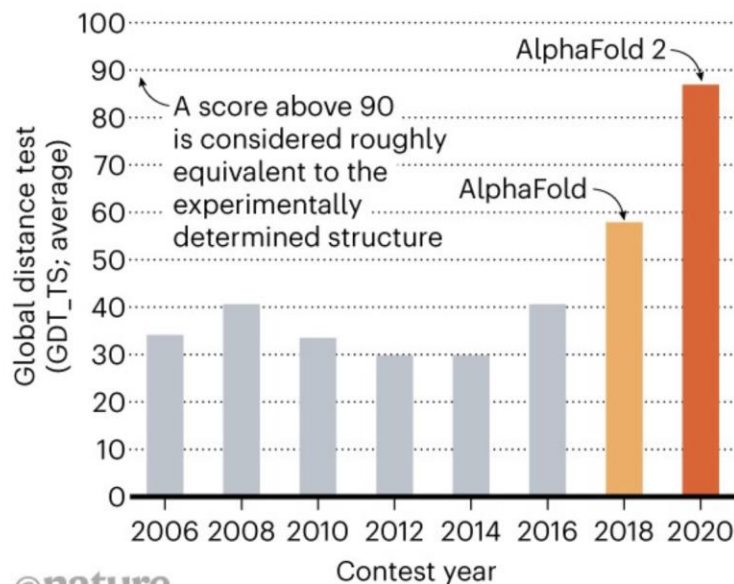
- Zatím žádná obecná inteligence srovnatelná s člověkem neexistuje – je ale bláhové předpokládat, že tomu tak bude napořád.
- Strojová inteligence zjevně neroste podle Mooreova zákona,
- ale daří se stavět pevné teoretické základy (racionální rozhodování, statistické učení, vnímání, zpracování přirozeného jazyka na bázi statistických vyvozování, Bayesovská vyvozování) a ...
- rychlý pokrok (hluboké učení ve zpracování jazyka, vidění a rozpoznávání mluveného, obecné pravděpodobnostní jazyky a dlouhodobé hierarchicky strukturované chování) ...
- ... celou oblast rychle posunují k inteligentnějším systémům bez viditelného omezení dalšího vylepšování.

Několik pozoruhodných speciálních případů

- Šachy: IBM DeepBlue porazil Kasparova
- Go: Google AlphaGo pro Go
- Go a šachy: Googles AlphaZero zvládl šachy i Go jen z pravidel hry, bez dalších zabudovaných znalostí (Kasparov: „AlphaZero otřásl hrou v šachy v základech“, předmluva ke knize Game Changer, 2019)
- Předpověď struktury proteinů: AlphaFold lepší než jiné přístupy v určování 3D struktury of proteinů na základě posloupností bází aminokyselin (Nature, 30. 11. 2020)
- Halicin – širokospektrální antibiotikum objevené pomocí AI a databází organických molekul (Cell a Nature, obojí 20. 2. 2020)
- Rozpoznávání obličejů a monitoring lidí
- GPT / ChatGPT, Bing, DeepL, ...
- Dall-E, Midjourney

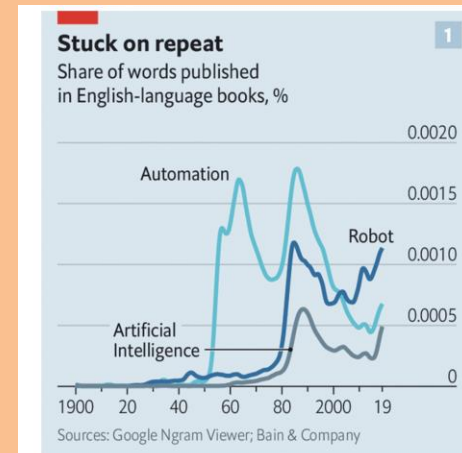
STRUCTURE SOLVER

DeepMind's AlphaFold 2 algorithm significantly outperformed other teams at the CASP14 protein-folding contest — and its previous version's performance at the last CASP.

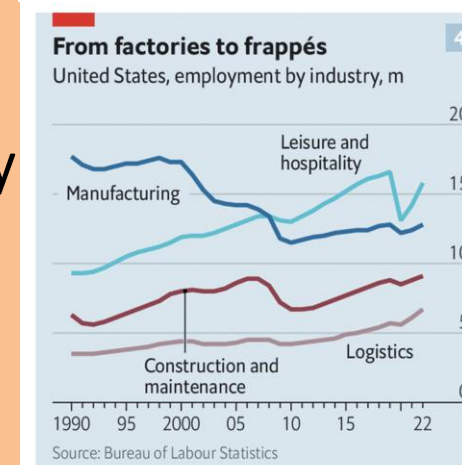


AI převyšuje schopnostmi člověka a mění strukturu pracovních míst

- Autonomní vozidla – byť zatím jen na úrovni 4 (oblasti pokryté mapami) jako velký přínos
 - Lékařská rentgenová vyšetření
 - Automatické generování novinových článků
 - Atd.
-
- Prakticky všechna pracovní místa mají nějakou složku, kterou mohou stroje dělat lépe – důsledky pro restrukturalizaci firem/společností, atd.
 - Jedinečná úloha zůstane “humanistickým” činnostem – péče o děti a o seniory, lékařské a sociální služby, sociální zabezpečení včetně např. univerzálního zaručeného příjmu, vzdělávání, atd.



The Economist



The Economist

Zdroj: Don't fear an AI-induced jobs apocalypse just yet, The Economist, 6. 3. 2023