MUNI
ICS

13. Umělá inteligence, velké jazykové modely a co s tím

Aleš Křenek
ljocha@ics.muni.cz

Ústav výpočetní techniky, MU

podzim 2023

Take-home message

- AI není magie, i když to tak občas vypadá
  - a není to nic nového (50–70 let)
- Synergie několika dobrých nápadů, rozvoje technologie, dostupnosti dat
  - plus atraktivita, která tentokrát přitáhla velký kapitál
- Je to už inteligence? A co je to ta naše inteligence?
- Dobrý sluha, ale zlý pán

Co je to umělá inteligence

- *„artificial intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings“* (Encyclopedia Britannica)

Co je to umělá inteligence

- *„artificial intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings“* (Encyclopedia Britannica)
- Mýtické a literární představy
  - Talos, Golem, …, Marvin (D. Adams)

Co je to umělá inteligence

- *„artificial intelligence (AI), the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings“* (Encyclopedia Britannica)
- Mýtické a literární představy
  - Talos, Golem, …, Marvin (D. Adams)
- Předpoklad, že postup lidského myšlení lze vyjádřit mechanicky
  - Aristoteles, Euclides, Al-Khwarizmi, …, K. Gödel
- Dartmouth Workshop 1956
  „every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it“

Turingův test

- Rozhodčí komunikuje se dvěma účastníky
- Ví se, že jeden je stroj, druhý člověk
- Komunikuje se v přirozeném jazyce, v základu jen textovým chatem
- AI test splnila, když ji takto nelze od člověka spolehlivě odlišit
- Současné LLM?

Co všechno je AI?

$$AI \supset ML \supset NN \supset DeepNN \supset LLM$$

Jde to bez počítače?

- Klíč k určování hub
  - tvar třeně: hlízovitý, soudkovitý, kyjovitý, válcovitý … jdi na bod 2,3,4,…
  - povrch klobouku: hrbolkatý, jamkatý, vrásčitý, vláknitý, šupinkatý, …
- Otázky uspořádány do **rozhodovacího stromu**
- Zpravidla na základě netriviálních **vlastností** (features)
  - obtížný úkol vyhodnocení vlastností zůstává na člověku – jeho přirozené inteligenci
  - mechanické zpracování realizuje poměrně jednoduchý postup (i když může být obsáhlý)

Expertní systémy

- AI postavená na strukturované **expertní znalosti**
- Zpravidla obsáhlý soubor faktů a pravidel
- Aplikace na konkrétní vstup
- Zřejmě nejúspěšnější použití v lékařských systémech
- Posloupnost logických kroků – **řešení** je současně **vysvětlení**

Expertní systémy

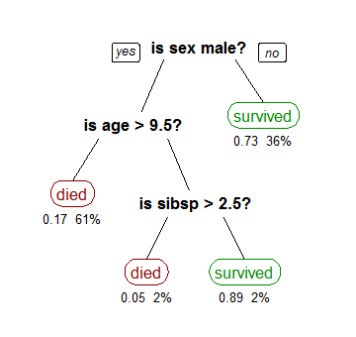
- AI postavená na strukturované **expertní znalosti**
- Zpravidla obsáhlý soubor faktů a pravidel
- Aplikace na konkrétní vstup
- Zřejmě nejúspěšnější použití v lékařských systémech
- Posloupnost logických kroků – **řešení** je současně **vysvětlení**
- Pravidla
  - x je alkoholik, y je syn $x \Longrightarrow y$ má sklon k alkoholismu
  - z má sklon k alkoholismu a z má rád pivo $\Longrightarrow z$ je alkoholik
- Fakta
  - Pepa byl alkoholik, Franta a Honza jsou jeho synové, Franta má rád pivo
- Konkrétní dotaz
  - Jsou Franta nebo Honza alkoholici?

Strojové učení

- Rozhodovací strom
  - klíč k určování hub, zajímá mě pouze je-li houba jedlá

Strojové učení

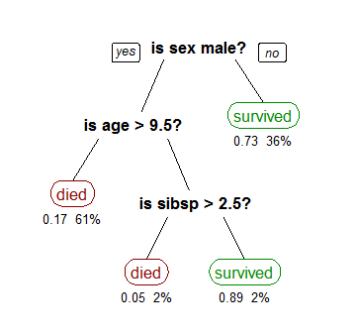
- Rozhodovací strom
  - klíč k určování hub, zajímá mě pouze je-li houba jedlá
- Lze zkonstruovat **pouze z dat**, bez expertní znalosti



(https://en.wikipedia.org/wiki/Recursive_partitioning)

Strojové učení

- Rozhodovací strom
  - klíč k určování hub, zajímá mě pouze je-li houba jedlá
- Lze zkonstruovat **pouze z dat**, bez expertní znalosti



(https://en.wikipedia.org/wiki/Recursive_partitioning)

- Náhodné rozhodovací lesy
  - více stromů, hlasují o výsledku

Support Vector Machine

- Vhodně reprezentovaná data
- Proložení přímkou, **co nejlépe** odděluje pozitivní a negativní případy
  - příklad: mohu už řídit po včerejším večírku?

Support Vector Machine

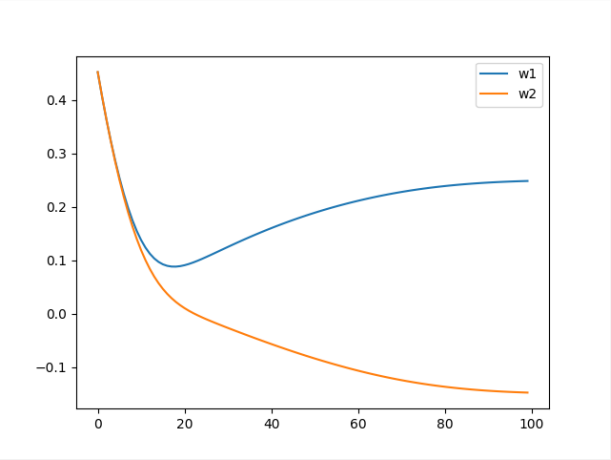
- Vhodně reprezentovaná data
- Proložení přímkou, **co nejlépe** odděluje pozitivní a negativní případy
  - příklad: mohu už řídit po včerejším večírku?
- Potřebná transformace dat
  - příklad: bezpečná úroveň radiace po zamoření

Neuronové sítě

- Volná inspirace strukturou nervového systému, ne jeho model
- **Neuron** – základní jednotka
  - pracuje se „signály“ – reálná čísla
  - jeden nebo více vstupů, každý má přiřazenu **váhu**
  - výstup je jejich transformovaným součtem, vstupem pro další neurony
  - zpravidla uspořádání do vrstev
- Aplikace modelu od vstupu k výstupu
- Učení modelu **zpětnou propagací** chyby
- Příklad: piva po večírku

Neuronové sítě

- Příklad – průběh učení
  - 1000 vzorků „promile v krvi po *x* pivech a *y* hodinách“
  - tréning v dávce po 20, zopakováno ve 100 epochách



Hrubá síla

- „Větší kladivo“ je v mnoha aplikacích IT účinné
- Šachy, tradičně spojovány s inteligencí, původní šachové programy expertní
- IBM Deep Blue vs. Gary Kasparov, 1997
- Strojové učení (neuronové sítě speciálně) jdou tímto směrem
- Máme už sílu na prozkoumání „všech“ možností
- Je to lepší než si uzavřít některé cesty a priori
- AlphaFold
  `https://deepmind.com/blog/article/putting-the-power-of-alphafold-into-the-worlds-hands`

Skutečné NN

- „Deep“ znamená mnoho vrstev
- Libovolný počet lineárních vrstev lze nahradit jedinou – omezená síla
- Nelinearita v podobě **aktivační funkce** aplikované na výstup
  - původně sigmoid $\frac{1}{1+e^{-x}}$ nebo tanh
  - v moderních systémech především **Rectified Linear Unit** (ReLU) a odvozené
- Komplexní architektury
  - konvoluční
  - rekurentní, LSTM
  - transformery
- Složitější postupy trénování

Rozpoznávání vlastností

- Rozpoznání vstupních vlastností (features) je významný problém
  - tvar klobouku a třeně houby, co zjišťovat o lidech na Titanicu, …
  - ručně pomalé, zatížené chybami
  - automatické vyžaduje samostatný expertní systém
- Nechme to hrubé síle, např. **konvoluční vrstvy**

- Příklad na zvířátkách
  `https://doi.org/10.1016/j.measen.2022.100611`

Realistický příklad – FaceId

- Nezávislá rekonstrukce
  `https://towardsdatascience.com/how-i-implemented-iphone-xs-faceid-using-deep-learning-in-python-d5dbaa128e1d`
- Vstupní data 200 $\times$ 200 bodů RGB-D
- Tzv. siamská neuronová síť
  - dvě stejné sítě vedle sebe na zpracování dvou obrázků, sdílejí váhy
  - cca. 10 konvolučních vrstev
  - výstupem 128-složkové vektory vlastností (redukce dimenzí 160000 : 128)
- Trénováno na kontrast (constrastive loss):
  - snímky stejnéno člověka jsou v 128-rozměrném výsledném prostoru co nejblíže
  - snímky různých lidí co nejdále
- To vše „ve výrobě“
  - velké datové sady, postupně zhoršující se podmínky (dvojčata, roušky, …)
- Nový telefon si při inicializaci „jen“ zapamatuje, kam se váš obličej do 128-rozměrného prostoru promítne

Čínský pokoj

- Člověk v zavřeném pokoji, neumí čínsky mluvit ani číst a psát
- Pod dveřmi mu podstrčíme papírek s otázkou v čínštině
- Má k dispozici veškerou čínsky psanou literaturu a neomezený čas
- Pravděpodobně dokáže najít a napsat korektní odpověď

Čínský pokoj

- Člověk v zavřeném pokoji, neumí čínsky mluvit ani číst a psát
- Pod dveřmi mu podstrčíme papírek s otázkou v čínštině
- Má k dispozici veškerou čínsky psanou literaturu a neomezený čas
- Pravděpodobně dokáže najít a napsat korektní odpověď
- Můžeme tvrdit, že umí čínsky?

Čínský pokoj

- Člověk v zavřeném pokoji, neumí čínsky mluvit ani číst a psát
- Pod dveřmi mu podstrčíme papírek s otázkou v čínštině
- Má k dispozici veškerou čínsky psanou literaturu a neomezený čas
- Pravděpodobně dokáže najít a napsat korektní odpověď

- Můžeme tvrdit, že umí čínsky?
- A jak se to liší od znalosti našeho mateřského jazyka?

Velké jazykové modely (LLM)

- Opravdu velké

  „Vesmír je velký. Fakticky velký. To byste nevěřili, jak je hrozivě obrovitánsky nepředstavitelně veliký. Myslíte si třeba, že drogerie ve vaší ulici je daleko, ale proti Vesmíru je to úplný houby.“ (D. Adams)

- GPT-4: 1.75T parametrů, kontext 25k slov, trénovací sada 13T tokenů

Velké jazykové modely (LLM)

- Opravdu velké

  „Vesmír je velký. Fakticky velký. To byste nevěřili, jak je hrozivě obrovitánsky nepředstavitelně veliký. Myslíte si třeba, že drogerie ve vaší ulici je daleko, ale proti Vesmíru je to úplný houby.“ (D. Adams)

- GPT-4: 1.75T parametrů, kontext 25k slov, trénovací sada 13T tokenů

- Vhodná reprezentace jazyka (embedding)

  „AI can do virtually anything for you … with the right embedding“ (T. Hoefler)

Velké jazykové modely (LLM)

- Opravdu velké

  „Vesmír je velký. Fakticky velký. To byste nevěřili, jak je hrozivě obrovitánsky nepředstavitelně veliký. Myslíte si třeba, že drogerie ve vaší ulici je daleko, ale proti Vesmíru je to úplný houby.“ (D. Adams)

- GPT-4: 1.75T parametrů, kontext 25k slov, trénovací sada 13T tokenů

- Vhodná reprezentace jazyka (embedding)

  „AI can do virtually anything for you … with the right embedding“ (T. Hoefler)

- Jinak je to „jen“ neuronová síť (s chytrou architekturou)

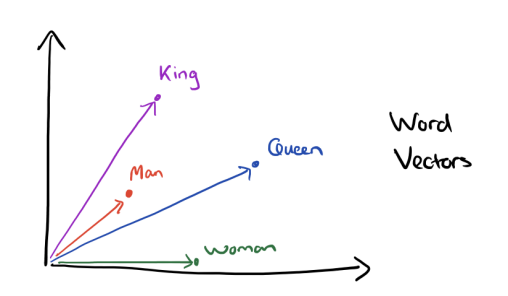
LLM: tokeny a slovník

- Slova
  - přirozený jazyk – statisíce až miliony slov
  - to je příliš mnoho
- Písmena (v západních jazycích), slabiky
  - příliš málo, nenesou dostatek kontextového významu
- Korpusové slovníky
  - získané analýzou rozsáhlých textů (četnost výskytu atd.)

    `['this', 'is', 'an', 'example', 'of', 'the', 'bert', 'token', '##izer']`
    `[101, 2023, 2003, 2019, 2742, 1997, 1996, 14324, 19204, 17629, 102]`
    (`https://huggingface.co/bert-base-uncased`)
  - řádově desetitisíce položek

LLM: Embedding

- Reprezentace tokenu vektorem



(https://blog.acolyer.org/2016/04/21/the-amazing-power-of-word-vectors/)

- Významem blízká slova jsou poblíž, ideálně funguje i „aritmetika“

LLM: Embedding

- Stovky až tisíce dimenzí, např. „example“ (2742):

```
[ 7.0699e-03,  3.9590e-02, -6.2164e-02, -8.4340e-02, -1.2362e-02,
  1.0582e-02, -1.2302e-01, -6.6595e-03, -6.5421e-02,  2.0174e-03,
...
  -1.7751e-02, -2.9460e-03, -7.4038e-02]
```

  (768 v bert-base-uncased)

- Získané analýzou obsáhlých textů

  `https://www.tensorflow.org/text/tutorials/word2vec`

Generativní LLM

- Vstup: embedding tokenů otázky, začátku věty apod.
- Hrozivě obrovitánsky nepředstavitelně velká neuronová síť
- Výstup: samostatný „signál“ pro každou položku slovníku – pravděpodobnost výskytu tohoto slova jako pokračování konverzace

Generativní LLM: Učení

- Síti předložíme začátek konverzace
- Víme jak pokračuje, ale to na vstup už nepřijde
- Na výstupu očekáváme 1 pro signál tohoto slova, 0 pro všechny ostatní
- Zpětně propagujeme chybu a upravujeme všechny váhy

Generativní LLM: Použití

- Na vstup otázka uživatele apod.
- Výstupem je distribuce pravděpodobnosti přes celý slovník
- Podle ní náhodně vylosujeme další token
- Token přídáme ke vstupu a pokračujeme
- Celé několikrát, z výstupů se vybere ten nejlepší

Transformer

- Attention Is All You Need
- „**Kocour** nevylezl na strom, protože **byl** příliš líný.“

Transformer

- Attention Is All You Need
- „**Kocour** nevylezl na strom, protože **byl** příliš líný.“
- Attention head
  - tři nezávislé lineární transformace slova na vektory **query**, **key** a **value**
  - každé slovo z věty použijeme jako **query** pro všechna ostatní slova v roli **key**
  - skalární součiny těchto dvojic vygenerují **skóre**
  - skórem vynásobená a sečtené **value** všech slov ve větě jsou výstupem hlavy pro toto slovo

Transformer

- Celková kompozice
  - více hlav vedle sebe v jedné vrstvě
  - více vrstev nad sebou
  - proloženo lineární kombinací, normalizací atd.
- Další technikality
  - poziční embedding maskování neúplného vstupu, obtékání hlav, sdílení vah, …
- Na výstupu pravděpodobnosti pro každý token ze slovníku

Transformer: it just works

- Trénují se všechny parametry všech transformaci Q, K, V a mezivrstvy
- Vše je vyjádřeno jako operace na tenzorech
  - velmi efektivní na moderních CPU i GPU a dalších specializovaných akcelerátorech
- Hlavy v jedné vrstvě a většina výpočtu uvnitř hlavy jsou nezávislé
  - lze je provádět masivně paralelně
- Milionová myšlenka ale i důsledek 10–15 let intenzivního vývoje
- Další čtení
  - `https://dl.acm.org/doi/10.5555/3295222.3295349`
  - `https://jalammar.github.io/illustrated-transformer/`
  - `https://towardsdatascience.com/attention-is-all-you-need-discovering-the-transformer-paper-73e5ff5e0634`

LLM ještě není chatbot

- ChatGPT et al. jsou nadstavby nad LLM
- Důraz na **spokojenost** uživatele a **bezpečnost** odpovědi
- RLHF: reinforcement learning with human feedback
  - lidé hodnotí výstupy chatbotu: dobrá/špatná, bezpečná/nebezpečná, …
  - takových hodnocení je málo, trénuje se z nich **award model**
  - náhodně generované konverzace se jím ohodnotí a podle výsledku se dolaďuje LLM

Technologie a investice

- Hardware a nízkoúrovňové programování
  - GPU od 1970s, významnější v 1990s (SGI), rané pokusy o výpočty
  - 2000: první rozumně programovatelný čip Nvidia GeForce 3
  - 2007: standard CUDA
- Vhodné knihovny
  - TensorFlow 2015(17), PyTorch 2017 (na původním Torch 2002)
  - správná úroveň abstrakce: efektivní i použitelná
- Peníze
  - identifikovaná příležitost a investice gigantů (Google, Facebook, Microsoft, …)
  - odhadovaný roční obrat přes 1T USD (10× státní rozpočet ČR)

Společenské dopady

témata k diskusi

- Krádež identity
- Deep fake a manipulace s lidmi
- Odpovědnost za rozhodnutí
- Autorská práva
- Zneužití nedemokratickými režimy
- Bullshit jobs v. 2.0

Mohou stroje myslet?

- „Know-how“ získáváme zkušeností a učením
- Lidský mozek má zřejmě stále větší výpočetní kapacitu a umí ji efektivněji použít. Ale opravdu vymýšlíme něco nového?
- „Věc o sobě,“ kterou nemůžeme v její plné podstatě poznat vs. „fenomén“ (I. Kant)