Historie kybernetiky a umělé inteligence

06. Počátky umělé inteligence

Miloš Železný

Katedra kybernetiky
Fakulta aplikovaných věd
Západočeská univerzita v Plzni

4. listopadu 2024

Definice inteligence 1

Definice inteligence

- Inteligence je schopnost člověka abstraktně a rozumně přemýšlet a od toho odvozovat účelná jednání (přijímat účelná rozhodnutí).
- **Douglas R. Hofstadter (*1945)**, americký kognitivní vědec, nabízí tyto základní schopnosti inteligence:
  1. velmi pružně reagovat na situace,
  2. využít příznivých okolností,
  3. pochopit smysl nejednoznačných nebo vzájemně protichůdných sdělení,
  4. rozpoznat relativní důležitost různých aspektů dané situace,
  5. najít podobnosti mezi situacemi navzdory rozdílům, které by je od sebe mohly odlišit,
  6. rozlišit situace navzdory podobnostem, které by je mohly spojovat,
  7. vytvářet nové koncepty s využitím starých konceptů a jejich inovativního kombinování,
  8. přijít s novými myšlenkami

Definice inteligence 2

Definice inteligence

- Definice inteligence od Jeana Piageta (1896–1980), švýcarského vývojového psychologa, parafrázovaná Jozefem Kelemenem:
  Inteligence je nástrojem styku mezi inteligentní bytostí a jejím prostředím. Její úlohou je zabezpečovat „rovnováhu“ mezi bytostí a prostředím (především ve smyslu rovnováhy mezi aktivní činností subjektu a pasivním přizpůsobením se prostředí).
- existuje samozřejmě nepřeberné množství dalších definic, například:
  Inteligence je schopnost dobře vyplňovat IQ testy.

Definice umělé inteligence 1

Definice umělé inteligence

- dle zakladatele oboru M. Minskyho:
  Umělá inteligence je věda o vytvoření strojů nebo systémů, které budou při řešení určitého úkolu užívat takového postupu, který – kdyby ho dělal člověk – bychom považovali za projev jeho inteligence.
- dle ISO 2382:2015:
  Umělá inteligence je odvětví počítačových věd věnující se vývoji systémů pro zpracování dat, které provádějí činnosti obvykle spojované s lidskou inteligencí, jako například uvažování, učení a sebezdokonalování.
- dle Elaine Richové:
  Umělá inteligence se zabývá tím, jak počítačově řešit úlohy, které dnes zatím zvládají lidé lépe.

Definice umělé inteligence 2

Definice umělé inteligence

- Robotik Rodney Brooks trochu posměšně tvrdí, že:
  . . . podle toho, čemu se věnovali výzkumníci v raných dobách UI, lze soudit, že za projev inteligence se považuje především schopnost řešit úlohy, které výtečně vzdělaní vědci mužského pohlaví považují za náročné.
- zatímco podle samotného Brookse obecná inteligence spíše *vyvstane z interakce mezi percepcí a akcí.*

Definice umělé inteligence 3

Definice umělé inteligence

Věda o vytváření strojů, které:

myslí jako lidé	myslí racionálně
jednají jako lidé	jednají racionálně

Definice umělé inteligence 4

Latentní motto umělé inteligence

Mozek je ve své podstatě jen složitý stroj a jako takový se dá simulovat uměle.

- Ačkoliv si většina vědců, který se tímto problémem zabývali, vždy myslela, že taková simulace bude jednodušší, než se nakonec ukazuje, (zřejmě) nikdo z nich nebyl tak naivní, aby se pokoušel mozek simulovat parní strojem, spalovacím motorem či hodinovým strojkem.
- Rozvoj oboru UI byl tak nastartován až s příchodem prvních elektronických počítačů.

Zrození UI 1

Dartmouth Workshop

- Informatik John McCarthy přesvědčil v roce 1955 Marvina Minskyho a několik dalších vědců, aby mu v následujícím roce pomohli zorganizovat letní výzkumný workshop na Dartmouth College, kde by se:
  ... po dobu 2 měsíců 10 vědců věnovalo výzkumu umělé inteligence ...
  Směřování výzkumu bude založeno na naší domněnce, že každý aspekt učení nebo i jakéhokoliv jiného rysu inteligence může být v principu popsán tak přesně, že je možné vyrobit stroj, který tyto rysy dokáže simulovat. Pokusíme se zjistit, jak přimět stroje používat jazyk, tvořit abstrakce a pojmy, řešit takové úlohy, které jsou nyní vyhrazeny lidem, a samy sebe vylepšovat.
  Myslíme si, že může být dosaženo významného pokroku v jedné nebo více z uvedených oblastí, pokud na tomto tématu bude společně pracovat pečlivě vybraná skupina vědců po celé léto.

Zrození UI 2

Dartmouth Workshop

- Nakonec se na workshopu žádný zásadní vědecký průlom nekonal, ale došlo k důležitému seznámení lidí, kteří ještě mnoho let poté táhli výzkum UI na MIT, Stanfordu, Carnegie Mellon University (CMU) a v IBM.
- Na zmíněném workshopu zazářili především Allen Newell and Herbert Simon, kteří přivezli s sebou svůj program pro „strojové uvažování“ *Logic Theorist* (LT), který uměl dokázat 38 z 52 matematických vět (teorémů) z díla A. W. Whiteheada a B. Russella *Principia Mathematica* – jeden z nich dokonce elegantněji než autoři knihy.

John McCarthy

John McCarthy (1927 – 2011)

- matematik a informatik
- studoval a působil na Princetonu, MIT, Dartmouth College a Stanfordu
  - 1959 – založení MIT Computer Science Artificial Intelligence Lab (CSAIL)
  - 1962 – založení Stanford Artificial Intelligence Lab(SAIL)
- prosazoval využití matematické logiky v UI

John McCarthy

John McCarthy (1927 – 2011)

- 1958 – v té době na MIT - programovací jazyk Lisp:
  - navržený pro efektivní manipulaci se symboly (pozn.: schopnost symbolického uvažování je další z možných definic inteligence)
  - zabudovaná podpora práce se seznamy (Lisp = List processing)
  - druhý nejstarší dosud používaný vysokoúrovňový programovací jazyk (po Fortranu)
- 1958 – Advice Taker – teoretický koncept počítačového programu, který by používal logiku (konkrétně zřejmě predikátovou logiku 1. řádu) pro odvozování nových faktů o světě z uložených či průběžně zadávaných axiomů a konkrétních informací předložených tomuto programu,
- 1966 – McCarthy byl členem týmu, který připravoval software pro první souboj USA-SSSR v počítačovém šachu.

Marvin Minsky

Marvin Minsky (*1927)

- původem matematik, poté informatik a kognititivní vědec
- 1951 – SNARC (Stochastic Neural Analog Reinforcement Calculator)
  - jedna z prvních fyzických implementací umělé neuronové sítě
  - sestavený z elektronek, motorů a převodovek
  - simuloval chování krysy, která se snaží najít správnou cestu v bludišti
  - obsahoval ekvivalent cca 40 neuronů – část z nich byla mezi sebou propojena náhodně

Marvin Minsky

Marvin Minsky (*1927)

- 1963 – „head-mounted display“
- 1969 – kniha Perceptrons: an introduction to computational geometry (spoluator S. Papert)
  - rozebírá vlastnosti jednoho z typů neuronových sítí (perceptronové sítě - 1957 - Frank Rosenblatt) a především poukazuje na její slabiny,
  - kritici tuto knihu viní z toho, že (zbytečně) způsobila „zmrazení“ výzkumu neuronových sítí až do 80. let

Marvin Minsky

Marvin Minsky (*1927)

- 1975 – teorie rámců (frames)
  - základem této teorie je myšlenka, že pokud se člověk dostane do situace, ve které už někdy byl, vyvolá v mozku strukturu (rámec), která již obsahuje nějaká očekávání o dané situaci,
  - rámce pak obsahují tzv. sloty, které lze plnit konkrétními hodnotami + “algoritmy” pro plnění slotů či postupy, co dělat, je-li slot nevyplněn,
  - tato teorie ovlivnila později i objektově orientované programování a dodnes se v UI využívá

Marvin Minsky

Marvin Minsky (*1927)

- 1986 – kniha (a model) Society of Mind
  - základní myšlenkou této knihy je předpoklad, že za lidskou (a koneckonců i jakoukoliv jinou) inteligencí je místo nějakého obecného jednotícího principu/algoritmu spíše složitá interakce mezi jednotlivými agenty (funkčními oblastmi mozku či částmi počítačového kódu),
  - tito „agenti“ sami o sobě „mysl“ nemají
- 1968 poradce Stanleyho Kubricka při filmování 2001: Vesmírná odysea

Logic Theorist 1

Logic Theorist

- program využíval principy matematické logiky (přirozeně – takto je napsaná i Principia Mathematica) a metody prohledávání stromů, včetně tzv. heuristik (tento termín je znám v obecnějším smyslu, tj. jako zkusmé, intuitivní řešení problémů či stejně založený způsob rozhodování.)
- Simon byl z výsledků (pochopitelně) nadšen, dokonce se na začátku roku 1956 pochlubil svým studentům, že *„přes Vánoce s Alem Newellem vynalezli myslící stroj“* – později dokonce napsal *„vymysleli jsme počítačový program, který je schopný nečíselného myšlení, a tudíž jsme vyřešili dávný problém mysli a těla, vysvětlujíc, jak může systém sestávající z hmoty mít vlastnosti mysli.“*
- 1959 – stejní autoři představují General Problem Solver – zobecněnou verzi LT využívající stejné principy.

Logic Theorist 2

Logic Theorist

- Jedná se o program pro strojové uvažování, považovaný za první UI program vůbec.
- autoři:
  - Herbert A. Simon (1916 – 2001)
    - americký politolog, ekonom, sociolog, psycholog a informatik – laureát Nobelovy ceny za ekonomii,
    - jeho hlavním předmětem zájmu bylo rozhodování - a to ve všech zmíněných oblastech
  - Allen Newell (1927 – 1992)
    - informatik a psycholog, student H. Simona
  - Cliff Shaw (1922 – 1991)
    - programátor
- LT používal techniky prohledávání stromů a to včetně tzv. heuristik – metody používané pro řešení úloh v UI dodnes.

Herbert A. Simon

Herbert A. Simon (1916-2001)

- americký kognitivní psycholog, ekonom, počítačový vědec a filozof,
- sám o sobě tvrdil, že jej – přes evidentní šíři jeho zájmů – vlastně zajímá jen jediná věc a to rozhodování (decision-making),
- je autorem konceptů omezené racionality (bounded rationality) a satisficingu – viz dále,

Herbert A. Simon

Herbert A. Simon (1916-2001)

- za svoje práce v oboru rozhodování v ekonomických organizacích dostal Nobelovu cenu za ekonomii (1978) a za svoji práci v oboru UI Turingovu cenu (přezdívanou „Nobelova cena za výpočetní techniku“) v roce 1975,
- jeho psychologické teorie omezené lidské racionality položily základ tzv. behaviorální ekonomie, za jejíž rozvoj získali Nobelovu cenu za ekonomii Daniel Kahneman (2002) a Richard Thaler (2017).

Racionalita 1

Racionalita v rozhodování

- v ekonomii se především ve 20. století většinově prosadila tzv. **teorie racionální volby**
- ta předpokládá, že všichni lidé jednají racionálně, tj. přísně logicky a individualisticky a jejich cílem je maximalizovat vlastní užitek – koncept zvaný též *homo economicus*
- dále se také předpokládá, že takový člověk jakožto „racionální agent“:
  - zná veškeré alternativy (má úplné informace) – výsledky x_i
  - zná distribuce pravděpodobností každého výsledku - $p(x_i)$
  - dokáže určit nejlepší řešení z dostupných alternativ – tj. má pevně určené své preference a navíc je schopen vyjádřit je číselně (užitková funkce $u(x_i)$)

Racionalita 2

Racionalita v rozhodování

- pokud tohle všechno platí, homo economicus jednoduše maximalizuje očekávanou (střední) hodnotu užitkové funkce:

$$E[u(x)] = \sum_{i=1}^{n} p(x_i) \cdot u(x_i)$$

Př. Vsadím 100 Kč na jedno číslo na hrací kostce a v případě výhry získám 500 Kč. Jaká je očekávaná hodnota této sázky?

$$[E[u(x)] = \frac{1}{6} \cdot 400 + \frac{5}{6} \cdot (-100) = \frac{400 - 500}{6} = -16.66$$

- výsledkem takového rozhodovacího procesu je vždy optimum (časová náročnost hledání a zvažování všech alternativ se obvykle neuvažuje).

Racionalita 3

Opravdu to tak funguje?

- Asi každý z vlastní zkušenosti intuitivně tuší, že takto se lidé nechovají (a experimenty to potvrzují).
- Lidé se však nechovají racionálně — v dříve uvedeném smyslu — ani za podmínek, kdy je „racionální vyhodnocení situace“ velmi jednoduché.

Racionalita 4

Šance na vyšší výhru

- Chtěli byste raději:
  1. Dostat 10 000 Kč.
  2. Zúčastnit se hry, do které nemusíte vkládat žádnou sázku a máte 90% pravděpodobnost, že vyhrajete 12 000 Kč (a tudíž 10% pravděpodobnost, že nedostanete nic)?

Očekávaná střední hodnota užitkové funkce:

1. $E[u(x)] = 1 \cdot 10000 = 10000$
2. $E[u(x)] = 0,9 \cdot 12000 + 0,1 \cdot 0 = 10800$

Racionalita 5

Šance na nulovou pokutu

- Chtěli byste raději:
  1. Zaplatit pokutu 10 000 Kč.
  2. Zúčastnit se hry, ve které máte 90% pravděpodobnost, že zaplatíte pokutu 12 000 Kč a 10% pravděpodobnost, že vám pokuta bude odpuštěna?

Očekávaná střední hodnota užitkové funkce:

1. $E[u(x)] = 1 \cdot (-10000) = -10000$
2. $E[u(x)] = 0,9 \cdot (-12000) + 0,1 \cdot 0 = -10800$

Omezená racionalita

Omezená racionalita (bounded rationality)

- Tato teorie bere v úvahu, že:
  - svět kolem nás je příliš komplexní na to, abychom mu rozuměli v celé šíři,
  - zjišťování všech alternativ je extrémně časově náročné a proto obtížně proveditelné a nebo prostě jen drahé,
  - lidé mají své „kognitivní zkratky“ (heuristiky), které často využívají:
    - tyto heuristiky jsou na jednu stranu pro lidi určitým „požehnáním“, protože ve většině případů fungují podobně dobře jako dokonale racionální „decision-making“ a jsou mnohem rychlejší,
    - na druhou stranu mohou i škodit, protože vedou k celé řadě tzv. kognitivních zkreslení (cognitive bias),
    - Viz též: Daniel Kahmeman: Thinking, Fast and Slow (česky Myšlení, rychlé a pomalé).
- Teorie omezené racionality tedy tvrdí, že lidé se, striktně vzato, často rozhodují iracionálně — tato skutečnost ale není považována za jev nutně negativní, jen je třeba s ním počítat.

Satisficing

Satisficing

- Kombinace slov „satisfy“ a „suffice“ – čili postoj při rozhodování, kdy se „spokojím s dostatečně dobrým“ (aniž bych musel hledat optimální)
  - mám tedy nastavené nějaké minimální požadavky a pokud dané rozhodnutí (výrobek, akce, investice atd.) tyto požadavky splňuje, přestávám pátrat po alternativách.
- *Příklad (z Wikipedie): Pokud je mým úkolem přišít záplatu na kalhoty, výrobce kalhot může uvádět, že nejlépe to půjde s 10 cm jehlou. Mám-li tuto jehlu společně s tisícem dalších o délkách od 3 do 15 cm schovanou v kupce sena, teorie satisficingu praví, že mám záplatu přišít s první nalezenou jehlou, se kterou to jen trochu půjde – hledání té optimální je jen mrhání energií a časem.*
- je zřejmé, že satisficing úzce souvisí s výše uvedenou teorií omezené racionality.

Turingův test 1

Turingův test

- Ve svém článku „Computing Machinery and Intelligence“ (časopis Mind, Vol. 59, Issue 236, 1950) Turing navrhuje nahradit otázku „Mohou stroje přemýšlet?“ otázkou „Mohou stroje uspět v definované imitační hře?”
- Princip imitační hry na pozadí Turingova testu:
  - osoba C (rozhodčí) má za úkol rozlišit, který ze subjektů A a B je stroj a který člověk na základě několikaminutové komunikace,
  - komunikace probíhá v přirozeném jazyce, ale v psané formě, aby se odfiltroval vliv chyb v rozpoznávání a syntéze řeči,
  - stroj se snaží přesvědčit rozhodčího, že je člověkem – pokud se mu podaří, v testu uspěl,
  - důležitý, ale často opomíjený detail – člověk „za plentou“ by neměl lhát, tj. neměl by se vydávat za stroj.
- Z toho vyplývá další – a dlouho dobu poměrně „mainstreamová“ definice UI: *„Umělá inteligence je systém, který úspěšně projde Turingovým testem“*

Turingův test 2

Kritika Turingova testu

- Relevantnost Turingova testu pro posouzení „inteligence“ strojů je zpochybňována z mnoha různých pozic – např. filozofických (J. Searle a jeho Argument čínské komory) či psychologických.
- Zmiňme zde námitku „inženýrskou“:
  - hlavním cílem (inženýrské větve) UI výzkumu je vyvinout systém, který bude svoji schopnost „inteligence“ využívat k nějakým užitečným cílům (rozpoznávání řeči či obrázků, autonomní řízení vozidel),
  - oproti tomu při vývoji systému, který by měl být úspěšný v Turingově testu, je třeba zdaleka nejvíc práce vložit do modelování lidských nedokonalostí (počítač, který nabídne řešení složité rovnice ve zlomku sekundy, si asi nikdo s člověkem nesplete),
  - autoři soudobé učebnice UI Artificial Intelligence: A Modern Approach Stuart Russell a Petr Norvig tento přístup přirovnávají k situaci, kdy by se „letečtí inženýři ze všech sil snažili vyvinout stroje, které létají přesně stejně jako holubi do té míry, že i holuby dokážou zmást“.

Turingův test 3

Argument čínského pokoje

- Argument čínského pokoje byl předložen filosofem Johnem Searlem v roce 1980.
- Je to myšlenkový experiment, jehož cílem je ukázat, že samotná schopnost smysluplně odpovídat na položené otázky (hlavní princip Turingova testu) není dostatečná pro prokázání schopnosti vědomého porozumění, což očekáváme od tzv. silné umělé inteligence.
- V tomto experimentu je hypotetická osoba, jež neovládá čínštinu, uzavřena uvnitř místnosti, naplněné velkým množstvím čínských textů, ve kterých se nalézá smysluplná odpověď na každou čínskou otázku.
- Tato osoba má znalost klíče, podle kterého vždy dokáže nalézt na základě předaného textu (otázky) smysluplnou odpověď.
- Vnější tazatel brzy nabyde přesvědčení, že osoba uvnitř pokoje čínštině perfektně rozumí, přestože ve skutečnosti pouze mechanicky pracuje s pro ni neznámými symboly, takže by její práci mohl zastat i nemyslící stroj.

Zpět k Umělé inteligenci 1

Definice umělé inteligence

Věda o vytváření strojů, které:

myslí jako lidé	myslí racionálně
jednají jako lidé	jednají racionálně

Zpět k Umělé inteligenci 2

Jaký to má vztah k UI?

- „Dostatečně dobré“ se matematicky vyhodnocuje hůř, než „optimální“.
- Současný výzkum umělé inteligence aspirující na vytvoření systému „obecné umělé inteligence“ (AGI) tedy nejčastěji pracuje právě s paradigmatem „racionálního agenta“ - tj. agenta, který za všech okolností maximalizuje svůj očekávaný užitek.
- Komunita UI si nejspíše tenhle rozpor mezi omezeně racionálním člověkem a dokonale racionální umělou inteligencí uvědomuje a jejím cílem je tedy spíše vytvořit něco, co „lidské nedokonalosti“ překoná (viz lehce posměšné výroky o „imitování holubího letu“).
- Jelikož si ale důsledek dokonale racionálního jednání nedokážeme vždy představit, může se snadno stát, že zapomeneme pro AGI specifikovat všechny možné „omezující podmínky“, což následně může vést ke (zdánlivě) nepředvídatelným a nežádoucím koncům.

Děkuji za pozornost ...

Dotazy?