

IV130 Přínosy a rizika inteligentních systémů

Intelligence v lidech a strojích

Jiří Zlatuška

1. března 2024

Inteligence v lidech a strojích

Pravidla pro vyvozování závěrů

- Aristotelés (384—322 př.n.l.) – první pokus formulovat přesné zákony, kterými se řídí racionální myšlení (systém sylogismů)
- Aristotelovo *praktické uvažování*: „Neuvažujeme tedy o účelech a cílech, nýbrž o prostředcích. Neboť ani lékař neuvažuje, bude-li léčit, ani řečník, bude-li promlouvat [...]. Ale když si stanovili cíl, hledí k tomu, jak a čím ho dosíci, a jeví-li se, že ho lze dosíci více prostředky, přihlížejí k tomu, kterým prostředkem lze ho dosáhnout nejsnáze a nejkrásněji. Možno-li však uskutečnit jej jen jedním prostředkem, uvažují, jak jest možno uskutečnit jej tímto a čím zase onen, až přijdou k příčině, která v hledání jest poslední; [...] to, co při rozboru jest poslední, jest při uskutečňování první. A narazí-li na nemožnost, ustávají, například je-li potřeba peněz a není možno je opatřit; ale jeví-li se to možným, pokoušejí se jednat.“
(*Etika Nikomachova*, kniha III, 3, 1112b)
- Slabinou tohoto přístupu je pomíjení nejistoty, která je běžnou součástí rozhodování
- Implementováno jako systém GPS (General Problem Solver) v roce 1959 a metody založené na logickém plánování dominovaly počátkům AI

Pravidla pro vyvozování závěrů

- Ramon Llull (cca 1232—1315) – systém vyvozování popsáný v díle *Ars Magna* (1305), pokus o mechanickou implementaci pomocí systému rotačních papírových koleček vytvářejících různé permutace
- Leonardo da Vinci (1452—1519) – nerealizovaný, ale zřejmě funkčně správný návrh kalkulačky cca 1500
- Wilhelm Schickard (1592—1635) sestrojil mechanickou kalkulačku cca 1623
- Blaise Pascal (1623—1662) sestavil několik kalkulaček, model Pascaline z 1642 byl komerčně nabízen, Pascal analýzou činnosti soudil, že „produkuje efekty, které se blíží lidské mysli víc než jakékoli konání zvířat“.
- Gottfried Wilhelm Leibniz (1646—1716) rozšířil mechanismus kalkulačky o automatické násobení/dělení a měl vizi stroje provádějící operace s jinými objekty než jen s čísly.
- Thomas Hobbes (1588—1679) v knize *Leviathan* (1651) předvídal myšlenku myslícího stroje založeného na pružinách a spojích s tím, že „rozum není nic jiného než „počítání“ jako sčítání a odčítání“.

Pravidla pro vyvozování závěrů

- René Descartes (1596—1650) přišel ve snaze popsat fungování mysli s rozlišením mezi myslí a hmotou (čistě mechanická realizace myšlení by nedávala prostor pro svobodnou vůli - podléhala by fyzikálním zákonům jako hozený kámen) – odtud koncept *dualismu*, kde vědomí existuje mimo hmotný svět, vs. koncept *materialismu*, kde fyzikální zákony vytvářejí vědomí a svobodná vůle je jen dojemem.
- Fyzické vědomí má smyslové vjemy jako zdroj vědomostí, což jako základ tzv. *empiricismu* formuloval Francis Bacon (1561—1626) v aktualizaci Aristotelova *Organonu* v *Novum Organum* a dále rozvinul John Locke (1623—1704) v tvrzení, že „Nic není součástí chápání, pokud to dříve neprošlo přes smysly“.
- David Hume (1711—1776) v *Pojednání o lidské přirozenosti* (1739) zavedl princip dnes nazývaný *principem indukce* – že obecná pravidla se vytvářejí opakovaným vystavováním vztahu jejich prvků.
- Vídeňský kroužek (20.-30. léta 20. století) rozvinul logický pozitivismus jako charakterizace znalosti pomocí logických teorií a observačních sentencí odpovídajících smyslovým vjemům, což kombinuje racionalismus s empirismem

Pravidla pro vyvozování závěrů

- Daniel Bernoulli (1700—1782) navrhl koncept „užitku“ vysvětlující chování z preferencí jednotlivců, kde přímý vztah mezi ziskem a užitkem mizí se zvyšujícím se ziskem (člověk s příjmem 10 tisíc ročně bude považovat zvýšení příjmu o 100 za větší užitek než někdo s příjmem 50 tisíc), dnes jeden ze způsobů jak se rozhodovat při nejistotě a maximalizaci očekávaného užitku.
- V polovině 20. století John von Neumann (1903—1957) a Oskar Morgenstern (1902—1977) publikovali axiomatickou bázi pro teorii užitku: dokud preference zastávané nějakým jednotlivcem splňují nějaké základní axiomy vhodné pro každého racionálního aktéra, pak volby, které jednotlivec dělá, lze nutně popsat jako maximalizaci očekávané hodnoty funkce užitku. Racionální aktér tedy jedná tak, aby maximalizoval očekávaný užitek.
- Kombinací teorie pravděpodobnosti a teorie užitku získáváme obecnou *teorii rozhodování*, zejména pak *teorii her*.
- Dalšími moderními aplikacemi jsou operační výzkum (motivován optimalizací rozmístění britských radarů za 2. světové války) a také metody zpětnovazebního učení.

Matematizace vyvozování závěrů

- Ideu formálních logických pravidel lze vysledovat do starověkého Řecka a Číny
- George Boole (1815—1864) matematizoval výrokovou logiku v práci z roku 1847.
- Gottlob Frege (1848—1925) rozšířil Booleův přístup na logiku prvního řádu (s predikáty a kvantifikátory).
- Pokus o dokázání úplnosti predikátové logiky a existence algoritmu rozhodujícího pravdivé formule v Peanově aritmetice dovedl Kurta Gödela (1906—1978) k důkazu principiální neúplnosti formálních systémů obsahujících aritmetiku, resp. toho, že existují funkce, které nejsou popsateľné jako algoritmus.
- Alan Turing (1912—1954) charakterizoval funkce, které popsateľné jako algoritmus jsou pomocí Turingova stroje, a dokázal i existenci funkcí Turingovým strojem nerealizovatelných, *nevyčíslitelných*.
- Exponenciální časová složitost některých algoritmů z nich dělá algoritmy prakticky nezvládnutelné pro větší rozsah vstupů (a je možné, že to platí i pro třídu NP-úplných problémů) – s většími důsledky pro AI než (ne)vyčíslitelnost.

Zpracování vyvozování v mozku

- Jednobuněčné organismy optimalizují své chování na základě nějakého parametru v ve svém prostředí (např. koncentrace glukózy).
- Jednobuněčné organismy s akčními potenciály a vícebuněčné s neurony již dokážou nejen reagovat na prostředí, ale také se učit a vyvíjet.
- Mozek lze chápat jako komplexní ústrojí uspívající velmi dobře v děláni racionálních rozhodnutí,
- víme ale poměrně málo o tom, jak skutečně funguje.
- Mozek lze ve vztahu k inteligenci vidět v podobné pozici jako křídla ve vztahu k létání.
- Bez ohledu na další (neznámé) detaily formování vědomí nebo uvažování v mozku je zřejmé, že rozhodování vyžaduje paměťovou kapacitu, znalosti, učení se novým rysům, vytváření procedur pro postupy činností a také schopnost simulace.
- Santiago Ramon y Cajal (1846—1934) přišel s technikou obarvování umožňující sledovat jednotlivé neurony a zjišťovat, že souhrn jednoduchých buněk může vést k myšlení, činnosti a vědomí.
- John Searle to cca 1992 formuloval jako „Mozky jsou příčinou mysli“.

Psychologické důsledky

- Studiem chování vědomí se věnuje psychologie, která na makroúrovni ukazuje, jak lidé i zvířata myslí.
- Popis nižších organismů na základě behaviorismu se výrazně liší od chování lidí. Chápání chování zvířat má jen omezenou aplikovatelnost na myšlení člověka.
- Kognitivní psychologie chápe mozek jako zařízení zpracovávající informace a je zřejmě výrazně relevantnější než behaviorismus.
- Frederic Bartlett (1886—1969) a jeho student Kenneth Craik (1943) rozpracovali chování znalostních aktérů založené na tom, že jsou vjemy přeloženy do vnitřní reprezentace, tato reprezentace je zpracovávána kognitivními procesy do jiných reprezentací a ty se mohou promítat do vykonávaných činností.

Počítače

- Charles Babbage (1792—1871) – návrh dvou mechanických počítačů Difference Engine a Analytical Engine, druhý se vstupem a výstupem na děrné štítky; Ada Lovelace analyzovala schopnosti Analytical Engine s tím, že jde o „uvažující stroj“ schopný úvah „o všech subjektech ve vesmíru“, nicméně bez přehnaných očekávání o tom, že by stroj něco sám vymyslel.
- Alan Turing a spolupracovníci v roce 1943: elektromechanický počítač Heath Robinson a elektronkový Colossus, oba pro luštění německých šifer;
- Konrad Zuse v roce 1941: první programovatelný počítač Z-3 (programovací jazyk Plankalkül) v Německu;
- John Atanasoff a Clifford Berry: první elektronický počítač ABC v letech 1940-42 na Iowa State University.
- Miniaturizace a zvyšování výkonnosti (tzv. Mooreův zákon – výkon se zdvojnásobuje každých 18 měsíců od 60.let, speciální hardware pro AI v uplynulém desetiletí dokonce každých 100 dní.
- Turingův článek v roce 1950, *Computing Machinery and Intelligence*, vize inteligentních strojů se slavnou *imitační hrou* (Turingovým testem) jako vyvrácení námitek skeptiků.

Lingvistika

- Noam Chomsky v roce 1957 publikuje převratnou práci *Syntactic Structures*, kde vyvrací do té doby převládající behavioralistickou teorii učení jazyka (F. B. Skinner, *Verbal Behavior*, 1957) a přichází se syntaktickými modely vysvětlujícími tvůrčí schopnosti v jazyce a jeho osvojování.
- Komputační lingvistika resp. Zpracování přirozeného jazyka se stává průsečíkem lingvistiky a umělé inteligence, s významnou oblastí zabývající se reprezentací znalostí – to vše ve formě, která je zpracovatelná počítačem.

Komputační racionalita

- Využití popisu vhodného pro strojové zpracování pro popis „inteligentního“ chování.
- Lidé jsou inteligentní do té míry, v níž se dá očekávat, že pomocí svého konání dosáhnou svých záměrů.
- Běžné vymezení pro stroje se zpravidla míní analogicky jako: Stroje jsou inteligentní do té míry, v níž se dá očekávat, že pomocí svého konání dosáhnou svých záměrů;
- tzv. *standardním modelem* AI je stanovení nějaké nákladové funkce, jejíž minimalizace odpovídá plnění účelu vloženého do stroje.
- Norbert Wiener již v roce 1960 přišel s kritikou standardního modelu v článku *Some Moral and Technical Consequences of Automation*, kde uvádí problematický aspekt vkládání záměrů do stroje optimalizujícího výsledek: „Pokud pro dosažení svých záměrů uijeme mechanizovaného činitele, do jehož operací se nemůžeme volně vměšovat, [...] měli bychom si být raději hodně jistí tím, že záměry vložené do stroje jsou záměry, které skutečně chceme.“
- *Intelligence vs. prospěšnost* – modifikované vymezení podle Stuarta Russela: Stroje jsou *prospěšné* do té míry, v níž se dá očekávat, že pomocí svého konání dosáhnou *našich* záměrů.