

Život a umelá inteligencia – hľadanie analógií

Anton Markoš

Karlova univerzita Praha, Přírodovědecká fakulta
Viničná 7, 128 44 Praha 2
markos@natur.cuni.cz

Abstrakt

Pokus porovnať chovanie živých bytostí s umelou inteligenciou AlphaZero. Konštatujem, že porovnanie znesú organizmy autistické – či už z vrodených príčin, alebo umelo privedené do hlbokoj izolácie od biosféry. Kladiem otázku, či ďalším priblížením k životu by nebola „biosféra“ medzi sebou komunikujúcich umelých inteligencií ponechaných evolúcii.

1 Šachový úvod

Podnet k tejto úvahe mi dal bývalý šachový majster sveta Garry Kasparov v úvodníku k *Science* (7 dec. 2018); upozorňuje v ňom na článok v tom istom čísle, popisujúci, ako sa umelá inteligencia AlphaZero učí hrať zložité stolové hry – šachy, šogo a go (Silver et al. 2018). Kasparov sa vracia k svoju dávne mu prehranému súboju s počítačom Deep Blue (1997) a porovnáva ho s dnešnou šachovou hrou umelej inteligencie (AI). Vtedy počítač mal v pamäti kartotéku snád' všetkých majstrovských zápasov histórie, a programátori do neho vložili i preverené algoritmy hracích (a hrateľných) stratégií; stroj potom ťažil z rýchlosti vyhľadávania optimálneho postupu. Porovnajme, píše Kasparov: „Tak ako lietadlá nemávajú krídlami, ani strojová šachová hra nepostupuje ako šachista. Prvé stroje mali mizerný výkon – úspech sa dostavil až vďaka algoritmu ‚minimax‘ a Moorovmu zákonu – teda nie tak, že by stroj imitoval neopísateľné hlbiny kombinácie ľudskej predstavivosti a vizualizácie. Toto suché konštatovanie značne rozladilo ľudí z oblasti AI: uvedomili si, že k tomu, aby stroj porazil majstra sveta, nepotrebuje mať vôbec žiadne hlboké vhl'ady do svojho konania. [...] [A tak teraz] AlphaZero už začína len s pravidlami šachu, bez znalosti stratégií, aké používajú ľudia. Za pár hodín zohrá sama so sebou viac partii, než je obsah všetkých zaznamenaných partii v dejinách. A v priebehu toho sa učí a poráža staršie šachové programy ako je Stockfish.“ AlphaZero pritom „dáva prednosť aktívnej hre, nie nekonečnému taktizovaniu smerujúcemu k remíze“. A tým, že sa učí sama, neodráža priority programátorov,

ani nepotrebuje neustále prehľadávať obrovské databázy. Stroje sa môžu stať expertmi, nielen nástrojmi expertov ako dosiaľ. A predsa, konštatuje Kasparov, sú prípady, keď si umelá inteligencia „nevšimne“ výnimky zo svojich pravidiel – tam ešte stále dokáže byť človek pred ňou.

V tom istom čísle *Science* komentuje zmienený článok i M. Campbell, ktorý ešte viac zdôrazňuje schopnosť učenia posilňovaním (reinforcement learning) u AI. Samozrejme, upozorňuje Campbell, nejde o akési zavŕšenie úsilia, aj keď zverejnené partie spôsobili v šachovej komunite senzáciu. Ide o stolné hry, u ktorých je celý priebeh hry pozorovateľný pre hráčov, a tí sú len dvaja; hra je navyše deterministická, s nulovým súčtom, a figúry sú diskkrétne, 1 takže simulácia je oveľa ľahšia ako napríklad v prípade pokru. Samotný odborný článok je nad moje sily, a tak odcitujem len túto interpretáciu od jeho autorov: „AlphaZero nahrádza dosiaľ používané individuálne nadobúdanie znalostí a cvik zameraný na jedinú oblasť, a to (1) hlbkovými neurálnymi sieťami (t.j. sieťami prepojenými cez mnoho úrovní organizácie či popisu, bez stanovenia akejkoľvek základnej úrovne) a (2) vyhľadávacím algoritmom použiteľným všeobecne.“ (Silver et al. 2018, 1140).

Netreba snád' zdôrazňovať, že AI je dokonalé „zapuzdrené“ voči svetu, jedinými vstupmi a výstupmi sú symboly jednotlivých ťahov, ktoré vyvolávajú odpoveď tejto mnohovrstvej siete. „Nezaujímam ju nič iného – a ani ju nič zaujímať nemôže: ved' vlastne ani „nevie“, že hrá šach (takže akáže „inteligencia“).

2 „Idiots savants“

Emócie a schopnosť spríadať príbehy z poznaného a komunikovať s druhými sú definitorickými znakmi ľudí, žijúcich v pospolitosti. Čo však, ak v niektorých prípadoch tieto vlastnosti chýbajú?

V ľudskom svete akoby tomuto stavu odpovedali prípady ťažkých autistov, ktorí okrem neobyčajného, priamo zázračného majstrovstva v istej oblasti ľudskej činnosti trpia retardáciou mentálnou, citovou, aj spoločenskou. V *Šachovej novele* Stefana Zweiga je to

¹ „[Hra] mechanická svojimi danosťami, a predsa účinná len pomocou fantázie, ohraničená v geometricky strnulom priestore a pritom neobmedzená vo svojich kombináciách,

neprestajne sa vyvíjajúca, a predsa sterilná, myslenie, ktoré k ničomu nevedie, matematika, ktorá nič nevyráta, umenie bez diel, architektúra bez substancie...“ (Zweig 2013, 24)

prípád geniálneho šachového veľmajstra Mirka Czentovića: „Lebo len čo vstal od šachovnice, kde bol majstrom, akému nebolo páru, zmenil sa na grotesknú, ba takmer komickú figúru; pretože ... zostal celým svojím zjavom a spôsobmi ten istý neotesaný dedinský chlapec, ktorý vymetal izbu u vidieckeho farára.“ (Zweig 2013, 17-8) Nič ho nezaujímalo, nič nevedel, chýbal mu zmysel pre vtip, inotaje, náznaky – a pritom bez akéhokoľvek štúdia, iba pozorovaním niekoľkých amatérskych partii nasal do seba šachové znalosti a hral potom na veľmajstrovskej úrovni. Inak svet nepotrebuje, nezaujímá ho, nerozumie mu. „Neľudský šachový automat.“ (s. 39) Analógia AI?

Porovnajme s druhým protagonistom novely, doktorom B., ktorý vedie plnohodnotný život – až ho nacisti zavú na celé mesiace na samotku: „Nerobili nám nič, ibaže nás vsadili do dokonalého Nič, lebo je známe, že na tomto svete nič nedolieha na ľudskú dušu tak mocne ako Nič. [...] Čakal som vždy znova a znova. Nič sa nestalo. Čakal som, čakal, čakal, myslel som, myslel som, myslel, myslel, až kým ma nerozboleli sluchy. Nič sa nestalo. Zostal som sám. Sám. Sám. [...] [V koncentráku] bol by som videl tváre, bol by som mohol zízať na pole, na káry, na strom, na hviezdu, na čosi, čosi, zatiaľ čo tu stálo okolo mňa večne to isté, večne to isté, príšerne rovnaké.“ (Zweig 2013, 66-73) A v tomto stave na pokraji šílenstva sa mu podarí ukradnúť knižočku so 150 veľmajstrovskými partiami. Naučí sa ich naspamäť, naučí sa prehrávať si ich v duchu, a nakoniec začne hrať partie sám proti sebe. A odrazu sa ocitne opäť na hrane šílenstva, ktorému vďaka šachovej príručke na čas unikol – a bez lekárskej pomoci by do neho aj upadol. Analógia s Czentovićom? Žiadna. Dr. B. z nedostatku senzorických vstupov sklzal do rozpadu osobnosti, zatiaľ čo u Czentovića sa nemalo čo rozpadáť.

Oliver Sacks (napríklad 1993, 1995) sprístupňuje čitateľsky prístupnou formou celú škálu prípadov takýchto ťažkých autistov, pritom geniálnych počtárov, matematikov, hudobníkov, výtvarníkov... (a často kombinujúcich niekoľko takýchto schopností). Ľudí s geniálnou pamäťou, ktorí často nedokážu na nič zabudnúť (ale ani dať do súvislostí), ktorí po prvom počutí naspamäť reprodujú operu, knihu, nakreslia scenériu – ale nič im to nedáva, neprejavujú žiadne emócie, radosť, vtip, nedokážu abstrahovať z jednotlivostí, nechápu metafory, inotaje, gestá ani intonáciu reči, sú väčšinou neschopní komunikácie, nie sú zasiahnutí tradíciou ani kultúrou komunity, do ktorej sa narodili. Čo iní zvládajú implicitne, musí autista nacvičiť, algoritimizovať, pamätať si vo forme akýchsi podmienených reflexov (aj keď mu uniká zmysel celého). Vnímajú predmety tohoto sveta, vedia ich pomenovať, ale svoj vnútorný svet z toho zostrojil nedokážu, aj iní ľudia sú len predmety ako každá iná vec, a svet neprejavuje vôbec žiadnu usporiadanosť, nedáva im zmysel. Nedokážu filtrovať dôležité od ničotného. Aké je vnútro, osobnosť týchto ľudí? Čo sa stalo, že sa zrodili do Ničoho (ale so zvláštnym darom

nevysvetliteľných schopností)? Aká je ich pamäť? Na jednej strane nám neurofyziológia poskytuje predstavu dynamickej, neustále sa prestavujúcej pamäti, na strane druhej veľmi presné, priamo archiválne a nemenné záznamy v mozgu týchto ľudí. Koľko ‚terabytov‘ činí kapacita takýchto zázračných detí-idiotov? Sacks však popisuje aj šťastlivcov, ktorí (obvykle s obetavou pomocou okolia) sa v dospelosti zo svojej samoty aspoň čiastočne vynoria, a dokonca dokážu popísať, v čom žili a žijú. Aj tak ide väčšinou o nacvičené chovanie. I ľudia, ktorí sa profesionálne vypracovali, uvedomili si svoju inakosť a vedia „čo sa má“, zostávajú emočne plochí a pri komunikácii si musia dať veľký pozor: svoju neobyčajnú, jednosmernú inteligenciu s veľkým úsilím dokázali presmerovať, ale netušia, prečo sa majú chovať práve takto. V podstate odporujú chovanie iných tak, ako my odporujeme nálady a želania zvierat; robia vlastne akúsi etológiu ľudí.

Skúsme sa teraz odraziť od nášho druhu a hľadať analógie aj u iných živých bytostí než človek.

3 Nič

Tvrdím, že nemôže dlhšiu dobu existovať živá bytosť, ktorá by nebola v tesnej väzbe na zbytok biosféry, a to isté platí pre komunity (buniek či mnohobunecných organizmov). Takéto súžitie znamená, že si obyvatelia rôznych komunit rozumejú, a evolúcia biosféry nie je len otázkou mutácie a selekcie, ale tiež vzájomného zjedňovania (príklady napr. Markoš 2015, 2016; Markoš & Švorcová 2019; model pozri nižšie Kauffman 2003). Aj náš organizmus predstavuje komplikovaný (tzv. holobiotický) ekosystém zložený z nášho tela a stoviek rôznych druhov (mikro)organizmov v útrobach a iných telových dutinách, i na povrchu. Podobne je to u rastlín. V izolácii sú len rané embryá, disperzné štádiá, alebo výnimočne a krátkodobé útvary, napríklad kolónia plesní na povrchu lekváru. Všetky ostatné organizmy sú súčasťou štruktúr a tokov (energie, látok i informácie) a študovať ich oddelene znamená študovať *naše* artefakty, popisovať *modely živého*. V rôznom stupni oddelené od biosféry sú až naše modelové organizmy pestované bez väzieb na biosféru tam jej úlohu musí zastúpiť vedec, lekár, pekáč, pivovarník a pod.

1. Dokonalými predstaviteľmi izolovaných jedincov sú tak len tzv. *bezmikróbne* mnohobunecné organizmy, ktoré nikdy neprišli do styku s inými organizmami: zvieratá odobrané cisárskym rezom alebo z vajička a chované v sterilnom prostredí, a podobne i rastliny vypestované takto zo sterilných semien. Tam môžeme dosiahnuť dokonale izolované jedince, len s takými vstupmi a výstupmi, aké im dovolíme, t. j. môžeme ich prinútiť k životu v „dokonalejšom“ samote. AI (na súčasnej úrovni) by snáď mohla slúžiť ako model živorenia takýchto samotárov. Pozor však: tieto organizmy nezačali „od nuly“ ako AI. Nesú si pamäť a skúsenosť línie, nazbierané za 4 miliardy rokov evolúcie. Vyvinie sa druhovo špecifické mnohobunecné telo, s intenzívnou

diferenciáciou buniek, ktoré intenzívne komunikujú medzi sebou. Analógia teda dokonalá nie je.

2. Podobné sú kultúry mikroorganizmov (baktérie, kvasinky, protisty), a tiež bunecné kultúry izolované z mnohobunecných organizmov (embryonálne bunky, bunky tkanív, nádorové bunky), obsahujúce populáciu len jediného druhu, klonu, izolátu, mutanta. Ako v predchádzajúcom prípade však môžu jednotlivé indivíduá komunikovať aspoň medzi sebou, a tak veľmi často dochádza k del'be práce v rámci populácie – metabolickej či štruktúrnej (z našej skupiny pozri napr. Pátková et al 2012; Čepl et al. 2016). Len malá časť organizmov sa uspokojí s týmto stavom a sú ochotné sa v takejto „kultúre“ aj množiť.

3. Poslednou skupinou organizmov čiastočne odlúčených od biosféry sú domáce zvieratá a rastliny. Ich holobiotické vzťahy sú viac-menej zachované, ale boli šľachtené na to, aby v prostredí chovov a monokultúr vydržali – ich diví predkovia by to nedokázali, alebo by len biedne prežívali vo zverincoch podobní diktorovi B. z Zweigovej novely. Ich vnútro-populačné vzťahy (hierarchie, sezónne cykly) sú potlačené, sú nútené prežiť svoj život na obmedzenom priestore, s minimom senzorickej vstupov. Snáď len psy sa dokázali z takého stavu vymaniť, keď svoju včľiu národu presmerovali na intenzívnu komunikáciu s človekom. Do tejto skupiny však musíme zaradiť aj niektoré skupiny ľudskej: pracovné tábory, väznice, kasárne, školy (s tým rozdielom, že ide väčšinou o skupiny s krátkou dobou trvania, a účastníci sa pamätajú – a aj sa tešia – na lepšie časy.)

4 Modely živého

Klasická géocentrická teória dedičnosti i evolúcie predpokladá, že živá bytosť dostáva do vienu verziu „databázy“ a „programu“, ktoré v rozhodujúcej miere determinujú jej ontogénu. Epigenetický pohľad na evolúciu môže hľadiť na tento základ ako na pravidlá hry, ktorej varianty začne bunka rozvíjať na základe interpretácie svojej pamäti (bezprostrednej i historickej), svojej vlastnej skúsenosti, a tiež interpretácie vonkajších faktorov – a k tým patrí hlavne orientácia v komplikovaných biosférických vzťahoch. Pamäť ani skúsenosť teda nemusia byť iba zapísané, sú súčasťou celkového nastavenia organizmu a jeho porozumenia situácie, do ktorej sa zrodili.

Modely definujúce život sú početné a žiadny nie je uspokojivý a hlavne, v obave pred vitalistickou herézou sa ich autori boja vysloviť prostú pravdu, že živé je živé. A tak sú organizmy prirovnávané k chemickým reakciám, schopným darvinovskej evolúcie (tzv. „definícia NASA“), k disipatívnym štruktúram hnaným druhým termodynamickým zákonom, (napr. Kováč 2017) k počítačom bežiacim pod genetickým softwarom

(napr. Dawkins 1998) a podobne; všetky pri bližšom pohľade postihnú len isté vlastnosti života, ale nedokážu ho pojať v jeho celku.² Realistickejší je vitalistický model *autonómnych agentov* (Kauffman 2003). Autonómny agent je najprv definovaný ako entita, ktorá vie (1) uskutočňovať aspoň jeden termodynamický pracovný cyklus a (2) množiť sa. Rýchlo sa však ukáže, že potrebuje mať aj početné senzory a aj schopnosť interpretovať vstupy týchto sensorov, aby mohol efektívne komunikovať s inými agentmi v *biosfére* autonómnych agentov. Výsledkom takej komunikácie je zjednávanie – negociácia najbližších budúcich stavov biosféry. Zdá sa mi, že modely založené na AI by mohli byť ďalšou výzvou. Môžeme si predstaviť biosféru veľkého počtu AI, a navyše ich autonómu kvázibiologickú evolúciu intenzívne komunikujúcich línií, populácií, spoločenstiev?

Pre tých, čo rozumejú AI a chcú sa vydať týmto smerom, ponúkam antropomorfný scenár, na ktorom pracujem v posledných rokoch (pozri Markoš 2015, 2016; Markoš & Švorcová 2019). Je inšpirovaný Darwinom, Kauffmanom a biosemiotikou. Biosemiotika je vetvou semiotiky, náuky, ktorá sa zaoberá interpretáciou znakov a hľadaním významov – znakov a významov nie nutne jazykových. Uvediem ho analógiou ľudskej komunity, ktorá si určuje isté kodifikované normy – morálku, tabu, chovanie pri rôznych príležitostiach, ústavu, náboženstvo, dejiny, spisovný jazyk... Nikto z členov komunity z princípu nemôže striktno dodržiavať všetky tieto normy, väčšinou však stačí, že ich pozná a uznáva – aj keď interpretácia výsledného chovania môže byť rôzna v rôznych vrstvách spoločnosti a v rôznych dobách. Vďaka existencii takýchto noriem je možná i komunikácia s inými skupinami, ktoré sa definujú síce inak, ale aspoň rámcovo dokážu tomu druhému *nejako* porozumieť – vďaka spoločnej minulosti biologickej, kultúrnej, náboženskej a inej. Samozrejme takéto rozumenie môže vyústiť aj do neporozumenia, omylov a konfliktov. Ponúkame biologickú verziu: dnešná biosféra je produktom evolúcie z biosféry prvotnej, kedy sa ustalovali normy (metabolizmus, genetické procesy, signály atď.) Jednotlivé línie organizmov s týmito normami pracovali po svojom, predsa však sa interpretácie sveta do istej miery prekrývajú – a vďaka tomu biosféra môže fungovať ako celok. Živé bytosti teda budujú *svoj* model sveta (1) na základe zdedeného genetického zápisu; (2) na základe historického úzu platného v danej línií, ako sa má tento zápis interpretovať; a (3) na základe interpretácie vlastností tej komunity, do ktorej sa zrodil. Slovom Z. Neubauera: „Znak, chování nevzniká z nutnosti: a je-li zde, nastává hra, co všechno se s ním dá udělat – až do absurdit a monstrosit, které se konečně stávají někdy fatální i pro přežití. Vidíme všude hru a harmonii, fantasií a

² Z. Neubauer: „Nesluší se např. biologovi ptát se, co je to život: taková otázka je považována za nevědeckou nikoliv

proto, že věda na ni nemůže odpovědět, ale proto, že se implicitně předpokládá, že na ni věda samozřejmě odpověď dávno má!“ (Kováč & Neubauer 2019, 55)

velkolepost projektů, z nichž mnohé ztroskotávají, ještě než došlo k jejich realizaci. Evoluce – toť dějiny velkých ‚neúspěchů‘, omylů, doknquijotstství i velkých koncepcí, které stvořily epochu.“ (Kováč & Neubauer 2019, 85).

5 Závěr

Čo väčšina autorov prisudzuje iba človeku, skúsme „povolit“ každej forme života. Ešte raz Neubauer: „Zkrátka: je třeba nemilosrdně odstranit narcisistický tabuismus tzv. ‚anthropomorfismů‘. Pokrytectví zákazu anthropomorfního uvažování a anthropomorfních představ ve vědě spočívá v tom, že předstírá, že nesmíme vnucovat své lidské vlastnosti mimolidským jsoecnům. Ve skutečnosti však tato dojemná ohleduplnost přikazuje *upírat* přírodě vše a ponechat jí pouze pasivní použitelnost.“ (Kováč & Neubauer 2019, 66).

Pod'akovanie

Podporené programom Univerzitní výzkumná centra UK č. 204056.

Literatúra

Campbell, M. (2018) Mastering board games. *Science* 362, 1118; DOI: 10.1126/science.aav1175.

Čepl, J., et al. (2016) Variations and heredity in bacterial colonies. *Communicative & integrative biol.* 9, e1261228; DOI: 10.1080/19420889.2016.1261228

Dawkins, R. (1998 [1976]) *Sobecký gen*. Praha, Mladá fronta.

Kasparov, G. (2018) Chess, a *Drosophila* of reasoning. *Science* 362, 1087; DOI: 10.1126/science.aaw2221

Kauffman, S. (2003) *Čtvrtý zákon. Cesty k obecné biologii*. Praha, Paseka.

Kováč, L. (2017) *Konec lidské evoluce*. Červený Kostelec, Pavel Mervart.

Kováč, L. & Neubauer, Z. (2019) *Listy o biologii. Korespondence 1975-1990 a další texty*. L. Nováková & A. Markoš eds. Červený Kostelec, Pavel Mervart.

Markoš, A. (2015) *Znaky a významy v evoluci*. Praha, Nová beseda.

Markoš, A. (2016) *Evoluční tápání*. Červený Kostelec, Pavel Mervart.

Markoš, A. & Švorcová, J. (2019) *Epigenetic processes and the evolution of life*. CRC Press.

Pátková, I., et al. (2012) Developmental plasticity of bacterial colonies and consortia in germ-free and gnotobiotic settings. *BMC Microbiol* 12: 178; DOI: 10.1186/1471-2180-12-178.

Sacks, O. (1993) *Muž, který si pletl manželku s kloboukem*. Praha, Mladá fronta.

Sacks, O. (1995) *Antropoložka na Marsu*. Praha, Mladá fronta.

Silver, D., et al. (2018) A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* 362, 1140-1144; DOI: 10.1126/science.aar6404

Zweig, S. (2013 [1942]) *Šachová novela*. Slovart, Bratislava.