

Využitie umelej inteligencie pre potreby evolučnej psychológie v kontexte kognitívnych vied

Peter Novák, 5iui
FMFI UK
pno@whitestein.com

13. februára 2002

*Kľúčom nieje materiál z ktorého je stvorený náš mozog, ale vzory,
ktoré sa tvoria v hmote mozgu.*

(D. G. Hofstadter, GEB: an Eternal Golden Braid)

1 Evolučná psychológia

Na prvých stranách knihy D. G. Hofstadtera, GEB: an Eternal Golden Braid nájdeme kúsok textu použitý ako motto tejto práce. Dost' efektne vyjadruje moje pocity, ktoré som mal pri svojom prvom stretnutí s evolučnou psychológiou. Zvyšok tejto práce bude teda subjektívnym pohľadom na to čo ma (ako autora) napadlo pri čítaní práce [1], ktorá mi bola úvodom do evolučnej psychológie.

Z pohľadu evolučnej psychológie je ľudská myseľ spletencom modulov, či lepšie malých programov, ktoré svojou činnosťou a kooperáciou vytvárajú to čo nazývame správaním. Úlohy a funkcie týchto mini-programov sa vyvinuli podobným mechanizmom ako popisuje Charles Darwin vo svojej evolučnej teórii o pôvode druhov. Pokiaľ dobre chápem evolučnú psychológiu, snaží sa nám povedať, že mnohé z týchto programov a modulov niesú všeobecné mechanizmy schopné vykonávať svoju funkciu a tým riešiť úlohy nad ľudovoľnými doménami, ale že často medzi nimi nájdeme také, ktoré sú špecializované na jednu konkrétnu úlohu s konkrétnymi vstupmi. Tento fakt sa dá demonštrovať na experimente s riešením konfliktov v podmienených pravidlách. Ukazuje sa, že ľudia majú zvláštnu schopnosť rozoznávať v sociálnych situáciách klamárov a podvodníkov. Ukazuje sa ([1]), že v máme v našom mentálnom vybavení schopnosť

rýchlo riešiť podmienené vzťahy týkajúce sa spoločenskej výmeny - situácií v ktorých je napríklad sľúbená odmena za protislužbu. Táto naša schopnosť je však tak úzko špecializovaná, že dokonca už nie sme tak pohotoví v odhaľovaní tých, ktorí nám chcú dobre (sociálni altruisti), teda tých, ktorí za svoju službu nežiadajú odmenu. Tieto mechanizmy prekvapivo nefungujú v situáciách, kedy ide o tú istú logickú úlohu, ale nad inou doménou, napríklad keď sformulujeme úlohu ako logickú hádanku týkajúcu sa podmienených udalostí v čase.

Z tohto jednoduchého, no o to prekvapivejšieho experimentu môžeme odvodiť, že v mysli naozaj existujú isté moduly, ktoré sú prísne doménovo špecializované. Toto ma privádza na podobnosť názorov evolučnej psychológie s niektorými oblasťami umelej inteligencie a zväzda podiskutovať o tejto podobnosti v kontexte umelej inteligencie a jej využítí v kognitívnych vedách.

2 Subsymbolové prístupy umelej inteligencie

V umelej inteligencii sa stretávame s dvoma hlavnými prúdmi vedeckého výskumu. Rozdelenie na *symbolové* a *subsymbolové* prístupy k riešeniu problémov umelej inteligencie je hlavnou dichotómiou tohto vedného odboru.

Symbolové prístupy sa dajú charakterizovať snahou o navrhnutie všeobecných algoritmov a operácií s dátami, ktoré by mali byť nezávislé na doméne, založené na matematickej logike a je tu snaha udržiavať konzistentné teoretické pozadie navrhnutých metód. Toto všetko vedie k tomu, že nech takto navrhnutý program, modul či operátor odvodí akýkoľvek výsledok zo zadaných vstupov, vždy je možné z jeho postupu, teda zo záznamu o behu algoritmu, odôvodniť prečo a na základe akých argumentov daný algoritmus dospel k práve tomu výsledku, ku ktorému dospel. Nakoniec je možné toto odôvodnenie výsledku artikulovať v prirodzenom jazyku. Medzi takéto systémy patria napríklad expertné systémy, alebo systémy založené na logickom programovaní.

Subsymbolový prístup sa na druhej strane snaží o tvorbu algoritmov, ktoré, na rozdiel od symbolovo orientovaných metód, napriek tomu, že nie je možné odôvodniť výsledok, ktorý poskytnú, fungujú nad vstupmi z danej domény rýchlo a zvyčajne sa nemýlia viac ako je potrebné. Takéto programy proste zo vstupov vypočítajú výsledok, no v konečnom dôsledku nie je možné odôvodniť, prečo je k daným vstupom práve ten výstup, ktorý algoritmus vypočítal, ten správny. Sú rýchle a ak sú dobre navrhnuté, zvyčajne sú i dostatočne presné, no navonok vyzerajú ako akési čierne skrinky do ktorých nevidno a nevedno odkiaľ sa v nich vzal výsledok, ktorý vypočítajú. Systémami, ktoré spadajú do tejto kategórie sú napríklad neurónové siete a evolučné algoritmy.

3 Subsymbolové prístupy a evolučná psychológia

Ako už bolo možné vybrať z predchádzajúcej časti, v ďalšom sa budem zaoberať podobnosťou niektorých subsymbolových prístupov a niektorých myšlienok evolučnej psychológie a využitím takýchto prístupov ako výpočtových modelov pre potreby evolučnej psychológie.

Evolučná psychológia, na rozdiel od umelej inteligencie, ktorá je viac menej založená na matematike a logike, nemá explicitné nástroje na hlbšiu analýzu dôvodov a dôsledkov svojich hypotéz. Môže sa opierať len o empirické výsledky získané z experimentov. Možno by sme mohli dať do analógie niektoré problémy s ktorými sa potýka umelá inteligencia a tak získať nový vhl'ad do evolučnej teórie.

Pri návrhu softwérového systému sa inžinier vždy stretne s úlohou analýzy riešenej úlohy. Musí odôvodnene rozdeľovať problém na podúlohy a tie na ešte menšie podproblémy. Takto je možné nakoniec dopracovať sa k algoritmom a metódam riešiacim hlavný problém poskladaním čiastkových riešení podproblémov. Je jednoduché pobaďať, že táto dogma neplatí v subsymbolových prístupoch umelej inteligencie. Pri riešení problému subsymbolovým prístupom, zvyčajne návrhár vytvorí z riešeného problému sadu príkladov a snaží sa rôznymi učiacimi metódami "naučiť" svoj všeobecný program efektívne vypočítať výsledok z daných vstupov. Ak je jeho snaha korunovaná úspechom, na konci má program schopný riešiť dostatočne spoľahlivo danú úlohu nad danou doménou, no bolo by nesmiernym hazardom pokúšať sa použiť takto "vyemergovaný" program nasadiť na riešenie hoci aj podobnej úlohy nad inou doménou nieto ešte inej úlohy.

Vývoj správání, či inteligencie, za predpokladu, že inteligenciu priznáme všetkým živým organizmom, je možné vidieť v podobnom svetle. Príroda, či možno Boh, stáli pred úlohou naučiť akýsi prvý "živý" organizmus, schopný adaptácie, existovať v prostredí našej planéty. Tento tvorca spravil presne to, čo robí inžinier ak používa subsymbolový prístup na riešenie svojej úlohy. Vrhol populáciu nášho hypotetického praorganizmu do prostredia v ktorom naň už čakali rôzne úlohy na ktoré bol náš tvor nútený adaptovať a bol nútený adaptovať sa na riešenie rôznych špeciálnych problémov. Táto podobnosť samozrejme nie je náhodná, špeciálne evolučné algoritmy vznikli ako metóda inšpirovaná princípmi Darwinovej evolúcie.

Na príklade evolučných algoritmov však vidno ako je charakterizovaný výsledok tejto evolúcie. Výsledný evolučný algoritmus robí to čo robí, je však veľmi ťažké analyzovať, prečo dáva správne výsledky v danej doméne pre daný problém. Podobne sa mi vidí, že síce evolučná psychológia je schopná poukázat' na skutočnosť, že existujú konkrétne moduly, ktoré vznikli evolúciou, nie je však už schopná povedať, prípadne presvedčivo argumentovať prečo práve tieto psy-

chické mini-programy vznikli a ako fungujú. Tiež si neviem predstaviť, ako by bolo možné skúmať prepojenie medzi týmito modulmi, ich množstvo a kooperáciu, keďže ide o ten istý rekurzívny problém. Totiž prepojenia a kooperácia týchto modulov sú tiež len produktom evolúcie, čo podobne ako v umelej inteligencii vedie k tomu, že systém ako celok sa nám javí ako čierna skrinka, ktorá síce niečo robí, no nevieme povedať ako a prečo.

Samozrejme toto nieje problém evolučnej psychológie. Táto si ani nekladie také vysoké ciele ako je analýza inteligencie, či správania. Chce len na základe zistených a odvodených znalostí prísne logicky vysvetliť javy v spoločnosti a v konaní jednotlivcov ľudskej komunity. Vyššie uvedenými vetami som chcel poukázať len na to, že v kontexte skúmania a analýzy ľudskej mysle, čím sa podľa mojich znalostí zaoberajú kognitívne vedy, nám evolučná psychológia pomôže len tým, že pomerne presvedčivo tvrdí, že existujú akési vrodené mechanizmy v ľudskej psychike, ktoré sú už pri narodení doménovo závislé, prípadne, že máme vrodené akési sklony počas života nadobudnúť niektoré obsahovo špecifické mentálne schopnosti, ktoré nie sú produktom učenia.

Ak teda mám odvodiť nejaký záver pre túto časť, povedal by som, že subsymbolové prístupy ako analógia k myšlienkam evolučnej psychológie, nie sú celkom vhodní kandidáti, pretože nás nevedú k poznaniu štruktúry a mechanizmov myslenia, len k slabým tvrdeniam o existencii istých modulov. Výpočtové modely založené na subsymbolových prístupoch preto pokladám za slabý nástroj na progresívne využitie v evolučnej psychológii.

4 Multiagentové systémy

Ak sa máme predsa len zaoberať vhodnou alternatívou medzi symbolovými a subsymbolovými prístupmi, ktoré by sme mohli dať do analógie s myšlienkami podobnými nápadom z evolučnej psychológie, mohli by sme ešte na záver venovať priestor multiagentovým systémom. Tieto systémy sa tiež dajú charakterizovať ako prepojenie a kooperácia mnohých modulov, či mini-programov, no ich spojenie je založené na pomerne logickej štruktúre. Tieto moduly sú však produktom cieľavedomej analýzy návrhára systému a teda je možný vhl'ad do ich činnosti a dokonca je zvyčajne možné hľ'adať odôvodnenia výstupov takýchto systémov. Podobnosť s teóriou evolučnej psychológie je v konekcionizme multiagentových systémov.

Samozrejme umelé experimenty realizované pomocou multiagentových systémov môžu len potvrdiť, respektíve podporiť hypotézy evolučnej psychológie na výpočtových modeloch. Nie je možné samostatným skúmaním umelých multiagentových systémov, na rozdiel od neurónových sietí a evolučných algoritmov, odvodzovať hypotézy, ak už nie použiteľné, tak aspoň zmysuplné v evolučnej

psychológii. Podporu tejto myšlienky vidím v obsahu knihy *Spoločenstvo mysle* od Marvina Minskeho ([3]) v ktorej uvádza svoj model mysle ako spoločenstva poprepájaných, kooperujúcich agentov-modulov, výsledkom spoločnej činnosti ktorých je správanie.

5 Kritika evolučnej psychológie v kontexte kognitívnych vied

V predchádzajúcom texte som sa už čiastočne dotkol problému potreby zdôvodňovania činnosti hypotetických programov, z ktorých sa skladá myseľ (ak ju chápeme ako veľkú funkciu realizujúcu správanie v danom prostredí). Každá vedecká disciplína, či už ide o humanitnú, alebo prírodnú vedu nevyhnutne musí byť založená na dokazovaní a zdôvodňovaní dosiahnutých výsledkov. Inak by nešlo o exaktnú činnosť. Preto si dovoľím z tejto pozície kritizovať používanie niektorých sub-symbolových prístupov ako overovacích výpočtových modelov, v kontexte nielen evolučnej psychológie, ako zavádzajúce a vedúce buď k triviálnym výsledkom, alebo k bludným hypotézam.

Tiež si dovoľím na záver tejto práce zapochybovať o evolučnej psychológii ako o vednej disciplíne, ktorá by v mne známej podobe mohla viesť k dostatočnému vysvetlenie mechanizmov myslenia ľudskej bytosti. Nevytýkam jej metódy ani výsledky, ale charakter získaných výsledkov. Mám totiž pocit, že nie je schopná ísť ďalej ako k poukázaniu na existenciu vrodenných doménovo špecifických modulov v mysli a k objaveniu ďalších takýchto modulov. Dostatočne odôvodniť a bližšie objasniť ich štruktúru je však asi za hranicami ľudskej možnosti.

V tejto chvíli sa mi teda žiada ukončiť tento elaborát pohľadom na funkciu kognitívnych vied ako spojenia rôznych vedných disciplín s cieľom objasniť fungovanie ľudskeho myslenia. Mám pocit, že sme schopní konvergenciou aj výpočtových metód s psychologickými disciplínami čiastočne vhliaďnuť do toho *ako to funguje*, ale nie viac.

Referencie

- [1] Leda Cosmides, John Tooby, *Evolutionary Psychology: Primer*, '90-te roky, University of California, Santa Barbara
- [2] Douglas R. Hofstadter, *Gödel, Escher, Bach: an Eternal Golden Braid*, 1979 Basic Books Inc., New York
- [3] Marvin Minsky, *The Society Of Mind*, 1988, Simon & Schuster Inc., New York