Princípy dizajnu rozumejúcich agentov

Martin Takáč

Abstract

In this paper, we summarize the results of our ongoing research on autonomous construction of meanings in artificial agents. We propose design principles for building „understanding" agents: interaction-grounded incremental and continuous individual construction of meanings, necessity to represent dynamic environmental aspects and social coordination of meanings. We describe several implemented computational models proving the feasibility of the proposed principles.

Podstatným atribútom inteligencie je flexibilita, čiže schopnosť adaptovať sa na meniace sa podmienky. Schopnosť učiť sa významne akcelerovala rozvoj ľudstva, pretože biologická evolúcia (genetický prenos) je v porovnaní s učením (kultúrnym prenosom) omnoho pomalšia. Čím je prostredie dynamickejšie, tým menej inteligentný systém vystačí s nemennými, hotovými, vrodenými, či predprogramovanými vedomosťami. Z toho môžeme vyvodiť dôsledky pre dizajn umelých inteligentných systémov.

Autonómne agenty – softvérové entity vykonávajúce nejakú činnosť vo virtuálnom prostredí sa čoraz viac stávajú súčasťou internetových aplikácií, či už v oblasti e-commerce, e-learning, monitorovania bezpečnosti, či internetových prehľadávačov. Potreba automatického spracovania webovských stránok na základe ich sémantického obsahu motivovala výskum v oblasti Sémantického webu (Fensel et al., 2000). Mnoho vedeckého a programátorského úsilia je venovaného tvorbe ontológií a jazykov pre reprezentáciu významu webovských stránok. Internet je však dynamické a otvorené prostredie, ktoré žiadna fixná ontológia nemôže raz a navždy opísať. Preto sa pre dizajn inteligentných agentov ponúka alternatívna cesta: namiesto preddefinovaných ontológií vybaviť agenty silnými učiacimi mechanizmami, ktoré im umožnia samostatne konštruovať adekvátnu reprezentáciu relevantných aspektov prostredia až za behu (v run-time) na základe interakcií s konkrétnym prostredím.

Pre vzájomnú komunikáciu medzi agentami je nutné individuálne konštruované reprezentácie navzájom koordinovať – spoločný komunikačný systém a jeho sémantika vzniká ako emergentný výsledok distribuovaných lokálnych interakcií vo viacerých spriahnutých komplexných adaptívnych dynamických systémoch (Steels, 2000; Staab, 2002). Takéto agenty sú flexibilnejšie a robustnejšie, pretože sa dokážu naučiť „rozumieť" aj tým vlastnostiam prostredia, s ktorými dizajnéri nepočítali. Učenie je inkrementálne a kontinuálne, čo šetrí zdroje (programátorskú prácu), keďže pri zmenách prostredí nie je nutné zakaždým agenty preprogramovávať a nahradzovať novými verziami.

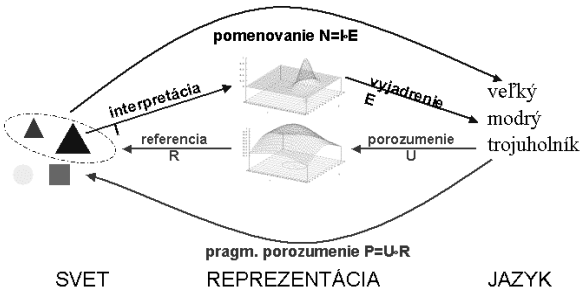
Náš doterajší výskum bol zameraný na charakterizáciu rozumenia a významov v umelých systémoch (Šefránek et al., v tlači) a formuláciu princípov dizajnu „rozumejúcich" agentov (Takáč, v tlači). Realizovateľnosť sformulovaných princípov sme prakticky overili vo výpočtových modeloch konštrukcie významov na základe senzomotorickej interakcie s prostredím (Takáč, 2006) a na základe sociálnej inštrukcie (Takáč, prijaté). Skúmali sme aj stabilitu naučených významov v iterovanom medzigeneračnom prenose (Takáč, 2007a). Významy vo všetkých výpočtových modeloch boli založené na tzv. sémantike identifikačných kritérií (Šefránek, 2002; Takáč, 2007b), ktorej rozpracovanie považujeme za náš najvýznamnejší teoretický prínos. Teraz v stručnosti zrekapitulujeme doteraz dosiahnuté výsledky (už bez uvádzania referencií).

Pre dizajn "rozumejúcich" agentov sú dôležité tieto princípy:

- ukotvenie významov (reprezentácií) v interakciách s prostredím,
- inkrementálna a kontinuálna konštrukcia významov,
- nutnosť reprezentácie dynamiky meniaceho sa prostredia a komplexných sémantických kategórií,
- sociálna koordinácia individuálne konštruovaných významov.

V súlade s týmito princípmi sme implementovali niekoľko výpočtových modelov konštrukcie významov reprezentovaných pomocou tzv. identifikačných kritérií. Každé identifikačné kritérium funguje ako detektor vyladený na nejaký konkrétny koncept: ak dostane na vstup dobrý príklad reprezentovaného konceptu, reaguje silnou aktivitou (blízkou 1), v opačnom prípade reaguje slabou aktivitou (blízkou 0). Aktivita teda vyjadruje mieru príslušnosti vstupu v reprezentovanej kategórii.

Dôležitou vlastnosťou identifikačných kritérií je, že môžu byť skonštruované na základe množiny príkladov reprezentovanej kategórie a ďalej „dolaďované" ďalšími prichádzajúcimi príkladmi. Umožňujú teda inkrementálne a permanentné učenie agentov. Výhodou navrhnutého učiaceho mechanizmu oproti iným modelom je jeho citlivosť na korelácie atribútov inštancií s príslušnosťou ku kategórii a na vzájomné korelácie medzi atribútmi. Skonštruovanú reprezentáciu možno priamo prepojiť s jazykom (obr. 1), čím sa rieši *Symbol Grounding Problem* (Harnad, 1990): významy jazykových výrazov sú ukotvené v interakčne skonštruovaných identifikačných kritériách. Na rozdiel od mnohých existujúcich prístupov, konštrukcia identifikačných kritérií je založená na medzi-situačných podobnostiach inštancií konceptov, a nie na rozdieloch medzi zvoleným objektom a ostatnými aktuálne prítomnými objektmi, čo je predpoklad pre situačne nezávislé používanie jazyka.



Obrázok 1. Pragmatické funkcie zabezpečujú prepojenie prostredia, vnútorných reprezentácií a jazyka.

Elementárne identifikačné kritériá klasifikujúce jediný vstupný objekt slúžia na reprezentáciu monadických konceptov indivíduí, tried a vlastností. Zložitejšie kritériá vzťahov medzi objektmi, zmien v čase, či situácií a udalostí sa konštruujú z jednoduchších pomocou transformácie vstupov. Napríklad kritérium vzťahu medzi dvoma objektmi dostane na vstupe objekt skonštruovaný ako rámec rozdielov v hodnotách relevantných atribútov dvoch pôvodne porovnávaných objektov. Pseudoformálne zapísané,

```
vľavo_od(objekt_1,objekt_2) := objekt_1.polohaX < objekt_2.polohaX.
```

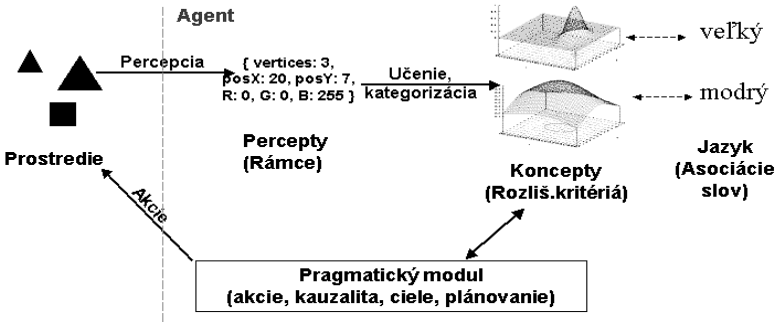
Zmeny objektu v čase sa redukujú na vzťah medzi vnímanými podobami objektu v dvoch po sebe nasledujúcich okamihoch. Kritérium situácie je hierarchický graf zložený z elementárnejších monadických a vzťahových kritérií.

Mechanizmus konštrukcie identifikačných kritérií je založený na indukcii z predložených príkladov konceptu, ktorý sa má reprezentovať. Agent si postupne

vytvára a dolaďuje mnoho kritérií, preto potrebuje rozhodnúť, ktoré vnímané objekty (resp. n-tice objektov) sú príkladmi ktorých kritérií. To môže urobiť viacerými spôsobmi:

a) klasterizáciou (bez učiteľa) na základe prirodzenej distribúcie vlastností objektov v prostredí (tzv. environmentálne relevantné koncepty),
b) vlastnou interakciou s prostredím a združovaním na základe podobných interakčných vzorcov (pragmatickú spätnú väzbu poskytuje prostredie – tzv. ekologicky relevantné koncepty)
c) s učiteľom – na základe jazykových inštrukcií (pomenovaní) iným agentom – agent indukuje význam pojmu z množiny všetkých referentov pomenovaných rovnakým slovom (kultúrny prenos kategórií).

Prvá séria implementovaných experimentov preverila funkčnosť navrhnutého mechanizmu individuálneho vytvárania kategórií na základe senzomotorickej interakcie s prostredím. Agent vykonával s objektmi v simulovanom prostredí rôzne činnosti a pozoroval ich dôsledky. Reprezentácie tried objektov, tried akcií a dôsledkov akcií a ich vzájomné asociácie agent konštruoval na základe rovnakých interakčných programov (afordancií). Vytvorená reprezentácia umožnila agentovi efektívne predikovať dôsledky vlastných akcií, čo tvorí dobrý základ pre rozšírenie modelu o autonómnu motiváciu (reprezentácia cieľov, zámerov a plánov). Architektúra agenta je na obr. 2.



Obrázok 2. Kognitívna architektúra agenta zahŕňa percepčnú, reprezentačnú, jazykovú a pragmatickú vrstvu.

V druhej sérii experimentov sme ukázali, ako si učiaci sa agent na základe verbálnej inštrukcie a neverbálnej referencie skonštruuje významy, ktoré sú pre plnenie pragmatických cieľov dostatočne podobné významom učiteľa. Vysoká podobnosť bola dosiahnutá rýchlo, čo je v súlade s pozorovaným fenoménom rýchleho učenia (fast mapping) u detí.

Tretia séria experimentov bola zameraná na skúmanie stability navrhnutej reprezentácie v medzigeneračnom prenose. Ukázalo sa, že medzigeneračný prenos významov môžeme chápať ako evolučný proces, v ktorom sú významy replikátormi súťažiacimi o prežitie, pričom selekčným tlakom je zúžený profil prenosu významov. Významy prejdú cez zúžený profil, ak sú relevantné prostrediu (teda ich inštancie sa v ňom vyskytujú dostatočne často). Ak necháme proces akvizície významov iterovať, významy neostanú nezmenené. Aj keď v každej generácii ostane vysoká podobnosť medzi význammi učiteľa a žiaka, medzigeneračne sa významy budú posúvať a vyvíjať. Tieto výsledky zodpovedajú tomu, že jazyky podliehajú historickým premenám bez toho, aby stratili svoju dorozumievaciu funkciu.

V súčasnosti je náš výskum zameraný na rozšírenie modelu individuálnej konštrukcie významov na základe interakčných vzorcov (afordancií) o autonómne plánovanie a jeho integráciu s modelom sociálnej koordinácie pomocou jazyka. V takomto modeli môžu agenty navzájom kooperovať pri dosahovaní cieľov. Potreba kooperácie pri dosahovaní cieľov je považovaná za hlavnú evolučnú hybnú silu pre situačne oddelené používanie jazyka (Gärdenfors, 2004). Validitu navrhnutých mechanizmov preveríme porovnaním efektívnosti dosahovania cieľov vo variantoch modelu bez komunikácie a so vzájomnou lingvistickou komunikáciou agentov.

Druhou veľkou témou je dopracovanie sémantickej reprezentácie udalostí založenej na identifikačných kritérách (reprezentácia udalostí bola v predchádzajúcich prácach nedopracovaná, resp. zjednodušená). Prvotným nápadom je reprezentovať udalosť ako dvojicu identifikačných kritérií situácií (*počiatočná situácia, cieľová situácia*), pričom v grafe počiatočnej situácie budú hrany zodpovedajúce kritériám elementárnych zmien, ktoré zapríčinili transformáciu počiatočnej situácie na cieľovú. Ďalším konštruktom bude rolový rámec - súbor metahrán určujúcich roly vykonanej akcie (agens, objekt, inštrument, atď.) a jej spôsob vykonania (identifikačné kritérium akcie). Takýto spôsob reprezentácie otvára možnosti na interakčné konštruovanie kritérií udalostí vykonávaním vlastných akcií a pozorovaním akcií iných agentov, pričom sa rieši *problém vymedzenia* (ktoré statické a dynamické aspekty zahrnúť do reprezentácie udalosti a ktoré nie).

Na záver chceme zdôrazniť, že problematika dizajnu „rozumejúcich“ agentov s interakčnou konštrukciou významov je komplexná, obsahuje mnoho otvorených problémov a je veľmi plodnou oblasťou pre dlhodobý výskum.

Poďakovanie

Tento príspevok bol podporený výskumným štipendiom pre mladých vedeckých pracovníkov FMFI v rámci projektu ESF č. JPD BA 3 2005/1-043.

Literatúra

Fensel, D., et al. (2000): Trends and Controversies: The Semantic Web and its Languages. *IEEE Intelligent Systems* 15 (6), s. 76–73.

Gärdenfors, P. (2004): Cooperation and the evolution of symbolic communication. In: Oller, K., Griebel, U. (Eds.): *The Evolution of Communication Systems*. MIT Press, Cambridge, MA, s. 237–256.

Harnad, S. (1990): The symbol grounding problem. *Physica D* 42, 335–346.

Steels, L. (2000). Language as a complex adaptive system. In: Schoenauer, M. (Ed.): *Proceedings of PPSN-VI*. Springer, Berlin, s. 17–26.

Staab, S., et al. (2002): Trends and Controversies: Emergent Semantics. *IEEE Intelligent Systems* 17 (1), s. 78–86.

Šefránek, J. (2002): Kognícia bez mentálnych procesov. In: Beňušková, Ľ. et. al. (Eds.): *Kognitívne vedy*. Kaligram, Bratislava, str. 200–256.

Šefránek, J., Takáč, M., Farkaš, I. (v tlači): Vznik inteligencie v umelých systémoch. In: Magdolen, D.: *Hmota, život, inteligencia: Vznik*. VEDA, Bratislava.

Takáč, M. (2006): Categorization by Sensory-Motor Interaction in Artificial Agents. In: Fum, D., Del Missier, F., Stocco, A. (eds.): *Proceedings of the 7th International Conference on Cognitive Modeling*. Edizioni Goliardiche, Trieste, Italy, s. 310-315.

Takáč, M. (2007a): Konštrukcia významov a jej dynamika v procese iterovaného učenia. In: Kelemen, J., Kvasnička, V., Pospíchal, J. (eds.): *Kognice a umělý život VII*. Slezská univerzita, Opava, s. 341-347.

Takáč, M. (2007b): Kognitívna sémantika komplexných kategórií založená na rozlišovacích kritériách. In: Kvasnička, V., Trebatický, P., Pospíchal, J., Kelemen, J.: *Myseľ, inteligencia a život*. Vydavateľstvo STU, Bratislava, s. 339-355.

Takáč, M. (prijaté): Autonomous Construction of Ecologically and Socially Relevant Semantics. *Cognitive Systems Research*.

Takáč, M. (v tlači): Construction of Meanings in Living and Artificial Agents. In: Trajkovski, G., Collins, S. G. (eds.): *Agent-Based Societies: Social and Cultural Interactions*, IGI Global, Hershey, PA.