Náčrt epizodicky ukotvenej sémantiky

Martin Takáč

Katedra aplikovanej informatiky FMFI UK
Mlynská dolina, 842 48 Bratislava
takac@ii.fmph.uniba.sk

Abstrakt

V príspevku navrhujeme modifikáciu sémantiky založenej na identifikačných kritériách, ktorá prinesie nasledujúce zlepšenia: reprezentácia udalostí a situácií, autoasociatívne dopĺňanie neúplnej informácie, možnosť reprezentácií indivíduí aj typov, robustnosť voči šumu, schopnosť zachytiť kategórie s multimodálnymi distribúciami vlastností, hierarchizácia a modularita vrstiev konceptov. Reprezentácie konceptov sú autonómne vytvárané inkrementálne a kontinuálne a organizované na základe konkrétnych epizód, v ktorých sa vyskytli. Autoasociatívne dopĺňanie umožňuje efekty "zrkadlových neurónov" ako predikovanie akcie iných, plánovanie, empatia (inferencia vnútorných stavov iných), či situované používanie jazyka.

1 (Veľmi stručný) úvod

V tomto článku sumarizujeme a kriticky analyzujeme výsledky nášho doterajšieho výskumu v oblasti autonómne konštruovateľnej sémantiky. Hlavným cieľom článku je zmapovať limity doterajšieho prístupu a načrtnúť cesty k ich prekonaniu.

2 Vývinový prístup k dizajnu umelých systémov

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.

Typickým príkladom veľmi dynamického a otvoreného prostredia je internet, resp. world wide web. Autonómne agenty – softvérové entity vykonávajúce nejakú činnosť vo virtuálnom prostredí sa čoraz viac stávajú súčasťou aplikácií v oblastiach 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 [4]. Mnoho vedeckého a programátorského úsilia je venovaného tvorbe ontológií a jazykov pre reprezentáciu významu webovských stránok. Dynamické a otvorené prostredie však žiadna fixná ontológia nemôže raz a navždy opísať. Preto sa pre dizajn inteligentných webovských 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.

Nový tzv. *vývinový prístup* k dizajnu umelých systémov [26] (čo do dôsledkov by sme ho smelo mohli nazvať vývinovou *paradigmou*) je inšpirovaný poznatkami z vývinovej psychológie. Pre flexibilitu a plné rozvinutie potenciálu ľudskej inteligencie je kľúčové to, že neprichádzame na svet „dokončení“: náš psychomotorický vývin je inkrementálny a pokračuje dlho po narodení. Pritom sú dôležité tieto faktory [14]:

1. Detská skúsenosť je inherentne multimodálna: redundantná a časovo koordinovaná informácia z viacerých senzorických systémov umožňuje učenie bez učiteľa či bez externe definovaných úloh: len na základe vnímania a konania vo svete.
2. Pre rozvinutie dostatočne komplexnej inteligencie je nutná počiatočná nezrelosť a inkrementálny vývin.
3. Detská inteligencia je distribuovaná medzi organizmom a prostredím, ktorého regularity organizujú percepciu, konanie a v konečnom dôsledku aj myslenie.
4. Deti experimentujú veľmi variabilným a často zdanlivo náhodným spôsobom, pričom objavujú nové problémy a nové riešenia.
5. Deti žijú v sociálnom svete, v ktorom zrelší partneri poskytujú podporné štruktúry pre učenie.
6. Deti sa učia jazyk – zdieľaný symbolický komunikačný systém, ktorý posúva kogníciu na kvalitatívne novú úroveň a umožňuje abstraktné myslenie.

V práci [17] sme sa venovali fenoménu vzniku takého správania umelých systémov, ktoré možno nazvať inteligentným. Pritom vyvstala potreba redefinovať pojmy ako „rozumenie“ či „význam“ neantropocentrickým spôsobom tak, aby boli použiteľné

aj pre umelé systémy, či predverbálne živé organizmy. Takáto redefinícia sa nemôže opierať o introspekciu, ani nemôže byť založená na prítomnosti intencionality či vedomia (ktoré je opäť ťažko uchopiť inak ako introspekciou). Vystavali sme ju preto v súlade s vývinovým prístupom na interakčnej spriahnutosti agenta s prostredím: agent rozumie svojmu svetu (*Umwelt*), resp. pripisuje mu významy, ak disponuje stelesnenou skúsenostne získanou znalosťou, ktorá mu umožňuje diferencovane reagovať na rôzne stavy (seba a prostredia) adekvátne k svojmu účelu/cieľu. Podrobne sme sa charakterizácii „rozumenia" venovali v práci [22], na tomto mieste vyslovíme princípy dôležité pre dizajn „rozumejúcich" agentov [23, 25]:

- 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.

3 Sémantika identifikačných kritérií

Dlhodobým cieľom nášho výskumu je vybudovať autonómne osvojiteľnú sémantiku s netriviálnou vyjadrovacou silou v súlade s vyššie uvedenými princípmi a jej autonómnu konštruovateľnosť verifikovať pomocou výpočtových modelov. Doteraz sme implementovali niekoľko výpočtových modelov konštrukcie významov: na základe senzomotorickej interakcie s prostredím [19] a na základe sociálnej inštrukcie [24], pričom sme skúmali aj ich stabilitu v iterovanom medzigeneračnom prenose [20]. Významy vo všetkých výpočtových modeloch boli založené na tzv. sémantike identifikačných kritérií [15, 21], ktorej rozpracovanie považujeme za náš najvýznamnejší teoretický prínos.

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.

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 (detaily pozri napr. v [22]).

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 kategórií a na vzájomné korelácie medzi atribútmi. Skonštruovanú reprezentáciu možno priamo prepojiť s jazykom: významy jazykových výrazov sú ukotvené v interakčne skonštruovaných identifikačných kritériách čím sa rieši *Symbol Grounding Problem* [6]. 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.

3.1 Nedostatky

1. V našich výpočtových modeloch sme využívali univerzálne rozhranie pre opis vnemov z vonkajšieho sveta v podobe rámcov – množín dvojíc atribút: hodnota, kde hodnota bola reálne číslo. Dôvodom pre takéto univerzálne rozhranie je skutočnosť, že rámcami možno zachytiť štruktúry virtuálnych svetov (napr. databázové záznamy či príznaky webovských stránok), ako aj reálnych (napr. rámec hodnôt nameraných na senzoroch, resp. poslaných na riadenie aktuátorov robota). Želaným zovšeobecnením by však bola možnosť spracovať aj symbolické, či *význačné* [18] hodnoty atribútov, príp. nominálne hodnoty na škále bez ordinálneho usporiadania, napr. *pohlavie*: {*muž*, *žena*}.
2. Indukcia identifikačných kritérií je založená na štatistických vlastnostiach spoločných *všetkým* príkladom, teda na vytvorení prieniku množín ich atribútov. Takýto prístup má niekoľko neželaných dôsledkov: Indukcia nie je robustná voči omylom/šumu – jeden nesprávny príklad s úplne odlišnou množinou atribútov môže nevratne „poškodiť" indukovaný prienik. V modeli osvojovania významov na základe slovných pomenovaní môže mať rovnaký efekt používanie homoným, ak sú rovnakým slovom označené navzájom nesúvisiace koncepty. Na dôvažok, mnoho významov označených rovnakým slovom je založených skôr na *rodinných podobnostiach* [28] jednotlivých členov ako na ich spoločných vlastnostiach [9]. Preto by bolo vhodné upraviť mechanizmus indukcie tak, aby množina atribútov nebola kritickým parametrom. Rámec popisujúci príklad nejakej kategórie by tak mohol mať premenlivý počet atribútov – v závislosti od

momentálneho štádia vývinu[1] resp. dostupnosti vnemov v danom kontexte. Bolo by vhodné, aby agent dokázal identifikovať inštanciu známeho konceptu aj na základe neúplnej množiny atribútov (napr. v prípade zlej viditeľnosti, šumu, či výpadku senzora) a nedostupné hodnoty autoasociatívne doplnil resp. predikoval.

3. V doterajších modeloch bola každá kategória (s výnimkou situácií) reprezentovaná práve jedným identifikačným kritériom, pričom všeobecnejšie kritérium sa vyznačovalo väčším receptívnym poľom (takým okolím prototypického príkladu, na ktoré reagovalo nadprahovou aktivitou). Takáto reprezentácia vylučovala koncepty, ktorých príklady by v geometrickom priestore tvorili nespojitú oblasť. Akékoľvek rozdelenie hodnôt atribútov – aj multimodálne – sa aproximovalo normálnym rozdelením. Ďalším neželaným dôsledkom princípu „jedna kategória – jedno kritérium“ bolo to, že oblasti bohato zastúpené interakčne získanými príkladmi, boli často reprezentované jedným alebo niekoľkými veľmi všeobecnými kritériami, čo znemožnilo presnejšiu predikciu hodnôt (napr. v modeli [19]. Prirodzenejšie by bolo, keby reprezentácia bola najhustejšia a umožňovala najšpecifickejšie predikcie práve v oblastiach najbohatšej interakčnej skúsenosti.
4. Sémantika zložitejších významových kategórií situácií a udalostí nebola dopracovaná do detailu – chýba mechanizmus osvojenia kvôli súvisiacemu problému vymedzenia: ktoré z aktuálne aktívnych kritérií považovať za relevantné pre indukované situáciu/udalosť? Alternatívou k budovaniu reprezentácie komplexných kategórií „zdola“ z kritérií objektov, vlastností a vzťahov je duálny prístup: základnou pamäťovou jednotkou by bola epizóda – udalosť a elementárnejšie významy by vznikli ako invariantné komponenty kategórií reprezentujúcich rôzne epizodické koncepty [1, 16].
5. Naučené kauzálne vzťahy medzi situáciami, akciami a ich dôsledkami by mali byť prirodzene využívané agentmi na plánovanie dosahovania cieľov. To v našich modeloch zatiaľ chýba (výnimkou je práca [8]).
6. Dva spôsoby učenia – individuálnou senzomotorickou interakciou a na základe jazyka – boli skúmané v separátnych modeloch. Bolo by vhodné ich prepojiť spôsobom analogickým používaniu zrkadlových neurónov [12, 27]: pozorovanie udalostí v prostredí, či akcií iných agentov by automaticky aktivovalo reprezentácie získané vlastnou skúsenosťou a naopak, pričom reprezentácie by sa autoasociatívne dopĺňali a navzájom ovplyvňovali. Tento istý mechanizmus by umožnil napr. aj „empatiu“ – predikciu vnútorného stavu iného agenta autoasociatívnym doplnením reprezentácie vlastnej skúsenosti vyvolanej aktuálne pozorovaným kontextom.

V ďalšom texte načrtneme modifikáciu reprezentácie identifikačných kritérií a mechanizmov jej budovania tak, aby sme dosiahli vylepšenie vyššie uvedených nedostatkov.

4 Epizodická reprezentácia

Z didaktických dôvodov – pre lepšiu motiváciu čitateľa – začneme od konca, teda zhora nadol, od úrovne reprezentácie udalostí. Všeobecne možno reprezentovať udalosť ako dvojicu identifikačných kritérií situácií[2] (*počiatočná situácia, cieľová situácia*), pričom v grafe počiatočnej situácie budú hrany zodpovedajúce kritériám elementárnych zmien, ktoré zapríčinili transformáciu počiatočnej situácie na cieľovú.[3] Ďalším konštruktom je *rolový rámec* – súbor metahrán určujúcich roly vykonanej akcie (agens, objekt, inštrument, atď.) a jej spôsob vykonania (identifikačné kritérium akcie). Problém vymedzenia (ktoré statické a dynamické aspekty zahrnúť do reprezentácie udalosti a ktoré nie) možno riešiť mechanizmom hradiel resp. prahov pre aktiváciu kritérií, pričom prahy by záviseli od momentálneho vnútorného stavu agenta (obsahu pamäte, zamerania pozornosti, aktuálneho cieľa a aktuálnych potrieb).

4.1 Vnímané epizódy

Pre účely výpočtového modelovania si situáciu zjednodušíme (a skonkretizujeme): agentova pozornosť bude štandardne zameraná na pozorovanie interakcií s tematickou (rolovou) štruktúrou: { *Subjekt_akcie, Objekt_akcie, Akcia, Dôsledky_pre_subjekt, Dôsledky_pre_objekt* }. Schopnosť chápať prežívanú situáciu pomocou tzv. *theta* rôl (kto komu čo urobil) je významnou súčasťou sociálnej kognície a je nutná pre disponovanie plnogramatickým jazykom (na rozdiel od protojazyka [2]). Nateraz budeme u agentov túto

[1] So zjemňovaním reprezentácie si agent môže všimnúť atribúty, ktoré predtým nevnímal.

[2] Kritérium situácie je definované ako graf, v ktorého vrcholoch sú indivíduá – vstupné argumenty a na hranách sú elementárne kritériá objektov, vlastností a vzťahov medzi objektmi, ktoré sú pre danú situáciu konštitutívne ([22], str. 114).

[3] Vo všeobecnosti je dôležité rozlíšiť krátkodobú situačne konštruovanú reprezentáciu konkrétnej situácie/udalosti od dlhodobej reprezentácie typu situácie/udalosti [11]. Pre účely tohto príspevku však vystačíme s jedinou (hybridnou) reprezentáciou pokrývajúcou indivíduá aj typy.

schopnosť predpokladať – agent dostane na vstupe perceptuálne rámce aj s označením jednotlivých rôl. Avšak vstup nemusí byť vždy kompletný – niekedy napr. vidíme udalosť – zmenu nejakého objektu, ale nevieme, aká akcia ju spôsobila a kto je jej subjektom (agentom). V prípade ak vnímame svoju vlastnú akciu, môžeme mať v rámci pre *Akciu* detailne špecifikované proprioceptívne a motorické parametre vykonanej akcie, zatiaľ čo ak ju vykonáva iný agent, jej rámec obsahuje iba pozorovateľné charakteristiky. Takisto, ak je agent sám subjektom alebo objektom nejakej akcie, môže príslušný rámec obsahovať atribúty opisujúce vnútorný stav, hodnoty potrieb, resp. ich zmeny (v dôsledkových rámcoch). Množina rôl a atribútov je otvorená. Agent kategorizuje (mechanizmom, ktorý opíšeme nižšie) a zapamätáva si celé rolové štruktúry a tieto využíva aj na autoasociatívne dopĺňanie (predikciu) nedostupných atribútov či celých rôl. Na základe zapamätanej epizódy s vlastným vnútorným stavom môže predikovať (vnímaniu nedostupný) vnútorný stav iného agenta v podobnej situácii, čo je výpočtová obdoba empatie.

Vedľajším efektom budovania klastrov na úrovni celých epizód je vznik implicitných kategórií objektov, akcií a zmien (ako komponentov epizód), ktoré sú inherentne prepojené a navzájom sa ovplyvňujú. Autoasociatívnym mechanizmom možno napr. doplniť k vstupnému rámcu objektu akcie, ktoré s ním možno vykonávať, teda jeho *afordancie* [5]. Objekty, s ktorými sa dá interagovať podobným spôsobom, resp. sa vyskytli v podobných epizodických kontextoch, budú mať tendenciu združovať sa. To je v súlade s poznatkami o objektových kategóriách základnej úrovne (*basic level*), ktoré sú organizované na základe rovnakých interakčných programov [9].

4.2 Rozšírenie o jazyk

V súlade s funkcionalistickým poňatím sémantiky [28] je možné chápať jazykovú výpoveď ako akciu s dôsledkami pre subjekt (hovorcu) a objekt akcie (adresáta) a ako také ich možno reprezentovať v horeuvedenej tematickej štruktúre. Takéto (zapamätané, indukované) štruktúry predstavujú agentovu situovanú znalosť jazyka a pomocou autoasociatívneho dopĺňania mu umožnia rozhodnúť, v akom kontexte je vhodná aká výpoveď, predikovať na základe výpovede vnútorný stav hovoriaceho, či rozlišovať jazykové štýly jednotlivých hovoriacich.

4.3 Časové asociácie a plánovanie

Pamäťový sklad epizód, tak ako sme ich opísali, umožňuje (proto-)plánovanie, keďže reprezentácia epizódy spája akciu s jej dôsledkami. Porovnaním zapamätaných dôsledkov s cieľmi agenta možno vybrať vhodné akcie.

Plánovací resp. kauzálno-prediktívny mechanizmus môžeme ešte posilniť zavedením časových asociácií nad pamäťovým skladom epizodických kategórií. Orientované asociácie sa vytvárajú medzi epizodickými štruktúrami aktívnymi v danom časovom kroku a štruktúrami aktívnymi v nasledujúcom kroku. Zvýšenie sily asociácie je úmerné hodnotám aktivity asociovaných štruktúr (Hebbovské pravidlo). V danom časovom kroku konfigurácia aktivít všetkých zapamätaných epizodických štruktúr (kritérií) pre daný vstup (epizódu) reprezentuje agentovo pochopenie, rozpoznanie, resp. klasifikáciu danej epizódy. Predikciu nasledujúceho stavu dostaneme šírením aktuálnych aktivít v smere orientovaných asociácií úmerne ich sile. Analogicky, abdukciu predchádzajúcej epizódy dostaneme šírením aktivít proti smeru orientovaných asociácií.

5 Základné elementy epizodickej reprezentácie

5.1 Lokálne vyladený detektor

Základnou jednotkou reprezentácie je *jadro* resp. *lokálne vyladený detektor*. Takýto detektor sa indukuje[4] z množiny príkladov $\{ x^{(1)}, ..., x^{(N)} \}$, kde každý príklad $x^{(i)}$ je rámec charakterizovaný svojou vlastnou množinou názvov atribútov $A_x^{(i)}$ a ich hodnôt $x^{(i)}.a_j \in [0,1]$[5] pre $a_j \in A_x^{(i)}$. Vzhľadom na nedostupnosť vnímaniu, šum, vývin vnímania a iné faktory je množina potenciálnych názvov atribútov otvorená a množiny atribútov jednotlivých rámcov $x^{(i)}$ nemusia byť totožné. Úlohou je z takýchto príkladov indukovať detektor, ktorý na akýkoľvek vektor reálnych čísel z intervalu [0,1] s pomenovanými zložkami reaguje aktivitou z intervalu [0,1] a dokáže autoasociatívne doplniť nedostupné zložky na základe zapamätanej reprezentácie.

Kvôli výpočtovej efektívnosti za základ zvolíme detektor citlivý na variancie jednotlivých atribútov (pozri [21])[6] a vhodne ho zmodifikujeme. Majme množinu príkladov $X=\{ x^{(1)}, ..., x^{(N)} \}$. Pre ľubovoľný atribút

[4] Mechanizmus indukcie je všeobecný – pracuje na akýchkoľvek rámcoch s reálnymi atribútmi – rolové rámce reprezentujúce epizódy sú len jednou z možných inštancií.

[5] Rozhodnutie pre interval [0,1] namiesto celého oboru reálnych čísel zdôvodňujeme v časti 5.2 aj s uvedením príslušnej transformácie.

[6] Kategóriu založenú na kovarianciách atribútov možno aproximovať väčším počtom jadier založených na varianciách [10].

$a_j \in \mathrm{A} = \bigcup_{i=1}^{N} A_{x^{(i)}}$ detektor eviduje trojicu f_j, p_j, σ_j^2, kde f_j je frekvencia výskytu atribútu a_j (t. j. počet príkladov z množiny X, v ktorých sa a_j vyskytuje), a p_j resp. σ_j^2 je iteratívne počítaný *lenivý*[7] priemer resp. rozptyl všetkých hodnôt atribútu a_j v množine X. Okrem toho si detektor pamätá celkový počet príkladov N. Prvý videný príklad teda inicializuje detektor a každý ďalší ho modifikuje iteratívnym spôsobom, pričom modifikácia prebieha permanentne a inkrementálne.

Indukovaný detektor r reaguje na vstup x aktivitou podľa vzorca $r(x) = \exp\left(-k \cdot \sqrt{\sum_{j \in \mathrm{A}} \frac{(x.a_j - p_j)^2}{s_j^2} \cdot \frac{f_j}{N}} \right)$, kde k je vhodne zvolená kladná konštanta a $s_j^2 = \sigma_j^2 + \varepsilon / f_j^d$ je odhad rozptylu.[8] Sumácia rozdielov prebieha cez atribúty zaznamenané v detektore. Ak sa niektorý zaznamenaný atribút a_j v x nevyskytuje, považujeme $x.a_j = 0$. Atribúty, ktoré sa vyskytujú v x, ale nie sú zapamätané v detektore, nemajú na výsledok vplyv (čo je to isté, akoby boli zapamätané s frekvenciou $f_j = 0$). Takýto detektor zohľadňuje frekvenciu výskytu jednotlivých atribútov: odlišnosť v hodnote atribútu videného zriedka nemá takú váhu ako odlišnosť v hodnote veľmi frekventovaného atribútu. Detektor teda už nie je založený na atribútoch spoločných pre všetky inštancie (prieniku množín ich atribútov) a je robustný voči šumu a premenlivej dostupnosti atribútov.

Okrem výstupnej aktivity $r(x)$ vypovedajúcej o miere príslušnosti vstupu x do reprezentovanej kategórie detektor automaticky vykonáva autoasociatívne doplnenie tým, že obsahuje zapamätané atribúty aj s ich zapamätanými priemernými hodnotami (ktoré sa nemusia práve vyskytovať v aktuálnom vstupe). Pokiaľ vopred poznáme množinu nedostupných atribútov B, ktorých hodnoty chceme predikovať, použijeme na výpočet aktivity vzorec

$$r(x, \mathrm{B}) = \exp\left(-k \cdot \sqrt{\sum_{j \in \mathrm{A/B}} \frac{(x.a_j - p_j)^2}{s_j^2} \cdot \frac{f_j}{N}} \right),$$

v ktorom atribúty z množiny B neprispievajú do celkového rozdielu.

5.2 Vstupná transformácia

Predpoklad, že lokálne vyladené detektory dostávajú na vstupe vektor reálnych čísel z intervalu [0,1] s pomenovanými zložkami (a na výstupe vracajú aktivitu taktiež z intervalu [0,1]) podporuje modularitu celkovej reprezentácie: jednotlivé detektory možno združovať do vrstiev a vrstvy možno spájať do hierarchií (vstupom pre detektory vyššej vrstvy je vektor charakterizujúci[9] konfiguráciu aktivít detektorov nižšej úrovne).

Od najnižšej vrstvy, ktorá priamo reaguje na vnímané epizódy (opísané v časti 4.1) však očakávame, že bude reagovať na množinu tematicky pomenovaných rámcov s atribútmi, ktorých hodnoty sú (ľubovoľné) reálne čísla.[10] Potrebujeme teda zabezpečiť transformáciu vnímanej epizódy na jediný rámec vyššie uvedenej podoby.

5.2.1 Transformácia atribútu

Najprv opíšeme transformáciu (jedného) atribútu s reálnočíselnými hodnotami na interval [0,1] pomocou konštruktu tzv. *priestorov význačných hodnôt* (PVH) inšpirovaného kvalitatívnym usudzovaním [18]. Pri dodatočných doménových znalostiach môže byť niekedy vhodné význačné hodnoty v PVH stanoviť vopred (manuálne vytvoriť vstupnú vrstvu), my však využijeme samoorganizáciu – význačnými hodnotami budú centrá (prototypy) klastrov hodnôt atribútu, s ktorými sa agent stretol. Pre jeden konkrétny atribút sa bude vytvárať a aktualizovať vždy jeden a ten istý priestor význačných hodnôt bez ohľadu na to, v akom rámci sa atribút vyskytol (napr. pre atribút *hmotnosť* bude jeden PVH bez ohľadu na to, či ide o hmotnosť subjektu alebo objektu akcie, atď.). Algoritmus samoorganizácie PVH nad oborom reálnych čísel je opísaný v Prílohe 1.

Vo všeobecnosti je priestor význačných hodnôt PVH_a pre nejaký atribút a charakterizovaný množinou jadier $\{a\_0, a\_1, ..., a\_k\}$. Každé jadro reprezentuje jednu význačnú hodnotu a je funkciou, ktorá pre daný vstup (reálne číslo) vracia hodnotu z intervalu [0,1] vyjadrujúcu, nakoľko je vstup blízky reprezentovanej

---

[7] V angličtine *sliding* [7] alebo *amnestic* [26] *average* $p := (1 - \alpha/f) * p + \alpha/f * x$, kde p je doterajší priemer, x je nový príklad, f je počet videných príkladov, môže prideľovať novoprichádzajúcim príkladom rôznu váhu v závislosti od nezáporného parametra α: pre $\alpha=1$ počíta štandardný aritmetický priemer, pre $\alpha>1$ sa priemer nachyľuje v prospech novo-videných príkladov, a pre $\alpha<1$ si udržuje určitú „konzervatívnosť".

[8] Keďže po videní prvej hodnoty atribútu je jej rozptyl vždy nula, takýto detektor by reagoval len na jediné indivíduum. Preto sa spočiatku zapamätaná hodnota rozptylu nadhodnotí o kladné ε, a s rastúcim počtom videných príkladov f_j klesá k skutočnej vypočítanej hodnote σ_j^2 v závislosti od kladnej rýchlosti d.

[9] Názvami zložiek vektora aktivít sú jednoznačné identifikátory detektorov nižšej úrovne.

[10] Spôsob transformácie atribútov s *nenumerickými* hodnotami naznačujeme v časti 5.2.3.

význačnej hodnote. Transformáciou vstupnej hodnoty *h* atribútu *a* je rámec

$PVH_a(h) = \{ a\_0: a\_0(h), a\_1: a\_1(h), ..., a\_k: a\_k(h) \},$

teda vektor hodnôt z intervalu [0,1] s pomenovanými zložkami. Takýto vektor môžeme vnímať ako konfiguráciu aktivít časti vstupnej vrstvy siete jadier alebo ako (nenormovanú) pravdepodobnostnú distribúciu náhodnej premennej *a* s diskrétnymi hodnotami. Ide teda o štruktúru vyjadrujúcu agentovo presvedčenie o hodnote vnímaného atribútu v termínoch hodnôt, ktoré už pozná.[11] Pre účely ďalšieho spracovania môžeme zaviesť aj konštrukty

$$PVH_a^{\theta}(h) = \{ a\_i: a\_i(h) \mid a\_i(h) > \theta \},$$

teda konfiguráciu iba tých detektorov, ktoré vracajú nadprahovú aktivitu, alebo

$$PVH_a^{*}(h) = \{ a\_i^{*}: a\_i^{*}(h) \mid i^{*} = \mathrm{argmax}_i(a\_i(h)) \},$$

teda vrátenie iba detektora, ktorý na hodnotu *h* reaguje najviac.

5.2.2 Transformácia celej epizódy

Tým sme opísali vstupnú transformáciu jedného atribútu. Vstupnú transformáciu perceptuálneho rámca zloženého z viacerých atribútov dostaneme jednoducho zjednotením výsledkov aplikácie príslušných PVH na všetky prítomné atribúty.

Vstupnú transformáciu celej epizodickej štruktúry (zloženej z rámcov pomenovaných tematickými rolami) dostaneme tak, že najskôr transformujeme jednotlivé perceptuálne rámce v tematických rolách a výsledky týchto transformácií zlúčime do jediného rámca vhodným premenovaním zložiek.

Príklad

Vstup:

{ *Subjekt_akcie* = { *poloha*: 5 }, *Akcia* = { *posuň_sa*: 1, *vlavo*: 10 }, *Dôsledky_pre_Subjekt* = { *poloha*: -10 } }[12]

Transformácie jednotlivých rámcov:

Subjekt_akcie: { *poloha_2*: 0.8 } *Akcia*: { *posuň_sa_0*: 1.0, *vlavo_1*: 0.7 } *Dôsledky_pre_Subjekt*: { *poloha_0*: 0.6 }[13]

Celkový výsledok:

{ *Subjekt_akcie-poloha_2*: 0.8, *Akcia-posuň_sa_0*: 1.0, *Akcia-vlavo_1*: 0.7, *Dôsledky_pre_Subjekt-poloha_0*: *0.6* }

5.2.3 Nenumerické atribúty

Všimnime si, že malou úpravou definície jadier PVH by sme dokázali vytvoriť aj vstupnú vrstvu transformujúcu na vektor aktivít atribút s nenumerickými, resp. nominálnymi hodnotami (bez definovaného usporiadania), príp. atribút vyjadrujúci vlastnosť charakterizovanú prítomnosťou alebo neprítomnosťou viacerých čŕt. Stačilo by predpokladať, že jednotlivé jadrá vstupnej vrstvy vedia adekvátne reagovať na nenumerický vstup. Napríklad pre určenie hodnoty atribútu *pohlavie* by sme mali v PVH dva detektory *pohlavie_muž* a *pohlavie_žena*, z ktorých každý na základe nejakých pozorovateľných príznakov vracia pravdepodobnosť, že na vstupe je muž, resp. žena, teda hodnoty z intervalu [0,1], čo znamená, že na vyšších vrstvách už môžeme použiť lokálne detektory v štandardnej podobe tak, ako sme ich definovali.

6 Algoritmus budovania epizodických kategórií

Opísali sme, ako vstupná vrstva zabezpečí transformáciu vnímaných epizód na vektory reálnych čísel z intervalu [0,1] s pomenovanými zložkami. Ostáva nám uviesť, ako sa takto reprezentované epizódy združujú do kategórií, resp. ako prebieha indukcia príslušných identifikačných kritérií z takýchto príkladov.

Opustíme doteraz platný princíp „jedna kategória – jedno kritérium“. Ten zodpovedal prototypovému kategoriálnemu systému. Nahradíme ho zmiešaným exemplárovo-prototypovým kategoriálnym systémom, v ktorom niektoré kritériá reprezentujú indivíduá, iné typy a celkové agentovo pochopenie danej situácie je definované konfiguráciou aktivít viacerých jadier (možností je viac: všetkých, nadprahových alebo len jedného – najviac reagujúceho).

Na začiatku je systém kategórií prázdny. Induktívny algoritmus funguje takto: Pre daný vstupný rámec sa vypočíta konfigurácia *K* všetkých kritérií, ktoré vracajú aktivitu väčšiu ako θ_m – prah príslušnosti do kategórie (*K* je zoznam dvojíc *identifikátor_kritéria*: *aktivita* usporiadaný zostupne podľa aktivity). Každé kritérium v *K* sa aktualizuje v smere vstupného rámca úmerne svojej aktivite.[14] Následne prebehne zapamätanie nového

[11] Pre čitateľov zorientovaných v sémantike *semiotických schém* [13], výstup činnosti priestoru význačných hodnôt zodpovedá konštruktu *analog belief*.

[12] Pre jednoduchosť uvažujeme veľmi malý počet atribútov a tiež sme vynechali objekt (nie všetky tematické roly musia byť vždy dostupné). Dôsledky pre subjekt sú vyjadrené ako *zmeny* – rozdiely hodnôt atribútov pred a po vykonaní akcie.

[13] Predpokladáme existenciu viacerých jadier v príslušných PVH, za transformáciu sme si zvolili PVH* verziu vracajúcu jediné jadro s maximálnou reakciou, indexy jadier aj hodnoty aktivít boli vymyslené náhodne.

[14] Aktivita sa dosadí za parameter α pre výpočet lenivého priemeru a rozptylu (pozri poznámku 7). Toto vytvára „*rich get richer*“ dynamiku, ktorá napomáha samorganizácii. Čím jadrá reagujú na daný vstup menej, tým menšia je pravdepodobnosť,

indivídua: pokiaľ bola množina *K* prázdna, alebo aktivita maximálne reagujúceho detektora je nižšia ako θ_i – prah identickosti indivíduí,[15] pridá sa nové kritérium aktualizované vstupným rámcom ako svojím prvým príkladom.

Prahy môžu byť stanovené pevne alebo sa dynamicky meniť, napr. na základe tzv. *vigilancie* – bdelosti [3]. Ak kategoriálny systém agentovi neprináša požadované výsledky (napr. predikcie o dôsledkoch akcií sú nesprávne), môže adaptívne zvýšiť prahy a tým zjemniť celkovú reprezentáciu.

Pridávaním pamäťových stôp pre každé indivíduum nám môže veľkosť kategoriálneho systému neúmerne narastať, preto je vhodné (a kognitívne plauzibilné) zaviesť do systému zabúdanie. To sa dá urobiť tak, že každé kritérium bude ohodnotené hodnotou *freshness*, ktorá aditívne narastie vždy, ak kritérium na nejaký vstup vráti aktivitu väčšiu ako θ_m (teda nastal „hit" – restimulácia). Zároveň na konci každého časového kroku *freshness* všetkých kritérií multiplikatívne klesne (prenásobením konštantou < 1) a vymažú sa všetky kritéria, ktorých *freshness* klesne pod stanovený prah zabúdania. Takisto je vhodné zaviesť do systému spájanie (merging) kritérií, ktoré začínali ako oddelené indivíduá, ale časom k sebe navzájom skonvergovali – splynuli v jedinú kategóriu.

Takto zavedená reprezentácia má vlastnosť požadovanú v bode 3 časti 3.1, že je najhustejšia a najšpecifickejšia tam, kde má agent najviac skúseností a teda môže dávať najpresnejšie predikcie.

7 Záver

Hlavným prínosom navrhnutých výpočtových modelov sémantiky identifikačných kritérií je dodržiavanie vývinového prístupu, ktorý dbá na to, aby systémy namiesto hotovej znalosti boli vybavené učiacimi mechanizmami a konkrétne poznatky resp. významy si osvojili samostatne v interakciách s prostredím.

Cieľom tohto článku bolo predstaviť teoretický návrh konštrukcií, ktoré môžu značne vylepšiť niektoré nedostatky sémantiky identifikačných kritérií a posunúť ju bližšie k realite: robustnosť voči šumu, epizodická reprezentácia, autoasociatívne dopĺňanie, predikcia, „zrkadlové" učenie pozorovaním iných agentov, či situované používanie jazyka. Ide o návrh, ktorý treba v mnohých detailoch dotiahnuť a hlavne verifikovať jeho zmysluplnosť a realizovateľnosť počítačovými simuláciami. V súčasnej dobe preto pracujeme na pilotnej implementácii[16] návrhu. Výsledky simulácií budú známe čoskoro.

Poďakovanie

Tento výskum bol čiastočne podporený grantami VEGA 1/3105/06 a VEGA 1/0361/08.

Príloha 1: Samoorganizácia PVH

Konkrétnou implementáciou PVH pre spojitý obor reálnych hodnôt atribútu *a* je štruktúra charakterizovaná celkovým počtom *N* videných hodnôt atribútu, ich iteratívne počítaným rozptylom σ^2_a a množinou jadier $\{a\_0, a\_1, ..., a\_k\}$, kde každé jadro je reprezentované dvojicou (v_i , f_i), teda priemernou hodnotou jadra *a_i* a počtom jeho aktualizácií. Na daný vstup $h \in \mathrm{R}$ reagujú jednotlivé jadrá aktivitou

$$a\_i(h) = \exp\left(-k \cdot \sqrt{\frac{(v_i - h)^2}{s_a^2}} \right),$$

kde *k* je kladná konštanta a $s_a^2 = \sigma_a^2 + \varepsilon / N^d$ je odhad rozptylu.[17] Zhliadnutie nového vstupu *h* zároveň modifikuje PVH_a: ak aspoň jedno jadro reaguje dostatočnou aktivitou, jeho hodnota sa posunie v smere vstupu *h* úmerne jeho aktivite (lenivý priemer), inak sa vytvorí nové jadro s hodnotou rovnou *h*.[18] Pseudokód algoritmu induktívnej úpravy $PVH_a = (\sigma^2_a, V)$ je nasledovný (na začiatku je $\sigma^2_a = 0$ a $V = \emptyset$, teda prázdna množina).

Majme $PVH_a = (\sigma^2_a, V)$ a nový vstup $h \in \mathrm{R}$.

$N := N+1$; iteratívne uprav σ^2_a

ak $V=\emptyset$ tak pridaj nové jadro v *h*

inak {
  $i^* = \mathrm{argmax}_i(a\_i(h))$;
  ak $a\_i^*(h) < \theta$ a nebola presiahnutá max. kapacita PVH, tak pridaj nové jadro v *h*
  inak { $f_i^* := f_i^* + 1$; $v_i^* := (1 - \alpha/f_i^*) v_i^* + \alpha/f_i^* h$, kde $\alpha = a\_i^*(h)$ }
}

že naozaj ide o „ich" inštanciu a teda sa až tak neposunú v jeho smere.

[15] Typicky $\theta_m \le \theta_i$, napr. $\theta_m = 0.5$ a $\theta_i = 0.9$.

[16] V jazyku JAVA, pričom využívame vlastné simulačné prostredie Jarmila vytvorené v rámci doterajšieho výskumu sémantiky identifikačných kritérií.

[17] Pozri poznámku 8.

[18] Celkový počet jadier môže byť limitovaný – ak sa dosiahne maximálna kapacita, nepridáva sa nové jadro, ale sa aktualizuje najviac reagujúce jadro (aj v prípade, že vracia aktivitu nižšiu ako stanovený prah).

Literatúra

[1] L. W. Barsalou: Perceptual Symbols Systems. *Behavioral and Brain Sciences* 22 (1999) 577 – 660.

[2] D. Bickerton: Catastrophic evolution: The case for a single step from protolanguage to full human language. In *Approaches to the Evolution of Language: Social and Cognitive Bases* (J. R. Hurford, M. Studdert-Kennedy, C. Knight, eds.), Cambridge University Press, Cambridge, UK, 1998.

[3] G. A. Carpenter, S. Grossberg: Adaptive Resonance Theory. In: *The Handbook of Brain Theory and Neural Networks, Second Edition* (M. A. Arbib, ed.), MIT Press, Cambridge, MA, 2003: 87 – 90.

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

[5] J. J. Gibson: *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston, 1979.

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

[7] N. Hulth, P. Grenholm: A Distributed Clustering Algorithm. *Lund University Cognitive Studies* 74 (1998).

[8] P. Jankovič: *Sémantická reprezentácia pragmatických znalostí*. Nepublikovaná diplomová práca, FMFI UK Bratislava, 2007.

[9] G. Lakoff: *Women, fire, and dangerous things: What categories reveal about the mind*, University of Chicago Press, Chicago, IL, 1987.

[10] T. Poggio, F. Girossi: Networks and the best approximation property. *Biol. Cybern.* 63 (1990) 169 – 176.

[11] D. Retová, J. Šilliková, J. Šefránek: Opice, psy, sémantika a logika. In: In: *Kognice a umělý život VII* (J. Kelemen, V. Kvasnička, J. Pospíchal, eds.), Slezská univerzita, Opava, 2007: 287 – 294.

[12] G. Rizolatti, et al.: Premotor cortex and the recognition of motor actions. *Cognitive Brain Research* 3 (1996) 131 – 141.

[13] D. Roy: Semiotic schemas: a framework for grounding language in action and perception. *Artificial Intelligence,* 167 (1–2) (2005) 170 – 205.

[14] L. Smith, M. Gasser: The Development of Embodied Cognition: Six Lessons from Babies. *Artificial Life, 11* (1–2) (2005) 13 – 30.

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

[16] J. Šefránek: osobná komunikácia.

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

[18] M. Takáč: *Kvalitatívne modelovanie a simulácia,* Vydavateľstvo UK, Bratislava, 2003.

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

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

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

[22] M. Takáč: *Construction of meanings in living and artificial agents.* Nepublikovaná dizertačná práca, FMFI UK Bratislava, podané 2007.

[23] M. Takáč: Princípy dizajnu rozumejúcich agentov. In: *Zborník príspevkov štipendistov z projektu JPD 3 BA 2005/1-043*, Centrum projektovej podpory FMFI UK, Bratislava, 2008.

[24] M. Takáč: Autonomous Construction of Ecologically and Socially Relevant Semantics. *Cognitive Systems Research,* prijaté.

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

[26] J. Weng, W. Hwang, Y. Zhang, C. Yang, R. Smith: Developmental Humanoids: Humanoids that Develop Skills Automatically. In: *Proceedings of the First IEEE-RAS International Conference on Humanoid Robots*. Cambridge, MA, 2000.

[27] J. Wiedermann: Nástin architektury vědomého kognitivního agenta se dvěma vnitřnými modely světa. In: *Kognice a umělý život VII* (J. Kelemen, V. Kvasnička, J. Pospíchal, eds.), Slezská univerzita, Opava, 2007: 377 – 383.

[28] L. Wittgenstein: *Philosophical Investigations,* Macmillan, New York, 1953.