Reprezentácia epizód v pracovnej pamäti, problém viazania a vytváranie očakávaní

Martin Takáč

Centrum pre kognitívnu vedu KAI FMFI UK
Mlynská dolina, 842 48 Bratislava
Email: takac@ii.fmph.uniba.sk

Alistair Knott

Department of Computer Science, University of Otago
PO Box 56, Dunedin 9054, New Zealand
Email: alik@cs.otago.ac.nz

Abstrakt

V článku predstavíme vylepšenú verziu konekcionistického modelu pracovnej pamäti pre epizódy (Takáč a Knott, 2013). Epizódy, ako aj individuálne objekty v nich participujúce, sú reprezentované ako plány vykonateľných senzomotorických a pozornostných operácií. V modeli riešime tzv. „problém viazania“ (angl. *binding problem*)—priradenie objektov k jednotlivým tématickým rolám epizódy (agens, paciens) originálnym spôsobom: plán prehratia epizódy obsahuje oddelené pozične kódované reprezentácie agensa a paciensa, ktoré postupne aktivujú reprezentácie zúčastnených objektov v inom médiu. Tým sú agens aj paciens reprezentované v tom istom médiu, ale v rôznych časoch. Takýto prístup umožňuje lepšiu schopnosť zovšeobecňovania, ako aj možnosť reprezentovať „top-down“ očakávania o tom, aké epizódy sa môžu vyskytnúť, resp. akú akciu by bolo vhodné v danej situácii vykonať, či očakávané umiestnenia a vlastnosti zúčastnených objektov.

1 Úvod

V tomto článku predstavíme konekcionistický model toho, ako mozog kóduje významy viet. Zameriame sa na reprezentáciu významov v pracovnej pamäti (angl. *working memory*, ďalej WM). Tieto reprezentácie hrajú úlohu pri spracovaní viet: počas generovania/produkcie, ale aj porozumenia/interpretácie viet, sa význam vety udržuje resp. konštruuje v sémantickej WM (Fiebach a spol., 2007).

Toto médium sa tiež využíva v mimojazykových úlohách, napr. na reprezentáciu plánovaných akcií (Zwaan a Taylor, 2006) a na dočasné uloženie vnímaných epizód pred prehrávaním do dlhodobej pamäte (Baddeley, 2000). V našom článku sa zameriame na epizódy obsahujúce akcie/udalosti, nie statické fakty.

Vo formálnej logike sa propozície štandardne reprezentujú ako predikáty s argumentmi, napr. epizóda *Mačka naháňa myš* je reprezentovaná predikátom NAHÁŇA s dvoma argumentmi, ktoré majú odlišné sémantické roly, konkrétne MAČKA je v role agensa—pôvodcu činnosti a MYŠ v role paciensa—zasahovanej entity (Páleš, 1994). Implementácia tohto viazania resp. priradenia argumentov k rolám v konekcionistických modeloch je problematická—ak by sme jednoducho naraz aktivovali reprezentácie mačky, myši a naháňania, nebolo by jasné, *kto koho* naháňa. Mechanizmus viazania musí umožňovať: a) simultánne priradenie argumentov k viacerým rolám, b) hierarchické reprezentácie, pretože argumenty viazané k rolám môžu mat vlastnú vnútornú štruktúru–vlastné role a ich argumenty (Jackendoff, 2002).

Medzi najznámejšie modely viazania, ktoré spĺňajú tieto podmienky, patria van der Velde a de Kamps (2006); Stewart a Eliasmith (2012); Mayberry a Miikkulainen (2008). V tomto článku predstavíme model WM s novým riešením viazania založeným na myšlienke, že sémantické reprezentácie majú sekvenčnú štruktúru.

Výsledný model WM má zaujímavé dôsledky pre pochopenie jazykových aj mimojazykových procesov: v jazyku ho možno využiť na reprezentáciu významov generovaných alebo interpretovaných viet, a z mimojazykových procesov modeluje generovanie očakávaní o nadchádzajúcich epizódach tým, že aktivita v rôznych častiach siete reprezentuje rozdelenia pravdepodobnosti epizód a ich participantov, ktoré sa s prichádzajúcimi novými informáciami priebežne aktualizujú.

V ďalšom texte najprv predstavíme a zdôvodníme myšlienku sekvenčnej štruktúry sémantických reprezentácií, potom opíšeme novú schému viazania argumentov k rolám. Ďalej predstavíme samotný model a výsledky experimentov s ním.

2 Sekvenčná štruktúra sémantických reprezentácií

Náš model je postavený na základnom predpoklade, že komplexné sémantické reprezentácie nie sú statickými konfiguráciami neurálnej aktivity, ale *sekvenciami* jednoduchších reprezentácií. Túto myšlienku podrobne rozvíjame v Knott (2012) a Takáč a spol. (2012); tu uvedieme len stručné zhrnutie.

Sekvenčná povaha reprezentácií je motivovaná pozorovaním, že priame vnímanie epizód a ich účastníkov zahŕňa sekvenciu sezomotorických operácií

v kanonickom poradí (Ballard a spol., 1997), ktorých dôsledkom je sekvenčná aktivácia senzorických a motorických reprezentácií v mozgu. Napríklad pri vnímaní epizódy, ktorá má niekoľko účastníkov, pozorovateľ zameria na nich pozornosť postupne a nie paralelne. Ak ide o epizódu s agensom aj paciensom, pozornosť sa zameria najprv na agens, potom na paciens, a nakoniec sa aktivuje reprezentácia vnímanej alebo vykonávanej motorickej akcie (Knott, 2012). Knott argumentuje, že sémantické roly agens a paciens sú *definované* sériovou pozíciou v postupnosti senzomotorických operácií a nie explicitným označením.

Percepcia objektov—účastníkov epizód je takisto výsledkom kanonickej postupnosti senzomotorických operácií. Ak chce pozorovateľ klasifikovať objekt, musí najprv zamerať fokálnu pozornosť na pozíciu v priestore, kde sa daný objekt nachádza (Treisman a Gelade, 1980; Zhang a spol., 2011). Podľa Walles a spol. (2014) je medzi zameraním fokálnej pozornosti a klasifikáciou objektu ešte ďalšia operácia—zvolenie *priestorovej škály* klasifikácie, ktorá určí, či sa má klasifikovať globálna alebo lokálna formu podnetu na zvolenej priestorovej pozícii. Táto operácia determinuje, či sa bude klasifikovať jednotlivý objekt alebo homogénna skupina objektov, čo v jazyku zodpovedá číslu (jednotné/singulár, množné/plurál).

Proces percepcie objektu je teda kanonickou sekvenciou troch operácií:

1. výberu salientnej priestorovej pozície,
2. zvolenia klasifikačnej škály,
3. samotnej klasifikácie (ktorej výsledkom je aktivácia reprezentácie kategórie objektu).

Proces vnímania epizódy, ktorá obsahuje tranzitívnu akciu, je tiež sekvenciou troch operácií:

1. vnímanie agensa,
2. vnímanie paciensa,
3. aktivácia motorického programu.

Podľa Knott (2012) sémantické reprezentácie epizód a objektov v pracovnej pamäti odrážajú sekvenčnú štruktúru senzomotorických rutín, ktorými boli zaznamenané: epizódy a objekty sú reprezentované v pracovnej pamäti ako plány—pripravené sekvencie senzomotorických operácií, ktoré je možno jednu po druhej vykonať. Náš model sémantických reprezentácií je explicitne simulacionistický: sémantická reprezentácia v pracovnej pamäti je vykonateľný plán, ktorý po spustení zreprodukuje senzomotorickú skúsenosť postupnou aktiváciou senzomotorických reprezentácií. Napr. epizóda vyjadrená vetou *Ján naháňal Máriu* je uložená v pracovnej pamäti ako plánovaná sekvencia operácií: aktivovania reprezentácie Jána, potom aktivovania reprezentácie Márie a nakoniec aktivácia motorického programu pre naháňanie. Spustením tejto plánovanej sekvencie sa zreprodukuje sekvencia senzomotorických operácií, ktorá viedla k zapamätaniu tejto epizódy.

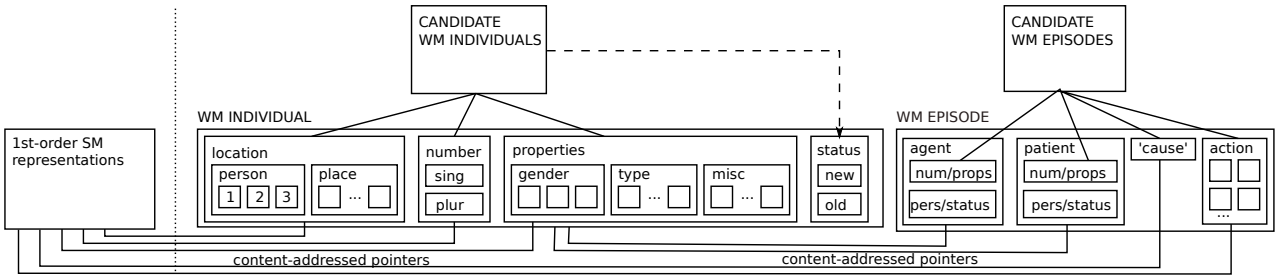
3 Mechanizmus viazania sekvenčne štruktúrovaných sémantických reprezentácií

Ako sme už spomínali, jednou z kľúčových požadovaných vlastností mechanizmu viazania je podpora simultánneho viazania viacerých rôl. Vo väčšine existujúcich modelov (napr. Chang, 2002; van der Velde a de Kamps, 2006; Stewart a Eliasmith, 2012; Mayberry a Miikkulainen, 2008) sú sémantické reprezentácie tvorené *statickými* konfiguráciami neurálnej aktivity, pričom sú všetky štruktúry viažuce roly aj ich argumenty aktívne *naraz*. V našom modeli je priradenie rôl implicitne prítomné v plánovanej senzomotorickej sekvencii uloženej v pracovnej pamäti a stáva sa explicitným *postupne* pri aktívnom prehrávaní plánovanej sekvencie. To umožňuje oveľa jednoduchšiu implementáciu viazania založenú na priamych asociatívnych spojeniach. Plánovaná senzomotorická sekvencia je v našom modeli kódovaná priestorovo, čo znamená, že rôzne neurálne polia aktivujú reprezentácie s nimi asociované postupne po jednom, v preddefinovanom poradí a tým spôsobia sekvenciu vzorcov aktivít v ďalších oblastiach. Tieto neurálne polia tak implementujú ukazovateľ (pointer) do iných oblastí. Pritom je veľmi dôležité, že rôzne polia môžu ukazovať do *toho istého* média, napr. neurálne pole, ktoré sa aktivuje ako prvé, cez asociatívne spojenia aktivuje reprezentáciu agensa v oblasti pre reprezentáciu objektov, a neurálne pole, ktoré sa aktivuje ako druhé, aktivuje reprezentáciu paciensa v *tej istej oblasti*, ale *v inom čase*, takže sa reprezentácie nepomiešajú.

Druhou požiadavkou na mechanizmus viazania je, aby umožňoval hierarchické reprezentácie, teda argumenty s vlastnou rolovou štruktúrou. V našom modeli túto požiadavku implementujeme tým, že priestorovo kódované neurálne polia v plánovacej oblasti môžu aktivovať reprezentácie, ktoré sú samy o sebe sekvenčným plánom v *inej* plánovacej oblasti. Napríklad naše reprezentácie agensa a paciensa, ktoré sa po jednom postupne aktivujú počas vykonávania plánu reprezentujúceho celú epizódu, sú samy o sebe sekvenčnými plánmi, takže jednotlivé kroky prehrávania epizódy majú vlastnú sekvenčnú štruktúru.

4 Model pracovnej pamäti pre epizódy a indivíduá

Architektúra nášho modelu je na obr. 1. Médiá tvoriace pracovnú pamäť sú vpravo od zvislej bodkovanej čiary.



Obr. 1: Architektúra modelu pracovnej pamäti pre epizódy a indivíduá.

Jedno reprezentuje indivíduá (jednotlivé objekty alebo homogénne skupiny objektov), druhé epizódy. V médiu `WM individual` je aktuálne vybraté indivíduum reprezentované ako sekvencia pozostávajúca z umiestnenia (`location`), čísla (`number`, klasifikačná škála) a množiny vlastností (`properties`), ktorá obsahuje distribuovanú reprezentáciu typu indivídua a jeho atribútov. Tieto reprezentácie sú aktívne v plánovacom médiu *naraz*, avšak keď sa plánovaná sekvencia vykonáva/prehráva, každá z nich postupne aktivuje príslušné asociované primárne reprezentácie v pozornostnom/perceptuálnom systéme (`1st order SM representations`) *po jednom*. To je v súlade s empirickými poznatkami o reprezentácii plánovaných senzomotorických sekvencií u makakov v prefrontálnych oblastiach (pozri referencie v Takáč a Knott, 2015). Na reprezentáciu umiestnenia, čísla a typu/vlastností môžeme nazerať ako na obsahom adresované ukazovatele do primárnych oblastí, podobne ako je to v modeli konvergenčných zón (Damasio a Damasio, 1994).

Vrstva `WM individual` poskytuje vstup ďalšej vrstve na obrázku nazvanej `candidate WM individuals` (ďalej `cWM-ind`), v ktorej sú krátkodobo uložené kombinácie umiestnenia, čísla a typu/vlastností zodpovedajúce indivíduám, ktoré boli nedávno v centre pozornosti. Neúplná reprezentácia vo vrstve `WM individual` môže slúžiť ako dotaz vrstve `cWM-ind`: ak napr. špecifikujeme umiestnenie, `cWM-ind` doplní číslo (jednotné/množné) a typ, a naopak. Ak sa podarí nájsť v `cWM-ind` indivíduum zodpovedajúce obsahu `WM individual`, klasifikuje sa ako známe, inak sa do `cWM-ind` pridá kombinácia zodpovedajúca aktuálnemu obsahu `WM individual` a označí sa ako nové indivíduum (atribút `old`/`new` v poli `status`, ktoré nie je súčasťou plánovanej sekvencie). Vrstva `cWM-ind` hrá úlohu v spracovaní jazyka aj pri mimojazykových procesoch. V jazyku reprezentuje významné referenty prebiehajúceho diskurzu, mimojazykovo reprezentuje zapamätané (a teda očakávané) umiestnenia a vlastnosti objektov na aktuálnej scéne.

Oblasť `WM episode` je formálne podobná oblasti `WM individual`: reprezentuje aktuálnu epizódu ako plánovanú sekvenciu operácií aktivujúcich agens, paciens a akciu, ktorá môže byt kauzatívna alebo nekauzatívna, viď Lee-Hand a Knott (2013). Tieto plánované reprezentácie sú opäť obsahom adresované ukazovatele na operácie v iných médiách: v plánovacom médiu sú aktívne naraz, ale počas prehrávania epizódy aktivujú reprezentácie, na ktoré ukazujú, po jednom. Kľúčovou myšlienkou je, že vrstvy `agent` aj `patient` obsahujú ukazovatele do vrstvy `WM individual` a nie priamo do primárnych senzomotorických oblastí. Počas prehrávania epizódy uloženej vo `WM episode` sú indivíduá reprezentujúce agens a paciens aktivované postupne, takže každé z nich umožňuje osobitné prehratie príslušnej podpostupnosti.

Plánované operácie uložené vo vrstve `WM episode` slúžia ako vstup vrstve ktorá reprezentuje viacero epizód—`candidate WM episodes` (ďalej `cWM-ep`). Táto vrstva je implementovaná ako samoorganizujúca sa mapa—SOM (Kohonen, 1982). Každý neurón si vo váhach pamätá konkrétnu kombináciu reprezentácií v oblastiach `agent`, `patient` a `action`, takže SOM po natrénovaní reprezentuje epizódy lokalisticky, pričom neuróny reprezentujúce podobné epizódy sú v mape blízko seba. Prirodzene, nie je reálne reprezentovať každú principiálne možnú epizódu lokalisticky, ale to ani nie je účelom média `cWM-ep`: jeho úlohou je reprezentovať *očakávané* alebo *želané* epizódy a spôsobom zhora nadol ovplyvňovať senzomotorické spracovanie počas percepcie alebo vykonávania akcie.

Dôležitým dôsledkom lokalistického kódovania je, že `cWM-ep` dokáže reprezentovať viacero epizód súčasne, takže profil aktivít mapy možno interpretovať ako pravdepodobnostné rozdelenie epizód. V mimojazykovom spracovaní to možno využiť na výber najpravdepodobnejšej alebo najželanejšej epizódy, alebo na priebežné generovanie rozdelenia pravdepodobnosti v oblasti `WM individual` počas procesu spracovania epizódy. Pri porozumení jazyka sa rozdelenie pravdepodobnosti epizód dá využiť na reprezentáciu alternatívnych interpretácií nejednoznačnej vety.

Ďalšou výhodou navrhnutej reprezentačnej schémy je, že `cWM-ep` sa dokáže naučiť *zovšeobecnenia* epizód. Jedno zovšeobecnenie je zabudované priamo do štruktúry modelu: SOM reprezentuje agens a paciens iba pomocou ich čísla a vlastností, a informáciu o pozícii a statuse ignoruje,

pretože táto nie je dôležitá pre generovanie očakávaní o epizódach. Na generovanie očakávaní o umiestneniach objektov slúži médium cWM−ind, ako ukážeme ďalej. To významne redukuje kombinatorické nároky na reprezentované epizódy. Ďalším zovšeobecnením je to, že pokiaľ je SOM vystavená veľkej a diverzifikovanej množine epizód, podobné epizódy môžu byt reprezentované tým istým neurónom. Keďže sú reprezentácie agensa aj paciensa distribuované, SOM sa môže naučiť abstrahovať od detailov, napr. reprezentovať *typy* indivíduí namiesto konkrétnych jednotlivcov.

Teraz popíšeme technické detaily nášho modelu. Vrstva WM individual pozostáva z niekoľkých sád lokalisticky reprezentovaných čŕt: osoba (1, 2, 3), číslo (singulár, plurál), rod (mužský, ženský, stredný) a status (známy, nový). Každá sada môže reprezentovať hodnotu vlastnosti jednoznačne (jeden neurón v sade úplne aktívny, ostatné úplne neaktívne), alebo ako rozdelenie pravdepodobnosti nad hodnotami reprezentovanými neurónmi v rámci jednej sady. Oblasť vlastností obsahuje okrem rodu ešte sadu pre životnosť (životné, neživotné), sadu pre typ objektu (človek, pes, mačka, vták, pohár, lopta, kreslo). Pozícia objektov, ktoré sú v reále situované na mriežke 100x100, je kódovaná populáciou 6x6 neurónov s čiastočne sa prekrývajúcimi gaussovskými receptívnymi poľami rovnomerne pokrývajúcimi mriežku. Oblasť misc reprezentuje farbu objektu, ktorá je nezávislá od typu (viď nižšie). Farba je kódovaná populáciou 11 neurónov s gaussovskými receptívnymi poľami v trojrozmernom RGB priestore maximálne reagujúcimi v oblasti 11 základných farieb (viď obr. 4). Takéto populačné kódovanie je biologicky plauzibilné a poskytuje matematicky korektný spôsob výpočtu vierohodnosti (angl. *likelihood*) rôznych stimulov pre konkrétny vzorec aktivít neurónov v populácii (Jazayeri a Movshon, 2006). Primárne senzomotorické reprezentácie (na obr. 1 vľavo od zvislej bodkovanej čiary) sú izomorfné s oblasťami vo WM individual a WM episode, s ktorými sú prepojené. Podobne aj oblasti agent a patient vo WM episode sú izomorfné s príslušnými oblasťami vo WM individual: num/props s oblasťou pre číslo a vlastnosti vo WM individual a pers/status s oblasťami pre osobu a status vo WM individual. Cause je jediný neurón, ktorý je aktívny pre kauzatívne akcie a neaktívny pre bežné akcie. Oblasť action pozostáva z 22 neurónov kódujúcich typ akcie lokalisticky (pozri legendu osi x na obr. 2) a z 11 neurónov kódujúcich vlastnosti akcií distribuovanými črtami. Vrstva cWM−ind je konvergenčná zóna variabilnej veľkosti plne prepojená s vrstvou WM individual: keď je na vstupe nové (nerozpoznané) indivíduum, vo vrstve cWM−ind sa regrutuje nový neurón a aktivity neurónov vo vrstve WM individual sa skopírujú do jeho váh (tzv. *one-shot learning*, čiže okamžité učenie). Vrstva cWM−ep je SOM so 400 neurónmi. Každý neurón v SOM má okrem štandardných vstupných váh ešte jednu skalárnu váhu, ktorá reflektuje počet/frekvenciu „zásahov“ tohto neurónu, teda koľko krát bol tento neurón víťazom (mal najvyššiu aktivitu). Z týchto frekvencií sa odvodia apriórne pravdepodobnosti pre výpočet Bayesovskej pravdepodobnosti, že aktuálny vstup zodpovedá epizóde reprezentovanej daným neurónom (detaily pozri v Takáč a Knott, 2015).

5 Trénovanie a testovanie modelu

5.1 Trénovanie

Model je trénovaný na epizódach v simulovanom virtuálnom svete. Svet pozostáva z objektov, ktoré vystupujú v epizódach. Každý objekt je určitého typu a má náhodne generované číslo,[1] pozíciu a farbu. Pozície sú náhodne generované na mriežke 100x100, farby sú stochasticky generované z gaussovských distribúcií centrovaných okolo 11 základných farieb. Následne generujeme množinu epizód, v ktorých vystupujú tieto objekty. Každá epizóda je prezentovaná WM systému ako sekvencia primárnych senzomotorických reprezentácií. Epizódy sú troch typov: tranzitívne (agens→paciens→tranzitívna-akcia), intranzitívne (agens→intranzitívna-akcia) a kauzatívne (agens→paciens→kauzatívna-akcia). Agens a paciens majú vlastnú vnútornú sekvenčnú štruktúru, konkrétne pozícia→číslo→typ/vlastnosti. Počas prezentácie epizódy sa komponenty týchto plánovaných podpostupnosti aktivujú postupne, z primárnych senzomotorických oblastí do častí vrstvy WM individual. Keď je reprezentácia vo vrstve WM individual úplná, aktivuje najprv ako dotaz vrstvu cWM−ind, čím sa zistí, či indivíduum, ktorého reprezentácia je momentálne obsahom WM individual, je nové, alebo sa už v momentálne zapamätanom kontexte vyskytuje. Každý neurón v cWM−ind (kandidát) sa aktivuje priamo úmerne priemernej Kullback-Leiblerovej divergencii (Kullback a Leibler, 1951) medzi zodpovedajúcimi si sadami čŕt vo WM individual a vo váhach kandidáta, čo zodpovedá vierohodnosti (likelihood), že aktuálny obsah vrstvy WMind je reprezentáciou indivídua zapamätaného vo váhach kandidáta (Jazayeri a Movshon, 2006). Ak je vierohodnosť víťazného kandidáta dostatočná, jeho váhy sa aktualizujú podľa obsahu WM individual a status WM individual sa nastaví na hodnotu old, v opačnom prípade sa vytvorí a pridá nový kandidát s váhami podľa obsahu WM individual, a status sa nastaví na hodnotu new. Kandidáti, ktorých váhy neboli aktualizované viac ako posledných 20 epizód, sa z cWMind odstránia, čo zodpovedá zabúdaniu.

Následne sa obsah WM individual (vrátane

[1] Ak je číslo plurál, objekt je vlastne *skupina* rovnakých objektov.

statusu) prekopíruje do izomorfnej oblasti v médiu WM episode, buď do časti agent, ak išlo o agens epizódy, alebo patient, ak to bol paciens. Po prezentácii kompletnej epizódy sa prešíri aktivita z vrstvy WM episode do SOM cWM-ep. Táto SOM je trénovaná štandardným spôsobom (Kohonen, 1982), teda učenie v cWM-ep je postupné (zatiaľ čo v cWM-ind je okamžité, one-shot).

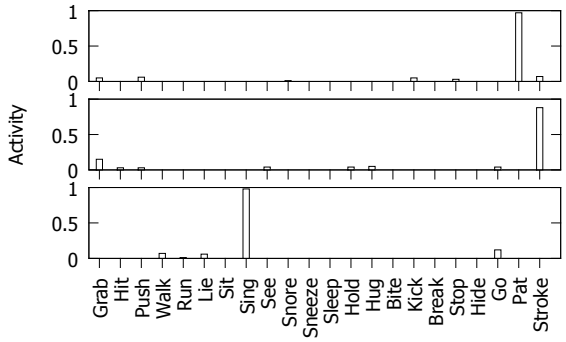
5.2 Testovanie mechanizmu viazania

Najprv ukážeme, ako náš model dokáže *znovuprehrať* práve vnímanú epizódu: po prezentovaní celej epizódy slúži obsah média WM episode ako vstup procesu znovuprehrávania, pričom zložky tohto média (agent, patient, action) sú aktivované postupne. Zložka agent (a neskôr patient) aktivuje asociovanú oblasť v médiu WM individual, čo spustí analogickú rutinu v tomto médiu (teda sekvenčné zložky WM individual aktivujú asociované primárne oblasti postupne). Ak navrhnutá schéma viazania funguje správne, mali by sme prehrať rovnakú sekvenciu SM signálov ako bola sieti prezentovaná počas vnímania epizódy. Pri testovaní sme dosiahli 99,6% úspešnosť, čo potvrdzuje, že navrhnutý mechanizmus viazania pracuje správne.

5.3 Testovanie predikčných a generalizačných schopností modelu

Navrhnutý model dokáže produkovať niekoľko typov predikcií. Po prvé, SOM cWM-ep dokáže predikovať, aké epizódy sa pravdepodobne vyskytnú, pričom predikcia sa s prichádzajúcou informáciou počas vnímania epizódy koriguje. Tieto predikcie je najľahšie demonštrovať, pretože sieť kóduje akcie priamo. Aby sme to mohli verifikovať, zaviedli sme do mechanizmu generovania trénovacích epizód nasledujúce pravidelnosti: Vtáky vždy spievali (t.j. ak je agensom vták, tak akciou je spievať), ďalej pri všetkých interakciách ľudí so psami a mačkami, človek potľapkal psa a poškrabkal mačku (person→dog→pat, person→cat→stroke). Natrénovanej SOM cWM-ep sme prezentovali na vstupe (WM episode) čiastočnú reprezentáciu epizódy—reprezentáciu vtáka v časti agent, alebo človeka v časti agent a psa resp. mačky v časti patient, a distribúciu očakávaných epizód sme zrekonštruovali ako lineárnu kombináciu váhových vektorov 10 najaktívnejších neurónov (s koeficientami úmernými aktivite). Na obr. 2 vidieť, že sieť predikuje s vyššou pravdepodobnosťou pravidelnosti, ktoré sa vyskytovali v trénovacích epizódach.

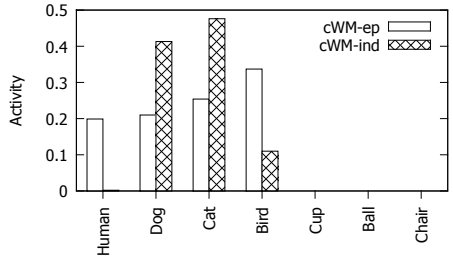
SOM cWM-ep tiež dokáže generovať predikcie o agense a paciense v epizódach. Mechanizmus týchto predikcií je zložitejší, pretože zahŕňa prehrávanie v systéme WM individual, čím sa predikcia aktualizuje o znalosti reprezentované týmto systémom. Najprv



Obr. 2: Typ akcie predpovedaný médiom cWM-ep pre tri fragmenty epizód. Zhora nadol: človek→pes→?, človek→mačka→?, vták→?.

sa zameriame na predikcie o agense. Aby sme ich mohli testovať, zaviedli sme v generovaní trénovacích epizód ďalšiu pravidelnosť—všetky epizódy majú životné agensy (teda ľudí a zvieratá, nie však veci). Pri testovaní sme najprv vygenerovali aktivitu v SOM cWM-ep na základe apriórneho rozdelenia pravdepodobnosti epizód (podľa frekvencie „zásahov“ každého neurónu v SOM), následne sme prešírili aktivitu zhora nadol do média WM episode ako aktivitami váhovanú kombináciu váhových vektorov 10 najaktívnejších neurónov v SOM. Výslednú aktivitu v časti agent sme prešírili/prekopírovali do média WM individual, kde slúži ako vstup pre sieť cWM-ind. Pretože obsah WM individual teraz reprezentuje očakávania, každý neurón vrstvy cWM-ind (reprezentujúci jedno zapamätané indivíduum) je aktivovaný priamo úmerne Kullback-Leiblerovej divergencii medzi jeho váhovým vektorom a obsahom WM individual. Váhové vektory 10 najaktívnejších neurónov sa skombinujú a prešíria výsledok zhora nadol, čím sa očakávanie v systéme WM individual aktualizuje o vedomosti o nedávno vnímaných/zapamätaných indivíduách. Na obr. 3 vidíme očakávanie/predikciu typu agensa v systéme WM episode aj WM individual. Oba systémy správne predikujú nulovú pravdepodobnosť výskytu neživotného agensa. Pretože však v momentálnom kontexte predikcie bolo v systéme cWM-ind zapamätaných oveľa viac psov a mačiek ako ľudí, predikcia typu v systéme WM individual je vychýlená v ich prospech.

Ďalším výsledkom interakcie systémov pre WM epizódy a indivíduá je generovanie predikcií o umiestnení a vlastnostiach jednotlivých objektov. Na testovanie týchto sme zaviedli ďalšiu pravidelnosť—vo všetkých epizódach, kde človek (v role agensa) interagoval so psom, bol pes čierny, ak bol agensom muž, a biely, ak žena. Ďalej, ľudia boli v prostredí vždy umiestnení v ľavom hornom a zvieratá v pravom hornom kvadrante mriežky (a veci v dolnej polovici mriežky). Pri testovaní sme najprv aktivovali



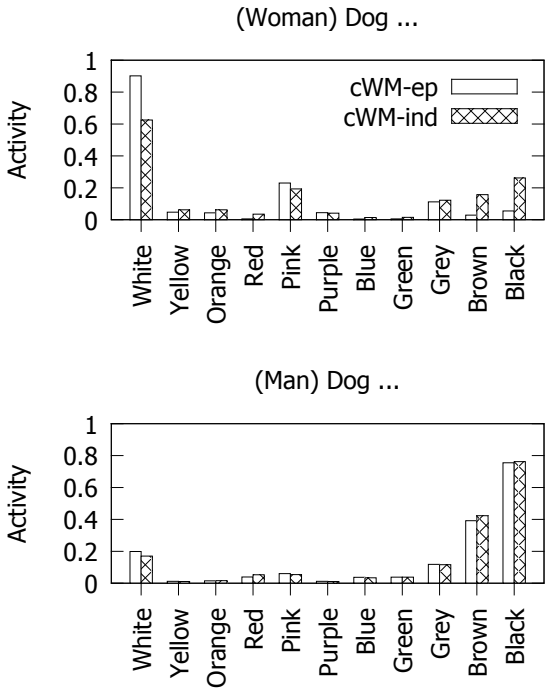
Obr. 3: Apriórne očakávanie typu agensa generované zhora nadol médiami `cWM-ep` a `cWM-ind`.

čiastočnú reprezentáciu epizódy vo `WM episode`— v časti `agent` sme aktivovali neuróny reprezentujúce typ človek a jeho rod (muž alebo žena), a v časti `patient` sme aktivovali neurón pre typ pes (všetky ostatné neuróny vrátane farby ostali neaktívne). Túto reprezentáciu sme použili ako vstup generujúci aktivitu v médiu `cWM-ind` a následne sme prešírili kombináciu váh 10 najaktívnejších neurónov v SOM zhora nadol spätne do média `WM episode`. Obsah časti `patient` sme prešírili/prekopírovali do média `WM individual`, ktoré sme následne aktualizovali pomocou znalostí v systéme `cWM-ind` spôsobom opísaným vyššie. Na obr. 4 vidno výslednú aktivitu neurónov kódujúcich farbu v médiu `WM individual`. Systém správne predikuje najvyššiu aktivitu v RGB oblasti zodpovedajúcej čiernej v prípade interakcií muž→pes, a bielej v prípade žena→pes. Vrstva `cWM-ind` dokáže generovať aj očakávanie o umiestnení psa (obr. 5). Vidíme všeobecné očakávanie výskytu v pravom hornom kvadrante, pretože zvieratá sa vždy vyskytujú v tomto kvadrante, ale vidíme aj špecifickú predpoveď na základe umiestnení nedávno zaznamenaných bielych a čiernych psov.

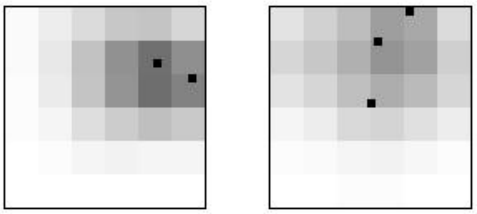
6 Diskusia

Prezentovaný konekcionistický model predstavuje novú originálnu schému viazania rôl, založenú na predpoklade, že sémantické reprezentácie sú v pracovnej pamäti uložené ako plánované sekvencie vykonateľných operácií. Takéto ponímanie sémantických reprezentácií je prítomné v mnohých simulacionistických teóriách významu. Náš model je však originálny v tom, že umožňuje, aby sémantické reprezentácie, ktoré sú obsahom jednotlivých rôl, mali vlastnú sekvenčnú štruktúru. Ďalšou výhodou modelu je, že obsahom indexované sémantické roly umožňujú reprezentovať celé pravdepodobnostné rozdelenie epizód naraz, nielen jednu epizódu.

Prezentovaná verzia modelu je predbežná, a tak zatiaľ ostáva veľa otázok otvorených. Dôležitá je napr. otázka priestorových nárokov modelu, pretože



Obr. 4: Predikcie o farbe paciensa generované médiami `cWM-ep` a `cWM-ind` pre epizódy žena→pes (hore) a muž→pes (dolu).



Obr. 5: Očakávané umiestnenie paciensa generované pre tieto epizódy systémom `cWM-ind` (tmavšia farba znamená silnejšie očakávanie). Čierne bodky reprezentujú skutočné umiestnenie aktuálne prítomných bielych (vľavo) a čiernych (vpravo) psov.

naše priestorové kódovanie indexov agensa a paciensa si vyžaduje udržiavanie niekoľkých oddelených kópií média `WM individual`. Počet kópií je obmedzený počtom rôl v epizóde, teda malý, aj tak je však dôležité zaoberať sa otázkou kapacitnej efektivity.

Teraz sa zameriame na problém, ktorý považujeme za ešte závažnejší: aj keď náš model podporuje *hierarchické* reprezentácie, tieto nie sú *rekurzívne*. Náš model dokáže reprezentovať význam menných fráz vo vetných frázach, ale menné frázy môžu obsahovať vnorené vetné frázy, ako je to v podraďovacích súvetiach (napr. *Muž, ktorého pohryzol pes, utiekol.*). Dá sa náš model rozšíriť tak, aby dokázal reprezentovať rekurzívne štruktúry tohto typu? Tvrdíme, že náš model na to má všetky predpoklady: Vo väčšine konekcionistických modelov sú totiž sémantické reprezentácie *statickými* vzorcami neurálnej aktivity—v týchto modeloch musí byť teda nadradená a podradená vetná fráza aktívna *naraz*. V našom modeli je prirodzené reprezentovať vnorené frázy v médiu `WM episode` *v rôznych časoch*, rovnako ako sú agens a paciens aktívne v tom istom médiu `WM individual`, ale postupne. Tento spôsob umožňuje napr. odložiť samotný *výber* podradenej frázy až do momentu, kedy sa príslušné indivíduum stane aktívnym v médiu `WM individual`. Vezmime si napr. vetu s podradenou referenčnou frázou *Pes [ktorý naháňal Máriu] ma pohrýzol*. Pri generovaní tejto vety sa najprv vo `WM episode` prehráva hlavná fráza *Pes ma pohrýzol*, počínajúc aktiváciou agensa `PES` v médiu `WM individual`. Predstavujeme si, že na konci tohto procesu sa vykoná operácia hľadania takých vlastností tohto psa, ktoré by ho jednoznačne referenčne identifikovali, výsledkom čoho môže byť vlastnosť participácie v epizóde, kde naháňal Máriu (táto epizóda sa získa nie z pracovnej pamäti, ale z nejakého dlhodobého pamäťového média). V tomto momente sa prehrávanie hlavnej epizódy dočasne pozastaví, a v médiu `WM episode` sa prehrá podradená epizóda, a potom sa pokračuje v prehrávaní hlavnej epizódy. Táto schéma je podobná riešeniu navrhnutému v Miikkulainen (1996) s použitím konekcionistickej rekurzívnej autoasociatívnej pamäte RAAM na implementáciu zásobníka epizód. V budúcnosti plánujeme rozšíriť náš model takýmto spôsobom.

Poďakovanie

Tento príspevok bol podporený grantom VEGA 1/0898/14 a grantom 13-UOO-048 grantovej agentúry NZ Marsden Fund.

Literatúra

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *TICS*, 4(11):417–423.

Ballard, D., Hayhoe, M., Pook, P. a Rao, R. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20(4):723–767.

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, 26:609–651.

Damasio, A. a Damasio, H. (1994). Cortical systems for retrieval of concrete knowledge: The convergence zone framework. Koch, C. a Davis, J. (zost.), V *Large-scale Neuronal Theories of the Brain*. MIT Press, Cambridge, MA.

Fiebach, C., Friederici, A., Smith, E. a Swinney, D. (2007). Lateral inferotemporal cortex maintains conceptual–semantic representations in verbal working memory. *Journal of Cognitive Neuroscience*, 19(12):2035–2049.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press, Oxford.

Jazayeri, M. a Movshon, A. (2006). Optimal representation of sensory information by neural populations. *Nature Neuroscience*, 9(5):690–696.

Knott, A. (2012). *Sensorimotor Cognition and Natural Language Syntax*. MIT Press, Cambridge, MA.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69.

Kullback, S. a Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86.

Lee-Hand, J. a Knott, A. (2013). A model of causative actions: from motor learning to syntactic structure. V *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, Berlin.

Mayberry, M. a Miikkulainen, R. (2008). Incremental nonmonotonic sentence interpretation through semantic self-organization. Technická správa AI08-12, Department of Computer Sciences, The University of Texas at Austin.

Miikkulainen, R. (1996). Subsymbolic case-role analysis of sentences with embedded clauses. *Cognitive Science*, 20:47–73.

Páleš, E. (1994). *SAPFÓ: Parafrázovač slovenčiny*. VEDA, Bratislava.

Stewart, T. a Eliasmith, C. (2012). Compositionality and biologically plausible models. Werning, M. a Hinzen, W. (zost.), V *The Oxford Handbook of Compositionality*. OUP, New York.

Takáč, M., Benušková, L. a Knott, A. (2012). Mapping sensorimotor sequences to word sequences: A connectionist model of language acquisition and sentence generation. *Cognition*, 125:288–308.

Takáč, M. a Knott, A. (2013). Konekcionistický model epizodickej pracovnej pamäti. Kelemen, J., Rybár, J., Farkaš, I. a Takáč, M. (zost.), V *Kognitívní věda a umělý život*, str. 265–272. Slezská univerzita, Opava.

Takáč, M. a Knott, A. (2015). A neural network model of episode representations in working memory. *Cognitive Computation*, DOI: 10.1007/s12559-015-9330-3.

Treisman, A. a Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12:97–136.

van der Velde, F. a de Kamps, M. (2006). Neural blackboard architectures of combinatorial structures in cognition. *Behavioral and Brain Sciences*, 29:37–108.

Walles, S., Robins, A. a Knott, A. (2014). A perceptually grounded model of the singular-plural distinction. *Language and Cognition*, 6:1–43.

Zhang, Y., Meyers, E., Bichot, N., Serre, T., Poggio, T. a Desimone, R. (2011). Object decoding with attention in inferior temporal cortex. *Proceedings of the National Academy of Sciences of the USA*, 108(21):8850–8855.

Zwaan, R. a Taylor, L. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, 135(1):1–11.