



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky



RNDr. Kristína Rebrová

Autoreferát dizertačnej práce

**Grounding the meaning in sensorimotor cognition:
a connectionist approach**

(Ukotvenie významov v senzomotorickej kognícii: konekcionistický prístup)

na získanie akademického titulu philosophiae doctor
v odbore doktorandského štúdia:

9.2.1 Informatika

Bratislava 2013

Dizertačná práca bola vypracovaná v dennej forme doktorandského štúdia na Katedre aplikovanej informatiky Fakulty matematiky, fyziky a informatiky Univerzity Komenského

Predkladateľ: RNDr. Kristína Rebrová
Katedra aplikovanej informatiky
Fakulta matematiky, fyziky a informatiky
Univerzita Komenského
Mlynská dolina
842 48 Bratislava

v Bratislave. **Školiteľ:** doc. Ing. Igor Farkaš, PhD.
Katedra aplikovanej informatiky FMFI UK, Bratislava

Oponenti:
.....
.....
.....
.....
.....

Obhajoba dizertačnej práce sa koná o h pred komisiou pre obhajobu dizertačnej práce v odbore doktorandského štúdia vymenovanou predsedom odborovej komisie 9.2.1 Informatika na

Predseda odborovej komisie:
prof. RNDr. Branislav Ročan, PhD.
Fakulta matematiky, fyziky a informatiky
Univerzita Komenského

1 Úvod

Teória ukotvenej kognície predpokladá, že koncepty sú v mozgu/mysli zakódované na základe senzoričných a motorických reprezentácií, ktoré s nimi súvisia. V doméne porozumenia akciám bolo mnohokrát empiricky overené, že pozorovanie pohybu vyvoláva v mozgu aktivitu nielen vo vizuálnych oblastiach, ale aj v motorickej kôre. Tejto problematike sa venujeme v práci (Rebrová, 2012). Tieto poznatky možno aplikovať v doméne kognitívnej robotiky pri tvorbe riadiacich architektúr pre humanoidné roboty. Prepojenie percepcie a akcie na vyššej úrovni umožní agentovi oddeliť vykonávanie akcie od jej pomenovania, čo je problém, ktorým trpí mnoho modelov v tejto problematike.

Úloha prepájania perceptuálnych dôsledkov akcií s ich motorickými reprezentáciami sa prikladá takzvaným zrkadliacim neurónom (Pellegrino a kol., 1992). Tento fenomén je v súčasnosti jednou z najdiskutovanejších tém na poli kognitívnych vied a kognitívneho modelovania. Ide o bunky vyššej motorickej kôry, objavené v oblasti F5 u makakov, ktoré reagujú nie len pri vykonávaní, ale aj pri pozorovaní akcií. Predpokladá sa, že úlohou zrkadliacich neurónov je párovanie pozorovaného deja a vlastnej motorickej skúsenosti za účelom porozumenia pozorovanej akcii a imitovania akcií. Nedávny výskum ukázal, že oblasť F5 prijíma informácie z vizuálnej oblasti STS mozgu makaka cez dve rôzne neurálne dráhy prepájajúce anatomicky a funkčne odlišné časti týchto oblastí Nelissen a kol. (2011).

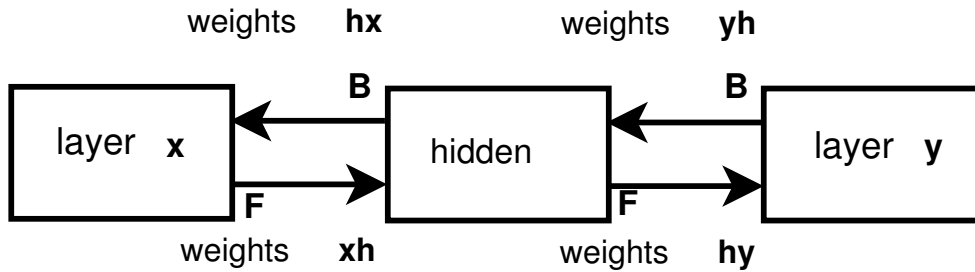
Spolu s objavom zrkadliacich neurónov a ich vlastností vzniklo množstvo výpočtových modelov systému zrkadliacich neurónov. Oztóp a kol. (2013) tvrdia, že práve výpočtové modely nám ponúkajú silný nástroj na skúmanie teórií o zrkadliacich neurónoch. Tieto modely zväčša modelujú presné časti mozgu opíc (či ľudí) a slúžia na konceptuálnu podporu teórií o motorickom porozumení pozorovanému dej. Najznámejšie z nich pochádzajú z dielne Arbiba, Oztópa a ďalších, napríklad FARS (Fagg a Arbib, 1998), MNS1 (Oztóp a Arbib, 2002), MSI (Oztóp a kol., 2005), MNS2 (Bonaiuto a kol., 2007) a iné (pozri (Oztóp a kol., 2006)). Väčšina z modelov systému zrkadliacich neurónov však uvažuje, že do nich vstupujú predspracované informácie, ktoré obsahujú pozične invariantné informácie o pozorovanom pohybe. Tento predpoklad je v rozpore so súčasnými poznatkami z neurovedy, ktoré hovoria o tom, že aj zrkadliace neuróny reagujú na to, z akého uhla pohľadu je pohyb pozorovaný. Niektoré výpočtové modely MNS možno zaradiť aj do oblasti kognitívnej robotiky. Takéto modely využívajú vlastnosti zrkadliacich neurónov, aby obohatili správanie robotického agenta o nové funkcie. Vhodným príkladom je RNNPB (Tani a kol., 2004).

2 Základné pojmy a označenia

Modulárna architektúra navrhovaného modelu systému zrkadliacich neurónov využíva nasledovné neurónové siete a mechanizmy učenia.

2.1 Algoritmus BAL

Pri výbere učiaceho algoritmu pre asociáciu vyšších motorických a vizuálnych oblastí sme dbali na biologickú plauzibilitu. Za týmto účelom sme navrhli algoritmus obojsmerného učenia na základe aktivačných fáz (bidirectional activation-based learning, BAL) založený na zovšeobecnenej recirkulácii (Generalized Recirculation, GeneRec) (O'Reilly, 1996). Tento učiaci algoritmus síce spadá do kategórie učenia s učiteľom, no na rozdiel od štandardného šírenia chyby funguje na báze rozdielu medzi dvoma aktivačnými stavmi siete, plusovou a mínusovou. V mínusovej fáze sa sieťou šíri aktivácia od vstupov k výstupom a zaznamenáva sa odhad siete na výstupe. Tento proces je schematicky zobrazený na Obr. 1. V druhej, plusovej fáze, sa na výstup „zafixuje“ (angl. clamp) želaný vzor (target) a aktivácia sa šíri naopak (od výstupu k vstupu). O'Reillyho model, pôvodne určený na jednosmernú asociáciu vstupov a výstupov, sme adaptovali tak, aby mohol slúžiť na obojsmernú asociáciu. Viac o GeneRec modeli a BAL v (Farkaš a Rebrová, 2013).



Obr. 1: Schéma obojsmerného učenia na báze aktivačných fáz.

Náš algoritmus pre obojsmernú asociáciu sa odvíja od GeneRec modelu, no tiež sa od neho líši. Náš model je perceptrón so 4 maticami váh, 2 pre každý smer toku aktivácie. Keďže je náš algoritmus obojsmerný, rozlišovanie medzi vstupnou a výstupnou vrstvou stráca význam. Vstupno-výstupné vrstvy teda označujeme ako \mathbf{x} a \mathbf{y} , a skrytú vrstvu \mathbf{h} (Obr. 1). Výpočet aktivácie na jednotlivých vrstvách v dvoch aktivačných fázach zobrazuje Tabuľka 1. Najprv aktivujeme vstupy na vizuálnej vrstve v smere $\mathbf{x}^F \rightarrow \mathbf{h}^F \rightarrow \mathbf{y}^F$ a získavame aktivácie na vrstvách v tomto smere, ktorý my označujeme písmenom F (zodpovedá mínusovej fáze)¹. V nasledujúcej fáze zafixujeme na vstup želaný výstup, ktorý sieť odhadovala v prvej fáze. V tomto prípade už o ňom uvažujeme ako o vstupe pre spätný smer, teda $\mathbf{y}^B \rightarrow \mathbf{h}^B \rightarrow \mathbf{x}^B$ označený písmenom B (plusová fáza). Výstup zo spätného smeru reprezentuje odhad siete o vstupe \mathbf{x}^F .

Na základe rozdielov v aktivačných fázach upravujeme váhové matice siete. Všeobecný predpis pre úpravu váh v doprednom F smere popisuje rovnica 1 a úpravu váh v spätnom B smere rovnica 2, kde λ je rýchlosť učenia. Všetky vrstvy obsahujú trénovateľné biasové neuróny, ktorých váhy sa upravujú analogicky k ostatným neurónom a ako vstup dostávajú konštantne hodnotu 1.

$$\Delta w_{ij}^F = \lambda \cdot a_i^F (a_j^B - a_j^F) \quad (1)$$

¹V našom prípade máme vstupy aj výstupy na oboch stranách. Rozlišujeme preto medzi smermi aktivácie, nie medzi odhadovacou mínusovou a zafixovanou plusovou fázou ako O'Reilly.

Tabuľka 1: Aktivačné fázy a stavy v modeli BAL. Obe aktivačné fázy sa skladajú z dvoch krokov, pričom vždy na začiatku aktivujeme vzor na vstupe a dostávame odhad siete na výstupe. Vrstvy \mathbf{x} a \mathbf{y} si navzájom striedajú úlohu vstupnej a výstupnej vrstvy.

Vrstva	Fáza	Vstup (net)	Aktivácia
\mathbf{x}	F	-	v_i^F (vstup)
\mathbf{h}	F	$\eta_j^F = \sum_i w_{ij}^{\mathbf{xh}} x_i^F$	$h_j^F = \sigma(\eta_j^F)$
\mathbf{y}	F	$\eta_k^F = \sum_j w_{jk}^{\mathbf{hy}} h_j^F$	$m_k^F = \sigma(\eta_k^F)$
\mathbf{y}	B	-	m_k^B (vstup)
\mathbf{h}	B	$\eta_j^B = \sum_k w_{kj}^{\mathbf{yh}} y_k^B$	$h_j^B = \sigma(\eta_j^B)$
\mathbf{x}	B	$\eta_i^B = \sum_j w_{ji}^{\mathbf{hx}} h_j^B$	$v_i^B = \sigma(\eta_i^B)$

$$\Delta w_{ij}^B = \lambda \cdot a_i^B (a_j^F - a_j^B) \quad (2)$$

2.2 Samoorganizujúca sa mapa s laterálnou excitáciou

Pri vytváraní invariantných reprezentácií na najvyššej úrovni nášho modelu sme použili štandardnú samoorganizujúcu sa sieť (SOM) (Kohonen, 1997), ktorú sme obohatili o mechanizmus laterálnej excitácie. Architektúra SOM pozostáva zo vstupného vektora, ktorý je váhovou maticou prepojený s výstupnou mapou neurónov (štandardne sa používajú 2D mapy). Hlavnou výhodou samoorganizujúcej sa mapy je to, že dokáže nájsť podobnosti a klasterizovať vysoko-dimenzionálne vstupné dáta, pričom váhový vektor medzi vstupom a neurónom i predstavuje prototyp jednej kategórie vstupných dát. Tým, že SOM zachováva topologickú organizáciu, neuróny, ktoré sú blízko seba, kódujú podobné vstupy.

Základné tréningovanie SOM pozostáva z dvoch krokov. Prvý krok je hľadanie víťazného neurónu i^* , ktorého váhový vektor má najmenšiu Euklidovskú vzdialenosť prezentovaného vstupu \mathbf{x} :

$$i^* = \arg \min_i \|\mathbf{x}(t) - \mathbf{w}_i\|, \quad (3)$$

\mathbf{w}_i je váhový vektor i -teho neurónu, a $\|\cdot\|$ označuje Euklidovskú normu. Následne sú pomocou víťazného neurónu i^* upravené váhy všetkých neurónov ako:

$$\Delta w_i = \alpha(t) h(i, i^*) (\mathbf{x}(t) - \mathbf{w}_i), \quad (4)$$

kde $\alpha(t)$ je rýchlosť učenia v čase t , ktorá v priebehu tréningovania siete klesá. Aby došlo k topologickej organizácii, váhy neurónov sú upravované na základe ich vzdialenosti od víťazného neurónu i^* . Túto vzdialenosť vyjadruje funkcia okolia $h(x, y)$. Na tento účel sa najbežnejšie používa Gaussovská vzdialenosť:

$$h(i, i^*) = \exp -d(i, i^*)^2 / \sigma^2(t), \quad (5)$$

kde σ predstavuje varianciu a v princípe určuje šírku tohto okolia. Pre vyhranenie reprezentácií ku koncu učenia σ linerárne klesá počas celého učenia.

Navrhovaný mechanizmus laterálnej excitácie sa aplikujeme na každý neurón na základe jeho pozície na mape (susednosti s ďalšími neurónmi) a na základe jeho aktivácie na mape

po prezentovaní vstupného vektora. Sila laterálnych spojení je udávaná váhovou maticou \mathbf{w}^{lat} . Výsledná aktivácia a neurónu i po uplatnení laterálnej excitácie sa vypočíta ako:

$$a_i = a_i^{\text{SOM}} + \sum_{j=0}^N w_{ij}^{\text{lat}} a_j, \quad (6)$$

kde aktivácia a^{SOM} neurónu i sa ráta pomocou vzdialenosti d_i z rovnice 3 ako:

$$a_i^{\text{SOM}} = \exp -d_i. \quad (7)$$

Váhy neurónu i s ostatnými neurónmi na mape sú upravované v každom kroku ako:

$$\Delta w_{ij}^{\text{lat}} = \beta(\lambda d(i, j) + (1 - \lambda)a_j)a_j, \quad (8)$$

kde $d(i, j)$ je Gaussovská vzdialenosť (rovnica 5) medzi neurónmi i and j , β je rýchlosť učenia a λ je parameter určujúci pomer vplyvu vzdialenosti medzi neurónmi i a j a vplyvu aktivácie neurónu j na váhu w_{ij}^{lat} . Tento mechanizmus zabezpečuje prešírenie aktivácie z jedného neurónu na jeho susedov s dôrazom na aktiváciu týchto neurónov, teda aktívnejšie neuróny budú posilnené viac ako menej aktívne.

2.3 Zlučovacia samoorganizujúca sa mapa (MSOM)

MSOM alebo Merge SOM (Strickert a Hammer, 2005), čiže zlučovacia samoorganizujúca sa mapa, je klasická SOM, ktorá má navyše kontextovú vrstvu ako rekurentná sieť, čo jej umožňuje spracovávať sekvencie dát. Na rozdiel od svojich predchodcov (napr. RecSOM či SOMSD, pozri porovnanie v (Vančo a Farkaš, 2010)) má MSOM kontextovú vrstvu veľkosti vstupu, nie celej mapy (Obr. 2). Do tejto vrstvy sa teda vkladá menšie množstvo informácie, konkrétne ide o lineárnu kombináciu váhy a kontextu aktuálneho (teda posledného) víťazného neurónu. Na rozdiel od klasickej SOM má každý neurón mapy $i \in \{1, 2, \dots, N\}$ dva váhové vektory:

1. $\mathbf{w}_i^{\text{inp}} \in \mathcal{R}^n$, ktorý sa viaže s n -dimenzionálnym vstupným vektorom $\mathbf{s}(t)$ v čase t a
2. $\mathbf{w}_i^{\text{ctx}} \in \mathcal{R}^n$, ktorý sa viaže s tzv. kontextovým deskriptorom $\mathbf{q}(t)$ popísaným nižšie.

Aktivácia neurónu i v čase t sa vypočíta ako:

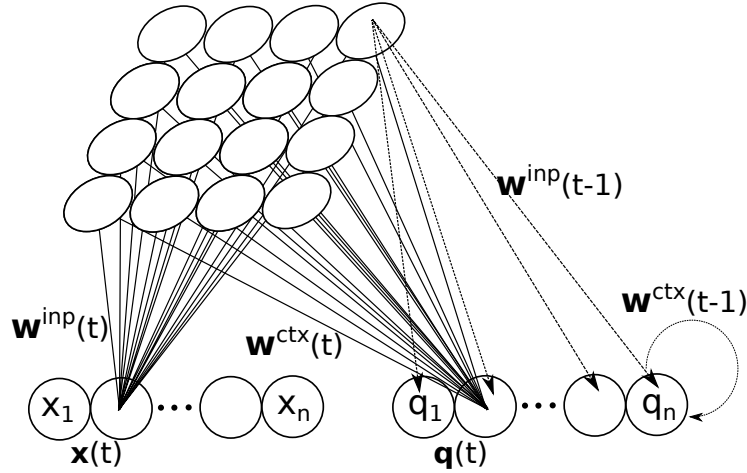
$$y_i(t) = \exp(-d_i(t)), \quad (9)$$

kde

$$d_i(t) = (1 - \alpha) \cdot \|\mathbf{s}(t) - \mathbf{w}_i^{\text{inp}}(t)\|^2 + \alpha \cdot \|\mathbf{q}(t) - \mathbf{w}_i^{\text{ctx}}(t)\|^2 \quad (10)$$

Kde parameter $0 < \alpha < 1$ reguluje pomer vplyvu kontextu a aktuálneho vstupu na aktiváciu neurónu a $\|\cdot\|$ označuje Euklidovskú normu vektora. Aktivácia neurónov kontextového deskriptora v čase t sa vypočíta ako:

$$\mathbf{q}(t) = (1 - \beta) \cdot \mathbf{w}_{i^*(t-1)}^{\text{inp}}(t) + \beta \cdot \mathbf{w}_{i^*(t-1)}^{\text{ctx}}(t), \quad (11)$$



Obr. 2: Schéma MSOM.

kde $i^*(t-1) = \arg \min_i \{d_i(t-1)\}$ je víťaz z minulého kroku a parameter $0 < \beta < 1$ reguluje vplyv kontextu a aktuálneho vstupu na aktuálnu aktiváciu kontextového deskriptora.

Trénovacie sekvencie sú sieti prezentované v prirodzenom poradí, vždy jeden vektor v jednom časovom kroku. Po prezentácii vstupu sú sieti v každom kroku upravené váhy na základe Hebbovského pravidla:

$$\Delta \mathbf{w}_i^{\text{inp}}(t) = \gamma \cdot h_{ii^*} \cdot (\mathbf{s}(t) - \mathbf{w}_i^{\text{inp}}(t)), \quad (12)$$

$$\Delta \mathbf{w}_i^{\text{ctx}}(t) = \gamma \cdot h_{ii^*} \cdot (\mathbf{q}(t) - \mathbf{w}_i^{\text{ctx}}(t)), \quad (13)$$

kde i^* je index víťaza v časovom kroku t a $0 < \gamma < 1$ je rýchlosť učenia a h_{ii^*} je funkcia okolia popísaná vyššie. Po natrénovaní sa odpovede neurónov siete zorganizujú podľa charakteristík vstupných sekvencií a sú biasované smerom ku koncom dátových sekvencií.

3 Ciele dizertačnej práce

Hlavným cieľom predkladanej dizertačnej práce bolo navrhnutie a vyhodnotenie konekcionistického modelu na obojsmerné prepojenie medzi senzoričkými a motorickými reprezentáciami slúžiaceho ako základ pre porozumenie akciám prostredníctvom systému zrkadliacich neurónov. Vzhľadom na modulárnu architektúru navrhnutého modelu možno ciele rozdeliť na nasledujúce úlohy:

- sumarizovať a kriticky zhodnotiť modely systému zrkadliacich neurónov a ukotvenia významu v senzomotorickej interakcii,
- navrhnuť, implementovať a zhodnotiť neurálnu architektúru pre vytváranie reprezentácií vyššej úrovne (model MSOM),
- navrhnuť, implementovať a zhodnotiť algoritmus pre obojsmernú asociáciu vyšších motorických a vizuálnych reprezentácií (algoritmus BAL).

4 Výsledky a prínosy práce

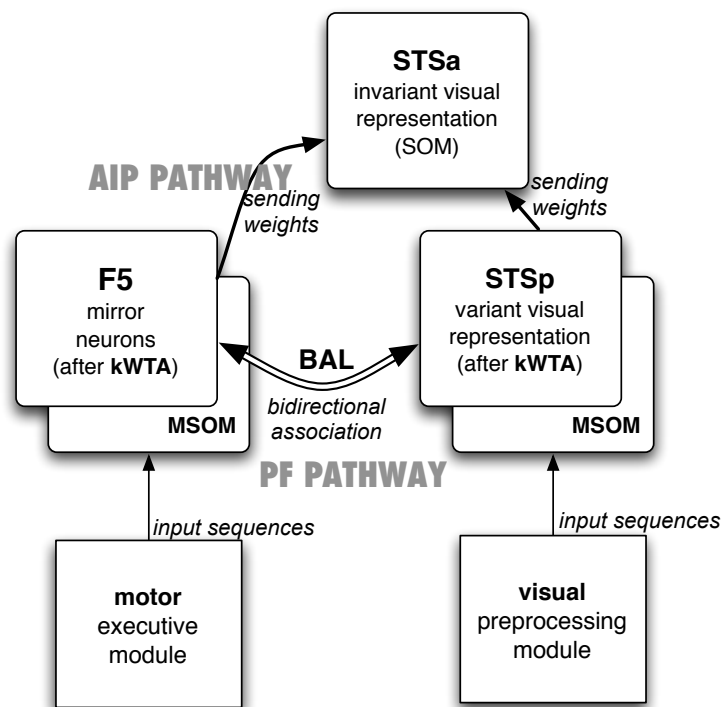
Jadro predkladanej dizertačnej práce tvorí konekcionistický model systému zrkadliacich neurónov zostavený z rôznych, vzájomne prepojených, umelých nerónových sietí navrhnutý pre kognitívneho robota iCub (Rebrová a Farkaš, 2012; Rebrová a kol., 2013; Rebrová a Farkaš, 2013). Pálenie zrkadliacich neurónov modelujem pomocou obojsmernej asociácie vysokoúrovňových vizuálnych a motorických reprezentácií. Hlavným prínosom práce je okrem samotného modelu – riadiacej architektúry pre robota iCub tiež nový obojsmerný biologicky motivovaný učiaci algoritmus BAL pre viacvrstvové neurónové siete, popísaný v časti 2.1.

Konkrétny návrh modulárnej architektúry je zobrazený na obrázku 3. Vzhľadom na architektúru a metódu učenia možno model rozdeliť do štyroch vrstiev, konkrétne:

1. nízkoúrovňové motorické a vizuálne moduly: riadiaci modul robota iCub, ktorý vykonáva 3 typy úchopov objektov modelovaný pomocou viacvrstvého perceptrónu trénovaného ekologicky plauzibilným učením s posilňovaním (RL) v spojitých priestoroch CACLA (van Hasselt, 2012), tieto moduly sme neimplementovali, len sme spracovávali výstupné dáta z predošlých experimentov so simulovaným iCub-om;
2. reprezentácie motorických a vizuálnych sekvencií vyššej úrovne (oblasti STSp and $F5_{\text{mir}}$), ktoré dostávajú sekvenčné vstupy z úrovne 1, tvoria dve MSOM, ktorých výstupy sú navyše binarizované metódou k -WTA;
3. dráha PF, spájajúca STSp a $F5_{\text{mir}}$ modelovaná pomocou viacvrstvého perceptrónu trénovaného algoritmom BAL, pričom vzniká asociácia medzi binárnymi výstupmi z MSOM predstavujúcich STSp a $F5_{\text{mir}}$, najprv medzi motorikou a pohľadom na vlastné akcie robota a potom asociácia aj s ďalšími perspektívami;
4. dráha AIP, spájajúca STSa s STSp a $F5_{\text{mir}}$, ktorá je modelovaná pomocou SOM s laterálnou excitáciou popísanou v 2.2.

4.1 Topologická reprezentácia dátových sekvencií pomocou MSOM

Na základe systematických experimentov sme našli optimálne parametre α a β pre obe MSOM v našom modeli a vyhodnotili víťazné neuróny. Ďalej sme experimentovali s parametrom k určujúcim počet pozitívnych bitov vo výslednom vektore. Pri analýze výstupov z MSOM sme zistili, že naša trénovacia množina pochádzajúca z pohybov robota iCub nebola úplne vyhovujúca, vzhľadom na to, že inštancie úchopov boli vygenerované z jedného prototypového pohybu pridaním šumu. Tým pádom neboli vo vzniknutých dátach očakávané štruktúry a v celom procese vznikli výstupné dáta asociované spôsobom m -na- n , čiže rôzne inštancie pohybov boli nekonzistentne previazané s ich vizuálnymi reprezentáciami. Kvôli tomu sme dáta zredukovali dostávajúc 29 trénovacích inštancií troch pohybov, ktoré považujeme za postačujúce. V budúcnosti plánujeme pracovať aj s modulmi na najnižšej úrovni a vygenerovať nové dáta vo viac ekologicky plauzibilnom scenári.



Obr. 3: Náčrt nášho robotického MNS modelu.

Význam výsledkov: Hlavným prínosom tejto úlohy je implementácia a experimentálna evaluácia relatívne novej architektúry neurónových sietí MSOM. Taktiež je dôležité, že úspešne natrénované moduly nášho modelu spĺňajú svoju funkciu. Použitie samoorganizácie v kontexte spracovania senzomotorických sekvencií v robotike je pomerne nový koncept a zatiaľ nebol modelovaný pomocou MSOM.

4.2 Obojsmerná asociácia binárnych vektorov pomocou BAL

V predkladanej dizertačnej práci ponúkame dva typy experimentov s novým algoritmom BAL, ktorý sme navrhli. V prvom rade prezentujeme všeobecné experimenty s týmto novým učiacim algoritmom, ktoré ukázali, že algoritmus konverguje k dobrým riešeniam a tiež, že je porovnateľný so symetrickou verziou algoritmu GeneRec (O'Reilly, 1996), z ktorého je odvodený. Konvergencia algoritmu, jeho ďalšie vlastnosti a reprezentácie na skrytej vrstve sú hodné ďalšieho skúmania. Myslíme si, že náš algoritmus má potenciál pre ďalšie využitie v oblasti výpočtovej kognitívnej neurovedy a neurónových sietí všeobecne. Ďalej sme experimentovali s BAL v kontexte nášho MNS modelu. V tomto prípade algoritmus funguje perfektne ak asociujeme motorické reprezentácie s vizuálnymi reprezentáciami z vlastnej perspektívy. V prípade, že rozšírime model na všetky štyri perspektívy, čiže asociujeme dáta spôsobom 1-na-4, sieť sa nezvládne naučiť jeden smer asociácie. Toto správanie je však pochopiteľné, pretože úloha vybrať správne riešenie pre jeden prezentovaný vzor, ktorý je asociovanými so štyrmi inými je neriešiteľná úloha. Asociácia v smere z vi-

zuálnych reprezentácií do motorických dosiahla dostatočnú úspešnosť, preto aj túto časť modelu považujeme za úspešne namodelované.

Význam výsledkov: Prínos tejto časti predkladanej dizertačnej práce spočíva opäť v dvoch bodoch. V prvom rade sme vytvorili, naimplementovali a otestovali nový učiaci algoritmus a odprezentovali výsledky v (Farkaš a Rebrová, 2013) a (Rebrová a Farkaš, 2013). Ďalej sme overili funkčnosť nášho algoritmu v prípade praktickej úlohy s reálnymi (nie umelo vytvorenými) dátami v kontexte nášho modelu. Výsledky z experimentov ukázali, že BAL je vhodný na asociáciu rôznych párov vstupných dát, avšak je citlivý na ich konzistentnosť. V prípade nejasných viacnásobných asociácií BAL na výstupe produkuje vzory, ktoré sú prienikom asociovaných vzorov s tendenciou vystihnúť prienik medzi viacnásobne naviazanými reprezentáciami.

4.3 Vznik invariantných reprezentácií na základe samoorganizácie s laterálnou excitáciou

Predbežné výsledky z tejto časti nášho modelu ukazujú, že navrhnutý laterálny mechanizmus pomáha pri vytváraní želaných invariantných reprezentácií na najvyššej úrovni. Experimenty s rôznym pomerom vplyvu aktivácie a susednosti neurónov ukázali, že tento mechanizmus pomáha cieľiť šírenie aktivácie na centrálnu, nie okrajové neuróny, ako sme mali pôvodne v úmysle. Navrhnutý algoritmus laterálnej excitácie je síce jednoduchý, ale produkuje zaujímavé výsledky a je rozhodne hodný ďalšieho skúmania rovnako ako algoritmus BAL a náš robotický MNS model celkovo.

Význam výsledkov: Podobne ako v predošlých častiach aj tu zavádzame nový princíp učenia neurónových sietí, ktorý je potenciálne škálovateľný a atraktívny pre ďalšie použitie vo výpočtovom kognitívnom modelovaní.

Zoznam použitej literatúry

- J. Bonaiuto, E. Rosta a M. Arbib. Extending the mirror neuron system model, I: Audible actions and invisible grasps. *Biological Cybernetics*, 96:9–38, 2007.
- A. Fagg a M. Arbib. Modeling parietal-premotor interactions in primate control of grasping. *Neural Networks*, 11:1277–1303, 1998.
- T. Kohonen. *Self-organizing Maps*. Springer, 1997. ISBN 3540620176.
- K. Nelissen, E. Borra, M. Gerbella, S. Rozzi, G. Luppino, W. Vanduffel, G. Rizzolatti a G. A. Orban. Action observation circuits in the macaque monkey cortex. *Journal of Neuroscience*, 31(10):3743–3756, 2011.
- R. O'Reilly. Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Computation*, 8(5):895–938, 1996.

- E. Oztop a M. Arbib. Schema design and implementation of the grasp-related mirror neuron system. *Biological Cybernetics*, 87:116–140, 2002.
- E. Oztop, M. Kawato a M. Arbib. Mirror neurons and imitation: A computationally guided review. *Neural Networks*, 19(3):254–271, 2006.
- E. Oztop, M. Kawato a M. Arbib. Mirror neurons: Functions, mechanisms and models. *Neuroscience Letters*, 2013.
- E. Oztop, D. Wolpert a M. Kawato. Mental state inference using visual control parameters. *Cognitive Brain Research*, 22:129–151, 2005.
- G. Pellegrino, L. Fadiga, L. Fogassi, V. Gallese a G. Rizzolatti. Understanding motor events: a neurophysiological study. *Experimental Brain Research*, 91(1):176–180, 1992.
- M. Strickert a B. Hammer. Merge SOM for temporal data. *Neurocomputing*, 64:39–71, 2005.
- J. Tani, M. Ito a Y. Sugita. Self-organization of distributedly represented multiple behavior schemata in a mirror system: reviews of robot experiments using RNNPB. *Neural Networks*, 17(8-9):1273–1289, 2004.
- H. van Hasselt. Reinforcement learning in continuous state and action spaces. *Reinforcement Learning*, 207–251, 2012.
- P. Vančo a I. Farkaš. Experimental comparison of recursive self-organizing maps for processing tree-structured data. *Neurocomputing*, 73(7-9):1362–1375, 2010.

Zoznam relevantných publikovaných prác autorky

- I. Farkaš, T. Malík a K. Rebrová. Grounding the meanings in sensorimotor behavior using reinforcement learning. *Frontiers in Neurorobotics*, 6(1), 2012. Doi: 103389/fnbot201200001.
- I. Farkaš, M. Malý a K. Rebrová. Mirror neurons – theoretical and computational issues. Výskumná správa, (TR-2011-28) Comenius University in Bratislava, 2011a.
- I. Farkaš, M. Malý a K. Rebrová. Porozumenie motorickým akciám – hypotéza kontinua. 61–68. Slezská univerzita v Opavě, Opava, ČR, 2011b.
- I. Farkaš a K. Rebrová. Bidirectional activation-based neural network learning algorithm. V *Proceedings of the International Conference on Artificial Neural Networks (ICANN)*, Sofia, Bulgaria. 2013. (in press).
- K. Rebrová. Stelesnené porozumenie a ideomotorická teória. V J. Rybár, redaktor, *Kognitívne paradigmy*, 127–150. Vydavateľstvo Európa, 2012.

- K. Rebrová a I. Farkaš. Robotický model systému zrkadliacich neurónov. V N. P. Kelemen J., redaktor, *Kognice a umělý život XII*, 231–238. Slezská univerzita, Opava, 2012.
- K. Rebrová a I. Farkaš. Robotický model systému zrkadliacich neurónov: experimentálna analýza. V J. Kelemen, J. Rybár, I. Farkaš a M. Takáč, redaktori, *Kognice a umělý život XIII*, 223–230. Slezská univerzita, Opava, 2013.
- K. Rebrová, M. Pecháč a I. Farkaš. Towards a robotic model of the mirror neuron system. V *Proceedings of the 3rd joint IEEE International Conference on Development and Learning and on Epigenetic Robotics*. Osaka, Japan, 2013. (in press).
- Rebrová K. and Farkaš I. Neurálne modely v kognitívnej robotike: porozumenie a pomenovávanie akcií. V K. V. Kelemen J., redaktor, *Kognice a umělý život XI*, 231–238. Slezská univerzita, Opava, 2011.
- J. Šilar, M. Kokoška, K. Rebrová a I. Farkaš. Motor resonance based desynchronization of the EEG mu rhythm. *Activitas Nervosa Superior Rediviva*, 53, 2011. Abstract.

Summary

My thesis discusses how a control architecture (be it a brain or an ANN) might utilize its motor systems to mediate understanding of the actions of others. The claim that perception and action are on a high level represented in a common framework is well rooted in empirical evidence. Embodied cognitive science suggests that motor representations in the brain which can be activated without the actual movement production might serve as simulation mechanisms allowing us to “step into the shoes” of the observed agent. Such mechanism is assumed to be a primary role of mirror neurons in area F5 of the macaque brain.

The discovery of mirror neurons gave rise to various computational models, mainly based on artificial neural networks. Since the computational models of MNS are considered a prominent tool for explaining the mirror neuron function and emergence, most of the models aim at capturing the actual neural circuitry, by having components that directly represent particular parts of the monkey’s brain. On the other hand, there are models that are closer to the paradigm of cognitive robotics, which use properties of the mirror neurons in a specific architecture. Such models do not aim to encompass the neural circuitry, but rather endow the agent with some special capabilities. The robotic MNS model proposed in this thesis also belongs to this category. However, it also aims on faithfully encompassing the MNS and its core brain areas.

The proposed model consists of various mutually interconnected neural networks. To provide a way to form bidirectional mappings between sensory and motor representations in a more biologically plausible way than standard BP, a new learning algorithm BAL was designed. Results from experiments with neural networks comprising this model showed that the model is successfully able to form coherent high-level representations and associate

them in 1-to-1 manner. Since bidirectional mapping of 1-to-4 patterns is an impossible task, adding more visual representations to the model reduces the success rate. However, in this way, various visual representations can trigger motor representations hence fulfilling the desired emergence of mirror neuron activity.

In the future, the model should be further evaluated and experimented with. The emergence of invariant representations on the basis of lateral interactions should be studied and possibly implemented in other parts of the model. Also the BAL algorithm should be further studied and possibly enhanced to drive the representations on the hidden layer to reorganize and display something like a mirroring activity, since mirror neurons were found also in the PF area. Last, but not least, a model providing a sensorimotor binding might be scaled to mediate simple language grounding. Concepts connected to action should be connected with motor areas as well. I believe that the common coding theory is most beneficial from the viewpoint of utilizing a theory in cognitive robotics. A common representational ground for action, perception, and meaning might extend to benefit the human-robot interaction.