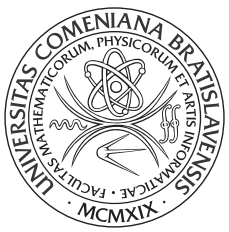
Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky



RNDr. Michal Malý

Autoreferát dizertačnej práce

Reinforcement Learning with Abstraction
(Učenie posilňovaním s abstrakciou)

na získanie akademického titulu philosophiae doctor
v odbore doktorandského štúdia:
9.2.1 Informatika

Bratislava 2012

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

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

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 Rovan, PhD.
Fakulta matematiky, fyziky a informatiky
Univerzita Komenského

1 Úvod

Učenie posilňovaním je moderná metóda učenia. Pomáha riešiť mnohé problémy, ktoré zahŕňajú nejaký koncept krátkodobej alebo dlhodobej odmeny. Učenie posilňovaním bolo úspešne aplikované na úlohy ako sú napríklad ovládanie robotov, plánovanie výťahov, rôzne úlohy v telekomunikáciách, alebo šach. Existujú však úlohy, ktoré napriek určitému konceptu odmeny nedokážeme zatiaľ dobre riešiť a existujúca teória poskytuje len málo rád. Táto práca má za cieľ pokúsiť sa riešiť problémy, ktoré sú len čiastočne pozorovateľnými Markovovskými procesmi, resp. nemajú Markovovskú vlastnosť. Zameriavame sa v nej na také úlohy, pri ktorých je vhodné, aby si agent sám odvodil model sveta.

V práci predstavujeme vlastný framework, ktorý umožňuje agentovi odvodiť si model sveta pomocou abstrakcie a ďalej tento model použiť pri svojom rozhodovaní. Ďalej predstavujeme implementáciu (nazývanú RLA - reinforcement learning with abstraction) a demonštrujeme jej funkčnosť a flexibilitu ju na praktických príkladoch. Porovnávame našu metódu s inými metódami.

1.1 Motivácia

Takmer všetky existujúce prístupy učenia posilňovaním sú založené na implicitne definovanom stavovom priestore. Priestor je obvykle určený rozsahom možných pozorovaní, ten môže byť diskrétny alebo spojitý.

Napríklad v hre piškvorky agent môže sledovať hraciu plochu a značky na nej umiestnené. Všetky možné usporiadania značiek tvoria stavový priestor. Niektoré rozloženia (ako napríklad štyri značky X a len jedna značka O) sa môžu ukázať počas trénovania ako nedostupné, alebo môžu byť už vylúčené zo stavového priestoru už počas návrhovej fázy.

Ale čo ak pozorovania nepokrývajú celý stavový priestor? Napríklad, ak si predstavíme problém navigácie v bludisku: Ak sú súradnice (x, y) pozorovateľné, dávajú agentovi plnú informáciu o jeho polohe v bludisku a umožňujú použiť klasické metódy, ktoré aproximujú ohodnotenia stavov a akcií. Ak však agent môže pozorovať len svoje okolie – povedzme pole, kde stojí, a susedné 4 polia) a nevie vopred rozmery bludiska, aká má byť definícia stavového priestoru? V bludisku môže existovať viacero polí s rovnakou konfiguráciou susedných polí a stien. Agent vo všetkých bude mať rovnaký vnem.

1.2 Cieľ

Cieľom je vytvoriť agenta, ktorý je schopný vytvoriť si model sveta zo svojich pozorovania a akcií. Tento model mu potom umožní rozlíšiť medzi stavmi s rovnakým pozorovaním. Tento postup môže byť tiež prínosný, ak priestor pozorovaní je zhodný so stavovým priestorom, ale je predpoklad, že sa v prostredí vyskytujú zaujímavé a užitočné vzťahy medzi stavmi, pričom tieto vzťahy dopredu nepoznáme alebo ich nechceme zadávať explicitne, lebo by to bolo prácne.

2 Framework pre učenie posilňovaním s abstrakciou

Všeobecný framework pre učenie posilňovaním s využitím abstrakcie pracuje v nasledovných krokoch:

1. Zo senzorov prichádza vstup, ten sa predspracuje a filtruje.
2. Senzorický vstup sa uloží do histórie.
3. Bezprostredne po každom vstupe, alebo vo vopred definovaných časových krokoch sa spúšťa abstrakčný modul.
4. Abstrakčný modul vyrobí model(y) sveta (v minimálnej popisnej dĺžke).
5. Metaplánovací modul analyzuje modely a spočíta rozdiely. Priradí interné odmeny akciám, ktoré sú schopné rozlíšenia medzi modelmi („systematický prieskum").
6. Modul učenia posilňovaním použije najlepší model sveta a rozhodne o ďalšej akcii, pričom berie do úvahy interné odmeny.
7. Akcia sa pošle na motorický výstup.

2.1 Popis komponentov

2.1.1 Senzory

Senzory pre agenta môže byť navrhnuté podľa potreby. Musí byť špecifikovaný rozsah možných výstupov pre každý senzor. Predspracovanie a filtrácia zabezpečujú technické detaily spracovania senzorového vstupu, ako napríklad kalibráciu, alebo grafické algoritmy.

2.1.2 História

Tento modul ukladá všetky akcie, pozorovania a odmeny spolu s časom. Celá sekvencia histórie je zasielaná abstrakčnému modulu.

2.1.3 Abstrakčný modul

To je najdôležitejší modul. Zoberie históriu a snaží sa nájsť model(y), čo možno najlepšie vysvetľuje(ú) skúsenosti agenta. Na tento účel navrhujeme použiť gramatickú indukciu alebo inferenciu automatu.

Očakávaným výstupom tohto modulu je model sveta, alebo automat, ktorý tento model sveta reprezentuje. Ak sa používa metaplánovací modul, výstupom je množina modelov spolu s ich ohodnotením podľa popisnej dĺžky a počtu chýb.

Vytvorený model alebo modely nemusia byť nutne minimálne, pre silné formalizmy to ani nie je ani možné. Aj pseudo-minimálny model (najlepší nájdený) je užitočný. Modul môže využívať heuristiky. Takisto môže využiť model z predchádzajúceho kroku a pokúsiť sa ho upraviť, aby vyhovoval novým poznatkom.

2.1.4 Modul učenia posilňovaním

Tento modul využíva model sveta, aby zistil, ktorá akcia je najvhodnejšia. To sa dá dosiahnuť napríklad použitím dynamického programovania.

2.1.5 Metaplánovací modul

Tento modul analyzuje modely navrhnuté abstrakčným modelom a riadi agenta tak, aby bolo možné rozhodnúť, ktorý z modulov je ten správny. To umožňuje systematické skúmanie prostredia. Modul počíta, pri akej postupnosti akcií sa predpovede modelov budú líšiť, a môže ovplyvniť ohodnotenie akcií tak, aby sa táto postupnosť vykonala.

2.2 Očakávané vlastnosti frameworku

Agent postavený na popísanom frameworku bude systematicky spoznávať prostredie a optimalizovať svoje správanie, aby získal čo najväčšiu odmenu. Počas prieskumu si dokáže vytvoriť model sveta, ktorý môže obsahovať znalosti, ktoré nie sú explicitne obsiahnuté v pozorovaniach. Samozrejme, prvé kroky agenta sú náhodné a nesystematické, ale ako čas postupuje a agent skúma prostredie, jeho model sveta by mal byť stále bližšie k realite. Ak sa použije dostatočne silný formalizmus, môže agent odvodzovať svoje vlastné koncepty, ktoré neboli explicitne formulované pri jeho návrhu. Ak má agent pochybnosti o tom, ktorý model vybrať, môže zvoliť postupnosť akcií, aby svoje pochybnosti vyriešil. To umožňuje agentovi systematicky skúmať svet.

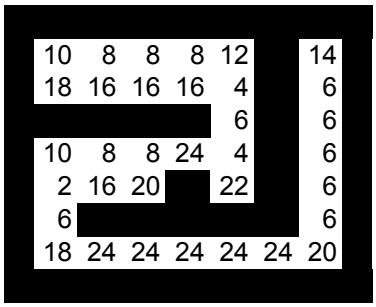
3 Implementácia

Vytvorili sme implementáciu predstaveného frameworku. Predstavujeme typické úlohy, na ktorých sme testovali našu implementáciu.

Agent beží v cykle „pozorovanie-učenie-konanie". Počas „pozorovania" a „konania" agent interaguje s prostredím. Časť "učenie" pozostáva zo zaznamenania skúsenosti agenta, vytvorenia modelu sveta, a vypočítania najlepšej akcie vzhľadom na novovypočítaný model. Vytvorenie modelu sveta pozostáva z vnútorného cyklu, v ktorom je iteratívne volaný SAT solver, ktorý má za úlohu nájsť riešenie formuly, obsahujúcej prepis celej skúsenosti agenta transformovanej do logických propozícií. Na základe výsledku zo SAT solvera si agent vytvorí alebo upraví model sveta, a tento sa použije na určenie nasledujúcej akcie.

3.1 Problém bludiska

Prvá úloha – problém bludiska — predstavuje jednoduchý problém, ktorý má vlastnosti čiastočne pozorovateľného Markovovského procesu. Na prvý pohľad by sa mohlo zdať, že tento problém je možné riešiť pomocou štandardných traverzovacích algoritmov alebo



Obr. 1: Príklad bludiska. Čísla reprezentujú pozorovanie, ktoré agent dostáva.

napríklad pomocou lokalizačných metód ako je SLAM (simultánna lokalizácia a mapovanie), ale nie je tomu tak. Navyše bludisko predstavuje len jednu konkretizáciu všeobecného problému nekompletne pozorovateľného sveta.

Predpokladajme, že máme bludisko, ktoré agent vopred nepozná. Ako prejsť toto bludisko a vytvoriť jeho mapu? Samozrejme, existujú štandardné prístupy k riešeniu tohto problému, za predpokladu, že je možné umiestniť do prostredia značku, aby sme vedeli, že sme navštívili konkrétne pole (Tremauxov algoritmus, Tarryho prieskum).

V našom prípade nie je možné umiestňovať značky ani nebudeme agentovi dávať informáciu o súradniciach. Tiež nechceme, aby mal agent explicitnú informáciu, že je v bludisku určitého typu. Požadujeme, aby vyriešil túto úlohu bez toho, aby dostal tieto informácie. To tiež bude demonštrovať jeho schopnosť riešiť aj rozsiahlejšie problémy.

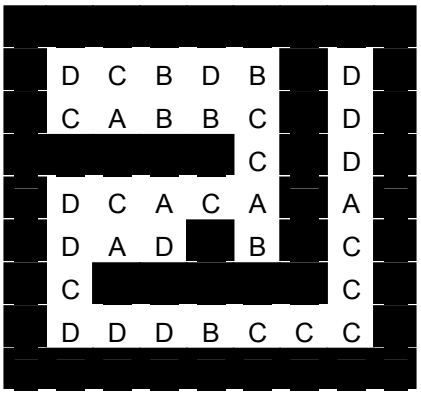
Inými slovami, chceme rovnaký algoritmus, ktorý bude schopný prejsť bludisko, aj keď bludisko obsahuje teleporty alebo jednosmerné dvere. Agent má k dispozícii len pozorovania z určenej množiny O a môže vykonať akcie z množiny A. Pozorovanie dáva agentovi informácie o jeho okolí. Akcie a pozorovania sú atomické a nenesú žiadnu ďalšiu informáciu, ako napríklad súradnice alebo smer. Ak primerane zmeníme množiny O, A, môžeme rovnakého agenta použiť v 2D alebo v 3D bludisku.

V princípe sa nedá bez možnosti umiestňovať značky presne určiť, ako bludisko vyzerá – nemožno napríklad rozlíšiť dva regulárne grafy (trojuholník a štvorec), pretože každý vrchol má dve hrany a teda rovnaké pozorovanie.

Bez ohľadu na tento výsledok má zmysel riešiť problém bludisko z pohľadu učenia posilňovaním. Ak sú možné dva modely sveta, a nie sme schopní rozoznať ich, je rozumné predpokladať najjednoduchší model. Ak sa začnú pozorovania líšiť od našich očakávaní, sme schopní vybrať si iný model.

3.2 Jednoduché bludisko

Bludisko môže vyzerať ako na Obr. 1. Pozorovanie agenta je obmedzené na 4-bitovú informáciu o okolí – či štyri susedné polia obsahujú stenu alebo nie.



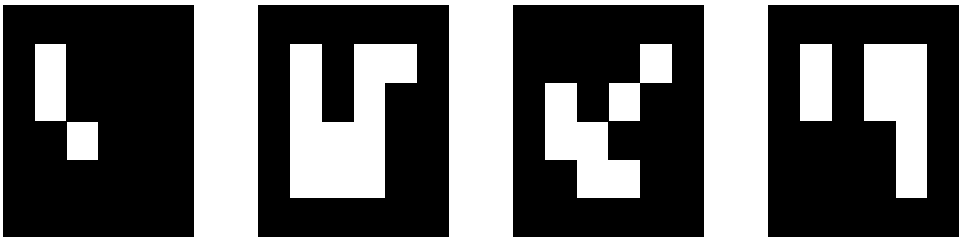
Obr. 2: "Písmenkové" bludisko. Písmená reprezentujú pozorovanie, ktoré dostane agent, ak sa nachádza na danom políčku.

3.3 "Písmenkové" bludisko

Framework a jeho implementácia boli navrhnuté tak, aby boli flexibilné a vytvárali model sveta podľa potreby. Preto algoritmus môže byť použitý bez zmeny napríklad na prieskum bludiska s poľami označenými písmenami – agent vidí len písmeno na aktuálnom poli a nemá žiadne ďalšie informácie (pozri Obr. 2). Je potrebné iba zmeniť počet pozorovaní, aby zodpovedal počtu možných písmen.

3.4 3D bludisko

Algoritmus môže byť tiež použitý na prieskum 3D bludisko, tu musíme upraviť počet akcií (6 namiesto 4). Príklad bludiska je na Obr. 1.



Tabuľka 1: Príklad 3D bludiska (27 voľných políčok) preskúmaného agentom na 199 krokov za 486 sekúnd.

3.5 „Objavovanie" protokolu

Princíp fungovania našej metódy nie je obmedzený na prechádzanie bludísk. Metóda môže byť použitá na rôzne úlohy, charakterizované čiastočným pozorovaním stavu a konečnosta-

vovou povahou prostredia. Jednoduchý príklad môže byť akási „hackerská" hra – objavovanie protokolu.

Predpokladajme, že nepoznáme presný protokol na zasielanie e-mailov (Simple Mail Transfer Protocol, SMTP), ale vieme napojiť agenta na SMTP server, aby mohol posielať požiadavky. Predpokladajme tiež, že dokážeme detekovať, či e-mail bol úspešne odoslaný.

Nášho agenta potom necháme naučiť sa protokol, skúšať rôzne akcie, ktoré vyvolajú rôzne odpovede od servera. Väčšina požiadaviek pravdepodobne spôsobí chybovú hlášku, ale niektoré povedú k odlišnej odpovedi a spôsobia zmenu interného stavu servera. Táto zmena je určená protokolom. Príklad jednoduchej SMTP komunikácie je na Obr. 3.

220 server ESMTP ready
HELO server
220 I am glad to meet you
MAIL FROM:<agent@rla.sk>
250 OK
RCPT TO:<user@rla.sk>
250 OK
DATA
354 End data with <CR><LF>.<CR><LF>
From: agent@rla.sk
To: user@rla.sk
Subject: first mail
Dear user,
this is my first mail.

Sincerely,
Your RLA agent
.
250 OK: queued as 71

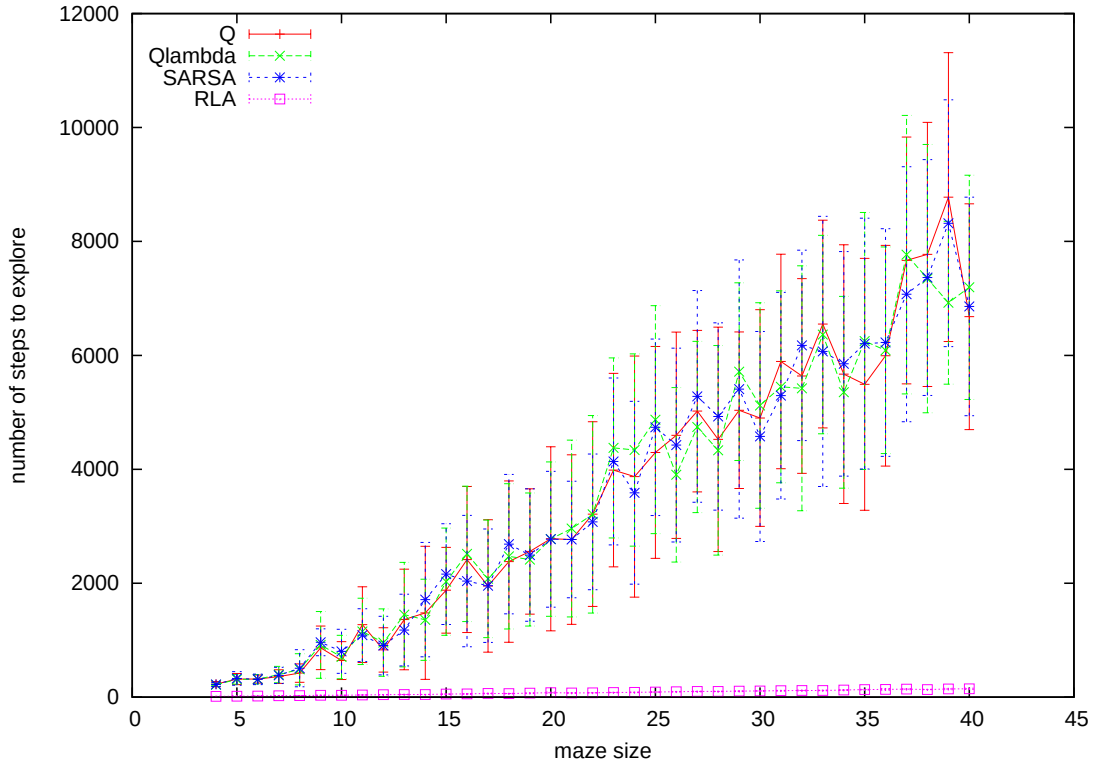
Obr. 3: Príklad SMTP komunikácie. **Boldom** sú značené odpovede servera, požiadavky klienta sú *kurzívou*.

Testovali sme RLA agenta v jednoduchom simulovanom SMTP prostredí. Modifikovali sme agenta tak, aby vedel posielať celé reťazce, ako napr. „HELO", a ďalšie reťazce, ktoré protokol vyžaduje. Agent sa dokázal naučiť SMTP protokol na 113 krokov.

4 Výkonnosť a porovnanie s inými metódami

Na porovnanie sme si zvolili tri generické metódy: Q-learning, Q(λ) a SARSA, implementované v knižnici PyBrain (Schaul et al., 2010), ktoré sme spolu s našou metódou spustili na

sade vstupov – bludísk s rôznou veľkosťou. Merali sme čas a počet krokov potrebných na úplné preskúmanie prostredia a vytvorenie modelu. V prípade RLA, model bol explicitný, pri ostatných metódach model vyplýva z ohodnotenia stavov. Generické metódy neboli navrhnuté na beh v čiastočne pozorovateľnom prostredí. Preto dostali plné pozorovanie. Naša metóda tak bola trochu „znevýhodnená", ale aj tak považujeme toto porovnanie za užitočné. Výsledky možno vidieť na Obr. 4.



Obr. 4: Počet krokov, ktoré potrebovali metódy SARSA, Q-learning and Q(λ) na objavenie prostredia, v závislosti od veľkosti bludiska (počtu voľných políčok). Vertikálne čiary znázorňujú štandardnú odchýlku z 10 behov. Na porovnanie je pridaná naša metóda (RLA).

Vidíme, že metóda RLA je v porovnaní s generickými metódami lepšia, čo sa týka počtu krokov. Čas pre výpočet pri RLA však rastie exponenciálne. To sa dá očakávať, keďže RLA používa SAT solver na výpočet modelov. S dnes dostupnými výkonmi procesorov je použiteľná hranica asi 50–60 stavov (cca 1 h).

4.1 Ako sa správajú generické metódy v čiastočne pozorovateľnom svete?

Zmenili sme kód použitý v predchádzajúcom odseku tak, aby generické metódy dostali len čiastočné pozorovanie, 3×3 políčka okolo agenta. Mohli teda v rôznych stavoch mať rovnaké pozorovanie. Výsledok bol, že žiadna z generických metód nebola schopná vytvoriť si model sveta.

Generické metódy totiž priemerujú informáciu o odmene, súvisiacej s konkrétnym stavom. Ak však svet obsahuje viacero stavov s rovnakým pozorovaním, ale rôznou utilitou, tieto utility sa spriemerujú. Bludisko však obsahuje mnoho perceptuálne zameniteľných stavov, ktoré sú rozložené približne rovnomerne po ploche bludiska. Majú teda celkom uniformnú distribúciu utility, keďže táto v jednoduchom bludisku závisí od vzdialenosti k cieľu. „Model" generickej metódy bude teda pozostávať zo stavov s takmer rovnakou utilitou, rovnou priemeru všetkých stavov, plus malý šum. Akcie agenta používajúceho taký model budú veľmi podobné náhodnému kráčaniu.

4.2 Budúca práca

Bolo by možné použiť iný spôsob gramatickej inferencie alebo inferencie automatu, napríklad metódu ECGI (Rulot et al., 1989), ktorá produkuje regulárnu gramatiku, genetický algoritmus (Javed et al., 2004) alebo iné symbolické techniky (Alquezar, 1997; Rivest and Schapire, 1993). Bolo by tiež možné upraviť náš algoritmus na použitie SAT solvera na nájdenie silnejšej gramatiky (napr. bezkontextovej), alebo automatu s jedným, prípadne dvoma počítadlami. Avšak zložitosť - počet kláuz - by bola zrejme oveľa väčšia. Je tiež potrebné mať na pamäti limit nevypočítateľnosti – automat s dvoma počítadlami je pri vhodnom kódovaní ekvivalentný Turingovmu stroju.

5 Prínos práce

Táto práca sa zaoberá čiastočne pozorovateľnými problémami učení posilňovaním. Táto trieda problémov je stále otvorená a existujúce teórie poskytujú iba čiastočné riešenie. Je uvedený prehľad existujúcich prístupov spolu so všeobecným teoretickým pozadím učenia posilňovaním.

Navrhli sme novú metódu pre agenta používajúceho učenie posilňovaním, ktorá využíva SAT solver na inferenciu konečného automatu. Naša metóda, učenie posilňovaním s abstrakciou, je schopná preskúmať prostredie a vytvoriť si model sveta, vďaka čomu dokáže vyriešiť rôzne problémy: klasické bludisko, „písmenkové" bludisko (bludisko, kde agent vidí políčka iba ako písmená, ktoré im boli ľubovoľne pridelené), trojdimenzionálne bludisko, bludisko s teleportami, a tiež problém „objavovania protokolu".

Výkon našej metódy sme testovali na klasickom bludisku s rôznymi veľkosťami. Naša metóda bola lepšia ako metóda UDM (McCallum, 1992). Z pohľadu počtu krokov, potrebných na objavenie prostredia, bola dokonca lepšia ako generické metódy, ktoré dostali úplné

pozorovanie. Ak sme generickým metódam neumožnili úplné pozorovanie, neboli schopné vyriešiť problém. To značí, že pre niektoré úlohy s čiastočným pozorovaním je naša metóda jednou z mála aplikovateľných metód.

Literatúra

Alquezar, R. (1997). *Symbolic and connectionist learning techniques for grammatical inference*. PhD thesis, Universitat Politecnica de Catalunya.

Javed, F., Bryant, B., Črepinšek, M., Mernik, M., and Sprague, A. (2004). Context-free grammar induction using genetic programming. In *Proceedings of the 42nd annual Southeast regional conference*, pages 404–405. ACM.

McCallum, R. (1992). "first results with utile distinction memory for reinforcement learning". Technical Report 446, Department of Computer Science, University of Rochester.

Rivest, R. and Schapire, R. (1993). Inference of finite automata using homing sequences. *Machine Learning: From Theory to Applications*, pages 51–73.

Rulot, H., Prieto, N., and Vidal, E. (1989). Learning accurate finite-state structural models of words through the ECGI algorithm. In *International Conference on Acoustics, Speech, and Signal Processing*, pages 643–646. IEEE.

Schaul, T., Bayer, J., Wierstra, D., Sun, Y., Felder, M., Sehnke, F., Rückstieß, T., and Schmidhuber, J. (2010). PyBrain. *Journal of Machine Learning Research*.

UNIVERZITA KOMENSKÉHO
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

Zoznam publikačnej činnosti

RNDr. Michal Malý

AFC Publikované príspevky na zahraničných vedeckých konferenciách

AFC01 Malý, Michal 100%: Kognitívny assembler
Lit. 5 zázn.
In: Kognice a umělý život VIII. - Opava : Slezská univerzita, 2008. - S. 215-219. - ISBN 978-80-7248-462-1
[Kognice a umělý život 2008 : Český a slovenský seminář o kognici a umělém životě. 8., Praha, 26.-29.5.2008]

AFC02 Malý, Michal 100%: Cognitive assembler
Rozšírená práca
Lit. 5 zázn.
In: AKRR'08 : The 2nd International and Interdisciplinary Conference on Adaptive Knowledge Representation and Reasoning. - Helsinki: Helsinki University of Techonology, 2008. - S. 60-64. - ISBN 978-951-22-9525-8
[AKRR 2008 : Adaptive Knowledge Representation and Reasoning : International and Interdisciplinary Conference. 2nd, Porvoo, 17.-19.9.2008]

AFC03 Malý, Michal 100%: Využitie abstrakcie v učení posilňovaním
Lit. 7 zázn.
In: Kognice a umělý život X. - Opava : Slezská univerzita, 2010. - S. 235-238. - ISBN 978-80-7248-589-5
[Kognice a umělý život 2010 : Český a slovenský seminář o kognici a umělém životě. 10., Ostravice, 31.5.-4.6.2010]

AFC04 Malý, Michal 100%: Using a network of untrusted computers for secure computing [elektronický dokument]
Lit. 13 zázn.
In: ICAS 2011 : The Seventh International Conference on Autonomic and Autonomous Systems (USB kľúč). - [s.l.] : ICAS, 2011. - S. 57-61. - ISBN 978-1-61208-006-2
[ICAS 2011 : Autonomic and Autonomous Systems : International Conference. 17th, Venice, 22.-27.5.2011]

AFD Publikované príspevky na domácich vedeckých konferenciách

AFD01 Malý, Michal 100%: Semi-automatic creation of a stemming dictionary of an inflecting language using grammatical induction
Recenzované
Lit. 14 zázn.
In: Študentská vedecká konferencia FMFI UK, Bratislava 2010 : Zborník príspevkov. - Bratislava : Fakulta matematiky, fyziky a informatiky UK, 2010. - S. 295-300. - ISBN 978-80-89186-68-6
[Študentská vedecká konferencia 2010. Bratislava, 28.4.2010]
POZNÁMKA: Vyšlo aj na CD ROM - Študentská vedecká konferencia FMFI UK, Bratislava 2010 : Zborník príspevkov. - Bratislava : Fakulta matematiky, fyziky a informatiky UK, 2010. - S. 295-300. - ISBN 978-80-89186-69-3

AFD02 Farkaš, Igor 40% - Malý, Michal 30% - Rebrová, Kristína 30%: Porozumenie motorickým akciám - hypotéza kontinua

Recenzované
Lit. 29 zázn., 1 obr.
In: Kognice a umělý život XI. - Opava : Slezská univerzita, 2011. - S. 61-68. - ISBN 978-80-7248-644-1
[Kognícia a umelý život 2011 : konferencia. 11., Smolenice, 4.-7.4.2011]

AFD03 Malý, Michal 100%: Kognitívna mapa bludiska
Recenzované
Lit. 13 zázn., 2 obr.
In: Kognice a umělý život XI. - Opava : Slezská univerzita, 2011. - S. 141-147. - ISBN 978-80-7248-644-1
[Kognícia a umelý život 2011 : konferencia. 11., Smolenice, 4.-7.4.2011]

AFD04 Malý, Michal 50% - Kocun, Miroslav 50%: Fyzikálna Eliza
Recenzované
Lit. 7 zázn.
In: Kognice a umělý život XI. - Opava : Slezská univerzita, 2011. - S. 149-152. - ISBN 978-80-7248-644-1
[Kognícia a umelý život 2011 : konferencia. 11., Smolenice, 4.-7.4.2011]

AFJ Preprinty vedeckých prác vydané v domácich vydavateľstvách

AFJ01 Farkaš, Igor 34% - Malý, Michal 33% - Rebrová, Kristína 33%: Mirror neurons theoretical and computational issues [elektronický dokument]. - Bratislava : Faculty of Mathematics, Physics and Informatics Comenius University, 2011. - (Technical Reports in Informatics ; TR-2011-028)
Popis urobený 10.11.2011
Lit. 88 zázn., 5 obr.
URL: http://kedrigern.dcs.fmph.uniba.sk/reports/display.php?id=34http://kedrigern.dcs.fmph.uniba.sk/reports/index.php

Štatistika kategórií (Záznamov spolu: 9):
AFC Publikované príspevky na zahraničných vedeckých konferenciách (4)
AFD Publikované príspevky na domácich vedeckých konferenciách (4)
AFJ Preprinty vedeckých prác vydané v domácich vydavateľstvách (1)

11. 10. 2012

Summary

Reinforcement learning is a modern method of learning. It helps to solve many problems which comprise some notion of a short-term or long-term reward. It was successfully applied in the fields like robot control, elevator scheduling, problems in telecommunication, and chess. However, there are tasks involving the concept of reward, that we are not able to successfully solve. The existing theory provides only a few clues. The goal of this thesis is to try to solve the problems of only partially observable Markov processes or those that do not have the Markov property. We focus on tasks where a possibility to derive a world model brings an advantage.

First, we explain basic concepts of reinforcement learning, together with some other concepts, which are necessary for comprehension of the following text. We show limitations of basic methods of reinforcement learning. Then we present our own framework, which enables the agent to derive a world model using abstraction and use the model for decision making. Next, we present a prototypical implementation and demonstrate it on practical examples. Our method, reinforcement learning with abstraction (RLA) was able to solve a classical maze, a "letter" maze (a maze, where agent sees cells only as arbitrary letters assigned to them), three dimensional maze, maze with teleports, and was also able to solve a "protocol discovery" problem.

We have tested the performance on classical maze problems with varying size. Our method performed better than UDM method (McCallum, 1992). It performed better in terms of steps necessary to discover the environment even with generic method, which were allowed full observation. When deprived of full observation, generic methods are not able to solve the problem. Thus, for some partial observation tasks, RLA can be one of few methods available.

We conclude the work with potential contribution and possible improvements.