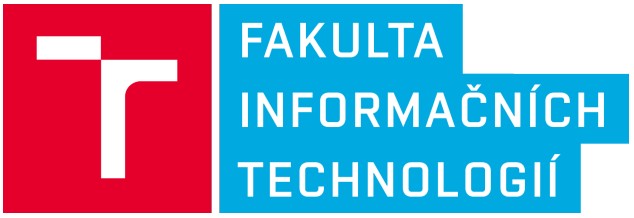


Vysoké učení technické v Brně

THE

Teorie her

Projekt

Rozhodování a zpětnovazebné učení opic v kompetativních hrách

---

Autor:
Petr Buchal

Login:
xbucha02

31. Leden, 2020

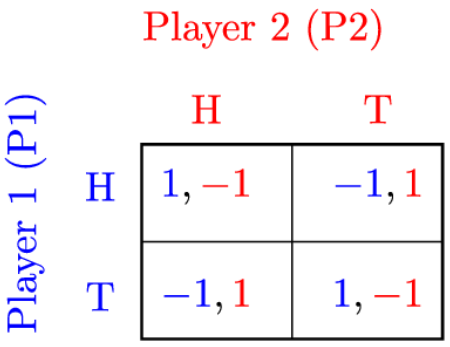
Obsah

1 Úvod

Z přednášek doktora Hrubého mně zaujaly experimenty prováděné na opicích, a tak jsem se rozhodl zpracovat jeden z nejznámějších článků "Reinforcement learning and decision making in monkeys during a competitive game"[3].

Rozhodování je proces vybírání určité akce z množiny alternativních voleb v dané situaci. Tento proces se dá studovat ze dvou různých pohledů. Ekonomové se zabývají matematickým aparátem na charakterizaci optimálních rozhodovacích pravidel. Psychologové a vědci zkoumající chování zvířat zase srovnávají predikce aparátu ekonomů se skutečným chováním živých stvoření.

Důležitý krok k pochopení mechanik rozhodování je skutečnost, jak se tyto procesy mění se zkušeností. Relativně jednoduché učící algoritmy jsou pro takový úkol dostačující, když je k dispozici malé omezené množství akcí a statické prostředí. V reálném světě je ovšem prostředí téměř vždy dynamické. Navíc u zvířat, která ve svém prostředí interagují s jinými, je problém komplikovanější kvůli faktu, že rozhodnutí jednoho může být ovlivněno rozhodnutími ostatních. Problém nalezení optimální rozhodovací strategie v multiagentním prostředím může být matematicky analyzován pomocí teorie her. Hra je definována pro konečný počet hráčů pomocí seznamu dostupných akcí každému hráči a funkcí užitku, která přiřazuje odměnu každému hráči jakožto funkce rozhodnutí všech hráčů [2]. Řešení takové hry je často jedno popřípadě vícero Nashových ekvilibrií. Nashovo ekvilibrium je konkrétní množina strategií všech hráčů, kde žádný hráč nemůže zvýšit svůj užitek individuální změnou strategie [2].



Obrázek 1: Hra matching pennies, převzato z [1].

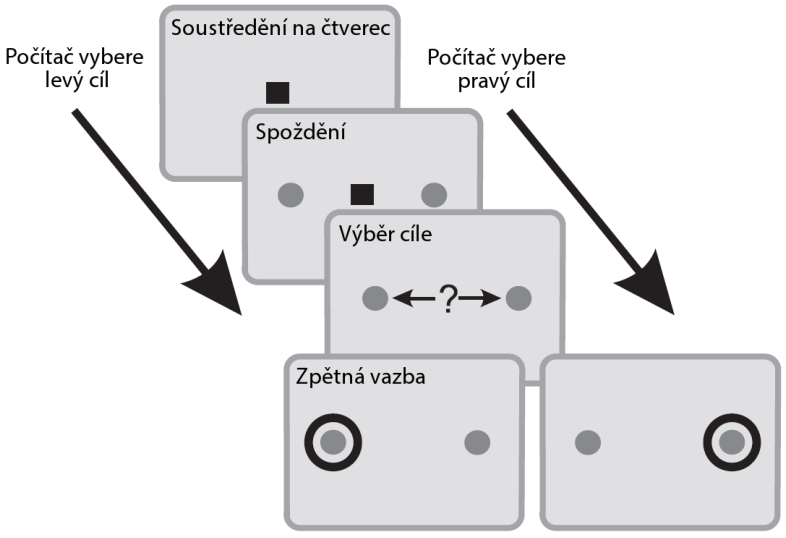
Studie v článku [3] zkoumá rozhodování opic v jednoduché hře s nulovým součtem známé jako "matching pennies", viz obrázek 1. Nashovo ekvilibrium v této hře je pro oba hráče stejné a to zahrát obě volby se stejnou pravděpodobností. V teorii her je taková strategie nazývaná jako smíšená a je definována funkcí hustoty

pravděpodobnosti přes jednotlivé strategie. Cílem studie [3] bylo určit, jak moc se blíží rozhodovací proces opic Nashovému ekvilibriu.

2 Metody

Ve studii byly použity tři opice (*Macaca mulatta*) s váhou 7-12 kg, které jsou dále označovány jako C, E a F. Zvířata byla v průběhu experimentu usazena v křesle pro primáty čelem k počítači, který byl vzdálen přibližně 57 cm od jejich očí. Všechny vizuální podněty byly prezentovány prostřednictvím monitoru. Pozice zvířecích očí byly zaznamenávány s frekvencí 250 Hz pomocí speciálního trackeru očí.

Na začátku každého pokusu bylo po opicích vyžadováno, aby se zaměřily na žlutý čtverec uprostřed obrazovky. Po půl vteřině soustředění na čtverec se na obrazovce objevily 2 zelené kruhy, jeden z každé strany čtverce. Zvíře si udržovalo soustředění na centrální čtverec se zpožděním 0.5 vteřiny. Poté se čtverec rozplynul a po zvířeti byl během následující vteřiny vyžadován sakadický pohyb (velmi rychlý pohyb [4]) očí na jeden z cílů (kruhů) a udržení pohledu po dobu 0.5 vteřiny. Po uplynutí požadované doby soustředění se po dobu 0.2 vteřiny zobrazil červený kruh kolem cíle, který vybral soupeř - počítač. Zvíře bylo odměněno, pokud zvítězilo, viz obrázek 1. Schéma experimentu lze vidět na obrázku 2.



Obrázek 2: Schéma experimentu, převzato z [3].

2.1 Algoritmus protihráče

Experiment probíhal řadu dní s tím, že každý den používal počítač k volbě cílů jeden ze tří algoritmů, označovaných dále jako 0,1 a 2.

V algoritmu 0 počítač vybíral cíl náhodně s pravděpodobností 50 % pro oba dva. Tato strategie je ve hře matching pennies Nashovým ekvilibriem. Pokud jeden z hráčů zahraje strategii, která je ekvilibriem, očekávaný zisk obou hráčů je fixní nezávisle na strategii druhého. Tento algoritmus byl tedy nasazen k prozkoumání počátečních strategií opic před tím, než počítač použil více sofistikované algoritmy.

V algoritmu 1 si počítač uchovával celou sekvenci rozhodnutí udělaných zvířetem během jednoho dne. V každé hře pak počítač použil tuto informaci, aby spočítal podmíněné pravděpodobnosti výběru konkrétních cílů zvířetem na základě zvolených cílů v N předchozích kolech (N=0 až 4). Algoritmus testoval nulovou hypotézu, že pravděpodobnost pro každou z těchto podmíněných pravděpodobností je 0.5 (binomický test, p ¡ 0.05). Pokud žádná z těchto hypotéz nebyla odmítnuta, algoritmus předpokládal, že zvíře zvolilo oba cíle se stejnou pravděpodobností nezávisle na předchozích hrách a počítač vybral cíl náhodně stejně jako v algoritmu 0. Pokud alespoň jedna hypotéza byla odmítnuta, počítač vybral svůj cíl pomocí podmíněné pravděpodobnosti s největší odchylkou od 0.5, která byla statisticky nezanedbatelná. Tohoto bylo dosaženo za pomocí výběru cíle s pravděpodobností 1-p, kde zvíře vybralo cíl s pravděpodobností p. Například pokud zvíře zvolilo pravý cíl s pravděpodobností 80 %, počítač by vybral stejný cíl s pravděpodobností 20 %. Po opicích tedy bylo vyžadováno, aby vybíraly oba cíle se stejnou pravděpodobností nezávisle na svých předchozích volbách.

V algoritmu 2 počítač použil historii voleb a odměn zvířete v daný den na predikci jeho volby v dalším kole. Algoritmus 2 počítá sérii podmíněných pravděpodobností, se kterými daná opice zvolí daný cíl, na základě jejích voleb a zisku v předchozích N kolech (N=*1* až 4). Stejně jako v algoritmu 1 byla každá z těchto podmíněných pravděpodobností testována proti nulové hypotéze, že podmíněná pravděpodobnost je 0.5. Výběr cíle zakládající se na odmítnutí některé hypotézy byl stejný jako v algoritmu 1. Algoritmus 2 tedy po opicích vyžadoval nejen zvolení cíle se stejnou pravděpodobností ale i to, že tato volba musela být nezávislá na kombinaci předešlých voleb obou hráčů.

Opice byly na začátku vycvičeny tak, že dokázaly provést zpožděný sakadický pohyb jako v úloze popsané výš s rozdílem, že byl dostupný pouze jeden náhodný cíl v každém kole hry. Následně hrály opice hru matching pennies proti algoritmu 0 od 2 do 4 dnů. Poté několik týdnů stejnou hru proti algoritmu 1 a nakonec proti algoritmu 2, viz obrázek 3. Pořadí algoritmů, proti kterým opice hrály, bylo ve všech případech stejné. Bylo tomu tak zejména kvůli tomu, že sofistikovanější algoritmy vyžadovaly

adaptaci zvířat a bylo zjištěno, že v okamžiku kdy zvíře přizpůsobilo své rozhodování sofistikovanějšímu algoritmu, nebylo pro něj snadné, vrátit se k původně užívané strategii, když počítač začal používat jednodušší algoritmus.

Algorithm	Animal	*N* days	*N* trials	Trials/day±S.D.
0	C	2	5327	2663±359
	E	2	1669	835±483
	F	4	4413	1103±155
	All	8	11,409	1426±812
1	C	36	50,143	1393±789
	E	63	70,111	1113±707
	F	26	35,504	1366±202
	All	125	155,758	1246±672
2	C	33	28,344	859±510
	E	41	45,769	1116±1213
	F	23	38,556	1676±373
	All	97	112,669	1162±909

Obrázek 3: Počty dní experimentování s jednotlivými algoritmy a počty her jednotlivých opic během nich, převzato z [3].

2.2 Analýza dat

2.2.1 Analýza pomocí logistická regrese

Pomocí logistické regrese bylo testováno, zdali bylo rozhodování opic ovlivněno volbami v předchozích hrách (volby opic i počítače). Pravděpodobnost $p_t(\mathrm{R})$, že zvíře vybere pravý cíl ve hře t, má co do činění s historií voleb obou hráčů

$$logit\{p_t(R)\} \equiv log(\frac{p_t(R)}{1-p_t(R)}) = a_0 + \sum_{i=1}^{5} a_i M_{t-i} + \sum_{i=1}^{5} b_i C_{t-i},$$

kde M_t a C_t reprezentují volby zvířete a počítače ve hře t (M_t nebo $C_t = 1$ pokud byl vybrán pravý cíl, jinak 0) a a_i a b_i jsou regresní koeficienty.

2.2.2 Model zpětnovazebního učení

V algoritmech zpětnovazebního učení poskytuje hodnotová funkce očekávanou odměnu. Hodnotová funkce ve hře t pro daný cíl x (x =L pro levý cíl, R pro pravý cíl) $V_t(x)$ je aktualizovaná po každé hře podle vzorce

$$V_{t+1}(x) = \alpha V_t(x) + \Delta_t(x),$$

kde α je diskontní faktor a $\Delta_t(x)$ reflektuje změnu v hodnotící funkci. Bylo předpokládáno že $V_1(R) = V_1(L) = 0$. V modelu $\Delta_t(x) = \Delta_1$, pokud zvíře vybere cíl x a je odměněno a $\Delta_t(x) = \Delta_2$, pokud zvíře vybere cíl x a není odměněno a $\Delta_t(x) = 0$, pokud zvíře nevybere cíl. Stejně jak je popsáno v podkapitole 2.2.1 pravděpodobnost, že zvíře vybrere pravý cíl je určena následující logitovou trasformací

$$logit\{p_t(R)\} \equiv log(\frac{p_t(R)}{1 - p_t(R)}) = V_t(R) - V_t(L).$$

2.2.3 Entropie a vzájemná informace

Míra náhodného rozhodování v sekvencích voleb byla kvantifikována pomocí entropie a vzájemné informace. Obě tato měření byla vyhodnocena za použití sekvence voleb dvou hráčů ve třech po sobě jdoucích hrách. Tohle umožnilo relativně spolehlivé odhady na základě limitování počtu možných výsledků. Výsledky byly kvantitativně podobné pro delší sekvence. Specificky pokud je k možných výsledků a i-tá událost má pravděpodobnost p_i, entropie H je definována následovně

$$H = -\sum_{i=1}^{k} p_i log_2(p_i).$$

Když je entropie počítána pouze na základě volby zvířete ve třech po sobě jdoucích kolech, je zde celkem 8 možných výsledků ($k = 2^3 = 8$). Entropie byla zároveň počítána na základě volby zvířete ve třech po sobě jdoucích kolech a prvních dvou voleb počítače z těchto tří ($k = 2^5 = 32$). Když je entropie určena s použitím pravděpodobností z konečného počtu vzorků, odhad entropie je ovlivněný. Kvůli tomu byla entropie určována následovně

$$H = -\sum_{i=1}^{k} \hat{p}_i log_2(\hat{p}_i) + \frac{k-1}{1.3863N},$$

kde $\hat{p}_i$ značí odhad maximální pravděpodobnosti pro p_i a N vzorků.

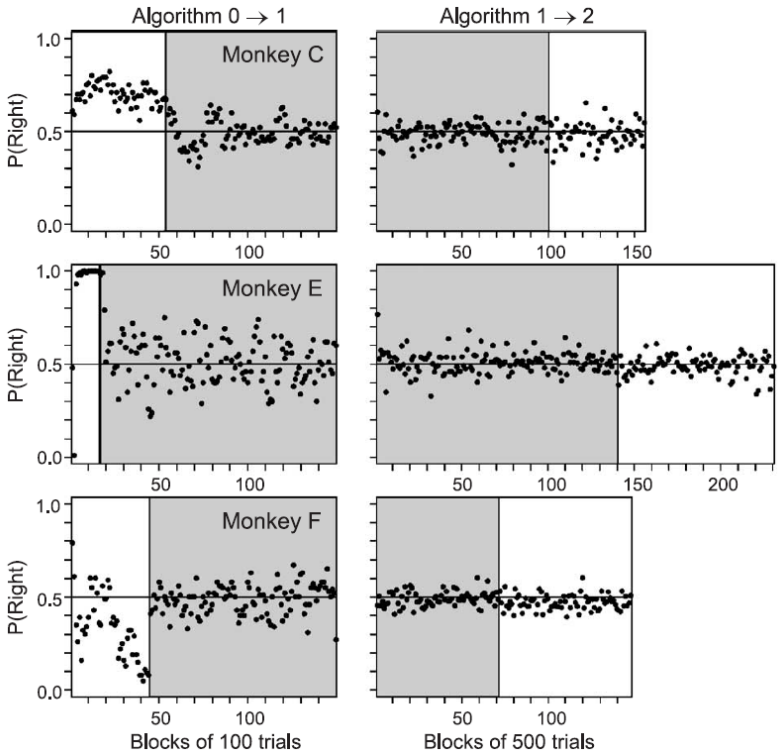
Vzájemná informace byla vypočítána na základě sekvencí voleb dvou hráčů ve třech po sobě jdoucích kolech (vstup) a volbou zvířete v dalším kole (výstup). Byla počítána následovně, aby bylo odstraněno ovlivnění plynoucí z konečného počtu vzorků,

$$H = -\sum_{i}^{r}\sum_{j}^{c}\hat{p}_{ij}log_2(\frac{\hat{p}_{ij}}{\hat{p}_i\hat{p}_j}) - \frac{(r-1)(c-1)}{1.3863N}$$

kde p_i je pravděpodobnost i-tého výsledku ve vstupní události ($r = 2^6 = 64$), p_j je pravděpodobnost j-tého výsledku ve výstupní události (c=2) a p_{ij} je společná pravděpodobnost pro i-tou vstupní událost a j-tou výstupní událost.

3 Výsledky

Celkem bylo analyzováno 11 409 kol s algoritmem 0, 155 758 kol s algoritmem 1 a 112 669 kol s algoritmem 2. Počet dní, které bylo každé zvíře testováno na jiný algoritmus, je vidět na obrázku 3.



Obrázek 4: Pravděpodobnost výběru pravého cíle opicemi pro různé algoritmy, převzato z [3].

3.1 Pravděpodobnost voleb a odměn

Když opice hrály proti algoritmu 0, výrazně častěji volily jeden cíl více než ten druhý, viz obrázek 4. Procenta kol ve kterých opice vybíraly pravý cíl byla 70.0%, 90.0% a 33.2% pro opice C, E a F. Ve všech případech byla odchylka od 0.5 statisticky významná (binomický test, $p < 10^{-12}$). Navzdory tendencím opic volit jeden cíl častěji, byla zvířata odměněna s 50% pravděpodobností, protože počítač vybíral cíl náhodně.

U algoritmu 1 tendence vybírat jeden cíl častěji rapidně poklesla u všech opic. Opice vybíraly pravý cíl s procentuální pravděpodobností 48.9%, 51.1% a 49.0%. Přestože jsou tyto strategie daleko blíže ekvilibriu ve smíšených strategiích (p=0.5) než u algoritmu 0, jejich odchylky od ekvilibria jsou stále statisticky významné u všech třech zvířat (binomický test, $p < 10^{-3}$). Pravděpodobnost odměny pro opice u algoritmu 1 byla o něco málo menší než 50 % (binomický test, $p < 0.05$), což odpovídá tomu, že počítač byl schopen do jisté míry předpovídat rozhodování zvířat.

Výsledky z algoritmu 2 byly podobné jako ty z algoritmu 1. Pravděpodobnost zvolení pravého cíle byla blíže k 0.5 u všech zvířat, nicméně odchylky od 0.5 byly stále statisticky významné u všech zvířat ($p < 10^{-5}$). Pravděpodobnost odměny opicím byla nižší u algoritmu 2 než u algoritmu 1 ($p < 0.001$), viz obrázek 5. Což bylo očekáváváno vzhledem k tomu, že počítač měl k dispozice historii odměn zvířete k určení jeho volby.

Algorithm	Animal	P(Right)	P(Reward)	P_{ind}(Same)	P(Same)	P(WSLS)
0	C	0.7002*	0.4969	0.5802	0.5726	0.6674*
	E	0.9017*	0.4985	0.8228	0.9808*	0.5081
	F	0.3320*	0.4892	0.5565	0.6727*	0.5718*
1	C	0.4886*	0.4894*	0.5003	0.5202*	0.6462*
	E	0.5110*	0.4911*	0.5002	0.4963	0.7314*
	F	0.4899*	0.4951*	0.5002	0.5043	0.6333*
2	C	0.4857*	0.4766*	0.5004	0.5137*	0.5478*
	E	0.4911*	0.4695*	0.5002	0.4878*	0.5345*
	F	0.4717*	0.4778*	0.5016	0.4693*	0.5650*

Obrázek 5: Pravděpodobnost různých strategií a odměn, převzato z [3].

3.2 Závislost na předchozích volbách

Pokud zvíře vybralo pravý cíl s pravděpodobností p_R, a pokud volby v po sobě jdoucích kolech byly nezávislé, pravděpodobnost zvolení stejného cíle ve dvou po sobě jdoucích kolech je určena následovně

$$p_{ind}(same) = p_R^2 + (1 - p_R)^2,$$

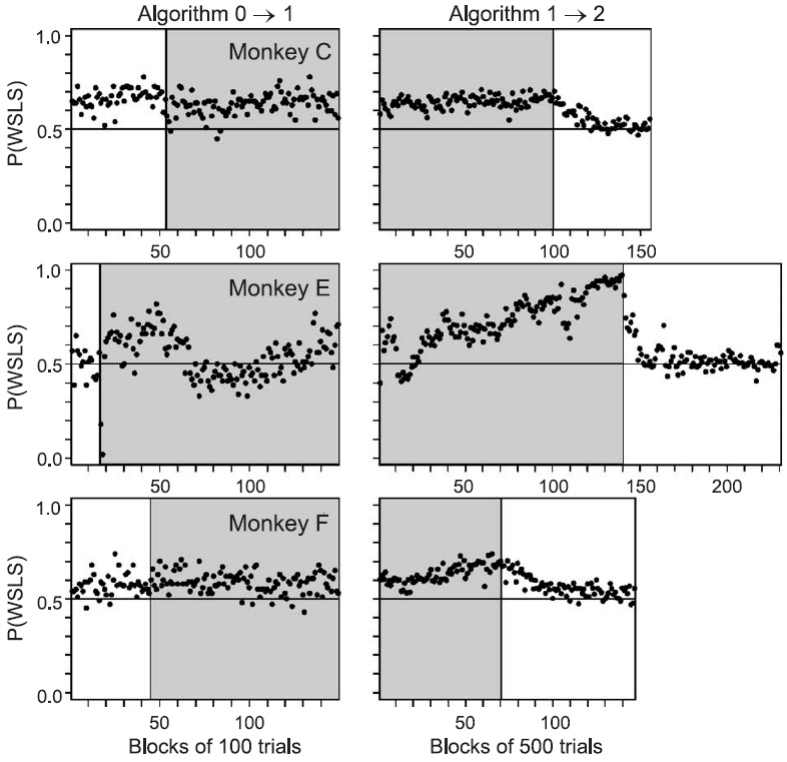
takže to zdali byla volba zvířete ovlivněna předchozí volbou více než se očekávalo z hodnoty p_R, bylo určeno na základě porovnání pravděpodobnosti vybrání stejného cíle ve dvou po sobě jdoucích tazích jako $p_{ind}(same)$. Rozdíl mezi těmito dvěma pravděpodobnostmi překročil 0.1 jen u opice E a F u algoritmu 0. Což pro tyto dva případy znamená, že zvíře opakovalo stejnou volbu více častěji, než se čekalo na základě jejich preferencí konkrétního cíle. Ve všech ostatních případech byl tento rozdíl menší než 0.05, nicméně v některých případech byl stále statisticky významný.

Při rozhodování se běžně využívá strategie "win-stay-lose-switch"(při výhře použij stejnou strategii, při prohře ji změň). Ve hře matching pennies je tato strategie ekvivalentní výběru stejného cíle jako to udělal soupeř v minulém kole. Pravděpodobnost, že zvíře vybere takovýto cíl byla výrazně větší než 0.5 ($p < 10^{-10}$) u všech zvířat a u všech algoritmů kromě algoritmu 0 u opice E. Tato skutečnost ovšem není překvapivá, jelikož opice E vybírala pravý cíl s relativně vysokou pravděpodobností. Pravděpodobnost WSLS strategie byla vysoká zejména u algoritmu 1, a to 64.6%, 73.1% a 63.3% pro opice C,E a F. I přes tyto tendence celková pravděpodobnost odměny zůstala blízká 0.5. Jinak řečeno, zvířata nebyla penalizována pro časté využívání WSLS strategie při použití algoritmu 1 počítačem, jelikož počítač nezkoumal historii odměn zvířat. Pravděpodobnost zahrání strategie WSLS u algoritmu 2 postupně klesala k 0.5 u všech zvířat, viz obrázek 6. U všech zvířat byla pravděpodobnost zahrání WSLS pro algoritmus významně menší než u algoritmu 1 (Z-test, $p < 10^{-79}$).

Pravděpodobnost hraní WSLS strategie se v průběhu použití algoritmu 1 postupně zvedala u všech zvířat. U opice C se pravděpodobnost WSLS strategie zvedla během prvních 10 000 kol algoritmu 1 z 0.633 na 0.664. U opice E se pravděpodobnost WSLS strategie u algoritmu 1 zvedla ještě více a to z 0.534 na 0.936. U opice F se pravděpodobnost WSLS strategie u algoritmu 1 zvedla z 0.588 na 0.679. Regresní analýza ukázala, že tato tendence byla statisticky významná u všech zvířat ($p < 10^{-5}$).

3.3 Analýza pomocí logistické regrese

Výsledky popsané v předchozí podkapitole 3.2 ukazují, že volba zvířete byla v konkrétním případě významně ovlivněna rozhodnutím počítače v předchozím kole podle strategie WSLS. Možnost, že volba zvířete mohla být ovlivněna předchozími volbami zvířete nebo stroje, byla otestována logistickou regresí. V této analýze indikují pozitivní koeficienty to, že zvíře pravděpodobněji vybere stejný cíl jako ono samo nebo jako počítač v některém z předchozích kol. Celkově byly koeficienty větší u algoritmu 1 než u algoritmu 2, což naznačuje, že efekty volby hráče v předchozím kole byly snížené tím, že počítač využil další informace o chování opice. Koeficienty



Obrázek 6: Pravděpodobnost WSLS strategie, převzato z [3].

u algoritmu 0 pro opici E se nechovaly dobře, což je konzistentní s faktem, že tahle opice silně preferovala pravý cíl.

Velikost koeficientů měla klesavou tendenci s herní prodlevou, tedy volby učiněné bezprostředně před současným volbou měli na rozhodování zvířete větší vliv. Tento trend byl více konzistentní pro koeficienty rozhodování počítače, což je konzistentní s použitím WSLS strategie.

3.4 Model zpětnovazebního učení

Možnost, že volba zvířete by mohla být spolehlivě předpovězena hodnotovou funkcí s dvěma cíli, byla testována modelem zpětnovazebního učení. V tomto modelu byla hodnotová funkce spojena s každým cílem a byla upravena podle, toho jaký cíl si vybralo zvíře a jestli dostalo odměnu či nikoliv. Model navíc zahrnoval diskontní faktor (), který určoval míru s jakou hodnotové funkce ztrácely významnost. Výsledky log likelihood ratio testu ukázaly, že model zpětnovazebního učení nemodeluje data tak dobře jako model logistické regrese. Nicméně rozdíl v hodnotových funkcích poskytl spolehlivou předpověď pro volbu zvířete ve všech algoritmech a to zejména když byl

rozdíl velký. Konzistentní s výše popsanými výsledky je rovněž to, že se zde objevily systematické rozdíly mezi algoritmem 1 a 2. Odhady maximální pravděpodobnosti modelu ukázaly, že velikost změn aplikovaných na hodnotovou funkci po každém kole (Δ_1 a Δ_2) byly větší u algoritmu 1 u všech zvířat než u algoritmu 2. Což je opět konzistentní s faktem, že pravděpodobnost WSLS byla větší u algoritmu 1. Navíc diskontní faktor byl menší u algoritmu 1, což naznačuje, že volba zvířete u algoritmu 1 byla převážně ovlivněná výsledkem předchozího kola. Efekt předchozích kol se u algoritmu 2 akumuloval přes několik kol, ale efekt jednotlivých kol byl menší.

3.5 Náhodnost v sekvenci voleb

Od dvou racionálních hráčů hrajících hru matching pennies se očekává, že budou hrát hru nezávisle na jejich předchozích hrách. Aby se dalo určit, jak blízce tohle chování odpovídá chování opic, bylo nutné určit náhodnost v sekvenci voleb. Ta byla určena pomocí entropie a vzájemné informace. Entropie byla počítána pro zvířecí rozhodnutí ve třech po sobě jdoucí kolech. Výsledky ukazují, že entropie byla relativně blízko k jejímu maximu a nebyl zde žádný významný rozdíl mezi algoritmy 1 a 2. Průměrná entropie byla větší než 2.95 pro všechna zvířata v obou algoritmech. Dále byla entropie počítána pro zvířecí volbu ve 3 po sobě jdoucích kolech a volbu počítače v prvních dvou z těchto tří kol. Hodnota entropie postupně klesala při použití algoritmu 1. Což je konzistentní s faktem, že pravděpodobnost WSLS strategie se v průběhu použití algoritmu 1 zvyšovala. Regrese ukázala, že tento trend byl statisticky významný pro všechna zvířata ($p < 0.001$). Průměrná hodnota entropie byla v době používání algoritmu 1 4.78, 4.44 a 4.85 pro opice C, E a F. Při používání algoritmu 2 se hodnoty zvedly na 4.89, 4.85 a 4.89. Rozdíl mezi těmito dvěma algoritmy byl statisticky významný u prvních dvou opic ($c < 10^{-7}$). Výsledky algoritmu 2 konzistentně s výsledky analýzy WSLS strategie a logistické regrese indikují, že volba zvířete se stala méně závislou na volbě počítače v předchozím kole.

4 Závěr

Studie zkoumala statistické vzorce v rozhodování opic při hře matching pennies proti počítači. Počítač používal 3 různé algoritmy s různou mírou informace o historii voleb a odměn zvířete. V algoritmu 0 počítač hrál strategii podle smíšeného Nashova ekvilibria a vybíral si mezi cíli náhodně se stejnou pravděpodobností. V tomto případě byl očekávaný výsledek nezávislý na strategii opic a není tudíž překvapivé, že se opice výrazně odchýlily od ekvilibria. Volby opic se přiblížily ekvilibriu v okamžiku, kdy začal počítač volit cíle podle algoritmu 1. To bylo opět očekávané, protože jakýkoliv systematický odklon opic od hraní ekvilibria byl počítačem využit k předpovědi volby zvířete a snížení odměny opice. Zároveň se ukázala tendence zvířat hrát takzvanou WSLS strategii a tato tendence se zvětšovala s dobou hraní algoritmu 1 počítačem. Hraní WSLS nebylo nijak penalizováno ani podporováno, jelikož algoritmus 1 nezkoumal historii odměn zvířete. Při použití algoritmu 2 ale pravděpodobnost hraní WSLS klesla blízko k 0.5. Tato strategie totiž nebyla dále efektivní proti počítači, který zkoumal jak historii voleb, tak historii odměn opic. Výsledky studie ukázaly, že opice byly schopny přizpůsobit své rozhodování strategii počítače. To dokazuje především fakt, že si po celou dobu hraní udrželi odměnu s pravděpodobností blízkou 0.5.

Reference

[1] Michele Budinich and Lance Fortnow. Repeated matching pennies with limited randomness. *Computing Research Repository - CORR*, 02 2011.

[2] Martin Hrubý. The: Nekooperativní hry v normální formě(non-cooperative normal-form games). http://www.fit.vutbr.cz/~hrubym/THE/2-normal-form-games.pdf. [Online; accessed 31-January-2020].

[3] Daeyeol Lee, Michelle L. Conroy, Benjamin P. McGreevy, and Dominic J. Barraclough. Reinforcement learning and decision making in monkeys during a competitive game. *Cognitive Brain Research*, 22(1):45 – 58, 2004.

[4] Wikipedia. Oční pohyb — Wikipedia, the free encyclopedia. http://cs.wikipedia.org/w/index.php?title=O%C4%8Dn%C3%AD%20pohyb&oldid=15761390, 2020. [Online; accessed 31-January-2020].