ŽILINSKÁ UNIVERZITA V ŽILINE
FAKULTA RIADENIA A INFORMATIKY

# DIZERTAČNÁ PRÁCA

Študijný odbor: **Aplikovaná informatika**

**Ing. Michal Chovanec**

**Aproximácia funkcie ohodnotení v algoritmoch Q-learning neurónovou sieťou**

Vedúci: **prof. Ing. Juraj Miček, PhD**

Reg.č. xxx/2008 Máj 2012

**Abstrakt**

PRIEZVISKO MENO: *Názov diplomovej práce* [Diplomová práca]

Žilinská Univerzita v Žiline, Fakulta riadenia a informatiky, Katedra matematických metód.

Vedúci: doc. RNDr. Štefan Peško, CSc.

Stupeň odbornej kvalifikácie: Inžinier v odbore .... Žilina.

FRI ŽU v Žiline, 2012 — ?? s.

Obsahom práce je...

**Abstract**

PRIEZVISKO MENO: *Name of the Diploma thesis* [Diploma thesis]
University of Žilina, Faculty of Management Science and Informatics, Department of mathematical methods.
Tutor: Assoc. Prof. RNDr. Štefan Peško, CSc.
Qualification level: Engineer in field ..... Žilina:
FRI ŽU v Žiline, 2009 — ?? p.

The main idea of this ...

## Prehlásenie

Prehlasujem, že som túto prácu napísal samostatne a že som uviedol všetky použité pramene a literatúru, z ktorých som čerpal.

V Žiline, dňa 15.5.2012

Meno Priezvisko

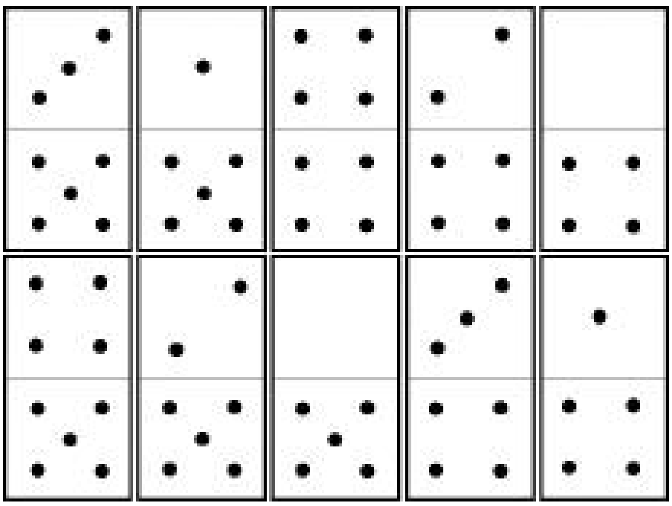
# Obsah

# Kapitola 1

# Súčasný stav problematiky

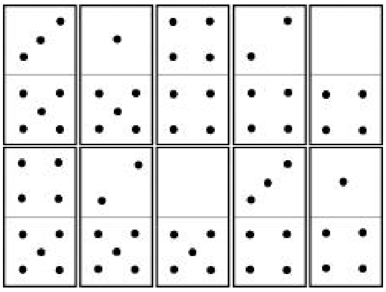
Tu je potrebné popísať doteraz získané poznatky z problematiky. Nezabudnúť dôsledne citovať autorov článkov, kníh aj internetových publikácií napr. monografia [2]. V prameňoch – spravidla posledná kapitola – treba uviesť všetku použitú literatúru. Nemala by obsahovať tie zdroje, ktoré nie sú v práci citované. A tiež nie je vhodné citovať nedôveryhodné zdroje ako sú Wikipédia ap.

Na obrázku 1.1 máme príklad zo zábavnej matematiky uvedený v práci Peško [3].

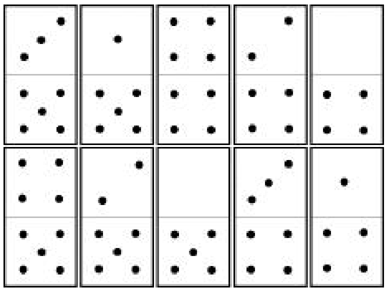


Obr. 1.1: Ako popreklápať kocky domina tak, aby rozdiel medzi súčtami horných a dolných políčok bol čo najmenší?

Alebo dva obrázky vedľa seba:



Obr. 1.2: Názov ľavého obrázku



Obr. 1.3: Názov pravého obrázku

Matematicky môžeme vzťahy označiť:

$$c_{i,j}^2 + c_{k,l} \leq \frac{c_{i,l}}{c_{k,j}} \tag{1.1}$$

a potom sa naň neskôr v texte (1.1) odvolávať. Ak sa pri reporte objavia symboly (??) treba zopakovať `pdflatex praca`.

Môžeme použiť aj iné matematické prostredia:

$$c_{i,j}^2 + c_{k,l} \leq \frac{c_{i,l}}{c_{k,j}}$$

resp.

$$c_{i,j}^2 + c_{k,l} \leq \frac{c_{i,l}}{c_{k,j}}$$

resp.

$$c_{i,j}^2 + c_{k,l} \leq \frac{c_{i,l}}{c_{k,j}}$$

resp.

$$c_{i,j}^2 + c_{k,l} \leq \frac{c_{i,l}}{c_{k,j}} \cdots \text{riadok číslo 1}$$

$$c_{i,j}^2 + c_{k,l} \leq \frac{c_{i,l}}{c_{k,j}} \cdots \text{riadok číslo 2}$$

$$c_{i,j}^2 + c_{k,l} \leq \frac{c_{i,l}}{c_{k,j}} \cdots \text{riadok číslo 3}$$

$$c_{i,j}^2 + c_{k,l} \leq \frac{c_{i,l}}{c_{k,j}} \cdots \text{riadok číslo 4}$$

resp.

$$c_{i,j}^2 + c_{k,l} \leq \frac{c_{i,l}}{c_{k,j}} \cdots \text{riadok číslo 1}$$
$$c_{i,j}^2 + c_{k,l} \leq \frac{c_{i,l}}{c_{k,j}} \cdots \text{riadok číslo 2}$$
$$c_{i,j}^2 + c_{k,l} \leq \frac{c_{i,l}}{c_{k,j}} \cdots \text{riadok číslo 3}$$
$$c_{i,j}^2 + c_{k,l} \leq \frac{c_{i,l}}{c_{k,j}} \cdots \text{riadok číslo 4} \quad (1.2)$$

resp.

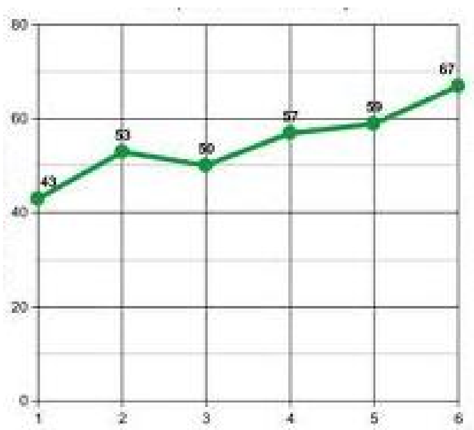
$$\begin{aligned} c_{i,j}^2 + c_{k,l} &\leq \frac{c_{i,l}}{c_{k,j}} \cdots \text{riadok číslo 1} \\ c_{i,j}^2 + c_{k,l} &\leq \frac{c_{i,l}}{c_{k,j}} \cdots \text{riadok číslo 2} \\ c_{i,j}^2 + c_{k,l} &\leq \frac{c_{i,l}}{c_{k,j}} \cdots \text{riadok číslo 3} \\ c_{i,j}^2 + c_{k,l} \leq &\; xxx \quad \frac{c_{i,l}}{c_{k,j}} \cdots \text{riadok číslo 4} \end{aligned}$$

resp.

$$c_{i,j}^2 + c_{k,l} \leq \frac{c_{i,l}}{c_{k,j}} \cdots \text{riadok číslo 1} \quad (1.3)$$
$$c_{i,j}^2 + c_{k,l} \leq \frac{c_{i,l}}{c_{k,j}} \cdots \text{riadok číslo 2} \quad (1.4)$$
$$c_{i,j}^2 + c_{k,l} \leq \frac{c_{i,l}}{c_{k,j}} \cdots \text{riadok číslo 3}$$
$$c_{i,j}^2 + c_{k,l} \quad xxx \quad \frac{c_{i,l}}{c_{k,j}} \cdots \text{riadok číslo 4} \quad (1.5)$$

Niekedy sa hodí pracovať s maticami alebo determinantami:

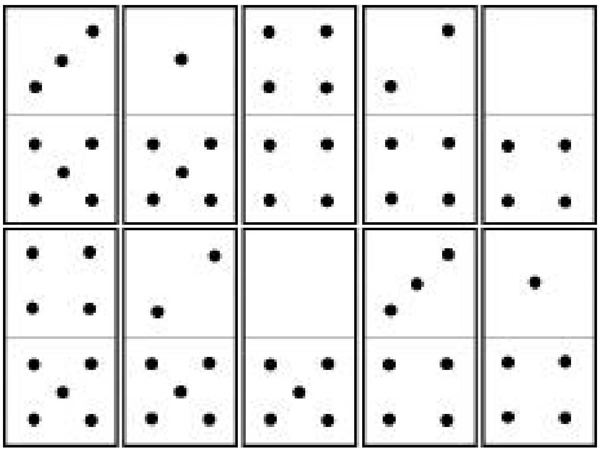
$$\mathbb{A} = \begin{pmatrix} -5 & 8 & 3 \\ 0 & -0 & 2 \\ 5 & 1 & -2 \end{pmatrix} \times \left[\begin{matrix} -5 & 8 & 3 \\ 0 & -0 & 2 \\ 5 & 1 & -2 \end{matrix}\right\} + \begin{vmatrix} -5 & 8 & 3 \\ 0 & -0 & 2 \\ 5 & 1 & -2 \end{vmatrix} + \begin{Vmatrix} -5 & 8,1 & 3,4 \\ x & -0,2 & 2 \\ 5 & 1 & -2 \end{Vmatrix}$$




Obr. 1.4: Obrázok

Tu je potrebné popísať doteraz získané poznatky z problematiky. Nezabudnúť dôsledne citovať autorov článkov, kníh aj internetových publikácií napr. monografia [3, 2].

Na obrázku 1.5 máme príklad zo zábavnej matematiky uvedený v práci Peško [3]. Porovnajte definíciu, zobrazenie a tiež umiestnenie obrázku 1.5 s obrázkom 1.1.



Obr. 1.5: Ako popreklápať kocky domina tak, aby rozdiel medzi súčtami horných a dolných políčok bol čo najmenší?

# Kapitola 2

# Q-larning algoritmus

Q-learning algoritmus je definovaný pre časovo diskrétne systémy. Agent ktorý prechádza stavový priestor vykonaním niektorej z vopred daných akcií získava za tieto prechody odmeny. Cieľom algoritmu je ohodnotiť všetky akcie v jednotlivých stavoch, tak aby bol dosiahnutý ustálený stav a v každom stave bolo možno vybrať akciu prinášajúcu najväčšiu odmenu, v globálnom zmysle.

## 2.1 Definícia algoritmu

Daná je množina stavov $\mathbb{S}$ a akcií $\mathbb{A}$, kde $\mathbb{S} \in \mathbb{R}^{n_s}$ a $\mathbb{A} \in \mathbb{R}^{n_a}$, kde $n_s$ a $n_a$ sú rozmery stavového vektora a vektora akcií.

Existuje prechodová funkcia

$$s(n+1) = \lambda(s(n), a(n)) \tag{2.1}$$

zo stavu $s(n) \in \mathbb{S}$ použitím akcie $a(n) \in \mathbb{A}$, táto funkcia je ale algoritmu neznáma.

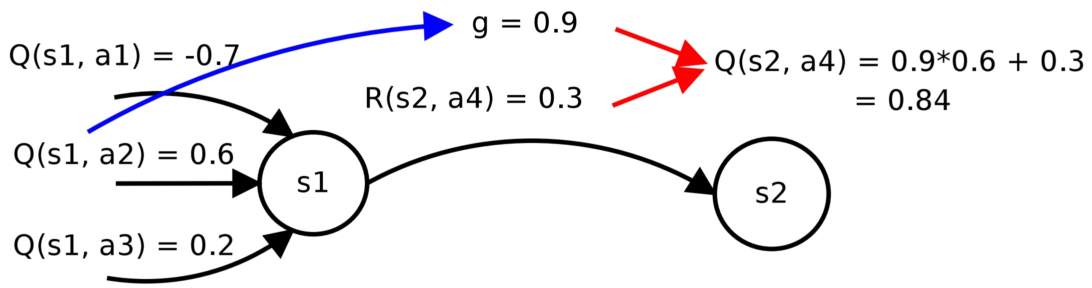
Ďalej je daná odmeňovacia funkcia $R(s(n), a(n))$, ktorá vyjadruje okamžité ohodnotenie konania agenta v $s(n)$ a $a(n)$. V reálnych aplikáciach táto funkcia nadobúda takmer v každom $s(n)$ a $a(n)$ hodnotu 0. Pre správnu funkciu algoritmu, musí byť aspoň jedna hodnota nenulová - napr. ohodnotenie dosiahnutia cieľového stavu (samotná existencia cieľového stavu však pre algoritmus nie je potrebná).

Funkcia ohdnotení je definovaná ako

$$Q(s(n),a(n)) = R(s(n),a(n)) + \gamma \max_{a(n-1)\in\mathbb{A}} Q(s(n-1),a(n-1)) \tag{2.2}$$

- $R(s(n),a(n))$ je odmeňovacia funkcia
- $Q(s(n-1),a(n-1))$ je odmeňovacia funkcia v stave $s(n-1)$ pre akciu $a(n-1)$
- $\gamma$ je odmeňovacia konštanta a platí $\gamma \in (0,1)$.

Funkcia 2.2 definuje ohodnotenie akcií vo všetkých stavoch : agent ktorý sa dostal do stavu $s(n)$ vykonaním akcie $a(n)$ zo stavu $s(n-1)$ získal odmenu $R(s(n),a(n))$ a zlomok najväčšieho možného ohodnotenia ktoré mohol získať dostaním sa do stavu $s(n-1)$, situáciu ilustruje obrázok 2.1.



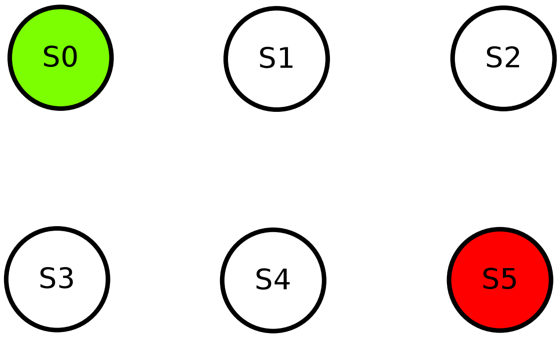

Obr. 2.1: Ilustrácia funkcie ohodnotení, pre $\gamma = 0.9$

Nasledujúce obrázky ilustrujú beh algoritmu pre systém so 6 stavmi. Na začiatku nie sú známe ani samotné prechody medzi stavmi 2.2, bol definovaný 1 cieľový stav $S5$, agent začína v stave $S0$ (môže však v ľubovolnom inom). Ďalej sa pre jednoduchosť predpokladá že

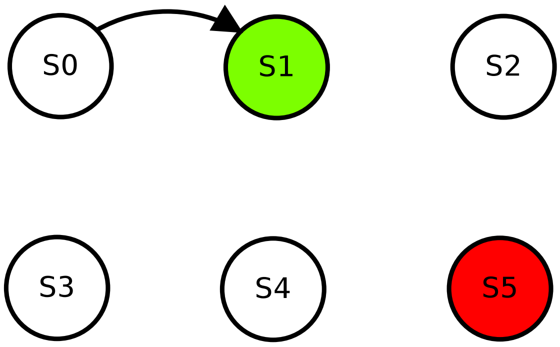
$$R(s(n),a(a)) = \begin{cases} 1 & s(n) = \text{S5} \\ 0 & inak \end{cases} \tag{2.3}$$

t.j. odmeňovacia funkcia nadobúda hodnotu 1 len ak sa agent dostal do stavu $S5$.

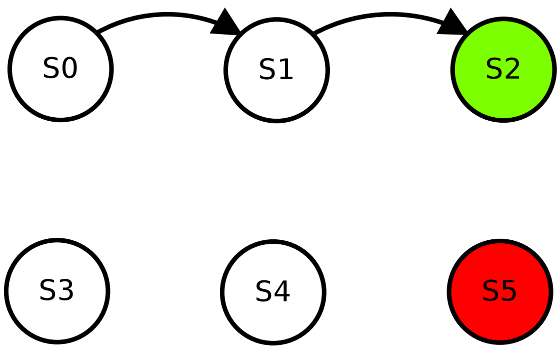
Agent v každom stave náhodne vyberá akcie (na výbere nezáleží, dôležité je aby každá akcia mala nenulovú pravdepodobnoť výberu, a rovnako bola nenulová prevdepodobnosť dosiahnutia ľubovolného stavu). Obrázky 2.3 a 2.4 ilustrujú jednu z možných ciest.



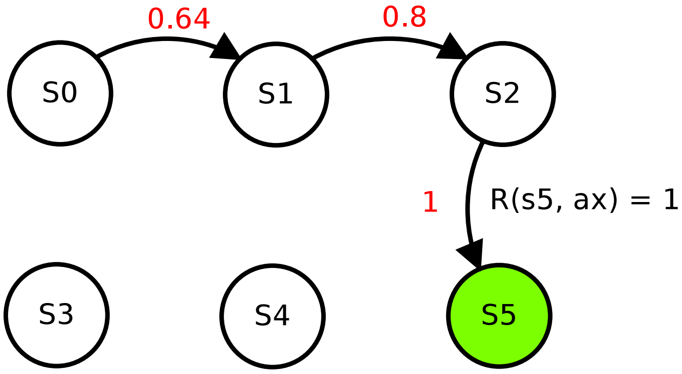

Obr. 2.2: Inicializácia



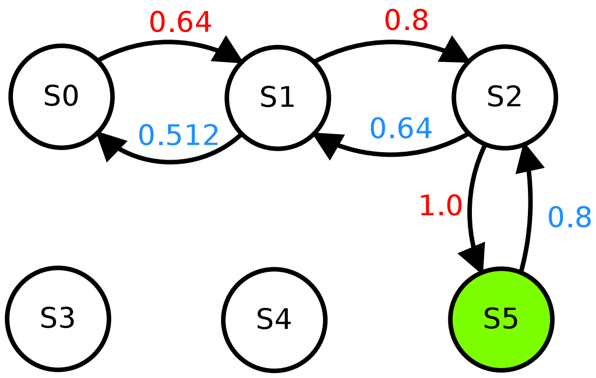

Obr. 2.3: Prechod do stavu S1



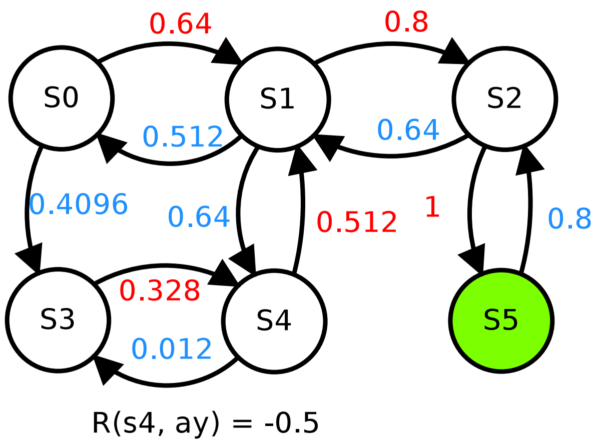

Obr. 2.4: Prechod do stavu S2




Obr. 2.5: Prechod do stavu S3




Obr. 2.6: Inicializácia




Obr. 2.7: Inicializácia

# Literatúra


[1] Bartsch H. J., *Matematické vzorce*, 3. revidované vydání, Praha, Mladá fronta 2000, ISBN 80-204-0607-7.

[2] Berman G. N., *Zbierka úloh z matematickej analýzy*, Bratislava, ŠNTL 1955.

[3] Peško, Š., *Operační systémi*, Knižnice výpočetní techniky, Nakladatelství technické literatury (1992), SNTL, ISBN 80-03-00269-9.

[4] World of mathematics, A Wolfram Web Resource, `http://mathworld.wolfram.com/`, WolframAlpha – computational knowledge engine, `http://www.wolframalpha.com/`.