# Spectral bandits for smooth graph functions

Tomáš Kocák

14. november 2018

# Multi-armed bandit problém


**Bandit 1**

Reward $r_1$

$\mu_1 = \mathbb{E}[r_1]$


**Bandit 2**

Reward $r_2$

$\mu_2 = \mathbb{E}[r_2]$

. . . . . .


**Bandit $N$**

Reward $r_N$

$\mu_N = \mathbb{E}[r_N]$

**Sekvenčný problém** - v každom kole sa opakuje:

- **Vyber akciu** - jeden z $N$ hracích automatov
- **Obdrž odmenu** - náhodná premenná s fixnou strednou hodnotou

**Cieľ: maximalizovať celkový zisk** (po $T$ kolách)

# Multi-armed bandit problém - aplikácie

- **Klinické štúdie**
  - Podať pacientovi jeden z $N$ liekov na chorobu.
  - Pozorovať výsledok liečby (odmena)
  - **Cieľ:** maximalizovať množstvo vyliečených pacientov
- **Odporúčacie systémy**
  - Odporučiť užívateľovi film, pesničku, ...
  - Obdržať hodnotenie filmu, pesničky, ...
  - **Cieľ:** maximalizovať spokojnosť užívateľa
- **Cielená reklama**
  - Zobraziť užívateľovi jednu z $N$ reklám
  - Obdržať spätnú väzbu od užívateľa (klik)
  - **Cieľ:** maximalizovať množstvo vyliečených pacientov
- **Rybárčenie**
  - Na začiatku dňa si vybrať jedno z $N$ miest
  - Na konci dňa pozorovať množstvo ulovených rýb
  - **Cieľ:** maximalizovať množstvo rýb po $T$ dňoch

# Multi-armed bandit problém

## Špecifikácia problému

- $T$ – počet opakovaní
- $N$ – počet akcií
- $r_{t,i}$ – odmena akcie $i$ v čase $t$

## V každom kole:

- Algoritmus zvolí akciu $I_t$
- Algoritmus obdrží odmenu $r_{t,I_t}$
- Algoritmus si upraví preferencie

**Cieľ** po $T$ kolách:

- **Maximalizovať celkovú odmenu**
- **Minimalizovať celkový očakávaný regret**

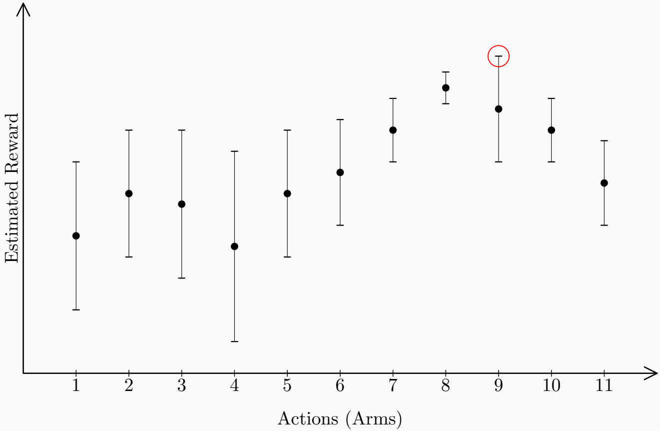
### Kumulatívny regret

$$R_T = \max_{j \in [N]} \mathbb{E}\left[\sum_{t=1}^{T} r_{t,j}\right] - \mathbb{E}\left[\sum_{t=1}^{T} r_{t,I_t}\right]$$

# Upper confidence bound algoritmus

**UCB algoritmus** - jeden z najznámejších algorithmov pre tento problém

- Zvoľme akciu s najvyššou hornou hranica intervalu spoľahlivosti
- Obdržme odmenu prislúchajúcu k zvolenej akcii
- Upravme interval spoľahlivosti



# Čo očakávame od regretu?

**Kumulatívny regret**

$$R_T = \max_{j \in [N]} \mathbb{E}\left[\sum_{t=1}^{T} r_{t,j}\right] - \mathbb{E}\left[\sum_{t=1}^{T} r_{t,I_t}\right]$$

Náhodný algoritmus:

$$R_T = \mathcal{O}(T)$$

Vševediaci algoritmus:

$$R_T = \mathcal{O}(0)$$

**UCB algoritmus:**

Horný odhad na regret:

$$R_T = \widetilde{\mathcal{O}}(\sqrt{NT})$$

Existujú ťažké problémy:

$$R_T = \Omega(\sqrt{NT})$$

# Odhady na regret

Horný odhad na regret:

$$R_T = \widetilde{\mathcal{O}}(\sqrt{NT})$$

Existujú ťažké problémy:

$$R_T = \Omega(\sqrt{NT})$$

## Dá sa odhad na regret vylepšiť?

- Vo všeobecnosti: **NIE!**
- S dodatočnými predpokladmi: **ÁNO!**

**Očakávanie na odhad na regret** (s využitím štruktúry problému):

$$R_T = \widetilde{\mathcal{O}}(\sqrt{DT})$$

- $D$ zachytáva štruktúru
- $D$ je menšie ako $N$

**Zvyčajný prístum:** kontextuálni banditi

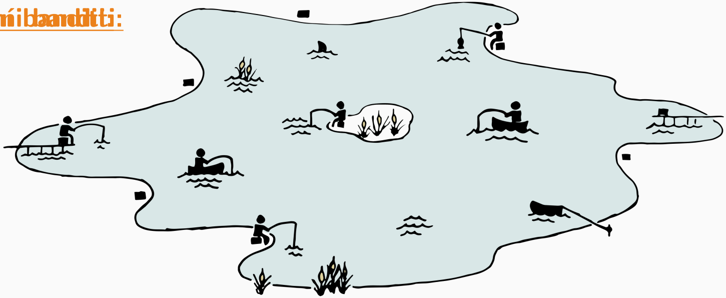
# Banditi s podobnosťami

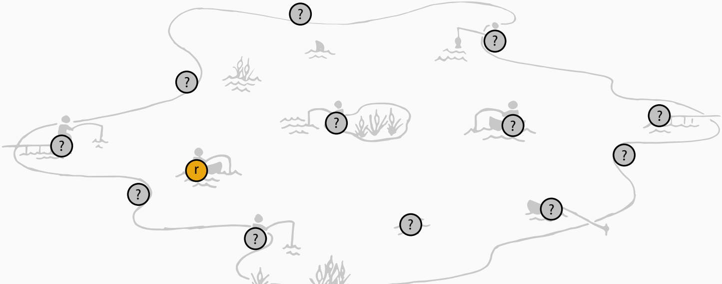
## Spectrálni banditi

# Rybársky príklad

**Základní bandit:**

**Spektrální bandit:**





# Hladkosť funkcie na grafe

**Hladkosť funkcie $\boldsymbol{r}$** (vzhľadom na graf $G$)

$$S_G(\boldsymbol{r}) = \frac{1}{2} \sum_{i,j \in [N]} s_{i,j}(r_i - r_j)^2 = \boldsymbol{r}^\top \mathcal{L} \boldsymbol{r}$$

- $r_i$ – odmena akcie $i$
- $\boldsymbol{r} = [r_1, r_2, \ldots, r_N]^\top$ – vektor odmien (funkcia na grafe)
- $s_{i,j}$ – váha na hrane spájajúcej akcie $i$ a $j$
- $\mathcal{L}$ – Laplaceova matica grafu $G$
- $S_G(\boldsymbol{r})$ – **hladkosť funkcie $\boldsymbol{r}$ na grafe** $G$

Funkcia $\boldsymbol{r}$ je hladká $\Leftrightarrow$ $S_G(\boldsymbol{r})$ je blízke nule

# Vlastnosti hladkých funkcií

$$S_G(\boldsymbol{r}) = \boldsymbol{r}^\top \mathcal{L}\boldsymbol{r} = \boldsymbol{r}^\top \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^\top\boldsymbol{r} = \boldsymbol{\alpha}^\top\boldsymbol{\Lambda}\boldsymbol{\alpha} = \|\boldsymbol{\alpha}\|_{\boldsymbol{\Lambda}} = \sum_{i=1}^{N} \lambda_i \alpha_i^2$$

- $\mathcal{L}$

**Čo znamená "veľké" $\lambda_i$?**

**Koľko koeficientov $\alpha_i$ je podstatných?**

$$\boldsymbol{r} = \begin{bmatrix} r_2 \\ \vdots \\ r_N \end{bmatrix} = \begin{bmatrix} | \\ q_1 \\ | \end{bmatrix}\alpha_1 + \begin{bmatrix} | \\ q_2 \\ | \end{bmatrix}\alpha_2 + \cdots + \begin{bmatrix} | \\ q_N \\ | \end{bmatrix}\alpha_N = \boldsymbol{Q}\boldsymbol{\alpha}$$

Veľké $\lambda_i$ $\Rightarrow$ malé $\alpha_i$

# Effective dimension

**Effective dimension**

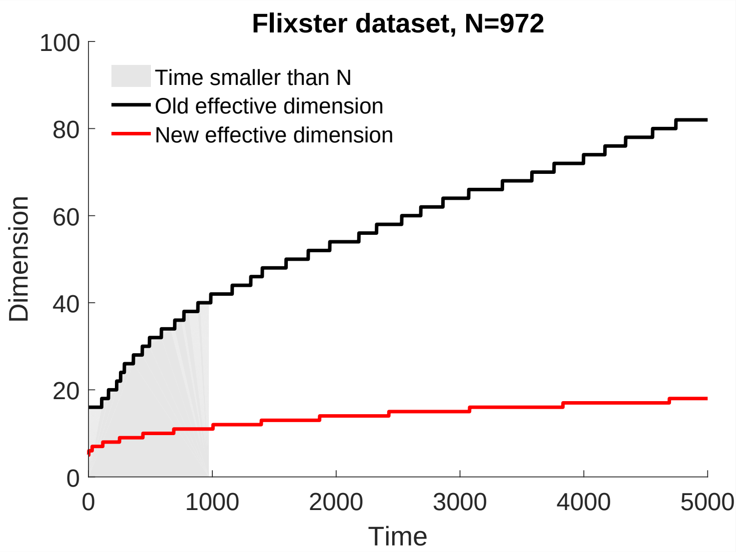
$$d = \left\lceil \frac{\max \log \prod_{i=1}^{N} \left(1 + \frac{t_i}{\lambda_i}\right)}{\log \left(1 + \frac{T}{K\lambda}\right)} \right\rceil$$

- $\{t_1, \ldots, t_N\}$ sú nezáporné celé čísla také, že $\sum_{i=1}^{N} t_i = T$.
- $\lambda$ – regularizácia
- $K$ – počet komponentov grafu $G$

**Vlastnosti effective dimension:**

- Zvyčajne **oveľa menšia ako** $N$
- **Závislá na čase:** veľké $T \Rightarrow$ viac dimenzií je podstatných
-

# Effective dimension – Flixster dataset



# Algoritmus pre spektrálnych banditov

## SpectralUCB algoritmus:

**Inicializácia:**

$\hat{\alpha}_1 = 0_N$ – počiatočný odhad na $\boldsymbol{\alpha}$

$\boldsymbol{V}_1 = \Lambda_{\mathcal{L}} + \lambda \mathbf{I}$

**V každom kole:**

$\hat{r}_i = \mathbf{x}_i^\mathsf{T} \hat{\alpha}$ – odhad odmeny $r_i$ akcie $i$

$I_t = \arg\max_a \left( \mathbf{x}_a^\mathsf{T} \hat{\alpha} + c \|\mathbf{x}_a\|_{\boldsymbol{V}_t^{-1}} \right)$ – akcia ktorú vyberieme

$\boldsymbol{V}_{t+1} = \boldsymbol{V}_t + \mathbf{x}_{I_t} \mathbf{x}_{I_t}^\mathsf{T}$

$\hat{\alpha}_{t+1} = \boldsymbol{V}_{t+1}^{-1} \sum_{s=1}^{t} \mathbf{x}_{I_t} r_s$ – odhad $\boldsymbol{\alpha}$ (lineárna regresia)

**Algoritmus podporuje skúmanie „správnych“ dimenzií pomocou vhodnej regularizácie**

# Banditové algoritmy

**Základní banditi:**

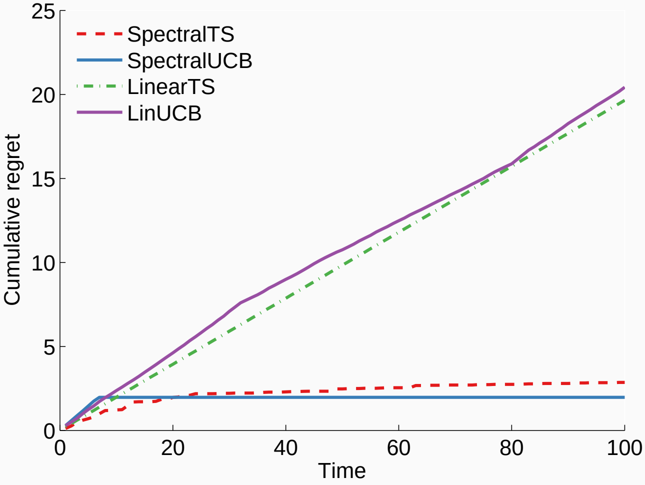
| **Názov algoritmu** | **Odhad na regret** | **Publikácia** |
|---|---|---|
| LinUCB | $R_T = \widetilde{\mathcal{O}}(N\sqrt{T})$ | Li et al., 2010 |
| LinearTS | $R_T = \widetilde{\mathcal{O}}(N\sqrt{T})$ | Agrawal and Goyal, 2013 |
| SupLinRel | $R_T = \widetilde{\mathcal{O}}(\sqrt{NT})$ | Auer, 2002 |

**Spektrálni banditi:**

| **Názov algoritmu** | **Odhad na regret** | **Publikácia** |
|---|---|---|
| SpectralUCB | $R_T = \widetilde{\mathcal{O}}(d\sqrt{T})$ | Valko et al., 2014 |
| SpectralTS | $R_T = \widetilde{\mathcal{O}}(d\sqrt{T})$ | Kocak et al., 2014 |
| SpectralEliminator | $R_T = \widetilde{\mathcal{O}}(\sqrt{dT})$ | Valko et al., 2014 |

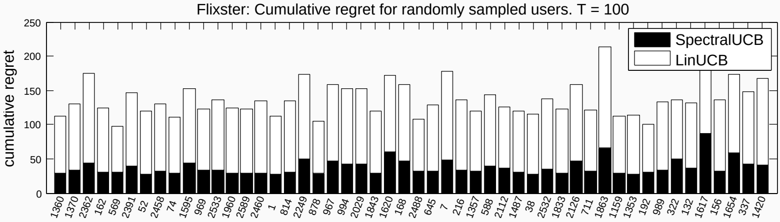
# Empirické výsledky algoritmov

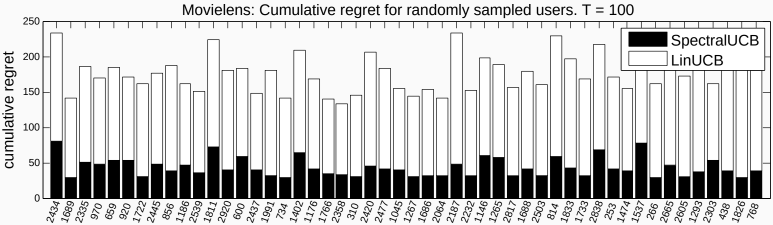
**Flixster** dataset: $N = 972$, $d < 10$




**Note:** SupLinRel and SpectralEliminator are not practical.

# Empirické výsledky algoritmov



# Zhrnutie

## Spectral bandits

- **Nový setting** pre problémy s hladkými funkciami na grafe
- **Effective dimension** popisujúca komplexitu problému
- **Algoritmy** využívyjúce hladkosť funkcie
  - Lepšie teoretické výsledky
  - Lepšie empirické výsledky

## References

**Tomáš Kocák**, Michal Valko, Rémi Munos, Shipra Agrawal: *Spectral Thompson Sampling*, The 28th AAAI Conference on Artificial Intelligence (AAAI 2014)

Michal Valko, Rémi Munos, Branislav Kveton, **Tomáš Kocák**: *Spectral Bandits for Smooth Graph Functions*, The 31th International Conference on Machine Learning (ICML 2014)

**Tomáš Kocák**, Michal Valko, Rémi Munos, Branislav Kveton, Shipra Agrawal: *Spectral Bandits*, to appear in Journal of Machine Learning Research (JMLR 2018)

**Ďakujem za pozornosť!**

**Čas na otázky...**