

Analýza nákupního košíku - historie a současnost

Radim Navrátil

Ústav matematiky a statistiky
Přírodovědecká fakulta MU, Brno

Robust
24. ledna 2018

O čem nebudu povídat

Penalizované a "shrinkage" R-odhady v modelu lineární regrese (společně s prof. Salehem, Carleton University, Ottawa).

- Zkoumání vychýlenosti a střední čtvercové chyby odhadů, je-li vektor parametrů řídký.
- Vytyčení oboru dominance pro jednotlivé odhady.
- Výsledky analogické těm pro odhady odvozené od LSE.

O čem nebudu povídat

Penalizované a "shrinkage" R-odhady v modelu lineární regrese (společně s prof. Salehem, Carleton University, Ottawa).

- Zkoumání vychýlenosti a střední čtvercové chyby odhadů, je-li vektor parametrů řídký.
- Vytyčení oboru dominance pro jednotlivé odhady.
- Výsledky analogické těm pro odhady odvozené od LSE.

O čem nebudu povídat

Penalizované a "shrinkage" R-odhady v modelu lineární regrese (společně s prof. Salehem, Carleton University, Ottawa).

- Zkoumání vychýlenosti a střední čtvercové chyby odhadů, je-li vektor parametrů řídký.
- Vytyčení oboru dominance pro jednotlivé odhady.
- Výsledky analogické těm pro odhady odvozené od LSE.

O čem nebudu povídat

Penalizované a "shrinkage" R-odhady v modelu lineární regrese (společně s prof. Salehem, Carleton University, Ottawa).

$$\begin{aligned} (i) \text{ ADRRE (RE:URE)} &= \left(1 + \frac{p_2}{p_1}\right) \left(1 + \frac{\Delta^2}{p_1}\right)^{-1} \\ (ii) \text{ ADRRE (LASSO:URE)} &= \left(1 + \frac{p_2}{p_1}\right) \left(1 + \frac{\Delta^2}{p_1}\right)^{-1} \\ (iii) \text{ ADRRE (RR:URE)} &= \left(1 + \frac{p_2}{p_1}\right) \left(1 + \frac{p_2 \Delta^2}{p_1(p_2 + \Delta^2)}\right)^{-1} \\ (iv) \text{ ADRRE (PTE:URE)} &= \left(1 + \frac{p_2}{p_1}\right) \left\{1 + \frac{p_2}{p_1} (1 - H_{p_2+2}(c_\alpha; \Delta^2)) \right. \\ &\quad \left. + \frac{\Delta^2}{p_1} (2H_{p_2+2}(c_\alpha; \Delta^2) - H_{p_2+4}(c_\alpha; \Delta^2))\right\}^{-1} \\ (v) \text{ ADRRE (JSE:URE)} &= \left(1 + \frac{p_2}{p_1}\right) \left\{1 + \frac{p_2}{p_1} - \frac{1}{p_1} (p_2 - 2) E[\chi_{p_2}^{-2}(\Delta^2)]\right\}^{-1} \\ (vi) \text{ ADRRE (PRSE:URE)} &= \left(1 + \frac{p_2}{p_1}\right) \left\{1 + \frac{p_2}{p_1} - \frac{1}{p_1} (p_2 - 2) E[\chi_{p_2}^{-2}(\Delta^2)] \right. \\ &\quad \left. - \frac{p_2}{p_1} E\left[(1 - (p_2 - 2)\chi_{p_2+2}^{-2}(\Delta^2))^2 I(\chi_{p_2+2}^2(\Delta^2) < (p_2 - 2))\right] \right. \\ &\quad \left. + \frac{\Delta^2}{p_1} \left[2E\left[(1 - (p_2 - 2)\chi_{p_2+2}^{-2}(\Delta^2)) I(\chi_{p_2+2}^2(\Delta^2) < (p_2 - 2))\right] \right. \right. \\ &\quad \left. \left. - E\left[(1 - (p_2 - 2)\chi_{p_2+4}^{-2}(\Delta^2))^2 I(\chi_{p_2+4}^2(\Delta^2) < (p_2 - 2))\right]\right]\right\}^{-1} \end{aligned}$$

O čem budu povídat

- 1 Úvod a motivace
- 2 Analýza nákupního košíku - základní pojmy
- 3 Zobecnění na složitější pravidla
- 4 Další rozšíření
- 5 Příklad
- 6 Výhled do budoucnosti

Asociační analýza

- Metoda pro objevování častých vzorů (vztahů) ukrytých ve velkých databázích.
- Tyto vzory jsou vyjadřovány pomocí asociačních pravidel, které popisují společně se vyskytující atributy.
- Rakesh Agrawal et al. (1993) poprvé uvedli asociační pravidla pro objevování zákonitostí mezi produkty v transakčních datech z pokladních systémů supermarketů.
- Analýza nákupního košíku (market basket analysis).
- Další metody: Web mining, Intrusion Detection System (odhalování podezřelé aktivity), Continuous production.

Jsou všechna pravidla zajímavá?

Asociační pravidla by měla být:

- **snadno pochopitelná:** jakmile je nějaký vztah nalezen, lze ho snadno ověřit,
- **použitelná:** obsahují užitečné informace, které mohou vést k dalším intervencím.

Asociační pravidla by neměla být:

- **triviální:** výsledky už stejně každý zná,
- **nevysvětlitelná:** neexistuje k nim žádné vysvětlení a nevedou k žádné akci.

Jsou všechna pravidla zajímavá?

Asociační pravidla by měla být:

- **snadno pochopitelná:** jakmile je nějaký vztah nalezen, lze ho snadno ověřit,
- **použitelná:** obsahují užitečné informace, které mohou vést k dalším intervencím.

Asociační pravidla by neměla být:

- **triviální:** výsledky už stejně každý zná,
- **nevysvětlitelná:** neexistuje k nim žádné vysvětlení a nevedou k žádné akci.

Matematické pozadí

Data (transakční):

ID	Položky transakce
1	x_1, x_3, x_4
2	x_1
3	x_2, x_3, x_4, x_5
4	x_3, x_4
.	.
.	.
n	x_2, x_3, x_4

Množina všech položek $I = \{x_1, \dots, x_m\}$.

Cíl: Vytvořit pravidla typu $x_i \Rightarrow x_j$.

Příklad - transakční data

ID	Transakce
1	pivo, pečivo
2	pečivo, voda
3	pečivo, máslo, voda
4	pečivo, máslo, zelenina
5	pivo
6	máslo, voda
7	pivo, pečivo, máslo, voda
8	pivo, voda
9	pečivo, máslo, zelenina
10	voda, zelenina

Příklad - transakční data zobrazená pomocí datové matice

ID	máslo	pivo	pečivo	voda	zelenina
1	0	1	1	0	0
2	0	0	1	1	0
3	1	0	1	1	0
4	1	0	1	0	1
5	0	1	0	0	0
6	1	0	0	1	0
7	1	1	1	1	0
8	0	1	0	1	0
9	1	0	1	0	1
10	0	0	0	1	1

Matematické pozadí

Podpora (support):

$$\text{supp}(x_i \Rightarrow x_j) = \frac{\text{Počet transakcí obsahující současně } x_i \text{ a } x_j}{\text{Celkový počet transakcí}},$$

$\text{supp}(x_i \Rightarrow x_j)$ = Relativní četnost transakcí obsahující současně x_i a x_j ,

$\text{supp}(x_i \Rightarrow x_j) = \hat{P}$ (v dané transakci se vyskytuje současně x_i a x_j).

Zřejmě:

$$\text{supp}(x_i \Rightarrow x_j) = \text{supp}(x_j \Rightarrow x_i).$$

Matematické pozadí

Podpora (support):

$$\text{supp}(x_i \Rightarrow x_j) = \frac{\text{Počet transakcí obsahující současně } x_i \text{ a } x_j}{\text{Celkový počet transakcí}},$$

$\text{supp}(x_i \Rightarrow x_j)$ = Relativní četnost transakcí obsahující současně x_i a x_j ,

$\text{supp}(x_i \Rightarrow x_j) = \hat{P}$ (v dané transakci se vyskytuje současně x_i a x_j).

Zřejmě:

$$\text{supp}(x_i \Rightarrow x_j) = \text{supp}(x_j \Rightarrow x_i).$$

Matematické pozadí

Podpora (support):

$$\text{supp}(x_i \Rightarrow x_j) = \frac{\text{Počet transakcí obsahující současně } x_i \text{ a } x_j}{\text{Celkový počet transakcí}},$$

$\text{supp}(x_i \Rightarrow x_j)$ = Relativní četnost transakcí obsahující současně x_i a x_j ,

$\text{supp}(x_i \Rightarrow x_j) = \hat{P}$ (v dané transakci se vyskytuje současně x_i a x_j).

Zřejmě:

$$\text{supp}(x_i \Rightarrow x_j) = \text{supp}(x_j \Rightarrow x_i).$$

Matematické pozadí

Podpora (support):

$$\text{supp}(x_i \Rightarrow x_j) = \frac{\text{Počet transakcí obsahující současně } x_i \text{ a } x_j}{\text{Celkový počet transakcí}},$$

$\text{supp}(x_i \Rightarrow x_j)$ = Relativní četnost transakcí obsahující současně x_i a x_j ,

$\text{supp}(x_i \Rightarrow x_j) = \hat{P}$ (v dané transakci se vyskytuje současně x_i a x_j).

Zřejmě:

$$\text{supp}(x_i \Rightarrow x_j) = \text{supp}(x_j \Rightarrow x_i).$$

Matematické pozadí

Spolehlivost (confidence):

$$\text{conf}(x_i \Rightarrow x_j) = \frac{\text{Počet transakcí obsahující současně } x_i \text{ a } x_j}{\text{Počet transakcí obsahující } x_i},$$

$\text{conf}(x_i \Rightarrow x_j)$ = Relativní četnost transakcí obsahující současně x_i a x_j , mezi těmi, které už x_i obsahují.

$\text{conf}(x_i \Rightarrow x_j)$ = \hat{P} (v dané transakci se vyskytuje současně x_i a x_j | v dané transakci se vyskytuje x_i).

Matematické pozadí

Spolehlivost (confidence):

$$\text{conf}(x_i \Rightarrow x_j) = \frac{\text{Počet transakcí obsahující současně } x_i \text{ a } x_j}{\text{Počet transakcí obsahující } x_i},$$

$\text{conf}(x_i \Rightarrow x_j)$ = Relativní četnost transakcí obsahující současně x_i a x_j , mezi těmi, které už x_i obsahují.

$\text{conf}(x_i \Rightarrow x_j)$ = \hat{P} (v dané transakci se vyskytuje současně x_i a x_j | v dané transakci se vyskytuje x_i).

Matematické pozadí

Spolehlivost (confidence):

$$\text{conf}(x_i \Rightarrow x_j) = \frac{\text{Počet transakcí obsahující současně } x_i \text{ a } x_j}{\text{Počet transakcí obsahující } x_i},$$

$\text{conf}(x_i \Rightarrow x_j)$ = Relativní četnost transakcí obsahující současně x_i a x_j , mezi těmi, které už x_i obsahují.

$\text{conf}(x_i \Rightarrow x_j)$ = \hat{P} (v dané transakci se vyskytuje současně x_i a x_j | v dané transakci se vyskytuje x_i).

Matematické pozadí

Lift (zlepšení):

$$\text{lift}(x_i \Rightarrow x_j) = \frac{\text{Počet transakcí obsahující současně } x_i \text{ a } x_j \cdot n}{\text{Počet transakcí obsahující } x_i \cdot \text{Počet transakcí obsahující } x_j},$$

$$\text{lift}(x_i \Rightarrow x_j) = \frac{\widehat{P}(\text{v dané transakci se vyskytuje současně } x_i \text{ a } x_j)}{\widehat{P}(\text{v dané transakci je } x_i) \widehat{P}(\text{v dané transakci je } x_j)}.$$

Zřejmě:

$$\text{lift}(x_i \Rightarrow x_j) = \text{lift}(x_j \Rightarrow x_i).$$

Matematické pozadí

Lift (zlepšení):

$$\text{lift}(x_i \Rightarrow x_j) = \frac{\text{Počet transakcí obsahující současně } x_i \text{ a } x_j \cdot n}{\text{Počet transakcí obsahující } x_i \cdot \text{Počet transakcí obsahující } x_j},$$

$$\text{lift}(x_i \Rightarrow x_j) = \frac{\widehat{P}(\text{v dané transakci se vyskytuje současně } x_i \text{ a } x_j)}{\widehat{P}(\text{v dané transakci je } x_i) \widehat{P}(\text{v dané transakci je } x_j)}.$$

Zřejmě:

$$\text{lift}(x_i \Rightarrow x_j) = \text{lift}(x_j \Rightarrow x_i).$$

Matematické pozadí

Lift (zlepšení):

$$\text{lift}(x_i \Rightarrow x_j) = \frac{\text{Počet transakcí obsahující současně } x_i \text{ a } x_j \cdot n}{\text{Počet transakcí obsahující } x_i \cdot \text{Počet transakcí obsahující } x_j},$$

$$\text{lift}(x_i \Rightarrow x_j) = \frac{\widehat{P}(\text{v dané transakci se vyskytuje současně } x_i \text{ a } x_j)}{\widehat{P}(\text{v dané transakci je } x_i) \widehat{P}(\text{v dané transakci je } x_j)}.$$

Zřejmě:

$$\text{lift}(x_i \Rightarrow x_j) = \text{lift}(x_j \Rightarrow x_i).$$

Postup při hledání pravidel

- Zvolíme si minimální hodnotu podpory α .
- Najdeme všechny dvojice položek x_i a x_j takové, že spolehlivost pravidla $x_i \Rightarrow x_j$ je alespoň α .
- Zvolíme si minimální hodnotu podpory β .
- Ze všech dvojic položek x_i a x_j vytvoříme pravidla $x_i \Rightarrow x_j$, resp. $x_j \Rightarrow x_i$.
- Z nich vybereme jen ta, která mají hodnotu spolehlivosti alespoň β .

Postup při hledání pravidel

- Zvolíme si minimální hodnotu podpory α .
- Najdeme všechny dvojice položek x_i a x_j takové, že spolehlivost pravidla $x_i \Rightarrow x_j$ je alespoň α .
- Zvolíme si minimální hodnotu podpory β .
- Ze všech dvojic položek x_i a x_j vytvoříme pravidla $x_i \Rightarrow x_j$, resp. $x_j \Rightarrow x_i$.
- Z nich vybereme jen ta, která mají hodnotu spolehlivosti alespoň β .

Postup při hledání pravidel

- Zvolíme si minimální hodnotu podpory α .
- Najdeme všechny dvojice položek x_i a x_j takové, že spolehlivost pravidla $x_i \Rightarrow x_j$ je alespoň α .
- Zvolíme si minimální hodnotu podpory β .
- Ze všech dvojic položek x_i a x_j vytvoříme pravidla $x_i \Rightarrow x_j$, resp. $x_j \Rightarrow x_i$.
- Z nich vybereme jen ta, která mají hodnotu spolehlivosti alespoň β .

Postup při hledání pravidel

- Zvolíme si minimální hodnotu podpory α .
- Najdeme všechny dvojice položek x_i a x_j takové, že spolehlivost pravidla $x_i \Rightarrow x_j$ je alespoň α .
- Zvolíme si minimální hodnotu podpory β .
- Ze všech dvojic položek x_i a x_j vytvoříme pravidla $x_i \Rightarrow x_j$, resp. $x_j \Rightarrow x_i$.
- Z nich vybereme jen ta, která mají hodnotu spolehlivosti alespoň β .

Postup při hledání pravidel

- Zvolíme si minimální hodnotu podpory α .
- Najdeme všechny dvojice položek x_i a x_j takové, že spolehlivost pravidla $x_i \Rightarrow x_j$ je alespoň α .
- Zvolíme si minimální hodnotu podpory β .
- Ze všech dvojic položek x_i a x_j vytvoříme pravidla $x_i \Rightarrow x_j$, resp. $x_j \Rightarrow x_i$.
- Z nich vybereme jen ta, která mají hodnotu spolehlivosti alespoň β .

Příklad - matice souslednosti

	máslo	pivo	pečivo	voda	zelenina
máslo	5	1	4	3	2
pivo	1	4	2	2	0
pečivo	4	2	6	3	2
voda	3	2	3	6	1
zelenina	2	0	2	1	3

Zvolíme minimální hodnotu podpory $\alpha = 0.3$.

- máslo, pečivo
- máslo, voda
- voda, pečivo

Příklad - matice souslednosti

	máslo	pivo	pečivo	voda	zelenina
máslo	5	1	4	3	2
pivo	1	4	2	2	0
pečivo	4	2	6	3	2
voda	3	2	3	6	1
zelenina	2	0	2	1	3

Zvolíme minimální hodnotu podpory $\alpha = 0.3$.

- máslo, pečivo
- máslo, voda
- voda, pečivo

Příklad - výsledná pravidla

Pravidlo	podpora	spolehlivost	zlepšení
máslo \Rightarrow pečivo	0.40	0.80	1.33
pečivo \Rightarrow máslo	0.40	0.67	1.33
máslo \Rightarrow voda	0.30	0.60	1.00
voda \Rightarrow máslo	0.30	0.50	1.00
voda \Rightarrow pečivo	0.30	0.50	0.83
pečivo \Rightarrow voda	0.30	0.50	0.83

Tale of beer and diapers

1990's: Americký obchodní řetězec Walmart na základě údajů o nákupech svých zákazníků zjistil, že prodej dětských plen a piva je silně korelovaný.

Tale of beer and diapers



Tale of beer and diapers

- 1990s: Americký obchodní řetězec Walmart na základě údajů o nákupech svých zákazníků zjistil, že prodej dětských plen a piva je silně korelovaný.
- Walmart zpracoval údaje ze svých věrnostních karet a transakčních dat a výsledek zpřesnil: Mladí Američané, kteří v pátek večer kupují dětské pleny, kupují také pivo.
- Vysvětlení: Mladí Američané jsou zvyklí v pátek večer popíjet s přáteli (pivo), ale pokud už mají děti, nemohou vysedávat s přáteli v barech, proto si koupí aspoň pivo, aby mohli popíjet doma.
- Po zjištění této souvislosti Walmart přesunul pivo hned vedle regálu s plenami. Prodej piva se rapidně zvýšil.
- Walmart nikdy ale statisticky své zdůvodnění neprokázal.
- Mnoho variací tohoto příběhu: 12 000 akademických článků, 14 000 knih, 1 000 000 internetových stránek.

Tale of beer and diapers

- 1990s: Americký obchodní řetězec Walmart na základě údajů o nákupech svých zákazníků zjistil, že prodej dětských plen a piva je silně korelovaný.
- Walmart zpracoval údaje ze svých věrnostních karet a transakčních dat a výsledek zpřesnil: Mladí Američané, kteří v pátek večer kupují dětské pleny, kupují také pivo.
- Vysvětlení: Mladí Američané jsou zvyklí v pátek večer popíjet s přáteli (pivo), ale pokud už mají děti, nemohou vysedávat s přáteli v barech, proto si koupí aspoň pivo, aby mohli popíjet doma.
- Po zjištění této souvislosti Walmart přesunul pivo hned vedle regálu s plenami. Prodej piva se rapidně zvýšil.
- Walmart nikdy ale statisticky své zdůvodnění neprokázal.
- Mnoho variací tohoto příběhu: 12 000 akademických článků, 14 000 knih, 1 000 000 internetových stránek.

Tale of beer and diapers

- 1990s: Americký obchodní řetězec Walmart na základě údajů o nákupech svých zákazníků zjistil, že prodej dětských plen a piva je silně korelovaný.
- Walmart zpracoval údaje ze svých věrnostních karet a transakčních dat a výsledek zpřesnil: Mladí Američané, kteří v pátek večer kupují dětské pleny, kupují také pivo.
- Vysvětlení: Mladí Američané jsou zvyklí v pátek večer popíjet s přáteli (pivo), ale pokud už mají děti, nemohou vysedávat s přáteli v barech, proto si koupí aspoň pivo, aby mohli popíjet doma.
- Po zjištění této souvislosti Walmart přesunul pivo hned vedle regálu s plenami. Prodej piva se rapidně zvýšil.
- Walmart nikdy ale statisticky své zdůvodnění neprokázal.
- Mnoho variací tohoto příběhu: 12 000 akademických článků, 14 000 knih, 1 000 000 internetových stránek.

Tale of beer and diapers

- 1990s: Americký obchodní řetězec Walmart na základě údajů o nákupech svých zákazníků zjistil, že prodej dětských plen a piva je silně korelovaný.
- Walmart zpracoval údaje ze svých věrnostních karet a transakčních dat a výsledek zpřesnil: Mladí Američané, kteří v pátek večer kupují dětské pleny, kupují také pivo.
- Vysvětlení: Mladí Američané jsou zvyklí v pátek večer popíjet s přáteli (pivo), ale pokud už mají děti, nemohou vysedávat s přáteli v barech, proto si koupí aspoň pivo, aby mohli popíjet doma.
- Po zjištění této souvislosti Walmart přesunul pivo hned vedle regálu s plenami. Prodej piva se rapidně zvýšil.
- Walmart nikdy ale statisticky své zdůvodnění neprokázal.
- Mnoho variací tohoto příběhu: 12 000 akademických článků, 14 000 knih, 1 000 000 internetových stránek.

Tale of beer and diapers

- 1990s: Americký obchodní řetězec Walmart na základě údajů o nákupech svých zákazníků zjistil, že prodej dětských plen a piva je silně korelovaný.
- Walmart zpracoval údaje ze svých věrnostních karet a transakčních dat a výsledek zpřesnil: Mladí Američané, kteří v pátek večer kupují dětské pleny, kupují také pivo.
- Vysvětlení: Mladí Američané jsou zvyklí v pátek večer popíjet s přáteli (pivo), ale pokud už mají děti, nemohou vysedávat s přáteli v barech, proto si koupí aspoň pivo, aby mohli popíjet doma.
- Po zjištění této souvislosti Walmart přesunul pivo hned vedle regálu s plenami. Prodej piva se rapidně zvýšil.
- Walmart nikdy ale statisticky své zdůvodnění neprokázal.
- Mnoho variací tohoto příběhu: 12 000 akademických článků, 14 000 knih, 1 000 000 internetových stránek.

Tale of beer and diapers

- 1990s: Americký obchodní řetězec Walmart na základě údajů o nákupech svých zákazníků zjistil, že prodej dětských plen a piva je silně korelovaný.
- Walmart zpracoval údaje ze svých věrnostních karet a transakčních dat a výsledek zpřesnil: Mladí Američané, kteří v pátek večer kupují dětské pleny, kupují také pivo.
- Vysvětlení: Mladí Američané jsou zvyklí v pátek večer popíjet s přáteli (pivo), ale pokud už mají děti, nemohou vysedávat s přáteli v barech, proto si koupí aspoň pivo, aby mohli popíjet doma.
- Po zjištění této souvislosti Walmart přesunul pivo hned vedle regálu s plenami. Prodej piva se rapidně zvýšil.
- Walmart nikdy ale statisticky své zdůvodnění neprokázal.
- Mnoho variací tohoto příběhu: 12 000 akademických článků, 14 000 knih, 1 000 000 internetových stránek.

Příklad - jak využít získaná pravidla

Pravidlo: panenka \Rightarrow sladkosti

- Dát je v obchodě do regálu blízko sebe.
- Dát je v obchodě daleko od sebe.
- Vytvořit balíček panenka+sladkosti.
- Vytvořit balíček panenka+sladkosti+špatně prodávané zboží.
- Snížit cenu jednoho zboží a zvýšit cenu druhého.
- Nabídnout příslušenství pro panenky jako dárek za nákup.
- Do slevových letáků nedávat současně panenku a sladkosti.
- Nabídnout sladkosti ve tvaru panenky.

Příklad - jak využít získaná pravidla

Pravidlo: panenka \Rightarrow sladkosti

- Dát je v obchodě do regálu blízko sebe.
- Dát je v obchodě daleko od sebe.
- Vytvořit balíček panenka+sladkosti.
- Vytvořit balíček panenka+sladkosti+špatně prodávané zboží.
- Snížit cenu jednoho zboží a zvýšit cenu druhého.
- Nabídnout příslušenství pro panenky jako dárek za nákup.
- Do slevových letáků nedávat současně panenku a sladkosti.
- Nabídnout sladkosti ve tvaru panenky.

Příklad - jak využít získaná pravidla

Pravidlo: panenka \Rightarrow sladkosti

- Dát je v obchodě do regálu blízko sebe.
- Dát je v obchodě daleko od sebe.
- Vytvořit balíček panenka+sladkosti.
- Vytvořit balíček panenka+sladkosti+špatně prodávané zboží.
- Snížit cenu jednoho zboží a zvýšit cenu druhého.
- Nabídnout příslušenství pro panenky jako dárek za nákup.
- Do slevových letáků nedávat současně panenku a sladkosti.
- Nabídnout sladkosti ve tvaru panenky.

Příklad - jak využít získaná pravidla

Pravidlo: panenka \Rightarrow sladkosti

- Dát je v obchodě do regálu blízko sebe.
- Dát je v obchodě daleko od sebe.
- Vytvořit balíček panenka+sladkosti.
- Vytvořit balíček panenka+sladkosti+špatně prodávané zboží.
- Snížit cenu jednoho zboží a zvýšit cenu druhého.
- Nabídnout příslušenství pro panenky jako dárek za nákup.
- Do slevových letáků nedávat současně panenku a sladkosti.
- Nabídnout sladkosti ve tvaru panenky.

Příklad - jak využít získaná pravidla

Pravidlo: panenka \Rightarrow sladkosti

- Dát je v obchodě do regálu blízko sebe.
- Dát je v obchodě daleko od sebe.
- Vytvořit balíček panenka+sladkosti.
- Vytvořit balíček panenka+sladkosti+špatně prodávané zboží.
- Snížit cenu jednoho zboží a zvýšit cenu druhého.
- Nabídnout příslušenství pro panenky jako dárek za nákup.
- Do slevových letáků nedávat současně panenku a sladkosti.
- Nabídnout sladkosti ve tvaru panenky.

Příklad - jak využít získaná pravidla

Pravidlo: panenka \Rightarrow sladkosti

- Dát je v obchodě do regálu blízko sebe.
- Dát je v obchodě daleko od sebe.
- Vytvořit balíček panenka+sladkosti.
- Vytvořit balíček panenka+sladkosti+špatně prodávané zboží.
- Snížit cenu jednoho zboží a zvýšit cenu druhého.
- Nabídnout příslušenství pro panenky jako dárek za nákup.
- Do slevových letáků nedávat současně panenku a sladkosti.
- Nabídnout sladkosti ve tvaru panenky.

Příklad - jak využít získaná pravidla

Pravidlo: panenka \Rightarrow sladkosti

- Dát je v obchodě do regálu blízko sebe.
- Dát je v obchodě daleko od sebe.
- Vytvořit balíček panenka+sladkosti.
- Vytvořit balíček panenka+sladkosti+špatně prodávané zboží.
- Snížit cenu jednoho zboží a zvýšit cenu druhého.
- Nabídnout příslušenství pro panenky jako dárek za nákup.
- Do slevových letáků nedávat současně panenku a sladkosti.
- Nabídnout sladkosti ve tvaru panenky.

Příklad - jak využít získaná pravidla

Pravidlo: panenka \Rightarrow sladkosti

- Dát je v obchodě do regálu blízko sebe.
- Dát je v obchodě daleko od sebe.
- Vytvořit balíček panenka+sladkosti.
- Vytvořit balíček panenka+sladkosti+špatně prodávané zboží.
- Snížit cenu jednoho zboží a zvýšit cenu druhého.
- Nabídnout příslušenství pro panenky jako dárek za nákup.
- Do slevových letáků nedávat současně panenku a sladkosti.
- Nabídnout sladkosti ve tvaru panenky.

Příklad - jak využít získaná pravidla

Pravidlo: panenka \Rightarrow sladkosti

- Dát je v obchodě do regálu blízko sebe.
- Dát je v obchodě daleko od sebe.
- Vytvořit balíček panenka+sladkosti.
- Vytvořit balíček panenka+sladkosti+špatně prodávané zboží.
- Snížit cenu jednoho zboží a zvýšit cenu druhého.
- Nabídnout příslušenství pro panenky jako dárek za nákup.
- Do slevových letáků nedávat současně panenku a sladkosti.
- Nabídnout sladkosti ve tvaru panenky.

Zobecnění analýzy nákupního košíku - víceprvková pravidla

- Budeme hledat pravidla typu:

$$x_i \wedge x_j \wedge x_k \Rightarrow x_l \wedge x_m.$$

- Analogicky jako v předchozím můžeme definovat podporu a spolehlivost.
- Pro hledání pravidel opět předem určíme minimální hodnotu podpory α a spolehlivosti β .

Zobecnění analýzy nákupního košíku - víceprvková pravidla

- Budeme hledat pravidla typu:

$$x_i \wedge x_j \wedge x_k \Rightarrow x_l \wedge x_m.$$

- Analogicky jako v předchozím můžeme definovat podporu a spolehlivost.
- Pro hledání pravidel opět předem určíme minimální hodnotu podpory α a spolehlivosti β .

Zobecnění analýzy nákupního košíku - víceprvková pravidla

- Budeme hledat pravidla typu:

$$x_i \wedge x_j \wedge x_k \Rightarrow x_l \wedge x_m.$$

- Analogicky jako v předchozím můžeme definovat podporu a spolehlivost.
- Pro hledání pravidel opět předem určíme minimální hodnotu podpory α a spolehlivosti β .

Zobecnění analýzy nákupního košíku - víceprvková pravidla

- Nejprve najdeme všechny dvouprvkové množiny, jejichž podpora je alespoň α .
- Poté se zaměříme na 3-prvkové, 4-prvkové, \dots , celou množinu I . Opět z nich vybereme takové, jejichž podpora je alespoň α .
- Celkový počet procházených množin:

$$\binom{m}{2} + \binom{m}{3} + \dots + \binom{m}{m} = 2^m - m - 1.$$

- Pro danou k -prvkovou množinu uvažujeme všechna možná pravidla. Vybereme z nich jen ta, jejichž spolehlivost je alespoň β .
- Celkový počet procházených pravidel:

$$\binom{k}{1} + \binom{k}{2} + \dots + \binom{k}{k-1} = 2^k - 2.$$

Zobecnění analýzy nákupního košíku - víceprvková pravidla

- Nejprve najdeme všechny dvouprvkové množiny, jejichž podpora je alespoň α .
- Poté se zaměříme na 3-prvkové, 4-prvkové, ..., celou množinu I . Opět z nich vybereme takové, jejichž podpora je alespoň α .
- Celkový počet procházených množin:

$$\binom{m}{2} + \binom{m}{3} + \dots + \binom{m}{m} = 2^m - m - 1.$$

- Pro danou k -prvkovou množinu uvažujeme všechna možná pravidla. Vybereme z nich jen ta, jejichž spolehlivost je alespoň β .
- Celkový počet procházených pravidel:

$$\binom{k}{1} + \binom{k}{2} + \dots + \binom{k}{k-1} = 2^k - 2.$$

Zobecnění analýzy nákupního košíku - víceprvková pravidla

- Nejprve najdeme všechny dvouprvkové množiny, jejichž podpora je alespoň α .
- Poté se zaměříme na 3-prvkové, 4-prvkové, ..., celou množinu I . Opět z nich vybereme takové, jejichž podpora je alespoň α .
- Celkový počet procházených množin:

$$\binom{m}{2} + \binom{m}{3} + \dots + \binom{m}{m} = 2^m - m - 1.$$

- Pro danou k -prvkovou množinu uvažujeme všechna možná pravidla. Vybereme z nich jen ta, jejichž spolehlivost je alespoň β .
- Celkový počet procházených pravidel:

$$\binom{k}{1} + \binom{k}{2} + \dots + \binom{k}{k-1} = 2^k - 2.$$

Zobecnění analýzy nákupního košíku - víceprvková pravidla

- Nejprve najdeme všechny dvouprvkové množiny, jejichž podpora je alespoň α .
- Poté se zaměříme na 3-prvkové, 4-prvkové, ..., celou množinu I . Opět z nich vybereme takové, jejichž podpora je alespoň α .
- Celkový počet procházených množin:

$$\binom{m}{2} + \binom{m}{3} + \dots + \binom{m}{m} = 2^m - m - 1.$$

- Pro danou k -prvkovou množinu uvažujeme všechna možná pravidla. Vybereme z nich jen ta, jejichž spolehlivost je alespoň β .
- Celkový počet procházených pravidel:

$$\binom{k}{1} + \binom{k}{2} + \dots + \binom{k}{k-1} = 2^k - 2.$$

Zobecnění analýzy nákupního košíku - víceprvková pravidla

- Nejprve najdeme všechny dvouprvkové množiny, jejichž podpora je alespoň α .
- Poté se zaměříme na 3-prvkové, 4-prvkové, \dots , celou množinu I . Opět z nich vybereme takové, jejichž podpora je alespoň α .
- Celkový počet procházených množin:

$$\binom{m}{2} + \binom{m}{3} + \dots + \binom{m}{m} = 2^m - m - 1.$$

- Pro danou k -prvkovou množinu uvažujeme všechna možná pravidla. Vybereme z nich jen ta, jejichž spolehlivost je alespoň β .
- Celkový počet procházených pravidel:

$$\binom{k}{1} + \binom{k}{2} + \dots + \binom{k}{k-1} = 2^k - 2.$$

Zobecnění analýzy nákupního košíku - víceprvková pravidla

- Velká časová náročnost – prořezávání, algoritmická řešení.
- Zřejmě platí

$$\text{supp}(x_i, x_j) \geq \text{supp}(x_i, x_j, x_k).$$

- **Důsledek:** Pokud daná množina prvků má podporu menší než α , není třeba procházet všechny množiny, které ji obsahují.

Zobecnění analýzy nákupního košíku - víceprvková pravidla

- Velká časová náročnost – prořezávání, algoritmická řešení.
- Zřejmě platí

$$\text{supp}(x_i, x_j) \geq \text{supp}(x_i, x_j, x_k).$$

- **Důsledek:** Pokud daná množina prvků má podporu menší než α , není třeba procházet všechny množiny, které ji obsahují.

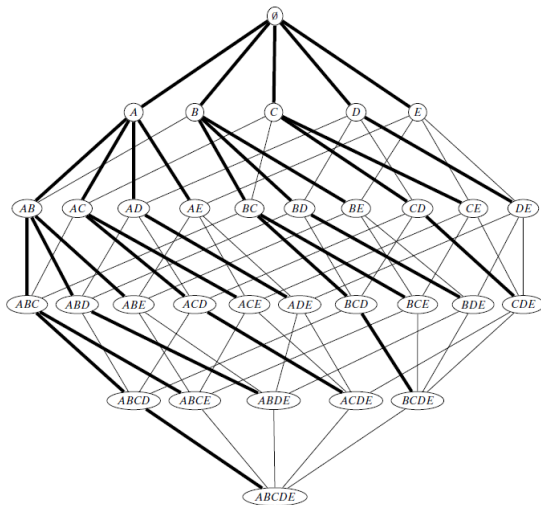
Zobecnění analýzy nákupního košíku - víceprvková pravidla

- Velká časová náročnost – prořezávání, algoritmická řešení.
- Zřejmě platí

$$\text{supp}(x_i, x_j) \geq \text{supp}(x_i, x_j, x_k).$$

- **Důsledek:** Pokud daná množina prvků má podporu menší než α , není třeba procházet všechny množiny, které ji obsahují.

Zobecnění analýzy nákupního košíku - prořezávání



Zobecnění analýzy nákupního košíku - prořezávání

- Podobně platí

$$\text{conf}(x_i \wedge x_j \Rightarrow x_k) \geq \text{conf}(x_i \Rightarrow x_j \wedge x_k).$$

Zobecnění analýzy nákupního košíku

- 1 **Pravidla:** $x_i \Rightarrow \text{NOT } x_j$.
 - Zavedou se nové položky NOT x_i indikující, že daná transakce neobsahuje položku x_i .
 - Nevýhodou je vyšší časová náročnost (dvakrát tolik položek).
 - Negované položky se vyskytují častěji než původní \Rightarrow horší interpretovatelnost pravidel.
- 2 **Využití hierarchie položek.**
 - Původní data bývají často řídká.
 - Existuje přirozená hierarchie položek: banán, jablko, pomeranč \rightarrow OVOCE.
 - Zavedou se nové položky popisující hierarchii.
 - Pravidla typu banán \Rightarrow OVOCE vs. OVOCE \Rightarrow banán.
- 3 **Aplikace na spojité a kategoriální náhodné veličiny.**
 - VĚK < 20 \Rightarrow poslouchá POP.
- 4 **Sekvenční analýza.**

Zobecnění analýzy nákupního košíku

- 1 **Pravidla:** $x_i \Rightarrow \text{NOT } x_j$.
 - Zavedou se nové položky NOT x_i indikující, že daná transakce neobsahuje položku x_i .
 - Nevýhodou je vyšší časová náročnost (dvakrát tolik položek).
 - Negované položky se vyskytují častěji než původní \Rightarrow horší interpretovatelnost pravidel.
- 2 **Využití hierarchie položek.**
 - Původní data bývají často řídká.
 - Existuje přirozená hierarchie položek: banán, jablko, pomeranč \rightarrow OVOCE.
 - Zavedou se nové položky popisující hierarchii.
 - Pravidla typu banán \Rightarrow OVOCE vs. OVOCE \Rightarrow banán.
- 3 **Aplikace na spojité a kategoriální náhodné veličiny.**
 - VĚK < 20 \Rightarrow poslouchá POP.
- 4 **Sekvenční analýza.**

Zobecnění analýzy nákupního košíku

- 1 **Pravidla:** $x_i \Rightarrow \text{NOT } x_j$.
 - Zavedou se nové položky NOT x_i indikující, že daná transakce neobsahuje položku x_i .
 - Nevýhodou je vyšší časová náročnost (dvakrát tolik položek).
 - Negované položky se vyskytují častěji než původní \Rightarrow horší interpretovatelnost pravidel.
- 2 **Využití hierarchie položek.**
 - Původní data bývají často řídká.
 - Existuje přirozená hierarchie položek: banán, jablko, pomeranč \rightarrow OVOCE.
 - Zavedou se nové položky popisující hierarchii.
 - Pravidla typu banán \Rightarrow OVOCE vs. OVOCE \Rightarrow banán.
- 3 **Aplikace na spojité a kategoriální náhodné veličiny.**
 - VĚK < 20 \Rightarrow poslouchá POP.
- 4 **Sekvenční analýza.**

Zobecnění analýzy nákupního košíku

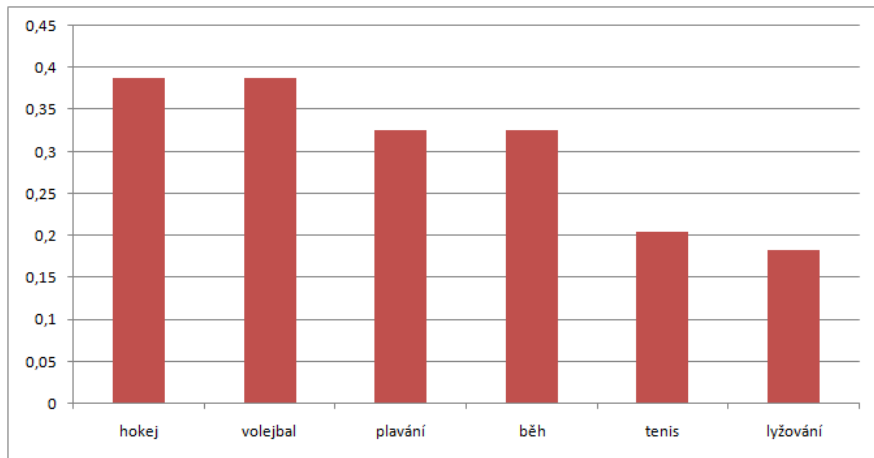
- 1 **Pravidla:** $x_i \Rightarrow \text{NOT } x_j$.
 - Zavedou se nové položky NOT x_i indikující, že daná transakce neobsahuje položku x_i .
 - Nevýhodou je vyšší časová náročnost (dvakrát tolik položek).
 - Negované položky se vyskytují častěji než původní \Rightarrow horší interpretovatelnost pravidel.
- 2 **Využití hierarchie položek.**
 - Původní data bývají často řídká.
 - Existuje přirozená hierarchie položek: banán, jablko, pomeranč \rightarrow OVOCE.
 - Zavedou se nové položky popisující hierarchii.
 - Pravidla typu banán \Rightarrow OVOCE vs. OVOCE \Rightarrow banán.
- 3 **Aplikace na spojité a kategoriální náhodné veličiny.**
 - VĚK < 20 \Rightarrow poslouchá POP.
- 4 **Sekvenční analýza.**

Oblíbené sporty

Data: 49 studentů uvedlo několik svých oblíbených sportů:

ID	Oblíbené sporty
1	běh,plavání,turistika,lezení
2	hokej,tanec,tenis
3	hokej,fotbal,biatlon,stolní tenis,karate
.	.
.	.
49	brusle,fitness,hokej,squash

Oblíbené sporty - nejoblíbenější



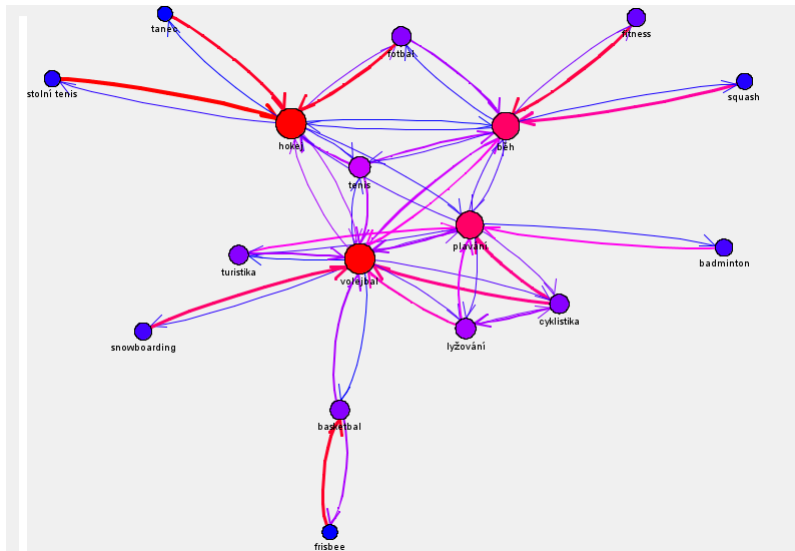
Oblíbené sporty - netradiční

- americký fotbal
- bříšní tanec
- karate
- kickbox
- pólo
- skialpinismus
- slamball
- šerm
- vodní fotbal

Oblíbené sporty - výsledná dvouprvková pravidla

Pravidlo	podpora (v %)	spolehlivost (v %)	zlepšení
běh ⇒ volejbal	16.3	50.0	1.29
fotbal ⇒ hokej	12.2	75.5	1.93
cyklistika ⇒ plavání	10.2	62.5	1.91
fitness ⇒ běh	10.2	71.4	2.19
cyklistika ⇒ volejbal	10.2	62.5	1.61
lyžování ⇒ volejbal	10.2	55.6	1.43

Oblíbené sporty - výsledná dvouprvková pravidla



Oblíbené sporty - výsledná víceprvková pravidla

Pravidlo	podpora (v %)	spolehlivost (v %)
lyžování \wedge cyklistika \Rightarrow volejbal	6.12	100
cyklistika \wedge volejbal \Rightarrow plavání	6.12	60.0
běh \wedge volejbal \Rightarrow plavání	6.12	37.5
běh \wedge turistika \Rightarrow plavání	4.08	100

Oblíbené sporty - pravidla s využitím hierarchie

Hierarchie: 4 kategorie sportů – míčové, zimní, pro volný čas, ostatní.

Pravidlo	podpora (v %)	spolehlivost (v %)
zimní sporty \Rightarrow míčové sporty	44.9	73.3
míčové sporty \Rightarrow zimní sporty	44.9	66.7
zimní sporty \Rightarrow hokej	38.8	63.3
míčové sporty \Rightarrow volejbal	38.8	57.6
hokej \Rightarrow míčové sporty	28.6	73.7
volejbal \Rightarrow zimní sporty	28.6	73.7
běh \Rightarrow míčové sporty	24.5	75.0

Pohled do budoucna

- V současné době hojně využíváno v komerční sféře.
- Youtube – "mohlo by se vám líbit"
- Internetové obchody – doporučené zboží.
- Optimalizace skladových zásob.
- Amazon si nechal patentovat službu *Method and system for anticipatory package shipping*, tj. Amazon vám zašle zboží, ještě než si ho stačíte objednat.

Pohled do budoucna

- V současné době hojně využíváno v komerční sféře.
- Youtube – "mohlo by se vám líbit"
- Internetové obchody – doporučené zboží.
- Optimalizace skladových zásob.
- Amazon si nechal patentovat službu *Method and system for anticipatory package shipping*, tj. Amazon vám zašle zboží, ještě než si ho stačíte objednat.

Pohled do budoucna

- V současné době hojně využíváno v komerční sféře.
- Youtube – "mohlo by se vám líbit"
- Internetové obchody – doporučené zboží.
- Optimalizace skladových zásob.
- Amazon si nechal patentovat službu *Method and system for anticipatory package shipping*, tj. Amazon vám zašle zboží, ještě než si ho stačíte objednat.

Pohled do budoucna

- V současné době hojně využíváno v komerční sféře.
- Youtube – "mohlo by se vám líbit"
- Internetové obchody – doporučené zboží.
- Optimalizace skladových zásob.
- Amazon si nechal patentovat službu *Method and system for anticipatory package shipping*, tj. Amazon vám zašle zboží, ještě než si ho stačíte objednat.

Pohled do budoucna

- V současné době hojně využíváno v komerční sféře.
- Youtube – "mohlo by se vám líbit"
- Internetové obchody – doporučené zboží.
- Optimalizace skladových zásob.
- Amazon si nechal patentovat službu *Method and system for anticipatory package shipping*, tj. Amazon vám zašle zboží, ještě než si ho stačíte objednat.

Poděkování

Děkuji za pozornost.

Práce byla spolufinancována grantem MUNI/A/1204/2017.