

Tato prezentace je součástí wiki-prezentace [Metoda GUHA a systém LISp-Miner](#)

Je dostupná z [této adresy](#)

Verze 1, 6. 7. 2019

Matice dat a atributy

prof. RNDr. Jan Rauch, CSc.

Katedra informačního a znalostního inženýrství

VŠE Praha

Matice dat a atributy

- Matice dat
- Nominální, ordinální a kardinální atributy
- Základní booleovský atribut
- Booleovský atribut

Matrice dat

řádek	A_1	A_2	...	A_K
o_1	3	9	...	6
o_2	7	5	...	4
...
o_n	3	7	...	1

- řádky matice odpovídají pozorovaným objektům o_1, \dots, o_n
- sloupce A_1, A_2, \dots, A_K odpovídají pozorovaným atributům, nazývají se základní atributy
- $A_1(o_1)$ je hodnota atributu A_1 pro objekt o_1 , platí $A_1(o_1) = 3$, atd.
- přípustné hodnoty základního atributu se nazývají kategorie
- Každý atribut (sloupec) má **konečný počet** přípustných hodnot

Poznámka: V jiném kontextu může být matice dat definována jinak, např. může mít nekonečně mnoho přípustných hodnot – reálných čísel.

Matice dat - příklad

Matice dat: HotelPlusExterni

pobyt	HPohlaví	HVek	HMesto	HStat	Mobloha	Dhodnocení	...
1	žena	⟨15;24⟩	Brno	ČR	slunečno	spokojen	
2	žena	⟨25;34⟩	Linec	Rakousko	zataženo	průměr	
3	muž	⟨45;54⟩	Gdaňsk	Polsko	srážky	nespokojen	
4	žena	⟨65;74⟩	Košice	Slovensko	slunečno	průměr	
...
2 000	muž	⟨15;24⟩	Mnichov	Německo	zataženo	spokojen	

Nominální, ordinální a kardinální atributy

Atributy – sloupce matice mají různé vlastnosti. Obecně rozlišujeme nominální, ordinální a kardinální atributy:

Nominální atribut má kategorie, které se nedají smysluplně uspořádat; příkladem je atribut *Stav* s kategoriemi *ženatý, svobodný, rozvedený, vdovec*.

Ordinální atribut má kategorie, které se dají smysluplně uspořádat; příkladem je atribut *Vzdělání* s kategoriemi *základní, vyučen, středoškolské, vysokoškolské*.

Kardinální atribut má přípustné hodnoty, které mají význam čísel, tedy je lze například vzájemně dělit; příkladem jsou atributy *Výška v m* a *Váha v kg*, ze kterých lze vypočítat hodnotu atributu *BMI* podle vzorce $BMI = \frac{\text{Váha v kg}}{(\text{Výška v m})^2}$.

Matice dat a základní booleovské atributy

řádek	A_1	A_2	...	A_K	$A_1(3)$	$A_2(5,7)$...
o_1	3	9	...	6	1	0	...
o_2	7	5	...	4	0	1	...
...
o_n	3	7	...	1	1	1	...

Základní atributy Základní booleovské atributy

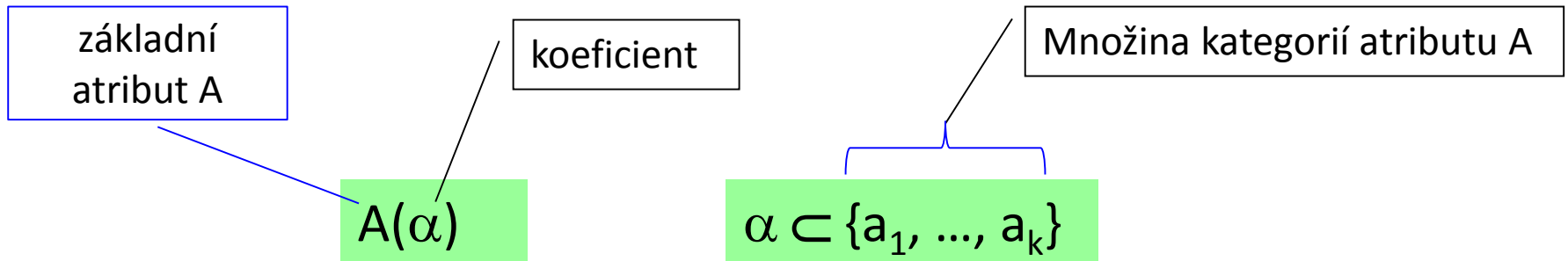
- $A(\alpha)$ je základní booleovský atribut, α je podmnožina množiny kategorií základního atributu A
- $A(\alpha)$ je pravdivý pro objekt o_i právě když $A(o_i) \in \alpha$
- místo $A(\{a_1, a_2\})$ píšeme pouze $A(a_1, a_2)$ i když „ a_1, a_2 “ není množina

Základní booleovské atributy – příklady

HVek	HStát	HVek⟨15;34⟩	HStat(ČR)	HStát(Rakousko, Německo)	...
⟨15;24⟩	ČR	1	1	0	...
⟨25;34⟩	Rakousko	1	0	1	...
⟨45;54⟩	Polsko	0	0	0	...
⟨65;74⟩	Slovensko	0	0	0	...
...
⟨15;24⟩	Německo	1	0	1	...

- HVek⟨15;34⟩ je zkrácený zápis pro HVek(⟨15;24⟩, ⟨25;34⟩)
- HStát(Rakousko, Německo) říká že se jedná o stát, ve kterém se mluví německy, kategorie atributu HStát jsou: ČR, Německo, Polsko, Rakousko, Slovensko

Základní booleovský atribut $A(\alpha)$



- $A(\alpha)$ je pravdivý pro objekt o_i právě když $A(o_i) \in \alpha$
říkáme také, že:
 - objekt o_i splňuje $A(\alpha)$
 - hodnota $A(\alpha)$ pro objekt o_i je 1
- místo $A(\{a_1, a_2\})$ píšeme pouze $A(a_1, a_2)$ i když „ a_1, a_2 “ není množina

Booleovský atribut

- Každý základní booleovský atribut je booleovský atribut
- Když φ a ψ jsou booleovské atributy pak $\neg\varphi$, $\varphi \wedge \psi$, $\varphi \vee \psi$ jsou booleovské atributy
- Platí obvyklé konvence o závorkách
- Booleovský atribut, který není základní, se nazývá odvozený

Booleovský atribut – příklady

- Základní booleovské atributy
HStat(ČR), HPohlavi(žena), POsob(1), POsob(1,2)
- Odvozené booleovské atributy
 - ¬ HStat(ČR)
 - HStat(ČR) ∧ POsob(1)
 - HStat(Slovensko) ∨ HPohlavi(žena)
 - HStat(Německo) ∧ (POsob(1,2) ∨ HPohlavi(muž))

Booleovský atribut je pravdivý

- Základní atribut $A(\alpha)$ je pravdivý pro objekt o_i právě když $A(o_i) \in \alpha$
- Pravdivost odvozených booleovských atributů je dána obvyklým způsobem:

φ	ψ	$\neg\varphi$	$\varphi \wedge \psi$	$\varphi \vee \psi$
1	1	0	1	1
1	0	0	0	1
0	1	1	0	1
0	0	1	0	0

- Jestliže je booleovský atribut φ pravdivý pro objekt o_i , pak také říkáme, že objekt o_i splňuje φ nebo že jeho hodnota pro objekt je 1 (viz tabulka)

Matice dat a atributy - literatura

Rauch J., Šimůnek M.: *Dobývání znalostí z databází, LISp-Miner a GUHA.*

Odstavce

3.1 Data a matice dat

3.2 Nominální, ordinální a kardinální atributy

4.2 Booleovské atributy