




PLATFORMA PREZENTACJI 4.0

## Zależności funkcyjne i przybliżone

Marzena Kryszkiewicz  
Politechnika Warszawska

## Definicje i własności




PLATFORMA PREZENTACJI 4.0

## Zależności funkcyjne

Niech  $D = (O, AT)$ , gdzie  $O$  jest zbiorem obiektów, a  $AT$  – zbiorem ich atrybutów. Niech  $A, B \subseteq AT$ .

- $A \rightarrow B$  jest definiowana jako *zależność funkcyjna*, jeśli  $\forall x \in O, [x]_A \subseteq [x]_B$ .
- Zależność funkcyjna  $A \rightarrow B$  jest zwana *minimalną*, jeśli  $\forall C \subset A, C \rightarrow B$  nie jest zależnością funkcyjną.
- Zależność  $A \rightarrow B$  jest zwana *trywialną*, jeśli  $B \subseteq A$ .

3



PLATFORMA PREZENTACJI 4.0


## Przykład: Zależności funkcyjne

$D = (O, AT)$ , gdzie  
 $O = \{1, 2, 3, 4, 5, 6\}$ ,  
 $AT = \{abcd\}$ .

Id	a	b	c	d
1	1	1	2	2
2	1	1	2	2
3	2	2	2	2
4	3	3	3	3
5	5	5	6	6
6	5	5	6	6

- $\{cd\} \rightarrow \{ab\}$  nie jest zależnością funkcyjną.
- $\{ab\} \rightarrow \{cd\}$  jest zależnością funkcyjną, ale nie jest minimalną.
- $\{a\} \rightarrow \{cd\}$  jest minimalną zależnością funkcyjną.
- $\{ab\} \rightarrow \{a\}$  jest zależnością trywialną.

4




PLATFORMA PREZENTACJI 4.0

## Własności zależności funkcyjnych

**Twierdzenie (Podstawowe reguły stosowane w algorytmie TANE).**  
 Niech  $a \in X$  i  $b \notin X$ .

- r1:** Jeśli  $X \setminus \{a\} \rightarrow \{a\}$  jest zależnością funkcyjną, to  $\forall Y \supset X, Y \setminus \{a\} \rightarrow \{a\}$  jest zależnością funkcyjną, ale nie jest minimalną.
- r2:** Jeśli  $X \setminus \{a\} \rightarrow \{a\}$  i  $X \rightarrow \{b\}$  są zależnościami funkcyjnymi, to  $\forall Y \supset X, Y \setminus \{b\} \rightarrow \{b\}$  jest zależnością funkcyjną, ale nie jest minimalną.

5



PLATFORMA PREZENTACJI 4.0

## Dowód własności r1 zależności funkcyjnych...

**Twierdzenie (reguła r1 stosowana w algorytmie TANE).** Niech  $a \in X$ .

- r1:** Jeśli  $X \setminus \{a\} \rightarrow \{a\}$  jest zależnością funkcyjną, to  $\forall Y \supset X, Y \setminus \{a\} \rightarrow \{a\}$  jest zależnością funkcyjną, ale nie jest minimalną.

**Dowód:** Ponieważ  $a \in X$ , to  $\forall Y \supset X, Y \setminus \{a\} \supset X \setminus \{a\}$ .  
 Stąd  $X \setminus \{a\} \rightarrow \{a\}$  jest zależnością funkcyjną.  
 W rezultacie  $\forall Y \supset X, Y \setminus \{a\} \rightarrow \{a\}$  jest zależnością funkcyjną, ale nie jest minimalną.

6

**Dowód własności r2 zależności funkcyjnych**

**Twierdzenie (reguła r2 stosowana w algorytmie TANE).** Niech  $a \in X$  i  $b \notin X$ .

- **r2:** Jeśli  $X \setminus \{a\} \rightarrow \{a\}$  i  $X \rightarrow \{b\}$  są zależnościami funkcyjnymi, to  $\forall Y \supset X, Y \setminus \{b\} \rightarrow \{b\}$  jest zależnością funkcyjną, ale nie jest minimalną.

**Dowód:** Jeśli  $X \setminus \{a\} \rightarrow \{a\}$  i  $X \rightarrow \{b\}$  są zależnościami funkcyjnymi, to  $X \setminus \{a\} \rightarrow X$  i  $X \rightarrow \{b\}$  są zależnościami funkcyjnymi, a zatem  $X \rightarrow \{b\}$  jest funkcyjne, ale nie jest minimalne. Ponieważ  $b \notin X$ , to  $X \setminus \{b\} = X$ . Stąd  $X \setminus \{b\} \rightarrow \{b\}$  jest funkcyjne, ale nie jest minimalne. W rezultacie  $\forall Y \supset X, Y \setminus \{b\} \rightarrow \{b\}$  jest funkcyjne, ale nie jest minimalne.

7

**Algorytm TANE: odkrywanie zależności funkcyjnych i/lub przybliżonych**

**Algorytm TANE: Odkrywanie zależności funkcyjnych**

- TANE służy do odkrywania **minimalnych nietrywialnych zależności funkcyjnych (i/lub przybliżonych) z jednym zależnym atrybutem**.
- W  $i$ -tej iteracji algorytmu TANE, tworzone są zależności kandydujące  $X \setminus \{a\} \rightarrow \{a\}$  ze zbiorów atrybutów  $X$  o liczności  $i$ .
- Nowe zbiory atrybutów o liczności  $i + 1$  są tworzone ze zbiorów atrybutów o liczności  $i$ , tak jak w algorytmie Apriori.
- Podstawowe reguły r1 i r2 są używane w celu uniknięcia wytwarzania zbiorów atrybutów o liczności  $i + 1$ , z których nie jest możliwe uzyskanie żadnych minimalnych nietrywialnych zależności funkcyjnych.

9

**TANE-przykład: Odkrywanie zależności funkcyjnych...**

Id	a	b	c	d	i=1	{a}	{b}	{c}	{d}
1	1	A	\$	F	C+	{abcd}	{abcd}	{abcd}	{abcd}
2	1	A	€	T	cand.	$\emptyset \rightarrow \{a\}$	$\emptyset \rightarrow \{b\}$	$\emptyset \rightarrow \{c\}$	$\emptyset \rightarrow \{d\}$
3	2	A	\$	D					
4	2	A	\$	F					
5	2	B	€	L					
6	3	B	\$	O					
7	3	C	€	F					
8	3	C	£	R					

10

**TANE-przykład: Odkrywanie zależności funkcyjnych...**

Id	a	b	c	d	i=2	{ab}	{ac}	{ad}	{bc}	{bd}	{cd}
1	1	A	\$	F	C+	{abcd}	{abcd}	{abcd}	{abcd}	{abcd}	{abcd}
2	1	A	€	T	cand.	$\{a\} \rightarrow \{b\}$	$\{b\} \rightarrow \{a\}$	$\{a\} \rightarrow \{d\}$	$\{b\} \rightarrow \{c\}$	$\{b\} \rightarrow \{d\}$	$\{c\} \rightarrow \{d\}$
3	2	A	\$	D		$\{b\} \rightarrow \{a\}$	$\{c\} \rightarrow \{a\}$	$\{d\} \rightarrow \{a\}$	$\{c\} \rightarrow \{b\}$	$\{d\} \rightarrow \{b\}$	$\{d\} \rightarrow \{c\}$
4	2	A	\$	F							
5	2	B	€	L							
6	3	B	\$	O							
7	3	C	€	F							
8	3	C	£	R							

11

**TANE-przykład: Odkrywanie zależności funkcyjnych...**

Id	a	b	c	d	i=3	{abc}	{abd}	{acd}	{bcd}
1	1	A	\$	F	C+	{abcd}	{abcd}	{abcd}	{abcd}
2	1	A	€	T	cand.	$\{ab\} \rightarrow \{c\}$	$\{ab\} \rightarrow \{d\}$	$\{ac\} \rightarrow \{d\}$	$\{bc\} \rightarrow \{d\}$
3	2	A	\$	D		$\{ac\} \rightarrow \{b\}$	$\{ad\} \rightarrow \{b\}$	$\{ad\} \rightarrow \{c\}$	$\{bd\} \rightarrow \{c\}$
4	2	A	\$	F		$\{bc\} \rightarrow \{a\}$	$\{bd\} \rightarrow \{a\}$	$\{cd\} \rightarrow \{a\}$	$\{cd\} \rightarrow \{b\}$
5	2	B	€	L					
6	3	B	\$	O					
7	3	C	€	F					
8	3	C	£	R					

delete b by rule r1      delete d by rule r2

12

**TANE-przykład: Odkrywanie zależności funkcyjnych**

Id	a	b	c	d
1	1	A	\$	F
2	1	A	€	T
3	2	A	\$	D
4	2	A	\$	F
5	2	B	€	L
6	3	B	\$	O
7	3	C	€	F
8	3	C	£	R

**i=3**

	{abc}	{abd}	{acd}	{bcd}
C <sup>+</sup>	{ac}	{ad}	{ad}	{d}
cand.	{ab}→{c}	{ab}→{d}	{ac}→{d}	{bc}→{d}
	{ac}→{b}	{ad}→{b}	{ad}→{c}	{bd}→{c}
	{bc}→{a}	{bd}→{a}	{cd}→{a}	{cd}→{b}

**i=4**

	{abcd}
C <sup>+</sup>	{ac}∩{ad}∩{ad}∩{d}=∅
cand.	-

13

**Przykład: Zależności funkcyjne a podziały**

D = (O, AT), gdzie  
 O = {1,2,3,4,5,6},  
 AT = {abcd}.

•  $\Pi_{\{ab\}} = \{\{1,2\}, \{3\}, \{4\}, \{5,6\}\}$   
 •  $\Pi_{\{cd\}} = \{\{1,2,3\}, \{4\}, \{5,6\}\}$

Id	a	b	c	d
1	1	1	2	2
2	1	1	2	2
3	2	2	2	2
4	3	3	3	3
5	5	5	6	6
6	5	5	6	6

•  $[1]_{\{ab\}} = \{1,2\} \in \Pi_{\{ab\}}$   
 •  $[1]_{\{cd\}} = \{1,2,3\} \in \Pi_{\{cd\}}$   
 •  $\neg([1]_{\{cd\}} \subseteq [1]_{\{ab\}})$   
 • Stąd  $\{cd\} \rightarrow \{ab\}$  nie jest zależnością funkcyjną.

14

**Zależności funkcyjne a podziały**

**Twierdzenie.**  
 $X \rightarrow Y$  jest funkcyjne  $\Leftrightarrow \Pi_X = \Pi_{X \cup Y} \Leftrightarrow |\Pi_X| = |\Pi_{X \cup Y}|$ .

Id	a	b	c	d
1	1	1	2	2
2	1	1	2	2
3	2	2	2	2
4	3	3	3	3
5	5	5	6	6
6	5	5	6	6

•  $\Pi_{\{ab\}} = \{\{1,2\}, \{3\}, \{4\}, \{5,6\}\}$   
 •  $\Pi_{\{cd\}} = \{\{1,2,3\}, \{4\}, \{5,6\}\}$   
 •  $\Pi_{\{abcd\}} = \{\{1,2\}, \{3\}, \{4\}, \{5,6\}\}$   
 Stąd:  
 •  $|\Pi_{\{ab\}}| = 4 = |\Pi_{\{abcd\}}|$ , więc  $\{ab\} \rightarrow \{cd\}$  jest zależnością funkcyjną.  
 •  $|\Pi_{\{cd\}}| = 3 \neq 4 = |\Pi_{\{abcd\}}|$ , więc  $\{cd\} \rightarrow \{ab\}$  nie jest zależnością funkcyjną.

15

**TANE: wyznaczanie podziałów**

• Własność.  
 $\Pi_{X \cup Y} = \Pi_X \cap \Pi_Y$

• W algorytmie TANE: Jeśli  $X = \{x_1, \dots, x_{n-1}, x_n\}$ , to  
 $\Pi_X = \Pi_{X \setminus \{x_n\}} \cap \Pi_{X \setminus \{x_{n-1}\}}$ .

16

**Podziały przystrzyżone**

•  $\widehat{\pi}_X$  jest nazywany *podziałem przystrzyżonym*, jeśli składa się tylko i wyłącznie ze wszystkich więcej niż 1-elementowych bloków w podziale  $\Pi_X$ .

• Przykład.  
 $\Pi_{\{ab\}} = \{\{1,2\}, \{3\}, \{4\}, \{5,6\}\}$ ,  
 $\widehat{\pi}_{\{ab\}} = \{\{1,2\}, \{5,6\}\}$ .

Id	a	b	c	d
1	1	1	2	2
2	1	1	2	2
5	5	5	6	6
6	5	5	6	6
3	2	2	2	2
4	3	3	3	3

17

**Zależności funkcyjne a podziały przystrzyżone...**

**Twierdzenie.**  
 $X \rightarrow Y$  jest zależnością funkcyjną  $\Leftrightarrow$   
 $|\Pi_X| = |\Pi_{X \cup Y}| \Leftrightarrow$   
 $||\widehat{\pi}_X|| - |\widehat{\pi}_X| = ||\widehat{\pi}_{X \cup Y}|| - |\widehat{\pi}_{X \cup Y}|$ ,  
 gdzie  
 $||\widehat{\pi}_Z|| = \sum_{c \in \widehat{\pi}_Z} |c|$ .

18

**Zależności funkcyjne a podziały przystrzyżone**

**Twierdzenie.**  $X \rightarrow Y$  jest zależnością funkcyjną  $\Leftrightarrow |\Pi_X| = |\Pi_{X \cup Y}| \Leftrightarrow ||\widehat{\pi_X}|| - |\widehat{\pi_X}| = ||\widehat{\pi_{X \cup Y}}|| - |\widehat{\pi_{X \cup Y}}|$ .

**Dowód:** Niech:

- $|D|$  oznacza liczbę obiektów,
- $n$  – oznacza liczbę jednoelementowych bloków w  $\Pi_X$ .

Wtedy,  $|\Pi_X| = |\widehat{\pi_X}| + n$  oraz  $n = |D| - |\widehat{\pi_X}|$ .

Zatem,  $|\Pi_X| = |\widehat{\pi_X}| + |D| - |\widehat{\pi_X}|$ .

oraz  $|\Pi_{X \cup Y}| = |\widehat{\pi_{X \cup Y}}| + |D| - |\widehat{\pi_{X \cup Y}}|$ .

Stąd  $|\Pi_X| = |\Pi_{X \cup Y}| \Leftrightarrow$

$$|\widehat{\pi_X}| + |D| - |\widehat{\pi_X}| = |\widehat{\pi_{X \cup Y}}| + |D| - |\widehat{\pi_{X \cup Y}}| \Leftrightarrow$$

$$||\widehat{\pi_X}|| - |\widehat{\pi_X}| = ||\widehat{\pi_{X \cup Y}}|| - |\widehat{\pi_{X \cup Y}}|.$$

19

**Przykład: Zależności funkcyjne a podziały przystrzyżone**

**Twierdzenie.**  $X \rightarrow Y$  jest zależnością funkcyjną  $\Leftrightarrow |\Pi_X| = |\Pi_{X \cup Y}| \Leftrightarrow ||\widehat{\pi_X}|| - |\widehat{\pi_X}| = ||\widehat{\pi_{X \cup Y}}|| - |\widehat{\pi_{X \cup Y}}|$ .

Id	a	b	c	d
1	1	1	2	2
2	1	1	2	2
3	2	2	2	2
4	3	3	3	3
5	5	5	6	6
6	5	5	6	6

- $\widehat{\pi_{\{ab\}}} = \{\{1,2\}, \{5,6\}\}$ ,  $\widehat{\pi_{\{cd\}}} = \{\{1,2,3\}, \{5,6\}\}$ .
- $\pi_{\{abcd\}} = \{\{1,2\}, \{5,6\}\}$ .

Stąd:

- $||\widehat{\pi_{\{ab\}}}| - |\widehat{\pi_{\{ab\}}}| = 4 - 2 = ||\widehat{\pi_{\{abcd\}}}| - |\widehat{\pi_{\{abcd\}}}|$ , czyli  $\{ab\} \rightarrow \{cd\}$  jest funkcyjne.
- $||\widehat{\pi_{\{cd\}}}| - |\widehat{\pi_{\{cd\}}}| = 5 - 2 \neq 4 - 2 = ||\widehat{\pi_{\{abcd\}}}| - |\widehat{\pi_{\{abcd\}}}|$ , czyli  $\{cd\} \rightarrow \{ab\}$  nie jest funkcyjne.

20

**Błąd zależności**

- $err(X \rightarrow Y) = \sum_{c \in \Pi_X} (|c| - \max\{|c'| \mid c' \subseteq c \wedge c' \in \Pi_{X \cup Y}\})$ .

Id	imię	pleć
1	Mary	F
2	Mary	F
3	Mary	M
4	Chris	M
5	Chris	M
6	Chris	F
7	Charlie	M
8	Charlie	F
9	Rob	M

$\Pi_{\{imię\}} = \{\{1,2,3\}, \{4,5,6\}, \{7,8\}, \{9\}\}$ .

$\Pi_{\{imię, pleć\}} = \{\{1,2\}, \{3\}, \{4,5\}, \{6\}, \{7\}, \{8\}, \{9\}\}$ .

$err(\{imię\} \rightarrow \{pleć\}) =$

$$(|\{1,2,3\}| - \max\{|\{1,2\}|, |\{3\}|\}) +$$

$$(|\{4,5,6\}| - \max\{|\{4,5\}|, |\{6\}|\}) +$$

$$(|\{7,8\}| - \max\{|\{7\}|, |\{8\}|\}) +$$

$$(|\{9\}| - \max\{|\{9\}|\}) =$$

$$(3 - 2) + (3 - 2) + (2 - 1) + (1 - 1) = 3.$$

21

**Zależności przybliżone**

- $X \rightarrow Y$  jest definiowana jako *zależność przybliżona*, jeśli  $err(X \rightarrow Y) \leq \epsilon$ , gdzie  $\epsilon$  jest wartością progową zadaną przez użytkownika.

Id	imię	pleć
1	Mary	F
2	Mary	F
3	Mary	M
4	Chris	M
5	Chris	M
6	Chris	F
7	Charlie	M
8	Charlie	F
9	Rob	M

Niech  $\epsilon = 3$ . Wtedy:

- $err(\{imię\} \rightarrow \{pleć\}) = 3 \leq \epsilon$ .
- $\{imię\} \rightarrow \{pleć\}$  jest zależnością przybliżoną.

22

**Błąd zależności a podziały przystrzyżone**

- **Własność.**  $err(X \rightarrow Y) = \sum_{c \in \widehat{\pi_X}} (|c| - \max\{|\{c'\}| \mid c' \in \widehat{\pi_{X \cup Y}} \wedge c' \subseteq c\} \cup \{1\}))$ .

Id	imię	pleć
1	Mary	F
2	Mary	F
3	Mary	M
4	Chris	M
5	Chris	M
6	Chris	F
7	Charlie	M
8	Charlie	F
9	Rob	M

$\widehat{\pi_{\{imię\}}} = \{\{1,2,3\}, \{4,5,6\}, \{7,8\}\}$ .

$\widehat{\pi_{\{imię, pleć\}}} = \{\{1,2\}, \{4,5\}\}$ .

$err(\{imię\} \rightarrow \{pleć\}) =$

$$(|\{1,2,3\}| - \max\{|\{1,2\}|, 1\}) +$$

$$(|\{4,5,6\}| - \max\{|\{4,5\}|, 1\}) +$$

$$(|\{7,8\}| - \max\{1\}) =$$


$$(3 - 2) + (3 - 2) + (2 - 1) = 3$$

23

**Adaptacja TANE do wyznaczania zależności przybliżonych**

- Sprawdzanie, czy  $X \rightarrow Y$  jest zależnością funkcyjną, jest zastąpione sprawdzaniem, czy  $err(X \rightarrow Y) \leq \epsilon$ .
- Reguła r2 jest używana tylko w przypadku, gdy  $err(X \rightarrow Y) = 0$ , czyli wtedy, gdy  $X \rightarrow Y$  jest zależnością funkcyjną.


24



### Literatura


- Ykä Huhtala, Juha Kärkkäinen, Pasi Porkka, Hannu Toivonen: TANE: An Efficient Algorithm for Discovering Functional and Approximate Dependencies. [Comput. J.](#) 42(2): 100-111 (1999)

25



### Ćwiczenia...

1. Wyznacz podziały  $\Pi_{\{a\}}$ ,  $\Pi_{\{b\}}$  i  $\Pi_{\{ab\}}$  na podstawie tabeli na slajdzie 10. Użyj tych podziałów do uzyskania odpowiedzi na następujące pytania:
  - Czy  $\{a\} \rightarrow \{b\}$  jest zależnością funkcyjną?
  - Czy  $\{b\} \rightarrow \{a\}$  jest zależnością funkcyjną?
  - Jaka jest wartość błędu zależności  $err(\{a\} \rightarrow \{b\})$ ?
  - Jaka jest wartość błędu zależności  $err(\{b\} \rightarrow \{a\})$ ?
  - Niech  $\varepsilon = 2$ . Czy  $\{a\} \rightarrow \{b\}$  jest zależnością przybliżoną?
  - Niech  $\varepsilon = 2$ . Czy  $\{b\} \rightarrow \{a\}$  jest zależnością przybliżoną?



### Ćwiczenia

2. Wyznacz podziały przyszyte  $\pi_{\{a\}}$ ,  $\pi_{\{b\}}$  i  $\pi_{\{ab\}}$  na podstawie tabeli na slajdzie 10. Użyj tych podziałów do uzyskania odpowiedzi na następujące pytania:
  - Czy  $\{a\} \rightarrow \{b\}$  jest zależnością funkcyjną?
  - Czy  $\{b\} \rightarrow \{a\}$  jest zależnością funkcyjną?
  - Jaka jest wartość błędu zależności  $err(\{a\} \rightarrow \{b\})$ ?
  - Jaka jest wartość błędu zależności  $err(\{b\} \rightarrow \{a\})$ ?
  - Niech  $\varepsilon = 1$ . Czy  $\{a\} \rightarrow \{b\}$  jest zależnością przybliżoną?
  - Niech  $\varepsilon = 1$ . Czy  $\{b\} \rightarrow \{a\}$  jest zależnością przybliżoną?