PLATFORMA PRZEMYSŁU 4.0

Metody reprezentacji reguł asocjacyjnych

Marzena Kryszkiewicz
Politechnika Warszawska

Odkrywanie silnych reguł asocjacyjnych

- Bezpośrednio:

Odkryj z danych częste zbiory (***F***) → Wyznacz silne reguły asocjacyjne (***AR***)

- Pośrednio:

Odkryj z danych zwięzłe reprezentacje częstych zbiorów → Wyznacz reprezentację (**R**) silnych reguł asocjacyjnych

Odkrywanie reprezentacji silnych reguł asocjacyjnych na podstawie reprezentacji zbiorów częstych

Częste generatory (*FG*)
Częste zbiory zamknięte (*FC*)
.....
→ Reprezentacja (*R*) silnych reguł ***AR*** |= ***AR***

Mechanizmy wnioskowania (|=):

- *Aksjomaty Armstronga* (*AA*)
- *Własność przechodniości zaufania* (ang. *confidence transitivity property, CTP*)
- *Operator pokrycia* (ang. *cover operator, C*)
- *Wnioskowanie regułowe domknięcie-domknięcie* (ang. *closure-closure rule inference, CCI*)

3

Klasy reprezentacji regułowych

Klasy reprezentacji regułowych

- *Reprezentacja regułowa* jest definiowana jako para (***R***, |=) , gdzie ***R*** jest podzbiorem ***AR***, a |= jest mechanizmem wnioskowania o regułach.

- (***R***, |=) jest definiowana jako *bezstratna,* jeśli ***R*** |= ***AR***.
- (***R***, |=) jest definiowana jako *poprawna,* jeśli ***R*** |≠ żadnej reguły *r* takiej, że *r*∉***AR***.
- (***R***, |=) jest definiowana jako (*ściśle*) *informacyjna,* jeśli dla każdej reguły *r* takiej, że ***R*** |= *r*, wartości jej wsparcia i zaufania są prawidłowo wyprowadzane.

5

Mechanizmy wnioskowania o regułach

Aksjomaty Armstronga (AA)

Odnoszą się tylko do *reguł pewnych, czyli* reguł o zaufaniu wynoszącym 100%:

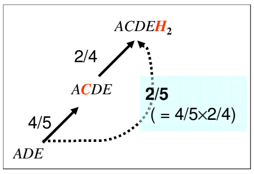
- Jeśli $X \supseteq Y$, to $conf(X \rightarrow Y) = 1$.
- Jeśli $conf(X \rightarrow Y) = 1$, to $conf(X \cup Z \rightarrow Y) = 1$.
- Jeśli $conf(X \rightarrow Y) = 1$ i $conf(Y \rightarrow Z) = 1$, to $conf(X \rightarrow Z) = 1$.

7

Własność przechodniości zaufania (CTP)

Niech $X \subset Y \subset Z$. Wtedy:

- $sup(X \rightarrow Z \setminus X) = sup(Y \rightarrow Z \setminus Y)$,
- $conf(X \rightarrow Z \setminus X) = conf(X \rightarrow Y \setminus X) \times conf(Y \rightarrow Z \setminus Y)$.



8

Operator pokrycia (C)

$$C(X \rightarrow Y) =$$
$$\{X \cup Z \rightarrow V \mid Z, V \subseteq Y \wedge Z \cap V = \emptyset \wedge V \neq \emptyset\}.$$

Id	*Transakcja*
T_1	{*abcde*}
T_2	{*abcdef*}
T_3	{*abcdehi*}
T_4	{*abe*}
T_5	{*bcdehi*}

#	reguła *r'* w *C*(*r*)	*sup*(*r'*)	*conf*(*r'*)
1.	**r: {*b*}→{*de*}**	**4**	**80%**
2.	{*b*}→{*d*}	4	80%
3.	{*b*}→{*e*}	5	100%
4.	{*bd*}→{*e*}	4	100%
5.	{*be*}→{*d*}	4	80%

9

Własności operatora pokrycia…

- Jeśli reguła $r' \in C(r)$, to $sup(r') \geq sup(r)$ i $conf(r') \geq conf(r)$.
- Reguła $(X' \rightarrow Z' \setminus X') \in C(X \rightarrow Z \setminus Y)$ wtedy i tylko wtedy $Z' \subseteq Z$ i $X' \supseteq X$.
- Jeśli reguła $r \in C(r')$ oraz reguła $r' \in C(r'')$, to $r \in C(r'')$.

#	rule *r'* in *C*(*r*)	*sup*(*r'*)	*conf*(*r'*)
1.	**r: {*b*}→{*de*}**	**4**	**80%**
2.	{*b*}→{*d*}	4	80%
3.	{*b*}→{*e*}	5	100%
4.	{*bd*}→{*e*}	4	100%
5.	{*be*}→{*d*}	4	80%

10

Własności operatora pokrycia

- Liczba reguł asocjacyjnych w pokryciu reguły $X \rightarrow Y$:

  $|C(X \rightarrow Y)| = 3^m - 2^m$, where $m = |Y|$.

#	rule *r'* in *C*(*r*)	*sup*(*r'*)	*conf*(*r'*)
1.	**r: {*b*}→{*de*}**	**4**	**80%**
2.	{*b*}→{*d*}	4	80%
3.	{*b*}→{*e*}	5	100%
4.	{*bd*}→{*e*}	4	100%
5.	{*be*}→{*d*}	4	80%

- $|C(\{b\} \rightarrow \{de\})| = 3^2 - 2^2 = 5$.
- $|C(\{ab\} \rightarrow \{cdefgh\})| = 3^6 - 2^6 = 665$.

11

Wnioskowanie regułowe *domknięcie-domknięcie* (CCI)

- $sup(X \rightarrow Y \setminus X) = sup(\gamma(Y))$.
- $conf(X \rightarrow Y \setminus X) = \frac{sup(\gamma(Y))}{sup(\gamma(X))}$.
- $(X \rightarrow Y \setminus X) \in$ ***AR***, jeśli $(\gamma(X) \rightarrow \gamma(Y) \setminus \gamma(X)) \in$ ***AR***.
- **Własność (wyznaczanie domknięcia zbioru):**
  Niech *X* będzie zbiorem pozycji. Domknięcie $\gamma(X)$ jest równe **najmniejszemu** (ze względu na zawieranie zbiorów) zbiorowi zamkniętemu zawierającemu zbiór *X*.

12

Reguły reprezentatywne

Reprezentatywne reguły asocjacyjne (RR)

- *Reprezentatywne reguły asocjacyjne* (krótko: *reguły reprezentatywne*, *RR*) są definiowane jako te silne reguły asocjacyjne, które nie należą do pokrycia żadnej innej silnej reguły asocyjacyjnej:

$$\boldsymbol{RR} = \{r \in \boldsymbol{AR} \mid \neg\exists r' \in \boldsymbol{AR}\ (r' \neq r \wedge r \in C(r'))\}.$$

- **Twierdzenie.** Jeśli reguła $X \rightarrow Z \backslash X \in \boldsymbol{RR}$, to jej poprzednik X jest częstym generatorem, a jej baza Z jest częstym zbiorem zamkniętym.

14

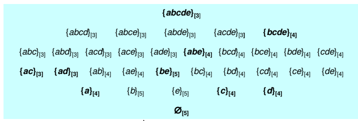
Własności reprezentacji regułowej (RR, C)

- Jeśli reguła $r \in \boldsymbol{RR}$, to jej pokrycie $C(r)$ zawiera wyłącznie silne reguły asocjacyjne (czyli $C(r) \subseteq \boldsymbol{AR}$).
- $\forall r \in \boldsymbol{AR}\ \exists$reguła $r' \in \boldsymbol{RR}$ taka, że $r \in C(r')$ oraz $sup(r) \geq sup(r')$ i $conf(r) \geq conf(r')$.
- Suma pokryć wszystkich reguł reprezentatywnych jest równa zbiorowi wszystkich silnych reguł asocjacyjnych (czyli $\boldsymbol{AR} = \bigcup_{r \in \boldsymbol{RR}} C(r)$).

- **Konkluzja:** (*RR*, *C*) jest poprawną i bezstratną reprezentacją regułową silnych reguł asocjacyjnych ***AR*** oraz umożliwia prawidłowe wyznaczanie pesymistycznego oszacowania ich wsparć i zaufań, ale nie gwarantuje wyznaczania ich wartości w sposób dokładny.

15

Przykład: Korzyści ze stosowania (RR, C)



→ 112 ***AR*s**

FC:
$\{abcde\}_{[3]}$
$\{bcde\}_{[4]}$
$\{abe\}_{[4]}$
$\{be\}_{[5]}$

\+

FG:
$\{ac\}_{[3]}$,
$\{ad\}_{[3]}$
$\{c\}_{[4]}$, $\{d\}_{[4]}$
$\{a\}_{[4]}$
$\emptyset_{[5]}$

→

4 ***RR*s**:
$\{ac\} \rightarrow \{bde\}$ [3, 1]
$\{ad\} \rightarrow \{bce\}$ [3, 1]
$\emptyset \rightarrow \{bcde\}$ [4, 4/5]
$\emptyset \rightarrow \{abe\}$ [4, 4/5]

16

Wnioskowanie z użyciem reguł reprezentatywnych (RR) i operatora pokrycia (C)

- Niech r będzie regułą, o której będziemy wnioskować z użyciem reprezentacji regułowej (***RR***, ***C***).
- Niech $\boldsymbol{R}$ będzie podzbiorem zbioru tych wszystkich reguł w ***RR***, których pokrycia zawierają r.

- Jeśli r nie należy do pokrycia żadnej reguły z ***RR*** (czyli, jeśli $\boldsymbol{R} = \emptyset$), to $r \notin \boldsymbol{AR}$.
- Jeśli r należy do pokrycia co najmniej jednej reguły z ***RR*** (czyli, jeśli $\boldsymbol{R} \neq \emptyset$), to $r \in \boldsymbol{AR}$ oraz
  - $sup(r) \geq \max\{sup(r') \mid r' \in \boldsymbol{R}\}$;
  - $conf(r) \geq \max\{conf(r') \mid r' \in \boldsymbol{R}\}$.

17

Przykład: Wnioskowanie z użyciem (RR, C)...

Niech $\boldsymbol{RR} = \{ \{ac\} \rightarrow \{bde\}$ [3, 1], $\{ad\} \rightarrow \{bce\}$ [3, 1], $\emptyset \rightarrow \{bcde\}$ [4, 4/5], $\emptyset \rightarrow \{abe\}$ [4, 4/5]}.

Pytania:
- Czy reguła r: $\{ab\} \rightarrow \{e\} \in \boldsymbol{AR}$?
- Jeśli tak, to jak oszacować jej wsparcie i zaufanie?

Odpowiedź:
- Reguła r należy do pokrycia następującej reguły z ***RR***:

$\emptyset \rightarrow \{abe\}$ [4, 4/5].

Stąd, $r \in \boldsymbol{AR}$ oraz $sup(r) \geq 4$ i $conf(r) \geq 4/5$.

18

Przykład: Wnioskowanie z użyciem (RR, C)

Niech $\boldsymbol{RR}$ = { {*ac*}→{*bde*} [3, 1], {*ad*}→{*bce*} [3, 1], ∅→{*bcde*} [4, 4/5], ∅→{*abe*} [4, 4/5]}.

Pytanie:

- Czy reguła *r*: {*b*}→{*f*} ∈ ***AR***?

Odpowiedź:

- Reguła *r* nie należy do pokrycia żadnej reguły z ***RR***. A zatem, *r* ∉ ***AR***.

Reguły minimalne nieredundantne

Minimalne nieredundantne reguły (MNR)

- *Minimalne nieredundantne reguły* (*MNR*) są definiowane jako te silne reguły asocjacyjne, które nie należą do pokrycia żadnej innej silnej reguły asocjacyjnej o takim samym wsparciu i takim samym zaufaniu:

$$\boldsymbol{MNR} = \{r \in \boldsymbol{AR} \mid \neg\exists r' \in \boldsymbol{AR}\ (r' \neq r \wedge r \in C(r') \wedge sup(r') = sup(r) \wedge conf(r') = conf(r))\}.$$

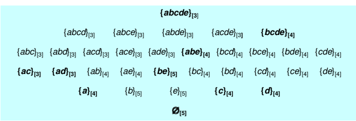
- **Twierdzenie.** Jeśli reguła $X \rightarrow Z \setminus X \in \boldsymbol{MNR}$, to jej poprzednik *X* jest częstym generatorem, a jej baza *Z* jest częstym zbiorem zamkniętym oraz:

$$\boldsymbol{MNR} = \{X \rightarrow Z \setminus X \mid Z \in \boldsymbol{FC} \wedge X \in \boldsymbol{FG} \wedge X \subset Z \wedge conf(X \rightarrow Z \setminus X) > minConf\}.$$

Własności reprezentacji regułowej (MNR, C)

- Jeśli reguła *r* ∈ ***MNR***, to jej pokrycie *C*(*r*) zawiera wyłącznie silne reguły asocjacyjne (czyli $C(r) \subseteq \boldsymbol{AR}$).
- ∀*r* ∈ ***AR*** ∃reguła *r′* ∈ ***MNR*** taka, że *r* ∈ *C*(*r′*) oraz *sup*(*r*) ≥ *sup*(*r′*) i *conf*(*r*) ≥ *conf*(*r′*).
- $\boldsymbol{AR} = \bigcup_{r \in \boldsymbol{MNR}} C(r)$.
- **Konkluzja:** (***MNR***, *C*) jest poprawną i bezstratną reprezentacją regułową silnych reguł asocjacyjnych ***AR*** oraz umożliwia prawidłowe wyznaczanie dokładnych wartości ich wsparć i zaufań (czyli jest ściśle informacyjna).

Przykład: Korzyści ze stosowania (MNR, C)



112 ***AR***s

FC:
{*abcde*}$_{[3]}$
{*bcde*}$_{[4]}$
{*abe*}$_{[4]}$
{*be*}$_{[5]}$

+

FG:
{*ac*}$_{[3]}$,
{*ad*}$_{[3]}$
{*c*}$_{[4]}$,
{*d*}$_{[4]}$
{*a*}$_{[4]}$
∅$_{[5]}$

→

8 ***MNR***s:
{*ac*}→{*bde*} [3, 1]
{*ad*}→{*bce*} [3, 1]
∅→{*bcde*} [4, 4/5]
{*c*}→{*bde*} [4, 1]
{*d*}→{*bce*} [4, 1]
∅→{*abe*} [4, 4/5]
{*a*} →{*be*} [4, 1]
∅→{*be*} [5, 1]

Wnioskowanie z użyciem minimalnych nieredundantnych reguł (MNR) i operatora pokrycia (C)

- Niech *r* będzie regułą, o której będziemy wnioskować z użyciem reprezentacji regułowej (***MNR***, ***C***).
- Niech ***R*** będzie podzbiorem zbioru tych wszystkich reguł w ***MNR***, których pokrycia zawierają *r*.

- Jeśli *r* nie należy do pokrycia żadnej reguły z ***MNR*** (czyli, jeśli ***R*** = ∅), to *r* ∉ ***AR***.
- Jeśli *r* należy do pokrycia co najmniej jednej reguły z ***MNR*** (czyli, jeśli ***R*** ≠ ∅), to *r* ∈ ***AR*** oraz
  - *sup*(*r*) = max{*sup*(*r′*) | *r′* ∈ ***R***};
  - *conf*(*r*) = max{*conf*(*r′*) | *r′* ∈ ***R***}.

Przykład: Wnioskowanie z użyciem (MNR, C)…

Niech ***MNR*** = { {*ac*}→{*bde*} [3, 1], {*ad*}→{*bce*} [3, 1],
∅→{*bcde*} [4, 4/5], {*c*}→{*bde*} [4, 1],
{*d*}→{*bce*} [4, 1], ∅→{*abe*} [4, 4/5],
{*a*}→{*be*} [4, 1], ∅→{*be*} [5, 1]}.

Pytania:

- Czy reguła *r*: {*ab*}→{*e*} ∈ ***AR***?
- Jeśli tak, to jak wyznaczyć jej wsparcie i zaufanie?

Odpowiedź:

- Reguła *r* należy do pokrycia następujących dwóch reguł z ***MNR***:

∅→{*abe*} [4, 4/5] oraz {*a*}→{*be*} [4, 1].

Stąd, *r* ∈ ***AR*** oraz *sup*(*r*) = 4 i *conf*(*r*) = 1.

25

Przykład: Wnioskowanie z użyciem (MNR, C)

Niech ***MNR*** = { {*ac*}→{*bde*} [3, 1], {*ad*}→{*bce*} [3, 1],
∅→{*bcde*} [4, 4/5], {*c*}→{*bde*} [4, 1],
{*d*}→{*bce*} [4, 1], ∅→{*abe*} [4, 4/5],
{*a*}→{*be*} [4, 1], ∅→{*be*} [5, 1]}.

Pytanie:

- Czy reguła *r*: {*b*}→{*f*} ∈ ***AR***?

Odpowiedź:

- Reguła *r* nie należy do pokrycia żadnej reguły z ***RR***. A zatem, *r* ∉ ***AR***.

26

Reguły reprezentatywne versus minimalne nieredundantne

RR versus MNR…

$\boldsymbol{RR} = \{r \in \boldsymbol{AR} \mid \neg\exists r' \in \boldsymbol{AR}\ (r' \neq r \wedge r \in C(r'))\}.$

$\boldsymbol{MNR} = \{r \in \boldsymbol{AR} \mid \neg\exists r' \in \boldsymbol{AR}\ (r' \neq r \wedge r \in C(r') \wedge sup(r') = sup(r) \wedge conf(r') = conf(r))\}.$

⇕

$\boldsymbol{RR} = \{r \in \boldsymbol{AR} \mid \forall r' \in \boldsymbol{AR}\ (r' = r \vee r \notin C(r'))\}.$

$\boldsymbol{MNR} = \{r \in \boldsymbol{AR} \mid \forall r' \in \boldsymbol{AR}\ (r' = r \vee r \notin C(r') \vee sup(r') \neq sup(r) \vee conf(r') \neq conf(r))\}.$

28

RR versus MNR

- ***RR*** ⊆ ***MNR***.
- ***RR*** = {*r*∈***MNR*** | ¬∃*r*′∈***MNR*** (*r*′≠*r* ∧ *r*∈*C*(*r*′))}.

4 ***RR***s:
{*ac*}→{*bde*} [3, 1]
{*ad*}→{*bce*} [3, 1]
∅→{*bcde*} [4, 4/5]
∅→{*abe*} [4, 4/5]

⊆

8 ***MNR***s:
{*ac*}→{*bde*} [3, 1]
{*ad*}→{*bce*} [3, 1]
∅→{*bcde*} [4, 4/5]
{*c*}→{*bde*} [4, 1]
{*d*}→{*bce*} [4, 1]
∅→{*abe*} [4, 4/5]
{*a*} →{*be*} [4, 1]
∅→{*be*} [5, 1]

29

Literatura…

- Yves Bastide, Nicolas Pasquier, Rafik Taouil, Gerd Stumme, Lotfi Lakhal: Mining Minimal Non-redundant Association Rules Using Frequent Closed Itemsets. Computational Logic 2000: 972-986
- Marzena Kryszkiewicz: Representative Association Rules. PAKDD 1998: 198-209
- Marzena Kryszkiewicz: Representative Association Rules and Minimum Condition Maximum Consequence Association Rules. PKDD 1998: 361-369
- Marzena Kryszkiewicz: Concise Representations of Frequent Patterns and Association Rules, Warsaw: Publishing House of Warsaw University of Technology (2002)
- Marzena Kryszkiewicz: Concise Representations of Association Rules. Pattern Detection and Discovery 2002: 92-109

30

Literatura

- Nicolas Pasquier, Yves Bastide, Rafik Taouil, Lotfi Lakhal: Closed Set Based Discovery of Small Covers for Association Rules. Proc. 15èmes Journées Bases de Données Avancées, BDA 1999: 361-381
- Mohammed Javeed Zaki: Generating non-redundant association rules. KDD 2000: 34-43

31

Literatura dodatkowa – reprezentacje reguł spełniających różnorakie kryteria kanoniczne

- Marzena Kryszkiewicz: A Lossless Representation for Association Rules Satisfying Multiple Evaluation Criteria. ACIIDS (2) 2016: 147-158
- Marzena Kryszkiewicz: Representative Rule Templates for Association Rules Satisfying Multiple Canonical Evaluation Criteria. ACIIDS (1) 2018: 550-561

32

Literatura dodatkowa – inne mechanizmy wyprowadzania reguł, w tym z danych niepełnych

- Marzena Kryszkiewicz, Henryk Rybinski: Incomplete Database Issues for Representative Association Rules. ISMIS 1999: 583-591
- Marzena Kryszkiewicz: Inducing Theory for the Rule Set. Rough Sets and Current Trends in Computing 2000: 391-398
- Marzena Kryszkiewicz: Mining with Cover and Extension Operators. PKDD 2000: 476-482
- Marzena Kryszkiewicz: Inferring Knowledge from Frequent Patterns. Soft-Ware 2002: 247-262
- Marzena Kryszkiewicz: Closed Set Based Discovery of Maximal Covering Rules. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 11 (Supplement-1): 15-30 (2003)

33

Ćwiczenia…

1. Wyznacz liczbę reguł asocjacyjnych należących do pokryć reguł: $\{ac\}\rightarrow\{bde\}$ oraz $\emptyset\rightarrow\{bcde\}$:
   - $|C(\{ac\}\rightarrow\{bde\})|$ =
   - $|C(\emptyset\rightarrow\{bcde\})|$ =
2. Wyznacz pokrycie reguły $\{ac\}\rightarrow\{bde\}$:
   - $C(\{ac\}\rightarrow\{bde\})$ =

34

Ćwiczenia

3. Rozważ reguły reprezentatywne ***RR*** ze slajdu 18. Dla każdej reguły $r \in \{\{a\}\rightarrow\{bd\}, \emptyset\rightarrow\{bd\}, \{acd\}\rightarrow\{e\}\}$, udziel odpowiedzi na poniższe pytania:
   - Które z reguł ***RR*** zawierają w swoim pokryciu regułę *r*?
   - Czy reguła $r \in$ ***AR***?
   - Jeśli tak, to oszacuj wsparcie i zaufanie reguły *r*.
4. Rozważ reguły minimalne nieredundantne ***MNR*** ze slajdu 25. Dla każdej reguły $r \in \{\{a\}\rightarrow\{bd\}, \emptyset\rightarrow\{bd\}, \{acd\}\rightarrow\{e\}\}$, udziel odpowiedzi na poniższe pytania:
   - Które z reguł ***MNR*** zawierają w swoim pokryciu regułę *r*?
   - Czy reguła $r \in$ ***AR***?
   - Jeśli tak, to oszacuj wsparcie i zaufanie reguły *r*.

35