WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME

Concise Representations of Association Rules

HUMAN CAPITAL

Project is co-financed by European Union within European Social Fund

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME

Discovering Association Rules

- Indirect:

Discover frequent itemsets (***F***) from data → Calculate ***AR***

- Direct:

Discover concise representation of ***F*** from data → Calculate representation of ***AR***

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME

Discovering AR Representation from Representation(s) of F

Frequent generators (*FG*)
Frequent closed itemsets *FC*)
.....

→ Representation of ***AR*** |= ***AR***

Inference mechanisms (|=):

- *Armstrong Axioms* (*AA*)
- *confidence transitivity property* (*CTP*)
- *cover operator* (*C*)
- *closure-closure rule inference* (*CCI*)

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME

Rule Representations Classes

Let ***R*** be a subset of ***AR***.

- (***R***, |=) is defined *lossless* if ***R*** |= ***AR***.
- (***R***, |=) is defined *sound* if ***R*** $|\neq r$ such that $r \notin$ ***AR***.
- (***R***, |=) is defined *informative* if for each rule *r* such that ***R*** |= *r*, confidence and support of *r* is determined correctly.

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME

Armstrong Axioms (AA)

Relate only to certain rules:

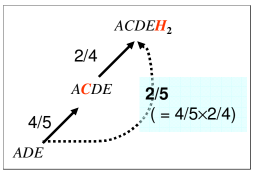
- If $X \supseteq Y$, then $conf(X \to Y) = 1$.
- If $conf(X \to Y) = 1$, then $conf(X \cup Z \to Y) = 1$.
- If $conf(X \to Y) = 1$ and $conf(Y \to Z) = 1$, then $conf(X \to Z) = 1$.

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME

Confidence Transitivity Property (CTP)

Let $X \subset Y \subset Z$. Then:

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



Cover Operator (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	*Transaction*
T_1	$\{abcde\}$
T_2	$\{abcdef\}$
T_3	$\{abcdehi\}$
T_4	$\{abe\}$
T_5	$\{bcdehi\}$

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

7

Properties of Cover Operator…

- If $r' \in C(r)$, then $sup(r') \geq sup(r)$ & $conf(r') \geq conf(r)$.
- $X' \rightarrow Z' \setminus X' \in C(X \rightarrow Z \setminus Y)$ if and only if $Z' \subseteq Z$ & $X' \supseteq X$.
- If $r \in C(r')$ & $r' \in C(r'')$, then $r \in C(r'')$.

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

8

Properties of Cover Operator

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

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

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

9

Closure-Closure Rule Inference (CCI)

- $sup(X \rightarrow Y \setminus X) = sup(\gamma(Y))$.
- $conf(X \rightarrow Y \setminus X) = sup(\gamma(Y)) / sup(\gamma(X))$.
- $(X \rightarrow Y \setminus X) \in$ ***AR*** if $(\gamma(X) \rightarrow \gamma(Y) \setminus \gamma(X)) \in$ ***AR***.
- **Property (closure determination):**
  Let *X* be an itemset. Closure $\gamma(X)$ is equal to **the smallest** (wrt. set inclusion) closed itemset containing *X*.

10

Representative Rules (RR, C)

- *Representative rules* (**RR**) are those strong association rules that are not covered by other strong association rules:

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

11

Properties of (RR, C)

- If $X \rightarrow Z \setminus X \in$ ***RR***, then $X \in$ ***FG*** and $Z \in$ ***FC***.
- If $r \in$ ***RR***, then $C(r) \subseteq$ ***AR***.
- $\forall r \in$ ***AR*** $\exists r' \in$ ***RR*** such that $r \in C(r')$.
- ***AR*** $= \bigcup_{r \in \textbf{\textit{RR}}} C(r)$.
- **Conclusion:** (***RR***, *C*) is sound and lossless representation of ***AR***, though, in general, it is not informative.

12

Benefits of applying (RR, C)

{*abcde*}$_{[3]}$

{*abcd*}$_{[3]}$ {*abce*}$_{[3]}$ {*abde*}$_{[3]}$ {*acde*}$_{[3]}$ **{*bcde*}$_{[4]}$**

{*abc*}$_{[3]}$ {*abd*}$_{[3]}$ {*acd*}$_{[3]}$ {*ace*}$_{[3]}$ {*ade*}$_{[3]}$ **{*abe*}$_{[4]}$** {*bcd*}$_{[4]}$ {*bce*}$_{[4]}$ {*bde*}$_{[4]}$ {*cde*}$_{[4]}$

{*ac*}$_{[3]}$ **{*ad*}$_{[3]}$** {*ab*}$_{[4]}$ {*ae*}$_{[4]}$ **{*be*}$_{[5]}$** {*bc*}$_{[4]}$ {*bd*}$_{[4]}$ {*cd*}$_{[4]}$ {*ce*}$_{[4]}$ {*de*}$_{[4]}$

{*a*}$_{[4]}$ {*b*}$_{[5]}$ {*e*}$_{[5]}$ **{*c*}$_{[4]}$** **{*d*}$_{[4]}$**

$\emptyset_{[5]}$

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

Reasoning with (RR, C)

Let *r* be a rule to be evaluated and **R** be a subset of rules in **RR** that cover *r*.

- If *r* is not covered by any rule in **RR** (that is, if **R** = ∅), then *r* ∉ **AR**.
- If *r* is covered by at least one rule in **RR** (that is, if **R** ≠ ∅), then *r* ∈ **AR** and
  - *sup*(*r*) ≥ max{*sup*(*r'*)| *r'* ∈ **R**};
  - *conf*(*r*) ≥ max{*conf*(*r'*)| *r'* ∈ **R**}.

Example: Reasoning with (RR, C)…

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

Task: *r*: {*ab*}→{*e*} ∈ **AR**?

Solution: *r* is covered by the following rule in RR: ∅→{*abe*} [4, 4/5].

Hence, *r* ∈ **AR** and *sup*(*r*) ≥ 4 and *conf*(*r*) ≥ 4/5.

Example: Reasoning with (RR, C)

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

Task: *r*: {*b*}→{*f*} ∈ **AR**?

Solution: *r* is not covered by any rule in **RR**.

Hence, *r* ≠ **AR**.

Minimal Non-Redundant Rules (MNR, C)

- *Minimal non-redundant rules* (**MNR**) are those strong association rules each of which is not covered by another strong association rule with the same support and confidence:

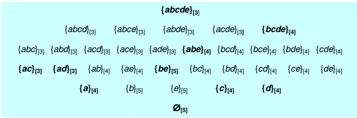
MNR = {*r* ∈ ***AR*** | ¬∃*r'* ∈ ***AR*** (*r'*≠*r* ∧ *r*∈ *C*(*r'*) ∧ *sup*(*r'*) = *sup*(*r*) ∧ *conf*(*r'*) = *conf*(*r*))}.

Properties of (MNR, C)

- ***MNR*** = {*X*→*Y**X* | *Y*∈ ***FC*** ∧ *X*∈ ***FG*** ∧ *X* ⊂ *Y* ∧ *conf*(*X*→*Y**X*) > *minConf*}.
- If *r* ∈ ***MNR***, then *C*(*r*) ⊆ ***AR***.
- ∀*r* ∈ ***AR*** ∃*r'* ∈ ***MNR*** such that *r*∈ *C*(*r'*).
- ***AR*** = $\cup_{r\in \boldsymbol{MNR}}$ *C*(*r*).
- **Conclusion:** (***MNR***, *C*) is sound, lossless and informative representation of ***AR***.

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME

Benefits of applying (MNR, C)



112 ***AR*s**

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

\+

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

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

19

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME

Reasoning with (MNR, C)

Let r be a rule to be evaluated and **R** be a subset of rules in **MNR** that cover r.

- If r is not covered by any rule in **MNR** (that is, if **R** $= \emptyset$), then $r \notin$ **AR**.
- If r is covered by at least one rule in **MNR** (that is, if **R** $\neq \emptyset$), then $r \in$ **AR** and
  - $sup(r) = \max\{sup(r') \mid r' \in \mathbf{R}\}$;
  - $conf(r) = \max\{conf(r') \mid r' \in \mathbf{R}\}$.

20

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME

Example: Reasoning with (MNR, C)…

Let MNR = { $\{ac\}\rightarrow\{bde\}$ [3, 1], $\{ad\}\rightarrow\{bce\}$ [3, 1], $\emptyset\rightarrow\{bcde\}$ [4, 4/5], $\{c\}\rightarrow\{bde\}$ [4, 1], $\{d\}\rightarrow\{bce\}$ [4, 1], $\emptyset\rightarrow\{abe\}$ [4, 4/5], $\{a\}\rightarrow\{be\}$ [4, 1], $\emptyset\rightarrow\{be\}$ [5, 1]}.

Task: r: $\{ab\}\rightarrow\{e\} \in$ **AR**?

Solution: r is covered by the following two rules in **MNR**: $\{a\}\rightarrow\{be\}$ [4, 1], $\emptyset\rightarrow\{abe\}$ [4, 4/5].

Hence, $r \in$ **AR** and $sup(r) = 4$ and $conf(r) = 1$.

21

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME

Example: Reasoning with (MNR, C)

Let MNR = { $\{ac\}\rightarrow\{bde\}$ [3, 1], $\{ad\}\rightarrow\{bce\}$ [3, 1], $\emptyset\rightarrow\{bcde\}$ [4, 4/5], $\{c\}\rightarrow\{bde\}$ [4, 1], $\{d\}\rightarrow\{bce\}$ [4, 1], $\emptyset\rightarrow\{abe\}$ [4, 4/5], $\{a\}\rightarrow\{be\}$ [4, 1], $\emptyset\rightarrow\{be\}$ [5, 1]}.

Task: r: $\{b\}\rightarrow\{f\} \in$ **AR**?

Solution: r is not covered by any rule in **MNR**.

Hence, $r \neq$ **AR**.

22

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME

RR versus MNR

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

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

$\Updownarrow$

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

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

23

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME

RR versus MNR

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

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]

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

24

References…

– Marzena Kryszkiewicz: Representative Association Rules and Minimum Condition Maximum Consequence Association Rules. PKDD 1998: 361-369
– Marzena Kryszkiewicz: Closed Set Based Discovery of Representative Association Rules. IDA 2001: 350-359
– Marzena Kryszkiewicz: Concise Representations of Frequent Patterns and Association Rules, Warsaw: Publishing House of Warsaw University of Technology (2002)

References

– 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
– 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