



A Lossless Representation for Association Rules Satisfying Multiple Evaluation Criteria

Marzena Kryszkiewicz Institute of Computer Science Warsaw University of Technology

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME



Rules

A *rule* is an expression associating two itemsets:

$$X \rightarrow Y$$

where:

- X is called an *antecedent* of $X \rightarrow Y$.
- Y is called a *consequent* of $X \rightarrow Y$.
- $X \cup Y$ is called the *base* of $X \rightarrow Y$.

2





Association Rules

 A rule X → Y is defined as an association rule if its antecedent and consequent are mutually exclusive:

$$X \cap Y = \emptyset$$

and
 $Y \neq \emptyset^*$.

* In some definitions of association rules, the condition $Y \neq \emptyset$ does not occur.





Standard Evaluation Measures of (Association) Rules

There are two basic measures for evaluating (association) rules:

- · support (or relative support) and
- · confidence

4

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME



Support of a Rule

 Support of X → Y is defined as the number of transactions that contain the base of X → Y; that is,

$$sup(X \rightarrow Y) = sup(X \cup Y)$$
.

• Relative support of $X \rightarrow Y$ is defined as the relative support of the base of $X \rightarrow Y$:

$$rSup(X \rightarrow Y) = rSup(X \cup Y) = P(XY)$$





Confidence of a Rule

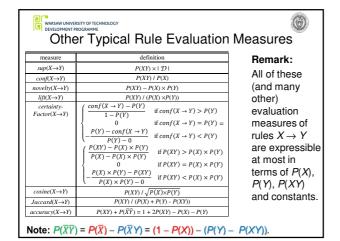
 Confidence of X → Y is defined as the ratio of the number of transactions that contain the base X ∪ Y to the number of transactions containing the antecedent X:

$$conf(X \rightarrow Y) = sup(X \rightarrow Y) / sup(X)$$

 Remark: conf(X → Y) can be regarded as the conditional probability that Y occurs in a transaction T provided X occurs in T:

$$conf(X \rightarrow Y) = P(Y|X) = P(XY) / P(X).$$

1







ACBC-Evaluation Measures of Rules

An evaluation measure of a rule $X \rightarrow Y$ is defined as an *ACBC-evaluation measure* if it can be expressed in terms of at most the following components:

- the probability *P*(*X*) of rule's *Antecedent X*,
- the probability P(Y) of rule's Consequent Y,
- the probability P(Z) of its Base $Z = X \cup Y$,
- Constants.

WARSAW UNIVERSITY OF TECHNOLOGY
DEVELOPMENT PROGRAMME



Property of ACBC-Evaluation Measures

Proposition. Let:

- $X \rightarrow Y$ and $Z \rightarrow V$ be rules such that P(X) = P(Z), P(Y) = P(V), P(XY) = P(ZV) and
- μ be an ACBC-evaluation measure. Then:

$$\mu(X \to Y) = \mu(Z \to V).$$





Rules Satisfying Multiple Evaluation Criteria

Let:

 M be a set {μ₁, ..., μ_n} of ACBC-evaluation measures

and

• $E = \{\varepsilon_1, ..., \varepsilon_n\}$ be a set of corresponding threshold values.

A rule $X \to Y$ is defined as an (M, E)-rule if: $\forall \mu_i \in M (\mu_i(X \to Y) > \varepsilon_i).$

10

WARSAW UNIVERSITY OF TECHNOLOGY
DEVELOPMENT PROGRAMME



Representing Rules Satisfying Multiple Evaluation Criteria

- The number of rules satisfying multiple evaluation criteria may be huge.
- Rule templates can be used as a concise lossless representation of association rules satisfying multiple ACBC-evaluation criteria.
- Rule templates are based on two types of itemsets:
 - closed itemsets and
 - key generators.







- Closure of an itemset X with non-zero support is defined as the intersection of all transactions containing X.
- Property. The closure of X is the greatest superset of X that occurs in the same (number of) transactions as X.
- An itemset X is called closed if X is its own closure; that is, if:

$$\forall Y \supset X, t(Y) \subset t(X)$$

(or equivalently, $\forall Y \supset X$, sup(Y) < sup(X)).





Important Property of Closed Itemsets

Let X be a an itemset and Y be a closed itemset being the closure of X. Then:

- X and Y occur in the same transactions.
- X and Y have the same support.
- Each itemset U such that $X \subseteq U \subseteq Y$ has Yas its closure (and thus, has the same support as X and Y).

Key Generators



- An itemset Y is called a *generator of X* if Y is a minimal subset of X such that X and Y occur in the same transactions.
- Property. Each itemset has one or more generators.
- An itemset X is called a *key generator* if X has itself as its (only) generator; that is, if

$$\forall Y \subset X, t(Y) \supset t(X)$$

(or equivalently, $\forall Y \subset X$, sup(Y) > sup(X)).

Example: Reasoning with Closed Itemsets and Key Generators

Transaction Id	Transaction		
#1	{abcde}		
#2	{abcdef}		
#3	{a <mark>bcde</mark> hi}		
#4	{abe}		
#5	{bcde <mark>hi</mark> }		

- {h} is a key generator.
- {bcdehi} is a closed itemset, which is the closure of {h}.
- Each itemset U such that $\{h\} \subseteq U \subseteq \{bcdehi\}$ has the same support as {h} and {bcdehi} (that is, 2). There are 32 such itemsets U.





Rule Templates

Any pair of two rules $(X \rightarrow Y, Z \rightarrow V)$ is defined a rule template if:

- X and Y are key generators, $X \cap Y = \emptyset$,
- Z and V are closed itemsets.
- Z is the closure of X and V is the closure of Y.

 $X \rightarrow Y$ is called a *lower rule* of $(X \rightarrow Y, Z \rightarrow V)$.

 $Z \rightarrow V$ is called an upper rule of $(X \rightarrow Y, Z \rightarrow V)$.



Rule Templates and their Properties...

Theorem. Let $(X \rightarrow Y, Z \rightarrow V)$ be a rule template. Then:

- a) t(X) = t(Z) and P(X) = P(Z).
- b) t(Y) = t(V) and P(Y) = P(V).
- c) t(XY) = t(ZV) and P(XY) = P(ZV).
- d) For any ACBC-evaluation measure $\mu: \mu(X \to Y) = \mu(Z \to V)$.

Proof: a) and b) comes from the fact that Z is the closure of Xand V is the closure of Y.

- b) $t(XY) = t(X) \cap t(Y) = /*$ by a) and b) $*/ = t(Z) \cap t(V) = t(ZV)$. Hence, P(XY) = P(ZV).
- c) Follows immediately by definition of an ACBC-measure and a), b) and c).





Rule Templates and their Properties...

A rule $U \rightarrow W$ is defined as covered by rule template $(X \rightarrow Y, Z \rightarrow V)$ if $X \subseteq U \subseteq Z$ and $Y \subseteq W \subseteq V$.

Theorem. Let $(X \rightarrow Y, Z \rightarrow V)$ be a rule template covering rule $U \rightarrow W$. Then:

- P(X) = P(U) = P(Z).
- P(Y) = P(W) = P(V).
- P(XY) = P(UW) = P(ZV).
- For any ACBC-evaluation measure μ:

$$\mu(X \to Y) = \mu(U \to W) = \mu(Z \to V).$$





Coverage Power of a Rule Template

Theorem. A rule template $(X \rightarrow Y, Z \rightarrow V)$ covers

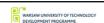
- $2^m \times 2^n \times 3^k$ distinct association rules, where: $\Delta_{Common} = (Z \cap V) \setminus (X \cup Y),$
- $\Delta_{\text{antecedent}} = Z \setminus (X \cup V);$
- $\Delta_{\text{consequent}} = V \setminus (Y \cup Z);$
- $k = |\Delta_{Common}|$; $m = |\Delta_{antecedent}|$; $n = |\Delta_{consequent}|$.

Example. Let *R* denote a rule template:

$$(\{a\} \rightarrow \{ij\}, \{abcdef\} \rightarrow \{efghij\}).$$

Then: $\Delta_{Common} = \{ef\}, \Delta_{antecedent} = \{bcd\}, \Delta_{consequent} = \{gh\}.$ So, rule template R covers $2^{\left|\frac{bcd}{a}\right|} \times 2^{\left|\frac{gh}{a}\right|} \times 3^{\left|\frac{ef}{a}\right|} =$

 $2^3 \times 2^2 \times 3^2 = 288$ association rules.





(M, E)-Rule Template

• Let M be a set of ACBC-evaluation measures. Then, a rule template $(X \rightarrow Y, Z \rightarrow V)$ is called an (M, E)-rule template if:

$$\forall \mu_i \in \mathsf{M}, \ \varepsilon_i \in E (\mu_i(X \to Y) > \varepsilon_i)$$
 (or eq., $\forall \mu_i \in \mathsf{M}, \ \varepsilon_i \in E \ (\mu_i(Z \to V) > \varepsilon_i)$).





(M, E)-Rule Templates as a Representation of Association Rules Satisfying Multiple ACBC-Criteria

Theorem. Let M be a set of ACBC-evaluation measures and $U \rightarrow W$ be an association rule.

If there is an (M, E)-rule template $(X \rightarrow Y, Z \rightarrow V)$ covering $U \rightarrow W$, then $U \rightarrow W$ is an (M, E)-association rule and

$$\forall \mu_i \in M (\mu_i(U \rightarrow W) = \mu_i(X \rightarrow Y) = \mu_i(Z \rightarrow V)).$$

Otherwise, $U \rightarrow W$ is not an (M, E)-association rule.



Example: (M, E)-Rule Templates as a Representation of (M, E)-Association Rules

xam	ole dataset D	(M, E)-rule template	sup	novelty	lift	certaintyFactor
TId	Transaction	$(\{h\} \rightarrow \{i\}, \{bcdehi\} \rightarrow \{bcdehi\})$	2	0.24	2.5	1
#1	{abcde}	$(\{h\} \rightarrow \{ai\}, \{bcdehi\} \rightarrow \{abcdehi\})$	1	0.12	2.5	0.375
#2	{abcdef}	$(\{i\} \rightarrow \{h\}, \{bcdehi\} \rightarrow \{bcdehi\})$	2	0.24	2.5	1
#3	{abcdehi}	$(\{i\}\rightarrow \{ah\}, \{bcdehi\}\rightarrow \{abcdehi\})$	1	0.12	2.5	0.375
#4	{abe}	$(\{ah\} \rightarrow \{i\}, \{abcdehi\} \rightarrow \{bcdehi\})$	1	0.12	2.5	1
#5	{bcdehi}	$(\{ai\} \rightarrow \{h\}, \{abcdehi\} \rightarrow \{bcdehi\})$	1	0.12	2.5	1

Let the set of ACBC-evaluation measure

- M = {sup, novelty, lift, certaintyFactor} and
- the set of their corresponding threshold values $E = \{0, 0.1, 2.0, 0.3\}.$

Then, there are 6 (M, E)-rule templates, which cover 486 (M, E)-association rules.





Summary...

- It has been identified a wide generic class of ACBCrule evaluation measures, which can be formulated at most in terms of:
 - the probability of the Antecedent of a rule,
 - the probability of its Consequent,
 - the probability of its Base and
 - Constants.
- A lossless representation of association rules satisfying multiple evaluation criteria expressible in terms of ACBC-evaluation measures has been presented.





Summary...

- The representation is based on rule templates each of which consists of two rules:
 - a lower rule built from disjoint key generators and
 - an upper rule built from closed itemsets their
- · Each association rule that is covered by a given rule template has the same values of ACBC-evaluation measures as the covering rule template.



