Formal Concept Analysis

Part III

Radim BĚLOHLÁVEK

Dept. Computer Science
Palacky University, Olomouc
radim.belohlavek@acm.org

Attribute Implications and Related Topics

Introducing attribute implications

Attribute implications (AIs) are expressions describing particular dependencies among attributes in relational data.

Examples:

$$\{\text{prime}, > 2\} \Rightarrow \{\text{odd}\}, \quad \{\text{flight No.}\} \Rightarrow \{\text{depart. time}, \text{arriv. time}\}.$$

AIs used in

- **formal concept analysis**
  - interpreted in formal contexts (tables with yes/no-attributes)
  - knowledge extraction
- **relational databases** (called functional dependencies)
  - interpreted in DB relations (tables with general attributes)
  - data redundancy, normalization, DB design
  - knowledge extraction
- **data mining** (called association rules)
  - interpreted in formal contexts (tables with yes/no-attributes)
  - validity modified by confidence, support (interestingness)
  - knowledge extraction

Introducing attribute implications

basic literature:

- formal concept analysis
  - Ganter, Wille: Formal Concept Analysis. Mathematical Foundations. Springer, 1999.
  - Carpineto C., Romano G.: Concept Data Analysis. Wiley, 2004.
- relational databases
  - Any textbook on databases.
  - Maier D.: The Theory of Relational Databases. Computer Science Press, 1983.
- data mining (association rules)
  - Any textbook on Data Mining.
  - Zhang , Zhang: Association Rule Mining. Springer, 2002.

Introducing attribute implications

AIs are interpreted in tables (formal contexts) $\mathcal{T} = \langle X, Y, I \rangle$ such as

table $\mathcal{T}$

	y_1	y_2	y_3	y_4
x_1	×	×	×	×
x_2	×		×	×
x_3		×	×	×
x_4		×	×	×
x_5	×		×	

$X = \{x_1, \ldots\}$ . . . **objects** (rows)

$Y = \{y_1, \ldots\}$ . . . **attributes** (columns)

× . . . **incidence** (object has attribute)

attribute implication . . . $A \Rightarrow B$ where $A, B \subseteq Y$ (sets of attributes)

$A \Rightarrow B$ is true in table $\mathcal{T}$ means
for each object x:
IF x has all attributes from A THEN x has all attr. from B

Example: $\{y_1\} \Rightarrow \{y_3\}$, $\{y_2, y_3\} \Rightarrow \{y_4\}$ are true in $\mathcal{T}$,
$\{y_1\} \Rightarrow \{y_2\}$ is not (x_2 as a counterexample)

Introducing attribute implications

What are we going to do with attribute implications?

- define validity, entailment and related basic notions,
- complete systems for reasoning with attribute implications (deriving mechanically new AIs from established AIs),
- non-redundant bases: how to extract minimal fully informative set of AIs from data?
- relationships to concept lattices,
- algorithms for AIs.

Our approach to attribute implications: logical approach

- AIs are formulas (statements about data),
- AIs can be evaluated in formal contexts, formal contexts (and rows of formal contexts) are our semantical structures,
- this brings us to ordinary logical framework where we can address entailment and further standard logical notions.

AIs – basic notions

Definition (attribute implication)

Let Y be a non-empty set (of attributes). An attribute implication over Y is an expression

$$A \Rightarrow B$$

where $A \subseteq Y$ and $B \subseteq Y$ (A and B are sets of attributes).

Example

- Let $Y = \{y_1, y_2, y_3, y_4\}$. Then $\{y_2, y_3\} \Rightarrow \{y_1, y_4\}$, $\{y_2, y_3\} \Rightarrow \{y_1, y_2, y_3\}$, $\emptyset \Rightarrow \{y_1, y_2\}$, $\{y_2, y_4\} \Rightarrow \emptyset$ are AIs over Y.
- Let $Y = \{$watches-TV, eats-unhealthy-food, runs-regularly, normal-blood-pressure, high-blood-pressure$\}$. Then $\{$watches-TV, eats-unhealthy-food$\} \Rightarrow \{$high-blood-pressure$\}$, $\{$runs-regularly$\} \Rightarrow \{$normal-blood-pressure$\}$ are attribute implications over Y.

AIs – validity

- Basic semantic structures in which we evaluate attribute implications are rows of tables (of formal contexts).
- Table rows can be regarded as sets of attributes. In table

	y_1	y_2	y_3	y_4
x_1	×	×	×	×
x_2	×			×
x_3				

, rows corresponding to x_1, x_2, and x_3 can be regarded as sets $M_1 = \{y_1, y_2, y_3, y_4\}$, $M_2 = \{y_1, y_4\}$, and $M_3 = \emptyset$.
- Therefore, we need to define a notion of a validity of an AI in a set M of attributes.

Definition (validity of attribute implication)

An attribute implication $A \Rightarrow B$ over Y is true (valid) in a set $M \subseteq Y$ iff

$$A \subseteq M \text{ implies } B \subseteq M.$$

AIs – validity

– We write

$$\|A \Rightarrow B\|_M = \begin{cases} 1 & \text{if } A \Rightarrow B \text{ is true in } M, \\ 0 & \text{if } A \Rightarrow B \text{ is not true in } M. \end{cases}$$

– Let M be a set of attributes of some object x. $\|A \Rightarrow B\|_M = 1$ says "if x has all attributes from A then x has all attributes from B", because "if x has all attributes from C" is equivalent to $C \subseteq M$.

Example

- Let $Y = \{y_1, y_2, y_3, y_4\}$.

$A \Rightarrow B$	M	$\|A \Rightarrow B\|_M$	why
$\{y_2, y_3\} \Rightarrow \{y_1\}$	$\{y_2\}$	1	$A \not\subseteq M$
$\{y_2, y_3\} \Rightarrow \{y_1\}$	$\{y_1, y_2\}$	1	$A \not\subseteq M$
$\{y_2, y_3\} \Rightarrow \{y_1\}$	$\{y_1, y_2, y_3\}$	1	$A \subseteq M$ and $B \subseteq M$
$\{y_2, y_3\} \Rightarrow \{y_1\}$	$\{y_2, y_3, y_4\}$	0	$A \subseteq M$ but $B \not\subseteq M$
$\{y_2, y_3\} \Rightarrow \{y_1\}$	$\emptyset$	1	$A \not\subseteq \emptyset$
$\emptyset \Rightarrow \{y_1\}$	$\{y_1, y_4\}$	1	$\emptyset \subseteq M$ and $B \subseteq M$.
$\emptyset \Rightarrow \{y_1\}$	$\{y_3, y_4\}$	0	$\emptyset \subseteq M$ but $B \not\subseteq M$.
$\{y_2, y_3\} \Rightarrow \emptyset$	any M	1	$\emptyset \subseteq M$

AIs – validity

- extend validity of $A \Rightarrow B$ to collections $\mathcal{M}$ of M's (collections of subsets of attributes), i.e. define validity of $A \Rightarrow B$ in $\mathcal{M} \subseteq 2^Y$.

Definition

Let $\mathcal{M} \subseteq 2^Y$ (elements of $\mathcal{M}$ are subsets of attributes). An attribute implication $A \Rightarrow B$ over Y is true (valid) in $\mathcal{M}$ if $A \Rightarrow B$ is true in each $M \in \mathcal{M}$.

- Again,

$$||A \Rightarrow B||_{\mathcal{M}} = \begin{cases} 1 & \text{if } A \Rightarrow B \text{ is true in } \mathcal{M}, \\ 0 & \text{if } A \Rightarrow B \text{ is not true in } \mathcal{M}. \end{cases}$$

Therefore, $||A \Rightarrow B||_{\mathcal{M}} = \min_{M \in \mathcal{M}} ||A \Rightarrow B||_M$.

AIs – validity

Definition (validity of attribute implications in formal contexts)

An attribute implication $A \Rightarrow B$ over Y is true in a table (formal context) $\langle X, Y, I\rangle$ iff $A \Rightarrow B$ is true in

$$\mathcal{M} = \{\{x\}^{\uparrow} \,|\, x \in X\}.$$

- We write $||A \Rightarrow B||_{\langle X,Y,I\rangle} = 1$ if $A \Rightarrow B$ is true in $\langle X, Y, I\rangle$.
- Note that, $\{x\}^{\uparrow}$ is the set of attributes of x (row corresponding to x). Hence, $\mathcal{M} = \{\{x\}^{\uparrow} \,|\, x \in X\}$ is the collection whose members are just sets of attributes of objects (i.e., rows) of $\langle X, Y, I\rangle$. Therefore, $||A \Rightarrow B||_{\langle X,Y,I\rangle} = 1$ iff $A \Rightarrow B$ is true in each row of $\langle X, Y, I\rangle$ iff for each $x \in X$:

  if x has all attributes from A then x has all attributes from B.

AIs – validity

Example

Consider attributes normal blood pressure (nbp), high blood pressure (hbp), watches TV (TV), eats unhealthy food (uf), runs regularly (r), and table

I	nbp	hbp	TV	uf	r
a	×				×
b	×			×	×
c		×	×	×	
d		×		×	
e	×				

Then

$A \Rightarrow B$	$\|\|A \Rightarrow B\|\|_{\langle X,Y,I \rangle}$	why
{r} ⇒ {nbp}	1	
{TV,uf} ⇒ {hbp}	1	
{TV} ⇒ {hbp}	1	
{uf} ⇒ {hbp}	0	*b* counterexample
{nbp} ⇒ {r}	0	*e* counterexample
{nbp,hbp} ⇒ {r,TV}	1	*A* never satisfied
{uf,r} ⇒ {r}	1	

AIs – theory, models, semantic consequence

- Previous example: {TV,uf} $\Rightarrow$ {hbp} intuitively follows from {TV} $\Rightarrow$ {hbp}. Therefore, provided we establish validity of {TV} $\Rightarrow$ {hbp}, AI {TV,uf} $\Rightarrow$ {hbp} is redundant.
  Another example: $A \Rightarrow C$ follows from $A \Rightarrow B$ and $B \Rightarrow C$ (for any A, B, C).
- Need to capture intuitive notion of entailment of attribute implications. We use standard notions of a theory and model.
- Eventually, we want to have a small set T of AIs which are valid in $\langle X, Y, I\rangle$ such that all other AIs which are true in $\langle X, Y, I\rangle$ follow from T.

Definition (theory, model)

A theory (over Y) is any set T of attribute implications (over Y).
A model of a theory T is any $M \subseteq Y$ such that every $A \Rightarrow B$ from T is true in M.

- $\mathrm{Mod}(T)$ denotes all models of a theory T, i.e.

$$\mathrm{Mod}(T) = \{M \subseteq Y \mid \text{for each } A \Rightarrow B \in T : A \Rightarrow B \text{ is true in } M\}.$$

- Intuitively, a theory is some "important" set of attribute implications. For instance, T may contain AIs established to be true in data (extracted from data).
- Intuitively, a model of T is (a set of attributes of some) object which satisfies every AI from T.
- Notions of theory and model do not depend on some particular $\langle X, Y, I\rangle$.

Example (theories over $\{y_1, y_2, y_3\}$)

- $T_1 = \{\{y_3\} \Rightarrow \{y_1, y_2\}, \{y_1, y_3\} \Rightarrow \{y_2\}\}$.
- $T_2 = \{\{y_3\} \Rightarrow \{y_1, y_2\}\}$.
- $T_3 = \{\{y_1, y_3\} \Rightarrow \{y_2\}\}$.
- $T_4 = \{\{y_1\} \Rightarrow \{y_3\}, \{y_3\} \Rightarrow \{y_1\}, \{y_2\} \Rightarrow \{y_2\}\}$.
- $T_5 = \emptyset$.
- $T_6 = \{\emptyset \Rightarrow \{y_1\}, \emptyset \Rightarrow \{y_3\}\}$.
- $T_7 = \{\{y_1\} \Rightarrow \emptyset, \{y_2\} \Rightarrow \emptyset, \{y_3\} \Rightarrow \emptyset\}$.
- $T_8 = \{\{y_1\} \Rightarrow \{y_2\}, \{y_2\} \Rightarrow \{y_3\}, \{y_3\} \Rightarrow \{y_1\}\}$.

Example (models of theories over $\{y_1, y_2, y_3\}$)

Determine $\mathrm{Mod}(T)$ of the following theories over $\{y_1, y_2, y_3\}$.

- $T_1 = \{\{y_3\} \Rightarrow \{y_1, y_2\}, \{y_1, y_3\} \Rightarrow \{y_2\}\}$.
  $\mathrm{Mod}(T_1) = \{\emptyset, \{y_1\}, \{y_2\}, \{y_1, y_2\}, \{y_1, y_2, y_3\}\}$,
- $T_2 = \{\{y_3\} \Rightarrow \{y_1, y_2\}\}$.
  $\mathrm{Mod}(T_2) = \{\emptyset, \{y_1\}, \{y_2\}, \{y_1, y_2\}, \{y_1, y_2, y_3\}\}$ (note: $T_2 \subset T_1$ but $\mathrm{Mod}(T_1) = \mathrm{Mod}(T_2)$),
- $T_3 = \{\{y_1, y_3\} \Rightarrow \{y_2\}\}$.
  $\mathrm{Mod}(T_3) = \{\emptyset, \{y_1\}, \{y_2\}, \{y_3\}, \{y_1, y_2\}, \{y_2, y_3\}, \{y_1, y_2, y_3\}\}$ (note: $T_3 \subset T_1$, $\mathrm{Mod}(T_1) \subset \mathrm{Mod}(T_2)$),
- $T_4 = \{\{y_1\} \Rightarrow \{y_3\}, \{y_3\} \Rightarrow \{y_1\}, \{y_2\} \Rightarrow \{y_2\}\}$.
  $\mathrm{Mod}(T_4) = \{\emptyset, \{y_2\}, \{y_1, y_3\}, \{y_1, y_2, y_3\}\}$
- $T_5 = \emptyset$. $\mathrm{Mod}(T_5) = 2^{\{y_1, y_2, y_3\}}$. Why: $M \in \mathrm{Mod}(T)$ iff for each $A \Rightarrow B$: if $A \Rightarrow B \in T$ then $||A \Rightarrow B||_M = 1$.
- $T_6 = \{\emptyset \Rightarrow \{y_1\}, \emptyset \Rightarrow \{y_3\}\}$. $\mathrm{Mod}(T_6) = \{\{y_1, y_3\}, \{y_1, y_2, y_3\}\}$.
- $T_7 = \{\{y_1\} \Rightarrow \emptyset, \{y_2\} \Rightarrow \emptyset, \{y_3\} \Rightarrow \emptyset\}$. $\mathrm{Mod}(T_7) = 2^{\{y_1, y_2, y_3\}}$.
- $T_8 = \{\{y_1\} \Rightarrow \{y_2\}, \{y_2\} \Rightarrow \{y_3\}, \{y_3\} \Rightarrow \{y_1\}\}$.
  $\mathrm{Mod}(T_8) = \{\emptyset, \{y_1, y_2, y_3\}\}$.

AIs – theory, models, semantic consequence

Definition (semantic consequence)

An attribute implication $A \Rightarrow B$ follows semantically from a theory T, which is denoted by

$$T \models A \Rightarrow B,$$

iff $A \Rightarrow B$ is true in every model M of T,

- Therefore, $T \models A \Rightarrow B$ iff for each $M \subseteq Y$: if $M \in \mathrm{Mod}(T)$ then $||A \Rightarrow B||_M = 1$.
- Intuitively, $T \models A \Rightarrow B$ iff $A \Rightarrow B$ is true in every situation where every AI from T is true (replace "situation" by "model").
- Later on, we will see how to efficiently check whether $T \models A \Rightarrow B$.
- Terminology: $T \models A \Rightarrow B$ ... $A \Rightarrow B$ follows semantically from T ... $A \Rightarrow B$ is semantically entailed by T ... $A \Rightarrow B$ is a semantic consequence of T.

How to decide by definition whether $T \models A \Rightarrow B$?

1. Determine $\mathrm{Mod}(T)$.
2. Check whether $A \Rightarrow B$ is true in every $M \in \mathrm{Mod}(T)$; if yes then $T \models A \Rightarrow B$; if not then $T \not\models A \Rightarrow B$.

Example (semantic entailment)

Let $Y = \{y_1, y_2, y_3\}$. Determine whether $T \models A \Rightarrow B$.

- $T = \{\{y_3\} \Rightarrow \{y_1, y_2\}, \{y_1, y_3\} \Rightarrow \{y_2\}\}$, $A \Rightarrow B$ is $\{y_2, y_3\} \Rightarrow \{y_1\}$.
  1. $\mathrm{Mod}(T) = \{\emptyset, \{y_1\}, \{y_2\}, \{y_1, y_2\}, \{y_1, y_2, y_3\}\}$.
  2. $||\{y_2, y_3\} \Rightarrow \{y_1\}||_{\emptyset} = 1$, $||\{y_2, y_3\} \Rightarrow \{y_1\}||_{\{y_1\}} = 1$, $||\{y_2, y_3\} \Rightarrow \{y_1\}||_{\{y_2\}} = 1$, $||\{y_2, y_3\} \Rightarrow \{y_1\}||_{\{y_1,y_2\}} = 1$, $||\{y_2, y_3\} \Rightarrow \{y_1\}||_{\{y_1,y_2,y_3\}} = 1$.

  Therefore, $T \models A \Rightarrow B$.
- $T = \{\{y_3\} \Rightarrow \{y_1, y_2\}, \{y_1, y_3\} \Rightarrow \{y_2\}\}$, $A \Rightarrow B$ is $\{y_2\} \Rightarrow \{y_1\}$.
  1. $\mathrm{Mod}(T) = \{\emptyset, \{y_1\}, \{y_2\}, \{y_1, y_2\}, \{y_1, y_2, y_3\}\}$.
  2. $||\{y_2\} \Rightarrow \{y_1\}||_{\emptyset} = 1$, $||\{y_2\} \Rightarrow \{y_1\}||_{\{y_1\}} = 1$, $||\{y_2\} \Rightarrow \{y_1\}||_{\{y_2\}} = 0$, we can stop.

  Therefore, $T \not\models A \Rightarrow B$.

exercise

Let $Y = \{y_1, y_2, y_3\}$. Determine whether $T \models A \Rightarrow B$.

- $T_1 = \{\{y_3\} \Rightarrow \{y_1, y_2\}, \{y_1, y_3\} \Rightarrow \{y_2\}\}$.
  $A \Rightarrow B$: $\{y_1, y_2\} \Rightarrow \{y_3\}$, $\emptyset \Rightarrow \{y_1\}$.
- $T_2 = \{\{y_3\} \Rightarrow \{y_1, y_2\}\}$.
  $A \Rightarrow B$: $\{y_3\} \Rightarrow \{y_2\}$, $\{y_3, y_2\} \Rightarrow \emptyset$.
- $T_3 = \{\{y_1, y_3\} \Rightarrow \{y_2\}\}$.
  $A \Rightarrow B$: $\{y_3\} \Rightarrow \{y_1, y_2\}$, $\Rightarrow \emptyset$.
- $T_4 = \{\{y_1\} \Rightarrow \{y_3\}, \{y_3\} \Rightarrow \{y_2\}, \}$.
  $A \Rightarrow B$: $\{y_1\} \Rightarrow \{y_2\}$, $\{y_1\} \Rightarrow \{y_1, y_2, y_3\}$.
- $T_5 = \emptyset$.
  $A \Rightarrow B$: $\{y_1\} \Rightarrow \{y_2\}$, $\{y_1\} \Rightarrow \{y_1, y_2, y_3\}$.
- $T_6 = \{\emptyset \Rightarrow \{y_1\}, \emptyset \Rightarrow \{y_3\}\}$.
  $A \Rightarrow B$: $\{y_1\} \Rightarrow \{y_3\}$, $\emptyset \Rightarrow \{y_1, y_3\}$ $\{y_1\} \Rightarrow \{y_2\}$.
- $T_7 = \{\{y_1\} \Rightarrow \emptyset, \{y_2\} \Rightarrow \emptyset, \{y_3\} \Rightarrow \emptyset\}$.
  $A \Rightarrow B$: $\{y_1, y_2\} \Rightarrow \{y_3\}$, $\{y_1, y_2\} \Rightarrow \emptyset$.
- $T_8 = \{\{y_1\} \Rightarrow \{y_2\}, \{y_2\} \Rightarrow \{y_3\}, \{y_3\} \Rightarrow \{y_1\}\}$.
  $A \Rightarrow B$: $\{y_1\} \Rightarrow \{y_3\}$, $\{y_1, y_3\} \Rightarrow \{y_2\}$.

Armstrong rules and reasoning with AIs

- some attribute implications semantically follow from others,
- example: $A \Rightarrow C$ follows from $A \Rightarrow B$ and $B \Rightarrow C$ (for every $A, B, C \subseteq Y$), i.e. $\{A \Rightarrow B, B \Rightarrow C\} \models A \Rightarrow C$.
- therefore, we can introduce a deduction rule
  (Tra) from $A \Rightarrow B$ and $B \Rightarrow C$ infer $A \Rightarrow C$,
- we can use such rule to derive new AI such as
  - start from $T = \{\{y_1\} \Rightarrow \{y_2, y_5\}, \{y_2, y_5\} \Rightarrow \{y_3\}, \{y_3\} \Rightarrow \{y_2, y_4\}\}$,
  - apply (Tra) to the first and the second AI in T to infer $\{y_1\} \Rightarrow \{y_3\}$,
  - apply (Tra) to $\{y_1\} \Rightarrow \{y_3\}$ and the second AI in T to infer $\{y_1\} \Rightarrow \{y_2, y_4\}$.

question:

- Is there a collection of simple deduction rules which allow us to determine whether $T \models A \Rightarrow B$?, i.e., rules such that
- 1. if $A \Rightarrow B$ semantically follows from T then one can derive $A \Rightarrow B$ from T using those rules (like above) and
- 2. if one can derive $A \Rightarrow B$ from T then $A \Rightarrow B$ semantically follows from T.

Armstrong rules and reasoning with AIs

Armstrong rules for reasoning with AIs

Our system for reasoning about attribute implications consists of the following (schemes of) deduction rules:

(Ax) infer $A \cup B \Rightarrow A$,

(Cut) from $A \Rightarrow B$ and $B \cup C \Rightarrow D$ infer $A \cup C \Rightarrow D$,

for every $A, B, C, D \subseteq Y$.

- (Ax) is a rule without the input part "from ...", i.e. $A \cup B \Rightarrow A$ can be inferred from any AIs.
- (Cut) has both the input and the output part.
- Rules for reasoning about AIs go back to Armstrong's research on reasoning about functional dependencies in databases: Armstrong W. W.: Dependency structures in data base relationships. IFIP Congress, Geneva, Switzerland, 1974, pp. 580–583.
- There are several systems of deduction rules which are equivalent to (Ax), (Cut), see later.

Armstrong rules and reasoning with AIs

Example (how to use deduction rules)

(Cut)
If we have two rules which are of the form $A \Rightarrow B$ and $B \cup C \Rightarrow D$, we can derive (in a single step, using deduction rule (Cut)) a new AI of the form $A \cup C \Rightarrow D$.
Consider AIs $\{r, s\} \Rightarrow \{t, u\}$ and $\{t, u, v\} \Rightarrow \{w\}$.
Putting $A = \{r, s\}$, $B = \{t, u\}$, $C = \{v\}$, $D = \{w\}$, $\{r, s\} \Rightarrow \{t, u\}$ is of the form $A \Rightarrow B$, $\{t, u, v\} \Rightarrow \{w\}$ is of the form $A \cup C \Rightarrow D$, and we can infer $A \cup C \Rightarrow D$ which is $\{r, s, v\} \Rightarrow \{w\}$.

(Ax)
We can derive (in a single step, using deduction rule (Ax), with no assumptions) a new AI of the form $A \cup B \Rightarrow A$.
For instance, we can infer $\{y_1, y_3, y_4, y_5\} \Rightarrow \{y_3, y_5\}$. Namely, putting $A = \{y_3, y_5\}$ and $B = \{y_1, y_4\}$, $A \cup B \Rightarrow A$ becomes $\{y_1, y_3, y_4, y_5\} \Rightarrow \{y_3, y_5\}$.

Armstrong rules and reasoning with AIs

How to formalize the concept of a derivation of new AIs using our rules?

Definition (proof)

A proof of $A \Rightarrow B$ from a set T of AIs is a sequence

$$A_1 \Rightarrow B_1, \ldots, A_n \Rightarrow B_n$$

of AIs satisfying:

1. $A_n \Rightarrow B_n$ is just $A \Rightarrow B$,
2. for every $i = 1, 2, \ldots, n$:
   - either $A_i \Rightarrow B_i$ is from T ("assumption"),
   - or $A_i \Rightarrow B_i$ results by application of (Ax) or (Cut) to some of preceding AIs $A_j \Rightarrow B_j$'s ("deduction").

In such case, we write $T \vdash A \Rightarrow B$ and say that $A \Rightarrow B$ is provable (derivable) from T using (Ax) and (Cut).

– proof as a sequence?: makes sense: informally, we understand a proof to be a sequence of our arguments which we take from 1. assumptions (from T) of 2. infer pro previous arguments by deduction steps.

Armstrong rules and reasoning with AIs

Example (simple proof)

Proof of $P \Rightarrow R$ from $T = \{P \Rightarrow Q, Q \Rightarrow R\}$ is a sequence:

$$P \Rightarrow Q, Q \Rightarrow R, P \Rightarrow R$$

because: $P \Rightarrow Q \in T$; $Q \Rightarrow R \in T$; $P \Rightarrow R$ can be inferred from $P \Rightarrow Q$ and $Q \Rightarrow R$ using (Cut). Namely, put $A = P$, $B = Q$, $C = Q$, $D = R$; then $A \Rightarrow B$ becomes $P \Rightarrow Q$, $B \cup C \Rightarrow D$ becomes $Q \Rightarrow R$, and $A \cup C \Rightarrow D$ becomes $P \Rightarrow R$.

Note that this works for any particular sets P, Q, R. For instance for $P = \{y_1, y_3\}$, $Q = \{y_3, y_4, y_5\}$, $R = \{y_2, y_4\}$, or
$P = \{$watches-TV,unhealthy-food$\}$, $Q = \{$high-blood-pressure$\}$,
$R = \{$often-visits-doctor$\}$.

In the latter case, we inferred:
$\{$watches-TV,unhealthy-food$\} \Rightarrow \{$often-visits-doctor$\}$ from
$\{$watches-TV,unhealthy-food$\} \Rightarrow \{$high-blood-pressure$\}$ and
$\{$high-blood-pressure$\} \Rightarrow \{$often-visits-doctor$\}$.

Armstrong rules and reasoning with AIs

remark

The notions of a deduction rule and proof are syntactic notions. Proof results by "manipulation of symbols" according to deduction rules. We do not refer to any data table when deriving new AIs using deduction rules. A typical scenario: (1) We extract a set T of AIs from data table and then (2) infer further AIs from T using deduction rules. In (2), we do not use the data table.

Next:

- Soundness: Is our inference using (Ax) and (Cut) sound? That is, is it the case that IF $T \vdash A \Rightarrow B$ ($A \Rightarrow B$ can be inferred from T) THEN $T \models A \Rightarrow B$ ($A \Rightarrow B$ semantically follows from T, i.e., $A \Rightarrow B$ is true in every table in which all AIs from T are true)?
- Completeness: Is our inference using (Ax) and (Cut) complete? That is, is it the case that IF $T \models A \Rightarrow B$ THEN $T \vdash A \Rightarrow B$?

Armstrong rules and reasoning with AIs

Definition (derivable rule)

Deduction rule

$$\text{from } A_1 \Rightarrow B_1, \ldots, A_n \Rightarrow B_n \text{ infer } A \Rightarrow B$$

is derivable from (Ax) and (Cut) if $\{A_1 \Rightarrow B_1, \ldots, A_n \Rightarrow B_n\} \vdash A \Rightarrow B$.

- Derivable rule = new deduction rule = shorthand for a derivation using the basic rules (Ax) and (Cut).
- Why derivable rules: They are natural rules which can speed up proofs.
- Derivable rules can be used in proofs (in addition to the basic rules (Ax) and (Cut)). Why: By definition, a single deduction step using a derivable rule can be replaced by a sequence of deduction steps using the original deduction rules (Ax) and (Cut) only.

Theorem (derivable rules)

The following rules are derivable from (Ax) and (Cut):

(Ref) *infer* $A \Rightarrow A$,

(Wea) *from* $A \Rightarrow B$ *infer* $A \cup C \Rightarrow B$,

(Add) *from* $A \Rightarrow B$ *and* $A \Rightarrow C$ *infer* $A \Rightarrow B \cup C$,

(Pro) *from* $A \Rightarrow B \cup C$ *infer* $A \Rightarrow B$,

(Tra) *from* $A \Rightarrow B$ *and* $B \Rightarrow C$ *infer* $A \Rightarrow C$,

for every $A, B, C, D \subseteq Y$.

Proof.

In order to avoid confusion with symbols A, B, C, D used in (Ax) and (Cut), we use P, Q, R, S instead of A, B, C, D in (Ref)–(Tra).

(Ref): We need to show $\{\} \vdash P \Rightarrow P$, i.e. that $P \Rightarrow P$ is derivable using (Ax) and (Cut) from the empty set of assumptions.
Easy, just put $A = P$ and $B = P$ in (Ax). Then $A \cup B \Rightarrow A$ becomes $P \Rightarrow P$. Therefore, $P \Rightarrow P$ can be inferred (in a single step) using (Ax), i.e., a one-element sequence $P \Rightarrow P$ is a proof of $P \Rightarrow P$. This shows $\{\} \vdash P \Rightarrow P$.

cntd.

(Wea): We need to show $\{P \Rightarrow Q\} \vdash P \cup R \Rightarrow Q$.
A proof (there may be several proofs, this is one of them) is:

$$P \cup R \Rightarrow P, P \Rightarrow Q, P \cup R \Rightarrow Q.$$

Namely, 1. $P \cup R \Rightarrow P$ is derived using (Ax), 2. $P \Rightarrow Q$ is an assumption, $P \cup R \Rightarrow Q$ is derived from $P \cup R \Rightarrow P$ and $P \Rightarrow Q$ using (Cut) (put $A = P \cup R$, $B = P$, $C = P$, $D = Q$).

(Add): EXERCISE.

(Pro): We need to show $\{P \Rightarrow Q \cup R\} \vdash P \Rightarrow Q$.
A proof is:

$$P \Rightarrow Q \cup R, Q \cup R \Rightarrow Q, P \Rightarrow Q.$$

Namely, 1. $P \Rightarrow Q \cup R$ is an assumption, 2. $Q \cup R \Rightarrow Q$ by application of (Ax), 3. $P \Rightarrow Q$ by application of (Cut) to $P \Rightarrow Q \cup R, Q \cup R \Rightarrow Q$ (put $A = P$, $B = C = Q \cup R$, $D = Q$).

(Tra): We need to show $\{P \Rightarrow Q, Q \Rightarrow R\} \vdash P \Rightarrow R$. This was checked earlier. □

Armstrong rules and reasoning with AIs

- (Ax) ... "axiom", and (Cut) ... "rule of cut",
- (Ref) ... "rule of reflexivity", (Wea) ... "rule of weakening", (Add) ... "rule of additivity", (Pro) ... "rule of projectivity", (Ref) ... "rule of transitivity".

Alternative notation for deduction rules: rule "from $A_1 \Rightarrow B_1, \ldots, A_n \Rightarrow B_n$ infer $A \Rightarrow B$" displayed as

$$\frac{A_1 \Rightarrow B_1, \ldots, A_n \Rightarrow B_n}{A \Rightarrow B}.$$

So, (Ax) and (Cut) displayed as

$$\frac{}{A \cup B \Rightarrow A} \quad \text{and} \quad \frac{A \Rightarrow B, B \cup C \Rightarrow D}{A \cup C \Rightarrow D}.$$

Armstrong rules and reasoning with AIs

Definition (sound deduction rules)

Deduction rule "from $A_1 \Rightarrow B_1, \ldots, A_n \Rightarrow B_n$ infer $A \Rightarrow B$" is sound if

$$\{A_1 \Rightarrow B_1, \ldots, A_n \Rightarrow B_n\} \models A \Rightarrow B.$$

- Soundness of a rule: if $A_1 \Rightarrow B_1, \ldots, A_n \Rightarrow B_n$ are true in a data table, then $A \Rightarrow B$ needs to be true in that data table, too.
- Meaning: Sound deduction rules do not allow us to infer "untrue" AIs from true AIs.

Theorem

(Ax) and (Cut) are sound.

Proof.

(Ax): We need to check $\{\} \models A \cup B \Rightarrow A$, i.e. that $A \cup B \Rightarrow A$ semantically follows from an empty set T of assumptions. That is, we need to check that $A \cup B \Rightarrow A$ is true in any $M \subseteq Y$ (notice: any $M \subseteq Y$ is a model of the empty set of AIs). This amounts to verifying

$$A \cup B \subseteq M \text{ implies } A \subseteq M,$$

which is evidently true.

(Cut): We need to check $\{A \Rightarrow B, B \cup C \Rightarrow D\} \models A \cup C \Rightarrow D$. Let M be a model of $\{A \Rightarrow B, B \cup C \Rightarrow D\}$. We need to show that M is a model of $A \cup C \Rightarrow D$, i.e. that

$$A \cup D \subseteq M \text{ implies } D \subseteq M.$$

Let thus $A \cup C \subseteq M$. Then $A \subseteq M$, and since we assume M is a model of $A \Rightarrow B$, we need to have $B \subseteq M$. Furthermore, $A \cup C \subseteq M$ yields $C \subseteq M$. That is, we have $B \subseteq M$ and $C \subseteq M$, i.e. $B \cup C \subseteq M$. Now, taking $B \cup C \subseteq M$ and invoking the assumption that M is a model of $B \cup C \Rightarrow D$ gives $D \subseteq M$. □

Armstrong rules and reasoning with AIs

Corollary (soundness of inference using (Ax) and (Cut))

If $T \vdash A \Rightarrow B$ *then* $T \models A \Rightarrow B$.

Proof.

Direct consequence of previous theorem: Let

$$A_1 \Rightarrow B_1, \ldots, A_n \Rightarrow B_n$$

be a proof from T. It suffices to check that every model M of T is a model of $A_i \Rightarrow B_i$ for $i = 1, \ldots, n$. We check this by induction over i, i.e., we assume that M is a model of $A_j \Rightarrow B_j$'s for $j < i$ and check that M is a model of $A_i \Rightarrow B_i$. There are two options:

1. Either $A_i \Rightarrow B_i$ if from T. Then, trivially, M is a model of $A_i \Rightarrow B_i$ (our assumption).

2. Or, $A_i \Rightarrow B_i$ results by (Ax) or (Cut) to some $A_j \Rightarrow B_j$'s for $j < i$. Then, since we assume that M is a model of $A_j \Rightarrow B_j$'s, we get that M is a model of $A_i \Rightarrow B_i$ by soundness of (Ax) and (Cut). □

Armstrong rules and reasoning with AIs

Corollary (soundness of derived rules)

(Ref), (Wea), (Add), (Pro), (Tra) are sound.

Proof.

As an example, take (Wea). Note that (Wea) is a derived rule. This means that $\{A \Rightarrow B\} \vdash A \cup C \Rightarrow B$. Applying previous corollary yields $\{A \Rightarrow B\} \models A \cup C \Rightarrow B$ which means, by definition, that (Wea) is sound. □

- We have two notions of consequence, semantic and syntactic.
- Semantic: $T \models A \Rightarrow B$ ... $A \Rightarrow B$ semantically follows from T.
- Syntactic: $T \vdash A \Rightarrow B$ ... $A \Rightarrow B$ syntactically follows from T (is provable from T).
- We know (previous corollary on soundness) that $T \vdash A \Rightarrow B$ implies $T \models A \Rightarrow B$.
- Next, we are going to check completeness, i.e. $T \models A \Rightarrow B$ implies $T \vdash A \Rightarrow B$.

Armstrong rules and reasoning with AIs

Definition (semantic closure, syntactic closure)

- Semantic closure of T is the set
$$sem(T) = \{A \Rightarrow B \mid T \models A \Rightarrow B\}$$
of all AIs which semantically follow from T.
- Syntactic closure of T is the set
$$syn(T) = \{A \Rightarrow B \mid T \vdash A \Rightarrow B\}$$
of all AIs which syntactically follow from T (i.e., are provable from T using (Ax) and (Cut)).
- T is semantically closed if $T = sem(T)$.
- T is syntactically closed if $T = syn(T)$.

- It can be checked that $sem(T)$ is the least set of AIs which is semantically closed and which contains T.
- It can be checked that $syn(T)$ is the least set of AIs which is syntactically closed and which contains T.

Armstrong rules and reasoning with AIs

Lemma

T is syntactically closed iff for any $A, B, C, D \subseteq Y$

1. $A \cup B \Rightarrow B \in T$,
2. *if* $A \Rightarrow B \in T$ *and* $B \cup C \Rightarrow D \in T$ *implies* $A \cup C \Rightarrow D \in T$.

Proof.

"$\Rightarrow$": If T is syntactically closed then any AI which is provable from T needs to be in T. In particular, $A \cup B \Rightarrow B$ is provable from T, therefore $A \cup B \Rightarrow B \in T$; if $A \Rightarrow B \in T$ and $B \cup C \Rightarrow D \in T$ then, obviously, $A \cup C \Rightarrow D$ is provable from T (by using (Cut)), therefore $A \cup C \Rightarrow D \in T$.
"$\Leftarrow$": If 1. and 2. are satisfied then, obviously, any AI which is provable from T needs to belong to T, i.e. T is syntactically closed. □

This says that T is syntactically closed iff T is closed under deduction rules (Ax) and (Cut).

Armstrong rules and reasoning with AIs

Lemma

If T is semantically closed then T is syntactically closed.

Proof.

Let T be semantically closed. In order to see that T is syntactically closed, it suffices to verify 1. and 2. of previous Lemma.
1.: We have $T \models A \cup B \Rightarrow B$ (we even have $\{\} \models A \cup B \Rightarrow B$). Since T is semantically closed, we get $A \cup B \Rightarrow B \in T$.
2.: Let $A \Rightarrow B \in T$ and $B \cup C \Rightarrow D \in T$. Since $\{A \Rightarrow B, B \cup C \Rightarrow D\} \models A \cup C \Rightarrow D$ (cf. soundness of (Cut)), we have $T \models A \cup C \Rightarrow D$. Now, since T is semantically closed, we get $A \cup C \Rightarrow D \in T$, verifying 2. □

Armstrong rules and reasoning with AIs

Lemma

If T is syntactically closed then T is semantically closed.

Proof.

Let T be syntactically closed. In order to show that T is semantically closed, it suffices to show $sem(T) \subseteq T$. We prove this by showing that if $A \Rightarrow B \notin T$ then $A \Rightarrow B \notin sem(T)$. Recall that since T is syntactically closed, T is closed under all (Ref)–(Tra).

Let thus $A \Rightarrow B \notin T$. To see $A \Rightarrow B \notin sem(T)$, we show that there is $M \in \mathrm{Mod}(T)$ which is not a model of $A \Rightarrow B$. For this purpose, consider $M = A^+$ where A^+ is the largest one such that $A \Rightarrow A^+ \in T$. A^+ exists. Namely, consider all AIs $A \Rightarrow C_1, \ldots, A \Rightarrow C_n \in T$. Note that at least one such AI exists. Namely, $A \Rightarrow A \in T$ by (Ref). Now, repeated application of (Add) yields $A \Rightarrow \bigcup_{i=1}^n C_i \in T$ and we have $A^+ = \bigcup_{i=1}^n C_i$.

□

cntd.

Now, we need to check that (a) $||A \Rightarrow B||_{A^+} = 0$ (i.e., A^+ is not a model of $A \Rightarrow B$) and (b) for every $C \Rightarrow D \in T$ we have $||C \Rightarrow D||_{A^+} = 1$ (i.e., A^+ is a model of T).

(a): We need to show $||A \Rightarrow B||_{A^+} = 0$. By contradiction, suppose $||A \Rightarrow B||_{A^+} = 1$. Since $A \subseteq A^+$, $||A \Rightarrow B||_{A^+} = 1$ yields $B \subseteq A^+$. Since $A \Rightarrow A^+ \in T$, (Pro) would give $A \Rightarrow B \in T$, a contradiction to $A \Rightarrow B \notin T$.

(b): Let $C \Rightarrow D \in T$. We need to show $||C \Rightarrow D||_{A^+} = 1$, i.e.

$$\text{if } C \subseteq A^+ \text{ then } D \subseteq A^+.$$

To see this, it is sufficient to verify that

$$\text{if } C \subseteq A^+ \text{ then } A \Rightarrow D \in T.$$

Namely, since A^+ is the largest one for which $A \Rightarrow A^+ \in T$, $A \Rightarrow D \in T$ implies $D \subseteq A^+$. So let $C \subseteq A^+$. We have
(b1) $A \Rightarrow A^+ \in T$ (by definition of A^+),
(b2) $A^+ \Rightarrow C \in T$ (this follows by (Pro) from $C \subseteq A^+$),
(b3) $C \Rightarrow D \in T$ (our assumption).
Therefore, applying (Tra) to (b1), (b2), (b3) twice gives $A \Rightarrow D \in T$. □

Theorem (soundness and completeness)

$T \vdash A \Rightarrow B$ iff $T \models A \Rightarrow B$.

Proof.

Clearly, it suffices to check $syn(T) = sem(T)$. Recall: $A \Rightarrow B \in syn(T)$ means $T \vdash A \Rightarrow B$, $A \Rightarrow B \in sem(T)$ means $T \models A \Rightarrow B$.

"$sem(T) \subseteq syn(T)$": Since $syn(T)$ is syntactically closed, it is also semantically closed (previous lemma). Therefore, $sem(syn(T)) = syn(T)$ (semantic closure of $syn(T)$ is just $syn(T)$ because $syn(T)$ is semantically closed). Furthermore, since $T \subseteq syn(T)$, we have $sem(T) \subseteq sem(syn(T))$. Putting this together gives

$$sem(T) \subseteq sem(syn(T)) = syn(T).$$

"$syn(T) \subseteq sem(T)$": Since $sem(T)$ is semantically closed, it is also syntactically closed (previous lemma). Therefore, $syn(sem(T)) = sem(T)$. Furthermore, since $T \subseteq sem(T)$, we have $syn(T) \subseteq syn(sem(T))$. Putting this together gives

$$syn(T) \subseteq syn(sem(T)) = sem(T).$$

□

Armstrong rules and reasoning with AIs

Summary

- (Ax) and (Cut) are elementary deduction rules.
- Proof ... formalizes derivation process of new AIs from other AIs.
- We have two notions of consequence:
  - $T \models A \Rightarrow B$ ... semantic consequence ($A \Rightarrow B$ is true in every model of T).
  - $T \vdash A \Rightarrow B$ ... syntactic consequence ($A \Rightarrow B$ is provable T, i.e. can be derived from T using deduction rules).
- Note: proof = syntactic manipulation, no reference to semantic notions; in order to know what $T \vdash A \Rightarrow B$ means, we do not have to know what it means that an AI $A \Rightarrow B$ is true in M.
- Derivable rules (Ref)–(Tra) ... derivable rule = shorthand, inference of new AIs using derivable rules can be replaced by inference using original rules (Ax) and (Cut).
- Sound rule ... derives true conclusions from true premises; (Ax) and (Cut) are sound; in detail, for (Cut): soundness of (Cut) means that for every M in which both $A \Rightarrow B$ and $B \cup C \Rightarrow D$ are true, $A \cup C \Rightarrow D$ needs to be true, too.

Armstrong rules and reasoning with AIs

- Soundness of inference using sound rules: if $T \vdash A \Rightarrow B$ ($A \Rightarrow B$ is provable from T) then $T \models A \Rightarrow B$ ($A \Rightarrow B$ semantically follows from T), i.e. if $A \Rightarrow B$ is provable from T then $A \Rightarrow B$ is true in every M in which every AI from T is true. Therefore, soundness of inference means that if we take an arbitrary M and take a set T of AIs which are true in M, then evey AI $A \Rightarrow B$ which we can infer (prove) from T using our inference rules needs to be true in M.
- Consequence: rules, such as (Ref)–(Tra), which can be derived from sound rules are sound.
- $sem(T)$ ... set of all AIs which are semantic consequences of T, $syn(T)$ ... set of all AIs which are syntactic consequences of T (provable from T).
- T is syntactically closed iff T is closed under (Ax) and (Cut).
- (Syntactico-semantical) completeness of rules (Ax) and (Cut): $T \vdash A \Rightarrow B$ iff $T \models A \Rightarrow B$.

Armstrong rules and reasoning with AIs

Example

- Explain why $\{\} \models A \Rightarrow B$ means that (1) $A \Rightarrow B$ is true in every $M \subseteq Y$, (2) $A \Rightarrow B$ is true in every formal context $\langle X, Y, I \rangle$.
- Explain why soundness of inference implies that if we take an arbitrary formal context $\langle X, Y, I \rangle$ and take a set T of AIs which are true in $\langle X, Y, I \rangle$, then evey AI $A \Rightarrow B$ which we can infer (prove) from T using our inference rules needs to be true in $\langle X, Y, I \rangle$.
- Let $\mathcal{R}_1$ and $\mathcal{R}_2$ be two sets of deduction rules, e.g. $\mathcal{R}_1 = \{$(Ax), (Cut)$\}$. Call $\mathcal{R}_1$ and $\mathcal{R}_2$ equivalent if every rule from $\mathcal{R}_2$ is a derived rule in terms of rules from $\mathcal{R}_1$ and, vice versa, every rule from $\mathcal{R}_1$ is a derived rule in terms of rules from $\mathcal{R}_2$.
  For instance, we know that taking $\mathcal{R}_1 = \{$(Ax), (Cut)$\}$, every rule from $\mathcal{R}_2 = \{$(Ref),..., (Tra)$\}$ is a derived rule in terms of rules of $\mathcal{R}_1$.
  Verify that $\mathcal{R}_1 = \{$(Ax), (Cut)$\}$ and $\mathcal{R}_2 = \{$(Ref),(Wea), (Cut)$\}$ are equivalent.

Armstrong rules and reasoning with AIs

Example

- Explain: If $\mathcal{R}_1$ and $\mathcal{R}_2$ are equivalent sets of inference rules then $A \Rightarrow B$ is provable from T using rules from $\mathcal{R}_1$ iff $A \Rightarrow B$ is provable from T using rules from $\mathcal{R}_2$.
- Explain: Let $\mathcal{R}_2$ be a set of inference rules equivalent to $\mathcal{R}_1 = \{(\text{Ax}), (\text{Cut})\}$. Then $A \Rightarrow B$ is provable from T using rules from $\mathcal{R}_2$ iff $T \models A \Rightarrow B$.
- Verify that $sem(\cdots)$ is a closure operator, i.e. that $T \subseteq sem(T)$, $T_1 \subseteq T_2$ implies $sem(T_1) \subseteq sem(T_2)$, and $sem(T) = sem(sem(T))$.
- Verify that $syn(\cdots)$ is a closure operator, i.e. that $T \subseteq syn(T)$, $T_1 \subseteq T_2$ implies $syn(T_1) \subseteq syn(T_2)$, and $syn(T) = syn(syn(T))$.
- Verify that for any T, $sem(T)$ is the least semantically closed set which contains T.
- Verify that for any T, $syn(T)$ is the least syntactically closed set which contains T.

Models of attribute implications

For a set T of attribute implications, denote

$$\mathrm{Mod}(T) = \{M \subseteq Y \mid ||A \Rightarrow B||_M = 1 \text{ for every } A \Rightarrow B \in T\}$$

That is, $\mathrm{Mod}(T)$ is the set of all models of T.

Definition (closure system)

A closure system in a set Y is any system $\mathcal{S}$ of subsets of Y which contains Y and is closed under arbitrary intersections.

That is, $Y \in \mathcal{S}$ and $\bigcap \mathcal{R} \in \mathcal{S}$ for every $\mathcal{R} \subseteq \mathcal{S}$ (intersection of every subsystem $\mathcal{R}$ of $\mathcal{S}$ belongs to $\mathcal{S}$).

$\{\{a\}, \{a, b\}, \{a, d\}, \{a, b, c, d\}\}$ is a closure system in $\{a, b, c, d\}$ while $\{\{a, b\}, \{c, d\}, \{a, b, c, d\}\}$ is not.

There is a one-to-one relationship between closure systems in Y and closure operators in Y. Given a closure operator C in Y, $\mathcal{S}_C = \{A \in 2^X \mid A = C(A)\} = \mathrm{fix}(C)$ is a closure system in Y.

Models of attribute implications

Given a closure system in Y, putting

$$C_{\mathcal{S}}(A) = \bigcap\{B \in \mathcal{S} \mid A \subseteq B\}$$

for any $A \subseteq X$, $C_{\mathcal{S}}$ is a closure operator on Y. This is a one-to-one relationship, i.e. $C = C_{\mathcal{S}_C}$ and $\mathcal{S} = \mathcal{S}_{C_{\mathcal{S}}}$ (we omit proofs).

Lemma

For a set T of attribute implications, $\mathrm{Mod}(T)$ is a closure system in Y.

Proof.

First, $Y \in \mathrm{Mod}(T)$ because Y is a model of any attribute implication.
Second, let $M_j \in \mathrm{Mod}(T)$ $(j \in J)$. For any $A \Rightarrow B \in T$, if $A \subseteq \bigcap_j M_j$ then for each $j \in J$: $A \subseteq M_j$, and so $B \subseteq M_j$ (since $M_j \in \mathrm{Mod}(T)$, thus in particular $M_j \models A \Rightarrow B$), from which we have $B \subseteq \bigcap_j M_j$.
We showed that $\mathrm{Mod}(T)$ contains Y and is closed under intersections, i.e. $\mathrm{Mod}(T)$ is a closure system. □

Models of attribute implications

remark

(1) If T is the set of all attribute implications valid in a formal context $\langle X, Y, I\rangle$, then $\mathrm{Mod}(T) = \mathrm{Int}(X, Y, I)$, i.e. models of T are just all the intents of the concept lattice $\mathcal{B}(X, Y, I)$ (see later).
(2) Another connection to concept lattices is: $A \Rightarrow B$ is valid in $\langle X, Y, I\rangle$ iff $A^{\downarrow} \subseteq B^{\downarrow}$ iff $B \subseteq A^{\downarrow\uparrow}$ (see later).

Since $\mathrm{Mod}(T)$ is a closure system, we can consider the corresponing closure operator $C_{\mathrm{Mod}(T)}$ (i.e., the fixed points of $C_{\mathrm{Mod}(T)}$ are just models of T). Therefore, for every $A \subseteq Y$ there exist the least model of $\mathrm{Mod}(T)$ which contains A, namely, such least model is just $C_{\mathrm{Mod}(T)}(A)$.

Theorem (testing entailment via least model)

For any $A \Rightarrow B$ and any T, we have

$$T \models A \Rightarrow B \quad \textit{iff} \quad ||A \Rightarrow B||_{C_{\mathrm{Mod}(T)}(A)} = 1,$$

i.e., $A \Rightarrow B$ semantically follows from T iff $A \Rightarrow B$ is true in the least model $C_{\mathrm{Mod}(T)}(A)$ of T which contains A.

Proof.

"$\Rightarrow$": If $T \models A \Rightarrow B$ then, by definition, $A \Rightarrow B$ is true in every model of T. Therefore, in particular, $A \Rightarrow B$ is true in $C_{\mathrm{Mod}(T)}(A)$.
"$\Leftarrow$": Let $A \Rightarrow B$ be true in $C_{\mathrm{Mod}(T)}(A)$. Since $A \subseteq C_{\mathrm{Mod}(T)}(A)$, we have $B \subseteq C_{\mathrm{Mod}(T)}(A)$. We need to check that $A \Rightarrow B$ is true in every model of T. Let thus $M \in \mathrm{Mod}(T)$. If $A \not\subseteq M$ then, clearly, $A \Rightarrow B$ is true in M. If $A \subseteq M$ then, since M is a model of T containing A, we have $C_{\mathrm{Mod}(T)}(A) \subseteq M$. Putting together with $B \subseteq C_{\mathrm{Mod}(T)}(A)$, we get $B \subseteq M$, i.e. $A \Rightarrow B$ is true in M. □

Models of attribute implications

- Previous theorem $\Rightarrow$ testing $T \models A \Rightarrow B$ by checking whether $A \Rightarrow B$ is true in a single particular model of T. This is much better than going by definition $\models$ (definition says: $T \models A \Rightarrow B$ iff $A \Rightarrow B$ is true in every model of T).
- How can we obtain $C_{\mathrm{Mod}(T)}(A)$?

Definition

For $Z \subseteq Y$, T a set of implications, put

1. $Z^T = Z \cup \bigcup\{B \mid A \Rightarrow B \in T, A \subseteq Z\}$,
2. $Z^{T_0} = Z$,
3. $Z^{T_n} = (Z^{T_{n-1}})^T$ (for $n \geq 1$).

Define define operator $C : 2^Y \to 2^Y$ by

$$C(Z) = \bigcup_{n=0}^{\infty} Z^{T_n}$$

Models of attribute implications

Theorem

Given T, C (defined on previous slide) is a closure operator in Y such that

$$C(Z) = C_{\mathrm{Mod}(T)}(Z).$$

Proof.

First, check that C is a closure operator.
$Z = Z^{T_0}$ yields $Z \subseteq C(Z)$.
Evidently, $Z_1 \subseteq Z_2$ implies $Z_1^T \subseteq Z_2^T$ which implies $Z_1^{T_1} \subseteq Z_2^{T_1}$ which implies $Z_1^{T_2} \subseteq Z_2^{T_2}$ which implies $\ldots Z_1^{T_n} \subseteq Z_2^{T_n}$ for any n. That is, $Z_1 \subseteq Z_2$ implies $C(Z_1) = \bigcup_{n=0}^{\infty} Z_1^{T_n} \subseteq \bigcup_{n=0}^{\infty} Z_2^{T_n} = C(Z_2)$.
$C(Z) = C(C(Z))$: Clearly,

$$Z^{T_0} \subseteq Z^{T_1} \subseteq \cdots Z^{T_n} \subseteq \cdots.$$

Since Y is finite, the above sequence terminates after a finite number n_0 of steps, i.e. there is n_0 such that

$$C(Z) = \bigcup_{n=0}^{\infty} Z^{T_n} = Z^{T_{n_0}}.$$

This means $(Z^{T_{n_0}})^T = Z^{T_{n_0}} = C(Z)$ which gives $C(Z) = C(C(Z))$. □

cntd.

Next, we check $C(Z) = C_{\mathrm{Mod}(T)}(Z)$.

1. $C(Z)$ is a model of T containing Z:

Above, we checked that $C(Z)$ contains Z. Take any $A \Rightarrow B \in T$ and verify that $A \Rightarrow B$ is valid in $C(Z)$ (i.e., $C(Z)$ is a model of $A \Rightarrow B$). Let $A \subseteq C(Z)$. We need to check $B \subseteq C(Z)$. $A \subseteq C(Z)$ means that for some n, $A \subseteq Z^{T_n}$. But then, by definition, $B \subseteq (Z^{T_n})^T$ which gives $B \subseteq Z^{T_{n+1}} \subseteq C(Z)$.

2. $C(Z)$ is the least model of T containing Z:

Let M be a model of T containing Z, i.e. $Z^{T_0} = Z \subseteq M$. Then $Z^T \subseteq M^T$ (just check definition of $(\cdots)^T$). Evidently, $M = M^T$. Therefore, $Z^{T_1} = Z^T \subseteq M$. Applying this inductively gives $Z^{T_2} \subseteq M$, $Z^{T_3} \subseteq M$, ... Putting together yields $C(Z) = \bigcup_{n=0}^{\infty} Z^{T_n} \subseteq M$. That is, $C(Z)$ is contained in every model M of T and is thus the least one. □

Models of attribute implications

- Therefore, C is the closure operator which computes, given $Z \subseteq Y$, the least model of T containing Z.
- As argued in the proof, since Y is finite, $\bigcup_{n=0}^{\infty} Z^{T_n}$ "stops" after a finite number of steps. Namely, there is n_0 such that $Z^{T_n} = Z^{T_{n_0}}$ for $n > n_0$.
- The least such n_0 is the smallest n with $Z^{T_n} = Z^{T_{n+1}}$.
- Given T, $C(Z)$ can be computed: Use definition and stop whenever $Z^{T_n} = Z^{T_{n+1}}$. That is, put

$$C(Z) = Z \cup Z^{T_1} \cup Z^{T_2} \cup \cdots \cup Z^{T_n}.$$

- There is a more efficient algorithm (called LinClosure) for computing $C(Z)$. See Maier D.: The Theory of Relational Databases. CS Press, 1983.

Models of attribute implications

Example

Back to one of our previous examples: Let $Y = \{y_1, y_2, y_3\}$. Determine whether $T \models A \Rightarrow B$.

- $T = \{\{y_3\} \Rightarrow \{y_1, y_2\}, \{y_1, y_3\} \Rightarrow \{y_2\}\}$, $A \Rightarrow B$ is $\{y_2, y_3\} \Rightarrow \{y_1\}$.
  1. $\mathrm{Mod}(T) = \{\emptyset, \{y_1\}, \{y_2\}, \{y_1, y_2\}, \{y_1, y_2, y_3\}\}$.
  2. By definition: $||\{y_2, y_3\} \Rightarrow \{y_1\}||_\emptyset = 1$, $||\{y_2, y_3\} \Rightarrow \{y_1\}||_{\{y_1\}} = 1$, $||\{y_2, y_3\} \Rightarrow \{y_1\}||_{\{y_2\}} = 1$, $||\{y_2, y_3\} \Rightarrow \{y_1\}||_{\{y_1,y_2\}} = 1$, $||\{y_2, y_3\} \Rightarrow \{y_1\}||_{\{y_1,y_2,y_3\}} = 1$. Therefore, $T \models A \Rightarrow B$.
  3. Now, using our theorem: The least model of T containing $A = \{y_2, y_3\}$ is $C_{\mathrm{Mod}(T)}(A) = \{y_1, y_2, y_3\}$. Therefore, to verify $T \models A \Rightarrow B$, we just need to check whether $A \Rightarrow B$ is true in $\{y_1, y_2, y_3\}$. Since $||\{y_2, y_3\} \Rightarrow \{y_1\}||_{\{y_1,y_2,y_3\}} = 1$, we conclude $T \models A \Rightarrow B$.

Models of attribute implications

Example (cntd.)

- $T = \{\{y_3\} \Rightarrow \{y_1, y_2\}, \{y_1, y_3\} \Rightarrow \{y_2\}\}$, $A \Rightarrow B$ is $\{y_2\} \Rightarrow \{y_1\}$.
  1. $\mathrm{Mod}(T) = \{\emptyset, \{y_1\}, \{y_2\}, \{y_1, y_2\}, \{y_1, y_2, y_3\}\}$.
  2. By definition: $||\{y_2\} \Rightarrow \{y_1\}||_\emptyset = 1$, $||\{y_2\} \Rightarrow \{y_1\}||_{\{y_1\}} = 1$, $||\{y_2\} \Rightarrow \{y_1\}||_{\{y_2\}} = 0$, we can stop.
  Therefore, $T \not\models A \Rightarrow B$.
  3. Now, using our theorem: The least model of T containing $A = \{y_2\}$ is $C_{\mathrm{Mod}(T)}(A) = \{y_2\}$. Therefore, to verify $T \models A \Rightarrow B$, we need to check whether $A \Rightarrow B$ is true in $\{y_2\}$. Since $||\{y_2\} \Rightarrow \{y_1\}||_{\{y_2\}} = 0$, we conclude $T \not\models A \Rightarrow B$.

Example

Let $Y = \{y_1, \ldots, y_{10}\}$,
$T = \{\{y_1, y_4\} \Rightarrow \{y_3\}, \{y_2, y_4\} \Rightarrow \{y_1\}, \{y_1, y_2\} \Rightarrow \{y_4, y_7\}, \{y_2, y_7\} \Rightarrow \{y_3\}, \{y_6\} \Rightarrow \{y_4\}, \{y_2, y_8\} \Rightarrow \{y_3\}, \{y_9\} \Rightarrow \{y_1, y_2, y_7\}\}$

1. Decide whether $T \models A \Rightarrow B$ for $A \Rightarrow B$ being $\{y_2, y_5, y_6\} \Rightarrow \{y_3, y_7\}$. We need to check whether $||A \Rightarrow B||_{C_{\mathrm{Mod}(T)}(A)} = 1$. First, we compute $C_{\mathrm{Mod}(T)}(A) = \bigcup_{n=0}^{\infty} A^{T_n}$. Recall:

$$A^{T_n} = A^{T_{n-1}} \cup \bigcup\{D \mid C \Rightarrow D \in T, C \subseteq A^{T_n}\}.$$

- $A^{T_0} = A = \{y_2, y_5, y_6\}$.
- $A^{T_1} = A^{T_0} \cup \bigcup\{\{y_4\}\} = \{y_2, y_4, y_5, y_6\}$.
  Note: $\{y_4\}$ added because for $C \Rightarrow D$ being $\{y_6\} \Rightarrow \{y_4\}$ we have $\{y_6\} \subseteq A^{T_0}$.
- $A^{T_2} = A^{T_1} \cup \bigcup\{\{y_1\}, \{y_4\}\} = \{y_1, y_2, y_4, y_5, y_6\}$.
- $A^{T_3} = A^{T_2} \cup \bigcup\{\{y_3\}, \{y_1\}, \{y_4\}\} = \{y_1, y_2, y_3, y_4, y_5, y_6\}$.
- $A^{T_4} = A^{T_3} \cup \bigcup\{\{y_3\}, \{y_1\}, \{y_4, y_7\}, \{y_4\}\} = \{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}$.

Example (cntd.)

- $A^{T_5} = A^{T_4} \cup \bigcup\{\{y_3\}, \{y_1\}, \{y_4, y_7\}, \{y_4\}\} = \{y_1, y_2, y_3, y_4, y_5, y_6, y_7\} = A^{T_4}$, STOP.

Therefore, $C_{\mathrm{Mod}(T)}(A) = \{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}$. Now, we need to check if $||A \Rightarrow B||_{C_{\mathrm{Mod}(T)}(A)} = 1$, i.e. if

$$||\{y_2, y_5, y_6\} \Rightarrow \{y_3, y_7\}||_{\{y_1,y_2,y_3,y_4,y_5,y_6,y_7\}} = 1.$$

Since this is true, we conclude $T \models A \Rightarrow B$.

2. Decide whether $T \models A \Rightarrow B$ for $A \Rightarrow B$ being $\{y_1, y_2, y_8\} \Rightarrow \{y_4, y_7\}$. We need to check whether $||A \Rightarrow B||_{C_{\mathrm{Mod}(T)}(A)} = 1$. First, we compute $C_{\mathrm{Mod}(T)}(A) = \bigcup_{n=0}^{\infty} A^{T_n}$.

- $A^{T_0} = A = \{y_1, y_2, y_8\}$.
- $A^{T_1} = A^{T_0} \cup \bigcup\{\{y_3\}\} = \{y_1, y_2, y_3, y_8\}$.
- $A^{T_2} = A^{T_1} \cup \bigcup\{\{y_7\}, \{y_3\}\} = \{y_1, y_2, y_3, y_7, y_8\}$.
- $A^{T_3} = A^{T_2} \cup \bigcup\{\{y_7\}, \{y_3\}\} = \{y_1, y_2, y_3, y_7, y_8\} = A^{T_2}$, STOP.

Thus, $C_{\mathrm{Mod}(T)}(A) = \{y_1, y_2, y_3, y_7, y_8\}$. Now, we need to check if $||A \Rightarrow B||_{C_{\mathrm{Mod}(T)}(A)} = 1$, i.e. if $||\{y_1, y_2, y_8\} \Rightarrow \{y_4, y_7\}||_{\{y_1,y_2,y_3,y_7,y_8\}} = 1$. Since this is not true, we conclude $T \not\models A \Rightarrow B$.

Non-redundant bases of attribute implications

Definition (non-redundant set of AIs)

A set T of attribute implications is called non-redundant if for any $A \Rightarrow B \in T$ we have

$$T - \{A \Rightarrow B\} \not\models A \Rightarrow B.$$

That is, if T' results from T be removing an arbitrary $A \Rightarrow B$ from T, then $A \Rightarrow B$ does not semantically follow from T', i.e. T' is weaker than T.
How to check if T is redundant or not? Pseudo-code:

```
for A ⇒ B ∈ T do
    T' := T − {A ⇒ B};
    if T' ⊨ A ⇒ B then
        output(``REDUNDANT'');
        stop;
    endif;
endfor;
output(``NONREDUNDANT'').
```

Non-redundant bases of attribute implications

- Checking $T' \models A \Rightarrow B$: as described above, i.e. test whether $||A \Rightarrow B||_{C_{\mathrm{Mod}(T')}(A)} = 1$.
- Modification of this algorithm gives an algorithm which, given T, returns a non-redundant subset nrT of T which is equally strong as T, i.e. for any $C \Rightarrow D$,

  $$T \models C \Rightarrow D \quad \text{iff} \quad nrT \models C \Rightarrow D.$$

  Pseudo-code:

```
nrT := T;
for A ⇒ B ∈ nrT do
    T' := nrT − {A ⇒ B};
    if T' ⊨ A ⇒ B then
        nrT := T';
    endif;
endfor;
output(nrT).
```

Non-redundant bases of attribute implications

Definition (complete set of AIs)

Let $\langle X, Y, I \rangle$ be a formal context, T be a set of attribute implications over Y. T is called complete in $\langle X, Y, I \rangle$ if for any attribute implication $C \Rightarrow D$ we have

$$C \Rightarrow D \text{ is true in } \langle X, Y, I \rangle \quad \text{IFF} \quad T \models C \Rightarrow D.$$

- This is a different notion of completeness (different from syntactico-semantical completeness of system (Ax) and (Cut) of Armstrong rules).
- Meaning: T is complete iff validity of any AI $C \Rightarrow D$ in data $\langle X, Y, I \rangle$ is encoded in T via entailment: $C \Rightarrow D$ is true in $\langle X, Y, I \rangle$ iff $C \Rightarrow D$ follows from T. That is, T gives complete information about which AIs are true in data.
- Definition directly yields: If T is complete in $\langle X, Y, I \rangle$ then every $A \Rightarrow B$ from T is true in $\langle X, Y, I \rangle$. Why: because $T \models A \Rightarrow B$ for every $A \Rightarrow B$ from T.

Non-redundant bases of attribute implications

Theorem (criterion for T being complete in $\langle X, Y, I\rangle$)

T is complete in $\langle X, Y, I\rangle$ iff $\mathrm{Mod}(T) = \mathrm{Int}(X, Y, I)$, i.e. models of T are just intents of formal concepts from $\mathcal{B}(X, Y, I)$.

Proof.

Omitted. □

Non-redundant bases of attribute implications

Definition (non-redundant basis of $\langle X, Y, I\rangle$)

Let $\langle X, Y, I\rangle$ be a formal context. A set T of attribute implications over Y is called a non-redundant basis of $\langle X, Y, I\rangle$ iff

1. T is complete in $\langle X, Y, I\rangle$,
2. T is non-redundant.

– Another way to say that T is a non-redundant basis of $\langle X, Y, I\rangle$:
(a) every AI from T is true in $\langle X, Y, I\rangle$;
(b) for any other AI $C \Rightarrow D$: $C \Rightarrow D$ is true in $\langle X, Y, I\rangle$ iff $C \Rightarrow D$ follows from T;
(c) no proper subset $T' \subseteq T$ satisfies (a) and (b).

Non-redundant bases of attribute implications

Example (testing non-redundancy of T)

Let $Y = \{ab2, ab6, abs, ac, cru, ebd\}$ with the following meaning of attributes: $ab2$ ... has 2 or more airbags, $ab6$ ... has 6 or more airbags, abs ... has ABS, ac ... has air-conditioning, ebd ... has EBD.
Let T consist of the following attribute implications: $\{ab6\} \Rightarrow \{abs, ac\}$, $\{\} \Rightarrow \{ab2\}$, $\{ebd\} \Rightarrow \{ab6, cru\}$, $\{ab6\} \Rightarrow \{ab2\}$.
Determine whether T is non-redundant.
We can use the above algorithm, and proceed as follows: We go over all $A \Rightarrow B$ from T and test whether $T' \models A \Rightarrow B$ where $T' = T - \{A \Rightarrow B\}$.

- $A \Rightarrow B = \{ab6\} \Rightarrow \{abs, ac\}$. Then, $T' = \{\{\} \Rightarrow \{ab2\}, \{ebd\} \Rightarrow \{ab6, cru\}, \{ab6\} \Rightarrow \{ab2\}\}$. In order to decide whether $T' \models \{ab6\} \Rightarrow \{abs, ac\}$, we need to compute $C_{\mathrm{Mod}(T')}(\{ab6\})$ and check $||\{ab6\} \Rightarrow \{abs, ac\}||_{C_{\mathrm{Mod}(T')}(\{ab6\})}$. Putting $Z = \{ab6\}$, and denoting $Z^{T_i'}$ by Z^i,

Example (testing non-redundancy of T, cntd.)

we get $Z^0 = \{ab6\}$, $Z^1 = \{ab2, ab6\}$, $Z^2 = \{ab2, ab6\}$, we can stop and we have $C_{\mathrm{Mod}(T')}(\{ab6\}) = \bigcup_{i=0^1} Z^i = \{ab2, ab6\}$. Now, $||\{ab6\} \Rightarrow \{abs, ac\}||_{C_{\mathrm{Mod}(T')}(\{ab6\})} = ||\{ab6\} \Rightarrow \{abs, ac\}||_{\{ab2,ab6\}} = 0$, i.e. $T' \not\models \{ab6\} \Rightarrow \{abs, ac\}$. That is, we need to go further.

- $A \Rightarrow B = \{\} \Rightarrow \{ab2\}$. Then, $T' = \{\{ab6\} \Rightarrow \{abs, ac\}, \{ebd\} \Rightarrow \{ab6, cru\}, \{ab6\} \Rightarrow \{ab2\}\}$. In order to decide whether $T' \models \{\} \Rightarrow \{ab2\}$, we need to compute $C_{\mathrm{Mod}(T')}(\{\})$ and check $||\{\} \Rightarrow \{ab2\}||_{C_{\mathrm{Mod}(T')}(\{\})}$. Putting $Z = \{\}$, and denoting $Z^{T_i'}$ by Z^i, we get $Z^0 = \{\}$, $Z^1 = \{\}$ (because there is no $A \Rightarrow B \in T'$ such that $A \subseteq \{\}$), we can stop and we have $C_{\mathrm{Mod}(T')}(\{\}) = Z^0 = \{\}$. Now, $||\{\} \Rightarrow \{ab2\}||_{C_{\mathrm{Mod}(T')}(\{\})} = ||\{\} \Rightarrow \{ab2\}||_{\{\}} = 0$, i.e. $T' \not\models \{\} \Rightarrow \{ab2\}$. That is, we need to go further.
- $A \Rightarrow B = \{ebd\} \Rightarrow \{ab6, cru\}$. Then, $T' = \{\{ab6\} \Rightarrow \{abs, ac\}, \{\} \Rightarrow \{ab2\}, \{ab6\} \Rightarrow \{ab2\}\}$. In order to decide whether $T' \models \{ebd\} \Rightarrow \{ab6, cru\}$, we need to compute

Example (testing non-redundancy of T, cntd.)

$C_{\mathrm{Mod}(T')}(\{ebd\})$ and check $||\{ebd\} \Rightarrow \{ab6, cru\}||_{C_{\mathrm{Mod}(T')}(\{ebd\})}$. Putting $Z = \{ebd\}$, and denoting $Z^{T_i'}$ by Z^i, we get $Z^0 = \{ebd\}$, $Z^1 = \{ab2, ebd\}$, $Z^2 = \{ab2, ebd\}$, we can stop and we have $C_{\mathrm{Mod}(T')}(\{ebd\}) = Z^0 = \{ab2, ebd\}$. Now, $||\{ebd\} \Rightarrow \{ab6, cru\}||_{C_{\mathrm{Mod}(T')}(\{ab2,ebd\})} = ||\{ebd\} \Rightarrow \{ab6, cru\}||_{\{ab2,ebd\}} = 0$, i.e. $T' \not\models \{ebd\} \Rightarrow \{ab6, cru\}$. That is, we need to go further.

- $A \Rightarrow B = \{ab6\} \Rightarrow \{ab2\}$. Then, $T' = \{\{ab6\} \Rightarrow \{abs, ac\}, \{\} \Rightarrow \{ab2\}, \{ebd\} \Rightarrow \{ab6, cru\}\}$. In order to decide whether $T' \models \{ab6\} \Rightarrow \{ab2\}$, we need to compute $C_{\mathrm{Mod}(T')}(\{ab6\})$ and check $||\{ab6\} \Rightarrow \{ab2\}||_{C_{\mathrm{Mod}(T')}(\{ab6\})}$. Putting $Z = \{ab6\}$, and denoting $Z^{T_i'}$ by Z^i, we get $Z^0 = \{ab6\}$, $Z^1 = \{ab2, ab6, abs, ac\}$, $Z^2 = \{ab2, ab6, abs, ac\}$, we can stop and we have $C_{\mathrm{Mod}(T')}(\{ab6\}) = \bigcup_{i=0^1} Z^i = \{ab2, ab6, abs, ac\}$. Now, $||\{ab6\} \Rightarrow \{ab2\}||_{C_{\mathrm{Mod}(T')}(\{ab6\})} = ||\{ab6\} \Rightarrow \{ab2\}||_{\{ab2,ab6,abs,ac\}} = 1$, i.e. $T' \models \{ab6\} \Rightarrow \{ab2\}$. Therefore, T is redundant (we can remove $\{ab6\} \Rightarrow \{ab2\}$).

Example (testing non-redundancy of T, cntd.)

We can see that T is redundant by observing that $T' \vdash \{ab6\} \Rightarrow \{ab2\}$ where $T' = T - \{\{ab6\} \Rightarrow \{ab2\}\}$. Namely, we can infer $\{ab6\} \Rightarrow \{ab2\}$ from $\{\} \Rightarrow \{ab2\}$ by (Wea). Syntactico-semantical completeness yields $T' \models \{ab6\} \Rightarrow \{ab2\}$, hence T is redundant.

Non-redundant bases of attribute implications

Example (deciding whether T is complete w.r.t $\langle X, Y, I\rangle$)

Consider attributes normal blood pressure (nbp), high blood pressure (hbp), watches TV (TV), eats unhealthy food (uf), runs regularly (r), persons $a, \ldots, e$, and formal context (table) $\langle X, Y, I\rangle$

I	nbp	hbp	TV	uf	r
a	×				×
b	×			×	×
c		×	×	×	
d		×		×	
e	×				

Decide whether T is complete w.r.t. $\langle X, Y, I\rangle$ for sets T described below.

Due to the above theorem, we need to check $\mathrm{Mod}(T) = \mathrm{Int}(X, Y, I)$. That is, we need to compute $\mathrm{Int}(X, Y, I)$ and $\mathrm{Mod}(T)$ and compare.

We have $\mathrm{Int}(X, Y, I) = \{\{\}, \{nbp\}, \{uf\}, \{uf, hbp\},$ $\{nbp, r\}, \{uf, hbp, TV\}, \{nbp, r, uf\}, \{hbp, nbp, r, TV, uf\}\}$

Example (deciding whether T is complete w.r.t $\langle X, Y, I\rangle$, cntd.)

1. T consists of $\{r\} \Rightarrow \{nbp\}$, $\{TV, uf\} \Rightarrow \{hbp\}$, $\{r, uf\} \Rightarrow \{TV\}$. T is not complete w.r.t. $\langle X, Y, I\rangle$ because $\{r, uf\} \Rightarrow \{TV\}$ is not true in $\langle X, Y, I\rangle$ (person b is a counterexample). Recall that if T is complete, every AI from T is true in $\langle X, Y, I\rangle$.
2. T consists of $\{r\} \Rightarrow \{nbp\}$, $\{TV, uf\} \Rightarrow \{hbp\}$, $\{TV\} \Rightarrow \{hbp\}$. In this case, every AI from T is true in $\langle X, Y, I\rangle$. But still, T is not complete. Namely, $\mathrm{Mod}(T) \not\subseteq \mathrm{Int}(X, Y, I)$. For instance, $\{hbp, TV\} \in \mathrm{Mod}(T)$ but $\{hbp, TV\} \notin \mathrm{Int}(X, Y, I)$.

   In this case, T is too weak. T does not entail all attribute implications which are true in $\langle X, Y, I\rangle$. For instance $\{hbp, TV\} \Rightarrow \{uf\}$ is true in $\langle X, Y, I\rangle$ but $T \not\models \{hbp, TV\} \Rightarrow \{uf\}$. Indeed, $\{hbp, TV\}$ is a model of T but $||\{hbp, TV\} \Rightarrow \{uf\}||_{\{hbp, TV\}} = 0$.

Example (deciding whether T is complete w.r.t $\langle X, Y, I\rangle$, cntd.)

3. T consists of $\{r\} \Rightarrow \{nbp\}$, $\{TV, uf\} \Rightarrow \{hbp\}$, $\{TV\} \Rightarrow \{uf\}$, $\{TV\} \Rightarrow \{hbp\}$, $\{hbp, TV\} \Rightarrow \{uf\}$, $\{nbp, uf\} \Rightarrow \{r\}$, $\{hbp\} \Rightarrow \{uf\}$, $\{uf, r\} \Rightarrow \{nbp\}$, $\{nbp, TV\} \Rightarrow \{r\}$, $\{hbp, nbp\} \Rightarrow \{r, TV\}$.
   One can check that $\mathrm{Mod}(T)$ consists of $\{\}$, $\{nbp\}$, $\{uf\}$, $\{uf, hbp\}$, $\{nbp, r\}$, $\{uf, hbp, TV\}$, $\{nbp, r, uf\}$, $\{hbp, nbp, r, TV, uf\}\}$.
   Therefore, $\mathrm{Mod}(T) = \mathrm{Int}(X, Y, I)$. This implies that T is complete in $\langle X, Y, I\rangle$.

   (An easy way to check it is to check that every intent from $\mathrm{Int}(X, Y, I)$ is a model of T (there are 8 intents in our case), and that no other subset of Y is a model of T (there are $2^5 - 8 = 24$ such subsets in our case). As an example, take $\{hbp, uf, r\} \notin \mathrm{Int}(X, Y, I)$. $\{hbp, uf, r\}$ is not a model of T because $\{hbp, uf, r\}$ is not a model of $\{r\} \Rightarrow \{nbp\}$.)

Example (reducing T to a non-redundant set)

Continuing our previous example, consider again T consisting of $\{r\} \Rightarrow \{nbp\}$, $\{TV, uf\} \Rightarrow \{hbp\}$, $\{TV\} \Rightarrow \{uf\}$, $\{TV\} \Rightarrow \{hbp\}$, $\{hbp, TV\} \Rightarrow \{uf\}$, $\{nbp, uf\} \Rightarrow \{r\}$, $\{hbp\} \Rightarrow \{uf\}$, $\{uf, r\} \Rightarrow \{nbp\}$, $\{nbp, TV\} \Rightarrow \{r\}$, $\{hbp, nbp\} \Rightarrow \{r, TV\}$.
From the previous example we know that T is complete in $\langle X, Y, I\rangle$.
Check whether T is non-redundant. If not, transform T into a non-redundant set nrT. (Note: nrT is then a non-redundant basis of $\langle X, Y, I\rangle$.)

Using the above algorithm, we put $nrT := T$ and go through all $A \Rightarrow B \in nrT$ and perform: If for $T' := nrT - \{A \Rightarrow B\}$ we find out that $T' \models A \Rightarrow B$, we remove $A \Rightarrow B$ from nrT, i.e. we put $nrT := T'$.
Checking $T' \models A \Rightarrow B$ is done by verifying whether $||A \Rightarrow B||_{C_{\mathrm{Mod}(T')}(A)}$.

- For $A \Rightarrow B = \{r\} \Rightarrow \{nbp\}$: $T' := nrT - \{\{r\} \Rightarrow \{nbp\}\}$, $C_{\mathrm{Mod}(T')}(A) = \{r\}$ and $||A \Rightarrow B||_{\{r\}} = 0$, thus $T' \not\models A \Rightarrow B$, and nrT does not change.

Example (reducing T to a non-redundant set, cntd.)

- For $A \Rightarrow B = \{TV, uf\} \Rightarrow \{hbp\}$: $T' := nrT - \{\{TV, uf\} \Rightarrow \{hbp\}\}$, $C_{\mathrm{Mod}(T')}(A) = \{TV, uf, hbp\}$ and $||A \Rightarrow B||_{\{TV,uf,hbp\}} = 1$, thus $T' \models A \Rightarrow B$, and we remove $\{TV, uf\} \Rightarrow \{hbp\}$ from nrT. That is, $nrT = T - \{\{TV, uf\} \Rightarrow \{hbp\}\}$.
- For $A \Rightarrow B = \{TV\} \Rightarrow \{uf\}$: $T' := nrT - \{\{TV\} \Rightarrow \{uf\}\}$, $C_{\mathrm{Mod}(T')}(A) = \{TV, hbp, uf\}$ and $||A \Rightarrow B||_{\{TV,hbp,uf\}} = 1$, thus $T' \models A \Rightarrow B$, and we remove $\{TV\} \Rightarrow \{uf\}$ from nrT. That is, $nrT = T - \{\{TV, uf\} \Rightarrow \{hbp\}, \{TV\} \Rightarrow \{uf\}\}$.
- For $A \Rightarrow B = \{TV\} \Rightarrow \{hbp\}$: $T' := nrT - \{\{TV\} \Rightarrow \{hbp\}\}$, $C_{\mathrm{Mod}(T')}(A) = \{TV\}$ and $||A \Rightarrow B||_{\{TV\}} = 0$, thus $T' \not\models A \Rightarrow B$, nrT does not change. That is, $nrT = T - \{\{TV, uf\} \Rightarrow \{hbp\}, \{TV\} \Rightarrow \{uf\}\}$.
- For $A \Rightarrow B = \{hbp, TV\} \Rightarrow \{uf\}$: $T' := nrT - \{\{hbp, TV\} \Rightarrow \{uhf\}\}$, $C_{\mathrm{Mod}(T')}(A) = \{hbp, TV, uf\}$ and $||A \Rightarrow B||_{\{hbp,TV,uf\}} = 1$, thus $T' \models A \Rightarrow B$, we remove $\{hbp, TV\} \Rightarrow \{uf\}$ from nrT. That is, $nrT = T - \{\{TV, uf\} \Rightarrow \{hbp\}, \{TV\} \Rightarrow \{uf\},$

Example (reducing T to a non-redundant set, cntd.)

- For $A \Rightarrow B = \{nbp, uf\} \Rightarrow \{r\}$: $T' := nrT - \{\{nbp, uf\} \Rightarrow \{r\}\}$, $C_{\mathrm{Mod}(T')}(A) = \{nbp, uf\}$ and $||A \Rightarrow B||_{\{nbp,uf\}} = 0$, thus $T' \not\models A \Rightarrow B$ and nrT does not change. That is, $nrT = T - \{\{TV, uf\} \Rightarrow \{hbp\}, \{TV\} \Rightarrow \{uf\}, \{hbp, TV\} \Rightarrow \{uf\}\}$.
- For $A \Rightarrow B = \{hbp\} \Rightarrow \{uf\}$: $T' := nrT - \{\{hbp\} \Rightarrow \{uf\}\}$, $C_{\mathrm{Mod}(T')}(A) = \{hbp\}$ and $||A \Rightarrow B||_{\{hbp\}} = 0$, thus $T' \not\models A \Rightarrow B$ and nrT does not change. That is, $nrT = T - \{\{TV, uf\} \Rightarrow \{hbp\}, \{TV\} \Rightarrow \{uf\}, \{hbp, TV\} \Rightarrow \{uf\}\}$.
- For $A \Rightarrow B = \{uf, r\} \Rightarrow \{nbp\}$: $T' := nrT - \{\{uf, r\} \Rightarrow \{nbp\}\}$, $C_{\mathrm{Mod}(T')}(A) = \{uf, r, nbp\}$ and $||A \Rightarrow B||_{\{uf,r,nbp\}} = 1$, thus $T' \models A \Rightarrow B$ and we remove $\{uf, r\} \Rightarrow \{nbp\}$ from nrT. That is, $nrT = T - \{\{TV, uf\} \Rightarrow \{hbp\}, \{TV\} \Rightarrow \{uf\}, \{hbp, TV\} \Rightarrow \{uf\}, \{uf, r\} \Rightarrow \{nbp\}\}$.

Example (reducing T to a non-redundant set, cntd.)

- For $A \Rightarrow B = \{nbp, TV\} \Rightarrow \{r\}$: $T' := nrT - \{\{nbp, TV\} \Rightarrow \{r\}\}$, $C_{\mathrm{Mod}(T')}(A) = \{nbp, TV, hbp, uf, r\}$ and $||A \Rightarrow B||_{\{nbp,TV,hbp,uf,r\}} = 1$, thus $T' \models A \Rightarrow B$ and we remove $\{nbp, TV\} \Rightarrow \{r\}$ from nrT. That is, $nrT = T - \{\{TV, uf\} \Rightarrow \{hbp\}, \{TV\} \Rightarrow \{uf\}, \{hbp, TV\} \Rightarrow \{uf\}, \{uf, r\} \Rightarrow \{nbp\}, \{nbp, TV\} \Rightarrow \{r\}\}$.
- For $A \Rightarrow B = \{hbp, hbp\} \Rightarrow \{r, TV\}$: $T' := nrT - \{\{hbp, nbp\} \Rightarrow \{r, TV\}\}$, $C_{\mathrm{Mod}(T')}(A) = \{hbp, nbp, uf, r\}$ and $||A \Rightarrow B||_{\{hbp,nbp,uf,r\}} = 0$, thus $T' \not\models A \Rightarrow B$ and nrT does not change. That is, $nrT = T - \{\{TV, uf\} \Rightarrow \{hbp\}, \{TV\} \Rightarrow \{uf\}, \{hbp, TV\} \Rightarrow \{uf\}, \{uf, r\} \Rightarrow \{nbp\}, \{nbp, TV\} \Rightarrow \{r\}\}$.

We obtained $nrT = \{\{r\} \Rightarrow \{nbp\}, \{TV\} \Rightarrow \{hbp\}, \{nbp, uf\} \Rightarrow \{r\}, \{hbp\} \Rightarrow \{uf\}, \{hbp, nbp\} \Rightarrow \{r, TV\}\}$. nrT is a non-redundant set of AIs.

Since T is complete in $\langle X, Y, I\rangle$, nrT is complete in $\langle X, Y, I\rangle$, too (why?). Therefore, nrT is a non-redundant basis of $\langle X, Y, I\rangle$.

Non-redundant bases of attribute implications

In the last example, we obtained a non-redundant basis nrT of $\langle X, Y, I\rangle$, $nrT = \{\{r\} \Rightarrow \{nbp\}, \{TV\} \Rightarrow \{hbp\}, \{nbp, uf\} \Rightarrow \{r\}, \{hbp\} \Rightarrow \{uf\}, \{hbp, nbp\} \Rightarrow \{r, TV\}\}$.
How to compute non-redundant bases from data?

We are going to present an approach based on the notion of a pseudo-intent. This approach is due to Guigues and Duquenne. The resulting non-redundant basis is called a Guigues-Duquenne basis.

Two main features of Guigues-Duquenne basis are

- it is computationally tractable,
- it is optimal in terms of its size (no other non-redundant basis has is smaller in terms of the number of AIs it contains).

Non-redundant bases of attribute implications

Definition (pseudo-intents)

A pseudo-intent of $\langle X, Y, I\rangle$ is a subset $A \subseteq Y$ for which

1. $A \neq A^{\downarrow\uparrow}$,
2. $B^{\downarrow\uparrow} \subseteq A$ for each pseudo-intent $B \subset A$.

–

Theorem (Guigues-Duquenne basis)

The set $T = \{A \Rightarrow A^{\downarrow\uparrow} \mid A$ *is a pseudointent of* $(X, Y, I)\}$ *of implications is a non-redundant basis of* $\langle X, Y, I\rangle$.

Proof.

We show that T is complete and non-redundant.

Complete: It suffices to show that $\mathrm{Mod}(T) \subseteq \mathrm{Int}(X, Y, I)$. Let $C \in \mathrm{Mod}(T)$. Assume $C \neq C^{\downarrow\uparrow}$. Then C is a pseudo-intent (indeed, if $P \subset C$ is a pseudo-intent then since $||P \Rightarrow P^{\downarrow\uparrow}||_C = 1$, we get $P^{\downarrow\uparrow} \subseteq C$). But then $C \Rightarrow C^{\downarrow\uparrow} \in T$ and so $||C \Rightarrow C^{\downarrow\uparrow}||_C = 1$. But the last fact means that if $C \subseteq C$ (which is true) then $C^{\downarrow\uparrow} \subseteq C$ which would give $C^{\downarrow\uparrow} = C$, a contradiction with the assumption $C^{\downarrow\uparrow} \neq C$. Therefore, $C^{\downarrow\uparrow} = C$, i.e. $C \in \mathrm{Int}(X, Y, I)$.

Non-redundant: Take any $P \Rightarrow P^{\downarrow\uparrow}$. We show that $T - \{P \Rightarrow P^{\downarrow\uparrow}\} \not\models P \Rightarrow P^{\downarrow\uparrow}$. Since $||P \Rightarrow P^{\downarrow\uparrow}||_P = 0$ (obvious, check), it suffices to show that $||T - \{P \Rightarrow P^{\downarrow\uparrow}\}||_P = 1$. That is, we need to show that for each $Q \Rightarrow Q^{\downarrow\uparrow} \in T - \{P \Rightarrow P^{\downarrow\uparrow}\}$ we have $||Q \Rightarrow Q^{\downarrow\uparrow}||_P = 1$, i.e. that if $Q \subseteq P$ then $Q^{\downarrow\uparrow} \subseteq P$. But this follows from the definition of a pseudo-intent (apply to P). □

Lemma

If P, Q are intents or pseudo-intents and $P \not\subseteq Q$, $Q \not\subseteq P$, then $P \cap Q$ is an intent.

Proof.

Let $T = \{R \Rightarrow R^{\downarrow\uparrow} \mid R \text{ a pseudo-intent}\}$ be the G.-D. basis. Since T is complete, it is sufficient to show that $P \cap Q \in \mathrm{Mod}(T)$ (since then, $P \cap Q$ is a model of any implication which is true in $\langle X, Y, I \rangle$, and so $P \cap Q$ is an intent).
Obviously, P, Q are models of $T - \{P \Rightarrow P^{\downarrow\uparrow}, Q \Rightarrow Q^{\downarrow\uparrow}\}$, whence $P \cap Q$ is a model of $T - \{P \Rightarrow P^{\downarrow\uparrow}, Q \Rightarrow Q^{\downarrow\uparrow}\}$ (since the set of models is a closure system, i.e. closed under intersections).
Therefore, to show that $P \cap Q$ is a model of T, it is sufficient to show that $P \cap Q$ is a model of $\{P \Rightarrow P^{\downarrow\uparrow}, Q \Rightarrow Q^{\downarrow\uparrow}\}$. Due to symmetry, we only verify that $P \cap Q$ is a model of $\{P \Rightarrow P^{\downarrow\uparrow}$: But this is trivial: since $P \not\subseteq Q$, the condition "if $P \subseteq P \cap Q$ implies $P^{\downarrow\uparrow} \subseteq P \cap Q$" is satisfied for free. The proof is complete. □

Lemma

If T is complete, then for each pseudo-intent P, T contains $A \Rightarrow B$ with $A^{\downarrow\uparrow} = P^{\downarrow\uparrow}$

Proof.

For pseudointent P, $P \neq P^{\downarrow\uparrow}$, i.e. P is not an intent. Therefore, P cannot be a model of T (since models of a complete T are intents). Therefore, there is $A \Rightarrow B \in T$ such that $||A \Rightarrow B||_P = 0$, i.e. $A \subseteq P$ but $B \not\subseteq P$. As $||A \Rightarrow B||_{\langle X,Y,I \rangle} = 1$, we have $B \subseteq A^{\downarrow\uparrow}$ (Thm. on basic connections ...). Therefore, $A^{\downarrow\uparrow} \not\subseteq P$ (otherwise $B \subseteq P$, a contradiction). Therefore, $A^{\downarrow\uparrow} \cap P$ is not an intent (). By the foregoing Lemma, $P \subseteq A^{\downarrow\uparrow}$ which gives $P^{\downarrow\uparrow} \subseteq A^{\downarrow\uparrow}$. On the other hand, $A \subseteq P$ gives $A^{\downarrow\uparrow} \subseteq P^{\downarrow\uparrow}$. Altogether, $A^{\downarrow\uparrow} = P^{\downarrow\uparrow}$, proving the claim. □

Theorem (Guigues-Duquenne basis is the smalest one)

If T is the Guigues-Duquenne base and T' is complete then $|T| \leq |T'|$.

Proof.

Direct corollary of the above Lemma. □

Non-redundant bases of attribute implications

$\mathcal{P}$ ... set of all pseudointents of $\langle X, Y, I\rangle$
THE base we need to compute: $\{A \Rightarrow A^{\downarrow\uparrow} \mid A \in \mathcal{P}\}$

Q: What do we need? A: Compute all pseudointents.

We will see that the set of all $P \subseteq Y$ which are intents or pseudo-intents is a closure system.

Q: How to compute the fixed points (closed sets)?

For $Z \subseteq Y$, T a set of implications, put
$Z^T = Z \cup \bigcup\{B \mid A \Rightarrow B \in T, A \subset Z\}$
$Z^{T_0} = Z$
$Z^{T_n} = (Z^{T_{n-1}})^T$ $(n \geq 1)$

define $C_T : 2^Y \to 2^Y$ by
$C_T(Z) = \bigcup_{n=0}^{\infty} Z^{T_n}$ (note: terminates, Y finite)

Note: this is different from the operator computing the least model $C_{\mathrm{Mod}(T)}(A)$ of T containing A (instead of $A \subseteq Z$, we have $A \subset Z$ here).

Non-redundant bases of attribute implications

Theorem

Let $T = \{A \Rightarrow A^{\downarrow\uparrow} \mid A \in \mathcal{P}\}$ (G.-D. base). Then

1. *C_T is a closure operator,*
2. *P is a fixed point of C_T iff $P \in \mathcal{P}$ (P is a pseudo-intent) or $P \in \mathrm{Int}(X, Y, I)$ (P is an intent).*

Proof.

1. easy (analogous to the proof concerning the closure operator for $C_{\mathrm{Mod}(T)}(A)$).

2. $\mathcal{P} \cup \mathrm{Int}(X, Y, I) \subseteq \mathrm{fix}(C_T)$: easy. $\mathrm{fix}(C_T) \subseteq \mathcal{P} \cup \mathrm{Int}(X, Y, I)$: It suffices to show that if $P \in \mathrm{fix}(C_T)$ is not an intent ($P \neq P^{\downarrow\uparrow}$) then P is an pseudo-intent. So take $P \in \mathrm{fix}(C_T)$, i.e. $P = C_T(P)$, which is not an intent. Take any pseudointent $Q \subset P$. By definition (notice that $Q \Rightarrow Q^{\downarrow\uparrow} \in T$), $Q^{\downarrow\uparrow} \subseteq C_T(P) = P$ which means that P is a pseudo-intent. □

So: $\mathrm{fix}(C_T) = \mathcal{P} \cup \mathrm{Int}(X, Y, I)$

Therefore, to compute $\mathcal{P}$, we can compute $\mathrm{fix}(C_T)$ and exclude $\mathrm{Int}(X, Y, I)$, i.e. $\mathcal{P} = \mathrm{fix}(C_T) - \mathrm{Int}(X, Y, I)$.

computing $\mathrm{fix}(C_T)$: by Ganter's NextClosure algorithm.

Caution! In order to compute C_T, we need T, i.e. we need $\mathcal{P}$, which **we do not know in advance.** Namely, recall what we are doing:

- Given input data $\langle X, Y, I \rangle$, we need to compute G.-D. basis $T = \{A \Rightarrow A^{\downarrow\uparrow} \mid A \in \mathcal{P}\}$.
- For this, we need to compute $\mathcal{P}$ (pseudo-intents of $\langle X, Y, I \rangle$).
- $\mathcal{P}$ can be obtained from $z\mathrm{fix}(C_T)$ (fixed points of C_T).
- But to compute C_T, we need T (actually, we need only a part of T).

But we are not in *circulus vitiosus*: The part of T (or $\mathcal{P}$) which is needed at a given point is already available (computed) at that point.

Non-redundant bases of attribute implications

Computing G.-D. basis manually is tedious.

Algorithms available, e.g. Peter Burmeister's ConImp (see the course web page).

Association rules

- classic topic in mining relational data
- available in most data mining software tools
- association rules = attribute implications + criteria of interestingness (support, confidence)
- introduced in 1993 (Agrawal R., Imielinski T., Swami A. N.: Mining association rules between sets of items in large databases. *Proc. ACM Int. Conf. of management of data*, pp. 207–216, 1993)
- but see GUHA method (in fact, association rules with elaborated statistics):
  - developed in 1960s by P. Hájek et al.
  - GUHA book available at `http://www.cs.cas.cz/~hajek/guhabook/`: Hájek P., Havránek T.: Mechanizing Hypothesis Formation. Mathematical Foundations for General Theory. Springer, 1978.

Association rules

- Good book: Adamo J.-M.: Data Mining for Association Rules and Sequential Patterns. Sequential and Parallel Algorithms. Springer, New York, 2001.
- Good overview: Dunham M. H.: Data Mining. Introductory and Advanced Topics. Prentice Hall, Upper Saddle River, NJ, 2003.
- Overview in almost any textbook on data mining.

Main point where association rules (ARs) differ from attribute implications (AIs): ARs consider statistical relevance. Therefore, ARs are appropriate when analyzing large data collections.

Association rules – basic terminology

Definition (association rule)

An association rule (over set Y of attributes) is an expression $A \Rightarrow B$ where $A, B \subseteq Y$ (sometimes one assumes $A \cap B = \emptyset$).

Note: Association rules are just attribute implications in sense of FCA.

Data for ARs (terminology in DM community): a set Y of **items**, a **database** D **of transactions**, $D = \{t_1, \ldots, t_n\}$ where $t_i \subseteq Y$, i.e., transaction t_i is a set of (some) items.

Note: one-to-one correspondence between databases D (over Y) and formal contexts (with attributes from Y):
Given D, the corresponding $\langle X, Y, I\rangle_D$ is given by

$$\langle X, Y, I\rangle_D \ldots X = D, \ \langle t_1, y\rangle \in I \Leftrightarrow y \in t_1;$$

given $\langle X, Y, I\rangle$, the corresponding $D_{\langle X, Y, I\rangle}$ is given by

$$D_{\langle X, Y, I\rangle} = \{\{x\}^{\uparrow} \mid x \in X\}.$$

Association rules – why items and transactions?

original motivation:
item = product in a store
transaction = cash register transaction (set of items purchased)
association rule = says: when all items from A are purchased then also all items from B are purchased

Example transactions $X = \{x_1, \ldots, x_8\}$, items $Y = \{be, br, je, mi, pb\}$ (beer, bread, jelly, milk, peanut butter)

I	be	br	je	mi	pb
x_1		X	X		X
x_2		X			X
x_3		X		X	X
x_4	X	X			
x_5	X			X	
x_6		X	X		X
x_7		X	X		X
x_8		X		X	X

For instance: a customer realizing transaction x_3 bought bread, milk, and peanut butter.

Association rules – support

Definition (support of AR)

Support of $A \Rightarrow B$ is denoted by $\mathrm{supp}(A \Rightarrow B)$ and defined by

$$\mathrm{supp}(A \Rightarrow B) = \frac{|\{x \in X \mid \text{for each } y \in A \cup B : \langle x, y\rangle \in I\}|}{|X|},$$

i.e. $\mathrm{supp}(A \Rightarrow B) \cdot 100\%$ of transactions contain $A \cup B$ (percentage of transactions where customers bought items from $A \cup B$).

Note that (in terms of FCA)

$$\mathrm{supp}(A \Rightarrow B) = \frac{|(A \cup B)^{\downarrow}|}{|X|}.$$

We use both "support is 0.3" and "support is 30%".

Association rules – confidence

Definition (confidence of AR)

Confidence of $A \Rightarrow B$ is denoted by $\mathrm{conf}(A \Rightarrow B)$ and defined by

$$\mathrm{conf}(A \Rightarrow B) = \frac{|\{x \in X \mid \text{for each } y \in A \cup B : \langle x, y \rangle \in I\}|}{|\{x \in X \mid \text{for each } y \in A : \langle x, y \rangle \in I\}|},$$

i.e. $\mathrm{conf}(A \Rightarrow B) \cdot 100\%$ of transactions containing all items from A contain also all items from B (percentage of customers which by also (all from) B if they buy (all from) A.

Note that (in terms of FCA)

$$\mathrm{conf}(A \Rightarrow B) = \frac{|(A \cup B)^{\downarrow}|}{A^{\downarrow}}.$$

We use both "confidence is 0.3" and "confidence is 30%".

Lemma

$\mathrm{supp}(A \Rightarrow B) \leq \mathrm{conf}(A \Rightarrow B)$.

Proof.

Directly from definition observing that $|X| \geq |A^{\downarrow}|$ □

Lemma (relating confidence and validity of AIs)

$\mathrm{conf}(A \Rightarrow B) = 1$ iff $||A \Rightarrow B||_{\langle X,Y,I\rangle} = 1$. That is, attribute implications which are true in $\langle X, Y, I\rangle$ are those which are fully confident.

Proof.

$\mathrm{conf}(A \Rightarrow B) = 1$ iff $|(A \cup B)^{\downarrow}| = |A^{\downarrow}|$. Since $(A \cup B)^{\downarrow} \subseteq A^{\downarrow}$ is always the case, $|(A \cup B)^{\downarrow}| = |A^{\downarrow}|$ is equivalent to $(A \cup B)^{\downarrow} \supseteq A^{\downarrow}$ which any object which has all attributes from A (object from $A^{\downarrow}$) has also all attributes from $A \cup B$ (object from $(A \cup B)^{\downarrow}$), thus, in particular, all attributes from B which is equivalent to $||A \Rightarrow B||_{\langle X,Y,I\rangle} = 1$. □

Example (support and confidence)

Consider data table fro previous example (be, br, je, mi, pb denote beer, bread, jelly, milk, peanut butter).

I	be	br	je	mi	pb
x_1		X	X		X
x_2		X			X
x_3		X		X	X
x_4	X	X			
x_5	X			X	
x_6		X	X		X
x_7		X	X		X
x_8		X		X	X

Determine support and confidence of the following association rules:

- $A \Rightarrow B$ is $\{br\} \Rightarrow \{pb\}$:
  $\text{supp}(A \Rightarrow B) = \frac{|(A \cup B)^{\downarrow}|}{|X|} = \frac{|\{br, pb\}^{\downarrow}|}{8} = \frac{|\{x_1, x_2, x_3, x_6, x_7, x_8\}^{\downarrow}|}{8} = \frac{6}{8} = 0.75.$
  $\text{conf}(A \Rightarrow B) = \frac{|(A \cup B)^{\downarrow}|}{|A^{\downarrow}|} = \frac{|\{br, pb\}^{\downarrow}|}{|\{br\}^{\downarrow}|} = \frac{|\{x_1, x_2, x_3, x_6, x_7, x_8\}|}{|\{x_1, x_2, x_3, x_4, x_6, x_7, x_8\}|} = \frac{6}{7} =$ 0.857.

Example (support and confidence, cntd.)

I	be	br	je	mi	pb
x_1		X	X		X
x_2		X			X
x_3		X		X	X
x_4	X	X			
x_5	X			X	
x_6		X	X		X
x_7		X	X		X
x_8		X		X	X

- $A \Rightarrow B$ is $\{mi, pb\} \Rightarrow \{br\}$:
  $\mathrm{supp}(A \Rightarrow B) = \frac{|(A \cup B)^{\downarrow}|}{|X|} = \frac{|\{mi,pb,br\}^{\downarrow}|}{8} = \frac{|\{x_3,x_8\}|}{8} = \frac{2}{8} = 0.25.$
  $\mathrm{conf}(A \Rightarrow B) = \frac{|(A \cup B)^{\downarrow}|}{|A^{\downarrow}|} = \frac{|\{mi,pb,br\}^{\downarrow}|}{|\{mi,pb\}^{\downarrow}|} = \frac{|\{x_3,x_8\}|}{|\{x_3,x_8\}|} = \frac{2}{2} = 1.0.$
- $A \Rightarrow B$ is $\{br, je\} \Rightarrow \{pb\}$:
  $\mathrm{supp}(A \Rightarrow B) = \frac{|(A \cup B)^{\downarrow}|}{|X|} = \frac{|\{br,je,pb\}^{\downarrow}|}{8} = \frac{|\{x_1,x_6,x_7\}|}{8} = \frac{3}{8} = 0.375.$
  $\mathrm{conf}(A \Rightarrow B) = \frac{|(A \cup B)^{\downarrow}|}{|A^{\downarrow}|} = \frac{|\{br,je,pb\}^{\downarrow}|}{|\{br,je\}^{\downarrow}|} = \frac{|\{x_1,x_6,x_7\}|}{|\{x_1,x_6,x_7\}|} = \frac{3}{3} = 1.0.$

  Both $\{mi, pb\} \Rightarrow \{br\}$ and $\{br, je\} \Rightarrow \{pb\}$ are fully confident (true) but $\{br, je\} \Rightarrow \{pb\}$ is supported more by the data (occurred more frequently).

Association rules

Definition (association rule problem)

For prescribed values s and c, list all association rules of $\langle X, Y, I\rangle$ with $\mathrm{supp}(A \Rightarrow B) \geq s$ and $\mathrm{conf}(A \Rightarrow B) \geq c$.

– such rules = interesting rules

– common technique to solve AR problem: via frequent itemsets

1. find all frequent itemsets (see later),
2. generate rules from frequent itemsets

Definition (support of itemset, frequent itemset)

– Support $\mathrm{supp}(B)$ of $B \subseteq Y$ in table $\langle X, Y, I\rangle$ is defined by

$$\mathrm{supp}(B) = \frac{|B^{\downarrow}|}{|X|}.$$

– For given s, an itemset (set of attributes) $B \subseteq Y$ is called frequent (large) itemset if $\mathrm{supp}(B) \geq s$.

Association rules

Note: $\mathrm{supp}(A \Rightarrow B) = \mathrm{supp}(A \cup B)$.

Example

List the set L of all frequent itemsets of the following table $\langle X, Y, I \rangle$ for $s = 0.3$ (30%).

I	be	br	je	mi	pb
x_1		X	X		X
x_2		X			X
x_3		X		X	X
x_4	X	X			
x_5	X			X	

$$L = \{\{be\}, \{br\}, \{mi\}, \{pb\}, \{br, pb\}\}.$$

step 2.: from frequent itemsets to confident ARs

```
"input" <X,Y,I>, L (set of all frequent itemsets), s
(support), c (confidence)

"output" R (set of all asociation rules satisfying s and c)

"algorithm (ARGen)"
1.  R:=O; //empty set
2.  for each l in L do
3.     for each nonempty proper subset k of l do
4.        if supp(l)/supp(k) >= c then
5.              add rule k=>(l-k) to R
```

Observe: $\mathrm{supp}(l)/\mathrm{supp}(k) = \mathrm{conf}(k \Rightarrow l - k)$ (verify)

Note: k is a proper subset of l if $k \subset l$, i.e. $k \subseteq l$ and there exists $y \in l$ such that $y \notin k$.

step 2.: from frequent itemsets to confident ARs

Example

Consider the following table and parameters $s = 0.3$ (support) and $c = 0.8$ (confidence).

I	be	br	je	mi	pb
x_1		X	X		X
x_2		X			X
x_3		X		X	X
x_4	X	X			
x_5	X			X	

From previous example we know that the set L of all frequent itemsets is

$$L = \{\{be\}, \{br\}, \{mi\}, \{pb\}, \{br, pb\}\}.$$

Take $l = \{br, pb\}$; there are two nonempty subsets k of l: $k = \{br\}$ and $k = \{pb\}$.

Rule $br \Rightarrow pb$ IS NOT interesting since

$$\mathrm{supp}(\{br, pb\})/\mathrm{supp}(\{br\}) = 0.6/0.8 = 0.75 \not\geq c$$

while $pb \Rightarrow br$ IS interesting since

$$\mathrm{supp}(\{pb, br\})/\mathrm{supp}(\{pb\}) = 0.6/0.6 = 1.0 \geq c.$$

step 1.: generating frequent itemsets

Generating frequent itemsets is based on

Theorem (apriori principle)

Any subset of a frequent itemset is frequent. If an itemset is not frequent then no of its supersets is frequent.

Proof.

Obvious. □

step 1.: generating frequent itemsets

basic idea of apriori algorithm:

- Given $\langle X, Y, I\rangle$ and s (support), we want to generate the set L of all frequent itemsets, i.e. $L = \{B \subseteq Y \mid \mathrm{supp}(B) \geq s\}$.
- Think of L as
$$L = L_1 \cup L_2 \cup \cdots \cup L_{|Y|}$$
where $L_i = \{B \subseteq Y \mid \mathrm{supp}(B) \geq s \text{ and } |B| = i\}$, i.e. L_i is the set of all frequent itemsets of size i.
- Apriori generates L_1, then L_2, then $\ldots L_{|Y|}$.
- Generating L_i from L_{i-1} – using set C_i of all itemsets of size i which are candidates for being frequent (see later):
    1. in step i, compute C_i from L_{i-1} (if $i = 1$, put $C_1 = \{\{y\} \mid y \in Y\}$);
    2. scanning $\langle X, Y, I\rangle$, generate L_i, the set of all those candidates from C_i which are frequent.

step 1.: generating frequent itemsets

How to get candidates C_i from frequent items L_{i-1}?

- what means "a candidate": candidates are constructed by union of two frequent sets; the underlying idea: proper subsets of candidate shall be frequent,
- this is drawn from the above apriori principle (all subsets of a frequent itemset are frequent),

Getting C_i from L_{i-1}:
find all $B_1, B_2 \in L_{i-1}$ such that $|B_1 - B_2| = 1$ and $|B_2 - B_1| = 1$ (i.e. $|B_1 \cap B_2| = i - 2$), and add $B_1 \cup B_2$ to C_i.

Is this correct? Next lemma says that C_i is guaranteed to contain L_i (all frequent subsets of size i).

Lemma (getting C_i from L_i)

If L_{i-1} is the set of all frequent itemsets of size $i-1$ then for every $B \in L_i$ we have $B = B_1 \cup B_2$ for some $B_1, B_2 \in L_{i-1}$ such that $|B_1 - B_2| = 1$ and $|B_2 - B_1| = 1$. Moreover, $|B_1 - B_2| = 1$ and $|B_2 - B_1| = 1$ iff $|B_1 \cap B_2| = i - 2$.

Proof.

First, check $|B_1 - B_2| = 1$ and $|B_2 - B_1| = 1$ iff $|B_1 \cap B_2| = i - 2$: We have $|B_1| = |B_2| = i - 1$. $|B_1 - B_2| = 1$ means exactly on element from B_1 is not in B_2 (all other $i - 2$ elements of B_1 are in B_2). $|B_2 - B_1| = 1$ means exactly on element from B_2 is not in B_1 (all other $i - 2$ elements of B_2 are in B_1). As a result B_1 and B_2 need to have $i - 2$ elements in common, i.e. $|B_1 \cap B_2| = i - 2$.

Second: Let $B \in L_i$ (B is frequent and $|B| = i$). Pick distinct $y, z \in B$ and consider $B_1 = B - \{y\}$ and $B_2 = B - \{z\}$. Evidently, $B_1, B_2 \in L_{i-1}$ (B_1 and B_2 are frequent itemsets of size $i - 1$) satisfying $|B_1 - B_2| = 1$ and $|B_2 - B_1| = 1$, and $B = B_1 \cup B_2$. □

Association rules

the resulting algorithm:

```
"input" L(i-1) //all frequent itemsets of size i-1

"output" C(i) //candidates of size i

"algorithm (Apriori-Gen)"
1.  C(i):=0; //empty set
2.  for each B1 from L(i-1) do
3.       for each B2 from L(i-1) different from B1 do
4.            if intersection of B1 and B2 has just i-2
elements then
5.                 add union of B1 and B2 to C(i)
```

Example

Consider the following table $\langle X, Y, I \rangle$ and $s = 0.3$, $c = 0.5$.

I	be	br	je	mi	pb
x_1		X	X		X
x_2		X			X
x_3		X		X	X
x_4	X	X			
x_5	X			X	

Construct L using algorithm Apriori-Gen.

step 1:
$C_1 = \{\{be\}, \{br\}, \{je\}, \{mi\}, \{pb\}\}$
$L_1 = \{\{be\}, \{br\}, \{mi\}, \{pb\}\}$

step 2:
$C_2 = \{\{be, br\}, \{be, mi\}, \{be, pb\}, \{br, mi\}, \{br, pb\}, \{mi, pb\}\}$
$L_2 = \{\{br, pb\}\}$

stop (not itemset of size 3 can be frequent).

Association rules - apriori algorithm

down(B) means $B^{\Downarrow}$

```
"input" <X,Y,I> //data table, s //prescribed support
"output" L //set of all frequent itemsets

"algorithm (Apriori)"
k:=0; //scan (step) number
L:=0; //emptyset
C(0):={{y}|y from Y}
repeat
     k:=k+1;
     L(k):=0;
     for each B from C(k) do
          if |down(B)| >= s x |X| do // B is frequent
               add B to L(k)
    add all B from L(k) to L;
    C(k+1):=Apriori-Gen(L(k))
until C(k+1)=0; //empty set
```

Association rules and maximal and closed itemsets

- frequent itemsets are crucial for mining association rules,
- restricting attention to particular frequent itemsets is usefull,
- two main particular cases (both connected to FCA):
  - maximal frequent itemsets,
  - closed frequent itemsets.
- next: brief overview of maximal and closed frequent itemsets.

Definition (maximal frequent itemset)

A frequent itemset is called a maximal frequent itemset (MFI) if none of its proper supersets is frequent.

- Main advantage: MFIs provide small representation of all frequent itemsets because

  $$A \text{ is a frequent itemset iff } A \subseteq M \text{ for some MFI } M.$$

- Representing all frequent itemsets is useful in various data mining tasks, including mining of association rules.
- Algorithms exist to compute all MFIs from data (directly, without computing all frequent itemsets first), e.g. Gouda K., Zaki M.: GenMax: An Efficient Algorithm for Mining Maximal Frequent Itemsets. Data Mining and Knowledge Discovery, 2005, 1–20. `http://www.cs.rpi.edu/~zaki/PS/DMKD05.pdf`
- Disadvantage of representation using MFIs: MFIs do not contain information about support of their subsets. Consequently, a scan through data is needed to compute a support of an itemset which is not maximal.
- Drawback disappears with closed frequent itemsets, see next.

Definition (closed frequent itemset)

An itemset A is called a closed itemset (CI) if $A = A^{\downarrow\uparrow}$. An itemset A is called a closed frequent itemset (CFI) if it is closed and frequent.

Remark (alternative definition)

In data mining literature, the following definition is often used: An itemset A is called closed if non of its supersets has the same support as A.

Lemma

Both definitions of a closed itemset are equivalent.

Proof.

$A = A^{\downarrow\uparrow}$ means that A is the set of all items which share all objects from $A^{\downarrow}$. This is equivalent to: for any $B \supset A$, $B^{\downarrow}$ has less objects than $A^{\downarrow}$, i.e. $\mathrm{supp}(B) < \mathrm{supp}(A)$. □

Association rules and closed itemsets

Lemma (maximal frequent implies closed frequent)

If A is a maximal frequent itemset then A is a closed frequent itemset.

Proof.

Let A be maximal frequent. Suppose A is not closed. Then $A \subset A^{\downarrow\uparrow}$ but since $\mathrm{supp}(A) = \mathrm{supp}(A^{\downarrow\uparrow})$, $A^{\downarrow\uparrow}$ is frequent too. Hence, A is not maximal frequent, a contradiction with the assumption. □

- Algorithms exist to compute closed frequent itemsets from data (see references later).
- Support of non-closed frequent itemsets can be determined from supports of closed frequent itemsets.

Association rules and closed itemsets

Zaki's result:

- closed frequent itemsets can be used for mining non-redundant association rules,
- Zaki M.: Mining non-redundant association rules. Data Mining and Knowledge Discovery, 2004.
  http://www.cs.rpi.edu/~zaki/PS/DMKD04.pdf
- quote from Zaki's paper: "The traditional association rule mining framework produces many redundant rules. The extent of redundancy is a lot larger than previously suspected. We present a new framework for associations based on the concept of closed frequent itemsets. The number of non-redundant rules produced by the new approach is exponentially (in the length of the longest frequent itemset) smaller than the rule set from the traditional approach. Experiments using several hard as well as easy real and synthetic databases confirm the utility of our framework in terms of reduction in the number of rules presented to the user, and in terms of time."

Association rules and closed itemsets

- An association rule $A \Rightarrow B$ is called redundant if there is another association rule $A_1 \Rightarrow B_1$ with the same support and confidence as $A \Rightarrow B$ such that $A_1 \subseteq A$ and $B_1 \subseteq B$.
- Example:

  {watches-TV, unhealthy-food, does-not-move} ⇒ {high-blood-pressure}

  is redundant if there is rule

  { unhealthy-food, does-not-move } ⇒ { high-blood-pressure}

  with the same support and confidence.
- Main result: if we use closed frequent itemsets instead of all frequent itemsets, redundant rules are not generated. That is, the output set of association rules contains only non-redundant rules. Moreover, all interesting association rules can be generated from non-redundant ones.

Association rules and closed itemsets

Some references on closed and maximal itemsets.

- Pasquier et al.: Discovering frequent closed itemsets for association rules. ICDT 1999, 398–416.
- Zaki M.: Mining non-redundant association rules. Data Mining and Knowledge Discovery, 2004. http://www.cs.rpi.edu/~zaki/PS/DMKD04.pdf
- M. Zaki's web page

Association rules and closed itemsets

Useful sources on association rules.

- referential data (testing):
  Machine Learning Repository
  http://www.ics.uci.edu/~mlearn/MLRepository.html,
  UCI KDD Archive
  http://kdd.ics.uci.edu,
- software:
  overview at
  http://www.kdnuggets.com/software/associations.html,
  free (GNU General Public License): ARTool,
  http://www.cs.umb.edu/~laur/ARtool/.