Attribute Implications and Other Data Dependencies

Radim Belohlavek

DAMOL
DATA ANALYSIS AND MODELING LAB
Palacky University, Olomouc, Czech Republic

european social fund in the czech republic
EUROPEAN UNION
MINISTRY OF EDUCATION, YOUTH AND SPORTS
OP Education for Competitiveness
INVESTMENTS IN EDUCATION DEVELOPMENT

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	×		×	

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

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

$\times$... **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)

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, 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 $M_3 = \emptyset$.

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

- 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

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

- 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	$\emptyset \subseteq M$

- 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,

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

Note that a formal context $\langle X, Y, I\rangle$ is a triplet where X and Y are non-empty sets (of objects and attributes) and $I \subseteq X \times Y$ is a binary relation interpreted as: $\langle x, y\rangle \in I$ if and only if object x has attribute y.

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

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

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

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$\} \Rightarrow \{$nbp$\}$	1	
$\{$TV,uf$\} \Rightarrow \{$hbp$\}$	1	
$\{$TV$\} \Rightarrow \{$hbp$\}$	1	
$\{$uf$\} \Rightarrow \{$hbp$\}$	0	b counterexample
$\{$nbp$\} \Rightarrow \{$r$\}$	0	e counterexample
$\{$nbp,hbp$\} \Rightarrow \{$r,TV$\}$	1	A never satisfied
$\{$uf,r$\} \Rightarrow \{$r$\}$	1	

AIs – theory, models, semantic consequence

- Previous example: {TV,uf} ⇒ {hbp} intuitively follows from {TV} ⇒ {hbp}. Therefore, provided we establish validity of {TV} ⇒ {hbp}, AI {TV,uf} ⇒ {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.

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

- 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 $||A \Rightarrow B||_M = 1$.
- $T_6 = \{\emptyset \Rightarrow \{y_1\}, \emptyset \Rightarrow \{y_3\}\}$. $\mathrm{Mod}(T_6) = \{\{y_1, y_3\}, \{y_1, y_2, y_3\}\}$.
- $T_7 = \{\{y_1\} \Rightarrow \emptyset, \{y_2\} \Rightarrow \emptyset, \{y_3\} \Rightarrow \emptyset\}$. $\mathrm{Mod}(T_7) = 2^{\{y_1, y_2, y_3\}}$.
- $T_8 = \{\{y_1\} \Rightarrow \{y_2\}, \{y_2\} \Rightarrow \{y_3\}, \{y_3\} \Rightarrow \{y_1\}\}$. $\mathrm{Mod}(T_8) = \{\emptyset, \{y_1, y_2, y_3\}\}$.

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").
- Terminology: $T \models A \Rightarrow B \dots A \Rightarrow B$ follows semantically from $T \dots A \Rightarrow B$ is semantically entailed by $T \dots A \Rightarrow B$ 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 $T \not\models A \Rightarrow B$.

Later on, we will see how to efficiently check whether $T \models A \Rightarrow B$.

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

$\mathrm{Mod}(T)$ as a closure system

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. Why?

Definition (closure operator)

A closure operator in a set Y is any mapping $C : 2^Y \to 2^Y$ satisfying:
(i) $A \subseteq C(A)$;
(ii) $A \subseteq B$ implies $C(A) \subseteq C(B)$;
(iii) $C(A) = C(C(A))$,
for every $A, B \in 2^Y$.

There is a one-to-one relationship between closure systems and closure operators in Y.

(1) 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.

(2) 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 $\mathcal{S} = \mathcal{S}_{C_{\mathcal{S}}}$.

Observe: If $\mathcal{S}$ is a closure system and C the corresponding closure operator in Y then for every $A \subseteq Y$, $C(A)$ is the least element from $\mathcal{S}$ of which is a subset.

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

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. ☐

Since $\mathrm{Mod}(T)$ is a closure system, we can consider the corresponding 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 $C_{\mathrm{Mod}(T)}(A) \subseteq M$ (because $C_{\mathrm{Mod}(T)}(A)$ is the least model of T containing A). Putting together with $B \subseteq C_{\mathrm{Mod}(T)}(A)$, we get $B \subseteq M$, i.e. $A \Rightarrow B$ is true in M. □

Checking entailment via least model

- Previous theorem implies that we may test $T \models A \Rightarrow B$ by checking whether $A \Rightarrow B$ is true in a single particular model of T, namely in $\mathrm{Mod}(T)(A)$. 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

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

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 $Z^{T_2} \subseteq M$, $Z^{T_3} \subseteq M$, $\ldots$. 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. ☐

- 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.

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 $T \models A \Rightarrow B$.

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:

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

- $A^{T_0} = A = \{y_2, y_5, y_6\}$.
- $A^{T_1} = A^{T_0} \cup \bigcup \{\{y_4\}\} = \{y_2, y_4, y_5, y_6\}$.
  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 $T \not\models A \Rightarrow B$.

Further topics (future seminars)

- Bases of implications (how to obtain the least fully informative set of implications from data).
- Logic of implications (how to obtain a syntactico-semantically complete set of implications).
- Association rules and GUHA rules (how to obtain rules that allow exceptions, e.g. rules true in 90% cases).
- What can fuzzy logic add to such rules?

HOMEWORK/EASY (send solutions to radim.belohlavek@acm.org)

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

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

HOMEWORK/MODERATE (send solutions to radim.belohlavek@acm.org)

1. Prove the following claim: Given a closure operator C in Y, $\mathcal{S}_C = \{A \in 2^X \mid A = C(A)\}$ is a closure system in Y.
2. Prove the following claim: 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.
3. Prove the following claim: The mappings defined in 1. and 2. are mutually inverse bijections between the set of all closure operators in Y and the set of all closure systems in Y.
4. Prove or disprove (by showing counterexamples) the following claims:
   - 4a. If $T_1 \subseteq T_2$ then $\mathrm{Mod}(T_1) \supseteq \mathrm{Mod}(T_2)$.
   - 4b. If $\mathrm{Mod}(T_1) \supseteq \mathrm{Mod}(T_2)$ then $T_1 \subseteq T_2$.
5. Let for a set $\mathcal{M} \subseteq 2^Y$ of sets of attributes denote by $\mathrm{Fml}(\mathcal{M})$ the set of all attribute implications which are in in every $M \in \mathrm{M}$. Prove that the pair $\langle \mathrm{Mod}, \mathrm{Fml} \rangle$ forms a Galois connection between the of all attribute implications over Y and the set of all subsets of Y.

HOMEWORK/HARD (send solutions to radim.belohlavek@acm.org)

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 $\mathcal{B}(X, Y, I)$.
2. Consider the problem of computing $C_{\mathrm{Mod}(T)}(A)$ (output) from a set T of implications over Y and a subset A of Y (input). Consider the direct implmentation of the algorithm for computing $C_{\mathrm{Mod}(T)}$ given by the definition: compute A^{T_1}, A^{T_2}, A^{T_n} until $A^{T_n} \neq A^{T_{n+1}}$ and output A^{T_n}, which is $C_{\mathrm{Mod}(T)}(A)$. Prove that the worst case time complexity $t(n)$ of this algorithm is $\Theta(n^2)$, provided the input size is considered to be the number $n = |T|$ of implications in T. That is, show that $t(n) = O(n^2)$ and that $t(n) = \Omega(n)$.
   Find an algorithm of worst case time complexity $\Theta(n)$.