Attribute Dependencies for Data with Grades II

Radim Belohlavek, Vilem Vychodil (Palacký University, Olomouc)

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

Problem Setting

input: data table $\langle X, Y, I \rangle$

output: collection T of graded attribute implications such that:

1. T is reasonably small (simple, ...), and
2. T conveys information about exactly all graded implications true in data

questions:

- what is a *collection?* (a set / $\mathbf{L}$-set of formulas)
- what is *reasonably small?* (in terms of size, redundancy, ...)
- what is meant by *true in data?* (fully true, true to a degree, ...)

in case you missed it: goto DAMOL webpage

Belohlavek, Vychodil: *Attribute Dependencies for Data with Grades I*

Completeness in $\langle X, Y, I \rangle$

Definition (completeness in data)

A set T of graded attribute implications is called **complete in $\langle \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{I} \rangle$** if

$$||A \Rightarrow B||_T = ||A \Rightarrow B||_{\langle X,Y,I \rangle}$$

for every implication $A \Rightarrow B$.

verbally: T being complete in $\langle X, Y, I \rangle$ means:

> *"each graded attribute implication $A \Rightarrow B$ semantically follows from T to the degree to which it is true in $\langle X, Y, I \rangle$"*

notes:

- T is an (ordinary) set
- completeness characterizes *all degrees* (strong property)

Theorem

T is complete in $\langle X, Y, I\rangle$ iff $\mathrm{Mod}(T) = \mathrm{Int}(X^*, Y, I)$.

Proof (sketch).

Let T be complete. If $M \in \mathrm{Mod}(T)$, then for $M \Rightarrow M^{\downarrow\uparrow}$,

$$||M \Rightarrow M^{\downarrow\uparrow}||_M \geq ||M \Rightarrow M^{\downarrow\uparrow}||_T = ||M \Rightarrow M^{\downarrow\uparrow}||_{\langle X,Y,I\rangle} = S(M^{\downarrow\uparrow}, M^{\downarrow\uparrow}) = 1,$$

i.e., $M \in \mathrm{Int}(X^*, Y, I)$ because $M^{\downarrow\uparrow} \subseteq M$.

Conversely, for $M \in \mathrm{Int}(X^*, Y, I)$ and $A \Rightarrow B \in T$:

$$||A \Rightarrow B||_M \geq ||A \Rightarrow B||_{\langle X,Y,I\rangle} = ||A \Rightarrow B||_T = 1,$$

i.e., $M \in \mathrm{Mod}(T)$.

The "$\Leftarrow$"-part is easy to see:

$$\begin{aligned} ||A \Rightarrow B||_T &= \bigwedge\{||A \Rightarrow B||_M \mid M \in \mathrm{Mod}(T)\} \\ &= \bigwedge\{||A \Rightarrow B||_M \mid M \in \mathrm{Int}(X^*, Y, I)\} = ||A \Rightarrow B||_{\langle X,Y,I\rangle}. \end{aligned}$$

□

Example (Complete Sets)

Let $\langle X, Y, I\rangle$ be any data table with graded attributes.

Then $T = \{A \Rightarrow A^{\downarrow\uparrow} \mid A \in L^Y\}$ is complete in $\langle X, Y, I\rangle$.

Apply the previous Theorem:

$\mathrm{Mod}(T) \subseteq \mathrm{Int}(X^*, Y, I)$-part:

for $M \Rightarrow M^{\downarrow\uparrow} \in T$, we get $1 = S(M, M)^* \leq S(M^{\downarrow\uparrow}, M)$, i.e., $M^{\downarrow\uparrow} \subseteq M$.

$\mathrm{Mod}(T) \supseteq \mathrm{Int}(X^*, Y, I)$-part:

if $M = M^{\downarrow\uparrow}$, then $S(A, M)^* \leq S(A^{\downarrow\uparrow}, M^{\downarrow\uparrow}) = S(A^{\downarrow\uparrow}, M)$, i.e., $||A \Rightarrow A^{\downarrow\uparrow}||_M = 1$.

Furthermore, $T = \{A \Rightarrow B \mid ||A \Rightarrow B||_{\langle X,Y,I\rangle} = 1\}$ is complete in $\langle X, Y, I\rangle$

both "$\subseteq$" and "$\supseteq$" are easy to see

1-Completeness in $\langle X, Y, I \rangle$

Definition (1-completeness in data)

A set T of graded attribute implications is called a **1-complete in $\langle X, Y, I \rangle$** if

$$||A \Rightarrow B||_T = 1 \text{ iff } ||A \Rightarrow B||_{\langle X,Y,I \rangle} = 1$$

for every implication $A \Rightarrow B$.

interesting characterization of completeness:

"it suffices to characterize all $A \Rightarrow B$ true to degree 1":

Theorem

T is complete in $\langle X, Y, I \rangle$ iff T is 1-complete in $\langle X, Y, I \rangle$.

Proof (sketch).

"$\Rightarrow$" is trivial, "$\Leftarrow$": is proven by showing that $\mathrm{Mod}(T) = \mathrm{Int}(X^*, Y, I)$. □

Redundant Theories

question:

"What is a reasonably small T?"

Definition (redundant sets of graded implications)

A set T of graded implications is **redundant** if there exists $A \Rightarrow B \in T$ such that $||A \Rightarrow B||_{T-\{A \Rightarrow B\}} = 1$. Otherwise, call T is **non-redundant**.

Lemma

The following conditions are equivalent:

1. *T is a non-redundant set of implications;*
2. *For every $A \Rightarrow B \in T$: $\mathrm{Mod}(T) \subset \mathrm{Mod}(T - \{A \Rightarrow B\})$;*
3. *For every $A \Rightarrow B \in T$ there are C, D s. t. $||C \Rightarrow D||_{T-\{A \Rightarrow B\}} < ||C \Rightarrow D||_T$.* □

Non-redundant Bases of $\langle X, Y, I \rangle$

Definition (non-redundant bases)

A set T of graded implications is called a **non-redundant base of $\langle X, Y, I \rangle$** if T is complete in $\langle X, Y, I \rangle$ and no proper subset of T is complete in $\langle X, Y, I \rangle$.

based on characterization of redundant sets:

Corollary

T is a non-redundant base of $\langle X, Y, I \rangle$ if and only if

- *T is complete in $\langle X, Y, I \rangle$, and*
- *T is non-redundant as a set of implications.* □

Systems of Pseudo-Intents

naive approach to getting non-redundant bases:

- start with a complete set,
- remove all redundant implications (based on checking entailment degrees)

alternatively describe (certain) non-redundant bases directly:

Definition (system of pseudo-intents)

$\mathcal{P} \subseteq L^Y$ is a **system of pseudo-intents of $\langle X, Y, I\rangle$** if for each $P \in L^Y$:

$$P \in \mathcal{P} \quad \text{iff} \quad P \neq P^{\downarrow\uparrow} \text{ and } ||Q \Rightarrow Q^{\downarrow\uparrow}||_P = 1 \text{ for each } Q \in \mathcal{P} \text{ with } Q \neq P.$$

equivalent formulations: $\mathcal{P} \subseteq L^Y - \mathrm{Int}(X^*, Y, I)$ such that

$$P \in \mathcal{P} \quad \text{iff} \quad P \in \mathrm{Mod}(\{Q \Rightarrow Q^{\downarrow\uparrow} \,|\, Q \in \mathcal{P} \text{ and } Q \neq P\})$$

$$\text{iff} \quad P \in \mathrm{Mod}(T - \{P \Rightarrow P^{\downarrow\uparrow}\}), \text{ where } T = \{Q \Rightarrow Q^{\downarrow\uparrow} \,|\, Q \in \mathcal{P}\}$$

Example (Systems of Pseudo-Intents)

multiple systems of pseudo-intents:

Structure of degrees and data:

- $L = \{0, 0.5, 1\}$, linearly ordered, Gödel $\otimes$ (i.e., $\otimes = \wedge$); * is identity;
- $X = \{x\}$, $Y = \{y, z\}$, $I(x, y) = I(x, z) = 0$

Systems of pseudo-intents:

$$\mathcal{P}_1 = \{\{z\}, \{{}^{0.5}/y, {}^{0.5}/z\}, \{y\}\}, \qquad \mathcal{P}_3 = \{\{y\}, \{{}^{0.5}/z\}\},$$
$$\mathcal{P}_2 = \{\{z\}, \{{}^{0.5}/y\}\}, \qquad \mathcal{P}_4 = \{\{{}^{0.5}/y\}, \{{}^{0.5}/z\}\}.$$

no system of pseudo-intents:

Structure of degrees:

- L on $[0, 1]$ given by left-continuous t-norm; * is globalization;
- $X = \{x\}$, $Y = \{y\}$, $I(x, y) = 0$

Theorem

If $\mathcal{P}$ is a system of pseudo-intents of $\langle X, Y, I\rangle$ then

$$T = \{P \Rightarrow P^{\downarrow\uparrow} \mid P \in \mathcal{P}\}$$

is a non-redundant base of $\langle X, Y, I\rangle$.

Proof (sketch).

T is complete: each intent is obviously a model of T (previous example). Conversely, if $M \in \mathrm{Mod}(T)$, then $||P \Rightarrow P^{\downarrow\uparrow}||_M = 1$ for all $P \in \mathcal{P}$. Now, if $M \in \mathcal{P}$, we get $M \notin \mathrm{Mod}(T)$ since $||M \Rightarrow M^{\downarrow\uparrow}||_M < 1$. Therefore, $M \notin \mathcal{P}$, i.e., $M = M^{\downarrow\uparrow}$ by definition of $\mathcal{P}$.

T is non-redundant: take $M \in \mathcal{P}$; directly from $||P \Rightarrow P^{\downarrow\uparrow}||_M = 1$ for all $P \in \mathcal{P}$ such that $P \neq M$, we get $\mathrm{Mod}(T) \subset \mathrm{Mod}(T - \{M \Rightarrow M^{\downarrow\uparrow}\})$. □

Graphs Induced by $\langle X, Y, I\rangle$

Definition (graph induced by data)

For $\langle X, Y, I\rangle$, define

$$V = \{P \in L^Y \mid P \neq P^{\downarrow\uparrow}\}.$$

If $V \neq \emptyset$, define

$$E = \{\langle P, Q\rangle \in V \times V \mid P \neq Q \text{ and } ||Q \Rightarrow Q^{\downarrow\uparrow}||_P \neq 1\}$$

and the corresponding **graph** $\mathbf{G} = \langle \boldsymbol{V}, \boldsymbol{E} \cup \boldsymbol{E}^{-1}\rangle$.

For any $Q \in V$ and $\mathcal{P} \subseteq V$ define the following subsets of V:

$$\mathrm{Pred}(Q) = \{P \in V \mid \langle P, Q\rangle \in E\},$$
$$\mathrm{Pred}(\mathcal{P}) = \bigcup\{\mathrm{Pred}(Q) \mid Q \in \mathcal{P}\}.$$

trivial case: if $V = \emptyset$ then $\mathcal{P} = \emptyset$ (and $T = \emptyset$ is the non-redundant base)

Graph-Theoretic Interpretation of Pseudo-Intents

Theorem

Let $\mathcal{P} \subseteq V$. $\mathcal{P}$ is a system of pseudo-intents iff $V - \mathcal{P} = \mathrm{Pred}(\mathcal{P})$.

Lemma

Let $\emptyset \neq \mathcal{P} \subseteq V$. If $V - \mathcal{P} = \mathrm{Pred}(\mathcal{P})$ then $\mathcal{P}$ is a maximal independent set in $\mathbf{G}$.

Corollary

The following are equivalent:

1. *$\mathcal{P} \neq \emptyset$ is a system of pseudo-intents;*
2. *$\mathcal{P}$ is a maximal independent set in $\mathbf{G}$ such that $V - \mathcal{P} = \mathrm{Pred}(\mathcal{P})$.* □

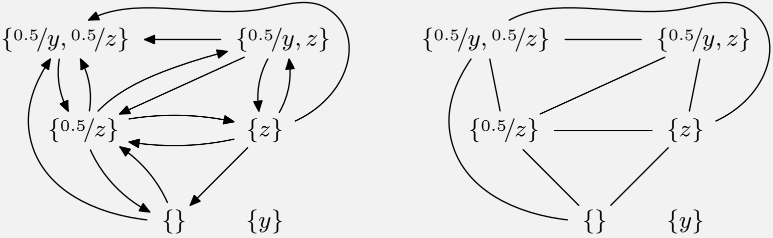
getting all systems of pseudo-intents:

- list all maximal independent sets $\mathcal{P}$ in $\mathbf{G}$ satisfying $V - \mathcal{P} = \mathrm{Pred}(\mathcal{P})$

Example (Graph-Theoretic Interpretation of Pseudo-Intents)

Structure of degrees and data:

- $L = \{0, 0.5, 1\}$, linearly ordered, Łukasiewicz $\otimes$; * is identity;
- $X = \{x\}$, $Y = \{y, z\}$, $I(x, y) = 0.5$, $I(x, z) = 1$.



Maximal independent sets and **systems of pseudo-intents**:

$\mathcal{P}_1 = \{\{\}, \{{}^{0.5}/y, z\}, \{y\}\}$,

$\mathcal{P}_2 = \{\{{}^{0.5}/z\}, \{y\}\}$,

$\mathcal{P}_3 = \{\{z\}, \{y\}\}$,

$\mathcal{P}_4 = \{\{{}^{0.5}/y, {}^{0.5}/z\}, \{y\}\}$.

Notes on Pseudo-Intents and Globalization

systems of pseudo-intents:

$$P \in \mathcal{P} \quad \text{iff} \quad P \neq P^{\downarrow\uparrow} \text{ and } ||Q \Rightarrow Q^{\downarrow\uparrow}||_P = 1 \text{ for each } Q \in \mathcal{P} \text{ with } Q \neq P$$

in case of globalization:

$$\begin{aligned} ||Q \Rightarrow Q^{\downarrow\uparrow}||_P = 1 \quad & \text{iff} \quad \text{either } Q \nsubseteq P \text{ or } Q \subseteq P \text{ and } Q^{\downarrow\uparrow} \subseteq P \\ & \text{iff} \quad Q \subseteq P \text{ implies } Q^{\downarrow\uparrow} \subseteq P \end{aligned}$$

definition of $\mathcal{P}$ simplifies to:

$$P \in \mathcal{P} \quad \text{iff} \quad P \neq P^{\downarrow\uparrow} \text{ and } Q^{\downarrow\uparrow} \subseteq P \text{ for all } Q \subset P \text{ such that } Q \in \mathcal{P}$$

case of finite $\boldsymbol{Y}$ and $\boldsymbol{L}$:

$$P \in \mathcal{P} \quad \text{iff} \quad P \neq P^{\downarrow\uparrow} \text{ and } Q^{\downarrow\uparrow} \subseteq P \text{ for all } Q \subset P \text{ such that } Q \in \mathcal{P}$$

consequence: $\mathcal{P}$ exists and is unique

Minimal Bases (Case of Globalization)

Definition (minimal base of $\langle X, Y, I\rangle$)

A set T of graded implications is called a **minimal base of $\langle \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{I}\rangle$** if T is complete in $\langle X, Y, I\rangle$ and $|T| \leq |T'|$ for each set T' which is complete in $\langle X, Y, I\rangle$.

case of globalization:

Theorem

Let $\mathbf{L}$ be a finite residuated lattice with globalization.
If Y is finite, then $T = \{P \Rightarrow P^{\downarrow\uparrow} \mid P \in \mathcal{P}\}$ is minimal.

Proof (idea).

Using the fact that for each $P \in \mathcal{P}$, there is $A \Rightarrow B \in T'$ such that $A^{\downarrow\uparrow} = P^{\downarrow\uparrow}$, considering $f : \mathcal{P} \to T'$ such that $f(P) = A \Rightarrow B$ and proving it is injective. □

Algorithm (Graph-Based Method for Globalization)

Data: $E \subseteq V \times V$, list $\mathcal{S}$ of lectically ordered element from V
Result: system of pseudo-intents $\mathcal{P}$ (for L and Y finite)

set $\mathcal{P}$ **to** $\emptyset$
while $\mathcal{S} \neq \emptyset$ **do**
    $\mathcal{P} := \mathcal{P} \cup \mathrm{first}(\mathcal{S})$
    vacate $\mathcal{S}'$
    for B **in** $\mathrm{rest}(\mathcal{S})$ **do**
        if $\langle B, \mathrm{first}(\mathcal{S}) \rangle \notin E$ **then**
            append B **to** $\mathcal{S}'$
        end
    end
    $\mathcal{S} := \mathcal{S}'$
end
return $\mathcal{P}$

Further Properties (for Globalization)

closure operator: define cl_T by iteration of f (Tarski, Cousot & Cousot):

$$f(M) = M \cup \bigcup\{B \otimes S(A, M)^* \mid A \Rightarrow B \in T \text{ and } A \neq M\}$$
$$= M \cup \bigcup\{B \mid A \Rightarrow B \in T \text{ and } A \subset M\}$$

Theorem

For finite Y and L with globalization, $T = \{P \Rightarrow P^{\downarrow\uparrow} \mid P \in \mathcal{P}\}$, the set of all fixed points of cl_T if $\mathcal{P} \cup \mathrm{Int}(X^*, Y, I)$. □

computational issues:

- cl_T uses (subsets of) T to compute (furter formulas of) T
- fixpoints shall be enumerated in an order extending "$\subset$"

Algorithm (Enumerating Fixed Points of cl_T for Globalization)

Data: data table $\langle X, Y, I\rangle$
Result: system of pseudo-intents $\mathcal{P}$ (for L and Y finite)

$B := \emptyset$
$T := \emptyset$
while $B \neq Y$ **do**
  if $B \neq B^{\downarrow\uparrow}$ **then**
    $T := T \cup \{B \Rightarrow B^{\downarrow\uparrow}\}$
  end
  $B := \mathrm{NextClosure}(B, cl_T)$
end
return T

optional extension:

computes $\mathcal{P}$ *and* $\mathrm{Int}(X^*, Y, I)$ *(additional else-branch)*

Getting Non-redundant Bases by Reduction

alternative approach: for general * (Y and L finite), introduce $\mathcal{P}$ by putting

$$P \in \mathcal{P} \quad \text{iff} \quad P \neq P^{\downarrow\uparrow} \text{ and } Q^{\downarrow\uparrow} \subseteq P \text{ for all } Q \subset P \text{ such that } Q \in \mathcal{P}$$

consequence:

Theorem

Under the above notation, $T = \{P \Rightarrow P^{\downarrow\uparrow} \mid P \in \mathcal{P}\}$ *is complete.* □

two-step procedure:

- compute T (as in case of globalization but with general *)
- remove redundant implications from T

Summary

results on graded attribute implications:

- foundations,
- semantic entailment,
- complete ordinary-style and graded-style axiomatizations,
- reductions to ordinary attribute imlications,
- relationship to other types of dependencies,
- non-redundant bases,
- algorithms
- more results available, ...