Attribute dependencies in incomplete binary data

Michal Krupka, **Jan Laštovička**, and Petr Osička

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

Overview

Input

- incomplete binary data:

	male	female	sport	culture
Adam	×		u_1	×
Bea		×		×
Chris	u_2	u_3	×	

- dependencies between variables: $u_2 = u_3'$
- attribute dependencies:

  sport and culture are less important then male and female

Output

- incomplete concepts compatible with attribute dependencies

Formal concept analysis

– formal context $\langle X, Y, I \rangle$, $I \subseteq X \times Y$:

	male	female	sport	culture
Adam	×			×
Bea		×		×
Chris		×	×	

– pair of induced mappings:

$$A^{\uparrow} = \{y \in Y \mid \langle x, y \rangle \in I \text{ for all } x \in A\}$$
$$B^{\downarrow} = \{x \in X \mid \langle x, y \rangle \in I \text{ for all } y \in B\}$$

Definition (concept lattice)

$\mathscr{B}(X,Y,I) = \{\langle A, B \rangle \in 2^X \times 2^Y \mid A^{\uparrow} = B \text{ and } B^{\downarrow} = A\}$

partial order: $\langle A_1, B_1 \rangle \leq \langle A_2, B_2 \rangle$ iff $A_1 \subseteq A_2$ (iff $B_2 \subseteq B_1$)

$\mathscr{B}(X,Y,I)$ is a complete lattice.

Atrribute dependencies - Belohlavek and Vychodil (2009)

Let Y be set (attributes).

AD-formula: $C \sqsubseteq D$ where $C, D \subseteq Y$

Definition (validity)

Let $E \subseteq Y$.

$||C \sqsubseteq D||_E = 1$ iff $C \cap E \neq \emptyset$ implies $D \cap E \neq \emptyset$

Example

$||\{\text{sport}, \text{culture}\} \sqsubseteq \{\text{male}, \text{female}\}||_{\{\text{sport}\}} = 0$

Let

- $\langle X, Y, I \rangle$ be a formal context
- T be a set of AD-formulas

Definition (compatible formal concepts)

$\mathscr{B}_T(X,Y,I) = \{\langle A, B\rangle \in \mathscr{B}(X,Y,I) \mid ||C \sqsubseteq D||_B = 1$ for each $C \sqsubseteq D \in T\}$.

Incomplete contexts - Krupka and Lastovicka (2012)

Let $\mathbf{L}$ be a Boolean algebra with variables $U = \{u_1, \ldots, u_n\}$ ($U \subseteq L$, $\mathbf{L}$ is generated by U).

Definition (Incomplete $\mathbf{L}$-context)

$\langle X, Y, I \rangle$ where $I \in L^{X \times Y}$ such that $I(X \times Y) \subseteq U \cup \{0, 1\}$

	male	female	sport	culture
Adam	$\times$		u_1	$\times$
Bea		$\times$		$\times$
Chris	u_2	u_3	$\times$	

Definition

- $v : U \to \mathbf{2}$ ... assignment
- If v can be extended to a homomorphism $\bar{v} : \mathbf{L} \to \mathbf{2}$, it is called admissible.

Example

If $u_2 = u_3'$, then an assignment v with $v(u_2) = v(u_3) = 1$ is not admissible.

Concept lattices of incomplete contexts

Let $\langle X, Y, I\rangle$ be an incomplete $\mathbf{L}$-context.
Formal concept analysis in fuzzy setting:

$$A^{\uparrow}(y) = \bigwedge_{x \in X} A(x) \to I(x,y)$$
$$B^{\downarrow}(x) = \bigwedge_{y \in Y} B(y) \to I(x,y)$$

Incomplete formal $\mathbf{L}$-concept: $\langle A, B\rangle \in L^X \times L^Y$ with $A^{\uparrow} = B$ and $B^{\downarrow} = A$

Definition (total incomplete $\mathbf{L}$-concept lattice)

- $\mathscr{B}(X,Y,I)\ldots$ the set of all incomplete formal $\mathbf{L}$-concepts
- partial ordering: $\langle A_1, B_1\rangle \leq \langle A_2, B_2\rangle$ iff $A_1 \subseteq A_2$ (iff $B_2 \subseteq B_1$)

$\mathscr{B}(X,Y,I)$ is a complete lattice.

Validity of AD-formulas in incomplete data

Let

- $\mathbf{L}$ be a Boolean algebra with variables
- Y be a set (attributes)
- $B \in L^Y$

Definition (Validity of AD-formula)

$$||C \sqsubseteq D||_B = \Big(\bigvee_{y \in C} B(y)\Big) \to \Big(\bigvee_{y \in D} B(y)\Big)$$

$\mathbf{L}$-theory is an $\mathbf{L}$-set of AD-formulas.
Let T be an $\mathbf{L}$-theory.

Definition (Validity of $\mathbf{L}$-theory)

$$||T||_B = \bigwedge_{C \sqsubseteq D \in \mathrm{ADF(Y)}} T(C \sqsubseteq D) \to ||C \sqsubseteq D||_B$$

Compatible incomplete formal $\mathbf{L}$-concepts

Let

- $\langle X, Y, I \rangle$ be an incomplete $\mathbf{L}$-context
- T be an $\mathbf{L}$-theory and $F \subseteq L$

Definition (F-compatible concepts)

$\mathscr{B}_{F,T}(X,Y,I) = \{\langle A, B\rangle \in \mathscr{B}(X,Y,I) \mid ||T||_B \in F\}$

Let v be an admissible assignment.

$\langle X, Y, \bar{v} \circ I \rangle \ldots$ v-completion

Induced homomorphism $\bar{v}^{\mathscr{B}(X,Y,I)} : \mathscr{B}(X,Y,I) \to \mathscr{B}(X,Y,\bar{v} \circ I)$

$$\langle A, B \rangle \mapsto \langle \bar{v} \circ A, \bar{v} \circ B \rangle$$

Theorem

$$\bar{v}^{\mathscr{B}(X,Y,I)}(\mathscr{B}_{\bar{v}^{-1}(1),T}(X,Y,I)) = \mathscr{B}_{\bar{v} \circ T}(X,Y,\bar{v} \circ I)$$

Structure of $\mathscr{B}_{F,T}(X,Y,I)$?

Structure corresponding to context and theory

Let $\langle X,Y,I\rangle$ be an $\mathbf{L}$-context and T be an $\mathbf{L}$-theory.

An instance of first-order fuzzy logic:

Definition

The structure corresponding to $\langle X,Y,I\rangle$ and T is $\mathbf{L}$-structure $\mathbf{M}(\langle X,Y,I\rangle,T)$ with

- $M = \mathscr{B}(X,Y,I)$
- $r_T^{\mathbf{M}}(\langle A,B\rangle) = ||T||_B$
- $\leq^{\mathbf{M}} = \preceq \qquad (\langle A_1,B_1\rangle \preceq \langle A_2,B_2\rangle) = S(A_1,A_2) \quad (= S(B_2,B_1))$
- $c_0^{\mathbf{M}} = \langle Y^{\downarrow},Y\rangle$, $c_1^{\mathbf{M}} = \langle X,X^{\uparrow}\rangle$
- $\inf^{\mathbf{M}} = \wedge$, $\sup^{\mathbf{M}} = \vee$

Shorthands:

$x||y$ is $\neg(x \leq y) \Cap \neg(y \leq x) \ldots$ incomparability

$x \approx y$ is $(x \leq y) \Cap (y \leq x) \ldots$ equivalence

Transferring theorem

Let

- φ be a formula
- $\langle X, Y, I\rangle$ be an incomplete $\mathbf{L}$-context
- T be an $\mathbf{L}$-theory
- $\mathbf{M}_1 = \mathbf{M}(\langle X, Y, I\rangle, T)$
- $\mathbf{M}_v = \mathbf{M}(\langle X, Y, \overline{v} \circ I\rangle, \overline{v} \circ T)$

Theorem

For any $\mathbf{M}_1$-valuation e it holds

$$||\varphi||^{\mathbf{L}}_{\mathbf{M}_1,e} = 1 \text{ iff } ||\varphi||^{\mathbf{2}}_{\mathbf{M}_v,e_v} = 1 \text{ for each admissible assignment } v,$$

where $e_v = \bar{v}^{\mathscr{B}(X,Y,I)} \circ e$.

Structure of $\mathscr{B}_{F,T}(X,Y,I)$

Corollary

Let T be an $\mathbf{L}$-theory. Then, $\mathscr{B}_{F,T}(X,Y,I)$ with $1 \in F$ has the least element.

Proof.

$\varphi = r_T(c_0)$ □

Corollary

Let T be an $\mathbf{L}$-theory consisting of AD-formulas of the form $\{a\} \sqsubseteq \{b\}$. Then, $\mathscr{B}_{\{1\},T}(X,Y,I)$ is a complete lattice which is a $\bigvee$-sublattice of $\mathscr{B}(X,Y,I)$.

Proof.

$\varphi = (\forall x)(\forall y)((r_T(x) \wedge r_T(y)) \Rightarrow r_T(sup(x,y)))$ □

When $\mathscr{B}_{F,T}(X,Y,I)$ is a tree?

	male	female	sport	culture
Adam	$\times$		u_1	$\times$
Bea		$\times$		$\times$
Chris	u_2	u_3	$\times$	

$\{y_1\}^{\downarrow_I} \cap \{y_2\}^{\downarrow_I} = \emptyset \ldots$ disjoint

tree theory: if y_1, y_2 are not disjoint, then $\{y_1\} \sqsubseteq Y_2 \in T$ or $\{y_2\} \sqsubseteq Y_1 \in T$ where $y_1 \in Y_1$ and $y_2 \in Y_2$ and attributes in both Y_1 and Y_2 are pairwise disjoint.

Example

$T = \{\{\text{sport}\} \sqsubseteq \{\text{male}, \text{female}\}, \{\text{culture}\} \sqsubseteq \{\text{male}, \text{female}\}, \{\text{sport}\} \sqsubseteq \{\text{culture}\}\}$

Let T be a tree theory.

Corollary

For each $\langle A_1, B_1\rangle, \langle A_2, B_2\rangle \in \mathscr{B}_{\{1\},T}(X,Y,I)$ *it holds if* $\langle A_1, B_1\rangle \preceq \langle A_2, B_2\rangle = 0$ *and* $\langle A_2, B_2\rangle \preceq \langle A_1, B_1\rangle = 0$, *then* $\langle A_1, B_1\rangle \wedge_{\mathscr{B}_{\{1\},T}(X,Y,I)} \langle A_2, B_2\rangle = \langle Y^{\downarrow}, Y\rangle$.

$\varphi = (\forall x)(\forall y)((r_T(x) \curlywedge r_T(y) \curlywedge (x||y)) \Rightarrow (\forall z)(z \leq \inf(x,y)) \Rightarrow (\neg\, r_T(z) \curlyvee z \approx c_0)))$

Conclusions and Future Work

Conclusions

- $\mathscr{B}_{F,T}(X,Y,I)$ and its properities
- transferring theorem
- structure of $\mathscr{B}_{F,T}(X,Y,I)$

$\mathscr{B}_{F,T}(X,Y,I)$ is huge. It can not be used in practice.

Future work

- Find appropriate subset of $\mathscr{B}_{F,T}(X,Y,I)$ being subalgebra of $\mathscr{B}(X,Y,I)$.