Formal Concept Analysis & Logical Analysis of Data

Jan Konečný

DATA ANALYSIS AND MODELING LAB

Palacky University, Olomouc, Czech Republic

Outline

- Logical Analysis of Data (LAD)
- Fuzzy Rough Concept Analysis (FRCA) $=$ FCA $+$ Fuzzy sets $+$ Rough sets
- FRCA $\wedge$ LAD
- FRCA $\vee$ LAD

LOGICAL ANALYSIS OF DATA

Logical Analysis of Data

Sources:

I. Chikalov, V. Lozin, I. Lozina, M. Moshkov, H.S. Nguyen, A. Skowron, B. Zielosko
Three Approaches to Data Analysis:
Test Theory, Rough Sets and Logical Analysis of Data
Series: Intelligent Systems Reference Library, Vol. 41 2013, XVIII, 202 p.

G. Alexe, S. Alexe, T.O. Bonates, A. Kogan
Logical Analysis of Data — the Vision of Peter L. Hammer.
Annals of Mathematics and Artificial Intelligence, April 2007, 49(1-4), pp. 265–312.

Wikipedia:
Peter Ladislaw Hammer (December 23, 1936 – December 27, 2006) was an American mathematician native to Romania. He contributed to the fields of operations research and applied discrete mathematics through the study of pseudo-Boolean functions and their connections to graph theory and data mining.

LAD – Input: Dataset (Context)

Denote

$$\Omega^+ = \langle X^+, Y, I^+ \rangle, \Omega^- = \langle X^-, Y, I^- \rangle, \Omega = \langle X^+ \cup X^-, Y, I^+ \cup I^- \rangle.$$

		y_1	y_2	y_3	y_4	y_5
Ω^+	a	1	0	1	1	1
	b	0	0	0	1	1
	c	1	1	1	1	1
	d	1	1	1	0	1
	e	1	1	1	0	0
	p	1	0	0	1	0
	q	0	0	1	0	1
Ω^-	r	1	0	1	0	0
	s	1	0	0	0	0
	t	0	0	1	0	0

LAD – Overview

- Redundant variables in the original dataset we extract from it a subset S, capable of distinguishing the positive observations from the negative ones.
- Cover dataset Ω^+ with a family of possibly overlapping homogeneous subsets of $\{0,1\}^n$, each of these subsets having a significant intersection with with Ω^+, but being disjoint from Ω^-. Similarly handle dataset Ω^-.
- A subset of the positive (resp. negative) patterns, the union of which covers every observation in Ω^+ (resp. Ω^-) is identified. The collection of these two subsets of intervals is called a "model."
- A classification method is developed which defines the positive or negative intervals of the model, leaving as "unclassified" those observations which are not covered by this union.
- One of the standard validation methods is applied to verify the accuracy of the resulting classification system.

LAD – Terms

- *Term* over Y – conjunction of literals,
- *Literal* – either y or $\neg y$.

Example

$C = \neg y_1 y_3$

For term C, denote

- $\mathrm{Pos}(C)$ – positive literals of C,
- $\mathrm{Neg}(C)$ – negative literals of C,
- $\mathrm{Lit}(C)$ – all literals of C; $\mathrm{Pos}(C) \cup \mathrm{Neg}(C)$,
- $\mathrm{Mod}(C)$ – set of all models of C

  that is, evaluations w, s.t. $\|C\|_w = 1$.

LAD – Terms

$\mathrm{Mod}(C)$ forms n-dimensional subcube of $\{0,1\}^Y$; $n = |Y - \mathrm{Lit}(C)|$.

Example

Considering $Y = \{y_1, y_2, \ldots, y_n\}$ we can unify evaluation w with the string $w(y_1)w(y_2)\ldots w(y_n)$.

$$C = \neg y_1 y_3$$

$$
\begin{tikzcd}
 & \mathbf{0}1\mathbf{1}11 \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \\
\mathbf{0}1\mathbf{1}10 \arrow[d, dash] \arrow[dr, dash] & \mathbf{0}1\mathbf{1}01 \arrow[dl, dash] \arrow[dr, dash] & \mathbf{0}0\mathbf{1}11 \arrow[dl, dash] \arrow[d, dash] \\
\mathbf{0}1\mathbf{1}00 \arrow[dr, dash] & \mathbf{0}0\mathbf{1}10 \arrow[d, dash] & \mathbf{0}0\mathbf{1}01 \arrow[dl, dash] \\
 & \mathbf{0}0\mathbf{1}00 &
\end{tikzcd}
$$

Such subcubes of $\{0,1\}^Y$ are in one-to-one correspondence with terms over Y.

LAD – Patterns

- Basic notion in LAD
- Positive pattern is simply a subcube of $\{0,1\}^Y$ which intersect Ω^+ and is disjoint from Ω^-.

  Negative patterns have a similar definition.

Definition

A term C is called a *positive pattern* of a dataset Ω if

- $\|C\|_w = 0$ for every $w \in \Omega^-$,
- $\|C\|_w = 1$ for at least one vector $w \in \Omega^+$.

A term C is called a *negative pattern* of a dataset Ω if

- $\|C\|_w = 0$ for every $w \in \Omega^+$,
- $\|C\|_w = 1$ for at least one vector $w \in \Omega^-$.

Pareto-optimality of Patterns

Definition

Given a preorder $\preceq$ on the set of patterns, a pattern P will be called *pareto-optimal* with respect to $\preceq$, if there is no distinct pattern P' such that $P \preceq P'$.

Definition (Simplicity preference)

A pattern P_1 is *simplicity-wise preferred* to a pattern P_2 (denoted by $P_2 \preceq_\sigma P_1$) if $\mathrm{Lit}(P_1) \supseteq \mathrm{Lit}(P_2)$.

Pareto-optimal patters w.r.t. $\preceq_\sigma$ are called *prime*

Remark

Inspired by the Occam's razor.

Pareto-optimality of Patterns

Definition (Evidential preference)

A pattern P_1 is *evidentially preferred* to a pattern P_2 (denoted by $P_2 \preceq_\epsilon P_1$) if $\mathrm{Cov}(P_1) \supseteq \mathrm{Cov}(P_2)$.

$\mathrm{Cov}(P)$ denotes $\mathrm{Mod}(P) \cap \Omega$.

Evidentially Pareto-optimal patterns are called *strong*.

Definition (Evidential preference)

A pattern P_1 is *selectively-wise preferred* to a pattern P_2 (denoted by $P_2 \preceq_\Sigma P_1$) if and only if $\mathrm{Mod}(P_1) \subseteq \mathrm{Mod}(P_2)$.

Pareto-optimal patterns w.r.t. $\Sigma \wedge \epsilon$ are called *spanned*.

Classification with LAD

Lets have

- Γ^+ – collection of (selected) positive patterns, s.t. it covers Ω^+,
- Γ^- – collection of (selected) negative patterns, s.t. it covers Ω^-.

For collection of positive (or negative) patterns Γ and new observation w define

$$\delta(w, \Gamma) = \{\|P\|_w \mid P \in \Gamma\}$$

Diskriminant

$$\Delta(w) = \frac{|\delta(w, \Gamma^+)|}{|\Gamma^+|} - \frac{|\delta(w, \Gamma^-)|}{|\Gamma^-|}.$$

FUZZY CONCEPT ANALYSIS

more precisely...

Structure of truth degrees = complete residuated lattice
$\mathbf{L} = \langle L, \wedge, \vee, \otimes, \rightarrow, 0, 1\rangle$
$\langle L, \wedge, \vee, 0, 1\rangle$ ... complete lattice
$\langle L, \otimes, 1\rangle$ ... commutative monoid
$\langle \otimes, \rightarrow\rangle$ ... adjoint pair ($a \otimes b \leqslant c$ iff $a \leqslant b \rightarrow c$)

L-set A in universe U ... mapping A: $U \rightarrow L$
Interpretation of $A(u)$: "degree to which u belongs to A"

Operations with $\mathbf{L}$-sets defined component-wise

- $\wedge$-intersection $(A \cup B)(u) = A(u) \cup B(u)$
- complement $(\neg A)(u) = A(u) \rightarrow 0$

Set of all L-sets in U is denoted by L^U.

Binary L-relation R between sets U, V ... mapping $R : U \times V \rightarrow L$,
Interpretation of $R(u, v)$: "degree to which u and v are R-related"

Fuzzy Concept Analysis

Fuzzy Context – triple $\langle X, Y, I \rangle$

X ...(finite crisp) set of objects
Y ...(finite crisp) set of attributes
I ...L-relation $I : X \times Y \to L$

	y_1	y_2	y_3	y_4
x_1	0.1	1.0	0.2	0.3
x_2	1.0	0.5	0.0	0.6
x_3	0.5	0.8	1.0	0.6
x_4	0.0	0.0	1.0	1.0

Antitone $\mathbf{L}$-concept-forming operators: $(\cdot)^{\uparrow} : L^X \to L^Y$, $(\cdot)^{\downarrow} : L^Y \to L^X$.

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

Formal concept w.r.t. $\langle \uparrow, \downarrow \rangle$ is pair $\langle A, B \rangle$ s.t. $A^{\uparrow} = B, B^{\downarrow} = A$
A=extent, B=intent

Concept lattice

$$\mathcal{B}^{\uparrow\downarrow}(X, Y, I) = \{\langle A, B \rangle \mid A^{\uparrow} = B, B^{\downarrow} = A\}$$

Fuzzy Concept Analysis – Isotone case

Isotone $\mathbf{L}$-concept-forming operators $\langle \cap, \cup \rangle$: $(\cdot)^{\cap} : L^X \to L^Y$, $(\cdot)^{\cup} : L^Y \to L^X$.

$$A^{\cap}(y) = \bigvee_{x \in X} A(x) \otimes I(x, y) \quad \text{and} \quad B^{\cup}(x) = \bigwedge_{y \in Y} I(x, y) \to B(y)$$

Formal concept w.r.t. $\langle \cap, \cup \rangle$ is pair $\langle A, B \rangle$ s.t. $A^{\cap} = B, B^{\cup} = A$; A=extent, B=intent

Concept lattice

$$\mathcal{B}^{\cap\cup}(X, Y, I) = \{\langle A, B \rangle \mid A^{\cap} = B, B^{\cup} = A\}$$

Isotone $\mathbf{L}$-concept-forming operators $\langle \wedge, \vee \rangle$ $(\cdot)^{\wedge} : L^X \to L^Y$, $(\cdot)^{\vee} : L^Y \to L^X$.

$$A^{\wedge}(y) = \bigwedge_{x \in X} I(x, y) \to A(x) \quad \text{and} \quad B^{\vee}(x) = \bigvee_{y \in Y} B(y) \otimes I(x, y)$$

Formal concept w.r.t. $\langle \wedge, \vee \rangle$ is pair $\langle A, B \rangle$ s.t. $A^{\wedge} = B, B^{\vee} = A$; A=extent, B=intent

Concept lattice

$$\mathcal{B}^{\wedge\vee}(X, Y, I) = \{\langle A, B \rangle \mid A^{\wedge} = B, B^{\vee} = A\}$$

intermezzo: ROUGH SETS

Rough Sets

Pawlak approximation space – $\langle U, E \rangle$, where

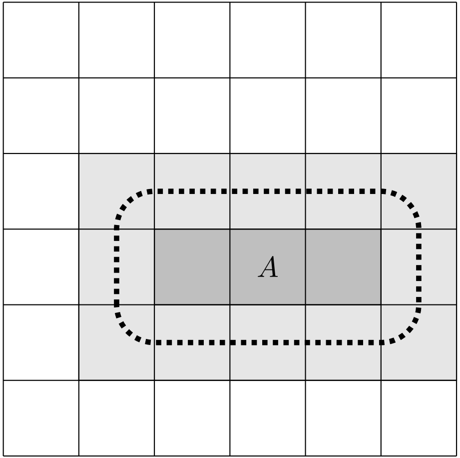
- U is a non-empty set of *objects* (universe),
- E is an equivalence relation on U,

the *rough approximation* of a crisp set $A \subseteq U$ by E is the pair $\langle A^{\Downarrow_E}, A^{\Uparrow_E} \rangle$ of sets in U defined by

$$
\begin{aligned}
x \in A^{\Downarrow_E} \quad & \text{iff} \quad ((\forall y \in U)\,\langle x, y \rangle \in E \text{ implies } y \in A), \\
x \in A^{\Uparrow_E} \quad & \text{iff} \quad ((\exists y \in U)\,\langle x, y \rangle \in E \text{ and } y \in A).
\end{aligned}
$$

$A^{\Downarrow_E}$ and $A^{\Uparrow_E}$ are called *lower and upper approximation* of the set A by E, respectively.

Rough Sets



set A

$A^{\Uparrow_E}$ – upper approx. of A

$A^{\Downarrow_E}$ – lower approx. of A

Fuzzy Rough Sets

In the fuzzy setting, one can generalize $\langle A^{\Downarrow_E}, A^{\Uparrow_E} \rangle$ as in

Didier Dubois and Henri Prade.
Rough fuzzy sets and fuzzy rough sets.
International Journal of General Systems, 17(2–3):191–209, 1990.

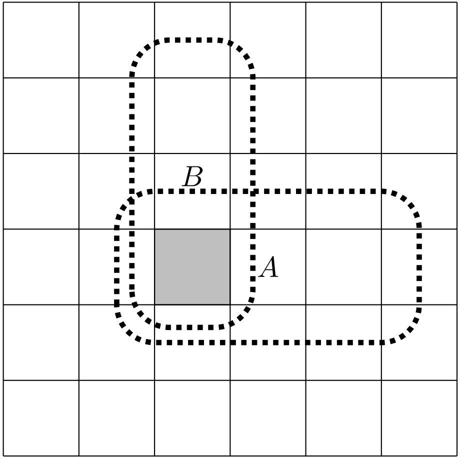
Didier Dubois and Henri Prade.
Putting rough sets and fuzzy sets together.
Intelligent Decision Support, volume 11 of *Theory and Decision Library*, pages 203–232., 1992.

Anna Maria Radzikowska and Etienne E. Kerre.
Fuzzy rough sets based on residuated lattices.
Transactions on Rough Sets II, volume 3135 of *Lecture Notes in Computer Science*, pages 278–296., 2005.

$$A^{\Downarrow_E}(x) = \bigwedge_{y \in U} (E(x,y) \to A(y)) \quad \text{and} \quad A^{\Uparrow_E}(x) = \bigvee_{y \in U} (A(y) \otimes E(x,y))$$

for $\mathbf{L}$-equivalence $E \in \mathbf{L}^{U \times U}$ and $\mathbf{L}$-set $A \in \mathbf{L}^U$.

Rough Sets – properties of approximations



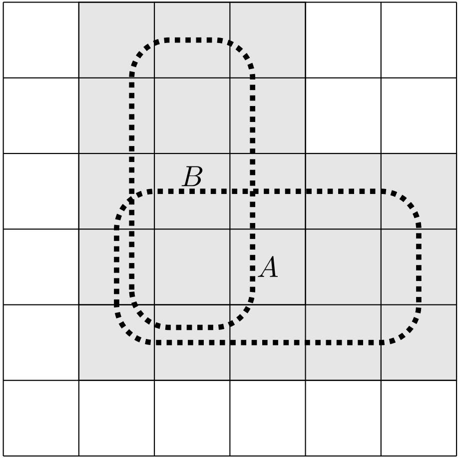
sets A, B

$(A \cap B)^{\Downarrow_E}$

We have:

$$(A \cap B)^{\Downarrow_E} = A^{\Downarrow_E} \cap B^{\Downarrow_E}$$

Rough Sets – properties of approximations



sets A, B

$(A \cup B)^{\Uparrow_E}$

We have:

$$(A \cup B)^{\Uparrow_E} = A^{\Uparrow_E} \cup B^{\Uparrow_E}$$

(Fuzzy Rough) Concept Analysis

Observation

Intents in $\mathcal{B}^{\cap\cup}(X, Y, I)$ behave like upper approximations in FRS.
Intents in $\mathcal{B}^{\uparrow\downarrow}(X, Y, I)$ behave like lower approximations in FRS.

Robert E. Kent
Rough Concept Analysis
Rough Sets, Fuzzy Sets and Knowledge Discovery
Workshops in Computing 1994, pp 248-255

Ming-Wen Shao, Min Liu, and Wen-Xiu Zhang.
Set approximations in fuzzy formal concept analysis.
Fuzzy Sets Syst., 158(23):2627–2640, December 2007.

More precisely,. . .

$$A^{\Uparrow_E} = A^{\cap_E} \text{ and } A^{\Downarrow_E} = A^{\wedge_E}$$

Let $E \in L^{Y \times Y}$ be Leibniz $\mathbf{L}$-equivalence induced by $I \subseteq L^{X \times Y}$, that is

$$E(y_1, y_2) = \bigwedge_{x \in X} I(x, y_1) \leftrightarrow I(x, y_1),$$

then E is compatible with I:

$$I = I \circ E = I \triangleright E.$$

where

$$\begin{aligned}
(A \circ B)(x, y) &= \bigvee_{f \in F} A(x, f) \otimes B(f, y), \\
(A \triangleright B)(x, y) &= \bigwedge_{f \in F} B(f, y) \to A(x, f).
\end{aligned}$$

$$I = I \circ E = I \triangleright E.$$

R. Belohlavek.
Fuzzy Relational Systems: Foundations and Principles.
Kluwer Academic Publishers, Norwell, USA, 2002.

From that we have

$$A^{\uparrow} = (A^{\uparrow})^{\wedge_E},$$
$$A^{\cap} = (A^{\cap})^{\cap_E}.$$

Belohlavek R., Konecny J.
Row and Column Spaces of Matrices over Residuated Lattices.
Fundamenta Informaticae 115(4)(2012), 279-295.

(Fuzzy Rough) Concept Analysis

Definition

Let $\langle X, Y, I\rangle$ be an $\mathbf{L}$-context. Define $\mathbf{L}$-*rough concept-forming operators* as

$$A^{\vartriangle} = \langle A^{\uparrow}, A^{\cap}\rangle \quad \text{and} \quad \langle \underline{B}, \overline{B}\rangle^{\triangledown} = \underline{B}^{\downarrow} \cap \overline{B}^{\cup}$$

for $A \in \mathbf{L}^X$, $\underline{B}, \overline{B} \in \mathbf{L}^Y$.

$\mathbf{L}$-*rough concept* is then a fixed point of $\langle \vartriangle, \triangledown\rangle$, i.e. a pair $\langle A, \langle \underline{B}, \overline{B}\rangle\rangle \in \mathbf{L}^X \times (\mathbf{L} \times \mathbf{L})^Y$ such that

$$A^{\vartriangle} = \langle \underline{B}, \overline{B}\rangle \quad \text{and} \quad \langle \underline{B}, \overline{B}\rangle^{\triangledown} = A.$$

$A^{\uparrow}$ and $A^{\cap}$ are called *lower intent approximation* and *upper intent approximation*, respectively.

Will be presented at CLA 2014.

THE LINK BETWEEN FRCA AND LAD

Crisp Case $L = \{0, 1\}$

- not very interesting
- similar results would be obtained using apposition of the context with its complement.

	y_1	y_2	y_3
x_1	0	1	0
x_2	1	1	0
x_2	1	0	1

$\Longrightarrow$

	y_1	y_2	y_3	$\neg y_1$	$\neg y_2$	$\neg y_3$
x_1	0	1	0	1	0	1
x_2	1	1	0	0	0	1
x_2	1	0	1	0	1	0

Still, it provides a connection to LAD.

(crisp) rough sets on Y correspond to subcubes of $\{0,1\}^Y$

Example

$\langle \underline{A}, \overline{A} \rangle = \langle \{y_3\}, \{y_2, y_3, y_4, y_5\} \rangle$

Characteristic vectors of sets for which $\langle \underline{A}, \overline{A} \rangle$ is their rough approximation:

$$C = \neg y_1 y_3$$

$$\begin{tikzcd}
 & \mathbf{0}1\mathbf{1}11 \arrow[dl, dash] \arrow[d, dash] \arrow[dr, dash] & \\
\mathbf{0}1\mathbf{1}10 \arrow[d, dash] \arrow[dr, dash] & \mathbf{0}1\mathbf{1}01 \arrow[dl, dash] \arrow[dr, dash] & \mathbf{0}0\mathbf{1}11 \arrow[dl, dash] \arrow[d, dash] \\
\mathbf{0}1\mathbf{1}00 \arrow[dr, dash] & \mathbf{0}0\mathbf{1}10 \arrow[d, dash] & \mathbf{0}0\mathbf{1}01 \arrow[dl, dash] \\
 & \mathbf{0}0\mathbf{1}00 &
\end{tikzcd}$$

Formal Rough (Crisp) Concept Analysis and Logical Analysis of Data

Denote

$$\Omega^+ = \langle X^+, Y, I^+\rangle, \Omega^- = \langle X^-, Y, I^-\rangle, \Omega = \langle X^+ \cup X^-, Y, I^+ \cup I^-\rangle.$$

Definition

For a term C define pair $\mathrm{rs}(C)$ of sets as

$$\mathrm{rs}(C) \mapsto \langle \mathrm{Pos}(C), Y - \mathrm{Neg}(C)\rangle.$$

Theorem

Term C is a positive pattern iff

$$\emptyset \neq \mathrm{rs}(C)^{\nabla\Omega} \subseteq X^+ \textit{ and } \mathrm{rs}(C)^{\nabla\Omega} \cap X^- = \emptyset.$$

Term C is a negative pattern *iff*

$$\emptyset \neq \mathrm{rs}(C)^{\nabla\Omega} \subseteq X^- \textit{ and } \mathrm{rs}(C)^{\nabla\Omega} \cap X^+ = \emptyset.$$

Denote

$$\hat{\Omega} = \langle X^+ \cup X^-, Y \cup \{d\}, I^+ \cup I^- \cup D \rangle.$$

Theorem

Term C *is a* positive pattern *iff*

$$\mathrm{rs}(C \cdot d)^{\nabla_{\hat{\Omega}}} = \mathrm{rs}(C)^{\nabla_{\hat{\Omega}}} \neq \emptyset.$$

Term C *is a* negative pattern *iff*

$$\mathrm{rs}(C \cdot \neg d)^{\nabla_{\hat{\Omega}}} = \mathrm{rs}(C)^{\nabla_{\hat{\Omega}}} \neq \emptyset.$$

Theorem

$P_1 \preceq_\Sigma P_2$ *iff* $\mathrm{rs}(P_1) \subseteq \mathrm{rs}(P_2)$.

Theorem

$P_1 \preceq_\epsilon P_2$ *iff* $\mathrm{rs}(P_1)^{\nabla\omega} \subseteq \mathrm{rs}(P_2)^{\nabla\omega}$.

Theorem

Pattern P *is spanned iff* $\mathrm{rs}(P)$ *is intent in* $\mathcal{B}^{\Delta\nabla}(\Omega)$.

Conclusions

We have some meeting points between FRCA and LAD.
What now?

- Algorithms for LAD based on FCA
- Fuzzy setting

Algorithm SPIC (for generating all spanned patterns)

Input: C_0: the collection of patterns spanned by each individual observation in Ω^+.
Initialize $C := C_0$
Repeat the following operation until the collection C cannot be furthermore enlarged.
    if their consensus P' exists and
        if it is not absorbed by a pattern already contained in C,
            then add it to C.

From the point of view of FCA this is a naïve generation of (part of) a concept lattice.

Algorithms for LAD based on FCA

Petr Krajca, Jan Outrata, Vilem Vychodil
Advances in algorithms based on CbO.
Proc. CLA 2010, 2010, pp. 325337.

Patterns as closure systems.

$$\text{close-positive}(A) = \begin{cases} A & \text{if } X^{-} \cap A = \emptyset, \\ X & \text{otherwise.} \end{cases}$$

Belohlavek R., Vychodil V.
Closure based constraints in formal concept analysis.
Discrete Applied Mathematics 161(13-14)(2013), 1894-1911. closures

Algorithms for LAD based on FCA

Do they know or not?

From

Alexe, Gabriela and Alexe, Sorin and Bonates, Tibérius O. and Kogan, Alexander.
Logical Analysis of Data — the Vision of Peter L. Hammer.
Annals of Mathematics and Artificial Intelligence, April 2007, 49(1-4), pp. 265–312.

For instance, Malgrange (ref) used a consensus-type approach to find all maximal submatrices consisting of ones of a 0-1 matrix (see also Kuznetsov and Obiedkov (ref) for references to algorithms with polynomial delay), while a concensus-type algorithm for finding all maximal bicliques of a graph was presented in (ref).

Kuznetsov, S.O., Obiedkov S.A.
Comparing performance of algorithms for generating concept lattices.
J. Exp. Theor. Artif. Intell. **14**, 189-216 (2002)

Fuzzy Setting

Since the concept-forming operators $\langle\vartriangle, \triangledown\rangle$ are defined in fuzzy setting, we have a direct lead to **fuzzy logical analysis of data**.

THANK YOU FOR YOUR ATTENTION.