Ordered Sets for Data Analysis: An Algorithmic Approach

Introduction to Formal Concept Analysis

Sergei O. Kuznetsov

National Research University Higher School of Economics, Moscow, Russia

October 11, 2022

[xcolor=table]beamer amsmathamsfontsamssymbgraphicx

When statistics and deep learning fail

- Comparison of case studies with relatively small cases (too small datasets to identify complex causation factors).
- Need of explainable/interpretable results
- Need of robust explanations of particular cases and their outcomes.

In which domains the problems usually arise?

- Social science (politics, economics, etc.).
- Linguistics.
- Natural science (chemistry, biology, geology, etc.).
- and so on.

Case Study: Banking

Problem: create standards of bank operations in order to prevent crimes in the international banking system.
Solution: establishment of the Wolfsberg Group of leading international banks to create basic banking principles.

Task from the perspective of data analysts: identifying the configuration of causal conditions for participation by top banks in such the Wolfsberg agreement.

Dataset

26 top international banks (11 of them make up the Wolfsberg Group)

Case	Country	BankType	Owner	Code	CorpC	FinLib	RegInt	BList	**Wolfs**
ABNAmro	NLD	Universal	Public	No	0.74	0.98	144	Yes	**No**
Barclays	GBR	Commercial	Public	Explicit	0.14	1	277	Yes	**Yes**
BNP Paribas	FRA	Universal	Public	Implicit	0.82	1	75	No	**No**
Carnegie	SWE	Investment	Public	No	0.71	0.95	83	No	**No**
Citigroup	USA	Commercial	Public	Explicit	0	1	426	Yes	**Yes**
Coutts & Co	GBR	Private	Public	Implicit	0.14	1	277	Yes	**No**
CS	CHE	Universal	Public	Explicit	0.44	0.95	83	Yes	**Yes**
Deutsche Bank	GER	Universal	Public	Explicit	0.95	0.9	45	No	**Yes**
Goldman Sachs	USA	Investment	Public	Explicit	0	1	426	Yes	**Yes**
HSBC	GBR	Commercial	Public	Implicit	0.14	1	277	Yes	**Yes**
ING	NLD	Universal	Public	Implicit	0.74	0.98	144	Yes	**No**
JPMorgan	USA	Commercial	Public	Explicit	0	1	426	Yes	**Yes**
Julius Baer	CHE	Private	Public	No	0.44	0.95	83	Yes	**No**
LCF	FRA	Private	Private	No	0.82	1	75	No	**No**
LODH	CHE	Private	Private	No	0.44	0.95	83	Yes	**No**
MeesPierson	NLD	Private	Public	No	0.74	0.98	144	Yes	**No**
Merrill Lynch	USA	Investment	Public	Implicit	0	1	426	Yes	**No**
Morgan S	USA	Investment	Public	Explicit	0.14	1	277	Yes	**No**
Nordea	SWE	Universal	Public	Implicit	0.71	0.95	83	No	**No**
Pictet & Cie	CHE	Private	Private	No	0.44	0.95	83	Yes	**No**
RBC	CAN	Universal	Public	Explicit	0.23	1	149	No	**No**
Rothschild	GBR	Investment	Private	No	0.14	1	277	Yes	**No**
Santander	ESP	Universal	Public	Explicit	0.77	1	53	No	**Yes**
SG	FRA	Universal	Public	Explicit	0.82	1	75	No	**Yes**
UBP	CHE	Private	Private	No	0.44	0.95	83	Yes	**No**
UBS	CHE	Universal	Public	Explicit	0.44	0.95	83	Yes	**Yes**

Attribute descriptions

Attributes	Intuition behind them / hypotheses
BankType the type of bank	banks focused on investment and private banking would have less incentive to join the WA
Owner	public ownership would increase the likelihood of joining the WA
Code an internal code of conduct	prior existence of self-imposed codes of conduct against money laundering a driver for joining the WA
CorpC the coordination of corporate relationships	higher coordination of corporate relationships will increase an individual bank's ability to participate in the WA
FinLib financial liberalization	liberalization (on the average of a particular indicator regarding financial reform) would be positively related to the need for multistakeholder agreements.
RegInt the stringency of regulations	regulatory intensity would lead to compliance with the WA
BList	presence on the IMF blacklist pressure on banks to participate in multistakeholder agreements

Concepts: clusters of banks

Case	Country	BankType	Owner	Code	CorpC	FinLib	RegInt	BList	**Wolfs**
ABNAmro	NLD	Universal	Public	No	0.74	0.98	144	Yes	**No**
Barclays	GBR	Commercial	Public	Explicit	0.14	1	277	Yes	**Yes**
BNP Paribas	FRA	Universal	Public	Implicit	0.82	1	75	No	**No**
Carnegie	SWE	Investment	Public	No	0.71	0.95	83	No	**No**
Citigroup	USA	Commercial	Public	Explicit	0	1	426	Yes	**Yes**
Coutts & Co	GBR	Private	Public	Implicit	0.14	1	277	Yes	**No**
CS	CHE	Universal	Public	Explicit	0.44	0.95	83	Yes	**Yes**
Deutsche Bank	GER	Universal	Public	Explicit	0.95	0.9	45	No	**Yes**
Goldman Sachs	USA	Investment	Public	Explicit	0	1	426	Yes	**Yes**
HSBC	GBR	Commercial	Public	Implicit	0.14	1	277	Yes	**Yes**
ING	NLD	Universal	Public	Implicit	0.74	0.98	144	Yes	**No**
JPMorgan	USA	Commercial	Public	Explicit	0	1	426	Yes	**Yes**
Julius Baer	CHE	Private	Public	No	0.44	0.95	83	Yes	**No**
LCF	FRA	Private	Private	No	0.82	1	75	No	**No**
LODH	CHE	Private	Private	No	0.44	0.95	83	Yes	**No**
MeesPierson	NLD	Private	Public	No	0.74	0.98	144	Yes	**No**
Merrill Lynch	USA	Investment	Public	Implicit	0	1	426	Yes	**No**
Morgan S	USA	Investment	Public	Explicit	0.14	1	277	Yes	**No**
Nordea	SWE	Universal	Public	Implicit	0.71	0.95	83	No	**No**
Pictet & Cie	CHE	Private	Private	No	0.44	0.95	83	Yes	**No**
RBC	CAN	Universal	Public	Explicit	0.23	1	149	No	**No**
Rothschild	GBR	Investment	Private	No	0.14	1	277	Yes	**No**
Santander	ESP	Universal	Public	Explicit	0.77	1	53	No	**Yes**
SG	FRA	Universal	Public	Explicit	0.82	1	75	No	**Yes**
UBP	CHE	Private	Private	No	0.44	0.95	83	Yes	**No**
UBS	CHE	Universal	Public	Explicit	0.44	0.95	83	Yes	**Yes**

Banks with high regulatory intensity (RegInt > 200):
Barclays, Citigroup, Coutts Co, Goldman Sachs, HSBC, JPMorgan, Merrill Lynch, Morgan S, Rothschild.
Their common properties: BList: Yes, FinLib: 1, RegInt > 200.

Implications between bank properties

Case	Country	BankType	Owner	Code	CorpC	FinLib	RegInt	BList	**Wolfs**
ABNAmro	NLD	Universal	Public	No	0.74	0.98	144	Yes	**No**
Barclays	GBR	Commercial	Public	Explicit	0.14	1	277	Yes	**Yes**
BNP Paribas	FRA	Universal	Public	Implicit	0.82	1	75	No	**No**
Carnegie	SWE	Investment	Public	No	0.71	0.95	83	No	**No**
Citigroup	USA	Commercial	Public	Explicit	0	1	426	Yes	**Yes**
Coutts & Co	GBR	Private	Public	Implicit	0.14	1	277	Yes	**No**
CS	CHE	Universal	Public	Explicit	0.44	0.95	83	Yes	**Yes**
Deutsche Bank	GER	Universal	Public	Explicit	0.95	0.9	45	No	**Yes**
Goldman Sachs	USA	Investment	Public	Explicit	0	1	426	Yes	**Yes**
HSBC	GBR	Commercial	Public	Implicit	0.14	1	277	Yes	**Yes**
ING	NLD	Universal	Public	Implicit	0.74	0.98	144	Yes	**No**
JPMorgan	USA	Commercial	Public	Explicit	0	1	426	Yes	**Yes**
Julius Baer	CHE	Private	Public	No	0.44	0.95	83	Yes	**No**
LCF	FRA	Private	Private	No	0.82	1	75	No	**No**
LODH	CHE	Private	Private	No	0.44	0.95	83	Yes	**No**
MeesPierson	NLD	Private	Public	No	0.74	0.98	144	Yes	**No**
Merrill Lynch	USA	Investment	Public	Implicit	0	1	426	Yes	**No**
Morgan S	USA	Investment	Public	Explicit	0.14	1	277	Yes	**No**
Nordea	SWE	Universal	Public	Implicit	0.71	0.95	83	No	**No**
Pictet & Cie	CHE	Private	Private	No	0.44	0.95	83	Yes	**No**
RBC	CAN	Universal	Public	Explicit	0.23	1	149	No	**No**
Rothschild	GBR	Investment	Private	No	0.14	1	277	Yes	**No**
Santander	ESP	Universal	Public	Explicit	0.77	1	53	No	**Yes**
SG	FRA	Universal	Public	Explicit	0.82	1	75	No	**Yes**
UBP	CHE	Private	Private	No	0.44	0.95	83	Yes	**No**
UBS	CHE	Universal	Public	Explicit	0.44	0.95	83	Yes	**Yes**

BankType: Commercial $\rightarrow$ Wolfsberg: Yes

Implications between bank properties

Case	Country	BankType	Owner	Code	CorpC	FinLib	RegInt	BList	**Wolfs**
ABNAmro	NLD	Universal	Public	No	0.74	0.98	144	Yes	**No**
Barclays	GBR	Commercial	Public	Explicit	0.14	1	277	Yes	**Yes**
BNP Paribas	FRA	Universal	Public	Implicit	0.82	1	75	No	**No**
Carnegie	SWE	Investment	Public	No	0.71	0.95	83	No	**No**
Citigroup	USA	Commercial	Public	Explicit	0	1	426	Yes	**Yes**
Coutts & Co	GBR	Private	Public	Implicit	0.14	1	277	Yes	**No**
CS	CHE	Universal	Public	Explicit	0.44	0.95	83	Yes	**Yes**
Deutsche Bank	GER	Universal	Public	Explicit	0.95	0.9	45	No	**Yes**
Goldman Sachs	USA	Investment	Public	Explicit	0	1	426	Yes	**Yes**
HSBC	GBR	Commercial	Public	Implicit	0.14	1	277	Yes	**Yes**
ING	NLD	Universal	Public	Implicit	0.74	0.98	144	Yes	**No**
JPMorgan	USA	Commercial	Public	Explicit	0	1	426	Yes	**Yes**
Julius Baer	CHE	Private	Public	No	0.44	0.95	83	Yes	**No**
LCF	FRA	Private	Private	No	0.82	1	75	No	**No**
LODH	CHE	Private	Private	No	0.44	0.95	83	Yes	**No**
MeesPierson	NLD	Private	Public	No	0.74	0.98	144	Yes	**No**
Merrill Lynch	USA	Investment	Public	Implicit	0	1	426	Yes	**No**
Morgan S	USA	Investment	Public	Explicit	0.14	1	277	Yes	**No**
Nordea	SWE	Universal	Public	Implicit	0.71	0.95	83	No	**No**
Pictet & Cie	CHE	Private	Private	No	0.44	0.95	83	Yes	**No**
RBC	CAN	Universal	Public	Explicit	0.23	1	149	No	**No**
Rothschild	GBR	Investment	Private	No	0.14	1	277	Yes	**No**
Santander	ESP	Universal	Public	Explicit	0.77	1	53	No	**Yes**
SG	FRA	Universal	Public	Explicit	0.82	1	75	No	**Yes**
UBP	CHE	Private	Private	No	0.44	0.95	83	Yes	**No**
UBS	CHE	Universal	Public	Explicit	0.44	0.95	83	Yes	**Yes**

BankType: Private $\rightarrow$ Wolfsberg: No

Association Rules between bank properties

Case	Country	BankType	Owner	Code	CorpC	FinLib	RegInt	BList	**Wolfs**
ABNAmro	NLD	Universal	Public	No	0.74	0.98	144	Yes	**No**
Barclays	GBR	Commercial	Public	Explicit	0.14	1	277	Yes	**Yes**
BNP Paribas	FRA	Universal	Public	Implicit	0.82	1	75	No	**No**
Carnegie	SWE	Investment	Public	No	0.71	0.95	83	No	**No**
Citigroup	USA	Commercial	Public	Explicit	0	1	426	Yes	**Yes**
Coutts & Co	GBR	Private	Public	Implicit	0.14	1	277	Yes	**No**
CS	CHE	Universal	Public	Explicit	0.44	0.95	83	Yes	**Yes**
Deutsche Bank	GER	Universal	Public	Explicit	0.95	0.9	45	No	**Yes**
Goldman Sachs	USA	Investment	Public	Explicit	0	1	426	Yes	**Yes**
HSBC	GBR	Commercial	Public	Implicit	0.14	1	277	Yes	**Yes**
ING	NLD	Universal	Public	Implicit	0.74	0.98	144	Yes	**No**
JPMorgan	USA	Commercial	Public	Explicit	0	1	426	Yes	**Yes**
Julius Baer	CHE	Private	Public	No	0.44	0.95	83	Yes	**No**
LCF	FRA	Private	Private	No	0.82	1	75	No	**No**
LODH	CHE	Private	Private	No	0.44	0.95	83	Yes	**No**
MeesPierson	NLD	Private	Public	No	0.74	0.98	144	Yes	**No**
Merrill Lynch	USA	Investment	Public	Implicit	0	1	426	Yes	**No**
Morgan S	USA	Investment	Public	Explicit	0.14	1	277	Yes	**No**
Nordea	SWE	Universal	Public	Implicit	0.71	0.95	83	No	**No**
Pictet & Cie	CHE	Private	Private	No	0.44	0.95	83	Yes	**No**
RBC	CAN	Universal	Public	Explicit	0.23	1	149	No	**No**
Rothschild	GBR	Investment	Private	No	0.14	1	277	Yes	**No**
Santander	ESP	Universal	Public	Explicit	0.77	1	53	No	**Yes**
SG	FRA	Universal	Public	Explicit	0.82	1	75	No	**Yes**
UBP	CHE	Private	Private	No	0.44	0.95	83	Yes	**No**
UBS	CHE	Universal	Public	Explicit	0.44	0.95	83	Yes	**Yes**

BankType: Universal & Owner: Public & Code: Explicit $\rightarrow$ Wolfsberg: Yes

confidence (conditional probability) $= 5/6$

Formal Concept Analysis. 1

[Ganter, Wille 1999]

Let two sets G and M be given. Elements of G are called **objects**, elements of M are called **attributes**.

Let $I \subseteq G \times M$ be a binary relation. If $(g, m) \in I$, one says that object g has attribute m. Triple $\mathbb{K} \stackrel{\text{def}}{=} (G, M, I)$ is called a **(formal) context**.

Formal Concept Analysis. 2

[Ganter, Wille 1999]

For a given context $\mathbb{K} := (G, M, I)$ consider two mappings $\varphi\colon 2^G \to 2^M$ and $\psi\colon 2^M \to 2^G$:

$$\varphi(A) \stackrel{\text{def}}{=} \{m \in M \mid gIm \ \text{ for all } g \in A\},$$

$$\psi(B) \stackrel{\text{def}}{=} \{g \in G \mid gIm \ \text{ for all } m \in B\}.$$

The following properties hold for all $A_1, A_2 \subseteq G$, $B_1, B_2 \subseteq M$

1. $A_1 \subseteq A_2 \Rightarrow \varphi(A_2) \subseteq \varphi(A_1)$
2. $B_1 \subseteq B_2 \Rightarrow \psi(B_2) \subseteq \psi(B_1)$
3. $A_1 \subseteq \psi\varphi(A_1)$ and $B_1 \subseteq \varphi\psi(B_1)$

Mappings φ and ψ define **Galois connection** between $(2^G, \subseteq)$ and $(2^M, \subseteq)$, since $A \subseteq \psi(B) \Leftrightarrow B \subseteq \varphi(A)$.

Formal Concept Analysis. 3

[Ganter, Wille 1999]

In FCA, instead of two notations φ and ψ, a unified notation $(\cdot)'$ is used, so for arbitrary $A \subseteq G$, $B \subseteq M$

$$A' \stackrel{\text{def}}{=} \{m \in M \mid gIm \text{ for all } g \in A\},$$

$$B' \stackrel{\text{def}}{=} \{g \in G \mid gIm \text{ for all } m \in B\}.$$

(Formal) concept is a pair (A, B):

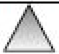
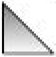
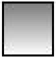
$$A \subseteq G,\ B \subseteq M,\ A' = B, \text{ and } B' = A.$$

A is called a **(formal) extent**, and B is called a **(formal) intent** of a concept (A, B).

Concepts are partially ordered by relation

$(A_1, B_1) \geq (A_2, B_2) \iff A_1 \supseteq A_2 \quad (B_2 \supseteq B_1)$.

Example. Context

	G \ M	a	b	c	d
1		×			×
2		×		×	
3			×	×	
4			×	×	×

Objects:

1 – equilateral triangle

2 – right triangle

3 – rectangle

4 – square

Attributes:

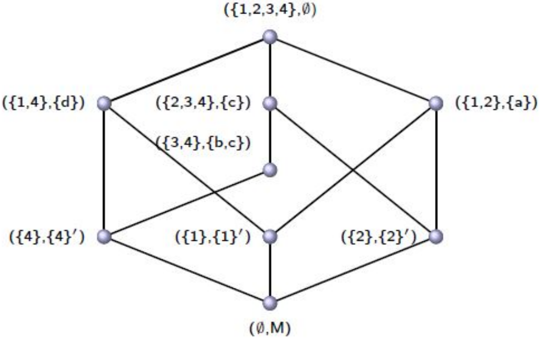
a – has 3 vertices

b – has 4 vertices

c – has a direct angle

d – equilateral

Example. Diagram of the ordered set of concepts



	G \ M	a	b	c	d
1		×			×
2		×		×	
3			×	×	
4			×	×	×

a – has 3 vertices

b – has 4 vertices

c – has a direct angle

d – equilateral

Properties of operation $(\cdot)'$

Let $\mathbb{K} = (G, M, I)$ be a formal context, $A, A_1, A_2 \subseteq G$ be subsets of objects, $B \subseteq M$ be subsets of attributes, then

1. If $A_1 \subseteq A_2$, then $A_2' \subseteq A_1'$;
2. If $A_1 \subseteq A_2$, then $A_1'' \subseteq A_2''$
3. $A \subseteq A''$
4. $A''' = A'$ (hence, $A'''' = A''$);
5. $(A_1 \cup A_2)' = A_1' \cap A_2'$;
6. $A \subseteq B' \Leftrightarrow B \subseteq A' \Leftrightarrow A \times B \subseteq I$.

Similar properties hold for subsets of attributes.

Closure operator on a set

A **closure operator** on set G is a mapping $\mathrm{cl}\colon \mathcal{P}(G) \to \mathcal{P}(G)$ with the following properties:

1. $\mathrm{cl}(\mathrm{cl}(X)) = \mathrm{cl}(X)$ (**idempotency**)
2. $X \subseteq \mathrm{cl}(X)$ (**extensity**)
3. $X \subseteq Y \Rightarrow \mathrm{cl}(X) \subseteq \mathrm{cl}(Y)$ (**monotonicity**)

For a closure operator cl the set $\mathrm{cl}(X)$ is called **closure** of X.
A subset $X \subseteq G$ is called **closed** if $\mathrm{cl}(X) = X$.

Example. Let (G, M, I) be a context, then operators $(\cdot)''\colon 2^G \to 2^G$, $(\cdot)''\colon 2^M \to 2^M$ are closure operators.

*Supremum- and infimum- dense subsets

A subset $X \subseteq L$ of lattice $(L, \leq)$ is called **supremum-dense** if any lattice element $v \in L$ can be represented as

$$v = \bigvee \{x \in X \mid x \leq v\}.$$

Dually for **infimum-dense** subsets.

Basic Theorem of Formal Concept Analysis

[Ganter, Wille 1999]

Concept lattice $\underline{\mathfrak{B}}(G, M, I)$ is a complete lattice. For arbitrary sets of formal concepts

$$\{(A_j, B_j) \mid j \in J\} \subseteq \underline{\mathfrak{B}}(G, M, I)$$

infimums and supremums are given in the following way:

$$\bigwedge_{j \in J}(A_j, B_j) = (\bigcap_{j \in J} A_j, (\bigcup_{j \in J} B_j)''),$$
$$\bigvee_{j \in J}(A_j, B_j) = ((\bigcup_{j \in J} A_j)'', \bigcap_{j \in J} B_j).$$

A complete lattice V is isomorphic to a lattice $\underline{\mathfrak{B}}(G, M, I)$ iff there are mappings $\gamma\colon G \to V$ and $\mu\colon M \to V$ such that $\gamma(G)$ is supremum-dense in V, $\mu(M)$ is infimum-dense in V, and $gIm \Leftrightarrow \gamma g \leq \mu m$ for all $g \in G$ and all $m \in M$.
In particular, $V \cong \underline{\mathfrak{B}}(V, V, \leq)$.

*Irreducible lattice elements

Let $(L, \leq)$ be a complete lattice. Element $x \in L$ is called **supremum (join)-irreducible** if

$$x \neq \bigvee \{y \mid y < x\}.$$

Dually, element $x \in L$ is calledd **infimum (meet)-irreducible** if

$$x \neq \bigwedge \{y \mid y > x\}.$$

The sets of join- and meet- irreducible elements of lattice $(L, \leq)$ are denoted by $J(L)$ and $M(L)$, respectively.
The set $J(L)$, as well as any its superset, is meet-dense.
The set $M(L)$, as well as any its superset, is join-dense.
The basic theorem of FCA implies the following isomorphism:

$$L \cong \underline{\mathfrak{B}}(J(L), M(L), \leq).$$

*Reducing attributes

Attribute $m \in M$, $\mathbb{K} = (G, M, I)$ is **reducible** if

$$m' = G \text{ or } m' = \bigcap \{n' \mid n \in M \;\&\; n' \supset m'\}.$$

If m is reducible, then
$\underline{\mathfrak{B}}(G, M, I) \cong \underline{\mathfrak{B}}(G, M \setminus \{m\}, I \cap (G \times (M \setminus \{m\})))$

Dually for objects. An irreducible attribute corresponds to a meet-irreducible lattice element, an irreducible object corresponds to a join-irreducible lattice element.

Example. Attribute m_k is reducible, since $m'_k = m'_i \cap m'_j$

$G \setminus M$	$\dots$	m_i	$\dots$	m_j	$\dots$	m_k	$\dots$
g_1	$\dots$	$\times$	$\dots$		$\dots$		$\dots$
g_2	$\dots$	$\times$	$\dots$	$\times$	$\dots$	$\times$	$\dots$
g_3	$\dots$	$\times$	$\dots$	$\times$	$\dots$	$\times$	$\dots$
g_4	$\dots$		$\dots$	$\times$	$\dots$		$\dots$

Duality principle for concept lattices

Let (G, M, I) be a formal context, then (M, G, I^{-1}) is also a formal context. One has

$$\underline{\mathfrak{B}}(M, G, I^{-1}) \cong \underline{\mathfrak{B}}(G, M, I)^d,$$

and the mapping $(B, A) \to (A, B)$

defines an antiisomorphism: interchanging objects and attributes we obtain a **dual concept lattice**.

Implications on subsets of attributes

Implication $A \to B$, where $A, B \subseteq M$ holds in context (G, M, I) if $A' \subseteq B'$, i.e., each object having all attributes from A also has all attributes from B.

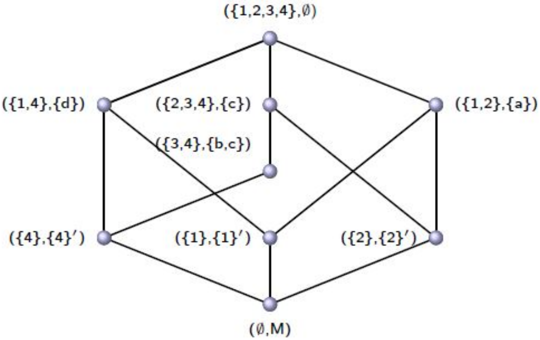
Implications satisfy **Armstrong rules**:

$$\frac{}{X \to X} \quad , \quad \frac{X \to Y}{X \cup Z \to Y} \quad , \quad \frac{X \to Y, Y \cup Z \to W}{X \cup Z \to W},$$

An **implication cover** is a subset of implications from which all other implications can be derived by means of Armstrong rules.
An **implication base** is a minimal (by inclusion) implication cover.

Concept lattice and implications

For $A \to B$ the meet of all attribute concepts for attributes from A lies below the meet of all attribute concepts of attributes in B.



	G \ M	a	b	c	d
1		×			×
2		×		×	
3			×	×	
4			×	×	×

Implications:

$abc \to d$

$b \to c$

$cd \to b$

Generator implication cover

A subset of attributes $D \subseteq M$ is a **generator** of a closed subset of attributes $B \subseteq M$, $B'' = B$ if $D \subseteq B$, $D'' = B = B''$.
A subset $D \subseteq M$ is a **minimal generator** if for any $E \subset D$ one has $E'' \neq D'' = B''$.
Generator $D \subseteq M$ is called **nontrivial** if $D \neq D'' = B''$.
Denote the set of all nontrivial minimal generators of B by $\text{nmingen}(B)$.

Generator implication cover looks as follows:
$\{F \to (F'' \setminus F) \mid F \subseteq M, F \in \text{ nmingen } (F'')\}$.

Direct (Proper premise) base

Definition

A subset $A \subseteq M$ is a **proper premise** if

$$A'' \neq A \cup \bigcup_{n \in A} (A \setminus \{n\})''.$$

The closure of a proper premise cannot be deduced from the union of closures of proper subsets of A.
Every proper premise is a minimal generator, but not vice versa.

Theorem

The set of implications

$$\mathcal{D} = \{A \to A'' \mid A \textit{ is a proper premise}\}$$

is complete, i.e., all implications of the context can be deduced from it by means of Armstrong rules.

Minimum implication basis

Duquenne-Guigues base is an implication base where each implication is a pseudo-intent.
A subset of attributes $P \subseteq M$ is called a **pseudo-intent** if $P \neq P''$ and for any pseudo-intent Q such that $Q \subset P$ one has $Q'' \subset P$.

Duquenne-Guigues base looks as follows:
$\{P \to (P'' \setminus P) \mid P$ - pseudo-intent $\}$.
Duquenne-Guigues base is a minimum (cardinality minimal) implication base.

Attribute exploration
Algorithm

- Start with any (possibly empty) set of objects.
- Generate an implication valid in the current subcontext.
- If the implication is not valid in the entire context, provide an object that violates it.
- Go to the next implication, etc.

Many-valued context

A **many-valued context** (relational databse) is a quadruple (G, M, W, J), where G is a set of objects, M is a set of **attributes**, W is a set of **attribute values**, and ternary **incidence relation** $J \subseteq G \times M \times W$ is such that $(g, m, w) \in J$, and $(g, m, v) \in J$ **implies** $w = v$.

Many-valued context

- Attribute m is **complete** if $\forall g \in G \exists w \in W$ such that $(g, m, w) \in J$.
- A many-valued context is **complete** if all its attributes are complete.

For complete many-valued contexts the value of attribute m for object g is denoted by $m(g)$, so $(g, m, m(g)) \in J$.

Functional dependency, an important tool of Data Science

For many-valued attributes X, Y **functional dependency** $X \to_{fd} Y$ is valid in complete many-valued context (G, M, W, J) if the following holds for every pair of objects $g, h \in G$:

$$(\forall m \in X \ \ m(g) = m(h)) \Rightarrow (\forall n \in Y \ \ n(g) = n(h)).$$

Example. Many-valued context (relational datatable) $K = (G, M, W, J)$, where functional dependency $b \to c$ is valid.

G \ M	a	b	c
1	a_1	b_1	c_1
2	a_2	b_1	c_1
3	a_1	b_2	c_2
4	a_2	b_3	c_2

$=$

G \ M	a	b
1	a_1	b_1
2	a_2	b_1
3	a_1	b_2
4	a_2	b_3

$\bowtie$

G \ M	b	c
1	b_1	c_1
2	b_2	c_2
3	b_3	c_2

Functional Dependencies and decomposability of data

Example. Many-valued context (relational datatable) $K = (G, M, W, J)$, where functional dependency $b \to c$ is not valid.

G \ M	a	b	c
1	a_1	b_1	c_1
2	a_2	b_1	c_2
3	a_1	b_2	c_2
4	a_2	b_3	c_2

$\neq$

G \ M	a	b
1	a_1	b_1
2	a_2	b_1
3	a_1	b_2
4	a_2	b_3

$\bowtie$

G \ M	b	c
1	b_1	c_1
2	b_2	c_2
3	b_3	c_2

Functional Dependencies as Implications

Proposition A. For a many-valued context $\mathbb{K} = (G, M, W, J)$, consider the context $\mathbb{K}_N := (\mathcal{P}_2(G), M, I)$, where $\mathcal{P}_2(G)$ is the set of all pairs of different objects from G and relation I is defined by

$$\{g, h\} Im :\Leftrightarrow m(g) = m(h).$$

Then $X \rightarrow_{fd} Y$ is valid in $\mathbb{K}$ for $X, Y \subseteq M$ iff the implication $X \rightarrow Y$ holds in the context $\mathbb{K}_N$.

Example. Consider the following many-valued context $\mathbb{K}_W$:

G \ M	a	b	c	d
1	α	α	β	β
2	α	γ	α	α
3	δ	α	δ	α
4	α	γ	α	δ

and the following formal context $\mathbb{K} = (G, M, I)$:

G \ M	a	b	c	d
$\{1, 2\}$	$\times$			
$\{1, 3\}$		$\times$		
$\{1, 4\}$	$\times$			
$\{2, 3\}$				$\times$
$\{2, 4\}$	$\times$	$\times$	$\times$	
$\{3, 4\}$				

Implications as functional dependencies

Proposition B. For a context $\mathbb{K} = (G, M, I)$ one can construct a many-valued context $\mathbb{K}_W = (G, M, W, J)$ such that an implication $X \to Y$ holds in $\mathbb{K}$ if and only if $X \to_{fd} Y$ is valid in $\mathbb{K}_W$.

Example. Consider the following formal context $\mathbb{K} = (G, M, I)$:

G \ M	a	b	c	d
1	×	×		
2	×		×	×
3		×		×
4	×		×	

and the following many-valued context $\mathbb{K}_W$:

G \ M	a	b	c	d
0	0	0	0	0
1	0	0	1	1
2	0	2	0	0
3	3	0	3	0
4	0	4	0	4

Literature

B. Ganter and R. Wille, *Formal Concept Analysis: Mathematical Foundations*, Springer, 1999.

B.A. Davey and H.A. Priestly. *Introduction to Lattices and Order (2nd edition)*. Cambridge Mathematical Textbooks. Cambridge University Press, 2002.

D. Maier, *The Theory of Relational Databases*. Computer Science Press, 1983.

C.J. Date, *Database Design and Relational Theory: Normal Forms and All That Jazz*. Apress, 2019.