Analogical Proportions, Multivalued Dependencies and Explanations

Sebastian Link[1] Henri Prade[2] Gilles Richard[2]

1- School of Computer Science, The University of Auckland, Auckland, New Zealand

2. IRIT - CNRS & Université Paul Sabatier, Toulouse, France

Two unrelated fields

- It is always striking to see that *the same concept* has been introduced *independently* and for *different purposes* in two *unrelated* fields
- analogical reasoning
  and analogical proportion
- database design
  and (weak) *multivalued dependencies*

Analogical proportions - 1

- " a is to b as c is to d"
  a *differs* from b as c *differs* from d
  and b *differs* from a as d *differs* from c".
- $a : b :: c : d \triangleq$
  $((a \wedge \neg b) \equiv (c \wedge \neg d)) \wedge ((\neg a \wedge b) \equiv (\neg c \wedge d))$
  it uses dissimilarity indicators only
- $a : b :: c : d$ satisfies the key properties of an analogical proportion, namely
  - reflexivity: $a : b : a : b$
  - symmetry: $a : b :: c : d \Rightarrow c : d :: a : b$
  - central permutation: $a : b :: c : d \Rightarrow a : c :: b : d$
  - also satisfies $a : a :: b : b$
    and *external permutation* $a : b :: c : d \Rightarrow d : b :: c : a$

Nominal values

- a, b, c, d belong to a *finite* attribute domain $\mathcal{A}$

Then, $a : b :: c : d$ holds true only for the 3 patterns

$(a, b, c, d) \in \{(g, g, g, g), (g, h, g, h), (g, g, h, h)\}$

This *generalizes the Boolean case* $\mathcal{A} = \{0, 1\}$.

- *Items* are represented by

tuples of n attribute values:

$a = (a_1, \cdots, a_n)$ a_i is the value of attribute i

$a : b :: c : d$ true iff $\forall i \in \{1, ..., n\}, a_i : b_i :: c_i : d_i$ true

- *graded* version of analogical proportions in case of *numerical attributes*

Example and analogical inference

	course	teacher	time
a	Maths	Peter	8 am
b	Maths	Peter	2 pm
c	Maths	Mary	8 am
d	Maths	Mary	2 pm

Assuming that the AP $a : b :: c : d$ is true,
one can *calculate* d from a, b, c (except if $a \neq b = c$)

$$\frac{\forall i \in \{1, ..., n\}, \quad a_i : b_i :: c_i : d_i \text{ holds}}{a_{n+1} : b_{n+1} :: c_{n+1} : d_{n+1} \text{ holds}}$$

If a_{n+1}, b_{n+1}, c_{n+1} are known, this enables the *prediction of* d_{n+1}, provided that $a_{n+1} : b_{n+1} :: c_{n+1} : x$ is *solvable*

Functional dependencies

- relation r - a finite set of tuples
  over a set of attributes R
  - $t[X]$: restriction of a tuple t to attributes in $X \subseteq R$
  $t[XY]$ is short for $t[X \cup Y]$.
- A functional dependency $X \rightarrow Y$ $X, Y \subseteq R$
  for any pair of tuples t_1 and t_2
  obeying the relational schema R
  if $t_1[X] = t_2[X]$ then $t_1[Y] = t_2[Y]$
  which reads "X determines Y"
- "The value of X *explains* the value of Y"

Multi-valued dependencies (R. Fagin, 1977)

- $X \twoheadrightarrow Y$ "X multi-determines Y" holds on R if, for *all pairs of tuples* t_1 *and* t_2 in r such that $t_1[X] = t_2[X]$, there exists a tuple t_3 in r such that $t_3[XY] = t_1[XY]$ and $t_3[X(R \setminus Y)] = t_2[X(R \setminus Y)]$ there also exists a tuple t_4 in r such as $t_4[XY] = t_2[XY]$ and $t_4[X(R \setminus Y)] = t_1[X(R \setminus Y)]$
- (x, y, z) the tuple having values x, y, z for subsets X, Y, $R \setminus (X \cup Y)$ respectively, then whenever the tuples (p, q, r) and (p, s, u) exist in r, the tuples (p, q, u) and (p, s, r) should also exist in r.

Multi-valued dependencies (R. Fagin, 1977)

- $X \twoheadrightarrow Y$ "X multi-determines Y"
  holds on R if, for *all pairs of tuples* t_1 *and* t_2 in r
  such that $t_1[X] = t_2[X]$, there exists a tuple t_3 in r
  such that $t_3[XY] = t_1[XY]$ and
  $t_3[X(R \setminus Y)] = t_2[X(R \setminus Y)]$
  there also exists a tuple t_4 in r such as
  $t_4[XY] = t_2[XY]$ and $t_4[X(R \setminus Y)] = t_1[X(R \setminus Y)]$
- (x, y, z) the tuple having values x, y, z for subsets X, Y, $R \setminus (X \cup Y)$ respectively, then whenever the tuples (p, q, r) and (p, s, u) exist in r, the tuples (p, q, u) and (p, s, r) should also exist in r.

Analogical proportion!

	X	Y	$Z = R \setminus (X \cup Y)$
t_1	p	q	r
t_4	p	s	r
t_3	p	q	u
t_2	p	s	u

the set of values of Y is *logically independent* of set Z and vice versa (in the context $X = p$).

Example

course	*teacher*	*time*
Maths	Peter	8 am
Maths	Peter	2 pm
Maths	Mary	8 am
Maths	Mary	2 pm
Maths	Paul	8 am
Maths	Paul	2 pm
Comp. Sci.	Peter	8 am
Comp. Sci.	Mary	8 am

  Table: Multivalued dependencies: {course} $\twoheadrightarrow$ {teacher}; {course} $\twoheadrightarrow$ {time}

course	*teacher*	*time*
Maths	{Peter, Mary, Paul }	{8 am, 2 pm }
Comp. Sci.	{Peter, Mary}	{8 am}

teachers & time are *logically independent* of each other

Weak multi-valued dependencies

- A ***weak*** *multivalued dependency* (Fischer,1984) $X \twoheadrightarrow_w Y$ holds on R if, for *all tuples* t_1, t_2, t_3 in r such that $t_1[XY] = t_2[XY]$ and $t_1[X(R \setminus Y)] = t_3[X(R \setminus Y)]$ there is some tuple t_4 in r such that $t_4[XY] = t_3[XY]$ and $t_4[X(R \setminus Y)] = t_2[X(R \setminus Y)]$
- if $X \twoheadrightarrow Y$ then $X \twoheadrightarrow_w Y$
- $t_1 : t_2 :: t_3 : t_4$
  t_4 *can be computed* from t_1, t_2, t_3

Is Result for d fair (and if no, why)?

	X (shared values)	Y (diploma)	Z (sex)	Result
a	s	yes	F	P
b	s	yes	M	P
c	s	no	F	N
d	s	no	M	P

- The answer to the question “ why *Result*(d) is not N?” is to be found in the value of the *sex* attribute for d
- Estimating fairness is a matter of *conditional stochastic independence*
- Multi-valued dependencies and thus analogical proportions exhibit logical independence relations
- The *violation* of an analogical proportion (and thus of a multivalued dependency) suggests that *Result*(d) is unfair

Concluding remarks

- Further developments (besides fairness)?
- The axiomatic characterization of dependencies may bring some new light on analogical proportions
- Impact on explanation capabilities ?
- Impact on analogical querying?
- Can we handle uncertain data with analogical proportions, as in database design?