Machine Learning with Analogical Proportions

A novel approach

Laurent Miclet[1], Henri Prade[2,3] and Gilles Richard[2]

1. IRISA, ENSSAT, Lannion and Univ. of Rennes, Rennes, France
2. IRIT, Univ. Paul Sabatier, Toulouse, France
3. QCIS, Univ. of Technology, Sydney, Australia

Tutorial ECML / PKDD, Nancy, Sept. 19, 2014

To start : A "zoological" example of analogical proportion

Calf

Cow

Foal

To start : A "zoological" example of analogical proportion

Calf

Cow

Foal

Mare

Calf is to Cow as Foal is to Mare

4 items : similarities and dissimilarities

The missing one can be predicted !

To start : A mathematical example

30 is to 66 as 70 is to 154

$$30 \ : \ 66 \ :: \ 70 \ : \ 154$$

$$\frac{\mathbf{30}}{\mathbf{66}} = \frac{\mathbf{70}}{\mathbf{154}}$$

30	66
70	154

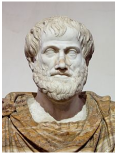
$$\mathbf{30} \times \mathbf{154} = \mathbf{66} \times \mathbf{70} = 4620$$

The missing one can still be predicted : Rule of three !

Contents

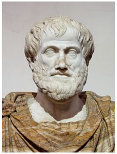
Historical introduction

- Western world : Aristotle (384-322 BC)



Historical introduction

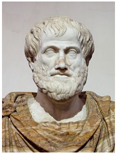
- Western world : Aristotle (384-322 BC)



- Eastern world : Mencius (A follower of Confucius : 372-289 BC)

Historical introduction

- Western world : Aristotle (384-322 BC)



- Eastern world : Mencius (A follower of Confucius : 372-289 BC)

- Idea of *analogical proportion*
- Use as a rhetorical argument
- Metaphor : "Messi is the Mozart of soccer"

Analogy

A matter of ...

- Things having *common* properties "with some *differences*"
- Similarity between relations (Diderot and D'Alembert Encyclopédie 1751-1772)
- No formal definition ... not yet !

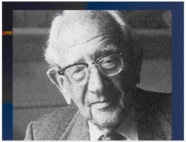
Analogy

A matter of ...

- Things having *common* properties "with some *differences*"
- Similarity between relations (Diderot and D'Alembert Encyclopédie 1751-1772)
- No formal definition ... not yet !

Analogy in science

- Analogy : a tool to transfer meaning from a source to a target (linguistics, psychology)
- Link with creativity (think out of the box !) (especially in science)
- Link with problem solving (see "How to solve it" : Polya - 1945)



First models

Psychology/Cognitive sciences

- K. Holyoak and P. Thagard → ACME
- D. Gentner → SMT

More AI-oriented

- Th. Evans → ANALOGY
- D. Hofstadter and M. Mitchell → COPYCAT
- T. Davies and S. Russell → determination rule
- Case-Based Reasoning (R. Schank, then J. Kolodner)

ACME

K. Holyoak and P. Thagard

- Date : 1989
- Place : UCLA and Princeton
- Result : Analogical Constraint Mapping Engine (ACME)

Main ideas

- Structural similarity : share the same relational structure
- Semantic similarity : identity of symbols, or predicates sharing
- Analogical mapping by constraint satisfaction
- Connectionist approach : mapping as the result of a constraint-satisfaction spreading activation network

SMT : Structure Mapping Theory

D. Gentner

- Date : 1983
- Place : Northwestern University, Department of Psychology
- Result : Structure Mapping Theory and Structure Mapping Engine

Main ideas

- Source and target : relation - properties - functions (structural view)
- Mapping between 2 structures
- Set of candidate inference rules about the target + evaluation scores

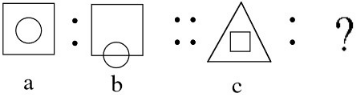
Evans' *ANALOGY* program

Th. G. Evans

- Date : 1964
- Place : MIT
- Result : ANALOGY written in LISP

Main ideas

- Pb : "fig. A is to fig. B as fig. C is to fig. X ?"
  X belonging to a **given** *set of candidate figures*
- Recognition and transformation of geometric figures



- Primitive input : description of the figures
- Find an appropriate transformation rule

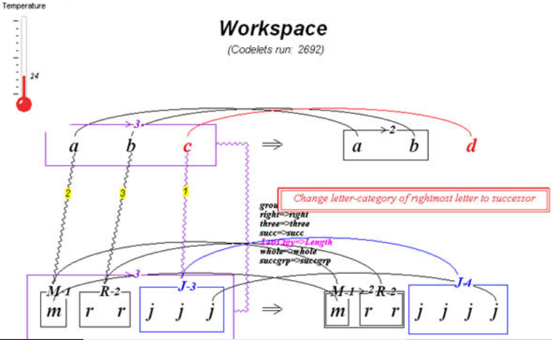
The *Copycat* project

D. Hofstadter and M. Mitchell

- Date : 1988
- Place : Center for Research on Concepts and Cognition, Indiana Univ.
- Result : Copycat written in LISP

Main ideas

- Pb : "*abc* : *abd* : : *ijk* : ?" (assumed to be representative problems)
- Many independent processes ("codelets") running in parallel



A logical approach

T. Davies and S. Russell

- Date : 1987
- Place : Berkeley and Stanford
- Result : determination rule

Main ideas

- First order logic
- Analogical jump $\dfrac{P(s) \quad P(t) \quad Q(s)}{Q(t)}$
- Find a *side condition*
  allowing a deductively safe "jump" !
- Too rigid framework

Case-Based Reasoning

R. Schank, then J. Kolodner, M. Lebowitz

- Date : 1977
- Place : Columbia Univ.
- Result : A subfield in AI !

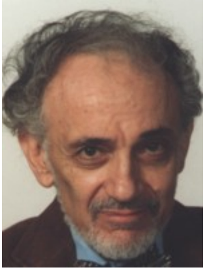
Main ideas

- Solving new problems based on the solutions of similar known problems
- A repository R of cases (problem, solution)
- Four steps process : retrieve (similarity) - reuse (mapping of solution) - revise (adaptation) - retain
- problem 1 : solution 1 : : new problem : solution X
- New problem compared to each problem in R, *one by one*

See also *Erica Melis*, *Manuela Veloso* "Analogy in Problem Solving" - 1998

An outsider !

Sheldon Klein (1935 - 2005) - pages.cs.wisc.edu/ sklein/sklein.html



- B.A. (anthropology - 1956) Ph.D. (linguistics - 1963)
  Prof. of CS and Linguistics University of Wisconsin
- Culture, mysticism & social structure and the calculation of behavior. Proc. Europ. Conf. in AI (ECAI'82), Orsay, 141-146, 1982
- A procedure for computing X such as $A : B :: C : X$, once A, B, C are encoded in a binary way feature by feature : $X = C \equiv (A \equiv B)$

A meeting that you probably missed...

Aristotelian Society meeting

- 21 Bedford Square - London - WC1
- Speaker : Mary Brenda Hesse (b. in 1924) Philosophy of Sciences
- Title of the talk : "On defining analogy"

A meeting that you probably missed...

Aristotelian Society meeting

- 21 Bedford Square - London - WC1
- Speaker : Mary Brenda Hesse (b. in 1924) Philosophy of Sciences
- Title of the talk : "On defining analogy"
- **14th December 1959 - 7h30 pm**

A meeting that you probably missed...

Aristotelian Society meeting

- 21 Bedford Square - London - WC1
- Speaker : Mary Brenda Hesse (b. in 1924) Philosophy of Sciences
- Title of the talk : "On defining analogy"
- **14th December 1959 - 7h30 pm**

Aims : to overcome 2 standard objections about analogy

- "*A trivial relation holding between certain pairs of terms*"
(relations have more than one example, in general !)
- "*The relation exists between any 2 pairs of terms*"

- investigates a four terms analogy relation
  having the "desirable properties of analogy"
- outlines a "find the missing guy" algorithm

Contents

Axioms of Analogical Proportion (**AP**)

Definition (Analogical Proportion)

An analogical proportion on a set $\mathbb{E}$ *is a relation on* $\mathbb{E}^4$ *such as, for all 4-tuples* A, B, C *and* D *in relation in this order (denoted* $A : B :: C : D$*) :*

1. $A : B :: A : B$
2. $A : B :: C : D \Leftrightarrow C : D :: A : B$
3. $A : B :: C : D \Leftrightarrow A : C :: B : D$

Five other proportions are equivalent :

$B : A :: D : C$ $\quad$ $D : B :: C : A$ $\quad$ $D : C :: B : A$

$B : D :: A : C$ $\quad$ $C : A :: D : B$

But these are not : $A : B :: D : C$ and $A : C :: D : B$

Just as mathematical proportions

Relation in $\mathbb{N}$: $A : B :: C : D$ iff $A \times D = B \times C$

$$30 \quad : \quad 66 \quad :: \quad 70 \quad : \quad 154$$

since $30 \times 154 = 66 \times 70 = 4620$

$$\frac{30}{66} = \frac{70}{154}$$

30	66
70	154

AP between *subsets*

Four subsets A, B, C and D are in AP ($A : B :: C : D$) when the *differences* between A and B are the same as between C and D.

$$A \backslash B = C \backslash D \quad \text{and} \quad B \backslash A = D \backslash C$$
$$\Leftrightarrow$$
$$A \cup D = B \cup C \quad \text{and} \quad A \cap D = B \cap C \; \textbf{!}$$

$A = \{a, b, c, h\}$, $B = \{a, b, d, e, h\}$, $C = \{f, c, h\}$ and $D = \{f, d, e, h\}$

$$A \backslash B = C \backslash D = \{c\} \quad \text{and} \quad B \backslash A = D \backslash C = \{d, e\}$$

	a	b	c	d	e	f	h
A	$\times$	$\times$	$\times$				$\times$
B	$\times$	$\times$		$\times$	$\times$		$\times$
C			$\times$			$\times$	$\times$
D				$\times$	$\times$	$\times$	$\times$

$$A \cup D = B \cup C = \{a, b, c, d, e, f, h\} \quad \text{and} \quad A \cap D = B \cap C = \{h\}$$

AP in Propositional Logic

$$a : b :: c : d$$

$$a : b :: c : d \quad \Leftrightarrow \quad (a \wedge \overline{b} \equiv c \wedge \overline{d}) \wedge (\overline{a} \wedge b \equiv \overline{c} \wedge d)$$

$$a : b :: c : d \quad \Leftrightarrow \quad (a \vee d \equiv b \vee c) \wedge (a \wedge d \equiv b \wedge c)$$

6 patterns make $a : b :: c : d$ true :

0 : 0 :: 0 : 0	0 : 0 :: 1 : 1	0 : 1 :: 0 : 1
1 : 1 :: 1 : 1	1 : 1 :: 0 : 0	1 : 0 :: 1 : 0

BUT

1 : 0 :: 0 : 1 and 0 : 1 :: 1 : 0 are not AP's

Generally speaking, $\{u\} : \{v\} :: \{v\} : \{u\}$ is **false** for $u \neq v$

AP for Boolean vectors

$$\vec{a} : \vec{b} :: \vec{c} : \vec{d} \text{ iff } \forall i \in [1, n], a_i : b_i :: c_i : d_i$$

a calf *is to* **a cow** *as* **a foal** *is to* **a mare**

	mammal	young	equine	adult female	bovine	adult male
A : calf	1	1	0	0	1	0
B : cow	1	0	0	1	1	0
C : foal	1	1	1	0	0	0
D : mare	1	0	1	1	0	0

The columns are all binary analogical proportions.

$$A \setminus B = \{ \text{ bovine } \} = C \setminus D$$

$$B \setminus A = \{ \text{ equine } \} = D \setminus C$$

Properties of Boolean analogical proportions

- reflexivity : $a : b :: a : b$
- symmetry : $a : b :: c : d \equiv c : d :: a : b$
- central and external permutations :
  $a : b :: c : d \equiv a : c :: b : d$
  $a : b :: c : d \equiv d : b :: c : a$
- code independency : $a : b :: c : d \equiv \neg a : \neg b :: \neg c : \neg d$
- transitivity :
  $(a : b :: c : d) \wedge (c : d :: e : f) \Rightarrow a : b :: e : f$

Properties still valid for Boolean vectors

What an Analogical Proportion means

The differences between A and B
are the same as
the differences between C and D.

The differences between A and C
are the same as
the differences between B and D.

Two possible formalizations for some operation $\oplus$

$$
\begin{array}{c||ccc|}
 & B & & C \\
\hline
A & x_1 & \oplus & x_2 \\
 & \oplus & & \oplus \\
D & t_2 & \oplus & t_1 \\
\hline
\end{array}
$$

$$A \oplus D = B \oplus C$$

A stronger definition : *factorization* (Stroppa et Yvon)

Let $(S, \oplus)$ be an Abelian semigroup : $\oplus$ associative and commutative

In an Abelian semigroup :

$a : b \; :: c \; : d \Leftrightarrow$

- Either $(b, c) \in \{(a, d), (d, a)\}$
- Or $\exists (x_1, x_2, t_1, t_2) : a = x_1 \oplus x_2, b = x_1 \oplus t_2, c = t_1 \oplus x_2, d = t_1 \oplus t_2$

Can be displayed as :

	b	c
a	x_1	x_2
d	t_2	t_1

For example, in $(\mathbb{N}, \times)$:

	66	70
30	6	5
154	11	14

Example : Multiplicative AP in $\mathbb{N}$

Transforming 30 into 66 is the same as transforming 70 into 154.

$$30 \; : \; 66 \; :: \; 70 \; : \; 154$$

$$30 \times 154 = 66 \times 70 = 4620$$

$$\frac{30}{66} = \frac{70}{154}$$

30	66
70	154

$$\begin{array}{rcccccccccccc} 30 & = & 1 & \times & 2 & \times & 3 & \times & 5 & \times & 1 & \times & 1 \\ 66 & = & 2 & \times & 1 & \times & 3 & \times & 1 & \times & 1 & \times & 11 \\ 70 & = & 1 & \times & 2 & \times & 1 & \times & 5 & \times & 7 & \times & 1 \\ 154 & = & 2 & \times & 1 & \times & 1 & \times & 1 & \times & 7 & \times & 11 \end{array}$$

$$\mathrm{lcm}(30, 154) = \mathrm{lcm}(66, 70) = 2310$$
$$\gcd(30, 154) = \gcd(66, 70) = 2$$

Another definition for Abelian semigroups : composing extremes and means

Weak Analogical Proportion

Let $(S, \oplus)$ be an Abelian semigroup.
(a, b, c and d) are by definition in *Weak Analogical Proportion (WAP)* when

$$a \oplus d = b \oplus c$$

NB This is reminiscent of *Piaget*'s definition of a *logical proportion* (*without* reference to *analogy*!) in a Boolean setting as $(a \wedge d \equiv b \wedge c) \wedge (a \vee d \equiv b \vee c)$!!

We denote :

$$a \ : \ b \overset{wap}{::} c \ : \ d \quad \text{or} \quad (a, d) \boxtimes (b, c)$$

$\boxtimes$ is an equivalence relation.

Analogical equation for Boolean variables

- Solving $a : b :: c : x$ just as the *rule of three*
- There are six analogical proportions between Boolean variables :

0	0	0	0	1	1	1	1
0	0	1	1	1	1	0	0
0	1	0	1	1	0	1	0

- This can be written as :

$$a : b :: c : d \quad \Leftrightarrow \quad (a \wedge \overline{b} \equiv c \wedge \overline{d}) \wedge (\overline{a} \wedge b \equiv \overline{c} \wedge d)$$

- Equation $a : b :: c : x$ may have *no solution*, e.g. $1 : 0 :: 0 : x$
  Equation $a : b :: c : x$ has a solution iff

$$(a \equiv b) \vee (c \equiv d) \text{ holds true}$$

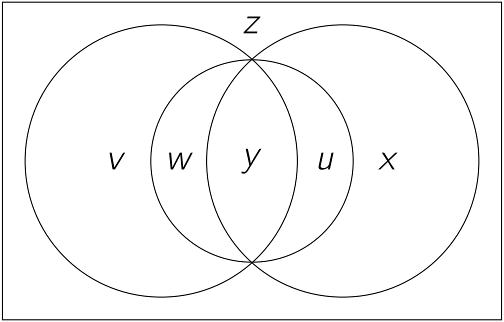
The solution is then unique and has the value $a \equiv (b \equiv c)$

Analogical equation for lattice of subsets

Proposition (Y. Lepage)

In the Boolean lattice $(\wp(\Sigma), \cup, \cap, \Sigma \subseteq)$, a 4-tuple (A, B, C, D) is in analogical proportion $(A : B :: C : D)$ iff there exists 6 subsets (u, v, w, x, y, z) partitioning $\wp(\Sigma)$ such that

$$A = u \cup w \cup y,\ B = v \cup w \cup y,\ C = u \cup x \cup y,\ D = v \cup x \cup y$$



The analogical equation in D :

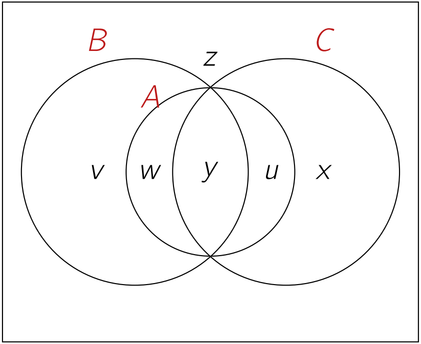
$$(A : B :: C : D)$$

has a solution iff

$$B \cap C \subseteq A \subseteq B \cup C$$

The solution is then unique and has the value

$$D = ((B \cup C) \backslash A) \cup (B \cap C)$$



$$D = v \cup x \cup y$$

Multiple-valued extensions

Aim : handling real life data

- Properties whose satisfaction is a matter of level
  e.g. $0.9 : 0 :: 1 : 0$
  (for dealing with *numerical* variables)
- which may not apply
- for which information is missing

Multiple-valued extensions

Aim : handling real life data

- Properties whose satisfaction is a matter of level
  e.g. $0.9 : 0 :: 1 : 0$
  (for dealing with *numerical* variables)
- which may not apply
- for which information is missing

2 questions (at least)

- what are the valuations that correspond to a "perfect" analogy ?
- are there valuations that could be regarded as "approximate" analogies (i.e. with a truth value distinct from 0 and 1) ?

NOTATION : $A(a, b, c, d)$ is the truth value of $a : b :: c : d$

A first definition A

Linearly ordered scale $\mathcal{L} \subseteq [0,1]$ (finite or not)

- $A(a,b,c,d) = (a \wedge \neg b \equiv c \wedge \neg d) \wedge (\neg a \wedge b \equiv \neg c \wedge d)$
- central $\wedge$ equal to min ;
  $s \equiv t = \min(s \rightarrow_{Luka} t, t \rightarrow_{Luka} s) = 1 - |s - t|$;
  $s \wedge \neg t = \max(0, s - t) = 1 - (s \rightarrow_{Luka} t)$: bounded difference
- $A(a,b,c,d) = 1 - \mid (a-b) - (c-d) \mid$ if $a \geq b$ and $c \geq d$,
  or $a \leq b$ and $c \leq d$
  $A(a,b,c,d) = 1 - max(|a-b|, |c-d|)$ if $a \leq b$ and $c \geq d$,
  or $a \geq b$ and $c \leq d$
- fully true for 19 patterns in the 3-valued case $\{0, 1/2, 1\}$.
  For instance, $(\alpha,\alpha,\alpha,\alpha), (1,\alpha,1,\alpha)$, etc. with $\alpha = 1/2$
  INCLUDING $(1,\alpha,\alpha,0)$; $(0,\alpha,\alpha,1)$; $(\alpha,1,0,\alpha)$; $(\alpha,0,1,\alpha)$

Another candidate definition A^*

- In the Boolean case,
  there is another equivalent expression of the analogical proportion :

$$a : b :: c : d = (a \wedge d \equiv b \wedge c) \ \wedge \ (a \vee d \equiv b \vee c)$$

- Taking $\wedge = \min$, $\vee = \max$, and $s \equiv t = 1 - |s - t|$
  $A^*(a, b, c, d) =$
  $\min(1 - |\max(a, d) - \max(b, c)|, 1 - |\min(a, d) - \min(b, c)|)$
- $A^*(a, b, c, d) = 1$
  only for the 15 patterns with at most two distinct values
  for which $A(a, b, c, d) = 1$,
  while $A^*(a, b, c, d) = \alpha$ for the 4 other patterns for which
  $A(a, b, c, d) = 1$,
  namely for $(1, \alpha, \alpha, 0)$; $(0, \alpha, \alpha, 1)$; $(\alpha, 1, 0, \alpha)$; $(\alpha, 0, 1, \alpha)$

Comparing A and A^*

- A^* is smoother than A
  in the sense that more patterns have intermediary truth values with A^* than with A
- $A(1,1,u,v) = 1 - |u - v| = A^*(1,1,u,v) = A(0,0,u,v) = A^*(0,0,u,v)$
- Both A and A^* still satisfy the *symmetry property*
- Both A and A^* still satisfy the *code independency* property
  with respect to $\overline{a} = 1 - a$.
- Only A^* still enjoys
  the *means permutation* properties and the *extremes permutation* properties.
- Only A **does not depend on the departure point of the change**

Contents

Analogy-based classification

Main idea to be implemented

- Rule of three is a *prediction* tool
- Predict *classes* instead of numbers

Considering a, b, c, d as logical formulas

- Take inspiration from the numerical case
- $\frac{a}{b} = \frac{c}{d}$ means $a \times d = b \times c$

Analogy-based classification

Main idea to be implemented

- Rule of three is a *prediction* tool
- Predict *classes* instead of numbers

Considering a, b, c, d as logical formulas

- Take inspiration from the numerical case
- $\frac{a}{b} = \frac{c}{d}$ means $a \times d = b \times c$

Leads to a general logical definition

- $a : b :: c : d$ iff $(a \wedge d \equiv b \wedge c) \wedge (a \vee d \equiv b \vee c)$
- Keeping basic properties (permutations, symmetry, etc.)
- Moving to nominal attributes : 3 valid patterns :

  $s : s :: s : s \quad s : t :: s : t \quad s : s : t : t$

Boolean view : a recap !

When using propositional variables

- Truth table (16 lines) - 6 lines leading to 1
- Equivalent to $(a \wedge \neg b \equiv c \wedge \neg d) \wedge (b \wedge \neg a \equiv d \wedge \neg c)$
- Equation solving process : find x such that $a : b :: c : x$ holds

Boolean view : a recap !

When using propositional variables

- Truth table (16 lines) - 6 lines leading to 1
- Equivalent to $(a \wedge \neg b \equiv c \wedge \neg d) \wedge (b \wedge \neg a \equiv d \wedge \neg c)$
- Equation solving process : find x such that $a : b :: c : x$ holds

Extension to Boolean vectors

- Easy game : $\vec{a} : \vec{b} :: \vec{c} : \vec{d}$ iff $\forall i \in [1, n], a_i : b_i :: c_i : d_i$
- Standard properties still valid (permutations, symmetry, etc.)
- Equation solving process $\vec{a} : \vec{b} :: \vec{c} : \vec{x}$: componentwise
- BUT NOW $\vec{a} : \vec{b} :: \vec{c} : \vec{d}$ may hold with 4 *distinct* vectors

a	1	1	0	1
b	1	0	1	1
c	1	1	0	0
d	1	0	1	0

The analogical credo (*learning bias*)

Main inference principle

- *Continuity* principle for classification $\frac{\vec{a}:\vec{b}::\vec{c}:\vec{d}}{cl(\vec{a}):cl(\vec{b})::cl(\vec{c}):cl(\vec{d})}$
- Generalized *continuity* principle

$$\frac{\forall i \in [1,n] \setminus J, a_i : b_i :: c_i : d_i}{\forall j \in J, a_j : b_j :: c_j : d_j}$$

- Useful for dealing with/predicting missing values

The analogical credo (*learning bias*)

Main inference principle

- *Continuity* principle for classification $\frac{\vec{a}:\vec{b}::\vec{c}:\vec{d}}{cl(\vec{a}):cl(\vec{b})::cl(\vec{c}):cl(\vec{d})}$
- Generalized *continuity* principle

$$\frac{\forall i \in [1,n] \setminus J, a_i : b_i :: c_i : d_i}{\forall j \in J, a_j : b_j :: c_j : d_j}$$

- Useful for dealing with/predicting missing values

General (naive !) algorithm for classification

1. $\vec{d}$ new item to be classified
2. Find $\vec{a}, \vec{b}, \vec{c}$ with solvable class equation such that : $\vec{a} : \vec{b} :: \vec{c} : \vec{d}$
3. Solve the class equation $cl(\vec{a}) : cl(\vec{b}) :: cl(\vec{c}) : x$
4. Allocate to $\vec{d}$ the class solution (majority vote if necessary)

Simple relaxing ideas

Aim : Relax the inductive principle

- Analogical proportion may not hold everywhere
- Diverse options (with majority votes)
  1. Consider only the best fit as voters i.e the triple(s) $(\vec{a}, \vec{b}, \vec{c})$ with solvable class equation such that
  $$card\{i \in [1, n] \mid a_i : b_i :: c_i : d_i = 1\} \text{ is maximum}$$
  2. Give a threshold p regarding this cardinal and consider all the triples $(\vec{a}, \vec{b}, \vec{c})$ with solvable class equation such that
  $$card\{i \in [1, n] \mid a_i : b_i :: c_i : d_i = 1\} \geq p$$
  3. Fix a number of voters v and consider the v best triples $(\vec{a}, \vec{b}, \vec{c})$ with solvable class equation
  4. Consider the nearest neighbor $\vec{c}$ of $\vec{d}$ and search as voters only the triple(s) $(\vec{a}, \vec{b}, \vec{c})$ with solvable class equation including $\vec{c}$

Current results

Some experiments (option 4 above, Bounhas *et al.*, 2014)

	non binarized			binarized		
	r=1	r=2	r=3	r=1	r=2	r=3
Balance	86 ± 4	**88** ± 2	72 ± 5	84 ± 4	87 ± 3	74 ± 5
Car	**95** ± 3	89 ± 3	72 ± 6	95 ± 3	94 ± 5	77 ± 6
TicTacToe	**98** ± 5	96 ± 5	98 ± 5	98 ± 5	97 ± 5	98 ± 5
Monk1	**99** ± 1	99 ± 1	90 ± 4	99 ± 1	99 ± 1	99 ± 1
Monk2	**99** ± 1	97 ± 3	91 ± 5	60 ± 7	99 ± 1	94 ± 5
Monk3	**99** ± 1	97 ± 2	91 ± 5	99 ± 1	99 ± 1	98 ± 2

r : number of features on which $\vec{c}$ and $\vec{d}$ differ

Current results

Some experiments (option 4 above, Bounhas *et al.*, 2014)

	non binarized			binarized		
	r=1	r=2	r=3	r=1	r=2	r=3
Balance	86 $\pm$ 4	**88** $\pm$ 2	72 $\pm$ 5	84 $\pm$ 4	87 $\pm$ 3	74 $\pm$ 5
Car	**95** $\pm$ 3	89 $\pm$ 3	72 $\pm$ 6	95 $\pm$ 3	94 $\pm$ 5	77 $\pm$ 6
TicTacToe	**98** $\pm$ 5	96 $\pm$ 5	98 $\pm$ 5	98 $\pm$ 5	97 $\pm$ 5	98 $\pm$ 5
Monk1	**99** $\pm$ 1	99 $\pm$ 1	90 $\pm$ 4	99 $\pm$ 1	99 $\pm$ 1	99 $\pm$ 1
Monk2	**99** $\pm$ 1	97 $\pm$ 3	91 $\pm$ 5	60 $\pm$ 7	99 $\pm$ 1	94 $\pm$ 5
Monk3	**99** $\pm$ 1	97 $\pm$ 2	91 $\pm$ 5	99 $\pm$ 1	99 $\pm$ 1	98 $\pm$ 2

r : number of features on which $\vec{c}$ and $\vec{d}$ differ

And the competitors

Datasets	**SVM**	**IBk(k=1, k=10)**	**JRip**	**C4.5**	**WAPC**
Balance	90	84, 84	72	64	86
Car	92	92, 92	88	90	n/a
Tic tac toe	98	99, 99	98	85	n/a
Monk1	75	99,96	94	96	98
Monk2	67	60, 63	66	67	100
Monk3	100	99, 98	99	100	96

So what ?

First comments

- More or less same performances in terms of accuracy whatever the technique
- At least as performant than standard methods
- No optimization process
- Significantly better than *k-NN*

So what ?

First comments

- More or less same performances in terms of accuracy whatever the technique
- At least as performant than standard methods
- No optimization process
- Significantly better than *k-NN*

Some questions

- Is analogy a “natural” regularity ? DO NOT KNOW
- Do we have datasets “more suitable” for analogical learning ? DO NOT KNOW
- Can we imagine other kind of “proportional regularities” ? YES

Relax... with Analogical Dissimilarity

Usual aim : Relax the inductive principle !

- Analogical proportion may not hold everywhere
- Distance to a perfect analogical proportion for Boolean values
- AD(a,b,c,d) = minimal number of bits that have to be switched to get a perfect proportion
- Perfect analogy : AD(a,b,c,d) = 0 !
- Imperfect analogies : AD(1,0,0,0) = 1, AD(1,0,0,1) = 2
- 2 distinct failure cases
- Generalization to Boolean vectors : adding AD componentwise
- With $\mathbb{B}^n$, $AD(\vec{a}, \vec{b}, \vec{c}, \vec{d}) \in [0, 2n]$
- $AD(a, b, c, d) < n$: at least one analogy for one component
- $AD(a, b, c, d) \geq n$: can be the case that not even one analogy holds !

AD implementation (WAPC)

Naive implementation

1. $\vec{d}$ New item to be classified
2. Compute $AD(\vec{a}, \vec{b}, \vec{c}, \vec{d})$ for every triple $(\vec{a}, \vec{b}, \vec{c})$ with solvable class equation
3. Fix v number of voters
4. Solve the class equation $cl(\vec{a}) : cl(\vec{b}) :: cl(\vec{c}) : x$ for the triples having the v lowest AD
5. Majority vote

AD implementation (WAPC)

Naive implementation

1. $\vec{d}$ New item to be classified
2. Compute $AD(\vec{a}, \vec{b}, \vec{c}, \vec{d})$ for every triple $(\vec{a}, \vec{b}, \vec{c})$ with solvable class equation
3. Fix v number of voters
4. Solve the class equation $cl(\vec{a}) : cl(\vec{b}) :: cl(\vec{c}) : x$ for the triples having the v lowest AD
5. Majority vote

Results

- See last column of the previous table
- Historically, first analogy-based classifier in 2007
- Outperforms the best known classifiers (at that time) on some data sets

Moving to real valued attributes (basic method)

Main idea : considering a normalized value as a truth value !

1. Normalized dataset of real value vectors $\in [0,1]^n$
2. For each triple $\vec{a}, \vec{b}, \vec{c}$ with solvable class equation, compute a new real value vector (*vector of truth values*) as
$$A(a_1, b_1, c_1, d_1), \ldots, A(a_n, b_n, c_n, d_n)$$
3. Sort the truth value vectors
4. Allocate to $cl(\vec{d})$ the solution of the class equation for the best truth value vector
5. Majority vote if needed

Moving to real valued attributes (basic method)

Main idea : considering a normalized value as a truth value !

1. Normalized dataset of real value vectors $\in [0,1]^n$
2. For each triple $\vec{a}, \vec{b}, \vec{c}$ with solvable class equation, compute a new real value vector (*vector of truth values*) as
$$A(a_1, b_1, c_1, d_1), \ldots, A(a_n, b_n, c_n, d_n)$$
3. Sort the truth value vectors
4. Allocate to $cl(\vec{d})$ the solution of the class equation for the best truth value vector
5. Majority vote if needed

Which ordering on vectors ?

- Vectors of truth values sorted in decreasing order of the component values ($\in [0,1]$)
- Lexicographic order on the sorted vectors

So what ?

Current results, when extending the idea of taking $\vec{c}$ as a nearest neighbor of $\vec{d}$, for A and A^* (Bounhas *et al.*, 2014)

Datasets	Algo.	**A**				**A***			
value of k		**1**	**3**	**5**	**11**	**1**	**3**	**5**	**11**
Diabetes	Algo 1	65.4±4.4	64.8± 4.1	65.7± 4.4	65.2±4.5	65.4±4.6	64.8±5.3	65.0±5.2	64.3± 5.0
	Algo 2		68.5±4.6	71.0±4.3	**73.0±4.8**		67.5±5.0	69.7±4.7	71.7± 5.2
W. B. Cancer	Algo 1	96.0±1.9	95.2±2.0	95.1 ±1.9	94.7±2.3	96.2±1.8	96.0±2.0	95.8±2.1	95.5 ±2.4
	Algo 2		96.7±2.0	96.7±1.9	96.6±2.3		**97.0±2.0**	96.8±1.9	96.8±2.1
Heart	Algo 1	73.3±7.1	71.7±8.7	72.2±7.9	72.4±7.3	72.9±7.9	71.4±8.5	70.9±7.9	70.6±7.6
	Algo 2		77.1±6.8	78.2±6.9	**82.1±6.1**		77.3±6.9	78.7±6.7	79.8±6.1
Iris	Algo 1	94.2±5.3	95.7±4.6	94.5±5	93.0±5.5	94.2±5.0	93.4±5.7	93.1±5.8	93.2±4.9
	Algo 2		95.8±4.8	95.3±5.1	**96.9 ±4.5**		95.7±4.5	95.2±4.9	94.9±4.9
Wine	Algo 1	95.3±4.0	96.1± 3.6	96.2± 4.2	95.8±4.3	95.8± 4	95.8± 3.9	95.3± 4.3	95.9±3.8
	Algo 2		96.6±3.2	96.9 ± 3.3	**98.2±2.7**		97.1±3.5	97.3±3.4	97.1±3.5
Sat. Image	Algo 1	94.1± 3.6	**95.3±3.4**	95.1±3.2	94.8±2.9	93.5±3.8	94.2±3.8	94.4±3.8	94.7±4.0
	Algo 2		95.1±3.9	94.4±4.1	94.5±3.9		94.8±3.7	94.1±4.2	93.5±4.3
Glass	Algo 1	71.7±8.9	70.2±8.6	70.7±8.6	71.3±9.1	73.7±8.9	73.4±8	73.8±7.8	74.4±8.2
	Algo 2		72.0±8.2	74.1±8	72.1±9.8		74.2±8.4	**74.6±8.7**	73.6±9.3
Ecoli	Algo 1	79.6±6.8	77.4±7.8	77.2±7.4	76.7±5.8	79.7±5.5	78.9±6.2	78.2±6.5	78.6±6.2
	Algo 2		82.3±6.6	84.6±5.6	**86.8±6.0**		81.7±5.7	83.1±5.9	83.9±5.7

Comparison with classification results of some well-known classifiers

Datasets	SVM	JRip	IBK(k=1,k=10)	Algo2 with A & k=11
Diabetes	77.3	76.0	70.0, 71.1	73.0
Cancer	97.1	96.0	96.2 , 96.9	96.6
Heart	83.7	81.1	74.8, 81.4	82.1
Iris	96.0	95.3	95.3, 96.0	96.9
Wine	98.3	92.7	94.9, 95.5	98.2
Sat. Image	94.2	93.9	94.2, 92.2	94.5
Glass	57.9	69.1	70.5, 64.5	72.1
Ecoli	84.2	81.2	80.3, 86.0	86.8

And what about complexity ?

Structural complexity

- good news : polynomial (in the size of the training set)
- bad news :
  1. TS^3 ! (i.e. cubic then not really scalable)
  2. $TS^2 + TS = TS^2$ with the option where c is the nearest neighbor

And what about complexity ?

Structural complexity

- good news : polynomial (in the size of the training set)
- bad news :
  1. TS^3 ! (i.e. cubic then not really scalable)
  2. $TS^2 + TS = TS^2$ with the option where c is the nearest neighbor

What can we do ?

1. Offline processing : suppress "non solvable triples"
2. Offline processing : choose base prototypes (inspired from k-NN)
3. Offline processing : organize the remaining pairs $(\vec{a}, \vec{b})$ in Hamming bags $bag_i = \{(\vec{a}, \vec{b}) | d_{Hamming}(\vec{a}, \vec{b}) = i\}$
4. Add some weight on attributes (2007)

A Kolmogorov complexity approach

A. Cornuejols view

- Analogical bias based on an “economy principle"
- The simpler $P : (a, b) \rightsquigarrow (c, d)$ the better $a : b :: c : d$
- Measure of simplicity : K

A Kolmogorov complexity approach

A. Cornuejols view

- Analogical bias based on an "economy principle"
- The simpler $P : (a, b) \rightsquigarrow (c, d)$ the better $a : b :: c : d$
- Measure of simplicity : K

Implementation via Google

- "black is to white as 0 is to 1" : natural language analogies
- Inspiration : Cornuejols + Li and Vitanyi works
- $a : b :: c : d$ iff $K(b|a) = K(d|c)$ and $K(a|b) = K(c|d)$
- $p(x) = 2^{-K(x)}$ then knowing p we get an estimation for K !
- p is the Google distribution
- Imperfect world... 75% of accuracy

Contents

Analogy-based problem solving

- Natural language processing
- Handwritten character generation
- Solving IQ tests
- Prediction of missing values (matrix abduction)

AP between words

$abc : aba :: acbbb : ababb$

abc	$=$	a	$\cdot$	$\sim$	$\cdot$	b	$\cdot$	c	$\cdot$	$\sim$	$\cdot$	$\sim$	$\cdot$	$\sim$
aba	$=$	a	$\cdot$	$\sim$	$\cdot$	b	$\cdot$	$\sim$	$\cdot$	$\sim$	$\cdot$	a	$\cdot$	$\sim$
$acbbb$	$=$	$\sim$	$\cdot$	a	$\cdot$	$\sim$	$\cdot$	c	$\cdot$	b	$\cdot$	$\sim$	$\cdot$	bb
$ababb$	$=$	$\sim$	$\cdot$	a	$\cdot$	$\sim$	$\cdot$	$\sim$	$\cdot$	b	$\cdot$	a	$\cdot$	bb

$$|ababb| = |aba| + |acbbb| - |abc|$$

abc	aba
$acbbb$	$ababb$

$$ababb \in (aba \bullet acbbb) \backslash abc$$

$\bullet$: shuffle product
(the union of all words obtained by interlacing the two operands)

Applications in computational linguistics

$F - AP$ in a free monoid

A factorization f_w of a word w is a set of n words $(w_1, \cdots w_n)$ such that

$$w = w_1 \cdot w_2 \cdots w_n$$

FAP in a free monoid

$$x \quad : \quad y \quad \overset{fap}{::} \quad z \quad : \quad t \Leftrightarrow \text{There exists four factorizations}$$
$$(f_x, f_y, f_z, f_t) \;\; \text{such as :}$$
$$\forall i \in \{1, n\} : (f_y(i), f_z(i)) \in \{(f_x(i), f_t(i)), (f_t(i), f_x(i))\}$$

For example :

$$\begin{array}{rcccccccccccc}
x = abc & = & a & \cdot & b & \cdot & c & \cdot & \varepsilon & \cdot & \varepsilon & \cdot & \varepsilon \\
y = aba & = & a & \cdot & b & \cdot & \varepsilon & \cdot & \varepsilon & \cdot & a & \cdot & \varepsilon \\
z = acbbb & = & a & \cdot & \varepsilon & \cdot & c & \cdot & b & \cdot & \varepsilon & \cdot & bb \\
t = ababb & = & a & \cdot & \varepsilon & \cdot & \varepsilon & \cdot & b & \cdot & a & \cdot & bb
\end{array}$$

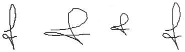
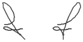
Knowledge-based generation of handwritten characters

Problem : scriptor adaptation of handwriting recognition systems

The scriptor provides few copies of letters :

need for generating more (plausible) characters

Distortions

1. Scale :  $\Rightarrow$ 
2. Verticality :  $\Rightarrow$ 

On-line distortion :

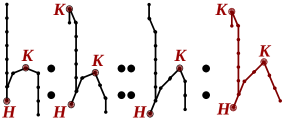
1. Velocity :  &  $\Rightarrow$  & 
2. Curvature :  &  $\Rightarrow$  & 

Analogy-based generation

$h_1 = 9 \backsim 9 \backsim 9\ 9\ 9\ 9\ 9 \backsim H\ 1\ 2 \backsim 4\ K\ 6\ 9\ 9\ 9$
$h_2 = 1\ K \backsim 8\ 9\ 9\ 9\ 9\ 9\ 10\ H \backsim 2\ 2\ 4\ K \backsim 8\ 8\ 9$
$h_3 = \backsim \backsim 9\ 8\ 9\ 9\ 9\ 9\ 9\ 10\ H\ 2\ 2\ 3\ 3\ K\ 8\ 9\ 9 \backsim$

Analogy-based generation

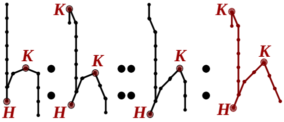
$$\begin{array}{l}
h_1 = 9 \backsim 9 \backsim 9\;9\;9\;9\;9 \backsim H\;1\;2 \backsim 4\;K\;6\;9\;9\;9 \\
h_2 = 1\;K \backsim 8\;9\;9\;9\;9\;9\;10\;H \backsim 2\;2\;4\;K \backsim 8\;8\;9 \\
h_3 = \backsim \backsim 9\;8\;9\;9\;9\;9\;9\;10\;H\;2\;2\;3\;3\;K\;8\;9\;9 \backsim \\
x = 1\;K \backsim 8\;9\;9\;9\;9\;9\;10\;H\;2\;2\;3\;3\;K\;8\;8\;8 \backsim
\end{array}$$



Freeman encoding (directions of move)
+ breaking points (e.g. pen up / down), which help synchronization
Minimization of analogical dissimilarity AD
(choice among several possible solutions)

Analogy-based generation

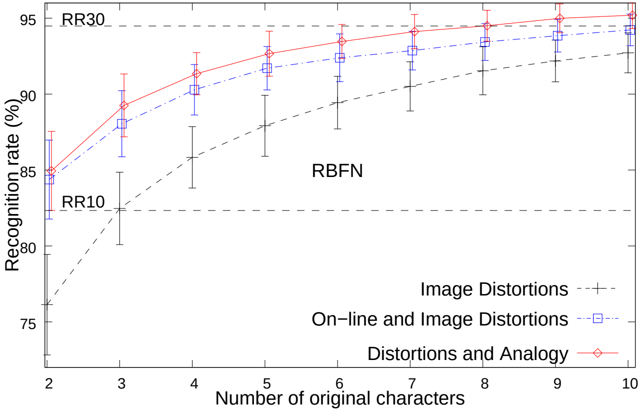
$$
\begin{aligned}
h_1 &= 9 \backsim 9 \backsim 9\ 9\ 9\ 9\ 9 \backsim H\ 1\ 2 \backsim 4\ K\ 6\ 9\ 9\ 9\\
h_2 &= 1\ K \backsim 8\ 9\ 9\ 9\ 9\ 9\ 10\ H \backsim 2\ 2\ 4\ K \backsim 8\ 8\ 9\\
h_3 &= \backsim \backsim 9\ 8\ 9\ 9\ 9\ 9\ 9\ 10\ H\ 2\ 2\ 3\ 3\ K\ 8\ 9\ 9 \backsim\\
x &= 1\ K \backsim 8\ 9\ 9\ 9\ 9\ 9\ 10\ H\ 2\ 2\ 3\ 3\ K\ 8\ 8\ 8 \backsim
\end{aligned}
$$



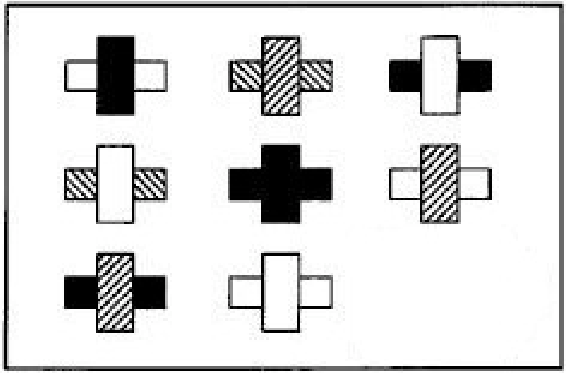
Freeman encoding (directions of move)
+ breaking points (e.g. pen up / down), which help synchronization
Minimization of analogical dissimilarity AD
(choice among several possible solutions)

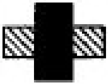
Examples of characters generated by analogy

Results


RR30
RR10
RBFN
Recognition rate (%)
95
90
85
80
75
2
3
4
5
6
7
8
9
10
Number of original characters
Image Distortions
On−line and Image Distortions
Distortions and Analogy

A Raven IQ test and its solution





Analogical approach

(a,b) : f(a,b) :: (c,d) : f(c,d)

$(pi[1,1], pi[1,2]) : pi[1,3] :: (pi[2,1], pi[2,2]) : pic[2,3]) ::$
$(pi[3,1], pic[3,2]) : pi[3,3])$

$(pi[1,1], pi[2,1]) : pi[3,1] :: (pi[1,2], pi[2,2]) : pi[3,2]) ::$
$(pi[1,3], pi[2,3]) : pi[3,3])$

	1	2	3
1	*WB*	*GG*	*BW*
2	*GW*	*BB*	*WG*
3	*BG*	*WW*	?i?ii

i?ii $= GB$

- for the horizontal bars :

(W,G) : B : : (G, B) : W (horizontal analysis)
(W,G) : B : : (B,W) : ?i (horizontal analysis)
(W,G) : B : : (G, B) : W (vertical analysis)
(W,G) : B : : (B,W) : ?i (vertical analysis)

- for the vertical bars :

(B,G) : W : : (W, B) : G
(B,G) : W : : (G,W) : ?ii
(B,W) : G : : (G, B) : W
(B,W) : G : : (W,G) : ?ii

Matrix abduction

	P	C	I	R	D	S
*screen*1	0	1	0	1	0	1
*screen*2	0	0	1	1	0	1
*screen*3	0	0	0	0	1	?
*screen*4	1	1	0	0	1	1

Prediction can be only “plausible", whatever the method

Completing missing values in databases

(M. Abraham, D. M. Gabbay, and U. J. Schild. Analysis of the talmudic argumentum a fortiori inference rule (kal vachomer) using matrix abduction. Studia Logica, 92(3) : 281-364, 2009)

Contents

Current researches and perspectives

- From analogical proportions to logical proportions
- Heterogeneous proportions with application to anomaly detection
- Analogy and Formal Concept Analysis
- Conclusion

Indicators as comparative descriptors

Similarity and dissimilarity indicators

- Given a collection of Boolean properties
  an object A viewed as (equated to) the set of properties that this object has
- two types of comparison indicators :
  - *similarity indicators* : $A \cap B$ and $\overline{A} \cap \overline{B}$ telling us about properties that both A and B have,
  or that both A and B do not have
  - *dissimilarity indicators* : $A \cap \overline{B}$ and $\overline{A} \cap B$ telling us about properties that only one among A and B has

Logical proportions : Definition

- Comparing a pair of Boolean variables (a, b) to another pair (c, d) is done via a pair of equivalences between indicators
- It leads to consider all the conjunctions of 2 equivalences between indicators : they are called *logical proportions*
- A logical proportion $T(a, b, c, d)$ is the conjunction of 2 distinct equivalences between indicators of the form

$$(I_{(a,b)} \equiv I_{(c,d)}) \wedge (I'_{(a,b)} \equiv I'_{(c,d)})$$

using *similarity indicators* $a \wedge b$ and $\overline{a} \wedge \overline{b}$
or *dissimilarity indicators* $a \wedge \overline{b}$ and $\overline{a} \wedge b$

Two important facts

- There are 120 distinct logical proportions
- Logical proportion: Boolean formula with 4 variables and as such, has a truth table with 16 lines
  Any logical proportion is true for exactly 6 lines of its truth table (and false for the 10 others)
- Example

a	b	c	d	$(a \land b \equiv c \land \neg d) \land (\neg a \land \neg b \equiv \neg c \land d)$
1	1	1	0	1
0	0	0	1	1
1	0	1	1	1
0	1	0	0	1
1	0	0	0	1
0	1	1	1	1

Different types of logical proportions. 1

4 *homogeneous* proportions that involve only dissimilarity, or only similarity indicators :

- **analogy** : $A(a, b, c, d) =$
  $(a \wedge \neg b) \equiv (c \wedge \neg d) \wedge (\neg a \wedge b) \equiv (\neg c \wedge d)$
- **reverse analogy** : $R(a, b, c, d) =$
  $(a \wedge \neg b) \equiv (\neg c \wedge d) \wedge (\neg a \wedge b) \equiv (c \wedge \neg d)$
- **paralogy** : $P(a, b, c, d) =$
  $(a \wedge b) \equiv (c \wedge d) \wedge (\neg a \wedge \neg b) \equiv (\neg c \wedge \neg d)$
- **inverse paralogy** : $I(a, b, c, d) =$
  $(a \wedge b) \equiv (\neg c \wedge \neg d) \wedge (\neg a \wedge \neg b) \equiv (c \wedge d)$

Picturing analogy with parallelograms

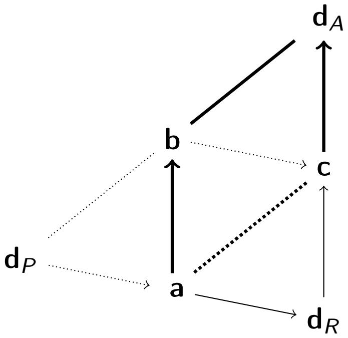
parallel between pairs (a, b) and (c, d) : reminiscent of vector equality

a, b, c and d viewed as elements of the real plan $\mathbb{R}^2$

interpret $a : b :: c : d$ as $\overrightarrow{ab} = \overrightarrow{cd}$.

holds because the coordinates of the 4 points a, b, c, d satisfy an *arithmetic* analogy : $\forall i \in \{1, 2\}, a_i - b_i = c_i - d_i$.

It simply means that the quadrilateral *abcd* is a parallelogram.



Different types of logical proportions. 2

16 *conditional* proportions defined as the conjunction of an equivalence between similarity indicators and of an equivalence between dissimilarity indicators, as, e.g., $((a \wedge b) \equiv (c \wedge d)) \wedge ((a \wedge \neg b) \equiv (c \wedge \neg d))$, which expresses that the two conditional objects[1] $b|a$ and $d|c$ have the same examples (1st condition) and the same counter-examples (2nd condition).

1. A conditional object $b|a$ can take 3 truth values : *true* if $a \wedge b$ is true, *false* if $a \wedge \neg b$ is true, *not applicable* if a is false ; it intuitively represents "if a then b".

Different types of logical proportions. 3

- **20** *hybrid* proportions obtained as the conjunction of two equivalences between similarity and dissimilarity indicators, as in the following example :
  $$a \wedge b \equiv \overline{c} \wedge d \text{ and } \overline{a} \wedge \overline{b} \equiv c \wedge \overline{d}.$$
- **32** *semi-hybrid* proportions for which one half of their expressions involve indicators of the same kind, while the other half requires equivalence between indicators of opposite kinds, as, e.g.,
  $$a \wedge b \equiv c \wedge d \text{ and } \overline{a} \wedge \overline{b} \equiv c \wedge \overline{d}.$$
- **48** *degenerated* proportions whose definition involves 3 distinct indicators only.

How many logical proportions are univocal ?

univocal :
$T(a, b, c, x) = 1$ has a unique solution x when it exists

Answer : 64 !

They might be of interest for learning
(they encode other forms of regularities)

Heterogeneous proportions

- Proportions with an odd number of 1 for valid valuations
- For instance H_a :

a	b	c	d	$(a \wedge \neg b \equiv c \wedge d) \wedge (\neg a \wedge b \equiv \neg c \wedge \neg d)$
1	1	1	0	1
0	0	0	1	1
1	1	0	1	1
0	0	1	0	1
1	0	1	1	1
0	1	0	0	1

- H_a heterogeneous
- There is an intruder and this is not a !

Anomaly detection with heterogeneous proportions

Pick the odd one out in {bus, bicycle, car, truck}

	hasEngine	*canMove*	*canFly*	*canDrive*	*has4Wheels*
A : *bus*	1	1	0	1	1
B : *bicycle*	0	1	0	1	0
C : *car*	1	1	0	1	1
D : *truck*	1	1	0	1	1

For each $x \in a, b, c, d$ compute

$$N(x) = card(\{i \in [1, n] s.t. H_x(A_i, B_i, C_i, D_i) = 0\})$$

$$intruder = argmax_x N(x) \ (= \mathrm{B})$$

Analogical Proportion, Proportional Analogy, and Analogy

Analogical Proportion : four objets of the same kind

A foal is to a mare as a calf is to a cow.
fins are to scales as wings are to feathers.

Proportional Analogy : two couples of objects of the same kind

A foal is to equines as a calf is to bovines.
fins are to fishes as wings are to birds.

Analogy : Proportional Analogy shortened as a kind of metaphor

fins are the wings of fishes.

Metaphor

fins are like wings. life is a journey.

Formal Concept Analysis - Example of four concepts in WAP

	1	2	3	4
a			×	×
b	×		×	
c		×		×
d	×	×		

a	Foal
b	Mare
c	Calf
d	Cow

1	Female and adult
2	Bovine
3	Equine
4	Young

A Foal (Young Equine) is to a Mare (Female adult Equine)

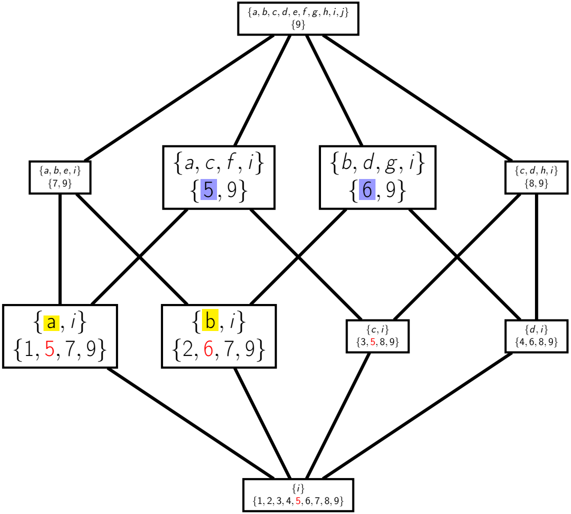
as

A Calf (Young Bovine) is to a Cow (Female adult Bovine)

Fins are to Fishes as Wings are to Birds.

i :	Part of an animal	9 :	Part of an animal
a :	Fins	5 :	Part of a Fish
b :	Wings	6 :	Part of a Bird
c :	Scales	7 :	Mobility part
d :	Feathers	8 :	Covering part
e :	Gills	1 :	Part of a Whale
f :	Beak	2 :	Part of a Bat
g :	Hooves	3 :	Part of a Snake
h :	Thick fur	4 :	Part of a Deinonychus

Fins are to Fishes as Wings are to Birds. *a* is to 5 as *b* is to 6.



	1	2	3	4	5	6	7	8	9
a	×				×		×		×
b		×				×	×		×
c			×		×			×	×
d				×		×		×	×
e					×				×
f						×			×
g							×		×
h								×	×
i	×	×	×	×	×	×	×	×	×

Concept Lattices

Weak Analogical Proportion between Concepts

$$\boxed{\begin{matrix}\{a, i\} \\ \{1, 5, 7, 9\}\end{matrix}} : \boxed{\begin{matrix}\{b, i\} \\ \{2, 6, 7, 9\}\end{matrix}} :: \boxed{\begin{matrix}\{c, i\} \\ \{3, 5, 8, 9\}\end{matrix}} : \boxed{\begin{matrix}\{d, i\} \\ \{4, 6, 8, 9\}\end{matrix}}$$

Proportional Analogy : two objects, two attributes

a	is to	5	as	b	is to	6
a	$\updownarrow$	5	$\Updownarrow$	b	$\updownarrow$	6
Fins	are to	Fishes	as	Wings	are to	Birds

Conclusion

Present

- Analogy *formalized* in terms of *analogical proportion*
- *Powerful* tool for different tasks
- *Competitive* results in classification
- *Shift of paradigm* : consider pairs of examples, can work with few data

Future

- This is just a beginning !
- Scalability
- Particularly appropriate for what class of data ?
- Analogical regression ?
- *Theoretical issues* :
  - better understanding of why / how analogical classification works
  - (other) logical proportions : potential use ?
  - link / hybridization with other machine learning paradigms
  - joint use of analogical proportions and formal concept analysis
- back to cognitive sciences

Some references. 1

Analogy at large (books)

- M. Hesse. Models and Analogies in Science. 1st ed. Sheed & Ward, London, 1963 ; 2nd augmented ed. University of Notre Dame Press, 1966.
- S. J. Russell. The Use of Knowledge in Analogy and Induction. Pitman, UK, 1989.
- B. Indurkhya. Metaphor and Cognition. An Interactionist Approach. Springer, 1992.
- D. Gentner, K. J. Holyoak, B. N. Kokinov (eds.). The Analogical Mind : Perspectives from Cognitive Science. Cognitive Science, and Philosophy. MIT Press, 2001
- Y. Lepage. De l'analogie rendant compte de la commutation en linguistique. Habil. Dir. Rech., Grenoble, 2003. http ://tel.archives-ouvertes.fr/tel-00004372/en
- D. Hofstadter, E. Sander. Surfaces and Essences : Analogy as the Fuel and Fire of Thinking. Basic Books, 2013.
- H. Prade, G. Richard (eds.). Computational Approaches to Analogical Reasoning. Current Trends. Studies in Computational Intelligence, Vol. 548, Springer, 2014.

Kolmogorov complexity approach

- M. Li, P. Vitanyi, An Introduction to Kolmogorov Complexity and its Applications, Springer Verlag, 2008.
- A. Cornuéjols. Analogy as minimization of description length. In : Machine Learning and Statistics : The Interface, (G. Nakhaeizadeh, C. Taylor, eds.), Wiley, 321-336, 1996
- M. Bayoudh, H. Prade, G. Richard. Evaluation of analogical proportions through Kolmogorov complexity. Knowledge-Based Systems, 29 : 20-30, 2012.

Some references. 2

Analogical proportions

- M. Hesse. On defining analogy. Proc. of the Aristotelian Society, 60 : 79-100, 1959.
- S. Klein. Culture, mysticism & social structure and the calculation of behavior. In Proc. 5th Europ. Conf. in Artificial Intelligence (ECAI'82), Orsay, 141-146, 1982.
- Y. Lepage. Analogy and formal languages. Electr. Notes Theor. Comput. Sci., 53, 2001
- N. Stroppa and F. Yvon. Analogical learning and formal proportions : Definitions and methodological issues. ENST technical report, Paris, June 2005.
- L. Miclet and H. Prade. Handling analogical proportions in classical logic and fuzzy logics settings. Proc. 10th Eur. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'09), (C. Sossai, G. Chemello, eds.), Verona, Springer, LNCS 5590, 638-650, 2009.
- H. Prade and G. Richard. Analogical proportions and multiple-valued logics. Proc. 12th Eur. Conf. on Symb. and Quantit. Appr. to Reasoning with Uncertainty (ECSQARU'13), (L. C. van der Gaag, ed.), Utrecht, Jul. 7-10, Springer, LNAI 7958, 497-509, 2013.
- N. Barbot, L. Miclet and H. Prade. Analogical proportions and the factorization of information in distributive lattices.CEUR Workshop Proc. 10th Int. Conf. Concept Lattices & Applications (CLA'13), La Rochelle, (M. Ojeda-Aciego, J. Outrata, eds.),175-186, 2013.

Some references. 3

Analogical proportions in classification

- S. Bayoudh, L. Miclet, A. Delhay. Learning by analogy : A classification rule for binary and nominal data. Proc. Int. Join. Conf. on Artificial Intelligence (IJCAI'07), 678-683, 2007.
- L. Miclet, S. Bayoudh, and A. Delhay. Analogical dissimilarity : definition, algorithms and two experiments in machine learning. JAIR, 32, 793-824, 2008.
- H. Prade, G. Richard, and B. Yao. Enforcing regularity by means of analogy-related proportions - A new approach to classification. Int. J. of Computer Information Systems and Industrial Management Applications, 4 : 648-658, 2012.
- R. M. Moraes, L. S. Machado, H. Prade, G. Richard. Classification based on homogeneous logical proportions. Proc. 33rd SGAI Int. Conf. on Innovative Techniques and Applications of Artificial Intelligence, (M. Bramer, M. Petridis, eds.), Cambridge, UK, Dec. 10-12, Springer, 53-60, 2013.
- M. Bounhas, H. Prade, G. Richard. Analogical classification : A new way to deal with examples. Proc. 21st Europ. Conf. on Artif. Intellig. (ECAI'14), (T. Schaub, G. Friedrich, B. O'Sullivan, eds.), Prague, Aug. 18-22, IOS Press, 135-140, 2014.
- M. Bounhas, H. Prade, G. Richard. Analogical classification : Handling numerical data. Proc. 8th Conf. on Scalable Uncertainty Management (SUM'14), (U. Straccia and A. Cali, eds.), Oxford, Sept. 15-17, Springer, LNAI 8720, 66-79, 2014.

Some references. 4

Analogical problem solving. 1

- E. Melis, M. Veloso. Analogy in problem solving.In Handbook of Practical Reasoning : Computational and Theoretical Aspects. Oxford Univ. Press, 1998.
- H. Gust, K. Kühnberger, and U. Schmid. Metaphors and heuristic-driven theory projection (HDTP). Theoretical Computer Science, 354(1) : 98-117, 2006.
- Th. Boy de la Tour, N. Peltier. Analogy in automated deduction : A survey. In : Computational Approaches to Analogical Reasoning. Current Trends, (H. Prade, G. Richard, eds.), Springer, 103-130, 2014.

- A. Lovett, K. Forbus, J. Usher. A structure-mapping model of Raven's progressive matrices. Proc. 32nd Annual Conf. of the Cognitive Science Soc., Portland, OR, 2010.
- W. Correa, H. Prade, and G. Richard. When intelligence is just a matter of copying. Proc. 20th Eur. Conf. on Artificial Intelligence, Montpellier, Aug. 27-31, IOS Press, 276-281, 2012. 2012.
- M. Ragni and S. Neubert. Solving Raven's IQ-tests : An AI and cognitive modeling approach. Proc. 20th Europ. Conf. on Artificial Intelligence (ECAI'12), (L. De Raedt, C.Bessière, D. Dubois, P. Doherty, P. Frasconi, F. Heintz, and P. J. F. Lucas, eds.), Montpellier, Aug. 27-31, 666-671, 2012.

Some references. 5

Analogical problem solving. 2

- Y. Lepage. Solving analogies on words : An algorithm. Proc. COLING-ACL, Montreal, 728-733, 1998.
- N. Stroppa and F. Yvon. An analogical learner for morphological analysis. Online Proc. 9th Conf. Comput. Natural Language Learning (CoNLL'05), 120-127, 2005.
- Ph. Langlais, A. Patry. Translating unknown words by analogical learning. Proc. Joint Con- ference on Empirical Methods in Natural Language Processing (EMNLP) and Conference on Computational Natural Language Learning (CONLL), Prague, 877-886, 2007.
- S. Bayoudh, H. Mouchère, L. Miclet, E. Anquetil. Learning a classifier with very few examples : Analogy based and knowledge based generation of new examples for character recognition. Proc.18th Europ. Conf. on Machine Learning (ECML'07), Warsaw, Sept., Springer, LNCS 4701, 527-534, 2007.

- W. Correa Beltran, H. Jaudoin, O. Pivert. Estimating null values in relational databases using analogical proportions. Proc. 15th Int. Conf. Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'14), (A. Laurent, O. Strauss, B. Bouchon-Meunier, R. R. Yager, eds.), Montpellier, July 15-19, Part III. Springer, CCIS series, 110-119, 2014.

Some references. 6

Logical proportions

- J. Piaget. Logic and Psychology. Manchester Univ. Press, 1953.
- H. Prade and G. Richard. From analogical proportion to logical proportions. Logica Universalis, 7(4) : 441-505, 2013.
- H. Prade and G. Richard. Homogeneous and heterogeneous logical proportions. The IfCoLog J. of Logics and their Applications, 1 (1), 1-51, 2014.

Formal Concept Analysis / Analogical proportion

- B. Ganter and R. Wille. *Formal Concept Analysis. Mathematical Foundations*. Springer Verlag, 1999.
- R. Bělohlávek. Introduction to formal context analysis. Internal report. Dept of Comp. Sc. Palacký Univ., Olomouk, Czech Rep. 2008.
- L. Miclet, N. Barbot, H. Prade. From analogical proportions in lattices to proportional analogies in formal concepts. Proc. 21st Europ. Conf. on Artif. Intellig. (ECAI'14), (T. Schaub, G. Friedrich, B. O'Sullivan, eds.), Prague, Aug. 18-22, IOS Press, 627-632, 2014.

Thank you for your time