A discussion of analogical-proportion based inference

Henri Prade[1,2] and Gilles Richard[1]

1. IRIT, Toulouse University, France
2. QCIS, University of Technology, Sydney, Australia
{prade, richard}@irit.fr,

Abstract. The Boolean expression of an analogical proportion, i.e., a statement of the form "a is to b as c is to d", expresses that "a differs from b as c differs from d, and vice-versa. This is the basis of an analogical inference principle, which is shown to be a particular instance of the analogical "jump": from $P(s)$, $P(t)$, and $Q(s)$, deduce $Q(t)$. Roughly speaking, an analogical proportion sounds like a sort of qualitative derivative. A counterpart of a first order Taylor-like formula indeed exists for affine Boolean functions. Affine functions can be predicted without error by means of analogical proportions. These affine functions are essentially the constants, the projections, the xor-based functions, and their complements. We discuss how one might take advantage of this state of fact for refining the scope of application of the analogical-proportion based inference to subparts of a Boolean function that may be assumed to be "locally" linear.

1 Introduction

Analogical proportions are statements of the form "a is to b as c is to d" that have been introduced at the time of Aristotle by mimicking numerical proportions. Such statements are appealing since they relate comparisons inside pair (a, b) to comparisons inside pair (c, d), by suggesting that "a differs from b in the same way as c differs from d", and for symmetry reason that "b differs from a in the same way as d differs from c".

Following a series of works aiming at formalizing the idea of analogical proportion in different settings, a Boolean logic modeling has been proposed almost a decade ago. This modeling formally acknowledges the above intuitive reading of an analogical proportion. The analogical-proportion based inference amounts to postulating that if analogical proportions hold on a series of features used to describe four situations a, b, c, d, such a proportion may also hold for other related attributes as well.

It turns out that such a view has been proved to be quite effective for classification tasks in particular. A natural question is then to try to understand why and in what respect. This question is not straightforward. An interesting clue has been recently obtained when showing that if (and only if) the classification function is an affine Boolean function, then the analogical-proportion based inference always predicts the right class [5]. This confirms previous experimental observations.

It also echoes some informal remarks pointing out the fact that an analogical proportion may be reminiscent of a qualitative notion of derivative. Indeed affine Boolean functions satisfy a first order Taylor-like formula as recalled in this paper. Since any Boolean function is piecewise linear (in terms of affine Boolean functions), one may wonder if one cannot take advantage of these facts for a better adjustment of the scope of the analogical-proportion based inference inside areas where the classification functions would be presumably linear.

The Boolean logic modeling of analogical proportions provides a simple basis for computing the result of an analogical inference. The paper mainly aims at pointing out the interest of a *functional* view of Boolean expressions when discussing analogical proportion-based inference. The paper first restates the notion of analogical proportion and its Boolean logic modeling. It then discusses the nature of the analogical proportion-based inference, first showing that it is a particular instance of a general "analogical jump" pattern, then explaining its link with affine Boolean functions, before discussing how we might take advantage of this situation for a better focusing of the analogical proportion-based inference.

2 Analogical proportion

Analogical proportions are statements of the form "a is to b as c is to d" like "Queen is to King as Woman is to Man", or "Paris is to France as Madrid is to Spain". In this paper, we assume that i) a, b, c, d are described in terms of Boolean features and they can be represented by the vector of their values on these features, and that ii) the relevant features are the same for a, b, c, and d. This second hypothesis is obviously not satisfied in the second example, indeed a and c belong to a conceptual universe (the one of cities) distinct from the one of b and d (the one of countries). Such more tricky proportions are discussed in [14].

2.1 Boolean logic modeling

The analogical proportion "a is to b as c is to d", denoted $a : b :: c : d$ in the following, intuitively suggests that a differs from b as c differs from d and b differs from a as d differs from c". In this subsection, a, b, c, and d are just Boolean variables, which pertain to the same unique feature for four items. The analogical proportion is logically expressed as [16] by the quaternary connective:

$$a : b :: c : d \triangleq ((a \wedge \neg b) \equiv (c \wedge \neg d)) \wedge ((\neg a \wedge b) \equiv (\neg c \wedge d)) \quad (1)$$

Note that this logical expression of an analogical proportion put forward dissimilarity, in agreement with the idea that analogy is as much a matter of dissimilarity as a matter of similarity. Similarity appears in the logically equivalent expression

$$a : b :: c : d = ((a \wedge d) \equiv (b \wedge c)) \wedge ((\neg a \wedge \neg d) \equiv (\neg b \wedge \neg c)) \quad (2)$$

This latter expression states that what a and d have in common (positively or negatively), b and c have it also in common.

Table 1 gives the truth table of $a : b :: c : d$. We can see that $a : b :: c : d$ is true for 6 patterns: 0000, 1111, 0011, 1100, 0101 and 1010 (in bold in Table 1) .

It is easy to see that the logical expression of $a : b :: c : d$ satisfies the key properties of an analogical proportion, namely

$a\ b\ c\ d$	$a:b::c:d$	$a\ b\ c\ d$	$a:b::c:d$
0 0 0 0	**1**	1 0 0 0	0
0 0 0 1	0	1 0 0 1	0
0 0 1 0	0	1 0 1 0	**1**
0 0 1 1	**1**	1 0 1 1	0
0 1 0 0	0	1 1 0 0	**1**
0 1 0 1	**1**	1 1 0 1	0
0 1 1 0	0	1 1 1 0	0
0 1 1 1	0	1 1 1 1	**1**

Table 1. Boolean valuations for $a : b :: c : d$

- reflexivity: $a : b : a : b$
- symmetry: $a : b :: c : d \Rightarrow c : d :: a : b$
- central permutation: $a : b :: c : d \Rightarrow a : c :: b : d$

Moreover, it is also worth noticing that the analogical proportion is independent with respect to the positive or negative encoding of a considered feature: $a : b :: c : d = \neg a : \neg b :: \neg c : \neg d$. Besides, with this definition, the analogical proportion is transitive in the following sense: $(a : b :: c : d) \wedge (c : d :: e : f) \Rightarrow a : b :: e : f$.

A simple extension of the definition of analogical proportion to Boolean vectors in $\mathbb{B}^n$ of the form $\boldsymbol{a} = (a_1, ..., a_n)$ is as follows: $\boldsymbol{a} : \boldsymbol{b} :: \boldsymbol{c} : \boldsymbol{d}$ iff $\forall i \in [1, n],\ a_i : b_i :: c_i : d_i$.

2.2 Equation solving

It is an acknowledged property of analogy to be creative. In this modeling, this is related to the following equation solving problem: find x such as $a : b :: c : x$ holds true, a, b, and c being given. It is easy to see that the equation has no solution in two cases: $1 : 0 :: 0 : x$ and $0 : 1 :: 1 : x$. When it exists the solution is clearly unique. It was first suggested by [11,12] that x can be computed as

$$x \triangleq c \equiv (a \equiv b)$$

where $\equiv$ is the equivalence connective $s \equiv t \triangleq (\neg s \vee t) \wedge (\neg t \vee s)$. Moreover, note that $s \equiv t = \neg((s \wedge \neg t) \vee (\neg s \wedge t)) = \neg(s \oplus t)$ where $\oplus$ is the xor connective (exclusive or). Thus it is clear that $c \equiv (a \equiv b)$ can be rewritten as $c \equiv (a \equiv b) = \neg(c \oplus \neg(a \oplus b)) = a \oplus b \oplus c$ since $\neg s = s \oplus 1$ and $1 \oplus 1 = 0$. Connectives $\equiv$ and $\oplus$ are associative operators. Thus, we can write $x = a \oplus b \oplus c$ and Table 2 shows the values of x in the 6 cases where equation $a : b :: c : x$ has a solution, as well as in the two remaining cases where there is no analogical solution for $a : b :: c : x$. In these two latter cases corresponding to patterns 0110 and 1001, we have a reverse analogy [19,20], where "b is to a as c is to d" holds rather than "a is to b as c is to d".

$\boldsymbol{a}$	0	1	0	1	0	1	0	1
$\boldsymbol{b}$	0	1	0	1	1	0	1	0
$\boldsymbol{c}$	0	1	1	0	0	1	1	0
$\boldsymbol{x}$	**0**	**1**	**1**	**0**	**1**	**0**	0	1

Table 2. Solving $a : b :: c : x$

Remark Interestingly enough, the eight patterns appearing in Table 2 with an *even* number of 0 and of 1 are involved in the four homogeneous logical proportions (which includes the analogical proportion and the reverse analogical proportion) [19]. The eight remaining patterns among the $2^4 = 16$ patterns of Table 1, which have an *odd* number of

0 and of 1 (and are at the basis of heterogeneous logical proportions [20]), appear in the columns of Table 3. The computation of the fourth line in such a case from the first three lines, in an equation now denoted $a/b//c/x$, is then given by $x = a \oplus b \oplus c \oplus 1$. This is an interesting operator that takes the majority value in a, b, c (the 6 first columns), provided that it does not lead to unanimity (the last two columns).

$\boldsymbol{a}$	1	0	0	1	0	1	0	1
$\boldsymbol{b}$	0	1	1	0	0	1	0	1
$\boldsymbol{c}$	0	1	0	1	1	0	0	1
$\boldsymbol{x}$	0	1	0	1	0	1	1	0

Table 3. Solving $a/b//c/x$

2.3 Analogical proportions induced by the comparison of two objects

As soon as we have two *distinct* Boolean vectors $\boldsymbol{a}$ and $\boldsymbol{d}$, it is possible to find two other vectors $\boldsymbol{b}$ and $\boldsymbol{c}$ such that $\boldsymbol{a} : \boldsymbol{b} :: \boldsymbol{c} : \boldsymbol{d}$. Indeed, let $Agr(\boldsymbol{a}, \boldsymbol{d})$ be the set of indices where $\boldsymbol{a}$ and $\boldsymbol{d}$ agree and $Dis(\boldsymbol{a}, \boldsymbol{d})$ the set of indices where the two vectors differ. Let us now consider two new vectors $\boldsymbol{b}$ and $\boldsymbol{c}$ such that : $\forall i \in Agr(\boldsymbol{a}, \boldsymbol{d}), a_i = b_i = c_i = d_i$ (all equal to 1 or all equal to 0) and $\forall i \in Dis(\boldsymbol{a}, \boldsymbol{d})(b_i = a_i$ and $c_i = d_i)$ or $(b_i = \neg a_i$ and $c_i = \neg d_i)$.

For instance, $\boldsymbol{a} = 0110, \boldsymbol{d} = 0011$, $Agr(\boldsymbol{a}, \boldsymbol{d}) = \{1, 3\}$ and $Dis(\boldsymbol{a}, \boldsymbol{d}) = \{2, 4\}$. Then $\boldsymbol{b} = 0111$ and $\boldsymbol{c} = 0010$ make $\boldsymbol{a} : \boldsymbol{b} :: \boldsymbol{c} : \boldsymbol{d}$ true. This may be viewed as instances of the equation solving problem $a : x :: x' : d$ with two unknowns x and x'. Obviously, we have always a solution: $x = a$ and $x' = d$ or $x = d$ and $x' = a$. But as soon as $Dif(\boldsymbol{a}, \boldsymbol{d})$ contains at least two indices as in the above example, we have solutions where the four vectors $\boldsymbol{a}, \boldsymbol{x}, \boldsymbol{x}', \boldsymbol{d}$ are distinct, as shown in the example. The creation of $(\boldsymbol{b}, \boldsymbol{c})$ from $\boldsymbol{a}$ and $\boldsymbol{d}$ is illustrated in [10] on images, using a non logical approach.

2.4 Non Boolean attributes

Real life datasets rarely involve Boolean features only. There may be a mix between Boolean and nominal feature (like $color$), or real-valued features. In the case of nominal attributes, it is quite common to binarise in the following way: for instance $color$ can take three values $red, green, blue$ which will be coded as $100, 010, 001$ (using features as $isRed, isGreen, isBlue$). Then we are back to the Boolean case. The case of real-valued features is more sophisticated and needs the tool of multi-valued logic to be properly handled. We refer the interested reader to [21,7] for a comprehensive development. Nevertheless, in this paper, we strictly stick to the Boolean case.

Another important issue is to get the relevant feature to code a given problem. We do not focus on this issue here as we consider the vectors coming from existing datasets, so the coding has already been done.

3 Analogical proportion-based inference

We have seen that we can obtain the solution x, when it exists, of an analogical proportion equation $a : b :: c : x$ as $x = a \equiv b \equiv c = a \oplus b \oplus c$. The analogical proportion-based inference principle [23] can now be stated as follows:

$$\frac{\forall i \in \{1, ..., n\},\ \ a_i : b_i :: c_i : d_i \text{ holds}}{\forall j \in \{n+1, ..., p\},\ \ a_j : b_j :: c_j : d_j \text{ holds}}$$

This is a form of analogical reasoning where we transfer knowledge from some components of our vectors to their remaining components, tacitly assuming that the values of the n first components determine the values of the others. An important particular case of this pattern is when $p = n + 1$, which corresponds to the situation in classification where the $(n + 1)$th component corresponds to the class of the item described by the n first features. Note also that this pattern is a tool for guessing missing values in a table, a problem, which has been considered for a long time [1]. Let us now examine how this inference pattern can be related to a more usual "analogical jump" pattern.

3.1 An instance of a general "analogical jump" pattern

In its simplest form, analogical reasoning, without any reference to the notion of proportion, is usually viewed as a way to infer some new fact on the basis of a single observation. Analogical reasoning has been mainly formalized in the setting of first order logic [6,13], and in second order logic [9]. A basic pattern for analogical reasoning is then to consider 2 terms s and t, to observe that they share a property P, and knowing that another property Q also holds for s, to infer that it holds for t as well. This is known as the "analogical jump" and can be described with the following simplified inference pattern, leading (possibly) to a wrong conclusion:

$$\frac{P(s) \;\; P(t) \;\; Q(s)}{Q(t)} \quad (AJ)$$

Making such an inference pattern valid would require the implicit hypothesis that P determines Q inasmuch as $\nexists u \; P(u) \wedge \neg Q(u)$. This may be ensured if there exists an underlying functional dependency, or more generally, if it is known for instance that when something is true for an object of a certain type, then it is true for all objects of that type. Otherwise, without such guarantees, the result of an analogical inference may turn to be definitely wrong.

To link the above analogical pattern with the concept of analogical proportion, it is tempting to write something like: $P(s) : P(t) :: Q(s) : Q(t)$ since we have 4 terms which obey, at least from a syntactic viewpoint, the structure of an analogical proportion. Indeed, it is sufficient to encode each piece of information in a binary way according to the presence or the absence of P, Q, s, or t in the corresponding term, and we get the encoding d of $Q(t)$ via the equation solving process as in Table 4. In that

	P	Q	s	t	
$\boldsymbol{a}$	1	0	1	0	$P(s)$
$\boldsymbol{b}$	1	0	0	1	$P(t)$
$\boldsymbol{c}$	0	1	1	0	$Q(s)$
$\boldsymbol{d}$	0	1	0	1	$Q(t)$

Table 4. A syntactic view of analogical jump

case, $\boldsymbol{a} = P(s), \boldsymbol{b} = P(t), \boldsymbol{c} = Q(s), \boldsymbol{d} = Q(t)$ are encoded as Boolean vectors where the semantics carried by the predicate symbols P and Q is not considered.

In [26,2,15,3], the authors take a similar inspiration where, starting from Boolean datasets and focusing on binary classification problem, they apply the following inference principle (and obtain competitive results on benchmark data sets):

$$\frac{\boldsymbol{a} : \boldsymbol{b} :: \boldsymbol{c} : \boldsymbol{d}}{cl(\boldsymbol{a}) : cl(\boldsymbol{b}) :: cl(\boldsymbol{c}) : cl(\boldsymbol{d})} \quad AP$$

It means that if 4 Boolean vectors build a valid analogical proportion, then it should be true that their classes build also a valid proportion. Starting from this viewpoint, in the case where $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ are in a sample set, i.e., their classes are known, and $\boldsymbol{d}$ being the object to be classified, if the equation $cl(\boldsymbol{a}) : cl(\boldsymbol{b}) :: cl(\boldsymbol{c}) : x = 1$ is solvable (in that case, we say that the triple $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c})$ is *class solvable*), they allocate its solution to $cl(\boldsymbol{d})$ just by applying the previous principle. Experiments highlight the predictive power of this principle. Let us understand why AP principle is just a particular instance of (AJ):

- Considering $\boldsymbol{a}$ and $\boldsymbol{b}$ as Boolean vectors in $\mathbb{B}^n$, the vector $\boldsymbol{k} = \boldsymbol{a} - \boldsymbol{b}$ belongs to $\{-1, 0, 1\}^n$ and summarizes the result of the comparison between $\boldsymbol{a}$ and $\boldsymbol{b}$. So given such a vector k, we define the predicate $P_k(\boldsymbol{a}, \boldsymbol{b}) := (\boldsymbol{a} - \boldsymbol{b} = \boldsymbol{k})$.
- Then we can consider 3 predicate symbols Q_1, Q_2, Q_3 defined as follows:
  1. $Q_1(\boldsymbol{a}, \boldsymbol{b}) := (cl(\boldsymbol{a}) = cl(\boldsymbol{b}))$
  2. $Q_2(\boldsymbol{a}, \boldsymbol{b}) := (cl(\boldsymbol{a}) = 0) \wedge (cl(\boldsymbol{b}) = 1))$
  3. $Q_3(\boldsymbol{a}, \boldsymbol{b}) := (cl(\boldsymbol{a}) = 1) \wedge (cl(\boldsymbol{b}) = 0))$

Let us note that the Q_i's are pairwise mutually exclusive predicates. Using these predicate symbols, the following rule:

$$\frac{P_k(\boldsymbol{a}, \boldsymbol{b}) \;\; P_k(\boldsymbol{c}, \boldsymbol{d}) \;\; Q_i(\boldsymbol{a}, \boldsymbol{b})}{Q_i(\boldsymbol{c}, \boldsymbol{d})}$$

is just an instance of (AJ). Moreover, it states that when the difference $\boldsymbol{a} - \boldsymbol{b}$ equals to $\boldsymbol{c} - \boldsymbol{d}$, then the relation between $cl(\boldsymbol{a})$ and $cl(\boldsymbol{b})$ is the same as the relation between $cl(\boldsymbol{c})$ and $cl(\boldsymbol{d})$. If we notice that $\boldsymbol{a} : \boldsymbol{b} :: \boldsymbol{c} : \boldsymbol{d}$ is just equivalent to $P_k(\boldsymbol{a}, \boldsymbol{b}) \wedge P_k(\boldsymbol{c}, \boldsymbol{d})$ (in the exact sense of the formal definition of the analogical proportion applied componentwise), and $Q_i(\boldsymbol{a}, \boldsymbol{b}) \wedge Q_i(\boldsymbol{c}, \boldsymbol{d})$ entails $cl(\boldsymbol{a}) : cl(\boldsymbol{b}) :: cl(\boldsymbol{c}) : cl(\boldsymbol{d})$, we obtain:

$$\frac{\boldsymbol{a} : \boldsymbol{b} :: \boldsymbol{c} : \boldsymbol{d}}{cl(\boldsymbol{a}) : cl(\boldsymbol{b}) :: cl(\boldsymbol{c}) : cl(\boldsymbol{d})}$$

which is exactly the expression of AP used in [2].

In pattern AP, we transfer the identity of differences pertaining to pairs $(\boldsymbol{a}, \boldsymbol{b})$ and $(\boldsymbol{c}, \boldsymbol{d})$ to the relation between their classes. It enables us to predict the missing information about $\boldsymbol{d}$, using AP as an extrapolation principle. This is clearly a form of reasoning that is not sound, but which may be useful for trying to guess unknown values.

3.2 Link with affine Boolean functions

As already mentioned [4], an analogical proportion of the form "$cl(\boldsymbol{b})$ is to $cl(\boldsymbol{a})$ as $\boldsymbol{b}$ is to $\boldsymbol{a}$" sounds a bit like the expression of the *qualitative derivative* of a function cl underlying the classification process, since the derivative of a function f in $\boldsymbol{a}$ is the limit when $x \longrightarrow a$ of the ratio $\frac{f(\boldsymbol{x})-f(\boldsymbol{a})}{\boldsymbol{x}-\boldsymbol{a}}$, which is a matter of comparing two algebraic differences[1]. Moreover, considering the 6 patterns that make true the analogical proportion, it can be also noticed that if there is a change from $\boldsymbol{a}$ to $\boldsymbol{b}$, there should a change in the same direction from $\boldsymbol{c}$ to $\boldsymbol{d}$. Besides, it has been observed that once extended from Boolean to graded scales, the analogical proportion-based inference provides a *linear* interpolation mechanism [7]. This is due to the fact that in this case $a : b :: c : d = 1$ if and only if $d - c = b - a$ (where a, b, c, $d \in [0, 1]$) and the solution of the equation $a : x :: x : d = 1$ is $x = \frac{a+d}{2}$.

In fact, it has been recently formally proved [5] that the AP principle is sound as soon as the labeling function is an affine Boolean function (this means in practice that the function is a constant, a projection, an xor function, or an $\equiv$ function, over some subsets of the n attributes). Moreover it can be shown that there is no other Boolean function for which this is true [5].

It is well-known that any Boolean function can be put in a polynomial form where the sum is taken as $\oplus$ and the product as the $\min$ [22]. With a functional view of Boolean expressions in mind, one can use the notion of qualitative derivative and define a Taylor-like development of a Boolean function; see, e.g., [17,25]. For instance, consider the linear function $f(x, y) = x \oplus y$ (indeed a polynomial of degree 1 and arity 2). Then it can be checked (see Table 5), that we can write a Taylor-like development of the form:

$$f(x, y) = f(a, b) \oplus \partial_x^f(a) \wedge (x \ominus a) \oplus \partial_y^f(b) \wedge (y \ominus b) \triangleq \Sigma(x, y)$$

where $s \ominus t = s \oplus t$ since $s \ominus t = x \Leftrightarrow s = t \oplus x$, and where all the partial derivatives are equal to 1 here. One can also rewrite the above equality as $f(x, y) \ominus f(a, b) = (x \ominus a) \oplus (y \ominus b)$, which is indeed the Boolean counterpart of what holds for affine functions in $\mathbb{R}^n$.

$x\ y\ a\ b$	$f(x,y) = x \oplus y$	$f(a,b) = a \oplus b$	$\partial_x^f(a) \wedge (x \ominus a)$	$\partial_y^f(b) \wedge (y \ominus b)$	$\Sigma(x,y)$
0 0 0 0	0	0	$1 \wedge (0 \ominus 0) = 0$	$1 \wedge (0 \ominus 0) = 0$	0
0 0 0 1	0	1	$1 \wedge (0 \ominus 0) = 0$	$1 \wedge (0 \ominus 1) = 1$	0
0 0 1 0	0	1	$1 \wedge (0 \ominus 1) = 1$	$1 \wedge (0 \ominus 0) = 0$	0
0 0 1 1	0	0	$1 \wedge (0 \ominus 1) = 1$	$1 \wedge (0 \ominus 1) = 1$	0
0 1 0 0	1	0	$1 \wedge (0 \ominus 0) = 0$	$1 \wedge (1 \ominus 0) = 1$	1
0 1 0 1	1	1	$1 \wedge (0 \ominus 0) = 0$	$1 \wedge (1 \ominus 1) = 0$	1
0 1 1 0	1	1	$1 \wedge (0 \ominus 1) = 1$	$1 \wedge (1 \ominus 0) = 1$	1
0 1 1 1	1	0	$1 \wedge (0 \ominus 1) = 1$	$1 \wedge (1 \ominus 1) = 0$	1
1 0 0 0	1	0	$1 \wedge (1 \ominus 0) = 1$	$1 \wedge (0 \ominus 0) = 0$	1
1 0 0 1	1	1	$1 \wedge (1 \ominus 0) = 1$	$1 \wedge (0 \ominus 1) = 1$	1
1 0 1 0	1	1	$1 \wedge (1 \ominus 1) = 0$	$1 \wedge (0 \ominus 0) = 0$	1
1 0 1 1	1	0	$1 \wedge (1 \ominus 1) = 0$	$1 \wedge (0 \ominus 1) = 1$	1
1 1 0 0	0	0	$1 \wedge (1 \ominus 0) = 1$	$1 \wedge (1 \ominus 0) = 1$	0
1 1 0 1	0	1	$1 \wedge (1 \ominus 0) = 1$	$1 \wedge (1 \ominus 1) = 0$	0
1 1 1 0	0	1	$1 \wedge (1 \ominus 1) = 0$	$1 \wedge (1 \ominus 0) = 1$	0
1 1 1 1	0	0	$1 \wedge (1 \ominus 1) = 0$	$1 \wedge (1 \ominus 1) = 0$	0

Table 5. Taylor-like expression of the linear function $f(x, y) = x \oplus y$

[1] We may also remember that the idea of differential has inspired adaption methods in case-based reasoning for solving numerical problems [8], even if case-based reasoning deals with cases one by one rather than handling triples of cases.

3.3 Adjusting analogical proportion-based inference

Since analogical proportion-based inference works perfectly for predicting affine Boolean functions, it is natural to wonder if given a training set $\mathcal{T}$ it would not be possible to cover it with a *piecewise* affine Boolean function.

The answer is yes, ... and in many different ways! It is easy to see that we just need two pieces. Indeed let $f(x_1, \cdots, x_n)$ be an affine function. In a binary classification problem, any function (be affine or not) partitions $\mathcal{T}$ into two parts $\mathcal{T}_f$ and $\mathcal{T}_{f\oplus 1}$ on which respectively f and $f \oplus 1$ correctly predict the class (since $f \oplus 1$ is just $\neg f$).

If the classification in $\mathcal{T}$ obeys an affine Boolean function f then either $\mathcal{T}_f$ or $\mathcal{T}_{f\oplus 1}$ is empty. This also means that the application of the analogical-proportion principle will amount to apply function f to a new item to be classified (or function $f \oplus 1$ if $\mathcal{T}_f$ is empty). So if the training set $\mathcal{T}$ is covered by an "almost" linear function, this means that one of the two subsets of the partition of $\mathcal{T}$ is very large with respect to the other.

When the training set is not covered by a unique affine Boolean function, this means that there exist triples that lead to false predictions when applying AP. Then we might think that it happens more often when a, b, c do not all belong to the same subset in some partition induced by a function f. So an idea for "almost" linear functions, would be to look for the affine Boolean functions such as $\mathcal{T}_f$ (or $\mathcal{T}_{f\oplus 1}$) is the largest possible subset, which would make easier the finding of triples such as all the three $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ are in it. However another issue is to wonder if some partitions are more appropriate than others for guessing the class of a particular new item.

Generally speaking, the issue is to find a way to identify those triples that are "suspect", i.e., likely to yield a faulty prediction, among a set of candidate triples that enables you to apply AP. Indeed in case of multiple triples, which is the usual situation, we apply a voting procedure among the predictions of the applicable triples, where sometimes the faulty triples are the majority. How to restrict this voting procedure to "good triples"? Another idea may come from a careful examination of the way triples are built and of the meaning of pairs inside, as first suggested in [4].

In Table 6, we have reordered the vectors in a particular way. Indeed the table shows that building the analogical proportion $\boldsymbol{a} : \boldsymbol{b} :: \boldsymbol{c} : \boldsymbol{d}$ is a matter of pairing the pair $(\boldsymbol{a}, \boldsymbol{b})$ with the pair $(\boldsymbol{c}, \boldsymbol{d})$. More precisely, on features or attributes $\mathcal{A}_1$ to $\mathcal{A}_{j-1}$, the four vectors are equal; on attributes $\mathcal{A}_j$ to $\mathcal{A}_{r-1}$, $\boldsymbol{a} = \boldsymbol{b}$ and $\boldsymbol{c} = \boldsymbol{d}$, but $(\boldsymbol{a}, \boldsymbol{b}) \neq (\boldsymbol{c}, \boldsymbol{d})$. In other words, on attributes $\mathcal{A}_1$ to $\mathcal{A}_{r-1}$ $\boldsymbol{a}$ and $\boldsymbol{b}$ agree and $\boldsymbol{c}$ and $\boldsymbol{d}$ agree as well. This contrasts with attributes $\mathcal{A}_r$ to $\mathcal{A}_n$, for which we can see that $\boldsymbol{a}$ differs from $\boldsymbol{b}$ as $\boldsymbol{c}$ differs from $\boldsymbol{d}$ (and vice-versa). In columns we recognize the 6 patterns that makes the analogical proportion true. There are two cases, either $cl(\boldsymbol{a}) = cl(\boldsymbol{b})$ (and then $cl(\boldsymbol{x}) = cl(\boldsymbol{c})$), or $cl(\boldsymbol{a}) \neq cl(\boldsymbol{b})$ (and then $cl(\boldsymbol{x}) = cl(\boldsymbol{b})$). In the first case, it suggests that the particular change observed between $\boldsymbol{a}$ and $\boldsymbol{b}$ on features from $\mathcal{A}_r$ to $\mathcal{A}_n$ does not affect cl in the context defined by the values of the features from $\mathcal{A}_1$ to $\mathcal{A}_{r-1}$ where $\boldsymbol{a}$ and $\boldsymbol{b}$ are equal. Applying AP amounts to assuming that this absence of effect is true in other contexts of values of features from $\mathcal{A}_1$ to $\mathcal{A}_{r-1}$. So the smaller the number of features from $\mathcal{A}_r$ to $\mathcal{A}_n$, the more cautious. A similar reasoning can be done when $cl(\boldsymbol{a}) \neq cl(\boldsymbol{b})$ where the change on the features from $\mathcal{A}_r$ to $\mathcal{A}_n$ should be responsible of the change of class in the context of the values of the other attributes. Observe also that if we have two pairs $(\boldsymbol{a}, \boldsymbol{b})$ and $(\boldsymbol{a'}, \boldsymbol{b'})$ such as $\boldsymbol{a'} : \boldsymbol{b'} :: \boldsymbol{a} : \boldsymbol{b}$, while $\boldsymbol{a} : \boldsymbol{b} :: \boldsymbol{c} : \boldsymbol{x}$, then by transitivity we have $\boldsymbol{a'} : \boldsymbol{b'} :: \boldsymbol{c} : \boldsymbol{x}$. Thus transitivity agrees with the idea that if a change has an effect (or no effect) in some context, then it may be the same elsewhere.

	$\mathcal{A}_1$	...	$\mathcal{A}_{i-1}$	$\mathcal{A}_i$	...	$\mathcal{A}_{j-1}$	$\mathcal{A}_j$	...	$\mathcal{A}_{k-1}$	$\mathcal{A}_k$	...	$\mathcal{A}_{r-1}$	$\mathcal{A}_r$	...	$\mathcal{A}_{s-1}$	$\mathcal{A}_s$	...	$\mathcal{A}_n$	cl
$\boldsymbol{a}$	1	...	1	0	...	0	1	...	1	0	...	0	1	...	1	0	...	0	$cl(\boldsymbol{a})$
$\boldsymbol{b}$	1	...	1	0	...	0	1	...	1	0	...	0	0	...	0	1	...	1	$cl(\boldsymbol{b})$
$\boldsymbol{c}$	1	...	1	0	...	0	0	...	0	1	...	1	1	...	1	0	...	0	$cl(\boldsymbol{c})$
$\boldsymbol{x}$	1	...	1	0	...	0	0	...	0	1	...	1	0	...	0	1	...	1	$cl(\boldsymbol{x})$?

Table 6. Pairing pairs (a, b) and (c, d)

A last idea would be to consider "continuous" analogical proportion and to solve interpolative equation $\boldsymbol{a} : \boldsymbol{x} :: \boldsymbol{x} : \boldsymbol{b}$. In a Boolean setting, such an equation has no solution, except in the trivial situation where $\boldsymbol{a} = \boldsymbol{b}$, then $\boldsymbol{x} = \boldsymbol{a}$. In the associated class equation one has necessarily $cl(\boldsymbol{a}) = cl(\boldsymbol{x}) = cl(\boldsymbol{b})$. Then one may relax $\boldsymbol{a} : \boldsymbol{x} :: \boldsymbol{x} : \boldsymbol{b}$ to a subset of features and makes sure that $\boldsymbol{x}$ is between $\boldsymbol{a}$ and $\boldsymbol{b}$ in the sense that $\max(h(\boldsymbol{a}, \boldsymbol{x}), h(\boldsymbol{x}, \boldsymbol{b})) \leq h(\boldsymbol{a}, \boldsymbol{b})$, where h is the Hamming distance. In such a case, we have a variant of nearest neighbors methods.

We have emphasized the role played by affine Boolean functions, suggesting that the training set in a classification problem might be restricted to subsets of examples more relevant for a new item to be classified. These subsets of examples should be covered by some affine Boolean function. Finding them remains an open question.

4 Conclusion

The paper has intended to provide an advanced discussion of the analogical proportion-based inference principle in the Boolean case, in a classification perspective. As already said, analogical proportion-based inference is also available for nominal and real valued data. The application of analogical proportions to regression is an open problem; then the agreement between a qualitative and a quantitative view of these proportions is crucial (see, e.g., [24] on such issue in learning); in that respect the main gradual extension [21,7] clearly distinguishes between situations where the changes from $\boldsymbol{a}$ to $\boldsymbol{b}$ and from $\boldsymbol{c}$ to $\boldsymbol{d}$ are in the same direction, and where the changes are in opposite directions.

Generally speaking, some authors, e.g., [18], view qualitative reasoning as made of components such as: comparison, categorization, identification of relations, and emergence of a meaning. Analogical proportions seem to offer an interesting mixture of at least two or three of these ingredients [14]; the proper understanding of their interrelationships is still to be further explored.

References

1. 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.
2. S. Bayoudh, L. Miclet, and A. Delhay. Learning by analogy: A classification rule for binary and nominal data. *Proc. Int. Joint Conf. on Artif. Intellig. (IJCAI'07)*, pages 678–683, 2007.
3. M. Bounhas, H. Prade, and G. Richard. Analogical classification: A new way to deal with examples. In *Proc. 21st Eur.Conf. Artif. Intell. (ECAI'14), Prague*, IOS Press,135-140, 2014.
4. W. Correa, H. Prade, and G. Richard. Trying to understand how analogical classifiers work. In E. Hüllermeier et al., editor, *Proc. 6th Inter. Conf. on Scalable Uncertainty Management (SUM'12), Marburg, Germany*, LNAI 7520, pages 582–589. Springer Verlag, 2012.

5. M. Couceiro, N. Hug, H. Prade, and G. Richard. Analogy-preserving functions: A way to ext-end Boolean samples. In *Proc.26th Int. J. Conf. Artif. Intel. IJCAI'17, Melbourne*, 2017.
6. T. R. Davies and Russell S. J. A logical approach to reasoning by analogy. In J. P. McDermott, editor, *Proc. of the 10th International Joint Conference on Artificial Intelligence (IJCAI'87), Milan*, pages 264–270. Morgan Kaufmann, 1987.
7. D. Dubois, H. Prade, and G. Richard. Multiple-valued extensions of analogical proportions. *Fuzzy Sets and Systems*, 292:193–202, 2016.
8. B. Fuchs, J. Lieber, A. Mille, and A. Napoli.Differential adaptation: An operational approach to adaptation for solving numerical problems with CBR. *Know.-Bd. Sys.*, 68:103–114, 2014.
9. H. Gust, K. Kühnberger, and U. Schmid. Metaphors and heuristic-driven theory projection (HDTP). *Theoretical Computer Science*, 354(1):98 – 117, 2006.
10. L. Jing, Y. Lu, H. Gang, and K. Sing Bing. Visual attribute transfer through deep image analogy. In *ACM Transactions on Graphics, (Proc. of Siggraph)*, volume 36, 2017.
11. S. Klein. Culture, mysticism & social structure and the calculation of behavior. In *Proc. 5th Europ. Conf. in Artificial Intelligence (ECAI'82), Orsay, France*, pages 141–146, 1982.
12. S. Klein. Analogy and mysticism and the structure of culture (and Comments & Reply). *Current Anthropology*, 24 (2):151–180, 1983.
13. E. Melis and M. Veloso. Analogy in problem solving. In *Handbook of Practical Reasoning: Computational and Theoretical Aspects*. Oxford Univ. Press, 1998.
14. L. Miclet, N. Barbot, and H. Prade. From analogical proportions in lattices to proportional analogies in formal concepts. In T. Schaub, G. Friedrich, and B. O'Sullivan, editors, *Proc. 21st Europ. Conf. on Artificial Intelligence (ECAI'14), Prague, Aug. 18-22*, volume 263 of *Frontiers in Artificial Intelligence and Applications*, pages 627–632. IOS Press, 2014.
15. L. Miclet, S. Bayoudh, and A. Delhay. Analogical dissimilarity: Definition, algorithms and two experiments in machine learning. *J. Artif. Intell. Res. (JAIR)*, 32:793–824, 2008.
16. L. Miclet and H. Prade. Handling analogical proportions in classical logic and fuzzy logics settings. In *Proc. 10th Eur. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'09),Verona*, pages 638–650. Springer, LNCS 5590, 2009.
17. E. Mizraji. Differential and integral calculus for logical operations. a matrix-vector approach. *J. Log. Comput.*, 25(3):613–638, 2015.
18. A. Mucchielli. Les processus intellectuels fondamentaux sous-jacents aux techniques et mthodes qualitatives. *RecherchesQualitatives*, 3:1–27, 2006.
19. H. Prade and G. Richard. From analogical proportion to logical proportions. *Logica Universalis*, 7(4):441–505, 2013.
20. H. Prade and G. Richard. Homogenous and heterogeneous logical proportions. *IfCoLog J. of Logics and their Applications*, 1(1):1–51, 2014.
21. H. Prade and G. Richard. Multiple-valued logic interpretations of analogical, reverse analogical, and paralogical proportions. In *Proc. 40th IEEE Int. Symp. on Multiple-Valued Logic (ISMVL'10)*, pages 258–263, Barcelona, 2010.
22. M. H. Stone. The theory of representation for Boolean algebras. *Trans. of the American Mathematical Society*, 40(1):37111, 1936.
23. N. Stroppa and F. Yvon. Analogical learning and formal proportions: Definitions and methodological issues. Technical Report D004, ENST-Paris, 2005.
24. Dorian Suc, Daniel Vladusic, and Ivan Bratko. Qualitatively faithful quantitative prediction. *Artificial Intelligence*, 158(2):189 – 214, 2004.
25. S. N. Yanushkevich and V. P. Shmerko. Taylor expansion of logic functions: From conventional to nanoscale design. Int. TICSP Wksp Spectral Meth. & Multirate Signal Process.,2004.
26. F. Yvon and N. Stroppa. Formal models of analogical proportions. Technical Report D008, Ecole Nationale Supérieure des Télécommunications, Paris, France, 2006.