Behavior of Analogical Inference w.r.t. Boolean Functions

Miguel Couceiro[1], Nicolas Hug[2], Henri Prade[2,3], Gilles Richard[2]
1. University of Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France
2. IRIT, University of Toulouse, France
3. QCIS, University of Technology, Sydney, Australia
miguel.couceiro@inria.fr, nh.nicolas.hug@gmail.com, {prade, richard}@irit.fr

Abstract

It has been observed that a particular form of analogical inference, based on analogical proportions, yields competitive results in classification tasks. Using the algebraic normal form of Boolean functions, it has been shown that analogical prediction is always exact iff the labeling function is affine. We point out that affine functions are also meaningful when using another view of analogy. We address the accuracy of analogical inference for arbitrary Boolean functions and show that if a function is ε-close to an affine function, then the probability of making a wrong prediction is upper bounded by 4ε. This result is confirmed by an empirical study showing that the upper bound is tight. It highlights the specificity of analogical inference, also characterized in terms of the Hamming distance.

1 Introduction

There have been several proposals for formalizing analogical inference in logical settings [Davies and Russell, 1987; Russell, 1989; Gust *et al.*, 2006]. More recently, a propositional logic modeling of analogical proportions has been developed [Miclet and Prade, 2009; Prade and Richard, 2013]. This latter view is at the basis of a successful classification method for which empirical evidence was provided in classification problems [Miclet *et al.*, 2008; Bounhas *et al.*, 2017].

Analogical proportions are statements of the form "a is to b as c is d", usually denoted $a : b :: c : d$. Such a proportion expresses that the ways a and b differ are the same as the ways c and d differ. In this paper, one should think of a, b, c and d as being represented by tuples of m Boolean feature values, i.e., as elements of the m-dimensional Boolean space $\mathbb{B}^m = \{0, 1\}^m$. How does analogy-based inference work?

Let f be a Boolean function from $\mathbb{B}^m$ to $\mathbb{B}$ and $x \in \mathbb{B}^m$ such that $f(x)$ is unknown ($f(x)$ can be viewed as a class label). The inference amounts to say that if there exist three items a, b, c with known labels such that $a : b :: c : x$ holds, then $f(x)$ may be the solution (if any) of the analogical proportion $f(a) : f(b) :: f(c) : f(x)$ [Stroppa and Yvon, 2005].

The notion of analogical proportion and the associated inference mechanism may appear a bit uncommon at first glance. First, as far as we know, the relationship between analogical proportion and Hamming distance has never been investigated despite the fact that the latter is a key notion for appreciating differences in a Boolean framework. Second, analogical proportion seems a quite sophisticated relation between 4 items and one may wonder whether it occurs rarely in practice.

In this paper, we shed some light into these concerns. Another important issue is to understand why it works in practice. Beyond the general good reputation of analogy, no explanation for its good results has been provided until now, with the exception of the recent contribution [Couceiro *et al.*, 2017] where a remarkable behavior w.r.t. affine Boolean functions is established. Namely, given a sample S of elements where a function f is known, it is shown that we can guess with certainty the value of $f(x)$ for $x \notin S$ for some classes of functions f, using analogical inference. The key result is that analogical inference does not make any error iff f is an affine Boolean function. Nevertheless, no theoretical or experimental results were provided regarding the possible link between the accuracy of the inference process and the distance between the target labeling function f and the set of affine functions. Here, we address the latter concern.

The paper is organized as follows. Section 2 gives the necessary background on two models (M and K) of Boolean analogical proportions. In Section 3, we link analogical proportions with Hamming distance. We show that, in general, it is possible to find non trivial intermediary items b and c given any a and d such that $a : b :: c : d$ holds. Section 4 surveys Boolean functions and their algebraic normal forms. In Section 5, we compare the class of model-preserving functions, based on M and K definitions, and show that both coincide with affine functions. In Section 6, we consider the case where the target function is not affine, and show that if a function is ε-close to an affine function, then the probability of making wrong predictions based on analogical inference is upper bounded linearly w.r.t. ε.

2 Analogical Proportions

An analogy A (or analogical proportion[1]) over a nonempty set X is a quaternary relation over X that satisfies four axioms [Miclet *et al.*, 2008] (all formulas are universally quantified):

1. **Reflexivity:** $A(a, b, a, b)$.

[1]In the paper, the term *analogy* is short for *analogical proportion*

2. **Symmetry:** $A(a, b, c, d) \implies A(c, d, a, b)$.
3. **Central permutation:** $A(a, b, c, d) \implies A(a, c, b, d)$.
4. **Unicity:** $A(a, b, a, x) \implies x = b$.

In this paper, we restrict the interpretation domain X to be the Boolean set $\mathbb{B} = \{0, 1\}$ and $A(a, b, c, d)$ will be denoted $a : b :: c : d$. It leads to 2 remarkable Boolean models:

$$M = \{0000, 1111, 1010, 0101, 1100, 0011\},$$
$$K = \{0000, 1111, 1010, 0101, 1100, 0011, 0110, 1001\}.$$

The K model was proposed, although very informally, by the American researcher S. Klein [Klein, 1982] more than 25 years before the introduction of the Boolean modeling of analogical proportion. It obeys the 4 postulates of analogical proportion but it has what may be considered as an undesirable behavior: 'a is to b as c is to d' is the same as 'b is to a as c is to d'. Still it is of interest to keep it for comparison with genuine analogical proportion. It is noteworthy that K is such that any added element of $\mathbb{B}^4$ would contradict *unicity*.

M is the smallest Boolean model obeying the 4 axioms above. We denote by $a : b :: c : d$ an analogical proportion w.r.t. M. It can be defined by the logical expression:

$$a : b :: c : d \triangleq (a \wedge \neg b \leftrightarrow c \wedge \neg d) \wedge (\neg a \wedge b \leftrightarrow \neg c \wedge d)$$

where $\wedge, \neg, \leftrightarrow$ are respectively the logical connectives for conjunction, negation and equivalence. It expresses the fact that a differs from b as c differs from d (and vice-versa: b differs from a as d from c). Considering $\mathbb{B}$ as a subset of $\mathbb{R}$, another equivalent formulation is given by:

$$a : b :: c : d \text{ if and only if } a - b = c - d.$$

Note however that the latter reformulation is not suitable in the Boolean setting since $a - b$ may not belong to $\mathbb{B}$ when $a, b \in \mathbb{B}$. It can be checked that M (and K) satisfies the *code independence property*:

$$a : b :: c : d \quad \text{if and only if} \quad \neg a : \neg b :: \neg c : \neg d,$$

which guarantees that 0 and 1 play symmetric roles in the encoding of the Boolean features. An M-analogy over $\mathbb{B}^m$ is defined componentwise, i.e., $\forall \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{B}^m$,

$$\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d} \triangleq \bigwedge_{i=1}^{m} a_i : b_i :: c_i : d_i.$$

When an analogy is defined over a set X, given 3 elements a, b, c of X, the requirement that the relation $A(a, b, c, x)$ holds defines an equation where x is the unknown. In the Boolean case, with M or K as models, there is at most one solution. When equation $a : b :: c : x$ is solvable, we abbreviate it with $solvable(a, b, c)$. In this case, $sol(a, b, c)$ denotes the solution. K-analogies extend to $\mathbb{B}^m$ as well, and they are always solvable.

Throughout this paper, we will adopt model M.

3 A New View on Analogies

We first show that an analogy is valid only if certain constraints (in terms of Hamming distances) on their 4 items hold. Then we highlight that analogies emerge as soon as 2 items are compared. We focus on M-analogies here.

3.1 Links with the Hamming Distance

A standard distance over $\mathbb{B}^m$ is the *Hamming distance*: $H_m(\mathbf{x}, \mathbf{y}) = |\{i \in [1, m] : x_i \neq y_i\}|$, which satisfies:

$$H_{m+1}(\mathbf{x}, \mathbf{y}) = H_m(\mathbf{x}/m, \mathbf{y}/m) + H_1(x_{m+1}, y_{m+1})$$

where $\mathbf{x}/m$ denotes the projection of $\mathbf{x}$ over $\mathbb{B}^m$. We write H when there is no ambiguity about the dimension. A first link between Boolean analogy and Hamming distance is given by:

Property 1 $\forall \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{B}^m$ *s.t.* $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$*, we have:*

$$H(\mathbf{a}, \mathbf{b}) = H(\mathbf{c}, \mathbf{d}), H(\mathbf{a}, \mathbf{c}) = H(\mathbf{b}, \mathbf{d}), H(\mathbf{a}, \mathbf{d}) = H(\mathbf{b}, \mathbf{c}).$$

Proof: We prove the first equality $H(\mathbf{a}, \mathbf{b}) = H(\mathbf{c}, \mathbf{d})$ by induction on m. Clearly, it holds for $m = 1$. By definition of analogy on $\mathbb{B}^{m+1}$, given $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{B}^{m+1}$, $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$ iff

$$\mathbf{a}/m : \mathbf{b}/m :: \mathbf{c}/m : \mathbf{d}/m \text{ and } a_{m+1} : b_{m+1} :: c_{m+1} : d_{m+1}$$

The induction hypothesis leads to $H(\mathbf{a}/m, \mathbf{b}/m) = H(\mathbf{c}/m, \mathbf{d}/m)$ and $H(a_{m+1}, b_{m+1}) = H(c_{m+1}, d_{m+1})$. Adding these two equations leads to the expected result. The same reasoning applies to the two remaining equalities. □

However the reverse property does not hold for $0 : 1 :: 1 : 0$ and for $1 : 0 :: 0 : 1$. A stronger relation can be established:

Property 2 $\forall \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$*, such that* $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$*, we have*

$$H(\mathbf{a}, \mathbf{b}) + H(\mathbf{a}, \mathbf{c}) = H(\mathbf{a}, \mathbf{d}).$$

Similarly, we have $H(\mathbf{d}, \mathbf{c}) + H(\mathbf{d}, \mathbf{b}) = H(\mathbf{d}, \mathbf{a})$.

Proof: The first equality is easily proved by induction on m. It is true for $m = 1$. Since we have $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$ iff :

$$\mathbf{a}/m : \mathbf{b}/m :: \mathbf{c}/m : \mathbf{d}/m \text{ and } a_{m+1} : b_{m+1} :: c_{m+1} : d_{m+1}$$

we can apply the induction hypothesis to get $H(\mathbf{a}/m, \mathbf{b}/m) + H(\mathbf{a}/m, \mathbf{c}/m) = H(\mathbf{a}/m, \mathbf{d}/m)$ and $H(a_{m+1}, b_{m+1}) + H(a_{m+1}, c_{m+1}) = H(a_{m+1}, d_{m+1})$. And we are done by adding these two equations. The second equality is deduced from the first one using Property 1. □

As Prop. 2 rules out $0 : 1 :: 1 : 0$ and $1 : 0 :: 0 : 1$, we may wonder if it characterizes analogy.

Property 3 *When* $m = 1$, $H(\mathbf{a}, \mathbf{b}) + H(\mathbf{a}, \mathbf{c}) = H(\mathbf{a}, \mathbf{d})$ *(or* $H(\mathbf{d}, \mathbf{c}) + H(\mathbf{d}, \mathbf{b}) = H(\mathbf{d}, \mathbf{a})$*) implies* $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$.

Proof: If $H(\mathbf{a}, \mathbf{d}) = 0$, then $H(\mathbf{a}, \mathbf{b}) = H(\mathbf{a}, \mathbf{c}) = 0$ and $a = b = c = d$. If $H(\mathbf{a}, \mathbf{d}) = 1$, then either $H(\mathbf{a}, \mathbf{b}) = 0$ or $H(\mathbf{a}, \mathbf{b}) = 1$. In the first case, $\mathbf{a} = \mathbf{b}$ and $H(\mathbf{a}, \mathbf{c}) = 1$, i.e. $\mathbf{c} = \mathbf{d}$ since $m = 1$ and variables can only take values 0 or 1. Similarly if $H(\mathbf{a}, \mathbf{b}) = 1$, $\mathbf{b} = \mathbf{d}$ and $H(\mathbf{a}, \mathbf{c}) = 0$, i.e. $\mathbf{a} = \mathbf{c}$. The other relation follows by permutation. □

When $m > 1$, $H(\mathbf{a}, \mathbf{b}) + H(\mathbf{a}, \mathbf{c}) = H(\mathbf{a}, \mathbf{d})$ is not enough to ensure $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$ as we can see by taking $\mathbf{a} = 000, \mathbf{b} = 100, \mathbf{c} = 110, \mathbf{d} = 111$. Indeed $a_i : b_i :: c_i : d_i$ holds only for $i = 2$, and not for $i = 1$ or 3, while both $H(\mathbf{a}, \mathbf{b}) + H(\mathbf{a}, \mathbf{c}) = H(\mathbf{a}, \mathbf{d})$ and $H(\mathbf{d}, \mathbf{c}) + H(\mathbf{d}, \mathbf{b}) = H(\mathbf{d}, \mathbf{a})$ hold. So, we should require that the relation holds componentwise for characterizing analogy:

Property 4 $\forall \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{B}^m$, $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$ *if and only if* $\forall k = 1, \ldots, m$, $H(a_k, b_k) + H(a_k, c_k) = H(a_k, d_k)$.

If we consider $\mathbb{B}^m$ as a subset of $\mathbb{R}^m$, $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$ means that the 4 vectors are the vertices of a parallelogram. Note that this parallelogram can be flat ($\mathbf{a} : \mathbf{b} :: \mathbf{a} : \mathbf{b}$), or even reduce to a single vertex ($\mathbf{a} : \mathbf{a} :: \mathbf{a} : \mathbf{a}$). With this interpretation, it is quite clear that given 3 vectors $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{B}^m$, it is not always possible to find a fourth vector $\mathbf{d} \in \mathbb{B}^m$ such that $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$ (although this 4th vector always exists in $\mathbb{R}^m$).

3.2 Analogy and Comparison of Two Items

Analogy is implicitly present when we compare two objects. Indeed, for any two distinct objects represented as Boolean vectors $\mathbf{a}$ and $\mathbf{d}$, there is a way to fill in the gap between them in a "smooth" way. More precisely, it is possible to find two other vectors $\mathbf{b}$ and $\mathbf{c}$ such that $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$. To formalize this idea, let $Agr(\mathbf{a}, \mathbf{d})$ be the set of indices where $\mathbf{a}$ and $\mathbf{d}$ agree and $Dis(\mathbf{a}, \mathbf{d})$ the set of indices where the two vectors differ. Then we have the following easy-to-check result.

Property 5 *Let* $\mathbf{a}, \mathbf{d} \in \mathbb{B}^m$ *and take* $\mathbf{b}, \mathbf{c} \in \mathbb{B}^m$ *such that:*

1. $\forall i \in Agr(\mathbf{a}, \mathbf{d}), a_i = b_i = c_i = d_i$
2. $\forall i \in Dis(\mathbf{a}, \mathbf{d})$, *either* ($b_i = a_i$ *and* $c_i = d_i$) *or* ($b_i = \neg a_i$ *and* $c_i = \neg d_i$).

Then $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$. *Moreover if* $|Dis(\mathbf{a}, \mathbf{d})| \geq 2$, *then we can find* $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ *that are pairwise distinct.*

Example: Let $\mathbf{a} = 01100$ and $\mathbf{d} = 00111$ in $\mathbb{B}^5$. Then $Agr(\mathbf{a}, \mathbf{d}) = \{1, 3\}$ and $Dis(\mathbf{a}, \mathbf{d}) = \{2, 4, 5\}$. By 1., we have $\mathbf{b} = 0b_21b_4b_5$ and $\mathbf{c} = 0c_21c_4c_5$, while 2. gives us more freedom, for instance, $\mathbf{b} = 01110$ and $\mathbf{c} = 00101$ are solutions. Indeed, 01100 : 01110 :: 00101 : 00111. We can check that $H(\mathbf{a}, \mathbf{b}) = H(\mathbf{c}, \mathbf{d}) = 1$, $H(\mathbf{a}, \mathbf{c}) = H(\mathbf{b}, \mathbf{d}) = 2$, while $H(\mathbf{a}, \mathbf{d}) = H(\mathbf{b}, \mathbf{c}) = 3 = |Dis(\mathbf{a}, \mathbf{d})|$.

Analogy allows us to compare $\mathbf{a}$ to $\mathbf{d}$ via some transitional steps involving intermediary values $\mathbf{b}$ to $\mathbf{c}$, which are closer to $\mathbf{a}$ (or to $\mathbf{d}$) than $\mathbf{a}$ is to $\mathbf{d}$. In a visual context, analogies were recently used for expressing constraints in multi-class categorization tasks and attribute transfer [Hwang *et al.*, 2013; Jing *et al.*, 2017]. The latter reference illustrates the creation of "intermediate" images b and c building an analogy with two given images a and d (in an approximative way).

4 Boolean Functions: Brief Overview

We now survey some useful results on Boolean functions.

4.1 $\mathbb{B}$ as a Field and $\mathbb{B}^m$ as a Vector Space

When it comes to represent a Boolean function, an algebraic view may be more relevant than sticking to the standard logical operators. In that case, multiplication in $\mathbb{B}$ is the usual one (corresponding to conjunction), but addition is addition modulo 2 (still denoted + and corresponding to XOR), making $\mathbb{B}$ a 2-element field such that $x + x = 0$. In this way, $\mathbb{B}^m$ becomes a vector space over the 2-element field $\mathbb{B}$: scalar multiplication distributes componentwise and addition of vectors is likewise defined componentwise as addition modulo 2. The Hamming weight $wt(\mathbf{x})$ is $|\{x_i | x_i = 1\}|$ so that the Hamming distance $H(\mathbf{x}, \mathbf{y})$ is just $wt(\mathbf{x} + \mathbf{y})$. By endowing $\mathbb{B}$ with the total order $0 < 1$, $\mathbb{B}^m$ becomes totally ordered via the lexicographic order. It is sometimes useful to represent a Boolean function $f \in \mathcal{F}_m$ by the unique vector of $\mathbb{B}^{2^m}$ of its values: $wt(f)$ is the number of vectors $\mathbf{x}$ such that $f(\mathbf{x}) = 1$. A linear form over $\mathbb{B}^m$ is a Boolean function $g \in \mathcal{F}_m$ such that for every $\mathbf{x}, \mathbf{y} \in \mathbb{B}^m$ and $\alpha \in \mathbb{B}$ we have

$$g(\mathbf{x} + \mathbf{y}) = g(\mathbf{x}) + g(\mathbf{y}) \quad \text{and} \quad g(\alpha \mathbf{x}) = \alpha g(\mathbf{x}).$$

Note that the second equality reduces to $g(\mathbf{0}) = 0$ in this case. The set of linear forms over $\mathbb{B}^m$ is usually denoted $(\mathbb{B}^m)^*$ and called the dual space of $\mathbb{B}^m$. This is also a vector space over $\mathbb{B}$. $\mathbb{B}^m$ is naturally equipped with a dot product given by

$$< \mathbf{x}, \mathbf{y} >= \Sigma_{i=1}^m x_i y_i.$$

This dot product induces the following well known properties

- $\mathbb{B}^m$ and $(\mathbb{B}^m)^*$ are isomorphic.
- $\mathbf{x} \mapsto g_{\mathbf{x}}$ where $g_{\mathbf{x}}(\mathbf{y}) =< \mathbf{x}, \mathbf{y} >$ is an isomorphism.

Analogy over $\mathbb{B}$ (thought of as a field) can then be defined by:

$$a : b :: c : d \triangleq (a + d = b + c) \wedge (ad = bc).$$

Extension to $\mathbb{B}^m$ is made componentwise. These definitions are strictly equivalent to the ones given in Section 2. We have to note that $a : b :: c : d$ implies $d = a + b + c$ and, with the notations of Section 2, if $solvable(a, b, c)$ then $sol(a, b, c) = a + b + c$. This is also true with vectors in $\mathbb{B}^m$. The reverse property does not hold as can be seen with 0110 which is not an analogy but where the last element is still the sum of the 3 first elements.

4.2 Polynomial Representation of Boolean Functions

A monomial is a term of the form $\mathbf{x}_I \triangleq \prod_{i \in I} x_i$, for some (possibly empty) finite set of positive integers I with the convention that 1 is the empty monomial $\mathbf{x}_\emptyset$. The size $\mid I \mid$ is called the degree of $\mathbf{x}_I$denoted $d(\mathbf{x}_I)$. A polynomial is a sum of monomials $\sum_{I \subseteq \{1,\ldots,m\}} \omega_I \cdot \mathbf{x}_I$ where each ω_I belongs to $\mathbb{B}$ (addition is understood as addition modulo 2). Equipped with scalar multiplication over $\mathbb{B}$ and addition, the set $\mathbb{B}[x_1, \ldots, x_m]$ of polynomials is a ring over $\mathbb{B}$. The degree of a polynomial is the largest degree of its monomials. The monomial 0 is just $0 \cdot \mathbf{x}_\emptyset$. Note that 0 and 1 are the only monomials of degree 0. A projection (corresponding to the selection of one feature) is obviously reduced to a monomial of degree 1.

Each Boolean function $f \in \mathcal{F}_m$ is uniquely represented by a polynomial called ANF for Algebraic Normal Form [Stone, 1936; Zhegalkin, 1927] i.e.,

$$f(x_1, \ldots, x_m) = \sum_{I \subseteq \{1,\ldots,m\}} \omega_I \cdot \mathbf{x}_I.$$

The degree $d(f)$ of a Boolean function f is the degree of its ANF. With no danger of ambiguity, we confuse f with its ANF. The negation of a function f is just $f + 1$. When $d(f) \leq 1$, we say that f is affine. In other words, a function $f \in \mathcal{F}_m$ is affine iff

$$f(x_1, \ldots, x_m) = \sum_{i \in [1,m]} \omega_i \cdot x_i + \omega_0 =< \omega, \mathbf{x} > +\omega_0,$$

where $\omega = (\omega_1, \ldots, \omega_m)$. The set of affine functions of arity m is denoted by $\mathcal{L}_m$, so that $\mathcal{L} = \bigcup_m \mathcal{L}_m$ is the set of all affine functions. If the term ω_0 is 0, the affine function is just a linear form. Thus, affine functions are either linear forms (including the constant function 0) or their complements.

5 Analogy and Affine Functions

5.1 Predicting by Analogy

For $m \geq 1$, let $\mathcal{F}_m = \mathcal{F}(\mathbb{B}^m, \mathbb{B})$, denote the set of all Boolean functions of arity m, and let $\mathcal{F} = \bigcup_{m \in \mathbb{N}} \mathcal{F}_m$, i.e., the set of all Boolean functions. Given a sample set $S \subseteq \mathbb{B}^m$ and a function $f \in \mathcal{F}_m$, the *analogical root* of a given element $\mathbf{x} \in \mathbb{B}^m$, denoted by $\mathbf{R}_S(\mathbf{x}, f)$, can be defined as follows [Hug *et al.*, 2016]:

Definition 1 *For $f \in \mathcal{F}_m$, define $\mathbf{R}_S(\mathbf{x}, f)$ to be the set*

$$\{(\mathbf{a}, \mathbf{b}, \mathbf{c}) \in S^3 \,:\, \mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{x} \textit{ and } \textit{solvable}(f(\mathbf{a}), f(\mathbf{b}), f(\mathbf{c}))\}.$$

In other words, $\mathbf{R}_S(\mathbf{x}, f)$ is the set of triples over S that are analogically linked to $\mathbf{x}$ or, equivalently, the set of vertices $\mathbf{a}, \mathbf{b}, \mathbf{c} \in S^3$ such that $(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{x})$ is a parallelogram in the Boolean cube $\mathbb{B}^m$. The set $\mathbf{R}_S(\mathbf{x}, f)$ can be used to provide a prediction for the label $f(\mathbf{x})$ of $\mathbf{x}$ using analogical inference. For any element $\mathbf{x} \in \mathbb{B}^m \setminus S$ such that $\mathbf{R}_S(\mathbf{x}, f) \neq \emptyset$, the *analogical label* of $\mathbf{x}$ is defined as follows:

Definition 2

$$\overline{\mathbf{x}}_{S,f} \triangleq Mode(\{sol(f(\mathbf{a}), f(\mathbf{b}), f(\mathbf{c})) \,:\, (\mathbf{a}, \mathbf{b}, \mathbf{c}) \in \mathbf{R}_S(\mathbf{x}, f)\})$$

where $Mode(\Sigma)$ returns the most frequent element of a multiset Σ (i.e. $\overline{\mathbf{x}}_{S,f}$ is the most common label among all candidate labels). In case of a tie, the returned element is randomly chosen among the most frequent elements. It is clear that, for these elements, we do not necessarily have $\overline{\mathbf{x}}_{S,f} = f(\mathbf{x})$.

Definition 3 *A function $f \in \mathcal{F}_m$ is said to be **analogy preserving** (AP) if for every $\mathbf{x} \in \mathbb{B}^m \setminus S$ such that $\mathbf{R}_S(\mathbf{x}, f) \neq \emptyset$, we have $\overline{\mathbf{x}}_{S,f} = f(\mathbf{x})$.*

Note that, thanks to code independency property, f is AP iff $\neg f$ is AP. Proposition 1, which has been proved in [Couceiro *et al.*, 2017], gives an alternative description of AP functions.

Proposition 1 *A function $f : \mathbb{B}^m \to \mathbb{B}$ is AP if and only if for every $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{B}^m$, the following implication holds:*

$$\begin{aligned} \mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d} \quad \text{and} \quad & \textit{solvable}(f(\mathbf{a}), f(\mathbf{b}), f(\mathbf{c})) \\ \implies \quad & sol(f(\mathbf{a}), f(\mathbf{b}), f(\mathbf{c})) = f(\mathbf{d}). \end{aligned}$$

Analogical prediction may look quite similar, at least in principle, to a k-NN technique where a final vote among k candidates is performed. Nevertheless, we can ensure the accuracy of the prediction: this is not the case with k-NN algorithms.

5.2 From AP Functions to KP Functions

In this section we recall the explicit description (Proposition 3 below) of the class of AP functions, which was recently presented in [Couceiro *et al.*, 2017] by making use of tools in clone theory [Denecke and Wismath, 2002; Lau, 2006]. The intuition comes from the following result in [Szendrei, 1980].

Proposition 2 *A function $f \in \mathcal{F}_m$ is affine if and only if for every $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{B}^m$, we have*

$$f(\mathbf{a} + \mathbf{b} + \mathbf{c}) = f(\mathbf{a}) + f(\mathbf{b}) + f(\mathbf{c}).$$

Inspired by the latter result, the following explicit description of AP functions has been obtained:

Proposition 3 *A function f is AP if and only if it is affine.*

This result provides a clear picture of what can be expected from a prediction outside a given Boolean sample: only affine functions can ensure sound predictions. If the number of Boolean functions of arity m is 2^{2^m}, the number of affine functions of arity m is just 2^{m+1}. This shows that affine functions constitute a rather small subclass of functions in this landscape. This fact asks for an extension of our results to non affine functions.

In fact, everything which has been done in the previous sections, based on the smallest model M of analogy, can be done with the model K. Let us denote $K(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d})$ the fact that $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ belong to model K. Obviously

$$\mathbf{a} : \mathbf{b} : \mathbf{c} : \mathbf{d} \implies K(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d})$$

and the main difference with analogy is that every triple in $\mathbb{B}^m$ is uniquely K-solvable i.e.,

$$\forall \mathbf{a}, \mathbf{b}, \mathbf{c}, \exists! \mathbf{d} \text{ such that } K(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}).$$

In fact, $\mathbf{d} = \mathbf{a} + \mathbf{b} + \mathbf{c}$ is the unique solution of the equation $K(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{x})$. The definitions we have used for defining AP functions still apply and lead to the following definition for an KP (for K-preserving) function:

Definition 4 *f is KP if and only if $\forall \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{B}^m$:*

$$K(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) \implies K(f(\mathbf{a}), f(\mathbf{b}), f(\mathbf{c}), f(\mathbf{d})) \quad (1)$$

Using Proposition 2, we get the following result:

Proposition 4 *The class of KP functions is the class of affine functions.*

Proof: Observe that $K(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d})$ means $\mathbf{d} = \mathbf{a} + \mathbf{b} + \mathbf{c}$, and thus if f is KP, then $f(\mathbf{a} + \mathbf{b} + \mathbf{c}) = f(\mathbf{a}) + f(\mathbf{b}) + f(\mathbf{c})$ by (1). It then follows from Proposition 2 that f is affine. □

So considering analogical inference, using model M or K does not make any difference w.r.t. preserving functions.

6 Approximately Affine Functions

Applying analogical inference without caution may lead to incorrect labelling of instances $\mathbf{x} \notin S$: if f is not AP, we may have $\overline{\mathbf{x}}_{S,f} \neq f(\mathbf{x})$ for some $\mathbf{x}$. It is natural to ask whether the use of analogical inference to *approximately affine* functions can ensure labels which are also *approximately correct*. Let us formalize what we mean by "approximation".

6.1 The Meaning of Approximately Affine

The word *approximation* is understood in various ways within the Boolean algebra community. In this paper we will take a probabilistic approach. Given two functions $f, g \in \mathcal{F}_m$, the Hamming distance $H(f, g)$ between f and g is just $wt(f + g)$ when f and g are represented as vectors (see Section 4.1). In other words,

$$H(f, g) = |dis(f, g)| = |\{\mathbf{x} \in \mathbb{B}^m \,:\, f(\mathbf{x}) \neq g(\mathbf{x})\}|.$$

Let P denote the uniform distribution[2] over $\mathbb{B}^m$. Then $P(\mathbf{x} \in dis(f, g)) = \frac{H(f,g)}{2^m}$ is a distance over the set $\mathcal{F}_m$ of all Boolean functions, and we will denote it by $d(f, g)$.

[2]Using the uniform probability will allow to use relative frequencies as probability approximation.

Now, we say that a function g *approximates* f *at level* ε or that it is ε*-close to* f if $d(f,g) \leq \varepsilon$. This can also be read as "the proportion of $\mathbf{x} \in \mathbb{B}^m$ such that $f(\mathbf{x}) \neq g(\mathbf{x})$ does not exceed ε". Given a set $\mathcal{C} \subseteq \mathcal{F}_m$, $d(f,\mathcal{C})$ is defined as usual as $min_{g\in\mathcal{C}} d(f,g)$. Obviously, when it comes to computing the distance $d(f,\mathcal{L})$ of f to the set $\mathcal{L}$ of affine functions, we are sure that $d(f,\mathcal{L}) \leq \frac{1}{2}$. Indeed, if there is an affine function g such that $d(f,g) > \frac{1}{2}$, then $g+1$ is still affine and $d(f,g+1) < \frac{1}{2}$, by definition.

Definition 5 *A function f is ε-close to $\mathcal{C}$ if $d(f,\mathcal{C}) \leq \varepsilon$. If f is ε-close to the set $\mathcal{L}$ of affine functions for a threshold ε, then f is said ε-approximately affine (necessarily $\varepsilon \in [0,\frac{1}{2}]$).*

As we will see, this notion of ε-closeness allow us to assess the quality of the analogical inference.

6.2 Sound Analogical Inference

In this section, we mainly recall concepts introduced in [Couceiro *et al.*, 2017]. Given a subset $S \subseteq \mathbb{B}^m$, a function $f \in \mathcal{F}_m$, and an element $\mathbf{x} \in \mathbb{B}^m$, if $\mathbf{R}_S(\mathbf{x}, f) \neq \emptyset$, then an analogical label $\overline{\mathbf{x}}_{S,f}$ can be associated to $\mathbf{x}$. Let us then define the *analogical extension* $\mathbf{E}_S(f)$ of S w.r.t. f as the set of $\mathbf{x} \in \mathbb{B}^m$ such that $\mathbf{R}_S(\mathbf{x}, f) \neq \emptyset$. Formally:

Definition 6 $\mathbf{E}_S(f) \triangleq \{\mathbf{x} \in \mathbb{B}^m : \exists(\mathbf{a},\mathbf{b},\mathbf{c}) \in S^3, \mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{x}$ *and* $solvable(f(\mathbf{a}), f(\mathbf{b}), f(\mathbf{c}))\}$.

Intuitively, $\mathbf{E}_S(f)$ can be seen as the set of all $\mathbf{x} \in \mathbb{B}^m$ that are the fourth vertex of a parallelogram built with 3 vertices from the sample set S, provided that the equation related to the associated labels is solvable. It is clear that $S \subseteq \mathbf{E}_S(f)$ because $\mathbf{a} : \mathbf{a} :: \mathbf{a} : \mathbf{a}$ always holds. Define

$$\mathrm{err}_{S,f} = P(\{\mathbf{x} \in \mathbf{E}_S(f) \setminus S \,|\, \overline{\mathbf{x}}_{S,f} \neq f(\mathbf{x})\}).$$

Hence, we are interested in the elements $\mathbf{x} \in \mathbf{E}_S(f) \setminus S$ whose predicted label $\overline{\mathbf{x}}_{S,f}$ is $f(\mathbf{x})$.

Definition 7 *We say that $\mathbf{E}_S(f)$ is sound if $err_{S,f} = 0$.*

Due to the definition of the probability P, $\mathrm{err}_{S,f} = 0$ means the proportion of wrong analogical labels is zero. In this case, $\mathbf{E}_S(f)$ can be used as an extended sample set for classification tasks. In the opposite case, $\mathbf{E}_S(f)$ is said to be *unsound* (i.e. noisy) which is equivalent to $\mathrm{err}_{S,f} > 0$. Thus Proposition 1 can then be restated as follows:

Proposition 5 *A function $f \in \mathcal{F}_m$ is AP if and only if for every $S \subseteq \mathbb{B}^m$, $\mathbf{E}_S(f)$ is sound.*

6.3 Approximately Sound Extension

By Proposition 5, if f is not affine (and thus not AP), then there exists a sample S such that $\mathbf{E}_S(f)$ is unsound. It means that [3] $P(\mathrm{err}_{S,f} > 0) > 0$ where P relates to the probability over all possible choices of the sample S.

This probability provides a measure of the imperfection of f w.r.t. analogical inference. Given a function f ε-close to $\mathcal{L}$, can we determine an upper bound of this probability? The following proposition gives a positive answer.

[3]From a rigorous viewpoint, we should denote $P'(\mathrm{err}_{S,f} > 0)$ where P is the probability distribution over the universe $\mathbb{B}^m$, and P' is the induced distribution over the powerset $2^{\mathbb{B}^m}$. In our case, $\mathrm{err}_{S,f} > 0$ simply means $\{x \in \mathbf{E}_S(f) : |x \neq f(x)\} \neq \emptyset$.

Proposition 6 *Let $\varepsilon \in [0,\frac{1}{2}]$. If f is ε-approximately affine, then $P(err_{S,f} > 0) \leq 4\varepsilon$.*

Proof: Let be f ε-close to $\mathcal{L}$ and $g \in \mathcal{L}$ such that

$$d(f,g) \leq \varepsilon, \text{ i.e., } P(\{x \in \mathbb{B}^m | f(x) \neq g(x)\}) \leq \varepsilon.$$

Given a sample set S, an element $\mathbf{x} \in \mathbf{E}_S(f) \setminus S$, the occurrence of $\overline{\mathbf{x}}_{S,f} \neq f(\mathbf{x})$ implies that there are $\mathbf{a},\mathbf{b},\mathbf{c} \in S$:

$$\begin{array}{lll} \mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{x} & \text{and} & solvable(f(\mathbf{a}), f(\mathbf{b}), f(\mathbf{c})) \\ & \text{and} & sol(f(\mathbf{a}), f(\mathbf{b}), f(\mathbf{c})) \neq f(\mathbf{x}). \end{array}$$

This implies that there are $\mathbf{a},\mathbf{b},\mathbf{c} \in \mathbb{B}^m$ such that

$$\begin{array}{lll} \mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{x} & \text{and} & solvable(f(\mathbf{a}), f(\mathbf{b}), f(\mathbf{c})) \\ & \text{and} & sol(f(\mathbf{a}), f(\mathbf{b}), f(\mathbf{c})) \neq f(\mathbf{x}). \end{array}$$

Now, for the sake of contradiction, suppose that

$$f(\mathbf{a}) = g(\mathbf{a}),\ f(\mathbf{b}) = g(\mathbf{b}),\ f(\mathbf{c}) = g(\mathbf{c}) \text{ and } f(\mathbf{x}) = g(\mathbf{x}).$$

In this case, $solvable(g(\mathbf{a}), g(\mathbf{b}), g(\mathbf{c}))$, and thus $sol(g(\mathbf{a}), g(\mathbf{b}), g(\mathbf{c})) = g(\mathbf{x})$ since g is affine. Furthermore, $f(\mathbf{x}) = g(\mathbf{x})$ implies that

$$\begin{array}{rcl} f(\mathbf{x}) & \neq & sol(f(\mathbf{a}), f(\mathbf{b}), f(\mathbf{c})) \\ & = & sol(g(\mathbf{a}), g(\mathbf{b}), g(\mathbf{c})) = g(\mathbf{x}) = f(\mathbf{x}), \end{array}$$

which yields the desired contradiction. As a consenquence, one of the following conditions must hold:

$$\begin{array}{llllll} f(\mathbf{a}) \neq g(\mathbf{a}) & \text{or} & f(\mathbf{b}) \neq g(\mathbf{b}) & & & \\ & \text{or} & f(\mathbf{c}) \neq g(\mathbf{c}) & \text{or} & f(\mathbf{x}) \neq g(\mathbf{x}). \end{array}$$

But the probability of this event is upper bounded by

$$\begin{array}{rl} & P(f(\mathbf{a}) \neq g(\mathbf{a})) + P(f(\mathbf{b}) \neq g(\mathbf{b})) \\ + & P(f(\mathbf{c}) \neq g(\mathbf{c})) + P(f(\mathbf{x}) \neq g(\mathbf{x})) \leq 4\varepsilon. \end{array}$$

Hence, the occurrence of the event ($\mathbf{E}_S(f)$ is unsound) is upper bounded by 4ε, which completes the proof. □

This result tells us that if we can upper bound the distance between f and $\mathcal{L}$ by a small ε, then the probability that no noise will be introduced by using $\mathbf{E}_S(f)$ instead of S in our classifier[4], will be at least $1 - 4\varepsilon$.

From a practical point of view, testing that a function is approximately linear (i.e. affine where the final constant in its ANF is 0) can be checked via a simple test due to Blum, Luby, and Rubinfeld [Blum *et al.*, 1993] also known as ***BLR test***. It can be proved that if BLR test accepts f with a probability at least $1 - 4\varepsilon$, then f is ε-close to $\mathcal{L}$. The test can be performed with $\mathcal{O}(\frac{1}{\varepsilon})$ queries. See [Kaufman *et al.*, 2008] for a comprehensive analysis of the test.

[4]The factor 4 has no link with the VC-dim of an analogical classifier, which is infinite [Hug *et al.*, 2016]. This multiplicative scalar is due to the fact that analogical proportion involves 4 items.

6.4 Relaxing the Bounds

Our definition of approximately sound extension is a rather strict. For a sample S, having a probability $\text{err}_{S,f}$ strictly positive does not mean that the classifier will performs modestly starting from S. For instance, starting with any S, if this probability is in $]0, 0.001]$, then the probability of having $P(\text{err}_{S,f}) > 0$ is 1 despite the algorithm being highly satisfactory. This fact asks for a relaxation of our definition. A bad performance would be more accurately described by $\text{err}_{S,f} > \delta$, for some $\delta \in [0, 1[$.

Definition 8 *We say that $\mathbf{E}_S(f)$ is δ-sound if $err_{S,f} \leq \delta$. Similarly, f is δ-sound if $\forall S \subset \mathbb{B}^m$, $\mathbf{E}_S(f)$ is δ-sound*[5].

If f is δ-sound (for a small δ), then we can consider the analogical inference principle as a valuable inference principle for f since the error rate is quite low. With this view, we are interested in the probability of f not being δ-sound, which then becomes: $\eta_f(\delta) = P(\text{err}_{S,f} > \delta)$. In the case when $\delta = 0$ and f is ε-approximately affine, we have that $\eta_f(0) \leq 4\varepsilon$. In fact, we can prove the following:

Proposition 7 *Let $\varepsilon \in [0, \frac{1}{2}], \delta \in [0, 1]$. If f is ε-approximately affine, then $P(err_{S,f} > \delta) \leq 4\varepsilon \cdot (1 - \delta)$.*

Proof: Given a sample set S and $\mathbf{x} \in \mathbf{E}_S(f) \setminus S$, it follows from Prop. 6 that $P(\overline{\mathbf{x}}_{S,f} \neq f(\mathbf{x})) \leq 4\varepsilon$. For $\text{err}_{S,f} > \delta$ to occur, such an x should lie outside an area of radius δ and the probability of picking such an $\mathbf{x}$ is $1 - \delta$. Obviously if $\delta = 1$, then $\forall f,\ P(\text{err}_{S,f} > 1) = 0$ since $\forall S,\ \text{err}_{S,f} \leq 1$. ☐

6.5 Experiments

In order to investigate whether the bounds in Propositions 6 and 7 are tight, we implemented the following protocol.

For $X = \mathbb{B}^8$, we started from 3 affine functions:

1. $g_1(\mathbf{x}) = x_1$ which is the first projection.
2. $g_2(\mathbf{x}) = x_1 + x_2$ (all variables but x_1, x_2 are irrelevant)
3. $g_3(\mathbf{x}) = \Sigma_{i=1}^m x_i$ where all variables are relevant.

To get an ε-close function, we just flip a random fraction ε of the values of g_i on the universe X. Doing so, we are sure to get a function $f_{i,\varepsilon}$ such that $d(f_{i,\varepsilon}, g_i) \leq \varepsilon$ and then $d(f_{i,\varepsilon}, \mathcal{L}) \leq \varepsilon$. There is no need to go beyond 0.5 since

$$d(f, g) > 0.5 \implies d(f, g + 1) < 0.5,$$

where $g + 1 \in \mathcal{L}$ if $g \in \mathcal{L}$.

For each approximately affine function $f_{i,\varepsilon}$, we performed 100 experiments with a random S of size $\frac{|X|}{2} = 2^7$:

1. For each S, we estimated the probability $\text{err}_{S,f_{i,\varepsilon}}$ by

$$P(\{\mathbf{x} \in \mathbf{E}_S(f_{i,\varepsilon}) \setminus S \,:\, \overline{\mathbf{x}}_{S,f_{i,\varepsilon}} \neq f_{i,\varepsilon}(\mathbf{x})\}),$$

i.e., the proportion of elements $\mathbf{x}$ in $\mathbf{E}_S(f_{i,\varepsilon}) \setminus S$ whose analogical predicted label is not equal to $f_{i,\varepsilon}(\mathbf{x})$.

2. The probability of interest $\eta_{f_{i,\varepsilon}}(\delta) = P(\text{err}_{S,f_{i,\varepsilon}} > \delta)$ is taken as the proportion of errors over the 100 experiments.

[5]Note that f is 0-sound iff f is AP.

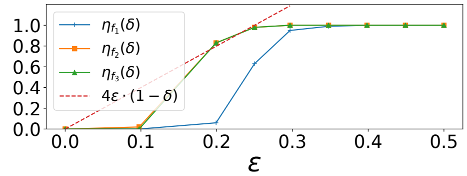

Figure 1: Values of $\eta_{f_i,\varepsilon}(\delta)$for $\delta = 0$

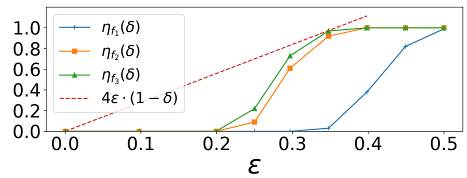

Figure 2: Values of $\eta_{f_i,\varepsilon}(\delta)$for $\delta = 0.3$

In figures 1, 2 and 3, we provide $\eta(\delta)$ as a function of ε for $\delta = 0,\ 0.3,\ 0.5$ respectively. The upper theoretical bound $4\varepsilon \cdot (1 - \delta)$ from Proposition 7 is shown as a dashed line. We observe that for all approximately affine functions f_1, f_2, f_3, the computed probability remains below the theoretical bound. Figures 1 and 2 show that the bounds are tight.

To give an idea of the behavior of 1-NN in the same context, we show (Fig. 4) a comparison between the evolution of $\omega = 1 - \text{err}$ and the proportion τ of properly classified elements in $\mathbf{E}_S(f_{i,\varepsilon}) \setminus S$ by 1-NN wrt ε. As expected, 1-NN becomes better when we move away from affine functions, i.e., the opposite behavior to analogy-based classifier.

7 Conclusion

Analogical inference allows to predict the exact labels of elements outside a given sample set S, provided that the labelling function f is affine. In the case when f is only approximately affine, we provided a probability of null error: it is less than 4 times the distance between the target function f and the set $\mathcal{L}$ of affine functions. As this distance can be evaluated via the BLR test, we have a way to upper bound the error associated with any Boolean function. Our experiments agree with this upper bound, and indicate that the bound may be tight. When relaxing the error requirement to a non null error bounded by δ, we get another theoretical upper bound.

We note that the upper bound does not rely on the size of S, just because the external probability P is considered over

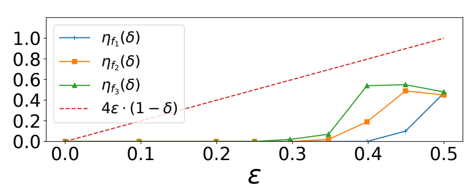

Figure 3: Values of $\eta_{f_i,\varepsilon}(\delta)$for $\delta = 0.5$

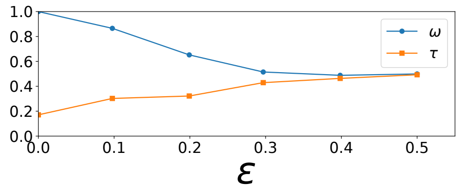

Figure 4: AP versus 1-NN - $m = 8$

the whole set of potential samples S. Introducing the size of S as a parameter could further refine this upper bound.

These results provide clues suggesting why (and when) analogical inference gives good prediction results. Still, there exist known cases of functions f that are far from being affine, but for which the analogical inference works well [Bayoudh *et al.*, 2007]. This is yet to be explained.

Also, in real life applications, we rarely deal with purely Boolean datasets: often they mix Boolean with symbolic attributes taking values in finite sets. A typical example of such an attribute is *color*. Obviously, on the set of values of *color*, there is no structure, no order and no known operator. Binarizing the attribute values, leads us to considering partial Boolean functions, which is a topic of ongoing research.

Acknowledgements

This work is partially supported by ANR-11-LABX-0040-CIMI (Centre International de Mathématiques et d'Informatique) within the program ANR-11-IDEX-0002-02, project ISIPA.

References

[Bayoudh *et al.*, 2007] S. Bayoudh, L. Miclet, and A. Delhay. Learning by analogy: A classification rule for binary and nominal data. In *Proc. 20th Int. Joint Conf. on Artificial Intelligence (IJCAI'07)*, pages 678–683, 2007.

[Blum *et al.*, 1993] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. Comput. Syst. Sci.*, 47(3):549–595, 1993.

[Bounhas *et al.*, 2017] M. Bounhas, H. Prade, and G. Richard. Analogy-based classifiers for nominal or numerical data. *Int. J. Approx. Reasoning*, 91:36–55, 2017.

[Couceiro *et al.*, 2017] M. Couceiro, N. Hug, H. Prade, and G. Richard. Analogy-preserving functions: A way to extend boolean samples. In *Proc. 26th Int. Joint Conf. on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, Aug. 19-25, 2017*, pages 1575– 1581, 2017.

[Davies and Russell, 1987] T. R. Davies and S. J. Russell. A logical approach to reasoning by analogy. In J. P. McDermott, editor, *Proc. 10th Int. Joint Conf. on Artificial Intelligence (IJCAI'87)*, pages 264–270. Morgan Kaufmann, 1987.

[Denecke and Wismath, 2002] K. Denecke and S. L. Wismath. *Universal Algebra and Applications in Theoretical Computer Science*. CRC / C&H, 2002.

[Gust *et al.*, 2006] H. Gust, K. Kühnberger, and U. Schmid. Metaphors and heuristic-driven theory projection (HDTP). *Theoretical Computer Science*, 354(1):98 – 117, 2006.

[Hug *et al.*, 2016] N. Hug, H. Prade, G. Richard, and M. Serrurier. Analogical classifiers: A theoretical perspective. In *Proc. 22nd Europ. Conf. on Artificial Intelligence (ECAI'16)*, pages 689–697. IOS Press, 2016.

[Hwang *et al.*, 2013] S. J. Hwang, K. Grauman, and F. Sha. Analogy-preserving semantic embedding for visual object categorization. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ICML'13, pages III–639–III–647, 2013.

[Jing *et al.*, 2017] 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 (4), pages 120:1–120:15, 2017.

[Kaufman *et al.*, 2008] T. Kaufman, S. Litsyn, and N. Xie. *Breaking the ϵ-Soundness Bound of the Linearity Test over GF(2)*, pages 498–511. Springer Berlin Heidelberg, 2008.

[Klein, 1982] 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.

[Lau, 2006] D. Lau. *Function Algebras on Finite Sets: Basic Course on Many-Valued Logic and Clone Theory*. Springer Monographs in Mathematics. Springer Berlin Heidelberg, 2006.

[Miclet and Prade, 2009] 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)*, pages 638–650. Springer, LNCS 5590, 2009.

[Miclet *et al.*, 2008] 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.

[Prade and Richard, 2013] H. Prade and G. Richard. From analogical proportion to logical proportions. *Logica Universalis*, 7(4):441–505, 2013.

[Russell, 1989] S. J. Russell. *The Use of Knowledge in Analogy and Induction*. Pitman, UK, 1989.

[Stone, 1936] M. H. Stone. The theory of representation for Boolean algebras. *Trans. of the American Mathematical Society*, 40(1):37–111, 1936.

[Stroppa and Yvon, 2005] N. Stroppa and F. Yvon. Analogical learning and formal proportions: Definitions and methodological issues. Technical Report D004, ENST-Paris, 2005.

[Szendrei, 1980] A. Szendrei. On closed classes of quasilinear functions. *Czechoslovak Math. J.*, 30(105):498 – 509, 1980.

[Zhegalkin, 1927] I. I. Zhegalkin. On the technique of calculating propositions in symbolic logic. *Mat. Sb*, 43:9–28, 1927.