Generalizing Analogical Inference from Boolean to Continuous Domains

Francisco Cunha[1], **Yves Lepage**[2], **Miguel Couceiro**[1,3] **Zied Bouraoui**[4],

[1]Dep. Mathematics, IST, U. Lisboa, Portugal
[2]Waseda University, IPS, Japan
[3]INESC-ID, DEI, IST, U.Lisboa, Portugal
[4]CRIL CNRS, U.Artois, France
francisco.vicente.e.cunha@tecnico.ulisboa.pt, yves.lepage@waseda.jp, miguel.couceiro@inesc-id.pt, bouraoui@cril.fr

Abstract

Analogical reasoning is a powerful inductive mechanism, widely used in human cognition and increasingly applied in artificial intelligence. Formal frameworks for analogical inference have been developed for Boolean domains, where inference is provably sound for affine functions and approximately correct for functions close to affine. These results have informed the design of analogy-based classifiers. However, they do not extend to regression tasks or continuous domains. In this paper, we revisit analogical inference from a foundational perspective. We first present a counterexample showing that existing generalization bounds fail even in the Boolean setting. We then introduce a unified framework for analogical reasoning in real-valued domains based on parameterized analogies defined via generalized means. This model subsumes both Boolean classification and regression, and supports analogical inference over continuous functions. We characterize the class of analogy-preserving functions in this setting and derive both worst-case and average-case error bounds under smoothness assumptions. Our results offer a general theory of analogical inference across discrete and continuous domains.

Extended version — https://arxiv.org/pdf/2511.10416

1 Introduction

Analogical reasoning seeks to identify structural similarities between different situations or objects, often formulated through analogical proportions of the form $a : b :: c : d$. Such reasoning has proven effective across domains ranging from case-based reasoning and classification to transfer learning and knowledge graph construction. In particular, it offers an appealing inductive principle in settings where direct supervision is limited, as it allows the inference of labels or attributes from analogically related examples.

The items a, b, c, d are supposed to be described by a set of related attributes and represented by tuples of attribute values. Attributes may be Boolean, nominal (i.e., finite attribute domains), or real-valued (e.g., embeddings). Modeling analogical reasoning has a long tradition in artificial intelligence, particularly in logic (Aristotle 2011; Prade and Richard 2013), cognitive modeling (Gentner 1983; Hofstadter and Sander 2013), and machine learning (Hu et al. 2023). In recent years, interest in formal analogical inference has grown, driven by its applications in few-shot learning (Hwang, Grauman, and Sha 2013), transfer learning (Badra 2020; Cornuéjols, Murena, and Olivier 2020), and interpretable AI (Hüllermeier 2020). Analogical reasoning has also played a central role in natural language processing, particularly in morphological analysis and word embedding evaluation, where analogies such as "king is to queen as man is to woman" have served as both tasks and benchmarks (Mikolov, Yih, and Zweig 2013; Bouraoui, Jameel, and Schockaert 2018; Gladkova, Drozd, and Matsuoka 2016). More recently, neural models have been designed to detect and explain analogical relations in textual data, including analogies between sentences, concepts, and procedures (Ushio et al. 2021; Jacob, Shani, and Shahaf 2023; Kumar and Schockaert 2023).

Foundational results characterizing the class of functions compatible with Boolean analogical inference were established in (Couceiro et al. 2017), and later works extended analogical classification techniques to nominal and numerical domains (Bounhas and Prade 2023; Couceiro et al. 2020). Theoretical results, notably those of (Couceiro et al. 2018), have established that analogical inference is exact (*i.e.,* error-free) if and only if the labeling function is affine. More generally, they showed that if a Boolean function is ε-close to an affine function, then the probability of making an incorrect analogical prediction is bounded above by 4ε, where the probability is taken over the random choice of training sets. This result has been influential in supporting the soundness of analogy-based classifiers in Boolean settings and later extended to a Galois theory of analogical classifiers (Couceiro and Lehtonen 2024).

Despite this progress, two important limitations remain. First, these results are limited to discrete attribute spaces and binary classification tasks. In practice, many real-world applications involve real-valued features and regression tasks, settings that are not covered by current theory. To the best of our knowledge, no prior work has addressed analogical reasoning in the context of regression. Secondly, the theoretical guarantees established in the Boolean setting such as the generalization bound proposed in (Couceiro et al. 2018; Couceiro and Lehtonen 2024) fail to extend to real-valued functions. This limits the application of these results to handle regression.

In this work, we address these shortcomings by revisiting analogical inference from both theoretical and practical perspectives. Firstly, we construct an explicit counterexample demonstrating that the main generalization bound proposed in (Couceiro et al. 2018) fails to hold, even for Boolean functions that are arbitrarily close to affine. This challenges a foundational assumption in the existing theoretical framework. Secondly, we extend analogical inference to real-valued domains by introducing a unified model based on parameterized analogies defined via generalized means. This formulation naturally subsumes classical notions such as additive and geometric analogies and supports analogical reasoning in regression tasks. Thirdly, we introduce functional distances tailored to this generalized setting and derive both uniform and probabilistic bounds on inference error, yielding worst-case and average-case guarantees under smoothness assumptions. Finally, we characterize the class of continuous functions that preserve analogical structures, showing that they correspond to a family of generalized-power functions with well-defined structural properties.

The remainder of this paper is organized as follows. Section 2 reviews foundational work on analogical inference, including Boolean and nominal settings, and discusses the concept of analogy-preserving functions. Section 3 presents a counterexample that falsifies a widely cited generalization bound, even in the Boolean case. Section 4 introduces a generalized framework for analogical proportions based on parameterized means and redefines analogy-preserving functions in continuous domains. In Section 5, we formalize analogical inference for regression, characterize the class of compatible functions, and establish performance guarantees. Section 6 shows how the Boolean case is recovered. We conclude with a summary and outline directions for future research.

2 Background and Related Work

Analogical inference builds on the foundational idea of analogical proportions, quaternary relations of the form $a : b :: c : d$, which express that the transformation from a to b is analogous to that from c to d. This concept has been studied both in logic and learning, where it forms the basis for inference schemes applicable to classification, regression, and relational reasoning. In formal settings, analogical proportions are evaluated componentwise, and their algebraic properties have been studied over Boolean and nominal domains.

2.1 Analogical Inference from Boolean to General Cases

In the Boolean case, the most well-known models of analogy are *Klein's model* (Klein et al. 1983) of Boolean analogy[1] and the so-called *minimal model* (Miclet and Prade 2009)[2].

$$K := \begin{pmatrix} 0&1&1&0&1&0&0&1 \\ 0&1&0&1&0&1&0&1 \\ 0&0&1&1&0&0&1&1 \\ 0&0&0&0&1&1&1&1 \end{pmatrix} \quad M := \begin{pmatrix} 0&1&1&0&0&1 \\ 0&1&0&1&0&1 \\ 0&0&1&0&1&1 \\ 0&0&0&1&1&1 \end{pmatrix}$$

In the Boolean case, the problem of finding an $x \in \{0,1\}$ such that $a : b :: c : x$ holds, does not always have a solution. For instance, neither $0 : 1 :: 1 : x$ nor $1 : 0 :: 0 : x$ has a solution in the minimal model M (since 0111, 0110, 1000, 1001 are not valid patterns for an analogical proportion). In fact, a solution exists if and only if $(a \equiv b) \vee (a \equiv c)$ holds. When a solution exists, it is unique and is given by $x = c \equiv (a \equiv b)$ (see also (Lepage 2023)). This corresponds to the original view advocated by S. Klein (Klein 1983), who however applied the latter formula even to the cases $0 : 1 :: 1 : x$ and $1 : 0 :: 0 : x$, where it yields $x = 0$ and $x = 1$ respectively.

In the nominal case, the situation is similar. The analogical proportion $a : b :: c : x$ may have no solution (*e.g.,* in the minimal model, $s : t :: t : x$ has no solution as soon as $s \neq t$), and otherwise (if $a = b$ or $a = c$) the solution is unique, and is given by $x = b$ if $a = c$ and $x = c$ if $a = b$. Namely, the solutions of $s : t :: s : x$, $s : s :: t : x$, and $s : s :: s : x$ are $x = t$, $x = t$, and $x = s$, respectively. This motivates the following inference pattern, first formalized by (Stroppa and Yvon 2006); see also (Bounhas, Prade, and Richard 2014):

$$\frac{\forall i \in \{1, \ldots, m\},\ \ a_i : b_i :: c_i : d_i \text{ holds}}{a_{m+1} : b_{m+1} :: c_{m+1} : d_{m+1} \text{ holds}} \tag{1}$$

which generalizes analogical inference over attribute vectors enabling to compute d_{m+1}, provided that $a_{m+1} : b_{m+1} :: c_{m+1} : x$ has a solution. This pattern expresses a rather bold inference which amounts to saying that if the representations of four items are in analogical proportion on m attributes, they should remain in analogical proportion with respect to their labels. We can restrict ourselves to binary labels, since a multiple class prediction can be obtained by solving a series of binary class problems. A key question, addressed in this paper, is to characterize the settings under which this inference is valid.

2.2 Analogy-preserving functions and analogy based classification.

The analogical inference pattern implicitly relies on the assumption that labels are functionally determined by the attribute values. More precisely, there exists some unknown function f such that for any item $\mathbf{e} = (e_1, \ldots, e_n)$, the label is given by $e_{n+1} = f(e_1, \ldots, e_n)$. The function f can be viewed of as a classifier that assigns a (unique) class to each item based on its n attributes. Since the solutions of analogical equations (when they exist) are unique, the inference

[1] Here the columns are precisely the tuples (a, b, c, d) for which the analogy holds.

[2] Note that M contains only patterns of the form $x : x :: y : y$ and $x : y :: x : y$, for $x, y \in \{0, 1\}$.

pattern (1) can equivalently be formulated as follows:

$$\begin{array}{cccccc} a_1 & \cdots & a_i & \cdots & a_n & f(\mathbf{a}) \\ b_1 & \cdots & b_i & \cdots & b_n & f(\mathbf{b}) \\ c_1 & \cdots & c_i & \cdots & c_n & f(\mathbf{c}) \\ \hline d_1 & \cdots & d_i & \cdots & d_n & f(\mathbf{d}) \end{array} \qquad (2)$$

where $\mathbf{a} = (a_1, \ldots, a_n)$, $\mathbf{b} = (b_1, \ldots, b_n)$, $\mathbf{c} = (c_1, \ldots, c_n)$ and $\mathbf{d} = (d_1, \ldots, d_n)$ are instances (objects, items, etc.).

Assuming that the latter four instances are in analogical proportion for each of the n attributes describing them, and that the class labels are known for $\mathbf{a}, \mathbf{b}, \mathbf{c}$ but unknown for $\mathbf{d}$, then one may infer that the label for $\mathbf{d}$ as a solution of an analogical proportion equation (Bounhas, Prade, and Richard 2017; Couceiro et al. 2017). The effectiveness of this analogical inference rule led to several studies that aimed to determine which classifiers were compatible with the Analogical Inference Principle (AIP) (Miclet, Bayoudh, and Delhay 2008; Bounhas, Prade, and Richard 2017; Couceiro et al. 2017, 2018, 2020; Bounhas and Prade 2024a; Couceiro and Lehtonen 2024).

In the case of Boolean attributes, a key result has been established in (Couceiro et al. 2017), where it was shown that the set of functions for which analogical inference is sound, i.e., no error occurs, are the *analogy-preserving* (AP) functions, which coincide exactly with the set of affine Boolean functions. More precisely, they showed the following result.

Theorem 1. *The class of AP functions is exactly the class $\mathcal{L}$ of affine functions, i.e., functions $f : \mathbb{B}^n \to \mathbb{B}$ of the form $f = a_1x_1 + \ldots + a_nx_n + b$, for $a_i \in \mathbb{B}$ and where $+$ is the addition over the 2-element field $\mathbb{B}$.*

Moreover, when the function is close to being affine, it was also shown that the prediction accuracy remains high (Couceiro et al. 2018). Assuming a uniform distribution and the Hamming distance d on the Boolean function space $\bigcup_{n>0} \mathbb{B}^{\mathbb{B}^n}$, the classification performance remained high for classifiers close to the AP functions.

Definition 2. *Given a sample set $S \subseteq \mathbb{B}^n$ and a function $f \in \mathbb{B}^{\mathbb{B}^n}$, the analogical root (Hug et al. 2016) of a given element $\mathbf{x} \in \mathbb{B}^n$, denoted by $\mathbf{R}_S(\mathbf{x}, f)$, is the set*

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

The analogical extension $\mathbf{E}_S(f)$ of S w.r.t. f is thus defined as the set of all those $\mathbf{x} \in \mathbb{B}^m$ such that $\mathbf{R}_S(\mathbf{x}, f) \neq \emptyset$.

Clearly, $S \subseteq \mathbf{E}_S(f)$ since $\mathbf{a} : \mathbf{a} :: \mathbf{a} : \mathbf{a}$ always holds. Define $\mathrm{err}_{S,f} = P(\{\mathbf{x} \in \mathbf{E}_S(f) \setminus S \,|\, \overline{\mathbf{x}}_{S,f} \neq f(\mathbf{x})\})$, where $\overline{\mathbf{x}}_{S,f}$ is the predicted label by AIP of $\mathbf{x}$. For classification purposes, we are thus interested in $\mathbf{x} \in \mathbf{E}_S(f) \setminus S$ whose predicted label $\overline{\mathbf{x}}_{S,f}$ is $f(\mathbf{x})$, and $\mathbf{E}_S(f)$ is is said to be *sound* if $\mathrm{err}_{S,f} = 0$.

Theorem 3. *Let $\varepsilon \in [0, \frac{1}{2}]$, and let $\delta \in [0, 1]$. Consider the uniform distance d' on $\mathbb{B}^{\mathbb{B}^m}$ given by*

$$\mathrm{d}(f, g) = \frac{|\{\mathbf{x} \in \mathbb{B}^m \ : \ f(\mathbf{x}) \neq g(\mathbf{x})\}|}{2^m}.$$

If $\mathrm{d}(f, \mathcal{L}) = \min_{g \in \mathcal{L}} \mathrm{d}(f, g) < \varepsilon$, then $P(err_{S,f} > \delta) \leq 4\varepsilon \cdot (1 - \delta)$.

Note that this result establishes a tight connection between inference errors and functional distances.

2.3 Limitations and Drawbacks

When attributes are valued on finite domains X_i, i.e. nominal case (which includes the Boolean case), the problem of characterizing analogy-preserving functions $f : \mathbf{X} \to X$, for $\mathbf{X} = X_1 \times \cdots \times X_n$, has been partially addressed for binary classification ($|X| \leq 2$). This has led to the definition of *hard AP functions* (HAP) $f : \mathbf{X} \to X$, which are either "essentially unary" or "quasi-linear", *i.e.,* for which there exist $\varphi : \{0, 1\} \to X$ and $\varphi_i : X_i \to \{0, 1\}$ ($1 \leq i \leq m$) such that $f = \varphi(\varphi_1(x_1) \oplus \cdots \oplus \varphi_n(x_n))$. However, this characterization is limited to nominal attributes, binary labels and constrained models of analogy, typically those satisfying the patterns (x, x, y, y) and (x, y, x, y). In addition, prior work often assumes a decomposable model of analogy, which may not hold in real-world data. These limitations have been critically examined in recent work, such as (Bounhas and Prade 2024a), which revisits analogical proportions and inference under relaxed assumptions.

These limitations motivate the reformulation of classification tasks using analogy-based regression. In the next section, we construct a counterexample that demonstrates the failure of the error bound from (Couceiro et al. 2018), even in the Boolean setting. We will then introduce an analogical inference principle tailored to regression and continuous domains, enabling extensions to multi-class settings. This framework yields new theoretical guarantees under reasonable smoothness and distributional assumptions.

3 An Illustrative Counter-example

In this section, we provide a counterexample to the generalization bound, which asserts that analogical inference remains accurate for Boolean functions close to the affine class $\mathcal{L}$. Let $\mathbb{B} = \{0, 1\}$, we exhibit a concrete counterexample to Theorem 3 for $\delta = 0$, by providing a Boolean function $f : \mathbb{B}^m \to \mathbb{B}$ such that $d(f, \mathcal{L}) \leq \varepsilon$ but $P(\mathrm{err}_{S,f} > 0) > 4\varepsilon$.

Definition 4. *Let $f : \mathbb{B}^4 \to \mathbb{B}$ be the function given by:*

$$f(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} = \mathbf{1} \\ 0 & \text{otherwise,} \end{cases} \qquad (3)$$

where $\mathbf{1}$ denotes the constant-1 tuple $(1, 1, 1, 1)$. Since the constant-0 Boolean function $h : \mathbb{B}^4 \to \mathbb{B}$ (i.e., $h(\mathbf{x}) = 0$, for every $\mathbf{x} \in \mathbb{B}^4$) is affine, it is not difficult to see that $\mathrm{d}(f, \mathcal{L}) = \min_{g \in \mathcal{L}} \mathrm{d}(f, g) = \mathrm{d}(f, h) = \frac{1}{16}$.

Failure of analogical inference. Let $S \subseteq \mathbb{B}^4 \setminus \{\mathbf{1}\}$, and suppose that $\mathbf{1} \in E_S(f)$. Then $\overline{\mathbf{1}}_{S,f} = 0$, since $f|_S \equiv 0$ and the only solution of $0 : 0 :: 0 : x$ for both Klein's and the minimal models is $x = 0$. This means that if for a given $S \subseteq \mathbb{B}^4$ we have $\mathbf{1} \in \mathbf{E}_S(f) \setminus S$, then $\mathbf{1} \in \{\mathbf{x} \in \mathbb{B}^4 \,|\, \overline{\mathbf{x}}_{S,f} \neq f(\mathbf{x})\}$ (in particular this set is nonempty), which means that $S \in \{T \in \mathcal{P}(\mathbb{B}^4) \,|\, \mathrm{err}_{T,f} > 0\}$. So we get

$$\{S \in \mathcal{P}(\mathbb{B}^4) \,|\, \mathrm{err}_{S,f} > 0\} \supseteq \{S \in \mathcal{P}(\mathbb{B}^4) \,|\, \mathbf{1} \in \mathbf{E}_S(f) \setminus S\}$$

which implies that

$$\{S \in \mathcal{P}(\mathbb{B}^4) \,|\, \mathrm{err}_{S,f} > 0\} \geq \{S \in \mathcal{P}(\mathbb{B}^4) \,|\, \mathbf{1} \in \mathbf{E}_S(f) \setminus S\}.$$

Thus $P\big(\{S \in \mathcal{P}(\mathbb{B}^4) \,|\, \mathrm{err}_{S,f} > 0\}\big)$ is greater or equal to

$$P\big(\{S \in \mathcal{P}(\mathbb{B}^4) \,|\, \mathbf{1} \in \mathbf{E}_S(f) \setminus S\}\big).$$

Algorithm 1: Estimate of the proportion of subsets with analogical error. Here, tupleslo(A, n) outputs all n-tuples over A in lexicographic order, subsets(A) outputs the set of subsets of A, whereas analogyQ receives as an input a quadruple of m-tuples, checks whether each component of the quadruple constitutes an analogy under the minimal model M, and outputs True if so, and False, otherwise.

Input: Integer n (dimension of Boolean vectors)
Output: Estimated probability $P(\text{err}_{S,f} > 0)$

Let $D \leftarrow$ tupleslo$([0, 1], n)$
Let $R \leftarrow$ subsets$(D \setminus \{\mathbf{1}\})$
$c \leftarrow 0$
for each $S \in R$ such that $|S| \geq 3$ **do**
  Let $T \leftarrow$ tupleslo$(S, 3)$
  stop $\leftarrow$ False, $i \leftarrow 0$
  while not *stop* and $i < |T|$ **do**
    stop $\leftarrow$ analogyQ$([T[i]|\mathbf{1}])$
    $i \leftarrow i + 1$
  end while
  if *stop* **then**
    $c \leftarrow c + 1$
  end if
end for
return $c/(2^{2^n})$

Hence,

$$P(\text{err}_{S,f} > 0) \geq \frac{\left|\left\{S \in \mathcal{P}(\mathbb{B}^4) \,\middle|\, \mathbf{1} \in \mathbf{E}_S(f) \setminus S\right\}\right|}{2^{16}}. \quad (4)$$

The first inequality follows from the monotony of P (which is the uniform probability measure over $\mathcal{P}(\mathbb{B}^4)$) and the second from its definition. We can compute this lower bound for $P(\text{err}_{S,f} > 0)$ using a brute force algorithm (Algorithm 1). It consists in checking, for each of the $2^{(2^n-1)}$ (in our case $n = 4$) subsets S_i of $\mathbb{B}^n \setminus \{\mathbf{1}\}$, whether there is any triple $(\mathbf{a}, \mathbf{b}, \mathbf{c}) \in S^3$ such that $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{1}$, and, if so, increasing the running total by 1.[3] Fortunately, since in this example $n = 4$, the algorithm is still manageable and takes about 30 seconds to run. We obtain from Algorithm 1 the lower bound

$$P(\text{err}_{S,f} > 0) \geq 0.42,$$

which contradicts the upper bound

$$P(\text{err}_{S,f} > 0) \leq 4\text{d}(f, \mathcal{L}) = 4\frac{1}{16} = 0.25$$

of Theorem 3 (using $\delta = 0$).

4 Analogy-Preserving Functions

The generalization of analogical reasoning from Boolean to continuous domains requires a more flexible notion of analogy that can capture numeric structure. To address this, we adopt a parameterized framework based on generalized means, originally studied by Hölder (Hölder 1889), and recently proposed for analogical inference in (Lepage and Couceiro 2024). This approach defines analogical proportions over real-valued tuples and supports analogical inference in both classification and regression settings.

4.1 Parameterized Analogical Proportions

For domains where inputs are represented as vectors, matrices or higher-order tensors, the unified view of analogical proportions is based on the *generalized mean in* p:

$$m_p(x_1, x_2, \ldots, x_n) = \lim_{r \to p} \left(\frac{1}{n} \sum_{i=1}^{n} x_i^r \right)^{1/r}. \quad (5)$$

With this, $(a, b, c, d) \in \mathbb{R}_+^4$ constitutes a valid analogy if there is a $p \in \mathbb{R}$ such that $m_p(a, d) = m_p(b, c)$, *i.e.*, the generalized mean in p of the *extremes* a and d equals the generalized mean in p of the *means* b and c.

This is denoted as "analogy in analogical power p" by $a : b ::^p c : d$, and it was shown that it has several desirable properties, in particular, that $::^p$ is transitive and that it constitutes an equivalence relation for $p \in \mathbb{R}$. One of the advantages of relying on this parameterized notion on p, is that it naturally subsumes well known mean notions, such as the commonly used arithmetic (for $p = 1$), geometric (for $p = 0$) or harmonic means (for $p = -1$). As consequence, for any four increasing positive real numbers a, b, c and d there exists a unique analogical power p such that $a : b ::^p c : d$ holds. Notice that such analogy can be reduced to an equivalent arithmetic analogy and that any analogical equation has a solution for increasing numbers.

4.2 Analogical Roots and Extensions

Definition 5. *Let* $\mathbf{x} \in \mathbb{R}_+^n$*, and* $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{R}_+^n$ *and* $q \in \mathbb{R}_+$. *Let* S *be a finite subset of* $\mathbb{R}_+^n$. *The* $(\mathbf{p}; q)$*-analogical root* $R_S(f, \mathbf{x})$ *of* $\mathbf{x}$ *with respect to* f *and* S *is defined as follows:*

$$R_S^{(\mathbf{p};q)}(f, \mathbf{x}) \quad := \quad \{(\mathbf{a}, \mathbf{b}, \mathbf{c}) \in S^3 \,|\, \mathbf{a} : \mathbf{b} ::^{(\mathbf{p})} \mathbf{c} : \mathbf{x} \text{ and } f(\mathbf{a})^q \leq f(\mathbf{b})^q + f(\mathbf{c})^q\}$$

Remark 6. *This definition seems rather different than that of Definition 2. However, it is a clear generalization: indeed, the analogical equation* $a : b ::^q c : d$ *has a solution if and only if* $a^q \leq b^q + c^q$.

Hence $R_S^{(\mathbf{p},q)}(f, \mathbf{x})$ is the set of the triples $(\mathbf{a}, \mathbf{b}, \mathbf{c}) \in S^3$ that form an analogy with $\mathbf{x}$ in powers $\mathbf{p} = (p_1, ..., p_n)$, and such that the analogical equation $f(\mathbf{a}) : f(\mathbf{b}) ::^q f(\mathbf{c}) : y$ can be solved in $\mathbb{R}_+$. We denote the set of all such solutions y by $\text{sol}_q(f(\mathbf{a}), f(\mathbf{b}), f(\mathbf{c}))$.

Definition 7. *Let* $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{R}_+^n$*,* $S \subseteq \mathbb{R}_+^n$ *and* $q \in \mathbb{R}_+$. *The* $(\mathbf{p}; q)$*-analogical extension of* S *with respect to* f*,* $E_S^{(\mathbf{p};q)}(f)$*, is defined as follows:*

$$E_S^{(\mathbf{p};q)}(f) := \{\mathbf{x} \in \mathbb{R}_+^n \,|\, R_S(f, \mathbf{x}) \neq \emptyset\}.$$

When $\mathbf{p}$ *and* q *are clear from the context we shall simply write* $E_S(f)$.

The analogical extension of S is the subset of $\mathbb{R}_+^n$ that can be valued using the Analogical Inference Principle. If $\mathbf{x} \in E_S(f)$, we can assign an *analogical value* to $\mathbf{x}$ (which will ideally coincide with $f(\mathbf{x})$) using the Analogical Inference Principle.

[3]Here $[T[i]|\mathbf{1}]$ is of the form $(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{1})$, for $T[i] = (\mathbf{a}, \mathbf{b}, \mathbf{c})$.

4.3 Characterizing Analogy-Preserving Functions

We now turn to the characterization of functions that preserve analogical proportions under the generalized setting.

Definition 8. *Let* $\mathbf{p}$*,* f *and* S *be as in the previous definition, and let* $\mathbf{x} \in E_S(f)$*. Define the analogical value* $\overline{\mathbf{x}}_{S,f}^{(\mathbf{p};q)}$ *of* $\mathbf{x}$ *w.r.t.* $\mathbf{p}$*,*q*,* f *and* S*, as*

$$m_q\big(sol_q(f(\mathbf{a}), f(\mathbf{b}), f(\mathbf{c})) \mid (\mathbf{a}, \mathbf{b}, \mathbf{c}) \in R_S^{(\mathbf{p};q)}(f, \mathbf{x})\big)$$

where m_q *denotes the* q*-generalized mean. When clear from the context, we will simply write* $\overline{\mathbf{x}}_{S,f}$*.*

Following the same steps as in Section 2, we first seek to describe the set of $AP_{(\mathbf{p};q)}$ of all $(\mathbf{p};q)$-analogy preserving functions, that is, functions f such that for all $S \subseteq \mathbb{R}_+^n$, $\overline{x}_{S,f} = f(\mathbf{x})$, for all $\mathbf{x} \in E_S(f)$.

Proposition 9. *If* f *is continuous, then the following statements are equivalent.*

1. $f \in AP_{(\mathbf{p};q)}$
2. f *maps analogies in powers* $\mathbf{p} = (p_1, ..., p_n)$ *to analogies in power* q*.*

Remark 10. *Observe that the underlying domain is the set of nonnegative real numbers. This might sound as a restriction but a good number of applications meet this condition. A large field of application is image processing where the values on the gray channel (see MNIST data for instance (LeCun et al. 1989)) or RGB channels are nonnegative real number (sometimes even natural numbers between 0 and 255). Any processing of images involving numerical analogy is thus possible on this kind of data. One may think about image completion, image reconstruction, etc. Another example of field where representations are nonnegative real numbers is, at least theoretically, word embedding models. It has been shown that vectors representing words trained from some specific models are concentrated in an orthant of the space, which means that a rotation can make all components in the vectors non-negative (Mimno and Thompson 2017).*

5 Analogy-based Regression

This section explicitly describes the class $AP_{(\mathbf{p};q)}$ of all $(\mathbf{p};q)$-analogy preserving functions under some natural assumptions (see Subsection 5.1), and provides performance guarantees that establish a tight connection between regression errors and distances to the class of analogy preserving functions in various functional spaces (Subsection 5.2). This not only corrects the estimates and performance guarantees provided in (Couceiro et al. 2018), but also generalizes the frameworks of, *e.g.*, (Bounhas, Prade, and Richard 2017; Bounhas and Prade 2023, 2024a; Couceiro et al. 2017, 2018, 2020), to positive reals and to both classification and regression tasks.

5.1 Explicit Description of AP functions

The main result of this section is the following explicit description of continuous analogy preserving functions. We will make use of the following auxiliary result that essentially states that we can restrict our quest to the case of arithmetic analogies (*i.e.,* for $p = 1$).

Lemma 11. *Let* $p'_j, q' \in \mathbb{R}_+$*. Define the mappings* $r : \mathbb{R}_+ \to \mathbb{R}_+$ *such that* $r(x) = \sqrt[q']{x}$*, and* $s : \mathbb{R}_+^n \to \mathbb{R}_+^n$ *such that* $s(x_1, \ldots, x_n) = (x_1^{p'_1}, \ldots, x_n^{p'_n})$*. Then the following statements are equivalent.*

1. $f : \mathbb{R}_+^n \to \mathbb{R}_+$ *maps analogies in powers* $\mathbf{p} = (p_1, \ldots, p_n)$ *to an analogy in power* q*.*
2. $g = r \circ f \circ s$ *maps analogies in powers* $\mathbf{p} \odot \mathbf{p}' = (p'_1 p_1, \ldots, p'_n p_n)$ *to an analogy in power* $q'q$*.*

Lemma 11 is entails the characterization of AP functions.

Theorem 12. *Suppose* $f : \mathbb{R}_+^n \to \mathbb{R}_+$ *is continuous. Then the following statements are equivalent.*

1. $f \in AP_{(\mathbf{p};q)}$*, for* $\mathbf{p} = (p_1, ..., p_n)$*.*
2. $f(x_1, ..., x_n) = \left(\sum_{j=1}^n a_j x_j^{p_j} + b\right)^{1/q}$*, for some matrix* $[a_j] \in M_{1\times n}(\mathbb{R}_+)$ *and some scalar* $b \in \mathbb{R}_+$*.*

Proof. To prove the theorem, we use Proposition 9, and show that f maps analogies in powers $\mathbf{p}$ to analogies in power q if and only if f has the form given in 2.

Applying the previous Lemma, it suffices to show the result for $\mathbf{p} = (1, ..., 1)$ and $q = 1$, which is an application of Cauchy's Functional Equation, which states that a continuous additive function must be linear. ☐

5.2 Performance Guarantees

Inspired by the statement of Theorem 3 that establishes a correspondence between the distance of a Boolean classifier to the class of AP Boolean functions and the probability of that classifier to make classification errors, we seek an analogous (and correct!) result in the setting of analogy based regression. We will make use of background on functional spaces and measure theory and we refer the reader to (Taylor and Kingman 2008) for further background.

Recall first the notion of "q-distance" between $x, y \in \mathbb{R}_+$, that is defined by

$$d_q(x, y) := (|x^q - y^q|)^{\frac{1}{q}}.$$

Formally speaking, this q-distance is not a distance, since it does not fulfill the triangle inequality for $q < 1$. However, it constitutes a semidistance and, as we will see, it is the natural candidate for "distance" when dealing with analogies in power q. The analogous distance for functional spaces can be defined as follows.

Definition 13. *Let* $D \subseteq \mathbb{R}_+^n$*, and consider* $f, g : D \to \mathbb{R}_+$*. Their uniform* q*-distance is defined by*

$$d_{q,\infty}(f, g) := \sup_{\mathbf{x}\in D}(d_q(f(\mathbf{x}), g(\mathbf{x}))).$$

To propose an analogue to the probablistic approach proposed in (Couceiro et al. 2017), we will need to assume that $D \in \mathcal{B}$, where $\mathcal{B}$ is the Borel σ-algebra over $\mathbb{R}_+^n$, and consider a probability space $(D, \mathcal{B}_D, \mathbb{P})$, where $\mathcal{B}_D$ consists of the elements of $\mathcal{B}$ that contain D (the ideal of $\mathcal{B}$ generated by D), and $\mathbb{P} : \mathcal{B}_D \to [0, 1]$ is a probability measure. For instance, if D is finite, $D \in \mathcal{B}$ and we can naturally choose

$\mathbb{P}$ to be the normalized counting measure (*i.e.,* uniform distribution).[4] This additional structure enables us to define our desired probabilistic distance.

Definition 14. *Let $f : D \to \mathbb{R}_+$ be a Borel-measurable function. Its q-expected value is defined as*

$$\mathbb{E}_q(f) := \left(\int_D f^q \, d\mathbb{P} \right)^{\frac{1}{q}}.$$

If $g : D \to \mathbb{R}_+$ is also Borel-measurable, then their probabilistic q-distance is given by

$$dist_q(f, g) := \mathbb{E}_q(x \mapsto d_q(f(x), g(x))).$$

Much like $d_{q,\infty}$, $dist$ is not necessarily a metric on $\mathbb{R}_+^D$, but it is a semimetric, and the most natural notion of distance when working with analogies in power q.

Remark 15. *It is noteworthy that $d_{q,\infty}(f, g)$ simplifies to $(||f^q - g^q||_\infty)^{\frac{1}{q}}$, whereas $dist_q(f, g)$ simplifies to $(\mathbb{E}(|f^q - g^q|))^{\frac{1}{q}}$. Moreover, if $D = \{\mathbf{d}_1, ..., \mathbf{d}_N\}$ and $\mathbb{P}$ is the normalized counting measure, $dist_q(f, g)$ simplifies to $\left(\sum_{i=1}^N |f(\mathbf{d}_i)^q - g(\mathbf{d}_i)^q|\right)^{\frac{1}{q}}$.*

Thus we have constructed two semimetrics, $d_{p,\infty}$ and $dist_p$, that will act as our distances in the functional space $(\mathbb{R}_+)^D$. On the one hand, the semimetric $d_{p,\infty}$ is a uniform distance and will enable us to obtain worst-case results, that is, upper bounds for the largest possible errors when using the AIP with functions $f \notin AP$ (see Proposition 16 and Corollary 17). On the other hand, $dist_p$ is a probabilistic distance and will enable us to obtain average-case results, *i.e.,* upper bounds for the expected value of the errors introduced by when using the AIP with functions $f \notin AP$ (see Proposition 20).

Proposition 16. *Let $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{R}_+^n$, and $f : D \to \mathbb{R}_+$ such that $d_{q,\infty}(f, AP_{(\mathbf{p};q)}) \leq \delta$, if $\mathbf{a} : \mathbf{b} ::^{\mathbf{P}} \mathbf{c} : \mathbf{d}$, then $d_q(f(\mathbf{x}), sol_{\mathbf{p}}(f(\mathbf{a}), f(\mathbf{b}), f(\mathbf{c}))) \leq \sqrt[q]{4}\delta$.*

Proof. The proof follows from simple manipulations that leverage the triangle inequality for the absolute value. For further details see the Appendix. □

In other words, if a function is δ-close to the class $AP_{(\mathbf{p};q)}$, then the regression errors are at most $\sqrt[q]{4}\delta$. The same applies for the other semimetric d_q.

Corollary 17. *Let $f : D \to \mathbb{R}_+$ be such that*

$$d_{q,\infty}(f, AP_{(\mathbf{p};q)}) \leq \delta.$$

Then, for every $x \in D$, $d_q(f(x), \overline{\mathbf{x}}_{S,f}) \leq \sqrt[q]{4}\delta$.

Proof. Follows from Proposition 16 and the monotonicity of the q-generalized mean m_q. □

[4]However, even in the discrete case, this is not the only setting that could be of interest (for example, we could have a domain with a Poisson distribution), so we will aim to develop a general theory that works for any probability space.

5.3 The Probabilistic View

To obtain a probabilistic counterpart to this result, we will need to introduce the notion of regularity for sample sets.

Definition 18. *A sample set $S \subseteq D$ is regular with respect to a function f if $E_S(f) = D$, and there is $m \in \mathbb{N}$, such that*

$$|R_S(f, \mathbf{x})| = m, \text{ for every } \mathbf{x} \in D.$$

Intuitively, a regular sample set is a set whose examples are representative and evenly distributed with respect to the feature space and output function. In such sets, nearby inputs lead to nearby outputs, and neighborhoods used for analogical reasoning are well-behaved: neither too sparse nor irregular. This regularity ensures that analogical inferences are stable and meaningful.

Practically, this assumption plays a role similar to the smoothness and density hypotheses in non-parametric regression: it guarantees that local relations in the data approximate the underlying functional dependencies. Regular sample sets thus provide the theoretical grounding ensuring the soundness and convergence of the analogical rule, and empirically correspond to well-sampled regions where analogical reasoning can be applied reliably.

To such a regular sample set, we can associate a mapping $S' : D \to M_{3\times m}(S)$ that maps each $\mathbf{x}$ to a matrix

$$\begin{bmatrix} a_1^{(\mathbf{x})} & a_2^{(\mathbf{x})} & \cdots & a_m^{(\mathbf{x})} \\ b_1^{(\mathbf{x})} & b_2^{(\mathbf{x})} & \cdots & b_m^{(\mathbf{x})} \\ c_1^{(\mathbf{x})} & c_2^{(\mathbf{x})} & \cdots & c_m^{(\mathbf{x})} \end{bmatrix}$$

where each column is a different element of $R_S(f, \mathbf{x})$.

Remark 19. *If S is regular, then for every $\mathbf{x} \in D$, we have*

$$\overline{\mathbf{x}}_{S,f} = \left(\frac{1}{m} \sum_{j=1}^m \left(f(b_j^{(\mathbf{x})})^q + f(c_j^{(\mathbf{x})})^q - f(a_j^{(\mathbf{x})})^q \right) \right)^{\frac{1}{q}}$$

Moreover, we have $d_q(\overline{\mathbf{x}}_{S,f}, f(\mathbf{x}))$ is equal to

$$\left(\frac{1}{m} \left| \sum_{j=1}^m \left(f(x)^q + f(a_j^{(\mathbf{x})})^q - f(b_j^{(\mathbf{x})})^q - f(c_j^{(\mathbf{x})})^q \right) \right| \right)^{\frac{1}{q}}.$$

The notion of regular sample will enable us to get probabilistic analogue of Proposition 16.

Proposition 20. *Let $S \subseteq D$ be a regular sample set with associated map S', and let $f, g : D \to \mathbb{R}_+$ such that $g \in AP_{(p_1,...,p_n;q)}$ and $dist_p(f, g) \leq \delta$. Suppose that*

$$\mathbb{E}\left(\mathbf{x} \mapsto |f^q - g^q|(S'(\mathbf{x})_{ji})\right) = \mathbb{E}(|f^q - g^q|),$$

for every $\mathbf{x} \in D$, $i \in \{1, ..., m\}$, $j \in \{1, 2, 3\}$[5]. Then

$$dist_q(\mathbf{x} \mapsto \overline{\mathbf{x}}_{S,f}, f) \leq \sqrt[q]{4}\delta.$$

Proof. This proof is a matter of unfolding the definitions, using Remark 15, and then applying the triangle inequality (as in Proposition 16) and the linearity of the expected value to the resulting four summands. The detailed proof is given in the extended version. □

[5]Losely speaking, this means that sampling is independent of the q-distance between f and g.

The assumption that S is regular may seem quite strong. However, the proof of Proposition 20 shows that we do not necessarily need S to be regular, as long as we can find a suitable value for m to construct a $S' : D \to M_{3\times m}(S)$. If we define $\overline{\mathbf{x}}_{S',f}$ as

$$m_q(sol_q(f(S'(\mathbf{x})_{1\,i}, f(S'(\mathbf{x})_{2\,i}), f(S'(\mathbf{x})_{3\,i})) \,|\, i \in \{1, ..., m\})$$

(which is in general different from $\overline{\mathbf{x}}_{S,f}$ if S is not regular, but still obtainable from S by applying the AIP), Proposition 20 is still valid if we substitute $\overline{\mathbf{x}}_{S',f}$ for $\overline{\mathbf{x}}_{S,f}$.

Proposition 20 also assumes that

$$\mathbb{E}\left(\mathbf{x} \mapsto |f^q - g^q|(S'(\mathbf{x})_{ji})\right) = \mathbb{E}(|f^q - g^q|),$$

for every $\mathbf{x} \in D$, $i \in \{1, ..., m\}, j \in \{1, 2, 3\}$. In other words, the way that S' is used to find $\overline{\mathbf{x}}_{S',f}$ does not depend on the distance between f and g. Since we are establishing an upper bound, this assumption can be relaxed to

$$\mathbb{E}\left(\mathbf{x} \mapsto |f^q - g^q|(S'(\mathbf{x})_{ji})\right) \leq \mathbb{E}(\mathbf{x} \mapsto |f^q - g^q|(\mathbf{x})),$$

which essentially states that the sample points we are choosing are more likely to be points where f is closer to g, and the proof would remain the same.

6 Boolean Case Revisited

We can now attempt to apply the previous results by revisiting the Boolean case. It has been shown that (for the two models of analogy M and K), the analogy preserving functions are precisely the affine functions $\mathbb{B}^n \to \mathbb{B}$ (see Theorem 1 that was obtained in (Couceiro et al. 2017)).

However, $\mathbb{B}$ is not a subring of $\mathbb{R}$, so some care is required when transferring these results. To illustrate this, take for example the triple $(0, 1, 1)$. If we see it in $\mathbb{R}$, $sol_1(0, 1, 1) = 2$. But if we see it in $\mathbb{B}$, with model K (see 2.1), $sol_1(0, 1, 1) = 0$. Moreover, if our model is M, this triple is not even solvable. With this in mind, for the sake of clarity, let us denote the sum in $\mathbb{B}$ as $\oplus$ and the sum in $\mathbb{R}$ as $+$.

Another problem we have to deal with is the fact that in $\mathbb{B}$, $\overline{\mathbf{x}}_{S,f} \in \mathbb{B}$ was originally defined as a mode. To make this notion compatible with the framework hitherto established, we will need to modify it slightly. Denote as $\iota : \mathbb{B} \to \mathbb{R}$ the set inclusion. For a regular sample $S \subseteq \mathbb{B}^n$, with corresponding map $S' : D \to M_{3\times m}(S)$, we can define

$$\tilde{\mathbf{x}}_{S,f} := \frac{1}{m} \sum_{i=1}^{m} \iota \left(\bigoplus_{j=1}^{3} f(S'(\mathbf{x})_{ji}) \right) \in [0, 1].$$

Note that $\bigoplus_{j=1}^{3} f(S'(\mathbf{x})_{ji}) = sol(S'(\mathbf{x})_{1i}, S'(\mathbf{x})_{2i}, S'(\mathbf{x})_{3i})$, and that $\overline{\mathbf{x}}_{S,f} \in \mathbb{B}$ can easily be recovered from $\tilde{\mathbf{x}}_{S,f}$. Moreover, if $f \in AP$, then $\tilde{\mathbf{x}}_{S,f} \in \{0, 1\}$ and

$$\tilde{\mathbf{x}}_{S,f} = \iota(\overline{\mathbf{x}}_{S,f}) = \iota f(\mathbf{x}).$$

However, in general $f \notin AP$, and thus $|\tilde{\mathbf{x}}_{S,f} - \iota(\overline{\mathbf{x}}_{S,f})|$ can be thought of as a measure of how confident the AIP classifies $\mathbf{x}$. By observing that, for $a, b \in \mathbb{B}$,

$$\iota(a \oplus b) = |\iota(a) - \iota(b)|,$$

and the inequality

$$\iota(a_1 \oplus \cdots \oplus a_n) \leq \iota(a_1) + ... + \iota(a_n),$$

we can easily adapt the proof of Proposition 20 to get the following (analogous) result.

Proposition 21. *Let $D \subseteq \mathbb{B}^n$, $f, g : D \to \mathbb{B}$ such that $g \in AP = \mathcal{L}$ and $\mathbb{E}(|\iota f - \iota g|) \leq \delta$. Suppose S is regular, and for all $x \in D$, $i \in \{1, ..., m\}$, $j \in \{1, 2, 3\}$,*

$$\mathbb{E}\left(x \mapsto |\iota f - \iota g|(S'(x)_{ji})\right) = \mathbb{E}(|\iota f - \iota g|).$$

Then $\mathbb{E}(|\mathbf{x} \mapsto \tilde{\mathbf{x}}_{S,f} - \iota f|) \leq 4\delta$.

Remark 22. *If the probability measure in use is the normalized counting measure (uniform distribution), $\mathbb{E}(|\iota f - \iota g|)$ is just the (normalized) Hamming distance between f and g, used in (Couceiro et al. 2018).*

The remarks following Proposition 20 also apply here.

7 Conclusion

We have revisited analogical inference from a foundational and unifying perspective. Our analysis reveals that previous generalization bounds for analogical classifiers fail even in the Boolean setting. To address this drawback, we explored a parameterized framework based on generalized means, allowing analogical inference to extend naturally to regression tasks. We characterized analogy-preserving functions in this setting and derived both worst-case and average-case guarantees based on functional distances. This framework bridges the gap between Boolean analogical classification and continuous regression, offering a general theory of analogical reasoning across domains.

While the focus of this paper is theoretical, aiming to invalidate incorrect claims and establish a sound unifying framework, it also sets the stage for empirical evaluation. Our formulation connects with SOTA analogical models (e.g., (Bounhas and Prade 2024b)) yet extends them to real-valued outputs through a principled analogical regression rule with formal guarantees (Prop. 16–20). This provides the first foundation for implementing analogical regression against methods such as kernel or k-NN regression.

Our framework yields implementable analogical regression rules with closed-form prediction. It also identifies the class of analogy-preserving functions (Th. 12), enabling learnable parametric models. These results provide both a unified theoretical basis and concrete algorithmic formulations for analogical reasoning in regression tasks. These highlighted new avenues for future work. As mentioned earlier, one of the limitations is the underlying assumption that the analogy model is decomposable with respect to each dimension. Despite not being of utmost importance for the current framework, it prevents taking into account synergetic relations between dimensions and thus a full account of emerging concepts in the analogical reasoning process. Also, we have considered a unified model of analogies based on generalized means. However, it would be worthwhile to investigate other recent variants, *e.g.,* based on generalized norms (Prade and Richard 2024).

From a practical perspective, future work will explore empirical evaluation, integration with neural architectures, and extensions to structured and probabilistic analogical forms.

8 Acknowledgments

This work was supported by ANR-22-CE23-0002 (ERIANA) and ANR-22-CE23-0023 (AT2TA). Francisco Cunha received support from Calouste Gulbenkian Foundation under "Programa Novos Talentos Científicos" Program. Yves Lepage was supported by a research grant from Waseda University, type Tokutei-kadai, number 2025R-034, entitled "Numerical analogy: mathematical definitions and applications to machine learning".

References

Aristotle. 2011. *Nicomachean Ethics*. University of Chicago Press.

Badra, F. 2020. A Dataset Complexity Measure for Analogical Transfer. In *IJCAI 2020*, 1601–1607.

Bounhas, M.; and Prade, H. 2023. Analogy-based classifiers: An improved algorithm exploiting competent data pairs. *Int. J. Approx. Reason.*, 158: 108923.

Bounhas, M.; and Prade, H. 2024a. Revisiting analogical proportions and analogical inference. *Int. J. Approx. Reason.*, 171: 109202.

Bounhas, M.; and Prade, H. 2024b. Revisiting analogical proportions and analogical inference. *Int. J. Approx. Reason.*, 171: 109202.

Bounhas, M.; Prade, H.; and Richard, G. 2014. Analogical classification: A new way to deal with examples. In *Proc. 21st Europ. Conf. on Artificial Intelligence (ECAI'14)*, 135–140. IOS Press.

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

Bouraoui, Z.; Jameel, S.; and Schockaert, S. 2018. Relation Induction in Word Embeddings Revisited. In *COLING 2018*, 1627–1637. ACL.

Cornuéjols, A.; Murena, P.; and Olivier, R. 2020. Transfer Learning by Learning Projections from Target to Source. In *IDA 2020*, volume 12080 of *LNCS*, 119–131. Springer.

Couceiro, M.; Hug, N.; Prade, H.; and Richard, G. 2017. Analogy-preserving functions: A way to extend Boolean samples. In *IJCAI 2017*, 1575–1581. ijcai.org.

Couceiro, M.; Hug, N.; Prade, H.; and Richard, G. 2018. Behavior of Analogical Inference w.r.t. Boolean Functions. In Lang, J., ed., *IJCAI 2018n*, 2057–2063. ijcai.org.

Couceiro, M.; and Lehtonen, E. 2024. Galois theory for analogical classifiers. *Ann. Math. Artif. Intell.*, 92(1): 29–47.

Couceiro, M.; Lehtonen, E.; Miclet, L.; Prade, H.; and Richard, G. 2020. When Nominal Analogical Proportions Do Not Fail. In Davis, J.; and Tabia, K., eds., *SUM 2020*, volume 12322 of *LNCS*, 68–83. Springer.

Gentner, D. 1983. Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2): 155–170.

Gladkova, A.; Drozd, A.; and Matsuoka, S. 2016. Analogy-based detection of morphological and semantic relations with word embeddings: what works and what doesn't. In *Proceedings of the NAACL-HLT SRW*, 47–54. San Diego, California.

Hofstadter, D.; and Sander, E. 2013. *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*. New York: Basic Books.

Hölder, O. L. 1889. Ueber einen Mittelwerthssatz. *Nachrichten von der königlichen Gesellschaft der Wissenschaft zu Göttingen und der Georg-Augusts-Universität zu Göttingen*, 1(2): 38–47.

Hu, X.; Storks, S.; Lewis, R. L.; and Chai, J. 2023. In-Context Analogical Reasoning with Pre-Trained Language Models. *arXiv preprint arXiv:2305.17626*.

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

Hüllermeier, E. 2020. Towards Analogy-Based Explanations in Machine Learning. In Torra, V.; Narukawa, Y.; Nin, J.; and Agell, N., eds., *MDAI 2020*, volume 12256 of *LNCS*, 205–217. Springer.

Hwang, S. J.; Grauman, K.; and Sha, F. 2013. Analogy-preserving Semantic Embedding for Visual Object Categorization. In *ICML 2013*, volume 28 of *JMLR Workshop and Conference Proceedings*, 639–647.

Jacob, S.; Shani, C.; and Shahaf, D. 2023. FAME: Flexible, Scalable Analogy Mappings Engine. In *EMNLP 2023*, 16426–16442. ACL.

Klein, S. 1983. Analogy and mysticism and the structure of culture (and Comments and Reply). *Current Anthropology*, 24 (2): 151–180.

Klein, S.; Bradshaw, J. L.; Harnad, S.; Ho, D. Y. F.; Holbrook, B.; Paredes, J. A.; Rubinstein, R. A.; TenHouten, W. D.; and Young, J. A. 1983. Analogy and Mysticism and the Structure of Culture [and Comments and Reply]. *Current Anthropology*, 24(2): 151–180.

Kumar, N.; and Schockaert, S. 2023. Solving Hard Analogy Questions with Relation Embedding Chains. In *EMNLP 2023*, 6224–6236. ACL.

LeCun, Y.; Boser, B.; Denker, J. S.; Henderson, D.; Howard, R. E.; Hubbard, W.; and Jackel, L. D. 1989. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4): 541–551.

Lepage, Y. 2023. Formulae for the solution of an analogical equation between Booleans using the Sheffer stroke (NAND) or the Pierce arrow (NOR). In *IARML@IJCAI 2023*, 3–14.

Lepage, Y.; and Couceiro, M. 2024. Towards a unified framework of numerical analogies: Open questions and perspectives. In *IARML@IJCAI 2024*.

Miclet, L.; Bayoudh, S.; and Delhay, A. 2008. Analogical Dissimilarity: Definition, Algorithms and Two Experiments in Machine Learning. *Journal of Artificial Intelligence Research*, 32: 793–824.

Miclet, L.; and Prade, H. 2009. Handling Analogical Proportions in Classical Logic and Fuzzy Logics Settings. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, 638–650. Berlin, Heidelberg: Springer.

Mikolov, T.; Yih, W.-T.; and Zweig, G. 2013. Linguistic Regularities in Continuous Space Word Representations. In *NAACL-HLT 2013*, 746–751. Atlanta, Georgia: ACL.

Mimno, D.; and Thompson, L. 2017. The strange geometry of skip-gram with negative sampling. In *EMNLP 2017*, 2873–2878. Copenhagen, Denmark: ACL.

Prade, H.; and Richard, G. 2013. From Analogical Proportion to Logical Proportions. *Logica Universalis*, 7: 441–505.

Prade, H.; and Richard, G. 2024. Frank's Triangular Norms in Piaget's Logical Proportions. In Destercke, S.; Martinez, M. V.; and Sanfilippo, G., eds., *Scalable Uncertainty Management - 16th International Conference, SUM 2024, Palermo, Italy, November 27-29, 2024, Proceedings*, volume 15350 of *Lecture Notes in Computer Science*, 369–377. Springer.

Stroppa, N.; and Yvon, F. 2006. Du quatrième de proportion comme principe inductif : une proposition et son application à l'apprentissage de la morphologie. *Traitement Automatique des Langues*, 47(2): 1–27.

Taylor, J. C.; and Kingman, J. F. C. 2008. *An Introduction to Measure and Probability*. Cambridge University Press. ISBN 9780521090322.

Ushio, A.; Anke, L. E.; Schockaert, S.; and Camacho-Collados, J. 2021. BERT is to NLP what AlexNet is to CV: Can Pre-Trained Language Models Identify Analogies? In *ACL/IJCNLP 2021, (Volume 1: Long Papers)*, 3609–3624. ACL.