Fonctions PAC-apprenables compatibles avec l'analogie numérique : premiers résultats théoriques et expérimentaux

Francisco Malaca[1], Yves Lepage[2], Miguel Couceiro[1]

[1] INESC-ID, IST, Université de Lisbonne, Portugal
[2] Université Waseda, Japon

francisco.malaca@tecnico.ulisboa.pt,
yves.lepage@waseda.jp,miguel.j.couceiro@tecnico.ulisboa.pt

Résumé

Cet article donne un rappel sur les analogies numériques qui couvrent le domaine du continu comme celui du discret. Il étend l'apprenabilité dans le cadre probablement approximativement correct (PAC) des fonctions préservant l'analogie – établie jusqu'alors seulement pour les contextes booléens, c'est-à-dire discret – au domaine du continu, et fournit un résultat d'apprenabilité de modèles à une dimension compatibles avec le raisonnement par analogie. Les résultats d'une étude empirique de modèles de regression cherchant à approcher ces fonctions sur des ensembles de données synthétiques et réelles confirment les résultats théoriques et fournissent des formulations algorithmiques concrètes pour leur mise en œuvre sur des jeux de données non-triviaux impliquant des traits à valeurs réelles.

Mots-clés

Analogie, moyenne généralisée, apprentissage PAC

Abstract

This article presents a review of the theory of analogies linking discrete and continuous domains. It extends learnability within the probably approximately correct (PAC) framework of analogy-preserving functions—previously established only for Boolean contexts—to continuous domains, and provides partial learnability results for models compatible with analogical reasoning. It also presents an empirical study of models approximating these functions on synthetic and real data sets. The results consolidate the theoretical foundations of analogical regression and provide concrete algorithmic formulations for its implementation in non-trivial real-valued environments.

Keywords

Analogy, generalized means, PAC learnability

1 Introduction

Le raisonnement analogique est un mécanisme inductif puissant, largement utilisé dans la cognition humaine et de plus en plus appliqué en intelligence artificielle. Ce type de raisonnement trouve ses racines dans la logique [3, 32]. Il a été étudié en psychologie [14] et en intelligence artificielle [16] pour la modélisation cognitive, l'apprentissage automatique [17] ou la traduction automatique [27]. Ces dernières années, l'intérêt pour une formalisation de l'inférence par analogie, introduite en [23] et en [24][1], s'est accru, stimulé par les applications en apprentissage avec peu d'exemples [19], en apprentissage par transfert [4], en raisonnement à partir de cas [28, 5], et même dans l'initialisation de graphes de connaissances [21, 30] ou en explicabilité de l'intelligence artificielle [18]. L'analogie a joué un rôle important en traitement automatique des langues dans l'évaluation des plongements lexicaux [29, 8] au moyen de bancs d'essai [15]. Plus récemment, des modèles neuronaux ont été conçus pour détecter et expliquer des analogies dans les données textuelles, y compris entre phrases, concepts ou procédures [33, 20, 22, 35, 36].

Des cadres formels pour l'inférence par analogie ont été développés pour le domaine booléen où la validité de l'inférence a été démontrée pour les fonctions affines et montrée approximativement correcte pour les fonctions proches des fonctions affines [9]. Ces cadres on été étendus aux domaines nominaux et numériques [7, 12].

Des résultats théoriques, notamment ceux de [10], ont établi que l'inférence par analogie est exempte d'erreurs si et seulement si la fonction d'étiquetage est affine. Ce résultat qui établit la validité des classificateurs fondés sur l'analogie pour le cas booléen a ensuite été étendu à une théorie de Galois des classificateurs analogiques [11]. Toujours dans le cas booléen, il a récemment été démontré [1] que les analogies proportionnelles sont efficacement apprenables dans le cadre PAC [34]. Ce résultat a été testé en pratique dans un scénario perceptif avec des images à quatre cellules ayant chacune une forme et une couleur. Cependant, aucun des cadres précédents ne s'étend directement aux tâches de régression sur des domaines continus.

Cette limitation a été étudiée dans [13] où les fondements de l'inférence par analogie on été réexaminés. Une démonstration par l'exemple donne les limites d'une formalisation précédente et réfute certaines limites de généralisation pour les classificateurs analogiques qui peuvent échouer même

1. Sous le nom d'homomorphisme entre structures analogiques, *op. cit.*, p. 191-205.

dans le cadre booléen. Techniquement cet article utilise le cadre des analogies numériques fondées sur les moyennes généralisées qui permet d'étendre naturellement l'inférence par analogie à la régression. Les fonctions préservant l'analogie dans ce cadre à valeurs réelles ont ainsi pu être caractérisées et des garanties fondées sur les distances fonctionnelles, pour le pire des cas et le cas en moyenne, ont pu être dérivées, comblant ainsi le fossé entre la classification analogique dans le cas booléen et la régression dans le cas continu. Ces résultats offrent une théorie générale du raisonnement analogique entre deux domaines et ils fournissent des garanties théoriques sur les performances de l'inférence fondée sur l'analogie. Cela fournit un cadre pour une évaluation empirique rigoureuse qui a permis, par exemple, de jeter les bases de la mise en œuvre de la régression par analogie et de comparer à la régression par noyau ou à la méthode des plus proches voisins. Elle a également permis d'obtenir des expressions analytiques des règles de régression par analogie implémentables. En résumé, elle fournit à la fois une base théorique unifiée et des formulations algorithmiques concrètes pour le raisonnement analogique dans les tâches de régression.

Dans cet article, nous nous appuyons sur toutes ces dernières avancées pour explorer de nouveaux domaines d'application et essayer d'étendre l'apprenabilité PAC des fonctions préservant les analogies au cas du continu. Cet article apporte trois contributions constituant chacune un paragraphe entier.

- Le paragraphe 2 rappellera brièvement la notion d'analogie numérique fondée sur les moyennes généralisées [25, 26] et définira formellement le principe d'inférence par analogie pour cette notion.
- À la suite des travaux récents cités plus haut, comme [1], le paragraphe 3 examinera la question de savoir si la classe des fonctions préservant l'analogie dans le cas continu introduite en [13] est apprenable dans le cadre PAC. Il présentera un premier résultat partiel qui répond par l'affirmative à cette question (Corollaire 3).
- Le paragraphe 4 présentera une étude empirique sur des ensembles de données synthétiques et réels avec des modèles qui approchent les fonctions préservant l'analogie identifiées dans le résultat précédent.

2 Prérequis théoriques

Cet article s'intéresse à des domaines continus, et non pas à des domaines discrets, typiquement booléens. Le problème qui y est abordé est celui de la régression considérée comme instanciation du principe d'inférence par analogie, plutôt que celui de la simple classification. La notion d'analogie pertinente à cette fin est donc celle d'analogie numérique fondées sur les moyennes généralisés introduite en [25] et dont il a été montré qu'elle étend, en particulier, le cas booléen.

2.1 Moyennes généralisées et analogie

La moyenne généralisée en $p \in \mathbb{R}$ de n nombres réels positifs appartenant à $\mathbb{R}_+^*$, $x_1, \ldots, x_N$, est définie par :

$$m_p(x_1, \ldots, x_N) = \lim_{r \to p} \sqrt[r]{\frac{1}{N} \sum_{i=1}^{N} {x_i}^r}. \tag{1}$$

Cette définition recouvre des moyennes classiques. Pour $p = 1$, elle correspond à la moyenne arithmétique, pour $p = 0$ à la moyenne géométrique, pour $p = 2$ à la moyenne quadratique.

Quatre nombres de $\mathbb{R}_+^*$ a, b, c et d, sont dits en analogie arithmétique si $a - b = c - d$, c'est-à-dire si $\frac{1}{2}(a + d) = \frac{1}{2}(b + c) \Leftrightarrow m_1(a, d) = m_1(b, c)$; ils sont dits en analogie en p, ce que l'on note $a : b ::^p c : d$, si $m_p(a, d) = m_p(b, c)$, ou, de manière équivalente, $a^p + d^p = b^p + c^p$ pour $p \neq 0$.

Par conséquent, pour tous $(a, b, c, d) \in \left(\mathbb{R}_+^*\right)^4$, la relation satisfait

— $a : b ::^p c : d \implies c : d ::^p a : b$, et
— $a : b ::^p c : d \implies a : c ::^p b : d$.

Sans perte de généralité, on peut supposer $a \leq b \leq c \leq d$.

Par extension, on dira que quatre vecteurs de $\mathbb{R}_+^n$, $\mathbf{a} = (a_1, \ldots, a_n)$, $\mathbf{a} = (b_1, \ldots, b_n)$, $\mathbf{c} = (c_1, \ldots, c_n)$, et $\mathbf{d} = (d_1, \ldots, d_n)$ sont en relation d'analogie en le vecteur $\mathbf{p} = (p_1, \ldots, p_n)$, si $m_{p_i}(a_i, d_i) = m_{p_i}(b_i, c_i)$, pour tous les i de 1 à n.

Cela posé, pour un $\mathbf{p}$ de $\mathbb{R}^n$ fixé et pour un q fixé de $\mathbb{R}$, le principe d'inférence par analogie s'énonce de la façon suivante :

$$\frac{\mathbf{a} : \mathbf{b} ::^{\mathbf{p}} \mathbf{c} : \mathbf{x}, \quad f(\mathbf{a}) : f(\mathbf{b}) ::^q f(\mathbf{c}) : y}{f(\mathbf{x}) = y}. \tag{2}$$

Dit autrement :

$$\left.\begin{array}{c} m_{\mathbf{p}}(\mathbf{a}, \mathbf{x}) = m_{\mathbf{p}}(\mathbf{b}, \mathbf{c}) \\ m_q(f(\mathbf{a}), y) = m_q(f(\mathbf{b}), f(\mathbf{c})) \end{array}\right\} \implies f(\mathbf{x}) = y. \tag{3}$$

2.2 Fonctions préservant l'analogie

On dira qu'une fonction préserve l'analogie si les prédictions fondées sur le principe d'inférence par analogie n'induisent pas en erreur. Comme on ne s'attend pas à ce que cela soit vrai pour toute application de ce principe, on vise à réduire le nombre d'erreurs en faisant la moyenne de plusieurs prédictions fondées sur le principe d'inférence par analogie. On formalisera donc le concept de fonction préservant l'analogie en commençant par déterminer quand il est possible d'avoir une solution.

$f(\mathbf{a}) : f(\mathbf{b}) ::^q f(\mathbf{c}) : \mathbf{s}$ a une solution si et seulement si

$$f(\mathbf{b})^q + f(\mathbf{c})^q - f(\mathbf{a})^q \geq 0. \tag{4}$$

Dans ce cas la solution est

$$\sqrt[q]{f(\mathbf{b})^q + f(\mathbf{c})^q - f(\mathbf{a})^q}. \tag{5}$$

Définition 1 (Ensemble des racines analogiques [13])**.** *Soient* $\mathbf{x} \in (\mathbb{R}_+^*)^n$, $\mathbf{p} = (p_1, \ldots, p_n) \in (\mathbb{R}_+^*)^n$, $q \in \mathbb{R}_+^*$ *et soit* S *un sous-ensemble fini de* $(\mathbb{R}_+^*)^n$*. L'ensemble des racines analogiques en* $(\mathbf{p}; q)$ *de* $\mathbf{x}$ *relativement à* S *et* f*, noté* $R_S^{(\mathbf{p};q)}(f, \mathbf{x})$*, est l'ensemble*

$$\{\ (\mathbf{a}, \mathbf{b}, \mathbf{c}) \in S^3 \mid \mathbf{a} : \mathbf{b} ::^{\mathbf{p}} \mathbf{c} : \mathbf{x} \ \textit{et}\ f(\mathbf{b})^q + f(\mathbf{c})^q - f(\mathbf{a})^q \geq 0\ \}.$$

On appellera $sol_S^{(\mathbf{p},q)}(f,\mathbf{x})$ l'ensemble des solutions de l'équation (5) pour tous les $(\mathbf{a},\mathbf{b},\mathbf{c})$ de $R_S^{(\mathbf{p};q)}(f,\mathbf{x})$.

$$sol_S^{(\mathbf{p},q)}(f,\mathbf{x}) = \Big\{ \sqrt[q]{f(\mathbf{b})^q + f(\mathbf{c})^q - f(\mathbf{a})^q}, \quad \forall (\mathbf{a},\mathbf{b},\mathbf{c}) \in R_S^{(\mathbf{p};q)}(f,\mathbf{x}) \Big\}$$

Définition 2 (Extension analogique [13])**.** *Soient* $\mathbf{p} = (p_1,\dots,p_n) \in (\mathbb{R}_+)^n$, $q \in \mathbb{R}_+^*$ *et* $S \subseteq \mathbb{R}_+^n$. *On définit l'extension analogique en* $(\mathbf{p};q)$ *de* S *relativement à* f, *notée* $E_S^{(\mathbf{p};q)}(f)$, *par*

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

Définition 3 (Valeur estimée par analogie [13])**.** *Soient* $\mathbf{p}$, q, f *et* S *comme dans la définition précédente. Soit* $\mathbf{x} \in E_S^{(\mathbf{p};q)}(f)$. *On définit la valeur estimée de* $\mathbf{x}$ *par analogie relativement à* $\mathbf{p}$, q, S *et* f, *notée* $\bar{\mathbf{x}}_{S,f}^{(\mathbf{p};q)}$, *par la moyenne généralisée en* q *des solutions de* (5) *:*

$$\bar{\mathbf{x}}_{S,f}^{(\mathbf{p};q)} = m_q(sol_S^{(\mathbf{p};q)}(f,\mathbf{x}))$$

Nous détaillons ci-dessous la signification des notions données par les définitions 1, 2 et 3.

- $R_S^{(\mathbf{p};q)}(f,\mathbf{x})$ est l'ensemble des triplets $(\mathbf{a},\mathbf{b},\mathbf{c}) \in S^3$ tels que $\mathbf{a} : \mathbf{b} ::^{\mathbf{p}} \mathbf{c} : \mathbf{x}$ qui vérifient l'équation (4), c'est-à-dire ceux qui permettront d'obtenir une solution.
- $E_S^{(\mathbf{p};q)}(f)$ est l'ensemble des $\mathbf{x}$ de $(\mathbb{R}_+^*)^n$ tels que $R_S^{(\mathbf{p};q)}(f,\mathbf{x}) \neq \emptyset$, autrement dit tels que l'ensemble des solutions est non vide. C'est donc l'ensemble des $\mathbf{x}$ pour lesquels on est assuré de pouvoir estimer une valeur.
- $\bar{\mathbf{x}}_{S,f}^{(\mathbf{p};q)}$ est la moyenne généralisée en q de toutes les solutions de l'équation $f(\mathbf{a}) : f(\mathbf{b}) ::^q f(\mathbf{c}) : \mathbf{s}$, dans laquelle $(\mathbf{a},\mathbf{b},\mathbf{c}) \in R_S^{(\mathbf{p};q)}(f,\mathbf{x})$ et $\mathbf{x} \in E_S^{(\mathbf{p};q)}(f)$.

Équipé de toutes les notions précédentes, il est possible de définir les fonctions préservant l'analogie en posant qu'elles doivent vérifier $f(\mathbf{x}) = \bar{\mathbf{x}}_{S,f}^{(\mathbf{p};q)}$.

Définition 4 (Fonctions préservant l'analogie)**.** *L'ensemble des fonctions préservant l'analogie en* $(\mathbf{p};q)$, *noté* $\mathcal{F}^{PA}(\mathbf{p},q)$, *est exactement l'ensemble des fonctions* f *telles que* $\bar{\mathbf{x}}_{S,f}^{(\mathbf{p};q)} = f(\mathbf{x})$, *pour tout* $\mathbf{x} \in E_S^{(\mathbf{p};q)}(f)$ *et pour tout* S *sous-ensemble de* $(\mathbb{R}_+^*)^n$.

Nous introduisons maintenant le résultat concernant les fonctions préservant l'analogie en un $(\mathbf{p};q)$ donné.

Théorème 1 ([13])**.** *Soit* f *une fonction continue de* $\mathbb{R}_+^n$ *dans* $\mathbb{R}_+^*$. *Soient* $\mathbf{p} = (p_1,\dots,p_n) \in (\mathbb{R}_+^*)^n$ *et* $q \in \mathbb{R}_+^*$. *Alors les deux propositions suivantes sont équivalentes.*

1. $f \in \mathcal{F}^{PA}(\mathbf{p},q)$.
2. $f(x_1,\dots,x_n) = \left(\sum_{j=1}^n w_j x_j^{p_j} + b\right)^{1/q}$ *pour une matrice de poids* $[w_j] \in M_{1\times n}(\mathbb{R}_+)$ *et un scalaire de biais* $b \in \mathbb{R}_+^*$.

Un certain nombre de remarques peuvent être faites qui établissent le lien avec des situations bien connues en apprentissage automatique.

D'abord, toute fonction linéaire de $\mathbb{R}_+^*$ dans $\mathbb{R}_+$ de coefficients positifs est un cas particulier de fonction préservant l'analogie puisque c'est une fonction préservant l'analogie en $(1;1)$. Ensuite, une fonction préservant l'analogie en $(p;1)$, avec p scalaire, plus précisément $p \in \mathbb{R}_+$, est la composition de deux fonctions, à savoir l'élévation à la puissance p suivie d'une fonction linéaire de $\mathbb{R}_+^*$ dans $\mathbb{R}_+$. Les deux remarques précédentes combinées montrent que l'application du principe d'inférence par analogie est une généralisation de la régression linéaire.

On peut aussi comparer les fonctions préservant l'analogie à la méthode des k plus proches voisins, l'idée étant que le principe d'inférence par analogie peut s'appliquer de façon restrictive au voisin de l'objet dont on veut estimer la valeur. Si, pour prédire $f(x)$, on se restreint à des triplets (a,b,c) dont les éléments sont proches de x, alors la solution de l'équation analogique donnée par la formule (5) est obtenue à partir des valeurs d'objets similaires à x. Mais il faut noter que, contrairement à le méthode des k plus proches voisins, la sortie n'est pas la moyenne *arithmétique* des valeurs, mais la moyenne *généralisée en* q de ces valeurs.

3 Apprenabilité dans le cadre PAC pour une dimension

Dans le paragraphe entier qui suit, et afin de coller au champ de l'apprenabilité dans le cadre probablement approximativement correct (PAC), nous nous limitons à l'ensemble des fonctions de domaine $[0,1]^n$ et d'image (ou codomaine) $[0,1]$.

Définition 5. *On appellera* $\mathcal{F}^n$ *l'ensemble des fonctions* $f : [0,1]^n \to [0,1]$ *définies par*

$$f(x_1,\dots,x_n) = \left(\sum_{j=1}^n w_j x_j^{p_j} + b\right)^{1/q}$$

avec $(b,q,w_j,p_j) \in (\mathbb{R}_+^*)^{(2+2\times n)}$.

Pour les définitions et les résultats concernant l'apprenabilité dans le cadre PAC, nous nous appuyons sur [2] aussi bien pour les résultats qui y sont originaux que pour ceux démontrés par ailleurs.

Définition 6 (Perte)**.** *Soit* $\mathcal{H}$ *une classe de fonctions*[2] *à valeurs dans* $[0,1]$ *définie sur un domaine* X. *Pour tout* h *de* $\mathcal{H}$, *on définit la fonction de perte* l_h *de la façon suivante.*

$$\begin{aligned} l_h : X \times [0,1] &\to [0,1] \\ (\mathbf{x},y) &\mapsto (h(\mathbf{x}) - y)^2 \end{aligned}$$

Définition 7 (Procédure d'apprentissage)**.** *Une procédure d'apprentissage est une application* M *d'une suite finie d'éléments de* $X \times [0;1]$ *dans* $\mathcal{H}$, *classe de fonctions à valeurs dans* $[0;1]$. *Elle produit une hypothèse* $\hat{h} = M((\mathbf{x}_i,y_i)_{i\in\{1,\dots,m\}})$ *pour tout échantillon*

2. $\mathcal{H}$ pour hypothèses.

$(\mathbf{x}_i, y_i)_{i\in\{1,\dots,m\}}$, *où les* $(\mathbf{x}_i, y_i)$ *sont des couples de* $X \times [0,1]$.

Définition 8 (Apprenabilité à ε près)**.** *Soit* $\varepsilon \in \mathbb{R}_+^*$. *On dira que* $\mathcal{H}$ *est apprenable à* ε *près s'il existe une procédure d'apprentissage* M *telle que*

$$\lim_{m\to\infty} \sup_{P\in\Pi} \mathbb{P}\big(P(l_{M((\mathbf{x}_i,y_i)_{i\in\{1,\dots,m\}})}) > \inf_{h\in\mathcal{H}} P(l_h) + \varepsilon\big) = 0.$$

$\mathbb{P}$ *est la probabilité relativement à l'échantillon, avec les* $(\mathbf{x}_i, y_i)$ *tirés indépendamment selon une distribution* P *sur* $X \times [0;1]$ *et le sup est pris sur toutes les distributions.*

On a apprenabilité si la procédure M est capable d'identifier la meilleure hypothèse h parmi les éléments de $\mathcal{H}$ avec une marge d'erreur de ε quand elle dispose de suffisamment de données d'entraînement (m). Si la taille des données d'entraînement est infini, le sup sur l'ensemble des distributions de données de la probabilité de ne pas trouver la meilleure hypothèse doit être nul.

Définition 9 (Apprenabilité)**.** $\mathcal{H}$ *est apprenable si et seulement s'il est apprenable à* ε *près quel que soit* $\varepsilon \in \mathbb{R}_+^*$ *(c'est-à-dire quel que soit* $\varepsilon > 0$*).*

Après avoir présenté les définitions relatives à l'apprenabilité, nous passons maintenant aux conditions suffisantes pour l'apprenabilité.

Définition 10 (Déchirure avec une marge γ [2])**.** *Soit* $\gamma \in \mathbb{R}_+^*$ *(c'est-à-dire* $\gamma > 0$*) que l'on appellera* marge. *On dira que* $\mathcal{H}$ *sépare un ensemble* $B \subseteq X$ *avec une marge* γ *s'il existe une constante* $\alpha \in \mathbb{R}$ *telle que, pour tout sous-ensemble* E *de* B *il existe une fonction* h_E *de* $\mathcal{H}$ *qui satisfasse :*

$$\begin{cases} h_E(\mathbf{x}) \geq \alpha + \gamma & \text{si } \mathbf{x} \in E, \\ h_E(\mathbf{x}) < \alpha - \gamma & \text{si } \mathbf{x} \in B \setminus E. \end{cases}$$

Autrement dit, la fonction h_E attribue à x une valeur à droite de α avec une marge de γ s'il appartient à B et une valeur à gauche de α avec une marge de γ dans le cas contraire.

Définition 11 (Dimension de Vapnik [2])**.** *La dimension de Vapnik de* $\mathcal{H}$ *(définie sur* X*) avec une marge de* γ *est définie comme le maximum des cardinaux des sous-ensembles de* X *séparables avec une marge de* γ.

Théorème 2 (Thèoreme 4.1. de [2])**.** *Si la dimension de Vapnik d'un ensemble* $\mathcal{H}$ *est finie pour tout* $\gamma > 0$*, alors* $\mathcal{H}$ *est apprenable dans le cadre PAC.*

Autrement dit, on a apprenabilité si la dimension de Vapnik n'est pas infinie.
Le théorème précédent permet de démontrer l'apprenabilité des fonctions préservant l'analogie en l'appliquant à l'ensemble de fonctions hypothèses $\mathcal{H} = \mathcal{F}^n$ dont la définition a été donnée au tout début de ce paragraphe. Ici, on ne considérera que le cas particulier de $n = 1$, c'est-à-dire le cas à une dimension. On montre d'abord la finitude de la dimension de Vapnik, ce qui constitue le lemme ci-dessous, pour en déduire par application du théorème 2, le corollaire d'après sur l'apprenabilité des fonctions de $\mathcal{F}^1$.

Lemme 1. *La dimension de Vapnik de* $\mathcal{F}^1$ *est égale à 1 pour tout* $\gamma > 0$.

Démonstration. On montre qu'aucune fonction de $\mathcal{F}^1$ ne peut déchirer, avec une marge de γ, un ensemble de deux éléments. Soient donc $B = \{x_1, x_2\}$, avec $x_1 < x_2$, un tel ensemble quelconque et $\gamma > 0$. Comme toute fonction dans $\mathcal{F}^1$ est croissante, il n'existe aucune fonction $f \in \mathcal{F}^1$ satisfaisant à la fois $f(x_1) \geq \alpha + \gamma$ et $f(x_2) \leq \alpha - \gamma$, car cela impliquerait $f(x_1) > f(x_2)$. □

Corollaire 3. $\mathcal{F}^1$ *est apprenable dans le cadre PAC.*

Par définition, le corollaire 3 implique l'existence d'une procédure qui trouve, avec une marge d'erreur donnée, l'une des fonctions de $\mathcal{F}^1$ qui correspond le mieux aux données. Comme il s'agit de fonctions de $\mathbb{R}_+^*$ dans $\mathbb{R}_+^*$, dans les expériences que nous allons maintenant présenter, un seul des traits des jeux de données sera utilisé pour la prédiction.

4 Expériences

En nous appuyant sur les résultats rappelés ou obtenus aux paragraphes précédents, nous nous proposons maintenant d'évaluer deux types de modèles de régression avec valeurs scalaires en entrée comme en sortie.

4.1 Modèles de régression

Pour créer des modèles de régression conformes aux fonctions de $\mathcal{F}^1$, il est nécessaire de déterminer les paramètres w, p, b et q. Comme la pente du modèle de régression peut être positive ou négative, on choisira $(wx^p + b)^{1/q}$ ou $(w(1-x)^p + b)^{1/q}$ selon le cas. conformément à ce qui a été vu en 2.2, pour prédire la sortie de x', on cherchera donc des triplets (a, b, c), tels que $a : b ::^p c : x'$ et tels qu'il existe une solution s à l'équation analogique $f(a) : f(b) ::^q f(c) : s$. L'application du principe d'inférence par analogie à plusieurs de ces triplets permettra le calcul de la moyenne généralisée en q de toutes les valeurs de s trouvées.
Nous testons deux stratégies possibles pour le choix des triplets. La première stratégie choisit a et b de manière aléatoire, tandis que la seconde sélectionne les voisins les plus proches de x. Pour les deux stratégies, le terme c est pris comme le ou les voisins les plus proches de $(a^p - b^p + (x')^p)^{1/p}$.
Comme nous avons affaire à des variables scalaires, au lieu d'évaluer si $a : b ::^p c : d$ est vrai, après avoir réorganisé les termes de manière à ce que $a \leq b \leq c$ et $a \leq d$, nous vérifions si le terme $A(a, b, c, d)$ défini comme suit

$$\begin{aligned} \Delta_p(a,b,c,d) &= \begin{cases} |(a^p - b^p) - (c^p - d^p)|, \text{ si } c \leq d \\ \max\{|a^p - b^p|, |c^p - d^p|\}, \text{ sinon} \end{cases} \\ \text{étendue}_p(X) &= \big(\max(X)\big)^p - \big(\min(X)\big)^p \end{aligned}$$

$$A_p(a,b,c,d) = 1 - \frac{\Delta_p(a,b,c,d)}{\text{étendue}_p(X)} \tag{6}$$

est supérieur à un certain seuil, θ, où X est l'ensemble d'apprentissage.[3] Pour un p donné, Δ_p est une dissimilarité qui évalue à quel point ses quatre arguments a, b, c et d sont près de former une analogie en p. Cette dissimilarité se normalise par l'« étendue » des valeurs possibles dans X, élevée à la puissance p. Dans le cas où les arguments sont proches de former une analogie en p en prenant en compte le contexte X, A_p sera proche de 1. On souhaite s'assurer que cette proximité est grande en utilisant un seuil θ inférieur à 1, mais très proche de 1. Autrement dit, une fois a, b et c choisis, nous vérifions après réordonnancement que

$$A_p(a,b,c,x') > \theta. \tag{7}$$

Algorithm 1 Sélection par tirage aléatoire ou plus proches voisins

Require: X ; y ; A ; θ ; k or s ; x'
  pente, _ = LinearRegression(X, y)
  if pente < 0 **then**
    p, q = CurveFit$(\lambda x.(w(1-x)^p + b)^{1/q}, 1 - X, y)$
  else
    p, q = CurveFit$(\lambda x.(wx^p + b)^{1/q}, X, y)$
  end if
  $sol_q = \emptyset$
  if cas tirage aléatoire **then**
    $IJ \leftarrow$ RandomChoice$(\lfloor s \times |X| \rfloor)$
  else if cas plus proches voisins **then**
    $IJ \leftarrow$ NearestNeighbors(x', k)
  end if
  for $i, j \in IJ$ **do**
    $a, f(a) \leftarrow X(i), y_s(i)$
    $b, f(b) \leftarrow X(j), y_s(j)$
    $\{l\} \leftarrow$ NearestNeighbors$((a^p - b^p + (x')^p)^{1/p}, 1)$
    $c, f(c) \leftarrow X(l), y_s(l)$
    if inégalités (4) and (7) vraies **then**
      $sol_q \leftarrow sol_q \cup \{$ solution calculée par (5) $\}$
    end if
  end for
  Output : $m_q(sol_q)$

L'algorithme 1 permet d'implémenter la méthode. RandomChoice(k) sélectionne k indices de l'ensemble d'apprentissage par tirage aléatoire uniforme, et NearestNeighbors(x, k) sélectionne les indices des k plus proches voisins de x dans l'ensemble d'apprentissage. Si aucun triplet ou aucune solution n'est trouvé, aucune prédiction ne peut être faite. Néanmoins, dans notre implémentation, la moyenne arithmétique des sorties sur l'ensemble d'apprentissage est prédite.

Afin d'évaluer les modèles de régression proposés ci-dessus, nous les comparons à différentes méthodes : la régression linéaire, les k plus proches voisins et arbres de décision avec les valeurs par défaut de la bibliothèque scikit-learn [31]. Nous testons en plus un modèle de régression appelé FIT dans les résultats, qui sert de référence de par le théorème 1. Il prédit x' comme $(w(x')^p + b)^{1/q}$, ou $(w(1-x')^p + b)^{1/q}$, pour les paramètres trouvés. Il est implémenté au moyen de la fonction `curve_fit` de la bibliothèque `scipy`, avec w borné par le quotient des maximums de la sortie et du trait sur le test d'entraînement.

Toutes les expériences ont été réalisées avec $\theta = 0{,}99$.

3. La fonction A définie dans l'équation (6) est la généralisation aux analogies non arithmétiques de la fonction P introduite dans [6].

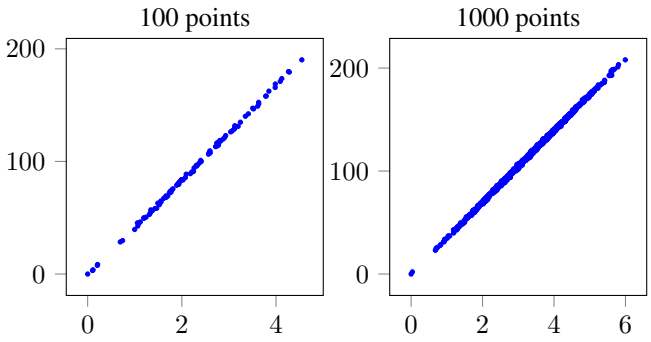

FIGURE 1 – Jeux de données artificielles de 100 et 1 000 points.

4.2 Données artificielles

Étant donné que nos modèles de régression sont des généralisations de la régression linéaire, nous les testons d'abord sur des jeux de données dont la sortie est produite par un modèle linéaire avec un bruit centré gaussien afin de vérifier leur qualité.

Nous utilisons deux jeux de données comprenant respectivement 100 et 1 000 échantillons, tous deux avec un bruit d'écart-type de 1, produit par la fonction `make_regression` de la bibliothèque `scikit-learn`. Les points ont été décalés de manière à ce que toutes les valeurs soient positives. Les graphes de ces deux jeux de données apparaissent dans la figure 1 avec la valeur d'entrée en x et la valeur de sortie en y.

L'erreur quadratique moyenne (EQM), l'erreur absolue moyenne (EAM) et le coefficient de détermination (r^2) des algorithmes ont été évalués sur le même découpage en 5 plis. Les résultats sont présentés dans les tableaux 2 et 3.

En ce qui concerne le jeu de données de 100 échantillons, lorsque le couple (a, b) provient des plus proches voisins, l'erreur quadratique moyenne est plus élevée mais avec une erreur absolue moyenne comparable aux méthodes traditionnelles lorsque 5 ou 7 voisins sont sélectionnés. Alors que les méthodes traditionnelles atteignent un r^2 d'au moins 99%, nos modèles de régression atteignent 94% avec 5 ou 7 voisins. En revanche, lorsque (a, b) est tiré au hasard, les trois mesures pour nos modèles de régression surpassent les méthodes traditionnelles, à l'exception de la régression linéaire par définition mieux adaptée à ces jeux de données.

Sur l'ensemble de données de 1 000 échantillons, lorsque le principe d'inférence par analogie est appliqué à tous les éléments du jeu de test, nos modèles de régression surpassent

TABLEAU 1 – Statistiques des jeux de données réelles

Jeu de données	Nbre d'instances utilisées	Nbre de traits utilisés
NY City - East River Bicycle Crossings	210	2
California Housing Prices	20,433	1

TABLEAU 2 – Moyennes et écarts-types des EQM et EAM et r^2 avec validation croisée à 5 plis sur notre jeu de données artificielles de 100 points.

Méthode		EQM	EAM	r^2	
Régression linéaire		0.978 ± 0.143	0.815 ± 0.059	0.999 ± 0.000	
1 plus proche voisin		4.780 ± 2.266	1.625 ± 0.285	0.997 ± 0.001	
3 plus proches voisins		8.613 ± 4.105	1.950 ± 0.372	0.995 ± 0.002	
5 plus proches voisins		13.940 ± 6.832	2.223 ± 0.348	0.992 ± 0.004	
Arbre de décision		4.780 ± 2.266	1.625 ± 0.285	0.997 ± 0.001	
FIT		1.004 ± 0.133	0.838 ± 0.059	0.999 ± 0.000	
(a, b)	Nbre de voisins	EQM	EAM	r^2	Pas d'analogie (% du jeu d'entraîn.)
Voisins	1	633.434 ± 116.561	11.318 ± 1.437	0.637 ± 0.103	19%
	3	226.035 ± 212.005	3.642 ± 2.029	0.879 ± 0.119	3 %
	5	102.824 ± 202.674	1.967 ± 1.896	0.936 ± 0.126	1 %
	7	102.635 ± 202.738	1.920 ± 1.921	0.936 ± 0.126	1 %
	% du jeu d'entraîn.				
Aléatoire	5%	1.544 ± 0.331	1.021 ± 0.095	0.999 ± 0.000	0%
	10 %	1.366 ± 0.369	0.953 ± 0.111	0.999 ± 0.000	0 %
	20 %	1.064 ± 0.163	0.865 ± 0.062	0.999 ± 0.000	0 %
	100 %	1.034 ± 0.179	0.849 ± 0.084	0.999 ± 0.000	0 %

TABLEAU 3 – Même chose que le tableau 2, mais sur un échantillon de 1 000 points.

Méthode		EQM	EAM	r^2	
Régression linéaire		0.945 ± 0.105	0.779 ± 0.036	0.999 ± 0.000	
1 plus proche voisin		2.043 ± 0.269	1.143 ± 0.057	0.998 ± 0.000	
3 plus proches voisins		1.881 ± 0.572	0.968 ± 0.025	0.998 ± 0.000	
5 plus proches voisins		2.175 ± 0.960	0.943 ± 0.041	0.998 ± 0.001	
Arbre de décision		2.043 ± 0.269	1.143 ± 0.057	0.998 ± 0.000	
FIT		0.945 ± 0.105	0.779 ± 0.036	0.999 ± 0.000	
(a, b)	Nbre de voisins	EQM	EAM	r^2	Pas d'analogie (% du jeu d'entraîn.)
Voisins	1	46.356 ± 41.265	1.594 ± 0.409	0.962 ± 0.032	0.5 %
	3	15.106 ± 17.778	1.111 ± 0.213	0.988 ± 0.014	0.2 %
	5	1.216 ± 0.148	0.884 ± 0.038	0.999 ± 0.000	0.0 %
	% du jeu d'entraîn.				
Aléatoire	0.5%	1.101 ± 0.131	0.842 ± 0.041	0.999 ± 0.000	0 %
	1%	0.995 ± 0.123	0.799 ± 0.039	0.999 ± 0.000	0 %
	10%	0.950 ± 0.105	0.781 ± 0.037	0.999 ± 0.000	0 %

des méthodes traditionnelles pour les trois mesures, à l'exception évidemment de la régression linéaire.

La méthode FIT a été utilisée pour servir de référence pour les performances des modèles de régression qui utilisent le principe d'inférence par analogie. Lorsque la prédiction pour tous les points de données est effectuée à l'aide de ce principe, le modèle de régression présente des scores similaires à celles de la méthode FIT, ce qui confirme la pertinence du théorème 1. Cependant, cette similitude n'apparaît pas dans les jeux de données réels, comme le montre la section suivante.

4.3 Données réelles

Nous testons maintenant nos modèles de régression sur deux jeux de données réelles : les passages à vélo sur l'East River à New York [4], et le prix des logements en Californie [5]. Ces deux jeux de test sont décrits dans le tableau 1.

Les expériences sur les données synthétiques nous ont permis de sélectionner des hyperparamètres adéquats pour que tous les points puissent être évalués par le principe d'inférence par analogie : θ_α doit être asssez proche de 1, s ne doit pas être trop grand, pour pouvoir prédire quelque chose (la complexité est en $O(n^3)$), k non plus pour coller à la notion de voisins.

4.3.1 Passages à vélo

Ce jeu de données donne le nombre total de cyclistes en fonction de deux traits : températures maximales et minimales. Nos modèles de régression étant unidimensionnels, nous évaluons sur chaque trait séparément. Ce jeu offre un troisème trait, les précipitations, mais nous ne l'utilisons pas car le principe d'inférence par analogie n'a pu s'appliquer que sur 25 % des points du jeu de test.

Les coefficients de détermination pour chaque modèle de régression et chaque trait sont présentés dans le tableau 4. Nos modèles de régression obtiennent ici de meilleurs résultats que la régression linéaire. Avec le trait température minimale comme donnée d'entrée, ils obtiennent de meilleurs scores que la méthode des k plus proches voisins avec $k = 7$, mais des résultats inférieurs pour cette même méthode avec $k = 5$, ou que les arbres de décision.

Lorsque a et b sont tirés au hasard, l'augmentation des points considérés diminue les performances, contrairement aux attentes. Bien que les scores se situent dans la même fourchette avec la température maximale, ce n'est pas le cas pour la température minimale, ce qui rend douteux les scores rapportées pour 1 %. Nos modèles de régression sont ici clairement meilleurs que la méthode FIT.

4.3.2 Prix des logements en Californie

Ce jeu de données, California Housing Prices, permet d'estimer la valeur médiane des logements pour les ménages d'un même quartier, en fonction de leur revenu médian. Dans nos expériences, nous n'avons pas pris en compte les lignes comportant des valeurs manquantes.

Les résultats sont présentés dans le tableau 5. Ici nous avons également testé la sélection de cinq valeurs de c au lieu d'une seule. Ces tests sont repérables dans le tableau avec le nombre c entre parenthèses. Cette augmentation du nombre de c nécessite une modification dans l'algorithme 1 :

$$L \leftarrow \text{NearestNeighbors}((a^p - b^p + (x')^p)^{1/p}, 5).$$

La méthode par arbre de décision affichant de très mauvais scores sur ce jeu de données, tous les modèles de régression proposés obtiennent un r^2 plus élevé. Néanmoins, nos modèles de régression avec plus proches voisins affichent des scores inférieurs à ceux de la méthode des k plus proches voisins, alors qu'avec sélection de type aléatoire ils obtiennent des scores du même ordre. En sélectionnant sur 5 c au lieu d'un seul, la moyenne r^2 augmente et l'écart-type diminue, le modèle de régression $0,05\,\%$ (5) obtenant même de meilleurs résultats qu'avec la méthode des k plus proches voisins pour $k = 11$. Malgré ces améliorations, aucun de nos modèles de régression n'arrive à battre la régression linéaire ou la méthode FIT.

5 Conclusion

Cet article a présenté un résultat nouveau sur l'apprenabilité dans le cadre PAC de fonctions de domaine et image de dimension un préservant l'analogie et compatibles avec le principe d'inférence par analogie. La définition de telles fonctions repose sur l'analogie numérique, type d'analogie qui permet de traiter le cas de valeurs continues. De telles fonctions définissent donc des modèles de régression fondés sur le principe d'inférence par analogie qui s'appliquent au cas du continu, et sont donc plus générales que des classificateurs définis dans un cadre discret booléen. Des premières expériences ont permis d'inspecter les performances de méthodes approchant de telles fonctions sur des jeux de données artificielles et réelles.

Cette approche présente bien sûr certaines limites identifiées dans notre étude empirique. Tout d'abord, le théorème 1 suppose qu'il existe une fonction avec un domaine $(\mathbb{R}_+^*)^n$, ce qui implique que la fonction est constante ou que son image est un intervalle non borné de $\mathbb{R}_+^*$. Pour aborder l'apprenabilité dans le cadre PAC, nous avons supposé que l'image soit $[0, 1]$, ce qui restreint le domaine d'application. Sur ce domaine, les fonctions de $\mathcal{F}^1$ ne sont peut-être pas les seules fonctions préservant l'analogie, et une étude plus complète des fonctions préservant l'analogie restreintes aux fonctions de domaine et d'image $[0, 1]$ est nécessaire. Dans les expériences, il n'y avait pas de garantie d'apprenabilité PAC pour la procédure trouvant les meilleures valeurs de p et q, pour tous les w et b possibles. Bien que les expériences n'aient pas démontré une meilleure performance des modèles de régression proposés par rapport aux méthodes traditionnelles, elles suggèrent qu'ils peuvent constituer une alternative dans certaines situations.

Afin de tester pleinement les méthodes proposées, une généralisation de l'apprenabilité PAC des fonctions définies

4. New York City - East River Bicycle Crossings : https://www.kaggle.com/datasets/new-york-city/nyc-east-river-bicycle-crossings

5. California Housing Prices : https://www.kaggle.com/datasets/camnugent/california-housing-prices

TABLEAU 4 – Moyennes et écarts-types de r^2 avec évaluation croisée à 5 plis pour le jeu des passages à vélo à New York.

Méthode	Sélection utilisée	Température maximale	Température minimale
Régression linéaire		0.515 ± 0.140	0.205 ± 0.081
5 plus proches voisins		0.808 ± 0.044	0.802 ± 0.089
7 plus proches voisins		0.755 ± 0.057	0.691 ± 0.098
Arbre de décision		0.842 ± 0.038	0.865 ± 0.051
FIT		0.498 ± 0.135	0.211 ± 0.071
(a, b)	Nbre de voisins		
Voisins	3	0.715 ± 0.087	0.776 ± 0.028
	5	0.724 ± 0.066	0.755 ± 0.037
	7	0.711 ± 0.066	0.714 ± 0.051
	% du jeu d'entraîn.		
Aléatoire	1%	0.706 ± 0.075	0.776 ± 0.028
	5 %	0.660 ± 0.070	0.481 ± 0.086
	10 %	0.617 ± 0.109	0.394 ± 0.092

TABLEAU 5 – Même chose que le tableau 4, mais pour le jeu des prix du logement en Californie. Le type aléatoire de modèles de régression avec un (k) suivant le pourcentage, par exemple $0,05$ % (5) indique les expériences où k et c sont pris en compte.

Méthode	Sélection utilisée	Revenu médian
Régression linéaire		0.474 ± 0.015
7 plus proches voisins		0.415 ± 0.018
9 plus proches voisins		0.432 ± 0.018
11 plus proche voisin		0.443 ± 0.020
Arbre de décision		0.134 ± 0.021
FIT		0.475 ± 0.015
(a, b)	Nbre de voisins	
Voisins	7	0.313 ± 0.023
	9	0.338 ± 0.020
	11	0.368 ± 0.020
	% du jeu d'entraîn. (nbre de c)	
Aléatoire	0.05 % (1)	0.438 ± 0.034
	0.10 % (1)	0.422 ± 0.069
	0.05 % (5)	0.465 ± 0.015
	0.10 % (5)	0.438 ± 0.032

sur des espaces continus à plusieurs dimensions ($\mathbb{R}^n$) est nécessaire. Quoi qu'il en soit, nos résultats préliminaires sont prometteurs quant à la possibilité de prouver l'apprenabilité PAC des classes de fonctions $\mathcal{F}^n$ compatibles avec le principe d'inférence par analogie.

Remerciements

Les travaux présentés ici ont bénéficié d'une subvention de recherche de l'université Waseda, 2025R-034, intitulée « Analogie numérique : formalisation mathématique et application pratique à l'apprentissage automatique ».
Ils ont été aussi partiellement soutenus par le projet ANR *Analogies : from theory to tools and applications* (AT2TA), ANR-22-CE23-0023.
Ils ont également bénéficié d'un financement national portugais par l'intermédiaire de la FCT, dans le cadre du projet UIDB/50021/2020 (DOI : 10.54499/UIDB/50021/2020), ainsi que du projet « Center for Responsible AI », référence C628696807-00454142.

Références

[1] Stergos AFANTENOS, Miguel COUCEIRO, Emiliano LORINI et Van-Duy NGO : Learning proportional analogies : Lightweight neural network vs large language models. *In Proceedings of the 18th International Conference on Agents and Artificial Intelligence, ICAART 2026 - To appear*, 2026.

[2] Noga ALON, Shai BEN-DAVID, Nicolò CESA-BIANCHI et David HAUSSLER : Scale-sensitive dimensions, uniform convergence, and learnability. *J. ACM*, 44(4):615–631, juillet 1997.

[3] ARISTOTE : *Éthique à Nicomaque*. Librairie philosophique J. Vrin, Paris, [1er tirage 1990] édition, 1997. Trad. J. Tricot.

[4] Fadi BADRA : A dataset complexity measure for analogical transfer. *In* Christian BESSIERE, éditeur : *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 1601–1607. ijcai.org, 2020.

[5] Fadi BADRA, Esteban MARQUER, Marie-Jeanne LESOT, Miguel COUCEIRO et David LEAKE : Energy-compress : A general case base learning strategy. *In IJCAI 2025*, pages 4339–4346. ijcai.org, 2025.

[6] Myriam BOUNHAS et Henri PRADE : Analogy-based classifiers : An improved algorithm exploiting competent data pairs. *Int. J. Approx. Reasoning*, 158(C), juillet 2023.

[7] Myriam BOUNHAS et Henri PRADE : Revisiting analogical proportions and analogical inference. *Int. J. Approx. Reason.*, 171:109202, 2024.

[8] Zied BOURAOUI, Shoaib JAMEEL et Steven SCHOCKAERT : Relation induction in word embeddings revisited. *In COLING 2018*, pages 1627–1637. ACL, 2018.

[9] Miguel COUCEIRO, Nicolas HUG, Henri PRADE et Gilles RICHARD : Analogy-preserving functions : A way to extend boolean samples. *In IJCAI 2017*, pages 1575–1581. ijcai.org, 2017.

[10] Miguel COUCEIRO, Nicolas HUG, Henri PRADE et Gilles RICHARD : Behavior of analogical inference w.r.t. boolean functions. *In* Jérôme LANG, éditeur : *IJCAI 2018*, pages 2057–2063. ijcai.org, 2018.

[11] Miguel COUCEIRO et Erkko LEHTONEN : Galois theory for analogical classifiers. *Ann. Math. Artif. Intell.*, 92(1):29–47, 2024.

[12] Miguel COUCEIRO, Erkko LEHTONEN, Laurent MICLET, Henri PRADE et Gilles RICHARD : When nominal analogical proportions do not fail. *In* Jesse DAVIS et Karim TABIA, éditeurs : *SUM 2020*, volume 12322 de *LNCS*, pages 68–83. Springer, 2020.

[13] Francisco CUNHA, Yves LEPAGE, Miguel COUCEIRO et Zied BOURAOUI : Generalizing analogical inference from boolean to continuous domains. *In* Sven KOENIG, Chad JENKINS et Matthew E. TAYLOR, éditeurs : *Fortieth AAAI Conference on Artificial Intelligence, Thirty-Eighth Conference on Innovative Applications of Artificial Intelligence, Sixteenth Symposium on Educational Advances in Artificial Intelligence, AAAI 2026, Singapore, January 20-27, 2026*, pages 19021–19029. AAAI Press, 2026.

[14] Dedre GENTNER : Structure-mapping : A theoretical framework for analogy. *Cognitive Science*, 7(2):155–170, 1983.

[15] Anna GLADKOVA, Aleksandr DROZD et Satoshi MATSUOKA : Analogy-based detection of morphological and semantic relations with word embeddings : what works and what doesn't. *In Proceedings of the NAACL-HLT SRW*, pages 47–54, San Diego, California, June 12-17 2016.

[16] Douglas HOFSTADTER et Emmanuel SANDER : *Surfaces and Essences : Analogy as the Fuel and Fire of Thinking*. Basic Books, New York, 2013.

[17] Xiaoyang HU, Shane STORKS, Richard L. LEWIS et Joyce CHAI : In-context analogical reasoning with pre-trained language models. *arXiv preprint arXiv :2305.17626*, 2023.

[18] Eyke HÜLLERMEIER : Towards analogy-based explanations in machine learning. *In* Vicenç TORRA, Yasuo NARUKAWA, Jordi NIN et Núria AGELL, éditeurs : *MDAI 2020*, volume 12256 de *LNCS*, pages 205–217. Springer, 2020.

[19] Sung Ju HWANG, Kristen GRAUMAN et Fei SHA : Analogy-preserving semantic embedding for visual object categorization. *In ICML 2013*, volume 28 de *JMLR Workshop and Conference Proceedings*, pages 639–647, 2013.

[20] Shahar JACOB, Chen SHANI et Dafna SHAHAF : FAME : flexible, scalable analogy mappings engine. *In EMNLP 2023*, pages 16426–16442. ACL, 2023.

[21] Lucas JARNAC, Miguel COUCEIRO et Pierre MONNIN : Relevant entity selection : Knowledge graph bootstrapping via zero-shot analogical pruning. *In ACM CIKM 2023, Birmingham, United Kingdom, October 21-25, 2023*, pages 934–944. ACM, 2023.

[22] Nitesh KUMAR et Steven SCHOCKAERT : Solving hard analogy questions with relation embedding chains. *In EMNLP 2023*, pages 6224–6236. ACL, 2023.

[23] Yves LEPAGE : Solving analogies on words : an algorithm. *In Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics (ACL'98) and the 17th International Conference on Computational Linguistics (COLING'98)*, volume I, pages 728–735, Montreal, août 1998. Association for Computational Linguistics.

[24] Yves LEPAGE : *De l'analogie rendant compte de la commutation en linguistique*. Mémoire d'habilitation à diriger les recherches, Université de Grenoble, mai 2003.

[25] Yves LEPAGE et Miguel COUCEIRO : Analogie et moyenne généralisée. *In* Jean-Guy MAILLY, François SCHWARZENTRUBER et Anaelle WILCZYNSKI, éditeurs : *Actes des 18es Journées d'Intelligence Artificielle Fondamentale et des 19es Journées Francophones sur la Planification, la Décision et l'Apprentissage pour la conduite de systèmes (JIAF 2024 et JFPDA 2024)*, pages 114–124, La Rochelle, France, Juillet 2024.

[26] Yves LEPAGE et Miguel COUCEIRO : L'analogie numérique revisitée et étendue, précisée, rétrécie ou agrandie. *In* Jean-Guy MAILLY, François SCHWARZENTRUBER et Anaelle WILCZYNSKI, éditeurs : *Actes des 19es Journées d'Intelligence Artificielle Fondamentale et des 20es Journées Francophones sur la Planification, la Décision et l'Apprentissage pour la conduite de systèmes (JIAF 2025 et JFPDA 2025)*, pages 133–142, Dijon, France, Juillet 2025. Association française pour l'Intelligence Artificielle.

[27] Yves LEPAGE et Etienne DENOUAL : Purest ever example-based machine translation : detailed presentation and assessment. *Machine Translation*, 19:251–282, 2005.

[28] Esteban MARQUER, Fadi BADRA, Marie-Jeanne LESOT, Miguel COUCEIRO et David LEAKE : Less is better : An energy-based approach to case base competence. *In* Lukas MALBURG et Deepika VERMA, éditeurs : *Proceedings of the Workshops at the 31st International Conference on Case-Based Reasoning (ICCBR-WS 2023) co-located with the 31st International Conference on Case-Based Reasoning (ICCBR 2023), Aberdeen, Scotland, UK, July 17, 2023*, volume 3438 de *CEUR Workshop Proceedings*, pages 27–42. CEUR-WS.org, 2023.

[29] Tomas MIKOLOV, Wen-Tau YIH et Geoffrey ZWEIG : Linguistic regularities in continuous space word representations. *In NAACL-HLT 2013*, pages 746–751, Atlanta, Georgia, June 2013. ACL.

[30] Pierre MONNIN, Cherif-Hassan NOUSRADINE, Lucas JARNAC, Laurel ZUCKERMAN et Miguel COUCEIRO : KGPRUNE : A web application to extract subgraphs of interest from wikidata with analogical pruning. *In* Ulle ENDRISS, Francisco S. MELO, Kerstin BACH, Alberto José Bugarín DIZ, Jose Maria ALONSO-MORAL, Senén BARRO et Fredrik HEINTZ, éditeurs : *ECAI 2024 - 27th European Conference on Artificial Intelligence, 19-24 October 2024, Santiago de Compostela, Spain - Including 13th Conference on Prestigious Applications of Intelligent Systems (PAIS 2024)*, volume 392 de *Frontiers in Artificial Intelligence and Applications*, pages 4495–4498. IOS Press, 2024.

[31] Fabian PEDREGOSA, Gaël VAROQUAUX, Alexandre GRAMFORT, Vincent MICHEL, Bertrand THIRION, Olivier GRISEL, Mathieu BLONDEL, Peter PRETTENHOFER, Ron WEISS, Vincent DUBOURG, Jake VANDERPLAS, Alexandre PASSOS, David COURNAPEAU, Matthieu BRUCHER, Matthieu PERROT et Édouard DUCHESNAY : Scikit-learn : Machine learning in python. *Journal of Machine Learning Research*, 12(85): 2825–2830, 2011.

[32] Henri PRADE et Gilles RICHARD : From analogical proportion to logical proportions. *Logica Universalis*, 7:441–505, 2013.

[33] Asahi USHIO, Luis Espinosa ANKE, Steven SCHOCKAERT et José CAMACHO-COLLADOS : BERT is to NLP what alexnet is to CV : can pre-trained language models identify analogies ? *In ACL/IJCNLP 2021, (Volume 1 : Long Papers)*, pages 3609–3624. ACL, 2021.

[34] Leslie G. VALIANT : A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, novembre 1984.

[35] Liyan WANG, Haotong WANG et Yves LEPAGE : AnaScore : understanding semantic parallelism in proportional analogies. *In* Luis CHIRUZZO, Alan RITTER et Lu WANG, éditeurs : *Proceedings of the 2025 Annual Conference of the North-American Chapter of the Association for Computational Linguistics (NAACL 2025)*, pages 1175–1188, Albuquerque, New Mexico, apr 2025.

[36] Liyan WANG, Bartholomäus WLOKA et Yves LEPAGE : Eliciting analogical reasoning from language models in retrieval-augmented translation under low-resource scenarios. *Neurocomputing*, 630(129680):18 pages, 2025.