

---

# Fonctions préservées par analogie : une méthode pour étendre des données booléennes

## Analogy-preserving functions: A way to extend Boolean samples

---

Miguel Couceiro<sup>1</sup>, Nicolas Hug, Henri Prade, Gilles Richard

1. LORIA, Campus Scientifique, 54506 Vandœuvre-lès-Nancy

2. IRIT, Université Paul Sabatier, 31062 Toulouse Cedex 9

miguel.couceiro@inria.fr, {nicolas.hug, henri.prade, gilles.richard}@irit.fr

### Résumé

L'extension des ensembles d'exemples est une question importante en apprentissage automatique. En effet quand les exemples disponibles sont en quantité limitée, les performances des méthodes classiques de classification peuvent diminuer, et il peut être désirable de construire des exemples additionnels. Dans cet article, nous considérons l'usage du raisonnement analogique, et plus particulièrement des proportions analogiques pour étendre les ensembles d'exemples. On fait l'hypothèse que les classes sont déterminées par une fonction (partiellement connue). On examine les conditions requises sur les fonctions pour garantir une extension sans erreur des ensembles d'exemples, dans un cadre booléen. Dans ce but, on introduit la notion de fonctions préservant l'analogie (fonctions PA), et nous établissons qu'elles coïncident avec la classe des fonctions booléennes affines. Ce résultat théorique est complété par une étude expérimentale de PA-fonctions approchées, ce qui suggère qu'elles peuvent rester adéquates pour étendre des ensembles d'exemples.

**Mots clés :** proportion analogique, extension d'ensembles de données d'entraînement.

### Abstract

Training set extension is an important issue in machine learning. Indeed when the examples at hand are in a limited quantity, the performances of standard classifiers may significantly decrease and it can be helpful to build additional examples. In this paper, we consider the use of analogical reasoning, and more particularly of analogical proportions for extending training sets. Here the ground truth labels are considered to be given by a (partially known) function. We examine the conditions that are required for such functions to ensure an error-free extension in a Boolean setting. To this end, we introduce the notion of Analogy Preserving (AP) functions, and we prove that their class is the class of affine Boolean functions. This theoretical result is comple-

ted with an empirical investigation of *approximate* functions AP, which suggests that they remain suitable for training set extension.

**Keywords :** proportion analogique, training set extension.

## 1 Introduction

La capacité à apprendre à partir de peu d'exemples est une faculté importante du cerveau humain, et joue un grand rôle dans l'élaboration des catégories cognitives chez l'enfant [14]. Cela contraste avec les algorithmes en apprentissage qui habituellement requièrent des ensembles de données suffisamment grands. La question d'apprendre à partir de quelques exemples n'est pas nouvelle, voir par exemple [12] dans le contexte de la reconnaissance des formes.

Le raisonnement analogique, qui est reconnu comme un moyen puissant d'établir des parallèles entre des entités apparemment indépendantes, peut également être employé pour construire de nouveaux exemples à partir d'un petit ensemble de données [2].

L'analogie a une longue histoire et a été principalement étudiée d'un point de vue psychologique [9, 13], mais des modélisations logiques ont aussi été introduites : ainsi [10] a donné une modélisation de l'analogie en logique du premier ordre qui apparaît trop restrictive, et une approche plus sophistiquée basée sur une logique d'ordre supérieure a été proposée [15], qui correspond à la l'idée de la théorie de la mise en parallèle de structures ("structure mapping theory") de [13].

En suivant des idées venant de l'anthropologie structurale [17] et de la linguistique computationnelle [18, 30], une modélisation en logique propositionnelle des proportions analogiques, i.e., des énoncés de la forme "**a** est à **b**

ce que  $\mathbf{c}$  est à  $\mathbf{d}$ ”, a été introduite par [23, 25]. L’approche du raisonnement analogique basée sur les proportions, dont l’intérêt cognitif a été souligné depuis longtemps [27], a aussi montré qu’elle pouvait présenter un intérêt en classification sur problèmes tests [1, 4].

Les auteurs de [16] ont montré que ce type de processus de classification analogique pouvait être formalisé en deux grandes étapes : Tout d’abord une *extension analogique* d’un ensemble de données est construite (comme expliqué en Section 3), et ensuite un algorithme de type  $k$ -NN est appliqué à cet ensemble de données étendu. Comme on peut s’y attendre, les performances du classifieur analogique dépendent grandement de la qualité de l’extension. Dans cet article, on introduit la classe PA des fonctions préservées par l’analogie qui garantissent une extension sans erreur.

L’article est organisé comme suit. En Section 2 nous proposons une vue d’ensemble des différentes méthodes couramment utilisées pour étendre un ensemble de données. En Section 3 on rappelle des éléments de base sur l’analogie et son expression en logique booléenne (sous forme de proportion analogique), en pointant l’existence de deux modélisations potentielles. On rappelle aussi le processus d’extension d’un ensemble de données par l’analogie et on introduit la notion de fonctions PA. La Section 4 porte sur la caractérisation théorique des fonctions PA. En section 5 on définit et on étudie empiriquement les fonctions PA *ap-prochées*. On montre ensuite leur intérêt pour l’extension des ensembles de données en pratique.

## 2 Extension d’un ensemble de données

L’extension d’un ensemble de données (“training set”) d’un référentiel donné  $\mathcal{X}$  est une idée simple pour améliorer les capacités de généralisation d’un classifieur. Il s’agit d’ajouter à un ensemble de données  $S$  de nouveaux exemples, mais en veillant à préserver la *qualité* de  $S$ .

Formellement, on part d’un ensemble  $S = \{x^{(i)} \in \mathcal{X} | i \in [1, n]\}$  d’exemples ( $n$  est supposé petit), où  $x^{(i)}$  est un élément d’un produit cartésien  $\mathcal{X} = X_1 \times \dots \times X_m$ . A chaque élément  $x^{(i)} \in S$ , on associe une valeur de la fonction cible  $f(x^{(i)}) = y^{(i)} \in Y$ . En cas de régression,  $y^{(i)} \in \mathbb{R}$ , et en cas de classification  $y^{(i)}$  appartient à un ensemble fini et est appelé “classe” ou **étiquette**.

Plusieurs méthodes ont été proposées pour étendre un ensemble de données avec de nouveaux exemples. On peut ainsi obtenir un nouvel exemple en partant de 1, 2 ou 3 exemples connus.

1. Avec un seul exemple, une façon naturelle de procéder est d’utiliser l’approche classique par voisinage : étant donné l’exemple  $(a, f(a))$ , on peut générer un nouvel exemple  $(b, f(b))$  où  $b$  est proche de  $a$  et  $f(b)$  est proche de  $f(a)$ . En classification,  $f(b)$  peut être choisi comme  $f(a)$ .

2. Avec deux exemples, l’option précédente reste applicable et conduit à interpoler le nouvel exemple à partir des deux qui sont donnés. Une option un peu différente est la procédure dite de “Feature Knockout” [32], qui consiste à construire un troisième exemple en modifiant un attribut choisi au hasard du premier exemple par sa valeur dans le second exemple. Ce procédé possède des propriétés intéressantes et est en fait équivalent à une technique populaire de régularisation (Tikhonov) en régression linéaire. Une idée apparentée est utilisée dans un proposition récente [5] qui introduit une mesure d’étrangeté (“oddness”) par rapport à une classe qui est calculée sur la base de paires constituées de deux plus proches voisins dans la même classe ; la méthode revient à remplacer les deux voisins par un représentant fictif de la classe.
3. Avec trois exemples  $(a, f(a)), (b, f(b)), (c, f(c))$ , les options précédentes restent possibles et conduisent à construire un quatrième exemple qui est dans un certain sens entre les trois autres : nous avons encore une forme d’interpolation. Une idée assez différente est d’extrapoler le quatrième item sur la base d’une proportion analogique [2]. Dans cette perspective, le quatrième élément n’est pas nécessairement dans le voisinage des trois autres.

Dans cet article, nous étudions cette dernière option en profondeur. Cet article est la version française de [6].

## 3 Proportions analogiques pour l’extension d’ensembles de données

Comme indiqué plus haut, les classifieurs analogiques [16] étendent d’abord l’ensemble des données, puis appliquent une procédure type  $k$ -NN à l’ensemble de données ainsi augmenté. Dans la suite, on se focalise sur la première étape, l’extension étant basée sur la notion de proportion analogique.

L’idée de **proportion** fut introduite dans la Grèce Ancienne dans un cadre numérique avec en particulier les exemples de :

1. la **proportion arithmétique**, où  $a, b, c, d$  sont en proportion si  $a - b = c - d$  ;
2. la **proportion géométrique**, où  $a, b, c, d$  sont en proportion si  $\frac{a}{b} = \frac{c}{d}$ .

Ces exemples capturent implicitement l’idée de proportion analogique qui correspond à un énoncé de la forme *a est à b comme c est à d*.

Dans cette section, on rappelle tout d’abord les propriétés de base des proportions analogiques, et on décrit ensuite le processus de construction de l’extension analogique d’un ensemble de données. Dans la suite,  $X$  dénote un ensemble non vide, et les éléments de  $X$  seront dénotés par des lettres minuscules  $a, b, c, d, \dots$ . De plus, les  $m$ -uplets (ou vecteurs)

sur  $X$  seront notés par lettres en caractères gras  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ , ... la  $i$ -ème composante d'un  $m$ -uplet  $\mathbf{a}$  sera noté  $a_i$ .

### 3.1 Proportions analogiques

Comme leur contrepartie numérique, une **proportion analogique**<sup>1</sup> sur un ensemble non vide  $X$  est une relation quaternaire  $A$  sur  $X$  qui satisfait les 3 axiomes suivants [11, 19] :

1. **Identité** : pour tout  $a, b \in X$ ,  $A(a, b, a, b)$ .
2. **Symétrie** : pour tout  $a, b, c, d \in X$ , si  $A(a, b, c, d)$ , alors  $A(c, d, a, b)$ .
3. **Permutabilité centrale** : pour tout  $a, b, c, d \in X$ , si  $A(a, b, c, d)$ , alors  $A(a, c, b, d)$ .

Il y a différentes manières de définir une analogie sur un ensemble  $X$ , selon la structure sous-jacente et les opérateurs disponibles. Dans [22, 30, 21], on trouvera des exemples pour des structures matricielles, des mots sur un alphabet, des treillis, etc. Quand  $X = \mathbb{B} = \{0, 1\}$ , des expressions logiquement équivalentes de la proportion analogique ont été données [23, 25] :

$$A(a, b, c, d) \text{ si } (a \wedge \neg b \leftrightarrow c \wedge \neg d) \wedge (\neg a \wedge b \leftrightarrow \neg c \wedge d) \\ \text{ou de manière équivalente, si} \\ (a \wedge d \leftrightarrow b \wedge c) \wedge (a \vee d \leftrightarrow b \vee c),$$

où  $\leftrightarrow$  est le connecteur d'équivalence :  $x \leftrightarrow x' = 1$  si  $x = x'$  et 0 sinon. La première expression exprime que  $a$  diffère de  $b$  comme  $c$  diffère de  $d$  et inversement  $b$  diffère de  $a$  comme  $d$  diffère de  $c$ . La seconde équivalence traduit le fait que l'analogie se comporte comme une proportion numérique pour ce qui est des extrêmes et des moyens.

Nous ferons référence dans la suite à ce modèle booléen de l'analogie comme le modèle **standard**. En fait, il y a d'autres modèles de l'analogie qui satisfont les trois axiomes qui peuvent être construits dans  $\mathbb{B}$  [24]. Parmi eux, un présente un intérêt particulier : il s'agit du modèle proposé dans [17], qui est obtenu en relaxant le modèle standard sous la forme  $(a \leftrightarrow b) \leftrightarrow (c \leftrightarrow d)$ . On référera aux modèles standard et de Klein par  $A_S$  et  $A_K$  respectivement. Contrairement au modèle standard, celui de Klein satisfait une propriété apparemment attractive :  $A_K(a, \neg a, b, \neg b)$ .

La Table 1 montre les 8 lignes où les proportions du modèle de Klein tiennent. Pour les 8 configurations de valeurs de  $a, b, c, d$  restantes qui n'apparaissent pas dans la table, aucune ne correspond à une proportion valide pour aucun des deux modèles. On peut voir que le modèle de Klein, quoique attractif à première vue, obéit à la propriété suivante qui ne semble pas naturelle pour une analogie :  $A_K(a, b, c, d) \iff A_K(b, a, c, d)$ . Il se trouve cependant que tous les résultats théoriques dans cet article seront valides à la fois pour le modèle standard et le modèle de

1. dans le reste de l'article, le terme *analogie* signifie toujours *proportion analogique*.

$a$	$b$	$c$	$d$	$A_S$	$A_K$
0	0	0	0	1	1
1	1	1	1	1	1
0	0	1	1	1	1
1	1	0	0	1	1
0	1	0	1	1	1
1	0	1	0	1	1
0	1	1	0	0	1
1	0	0	1	0	1

TABLE 1 – Valuations de  $a : b :: c : d$  pour le modèle standard et le modèle de Klein de l'analogie. Les valuations manquantes donnent 0 pour  $A_S$  et  $A_K$ .

Klein ; on utilisera la notation  $a : b :: c : d$  qui vaudra pour  $A_S(a, b, c, d)$  ou  $A_K(a, b, c, d)$  indifféremment. En Section 5, on verra cependant que, en dépit de propriétés théoriques similaires, les résultats empiriques montrent que le modèle standard est préférable.

On peut vérifier sur la Table 1 que  $A_S$  et  $A_K$  satisfont aussi une sorte de *propriété d'indépendance au codage* :

$$a : b :: c : d \iff \neg a : \neg b :: \neg c : \neg d,$$

ce qui garantit que 0 et 1 jouent des rôles symétriques. Notons par ailleurs que les analogies sur l'ensemble  $X$  induisent une analogie sur  $X^m$  où pour  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in X^m$ ,

$$\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d} \text{ si } a_i : b_i :: c_i : d_i \text{ pour chaque } i \in [1, m].$$

A partir de cette définition et en utilisant la Table 1 on peut déduire la propriété suivante qui sera utilisée dans de futures preuves :

**Property 1.** Pour tout  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{B}^m$  tel que  $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$ , on a  $h(\mathbf{a}, \mathbf{b}) = h(\mathbf{c}, \mathbf{d})$ ,  $h(\mathbf{a}, \mathbf{c}) = h(\mathbf{b}, \mathbf{d})$  et  $h(\mathbf{a}, \mathbf{d}) = h(\mathbf{b}, \mathbf{c})$ , où  $h(\mathbf{x}, \mathbf{x}')$  est la distance de Hamming, définie comme le nombre de composantes qu'on doit changer pour transformer  $\mathbf{x}$  en  $\mathbf{x}'$  (ou l'inverse).

### 3.2 Equation analogique et principe inférentiel

Une fois la notion de proportion analogique définie sur un ensemble  $X$ , étant donnés 3 éléments  $a, b, c$  de  $X$ , l'exigence que la relation  $a : b :: c : x$  soit vraie définit une équation où  $x$  est l'inconnue. Selon l'ensemble  $X$ , l'analogie  $A$  et les éléments  $a, b, c \in X$ , on peut rencontrer une des trois situations : l'équation n'a pas de solution, l'équation a une solution unique, ou l'équation a plusieurs solutions.

Dans le cas booléen,  $A_S(a, b, c, x)$  a une solution si et seulement si  $(a \leftrightarrow b) \vee (a \leftrightarrow c)$  est vrai. Ainsi  $A_S(1, 0, 0, x)$  et  $A_S(0, 1, 1, x)$  n'ont pas de solution. Quand la solution existe, elle est donnée par  $x = (a \leftrightarrow b) \leftrightarrow c$  ou plus simplement par  $x = c$  si  $a \leftrightarrow b$ , et par  $x = b$  si  $a \leftrightarrow c$ . Notons que pour  $A_K$ , l'équation a toujours une solution.

Le processus de résolution d'équation analogique joue un rôle essentiel dans le **principe d'inférence analogique** qui peut s'écrire schématiquement comme suit :

$$\frac{\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}}{f(\mathbf{a}) : f(\mathbf{b}) :: f(\mathbf{c}) : f(\mathbf{d})},$$

pour des  $m$ -uplets  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in X^m$  et une fonction  $f : X^m \rightarrow X$ . En essence, il pose que si des  $m$ -uplets  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in X^m$  sont en analogie, alors leurs images par  $f$  (et dans notre cas, leurs étiquettes) sont aussi en analogie. Ce principe peut être vu comme un cas particulier d'un autre principe appelé quelquefois *saut analogique* [10]; voir [26] sur ce point.

Le principe d'inférence analogique requière que la fonction  $f$  satisfasse deux conditions :

- (i) l'équation  $f(\mathbf{a}) : f(\mathbf{b}) :: f(\mathbf{c}) : y$  a une solution quand  $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$  tient,
- (ii) la solution  $y$  est égal à  $f(\mathbf{d})$ .

Dans la sous-section suivante, on rappelle comment ce principe peut être utilisé pour prédire des valeurs inconnues et étendre un jeu d'essai. Dans le reste de l'article on se focalise sur le cas booléen ( $X = \mathbb{B}$ ), et on recherche une caractérisation complète des fonctions booléennes qui sont complètement compatibles avec ce principe.

### 3.3 Extension d'un jeu d'essai par analogie

Etant donné un jeu d'essai  $S \subseteq \mathbb{B}^m$  et une fonction  $f : \mathbb{B}^m \rightarrow \mathbb{B}$ , on définit ce qu'on appelle **l'extension analogique** de  $S$  utilisant  $f$ , comme suit :

$$\mathbf{E}_S(f) = \{\mathbf{x} \in \mathbb{B}^m \mid \exists(\mathbf{a}, \mathbf{b}, \mathbf{c}) \in S^3, \mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{x} \text{ et } f(\mathbf{a}) : f(\mathbf{b}) :: f(\mathbf{c}) : y \text{ a une solution}\}$$

Intuitivement,  $\mathbf{E}_S(f)$  peut être vu comme l'ensemble de tous les  $\mathbf{x} \in \mathbb{B}^m$  qui sont en analogie avec au moins un 3-uplet dans  $S$ , pourvu que l'équation sur les étiquettes associées soit aussi soluble. Il est clair que  $S \subseteq \mathbf{E}_S(f)$ , et on dénote  $\mathbf{E}_S(f) \setminus S$  par  $\mathbf{E}_S^*(f)$ . Chaque élément de  $\mathbf{E}_S(f)$  reçoit une **étiquette analogique**  $\bar{\mathbf{x}}_f$ . Pour les éléments de  $\mathbf{E}_S^*(f)$ ,  $\bar{\mathbf{x}}_f$  est défini comme la prédiction la plus commune parmi les solutions candidates  $y$ . Pour les éléments de  $S$ , on rend simplement pour  $\bar{\mathbf{x}}_f$  la valeur  $f(\mathbf{x})$  qui est connue. Voici une description algorithmique du processus :

1. Tout d'abord, ajouter chaque  $\mathbf{x} \in S$  à  $\mathbf{E}_S(f)$ . Alors, pour tout  $\mathbf{a}, \mathbf{b}, \mathbf{c} \in S$  tel que  $f(\mathbf{a}) : f(\mathbf{b}) :: f(\mathbf{c}) : y$  est soluble et tel qu'il existe  $\mathbf{x} \in \mathbb{B}^m \setminus S$  with  $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{x}$ , ajouter  $\mathbf{x}$  à  $\mathbf{E}_S(f)$  et conserver  $y$  comme un candidat pour  $\bar{\mathbf{x}}_f$ . Techniquement  $\mathbf{x} \in \mathbf{E}_S^*(f)$ .
2. Puis pour chaque  $\mathbf{x} \in \mathbf{E}_S^*(f)$ , effectuer une **procédure de vote majoritaire** : prendre pour  $\bar{\mathbf{x}}_f$  le candidat le plus commun parmi toutes les solutions  $y$  en cas de d'ex-æquo, prendre une des valeurs au hasard). Pour les éléments de  $S$ ,  $\bar{\mathbf{x}}_f$  est simplement  $f(\mathbf{x})$ .

L'extension analogique peut alors être utilisée, comme un ensemble de données augmenté, par n'importe quel classifieur, en utilisant les étiquettes analogiques calculées comme si c'était la vérité terrain [16]. Il est donc naturel de s'attendre à ce que pour chaque  $\mathbf{x} \in \mathbf{E}_S^*(f)$  on ait  $\bar{\mathbf{x}}_f = f(\mathbf{x})$ . La notion de fonction PA va nous aider à formaliser cette espérance.

### 3.4 Fonctions préservant l'analogie

**Définition 1.** On dit que l'extension  $\mathbf{E}_S(f)$  est **saine** si  $\bar{\mathbf{x}}_f = f(\mathbf{x})$ , pour chaque  $\mathbf{x} \in \mathbf{E}_S^*(f)$ . Aussi, si  $\mathbf{E}_S(f)$  est sain pour tous les  $S \subseteq \mathbb{B}^m$ , on dit alors que  $f$  **préservé l'analogie** (on parle alors de "PA fonction").

La Proposition 1 fournit une définition équivalente des PA fonctions.

**Proposition 1.** Une fonction  $f : \mathbb{B}^m \rightarrow \mathbb{B}$  est PA si et seulement si pour tout  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{B}^m$ ,  $f$  satisfait la condition suivante :

$$\left\{ \begin{array}{l} \mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d} \text{ et} \\ f(\mathbf{a}) : f(\mathbf{b}) :: f(\mathbf{c}) : y \text{ est soluble} \end{array} \right. \implies y = f(\mathbf{d})$$

*Démonstration.* si  $f$  satisfait cette condition, alors il est clair d'après la sous-section 3.3 que pour tout  $S \subseteq \mathbb{B}^m$  et  $\mathbf{x} \in \mathbf{E}_S^*(f)$ , tous les candidats  $y$  pour  $\bar{\mathbf{x}}_f$  sont égal à  $f(\mathbf{x})$ , aussi  $\bar{\mathbf{x}}_f$  sera toujours  $f(\mathbf{x})$ , ce qui rend la fonction  $f$  PA.

Si  $f$  ne satisfait pas cette condition, alors il existe  $\mathbf{a}, \mathbf{b}, \mathbf{c}$ , et  $\mathbf{d} \in \mathbb{B}^m$  tels que  $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$  mais la solution  $y$  pour  $f(\mathbf{a}) : f(\mathbf{b}) :: f(\mathbf{c}) : y$  n'est pas égal à  $f(\mathbf{d})$ . En prenant  $S_0 = \{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$  on obtient  $\mathbf{E}_{S_0}^*(f) = \{\mathbf{d}\}$ , et puisque  $\bar{\mathbf{d}}_f = y \neq f(\mathbf{d})$ ,  $\mathbf{E}_{S_0}(f)$  n'est pas saine, alors  $f$  n'est pas PA.  $\square$

Si  $f$  est PA, alors l'extension  $\mathbf{E}_S(f)$  est saine pour tout  $S$ , c'est donc une extension pertinente de  $S$  qui peut être utilisée en toute confiance pour la classification. Dans la suite de l'article, on apporte une réponse au problème suivant :

**Problème 1.** Donner une description complète des fonctions  $f : \mathbb{B}^m \rightarrow \mathbb{B}$  qui assurent une extension  $\mathbf{E}_S(f)$  saine pour tout jeu d'essai  $S \subseteq \mathbb{B}^m$ . En d'autres termes, il s'agit d'identifier toutes les PA fonctions.

Notons tout d'abord que beaucoup de fonctions "naturelles" ne sont pas PA. Considérons par exemple la fonction binaire  $f(x_1, x_2) = x_1 \wedge x_2$ , avec  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{B}^2$  donnés dans la Table 2. On a  $\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d}$  et  $f(\mathbf{a}) : f(\mathbf{b}) :: f(\mathbf{c}) : y$  est soluble, mais la solution est  $y = 0$ , qui est différente de  $f(\mathbf{d}) = 1$ , donc  $f$  n'est pas PA. Cela est à relier au fait que les proportions analogiques ne sont pas stables par combinaison conjonctive. Ce serait aussi le cas pour la disjonction [25].

Dans la section suivante, on présente une description complète des fonctions PA et on répond ainsi au Problème 1.

	$x_1$	$x_2$	$f(\cdot)$
<b>a</b>	0	0	0
<b>b</b>	0	1	0
<b>c</b>	1	0	0
<b>d</b>	1	1	1

TABLE 2 –  $f(x_1, x_2) = x_1 \wedge x_2$  n'est pas PA.

## 4 La classe des fonctions PA

Il nous faut tout d'abord rappeler quelques notions de base de la théorie des variables essentielles pour les fonctions.

### 4.1 Variables essentielles et sections de fonctions

Pour  $i \in [1, m]$ ,  $\alpha \in \mathbb{B}^m$  et  $c \in \mathbb{B}$ , soit  $\alpha_i^c$   $m$ -uplet de  $\mathbb{B}^m$  obtenu en remplaçant  $\alpha_i$  par  $c$  dans  $\alpha$ . Une variable  $x_i$  est dite **inessentielle** dans  $f: \mathbb{B}^m \rightarrow \mathbb{B}$  si pour tout  $\alpha \in \mathbb{B}^m$  et  $c \in \mathbb{B}$ ,  $f(\alpha_i^c) = f(\alpha_i^{\neg c})$ . Sinon,  $x_i$  est dite **essentielle** dans  $f$ , ou encore que  $f$  dépend de  $x_i$ . En termes simples, une variable essentielle est une variable qui a la *capacité* de changer la valeur de  $f$ . Par exemple dans  $f(x_1, x_2, x_3) = x_1 \wedge x_3$ ,  $x_1$  et  $x_3$  sont des variables essentielles tandis que  $x_2$  est inessentielle. On notera  $\text{ess}(f)$  le nombre de variables essentielles de  $f$  (ou **arité essentielle**).

Deux fonctions  $f: \mathbb{B}^m \rightarrow \mathbb{B}$  et  $g: \mathbb{B}^n \rightarrow \mathbb{B}$  sont dites **équivalentes** s'il existe deux applications  $\sigma: [1, m] \rightarrow [1, n]$  et  $\sigma': [1, m] \rightarrow [1, n]$  telles que

$$\begin{aligned} f(x_1, \dots, x_m) &= g(x_{\sigma(1)}, \dots, x_{\sigma(m)}) \text{ et} \\ g(x_1, \dots, x_n) &= f(x_{\sigma'(1)}, \dots, x_{\sigma'(m)}). \end{aligned}$$

En d'autres termes,  $f$  et  $g$  sont équivalentes si l'une peut être obtenue à partir de l'autre par permutation des variables, addition de variables inessentiels, ou identification de variables inessentiels. Par exemple,  $f(x_1, x_2, x_3) = x_1 \wedge x_3$  et  $g(x_1, x_2) = x_1 \wedge x_2$  sont des fonctions équivalentes. Notons que deux fonctions équivalentes ont nécessairement le même nombre de variables essentielles. Pour plus de détails sur la théorie des variables essentielles de fonctions, voir [8, 7, 28, 31].

Dans nos démonstrations, on utilisera la propriété suivante :

**Property 2.** Soit  $f: \mathbb{B}^m \rightarrow \mathbb{B}$  et  $g: \mathbb{B}^n \rightarrow \mathbb{B}$  des fonctions équivalentes. Alors  $f$  est PA si et seulement si  $g$  est PA.

Ceci peut être vérifié en remarquant que l'analogie dans  $\mathbb{B}^m$  est définie composante par composante, que la permutation de variables n'a pas d'effet sur les équations et leur solution. De plus, la manipulation de variables inessentiels ne change pas la valeur de la fonction  $f$ , et donc la propriété PA tient encore.

Nous définissons maintenant le concept de **section** d'une fonction, aussi connu sous le nom de *restriction*, ou encore de manière équivalente comme le résultat d'une *application partielle* en informatique. Soit  $f$  une fonction  $\mathbb{B}^m \rightarrow \mathbb{B}$ , et  $(I, J)$  une partition de  $[1, m]$ . Pour  $\mathbf{x} \in \mathbb{B}^m$  et  $\alpha \in \mathbb{B}^{|J|}$ , la  $I$ -section (ou simplement la section)  $f_I^\alpha: \mathbb{B}^{|J|} \rightarrow \mathbb{B}$  est la fonction qui est obtenue en fixant toutes les variables de  $I$  aux valeurs des composantes de  $\alpha$ . Notons que l'arité de  $f_I^\alpha$  est  $|J|$ , et que  $\text{ess}(f_I^\alpha) \leq \text{ess}(f)$ . Par exemple, considérons la fonction  $f$  à trois variables :  $f(x_1, x_2, x_3) = (x_1 \wedge x_2) \vee x_3$ . La section  $f_{\{1,3\}}^{(1,0)}$  est définie par  $f_{\{1,3\}}^{(1,0)}(x_2) = (1 \wedge x_2) \vee 0 = x_2$ .

Un résultat important à propos des sections qui sera utilisé dans les autres preuves est énoncé par la Propriété 3, qu'on peut vérifier en notant que  $x : x :: x : x$  pour tout  $x \in \mathbb{B}$  :

**Property 3.** Si  $f: \mathbb{B}^m \rightarrow \mathbb{B}$  et PA, alors toute section de  $f$  est aussi PA.

Dans ce qui suit, l'opérateur ET, ' $\wedge$ ', sera noté ' $\cdot$ ' pour se conformer à une notation algébrique. De plus, ' $+$ ' dénotera maintenant l'addition modulo-2, équivalente à l'opérateur XOR. Notons que  $x + 1 = \neg x$ .

### 4.2 Les fonctions affines

Nous sommes maintenant en position d'examiner des exemples de fonctions PA. Nous allons montrer que toute fonction affine est PA.

**Proposition 2.** Soit  $L$  la classe de toutes les fonctions affines, i.e. les fonctions de la forme :

$$f(x_1, \dots, x_m) = \alpha_1 \cdot x_1 + \dots + \alpha_m \cdot x_m + \alpha,$$

avec  $\alpha_1, \dots, \alpha_m, \alpha \in \mathbb{B}$ . Toute fonction affine (aussi appelée **linéaire** quand  $\alpha = 0$ ) est PA.

*Démonstration.* Soit  $f: \mathbb{B}^m \rightarrow \mathbb{B} \in L$ . En remarquant que  $f$  est PA si et seulement si  $f + 1 = \neg f$  est PA, on peut supposer sans perte de généralité que  $\alpha = 0$ . De plus, en considérant que  $f$  dépend essentiellement de  $n \leq m$  variables ( $n$  est le nombre de  $\alpha_i$  égal à 1),  $f$  est équivalente à la fonction  $g: \mathbb{B}^n \rightarrow \mathbb{B}$  définie par  $g(x_1, \dots, x_n) = x_1 + \dots + x_n$ . En prenant en compte la Propriété 2, on a juste à prouver que  $g$  est PA pour montrer que  $f$  est aussi PA.

La fonction  $g$  a la propriété remarquable<sup>2</sup> que si on change la valeur de n'importe quel  $x_i$  cela change la valeur de  $g : \forall i, g(x_1, \dots, x_i, \dots, x_n) = \neg g(x_1, \dots, \neg x_i, \dots, x_n)$ . Compte tenu de cette propriété, il est facile de voir que :

$$\forall \mathbf{x}, \mathbf{x}' \in \mathbb{B}^n, g(\mathbf{x}) = g(\mathbf{x}') \iff h(\mathbf{x}, \mathbf{x}') \text{ est pair,}$$

2. Ceci est la raison pour laquelle les fonctions affines conduisent à des problèmes de classification qui sont, en fait, hautement **non** linéairement séparables.

où  $h$  est la distance de Hamming.

Soit  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{B}^n$  tels que les deux hypothèses de la définition de PA soient satisfaites, i.e.

$$\mathbf{a} : \mathbf{b} :: \mathbf{c} : \mathbf{d} \quad \text{et} \quad g(\mathbf{a}) : g(\mathbf{b}) :: g(\mathbf{c}) : y \quad \text{est soluble.}$$

Comme l'équation est soluble, la Table 1 nous indique qu'il y a trois cas possibles (on utilise la Propriété 1) :

1.  $g(\mathbf{a}) = g(\mathbf{b})$ , et dans ce cas la solution est  $y = g(\mathbf{c})$ . Comme  $g(\mathbf{a}) = g(\mathbf{b})$ , alors  $h(\mathbf{a}, \mathbf{b})$  est pair, ainsi que  $h(\mathbf{c}, \mathbf{d})$ . Alors,  $g(\mathbf{c}) = g(\mathbf{d})$  et  $y = g(\mathbf{d})$ .
2.  $g(\mathbf{a}) = g(\mathbf{c})$ , et dans ce cas la solution est  $y = g(\mathbf{b})$ . Comme  $g(\mathbf{a}) = g(\mathbf{c})$ , alors  $h(\mathbf{a}, \mathbf{c})$  est pair, ainsi que  $h(\mathbf{b}, \mathbf{d})$ . Donc  $g(\mathbf{b}) = g(\mathbf{d})$  et  $y = g(\mathbf{d})$ .
3.  $\neg g(\mathbf{a}) = g(\mathbf{b}) = g(\mathbf{c})$ , et dans ce cas la solution est  $y = g(\mathbf{a})$ . Comme  $g(\mathbf{b}) = g(\mathbf{c})$ , alors  $h(\mathbf{b}, \mathbf{c})$  est pair, ainsi que  $h(\mathbf{a}, \mathbf{d})$ . Donc  $g(\mathbf{a}) = g(\mathbf{d})$  et  $y = g(\mathbf{d})$ .

Le troisième cas est seulement pertinent pour le modèle de Klein. Dans tous les cas on a  $y = g(\mathbf{d})$ , ce qui montre que  $g$  est PA, et par suite tout  $f \in L$ .  $\square$

### 4.3 Une description complète des fonctions PA

Nous venons de voir que toute fonction affine est PA. Nous allons maintenant établir un résultat plus fort : les fonctions affines sont les **seules** fonctions PA.

Pour cela on va utiliser la représentation polynomiale des fonctions booléennes. Un **monôme** est un terme de la forme :

$$\mathbf{x}_I = \prod_{i \in I} x_i,$$

pour un ensemble quelconque possiblement vide d'entiers positifs  $I$ , où  $|I|$  est appelé le **degré** de  $\mathbf{x}_I$ . On prend la convention que 1 est le monôme vide  $\mathbf{x}_\emptyset$ . Un **polynôme** est une somme de monômes et son degré est le plus grand degré de ses monômes. Il est bien connu [29, 33] que toute fonction  $f : \mathbb{B}^m \rightarrow \mathbb{B}$  est représentée de façon unique par un tel polynôme, aussi appelé sa Forme Normale Algébrique :

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

où chaque  $a_I$  appartient à  $\mathbb{B}$ . Notons que la fonction constante 0 est représentée par  $a_\emptyset \cdot \mathbf{x}_\emptyset$  avec  $a_\emptyset = 0$ . Le degré d'une fonction  $f : \mathbb{B}^m \rightarrow \mathbb{B}$ , noté  $d(f)$ , est défini comme le degré du polynôme unique représentant  $f$ .

Notons que la classe des fonctions de degré au plus 1 est exactement la classe  $L$  des fonctions affines, qui sont PA. On va montrer que la classe des PA fonctions est la classe des fonctions affines en prouvant que si une fonction  $f$  est PA, alors  $d(f) \leq 1$ . On considère d'abord le cas où  $d(f) = 2$ .

**Property 4.** Soit  $f : \mathbb{B}^m \rightarrow \mathbb{B}$  avec  $d(f) = 2$ .  $f$  n'est pas PA.

*Démonstration.* Considérons  $f$  avec  $d(f) = 2$  et  $\text{ess}(f) \geq 2$ . On note  $\mathbf{x}_I$  un des monômes de  $f$  de degré 2. On considère la section  $f_J^0$ , où  $J = [1, m] \setminus I$ , où  $\mathbf{0}$  dénote le vecteur constant 0 in  $\mathbb{B}^{|J|}$ . Toutes les variables qui ne font pas partie du monôme  $\mathbf{x}_I$  ont été mises à 0. La section  $f_J^0$  a un monôme unique de degré 2 ( $\mathbf{x}_I$ ) et  $\text{ess}(f) = 2$ .  $f_J^0$  est nécessairement équivalente à une des fonctions suivantes :

$$\begin{cases} f_1(x_1, x_2) = x_1 \cdot x_2 + \alpha \\ f_2(x_1, x_2) = x_1 \cdot x_2 + x_1 + \alpha \\ f_3(x_1, x_2) = x_1 \cdot x_2 + x_1 + x_2 + \alpha \end{cases}$$

Il est assez facile de trouver des exemples de  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{B}^2$  pour montrer qu'aucune de ces fonctions n'est PA (voir la Table 2), so  $f_J^0$  ne peut pas être PA non plus. Comme  $f_J^0$  est une section  $f$ ,  $f$  n'est pas PA.  $\square$

Ce dont nous avons besoin maintenant c'est d'une propriété qui nous permettrait de décroître le degré d'une fonction sans changer sa propriété PA. C'est le rôle de la Propriété 5.

**Property 5.** Soit  $f : \mathbb{B}^m \rightarrow \mathbb{B}$  une fonction telle que  $d(f) = k \geq 2$ . Alors il y a une section  $g$  de  $f$  telle que  $d(g) = k - 1$ .

*Démonstration.* Supposons que  $d(f) = k \geq 2$ , et soit  $\mathbf{x}_I$  un monôme de  $f$  de degré maximum, i.e.  $|I| = k$ . A nouveau, considérons la section  $g = f_J^0$  où  $J = [1, m] \setminus I$  et  $\mathbf{0}$  dénote le vecteur constant 0 de  $\mathbb{B}^{|J|}$ . Il est clair que  $g$  est représenté par un polynôme qui a un unique monôme de degré maximal  $k$ , c'est-à-dire  $\mathbf{x}_I$ , et peut-être quelques autres monômes de degré strictement inférieur à  $k$ . Prenons un  $i \in I$  : alors  $g' = g_{\{i\}}^1$  est une section de  $g$  de degré (et d'arité)  $k - 1$ . Comme  $g'$  est une section de  $g$ , c'est aussi une section de  $f$ , ce qui complète la preuve.  $\square$

Nous sommes maintenant en position d'établir le résultat principal.

**Proposition 3.** La classe des fonctions PA est la classe  $L$  des fonctions affines.

*Démonstration.* Nous avons vu que toute fonction affine est PA, i.e. si  $d(f) \leq 1$ , alors  $f \in PA$ . D'autre part, la Propriété 4 nous dit que si  $d(f) = 2$ , alors  $f \notin PA$ . Aussi supposons que  $d(f) \geq 3$ . Par des applications successives de la Propriété 5, il s'ensuit que qu'il y a une section  $g$  de  $f$  avec  $d(g) = 2$ . Comme  $g$  n'est pas PA,  $f$  n'est pas PA non plus à cause de la Propriété 3. En résumé, si  $d(f) \geq 2$  alors  $f$  n'est pas PA, at donc la classe des fonctions PA est exactement  $L$ .  $\square$

On peut donc apporter une réponse finale au problème initial : **la classe des fonctions qui assurent une extension**

saine pour tout jeu d'essai  $S \subseteq \mathbb{B}^m$  est la classe  $L$  des fonctions affines. Si la fonction n'est pas affine, alors il existe un jeu d'essai  $S_0 \subseteq \mathbb{B}^m$  pour lequel  $\mathbf{E}_{S_0}(f)$  n'est pas sain.

Cependant, tandis que ce résultat théorique est intéressant en soi, il est évident que les fonctions purement affines ne sont pas représentatives de ce que l'on peut rencontrer dans la réalité. Ceci conduit au **Problème 2** : Que reste-t-il de la qualité de l'extension analogique  $\mathbf{E}_S(f)$  quand  $f$  s'écarte d'être PA de différentes manières ? Le but de la section suivante est de commencer à étudier empiriquement cette question.

## 5 Fonctions approximativement PA et expériences

Ici, nous étudions d'abord la qualité de l'extension analogique quand la fonction  $f$  n'est pas exactement affine. Etant donné un échantillon  $S$ , nous définissons  $\omega(S, f)$  (ou simplement  $\omega$ ) comme la **qualité** de l'extension  $\mathbf{E}_S(f)$  :

$$\omega(S, f) = P_{\mathbf{x} \in \mathbf{E}_S^*(f)} [\bar{\mathbf{x}}_f = f(\mathbf{x})],$$

où  $P_{\mathbf{x}}$  est la distribution uniforme sur  $\mathbb{B}^m$ .

Ici,  $\mathbf{x} \in \mathbb{B}^m$  est considérée comme une variable aléatoire. Par définition,  $\omega(S, f) = 1$  pour tout  $S$  si et seulement si  $f$  est PA. Il est clair que  $\omega(S, f) = \omega(S, \neg f)$  pour tout  $S$  et pour toute  $f$ , simplement à cause de la propriété de *code independance* (voir Section 3.1).

Etant données 2 fonctions booléennes  $f$  et  $g$ , on définit leur distance comme  $\text{dist}(f, g) = P_{\mathbf{x}} [f(\mathbf{x}) \neq g(\mathbf{x})]$ . On dit que  $f$  est  $\epsilon$ -proche de  $g$  si  $\text{dist}(f, g) \leq \epsilon$ , et que  $f$  est  $\epsilon$ -proche d'un ensemble  $\Sigma$  s'il  $\exists g \in \Sigma$  telle que  $f$  est  $\epsilon$ -proche de  $g$ .

Nous étudions les variations de  $\omega$  quand  $f$  est  $\epsilon$ -proche de  $L$ . En fait, comme l'ensemble des fonctions affines est constitué de l'ensemble des fonctions linéaires ainsi que de leurs compléments, des propriétés de  $\omega$ , on déduit qu'il suffit de considérer que  $f$  est  $\epsilon$ -proche de l'ensemble des fonctions linéaires. Ce qui est très utile car, en pratique, on peut utiliser le test BLR [3] qui permet d'interroger une fonction partiellement connue pour découvrir si elle est  $\epsilon$ -proche de l'ensemble des fonctions linéaires.

Partant de  $g: \mathbb{B}^8 \rightarrow \mathbb{B}$  linéaire définie comme  $g(\mathbf{x}) = x_1 + \dots + x_8$ , on introduit du bruit en niant certaines entrées  $g(\mathbf{x})$  avec une fréquence  $\epsilon$ . On obtient ainsi une fonction  $f_\epsilon$  qui est nécessairement  $\epsilon$ -proche de l'ensemble des fonctions linéaires.

Dans la Figure 1, nous présentons les valeurs de  $\omega(S, f_\epsilon)$  avec le modèle Standard pour différentes tailles de  $S$  (exprimées comme un pourcentage de  $|\mathbb{B}^m|$ ). Les résultats sont des moyennes sur 50 expériences. D'autres expériences ont été faites avec d'autres fonctions linéaires (i.e. avec moins de variables essentielles ou avec une arité différente). Les résultats obtenus sont similaires.

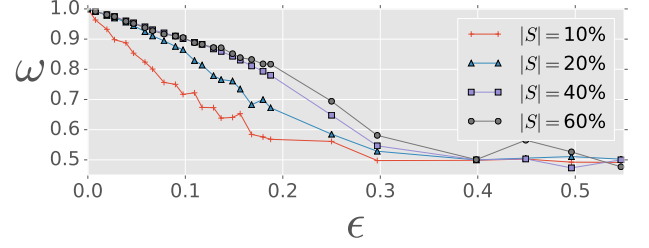


FIGURE 1 – Valeur de  $\omega$  pour des fonctions  $\epsilon$ -proches de  $L$ .

	$\omega_S$	$\omega_K$	$\beta_S$	$\beta_K$
Monk 1	.96	.96	.73	.62
Monk 2	.96	.84	.69	.60
Monk 3	.98	.95	.87	.77

TABLE 3 –  $\omega$  et  $\beta$  pour les modèles Standard et Klein sur les problèmes Monks.

Notons qu'il suffit de considérer  $\epsilon \in [0, \frac{1}{2}]$ , car si  $g$  est  $\epsilon$ -proche de  $L$ ,  $\neg g$  est  $(1 - \epsilon)$ -proche de  $L$  et les 2 fonctions ont donc le même  $\omega$ , ainsi les courbes sont symétriques par rapport à l'axe  $\epsilon = \frac{1}{2}$ .

Quand  $\epsilon = 0$ , on obtient  $\omega = 1$ , comme attendu par la Proposition 3. On observe une décroissance presque linéaire en  $\omega$  quand  $\epsilon$  varie de 0.3 – 0.4 puis conduisant à un plateau où  $\omega = \frac{1}{2}$ , indiquant que les labels analogiques  $\bar{\mathbf{x}}_f$  sont plus ou moins aléatoires. De plus,  $\omega$  décroît plus vite pour de petits échantillons  $S$ .

Cela est dû au fait que les labels  $\bar{\mathbf{x}}_f$  résultent d'un vote majoritaire et que le nombre de candidats devient plus petit quand  $|S|$  décroît, altérant ainsi la qualité de la prédiction. Reste à déterminer une dépendance fonctionnelle entre  $\omega$ ,  $\epsilon$  et  $|S|$ .

Maintenant, notons que, même si la fonction  $f$  est loin d'être PA, la qualité  $\omega$  de l'extension  $\mathbf{E}_S(f)$  peut rester haute.

Pour illustrer ce fait, définissons la valeur  $\beta$  qui sera un indicateur de combien  $f$  est loin d'être parfaitement PA. Pour chaque  $\mathbf{x} \in \mathbf{E}_S^*(f)$ , on définit  $\beta_{\mathbf{x}}$  comme la proportion de candidats  $y$  menant au label correct, i.e. la proportion de  $y$  tels que  $y = f(\mathbf{x})$ .  $\beta$  est définie comme la moyenne de tous les  $\beta_{\mathbf{x}}$ . Clairement, la fonction  $f$  est PA si et seulement si  $\beta = 1$  pour tout  $S$ , i.e. si  $\beta_{\mathbf{x}} = 1$  pour tout  $\mathbf{x} \in \mathbf{E}_S^*(f)$  et pour tout  $S$ .

La Table 3 donne les valeurs de  $\omega$  et  $\beta$  pour les modèles Standard et Klein (respectivement  $\omega_S$ ,  $\omega_K$ ,  $\beta_S$  et  $\beta_K$ ) sur 3 jeux de données du dépôt UCI, précisément les 3 problèmes 'Monk'<sup>3</sup> [20]. Les résultats sont une moyenne sur 100 expériences, où  $S$  est à chaque fois choisi aléatoirement comme 30% de l'univers  $X$ .

On observe que sur chaque jeu de données,  $\beta_S$  est significativement plus petit que 1. Cela suggère que les fonctions

3. Comme ils sont nominaux, nous les binarisons.

sous jacentes sont hautement non PA, car en moyenne, il y a une grande proportion (aux alentours de 20%) de candidats  $y$  qui prédisent le mauvais label. Cependant,  $\omega_S$  n'est jamais plus petit que 96%, ce qui génère des extensions de bonne qualité. C'est là que le vote majoritaire entre en jeu : dans certains cas, ce vote compense les fausses prédictions. C'est ce qui se produit dans 96%, 96% et 98% des cas respectivement.

Ici encore, nous cherchons à obtenir des garanties théoriques sur cette procédure de vote majoritaire.

Notons aussi que le modèle de Klein produit des résultats de la même qualité ou de qualité inférieure au modèle standard. Cela suggère que le modèle standard, qui définit la même classe de fonctions PA, est plus utile en pratique. Notons que la différence entre  $\omega_S$  et  $\omega_K$  a été observée de manière consistante au travers de diverses expériences que nous ne mentionnons pas ici faute de place. Cela pourrait s'expliquer par le fait que le modèle de Klein satisfait une propriété peu attendue d'une analogie :  $A_K(a, b, c, d) \iff A_K(b, a, c, d)$ .

## 6 Conclusion et pistes futures

Dans ce papier, nous avons démontré l'intérêt de l'usage des proportions analogiques pour étendre un échantillon de valeurs booléennes. Nous avons introduit la notion de fonctions PA, qui sont les fonctions pour lesquelles nous sommes sûrs d'obtenir une extension sans erreur. Après avoir identifié la classe des fonctions PA comme étant celle des fonctions linéaires, nous avons discuté comment un tel résultat théorique pourrait être utile dans les cas réels.

Nous avons étudié 2 manières de dévier du cas PA. Premièrement, en étudiant les fonctions approximativement linéaires et en observant les variations dans la qualité de l'extension analogique.

Deuxièmement, en étudiant des ensembles de données dont la fonction sous jacente est clairement non linéaire mais pour lesquels les classifieurs analogiques sont performants. Ces 2 investigations empiriques suggèrent que nous devons rechercher des garanties théoriques sur la qualité de l'extension analogique dans ces 2 contextes.

Il est aussi clair que ce type de résultat doit aussi être recherché pour des attributs nominaux (et pas seulement booléens.)

**Acknowledgements.** Ce travail bénéficie du soutien du Labex ANR-11-LABX-0040-CIMI (Centre International de Mathématiques et d'Informatique) dans le cadre du programme ANR-11-IDEX-0002-02, projet ISIPA.

## Références

[1] Bayouhd, S., L. Miclet et A. Delhay: *Learning by*

*analogy : A classification rule for binary and nominal data.* pages 678–683, 2007.

- [2] Bayouhd, S., H. Mouchère, L. Miclet et E. Anquetil: *Learning a Classifier with Very Few Examples : Analogy Based and Knowledge Based Generation of New Examples for Character Recognition.* Dans *Proc. 18th Europ. Conf. on Machine Learning (ECML'07)*, pages 527–534. Springer-Verlag, 2007.
- [3] Blum, M., M. Luby et R. Rubinfeld: *Self-testing/Correcting with Applications to Numerical Problems.* *J. Comput. Syst. Sci.*, 47(3) :549–595, 1993, ISSN 0022-0000.
- [4] Bounhas, M., H. Prade et G. Richard: *Analogical classification : A new way to deal with examples.* Dans *Proc. 21st Europ. Conf. on Artificial Intelligence (ECAI'14)*, pages 135–140. IOS Press, 2014.
- [5] Bounhas, M., H. Prade et G. Richard: *Not being at odds with a class : A new way of exploiting neighbors for classification.* Dans *Proc. 22nd Europ. Conf. on Artificial Intelligence (ECAI'16)*, tome 285, pages 1662–1663. IOS Press, 2016.
- [6] Couceiro, M., N. Hug, H. Prade et G. Richard: *Analogy-preserving functions : A way to extend Boolean samples.* Dans *Proc. 26th Int. J. Conf. Artif. Intel. IJCAI'17, Melbourne*, 2017.
- [7] Couceiro, M. et E. Lehtonen: *Generalizations of Swierczkowski's lemma and the arity gap of finite functions.* *Discrete Math.*, 309(20) :5905–5912, 2009.
- [8] Couceiro, M. et M. Pouzet: *On a quasi-ordering on Boolean functions.* *Theor. Comput. Sci.*, 396(1-3) :71–87, 2008.
- [9] Dastani, M., B. Indurkha et R. Scha: *Analogical projection in pattern perception.* *J. of Experimental and Theoretical Artificial Intelligence*, 15(4) :489–511, 2003.
- [10] Davies, T. R. et S. J. Russell: *A logical approach to reasoning by analogy.* Dans McDermott, J. P. (éditeur) : *Proc. 10th Int. Joint Conf. on Artificial Intelligence (IJCAI'87)*, pages 264–270. Morgan Kaufmann, 1987.
- [11] Dorolle, M.: *Le Raisonnement par Analogie.* PUF, Paris, 1949.
- [12] Fei-Fei Li, R. Fergus et P. Perona: *One-shot learning of object categories.* *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(4) :594–611, 2006.
- [13] Gentner, D.: *Structure-Mapping : A Theoretical Framework for Analogy.* *Cognitive Science*, 7(2) :155–170, 1983.
- [14] Gentner, D., K. J. Holyoak et B. N. Kokinov (éditeurs) : *The Analogical Mind : Perspectives from Cog-*



- nitive Science*. Cognitive Science, and Philosophy. MIT Press, Cambridge, MA, 2001.
- [15] Gust, H., K. U. Kühnberger et U. Schmid: *Metaphors and heuristic-driven theory projection (HDTP)*. Theoretical Computer Science, 354(1) :98 – 117, 2006.
- [16] Hug, N., H. Prade, G. Richard et M. Serrurier: *Analogical classifiers : A theoretical perspective*. Dans *Proc. 22nd Europ. Conf. on Artificial Intelligence (ECAI'16)*, pages 689–697. IOS Press, 2016.
- [17] Klein, S.: *Analogy and mysticism and the structure of culture (and Comments & Reply)*. Current Anthropology, 24 (2) :151–180, 1983.
- [18] Lepage, Y.: *Analogy and formal languages*. Electr. Notes Theor. Comput. Sci., 53, 2001.
- [19] Lepage, Y.: *De l'analogie rendant compte de la commutation en linguistique*. Habilit. à Diriger des Recher., Univ. J. Fourier, Grenoble, 2003.
- [20] Lichman, M.: *UCI Machine Learning Repository*, 2013. <http://archive.ics.uci.edu/ml>.
- [21] Miclet, L., S. Bayouh et A. Delhay: *Analogical Dissimilarity : Definition, Algorithms and Two Experiments in Machine Learning*. J. Artif. Intell. Res. (JAIR), 32 :793–824, 2008.
- [22] Miclet, L. et A. Delhay: *Relation d'analogie et distance sur un alphabet défini par des traits*. Rapport technique 1632, IRISA, July 2004.
- [23] Miclet, L. et H. Prade: *Handling analogical proportions in classical logic and fuzzy logics settings*. Dans *Proc. 10th Eur. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'09)*, pages 638–650. Springer, LNCS 5590, 2009.
- [24] Prade, H. et G. Richard: *Boolean analogical proportions - Axiomatics and algorithmic complexity issues*. Dans Antonucci, A., L. Cholvy et O. Papini (rédacteurs) : *Proc. 14th Eur. Conf. on Symb. and Quantit. Appr. to Reasoning with Uncertainty (ECSQARU'17)*, Lugano, Jul. 7-10, LNAI 10369. Springer, 2012.
- [25] Prade, H. et G. Richard: *From Analogical Proportion to Logical Proportions*. Logica Universalis, 7(4) :441–505, 2013.
- [26] Prade, H. et G. Richard: *A discussion of analogical-proportion based inference*. Dans *Proc. 2nd Computational Analogy Workshop at ICCBR-17, Trondheim*, 2017.
- [27] Rumelhart, D. E. et A. A. Abrahamson: *A model for analogical reasoning*. Cognitive Psychol., 5 :1–28, 2005.
- [28] Salomaa, A.: *On essential variables of functions, especially in the algebra of logic*. Ann. Acad. Sci. Fenn. Ser. A I. Math., 339 :3–11, 1963.
- [29] Stone, M. H.: *The Theory of Representation for Boolean Algebras*. Trans. of the American Mathematical Society, 40(1) :37–111, 1936.
- [30] Stroppa, N. et F. Yvon: *Analogical Learning and Formal Proportions : Definitions and Methodological Issues*. Rapport technique D004, ENST-Paris, 2005.
- [31] Willard, R.: *Essential arities of term operations in finite algebras*. Discrete Math., 149(1-3) :239–259, 1996.
- [32] Wolf, L. et I. Martin: *Regularization Through Feature Knock Out*. MIT Computer Science and Artificial Intelligence Laboratory, (CBCL Memo 242), 2004.
- [33] Zhegalkin, I. I.: *On the Technique of Calculating Propositions in Symbolic Logic*. Mat. Sb, 43 :9–28, 1927.