
Proportions analogiques et créativité: une étude préliminaire

Stergos Afantenos¹ Henri Prade¹ Gilles Richard¹ Leonardo Cortez Bernardes¹

¹IRIT – CNRS, 118, route de Narbonne, 31062 Toulouse Cedex 9, France

stergos.afantenos@irit.fr

henri.prade@irit.fr

gilles.richard@irit.fr

leonardo.cortez_bernardes@irit.fr

Résumé

Les proportions analogiques sont des énoncés de la forme “ a est à b comme c est à d ”, qui expriment que la comparaison des éléments de la paire (a, b) et de la paire (c, d) donne des résultats semblables. Les proportions analogiques sont créatives dans le sens où, étant donné 3 éléments distincts, on peut calculer la représentation d’un quatrième élément d , distinct des éléments précédents, qui forme avec eux une proportion analogique, si certaines conditions sont remplies. Après une introduction aux proportions analogiques et à leurs propriétés, l’article rapporte les résultats d’une expérience réalisée avec une base de données de descriptions d’animaux et de leur classes, où nous tentons de “créer” de nouveaux animaux à partir de ceux existants, ou de retrouver des animaux rares tels que l’ornithorynque. Nous effectuons une série d’expériences en utilisant des plongements lexicaux de mots ainsi que des caractéristiques booléennes afin de proposer de nouveaux animaux basés sur des proportions analogiques, en comparant l’approche symbolique avec des plongements lexicaux non contextuels, montrant que de tels plongements lexicaux de mots obtiennent également de bons résultats. En outre, nous fournissons une série d’expériences avec des plongements lexicaux de phrases utilisant des plongements lexicaux contextuels qui sont moins concluants. Enfin, nous proposons également un processus créatif plus sophistiqué basé sur les proportions analogiques où l’ensemble des paires (a, b) utilisées est choisi et élargi au moyen d’une opération de préservation des propriétés.

Abstract

Analogical proportions are statements of the form “ a is to b as c is to d ”, which express that the comparison of the elements in pair (a, b) and in pair (c, d) yield similar results. Analogical proportions are creative in the sense that given 3 distinct items, the representation of a 4th item d , distinct from the previous items, which forms an analogical proportion with them, can be calculated, provided certain conditions are met. After an introduction to analogical proportions

and their properties, the paper reports the results of an experiment made with a database of animal descriptions and their class, where we try to ‘create’ new animals from existing ones, retrieving rare animals such as platypus. We perform a series of experiments using word embeddings as well as Boolean features in order to propose novel animals based on analogical proportions, comparing the symbolic approach with non-contextual embeddings showing that such word embeddings obtain also good results. Furthermore we provide a series of experiments with sentence embeddings using contextual embeddings which are less conclusive. Finally, we also propose a more sophisticated creative process based on analogical proportions where the set of pairs (a, b) used is chosen and enlarged by means of a property-preserving operation.

Introduction

La créativité a suscité un grand intérêt depuis longtemps dans les sciences informatiques et en intelligence artificielle [2, 47, 6, 23] avec des applications dans de nombreux domaines. Le raisonnement analogique a toujours été reconnu pour favoriser la créativité, notamment dans la pensée créative et la résolution de problèmes [20, 16, 50]. En effet, le raisonnement analogique établit un parallèle entre deux situations, suggérant que ce qui est vrai ou applicable dans la première situation pourrait également être vrai ou applicable dans la seconde situation qui présente une certaine similitude avec la première.

Les proportions analogiques [38] (PA) sont des relations quaternaires notées $a : b :: c : d$ entre quatre éléments a, b, c, d , qui se lisent “ a est à b comme c est à d ”. Dans la suite, a, b, c, d sont représentés au moyen de vecteurs ; ces vecteurs peuvent être constitués soit de valeurs d’attributs soit de plongements lexicaux de mots en anglais “word embeddings”) [34]. Les proportions analogiques peuvent être

considérées comme des éléments constitutifs du raisonnement analogique. En effet, elles établissent des parallèles entre les paires ordonnées (a, b) et (c, d) . Par exemple, « le veau est à la vache comme le poulain est à la jument » met les bovidés sur un pied d'égalité avec les équidés. Dans de telles proportions analogiques, les quatre éléments sont décrits au moyen du même ensemble d'attributs. Les proportions analogiques ont été appliquées avec succès à la classification [32, 5, 4], à la prédiction de préférences [12, 3], ou pour résoudre des tests de QI de type Raven [8, 43].

Les travaux sur les plongements lexicaux [34] ont montré que les modèles de langage où les mots sont représentés par des vecteurs réels ont le potentiel de respecter les proportions analogiques dans un espace vectoriel, bien que des approches ultérieures aient montré que cela était dû au corpus limité d'analogies qui était utilisé, proposant de meilleures ressources pour tester les analogies [15, 52]. Récemment, [21] ont montré que les grands modèles de langage ont la capacité de résoudre des problèmes de Raven, pourvu qu'ils leur soient présentés dans une description en langage naturel, atteignant au moins le même niveau que les êtres humains. À notre connaissance, les capacités créatives des analogies à produire quelque chose de nouveau n'ont pas été explorées, notamment en langage naturel.

Cet article présente une étude du pouvoir créatif des proportions analogiques. Ce pouvoir repose sur leur capacité à produire un quatrième élément à partir de trois éléments (à condition que certaines conditions soient remplies). L'article est organisé en deux sections principales. La première rappelle les définitions et propriétés de base des proportions analogiques, ainsi qu'une nouvelle procédure avancée pour la créativité guidée. La seconde propose une série d'expériences utilisant l'ensemble de données Zoo¹. Initialement, nous générons des descriptions symboliques d'animaux à partir de celles existantes et vérifions si ces descriptions existent déjà dans la base de données. Nous menons ensuite deux autres séries d'expériences : l'une utilisant des représentations vectorielles de mots, l'autre traitant de phrases descriptives. Ces expériences utilisent des incorporations de mots, soit par le biais d'incorporations statiques GloVe [35], soit par des incorporations de phrases contextuelles avec BERT [10, 45]. Cela est complété par un exemple de matrice de Raven résolu au moyen de proportions analogiques, ainsi qu'une discussion conclusive.

Proportions Analogiques

Cette section est structurée en cinq sous-parties : i) rappelant la modélisation logique booléenne, les postulats et les propriétés des proportions analogiques (PA), ii) fournissant un exemple montrant leur pouvoir créatif, iii) traitant

des valeurs d'attributs nominaux, iv) introduisant des PA imbriquées, et v) traitant des PA de mots.

Tables de vérité, postulats et propriétés

Un modèle logique d'une PA « a est à b comme c est à d » où a, b, c, d sont des variables booléennes est donné par la formule suivante qui dit que a diffère de b comme c diffère de d et que b diffère de a comme d diffère de c [33] :

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

Cette expression est vraie pour les 6 valuations données dans la Table 1 et fausse pour les $2^4 - 6 = 10$ autres valuations possibles. C'est le modèle booléen minimal qui

a	b	c	d
0	0	0	0
1	1	1	1
0	0	1	1
1	1	0	0
0	1	0	1
1	0	1	0

TABLE 1 – Valuations booléennes validant $a : b :: c : d$

satisfait les trois postulats de base suivants (inspirés des proportions numériques) qu'une PA doit respecter [37] :

- *réflexivité* : $a : b : a : b$;
- *symétrie* : $a : b :: c : d \Rightarrow c : d :: a : b$;
- *permutation centrale* : $a : b :: c : d \Rightarrow a : c :: b : d$.

En conséquence, nous avons :

- $a : a : b : b$ (*identité*),
- $a : b :: c : d \Rightarrow b : a :: d : c$ (*inversion interne*), et
- $a : b :: c : d \Rightarrow d : b :: c : a$ (*permutation des extrêmes*).

Remarquablement, les proportions analogiques booléennes sont *indépendantes du code*, c.-à-d. que $a : b :: c : d \Rightarrow \neg a : \neg b :: \neg c : \neg d$. Ainsi, toute propriété utilisée pour décrire des éléments peut être encodée de manière positive ou négative.

Nous supposons que les éléments considérés sont représentés par des *vecteurs* booléens avec n composantes correspondant à n valeurs de caractéristiques, c'est-à-dire $\vec{a} = (a_1, \dots, a_n)$, etc. Une proportion analogique " \vec{a} est à \vec{b} comme \vec{c} est à \vec{d} ", notée $\vec{a} : \vec{b} :: \vec{c} : \vec{d}$, est définie composante par composante :

$$\vec{a} : \vec{b} :: \vec{c} : \vec{d} \text{ si et seulement si } \forall i \in [1, n], a_i : b_i :: c_i : d_i$$

Les proportions analogiques sont créatives

À titre d'illustration, considérons le problème d'analogie géométrique de la Figure 1 ci-après [8, 38]. Il peut être codé avec cinq prédicats booléens : *Rectangle* (R), *PointNoir* (PN), *Triangle* (T), *Cercle* (C), *Ellipse* (E), dans cet ordre

1. <https://archive.ics.uci.edu/dataset/111/zoo>

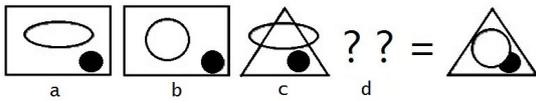


FIGURE 1 – Un problème d’analogie géométrique

dans le Tableau 2, où a, b, c sont codés. Chaque colonne est une équation de PA : $a_i : b_i :: c_i : x_i$.

Une équation $a : b :: c : x$ n’a pas toujours une solution. En effet, les équations $1 : 0 :: 0 : x$ et $0 : 1 :: 1 : x$ n’ont pas de solution. Si $a : b :: c : x$ a une solution, elle est *unique*. Elle est donnée par $x = c \equiv (a \equiv b)$, où \equiv représente le connecteur d’équivalence. C’est le cas dans l’exemple, où $\vec{x} = \vec{d} = (01110)$ dans la Table 2, représenté sur la droite de la Figure 2.

Il convient de souligner que dans ce cas, la description de \vec{d} est calculée directement à partir de celles de $\vec{a}, \vec{b}, \vec{c}$, ce qui contraste avec ce genre de tests de QI où la réponse doit être trouvée parmi un ensemble de solutions candidates. En effet, le premier programme pour résoudre de tels problèmes [11], dans les premières années de l’IA, consistait à sélectionner la solution parmi les solutions candidates sur la base d’une distance tenant compte de la similarité des transformations pour passer de \vec{a} à \vec{b} et de \vec{c} à \vec{x} (où \vec{x} est la solution candidate considérée). Le fait que la description de \vec{x} soit maintenant directement calculée à partir de $\vec{a}, \vec{b}, \vec{c}$ montre le *pouvoir créatif* de la proportion analogique. De plus, $\vec{d} = \vec{x}$ est distinct de \vec{a}, \vec{b} et \vec{c} . C’est général pour $\vec{a} : \vec{b} :: \vec{c} : \vec{d}$ dès que \vec{a} est distinct de \vec{b} et distinct de \vec{c} (sur au moins deux attributs).

	R	PN	T	C	E
\vec{a}	1	1	0	0	1
\vec{b}	1	1	0	1	0
\vec{c}	0	1	1	0	1
\vec{x}	?	?	?	?	?

TABLE 2 – Codage de l’exemple. $\vec{x} = (01110)$

Attributs nominaux

La description des éléments peut impliquer des attributs nominaux, c’est-à-dire des attributs avec un domaine fini ayant une cardinalité supérieure à 2. Alors $a : b :: c : d$ est vrai pour les variables nominales si et seulement si :

$$(a, b, c, d) \in \{(s, s, s, s), (s, t, s, t), (s, s, t, t) \mid s, t \in \mathcal{A}, s \neq t\}$$

où s, t représentent n’importe quelle valeur du domaine d’attributs \mathcal{A} (comme suggéré pour la première fois dans [36], voir aussi [5]). Cela généralise le cas booléen et préserve toutes les propriétés rapportées pour ce cas.

Un exemple de proportions analogiques imbriquées

Nous illustrons maintenant les valeurs nominales avec un exemple de PA (de [25]) entre quatre phrases $a =$ "les filles

détestent la lumière", $b =$ "les garçons aiment la lumière", $c =$ "les femmes détestent l’obscurité", $d =$ "les hommes aiment l’obscurité", vues comme des tuples ordonnés de valeurs nominales, par rapport aux 3 attributs "sujet", "verbe", "complément". Les 4 phrases a, b, c, d sont donc représentées par des vecteurs à 3 composantes $\vec{a}, \vec{b}, \vec{c}, \vec{d}$:

	sujet	verbe	complément
\vec{a}	filles	détestent	lumière
\vec{b}	garçons	aiment	lumière
\vec{c}	femmes	détestent	obscurité
\vec{d}	hommes	aiment	lumière

TABLE 3 – Proportion analogique entre phrases

Évidemment, les valeurs dans la colonne "sujet" du Tableau 3 ne forment pas une proportion analogique par elles-mêmes, puisque quatre valeurs distinctes sont impliquées (alors que *déteste : aime :: déteste : aime*, et *lumière : lumière :: obscurité : obscurité* sont des proportions analogiques claires en termes de valeurs nominales). Cependant, *filles : garçons :: femmes : hommes* peut être considéré comme une écriture compacte d’une PA entre des descriptions en termes d’une collection de caractéristiques booléennes comme dans le Tableau 4, où une PA est vérifiée dans chaque colonne. Ainsi, $\vec{a} : \vec{b} :: \vec{c} : \vec{d}$, et donc $a : b :: c : d$ sont vraies. La PA $\vec{a} : \vec{b} :: \vec{c} : \vec{d}$ est une PA *imbriquée* car elle implique un attribut dont les quatre valeurs ne forment pas l’un des trois schémas de base d’une proportion analogique nominale, alors que ces valeurs sont associées à des descriptions vectorielles qui forment elles-mêmes une PA.

	<i>male</i>	<i>femelle</i>	<i>jeune</i>	<i>adulte</i>	<i>humain</i>
filles	0	1	1	0	1
garçons	1	0	1	0	1
femmes	0	1	0	1	1
hommes	1	0	0	1	1

TABLE 4 – Les filles sont aux garçons ce que les femmes sont aux hommes

Proportions analogiques sur les mots

Dans l’exemple du Tableau 3, les valeurs des composantes du vecteur sont des mots. Les mots peuvent eux-mêmes être représentés par des plongements lexicaux [34]. Les proportions analogiques peuvent être directement définies en termes de telles représentations vectorielles, comme cela a été anticipé dès [46]. Alors, la proportion analogique $\vec{a} : \vec{b} :: \vec{c} : \vec{d}$ est vraie si et seulement si $\vec{a} - \vec{b} = \vec{c} - \vec{d}$, c’est-à-dire, $\forall i, a_i - b_i = c_i - d_i$. Remarquez que cela est conforme à la Table de vérité 1 dans le cas booléen : lorsque $a_i, b_i, c_i, d_i \in \{0, 1\}$, $a_i : b_i :: c_i : d_i \Leftrightarrow a_i - b_i = c_i - d_i$. Voir [39] pour plus de détails sur cette perspective. Ainsi, les plongements lexicaux des mots ‘filles’, ‘garçons’, ‘femmes’, et ‘hommes’ devraient être telles que $v_{\vec{filles}} - v_{\vec{garçons}} = v_{\vec{femmes}} - v_{\vec{hommes}}$.

De plus, cette perspective est conforme au calcul des plongements lexicaux des phrases comme la somme des plongements lexicaux des mots de la phrase. En effet, nous avons $(\vec{a} - \vec{b}) + (\vec{a}' - \vec{b}') = (\vec{c} - \vec{d}) + (\vec{c}' - \vec{d}') \Leftrightarrow (\vec{a} + \vec{a}') - (\vec{b} + \vec{b}') = (\vec{c} + \vec{c}') - (\vec{d} + \vec{d}')$, ce qui signifie que si deux 4-uplets $(\vec{a}, \vec{b}, \vec{c}, \vec{d})$ et $(\vec{a}', \vec{b}', \vec{c}', \vec{d}')$ de mots forment une PA, leur regroupement additif forme également une PA. En effet, pour confirmer cela, nous avons utilisé des plongements lexicaux de GloVe représentant chaque phrase comme la moyenne des plongements lexicaux de ses mots. Nous avons ensuite calculé $1 - \cos \frac{\vec{b}-\vec{a}}{\vec{d}-\vec{c}}$, qui représente la proximité des vecteurs dans la formation d'une analogie, obtenant une valeur de 0,8513, montrant que les 4 vecteurs sont proches de former un parallélogramme.

Créativité guidée

Étant donné un ensemble \mathcal{G} d'éléments existants, chacun représenté en termes du même ensemble d'attributs booléens \mathcal{A} , la créativité peut consister à produire un nouvel élément, non présent dans \mathcal{G} , mais décrit par le même ensemble d'attributs \mathcal{A} [23]. Vu sous cet angle, la créativité semble facile : nous pouvons choisir au hasard des valeurs pour les attributs dans \mathcal{A} et vérifier si le résultat n'est pas déjà dans \mathcal{G} . Cependant, avec un tel processus, nous n'avons aucun contrôle sur les valeurs d'attribut qui pourraient être souhaitables.

Dans la section précédente, nous avons proposé une autre option pour générer un nouvel élément à partir de trois éléments connus en utilisant des proportions analogiques. Comme souligné dans [40], la proportion analogique consiste à associer une paire (\vec{a}, \vec{b}) avec une paire (\vec{c}, \vec{d}) . Ainsi, à partir de l'ensemble d'éléments \mathcal{G} , on peut construire un ensemble de k paires ordonnées $\mathcal{P} = \{(\vec{a}^j, \vec{b}^j) \mid \vec{a}^j \in \mathcal{G}, \vec{b}^j \in \mathcal{G}, j = 1, \dots, k\}$. Cet ensemble de paires représente potentiellement les connaissances pouvant être obtenues par comparaison par paires des éléments de \mathcal{G} . En prenant un élément \vec{c} comme point de départ que nous aimerions modifier, nous pouvons obtenir un nouvel élément \vec{x} en résolvant, composante par composante, les équations $\vec{a}^j : \vec{b}^j :: \vec{c} : \vec{x}^j$ (lorsque la solution existe), c'est-à-dire $\vec{x}^j = \vec{c} \equiv (\vec{a}^j \equiv \vec{b}^j)$ pour toute paire j . En d'autres termes, nous recherchons l'ensemble des solutions.

$$\mathcal{S} = \{\vec{x} \mid \exists (\vec{a}^j, \vec{b}^j) \in \mathcal{P}, j \in \{1, \dots, k\} \text{ s.t. } \vec{a}^j : \vec{b}^j :: \vec{c} : \vec{x}\}$$

Notons que : i) si une propriété est acquise (resp. perdue) lors du passage de \vec{a}^j à \vec{b}^j , c'est-à-dire $a_i^j = 0, b_i^j = 1$ (resp. $a_i^j = 1, b_i^j = 0$) et que \vec{c} n'a pas la propriété (resp. a la propriété), alors cette propriété est également acquise (resp. perdue) par \vec{x}^j ; ii) s'il n'y a pas de changement de \vec{a}^j à \vec{b}^j sur l'attribut i , c'est-à-dire $a_i^j = b_i^j$, alors \vec{x}^j copiera la valeur de \vec{c}^j pour la valeur de l'attribut i .

Pour mieux contrôler le processus de génération de nouveaux éléments, on peut utiliser un ensemble de paires sélectionnées. Plus précisément, supposons que les éléments dans \mathcal{G} à partir desquels l'ensemble \mathcal{P} de paires est construit représentent des objets / profils / situations appartenant à un univers réel, et ensuite, que chaque paire ordonnée (\vec{a}^j, \vec{b}^j) de vecteurs représente des changements légitimes / faisables / autorisés / précieux de \vec{a}^j à \vec{b}^j .

Lorsqu'il n'y a pas de solution dans \mathcal{S} (parce que \mathcal{P} est trop petit), ou lorsque les éléments trouvés ne sont pas considérés comme suffisamment satisfaisants, nous devons envisager l'option d'élargir l'ensemble potentiel \mathcal{P} avec lequel nous commençons, en construisant de nouvelles paires ordonnées à partir des paires déjà présentes dans \mathcal{P} . Cela donne naissance à un processus d'inférence créative qui tente d'améliorer un élément ou une entité particulière, en tirant parti d'un ensemble de paires ordonnées d'éléments existants, en utilisant le mécanisme basé sur la proportion analogique.

Afin d'agrandir la base initiale de paires, nous calculons de nouvelles paires au moyen d'une opération, notée $\wedge \vee$, introduite dans [41, 40], qui produit une nouvelle paire ordonnée en combinant deux paires. Cette opération $\wedge \vee$ est définie de la manière suivante :

$$(\vec{a}, \vec{b}) \wedge \vee (\vec{c}, \vec{d}) = (\vec{a} \wedge \vec{c}, \vec{b} \vee \vec{d})$$

où la conjonction et la disjonction des éléments sont définies composante par composante :

$$\begin{aligned} \vec{a} \wedge \vec{b} &= (a_1 \wedge b_1, \dots, a_n \wedge b_n); \\ \vec{a} \vee \vec{b} &= (a_1 \vee b_1, \dots, a_n \vee b_n). \end{aligned}$$

Clairement, cet opérateur $\wedge \vee$ est commutatif, associatif et idempotent par construction. On peut vérifier que si (a_i, b_i) ou $(c_i, d_i) = (0, 1)$, $(a_i \wedge c_i, b_i \vee d_i) = (0, 1)$. Ainsi, si une propriété i est acquise lors du passage du premier élément d'une paire au deuxième élément de cette paire, alors la propriété est également acquise dans la nouvelle paire résultant de la combinaison par $\wedge \vee$. En revanche, $(a_i \wedge c_i, b_i \vee d_i) = (1, 0)$ seulement si $(a_i, b_i) = (c_i, d_i) = (1, 0)$. Ainsi, si une propriété i est perdue dans la nouvelle paire, c'était déjà le cas dans chacune des paires combinées. L'opération $\wedge \vee$ a le mérite de "cumuler" l'acquisition de caractéristiques en préservant les motifs $(0, 1)$.

Étendre l'ensemble initial \mathcal{P} de paires nous donne plus de chances de créer un nouvel élément d'intérêt, peut-être avec des caractéristiques plus désirables. Plus précisément, nous calculons l'ensemble \mathcal{S} précédemment défini où \mathcal{P} est remplacé par

$$\mathcal{P}' = \{(a^k, b^k) \mid (a^k, b^k) = (a^i, b^i) \wedge \vee (a^j, b^j)$$

tel que $((a^i, b^i), (a^j, b^j)) \in \mathcal{P}^2\}$.

Nous pouvons appliquer cette extension de \mathcal{P} de manière itérative à \mathcal{P}' et ainsi de suite. En raison de l'idempotence de $\wedge \vee$, $\mathcal{P} \subseteq \mathcal{P}'$.

Ensuite, étant donné un élément fixé, choisi, représenté par un vecteur $\vec{c} \in \mathcal{G}$, on peut chercher quel(s) nouvel(s) élément(s) pourrait/peuvent être obtenu(s) en appliquant un changement existant dans la base de paires ordonnées \mathcal{P}' .

Nous illustrons maintenant l'ensemble du processus avec un petit exemple librement inspiré de la créativité en littérature. Les éléments que nous considérons sont censés être des descriptions sommaires de romans. Nous avons 6 attributs qui indiquent si le roman i) est écrit en vers ou en prose (*ver.*); ii) est épistolaire ou non (*epi.*); iii) est un roman d'aventure ou non (*adv.*); iv) est un roman d'amour ou non (*rom.*); v) se déroule aujourd'hui ou non (*tod.*); vi) se déroule dans un pays exotique ou non (*exo.*).

Nous avons 2 paires constituant \mathcal{P} :

- $(\vec{a}_1, \vec{b}_1) = ([0, 0, 0, 0, 1, 0], [0, 1, 0, 0, 1, 1])$
- $(\vec{a}_2, \vec{b}_2) = ([0, 0, 0, 0, 1, 0], [1, 0, 0, 0, 1, 1])$.

Ainsi, \vec{a}_1 fait référence à un roman écrit en prose, non épistolaire, sans aventure, sans romance, se déroulant aujourd'hui dans un pays non exotique. \vec{b}_1 est similaire à \vec{a}_1 à l'exception qu'il est épistolaire et exotique et puisque $(\vec{a}_1, \vec{b}_1) \in \mathcal{P}$, \vec{b}_1 est considéré comme "plus intéressant" que \vec{a}_1 . La paire (\vec{a}_2, \vec{b}_2) peut être lue de manière similaire.

Lorsque nous étendons \mathcal{P} avec l'opérateur \wedge (mais sans effectuer la fermeture complète), nous ajoutons à \mathcal{P} la paire suivante : - $(\vec{a}_3, \vec{b}_3) = ([0, 0, 0, 0, 1, 0], [1, 1, 0, 0, 1, 1])$ puisque $(\vec{a}_3, \vec{b}_3) = (\vec{a}_1, \vec{b}_1) \wedge (\vec{a}_2, \vec{b}_2)$.

Parce que $\vec{a}_1 = \vec{a}_2$, évidemment $\vec{a}_1 = \vec{a}_3$, mais cela n'est pas du tout obligatoire, juste une question de simplicité dans l'exemple. Notez que \vec{b}_3 cumule les "avantages" de \vec{b}_1 et \vec{b}_2 par rapport à \vec{a}_1 . En partant de $\vec{c} = [0, 0, 0, 1, 0, 0]$, nous observons dans le Tableau 5 que les 3 équations analogiques correspondantes sont solvables. Ainsi, nous avons obtenu 3 nouveaux profils de romans possibles $\vec{x}_1, \vec{x}_2, \vec{x}_3$. Par exemple, la solution de la troisième équation est alors un nouveau type de roman, distinct des 5 vecteurs existants $\vec{a}_1, \vec{b}_1, \vec{b}_2, \vec{x}_1, \vec{x}_2$. Il s'agit d'un roman d'amour exotique écrit en vers et épistolaire, sans aventure, se déroulant dans le passé ou dans le futur. Voir le Tableau 5. Le processus que nous avons décrit garantit que i) les éléments obtenus sont nouveaux, et ii) ils sont obtenus à partir d'un \vec{c} existant sur la base de changements déjà existants, tels qu'observés sur des paires ordonnées d'éléments existants. Les résultats ainsi obtenus sont-ils "intéressants"? C'est une question complètement différente : tout ce qui est nouveau n'est pas nécessairement intéressant, mais au moins le fait d'être nouveau peut aider à penser différemment.

Dans la Table 5, nous ne donnons qu'un très petit exemple pour expliquer le mécanisme qui nous permet i) d'élargir un ensemble de paires qui présente des changements intéressants qui sont cumulés dans l'élargissement; ii) et ensuite de produire un nouvel élément amélioré à partir de l'ensemble obtenu de paires ordonnées et d'un élément de référence; cf. [42] pour une première mise en œuvre. Notons qu'une façon

	ver.	epi.	adv.	rom.	tod.	exo.
\vec{a}_1	0	0	0	0	1	0
\vec{b}_1	0	1	0	0	1	1
\vec{c}	0	0	0	1	0	0
x_1	0	1	0	1	0	1
	ver.	epi.	adv.	rom.	tod.	exo.
\vec{a}_2	0	0	0	0	1	0
\vec{b}_2	1	0	0	0	1	1
\vec{c}	0	0	0	1	0	0
x_2	1	0	0	1	0	1
	ver.	epi.	adv.	rom.	tod.	exo.
\vec{a}_3	0	0	0	0	1	0
\vec{b}_3	1	1	0	0	1	1
\vec{c}	0	0	0	1	0	0
x_3	1	1	0	1	0	1

TABLE 5 - $(\vec{a}_3, \vec{b}_3) = (\vec{a}_1, \vec{b}_1) \wedge (\vec{a}_2, \vec{b}_2)$.

d'étudier le pouvoir créatif de ce processus pourrait consister à comparer la distribution de probabilité des valeurs des caractéristiques dans les paires avec lesquelles nous commençons, à la distribution de probabilité des valeurs des caractéristiques dans l'ensemble des éléments produits (en faisant varier ou non l'élément de référence) à l'aide de la divergence de Kullback-Leibler [27].

L'inférence basée sur la proportion analogique fournit une approche à la créativité qui repose sur la recopie de ce qui est observé dans une paire, en termes de changement et de permanence, sur une autre paire, dont le premier élément est connu. De plus, nous avons montré qu'il est possible de combiner des paires ordonnées existantes en de nouvelles paires tout en préservant des caractéristiques souhaitables. Cependant, nous ne prétendons certainement pas que toute forme d'inférence analogique créative, prise au sens large, pourrait être capturée par le mécanisme de transfert que nous proposons.

Expériences

Nous présentons la procédure expérimentale pour le cas de représentations symboliques, ou par plongements lexicaux, avant d'expliquer l'implémentation et de présenter les résultats ².

Procédure

Nous commençons avec une base de données contenant les descriptions des animaux en termes de caractéristiques booléennes. Nous supposons également qu'une classe est connue pour chaque animal. Ainsi, un animal A est décrit par un vecteur \vec{a} avec sa classe $cl(\vec{a})$. L'idée, étant donné

2. Le code de nos expériences est accessible ici : https://github.com/Sasufox/TER_Analogy

un sous-ensemble S de la base de données, est de calculer les solutions des équations de la forme $\vec{a} : \vec{b} :: \vec{c} : \vec{x}$ où $\vec{a}, \vec{b}, \vec{c} \in S$, et de voir si \vec{x} est, ou non, dans la base de données. De plus, nous pouvons utiliser uniquement la restriction des vecteurs $\vec{a}, \vec{b}, \vec{c}$ à un sous-ensemble de caractéristiques. Plus précisément, nous recherchons de petits sous-ensembles de caractéristiques utilisées pour la description des animaux, comme dans l'exemple de la Table 6 (où la solution existe effectivement dans le règne animal). Dans

	<i>allaite leurs petits</i>	<i>pond des oeufs</i>
<i>scorpions</i>	0	0
<i>mammifères</i>	1	0
<i>oiseaux</i>	0	1
?	1	1

TABLE 6 – ? = monotrèmes

cet exemple, la résolution d'équations analogiques "crée" une espèce animale qui pond des œufs et allaite ses petits. Il s'avère que de tels animaux existent : les ornithorynques, les échidnés.

Notez que lors de la recherche de solutions analogiques, nous devons prendre en compte deux problèmes : 1. l'équation doit être solvable pour chaque caractéristique considérée; 2. $\vec{a} : \vec{b} :: \vec{c} : \vec{x}$ et $\vec{a} : \vec{c} :: \vec{b} : \vec{x}$ ont la même solution. De plus, d'un point de vue de la créativité, les vecteurs $\vec{a}, \vec{b}, \vec{c}$ (ou les sous-vecteurs utilisés) font référence à des animaux qui devraient être suffisamment différents. C'est pourquoi il peut être utile d'imposer les contraintes $cl(\vec{a}) \neq cl(\vec{b}), cl(\vec{a}) \neq cl(\vec{c}), cl(\vec{b}) \neq cl(\vec{c})$.

De plus, nous sommes également intéressés par la résolution d'équations analogiques lorsque les vecteurs sont des plongements de mots, c'est-à-dire en calculant $\vec{x} = \vec{c} + \vec{b} - \vec{a}$, et en recherchant les mots dont le plongement est proche de \vec{x} . La question est alors de voir si les résultats sont compatibles avec ceux obtenus à partir des représentations booléennes.

16 attributs	5 attributs les plus importants
22.64%	59.67%

TABLE 7 – Précision avec des vecteurs booléens

Implémentation

Approche symbolique

Nous utilisons l'ensemble de données *Zoo* (voir note de bas de page 2). Cet ensemble de données contient 101 animaux, chacun décrit par 16 caractéristiques et leur classe. Parmi ces 17 caractéristiques, 15 sont binaires tandis que deux sont nominales. La première caractéristique nominale décrit le nombre de pattes que possède l'animal, tandis que l'autre décrit sa classe (mammifère, reptile, etc.).

Les premières expériences ont pris en compte l'ensemble complet des 16 caractéristiques, en excluant la classe de

l'animal. Pour la liste complète des 101 animaux, nous avons calculé tous les triplets possibles $\binom{101}{3}$ et nous n'avons conservé que ceux pour lesquels les trois animaux appartenaient à une classe différente. Pour chaque triplet $(\vec{a}, \vec{b}, \vec{c})$, nous avons prédit l'existence, ou non, d'un quatrième élément représentant un animal "potentiel" en utilisant des proportions analogiques pour chaque caractéristique, comme décrit dans les sections précédentes.

Imposer des proportions analogiques pour les 16 caractéristiques complètes est assez restrictif, car il suffit qu'une caractéristique ne soit pas dans la PA pour écarter tout le triplet. Examiner tous les sous-ensembles possibles de caractéristiques de cardinalité 2 ou plus est computationnellement prohibé, car cela entraînerait $\binom{101}{3} \times 2^{15}$ instances. Nous voulions donc identifier le sous-ensemble des caractéristiques les plus "importantes" afin de réaliser nos expériences. Pour ce faire, nous avons identifié les caractéristiques partagées par la plupart des animaux et nous avons sélectionné les cinq premières. Ces caractéristiques étaient a des poils, pond des œufs, produit du lait, venimeux, et domestiqué. Nous avons appliqué la même procédure que précédemment pour ce sous-ensemble de caractéristiques.

Expériences avec GloVe

Dans une autre série d'expériences, nous avons examiné comment les plongements représentant des mots peuvent être utilisés pour la créativité à base d'analogies. Comme expliqué ci-dessus, chaque élément dans un triplet $(\vec{a}, \vec{b}, \vec{c})$ est représenté sous une forme vectorielle avec des valeurs réelles. Nous prédisons un quatrième élément \vec{d} tel que $\vec{d} = \vec{c} + \vec{b} - \vec{a}$. Nous représentons chaque animal A, B, C à l'aide de plongements statiques ou contextuels. Pour les plongements statiques, nous avons choisi les plongements GloVe [35] avec 300 dimensions ("common crawl" avec 840 milliards de jetons). Après avoir calculé \vec{d} , nous trouvons les mots représentés par les plongements GloVe les plus proches de ce vecteur en utilisant la distance euclidienne. Notre objectif dans cet ensemble d'expériences était d'examiner si ces mots représentent effectivement des animaux présents dans la base de données Zoo.

Expériences avec BERT

Enfin, nous voulions également tester les plongements contextuels créés à l'aide des architectures Transformer [49]. Nous avons opté pour une approche basée sur l'architecture BERT [10]. Les architectures basées sur Transformer fournissent des plongements qui ne sont pas statiques mais dépendent plutôt de leur contexte, s'appuyant fortement sur un mécanisme d'attention complexe sur un contexte de taille fixe de sous-mots. Pour cette raison, nous n'avons pas pu suivre la même procédure que les

plongements statiques de GloVe. Au lieu de cela, nous avons opté pour la création de phrases pour chaque animal dans la base de données, reflétant les caractéristiques qui décrivent chaque animal. Plus précisément, nous avons choisi les cinq caractéristiques mentionnées précédemment (a des poils, pond des œufs, produit du lait, venimeux, et domestiqué). Par exemple, pour l'*ornithorynque*, nous avons créé la phrase suivante, reflétant les caractéristiques qui décrivent cet animal :

L'ornithorynque est un animal qui a des poils, pond des œufs, produit du lait, n'est pas venimeux et n'est pas domestiqué.

Chaque phrase ainsi créée a été encodée à l'aide de Sentence-BERT [45]. En raison de la nature non statique des plongements BERT, nous n'avons pas pu rechercher parmi les phrases potentielles dans l'espace de plongement BERT celle qui était la plus proche du vecteur $\vec{d} = \vec{c} + \vec{b} - \vec{a}$. Nous avons donc opté pour la méthode suivante. Une fois que nous avons créé de tels plongements, pour un triplet donné (A, B, C) d'animaux, nous avons calculé la distance euclidienne, pour tout $D \notin \{A, B, C\}$ dans la base de données, entre $\vec{a} - \vec{b}$ et $\vec{c} - \vec{d}$. Nous décrivons les résultats dans la section suivante.

Les descriptions d'animaux mentionnées que nous avons créées incluent l'animal lui-même. En tant que tel, les plongements de phrases sBERT reposent indéniablement fortement sur cette information. Nous voulions donc tester si les plongements de phrases sBERT qui n'utilisent pas d'informations sur l'animal sont capables de simuler l'approche symbolique que nous avons décrite ci-dessus. Nous avons donc procédé en créant des descriptions d'animaux basées sur leurs caractéristiques sans mentionner explicitement l'animal. Nous avons créé deux variantes : a) des descriptions complètes et b) des descriptions omettant les caractéristiques absentes.

Par exemple, pour l'*ornithorynque*, une description complète serait :

C'est un animal qui a des poils, pond des œufs, produit du lait, n'est pas venimeux et n'est pas domestiqué.

tandis que la description omettant les caractéristiques absentes serait :

C'est un animal qui a des poils, pond des œufs et produit du lait.

Comme dans le cas précédent, pour un triplet donné (A, B, C) , nous avons créé des descriptions et avons ainsi obtenu des vecteurs sBERT \vec{a} , \vec{b} et \vec{c} à partir desquels nous avons calculé $\vec{d} = \vec{c} + \vec{b} - \vec{a}$. Nous avons ensuite procédé au calcul des 32 descriptions différentes pour les 5 caractéristiques mentionnées ci-dessus³ exactement de la même

3. C'est-à-dire pour l'ensemble de caractéristiques suivant : a des poils, pond des œufs, produit du lait, venimeux, et domestiqué.

manière que nous l'avons fait pour les animaux, et obtenu des vecteurs d_i sBERT avec $i \in [1 \dots 32]$ pour chacune des combinaisons possibles. Nous avons ainsi pu calculer dans un ordre décroissant la distance euclidienne entre d_i et $\vec{d} = \vec{c} + \vec{b} - \vec{a}$. De plus, pour chacun des 32 d_i , nous avons pu récupérer le sous-ensemble des animaux partageant exactement les mêmes caractéristiques. Cela nous a permis de comparer directement une approche basée sur BERT avec l'approche symbolique que nous avons décrite ci-dessus. Nous décrivons nos résultats dans la section suivante.

Résultats et discussion

Approche symbolique

Évaluer la créativité est un sujet de recherche ouvert. Il n'existe pas de mesures de créativité largement acceptées et la plupart des évaluations restent subjectives. Dans cet article, nous voulions avoir une estimation approximative de la fréquence à laquelle les animaux que nous avons proposés existaient effectivement dans le micro-monde de la base de données Zoo. Nous avons donc choisi de mesurer la Précision de nos résultats : le nombre de prédictions pour lesquelles au moins un animal existe dans la base de données Zoo sur l'ensemble de nos prédictions. Les résultats pour les vecteurs booléens sont présentés dans le Tableau 7.

Comme nous pouvons le voir lorsque nous utilisons la liste complète des 16 caractéristiques, 22,64% des animaux proposés existent déjà dans la base de données Zoo, tandis que pour les 5 caractéristiques les plus importantes, ce pourcentage est de 59,67%. Tenant compte de la subjectivité dans l'évaluation de la créativité, examinons quelques cas spécifiques. Prenons par exemple les trois animaux suivants : *serpent de mer*, *grenouille*, *tamanoir*. Leurs vecteurs pour le sous-ensemble des caractéristiques importantes sont les suivants : $\vec{a} = (0, 0, 0, 1, 0)$, $\vec{b} = (0, 1, 0, 1, 0)$, $\vec{c} = (1, 0, 1, 0, 0)$ respectivement. Lorsque nous appliquons les PA sur ces vecteurs, nous obtenons le vecteur $\vec{d} = (1, 1, 1, 0, 0)$, prédisant ainsi un animal poilu qui pond des œufs et allaite ses petits. Bien que cela puisse sembler étrange à première vue, un tel animal existe effectivement : l'*ornithorynque*. Considérons maintenant le triplet d'animaux (*ornithorynque*, *antilope*, *raie*) avec les vecteurs correspondants : $\vec{a} = (1, 1, 1, 0, 0)$, $\vec{b} = (1, 0, 1, 0, 0)$, $\vec{c} = (0, 1, 0, 1, 0)$. En appliquant la même procédure, nous obtenons $\vec{d} = (0, 0, 0, 1, 0)$, ce qui correspond à un être non poilu venimeux qui ne pond pas d'œufs (lors de la naissance) et n'allait pas ses petits. De tels animaux existent effectivement : les *scorpions* et les *serpents de mer*. Notre approche les identifie effectivement tous les deux.

Expériences avec GloVe

En ce qui concerne GloVe, le coût computationnel était prohibitif pour calculer $\binom{101}{3}$ instances. Puisque notre objectif dans cet article est d’explorer le potentiel des analogies pour la créativité, nous voulions comparer les résultats entre une approche basée sur la logique prédicative et une basée sur les plongements de mots. Nous avons donc décidé d’évaluer GloVe sur les résultats obtenus avec l’approche basée sur la logique prédicative et voir si GloVe pouvait les égaler. Après avoir calculé $\vec{d} = \vec{c} + \vec{b} - \vec{a}$, nous recherchons les 10 vecteurs GloVe les plus proches de \vec{d} selon la distance euclidienne. Nous éliminons les instances qui ne correspondent pas à des animaux selon les “synsets” de WordNet [31], tel que mis en œuvre par NLTK. Ensuite, nous calculons la précision à k , c’est-à-dire que pour les k premières propositions de GloVe, nous examinons si au moins une est présente comme animal dans la base de données Zoo, après avoir appliqué le “stemming” avec NLTK. Les résultats pour $k \in [1, 8]$ sont présentés dans la Table 8. Comme nous pouvons le voir, déjà pour $k = 2$, GloVe atteint une précision de 64,60, dépassant les résultats obtenus dans la méthode basée sur la logique prédicative, tandis que pour $P@8 = 83, 44$.

Expériences avec BERT

Concernant les plongements contextuels utilisant les architectures Transformer, comme mentionné précédemment, nous avons utilisé Sentence-BERT [45] après avoir créé des phrases pour chaque animal reflétant leurs caractéristiques. Bien que nous puissions calculer $\vec{d} = \vec{c} + \vec{b} - \vec{a}$ et ensuite rechercher une phrase qui est plus proche de \vec{d} , le faire pour chaque phrase potentielle est computationnellement prohibitif. Nous avons donc opté pour l’identification de \vec{d} pour les vecteurs représentant les animaux restreints à la base de données Zoo, en calculant les distances euclidiennes entre $\vec{a} - \vec{b}$ et $\vec{c} - \vec{d}$ pour tout $d \notin \{a, b, c\}$. Nous voulions ensuite évaluer les résultats de cette approche par rapport aux résultats de l’approche symbolique décrite ci-dessus. Plus précisément, pour tout triplet pour lequel l’approche symbolique fournissait une liste d’animaux non vide, nous avons calculé la $Precision@k$ entre cette liste et les $k = 8$ premiers animaux (représentés par le vecteur \vec{d}) calculés avec Sentence-BERT. Les résultats sont présentés dans la Table 8.

Comme nous pouvons le voir, il y a un certain chevauchement entre l’approche basée sur BERT et l’approche symbolique, mais nous ne pouvons pas prétendre qu’il y ait un chevauchement très étroit. Cela dit, notons cependant que, comme dans le cas symbolique, dans le cas de BERT, des animaux tels que l’*ornithorynque* sont également “prédits”. Par exemple, pour le triplet (*scorpion*, *vipère*, *tamanoir*), l’approche symbolique fournit l’*ornithorynque*

	P@1	P@2	P@3	P@4	P@5	P@6	P@7	P@8
GloVe	46.33	64.60	73.94	78.75	81.33	82.60	83.21	83.44
exp.	6.87	12.48	16.65	20.66	24.04	27.17	30.15	32.89
not exp.	0.0	0.0	0.0	0.0	0.0	0.2	1.4	2.6
om.	0.0	0.01	0.08	0.12	0.19	0.2	0.4	0.6

TABLE 8 – Précision@k pour les vecteurs GloVe et sBERT en pourcentages. Pour sBERT, les résultats reflètent le pourcentage d’animaux prédits qui ont également été prédits par l’approche symbolique. Les descriptions d’animaux mentionnent explicitement l’animal (exp.), ou non (not exp.), ou ne mentionnent pas explicitement l’animal et omettent des caractéristiques dont l’animal est dépourvu (om.).

comme seule prédiction, tandis que l’approche basée sur BERT fournit une fois de plus l’*ornithorynque* comme l’animal “analogique” le plus probable. Il en va de même pour le triplet (*scorpion*, *tamanoir*, *physalie*). En ce qui concerne le *scorpion*, celui-ci est donné comme l’animal “analogique” le plus probable par différents triplets par BERT, par exemple (*oryx*, *autruche*, *poney*), ainsi que par l’approche symbolique, par exemple (*ornithorynque*, *opossum*, *physalie*), bien qu’il n’y ait aucun triplet qui fournisse le *scorpion* à la fois pour l’approche symbolique et pour celle basée sur BERT.

Une deuxième approche consistait à comparer directement l’approche symbolique par rapport à une approche basée sur BERT. Nous avons donc créé, comme décrit dans la section précédente, des descriptions complètes ou des descriptions omettant les caractéristiques absentes des animaux sans fournir le nom des animaux. Cela nous a permis de générer des vecteurs sBERT et de calculer $\vec{d} = \vec{c} + \vec{b} - \vec{a}$. Nous avons ensuite procédé à la création de descriptions similaires pour toutes les combinaisons possibles des 5 caractéristiques sélectionnées, donnant ainsi $2^5 = 32$ descriptions et donc 32 vecteurs BERT $\vec{d}_i, i \in [1, \dots, 32]$. Nous avons ensuite ordonné dans un ordre décroissant chaque \vec{d}_i par rapport à \vec{d} . Comme nous avons pu récupérer le sous-ensemble des animaux qui satisfont chaque description, nous avons ainsi pu comparer directement à l’approche symbolique. Nous avons décidé d’utiliser la $Precision@k$ comme précédemment afin de mesurer le chevauchement entre l’approche symbolique et celle basée sur BERT. Comme le montrent les résultats de la Table 8, BERT s’appuie fortement sur la sémantique du mot lui-même pour décrire l’animal au lieu de la description de ses caractéristiques.

Un dernier exemple de créativité avec les PA

Les tests de Raven [44] sont des tests de QI bien connus. Ils se présentent sous la forme d’une série d’instances ayant le format d’une matrice $n \times n$ (où n est 2, 3 ou 4) dont les cellules contiennent diverses figures géométriques (voir

Figure 2 pour un exemple), à l'exception de la dernière cellule qui est vide et doit être complétée en sélectionnant une solution parmi 8 candidats (lorsque $n = 3$). Les séries de matrices d'un test sont de difficulté progressive (d'où la dénomination de Raven's Progressive Matrices, abrégé en RPM). Lorsqu'aucune solution n'est proposée, nous sommes confrontés à un exercice de créativité pour construire le contenu de la cellule vide.

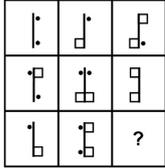


FIGURE 2 – Le problème

Décrivons notre ensemble d'images, où $im9$ doit être deviné, comme une matrice telle que :

$$\begin{pmatrix} im1 & im2 & im3 \\ im4 & im5 & im6 \\ im7 & im8 & im9 \end{pmatrix}$$

Chaque image i peut être codée par un vecteur avec 4 composantes :

- tl décrit l'élément en haut à gauche de la barre verticale ;
- tr décrit l'élément en haut à droite de la barre ;
- br décrit l'élément en bas à droite de la barre ;
- bl décrit l'élément en bas à gauche de la barre.

La valeur de chaque composante appartient à l'ensemble $X = \{point, carr, rien\} = \{p, c, r\}$ de sorte qu'une image est un vecteur dans X^4 .

Cela conduit à une nouvelle représentation de la matrice initiale :

$$\begin{pmatrix} nddn & ndns & nsds \\ dsdn & ddss & snns \\ dnsn & dssd & im9 \end{pmatrix}$$

En termes de proportion analogique, le problème peut être lu comme les PA suivantes : $(im1, im2) : im3 :: (im4, im5) : im6$ et $(im4, im5) : im6 :: (im7, im8) : im9$ ou avec la représentation vectorielle : $(nddn, ndns) : nsds :: (dsdn, ddss) : snns$ et $(dsdn, ddss) : snns :: (dnsn, dssd) : im9$, où $im9$ est alors la solution à trouver. En d'autres termes, la relation R entre les 2 premières images et la 3e dans une ligne est la même quelle que soit la ligne. Cette hypothèse est également valable pour les colonnes : la relation R entre les 2 premières images et la 3e dans une colonne est la même quelle que soit la colonne. La manière naturelle de considérer une relation entre 2 images et une troisième est de supposer que la troisième est une combinaison T des deux premières, où T est un opérateur $X \times X \rightarrow X$, défini composante par composante. T doit être défini sur 9 paires. Observer les deux premières lignes complètes, ainsi que les deux premières colonnes complètes de la matrice, nous donne une définition complète de l'opérateur T : Pour chaque ligne ou colonne, nous obtenons 4 équations :

- ligne 1 $T(n, n) = n, T(d, d) = s, T(d, n) = d, T(n, s) = s$;
- ligne 2 $T(d, d) = s, T(s, d) = n, T(d, s) = n, T(n, s) = s$;
- col. 1 $T(n, d) = d, T(d, s) = n, T(d, d) = s, T(n, n) = n$;

$$\text{col. 2 } T(n, d) = d, T(d, d) = s, T(n, s) = s, T(s, s) = d.$$

Cela définit entièrement T , et nous obtenons $im9 = ssdd$ puisque $T(d, d) = s, T(n, s) = s, T(s, s) = d$, et $T(n, d) = d$, conduisant ainsi à la solution unique ci-contre :



Cet exemple montre les capacités créatives de l'inférence basée sur l'AP, à partir de l'utilisation d'une représentation basée sur des caractéristiques (symboliques). Remarquablement, cette approche s'est avérée efficace sur une série de 36 tests RPM avancés [7, 8] ; dans 16 cas, l'approche donne un bon résultat même lorsqu'elle est appliquée au niveau granulaire du pixel.

Cet exemple illustre la puissance créative de l'inférence basée sur les proportions analogiques (PA). Remarquons que les PA sont ici de la forme $(x, y) : \mathcal{F}(x, y) :: (x', y') : \mathcal{F}(x', y')$ où \mathcal{F} s'applique aux paires de vecteurs et renvoie un vecteur. De tels PA sont proches des PA de la forme $x : f(x) :: x' : f(x')$, étudiés en détail par [18], qui peuvent être liés aux PA booléennes [1]. Dans cet exemple, \mathcal{F} est entièrement défini à travers les huit premières cellules de la matrice. Mais des exemples similaires où \mathcal{F} est seulement partiellement défini conduiraient à plusieurs solutions. Cette forme de créativité est une "génération nouvelle adaptée aux contraintes d'une tâche particulière" [30], comme c'est généralement le cas avec le raisonnement analogique [17].

Les tests RPM sont un exemple de la puissance de l'inférence basée sur les proportions analogiques. Cette approche s'écarte clairement de celles basées sur la théorie de la correspondance des structures ("structure mapping theory") [14, 13], ou de ses variantes [24], y compris l'algorithme d'inférence CWSG ("copy with substitution and generation") [19], qui ont également été appliquées aux problèmes de RPM, voir par exemple [29], ou [28, 48]. Mais ces approches nécessitent de disposer de solutions potentielles, ce qui n'est pas très conforme à l'idée de créativité. Voir [8] pour une discussion comparative.

Remarques de conclusion

Cet article présente une étude préliminaire sur le pouvoir créatif des PA, montrant des résultats encourageants. Dans son travail pionnier, Margaret Boden [2] distingue trois formes de créativité : la créativité combinatoire, exploratoire et transformationnelle. La créativité combinatoire est le résultat de la combinaison d'idées familières. Les deux autres types de créativité, selon Boden, présupposent l'existence d'un espace conceptuel à travers lequel de nouvelles idées sont réalisées. Ainsi, dans le deuxième type de créativité, l'agent créatif explore les différentes parties de l'espace conceptuel à la recherche d'instances qui peuvent être

considérées comme créatives. Le dernier type de créativité présuppose que l'agent "élargit", pour ainsi dire, les limites de l'espace conceptuel, permettant de nouvelles dimensions et donc une compréhension plus profonde du monde et donc la proposition de nouvelles idées créatives.

Notre travail va au-delà de la simple combinaison de vecteurs, nous proposons un mécanisme basé sur les PA, pour changer un vecteur \vec{c} en un vecteur \vec{d} , une fois que le *contexte approprié* d'une autre paire (\vec{a}, \vec{b}) est trouvé dans l'environnement, souscrivant ainsi à la définition de la créativité *exploratoire*, grâce aux nombreuses paires (\vec{a}, \vec{b}) et aux pivots \vec{c} qui peuvent être utilisés en général.

Nous concluons donc que le processus créatif que nous présentons dans cet article est du deuxième type. En d'autres termes, nous explorons l'espace conceptuel dans le micromonde de la base de données Zoo et proposons des instances nouvelles qui n'ont jamais existé auparavant. Les résultats montrent que la méthode proposée est capable d'identifier l'existence d'animaux dont on pourrait ne pas penser qu'ils existent, tels que des animaux qui pondent des œufs et allaitent leurs petits (ornithorynque) ou ceux qui donnent naissance à leurs petits par ovoviviparité mais ne les allaitent pas (scorpions).

Par ailleurs, nous voulions examiner comment cette approche se comporte pour les plongements lexicaux obtenus par différents modèles de langage. Nous avons donc examiné à la fois les plongements lexicaux statiques, en utilisant GloVe, et les plongements lexicaux contextuels en utilisant BERT. En ce qui concerne les résultats de GloVe, nous avons constaté que notre approche est capable de proposer des animaux "nouveaux" avec jusqu'à 83,44% (pour $P@8$) des animaux proposés dans la base de données originale, alors que même les résultats pour $P@2$ dépassent l'approche booléenne. Ces résultats ne sont à prendre que comme une indication du potentiel de cette approche. Comme montré dans [15], GloVe fonctionne vraiment bien lorsqu'il s'agit d'ensembles de données d'analogies simples, comme celui proposé par [34], mais seuls 30% des analogies sont capturées dans leur nouveau corpus BATS. Des preuves récentes [51] montrent que des modèles plus avancés basés sur les technologies des transformateurs sont capables d'identifier des analogies d'une manière comparable à celle des humains. Nous avons donc choisi d'examiner comment les plongements lexicaux contextuels basés sur les architectures de Transformateur, tels que BERT, se comparent à notre approche. Nous avons donc fourni des descriptions en langage naturel des animaux et calculé des plongements lexicaux de phrases pour chacun en utilisant SBERT. Nous avons réalisé deux séries d'expériences, l'une contenant le nom de l'animal et l'autre ne contenant pas le nom de l'animal. Bien que cette approche puisse en effet récupérer des animaux tels que les ornithorynques et les scorpions (lorsque les noms des animaux sont fournis), ils ont très peu de chevauchement avec l'approche symbolique.

Il reste beaucoup à faire pour comprendre la créativité en utilisant les PA pour la génération de nouvelles instances créatives. Tout d'abord, nous avons besoin de meilleures mesures pour évaluer la créativité qui ne prennent pas uniquement en compte l'existence ou non des instances proposées,⁴ mais aussi d'une certaine manière mesurent à quel point les instances proposées sont "intéressantes". À l'avenir, nous prévoyons de travailler sur de telles mesures ainsi que d'explorer d'autres approches plus sophistiquées des grands modèles de langage tels que GPT-3 et GPT-4. Des expériences initiales concernant l'évaluation de la partie "créative" des animaux proposés pourraient être réalisées en se basant sur la fréquence de la combinaison de leurs caractéristiques. Plus précisément, plus la combinaison de leurs caractéristiques est rare, tout en étant un animal "valide", plus on peut argumenter que la proposition de l'animal est créative.

De manière générale, le raisonnement génératif par le biais des PA ne garantit pas la valeur pratique des résultats, ni leur caractère imprévu. Peut-être peut-on s'inspirer des mesures utilisées dans les systèmes de recommandation [26, 22], pour évaluer l'imprévu dans la créativité. Cependant, le raisonnement basé sur les PA ne reste pas dans le voisinage de ce qui est connu, comme le montrent les exemples de cet article, et peut produire des résultats innovants. Lorsqu'il s'agit de représentations booléennes ou nominales, les résultats peuvent être expliqués, et comme le montre l'exemple de Raven, il n'est pas nécessaire de choisir entre des résultats candidats.

Certes la notion de créativité est délicate à cerner. On peut toutefois convenir que créer, c'est avant tout imaginer et concevoir quelque chose de nouveau, quelque chose qui n'a pas encore été vu ou réalisé. Ainsi, un enfant ayant observé une vache avec son veau, voyant ensuite une jument, pourrait imaginer à quoi ressemble un poulain, même s'il n'en a jamais vu auparavant. De façon similaire, un mathématicien peut concevoir qu'un certain résultat pourrait être valide, car le problème qu'il cherche à résoudre présente des similitudes avec un problème dans un domaine différent où un résultat similaire est déjà établi. Si l'idée du nouveau résultat émerge de façon analogique, sa validation formelle se fera à travers une démarche déductive. La déduction permet de révéler des faits jusque-là ignorés, mais elle ne constitue pas en soi un acte de créativité au sens évoqué précédemment. D'une manière générale, l'espace des faits formellement déductibles d'un ensemble de connaissances est très grand, voire infini. Un outil pour "traverser, parcourir" l'espace des combinaisons possibles est nécessaire. Cet outil peut être de nature probabiliste, car un changement aléatoire d'un des paramètres d'un problème peut relancer la recherche dans de nouvelles directions. Les

4. Dans l'expérience avec l'ensemble de données Zoo, nous avons bénéficié de la connaissance des animaux existants, mais de telles informations sont rarement disponibles pour juger de la créativité.

modèles de diffusion [53], essentiellement basés sur la théorie de probabilités, sont une excellente illustration de cette approche. Mais l’outil peut être aussi de nature analogique, car le mécanisme de la proportion analogique peut parfois être vu comme un processus de recopie partielle [7]. La création, au lieu d’être le résultat du hasard, peut aussi être une affaire de recombinaison d’idées/concepts ou d’attributs préalablement connus : peut-on imaginer un centaure si on ne sait pas ce qu’est un cheval et un homme [9] ? L’inférence à base de PA peut donner l’illusion d’être déductive, d n’est-il pas le résultat du calcul $c + b - a$? Mais pour faire ce calcul il faut partir d’un c et apprécier la différence entre un b et un a qu’il faut avoir choisis, et des b et des a il y en en général beaucoup qui pour un même c ne conduiront pas toujours au même d !

Remerciements

Cette recherche a été soutenue par le projet ANR “Analogy : from Theory to Tools and Applications” (AT2TA), ANR-22-CE23-0023.

Références

- [1] N. Barbot, L. Miclet, H. Prade, and G. Richard. A new perspective on analogical proportions. In *ECSQARU, LNCS, 11726, 163-174*. Springer, 2019.
- [2] M. Boden. *The Creative Mind : Myths and Mechanisms, 2nd edition*. Routledge, 2004.
- [3] M. Bounhas, M. Pirlot, H. Prade, and O. Sobrie. Comparison of analogy-based methods for predicting preferences. In *SUM, LNCS, 11940, 339-354*. Springer, 2019.
- [4] M. Bounhas and H. Prade. Analogy-based classifiers : An improved algorithm exploiting competent data pairs. *Int. J. Approx. Reason.*, 158 :108923, 2023.
- [5] M. Bounhas, H. Prade, and G. Richard. Analogy-based classifiers for nominal or numerical data. *Int. J. of Approximate Reasoning*, 91 :36 – 55, 2017.
- [6] S. Colton. Experiments in constraint-based automated scene generation. In P. Gervás, R. Pérez y Pérez, and T. Veale, editors, *Proc. Int. Conf. on Computational Creativity, Madrid*, pages 127–136, 2008.
- [7] W. F. Correa, H. Prade, and G. Richard. When intelligence is just a matter of copying. In *ECAI, Frontiers in Artificial Intelligence and Applications*, 242, pages 276–281. IOS Press, 2012.
- [8] W. Correa Beltran, H. Prade, and G. Richard. Constructive solving of Raven’s IQ tests with analogical proportions. *Int. J. Intell. Syst.*, 31(11) :1072–1103, 2016.
- [9] R. Descartes. *Méditations Métaphysiques*. 1641.
- [10] J. Devlin, Ming-Wei Chang, K. Lee, and K. Toutanova. BERT : Pre-training of deep bidirectional transformers for language understanding. In *Proc. 2019 Conf. of the North Amer. Chapter of the Assoc. for Comput. Ling. : Human Language Technologies, Vol.1 (Long and Short Papers)*, 4171-4186, Minneapolis, 2019.
- [11] T. G. Evans. A program for the solution of a class of geometric-analogy intelligence-test questions. In M. L. Minsky, editor, *Semantic Information Processing*, pages 271–353. MIT Press, Cambridge, 1968.
- [12] M. A. Fahandar and E. Hüllermeier. Learning to rank based on analogical reasoning. In *AAAI*, pages 2951–2958, 2018.
- [13] B. Falkenhainer, K. D. Forbus, and D. Gentner. The structure-mapping engine : Algorithm and examples. *Artif. Intell.*, 41(1) :1–63, 1989.
- [14] D. Gentner. Structure-mapping : A theoretical framework for analogy. *Cognit. Sci.*, 7(2) :155–170, 1983.
- [15] A. Gladkova, A. Drozd, and S. Matsuoka. Analogy-based detection of morphological and semantic relations with word embeddings : what works and what doesn’t. In *Proc. NAACL Student Res. Workshop*, San Diego, 8-15, 2016. Assoc. for Comput. Ling.
- [16] A. K. Goel. Design, analogy and creativity. *IEEE Expert*, 12 :62–70, 1997.
- [17] A.E. Green, D.J. Kraemer, J.A. Fugelsang, J.R. Gray, and K.N. Dunbar. Neural correlates of creativity in analogical reasoning. *J. Experim. Psychology : Learning, Memory, and Cognition*, 38 :264–272, 2012.
- [18] D. R. Hofstadter and Fluid Analogies Research Group. *Fluid Concepts and Creative Analogies : Computer Models of the Fundamental Mechanisms of Thought*. Basic Books, New York, 1995.
- [19] K. J. Holyoak. Analogy. In K. J. Holyoak and R. G. Morrison, editors, *The Cambridge Handbook of Thinking and Reasoning*, pages 117–142. Cambridge Univ. Press, 2005. Chap. 6.
- [20] K. J. Holyoak and P. Thagard. *Mental Leaps : Analogy in Creative Thought*. MIT Press, 1995.
- [21] Xiaoyang Hu, S. Storks, R. Lewis, and J. Chai. In-context analogical reasoning with pre-trained language models. In *Proc. 61st Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, Toronto, July 2023.
- [22] M. Kaminskis and D. Bridge. Diversity, serendipity, novelty, and coverage : A survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Trans. Interact. Intell. Syst.*, 7(1), 2016.
- [23] A. Kaufmann. L’imagination artificielle - (heuristique automatique). *Revue française d’automatique, d’informatique et de recherche opérationnelle (R.A.I.R.O.) - Recherche opérationnelle*, 3(3) : 5–24, 1969.

- [24] M. T. Keane, T. Ledgeway, and S. Duff. Constraints on analogical mapping : A comparison of three models. *Cognitive Science*, 18(3) :387–438, 1994.
- [25] S. Klein. Culture, mysticism & social structure and the calculation of behavior. In *Proc. 5th Europ. Conf. in Artif. Intel. (ECAI’82)*, Orsay, 141-146, 1982.
- [26] D. Kotkov, Shuaiqiang Wang, and J. Veijalainen. A survey of serendipity in recommender systems. *Knowledge-Based Systems*, 111 :180–192, 2016.
- [27] S. Kullback and R. Leibler. On information and sufficiency. *Annals Math. Statistics*, 22 :79–86, 1951.
- [28] M. Kunda, K. McGregor, and A. K. Goel. A computational model for solving problems from the Raven’s Progressive Matrices intelligence test using iconic visual representations. *Cognitive Systems Research*, 22 :47–66, 2013.
- [29] A. Lovett, K. Forbus, and J. Usher. A structure-mapping model of Raven’s progressive matrices. In *Proc. 32nd Annual Conf. of the Cognitive Science Soc., Portland, OR*, 2010.
- [30] R. E. Mayer. Fifty years of creativity research. In R.J. Sternberg, editor, *Handbook of Creativity*, pages 449–460. Cambridge Univ. Press, 1998.
- [31] J. P. McCrae, A. Rademaker, F. Bond, E. Rudnicka, and Ch. Fellbaum. English WordNet 2019 – an open-source WordNet for English. In *Proc. 10th Global Wordnet Conference*, pages 245–252, Wroclaw, 2019.
- [32] L. Miclet, S. Bayouth, and A. Delhay. Analogical dissimilarity : definition, algorithms and two experiments in machine learning. *JAIR*, 32, 793-824, 2008.
- [33] L. Miclet and H. Prade. Handling analogical proportions in classical logic and fuzzy logics settings. In *ECSQARU*, LNCS, 5590, 638-650. Springer, 2009.
- [34] T. Mikolov, I. Sutskever, K. Chen, G. S Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges et al., editor, *Advances in Neural Inform. Processing Syst.* 26., 3111-3119, Curran Assoc. Inc., 2013.
- [35] J. Pennington, R. Socher, and Ch. Manning. GloVe : Global vectors for word representation. In *Proc. 2014 Conf. on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, 2014.
- [36] V. Pirrelli and F. Yvon. Analogy in the lexicon : a probe into analogy-based machine learning of language. In *Proc. 6th Int. Symp. on Human Communic.*, 1999. Santiago de Cuba, 6 p.
- [37] H. Prade and G. Richard. Analogical proportions : From equality to inequality. *Int. J. of Approximate Reasoning*, 101 :234 – 254, 2018.
- [38] H. Prade and G. Richard. Analogical proportions : Why they are useful in AI. In *IJCAI*, pages 4568–4576, 2021.
- [39] H. Prade and G. Richard. Multiple analogical proportions. *AI Commun.*, 34(3) :211–228, 2021.
- [40] H. Prade and G. Richard. First steps towards a logic of ordered pairs. In *ECSQARU*, LNCS, 14294, 198-209. Springer, 2023.
- [41] H. Prade and G. Richard. Premiers pas vers une logique des paires ordonnées. In *JIAF*, pages 104–112, 2023.
- [42] H. Prade and G. Richard. Analogical proportion-based induction : From classification to creativity. *J. of Applied Logics — IfCoLog J. of Logics and their Applications*, 11 :51–87, 2024.
- [43] M. Ragni and S. Neubert. Analyzing Raven’s intelligence test : Cognitive model, demand, and complexity. In *Computational Approaches to Analogical Reasoning*, Studies in Computational Intelligence, 548, pages 351–370. Springer, 2014.
- [44] J. C. Raven. *Progressive Matrices*. The Psychological Corporation, New York, 1965.
- [45] N. Reimers and I. Gurevych. Sentence-bert : Sentence embeddings using siamese bert-networks, 2019. arXiv, cs.CL, 1908.10084.
- [46] D. E. Rumelhart and A. A. Abrahamson. A model for analogical reasoning. *Cognit. Psycho.*, 5 :1–28, 1973.
- [47] J. Schmidhuber. Formal theory of creativity, fun, and intrinsic motivation (1990-2010). *IEEE Trans. on Autonomous Mental Development*, 2(3) :230–247, 2010.
- [48] S. Shegheva. A computational model for solving Raven’s Progressive Matrices intelligence test, 2018. Master of Science dissertation, School of Computer Science, Georgia Institute of Technology, Aug., 45 p.
- [49] A. Vaswani, N. Shazeer, N. Parmar, J.Uszkoreit, Llion Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need, 2017. arXiv, cs.CL, 1706.03762.
- [50] T. Veale. An analogy-oriented type hierarchy for linguistic creativity. *Knowledge-Based Systems*, 19(7) :471 – 479, 2006.
- [51] T. Webb, K. J. Holyoak, and Hongjing Lu. Emergent analogical reasoning in large language models. *Nature Human Behaviour*, 7(9) :1526–1541, 2023.
- [52] T. Wijesiriwardene, R. Wickramarachchi, B. Gajera, S. Gowaikar, C. Gupta, A. Chadha, A. N. Reganti, A. Sheth, and A. Das. ANALOGICAL - a novel benchmark for long text analogy evaluation in large language models. In *Findings of the Association for Computational Linguistics : ACL 2023*, Toronto, 2023.
- [53] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui, and M.-H. Yang. Diffusion models : A comprehensive survey of methods and applications. *ACM Computing Surveys*, 56(4), 2023.