

Nouvelle mesure de complexité de l'information pour l'analyse du comportement de mobilité éco-responsable

Thomas Chambon¹ et Jeanne Lallement² et Arnaud Revel¹

¹ *Laboratoire Informatique, Image et Interaction (L3i), La Rochelle Université, La Rochelle, France*

² *Centre de REcherche en GEstion (CEREGE), La Rochelle Université, La Rochelle, France*

Le changement climatique est l'un des principaux enjeux du siècle : les scénarios les plus optimistes prédisent une augmentation des températures (CIEC 2014), avec des conséquences catastrophiques. Pour endiguer ce fléau, une réduction des émissions de gaz à effet de serre est nécessaire. Pour essayer de changer progressivement les comportements, nous essayons de rendre plus efficient le concept de nudge (coup de pouce), développé par Thaler et Sunstein. Nous souhaitons, à l'aide des technologies mobiles, analyser et modéliser le comportement des utilisateurs en leur transmettant des informations, et ainsi comprendre si elles sont susceptibles de modifier leurs habitudes de transport. La première étape de ce processus est tout d'abord de pouvoir quantifier la richesse d'une information, pour cela, nous proposons une nouvelle méthode de classement de l'information en fonction de leur complexité.

Mots-clefs : mobilité éco-responsable, prise de décision, complexité relative, complexité de Kolmogorov

1 Problématique

Pour faire évoluer les gens et les encourager à modifier en douceur leurs comportements, les chercheurs en économie, en gestion et en sciences du comportement ont introduit le concept de "nudging"[5] qui peut éventuellement être utilisé pour faire évoluer les gens dans leurs comportements de mobilité. Entre-temps, les technologies numériques se sont répandues dans la population et sont devenues un moyen intéressant de rendre le nudging plus efficace. Cependant, de nombreux facteurs contribuent à l'acceptabilité des techniques de nudging. Dans le contexte de la prise de décision, le concept de complexité pourrait être un facteur participant au processus de décision : lorsque nous devons décider entre deux choix, notre hypothèse est que nous préférons le moins complexe. Alors, plus que la complexité en elle-même, ce qui pourrait être pertinent pour la décision serait la complexité relative puisqu'elle permet de comparer plusieurs types d'informations. Le but de cet article est donc de proposer un nouveau calcul de la complexité relative de l'information afin de pouvoir ordonner le niveau de complexité de l'information fournie à une personne et de mieux comprendre son impact sur son comportement.

2 Méthodologie

Nous nous concentrons sur la complexité de l'information fournie à l'utilisateur que nous prétendons être primordiale pour faciliter l'acceptabilité du processus de nudging. Pour évaluer cet impact, il est d'abord important de quantifier et de qualifier le contenu de l'information produite. Or, la condition préalable à la résolution de ce problème est la définition même donnée au terme "information". Depuis plus d'un demi-siècle, deux grandes approches tentent de quantifier le contenu informationnel d'un objet numérique. C.E. Shannon a été le premier à fournir une définition de l'information basée sur la notion d'entropie (1948)[4]. Dix-sept ans plus tard, A.N. Kolmogorov inspiré par la théorie de Shannon tente de donner une définition de la complexité d'un objet [2].

Le but de notre proposition appelée Information Complexity Ratio (ICR) est de classer plusieurs informations de la plus simple à la plus complexe afin de comprendre leur impact relatif sur les personnes recevant

ces informations. ICR est basée sur deux notions fondamentales, la complexité intrinsèque d'un objet et la similarité entre deux objets. La complexité intrinsèque peut être calculée à l'aide de la complexité de Kolmogorov [2]. La similarité, quant à elle, peut être exprimée grâce à la distance de compression normalisée (NCD) [6]. Ces deux notions peuvent être approximées à l'aide d'algorithmes de compression.

$$ICR(x, y) = NCD(x, y) \times C(y) / C(x) \quad (1)$$

où $NCD(x, y)$ représente la distance de compression normalisée entre les objets x et y et où $C(y)$ et $C(x)$ représentent respectivement les complexités intrinsèques des objets y et x .

Pour aller plus loin et tester les performances de notre approche dans un contexte réaliste, nous avons expérimenté sa robustesse sur le jeu de données SAVOIAS [3]. Ce jeu de données contient **1400 images** réparties en 7 catégories. SAVOIAS présente comme avantage une diversité des images en termes de caractéristiques de bas niveau et de haut niveau. Nous avons confronté notre méthode à la vérité terrain et comparé au simple ratio de compression. Pour mettre en oeuvre cette comparaison, nous avons concaténé les images deux à deux horizontalement et verticalement pour calculer la NCD et également tester si le sens de concaténation avait un impact sur la complexité.

Catégories	CR vs VT	ICR vs VT	Catégories	CR vs VT	ICR vs VT
Scenes	0.30	0.40	Objects	0.16	0.16
Advertisement	0.56	0.58	Interior Design	0.68	0.67
Visualizations	0.55	0.31	Suprematism	0.60	0.73
Art	0.51	0.58			

TABLE 1. Résultats de la comparaison entre le ratio de compression et notre nouvelle méthode. Les valeurs de la première colonne ont été reportées de la colonne *Compression Ratio* (CR) du tableau 3 dans [3].

Comme présenté dans le tableau 1, les résultats de notre méthode sont comparables à la vérité terrain pour les catégories Publicité, Objets, Décoration intérieure. Pour les catégories *Scènes et Art*, l'ICR a même de meilleurs résultats. La catégorie Suprématisme présente les meilleurs résultats (**0,60** vs **0,73**). En revanche, pour la catégorie *Visualisations*, les résultats sont moins bons (**0,55** vs **0,31**). Cette différence pourrait s'expliquer par la grande diversité des images. Dans notre comparaison par paire, les résultats produits sont moins bons, car la similarité est faible, contrairement au taux de compression qui n'intègre pas cette notion. La différence obtenue sur la catégorie *Suprématisme* peut s'expliquer par la structure (formes géométriques) et les couleurs de ces images (palette de couleurs réduite). Cela renforce la capacité de notre mesure à mieux hiérarchiser la complexité en détectant les nombreuses répétitions contenues dans ces images.

3 Conclusion et Perspectives

Au vu de notre objectif de quantification de l'impact de l'information sur le comportement de mobilité, nous souhaitons intégrer notre méthode à une application de traçage que nous développons en parallèle. Le but de cette application est de recueillir des informations sur la mobilité domicile-université auprès d'une population de **800 étudiants** sur une période de **3 mois**. Compte tenu de la situation sanitaire (COVID-19), nous avons dû repousser cette expérimentation. Nous souhaitons également améliorer notre méthode en explorant d'autres pistes théoriques comme la profondeur logique de Bennett [1].

Références

- [1] BENNETT, C. H. Logical depth and physical complexity. In *A Half-Century Survey on The Universal Turing Machine* (USA, 1988), Oxford University Press, Inc., p. 227–257.
- [2] KOLMOGOROV, A. N. Three Approaches to Information. *Problemy Peredachi Informatsii* 1, 1 (1965), 3–11.
- [3] SARAEE, E., JALAL, M., AND BETKE, M. Savoias : A diverse, multi-category visual complexity dataset, 2018.
- [4] SHANNON, C. E. A Mathematical Theory of Communication. *Bell System Technical Journal* 27, 3 (jul 1948), 379–423.
- [5] THALER, R., AND SUNSTEIN, C. *Nudge : Improving decisions about health, wealth, and happiness*. 2008.
- [6] VITÁNYI, P. M., AND LI, M. Minimum description length induction, Bayesianism, and Kolmogorov complexity. *IEEE Transactions on Information Theory* 46, 2 (2000), 446–464.