

# Une extension de la décomposition tensorielle au phénotypage temporel

Hana Sebia\*, Thomas Guyet\*, Etienne Audureau\*\*

\* Inria, AIStroSight, Centre de Lyon, France  
{hana.sebia, thomas.guyet}@inria.fr,

\*\* AP-HP, Hôpital Henri Mondor, Université Paris Est Créteil, France

**Résumé.** La décomposition tensorielle a récemment fait l'objet d'une attention croissante dans la communauté de l'apprentissage automatique en raison de sa polyvalence dans le traitement des données à grande échelle. Cependant, cette tâche devient plus difficile lorsqu'il s'agit de prendre en compte la dimension temporelle. Dans cet article, nous étendons la décomposition tensorielle à l'extraction de phénotypes temporels, décrits comme un combinaison de caractéristiques sur une fenêtre de temps. Nous proposons un nouveau modèle de décomposition intégrant plusieurs régularisations pour améliorer l'interprétabilité des phénotypes extraits. Nous validons ce dernier sur des données synthétiques et réelles provenant de l'Assistance Publique – Hôpitaux de Paris (AP-HP). Les résultats montrent qu'il est plus performant que les modèles les plus récents de décomposition et qu'il découvre des phénotypes intéressants pour les cliniciens.

## 1 Introduction

Un tenseur est une représentation naturelle des données multidimensionnelles. La décomposition tensorielle est un outil statistique historique pour l'analyse de ces données complexes. La popularisation de techniques d'apprentissage automatique efficaces et évolutives l'a rendue attrayante pour les données du monde réel (Perros et al., 2017). Elle a donc été intensivement étudiée dans de nombreux domaines, tels que le traitement du signal, les neurosciences, la communication, la psychométrie, etc (Fanaee-T et Gama, 2016). Techniquement, la décomposition tensorielle est une méthode non-supervisée qui simplifie un tenseur multidimensionnel en tenseurs d'ordre inférieur (Anandkumar et al., 2014). Ceci revient à identifier des variables latentes. Ces variables latentes sont des caractéristiques non observées qui capturent les comportements cachés d'un système. Elles sont difficiles à extraire de données multidimensionnelles complexes en raison 1) des multiples interactions entre les dimensions et 2) de l'entrelacement des occurrences de comportements cachés.

Récemment, plusieurs approches basées sur la décomposition tensorielle ont montré leur efficacité et leur intérêt pour le phénotypage computationnel à partir des dossiers médicaux électroniques (DME) (Afshar et al., 2020, 2021; Chambard et al., 2021; Yin et al., 2019). Les motifs récurrents cachés qui sont découverts dans ces données sont appelés *phénotypes*. Ces phénotypes sont particulièrement intéressants pour 1) décrire les pratiques réelles des unités