

Partitionnement double-niveau de données médicales issues du post-traitement d'immunothérapie

Victor Dupriez*, Chanez Mokhtari*, Juba Agoun*, Mohand-Saïd Hacid*

* Univ Lyon, LIRIS UMR 5205 CNRS, Université Lyon 1, Lyon, France

1 Introduction

Le secteur de la santé produit quotidiennement de grandes quantités de données à travers les dossiers médicaux électroniques. Ces données représentent l'ensemble des informations recueillies lors des visites cliniques telles que les données démographiques, les diagnostics, un historique des opérations subies, les tests de laboratoire et les traitements. Une analyse exploratoire de ces données permet de fournir des informations capitales à la prédiction de l'évolution de l'état de santé des patients, la réduction des coûts de traitement et l'amélioration de la qualité de vie.

Notre étude¹ vise à générer des directives d'hygiène de vie afin de promouvoir la qualité de vie des patients suivant un traitement d'immunothérapie du cancer en se basant sur la caractérisation des groupes de patients. Nous disposons de :

- Données cliniques comprenant des résultats d'analyses et de bilans sanguins. Toutes ces données sont numériques (pour la plupart) ou textuelles.

- Données qualité & hygiène de vie contenant les réponses à un questionnaire rempli tous les 3 mois. Il comprend 6 parties couvrant notamment la consommation du patient, ses émotions et maladies au quotidien ainsi que son ressenti.

2 Approche proposée

Compte tenu de la différence sémantique de nos deux jeux de données, on procède à un partitionnement en deux niveaux, que l'on va ensuite croiser pour tirer des règles d'association.

2.1 Partitionnement des données cliniques

Lors du clustering de données complexes comme celles cliniques, le choix et l'évaluation de la qualité d'un tel partitionnement est complexe, un algorithme optimal n'existe pas a priori pour toutes les données cliniques (Pina et al., 2019).

En partant des données cliniques des patients, plusieurs méthodes de partitionnement sont évaluées suivant la méthodologie proposée par (Handl et al., 2005) : chaque algorithme de

1. This research is supported by the European Union's Horizon 2020 research and innovation program under grant agreement No 875171, project QUALITOP.