

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.

Partitionnement double-niveau des données patients

clustering est appliqué pour différents nombres de clusters. On évalue alors les partitionnements selon 2 techniques de validation par mesure interne pour trouver l'algorithme et le nombre de clusters offrant le meilleur compromis entre 2 propriétés de partitionnement que sont compacité et connectivité.

2.2 Partitionnement des données qualité de vie

Les données évaluant l'hygiène de vie de chaque patient correspondent aux réponses à un questionnaire (selon une échelle). Nous utilisons la méthodologie proposée ci-dessus pour le partitionnement. Afin de visualiser le partitionnement obtenu et d'extraire les caractéristiques de chaque groupement, nous employons la méthode de (Khoie et al., 2017), qui consiste à utiliser des cartes auto adaptatives de Kohonen. Chaque nœud de la SOM est associé à un certain nombre (plus ou moins élevé) de patients dont la répartition est utile pour la suite de cette étude.

2.3 Association des résultats

En repartant de la SOM réalisée sur les données QoL, on examine pour les patients associés à chaque nœud le cluster clinique associé. Dans les cas où un cluster est sur-représenté pour un nœud, ce dernier peut être colorisé. On obtient alors une SOM avec les frontières des clusters QoL dessinées, et une colorisation partielle des nœuds, uniquement lorsqu'un cluster clinique majoritaire est trouvé. De cette visualisation des recouvrements entre clusters cliniques et QoL se trouvent les associations.

3 Conclusion

Nous proposons une méthodologie permettant de catégoriser les patients selon leurs données cliniques et d'hygiène de vie à des fins de recommandation. Ces dernières seront validées par les experts du domaine médical afin de les proposer à leurs patients.

Références

- Handl, J., J. Knowles, et D. B. Kell (2005). Computational cluster validation in post-genomic data analysis. *Bioinformatics* 21(15), 3201–3212.
- Khoie, M., T. Sattari Tabrizi, E. Khorasani, et N. Rahimi, S.; Marhamati (2017). A hospital recommendation system based on patient satisfaction survey. *Applied Sciences* 7(10), 966.
- Pina, A., M. P. Macedo, et R. Henriques (2019). Clustering clinical data in r. *Mass Spectrometry Data Analysis in Proteomics*, 309–343.