

Évaluation des améliorations de prédiction d’hospitalisation par l’ajout de connaissances métier aux dossiers médicaux

Raphaël Gazzotti^{*,****}, Catherine Faron-Zucker^{*}, Fabien Gandon^{**},
Virginie Lacroix-Hugues^{***}, David Darmon^{***}

^{*}Université Côte d’Azur, Inria, CNRS, I3S, France, prénom.nom@unice.fr

^{**}Inria, Université Côte d’Azur, CNRS, I3S, France, prénom.nom@inria.fr

^{***}Université Côte d’Azur, Département de Médecine Générale,
vhugues@outlook.fr, david.darmon@unice.fr

^{****}SynchroNext, France

Résumé. Les dossiers médicaux électroniques (DME) contiennent des informations essentielles sur les différents épisodes symptomatiques qu’un patient a subis. Cependant, les connaissances disponibles à travers ces enregistrements restent limitées : les attributs extractibles à partir de ces textes pour un algorithme d’apprentissage ne contiennent pas toutes les informations implicites connues par un expert. Afin d’évaluer et de pallier ce problème, nous avons étudié l’impact de l’augmentation des textes et des informations textuelles en provenance des DMEs par des annotations ontologiques générées automatiquement à partir de leur analyse afin d’enrichir en amont les représentations vectorielles utilisées ensuite par des algorithmes d’apprentissage.

1 Introduction

Les dossiers médicaux électroniques (DME) contiennent des informations essentielles sur les différents épisodes symptomatiques traversés par un patient. Ils possèdent le potentiel d’améliorer le bien-être des patients et constituent une source de données précieuse pour les approches d’intelligence artificielle. Dans cet article, nous augmentons avec des connaissances ontologiques les champs textuels issus des DMEs et évaluons leur impact sur la tâche de prédiction d’une hospitalisation. Notre évaluation se base sur un jeu de données réelles extrait de la base PRIMEGE PACA (Lacroix-Hugues et al. (2017)) qui contient plus de 350 000 consultations par 16 médecins généralistes. Dans cette base, les descriptions textuelles rédigées par les généralistes sont disponibles avec les codes de classification internationaux des médicaments prescrits, antécédents personnels, familiaux, facteurs environnementaux, pathologies et raisons de consultation, ainsi que les valeurs numériques des différents résultats d’examen médicaux. Les connaissances disponibles dans un DME restent cependant limitées aux spécificités de chaque patient et les textes qui s’y trouvent reposent sur un certain nombre d’informations implicites connues des experts médicaux et à des niveaux de détail variables. Aussi, un algorithme d’apprentissage automatique reposant sur ces seules informations ne pourra pas exploiter de connaissances spécifiques implicites dans les documents à analyser ou devra les réapprendre, possiblement de façon incomplète et coûteuse.