Mesures de similarité entre trajectoires de soins issues de bases de données médico-administratives

Yann Rivault, Nolwenn Le Meur, Olivier Dameron

▶ **To cite this version:**

Yann Rivault, Nolwenn Le Meur, Olivier Dameron. Mesures de similarité entre trajectoires de soins issues de bases de données médico-administratives. 8 èmes rencontres scientifiques du réseau doctoral en santé publique, Mar 2017, Rennes, France. pp.1. hal-01915618

HAL Id: hal-01915618
https://hal.science/hal-01915618

Submitted on 7 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mesures de similarité entre trajectoires de soins issues de bases de données médico-administratives

Yann Rivault[1,3,4], Nolwenn Le Meur[1,4], Olivier Dameron[2,3,4]

[1] EHESP Rennes, Sorbonne Paris Cité, EA 7449 REPERES, Recherche en Pharmaco-Epidémiologie et Recours aux Soins, France
[2] Université de Rennes 1, 35000 Rennes, France
[3] IRISA équipe Dyliss, 35042 Rennes
[4] PEPS, Pharmaco-Epidémiologie des Produits de Santé

Introduction

- Nécessité d'**analyser et de comprendre les trajectoires de soins** pour la recherche en épidémiologie et en services de santé
- Possibilité de retracer des trajectoires de soins à partir de données de soins de ville et d'hôpital issues de bases médico-administratives
- Données **complexes** et **massives**
- Des comparaisons de trajectoires de soins ont déjà été réalisées en les considérant comme des séquences simples d'évènements :

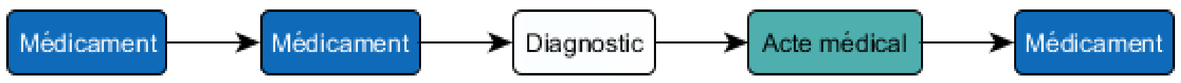

Fig. 1. Exemple d'une trajectoire vue avec le formalisme des séquences d'évènements

Enjeux et objectifs

- **Comparer des trajectoires de soins et quantifier leurs ressemblances** : une mesure qui reflète la part que deux trajectoires ont en commun
- Prendre en compte la complexité des données : massives, multidimensionnelles, simultanéité d'évènements, incertitudes sur l'ordre entre certains évènements, alphabet des évènements très important, très grande variabilité possible des trajectoires, etc

Méthodes

Calcul d'une similarité entre deux trajectoires de soins :

- Reconstitution de trajectoires de soins grâce à des données du Programme de Médicalisation des Systèmes d'Information (**PMSI**) et d'un échantillon du Système National d'Information Inter-Régime de l'Assurance Maladie (**SNIIRAM**)
- Les évènements sont composés de codes issus de trois vocabulaires contrôlés : la classification Anatomique Thérapeutique et Chimique (ATC) pour les médicaments, la Classification Commune des Actes Médicaux (CCAM) et la Classification Internationale des Maladies (CIM-10).
- Utilisation d'un formalisme plus proche de la complexité des données :

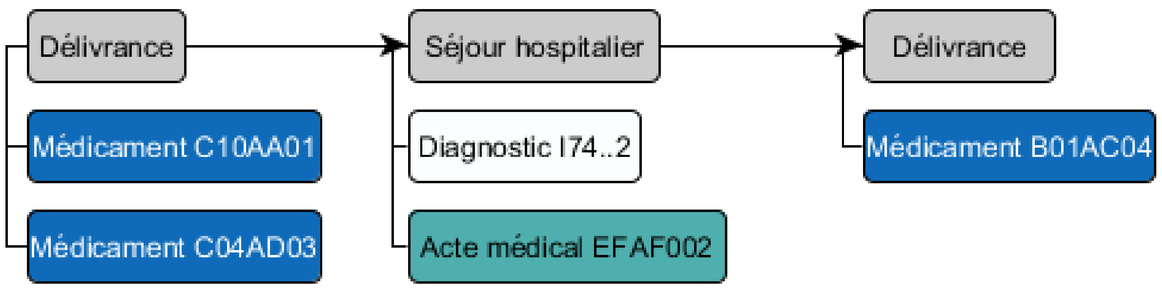

Fig. 2. Un formalisme de trajectoire de soins en séquence d'ensembles

- Généralisation du calcul de la longueur de la plus grande sous séquence commune (LCS) [1] au formalisme des séquences d'ensembles, par programmation dynamique [2]
- Proposition d'une première similarité :

$$sim(X,Y) = \frac{|LCS(X,Y)|}{max(|X|,|Y|)}$$

Enrichissement de la méthode par intégration de similarités sémantiques [3]:

- Calcul des ressemblances entre médicaments, entre diagnostics et entre actes, basé sur la structure hiérarchique des vocabulaires contrôlés avec la similarité de Wu et Palmer [4] : deux médicaments ont une similarité élevée s'ils sont proches dans la hiérarchie de l'ATC, faible s'ils sont éloignés.
- Intégration de ces similarités dans la méthode précédente

→ Plus grande sous-séquence similaire et non plus commune

Etude de la pertinence de l'approche

- Constitution d'un échantillon test de 288 trajectoires où trois groupes sont connus *a priori* : chirurgie du cristallin, chirurgie du sein et angioplastie
- Calcul des similarités entre toutes les trajectoires deux à deux, avec et sans prendre en compte les similarités sémantiques
- Classification ascendante hiérarchique avec un lien de Ward à partir des similarités entre trajectoires

Résultats

Séquences d'éléments atomiques ou séquence d'ensembles ? une comparaison :

- Temps de calculs améliorés avec le formalisme des séquences d'ensembles
- Meilleure expressivité du formalisme des séquences d'ensembles

Pertinence de l'enrichissement par similarités sémantiques :

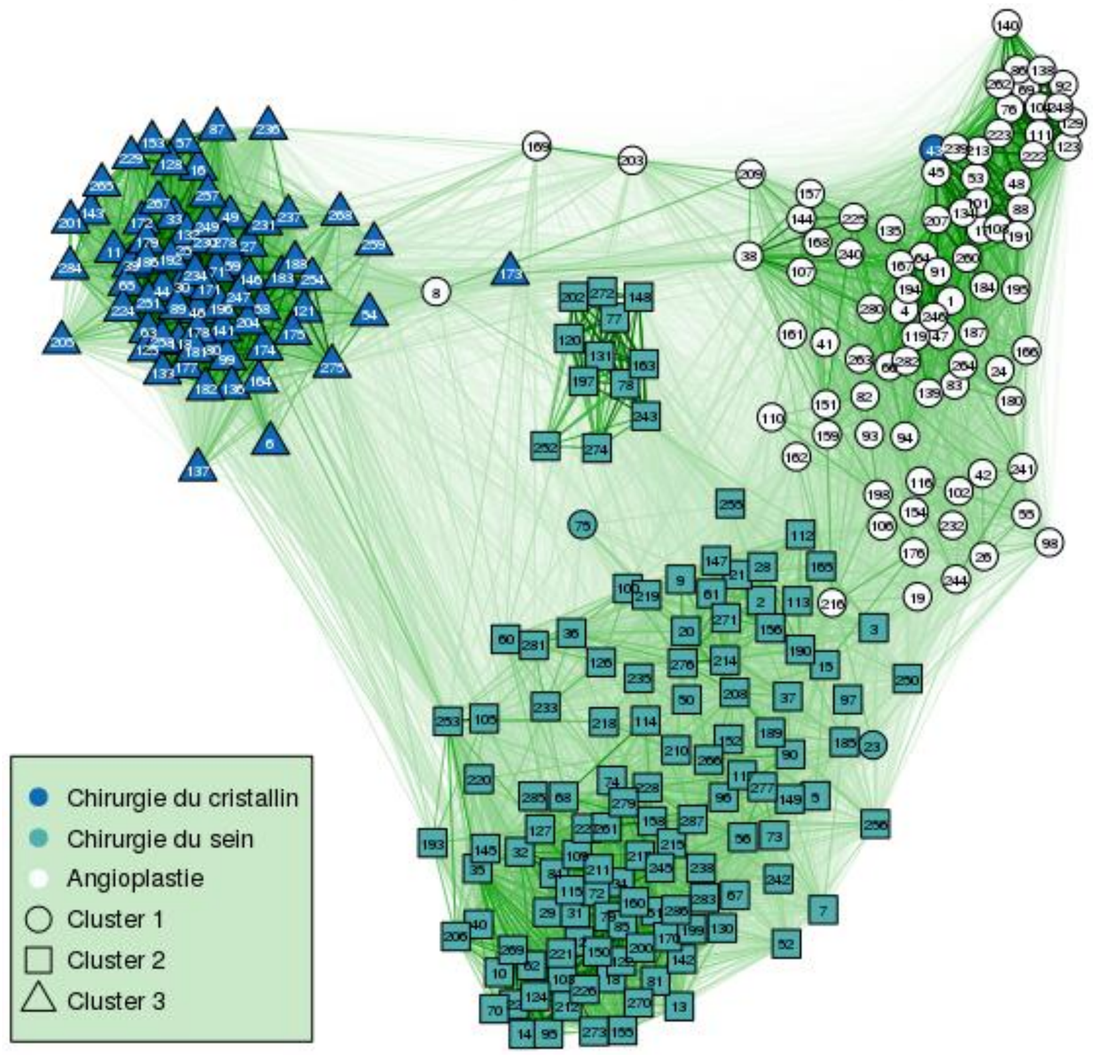

Fig. 3. Représentation des cluster et des similarités entre trajectoires

- Trois patients mal classés par la classification car ils ont des comorbidités fréquentes dans un autre groupe
- Etude du rapport entre la somme des similarités intra-classes et la somme des similarités interclasses pour toutes les trajectoires : en moyenne les similarités intra-classes augmentent plus que les similarités interclasses (test de rang de Wilcoxon p=0,002). Les classes ont gagné en homogénéité du fait de cet enrichissement.

Discussion

- Des modifications et une généralisation de la notion de la plus grande sous-séquence commune rendent la méthode adaptée à la complexité des données issues des bases médico-administratives
- Investiguer l'application de la méthode dans la comparaison de trajectoires observées à des trajectoires théoriques recommandées

Références

- Hirschberg, D.S.: A linear space algorithm for computing maximal common subsequences. Communications of the ACM. 18, 341–343 (1975).
- Bellman, R.: The theory of dynamic programming. Bull. Amer. Math. Soc. 60, 503–515 (1954).
- Pesquita, C., Faria, D., Falcão, A.O., Lord, P., Couto, F.M.: Semantic Similarity in Biomedical Ontologies. PLoS Computational Biology. 5, e1000443 (2009).
- Wu, Z., Palmer, M.: Verbs semantics and lexical selection. Presented at the Proceedings of the 32nd annual meeting on Association for Computational Linguistics June 27 (1994).