

Intelligence artificielle, enjeux réels !

Analyse de données de santé

Yann Dauxais

Unique au monde !

- Concerne l'ensemble des assurés français (> 60 000 000)
- 20 milliards de prestations disponibles (2015)
- Collecte progressive depuis 2002

Créé et principalement utilisé pour la gestion administrative



SID

Séquence



Étude des usages et effets des produits de santé sur des populations précises

Pharmaco-épidémiologie sur données SNDS

- Porté par le CHU de Rennes
- Financement ANSM (Agence National de Sécurité du Médicament)

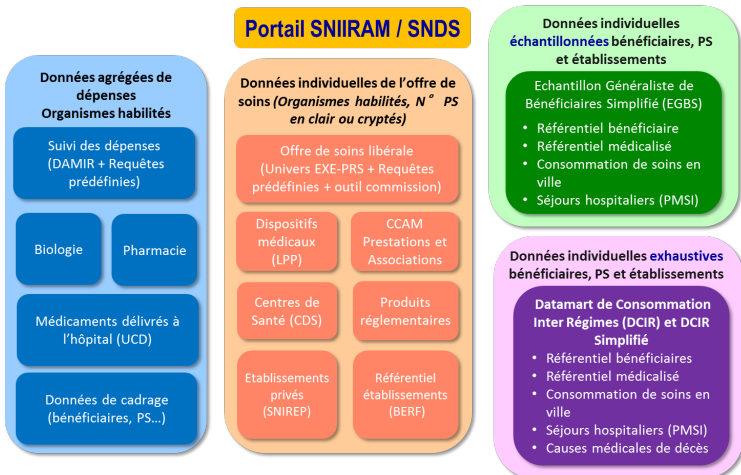
Avantages

- Large couverture de la population française
- Données déjà disponibles

Inconvénients

- Données médico-administratives \neq données médicales
- Données incomplètes ou incorrectes

Les produits de restitution par objectifs

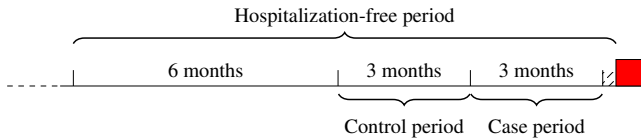


Hypothèse

- Une substitution d'anti-épileptique princeps à générique est-elle corrélée à une crise d'épilepsie ?

Protocole

- Sélection de patients épileptiques
 - Au moins une délivrance parmi une liste d'AE
- Retrait des patients "à risque" (cancer, grossesse)
- Sélection de patients stables ayant fait une crise
- Analyse statistique des différents types de substitution



Traitement lourd

- Un traitement des données pour une hypothèse

Amélioration

- Automatisation de la recherche d'hypothèses
→ **Extraction de motifs**

Analyse de données séquentielles

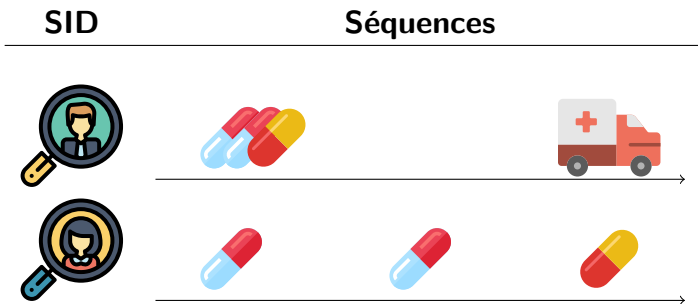
SID

Séquences



- Domaines d'application variés (Activité web, Panier de supermarché, ...)
- Données séquentielles = traces d'activités
 - Parcours de soin
- Hypothèse: ces activités sont régies par des comportements

Séquences temporelles



- Séquence temporelle = liste d'évènements temporels
 - Parcours de soin (un par patient)
- Évènement temporel = type d'évènement + une date
 - Délivrance de médicaments, hospitalisation, consultation, ...

Classification de séquences

SID

Séquences

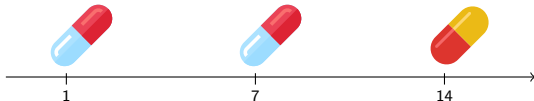


Expliquer à partir du comportement perçu l'appartenance à une population

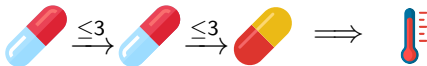
Motif temporel décrivant un comportement discriminant

SID

Sequences



Exemple



Pré-requis du modèle de comportements

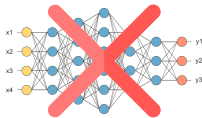
Temporel

- Description de la dimension temporelle des données



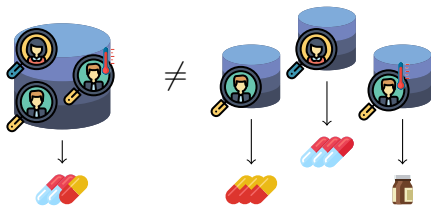
Interprétable

- Explicable
- Description des comportements



Discriminant

- Appartenance à une population

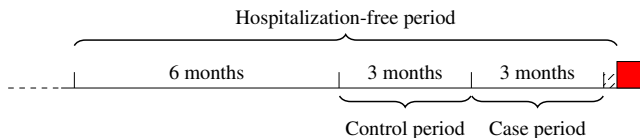


Trop de motifs / Motifs redondants

- Motifs basés sur des évènements potentiellement peu pertinents
 - Paracétamol, vaccination H1N1, ...
- Motifs décrivant un même comportement
 - Délivrance régulière de paracétamol, Délivrance régulière de paracétamol + anti-épileptique, ...

Définir l'intérêt

- Contraindre le type de motifs extraits
 - Fréquence (n'élimine pas le paracétamol)
 - Emergence/Discriminance (le paracétamol peut toujours s'accrocher à un sous-motif)
 - ?



Discriminer les périodes cas des contrôles

- 8379 patients
- 777 motifs extraits
- environ 30 motifs présentés aux cliniciens

Exemples de résultats

- Délivrances d'anti-épileptique < 30 jours
- Changement de traitement anti-épileptique
- Aspirine (anti-thrombotique) + anti-épileptique

Avantages

- Interprétabilité des motifs
- Corrélations réelles

Inconvénients

- Sélection manuelle des motifs intéressants
- Pas de test de causalité

Analyse de données de santé

- Fort potentiel d'impact sociétal
- Enjeu capital (rapport Villani)

Acceptation du modèle (par les cliniciens)

- Utilisation du modèle de motifs
 - Ouverture d'autres recherches basées sur ce modèle
- Reconnaissance de motifs

Perspectives

- Définition de l'intérêt / Sélection automatique de motifs
- Utilisation sur les données d'origine
- Beaucoup d'autres choses !

Merci pour votre attention !