

Apprentissage machine interactif pour une aide au codage PMSI

CDS 2.0 Pitching day

09/11/2016

Qui sommes-nous ?

- ▶ Médecins d'information médicale
- ▶ Assistance publique – Hôpitaux de Paris, CHU d'Île-de-France
 - ▶ Rémi Flicoteaux, Hôpital Saint-Louis, remi.flicoteaux@aphp.fr
 - ▶ Namik Taright, Siège de l'AP-HP, namik.taright@aphp.fr

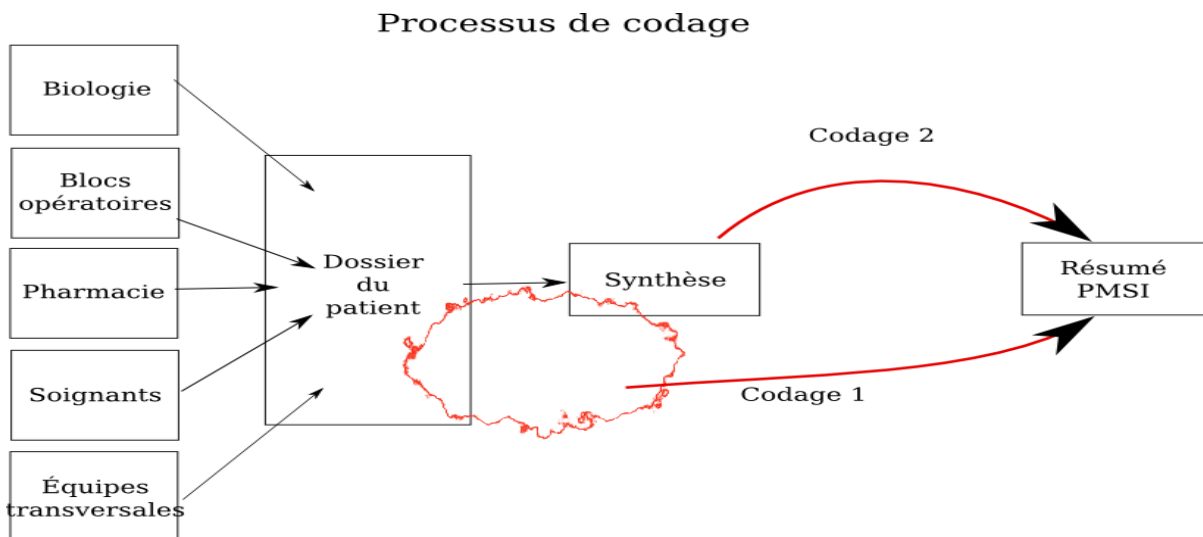
Contexte

- ▶ Les hôpitaux réalisant des soins aigus sont financés à partir d'un dispositif appelé tarification à l'activité (T2A)
- ▶ Le mécanisme est simple :
 - ▶ chaque séjour hospitalier fait l'objet d'un paiement
 - ▶ le montant du paiement est déterminé à partir d'un relevé d'activité
 - ▶ ce relevé comporte les codes correspondant aux maladies prises en charge et des procédures mises en œuvre (explorations, actes thérapeutiques)
 - ▶ les codes sont issus de classifications
- ▶ à l'AP-HP :
 - ▶ 1,3 millions de séjours par an
 - ▶ 2,7 Mds€ de recettes

Le mécanisme de codage

- ▶ introduit en France en 1989 par voie de circulaire, c'est le programme de médicalisation des systèmes d'information hospitaliers (PMSI)
- ▶ réalisé à partir des informations du dossier du patient
- ▶ utilise des outils de codage
 - ▶ pour les maladies : la classification internationale des maladies (version 10, OMS), autour de 15 000 codes
 - ▶ pour les procédures : la classification commune des actes médicaux (classification française, mise au point et maintenue par l'Assurance – maladie), environ 8 000 codes

Un processus fragile



- ▶ Codage 1 : réalisé à partir de la mémoire du cas pris en charge
- ▶ Codage 2 : réalisé à partir de la lecture de la synthèse (dont on a démontré qu'elle ne contient en moyenne que 50% des informations pertinentes)

Rôle des médecins d'information médicale

- ▶ Formations au codage
- ▶ Validation du contenu du codage
 - ▶ de 1 à 99 codes de maladies par résumé (en moyenne, 4,5 codes par résumés)
 - ▶ de 0 à 999 codes d'actes par résumé
- ▶ Enjeux sur les recettes
 - ▶ optimisation de la valorisation
 - ▶ réduction des contrôles de l'Assurance-maladie

Tâche

- ▶ Produire des codes à partir des informations issues des données du dossier informatique du patient (electronic medical/health record)
- ▶ Données
 - ▶ textuelles
 - ▶ semi-structurées : compte rendu (CR) d'hospitalisation, CR opératoire, CR de réunion de concertation multidisciplinaire, ...
 - ▶ texte libre : dossier de soins, observations médicales
 - ▶ des données structurées de biologie (chaque examen est identifié dans un catalogue par un code, une unité, une valeur),
 - ▶ des prescriptions médicamenteuses (structurées : multi-axiale)

Premières expériences de prédiction

Données

- ▶ comptes rendus d'hospitalisation de deux spécialités médicales
- ▶ base MIMIC
- ▶ certificats de décès du CepiDC (challenge Clef)
- ▶ comptes rendus opératoires

Objectif et méthodes

- ▶ prédiction d'un code
- ▶ NB, SVM, sLDA, labelledLDA

Premiers résultats

- ▶ de meilleures performances pour SVM
- ▶ prédiction acceptable en multi-code

Références

- ▶ M. Dermouche, J. Velcin, R. Flicoteaux, S. Chevret and N. Taright. Supervised Topic Models for Diagnosis Code Assignment to Discharge Summaries. Proceedings of the 17th International CICLING, Turkey, 2016.
- ▶ M. Dermouche, V. Looten, R. Flicoteaux, S. Chevret, J. Velcin, N. Taright. ECSTRA-INSERM @ CLEF eHealth2016-task 2: ICD10 Code Extraction from Death Certificates.

Propositions

Amélioration de la performance des méthodes déjà utilisées

- ▶ Tester d'autres méthodes d'apprentissage supervisé
- ▶ Enrichir les données

Apprentissage interactif (human in loop)

- ▶ questions algorithmiques
- ▶ validations

Données

- ▶ open data : base MIMIC III (patients de réanimation)
- ▶ entrepôt de données de santé de l'AP-HP
 - ▶ en cours de constitution
 - ▶ données sensibles
 - ▶ accès réglementé (travaux en cours)
 - ▶ perspective open data envisagée

Constitution d'un entrepôt de données de santé (EDS) Une opportunité pour l'AP-HP

