

Vers la fusion d'informations hétérogènes et partielles pour l'aide au codage diagnostique

Laurent Lecornu^{*,***}, Clara Le Guillou^{**,*}, Frédéric Le Saux^{*}, Matthieu Hubert^{**},
Julien Montagner^{*,***}, John Puentes^{*,***}, Jean-Michel Cauvin^{*,***}

* Telecom Bretagne , Technopole Brest Iroise, CS 83818
29238 Brest Cedex 3

Laurent.Lecornu@telecom-bretagne.eu

**CHRU de Brest, DIM, Boulevard Tanguy Prigent
29609 Brest Cedex

jean-michel.cauvin@chu-brest.fr

***Latim, Inserm U650

1 Contexte

En effet, depuis la mise en place du Programme de Médicalisation des Systèmes d'Information (PMSI), l'ensemble des informations médicales concernant le patient au cours de son séjour à l'hôpital doit être recueilli sous une forme standardisée et codée. Malheureusement, choisir un code diagnostique est une opération non-intuitive pour le praticien en raison de la quantité de codes prévue par les normes, et de la difficulté d'en établir une classification pertinente. Les erreurs sont fréquentes avec des conséquences graves tant au niveau de l'évaluation des soins de santé qu'au niveau financier. Pour aider le médecin-codageur, l'approche choisie consiste à lui présenter une liste de codes diagnostiques adaptée au cas du patient dont il code le séjour, ordonnée par ordre de pertinence, et correspondant aux éléments diagnostiques qu'il a en tête à partir des informations réparties au sein du Système d'Information Hospitalier (SIH) et sur des connaissances externes. Dans ce but, un ensemble de modules spécifiques d'aide au codage a été développé. Chacune d'entre eux analyse une partie spécifique des données du SIH et en extrait une information partielle sous la forme d'une liste ordonnée et contextualisée de codes diagnostiques classés selon une valeur de pertinence qui lui est propre. L'ensemble des listes est alors fusionné pour ne présenter au médecin qu'une seule liste ordonnée. Deux stratégies d'ordonnement de ces valeurs sont explorées (la 1^{ère} méthode est dérivée d'une fusion procédurale et la 2^{ème} utilise la théorie des possibilités).

2 Méthode

Le premier module développée s'appuie sur une étude statistique des codes diagnostiques en fonction de l'âge du patient, de son sexe, de la durée du séjour, de l'unité médicale fréquentée, des actes médicaux effectués et des diagnostics déjà codés (Lecornu (2009b)). Le second s'appuie sur le codage de séjours antérieurs en tenant compte du temps écoulé entre

les séjours. Le choix des diagnostics repose sur l'étude de la courbe de survie de la réapparition du code (Lecornu (2009a)). Enfin, le troisième module exploite les résultats d'analyses biologiques (Lecornu 2010) grâce à un ensemble de règles associées à une valeur de pertinence linguistique tel que « rare » ou « souvent ».

Deux méthodes de fusion ont été envisagées afin de résoudre le problème d'hétérogénéité de des valeurs de pertinence. La première approche développée s'appuie sur une méthode procédurale c'est-à-dire basée sur un système de vote qui permet d'effectuer un classement par le nombre de voix. Pour classer les codes ayant le même nombre de voix, on utilisera la valeur de pertinence du module considéré comme le meilleur (ici, AnterOcod, suivit de ReferOcod et d'AlertOcod). La deuxième s'appuie sur un tableau de conversion, déterminé empiriquement pour chaque module et utilisant la précision obtenue dans chaque module pour un intervalle de pertinence ou pour un terme linguistique. La méthode de fusion s'appuie sur la théorie des possibilités et les opérateurs de fusion qui peut y être associé.

3 Conclusion

Les résultats obtenus permettent de retrouver 80% des codes diagnostiques (sur un échantillon de 998 séjours anonymisés) qui avaient été codé par le médecin, malgré la difficulté d'associer à chaque valeur de probabilité, ou de terme linguistique, une valeur de nécessité cohérente. La seconde méthode de fusion apparait plus efficace car les codes à trouver sont mieux placés dans la liste.

Ce projet est un projet TECSAN soutenu par l'ANR sous le nom Med1dex associant le CHU de Brest, TELECOM Bretagne, le LaTIM Unité INSERM U650 et la société PRISMEDICA.

Références

- Lecornu, L., C. Le Guillou, G. Thillay, P. J. Garreau, H. Jantzem, J. M. Cauvin (2009a). C2i: A tool to gather medical indexed information, *ITAB 2009*. Larnaca, Cyprus, 5-7 Nov.
- Lecornu, L., G. Thillay, C. Le Guillou, P. Saliou, J. Puentes, H. Jantzem, P. J. Garreau, J. M. Cauvin (2009b). REFEROCOD : a probabilistic method to medical coding support, *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 02-06 sept. 2009, Minneapolis, United States, 3421-3424
- Lecornu, L., C. Le Guillou, F. Le Saux, M. Hubert, J. Puentes, J. M. Cauvin (2010). ANTEROCOD: Actuarial Survival Curves Applied to Medical Coding Support for Chronic Diseases, *32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, sept. 2010, Buenos Aires, Argentine.

Summary

To support the discharge abstract coding by the physician, we developed a coding-aid system composed of a set of modules connected to the Hospital Information System. A fusion module has been developed to synthesize the partial information extracted from each module.