Pertinence Medicale des Cooccurrences Diagnostic-Acte dans les Resumes Standardises de Sortie

N. GRIFFON [a], P. MASSARI [a 1], M. JOUBERT[b], P. STACCINI[b], S.J. DARMONI[a]
[a] *CISMeF, Rouen University Hospital, Rouen, France & TIBS, LITIS EA 4108, Institute of Biomedical Research, Rouen, France*
[b] *SESSTIM UMR 912, Université Aix-Marseille, Université Nice-Sophia Antipolis*

Abstract. The frequency with which a principal diagnosis (PD) and a procedure match in standardized summary of discharge (RSA) is supposed to reflect medical practices. Medically validated "diagnosis-procedure" associations could have various uses including: coding assistance, assessment of practices.... In French DRG system, many of these co-occurrences (CO) are artificial. The aim of this study is to measure the interest using semantic relationships in association with statistical methods to highlight CO that have medical relevance. Data are issued from the 798,916 RSA of PACA region for the year 2009. To assess the contribution of semantic relations provided by the CISMeF "super concepts", we compared the medical relevance of CO between: 1) a semantic sample of CO with PD and act connected to the same super concept and 2) a control sample. This comparison was performed at different levels of statistical association. Semantic sample includes 620 CO and control sample includes 560 CO. Overall, the medical relevance was significantly higher in the semantic sample (73.7% vs. 55.2%; p<10, Fischer test). This association was consistently found at every level of statistical association. Conclusion: using both statistical and semantic knowledge allows a more accurate selection of medically relevant co-occurrences between acts and diagnoses.

Keywords. Data mining - Vocabulary, controlled Clinical coding - Patient discharge - Electronic health record.

Introduction

La fréquence de l'association du diagnostic principal et d'un acte (ou cooccurrence) dans les résumés standardisés de sortie anonymisés (RSA) devrait être un reflet des pratiques, ces associations correspondant schématiquement à 3 circonstances :

- l'acte pratiqué permet de faire le diagnostic (positif, différentiel) ou contribue à celui-ci ;
- l'acte rentre dans le cadre du bilan (extension, gravité) de la pathologie du patient ;
- l'acte permet de traiter la pathologie représentée par le diagnostic codé.

[1] Correspondant : Philippe Massari, Unité d'informatique médicale, 76031 ROUEN Cedex

Mais la fréquence de certaines associations (ou cooccurrences) est augmentée artificiellement, pour des actes ayant un caractère systématique, lorsque le diagnostic justifiant l'acte est en position de diagnostic associé de façon justifiée ou du fait d'erreurs de codage. Il est alors difficile de détecter les associations ou cooccurrences ayant véritablement un sens médical.

Des cooccurrences "diagnostic-acte" validées peuvent avoir de nombreuses applications :

- évaluation des pratiques [1] ;
- proposition de codes diagnostiques après la pratique d'un acte ;
- présentation lors de la prescription d'une liste d'actes en fonction des diagnostics enregistrés pour le patient.

Le présent travail est basé sur l'analyse d'une banque de RSA. Son objectif est de mettre en évidence des cooccurrences pertinentes en combinant des outils statistiques de mesure de la cooccurrence [4] et en utilisant les relations sémantiques des codes actes et des codes diagnostiques grâce à des super-concepts : les métatermes CISMeF [2],[3].

1. Méthode

Les métatermes (MT) sont des super-concepts initialement définis pour représenter une spécialité médicale ou une science biologique au sein du MeSH. Le conservateur des bibliothèques de l'équipe CISMeF a ensuite créé des liens sémantiques entre les descripteurs MeSH et les métatermes. Cette démarche a ensuite été étendue à d'autres terminologies, en particulier à la CIM10 et à la CCAM : les relations sémantiques entre chacun des codes de ces deux terminologies et les métatermes ont été créées manuellement. Les MT sont utilisés à la fois pour la recherche d'information sur le Web [2] et aussi pour catégoriser les concepts dans le dossier médical [3].

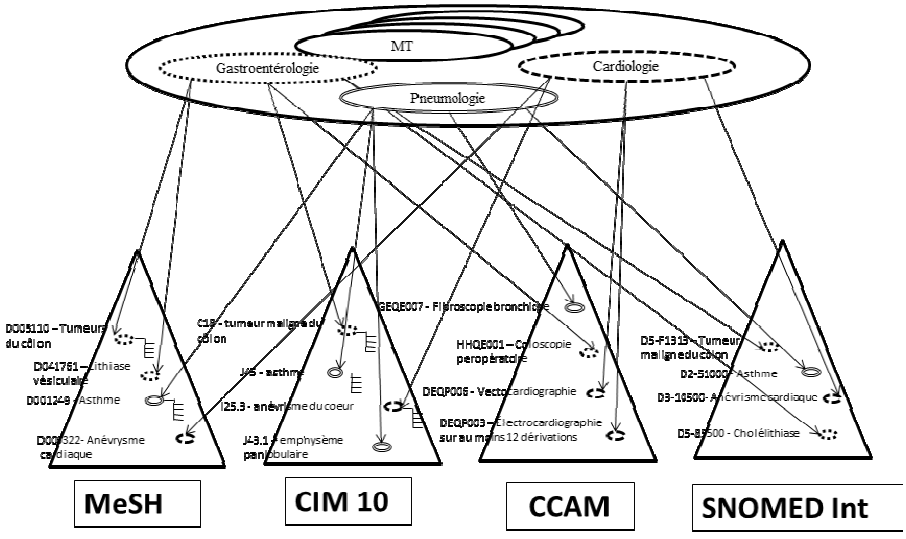

Figure 1. Les métatermes CISMeF.

Les données sources utilisées sont les RSA de la région PACA (Provence-Alpes-Côte d'Azur) pour l'année 2009. A partir de ces 798 916 RSA, un des co-auteurs (MJ) a constitué une base de données identifiant 354 795 cooccurrences entre un diagnostic principal (DP) et un acte. Chacun de ces couples a été associé, dans la base, à sa fréquence et aux fréquences spécifiques de l'acte et du diagnostic.

Notre approche a été de sélectionner un nombre analysable de cooccurrences susceptible d'avoir une pertinence médicale. Pour ce faire, les cooccurrences dont l'acte CCAM correspondait à un supplément (actes débutants par YY ou ZZ), ou présentes dans moins de 30 RSS ont été exclues. De même, les cooccurrences dont le diagnostic principal commençait par la lettre Z ont été exclues puisqu'elles correspondent généralement à des séances (radiothérapie, chimiothérapie…). Pour évaluer l'apport des relations sémantiques fournies par les métatermes CISMeF, deux groupes de cooccurrences ont été étudiés : (1) un "échantillon témoin" aléatoire représentatif de l'ensemble des cooccurrences respectant les critères de sélection et (2) un "échantillon métatermes" composé de l'ensemble des cooccurrences, pour lesquelles l'acte et le DP étaient rattachés aux métatermes « cardiologie » ou « chirurgie thoracique et cardiovasculaire ».

Pour chaque cooccurrence, la force de la cooccurrence a été mesurée à l'aide du coefficient de confiance [4]. Il s'agit de la probabilité d'observer l'acte dans un RSA conditionnellement à la présence du DP dans ce RSA :

$$confiance_{DP,Acte} = p(Acte/DP) \quad (1)$$

On a en outre calculé : la c*onfiance* $_{Acte,\ DP}$ *et le lift :*

$$confiance_{Acte,DP} = p(DP/Acte) \quad (2)$$

$$lift = \frac{confiance_{DP,Acte}}{p(Acte)} \quad (3)$$

La pertinence médicale de chaque couple "diagnostic-acte" cooccurrent des 2 échantillons a été évaluée par l'expert. Les couples ont été jugés médicalement pertinents quand il y avait un lien de cause à effet entre le diagnostic et l'acte. La précision correspond à la proportion de cooccurrences médicalement pertinentes. Les précisions de l'approche statistique et de l'approche combinant statistiques et relations sémantiques ont été comparées.

2. Results

Le nombre de cooccurrences présentes 30 fois ou plus et dont l'acte ne correspond pas à un supplément CCAM est de 10 394. Le DP et l'acte étaient reliés aux métatermes "cardiologie" ou "chirurgie thoracique et cardiovasculaire" dans 620 cas. L'échantillon témoin comprenait 560 cooccurrences.

Les moyennes des confiances$_{DP,Acte}$ n'étaient pas significativement différentes entre les deux échantillons : 0,18 pour l'échantillon métaterme vs. 0,19 pour l'échantillon témoin (p=0,41, test de Mann-Whitney). La précision au sein de l'échantillon témoin était de 55,2% [52%-60%]$_{95\%}$. Elle était significativement plus élevée ($p<10^{-3}$; test de Fisher) dans l'échantillon métatermes : 73,7% [70%-77%]$_{95\%}$. Le tableau 1 montre que

la précision au sein de l'échantillon métatermes est supérieure à la précision au sein de l'échantillon témoin quelque soit le niveau de confiance$_{DP,Acte}$. On observe également que ce gain en précision ne s'effectue pas au détriment du nombre d'occurrences.

Tableau 1, Comparaison des précisions entre les deux échantillons à différents seuils de niveau de confiance.

Seuils de niveau de confiance	Échantillon métatermes	Échantillon Témoin	p*
≥ 0,6	100% (20/20)	86,2% (25/29)	0.13
≥ 0,5	100% (65/65)	78,4% (40/51)	<10-3
≥ 0,4	100% (100/100)	71,8% (56/78)	<10-3
≥ 0,3	97,9% (143/146)	70,1% (82/117)	<10-3
≥ 0,2	91,5% (195/213)	66,7% (130/195)	<10-3
≥ 0,1	86,1% (278/323)	63,1% (190/301)	<10-3
≥ 0	73,7% (457/620)	55,7% (312/560)	<10-3

* : test de Fisher

Les distributions de la confiance $_{DP,Acte}$, de la confiance $_{Acte, DP}$ et du lift en fonction de la pertinence sont présentés dans la figure 2.

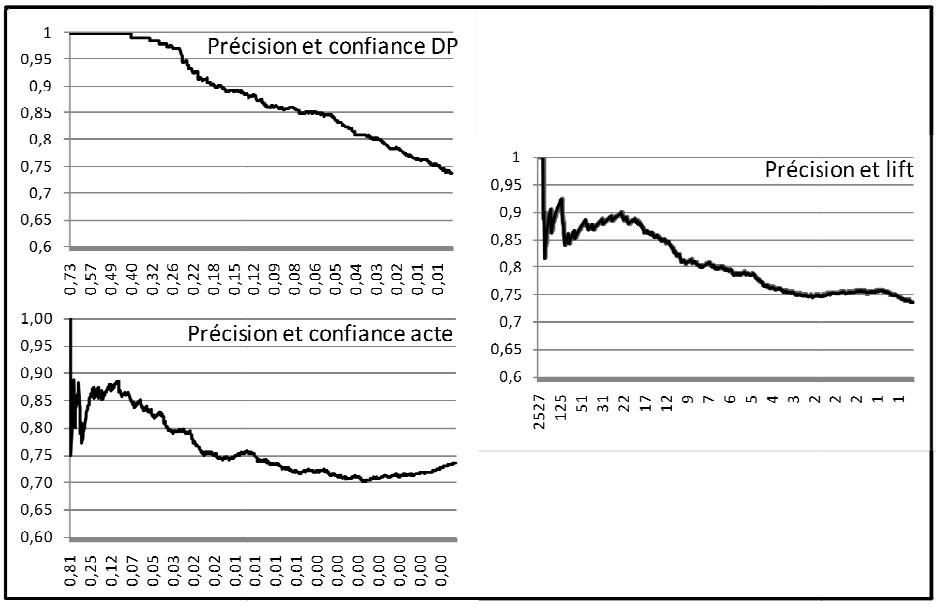

Figure 2. Précision et mesures statistiques pour l'échantillon métaterme

3. Discussion

De nombreuses cooccurrences DP - acte de l'échantillon témoin sont sans sens médical (44,3%). Ceci a déjà été observé par d'autres auteurs [5]. De nombreuses mesures relatives à l'intérêt des règles d'association ont été proposées [4], dans le domaine médical la définition de seuils permettant d'atteindre une précision suffisante est difficile. Adboune et coll [6] ont proposés un seuil du lift à 1 pour des cooccurrences de termes MeSH de notices bibliographiques. Ce critère leur paraît insuffisant pour

affirmer la pertinence médicale. Ils proposent, comme d'autres auteurs [5],[7], d'associer ces critères statistiques à des relations sémantiques. Nous avons choisi d'utiliser les relations sémantiques fournies par les métatermes CISMeF : "cardiologie" et "chirurgie cardio-vasculaire". Nous les avons associés car le traitement d'un bon nombre de pathologies cardiologiques est chirurgical ou tout du moins fait appel à un acte invasif. Les couples DP-acte dont les 2 éléments sont liés à ces 2 spécialités et présents dans 30 RSA ou plus ont un sens médical dans plus de 73% des cas. En tenant compte de la confiance DP il semble possible de choisir une précision adaptée à l'utilisation de ces cooccurrences, pour les métatermes que nous avons étudiés. Il serait nécessaire de valider cette méthode avec d'autres métatermes pour confirmer son intérêt.

References

[1] Wang MC, Laud PW, Macias M, Nattinger AB. Utility of a combined current procedural terminology and International Classification of Diseases, Ninth Revision, Clinical Modification code algorithm in classifying cervical spine surgery for degenerative changes.Spine (Phila Pa 1976). 2011 Oct 15;36(22):1843-8.

[2] Gehanno JF, Thirion B, Darmoni SJ. Evaluation of Meta-concepts for Information Retrieval in a Quality-Controlled Health Gateway.. AMIA Symp., Pages 269-73, IOS Press, 2007.

[3] Massari P, Pereira S, Thirion B, Derville A, Darmoni SJ. Use of super-concepts to customize electronic medical records data display. Studies in Health Technology and Informatics, Volume 136, Pages 845 - 850, 2008.

[4] Vaillant B, Meyer P, Prudhomme E, Lallich S, Lenca P, Bigaret S. Mesurer l'intérêt des règles d'association. Revue des Nouvelles Technologies de l'Information : Extraction et gestion des connaissances : état et perspectives, 2006, vol. RNTI-E-5, pp. 421-426. http://www.auroras.fr/vaillant/IMG/pdf/vaillant_etal_dkq2005.pdf (accédé le 26/09/2012)

[5] Avillach P, Joubert M, Fieschi M. Improving the quality of the coding of primary diagnosis in standardized discharge summaries.Health Care Manag Sci. 2008 Jun;11(2):147-51.

[6] Abdoune H, Soualmia L, Joubert M. Analyse de cooccurrences de concepts biomédicaux dans MEDLINE. 1ère édition du Symposium sur l'Ingénierie de l'Information Médicale SIIM 2011, Toulouse, 9 & 10 Juin 2011.http://www.irit.fr/SIIM/SIIM2011-6.pdf (accédé le 26/09/2012)

[7] Rassekh SR, Lorenzi M, Lee L, Devji S, McBride M, Goddard K. Reclassification of ICD-9 Codes into Meaningful Categories for Oncology Survivorship Research. J Cancer Epidemiol. 2010;2010:569517. Epub 2010 Dec 29.