Information Fusion for Diagnosis Coding Support

L. Lecornu, C. Le Guillou, F. Le Saux, M. Hubert, J. Puentes, J. Montagner, J. M. Cauvin

Abstract - **All patient-related medical information during a hospital stay in France, has to be collected and coded in the compilation of medical units discharge documents, according to a standardized approach. The process of describing a patient disease in terms of appropriate diagnostic codes is nevertheless, a non-intuitive operation for the physician. As a consequence, coding errors, inaccuracies and missing data are frequent, leading to potentially severe economical upshots. A coding support system developed to improve medical coding results, integrates three information processing methodologies, using the outputs from various Hospital Information System applications. Each methodology generates partial heterogeneous information, with considerable semantic variety. In order to properly synthesize these outputs, information fusion is required to produce enriched contextualized information, presented to the physician as an ordered list of suggested codes. This paper explores two information fusion approaches: voting system and possibilistic. Both methods are tested on a database of 1,000 discharge summaries, to show the interest of information fusion in this context. Results show that fusion methods perform better in most of the cases than partial information extraction methods.**

I. INTRODUCTION

The French medical information policy makes mandatory to code each procedure and diagnosis of the discharge summary document, which is a compilation of medical units discharge documents and does not include discharge letters. Nationwide hospital funding has been directly calculated relying on this information since 2004. However, coding is complex, time consuming and considered by physicians as a boring task. Since coding mistakes are common, any hospital may suffer the consequences of these errors, eventually leading to wrong statistics and hence insufficient funding. To cope with this problem, hospitals may consider working with professional coders (more expensive), apply automatic coding (complex to exploit), or use the emerging alternative of coding support systems (technologically challenging).

Diagnostic coding support usually involves different approaches and sources: e.g. easy navigation through the nomenclatures, use of previous codes associated to a patient,

detection of abnormalities, and probabilistic estimations, among others. While the medical coder needs one global and coherent list per patient to find the proper codes, it is not reasonable to expect that the coder will separately interpret each information source to build a list of codes before making her or his choices. This paper evaluates and compares two strategies, voting system and possibilistic, to accomplish information fusion, intended to ameliorate the medical coder task, by suggesting a single list of diagnosis codes adapted to the care situation, ranked by relevance and according to the coder expected account of a patient stay.

Before information fusion can be applied, three different methods analyze specific parts of various Hospital Information System (HIS) sources, concerning the patient hospital stay. Thereafter, information fusion consists in aggregating the partial information obtained by each method, taking into account the heterogeneity of multiple values, considered as relevant by the previous processing. Two information fusion methods are examined - procedural voting system and recall/precision combined with possibility theory – and tested on a database of 1,000 discharge summaries, containing 1 to 23 codes. Separate and aggregated results are independently evaluated and compared to illustrate the distinct behavior of each method, along with the advantages of information fusion.

In France, as increasingly in many other countries, hospitals are funded according to the evaluation of standardized information [1]. This information collected in the discharge summary only concerns hospitalizations, and is transmitted after anonymization, to governmental health services, to be periodically analyzed. The French hospital discharge document contains elementary discharge summaries, from all medical units that provided healthcare during the patient stay. Each elementary discharge summary contains patient demographics, a main diagnosis, associated diagnoses corresponding to secondary diseases and adverse events, complemented by annex diagnosis and/or therapeutic procedures. All these information elements are related to the consumption of healthcare resources.

Diseases are coded according to the International Classification of Diseases – 10th revision (ICD-10), published and maintained by the World Health Organization [2]. ICD-10 is mainly used in many countries for registering morbidity and mortality causes, as well as to facilitate the organization of healthcare services. The ICD-10 contains nearly 17,000 entries corresponding to diseases, traumatisms, symptoms, and other reasons for using health services, classified in 22 chapters. Each diagnosis code is composed of a letter, followed by 2 to 4 digits.

II. System Architecture

Diagnosis coding support objective is to generate a list of appropriate contextualized (adapted to a specific patient hospital stay) diagnosis codes. Those codes are defined after the fusion of heterogeneous information extracted from various applications of the HIS, or external knowledge sources. The system consists of three methods that independently extract specific information from structured documents stored in the HIS (Fig. 1). Two of these methods, related to the analysis of biological results (AlertOcod [3]) and the temporal pattern of previously existing codes (AnterOcod [4]), mainly use patient and linked hospital stay data. The third method (ReferOcod [5]), is based on the probabilistic relationships between diagnosis codes, observed in a large anonymized discharge summaries database. Information extraction methods briefly described in the next section, assert the relevance of partial extracted information differently, as linguistic label, recall rate, or probabilistic values.
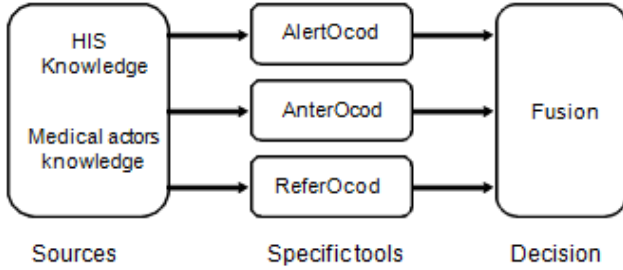

Fig. 1: Architecture of the information fusion system.

III. Partial Information Extraction Methods

A. Analysis of Biological Results

The AlertOcod method retrieves appropriate patient condition indicators and data of laboratory examinations, from distributed sources within the HIS, in order to propose diagnosis codes. It is structured in three parts:

- Rules to describe a patient's condition, the nature of the laboratory results that characterize it, and to propose a set of diagnosis codes for that particular patient condition-laboratory results data collection.
- Alerts to notify the user that the system has found a group of facts, agreeing with a rule applied to a specific patient case, enabling therefore to provide relevant coding.
- Management unit to automatically handle the interface between different data sources.

Each code proposal is generated with a comparative importance value, represented by a linguistic label, for example rare or often, previously defined by an expert [3].

B. Patterns of Previously Existing Codes

The AnterOcod method analyses diseases elements recurrence, to identify if previously existing codes reveal temporal occurrence patterns. It is assumed that a relatively chronic disease, likely to have been described in previous hospital stays, can be proposed as a pertinent coding option for the current stay. Adapted actuarial survival models are used to estimate the recurrence of diagnosis codes corresponding to chronic diseases, on a reference database of hospital stays discharge summaries. Calculated estimations constitute a knowledge base of recall rates for each diagnosis code, which depend on the elapsed time since it was previously used. Additional knowledge represented by specific codes assigned to the patient during the two years preceding the current hospital stay, is applied to construct the proposed partial list of codes. Results include, besides the ordered codes and associated labels, recall rates that define the code importance [4].

C. Probabilistic Relationships

The ReferOcod method makes use of a knowledge base, extracted from a large database of anonymous discharge summaries. Knowledge is formed by probabilistic predictions of diagnosis codes depending on patient age, sex, hospital stay length, diagnoses, and medical acts. To reduce the number of possible combinations, diagnoses probabilities are grouped according to the impact of age, stay length, diagnoses, and accomplished medical acts.

The probability of diagnosis D_j calculated using the previous elements, considers four information sources:

(1) Age, sex and stay length,
(2) Medical unit / functional unit (UM),
(3) Procedures already encoded ($proc_N$),
(4) Already coded diagnoses (D_j),

which are applied to calculate the following conditional probabilities:

$$\hat{P}\left(D_j \mid age, sex, stay\ length\right),\ \hat{P}\left(D_j \mid UM\right),$$
$$\hat{P}\left(D_j \mid proc_1,...proc_N\right)\ \text{and}\ \hat{P}\left(D_j \mid D_1,...D_M\right) \quad (1)$$

The diagnosis code probability is then computed by linear combination of the previous four conditional probabilities:

$$\hat{P}\left(D_j\right) = \beta_1 \cdot \hat{P}\left(D_j \mid age, sex, stay\ length\right)$$
$$+ \beta_2 \cdot \hat{P}\left(D_j \mid UM\right)$$
$$+ \beta_3 \cdot \hat{P}\left(D_j \mid proc_1,...proc_N\right)$$
$$+ \beta_4 \cdot \hat{P}\left(D_j \mid D_1,...D_M\right) \quad (2)$$

Where β_1, β_2, β_3 and β_4 depend on the performance of individual information sources, and were determined by experimental evaluation. A partial list of codes is generated, along with the respective labels and estimated probabilities, which indicate how pertinent each code is [5].

IV. Partial Information Fusion

To produce one enriched ordered list of suggested diagnosis codes, adapted to the facts that took place during the patient hospital stay (contextualized list) it is necessary to aggregate the three partial codes lists described before, complying with the respective relevance indicators. Because of partial results heterogeneity, our work focuses on suitable partial information fusion methods.

A. Procedural Approach

This method is based on a voting system [6], to determine the underlying codes classification, counting how many times any code was proposed by the information extraction methods (1 to 3 times). Given the complementary nature of partial extracted information, many codes obtain the same amount of votes, implying that information sources degree of credibility must be taken into account. Such analysis [3-5] yields the best prediction scores for patterns of previously existing codes, followed by probabilistic relationships and biological results. Credibility estimation is applied subsequently, to refine the ranking of codes having initially the same number of votes.

B. Possibility Theory Based Approach

This method orders diagnosis codes according to the different relevance values. Because of relevance values heterogeneity, it is required to find first a common ground for the analysis. Code accuracy, defined as the ratio between how many times each code is chosen by the physician, with respect to the number of times the code is proposed by the information extraction method, is applicable to this end. Thereafter, the intrinsic accuracy of each code is represented by a possibility value [7], in order to overcome the uncertainty of probability estimations, recall rates and especially linguistic labels, enabling to merge them accordingly. Accuracy is studied for each information extraction method in terms of the transformed possibilistic relevance (necessity / possibility pairs), from which a conversion table is experimentally defined. Although, the ideal solution in this case would be to find a monotonic conversion function to transform relevance values, since their range is divided in intervals associated to accuracy values, results are smoothed to achieve a monotonically decreasing trend. Values are merged using a conventional fusion operator, e.g. the max operator [8].

C. Combined Evaluation

The performance of the two fusion methods was evaluated on a sample of 1,000 discharge summaries, representing hospital stays of more than 24 hours, having at least 2 diagnosis codes and one medical procedure code. Such evaluation is intended to facilitate afterwards, the comparison of codes proposed by the diagnosis coding support system, to the codes produced by physicians alone.

Information fusion evaluation is based on known indicators (recall-precision) applied to assess the performance of retrieval systems, adapted to the multiplicity of diagnoses described in the discharge summary. Whereas recall is defined as the number of pertinent retrieved codes divided by the total number of existing pertinent codes, precision is defined as the number of retrieved pertinent codes divided by the total number of retrieved codes. Figure 2 shows the recall and precision rates curves, depending on the number of pertinent proposed codes. Both partial information fusion methods propose relatively the same codes in the first 15 ranks. However, a relevant diagnosis code, found by only one extraction method could be pushed down the proposed ordered list, for instance if the credibility factor is rather low.

V. Discussion

The two partial information fusion methods outperform the three partial information extraction methods, in terms of recall closely followed by probabilistic relationships.

Although the recall and precision curves of the two partial information fusion methods give the impression that results are identical, there are some particular differences, related to slight order of magnitude disparities not revealed by the graphic resolution. To better understand those differences a specific coding example is analyzed. Table I presents the coding of a hospital stay made by a physician, and table II presents the ordered lists of codes proposed by each information fusion method. We observe that 14 proposed codes are the same for both methods, while two (one per method) were expected but not proposed (n.p.). Under these conditions the valid codes order is expected to be different. Results of table II confirm this hypothesis: four codes, N18, Z49, Z94, and D64, are placed at positions 1, 3, 5, and 6 by the possibilistic fusion, and positions 1, 3, 10, and 5 by the vote fusion. These variations generate the differences not observed in Fig. 2, which have nevertheless an impact when the physician uses the diagnosis coding support system.
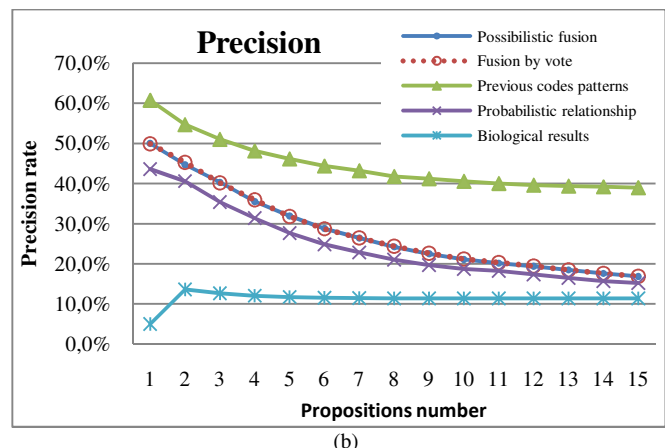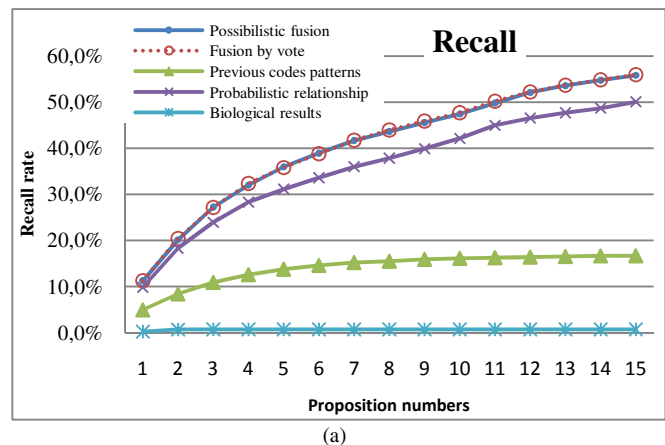

(a)

(b)

Fig. 2: Evaluation results of the recall (a) and precision (b) rates.

Discharge summary ICD10 codes and label

D64	Other anemias
N18	Chronic kidney disease (CKD)
T86	Failure and rejection of transplanted organs and tissues
Z00	General examination and investigation of persons without complaint and reported diagnosis
Z49	Encounter for care involving renal dialysis
Z94	Transplanted organ and tissue status

TABLE II: EXAMPLE OF THE PROPOSED CODING (FOR THE STAY OF TABLE I) BY THE TWO FUSION METHODS

Proposed ICD10 codes and label	Possibilistic		Vote	
	rank	value	rank	sources
D47 Other neoplasms of uncertain behavior of lymphoid, hematopoietic and related tissue	13	0,130	8	1
D64 Other anemias	**6**	0,632	**5**	1;2
D69 Purpura and other hemorrhagic conditions	8	0,437	6	1;3
E11 Type 2 diabetes mellitus	4	0,867	4	1;2
E79 Disorders of purine and pyrimidine metabolism	11	0,25	9	1
I10 Essential (primary) hypertension	12	0,25	12	2
I15 Secondary hypertension	n.p.		15	2
I25 Chronic ischemic heart disease	14	0,12	13	2
K74 Fibrosis and cirrhosis of liver	9	0,388	7	1
N17 Acute renal failure	2	1,87	2	1;2;3
N18 Chronic kidney disease (CKD)	**1**	1,87	**1**	1;2;3
N99 Intraoperative and postprocedural complications and disorders of genitourinary system, not elsewhere classified	7	0,5	n.p.	
Y43 Primarily systemic agents adverse effect	15	0,1	14	2
Z49 Encounter for care involving renal dialysis	**3**	1,042	**3**	1;2
Z51 Encounter for other aftercare	10	0,25	11	2
Z94 Transplanted organ and tissue status	**5**	0,68	**10**	2

Sources: 1: Previous codes patterns, 2: Probabilistic relationships, 3: Biological results, n.p. not proposed.

On the other hand, even though previous codes allow identifying a comparative smaller amount of proper codes, resulting in lower recall levels, most of those codes are relevant, producing therefore the highest precision results. In general, diagnosis coding results provided by both information fusion methods are somewhat equally distributed, between identical lists and either one of the methods being significantly better than the other. Codes obtained from biological laboratory examinations have nevertheless less influence on the final results, mainly because only a particular set of examinations was used, which is not widely representative of most known cases.

VI. CONCLUSION

Regardless of data and knowledge sources variety, which result in significant partial extracted information heterogeneity, information fusion for diagnosis coding support is suitable to generate enriched contextualized ordered lists of codes, to improve the physician coding task. The examined information fusion approaches handle as input partial lists of codes extracted by three complementary methods that process scattered data and information within the HIS and knowledge bases out of it. Obtained results show the superior recall performance of information fusion, as well as improved precision performance with respect to two information extraction methods. An experimental prototype is being tested under real clinical conditions, in order to analyze medical staff acceptance, as well as improve the proposed approach functionality and robustness.

Currently, information fusion has been examined using the outputs of only three information extraction modules, previously tested and characterized. A major challenge of this approach is then to operate at all levels of available information handling information imperfection, being user friendly at the same time. Other documents, like discharge letters could also be processed to extract complementary information, as it is being increasingly done in other projects (https://www.i2b2.org/pubs/index.html). Additionally, the significant dependency on external knowledge bases for some extraction methods like probabilistic relationships and previous codes patterns makes necessary to develop them further. Otherwise, sources credibility factors are prone to changes between groups of patients and over time. Learning methods should be expanded and further developed, even though case-based reasoning and natural language processing are not likely to be considered, as complementary methodologies. Finally, our approach could be applied to detect adverse medication effects, or monitoring of food or medicine consumption, related to epidemiological research.

REFERENCES

[1] P. Chopard. Evaluation de la qualité du codage dans les services hospitaliers des HUG. Journées Qualité des HUG 2001, http://qualite.hugge.ch/_library/pdf/journee_qualite_projets2001/11_PQ.pdf

[2] Organisation Mondiale de la Santé (OMS) (1998). Classification Internationale des Maladies : CIM-10. vol. 1, 2 et 3. In: Organisation Mondiale de la Santé (OMS) (ed). 1998.

[3] L. Lecornu, C. Le Guillou, G. Thillay, P.J. Garreau, H. Jantzem, J-M. Cauvin. C2i: A tool to gather medical indexed information. Proc. 9th International Conference on Information Technology and Applications in Biomedicine, 2009, DOI: 10.1109/ITAB.2009.5394351, 4 pp.

[4] L. Lecornu, C. Le Guillou, F. Le Saux, M. Hubert, J. Puentes, J-M. Cauvin. ANTEROCOD: Actuarial survival curves applied to medical coding support for chronic diseases, Proc. 32nd IEEE International Conference Engineering in Medicine and Biology Society, Buenos Aires, Argentine, 2010, pp. 1158-1161.

[5] L. Lecornu, G. Thillay, C. Le Guillou, P. Saliou, J. Puentes, H. Jantzem, P.J. Garreau, J-M. Cauvin. REFEROCOD: A probabilistic method to medical coding support. Proc. 31st IEEE International Conference Engineering in Medicine and Biology Society, Minneapolis, USA, 2009, pp. 3421-3424

[6] I. Bloch. Fusion d'informations en traitement du signal et des images, Lavoisier (eds), Hermes Science Publication, 2003.

[7] D. Dubois, H. Prade. Théorie de possibilité, théorie des probabilités et logiques multiple-évaluées : une clarification. Annales des Mathématiques et de l'Intelligence Artificielle, 2001, 32, pp. 35-66.

[8] M.H. Alsun, L. Lecornu, B. Solaiman, C. Le Guillou, J-M. Cauvin. Medical diagnosis by possibilistic classification reasoning, Proc. 13th Conference on Information Fusion, Edingburg, Scotland, 2010, 7 pp.