Case-based Interpretation of Best Medical Coding Practices — Application to Data Collection for Cancer Registries

Michael Schnell[1,2]

[1] Epidemiology and Public Health Research Unit, Department of Population Health, Luxembourg Institute of Health, 1A-B, rue Thomas Edison, L-1445 Strassen, Luxembourg, michael.schnell@lih.lu
[2] UL, CNRS, Inria, Loria, F-54000 Nancy, michael.schnell@loria.fr

1 Introduction

Currently, cancer remains one of the main causes of death. To assist in the fight against this disease, cancer registries are being used. A cancer registry is a systematic, continuous, exhaustive and non redundant collection of data about cancers diagnosed and/or treated in a country or region. The collected data is defined in international standards with common terminologies (e.g., the International Classification of Diseases, commonly abbreviated as ICD [7]) and associated to best medical coding practices. Unfortunately, these practices and standards are very complex, making it difficult for operators, i.e. medical staff in charge of collecting and coding the data, and coding experts to apply them efficiently and consistently.

The aim of this research is to tackle this complexity, by assisting both operators and coding experts in the interpretation of best medical coding practices.

For the Luxembourg National Cancer Registry (NCR), a ticketing system has been implemented for operators. When they encounter a difficult coding problem, they can ask questions through this system and coding experts provide individual answers. Interesting questions are later presented and discussed in regular training sessions for operators. Coding experts rely on their medical knowledge and their understanding of the coding standards and best medical coding practices to answer the questions. However, it is crucial for cancer registries to have a consistent coding of the data, meaning that two similar patients should be coded similarly. Thus, two similar coding questions should have similar answers. For that reason, the coding experts must also take into account previous questions to answer new questions.

Section 2 presents the research plan, followed by a review of the current progress in section 3. Finally, section 4 outlines the remaining work.

2 Research plan

The research plan contains the following main work packages: an analysis of the questions answering approach and design of a coding solving method and a question management tool, an implementation of the tool and an evaluation of the impact of the proposed tool.

The analysis and design package contains the following areas of work:

- a literature review for case-based reasoning and the combination with other methods, e.g., rule-based reasoning, belief merging or formal argumentation;
- a state of the art for coding assistants and decision support systems;
- a analysis of the questions asked by operators of the NCR and the answers provided by experts, as well as their solving process;
- functional design of a question management tool and technical choices for the implementation.

With the results of the analysis, a implementation of the designed software will be provided, starting with a prototype for a limited test usage before transitioning to a more general usage.

Lastly, an evaluation of the impact of the implemented assistant, notably to assess the evolution of the expert and operator workloads, is planned.

3 Current progress

The first step of this research is to analyze the coding process and the question solving process. Therefore, a dozen tricky questions and several easier questions have been discussed with the coding experts of the NCR. Given the similarities in the solving process of the experts and case-based reasoning, the proposed approach is adapted from the 4-R cycle presented in [1] and the knowledge containers presented in [6]. Other approaches have been considered (e.g., automatic coding [5]), but are not being pursued at the moment.

The approach proposed by this research uses arguments for the solving process. Indeed, when answering a case, experts often point out which arguments (pros and/or cons) have been identified and which solution they support. This is very helpful for operators, as it gives them insights into the reasoning process and allows them to learn more quickly. In order to incorporate this user guidance and solution explanation into our approach, the retrieval step uses the arguments of a source case to find the best match for the given target problem. The approach has been described in more detail in a paper submitted to ICCBR 2017. A prototype developed for the paper (shown in figures 1 and 2) focuses on structured questions asked by operators and the solutions provided by the system.

Arguments have already been used in CBR, but not to identify similar cases. For example, in [3] and [4] arguments are generated and used to explain the inferred solution of the target problem.

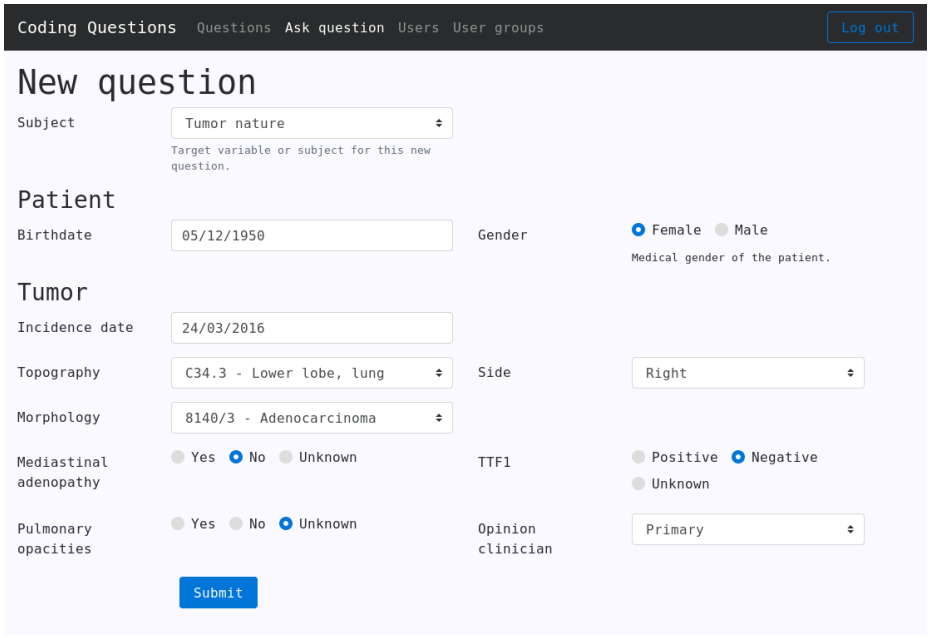

Fig. 1. Form used for the question asked by operators and to describe the target problem.

To facilitate the handling of the operators' questions, it was decided to structure the questions rather than apply natural language processing methods. Figure 1 shows an example of a structured question. Still, there are plenty of items that can be relevant for the various questions, making it very difficult to define every possible data item. Thus, only the most important subject are completely structured, i.e. the data asked of the operator is almost exhaustively defined. The assistant proposed by this research project will only handle structured questions. The remaining unstructured questions, i.e. questions where the operators describes his problem using free text, will continue to be answered by the coding experts, with little change to the current situation.

4 Future work

Once the prototype has been tested, a first version of the ticketing system for the coding questions will be developed. By the end of the year, this first version should be tested by the operators of the NCR. This version will be evaluated, notably to determine the impact of the system. Several types of criteria can be measured (e.g., quality of the coded data, workload of the coding experts). The final list of criteria will be determined during the implementation of the ticketing system. Alongside, other avenues are considered:

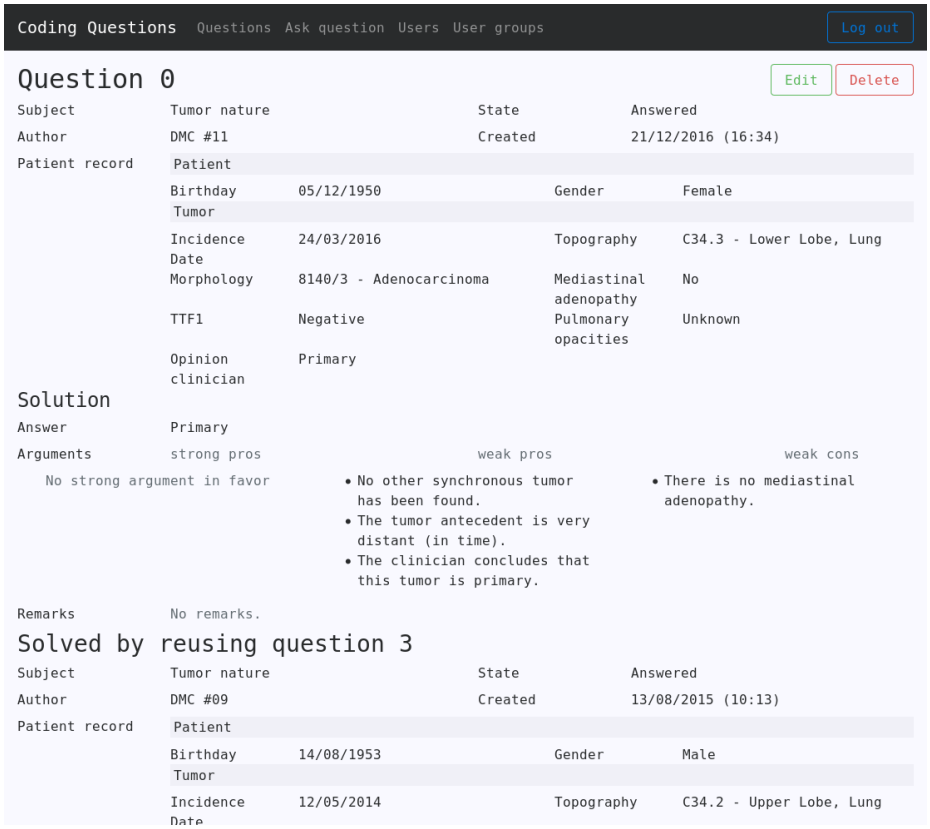

Fig. 2. Summary of the described target problem and the proposed solution. The retrieved source case is shown similarly to the target problem.

- increasing the efficiency of the solving algorithm (e.g., combining arguments using logic operators like *and*, reusing arguments from several source cases);
- adding a conversational approach where the system can ask for additional information about a given patient (to reduce the amount of information asked to only relevant data);

The following problems, though interesting, will probably not be researched in depth during this project:

- adding a confidence indication of the given result to help operators and coding experts distinguish between tentative solutions and validated solutions;
- taking into account the evolution of the coding standards [2].

References

1. Aamodt, A., Plaza, E.: Case-based reasoning: Foundational issues, methodological variations, and system approaches (1994)

2. Cardoso, S.D., Pruski, C., Da Silveira, M., Lin, Y.C., Groß, A., Rahm, E., Reynaud-Delaître, C.: Leveraging the Impact of Ontology Evolution on Semantic Annotations, pp. 68–82. Springer International Publishing, Cham (2016)
3. McSherry, D.: Explaining the Pros and Cons of Conclusions in CBR. In: European Conference on Case-Based Reasoning. pp. 317–330. Springer (2004)
4. Murdock, J.W., Aha, D.W., Breslow, L.A.: Assessing elaborated hypotheses: An interpretive case-based reasoning approach. In: International Conference on Case-Based Reasoning. pp. 332–346. Springer (2003)
5. Pestian, J.P., Brew, C., Matykiewicz, P., Hovermale, D.J., Johnson, N., Cohen, K.B., Duch, W.: A shared task involving multi-label classification of clinical free text. In: Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing. pp. 97–104. Association for Computational Linguistics, 1572411 (2007)
6. Richter, M.M., Weber, R.O.: Case-based reasoning: a textbook. Springer Science & Business Media (2013)
7. World Health Organisation: The ICD-10 classification of mental and behavioural disorders. Geneva: World Health Organisation (1992), `http://www.who.int/classifications/icd/en`