Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-231

Medical and Healthcare Curriculum Exploratory Analysis

Martin KOMENDA[a,1], Matěj KAROLYI[a], Andrea POKORNÁ[a,b], Christos VAITSIS[c]
[a] Institute of Biostatistics and Analyses, Faculty of Medicine, Masaryk University
[b] Department of Nursing, Faculty of Medicine, Masaryk University
[c] Department of Learning, Informatics Management and Ethics, Karolinska Institutet

Abstract. In the recent years, medical and healthcare higher education institutions compile their curricula in different ways in order to cover all necessary topics and sections that the students will need to go through to success in their future clinical practice. A medical and healthcare curriculum consists of many descriptive parameters, which define statements of what, when, and how students will learn in the course of their studies. For the purpose of understanding a complicated medical and healthcare curriculum structure, we have developed a web-oriented platform for curriculum management covering in detail formal metadata specifications in accordance with the approved pedagogical background, namely outcome-based approach. Our platform provides a rich database that can be used for innovative detailed educational data analysis. In this contribution we would like to present how we used a proven process model as a way of increasing accuracy in solving individual analytical tasks with the available data. Moreover, we introduce an innovative approach on how to explore a dataset in accordance with the selected methodology. The achieved results from the selected analytical issues are presented here in clear visual interpretations in an attempt to visually describe the entire medical and healthcare curriculum.

Keywords. Exploratory data analysis, medical and healthcare curriculum, data mining, outcome-based education.

1. Introduction

Medical and healthcare education (MHE) is a domain which constantly needs to be evaluated and accordingly reshaped while it tries to incorporate to the extent possible the growing body of medical evidence. Higher education institutions aim to ensure both transparency and effectiveness as an educational system but also to create health professionals able to cope with healthcare trends and demands [1, 2]. This is a significant multifaceted task with underlying challenges and with a considerable level of complexity consisting of a number of components which need to be properly instrumented to effectively address these challenges. Although there is significant progress in using different techniques and methods such as different analytics approaches to leverage data successfully in other sectors, there is effort to transfer these techniques in higher education [3] but more process is required particularly for the complex world of MHE [4]. In previous studies, we have demonstrated how such

[1] Corresponding author, Kamenice 126/3, Brno, 625 00, Czech Republic; E-mail: komenda@iba.muni.cz.

educational data can be leveraged with the use of Visual Analytics in different cases. First in an effort to represent a medical and healthcare curriculum (MHC) from the perspective of relations between competencies and learning outcomes addressed in an entire medical programme [5]. Then, we used the same method to go deeper to analyze and investigate a specific course and its alignment for the same programme [6]. Finally, we evaluated the possibilities of all these approaches to support decision making [7] in an outcome-based MHE context. We have also demonstrated in another case and from another perspective how different analytical approaches, data analysis methods, and data representations can be used to collect the dispersed and multi-structured educational data to create a blueprint and fully or partially model MHC that would allow us to understand and reason for its different components and how they fit together in the big picture [8].

One of the most important aspects in data analysis is to be able to ask questions and get meaningful answers from the data. To achieve this, a deep understanding of the examined MHC as an entity is equally needed in order to give us the expertise and intuition needed to exploit the data we possess [9]. We have identified the following analytical issues, which are focused on how to systematically explore medical and healthcare study programmes from the data point of view. Specifically, the aim of this study is to investigate: (i) How can we effectively apply data analysis methodologies and techniques to identify hidden relations between MHC's components? (ii) How can we visually represent the identified hidden components and measurements in order to provide insights on the used pedagogical approaches behind developing and delivering learning activities, and analytically describe the entire MHC as an objective to support decision making?

2. Methods

In this section, we introduce the selected methodological model for data mining as well as data analysis and visualization background. In order to systematically describe our study programmes including all forms of teaching and learning activities, we developed a specialized curriculum management system call OPTIMED [10]. OPTIMED is an original platform for optimizing and harmonizing MHC, while supporting the outcome-based approach to education. Moreover, it provides huge volume of well-organized data stored in entity-relation database (detailed formal metadata specification down to the level of medical sections, disciplines, courses, learning units, and interconnections to the learning outcomes).

Many different data mining methodologies have been developed and are well established in order to systematically approach and solve tasks similar to MHC data analysis, modeling and deeper understanding. We have adopted CRISP-DM [11], which is often used for medical and healthcare as well as higher education data mining tasks [13–15]. This model completely fits to the curriculum exploration, which among other things fully supports the following crucial steps: (i) The understanding of the curriculum innovation objectives from the academic perspective, which are defined by research questions. (ii) The understanding of initial MHC data arrangement including the data quality problems identification. (iii) The evaluation and refinement of mined results by medical experts in a manner more effective to higher education institutions.

The six-stage-sequence CRISP-DM process helps us identify, analyze, and visualize hidden relations between MHC components as following: (i) The first step is data extraction, where we accessed our PostgreSQL database to obtain raw data. (ii) The extracted dataset was then pre-processed, normalized and cleaned. Preprocessing phase is partially done by the database server and partially by server side application code written in PHP or R language. This produced a well-formed dataset with required data types and clear structure. (iii) At the next step we filtered the data to exclude entries irrelevant to the analysis purpose. What entries are irrelevant and therefore how the entire filtering approach is performed, is very much depended on the needs of the data reports recipients (e.g. curriculum designers, teachers), represented by the different visualizations. (iv) Following the data preparation we further analyzed the data with the use of available statistical software (SPSS, MS Excel, R). In this step, anomalies and discrepancies are frequently revealed in the dataset and if necessary the first three steps are repeated to refine the data and run the forth step again. This produces comprehensive static reports in a form of tables and graphs including legends and short data descriptions. (v) After evaluating the data reports, we automated the first three phases as described in previous subsection (phase iv) and thereafter this process enters a routine and executed daily. Thus, the generated reports are stored in a server where can be used in front-end modeling and visualization. (vi) The final step is modeling and visualization, where we produce two types of outputs: tables and graphs as static images, and web-based interactive graphs and tables with advanced filtering features. To plot the graphs, we used D3.js and NVD3 libraries and we followed the user-driven exploratory data analysis approach, which is based on additional filtering functionalities that enable the user to intuitively adjust visualizations and to explore the provided dataset by himself/herself.

3. Results

We created visual representations of the curriculum analysis, as a way to make the study programme more transparent and more easily understandable to students, teachers, faculty management, and potential employers. Graphs and tables show the connections and links between the various components of the curriculum. It enabled us to explore some general patterns within the study programme in relation to the promotion of generic skills. More than 25 various analytical reports have been prepared on the General medicine study programme. They show primarily basic overviews providing required summarizations in numbers. Below are two selected analytical reports from a pool of twenty-five analytical reports in total of explored hidden components and measurements in curriculum. We have integrated Bloom's educational activities in the curriculum conceptual data model including action verbs. Based on the presence of these meta-information in learning outcomes, we are able to categorize them in several levels and determine the requirements for students (Figure 1).

Figure 2 illustrates the ability to extract information about the annotation extent in a comprehensive manner. It is not humanly possible to go through the entire contents or to imagine how the sections, medical disciplines or courses are described by a set of plain text attributes, where a total length is not limited (e.g. meaning or annotation). We are able to determine their exact length in standard pages, compare individual curriculum components and their measurements, and provide the objective material to further discussion and support decision making.

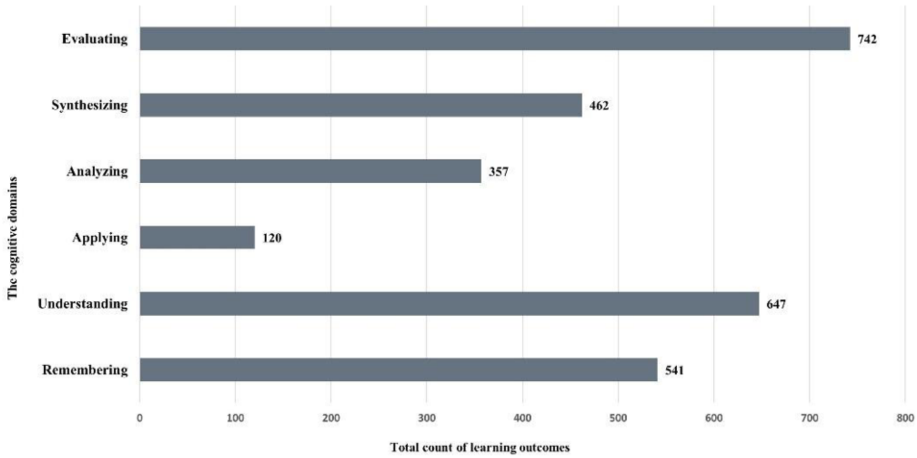

Figure 1. The content classification according to Bloom's taxonomy.

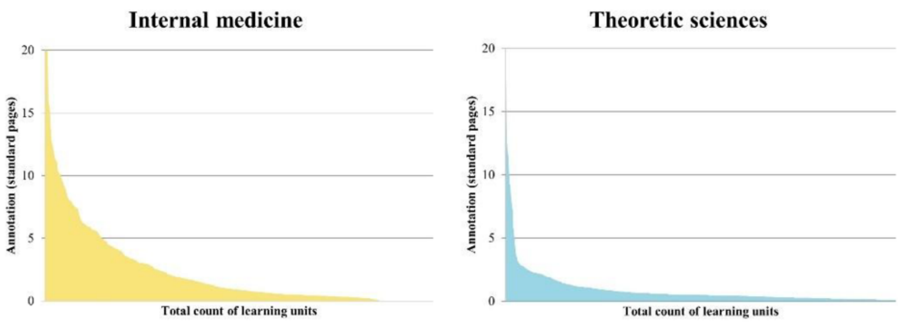

Figure 2. The annotation length of learning units according to medical sections.

4. Discussion

The exploratory analysis focused on curriculum evaluation should be always prepared very carefully in order to be of practical use for all stakeholders in the education process. We highlight with our results how analytical reports can be useful tools and how their usage in an appropriate way could contribute to the evaluation and improvement of the curriculum content in general. As we can see on Figure 1, there is wide range of Bloom´s taxonomy cognitive domains, which are covered in different learning units in analyzed curriculum. We could speak about success when evaluating the type of selected outcomes, because the ability to use Bloom's taxonomy by clinician-teachers was very low when we started to prepare the OPTIMED platform. Moreover, Figure 2 shows how we compare the annotation length or extent of learning units to different medical sections. Despite that the evaluation of the content is presumably not the main focus of this study, this information is very useful for guarantors of each medical section who could compare whether the learning units produced under their leading have the same basic components and to which amount.

We anticipate that our approach will support different stakeholders involved in MHE to build a deeper understanding of how MHE is currently conducted, use it for different education evaluation purposes and also to positively impact decisions concerning possible future reformations of MHE.

5. Conclusion

In general, the presented analytical reports show global overviews and identify whether the intended curriculum structure is actually being well-balanced. Moreover, we are able to demonstrate the overlapping or missing links among the different components of explored study programme. It helps mainly the staff by displaying these key elements of the curriculum and the relationships between them. Teachers and faculty management can be clear about their role in the big picture. Combining gained expert analytical knowledge with the CRISP-DM data mining process methodology can help reaching the core of a defined research questions. It can also advise the process of real data analysis and preparation, the features selection, the design and algorithms fine-tuning, and the evaluation and refinement of mined results in a manner more effective to higher education institutions.

References

[1] J. Frenk *et al.*, “Health professionals for a new century: transforming education to strengthen health systems in an interdependent world,” *The lancet*, vol. 376, no. 9756, pp. 1923–1958, 2010.

[2] N. Cho, C. Gilchrist, G. Costain, and N. Rosenblum, “Incorporating evidence-based medicine in the undergraduate medical curriculum: early exposure to a journal club may be a viable solution,” *Univ. Tor. Med. J.*, vol. 88, no. 3, pp. 154–155, 2011.

[3] C. Vaitsis, V. Hervatis, and N. Zary, “Introduction to Big Data in Education and Its Contribution to the Quality Improvement Processes,” 2016.

[4] S. Mennin, “Self-organisation, integration and curriculum in the complex world of medical education,” *Med. Educ.*, vol. 44, no. 1, pp. 20–30, 2010.

[5] C. Vaitsis, G. Nilsson, and N. Zary, “Big Data in Medical Informatics: Improving Education Through Visual Analytics,” *Stud. Health Technol. Inform.*, vol. 205, p. 1163, 2014.

[6] C. Vaitsis, G. Nilsson, and N. Zary, “Visual analytics in healthcare education: exploring novel ways to analyze and represent big data in undergraduate medical education,” *PeerJ*, vol. 2, p. e683, 2014.

[7] C. Vaitsis, G. Nilsson, and N. Zary, “Visual Analytics in Medical Education: Impacting Analytical Reasoning and Decision Making for Quality Improvement,” *Stud. Health Technol. Inform.*, vol. 210, no. Digital Healthcare Empowering Europeans, pp. 95–99, 2015.

[8] M. Komenda, D. Schwarz, J. Švancara, C. Vaitsis, N. Zary, and L. Dušek, “Practical use of medical terminology in curriculum mapping,” *Comput. Biol. Med.*, 2015.

[9] H. Cuesta, *Practical data analysis*. Packt Publishing Ltd, 2013.

[10] M. Komenda, D. Schwarz, C. Vaitsis, N. Zary, J. Štěrba, and L. Dušek, “OPTIMED Platform: Curriculum Harmonisation System for Medical and Healthcare Education,” *Stud. Health Technol. Inform.*, vol. 210, pp. 511–515, 2015.

[11] P. Chapman, J. Clinton, T. Khabaza, T. Reinartz, and R. Wirth, “The CRISP-DM process model,” *CRIP–DM Consort.*, vol. 310, 1999.

[12] N. Caetano, P. Cortez, and R. M. Laureano, “Using Data Mining for Prediction of Hospital Length of Stay: An Application of the CRISP-DM Methodology,” in *Enterprise Information Systems*, Springer, 2014, pp. 149–166.

[13] D. Asamoah and R. Sharda, “Adapting CRISP-DM Process for Social Network Analytics: Application to Healthcare,” 2015.

[14] C. Catley, K. Smith, C. McGregor, and M. Tracy, “Extending CRISP-DM to incorporate temporal data mining of multidimensional medical data streams: A neonatal intensive care unit case study,” in *22nd IEEE International Symposium on Computer-Based Medical Systems, 2009*, 2009, pp. 1–5.