CORDIET - Healthcare: Extracting knowledge from electronic patient record data

Jonas Poelmans, MSc, PhD[1,3], Herman Van der Mussele, RN, MSc[2], Gerda Verheyden, RN[2], Stijn Viaene, MSc, PhD[1], Guido Dedene, MSc, PhD[1], Sergei Kuznetsov, MSc, PhD[3], Alexei Neznanov, MSc, PhD[3]

[1]Katholieke Universiteit Leuven, Leuven, Vlaams-Brabant, Belgium; [2]GZA-hospitals, Antwerp, Belgium; [3]Higher School of Economics, Moscow, Russia

Abstract

In this paper we propose the software system CORDIET-Healthcare which we are currently developing in collaboration with the Katholieke Universiteit Leuven, Moscow Higher School of Economics and the GZA-hospital group located in Antwerp. The main aim of this system is to offer healthcare management staff a user-friendly and powerful data analysis environment. Using state of the art techniques from computer science and mathematics we show how CORDIET-Healthcare can be used to gain insight in existing care processes and reveal actionable knowledge which can be used to improve the current way of working.

Introduction

Since the introduction of computer systems in healthcare organizations the amount and complexity of information available in databases became too large to analyze manually. To really know what is happening on the operational working floor simple excel graphs and other traditional statistical data analysis methods used by healthcare management staff often turn out to be insufficient. To cope with these challenges we started in 2010 with the development of the "COncept Relation Discovery and Innovation Enabling Technology"- software system so called CORDIET-Healthcare) . The main idea behind the project was to use various kinds of visual models to represent the data available in databases. Amongst others, Hidden Markov Models were chosen to show the activities performed over time to certain patient groups and concept lattice diagrams[1] turned out to be an ideal instrument the quality of care delivered to patients. The main goal of this paper is not to describe in detail these various methods used, we aim at making the user familiar with the potential of our software by focusing on their application to healthcare data.

CORDIET-Healthcare software architecture

The hospital electronic patient record system (the patient care system – PCS) contains all information relevant for supporting clinical pathways in our organisation. Clinical pathways are a methodology for organizing multidisciplinary care (Campbell 1998, Vanhaecht et al. 2010). Each pathway has received a unique identifier and specific information can be stored for each of these paths. Structured forms are used for data collection, for registration of interventions, critical indicators and outcome indicators. The database is developed as a level three electronic patient record (EPR) system as Brennan (2001) described. It contains: "clinical, ward - based applications including order communications and results reporting, multidisciplinary clinical pathways, electronic prescribing and drug administration". At this moment 276 nursing care plans and 72 clinical pathways, are integrated in PCS. Performing statistic analysis on this database is made possible by an Excel and Access compatible tool. The latter is used to retrieve the required information from the database into a spreadsheet environment. The technical specifications of the underlying hardware are as follows: a VM/VSE operating system on an IBM mainframe working under CICS. The technical specifications of the database are as follows: DLI-database completed with several VSAM databases making use of a computer language known, as ADS (Application Development system). The pivotal role of the PCS system is to support an efficient care planning process by amongst others effectively organizing the communication between the different care providers, the ancillary services,… and clinical management systems (such as the nursing Minimal Dataset and the hospital discharge dataset).

The CORDIET-Healthcare software makes use of a three-layered and client-server based architecture. The data access layer was designed to work with the freely available PostgreSQL database, data input files are formatted in a proprietary XML format, however in the future the software can easily be expanded to connect with other data sources. In the middle layer we have the data analysis components including Formal Concept Analysis, Hidden Markov Models, Emergent Self Organizing Maps, simple clustering methods, etc. The user can dynamically compose the visual models he will use to gain new insights by selecting subgroups of patients with specific properties, introducing and selecting relevant attributes, etc. The presentation layer contains a manual and graph-based editor for the attribute space and multiple facilities have been introduced to make the data browsing experience using the methods implemented in the middle layer as ergonomically as possible. The prototype of the software system which was developed in 2010 consists of interconnected modules which were implemented using topic map technology, matlab and Java. The software system is expected to be finalised in November 2011 and will offer a better user interface and a more solid and expandable software design.

In Figure 1 an intuitive visual description of the knowledge acquisition process using CORDIET-Healthcare is given. Shortly described, starting from healthcare databases containing various kinds of data including medical reports, nursing activities performed to patients, financial information, etc. we extract a subset of patients which is of particular interest and define attributes to analyze them such as "length of stay" or "mastectomy surgery". We then create visual data analysis models such as a concept lattice diagram which may reveal quality of care issues such as a mastectomy patient who did not receive a prosthesis before discharge. We may be interested in the care trajectory through which such a patient went and a process model depicting the nursing activities performed to such a patient over time may give us important insights in the root causes of such an inefficiency. Another interesting research track consists of identifying properties of certain patient groups which may help us predict the progress of the patients' condition so that we are able to support early discharge decisions, provide all the necessary discharge documents, etc. For example, we are currently developing predictive models whether a stroke patient will go home or go to a revalidation centre, based on the medical records available for about 400 patients a year.
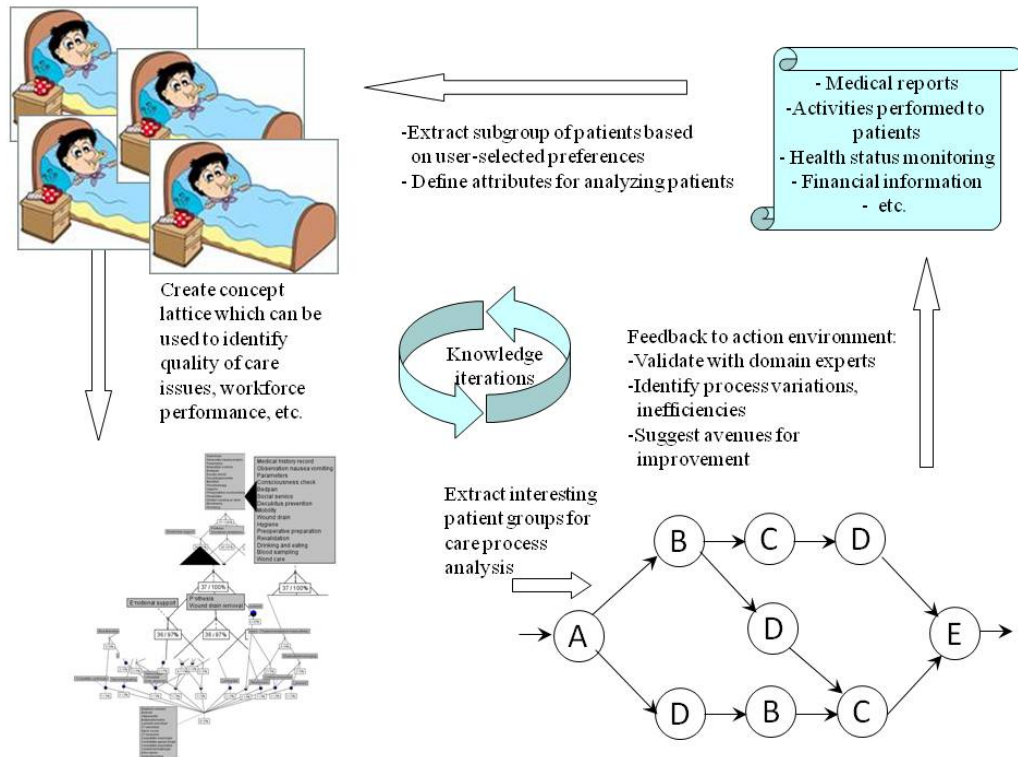

Figure 1. Knowledge discovery process with CORDIET-Healthcare software

A case study on breast cancer care

The CORDIET-Healthcare software was tested on the PCS-database of the GZA-hospital Antwerp. The data of 148 patients who were treated for primary operable breast cancer were analyzed. 60 of these patients received breast conserving surgery with lymph node removal, 37 of them received mastectomy with lymph node removal. We grouped the 469 types of activities performed to these patients in about 60 clusters and we created a concept lattice diagram for these two patient groups and used these 60 activity type clusters as attributes (see Poelmans et al. 2010 for details on Formal Concept Analysis).
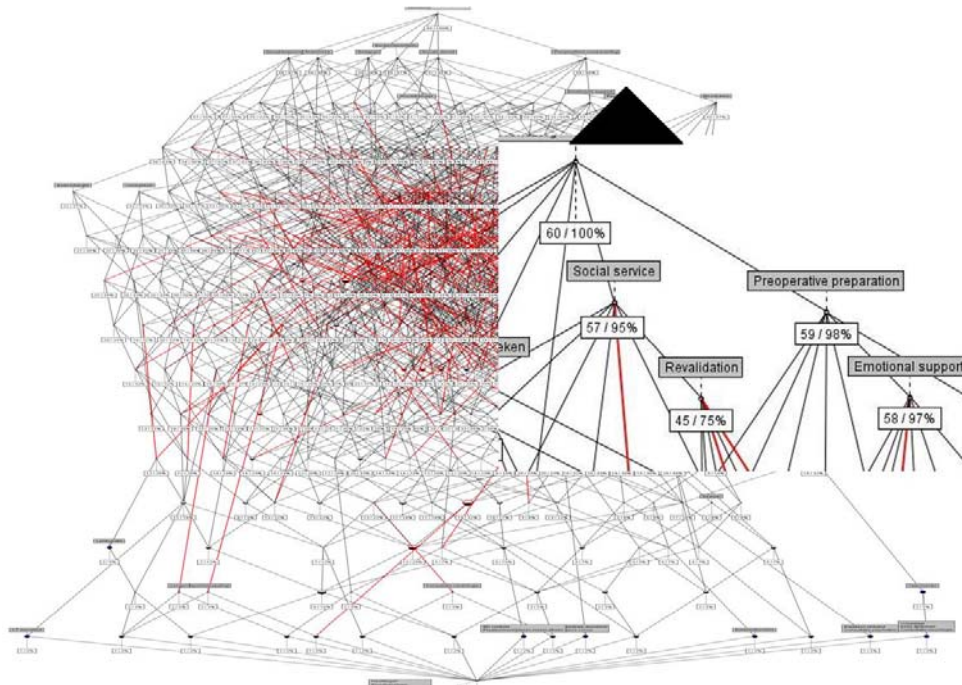

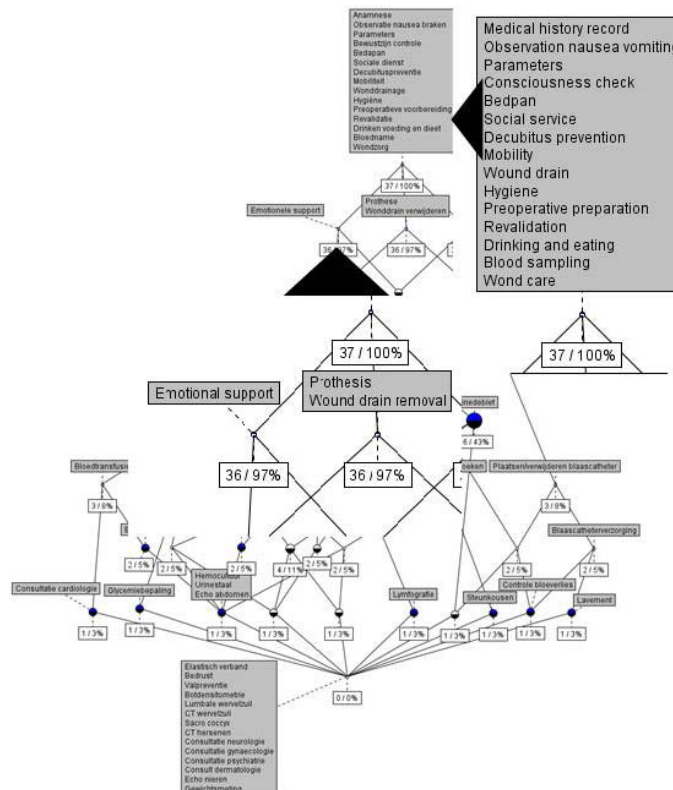**Figure 2.** Concept lattice diagram for breast conserving therapy with lymph node removal patients

Figure 3. Concept lattice diagram for mastectomy with lymph node removal patients

NI2012: 11th International Congress on Nursing Informatics

It is interesting to observe that the lattice structure of the breast conserving surgery with lymph node removal patients in Figure 2 is much more complex than the lattice structure of the mastectomy with lymph node removal patients in Figure 3. By analyzing both diagrams we found that this was caused by some key interventions which were not performed to patients such as revalidation (25% of patients did not receive physiotherapy) and emotional support (3% of patients). Discussion with nursing staff revealed that the original care pathway was written for an average length of stay of 5 days whereas this length of stay was decreased over the years to 3 days for patients who received breast conserving therapy. As a consequence it became impossible to execute the prescribed pathway in practice and currently modifications to cope with this problem are being implemented .

To analyze process variations within the group of breast conserving therapy with lymph node removal patients we subdivided them according to their length of stay. The process models we distilled from the data depict the activities performed to patients (the ovals, also called states, in the process diagrams) and the arrows describe the probability that one state is followed by another. Figure 4 contains the process model for the patients with a length of stay less than 4 days. We can see that this model is more complex than the model for patients with length of stay of 4 days displayed in Figure 5. This is because patients with a lower length of stay are receiving suboptimal care in which key interventions are missing. The model in Figure 4 contains suboptimal care process variations.
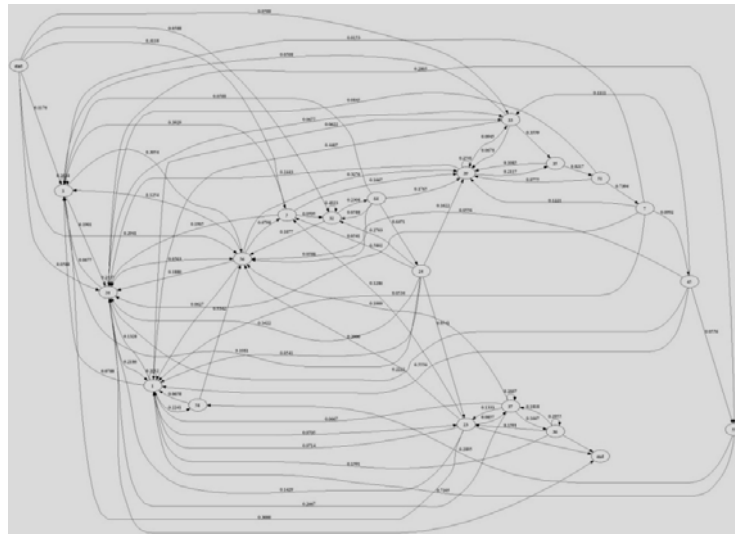

Figure 4. Process model for patients receiving breast conserving surgery with lymph node removal and LOS < 4 days
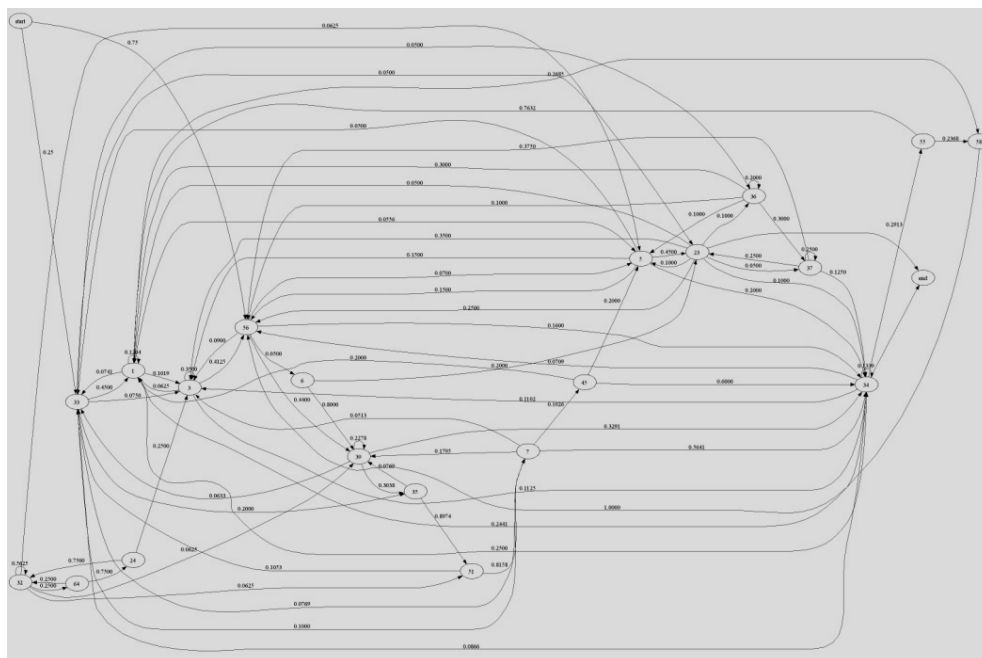

Figure 5. Process model for patients receiving breast conserving surgery with lymph node removal and LOS = 4 days

Figure 5 depicts the process model for patients with an average length of stay and mostly resembles the care pathway which was originally designed. The model in Figure 6 is also more complex than this average LOS process model. This was caused by patients with co morbidities and a sometimes suboptimal alignment between different units.
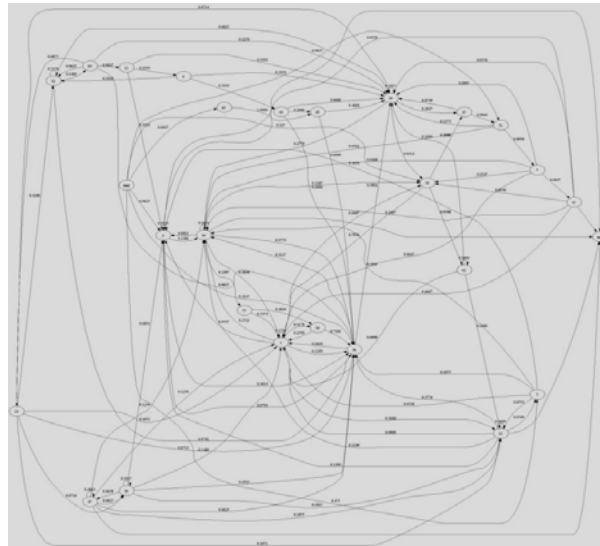
NI2012: 11th International Congress on Nursing Informatics

Figure 6. Process model for patients receiving breast conserving surgery with lymph node removal and LOS > 4 days

Conclusion

CORDIET-Healthcare is a promising software system for healthcare organizations who want to gain insight in what is happening at the operational working floor. It allows management staff to go much deeper into their data and gain truly valuable knowledge for improving their care processes. The software is essentially human centered and was designed to actively engage expert domain knowledge to guide the data exploration process. We showcased our approach with a small case study and we plan in the near future to expand this research on much larger data repositories.

References

1. Poelmans, J., Elzinga, P., Viaene, S., Dedene, G. (2010), Formal Concept Analysis in knowledge discovery: a survey. Lecture Notes in Computer Science, 6208, 139-153, 18th international conference on conceptual structures (ICCS): from information to intelligence. 26 - 30 July, Kuching, Sarawak, Malaysia. Springer.
2. Poelmans, J., Dedene, G., Verheyden, G., Van der Mussele, H., Viaene, S., Peters, E. (2010). Combining business process and data discovery techniques for analyzing and improving integrated care pathways. Lecture Notes in Computer Science, Advances in Data Mining. Applications and Theoretical Aspects, 10th Industrial Conference (ICDM), Leipzig, Germany, July 12-14, 2010. Springer
3. Campbell, H., Hotchkiss, R., Bradshaw, N., Porteous, M. (1998) Integrated care pathways. British Medical Journal 316, 133-137.
4. Brennan S. Medical illustration and the electronic patient record. J Audiovis Media Med 2000; 23(4): 171–2.
5. Vanhaecht, K., Sermeus, W., Van Zelm, R., Panella, M. (2010). Care Pathways are defined as complex interventions. BMC, 8, 31.