Process mining challenges in hospital information systems

Payam Homayounfar
Wrocław University of Economics,
ul. Komandorska 118/120
53-345 Wrocław, Poland
Email: p.homayounfar@gmail.com

Abstract—**Over the last years hospital information systems became more integrated. Hospital information systems with the wide variety of systems are highly suitable to use process mining for knowledge discovery and optimization of processes. Applying process mining in hospital information systems is a modern and recommendable approach in health care. But process mining techniques can only provide a high result quality if the structure of data is known and if the structure of the event logs are maintained appropriately. This paper describes process mining, hospital information systems and shows where the challenges are if the two areas are combined.**

I. INTRODUCTION

PROCESSES become in practice more and more a critical success factor in organizations. In health care organizations are the processes implemented in the hospital information systems. Process mining is a young discipline that is known since less than ten years. It has been evolved from business process discovery, workflow analysis and data mining. The goal is to analyze automatically event logs of systems to discover optimization potential. The process model can be developed within process mining or can be verified and analyzed if it exists a priory. Process mining is a process management technique that allows the analysis of business processes based on event logs. The basic idea is to extract knowledge from event logs recorded by an information system [1]. Process mining aims at improving this by providing techniques and tools for discovering knowledge about process, control, data, organization, and social structures from event logs [2], [3].

Many process mining algorithms were developed in the last years derived from data mining techniques [3]-[6]. Process mining enables an organization to discover the reality about existing processes.

Data Mining extracts patterns, explicit knowledge and information from data. The objective of data mining in this context is to support the medical doctor and the health care institution in decision making.

Like process mining, data mining analyzes data and extracts models that allow the interpretation and transformation of the raw data in the hospital information system into knowledge [7], [8]. This is the entry point for knowledge management to create tacit knowledge. Knowledge management is the system and managerial approach to the gathering, management, use, analysis, sharing, and discovery of knowledge. Data analysis in medicine depends more than in other areas on medical background knowledge. Further approaches, such as association and classification rules, joining the declarative nature of rules, and the availability of learning mechanisms are a great potential for effectively merging process, data, and knowledge mining.

This potential can only be used if the data quality in the event log of each system component of a hospital information system is sufficient. Simplified it can be described as the trustworthiness and quality of the result of process mining depends on the quality and consistency of the data in the event logs. Only with a clear understanding of the processes and the health care domain it is possible to use the process mining techniques in an appropriate way.

The topic has a high importance for hospitals as the economical pressure is constantly rising and competition is getting more difficult [9]. The application of process mining in hospital information systems is not related to high costs and the results can be developed in a relatively short timeframe. This makes it very attractive to use process mining in health care. Different case studies have proven the importance of process mining in health care [6], [10], [11].

Before describing the health care relevant challenges for process mining in hospital information systems it is important to describe the process mining fundamentals with examples from health care. It is also important to understand the general concept and specify of hospital information systems. This will lead to the challenges of process mining and finally to the conclusion.

The author's contribution is the analysis and commutation of currently existing literature and his experience in the field of computer science, systems design, development of software implementation methodologies and hospital information systems to show which challenges have to be faced when applying process mining in hospital information systems. Not in scope of this paper are the referred but not described process mining algorithms. Process mining needs automated tools for analyzing the event logs. Furthermore is out of scope of this paper the description of the tools and the software frameworks that are necessary for process mining.

II. PROCESS MINING FUNDAMENTALS FOR HEALTH CARE

Process mining is based on the usage of data mining technologies for processes. The extraction of patterns and explicit knowledge is the aim of data mining. Particularly in the complex and information-overflowing environment of health care process mining is able to support the medical doctor and the health care institution in decision making. Data mining analyzes data and extracts models that allow the interpretation and transformation of the raw data in the hospital information system into knowledge as foundation of decision making. Process mining uses the technique of data mining for extracting data from existing databases to understand the reality of processes. This knowledge helps to improve and to optimize processes.

The prerequisite for process mining is the existence of business processes mirrored in running systems. These systems again have to provide structured event log data. There is no limitation for the amount of connected systems or amount of event log files. This fact makes process mining especially usable in hospital information systems as the hospital information system is a commutation of many connected systems.

Process mining extracts information from event logs and delivers automatically the description of current processes by visualizing the process flow out of real data. The structures of event log files have to provide all relevant information for transactions. According to Weijters et al. 2006, each event has to refer to an activity (for example a detailed defined step in the process). Each event has to refer to a specific case like a process instance. Each event can have a performer also referred to the originator like the executing or initiating person of the activity. Events have a time stamp and have a specific order.

Some event logs contain more information on the case itself. Data elements in a hospital information system referring to properties of the case would contain also information about age, sex, diagnosis, etc. of a patient. Event logs of the underlying systems in a hospital information system contain transaction information with a low degree of abstraction. To enable process mining it is necessary to preprocess the data to obtain interpretable results [1].

After preprocessing the transaction data of event logs the process mining can be performed with different available algorithms on the event logs that can be assigned to different perspectives. Process mining can have different types of results and is not reduced only to the discovery of process models for optimizing inefficient interfaces or to unreveal shortcoming in a process.

Three different classes of process mining techniques can be differentiated – discovery, conformance analysis and extension. In the class of discovery there is no apriori model defined. The process model is based on the event logs. The process model can be discovered on low level events. The class of conformance analysis exists an apriori process model. The model will be compared and checked against data from the event log and discrepancies are made visible. Conformance checking may be used to detect deviations to

enrich the model with e.g. performance data. Also the class of extension has an apriori model. Based on the event logs the model will be extended with a new aspect or perspective e.g. to show where bottlenecks are arising.

More of interest than the classes of process mining techniques for healthcare and hospital information systems are the perspectives. According to Weijters et al. there are three different perspectives of process mining techniques: control flow perspective, organizational perspective, and performance perspective.

Algorithms assigned to the control flow perspective are control flow mining algorithm, heuristicminer algorithm, region mining algorithm. alpha algorithm etc. [1], [2], [4], [5]. The control flow perspective generates a process model that reflects the current observable process in reality. This enables the discovery of the process model by inferring the ordering relations between various tasks in the event log. Processes of the health care domain in hospital information systems have a lot of variants based on patients and diseases the result of process mining has a high complexity [12]. To reduce the complexity it is a common practice to break down an event log into sub event logs as long as the complexity it not simple enough to be able to analyze the data. Here are also different clustering techniques available like self clustering map technique or trace clustering [1]. The result is to have many clusters with a handelable size.

The organizational perspective is covered by process mining techniques for answering organizational questions like organizational process mining, social network mining, staff assignment analysis [1], [11]. These techniques can be used to determine the collaboration between departments in the hospital based on the data in the hospital information systems. Here it becomes visible how the handover of work happens between departments with detailed information about the frequency of used interfaces. A common result is the visibility of high involvement of central clinical departments like the laboratory and the radiology department. Optimization needs become obvious with the organizational perspective.

Also the performance perspective clusters different techniques with supporting time based options to identify the case duration of individual cases. The first and last events or the average cases duration in a process are used as performance metric.

III. GENERAL CONCEPT AND SPECIFITY OF HOSPITAL INFORMATION SYSTEMS

In the operation of a hospital a large number of different people including doctors, nurses, administrators or technicians work in various different areas like laboratory, radiology, ambulance, and administration. The different areas and people constantly exchange information. Therefore, in a hospital, as it is also the case in other companies, information processing within existing processes takes a very high priority [13].

A hospital information system is the entire information processing and information storage subsystem of a hospital, whereby it is not just about computer systems and networks

and the computer-based application systems that are installed on them, but it is about the information in a hospital as a whole [14]. To address this complex problem, there are a large number of interconnected and communicating subsystems. All of those systems are potential candidates for process mining, as the generate event logs for the case based transactions.

These can be grouped according to function in two groups, the clinical information system with all of the medical functionality and the hospital management system for managing the administrative, managerial and technical supply functions. Belonging to the clinical information systems are, depending on how they are viewed, the department systems such as the radiology information system for radiology. Due to the complexity and partly redundant functionality of these systems, such as patient management, a separate consideration is appropriate.

The use of electronic or computerized patient records as part of hospital information systems is also seen as highly significant. In scientific literature, the computerized patient record is counted by Hannan as belonging to clinical information systems. However, it seems useful to define a new group for the computerized patient record and other systems that are used by both clinical and administrative systems.

Therefore, the central systems which include global functions, which both the clinical information system and the hospital management access, are to be considered as a third group. The computerized patient record system is an important part of central system. It is a digital collection of documents and information to map and document the medical care of patients.

The hospital information system is an interface between the applications in the hospital and external applications of third parties outside the hospital, such as other hospitals or health insurance companies. In particular, the support of clinical applications and procedures in the hospital is important for a hospital information system. This way a hospital information system has a central and strategic importance for a hospital or a practice.

The main tasks of the hospital information system include the timely and context related provision of current patient data for an authorized group of people. The important thing is the appropriate form of the presentation of the information. Medical knowledge about diseases and drugs is also made available by a hospital information system. Information on the quality of patient care and the relationship between costs and benefits for hospital services are further responsibilities of a hospital information system. Further descriptions of the duties of a hospital information system and the new impetus on useable applications can be found in numerous scientific publications.

The optimization of business processes in hospital information systems is the key success factor for the integration of digital data into hospital facilities and equipment or interdepartmental treatment processes. This is primarily about uniting the hospital information systems areas of patient management, operations management,

financial management, computing, as well as testing and certification.

The most important tasks in hospital information systems can be summarized as follows:

- Storage and monitoring of patient's condition:
  - Accurate and electronically stored medical records of patients (e.g. drug allergies) are provided
  - Visual and auditory warning systems are generated in the event of abnormal test results or other important data
  - Time intervals and / or testing periods for tests on patients to be specified
  - Data Processing and analysis for statistical purposes and research oriented purposes
- Management and Data Flow:
  - Support automated patient data transfers between departments and institutions
  - Enable graphic or digitized diagnostic images from the hospital database based on the integrated retrieval system
  - Digital signatures, in order to create internal orders electronically
  - Communication by Laboratory Information System
  - Registration of human resources and their properties
- Financial Aspects:
  - Efficient administration of finances
  - Use and monitoring of medicines and effectivity of the ordering process
  - Expected and actual treatment costs are listed and reported
  - Automated representation of the needs of the nursing staff
  - Status analysis of bed occupancy and overall performance in the hospital information system

The relevance, completeness and timeliness of the information set quality criteria for the evaluation of the hospital information system. At the same time it applies that all patient information and that of the administrative processes for all activities are to be available in the hospital information system. Labor requirements, findings, tools to support the therapeutic and diagnostic process and the medical histories of patients are included, as well as search functions, or the documentation or the creation of letters. The linking of systems through interfaces and the integration of the subsystems are considered strategic and, at the same time, critical success factors for a hospital information system.

IV. HEALTH CARE RELAVANT CHALLENGES FOR PROCESS MINING IN HOSPITAL INFORMATION SYSTEMS

This chapter of the paper describes the diverse challenges of process mining in the health care domain by using event logs from hospital information systems. The nature of the medical domain is more complex and not comparable to other economical business organizations. The reason is based on the different types of patients, doctors, nurses, administrators and technicians involved in the diverse stakeholder process. This causes the high complexity that

also becomes obvious by definition of hospital information systems.

Especially in health care the multidisciplinarity causes problems based on the specify of the health care domain. Ad hoc actions and process changes, proper data collection, data redundancy, as well as incorrect and insufficient logging are the most important challenges of process mining in hospital information systems [1], [2], [6], [12], [15].

I. Complexity caused by heterogeneity

Beside of the already mentioned different involved parties in healthcare, the complexity of health care processes is based on heterogeneity of the diseases and the related treatments of patients [6], [12]. Furthermore are many different ways to treat different patients with same diseases. In companies of other branches a process for a specific task will be applied always in the same way, but in health care it can be different each time under consideration of the circumstances. For applying process mining techniques it becomes a challenge as processes are not easy to generalize and to find common or similar patterns in the processes. A treatment process in a hospital information system can first appear not complex, but the complexity can evolve dynamically. A patient can be registered in the system with one disease, but it can be discovered within the process that there are more than one different diseases that cause some symptoms of an other disease. Also here it is not easy to design the process information to generate similarities and to analyze the event logs.

Hospital information systems involve clinical and administrative processes with large volumes of data and a large number of people, patients and personnel. Additionally financial processes are involved in the processes. This relation to other not directly health related areas like governance, management, IT and finance rise the degree of complexity.

II. Multidisciplinarity

Hospitals departments are organized in silos and are highly specialized in their areas. On the other hand they are interlinked and have to work together across their fields and departments. A medical doctor treating a patient surely needs the involvement of the laboratory and the radiology in case of an surgical injury of a patient. Health care processes are multi-disciplinary in their origins and cause a higher complexity. Furthermore, these interfaces offer an opportunity for utilizing synergies and can be optimized by applying process mining with the right level of questions.

One of the biggest challenges of process mining in hospital information systems is to understand what the logged information means. The semantics of existing event logs can become difficult in complex hospital information systems. It depends on how technically or business driven the logs are. Before process mining can be performed it is necessary to understand the meaning of the fields and the content of event logs. Due to the multidisciplinary structure in health care and the many involved people in using hospital information systems it is very hard to align meaning of words as field description or in the content of the event logs.

III. Ad hoc actions and process changes

Processes in dynamic environments are not staying as they are defined and have an own life cycle. Ad hoc changes are those changes in a complex and dynamic process environment that occur at the time of execution. They are per definition not known and not implemented in a process model at the time of definition of the model [16]. Those changes deviate at the individual instance of the process from the process model. If the changes would be known it would lead to a variant of a process model and could be integrated in the basic set up of process mining. Variants are based on specific transformation rules.

Ad hoc changes occur often in hospital information systems with the tremendous amount of data that are always changing. This can be explained by the high turnover rate of knowledge in health care, but also with the continuously changing corresponding medical procedures. Also the administrative procedures, medical processes or technological developments in health care are changing. This causes mismatching data if event logs contain information about one process that has changed. Ad hoc actions and process changes may confuse process mining algorithms if the event logs are not clustered by the timestamp before and after the change.

It is a challenge for process mining in hospital information systems to build clear models and to automate the fast changing processes.

IV. Data collection

The event log data are the basis for process mining. Therefore it is obvious that a proper structure and content of the data are fundamentally relevant for applying process mining techniques. Hospital information systems contain and provide a wide variety of transactional based event logs. The practice in health care organizations is that also with the availability of e.g. the computerized patient record there are many process steps that are done manually and are not stored in the hospital information system, if they are cost neutral or if the cost causing activity has been documented in the event log.

Decision and transaction are performed by the previously described different human people involved in the health care processes. Automation of events and information flow enabled by more integrated hospital information systems supports all involved persons in their daily work. Automation has to be implemented and designed in the early development and implementation phases of hospital information systems. With the increasing automation, developments and integration initiatives the data quality and availability is continuously increasing.

V. Data redundancy

As described in advance, hospital information systems are combining many other systems with a lot of interfaces to internal and external systems. These systems support the users in different departments with a wide variety of

functions. In many cases data are held redundantly in the different systems. This redundancy can be implemented intentionally on purpose to enhance the technical stability, provide performance, or fulfill legal requirements. Not intended or unknown redundancy becomes a challenge in process mining as the result of the mining techniques. If redundancy of data is not known a further step in the preprocessing of the data is necessary to find redundant data. Afterwards the results of process mining are becoming more reliable as the basis of the data has a higher quality. This enables process mining to find the right conclusions.

VI. Incorrect or insufficient logging

Noisy data in the meaning of process mining are data which do not fit to comparable data as they provide exceptional values and an exceptional behavior. Those noisy data can be based on incorrect logging and deliver wrong information about the transactions. This can happen when the hospital information system has a high degree of manually documented process steps. Manually created data are generally less reliable than automatically performed data. In case of any doubts about the trustworthiness of the data, it is necessary to verify the data quality first before starting the process mining and to analyze the event logs.

Incorrect logging with noisy data describes the fact of wrong data. Insufficient logging is about missing data. To be able to describe the process instances based on the data from the event log, it is necessary to store at least the information about the case ID, the activity name or description, and the timestamp for each transaction or event.

Incorrect data can be found by specific preprocessing data, but insufficient data are not easy to find for adapting the process mining techniques and to build exceptions for areas with insufficient data. Missing data are often caused by the huge amount of data in hospital information systems and the need to create free spaces on the hard drives. Therefore some data are only hold for a period of time and will be deleted after the expiration data for the log file entry. The way of logging can also cause insufficient logging, depending on synchronous or asynchronous logging. Parts of hospital information systems have a synchronous logging where log files are created in real time. Other administrative parts of hospital information systems may maintain the event logs asynchronously in a batch mode, where data are logged once a day. The asynchronous logging has the danger of loosing important information between the batch runs as it is possible that only the delta between two days will be documented and not changes within a day. For using data mining techniques, it is not necessary to have access to the whole history data of a process. In difference to data mining it is necessary for process mining to have a view on the whole history of the process to provide better results.

V. Conclusion

Business processes in health care are complex, multidisciplinary, fast changing ad hoc and not easy to understand and structurable. This paper has described the different challenges that arise particularly if process mining

is used to discover knowledge in hospital information systems due to the characteristics of the medical domain.

Process mining delivers a true picture of the current processes in a hospital based on log files from different systems of a hospital information system. This transparency allows particularly hospitals to continuously monitor and improve their processes. It is apparent that the business processes in the medical domain are dynamic, ad hoc, unstructured and multidisciplinary in nature. Also in modern times, health care organizations place strong emphasis on medical and organization efficiency and effectiveness to control their health care expenditures. These process models as mentioned earlier can be analyzed to gain an insight into the reality. Process mining techniques help to understand what is actually going on in reality and if it is what is actually desired.

On the other hand it became obvious how important process mining in the health care domain can be. Most of the challenges can be handled better after process mining techniques have been applied to hospital information systems. After knowledge discovery it is possible to optimize the processes on different levels and with different perspectives. The next process mining cycle can then show the usage by demonstrating that the previously discovered findings have been improved.

In times of growing economical pressure on health care organizations, process mining in health care information systems has a great potential and will show fast progress in the near future due to many research activities in this field.

REFERENCES

[1] Weijters AJMM, van der Aalst WMP, Alves de Medeiros AK (2006) Process Mining with the HeuristicsMiner Algorithm. BETA Working Paper Series, WP 166, Eindhoven University of Technology, Eindhoven.

[2] van der Aalst (2012) http://www.processmining.org/research/start (06.04.2012)

[3] van der Aalst WMP, van Dongen BF, Herbst J, Maruster L, Schimm G, Weijters AJMM (2003) *Workflow Mining: A Survey of Issues and Approaches*. Data and Knowledge Engineering. 47(2): pp.237–267.

[4] van Dongen BF, Alves de Medeiros AK, Verbeek HMW, Weijters AJMM, van der Aalst WMP (2005) The ProM framework: A New Era in Process Mining Tool Support. In: Ciardo G, Darondeau P. [Eds.]: Application and Theory of Petri Nets 2005, volume 3536 of Lecture Notes in Computer Science, Springer Berlin, pp.444–454.

[5] van der Aalst WMP, Weijters AJMM, Maruster L (2004) Workflow Mining: Discovering Process Models from Event Logs. IEEE Transactions on Knowledge and Data Engineering. 16(9): pp.1128–1142.

[6] Mans RS, Schonenberg MH, Song M, van der Aalst WMP, Bakker PJM (2008) Application of Process Mining in Healthcare - A Case Study in a Dutch Hospital. In: Fred A, Fifipe J, Gamboa H [Eds.]: Biomedical Engineering Systems and Technologies, Communications in Computer and Information Science, Volume 25. Springer Berlin Heidelberg, pp.425-428.

[7] Homayounfar P, Owoc M.L. (2011a) Data Mining Research Trends in Computerized Patient Records. Publication in Preparation for FedCSIS in Szczecin: 1st International Workshop on Interoperable Healthcare Systems (HIS'2011).

[8] Homayounfar P, Owoc M.L. (2011b) Data Mining Methods as an Essential Part of Hospital Information Systems. Publication in Preparation for 2011 Conference in Katowice Knowledge Technologies in Public Management.

[9] Anyanwu K, Sheth A, Cardoso J, Miller J, Kochut K (2003). Healthcare Enterprise Process Development and Integration. Journal of Research and Practice in Information Technology, 35(2): pp.83–98.

[10] McGregor C, Catley C, James A (2010). A Process Mining Driven Framework for Clinical Guideline Improvement in Critical Care). In Learning from Medical Data Streams 13th Conference on Artificial Intelligence in Medicine (LEMEDS). http://ceur-ws.org/Vol-765/ (04.04.2012)

[11] Günther CW, Rinderle-Ma S.,Reichert M, van der Aalst WMP, Recker J (2008) Using process mining to learn from process changes in evolutionary systems. International Journal of Business Process Integration and Management, Special Issue on Business Process Flexibility, 3(1): pp.61–78.

[12] Lenz R, Elstner T, Siegele H, Kuhn K (2002). A Practical Approach to Process Support in Health Information Systems. Journal of the American Medical Informatics Association, 9(6): pp.571–585.

[13] Winter A, Ebert J (1997) Reference model for hospital information systems and their requirements. In: Zwierlein E [Ed.]: Management in Health Care, Urban & Schwarzenberg München, pp.548-562.

[14] Haux R, Schmücker P, Winter A (1996) Gesamtkonzept der Informationsverarbeitung im Krankenhaus. In: Haas P, Köhler CO, Kuhn K, Pietrzyk PM, Prokosch HU [Eds.]: Praxis der Informationsverarbeitung im Krankenhaus. Ecomed Landsberg, pp. 25-37.

[15] Gupta S (2007) Workflow and Process Mining in Healthcare. Master thesis, Technical Univeristy of Eindhoven, Department of Mathematics and Computer Science. http://www.processmining.org/_media/publications/shaifaligupta_masterthesis.pdf (06.04.2012)

[16] Kaiser S, Reichert M (2011) Dataflow variants in process models: scenarios, challenges, approaches. Ulm Informatic Reports, Nr. 2011-03.