Mercy Health System

St. Louis, MO

Process Mining of Clinical Workflows for Quality and Process Improvement

Paul Helmering, Executive Director, Enterprise Architecture

Pete Harrison, Data Analyst, Mercy Technology Services

Vidya Iyer, Solution Architect, Mercy Technology Services

Anil Kabra, Senior Application Architect, Mercy Technology Services

Jeff Van Slette, Biostatistician, Center for Innovative Care

PURPOSE AND GOALS

Moving into the future, clinical process re-engineering, standardization, and optimization will be required to gain the efficiencies, cost reductions, and safety improvements required by the new healthcare model. At present, the efforts required to manage, analyze, and monitor clinical processes are substantial, and mostly manual, which will greatly limit our goals for clinical process improvement. The purpose of our work is to make large scale clinical process management and optimization more cost effective and sustainable. The goals we seek to address: **Automate the method of documenting clinical workflows –** currently this is a manual effort, impacting the cost, abilities, and accuracy of analyzing and optimizing clinical workflows; visual documentation helps educate the organization, and assists in communication, implementation, and adoption. **Identify varying clinical workflows –** comparing workflows against one another and best practices identifies variances in the workflows, knowledge which can be used for standardization and confirmation of process execution to design. **Optimize clinical workflows –** time-, cost-, and role-based analysis of different workflows helps identify workflow efficiencies and inefficiencies.

OBJECTIVES

To accomplish these goals, the following objectives were defined:

- To document and analyze Electronic Health Record (EHR) data and database structures which are used to record user activity in the system (i.e., event logs and related data), and evaluate the data to determine its applicability and fit for automated process discovery and process mining.
- To create a process model framework that will allow meaningful hierarchical decomposition and organization of processes throughout the health system.
- To design and implement database queries to extract process data and related contextual data elements from the EHR database.
- To evaluate and select a commercially available process mining software package, build working knowledge of the software functions and operations, and use process mining algorithms.
- To visualize clinical processes in a user-friendly form, suitable for systems documentation and training, and communication with clinical operations.

DESIGN AND METHOD

The logical and physical models of the system architecture are presented in Figure 1. On the left of Figure 1, the logical view identifies three major clinical process functions. **Data Pre-Processing –** this function extracts event logs and related contextual data elements from the EHR database and formats and filters the data for import and utilization by the other functions. **Process Mining –** this function uses sophisticated mathematical models and algorithms to discover patterns, analyze, and compare process patterns. The function also provides one type of process visualization. **Clinical Process Management –** this function provides a process repository for documentation and change control management, process visualization, and simulation of process execution using process control variables.

On the right side of Figure 1, the technical components of the solution are depicted. Nearly 40 billion rows of log data have been recorded by the EHR since implementation. The database processing time required for extracting and filtering desired data sets was too long to effectively support data preparation for mining. To solve this issue, Infobright, a column-oriented database was implemented and utilized; query processing time was reduced from hours to minutes. Custom coded procedures were written in Java, and used to translate the data sets into the formats required by the selected process modeling and mining software packages, specifically Extensible Event Stream (XES) for process mining and Business Process Modeling Notation (BPMN) for process simulation. The Tibco Business Studio software package is used to simulate processes using varying scenarios of control variables. The ProM [1,2] process mining toolkit was selected because the open-source framework provides robust process mining utilities meeting our requirements.

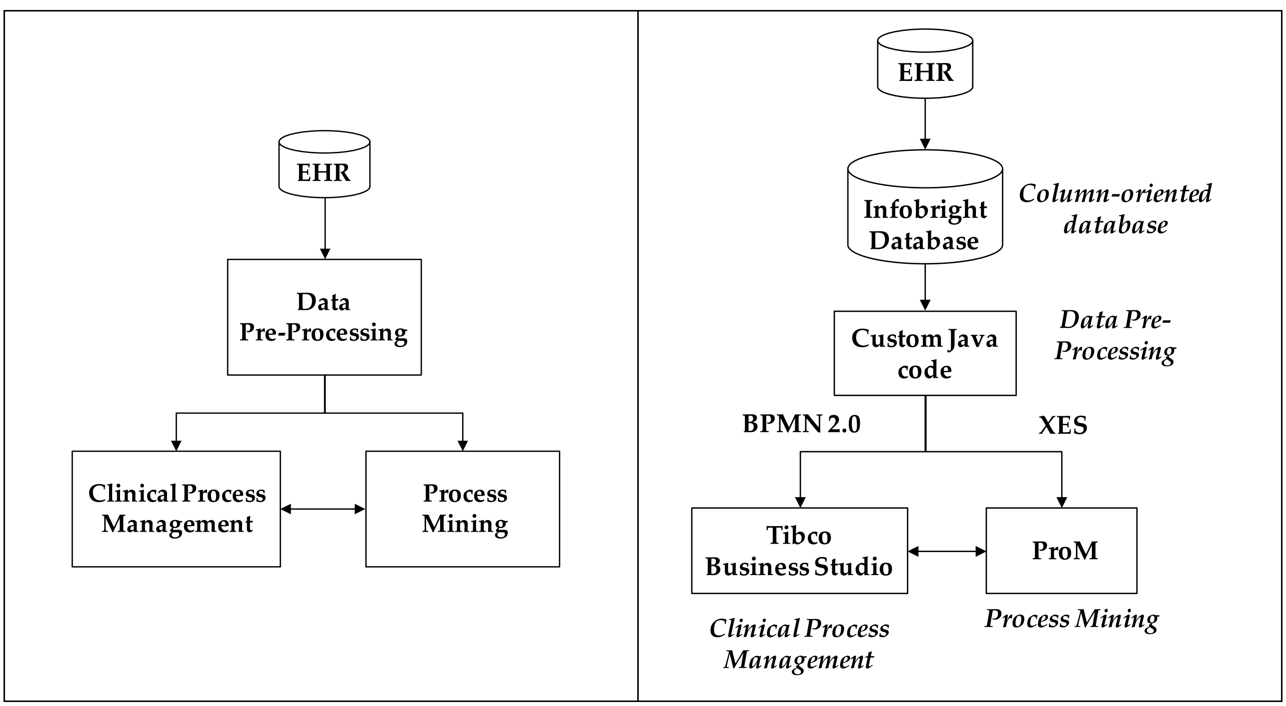

Figure 1. Logical and Technical Views of the System Architecture

RESULTS

The EHR event logs contain process steps and timestamps which are used to sequentially relate activities comprising a clinical process, and contextual attributes – such as department, diagnosis, case number, and system user – which allow grouping, classification, and analysis of processes by clinically relevant parameters. In our framework we selected diagnosis, department performing the process, and clinical procedure to decompose and organize the processes. This decomposition model was well aligned with clinical re-engineering use cases.

The data set used in the examples of this paper was extracted from the Mercy EHR event logs from the month of September 2011 for those patients diagnosed with Congestive Heart Failure (CHF), who underwent a radiological procedure. The resulting filtered data set comprised 91 patients, with 157 radiological procedures performed, and 2,791 total process steps executed.

In this paper we focus mainly on the discovery of processes. The first process mining algorithm applied to the data set was the ProM Analyze Structural Property of Petri net [3], which created a Petri net model of the processes using the alpha-algorithm [3]. The results are displayed in Figure 2a. This visualization is far too complex for effective analysis and communication, and therefore additional mining algorithms were applied. The next process mining algorithm applied to the data set was the ProM Heuristics Miner [4,5]. The mining showed that 42% of the workflows executed exactly one of three main paths, and among all paths, the first eight process steps were standard among 90% of the cases. The fitness of the discovered model was high with only 7 % of the cases in the data set failing to fit the discovered process model. The most common path is highlighted in Figure 2b. Finally the ProM Genetic Miner algorithm [6] was used to extract only the most common path, represented in Figure 2c.

The Performance Sequence Diagram algorithm [7] was used next to evaluate duration of processes and process steps. Sub-processes were identified from the most common path (Figure 2c) for further analysis; the process miner computed statistics on common paths of the sub-process and their durations. In one critical sub-process there were two main workflow paths, representing over 98% of all executed paths of the sub-process. However there was significant variance in the average duration of each path (176 and 342 minutes respectively). Upon analysis of the sub-process contextual data, it was determined that these two paths represented specifically two different geographic regions of the health system, i.e., the two regions had standardized on different workflows. This presented an opportunity for further analysis of efficiency and standardization potential.

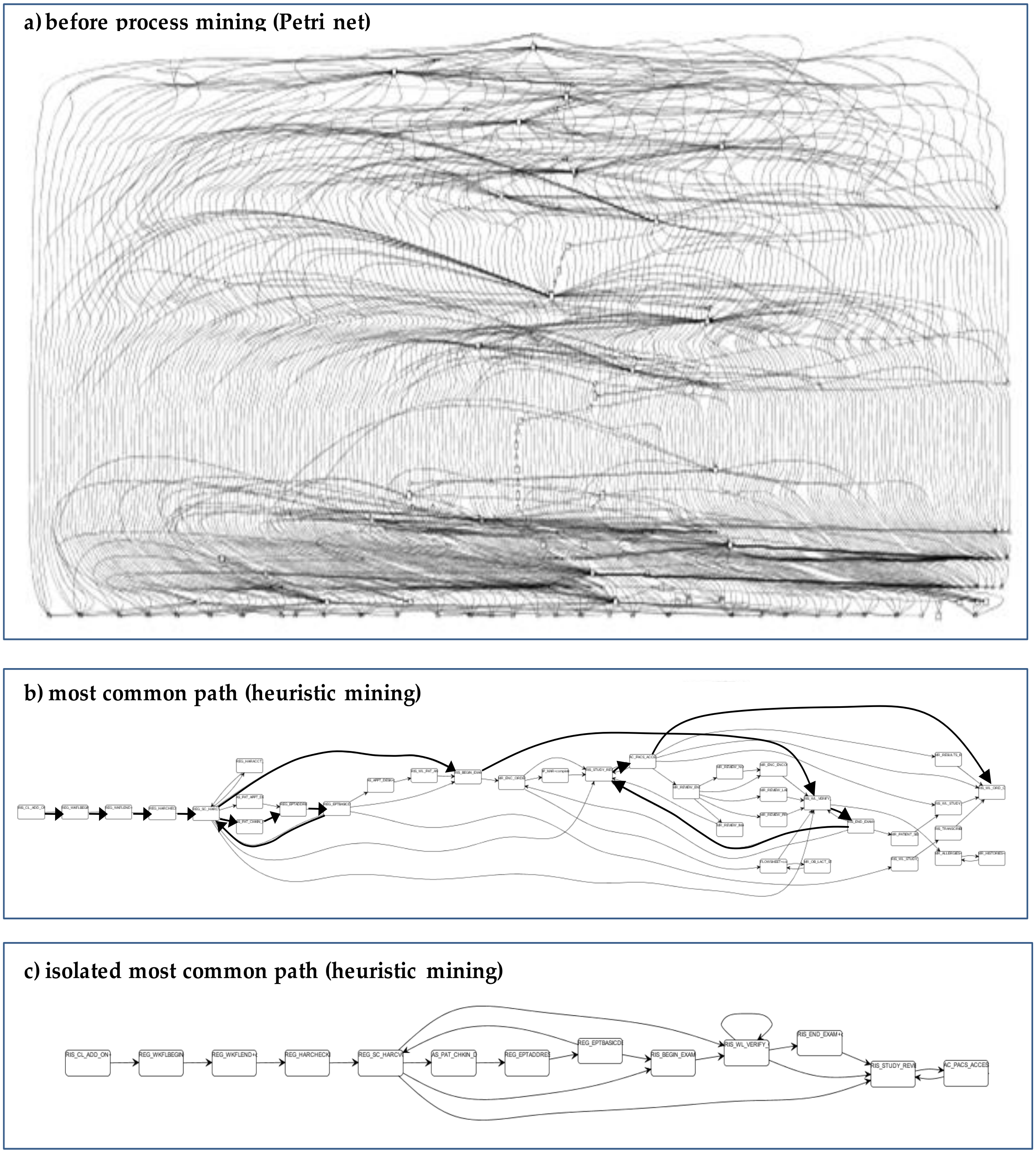

Figure 2. Process Mined Data Set

CONCLUSIONS

In this paper we presented reasons why automated process discovery and process mining are important to improving clinical processes, and a systems approach to implementing these capabilities. We then presented initial results from one pilot, where log files from the clinical EHR were used for process discovery and mining. Through this work we gained confidence that there is significant promise for process mining of clinical processes [8]. Rich data sets comprising process activities and related contextual data elements can be extracted from the EHR and formatted for process mining; a clinically relevant process framework can be established for organizing the processes; the resulting visualizations aid process understanding; and software capabilities and research are available in the industry [9,10].

However, the variation in the analyzed processes requires more research. As a next step, process variance metrics for additional procedures in the enterprise process inventory will be computed. Processes with high variance will be reviewed through observation and interviews to confirm the nature of the variance and whether or not the sequence of the process activities is thought to be relevant by clinical experts. Better conformance models are required in our analysis. Clinical processes may also require new approaches for organizing, joining, decomposing, and summarizing process activities and processes. And most importantly, a strong partnership and working relationship must be established between clinical operations, performance improvement, and process mining teams to ensure technical efforts and process interpretations are consistent with clinical practices, and that high priority process improvement areas are targeted for process mining efforts.

REFERENCES

1. *Analyze Structural Property of Petri Net Plug-in 1.3.6* [computer program]2010.
2. *Genetic Miner 1.0* [computer program]2010.
3. Mans RS, Schonenberg MH, Song M, Aalst WMP, Bakker PJM. Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital. Paper presented at: Biostec2008.
4. Medeiros A, Weijters T. ProM Framework Tutorial: Eindhoven, The Netherlands; 2009.
5. van der Aalst WMP. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*: Springer Berlin Heidelberg; 2011.
6. van der Aalst WMP, Andriansyah A, Alves de Medeiros AK, et al. Process Mining Manifesto: Springer-Verlag; 2012:169-194.
7. *ProM 6.0* [computer program]. Eindhoven Eindhoven University of Technology; 2011.
8. *HeuristicsMiner 6.0: Users Guide* [computer program].
9. *Performance Sequence Diagramer* [computer program]
10. Weijters AJMM, van Der Aalst WMP, Alves de Medeiros AK. Process mining with the HeuristicsMiner algorithm. 2006

AUTHOR BIOGRAPHIES

Paul Helmering has twenty years experience in business intelligence, data analysis, software development, process improvement, and enterprise architecture. He has an economics degree from Harvard University and participates in a wide variety of professional organizations including Forrester and Gartner leadership boards, and the HIMSS mid-west chapter. He currently supports enterprise business and technology strategies for the Mercy Health System. The visualization and analysis techniques being applied to clinical workflows are based on previous experience and development as represented in patent 5,963,922, System for graphically mapping related elements of a plurality of transactions.

Peter Harrison has twenty-five years of experience with software systems and data analysis. He has an M.A. from Southern Illinois University at Carbondale. Experience includes statistical, database, and reporting analyses for laboratory information systems. He currently is a Data Analyst with the Enterprise Architecture group at Mercy. He is a recent member of the HIMSS mid-west chapter.

Vidyalakshmi Iyer has over thirteen years experience in Information Systems Industry most of which is in the data warehousing area. She has an Engineering degree from Ramrao Adik Institute of Technology, Mumbai, India. She is the Solution Architect in Mercy presently the liaison between the business and the technology. For this project she brings the knowledge of Patient EHR System whose event logs were used to determine the clinical process workflow.

Anil Kabra has twenty years of experience in software development, architecture, programming and analysis. He has an Electronics and Communications engineering degree from MBM Engineering College in Jodhpur, India. He is currently an architect with the Enterprise Architecture department for Mercy Health System in St. Louis, Missouri. He has extensive experience with data transformation and specifically worked on preprocessing the data from the event logs to the Extensible Event Stream format required by the ProM tool.

Jeff VanSlette has seven years experience in the use of traditional statistical data analysis and data mining techniques. His experience includes operationalizing predictive models for real world use in diverse areas such as customer retention modeling, quote to sale conversion, and credit based insurance risk scores. He currently services as biostatistician for the Mercy Research Center and Center for Innovative Care, and is performing statistical analysis of EHR clinical data to identify and reduce hospital readmission risks for patients with congestive heart failure.