Innovative informatics methods for process mining in health care

Peer-reviewed author version

Munoz-Gama, Jorge; MARTIN, Niels; Fernandez-Llatas, Carlos; Johnson, Owen A. & Sepúlveda, Marcos (2022) Innovative informatics methods for process mining in health care. In: JOURNAL OF BIOMEDICAL INFORMATICS, (Art N° 104203).

DOI: 10.1016/j.jbi.2022.104203
Handle: http://hdl.handle.net/1942/38693

Innovative informatics methods for process mining in health care

Jorge Munoz-Gama[a], Niels Martin[b,c], Carlos Fernandez-Llatas[d,e], Owen A. Johnson[f], Marcos Sepúlveda[a]

[a] *Pontificia Universidad Catolica de Chile, Chile*
[b] *Hasselt University, Belgium*
[c] *Research Foundation Flanders (FWO), Belgium*
[d] *Universitat Politecnica de Valencia, Spain*
[e] *Karolinska Institutet, Sweden*
[f] *University of Leeds, United Kingdom*

Keywords: Process Mining, Health Care

1. Introduction

Process mining is an emerging discipline which encompasses a wide variety of methods to extract process-related knowledge from process execution data recorded by information systems [1]. Within a health care context, process mining can be used for a variety of use cases such as discovering treatment trajectories with their associated clinical outcomes, or assessing the adherence of a treatment trajectory to a normative model such as a clinical pathway [2, 3]. The data that is used as a starting point for process mining is typically readily available in health information systems such as electronic medical records [3]. Nevertheless, health care is a highly complex focus domain in process mining, among others due to the knowledge-intensive

character of health care processes. This implies that decisions regarding a patient's path in a clinical process are made by extensively trained healthcare professionals such as physicians and nurses based on a careful far-from-trivial assessments of all available data [4]. Despite its challenging nature, process mining in health care offers great potential given the significant societal relevance of methods aimed at improving the health care system [5]. In particular, the generated process-related insights can be leveraged by health care professionals to identify areas for process improvement [3]. As a consequence, the development of innovative methods for process mining in health care is of great interest to both research and society.

To advance the state-of-the-art on process mining in health care, the Journal of Biomedical Informatics organized a special issue on the topic "Innovative informatics methods for process mining in health care". After a thorough review process, twelve articles were accepted for inclusion in the special issue: nine original research articles, and three special communications. Table 1 provides an overview of the accepted articles, structured by the topic of their contribution. This overview shows that the special issue covers a wide spectrum of the field: a systematic literature review and characterization of the field to get a better understanding of the intersection of process mining and health care (two articles), methodologies to apply process mining in health care (two articles), innovative methods to use process mining to analyze the specific domains of clinical pathways (two articles) and capacity management (two articles), articles that explore process mining techniques beyond discovery, such as conformance and enhancement (two articles), and finally articles on how to gain explainable insights into health

care processes, one of the strong points of process mining compared to other computer science disciplines (two articles). The remainder of this editorial will elaborate on the variety of topics covered within this special issue.

Table 1: Articles included in the special issue

Reference	Title	First Author
	Understanding process mining in health care	
[6]	Process mining in healthcare – an updated perspective on the state of the art	De Roock, E.
[7]	Process mining for healthcare: characteristics and challenges	Munoz-Gama, J.
	Methodologies for process mining in health care	
[8]	Process mining-driven analysis of COVID-19's impact on vaccination patterns	Augusto, A.
[9]	Multi-level process mining methodology for exploring disease-specific care processes	Vathy-Fogarassy, A.
	Analyzing pathways with process mining	
[10]	Assessment of the feasibility of developing a clinical pathway using a clinical order log	Lim, J.
[11]	Signal from the noise: A mixed graphical and quantitative process mining approach to evaluate care pathways applied to emergency stroke care	Noshad, M.
	Process mining for capacity management	
[12]	Supporting capacity management decisions in healthcare using data-driven process simulation	van Hulzen, G.
[13]	Process data analytics for hospital case-mix planning	Andrews, R.
	Beyond discovery: conformance and enhancement	
[14]	Are we ready for conformance checking in healthcare? Measuring adherence to clinical guidelines: A scoping systematic literature review	Oliart, E.
[15]	How can interactive process discovery address data quality issues in real business settings? Evidence from a case study in healthcare	Benevento, E.
	The search for explainability	
[16]	A framework for the automatic description of healthcare processes in natural language: Application in an aortic stenosis integrated care process	Fontenla-Seco, Y.
[17]	Explainable process trace classification: an application to stroke	Leonardi, G.

2. Understanding process mining in health care

In order to successfully use process mining in the health care domain or to conduct relevant research, it is necessary to thoroughly understand this interdisciplinary research area. This involves understanding its specific

characteristics, where process mining has been already successfully applied in health care, which challenges have been addressed and what the existing research gaps are. The special issue includes two articles that contribute to the understanding of the area.

De Roock et al. [6] present an extensive systematic literature review on process mining in health care, in which 263 papers have been included in the literature analysis. This study not only updates previous reviews on the field, such as Rojas et al.[2], but also discusses new perspectives on the literature such as the Key Performance Indicators (KPI) used to measure the impact of process mining, the involvement of domain experts, and the process mining project stages that received attention. This study confirms process mining in health care as a research area in full development, identifying trends as well as open research challenges.

Munoz-Gama et al. [7] present the outcomes of a four-year community effort which involved 55 highly relevant researchers and practitioners in the field, in order to convey a unified view on process mining in health care as a research domain. To this end, they identify the distinguishing characteristics to be considered to successfully leverage process mining in the health care domain, as well as identifying open challenges that need to be addressed by the community in the future. In particular, ten key distinguishing characteristics arise from the study – from the use of medical guidelines and protocols, to the different abstraction levels of the health care data– as well as ten key challenges – from minding concept drift, to the need of process mining techniques beyond discovery.

3. Methodologies for process mining in health care

Methodologies and step-by-step procedures to support the application of process mining in health care are highly important given the complexity of the health care domain. This special issue contains two articles related to this topic.

Augusto et al. [8] show how process mining and traditional data mining can be applied to analyze a COVID-19-related health care scenario. To this end, they apply a three-step procedure: prepare the health care process data for conducting a process mining analysis; select and apply suitable process mining solutions for successfully executing the analysis; and extract valuable insights from the obtained results. Following these steps generate relevant clinical insights regarding the health care service utilization patterns during the COVID-19 pandemic.

Vathy-Fogarassy et al. [9] propose a novel methodology to apply process mining, called Methodology for Exploring Disease-specific Care Processes (MEDCP). In particular, MEDCP distinguishes itself by being multi-level, disease-specific, and involving expert knowledge in terms of taxonomies. Results show that the methodology can provide efficiency improvements of standard process mining methods, and generate conclusions that are easy to interpret by domain experts.

4. Analyzing pathways with process mining

The analysis of clinical and care pathways is one of the most extensively explored use case for process mining as pathways are one of the most common

forms of processes in the health care domain. The special issue include two innovative studies that contribute to the state-of-the-art in this direction.

Lim et al. [10] address the issue of variability of health care data when trying to define clinical pathways. Before applying process mining methods to establish a clinical pathway, it is important to assess whether a care process can be standardized. In this respect, the study proposes a novel method for assessing the clinical pathway feasibility based on the clinical complexity process. The proposed approach consisted of three steps: data preparation, activity and trace homogeneity evaluation, and process inspection using process mining. For each step, the article proposes novel metrics and visualization methods.

Noshad et al. [11] present a study that utilizes the Electronic Health Records (EHR) as an event data source for the application of process mining. The study illustrates the use of an unsupervised process mining algorithm which generates process graphs with the most common process patterns, but also enables calculating the conformity of clinical pathways.

5. Process mining for capacity management

In order to execute health care processes, various resources are typically required such as medical equipment (e.g. radiology devices) and staff (e.g. physicians and nurses). Capacity management relates to determining the required level of resources, as well as their allocation within the process [18]. Two articles in the special issue focus on the use of process mining within the context capacity management in health care.

van Hulzen et al. [12] use Data-Driven Process Simulation (DDPS) to

support capacity management decisions at the radiology department of a hospital. These decisions are situated within the context of the construction of new facilities and relate to the required radiology equipment, waiting room space and receptionists in the future new setting. Particular attention is attributed to the interaction between process execution data and domain expertise within the context of DDPS.

Andrews et al. [13] present an approach to use process mining in the patient case-mix planning of a hospital. The patient case-mix describes the type and volume of patients which are treated at a hospital and will have a direct impact on the resource utilization and, hence, on what can be achieved with a particular volume of resources. The proposed approach has been developed in collaboration with the Queensland Children's Hospital.

6. Beyond discovery: conformance and enhancement

Process mining techniques can be classified in three main types [1]: *discovery* (i.e., to build a process model from the event data), *conformance* (i.e., to compare an existing process model with the event data), and *enhancement* (i.e., to extend or modify an existing process model with information from the event data). Discovery is the most common process mining use case in health care, representing the majority of studies in the literature. Nonetheless, conformance and enhancement techniques are also highly relevant, as demonstrated by two studies included in this special issue.

With regard to conformance, Oliart et al. [14] study the role that process mining could play in measuring the degree to which health care organizations adherence to clinical guidelines. To this end, they review the criteria

used in literature to assess adherence over the past 20 years and explore the suitability of using process mining techniques. The 31 studies included in the review show that there are several threats to the applicability of process mining to measure adherence, including a lack of clear and concise rules, the fact that process models are not defined in most cases, and the availability of unstructured and incomplete data which require human interpretation.

With regard to enhancement, Benevento et al. [15] explore the idea of interactive process mining as a means to improve the quality of the process models obtained. Their approach combines the available event data with domain knowledge in order to improve the process modeling results when noisy and incomplete data is used, which is common in health care. The results of the study show that, in that context, interactive process mining constitutes a more suitable approach than traditional process discovery techniques.

7. The search for explainability

End-users value process mining as a means to understand what is going on in their process execution data. While other techniques in data science are often considered *black boxes* (i.e., an input is introduced, and an output with the suggested next step is produced as output, without understanding the reason for this suggestion), process mining techniques are considered *white box* (i.e., the final goal is not to produce an output, but the help you understand your data, in order to take a better decision). The special issue includes two studies that explore process mining as a mechanism to explain the data.

Fontenla-Seco et al. [16] explore a recent research trend: the combination

of process mining with Natural Language Processing. This is a promising research line in interdisciplinary domains, such as health care, where domain experts are more familiar with textual descriptions than other process modeling graphical notations. The work presents a novel framework for the automatic generation of natural language descriptions of health care processes, combining quantitative and qualitative data, and medical domain knowledge. The study shows promising results in terms of the understandability, usefulness and impact of the natural language descriptions on the medical experts' work.

Leonardi et al. [17] explore a different problem: the classification of process traces. Trace classification can be used to check or predict if an execution will fulfill some expected criteria (e.g.: it gets completed in less than a certain time). This information is useful both for a better planning of resources and to identify non-complying cases. As an alternative to more black-box input-output alternatives, this work proposes a novel concept – *trace saliency maps* – as a way of providing an explanation for the output, making it more easy to interpret for medical users.

8. Summary

The papers collected in this special issue cover a wide range of topics on process mining in health care. Besides papers which aim to structure the research field, various innovative methods are presented with the ambition of providing rich data-driven insights in health care processes. The variety of topics illustrates the broad applicability of process mining and its associated (research) challenges. Besides the variety in topics, the contribu-

tions to the special issue also exhibit a diversity in terms of the country of origin (from South-Korea and Australia, over Italy, Spain and Belgium, to the United States), the health care department considered (such as cardiology and radiology) and the health care organizations considered (such as children's hospitals and university hospitals). This illustrates the general attention for process mining in health care across continents and care contexts. While many open challenges remain and new challenges have been identified in the accepted papers, this special issue will move the state-of-the-art in this promising research domain forward.

Acknowledgment

This special issue is an initiative of the Process-Oriented Data Science for Healthcare Alliance, which is affiliated with the IEEE Task Force on Process Mining. The alliance aims to promote research and knowledge sharing between multidisciplinary stakeholders in the use of process mining approaches in the health care domain. The editors of this special issue would like to thank all reviewers for their valuable feedback, which certainly contributed towards improving the quality of the articles that were ultimately included in the special issue.

References

[1] W. M. P. van der Aalst, A. Adriansyah, A. K. A. de Medeiros, F. Arcieri, T. Baier, T. Blickle, R. P. J. C. Bose, P. van den Brand, R. Brandtjen, J. C. A. M. Buijs, A. Burattin, J. Carmona, M. Castellanos, J. Claes, J. Cook, N. Costantini, F. Curbera, E. Damiani, M. de Leoni, P. Delias,

B. F. van Dongen, M. Dumas, S. Dustdar, D. Fahland, D. R. Ferreira, W. Gaaloul, F. van Geffen, S. Goel, C. W. Günther, A. Guzzo, P. Harmon, A. H. M. ter Hofstede, J. Hoogland, J. E. Ingvaldsen, K. Kato, R. Kuhn, A. Kumar, M. L. Rosa, F. M. Maggi, D. Malerba, R. S. Mans, A. Manuel, M. McCreesh, P. Mello, J. Mendling, M. Montali, H. R. M. Nezhad, M. zur Muehlen, J. Munoz-Gama, L. Pontieri, J. Ribeiro, A. Rozinat, H. S. Pérez, R. S. Pérez, M. Sepúlveda, J. Sinur, P. Soffer, M. Song, A. Sperduti, G. Stilo, C. Stoel, K. D. Swenson, M. Talamo, W. Tan, C. Turner, J. Vanthienen, G. Varvaressos, E. Verbeek, M. Verdonk, R. Vigo, J. Wang, B. Weber, M. Weidlich, T. Weijters, L. Wen, M. Westergaard, M. T. Wynn, Process mining manifesto, Lecture Notes in Business Information Processing 99 (2011) 169–194.

[2] E. Rojas, J. Munoz-Gama, M. Sepúlveda, D. Capurro, Process mining in healthcare: A literature review, Journal of Biomedical Informatics 61 (2016) 224–236.

[3] N. Martin, J. D. Weerdt, C. Fernández-Llatas, A. Gal, R. Gatta, G. Ibáñez, O. A. Johnson, F. Mannhardt, L. Marco-Ruiz, S. Mertens, J. Munoz-Gama, F. Seoane, J. Vanthienen, M. T. Wynn, D. B. Boilève, J. Bergs, M. Joosten-Melis, S. Schretlen, B. B. V. Acker, Recommendations for enhancing the usability and understandability of process mining in healthcare, Artificial Intelligence in Medicine 109 (2020) 101962.

[4] C. Di Ciccio, A. Marrella, A. Russo, Knowledge-intensive processes: characteristics, requirements and analysis of contemporary approaches, Journal on Data Semantics 4 (1) (2015) 29–57.

[5] N. Martin, N. Wittig, J. Munoz-Gama, Using process mining in healthcare, in: W. M. P. van der Aalst, J. Carmona (Eds.), Process Mining Handbook, Springer, 2022, pp. 416–444.

[6] E. De Roock, N. Martin, Process mining in healthcare – an updated perspective on the state of the art, Journal of Biomedical Informatics 127 (2022).

[7] J. Munoz-Gama, N. Martin, C. Fernandez-Llatas, O. A. Johnson, M. Sepúlveda, E. Helm, V. Galvez-Yanjari, E. Rojas, A. Martinez-Millana, D. Aloini, I. A. Amantea, R. Andrews, M. Arias, I. Beerepoot, E. Benevento, A. Burattin, D. Capurro, J. Carmona, M. Comuzzi, B. Dalmas, R. de la Fuente, C. Di Francescomarino, C. Di Ciccio, R. Gatta, C. Ghidini, F. Gonzalez-Lopez, G. Ibanez-Sanchez, H. B. Klasky, A. Prima Kurniati, X. Lu, F. Mannhardt, R. Mans, M. Marcos, R. Medeiros de Carvalho, M. Pegoraro, S. K. Poon, L. Pufahl, H. A. Reijers, S. Remy, S. Rinderle-Ma, L. Sacchi, F. Seoane, M. Song, A. Stefanini, E. Sulis, A. H. ter Hofstede, P. J. Toussaint, V. Traver, Z. Valero-Ramon, I. van de Weerd, W. M. van der Aalst, R. Vanwersch, M. Weske, M. T. Wynn, F. Zerbato, Process mining for healthcare: Characteristics and challenges, Journal of Biomedical Informatics 127 (2022).

[8] A. Augusto, T. Deitz, N. Faux, J.-A. Manski-Nankervis, D. Capurro, Process mining-driven analysis of covid-19's impact on vaccination patterns, Journal of Biomedical Informatics 130 (2022).

[9] Ágnes Vathy-Fogarassy, I. Vassányi, I. Kósa, Multi-level process min-

ing methodology for exploring disease-specific care processes, Journal of Biomedical Informatics 125 (2022).

[10] J. Lim, K. Kim, M. Song, S. Yoo, H. Baek, S. Kim, S. Park, W.-J. Jeong, Assessment of the feasibility of developing a clinical pathway using a clinical order log, Journal of Biomedical Informatics 128 (2022).

[11] M. Noshad, C. C. Rose, J. H. Chen, Signal from the noise: A mixed graphical and quantitative process mining approach to evaluate care pathways applied to emergency stroke care, Journal of Biomedical Informatics 127 (2022).

[12] G. van Hulzen, N. Martin, B. Depaire, G. Souverijns, Supporting capacity management decisions in healthcare using data-driven process simulation, Journal of Biomedical Informatics 129 (2022).

[13] R. Andrews, K. Goel, P. Corry, R. Burdett, M. T. Wynn, D. Callow, Process data analytics for hospital case-mix planning, Journal of Biomedical Informatics 129 (2022).

[14] E. Oliart, E. Rojas, D. Capurro, Are we ready for conformance checking in healthcare? measuring adherence to clinical guidelines: A scoping systematic literature review, Journal of Biomedical Informatics 130 (2022).

[15] E. Benevento, D. Aloini, W. M. van der Aalst, How can interactive process discovery address data quality issues in real business settings? evidence from a case study in healthcare, Journal of Biomedical Informatics 130 (2022).

[16] Y. Fontenla-Seco, M. Lama, V. González-Salvado, C. Peña-Gil, A. Bugarín-Diz, A framework for the automatic description of healthcare processes in natural language: Application in an aortic stenosis integrated care process, Journal of Biomedical Informatics 128 (2022).

[17] G. Leonardi, S. Montani, M. Striani, Explainable process trace classification: An application to stroke, Journal of Biomedical Informatics 126 (2022).

[18] V. L. Smith-Daniels, S. B. Schweikhart, D. E. Smith-Daniels, Capacity management in health care services: review and future research directions, Decision Sciences 19 (4) (1988) 889–919.