

Analysis of patient pathways with contextual process mining

Fazle Rabbi¹, Bahareh Fatemi¹ and Wendy MacCaull²

¹Information Science and Media Studies, University of Bergen, Norway

²Department of Computer Science, St Francis Xavier University, Canada

Abstract

The volume and diversity of healthcare information available through the use of modern technology offers great potential for improving health services delivery. Giving healthcare practitioners and health system administrators the ability to investigate and analyse data from various perspectives can promote evidence-based decision making. Current analysis approaches often result in models where essential relations are difficult to depict and/or discern. Further it is not easy to dynamically change the level of detail in order to accommodate user requirements and allow users to analyse data from various perspectives. We outline an approach involving filtering and abstraction of event logs using ontologies, dimensional modeling and cluster analysis so healthcare professionals may access the information they need. Our approach enhances existing process mining techniques by incorporating data mining techniques for dividing patient records into closely related groups, and results in contextual process mining.

Keywords

healthcare information, abstraction, process mining, community detection, visualization, ontology, context

1. Introduction

Today's vast amount of healthcare information needs to be accessed easily, integrated intelligently and analysed appropriately to support better healthcare delivery. With intelligent integration and analysis, data can be used to gain deeper insights into patient health using historical healthcare information (including a patient's symptoms, appointments, exercise routines, lab results, vital signs, prescriptions, allergies, assessments, etc.) and also can be used to determine the effectiveness of treatment. Context relevant clinical information can enable improved decision making. Healthcare data analysis across disciplines can facilitate the discovery of new knowledge which can lead to streamlined workflows, greater efficiency and improved patient care. Systematic analysis of healthcare data can help to detect patterns so that healthcare providers can optimize their resource allocation and clinicians can optimize treatment plans for individuals leading to better health outcomes.

Healthcare systems consist of numerous health facilities that execute clinical and non-clinical activities. Up-to-date information can influence decision makers both at the level of the indi-

The International Health Data Workshop (HEDA-2022), June 19-24, 2022, Bergen, Norway

✉ fazle.rabbi@uib.no (F. Rabbi); Bahareh.Fatemi@uib.no (B. Fatemi); wmaccaul@stfx.ca (W. MacCaull)

ORCID [0000-0001-5626-0598](https://orcid.org/0000-0001-5626-0598) (F. Rabbi)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

vidual practitioner and at the system level. Clinicians require information regarding various treatments applied over different patient groups to choose the best treatment plan for the individual patient. At the system level, there is an increasing demand to optimize systems by the efficient use of the limited resources available. Both clinicians and administrators would benefit from more intuitive tools that would allow them to more fully utilize the data available to them without the need for sophisticated technical knowledge. They are particularly interested to find common pathways for patients, to ascertain how a process model be improved and to determine to what extent existing systems are following clinical guidelines. However it is not easy to meet these demands as healthcare processes are highly dynamic, complex, ad-hoc, and are increasingly multidisciplinary [1].

Data mining techniques present the opportunity to analyse and learn from healthcare information, from numerous viewpoints (i.e., contexts) such as with respect to patient populations with specific diseases, ages, gender, incidences of co-morbidity, type of healthcare service setting (e.g., clinic, hospital, nursing home), home location (urban or rural), procedures used, etc. Clustering [2, 3] and community detection [4, 5] can provide useful information in order to understand relationships between a patient's symptoms and other information. These data mining algorithms have the potential to identify similar groups of patients but do not include temporal information and therefore lack the capacity to determine patient pathways. On the other hand, process mining [6] techniques hold great potential to support health services by identifying common flows of patients, but they are not equipped with clustering and community detection methods. To adequately study the progression of diseases and the flow of patients from a tight-knit group we need an integrated approach combining data mining and process mining. As with much of the literature on network science and social network analysis, the terms clustering and community detection are used interchangeably. In the rest of the paper we use the term "community structure" to refer to the concept of a network structure where the nodes are densely connected internally.

In [7] we introduced the idea of using a new approach, called model based slicing, that utilizes ontologies and dimensional models to access the data required for improved data analysis. This approach exploits the use of structured information available in the healthcare industry such as standard ontologies (e.g., ICD (International Classification of Disease)-10 [8], designed to provide diagnostic codes for classifying diseases, and SNOMED-CT [9] which provides a comprehensive terminology for clinical health) as well as hospital organizational and other relevant hierarchies.

In this paper we propose a data mining approach that integrates with and augments process mining which also utilizes healthcare ontologies. The proposed approach has the potential to analyze patients with similar issues. We present a flexible approach for extracting healthcare data through a data preprocessor designed to pipe the data extracted according to a particular user's specifications to existing data analysis tools. The experiments described in this paper used data on patient admission, diagnosis, and procedural activities over the period of 6 months taken from the Norwegian Patient Registry (NPR data).

The rest of this paper is organized as follows. Section 2 outlines popular solutions available for analysing healthcare data and discuss some of the problems that remain. Section 3 describes how our data abstraction methods can be used to enhance output from community detection technology. Section 4 discusses related work and in section 5 we conclude the paper and give

directions for our future work.

2. Existing solutions and some problems that remain

In this section we outline popular techniques for analyzing healthcare data, both timed (i.e., associated with processes) and untimed, and point out several problems that remain.

2.1. Process mining

In many developed countries, most citizens use public healthcare services involving multiple service providers providing diverse services supported by numerous software applications. Various types of data analysis are required due to the various roles played by healthcare professionals e.g., while a clinician is interested to see the progression of diseases and to study the effects of different treatment procedures, healthcare managers are interested to study the patient flow from one clinic to another and the number of patients waiting to get services.

In principle, process mining can be used to acquire and explore new knowledge about patient condition, adherence to treatment program, effectiveness of treatment, etc. The main focus of process mining in healthcare is to provide evidence-based process analysis techniques for effective process management [10]. It combines data mining and process analysis techniques and is used to discover trends and patterns of process executions by analyzing the trace of activities performed in a system. These traces are referred to as 'event logs'. However, existing process mining techniques use event logs as input which typically consist of event information such as case-id, event time, event name and attributes of events. Since healthcare information is much more complex than information from most other domains, current process mining techniques cannot adequately mine the information from the event logs required by the various health professionals and therefore are not exploited to their full potential.

In a healthcare setting, patients may need to visit various health facilities for various health related issues and also may need to use certain online treatment programs. Hence, for effective process mining, event logs representing activities of various systems must be considered. Event logs from a variety of systems using various data definitions and formats must therefore be harmonized before they can be analysed by any process mining algorithm. Another issue with applying process mining in healthcare data is related to the diversity of information. Identifying common pathways for patient flow in healthcare systems is complicated by the large variety of patient conditions and diagnoses. In healthcare, the data preparation task is very critical as healthcare data are very sensitive and therefore semantics of the data must be preserved. Getting the right setup for data preparation is important to get the best understanding from the data as efficiently as possible [11]. The current situation in healthcare information systems makes this a complicated problem indeed.

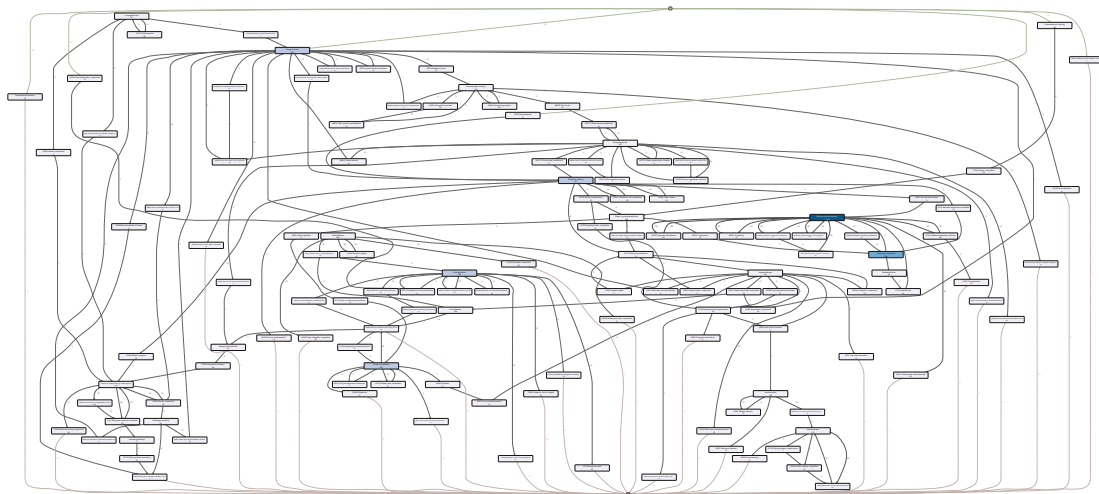
Imagine a situation where we have event logs representing patient information with respect to patient visits to various clinics (or service points in the hospital), and patient diagnoses. Table 1 shows a portion of such an event log. The first column of the event log shows case id; the second column shows when the event occurred; the third column shows event names, (such as check in at a department or clinic, or a clinical finding) and the fourth column shows the resource for the information. Due to the complexity in the process flow and diversity of

Table 1

Portion of a healthcare event log (Sample data)

Id	Event time	Event name	Responsible health worker
1	2017-03-20 13:30	Surgical Clinic	Kristi Salazar
1	2017-03-20 13:30	(N39_9) Disorder of urinary system	Darla Ramirez
2	2017-03-07 14:00	Radiology Department	Ricky Alvarado
2	2017-03-07 15:00	(N63) Unspecified lump in breast	Deborah Tyler
2	2017-03-07 15:15	(N64_5) Other symptoms in breast	Johanna Buchanan
3	2017-04-06 08:30	Division of Mental Health Protection	Henrietta King
3	2017-04-06 08:30	(F321) Depressive episode	Beatrice French

diagnosis information, the process model discovered from such an event log is very large in size and it is difficult to extract meaningful information from it. Even after filtering, the event log to keep, e.g., only the patient cases which are admitted to the Radiology Department, the model is still very large. Figure 1 shows a fragment of radiology patients' flow (using a screenshot from *Fluxicon Disco* [12]) from the NPR sample dataset as mentioned in section 1. While the *Fluxicon Disco* process mining tool is equipped with several filtering mechanisms, they are not always adequate. For example, *Fluxicon Disco* can ignore low frequency nodes to reduce the size of a diagram but this results in the loss of information that may be relevant. *Fluxicon Disco* and similar other process mining tools such as PROM [13] cannot provide an overall view of a large process model from a higher level of abstraction.

**Figure 1:** A fragment of Radiology Patients flow visualized in *Fluxicon Disco* Process mining tool

Another issue with existing process mining tools is that they do not provide a flexible mechanism to perform context aware process mining. While analyzing common pathways for patients, different contexts are required to allow clinicians to focus on different groups of patients and to visualize their careflows. Clinicians would be interested to study the comorbidity issues of patients and analyze the pathways of different patient groups. Existing process mining techniques may be used to visualize the flow of patients but the event logs need

to be prepared in different ways in order to support a wider variety of queries. In the following, we provide an example to illustrate this need. Considering the event logs from Table 1, suppose we wish to investigate the flow of those patients to various departments in the hospital who have been diagnosed with some mental disorders, i.e., they have an ICD-10 code in the range F00-F99.9. We wish to focus on only those patients who are very similar or, in other words, belong to a community structure. The event logs in Table 1 consists of some information at a very low level, i.e., visit to clinics and specific diagnosis information. We need to relate clinics to their department hierarchy and specific diagnosis information to their diagnosis group from ICD-10 ontology. Relating clinics to departments and diagnosis to diagnosis group allows us to choose the desired abstraction layer. This choice of abstraction layer for defining the context and integrate community detection algorithms are not supported by existing process mining tools. In [7] we argued that the current practice of process mining needs to be advanced by means of a rich information model which supports various perspectives needed to analyze event logs.

2.2. Contextualization using data mining

Healthcare data are received from various healthcare service providers and personal health applications such as patient monitoring applications and sensor devices. It is becoming a difficult task for health caregivers to analyze all these data from various sources and extract information in a meaningful way. Visualization of healthcare information can greatly help care providers to gain insights from patients data and make important decisions [14]. Data mining techniques such as community detection are used to identify useful information from healthcare data and the outcome of such data mining technologies can be useful to healthcare providers enabling them to understand effectiveness of treatment processes and to streamline healthcare processes and treatment protocols [2]. Community structure in patient records based on patient condition and symptoms provide useful information for efficiently detecting unknown and valuable information. Data mining approaches can also be used to help medical researchers in making efficient healthcare policies, constructing drug recommendation systems, and developing health profiles of individuals [2]. Community detection algorithms are easy to apply for healthcare data as they are based on unsupervised learning techniques, i.e., the data does not need to be labeled.

One of the challenges for revolutionizing current healthcare systems is to adapt the treatment procedure according to the various needs of a diverse group of patients. This requires better understanding of each patient's condition which include patient demographics, symptoms, preferences, interests, etc. In general, healthcare systems consist of various processes inside hospital organizations and several other processes outside the boundary of healthcare organization to support online treatments. The central motive of personalized medicine is the premise that an individual has a significant role in disease vulnerability and the selection of treatments [15]. Such personalization and adaptive treatment systems behave differently for different users. The decision on how these systems should behave for any particular user is based on a user model, which is a detailed representation of an individual user's information, such as *user preferences, interests, behavior, background, knowledge, individual traits, and others*. As these attributes are dynamic it is essential to create, maintain, and update the user model.

There exists several community detection algorithms to partition the dataset based on the association of its individuals. Patient demographic information along with other information can be effectively used to group patients using clustering or community detection algorithms such as k -Means Clustering Algorithm [16, 2], Community detection algorithm [5, 4]. [17] constructs a lower-dimensional feature matrix of the weighted network using a deep sparse autoencoder and then performs K-means algorithm for obtaining the community detection results. [18] proposes an embedding-based method that tackles the problem of overlapping communities. They design a Generative Adversarial Net (GAN) for optimization. [19] propose a Markov Random Field (MRF) method formalizing modularity as the energy function. However, such algorithms are not equipped with healthcare ontologies to provide perspectives from various abstraction levels. There are situations where domain experts would like to investigate patient problems on a specific area with a greater detail but at the same time only get an overview for other areas with an abstract level. For example, a psychologist would be interested to see the patients with specific mental problems and their co-morbidity with ailments such as diabetes, cardiovascular disease, cancer, infectious diseases, etc. In the next section we propose a solution that allows specifying contexts for mining patient information from a variety of abstraction levels, leading to improved process mining.

3. Contextual process mining

To exploit the vast amount of information available in healthcare information systems, we need to develop ways to support contextual data analysis. One may need to both filter information to select certain subpopulations of patients, or certain clinical contexts, and also to group information to allow patterns to emerge from a more abstract view. The main challenge in performing data analysis using existing process mining and clustering tools and techniques is the lack of support for the needed filtering and abstraction which would give the individual practitioner or administrator the ability to extract and explore data from whatever perspective they deem appropriate. In [7], Rabbi et al., presented a model based slicing technique for process mining which utilizes dimensional modeling and ontological representations of healthcare information. The slicing techniques allow domain experts to analyze pathways for patients from various contexts e.g., patients diagnosed with cancer, or patients admitted into the women's clinic, etc. The approach was based on filtering mechanism where we allowed filtering over the abstraction level on healthcare ontologies, but this approach was limited to selecting patient groups of similar kinds. For example, in [7] it is possible to filter patients who have been diagnosed with any 'mental disorder' but that may include patients who have been through a variety of other issues. Patient P_1 may be quite different than P_2 even though they had some commonalities. In this paper, we propose to utilize a community detection algorithm over patient records to identify patient groups with common issues. In order to support a variety of abstraction levels, we introduce a pre-processing step where we manipulate the input for existing community detection algorithms. The input to the community detection is manipulated based on a dimensional model which includes healthcare ontologies.

The NetworkX python library allows us to construct directional and bi-directional graphs with many nodes and edges. The library allows us to export the graphs in various formats such

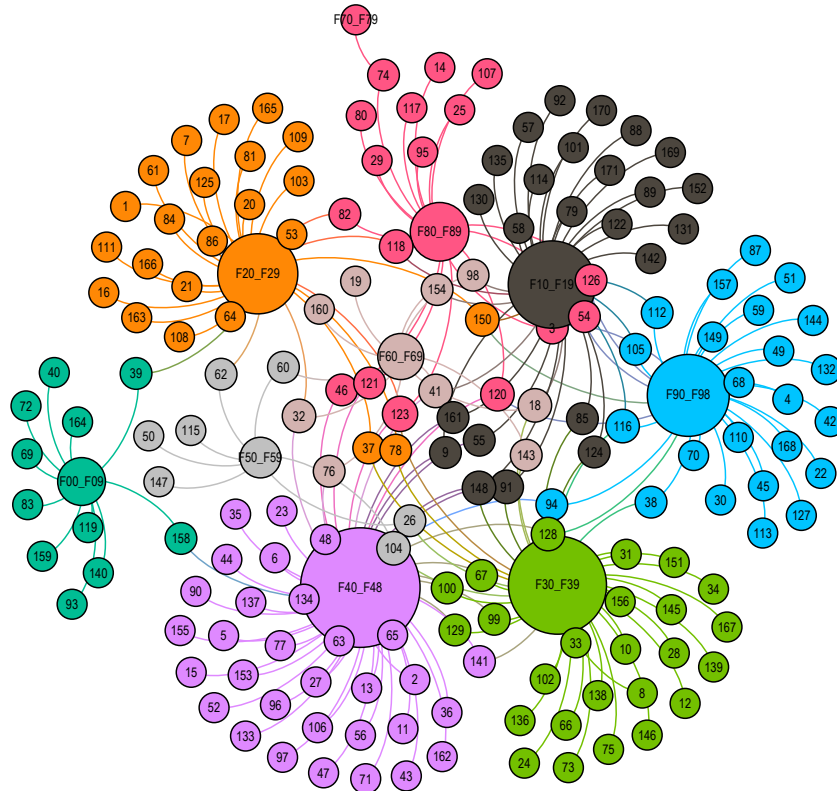


Figure 3: Clustering of patients based on their diagnosis (abstract view)

Although this visualization gives us an idea of overall clustering associated with diagnoses, and indicates that some patients are associated with more than one diagnoses, the fine grained detail prevents us from getting a quick visualization giving meaningful information on patients having more than one diagnoses.

We constructed another graph, see Figure 3, where we used the group code of ICD-10 diagnostics code (i.e., diagnoses in the range F00_F09 are grouped, etc). Here different colours denote different clusters and the size of the node representing the group code is determined by the degree of the node (i.e., the number of patients with that group code). This abstraction yields a better visualization: we can more easily pick the patients who have been diagnosed with multiple mental issues. From this figure we can see that many of the patients were diagnosed with F40_F48 (Anxiety, dissociative, stress-related, somatoform and other nonpsychotic mental disorders), F30_F39 (Mood [affective] disorders), F20_F29 (Schizophrenia, schizotypal and delusional disorders), F90_F98 (Unspecified mental disorder), F10_F19 (Mental and behavioural disorders due to psychoactive substance use) and the overlap is indicated by the patient nodes in the center of the graph with their association with multiple diagnosis nodes.

Now that we have seen how ontologies can be incorporated in community detection, we show how the selected group of patient event logs can be analyzed with process mining tools. An analyst may be interested in investigating the progression of diseases or the admission flow of patients in different departments. In our approach, appropriate events are prepared based on

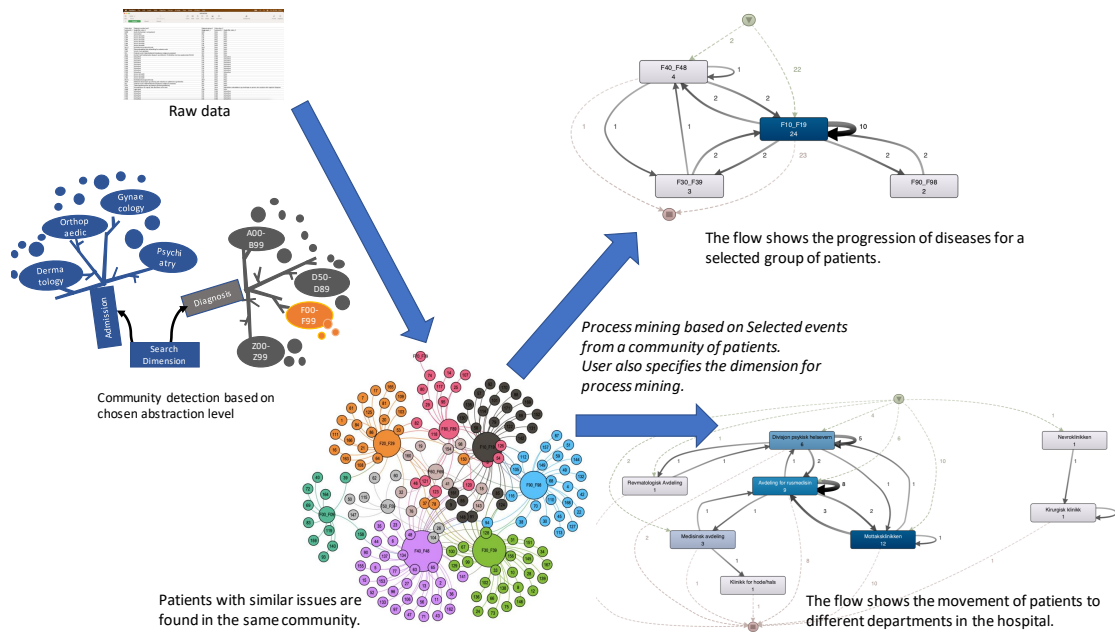


Figure 4: Integration of community detection with model based slicing for contextual process mining

the selection of event types e.g., diagnosis, admission, etc. We enrich the model based slicing approach presented in [7] with a two step approach where in the first step we contextualize the dataset with patients from specific community and in the next step we provide the contextual information to a process mining tool. Figure 4 shows the approach proposed in the paper. The raw data includes patient diagnosis and admission related information. We include a variety of hierarchical search dimensions (e.g., diseases, admissions) for the data analysis tasks. While applying community detection over patient records, we have the opportunity to specify which search dimension we would like to explore. In this example, we have specified a specific group of diagnosis code range ($F00 - F99$). The output of community detection algorithm is shown on the bottom left side of the figure. The communities can be further analyzed by means of a process mining technique. While applying process mining technique, we have the opportunity to specify which dimension we wish to explore. This selection is taken into account to prepare the event log for the process mining technique. The output of process mining technique is shown on the right side of the figure. In this example, we have shown flows for two different dimensions: the progression of diseases of a particular group of patients (top right side of the figure) and patient admission flow in various departments (bottom right side of the figure). The proposed approach is supported by a prototype implementation which extends the approach presented in [7].

4. Related work

van der Aalst presented four different analysis perspectives for process mining in [6] which include control-flow perspective, organizational perspective, case perspective, and time perspective. These perspectives are useful to understand the ordering of activities, the roles of resources, the attributes related to a particular case, and the frequency and timing of events. Although these perspectives can be used to derive useful insight by analyzing event logs from different points of view they lack an abstraction mechanism which will allow health professionals to both mine relevant information from highly discipline specific data sources and also to process event data from often highly individualistic patient pathways in order to discover common pathways.

Bistarelli et al. presented a prototype tool called PrOnto in [20] which can discover business processes from an event log and classify them w.r.t a business ontology. The tool takes an event log file as input and produces an UML based activity diagram in XML format. The aim of the approach is to raise the level of abstraction in process mining by utilizing business ontologies. They proposed an ontology representing the hierarchy of resources. In their approach, resources are the actors of the activities. Since the resources are given an ontological hierarchy, it is possible to define which level of abstraction will be used for process mining. They proposed to use integer numbers to define the level of abstraction. Defining a high level of abstraction would merge several activities being performed by all the actors that belong to the high level classification of resources. Our work is different from the approach presented in [20] in a sense that we proposed to use dimensional models and ontologies to classify event logs. Our approach is more flexible, since it is possible to be more specific in one portion of the process model while being more generic in another portion of the process model. The idea of combining dimensional modeling with ontologies is novel in this paper. We have shown that the mining process includes some preprocessing steps. In these preprocessing steps the user specifies the context (using dimensional modeling and ontologies) to define a patient group and also specifies the level of abstraction (using dimensional modeling and ontologies) that will be used to visualize the process mining output.

In [21], the authors discussed the application of process mining in healthcare and provided an overview of frequently asked questions by medical professionals in process mining projects. The questions reflect the medical professionals' interest both in learning common pathways of different patient groups, to determine their compliance with internal and external clinical guidelines, and also in gathering information about the throughput times for treating patients. The authors pointed out the need for accumulating data from different data sources and they claimed this to be a major challenge in healthcare. In the conclusion of the paper the authors suggested that ontologies can be used for defining appropriate scope and for identifying the cases from different data sources. They urged the exploitation of ontology-based process mining approaches in the healthcare domain.

In [22], the authors discussed the necessity of relating elements in event logs with their semantic concepts. By linking event logs with the concepts from an ontology they presented a process mining approach that performs concept-based analysis. The idea of using semantics makes it possible to automatically reason or infer relationships of concepts that are related. They distinguished between the application of process mining in two different levels: instance level and conceptual level. They illustrated their ideas with an example process model to

repair telephones in a company. That process model included three different ontologies: a Task ontology, a Role ontology and a Performer ontology. The idea of using an ontology for process mining presented in [22] is very similar to our approach. The idea of filtering based on ontological concepts and the idea of grouping nodes by a high level ontological concept is similar. However, in our approach we emphasize the benefits of an ontology based process mining for the healthcare domain. While in [22] the authors implemented their technique in ProM, our approach is more general as it offers a pre-processing step where we filter events for patients from a tightly-knit group using data mining technique and import the filtered events to *Fluxicon Disco* process mining tool.

A K-means clustering algorithm was presented in [23] where the authors enhanced the traditional K-means algorithm by means of a semantic model. The approach is similar to our community detection analysis where we utilize an abstract layer which gives semantics for our model. In our approach we focused on the usability of incorporating a flexible abstraction layer for healthcare data analysis. Rosvall et al. in [4] provided a list of approaches for community detection and briefly presented the potential of utilizing network abstraction for understanding an air traffic system. Again, they did not present any example of abstraction which can be applied in healthcare data to enhance clustering and its visualization.

Machine learning techniques have been employed in a variety of healthcare studies such as diagnostic code assignment [24], patient representation [25], etc. However, the potential for these machine learning algorithms and their integration with process mining techniques for analyzing healthcare information needs to be exploited. Healthcare data, due to its complex nature, can be modeled as a heterogeneous network. Representation learning methods [26] can be used to analyze the community structure of healthcare information represented as heterogeneous networks. According to [27] representation learning methods on heterogeneous networks could be divided into three categories: Path based, Semantic unit based and Other methods. Similar to node2vec [28] for homogeneous node embedding methods that preserves the random-walk probabilities in the feature space, for heterogeneous networks, one might try to preserve metapath probabilities, which is a more sophisticated variation of path. MetaPath2vec [29], MetaGraph2vec [30] fall into the Path based category which use path-based random-walks and a heterogeneous skip-gram model to learn node representation vectors. The basic principle of graph representation algorithms is to preserve the relationship and structural properties of network in a low dimensional vector space which therefore can essentially be used for extracting information about similarity of nodes from a network. Semantic unit based algorithms such as [31] define particular semantic units by means of capturing semantic information in the embedding space. Du et al. [32] presented an algorithm for network representation learning based on a graph partitioning strategy where a heterogeneous network is partitioned into homogeneous and bipartite subnetworks and the projective relations hidden in bipartite subnetworks are extracted by learning the projective embedding vectors. Although these machine learning based approaches are relevant for studying healthcare information, they need to be adapted to support the analysis of healthcare information which includes large healthcare ontologies, and temporal aspects.

5. Conclusion and Future Work

This paper proposes a model-based approach for determining patient pathways based on contextual process mining. We incorporate visualization techniques for filtering and analyzing patient records. Setting the context for process mining using community detection together with dimensional analysis enriched with healthcare ontologies is novel in this paper. The approach we presented here uses a flexible abstraction layer which can be tailored to meet the needs of the user. We envision a healthcare information system that supports an easy-to-use tool based on a graphical query language that permits a variety of healthcare professionals to investigate data from their personal and specialized perspective. In future we plan to develop methods and tools incorporating state-of-the art artificial intelligence techniques to apply our approach to a large healthcare dataset.

Acknowledgement

This research was supported by the Research Council of Norway with funding to INTROMAT (INtroducing personalized TReatment Of Mental health problems using Adaptive Technology), project number 259293.

References

- [1] E. Rojas, J. Munoz-Gama, M. Sepúlveda, D. Capurro, Process mining in healthcare: A literature review, *Journal of Biomedical Informatics* 61 (2016) 224 – 236.
- [2] R. A. Haraty, M. Dimishkieh, M. Masud, An enhanced k -means clustering algorithm for pattern discovery in healthcare data, *Int. J. Distributed Sens. Networks* 11 (2015) 615740:1–615740:11.
- [3] M. Liao, Y. Li, F. Kianifard, E. Obi, S. Arcona, Cluster analysis and its application to healthcare claims data: a study of end-stage renal disease patients who initiated hemodialysis, *BMC Nephrology* 17 (2016) 25. doi:10.1186/s12882-016-0238-2.
- [4] M. Rosvall, J. Delvenne, M. T. Schaub, R. Lambiotte, Different approaches to community detection, *CoRR abs/1712.06468* (2017). arXiv:1712.06468.
- [5] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment* 2008 (2008) P10008.
- [6] W. M. P. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, 1st ed., Springer Publishing Company, Incorporated, 2011.
- [7] F. Rabbi, Y. Lamo, W. MacCaull, A model based slicing technique for process mining healthcare information, in: Ö. Babur, J. Denil, B. Vogel-Heuser (Eds.), *Systems Modelling and Management - First International Conference, ICSMM 2020, Bergen, Norway, June 25-26, 2020, Proceedings*, volume 1262 of *Communications in Computer and Information Science*, Springer, 2020, pp. 73–81.
- [8] W. H. Organization., *ICD-10 : international statistical classification of diseases and related*

health problems / World Health Organization, 10th revision, 2nd ed. ed., World Health Organization Geneva, 2004.

- [9] O. Bodenreider, R. Cornet, D. J. Vreeman, Recent developments in clinical terminologies - snomed ct, loinc, and rxnorm., *Yearbook of medical informatics* 27 (2018) 129–139.
- [10] A. Partington, M. Wynn, S. Suriadi, C. Ouyang, J. Karnon, Process mining for clinical processes: A comparative analysis of four australian hospitals, *ACM Trans. Manage. Inf. Syst.* 5 (2015) 19:1–19:18.
- [11] D. Pyle, *Data Preparation for Data Mining*, 1st ed., Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.
- [12] C. Günther, A. Rozinat, Disco: discover your processes, in: N. Lohmann, S. Moser (Eds.), *Proceedings of the Demonstration Track of the 10th International Conference on Business Process Management (BPM 2012)*, CEUR Workshop Proceedings, CEUR-WS.org, 2012, pp. 40–44. *Demonstration Track of the 10th International Conference on Business Process Management, BPM Demos 2012 ; Conference date: 04-09-2012 Through 04-09-2012.*
- [13] B. F. van Dongen, A. K. A. de Medeiros, H. M. W. Verbeek, A. J. M. M. Weijters, W. M. P. van der Aalst, The prom framework: A new era in process mining tool support, in: G. Ciardo, P. Darondeau (Eds.), *Applications and Theory of Petri Nets 2005*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 444–454.
- [14] F. Rabbi, J. D. Wake, T. Nordgreen, Reusable data visualization patterns for clinical practice, in: Ö. Babur, J. Denil, B. Vogel-Heuser (Eds.), *Systems Modelling and Management - First International Conference, ICSMM 2020, Bergen, Norway, June 25-26, 2020, Proceedings*, volume 1262 of *Communications in Computer and Information Science*, Springer, 2020, pp. 55–72. URL: https://doi.org/10.1007/978-3-030-58167-1_5. doi:10.1007/978-3-030-58167-1_5.
- [15] U. Ozomaro, C. Wahlestedt, C. B. Nemeroff, Personalized medicine in psychiatry: problems and promises, *BMC medicine* 11 (2013) 132.
- [16] J. MacQueen, Some methods for classification and analysis of multivariate observations, in: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, University of California Press, Berkeley, Calif., 1967, pp. 281–297.
- [17] S. Li, L. Jiang, X. Wu, W. Han, D. Zhao, Z. Wang, A weighted network community detection algorithm based on deep learning, *Applied Mathematics and Computation* 401 (2021) 126012.
- [18] Y. Jia, Q. Zhang, W. Zhang, X. Wang, Communitygan: Community detection with generative adversarial nets, in: *The World Wide Web Conference, 2019*, pp. 784–794.
- [19] D. He, W. Song, D. Jin, Z. Feng, Y. Huang, An end-to-end community detection model: Integrating lda into markov random field via factor graph, in: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, International Joint Conferences on Artificial Intelligence Organization, 2019*, pp. 5730–5736.
- [20] S. Bistarelli, T. D. Noia, M. Mongiello, F. Nocera, Pronto: an ontology driven business process mining tool, *Procedia Computer Science* 112 (2017) 306 – 315.
- [21] R. S. Mans, W. M. P. van der Aalst, R. J. B. Vanwersch, A. J. Moleman, Process mining in healthcare: Data challenges when answering frequently posed questions, in: R. Lenz, S. Miksch, M. Peleg, M. Reichert, D. Riaño, A. ten Teije (Eds.), *Process Support and Knowledge Representation in Health Care*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 140–153.

- [22] A. K. A. de Medeiros, W. M. P. van der Aalst, Process mining towards semantics, in: *Advances in Web Semantics I*, volume 4891 of *Lecture Notes in Computer Science*, Springer, 2009, pp. 35–80.
- [23] Z. Liu, J. Bao, F. Ding, An improved k-means clustering algorithm based on semantic model, in: *Proceedings of the International Conference on Information Technology and Electrical Engineering 2018, ICITEE '18*, Association for Computing Machinery, New York, NY, USA, 2018.
- [24] V. Mayya, S. K. S., G. S. Krishnan, T. Gangavarapu, Multi-channel, convolutional attention based neural model for automated diagnostic coding of unstructured patient discharge summaries, *Future Generation Computer Systems* 118 (2021) 374–391.
- [25] X. Zhang, B. Qian, Y. Li, Y. Liu, X. Chen, C. Guan, C. Li, Learning robust patient representations from multi-modal electronic health records: a supervised deep learning approach, in: *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)*, SIAM, 2021, pp. 585–593.
- [26] W. L. Hamilton, R. Ying, J. Leskovec, Representation learning on graphs: Methods and applications, 2017. Cite arxiv:1709.05584Comment: Published in the IEEE Data Engineering Bulletin, September 2017; version with minor corrections.
- [27] Y. Xie, B. Yu, S. Lv, C. Zhang, G. Wang, M. Gong, A survey on heterogeneous network representation learning, *Pattern Recognition* 116 (2021) 107936.
- [28] A. Grover, J. Leskovec, Node2vec: Scalable feature learning for networks, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, ACM, New York, NY, USA, 2016, p. 855–864.
- [29] Y. Dong, N. V. Chawla, A. Swami, Metapath2vec: Scalable representation learning for heterogeneous networks, in: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '17*, Association for Computing Machinery, New York, NY, USA, 2017, p. 135–144.
- [30] D. Zhang, J. Yin, X. Zhu, C. Zhang, Metagraph2vec: Complex semantic path augmented heterogeneous network embedding, in: D. Phung, V. S. Tseng, G. I. Webb, B. Ho, M. Ganji, L. Rashidi (Eds.), *Advances in Knowledge Discovery and Data Mining*, Springer International Publishing, Cham, 2018, pp. 196–208.
- [31] G. Fu, B. Yuan, Q. Duan, X. Yao, Representation learning for heterogeneous information networks via embedding events, in: *International Conference on Neural Information Processing*, Springer, 2019, pp. 327–339.
- [32] W. Du, S. Yu, M. Yang, Q. Qu, J. Zhu, Gpsp: Graph partition and space projection based approach for heterogeneous network embedding, in: *WWW (Companion Volume)*, 2018, pp. 59–60.