

Multidimensional Process Mining with PMCube Explorer

Thomas Vogelgesang and H.-Jürgen Appelrath

Department of Computer Science
University of Oldenburg, Germany
thomas.vogelgesang@uni-oldenburg.de

Abstract. Process mining techniques allow process analysts to generate process models from recorded event logs. Typically, process mining considers the event log as a whole and creates a single model reflecting its behavior. However, the process may be influenced by several characteristics of the process instances, e.g., by the individual characteristics of a patient in the healthcare domain like age and sex. This leads to a wide range of process variations which can end up in complex and confusing models, blurring the behavior of specific process variants. Multidimensional process mining (MPM) aims to overcome this limitation by the notion of data cubes, spreading the data over multiple cells, each representing a group of cases with similar characteristics. This allows for the creation of separated process models for a homogenous set of cases. In this paper, we introduce PMCube Explorer, a novel tool for MPM, that allows for the analysis of a process from various views. It enables the analyst to specify OLAP queries to extract multiple cells from the data warehouse. Each cell contains a subset of event data which are mined separately to discover independent process models. To deal with the potentially high amount of resulting models, our tool provides some distinctive features like the visualization of model differences or the consolidation of multiple process models. We applied our tool in a case study to analyze the perioperative processes in a large German hospital.

1 Introduction

Process mining is a set of techniques that allow analysts to generate process models from event logs which contain event data recorded during the execution of the process. However, the behavior of a process is often influenced by several characteristics of the executed process instance. For instance, healthcare processes have to consider age, sex, and allergies of the patients. This leads to a wide range of process variations which can result in big and complex process models when analyzing them with process mining techniques. The notion of multidimensional process mining (MPM) aims to solve this problem by partitioning the underlying

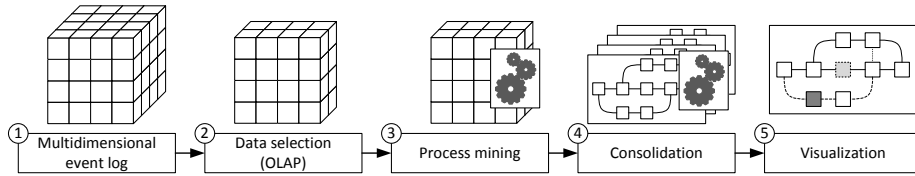


Fig. 1. Basic concept of PMCube

event log into subsets that consist of cases with homogenous features. These subsets (or sublogs) are mined separately to discover independent process models, each focusing on a limited feature combination of the cases. MPM adopts the concepts of OLAP and data cubes, that are commonly used in data warehouses (DWH), to the field of process mining. The intention is to partition and filter the event log in a dynamic and flexible way in order to provide customized views on the process.

In this paper, we demonstrate the *PMCube Explorer*, a novel tool for MPM. Section 2 briefly introduces the underlying concepts. In Section 3, we show the main features of our tool. A case study using the tool is discussed in Section 4. Finally, we briefly present the screencast of this demonstration in Section 5.

2 The PMCube concept

Figure 1 illustrates the PMCube approach, which is the underlying concept of our tool. The multidimensional event log (MEL) is a DWH that stores the data from the event log as a multidimensional data cube. In contrast to Event Cube [4], another approach for MPM, the cells of the MEL do not contain precomputed dependency measures, but raw event data forming a sublog of the event log. This is similar to Process Cubes [5], another approach for MPM. However, PMCube organizes event attributes and case attributes on different levels. While a Process Cube stores sets of events in each cell, the cells of PMCube’s MEL contain cases on the first level. On the second level, each case owns a sequence of events (so-called trace), forming a distinct cube. This structure of nested cubes allows analysts to define complex filtering and aggregation operations using OLAP queries to extract highly customized sublogs from the MEL, e.g., only selecting cases having events that in average exceed a given cost limit.

The result of an OLAP query is a set of independent sublogs. Process discovery techniques are applied to each sublog to discover a process model. Depending on the query, this may result in a high number of potentially complex process models which makes it hard for the analyst to interpret the results. To cope with this, PMCube provides an optional step of consolidation. It aims to reduce the number of process models by an automatic preselection of the most relevant process models. One consolidation approach is to cluster process models reflecting similar behavior and to select one representative process model per cluster. It is based on the heuristic that major differences between process models are more

relevant to the analyst than minor variations. After the consolidation, the results are visualized for interpretation. PMCube can arrange all process models side by side in a matrix to provide a general overview of the models. Alternatively, PMCube can also calculate the differences between two models and highlight them in a merged model.

3 Implementation

The PMCube Explorer is a prototypical implementation of the PMCube concept. It is written in C# using the Microsoft .NET framework. The MEL is stored in an external, relational database like Oracle or Microsoft SQL Server in an advanced snowflake schema reflecting the two distinct levels for cases and events. The analyst can query the MEL via a graphical user interface (GUI) to create a customized view of the data. For this purpose, the GUI provides multiple options to filter and aggregate the data cubes on both the case and the event level. For each cell defined by the OLAP query, a separated SQL query is created, which is sent to the MEL. While executing the SQL queries, the multidimensional data of the data cube is implicitly flattened into a table where each line represents an event of the sublog. Because the query results reflect the commonly used event log structure, arbitrary process discovery algorithms can be used without any adaptations to discover the process models. The resulting process models can be visualized side by side in a matrix or as a single model. In contrast to other MPM tools like Process Mining Cubes (PMC) [1], PMCube Explorer allows for the selection of two models to automatically visualize their difference. Additionally, it provides the novel process model consolidation, e.g, the filtering of process models by specific model features (like the existence of particular events) and the clustering-based consolidation, as a unique feature. To calculate the models' fitness, it is possible to replay the sublogs on the discovered process models or on an external reference model. The process models can be enhanced with statistical information like average or median duration between two consecutive activities.

The PMCube Explorer is highly extensible. All algorithms for process discovery, conformance checking, difference view calculation, and consolidation, as well as the process model languages (data structures and view models) and the database connectors are provided as plug-ins and loaded during run-time. Currently, PMCube Explorer provides plug-ins for Inductive Miner – infrequent [3], Flexible Heuristics Miner [6], and Fuzzy Miner [2] for process discovery.

Figure 2 presents some screenshots of PMCube Explorer. They show the preview of the resulting cells while creating the OLAP query (1), the dialog for selecting the dimension that should be used for slicing (2), the matrix view (3), and the time perspective dialog showing the distribution of waiting times (4).

4 Case study

We conducted a case study where we used the PMCube Explorer to analyze healthcare processes of a university hospital as a center of maximum care in

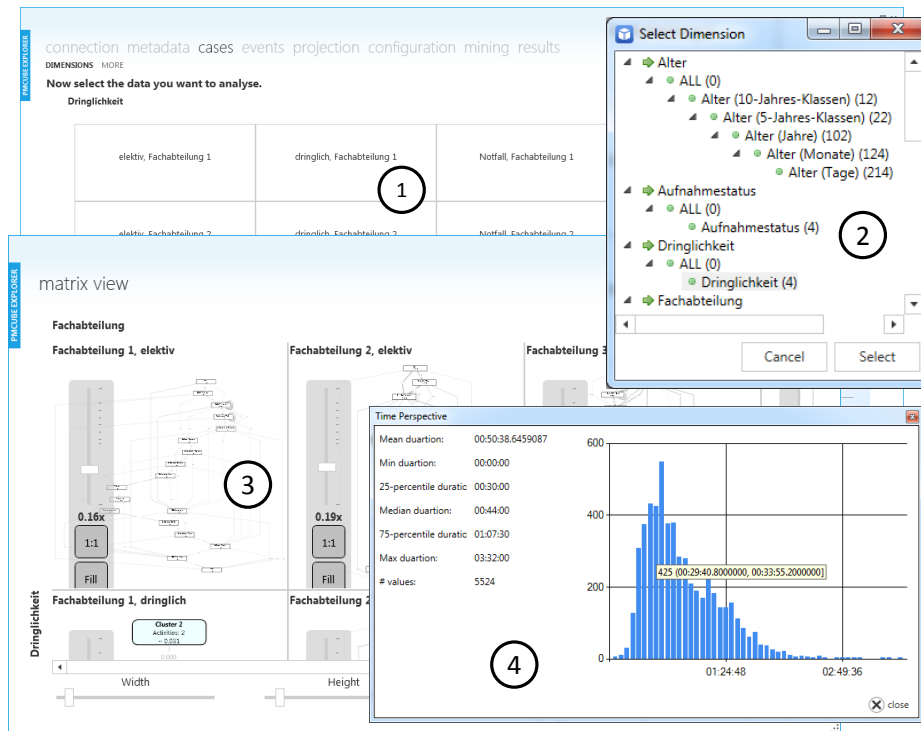


Fig. 2. Screenshots of PMCube Explorer

Germany. We started the evaluation study after the approval of the ethical committee of the Justus Liebig University (ethical review committee of the Faculty of Human Medicine at the Justus Liebig University Gießen, chairman Prof. Dr. Tillmanns, vote number 261/14) with an anonymized data set. This data set covers a random sample of 16,280 surgical interventions of four medical departments in 2012 and 2013 with a total of 388,395 events. We focused on the perioperative process, which comprises all activities in the periphery of surgical interventions, especially activities related to anesthesia. The event data was extracted from several clinical information systems and anonymized by the hospital IT, before we integrated it into the multidimensional structure of the MEL.

We applied multiple OLAP queries in an explorative way to analyze the processes from various points of view. We discussed the discovered models with a medical expert, who is familiar with the perioperative processes of that hospital. The case study showed that MPM provides a dynamic and flexible way to analyze processes from different views. Queries can be easily adjusted, which allows for the explorative analysis of the processes. However, the case study revealed that MPM can become quite complex and confusing, especially when comparing many process models. Although the consolidation and the different visualization techniques showed to be a helpful tool during analysis, they need to be

improved and complemented by more sophisticated techniques to deal with the high complexity of results. This should be tackled by future work.

5 Demonstration

A screencast that demonstrates the usage of the PMCube Explorer is available on the web (<http://youtu.be/CTXyIZp2BJw>). It gives a walk-through of an example process mining analysis, conducted on the data of the case study described in Section 4. Starting with the creation of an OLAP query, it presents the main features of the tool, like the clustering-based consolidation of process models, the matrix visualization, and the automatic visualization of differences between process models. Furthermore, it shows how to switch the view of the data cube.

Acknowledgments. The authors would like to thank all contributors to the PMCube Explorer and especially Rainer Röhrig, Lena Niehoff, Raphael W. Majeed, and Christian Katzer for their support during the case study.

References

1. Alfredo Bolt and Wil M.P. van der Aalst. Multidimensional Process Mining Using Process Cubes. In Khaled Gaaloul, Rainer Schmidt, Selmin Nurcan, Sergio Guerreiro, and Qin Ma, editors, *Enterprise, Business-Process and Information Systems Modeling*, volume 214 of *Lecture Notes in Business Information Processing*, pages 102–116. Springer International Publishing, 2015.
2. Christian W. Günther and Wil M. P. van der Aalst. Fuzzy mining: adaptive process simplification based on multi-perspective metrics. In *Proceedings of the 5th international conference on Business process management, BPM'07*, pages 328–343, Berlin, Heidelberg, 2007. Springer-Verlag.
3. Sander J.J. Leemans, Dirk Fahland, and Wil M.P. van der Aalst. Discovering Block-Structured Process Models from Event Logs Containing Infrequent Behaviour. In Niels Lohmann, Minseok Song, and Petia Wohed, editors, *Business Process Management Workshops*, volume 171 of *Lecture Notes in Business Information Processing*, pages 66–78. Springer International Publishing, 2014.
4. J. T. S. Ribeiro and A. J. M. M. Weijters. Event cube: another perspective on business processes. In *Proceedings of the 2011th Confederated international conference on On the move to meaningful internet systems - Volume Part I (OTM'11)*, pages 274–283, Berlin, Heidelberg, 2011. Springer-Verlag.
5. Wil M. P. van der Aalst. Process Cubes: Slicing, Dicing, Rolling Up and Drilling Down Event Data for Process Mining. In Minseok Song, MoeThandar Wynn, and Jianxun Liu, editors, *Asia Pacific Business Process Management*, volume 159 of *Lecture Notes in Business Information Processing*, pages 1–22. Springer International Publishing, 2013.
6. A. J. M. M. Weijters and J. T. S. Ribeiro. Flexible heuristics miner (FHM). Technical report, Technische Universiteit Eindhoven, 2011.