

# Multidimensionales Process-Mining für die Analyse medizinischer Versorgungsprozesse

Dissertation zur Erlangung des Grades eines  
Doktors der Naturwissenschaften

vorgelegt von

**Thomas Vogelgesang, M.Sc.**

Gutachter:

**Prof. Dr. Dr. h.c. H.-Jürgen Appelrath**

**Univ.-Prof. Dr. Stefanie Rinderle-Ma**

Tag der Disputation: 30. November 2015

## **Zusammenfassung**

Process-Mining ermöglicht u.a. die automatische Extraktion von Prozessmodellen aus sogenannten Eventlogs, die während einer Prozessausführung aufgetretene Events enthalten. Neben klassischen Geschäftsprozessen können damit auch Prozesse in der medizinischen Versorgung analysiert werden. Eine besondere Herausforderung stellt hierbei vor allem die Individualität der Patienten dar, die den konkreten Versorgungsprozess stark beeinflusst. So hängt der Behandlungsverlauf von vielen Faktoren wie z.B. Alter, Geschlecht und Nebenerkrankungen – und möglicherweise anderen, eher intransparenten Faktoren – des Patienten ab. Etablierte Process-Mining-Verfahren analysieren die Eventlogs aber als Ganzes, sodass die Versorgung aller Patienten in einem gemeinsamen Modell dargestellt wird, in dem die Einflüsse der Faktoren „verschwimmen“ bzw. unberücksichtigt bleiben.

Um jedoch den Einfluss dieser Faktoren gezielt untersuchen zu können, macht es Sinn Patienten mit ähnlichen Eigenschaften zu gruppieren und für jede Gruppe ein separates, gruppenspezifisches Prozessmodell zu extrahieren. Diese Modelle können dann einander gegenübergestellt und miteinander verglichen werden. Da die Aufteilung der Patienten in verschiedenen Gruppen vor allem von der zu untersuchenden Fragestellung abhängt, wird ein Ansatz benötigt, der eine einfache und flexible Definition der Patientengruppen erlaubt.

Im Rahmen dieser Arbeit wird ein Ansatz vorgestellt, der die Eigenschaften der Patienten als Dimensionen interpretiert, welche einen multidimensionalen Datenraum aufspannen. Mit einem solchen multidimensionalen Process-Mining lassen sich die Eventlogs auf einen Datenwürfel abbilden, der mittels OLAP-Operatoren manipuliert werden kann. Die dadurch definierten Zellen des Würfels repräsentieren jeweils eine Patientengruppe, die separat mittels Process-Mining analysiert werden, wobei für jede Gruppe ein eigenes Prozessmodell erstellt wird. In einem optionalen Schritt der Konsolidierung kann anschließend eine automatische Vorauswahl relevanter Modelle erfolgen, die schließlich visualisiert und miteinander verglichen werden können.

Durch eine prototypische Implementierung wird die technische Umsetzung der entwickelten Konzepte gezeigt. Die Anwendbarkeit des Ansatzes zur Analyse medizinischer Versorgungsprozesse wird im Rahmen einer Fallstudie in der Versorgungsforschung demonstriert. Durch verschiedene Messreihen wird zudem gezeigt, dass die entwickelten Konzepte deutlich leistungsfähiger sind als der State-of-the-Art für multidimensionales Process-Mining und mit zunehmender Anzahl der Events gut skalieren.

### **Abstract**

Process mining allows for the automatic discovery of process models from so-called event logs which capture events that occurred during process execution. Besides traditional business processes, it can also be used to analyze medical treatment processes. Especially the individual characteristics of the patients turn out to be very challenging, because they significantly influence the process behavior. The treatment may be influenced by many features like age, sex, comorbidity and other – possibly nontransparent – features of a patient. However, established process mining techniques consider the entire event log and create a single process model reflecting the treatment of all patients which may blur or even hide the influence of the individual characteristics.

To analyze the influence of these characteristics, it is desirable to group patients with similar characteristics and to discover a separated process model for each group of patients. Then, these models can be compared. As the grouping of the patients may depend on the underlying question of the analysis, an approach is required to define the grouping criteria in a simple and flexible way.

This work presents a novel approach that considers the patients' characteristics as dimensions forming a multidimensional data space. This so-called multidimensional process mining allows to map the event logs to a data cube which can be queried using OLAP operators. This way, a set of cells can be defined, each representing a group of patients. For each group, a separated process model is discovered. An optional step of consolidation allows for an automatic pre-selection of relevant process models. Finally, the models can be visualized and compared.

This approach is implemented as a prototype to show its feasibility. The applicability of the approach to the healthcare domain is demonstrated during a case study in health services research. Additional experiments show that this concept provides a significantly better performance than the state-of-the-art approach for multidimensional process mining and is scalable with the number of events.