Data-Based Process Variant Analysis

Jonas Cremerius
Hasso Plattner Institute
University of Potsdam
jonas.cremerius@hpi.de

Hendrik Patzlaff
Hasso Plattner Institute
University of Potsdam
hendrik.patzlaff@student.hpi.de

Vincent Rahn
Hasso Plattner Institute
University of Potsdam
vincent.rahn@student.hpi.de

Henrik Leopold
Kuhne Logistics
University
henrik.leopold@the-klu.org

Abstract

Processes in healthcare are complex and data-intensive. Process mining uses data recorded during process execution to obtain an understanding of the actual execution of a process. Due to the complexity of healthcare processes, it is useful to consider and analyse the process execution of certain cohorts, such as old and young patients, separately. While such analysis is facilitated by process variant analysis techniques, existing approaches for process variant analysis only consider a comparison based on the control flow and performance perspectives. Given the large amount of event data attributes available in healthcare settings, we propose the first data-based process variant analysis approach. Our approach allows comparing process variants based on differences in event data attributes by building on statistical tests. We applied our approach on the MIMIC-IV real-world data set on hospitalizations in the US, where we demonstrate that the approach is feasible and can actually provide relevant medical insights.

1. Introduction

Health insurance companies, governments, and healthcare reforms put more and more pressure on hospitals to work as efficiently as possible as a constant care demand increment is taking place [Mans et al., 2008]. Over the past years, healthcare organizations have shifted their focus on ways to streamline their processes to deliver high-quality care while reducing costs at the same time [Anyanwu et al., 2003]. Processes in healthcare are highly complex and distributed across many disciplines. Thus, it is often unknown what exactly happens in a treatment process for a group of patients, which in the medical domain is called a patient cohort [Mans et al., 2008].

A promising way to obtain insights into treatment processes is the use of process mining. Process mining provides a set of techniques that aim to understand the actual execution of processes. It builds on data extracted from information systems, so-called event logs, that has been recorded in the context of the process execution and, therefore, reflects what really happened. Among others, process mining can be used to discover a visual representation of the process, detect conformance issues, or predict process outcomes [van der Aalst, 2016]. By building on event logs extracted from health information systems, process mining can also be used to analyse treatment processes of patients.

One aspect that is particularly relevant in the context of analysing treatment processes is to compare different patient cohorts, such as young and old patients. A comparison of different subsets of an event log that differ with respect to certain characteristics is generally referred to as *process variant analysis* [Taymouri et al., 2021]. In principle, process variant analysis can enable health experts to understand how the process execution of different cohorts differ and, in this way, identify specific improvement opportunities for one or more considered cohorts [Rojas et al., 2016].

However, existing techniques for process variant analysis in the context of healthcare processes are limited to the control flow and performance dimension [Rojas et al., 2016, Taymouri et al., 2021]. This means that event data attributes, such as laboratory measurements, are not considered for the analysis. This is rather surprising given that such event data attributes can differ considerably among patient cohorts. For example, hypertension, which is characterized by increased blood pressure, is more prevalent in older patients [Lloyd-Jones et al., 2005]. Thus, the activity of measuring blood pressure in process executions relating to older patients will likely result in a higher blood pressure measurements. With that information at hand, medical experts can understand patient cohorts better, allowing them to derive cohort specific and process oriented treatment insights based on data.

Due to the lack of event data attribute comparison

in process variant analysis, such differences are not yet possible to identify. Thus, potentially important differences in patient cohorts cannot be detected. In this paper, we therefore propose an approach for *data-based process variant analysis*, which enables further insights into process variants by taking event data attributes from the event logs explicitly into account.

The remainder of this work is organized as follows. Section 2 shows an overview of related work and Section 3 illustrates an example describing the problem. In Section 4, we present our approach for data-based process variant analysis. With the help of an interactive graphical tool, we evaluate our approach on the MIMIC-IV real-world data set on hospitalizations in Section 5. We discuss the approach and its limitations in Section 6 before we conclude by discussing future research in Section 7.

2. Related Work

The analysis of process variants has been approached from different perspectives in the literature. The survey conducted in [Taymouri et al., 2021] shows, that process variant analysis is a cluttered field. Approaches for process variant analysis can be characterized by four factors: the required input data, the provided output, the type of analysis conducted, and the algorithms used. From the perspective of the *type of analysis conducted*, three categories can be distinguished.

Approaches from the first category look at process variants from a control flow perspective [Buijs et al., 2012, Swinnen et al., 2012]. This enables analysts to see how process variants differ with respect to the performed activities and the execution order. Approaches from the second category take performance data into account [Cuzzocrea et al., 2017, Gulden, 2017]. This allows for a more detailed analysis, where execution times of activities and other time-related measures are compared. Approaches from the third category consider data, which can be activity data or resource information [Low et al., 2017, Nguyen et al., 2018]. This is interesting for event logs containing event data attributes that can be leveraged for a more detailed comparison.

In this paper, we focus on the third category and propose a novel approach for data-based process variant analysis. While there already exist some approaches belonging to this category [Low et al., 2017, Nguyen et al., 2018], these approaches do not consider additional information from event data attributes. The approach proposed in [Low et al., 2017] uses data on resources to visualize information on resource allocation and activity scheduling. While the primary goal of this work is not process variant analysis but business process improvement, it is related due to the focus on resource data and the overall goal of providing insights for the purpose of process improvement. The approach from [Nguyen et al., 2018] uses event data attributes and statistical tests for investigating process variants. They propose perspective graphs as a multi-perspective view on process variants. Perspective graphs represent any entity referring to an attribute of the event log, where the connecting arcs represent an arbitrary relation between the nodes. Two perspective graphs are compared by generating a comparison graph, where statistic significant differences in frequency are represented in a matrix. The application of this approach focuses on categorical event data attributes. Their approach allows the comparison of detailed value changes, such that a handover from one country to another country is performed more frequently.

To summarize, existing approaches for process variant analysis hardly consider event data attributes. The recent review in [Taymouri et al., 2021] even explicitly identifies data- and resource-aware process variant analysis as a research gap, as existing work in the field has focused on control flow and performance data. Against this background, we propose an approach that takes differences in event data attributes associated with activities between process variants into account. By considering categorical as well as continuous event data attributes, our approach provides the required flexibility to obtain detailed insights into the differences between process variants.

In the next section, we take a look at the specific challenges that we need to tackle to accomplish this.

3. Problem Statement

To illustrate the problem of analysing process variants based on event data attributes, consider the event log depicted in Table 1. It shows three different cases related to three different individuals. We can see that the execution sequences of the activities are perfectly identical (i.e. A, B, C). However, the cases strongly differ with respect to the event data attributes. The individuals involved in the cases are of different age, have different heart rate ranges, and experienced different levels of pain.

From a medical perspective, the challenge is now to identify process variants that relate to potential patient cohorts. In practice, there might be numerous data attributes and, therefore, also numerous options to split the log based on available attributes. One possible split is depicted in Table 2 and Table 3 where the event log

Table 1. Example event log enriched with event data attributes.

Case	Age	Activity	Heart Rate	Pain
0	20y	A	67	low
		B	66	low
		C	-	low
1	80y	A	103	high
		B	100	medium
		C	90	low
2	23y	A	56	medium
		B	58	medium
		C	62	low

is split based on the case attribute age into younger ($\leq$ 60y) and older ($>$ 60y) patients.

Table 2. Event log split by subject age, including younger patients.

Case	Age	Activity	Heart Rate	Pain
0	20y	A	67	low
		B	66	low
		C	-	low
2	23y	A	56	medium
		B	58	medium
		C	62	low

Table 3. Event log split by subject age, including older patients.

Case	Age	Activity	Heart Rate	Pain
1	80y	A	103	high
		B	100	medium
		C	90	low

This paper tackles the problem that there is no approach available that allows to make statements regarding the actual value differences in process variants. As a result, we cannot answer whether heart rate or pain in older patients undergoing an activity is higher than in younger patients undergoing the same activity. This information is highly useful in healthcare, as it enables to make cohort specific and process oriented statements. For example, it could be identified, that younger patients are less likely to experience pain at the beginning of the process in activity A, which could represent triage in the emergency department. Thinking this further, this information could also improve the process by prioritizing older patients in the beginning of the process, as they are expected to experience more pain.

Another problem under investigation is that the changing behaviour of event data attributes cannot be compared yet. For example, we cannot say, whether the pain level or heart rate is changing throughout the process. It could be, that at some point in the process, these measurements increase for younger patients and decrease for older patients. In healthcare, this could happen due to different treatments for younger and older patients.

In the next section, we present our approach to identify differences in event data attributes for process variants.

4. Approach for Data-based Process Variant Analysis

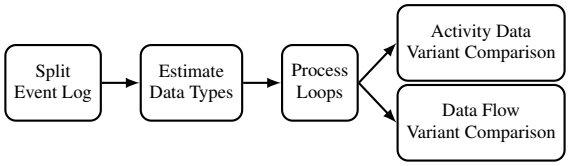

Figure 1. Overview of Proposed Approach

In this section, we introduce our approach for data-based process variant analysis. Figure 1 provides an overview of the proposed approach. It consists of five main steps. The first step is to split the event log into possible process variants. This split is expected to be done manually, for example, based on specific case or event data attributes. Then, we determine the data types of the event data attributes. This is important since categorical and continuous event data attributes need to be handled differently. Next, we pre-process the loops in the variants. This is important since loops may lead to repeated occurrences of activities and, therefore, also in event data attribute changes. Finally, we analyse the variants. We provide two specific mechanisms for this. Using the *activity data variant comparison*, we can identify differences between variants in terms of collected data values from activities. Using the *data flow variant comparison*, we can identify differences based on the changing flow of data values among activities. In the subsequent sections, we explain these steps in more detail.

4.1. Estimate Data Types

Our approach aims to compare the event data attributes of two process variants. From a mathematical point of view, these attributes can be either continuous or categorical variables. In Table 1, we observe two common measurements. Pain is a measurement describing how much pain the patient feels at the moment. It is a categorical value that can have the

values low, medium, or high. Heart rate, by contrast, is a continuous variable.

For our analysis, we need to differentiate between these two data types because they require different statistical tests. The separation in one of the two groups depends on the investigated data set. In statistics, it is often assumed that the statistical data types of variables are known. But in event logs retrieved from healthcare data, not all measurements might be well documented. Tagging event logs containing many measurements would require a lot of manual processing.

As shown by [Valera and Ghahramani, 2017, Cremerius and Weske, 2022], we can apply simple logic rules to distinguish between categorical and continuous variables. Thus, we apply a check that divides the number of distinct values by the total number of values and compares this fraction against a user-defined threshold. For example, if all values of an event data attribute are unique, the resulting fraction is 1, indicating a continuous attribute. If 4 of 100 values are unique, the resulting fraction is 0.04, which is likely to be categorical. It should be noted, that this can result in wrong classifications, depending on the threshold. Therefore, we see this as an automated guidance to identify the data types, where the user can change the threshold or edit the data type of single variables.

We perform the classification for the XES data types String, Date, Float, and Integer. We only consider String if all values of the variable contain numbers encoded as a string. If not, we consider it is as categorical. The data type Boolean is always categorical [XES, 2016].

4.2. Process Loops

Processes can contain loops, which means that within a single process execution one activity or transition occurs several times. Transitions considered are directly follows and eventually follows relations. As we compare activities and transitions in both variants, we need to define how to cope with loops.

If we do not require a fine-grained comparison of each occurrence of the activities, we can process each case by aggregating the values collected for each occurrence of one activity or transition. For example, we can calculate the mean, mode, minimum or maximum per case activity or transition. With this approach, however, we do not consider at what point in time an activity or transition happens within a trace.

If we require a fine-grained comparison, we propose to enumerate activities and transitions for both types of analysis. For example, the trace $< A, B, A, C, A, B >$ would become $< A_1, B_1, A_2, C_1, A_3, B_2 >$ for activity comparison. For data-flow comparison, we enumerate each transition, resulting in $(A, B)_1$, $(A, B)_2$, $(A, C)_1$, $(B, A)_1$, $(C, A)_1$ for directly follows relations. One can further differentiate between the occurrence of directly follows and eventually follows relations, which is dependent on the type of analysis.

4.3. Statistical Tests

To enable data-based process variant analysis, we require a measure of difference between two process variants. We need to quantify whether two samples differ on a significant level.

To accomplish this, we use statistical tests. They allow us to measure whether process variants differ with respect to their event data attributes and provide a measure of how two or multiple different populations behave. Choosing the statistical test depends on the data type, the distribution of the data, and the dependency between the samples [Parab and Bhalerao, 2010]. As described in Section 4.1, we differentiate between continuous and categorical variables. As we cannot make any assumptions about the distribution of the data, we need to employ so-called non-parametric tests. As for the sample comparison, we compare samples drawn from independent process variants, which means that we consider an unpaired sample setting.

For continuous data, we use the Mann-Whitney U test. It is a non-parametric test, which considers two samples that are independent of each other. The test results in a p-value, which describes how statistically independent both samples are. If the p-value is below a certain threshold, the difference between the samples is concluded to be statistically significant. Furthermore, we calculate the Rank-biserial correlation (RBC) between both samples. The RBC value is between -1 and 1, where high absolute numbers mean strong correlation, and therefore a strong difference between both samples [Mann and Whitney, 1947]. For example, if the RBC value is -1, all values in the second sample are lower than the values in the first sample. Thus, the RBC value expresses whether the values in one sample are higher or lower than in the other sample.

For categorical data, we use the Pearson Chi-Square Independence test. It evaluates if the given samples are independent by comparing the frequencies of each categorical variable present in both samples. The test results in a p-value. A small p-value shows the independence of both samples and, therefore, a difference in their distributions. Similar to the RBC value for the Mann-Whitney U test, the chi-square value indicates how much difference exists between both samples. Thus, the higher the chi-square value, the higher the difference in the categories [McHugh, 2013].

As described above, each statistical test results in a p-value indicating the statistical significance and a test-statistic indicating the degree of the difference. To evaluate whether both samples differ in a statistically significant way, we need to define a p-value threshold α, which is typically 0.05. Technically, statistical tests can be conducted on any sample size. However, to retrieve meaningful results, the sample size should be considered when interpreting the test results.

If multiple tests are conducted for the same samples, which happens if multiple event data attributes exist, α needs to be adjusted according to the number of tests performed. For this, we use the Bonferroni correction, which determines the p-value threshold α for statistical significance [Armstrong, 2014]. For example, if 10 tests are conducted on the same samples, α needs to be divided by 10, resulting in $\alpha = \frac{0.05}{10}$.

4.4. Activity Data Variant Comparison

Activity data variant comparison provides insights into whether a specific event data attribute (such as heart rate) differs between the two variants for a particular activity. Table 4 and Table 5 show the split of the event log by age, where variant 1 contains cases of young patients and variant 2 contains cases of old patients. Assume we would like to compare Activity A and consider the measurements heart rate and pain. Note that we, in comparison to Table 1, added additional cases to reach a sufficient sample size, which needs to be at least four for both variants to technically reach a statistical significant p-value.

In the example, we see that the heart rate of younger patients seems to be lower than the heart rate of older patients. Thus, laboratory values or other measurements collected from patients during their treatment process could exhibit statistical significant differences in different variants. Having that information, one could provide more precise process-oriented statements about different patient groups, such that older patients have a higher heart rate in the intensive care unit or during a procedure, instead of saying that this is the case in general.

To compare variant 1 and variant 2, we now pick all values for one measurement, such as the heart rate, and create two samples that can be compared. Based on the two samples for one laboratory value for variant 1 and variant 2, we then perform a statistical test between the two collected samples. As explained before, for continuous data, a Man-Whitney-U-Test is used, while for categorical data, a Chi-squared test is applied. Both tests result in a p-value and a second, test-specific measurement. The p-value shows, whether the measurement differs significantly between both variants.

In our example, the heart rate differs between variant 1 and variant 2, as the older patients have a considerably higher heart rate. Looking at Table 4 and Table 5, we have the values 56, 67, 45, 40 for variant 1 and 103, 100, 100, 102 for variant 2 as input for the statistical test. The resulting p-value is 0.0294 with an RBC value of 1.0, as all values are higher in variant 2. Therefore, we observe a difference in the heart rate values in the two process variants, which can be justified with statistical significance.

Afterwards, we only pick the measurements that show a statistically significant difference between both variants.

Table 4. Input for Activity Data Variant Comparison of Activity A for Variant 1.

Case	Age	Activity	Heart Rate	Pain
0	20y	A	67	low
2	23y	A	56	medium
3	25y	A	45	medium
4	30y	A	40	high

Table 5. Input for Activity Data Variant Comparison of Activity A for Variant 2.

Case	Age	Activity	Heart Rate	Pain
1	80y	A	103	high
5	85y	A	100	high
6	90y	A	100	high
7	82y	A	102	high

We propose this procedure for each activity with its respective event data attributes. In our example, we would do the same for the pain value for activity A and then with both measurements for activities B and C. Therefore, all values for the activity considered are collected per variant from all cases. This requires a case to contain the activity and the respective event data attribute under investigation. If loops are included, we provide different methods to deal with them, which are explained in Section 4.2.

4.5. Data Flow Variant Comparison

Besides activity data comparison, we also consider the behaviour of event data attributes and compare the transitions present in both process variants. The data shown in Table 1 contains a transition from activity A to activity B for all cases. We want to understand if the changes in measurements from activity A to activity B are statistically significant between variant 1 and

variant 2. This is especially interesting in the context of healthcare data, as it could show how treatments for different patient groups affect patients' measurements and, thus, the patients' wellbeing. This could help to compare and evaluate different treatment options for different patient groups.

In Table 6 and Table 7, we see a visualization of the transitions from A to B present in the data. We compute the change in measurements from activity A to activity B. This is possible for the continuous variable heart rate by calculating the difference. While continuous data, such as heart rate, allows simple subtractions of its values, categorical data requires different processing: the categorical value of the start and the ending activity is merged into a new categorical value. For the pain value, the transition from high to medium for Case 1 results in "high-medium".

We again collect all changes for one transition, such as the transition from A to B for one measurement, such as the heart rate. This is done for both activities to result in two samples that can be compared regarding their difference. We perform a statistical test on both samples to measure if there is a difference between variant 1 and variant 2. Here, heart rate for variant 2 decreases, whereas it mostly increases for variant 1. As illustrated in Table 6 and Table 7, the heart rate in variant 1 is almost only increasing with Δ values -1, 2, 3, 4 and in variant 2 decreasing with Δ values -3, -2, -5, -4, the resulting p-value would be 0.0285 with an RBC value of -1. Thus, the difference is statistically significant and the RBC value indicates that the changing values in variant 2 are all lower than in variant 1.

Table 6. Input for Data Flow Variant Comparison of Transition A $\rightarrow$ B for Variant 1.

Case	Age	Transition	Δ Heart Rate	Δ Pain
0	20y	A $\rightarrow$ B	-1	low $\rightarrow$ low
2	23y	A $\rightarrow$ B	+2	medium $\rightarrow$ medium
3	25y	A $\rightarrow$ B	+4	medium $\rightarrow$ high
4	30y	A $\rightarrow$ B	+3	high $\rightarrow$ high

Table 7. Input for Data Flow Variant Comparison of Transition A $\rightarrow$ B for Variant 2.

Case	Age	Transition	Δ Heart Rate	Δ Pain
1	80y	A $\rightarrow$ B	-3	high $\rightarrow$ medium
5	85y	A $\rightarrow$ B	-2	high $\rightarrow$ medium
6	90y	A $\rightarrow$ B	-5	high $\rightarrow$ medium
7	82y	A $\rightarrow$ B	-4	high $\rightarrow$ medium

As explained in Section 4.3, we perform a Man-Whitney-U-Test for continuous and a Chi-squared test for categorical data. We then filter the records based on their p-value to find those with statistically significant differences between both variants. This is done for each event data attribute and for each transition to compare all possible differences for statistical significance. Transitions under considerations are all directly follows and eventually follows relations available in the event log.

5. Evaluation

To evaluate our approach, we implemented it in the context of a Python application that provides the user with an interactive graphical representation. The overall goal of the evaluation was to demonstrate the feasibility of the proposed approach. To this end, we applied it on data of patients with acute kidney failure from the MIMIC IV dataset and show which insights on age-related data-based differences between patients can be obtained by using our approach. In the following sections, we elaborate on the details of our evaluation. We start by introducing the MIMIC IV dataset (Section 5.1) and the tool-based implementation (Section 5.2). Section 5.3 briefly elaborates on the split into variants before Section 5.4 reports on the results.

5.1. MIMIC IV Dataset

The MIMIC IV dataset is a medical dataset consisting of anonymized data of more than 40,000 patients. The data is retrospectively collected with the purpose to support research in healthcare. MIMIC IV contains a plethora of health measurements, such as blood pressure, heart rate, and laboratory values. Furthermore, it is enriched with context information of each patient, such as the type of health insurance, age, and many more.

The event log covers each step of the patient within the hospital, starting in the emergency department up to the patient's discharge. For some numerical measurements, like laboratory measurements, enriched categorical information is available. This enriched information can describe if a laboratory measurement is abnormally high, normal, or abnormally low, and is evaluated based on the individual patient data. For example, besides the numerical measurement of the phosphate value, there is also the categorical measurement "Abnormal Phosphate" present [Johnson et al., 2021, Goldberger et al., 2000].

Within MIMIC IV, there are several disease types available. Here, we focus on acute kidney failure patients because its disease progression can be measured by kidney specific laboratory measurements. We extracted an event log containing data of patients with

acute kidney failure. For the analysis in this paper, we limited the event log activities to the five most common. The full event log is available in our implementation[1].

The considered activities are Emergency Department, where the patient enters the hospital and undergoes triage with basic measurements, such as blood pressure. Afterwards, patients either undergo surgery in the Surgical Intensive Care Unit (ICU) or get their treatment in the Medical ICU. The Post-ICU Medicine department functions as a step after ICU treatments, and the last step, Discharge, describes the activity of letting patients leave the hospital.

5.2. Interactive Data-Based Process Comparator

We implemented our approach as a Python application named *Data-Based Process Comparator (DPC)*. This application allows analysing data-based variant differences between two event logs. The application is designed to be used in Jupyter Labs[2] and makes use of ipywidgets[3] for interactivity inside notebooks.

The default workflow starts with an event log. This event log can be manually split by the user into two variants. These variants are passed into the *DPC* which can then compute all relevant information. For a more advanced experience, *DPC* also comes with a visual tool that enables the user to quickly identify interesting characteristics of the two variants in respect to data values.

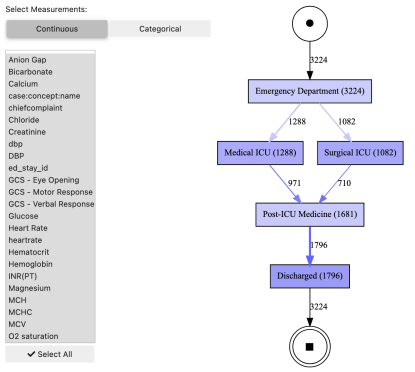

Figure 2. Visual Comparison example with the Data-Based Process Comparator.

The interactive tool is shown in Figure 2. It consists of a measurements section and a graph view. In the measurements section, one can select the measurements to be compared. At first, it can be switched between continuous and categorical measurements, and then specific measurements can be included or excluded by selecting or deselecting them. The graph view shows a Directly Follows Graph of the joined event logs of both variants. Each activity and each transition is coloured according to the highest RBC/Chi2 (test statistic) value, indicating the degree of difference between both variants. This means that the darker an activity is displayed, the higher the difference between the variants for at least one selected measurement. The same applies to transitions.

If the user wants to gain more information about the reason for the different colours, the activities and the transitions can be clicked, as shown in Figure 3. Then, an overview is displayed. It contains a list of all statistically significant measurements for this activity or transition, as well as the concrete results from the statistical tests. For example, Figure 3 illustrates the results when clicking on the Emergency Department activity, where statistical significant differences in heart rate and diastolic blood pressure (dbp) were identified. Furthermore, plots are generated, that help the user to understand the data distribution in both variants.

These plots will be further explained in the next paragraphs.

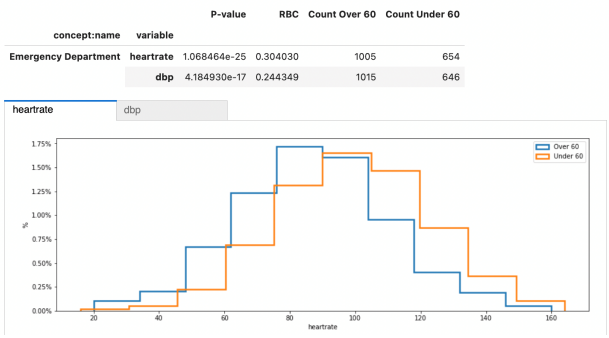

Figure 3. Comparison of Emergency Department values in Data-Based Process Comparator.

5.3. Variant Split

As mentioned above, the variant split for the proposed data-based process variant analysis is arbitrary. For the evaluation, we decided for an important factor in health care: the patients' age. The used split criterion is an age of above 60 years as variant 1 and an age smaller or equal to 60 years as variant 2. The age criteria of 60 years is a common risk factor in healthcare, which is the reason why we have chosen it [Setiati et al., 2020].

[1] https://github.com/bptlab/data-based-process-variant-analysis

[2] https://github.com/jupyterlab/jupyterlab

[3] https://github.com/jupyter-widgets/ipywidgets

5.4. Comparison Results

In the following, the results of the statistical tests are demonstrated for one laboratory measurement of Phosphate, which is associated with acute kidney failure [Lim et al., 2017]. We want to clarify that there are further interesting observations, which we cannot present due to the lack of space. Additionally, we cannot proof correctness for the test results. However, as the split is based on age, we can confirm the correctness of one statistical test, as the test on age was significant with an RBC value of -1, meaning, that all age values of younger patients are smaller than for older patients.

Evaluating the results based on the acute kidney failure dataset with the described split criterion, we see a dark blue colouring in the Surgical ICU activity, as shown in Figure 2. This indicates statistically significant differences between the two variants in this activity.

The Phosphate value is considerably interesting, with a statistical significance between younger and older patients, as shown in the graph in Figure 4. On the y-axis, we see the percentage of all values present for the value on the x-axis. The x-axis shows the absolute phosphate value.

The p-value is 3.8 x 10^{-14} and the RBC value is -0.26. This means that the Phosphate value distribution in the Surgical ICU activity differs between younger and older patients. Younger patients tend to have a lower Phosphate value compared to the older ones. The normal Phosphate value of adults is between 2.8 mg/dL and 4.5 mg/dL.

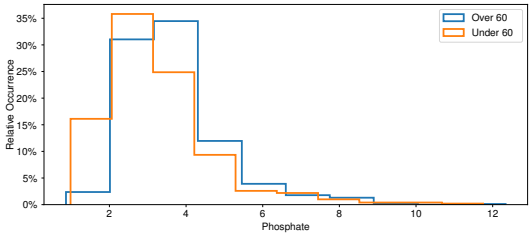

Figure 4. Comparison results for Phosphate in the Surgical ICU Activity.

The described result is also visible when looking at the categorical Abnormal Phosphate value, where we observe that young patients more often fall short of it, whereas old patients exceed the value, as shown in Figure 5. Comparing both variants, there is a statistical significance difference with a p-value of 2.29 x 10^{-15} and a chi2 value of 71.26.

Besides the coloured activities, we observe dark blue transitions in the graph in Figure 2. We made one interesting observation in the transition from Surgical

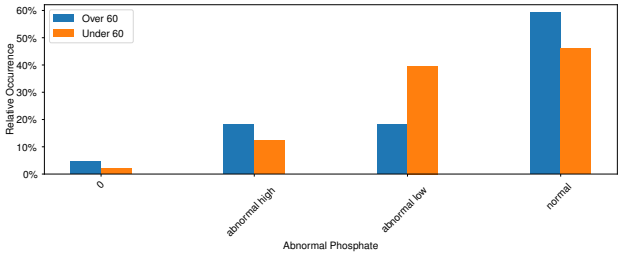

Figure 5. Comparison results for Abnormal Phosphate in the Surgical ICU Activity.

ICU to Post-ICU Medicine, where the Phosphate value behaviour differed between both variants. The comparison of the Phosphate value behaviour can be investigated in Figure 6. The x-axis shows the absolute change from the Surgical ICU Activity to the Post-ICU Surgery activity of the Phosphate value. On the y-axis, we see the percentage of all values present for the value on the x-axis.

For the younger patients, the Phosphate value increases from Surgical ICU to Post-ICU Medicine for the majority of patients. Contrary to that, the Phosphate value of the older patients mostly decreases in the same transition. We observe a p-value of 1.76 x 10^{-13} and an RBC value of 0.54. This is due to the previously discussed presence of rather low Phosphate values for younger patients and the rather high Phosphate values for older patients.

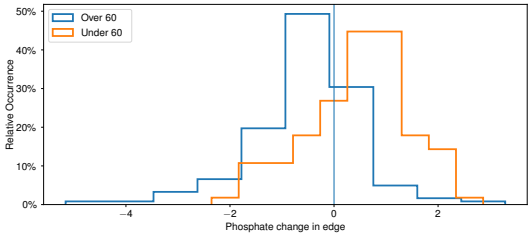

Figure 6. Comparison results for Phosphate lab value in the transition from Surgical ICU to Post-ICU Medicine.

The categorical variable Abnormal Phosphate shows a statistical significance in the transition from Surgical ICU to Post-ICU Surgery between the younger and older patients as well, with a p-value of 5.42 x 10^{-6} and a chi2 value of 38.77.

We observe a frequent change from abnormal low Phosphate values to normal Phosphate values for younger patients (30%), whereas older patients more often change from abnormal high Phosphate values to normal Phosphate values (10%). This is visualized in Figure 7.

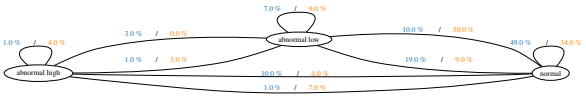

Figure 7. Comparison results for Abnormal Phosphate lab value in the transition from Surgical ICU to Post-ICU Medicine.

Table 8. Summary of statistical test results for phosphate

Analysis Type	Data Type	Activity/Transition	P	Test Statistic (**RBC**/*Chi2*)
Activity	Con.	Surgical ICU	3.8×10^{-14}	**-0.26**
Activity	Cat.	Surgical ICU	2.29×10^{-15}	*71.26*
Data Flow	Con.	Surgical ICU to Post-ICU Medicine	1.76×10^{-13}	**0.54**
Data Flow	Cat.	Surgical ICU to Post-ICU Medicine	2.29×10^{-15}	*38.77*

Furthermore, we see different behaviour between younger and older patients with regard to Phosphate values. The results of the statistical tests are summarized in Table 8 displaying the type of analysis with their respective p-value and test-statistic (RBC/Chi2). We could confirm already known characteristics of phosphate values in acute kidney patients, as a higher age is associated with higher phosphate values [Rubio-Aliaga, 2020]. Additionally, we could create new process oriented insights by comparing the behaviour of phosphate during the hospital treatment process, where phosphate increased for young patients and decreased for older patients.

Regarding the transition findings, there is Hyperphosphatemia, which describes an extensively high Phosphate value, and Hypophosphatemia, which describes a low Phosphate value. Both can be related to acute kidney injury. According to [Lim et al., 2017], Hypophosphatemia is common for critically ill patients and related to acute kidney patients. Hyperphosphatemia needs to be treated by giving Phosphate lowering medicine, as described in [Goyal and Jialal, 2021] and occurs together with acute kidney injury. This could be the reason for our observation, but has to be further evaluated by medical experts.

All in all, this evaluation has shown that event data attributes associated to activities holds additional information that can be used to derive further characteristics of process variants.

6. Discussion

Our approach enables data-based process variant analysis by analysing event data attributes, and thus enables using more of the available information in event logs to find distinguishing characteristics between process variants. Especially in healthcare, a process oriented, and data-based cohort comparison can be performed, which allows comparing treatment effects on different cohorts. However, the results should be interpreted with caution, as we only observe differences in data and cannot guarantee that the differences occur because of the performed activity or transition. We propose to compare event data attributes based on the values associated to activities and transitions, which allows deriving process-oriented insights about event data attribute behaviour in the process. However, there exist more perspectives to look at, such as understanding how event data attributes change over multiple activities or in specific trace variants. To evaluate, whether there is a statistically significant difference between transitions, we need a graph-based representation of the used event logs. Currently, the developed visual tool only supports Directly Follows Graphs for this. Typical event logs do not contain event data attributes for each event. Thus, one event can contain information about the heart rate in a medical context, and three activities later, the heart rate is measured again. Such changes would not be detected by only considering DFGs. A solution to this issue is the usage of Eventually Follows Graphs (EFGs), as they cover more possible transitions that can be compared for statistically significant differences. Integrating EFGs, one has to think about how to visualize and interpret them, as the paths could differ between the activities. One solution might be to analyse trace variants separately to cope with the complexity. Additionally, data flow variant comparison relies on similar control flow in the variants compared. Our approach for data-based variant comparison detects differences between variants by using tests for statistical significance. This approach works well for detecting an interesting finding, but is limited in meaningfulness. Even though there is a significant difference between two activity or data flow samples, we do not fully consider the amplitude of difference between variants. Thus, the p-value in combination with its test-specific measurement have a limited expression about the data under consideration. Nevertheless, statistical tests are broadly accepted in the medical domain, which might help to communicate results to medical experts. Thus, we see this approach as a white-box approach, which is one of the challenges proposed for healthcare process analytics [Munoz-Gama et al., 2022].

To summarize, our approach adds a data-based process variant analysis method that can be used across different domains independently of available data types.

7. Conclusion

In this paper, we presented an approach for data-based process variant analysis by determining statistical significant differences between two process

variants. We provide new insights for comparison by exploiting the increasing amount of data measurements available in healthcare processes. In this way, we enable users to analyse and compare two given process variants in a data-based fashion.

This is the first work in the field of data-based process variant analysis that considers continuous and categorical data associated to process activities. An evaluation of process data of kidney failure patients from the MIMIC-IV database demonstrated the feasibility of the approach and already showed interesting results for age-related differences. The results can help to make process-oriented statements regarding event data attribute behaviour during a process, such as comparing the development of the patient's state during a hospital treatment process.

The presented approach can be further enhanced by supporting loops in the variants and creating a loop-specific data-based visualization. Furthermore, the DPC could be extended to plot directly follows and eventually follows relations. Possible future work also includes the introduction of a data-based difference measure for variants. Utilizing this measure, an automatic selection of interesting variants could be proposed to allow users to obtain insights that do not require defining the variant split in the first place.

References

[XES, 2016] (2016). IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams. *IEEE Std 1849-2016*, pages 1–50.

[Anyanwu et al., 2003] Anyanwu, K. et al. (2003). Healthcare enterprise process development and integration. *Journal of research and practice in information technology*, 35(2):83–98.

[Armstrong, 2014] Armstrong, R. A. (2014). When to use the bonferroni correction. *Ophthalmic and Physiological Optics*, 34(5):502–508.

[Buijs et al., 2012] Buijs, J. C. A. M. et al. (2012). Towards cross-organizational process mining in collections of process models and their executions. In Daniel, F., Barkaoui, K., and Dustdar, S., editors, *Business Process Management Workshops*, pages 2–13, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Cremerius and Weske, 2022] Cremerius, J. and Weske, M. (2022). Supporting domain data selection in data-enhanced process models. In *Wirtschaftsinformatik 2022 Proceedings 3*.

[Cuzzocrea et al., 2017] Cuzzocrea, A. et al. (2017). Extensions, analysis and experimental assessment of a probabilistic ensemble-learning framework for detecting deviances in business process instances. In *ICEIS (1)*, pages 162–173.

[Goldberger et al., 2000] Goldberger, A. L. et al. (2000). Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220.

[Goyal and Jialal, 2021] Goyal, R. and Jialal, I. (2021). Hyperphosphatemia. In *StatPearls [Internet]*. StatPearls Publishing.

[Gulden, 2017] Gulden, J. (2017). Visually comparing process dynamics with rhythm-eye views. In Dumas, M. and Fantinato, M., editors, *Business Process Management Workshops*, pages 474–485, Cham. Springer International Publishing.

[Johnson et al., 2021] Johnson, A. et al. (2021). Mimic-iv.

[Lim et al., 2017] Lim, C. et al. (2017). Hypophosphatemia in critically ill patients with acute kidney injury treated with hemodialysis is associated with adverse events. *Clinical kidney journal*, 10(3):341–347.

[Lloyd-Jones et al., 2005] Lloyd-Jones, D. M., Evans, J. C., and Levy, D. (2005). Hypertension in adults across the age spectrum: current outcomes and control in the community. *JAMA*, 294(4):466–472.

[Low et al., 2017] Low, W. Z. et al. (2017). Change visualisation: Analysing the resource and timing differences between two event logs. *Information Systems*, 65:106–123.

[Mann and Whitney, 1947] Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60.

[Mans et al., 2008] Mans, R. et al. (2008). Process mining in healthcare - a case study. volume 1, pages 118–125.

[McHugh, 2013] McHugh, M. L. (2013). The chi-square test of independence. *Biochemia medica*, 23(2):143–149.

[Munoz-Gama et al., 2022] Munoz-Gama, J. et al. (2022). Process mining for healthcare: Characteristics and challenges. *Journal of Biomedical Informatics*, 127:103994.

[Nguyen et al., 2018] Nguyen, H. et al. (2018). Multi-perspective comparison of business process variants based on event logs. In *International Conference on Conceptual Modeling*, pages 449–459. Springer.

[Parab and Bhalerao, 2010] Parab, S. and Bhalerao, S. (2010). Choosing statistical test. *Int J Ayurveda Res*, 1(3):187–191.

[Rojas et al., 2016] Rojas, E. et al. (2016). Process mining in healthcare: A literature review. *Journal of Biomedical Informatics*, 61:224–236.

[Rubio-Aliaga, 2020] Rubio-Aliaga, I. (2020). Phosphate and Kidney Healthy Aging. *Kidney Blood Press Res*, 45(6):802–811.

[Setiati et al., 2020] Setiati, S. et al. (2020). Risk factors and laboratory test results associated with severe illness and mortality in COVID-19 patients: A systematic review. *Acta Med Indones*, 52(3):227–245.

[Swinnen et al., 2012] Swinnen, J. et al. (2012). A process deviation analysis – a case study. In *Business Process Management Workshops*, pages 87–98, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Taymouri et al., 2021] Taymouri, F. et al. (2021). Business process variant analysis: Survey and classification. *Knowledge-Based Systems*, 211:106557.

[Valera and Ghahramani, 2017] Valera, I. and Ghahramani, Z. (2017). Automatic discovery of the statistical types of variables in a dataset. In *International Conference on Machine Learning*, pages 3521–3529. PMLR.

[van der Aalst, 2016] van der Aalst, W. (2016). *Process Mining*. Springer Berlin Heidelberg.