

Behavioral Petri Net Mining and Automated Analysis for Human-Computer Interaction Recommendations in Multi-Application Environments

JULIAN THEIS and HOUSHANG DARABI, University of Illinois at Chicago

Process Mining is a famous technique which is frequently applied to Software Development Processes, while being neglected in Human-Computer Interaction (HCI) recommendation applications. Organizations usually train employees to interact with required IT systems. Often, employees, or users in general, develop their own strategies for solving repetitive tasks and processes. However, organizations find it hard to detect whether employees interact efficiently with IT systems or not. Hence, we have developed a method which detects inefficient behavior assuming that at least one optimal HCI strategy is known. This method provides recommendations to gradually adapt users' behavior towards the optimal way of interaction considering satisfaction of users. Based on users' behavior logs tracked by a Java application suitable for multi-application and multi-instance environments, we demonstrate the applicability for a specific task in a common Windows environment utilizing realistic simulated behaviors of users.

CCS Concepts: • **Human-centered computing** → **User models; User studies; User centered design; User interface design; User models; User studies; User centered design; User interface design; User studies; User centered design; User interface design; Human computer interaction (HCI); • Computing methodologies** → **Machine learning;**

Additional Key Words and Phrases: Multi-Application Environments, Software Process Mining, User Behavior Optimization, Human-Computer Interaction Recommendation, Behavioral Petri Nets

ACM Reference format:

Julian Theis and Houshang Darabi. 2019. Behavioral Petri Net Mining and Automated Analysis for Human-Computer Interaction Recommendations in Multi-Application Environments. *Proc. ACM Hum.-Comput. Interact.* 3, EICS, Article 13 (June 2019), 16 pages.
<https://doi.org/10.1145/3331155>

1 INTRODUCTION

Today's businesses are highly dependent on software which gets more and more complex, especially when applied in industrial environments [42]. Moreover, IT expertise is perceived as self-evident in many professions. In industries like Communication, Healthcare, Education, and Government, employees are often confronted with similar and repetitive software processes. Usually, the UIs of such software are developed to be used by a broad audience. Hence, there is a chance that the users in a given organization select to use a UI in a nonoptimal way. Therefore, organizations need to train their users to use the UI in the best possible way. How users deal with a corresponding UI can be seen via UI interaction event log analysis. Standard Human-Computer Interactions (HCI) such as mouse clicks, mouse movements, and keystrokes can be tracked and the called application

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2019 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.
2573-0142/2019/6-ART13 \$15.00
<https://doi.org/10.1145/3331155>

function information can be stored in log files. These logs, in turn, can be analyzed by Process Mining techniques to detect patterns.

Process Mining is the extraction of process models from event logs [39]. Analyzing logs originating from software environments aiming on optimizing software design, running behavior, and user-software interaction is also referred to as *Software Process Mining* [25, 34]. Software Process Mining plays an essential role in UI design. In many companies, users do not use the available UIs efficiently. In such cases, there is a need to train the users to use the company's UIs in a more efficient, and perhaps optimal, way. However, organizations do not always have resources to train each employee separately. A more cost-effective solution would be to train employees using individual recommendations provided by a software.

In this paper, we are proposing a method to detect inefficient user's behavior interactions in a multi-instance and multi-application environment assuming that at least one optimal way of interaction is known. Our method provides recommendations to gradually train employees or users. Potential applications can be found in areas in which interaction speeds are crucial - such as Electronic Health Record (EHR) systems. Doctors and nurses are often confronted with time issues when using EHR systems. Using such systems in a faster way saves time which can be spent on patients and/or reduce stress and workload. Similarly, our approach can be applied to train receptionist at hotels, helpdesk employees in financial institutions, or to optimize spreadsheet operations of users in virtually any industry.

To the best of our knowledge, we are the first ones applying Process Mining to provide HCI recommendations in such environments. Moreover, we provide recommendations ranked by average time saving such that a user gradually reaches to an optimal way of interaction.

After providing a brief introduction, we present the background and related work in Section 2. We provide relevant preliminaries and definitions in Section 3. The experimental and simulation environment for data acquisition purposes and the actual recommendation engine are introduced in Section 4. We discuss the obtained results in Section 5 and conclude the paper in Section 6.

2 BACKGROUND AND RELATED WORK

When working with UIs, one should always keep in mind the overall objectives. Naively, one might think the design of an application should be modern, generate interest, and be easy to use. Though these assumptions are usually true, UIs have a much larger impact since almost all businesses are highly dependent on software tools for their success [13, 32]. Intelligently designed UIs reduce the number of errors and their related cost [36]. They also reduce the task interaction time, resulting in an overall higher productivity [41]. In industrial applications, these task interaction times are production times due to reduced iterations and rework. Moreover, an efficient UI does not only affect the software's usability, but also leads to a smooth completion of any task at hand thereby making the interaction satisfactory as per the requirements of the user. In turn, a complicated or impractical UI can be an existing problem within an organization. Exemplary, an inefficient UI can slow down operations and eventually cause stress for employees [9]. It is also important to note that employees might not always complain about UI inefficiencies making the detection even harder. Therefore, companies are highly interested in conducting application log analysis to unveil frequent users' tasks and to understand users' behavior [14]. In any UI and HCI problem, one always wants to detect an optimal solution for a specific problem, i.e. an optimal way of interaction where the problem is the detection of an optimal user's behavior in a software environment. In this case an optimal user's behavior can be defined as one that has the maximum *Effectiveness*, *Efficiency*, and *Satisfaction* [11, 20, 30].

Formal methods, such as Process Mining techniques, have been used for usability analysis and user assistance for a long time to achieve effective, efficient and satisfying UIs. However, they have

not been applied to provide automatic HCI recommendations. For example, Hartman et al. proposed a method called AUGUR to assist users in navigating and entering data in form applications [19]. However, this method applies only to web-based applications and does not consider the interaction between multiple applications.

Palanque et al. proposed a formal method consisting of a set of different techniques to facilitate the evaluation of usability [31]. Their approach is to model applications using high-level PNs and to evaluate the application using observed user's logs. From the user's logs and replay on the application PN, one can observe failures of the task as well as usability issues. This approach can detect usability issues. However, it cannot be used to train the users. It cannot be used for user's behavior evaluation either. Moreover, this method requires a modeling of the application in a non-automated way.

Thimbleby et al. have applied Social Network Analysis theory to HCI [38]. Their work shows that an application's features can be interrelated using a social network. User's behavior represented as a social network graph is then able to unveil user's interactions and used features. Again, this application does not aim to provide automated user recommendations to interact with the application in an efficient way. It rather evaluates the work of the designer and gives hints which functions need more or less attention.

Bowen et al. utilize presentation interaction models (which are state transition systems) to describe graphical UIs and by modeling system manuals [7]. Their objective is to align and unveil inconsistencies between a graphical UI and according manuals. Though this work relates to user's training by providing correct training materials, it does not consider user's behavior logs and does not provide automated, step-wise training recommendations for users.

Dostal et al. developed a framework for the OpenOffice.org Suite which enables the logging of HCIs while interacting with OpenOffice [15]. This tool enables many opportunities to analyze user's behavior in OpenOffice applications. However, it is limited to OpenOffice only and does not track HCI between applications and instances outside the OpenOffice.org Suite. Unfortunately, no research has been conducted on user's behavior logs observed by OpenOffice in order to recommend user's interactions.

Liu et al. discovered user's behavior from a software execution logs utilizing Process Mining techniques [24]. However, the authors utilize the observed user's behaviors to detect patterns like commonly triggered user's operations in an application, specifically ProM, only.

In general, a significant amount of work has been done in the area of user's behavior analysis, but mainly in web analytics [22] and user activity tracking [3]. Also, a lot of work has been done in web usage mining [10, 28]. Atterer et al. investigated the tracking of user's interaction using web technologies [3]. The authors obtained meaningful statements by analyzing and refining the collected log files. Furthermore, there has been research conducted on utilization of Process Mining to detect abnormal behavior in social networks [35]. A further approach to the analysis of farmers' interaction with decision support systems was shown by Mărușter et al. [29].

There is also an area called *Software Runtime Analysis* focusing on model generation [26] and mining [2, 16]. However, these contributions do not deal with the user but with the runtime perspective.

All these areas utilize Process Mining to support the area of Software Development. To the best of our knowledge, no research has been conducted on supporting users to solve tasks in a more efficient way utilizing Process Mining techniques.

3 PRELIMINARIES

In this section, we introduce the preliminaries and definitions that are required throughout this paper. We define event logs and Petri nets (PNs), provide an introduction to Process Mining and

process discovery, and define HCI specific terms which are required to understand the proposed approach.

3.1 Event Log

The elementary component of *event logs* are *events*. An *event* can be any real-life action consisting of a *name* and an associated *timestamp* [39]. These two attributes are required whereas further attributes are optional. The term *attribute* is often used as a synonym for *resource* in Process Mining literature. Generally speaking, we define \mathcal{E} representing the event space of all possible events. Moreover, \mathcal{A} is the set of all possible attributes. Consequently, for $e \in \mathcal{E}$ and $a \in \mathcal{A}$: $\tau_a(e)$ is the value of the attribute a [17, 39]. We define a_{name} and a_{time} as the attribute of an event's name and timestamp respectively.

A finite sequence of *events* is defined as a *trace* g and is associated with an attribute *name*. In any trace, each *event* occurs only once and *events* are ordered according to their *timestamp* [17, 39]. The term *case* is used synonymously for a *trace*. We can define \mathcal{T} as the set of all possible traces. Therefore, $g \in \mathcal{T}$ and $a \in \mathcal{A}$: $\tau_a(g)$ is the value of the attribute a for trace g . Moreover, we define $|e|_g$ as the number of events in a trace g [17]. Therefore, the order of events in a trace has to satisfy

$$\forall_{1 \leq i < j \leq |e|_g} \tau_{a_{time}}(e_i) \leq \tau_{a_{time}}(e_j). \quad (1)$$

Finally, an *event log* L is a collection of *traces* $g_i \in \mathcal{T}$ for $0 \leq i$ such that each *trace* g_i occurs only once [17].

3.2 Process Mining

Process Mining has been successfully applied in different fields like Healthcare [12, 27], Insurance [37], Auditing [21], and also Software Development [23, 25, 34] and consists of three disciplines: (i) process discovery, (ii) conformance checking, and (iii) enhancement of process models [34, 39]. A process model can be discovered by considering a log file as input. The output of such a discovery algorithm is a process model, which is usually a PN or a Business Process Modeling Notation (BPMN), Event Driven Process Chains (EPCs), or Casual Net (CNs). We are focusing on PNs only.

Conformance Checking is used to evaluate if a discovered process model is a good representation of the system recorded by a log. Two commonly used quality measures are *fitness* and *precision* among others. The *fitness* metric measures if a process model can replay all events in each of an event log's traces. The *precision* metric measures how well a model generalizes the actual process by simulation. Each generated trace by the process model should be a real-world trace, i.e. should be found in the actual event log. Further measures include the complexity of a model. A process model is considered better the less complex it is.

Lastly, enhancement considers a discovered process model as well as the logs to improve or extend the model using additional information, e.g. to predict events or to provide recommendations.

The term *Software Process Mining* deals with analyzing software development processes, their runtime behavior as well as the user-software interaction [25]. Therefore, we are mainly interested in the enhancement discipline of Process Mining, specifically in analyzing the user-software interaction or HCI.

3.3 Petri Net

A PN is a commonly used modeling technique for process models. It consists of a set of places, graphically represented as circles and transitions represented as rectangles. Transitions usually correspond to events. The unidirectional interconnection between places to transitions and transitions to places is visualized using arcs. As such, a PN is defined as

$$PN = (\mathcal{P}, T, F, \pi) \quad (2)$$

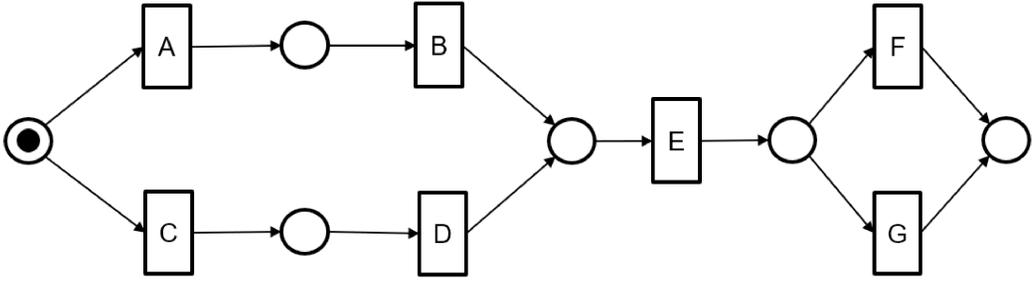


Fig. 1. Basic PN with seven transitions and six places. The first place holds a token.

in which \mathcal{P} is a set of places, T is a set of transitions and $F \subseteq (\mathcal{P} \times T) \cup (T \times \mathcal{P})$ is a set of directed arcs connecting places and transitions [1]. π is a function which maps transitions to event names. It is defined such that [1]

$$\forall e \in \mathcal{E} \exists t \in T \pi(t) = \tau_{aname}(e). \quad (3)$$

In general, each place can hold a non-negative integer number of tokens, however, in this paper the number of tokens in each place is either 1 or 0. We define the number of tokens in a place p as a function $\sigma(p)$ where $p \in \mathcal{P}$. For all $p \in \mathcal{P}$, the corresponding array of $\sigma(p)$ represents the marking m of the PN. Each marking m such that $m \in M$ represents the state of the PN [39]. We call the initial marking m_0 , whereas the final marking is represented as m_z . Therefore, the sum of the vector elements of m_0 and m_z must be each greater than 0. Furthermore, a transition $d \in T$ can only be fired if Equation (4) is satisfied for $p_n \in P : \exists(p_n \times d) \in F$ and if $p_n \neq \emptyset$.

$$\forall p_n : \sigma(p_n) \geq 1 \quad (4)$$

By firing a transition, a token is removed from each input place to that transition, while for each outgoing place from the transition the number of tokens will be increased by 1. An important property to maintain is *soundness* which describes the absence of livelocks, deadlocks, and other anomalies. Generally speaking, a PN is considered to be *sound* if and only if for each trace [40]:

- it is always possible to reach the final marking,
- there are no remaining tokens when the final marking is reached,
- and if it is possible to execute an arbitrary event starting from m_0 by following the appropriate route through the PN.

A basic PN can be seen in Figure 1. It holds a token in the first place, representing the initial marking m_0 .

A special kind of transition is a so-called *invisible transition*, graphically represented as a black rectangle. Invisible transitions can always be fired without corresponding events from a log as long as token requirements in incoming places are satisfied. Invisible transitions are used to model complex dependencies.

We call a PN that models user's behavior a *behavioral PN*.

3.4 Process Discovery Algorithm

Throughout this paper, we use a state-of-the-art discovery algorithm called *split miner* [4, 6] which is a recent technique to discover PNs from event logs. The method has been developed by Augusto et al. with the objective to detect models with high fitness and precision, yet low complexity. The algorithm consists of five steps: a directly-follows dependency graph will be created and short loops will be discovered at first. Afterwards, the algorithm searches for concurrency and

marks respective elements as such. In a third step, filtering will be applied such that each node is on a path from a single start to end node (maintaining soundness), the number of edges are minimal (minimizing complexity), and every path from start to end has the highest possible sum of frequencies (maximizing fitness). Afterwards, the algorithm adds split gateways in order to capture choice and concurrency. In a final step, joins will be discovered.

The *split miner* algorithm works well on a huge set of artificial and real-world event logs compared to existing methods [5] and achieves significant performance improvements.

3.5 Further Definitions

In this paper, we are differentiating between *high-level events* and *low-level events*. We define *low-level events* as events produced by software modules and libraries, such as Dynamic Link Libraries (DLLs). We call it *low-level* since the event is recorded in its raw format. Therefore, *low-level events* are usually hard to understand and to debug. In contrast, *high-level events* are easier to understand, and follow the format of event logs which can be utilized for Process Mining. Thus, *high-level events* consist of a *name* as well as a *timestamp* and optionally further attributes. *Low-level events* can be translated to *high-level events* using transfer functions.

Furthermore, we differentiate between *controllable* and *uncontrollable* events. *Controllable events* are events which a user physically performs, such as keystrokes, mouse movements, or clicks. In contrast, we define *uncontrollable events* as events which result from a sequence of *controllable events* and underlying software logic. Examples are the visual appearance of an application window or closing of an application's instance. Both of the examples can be triggered by different sequences of *controllable events*. Thus, *controllable events* are user controlled and the opposition of *uncontrollable* ones which are triggered by the system. Such representations are usually created by the transformation of events using transducers [8, 18] or transfer functions [33].

Finally, we are differentiating between *Intra Task* and *Inter Task* behavior. We are defining an Intra Task t_{intra} as a sequence of controllable events which end with a single uncontrollable event. We denote controllable events as c_i , where i is the i th controllable event in a sequence, and uncontrollable ones as u . Thus, $t_{intra} = \{c_1, \dots, c_i, u\}$ where $i > 0$. An example for an Intra Task could be changing the directory from a directory a to its subfolder b . An Inter Task, denoted as t_{inter} , instead deals with the sequence of Intra Tasks, as such only with the uncontrollable events. An example could be the task to open a Notepad instance. This can be achieved by the sequence of the Intra Tasks *open Windows menu*, *search for Notepad.exe*, and *open application*.

4 APPROACH

The approach consists of two steps. First, we consider the acquisition of data since no suitable datasets are publicly available. The data acquisition section describes how we derive and develop an HCI user's behavior simulator for a given task from real participant observations. In the second part of this section, we propose the actual recommendation engine which is based on the output of the HCI user's behavior simulation.

4.1 Data Acquisition

We are presenting the applicability of the proposed method based on a simple yet realistic real-world example in a classic desktop environment. We asked five real users/participants to solve a well-defined task while tracking their HCIs as low-level events. The resulting logs were used to create a simulation in order to create a larger amount of traces in a scalable way. The simulated traces were used to validate our approach. Figure 2 visualizes this procedure in a flow diagram.

Here, we describe the actual *Task*, the *HCI Logging*, and the final derived *Simulation* in detail.

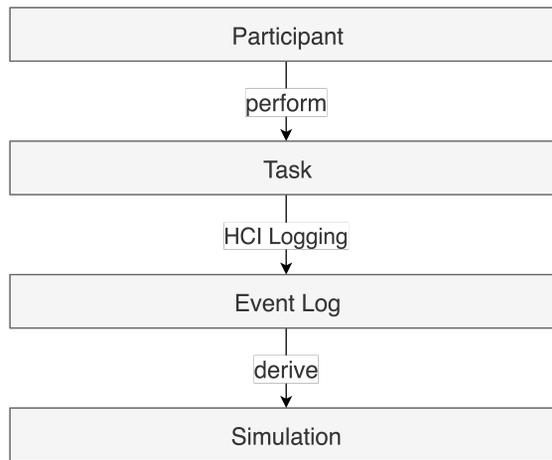


Fig. 2. Flow diagram of the data acquisition process.

4.1.1 Task. The participants are asked to summarize a company's performance data consisting of two files located in the *documents/company data* folder for 15 times. Each of those two files consists of information about revenues and expenses of a single product. In order to solve the problem, the user has to create a simple text file called *summary.txt* in the folder *documents/company data/summaries* and has to fill in the summarized information. The applications a user can use are the *Windows Calculator*, *Windows Explorer*, *Notepad* as well as the standard desktop. The user is provided a mouse, keyboard, screen, and a computer with a *Windows 10* operating system.

4.1.2 HCI Logging. To record the HCIs, we develop a Java application which tracks all inputs as well as visually shown on-screen events. The application tracks mouse movements, mouse clicks as well as keystrokes and leverages Window's DLLs to obtain information about opened, closed, maximized, and minimized applications. Additionally, our application tracks the hierarchy of visually on-screen shown applications and current folder locations of the *Windows Explorer*. Table 1 shows the details of the tracked events. Each event is associated with a corresponding timestamp. Moreover, in order to protect the privacy of participants, we only track the keys listed in Table 2. All other keystrokes are expressed as a *TEXT KEY* or *NUM KEY*, depending if the text key was an alphabetical character or numerical/mathematical operation key. For simplicity reasons, we have split the screen into 4x4 rectangles such that mouse movements and clicks are not recorded pixel-wise but within the defined regions. The logging application is used to obtain HCI traces of the participants while performing the above described task.

The source code of the logging application is available in our Github repository ¹.

4.1.3 Simulation. For Process Mining purposes, we need multiple traces per user in order to discover a process model. A small amount of traces would not provide sufficient behavioral insight. Therefore, we develop a logical simulation software with control parameters to create low-level event traces at scale. The development of this logical simulation is based on observations of the participant's event logs in order to create a larger set of traces. These large set of traces can be used to demonstrate the applicability of the proposed approach. The source code of the simulation is available in our Github repository, too ².

¹<https://github.com/Julian-Theis/ActivityTracker>

²<https://github.com/Julian-Theis/UserActivitySimulation>

Event Key	Parameter	Description
A1	[PID, ProcessName, WindowTitle, width, height, top_left_x, top_left_y]	Application opened
A2	[PID, ProcessName, WindowTitle]	Application closed
A3	[PID, ProcessName]	Application maximized
A4	[PID, ProcessName]	Application minimized
A5	[PID, ProcessName, WindowTitle]	Window title changed
A6	[PID, ProcessName, width, height, top_left_x, top_left_y]	Window position changed
A7	[PID#ProcessName;...;PID#ProcessName]	Window hierarchy order changed
A8	[PID, OldPath, NewPath]	Explorer path changed
K1	[KeyCode]	Key pressed
K2	[KeyCode]	Key released
K3	[KeyCode, x, y]	Mouse click pressed
K4	[KeyCode, x, y]	Mouse click released
K5	[-1 1]	Mouse wheel
M	[x, y]	Mouse position

Table 1. All low-level events of the developed logging application. Each event consists of an event key as well as a set of parameters where PID is the application's Process Identifier used by the operating system.

Type	Key Code	Description
Keyboard	29	Left ctrl
Keyboard	3613	Right ctrl
Keyboard	3675	Left meta
Keyboard	28	Enter
Keyboard	3665	Page down
Keyboard	3657	Page up
Keyboard	1	ESC
Keyboard	56	Left alt
Keyboard	3640	Right alt
Mouse	1	Left click
Mouse	2	Right click
Keyboard	42	Shift
Keyboard	54	Right shift
Keyboard	57419	Left arrow
Keyboard	57416	Top arrow
Keyboard	57421	Right arrow
Keyboard	57424	Down arrow
Keyboard	59-68	F1-F10
Keyboard	87-88	F11-F12
Keyboard	15	Tab
Keyboard	14	Backspace

Table 2. Overview about the recorded keys and key codes.

The obtained low-level event logs of each real participant encompasses between 15,000 and 30,000 events over a time period of 18 to 28 minutes for the 15 traces. We observe that each participant had an individual level of reactivity, which resulted in different speeds of mouse movements and frequencies of keystrokes. Therefore, one of the simulation parameter is *Reactivity*. Similarly, each participant had a different mouse precision which resulted in the parameter *MousePrecision*.

Next, some participants tend to open new application instances instead of reopening already existing ones and they tend to close or minimize applications. These behaviors can be controlled using the parameters *Minimize*, *AppClosing* and *AppOpenOrReopen*. Another observation is that some participants used the *Windows Search Functionality* instead of browsing the *Windows Explorer*. Therefore, we introduce a parameter *Search*. Furthermore, some of the participants utilized hotkeys, whereas others solved the problem using the mouse only. This behavior is controlled with the parameter *HotkeyUsage*. Finally, the task can be separated into subtasks such as *creating summary file*, *changing location to documents/data summaries*, *calculating*, etc. These tasks can be repeated multiple times and in different orders. To model this behavior, we introduce the two parameters *Repetition* and *Sequential*. All parameters are expressed as the likelihood of the corresponding behavior with a numerical value between 0 and 1. The logical path simulation is developed in Python and creates traces based on the low-level events defined in Table 1.

Moreover, we develop a log translator which translates the low-level events of Table 1 to high-level events. The reason for this is that even trivial switching between two applications produces multiple low-level events. Such a high-level event is always triggered by a user's interaction, which in turn is defined as a controllable event. User's interactions can be a mouse click or keystroke, or a sequence of clicks and keystrokes. Usually, a controllable event leads to an uncontrollable event which consists a sequence of the low-level events *A1* to *A7*. Switching between two applications leads for example to *A3*, *A6* and *A7* simultaneously. Since such sequences would add extra complexity to PN models, we are translating them to high-level events. These high-level events are human-understandable, e.g. *open explorer*, *minimize notepad*, *maximize calculator*, etc.. Later on, we will use high-level events to discover the behavioral PNs.

4.2 Proposed Recommendation Engine

In this section, we propose the actual recommendation engine. Figure 3 illustrates the procedure. We use the simulation described in the previous section to simulate a user's event log with specific characteristics, i.e. parameters. The user's event log has to consist of a statistically representative amount of traces of high-level events. The optimal log, however, consists only of one trace with high-level events. This trace is the optimal behavior we want to achieve. Therefore, the simulation parameters are set to best possible values. The user's event log is used to discover a behavioral PN using the *split miner*. Finally, the optimal log and the user's event log are replayed against the obtained PN in order to calculate metrics and user recommendations. Especially the calculation of metrics and recommendations is discussed in this section.

In the first step, a behavioral PN will be discovered on the user's event log. This model summarizes the diverse set of traces and removes noise while being a statistical valid and solid representation of the user's behavior. Therefore, we utilize the high-level event traces from the HCI user simulation and leverage the earlier introduced *split miner* discovery algorithm.

We replay the user's event log as well as the optimal log on the obtained behavioral PN. This is also called a *conformance* check. Therefore, we utilize a replay function which considers only *move on model* transitions. This means that the conformance metrics are calculated based on missing tokens in the PN only, such that we will never skip and penalize the actual events for fitness calculation purposes. The reason for this is that the user's behavior, and thus the traces, are the ground truth to be optimized rather than the PN model. If a user's behavior does not represent a specific optimal behavior, it should only move tokens within the model. Moreover, as long as the current state of the model is in a non-concurrency situation, there will be only one token in the system. If a deviation between a trace and the model exists, we are calculating the minimal distance between the current token position and the corresponding event observed in the log. The

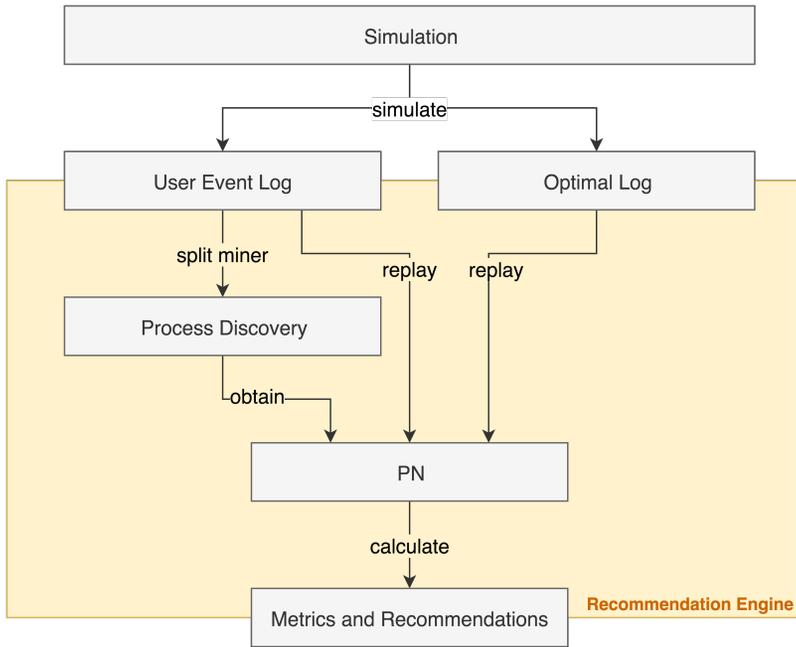


Fig. 3. Flow diagram of the proposed approach.

higher the distance, the higher the cost associated with the deviation. The cost for moving one token for one transition is 1.

The fitness of the user's behavior event logs is measured as a baseline. For comparison purposes, the fitness of the optimal trace is measured, too. Since the *split miner* discovery algorithm removes noisy behavior by filtering, we are not assuming a perfect fitness value. Instead we expect a value between 60% and 90%. The optimal trace fitness value will be lower assuming that the user's behavior is not optimal. The difference between the user's behavior and optimal fitness values provides a good indication of how close a user is to the optimal task solution.

Furthermore, the time span between transition enabling, i.e. the time Equation (4) is satisfied, and firing are measured. We also measure the frequency between transition enabling and actual transition firing. This information is important to obtain the reactivity of a user as well as to obtain his frequent behavior, in particular to differentiate between common and accidental behavior. The reactivity can be measured as the average firing time between two consecutive controllable events. In a very similar way, we can measure the mouse precision. The first and last mouse movement events in a consecutive sequence of mouse movements provide the start and end position of the cursor. Using this information, one can calculate the shortest path from start to end. The shortest path can be compared to the sequence replayed on the behavioral PN model. The paths are being compared to obtain a value corresponding to the user's mouse precision. Reactivity and mouse precision can be used to recommend specialized training. However, the reactivity value is an important metric to calculate the costs for further recommendations.

Moreover, we provide a user with Intra Task and Inter Task recommendations. An example for an Intra Task could be changing the directory from the *documents* folder to its subfolder *company data*. This Intra Task can be solved in a couple of ways using different sequences of controllable events. Figure 4 shows the path of a sample Intra Task in a PN. An Intra Task recommendation consists of

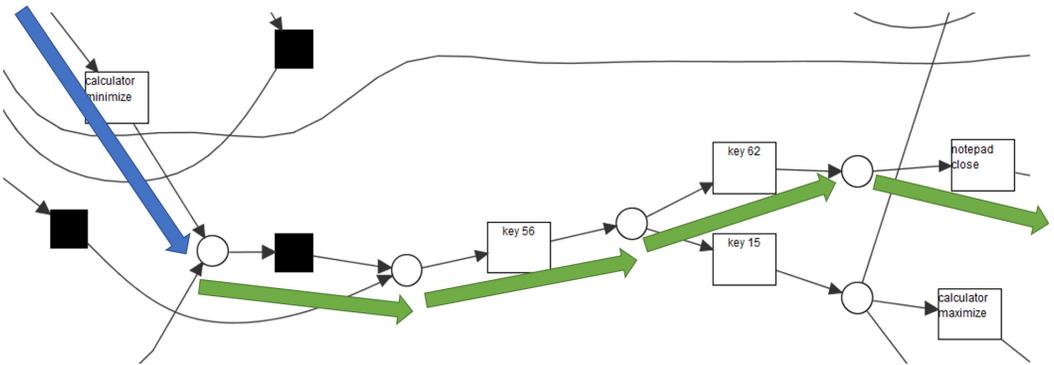


Fig. 4. This figure shows an extract of a discovered PN from user's event logs and visualizes an Intra Task. The blue arrow indicates the uncontrollable event of the previous Intra Task. Green shows the complete flow of an Intra Task ending with the uncontrollable event *notepad close*.

the final uncontrollable event as descriptor, a percental occurrence rate as the likelihood that the Intra Task occurs in a trace, the total number of occurrences averaged per trace, the average time saving based on the user's reactivity in milliseconds as well as the controllable event sequence a user should perform and the one the user should avoid, based on the optimal log. To obtain these recommendations, we first detect the Intra Tasks which deviate from the optimal log by replaying. We then calculate the cost which is the time a user could save by applying the optimal Intra Task behavior over the already applied user's behavior based on the earlier obtained user reactivity.

In comparison, an Inter Task is a sequence of Intra Tasks. Therefore, the shortest sequence of multiple t_{intra} tasks is supposed to be the optimal solution to execute a specific task. As such, we analyze the Inter Task sequence obtained from the user's event log replay compared to the optimal log for repetitions. We calculate the cost for skipping Inter Tasks in order to argument how valuable a recommendation is to the user. The structure of Inter Task recommendations is very similar to Intra Task recommendations. It consists of a sequence of uncontrollable events which a user should not repetitively trigger. Moreover, it consists of a percental occurrence rate as the likelihood of occurrence per trace as well as an average time saving in milliseconds.

The Java source code of the proposed approach has been published in our Github repository ³.

5 RESULTS

In order to evaluate our approach, we set up different user's behavior scenarios and create the corresponding event logs utilizing the developed simulation. Table 3 provides an overview. The simulation is based on random sampling and every run is independent and identically distributed. When defining different user's behaviors, the likelihoods have been set to 0.05 or 0.95 accordingly. Therefore, we have chosen a sample size of 825 traces to obtain a 95% confidence interval with an interval size smaller than 1.5 for each simulation parameter. Thus, we allow for enough random behavior, and obtain a statistical representative sample.

We run the above approach on the five different user's behaviors defined in Table 3. The measured reactivity of the optimal trace is 328ms. The optimal mouse precision is 1.0. Results are shown in Table 4.

For *User 1*, who is in general slower and less precise, a high user reactivity time of 674ms compared to the optimal 328ms is recorded. Also, the mouse precision of 0.328 is very low compared to 1.0.

³<https://github.com/Julian-Theis/HCI-PetriNets>

User	Traces	Events	Time [s]	Description
Optimal	1	145.0	65.56	High precision and reactivity, no repetitions, hotkeys
User 1	825	250.31	209.58	Lower reactivity, low mouse precision
User 2	825	189.00	75.58	Very low reactivity, no hotkey usage
User 3	825	159.04	81.00	Very low reactivity, repetitions
User 4	825	205.24	232.31	Low reactivity, repetitions, no hotkeys
User 5	825	446.74	191.03	High reactivity, low key precision, repetitions, no hotkeys

Table 3. Overview of the modeled high-level user's event logs. The *Events* and *Time* column shows the average number of events and average time of each trace.

User's Behavior	User's Fitness	Optimal Fitness	User's Reactivity [ms]	User's Mouse Precision
User 1	0.804	0.734	674	0.352
User 2	0.884	0.565	901	1.0
User 3	0.888	0.679	922	1.0
User 4	0.918	0.510	731	1.0
User 5	0.874	0.582	337	1.0

Table 4. Metrics obtained from the simulated user's behavior scenarios

The user's fitness of 0.804 is as expected, but not perfect due to the already mentioned reasons in section 4.2. The fitness value of the optimal trace has a value of 0.734 and is quite high relatively to the assumptions we made. However, it is lower than the user fitness value due to imprecise mouse events.

User 2 does not use any hotkeys. Therefore, the user's behavior fitness value is at a baseline of 0.884. The optimal trace has a very low fitness value of 0.565 which makes sense as many of the optimal event sequences cannot be replayed on the user's behavior model. Reactivity is very low and mouse precision high, as simulated.

The reactivity and mouse precision of *User 3* are as expected. However, this user's behavior is simulated such that he/she repeats Inter Tasks often. Therefore, the user's fitness as well as optimal fitness values are comparatively high with 0.888 and 0.679. This is due to the fact that both sequential strategies are very similar except for repetitions of Inter Tasks. Such repetitive behavior is not reflected in the fitness value.

The behavior of *User 4* is a mix of everything. This user is slow, repeats a lot of Inter Tasks and does not consider hotkeys. Therefore, the user's fitness baseline is 0.918. The fitness value of the optimal trace is very low with 0.510 as expected.

Finally, *User 5* does not consider hotkeys and is simulated with a high likelihood of repetitive Inter Tasks. Similarly, the user's fitness is high with 0.874 while the optimal one is comparatively low. Both, reactivity and precision values are acceptable.

Intra Task and/or Inter Task Recommendations have been obtained for all user's behaviors except *User 1*. As *User 1* was simply slower and imprecise, but all other parameters were optimal, no further recommendations can be provided. As such, for example for *User 2*, we observe the following Intra Task recommendation:

Task: explorer path to documents/summary

Occurrence Rate: 0.460

Total Occurrence Per Trace: 1.412

Average Time Saving: 8109

User should not do: [mouse click, key TEXT, key TEXT, key TEXT, key TEXT, key TEXT,

key TEXT, key TEXT, key 28]
User should do instead: [mouse doubleclick]

The user always typed the word *summary* in the application window of the Explorer and pressed *Enter* in order to move on into another directory. However, a simple double click would have been faster according to the user's reactivity. Another example of an Intra Task recommendation for *User 2* is the following:

Task: explorer maximize
Occurrence Rate: 0.162
Total Occurrence Per Trace: 2.725
Average Time Saving: 3604
User should not do: [mouse to 4,3, mouse to 3,3, mouse to 2,3, mouse to 1,3,
mouse to 1,2, mouse click]
User should do instead: [key 56, key 15]

Instead of maximizing the Explorer by moving the mouse and clicking, it would be faster for the user to use the combination of *ALT* and *TAB* on the keyboard. However, the occurrence rate as well as the average time saving are lower. Therefore, the first Intra Task has a higher priority than the second one due to the average time saving.

In this way, we obtained Intra Task recommendations for *User 2*, *User 4*, and *User 5*. Especially, since those users have not utilized hotkeys. *User 3*, however, has an optimal Intra Task behavior.

Similarly, for *User 3*, the algorithm recommends the following two Inter Task recommendations:

Task: notepad close
Occurrence Rate: 0.413
Average Time Saving: 39579
User should not repetitively do: notepad close

Task: calculator minimize
Occurrence Rate: 0.192
Average Time Saving: 4806
User should not repetitively do: calculator minimize

Both Inter Tasks were observed in the user's event logs repetitively. We also obtain suitable Inter Task recommendations for *User 4* and *User 5* which would fix the behavior introduced in the user profiles in Table 3. By incorporating and considering each of the recommendations, a user will get closer to the optimal behavior, i.e. will be trained gradually.

6 DISCUSSION AND CONCLUSION

We developed an approach to provide HCI user recommendations using Process Mining assuming that at least one optimal way of interaction is known. This approach deals with HCI traces in multi-instance and multi-application environments and provides users with useful recommendations. Applications can be found in any HCI user interaction improvement problem in which time is a crucial component and users are confronted with a large amount of repetitive tasks. We target supervisors or commissaries in organizations whose objective is to optimize HCI of employees in a cost-effective manner. We successfully applied our Process Mining approach on simulated user traces based on a realistic example which has been designed and run with real users in advance. Therefore, our approach can be used to detect users who do not interact optimally, i.e. their user

logs deviate from optimal ones. These users can be trained in a step-wise manner by providing them with either a single or few recommendations at a time.

A re-validation of their HCI behavior and the impact of providing recommendations could be investigated at a later time to monitor each user's progress. In conclusion, the advantages for organizations are twofold: firstly, users will not be overtaxed and employee's satisfaction is still ensured. Secondly, organizations themselves increase efficiency while decreasing cost.

Our approach has a few limitations. Process Mining requires large amount of traces. Therefore, organizations have to collect data over a few weeks or even months, depending on the frequency of repetitive user tasks. Although, the translation of low-level to high-level events is expensive and requires domain knowledge. Significant effort has to be spent to track essential events.

In the future, an investigation on the performance and limitations of the proposed approach should be performed based on a real-world scenario and real event logs. Moreover, our proposed approach requires a logging application which is able to track HCIs in multi-application environments and to translate low-level events to high-level events. Unfortunately, there are no standardized logging interfaces for HCI optimization available which could be incorporated with approaches like ours. Therefore, research studies on the design and profitability of such HCI logging interfaces should be performed.

REFERENCES

- [1] Arya Adriansyah, Boudewijn F van Dongen, and Wil MP van der Aalst. 2011. Conformance checking using cost-based fitness analysis. In *2011 IEEE 15th International Enterprise Distributed Object Computing Conference*. IEEE, 55–64.
- [2] Glenn Ammons, Rastislav Bodík, and James R. Larus. 2002. Mining Specifications. In *Proceedings of the 29th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL '02)*. ACM, New York, NY, USA, 4–16. <https://doi.org/10.1145/503272.503275>
- [3] Richard Atterer, Monika Wnuk, and Albrecht Schmidt. 2006. Knowing the User's Every Move: User Activity Tracking for Website Usability Evaluation and Implicit Interaction. In *Proceedings of the 15th International Conference on World Wide Web (WWW '06)*. ACM, New York, NY, USA, 203–212. <https://doi.org/10.1145/1135777.1135811>
- [4] Adriano Augusto, Raffaele Conforti, Marlon Dumas, and Marcello La Rosa. 2017. Split Miner: Discovering Accurate and Simple Business Process Models from Event Logs. In *2017 IEEE International Conference on Data Mining (ICDM)*. 1–10. <https://doi.org/10.1109/ICDM.2017.9>
- [5] Adriano Augusto, Raffaele Conforti, Marlon Dumas, Marcello La Rosa, Fabrizio Maria Maggi, Andrea Marrella, Massimo Mecella, and Allar Soo. 2019. Automated discovery of process models from event logs: Review and benchmark. *IEEE Transactions on Knowledge and Data Engineering* 31, 4 (2019), 686–705.
- [6] Adriano Augusto, Raffaele Conforti, Marlon Dumas, Marcello La Rosa, and Artem Polyvyanyy. 2018. Split miner: automated discovery of accurate and simple business process models from event logs. *Knowledge and Information Systems* (2018), 1–34.
- [7] Judy Bowen and Steve Reeves. 2012. Modelling User Manuals of Modal Medical Devices and Learning from the Experience. In *Proceedings of the 4th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '12)*. ACM, New York, NY, USA, 121–130. <https://doi.org/10.1145/2305484.2305505>
- [8] William Buxton. 1990. A three-state model of graphical input. In *Human-computer interaction-INTERACT*, Vol. 90. 449–456.
- [9] Fethi Calisir and Ferah Calisir. 2004. The relation of interface usability characteristics, perceived usefulness, and perceived ease of use to end-user satisfaction with enterprise resource planning (ERP) systems. *Computers in human behavior* 20, 4 (2004), 505–515.
- [10] Flavio Chierichetti, Ravi Kumar, Prabhakar Raghavan, and Tamas Sarlos. 2012. Are Web Users Really Markovian?. In *Proceedings of the 21st International Conference on World Wide Web (WWW '12)*. ACM, New York, NY, USA, 609–618. <https://doi.org/10.1145/2187836.2187919>
- [11] Stefan Cronholm. 2009. The Usability of Usability Guidelines: A Proposal for Meta-guidelines. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7 (OZCHI '09)*. ACM, New York, NY, USA, 233–240. <https://doi.org/10.1145/1738826.1738864>
- [12] Houshang Darabi, William L. Galanter, Janet Y. Lin, Ugo Buy, and Rupa Sampath. 2009. Modeling and integration of hospital information systems with Petri nets. In *2009 IEEE/INFORMS International Conference on Service Operations, Logistics and Informatics*. 190–195. <https://doi.org/10.1109/SOLI.2009.5203928>

- [13] Thomas H Davenport. 2000. *Mission critical: realizing the promise of enterprise systems*. Harvard Business Press.
- [14] Himel Dev and Zhicheng Liu. 2017. Identifying Frequent User Tasks from Application Logs. In *Proceedings of the 22Nd International Conference on Intelligent User Interfaces (IUI '17)*. ACM, New York, NY, USA, 263–273. <https://doi.org/10.1145/3025171.3025184>
- [15] Martin Dostál and Zdenek Eichler. 2011. A Research Framework for Performing User Studies and Rapid Prototyping of Intelligent User Interfaces Under the OpenOffice.Org Suite. In *Proceedings of the 3rd ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '11)*. ACM, New York, NY, USA, 153–156. <https://doi.org/10.1145/1996461.1996511>
- [16] Robert Gombotz and Schahram Dustdar. 2006. On Web Services Workflow Mining. In *Business Process Management Workshops*, Christoph J. Bussler and Armin Haller (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 216–228.
- [17] Qinlong Guo, Lijie Wen, Jianmin Wang, Zhiqiang Yan, and Philip S. Yu. 2015. Mining Invisible Tasks in Non-free-choice Constructs. In *Business Process Management*, Hamid Reza Motahari-Nezhad, Jan Recker, and Matthias Weidlich (Eds.). Springer International Publishing, Cham, 109–125.
- [18] Arnaud Hamon, Philippe Palanque, José Luís Silva, Yannick Deleris, and Eric Barboni. 2013. Formal description of multi-touch interactions. In *Proceedings of the 5th ACM SIGCHI symposium on Engineering interactive computing systems*. ACM, 207–216.
- [19] Melanie Hartmann, Daniel Schreiber, and Max Mühlhäuser. 2009. AUGUR: Providing Context-aware Interaction Support. In *Proceedings of the 1st ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '09)*. ACM, New York, NY, USA, 123–132. <https://doi.org/10.1145/1570433.1570458>
- [20] Julie A. Jacko. 2012. *Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, Third Edition* (3rd ed.). CRC Press, Inc., Boca Raton, FL, USA.
- [21] Mieke Jans, Michael Alles, and Miklos Vasarhelyi. 2013. The case for process mining in auditing: Sources of value added and areas of application. *International Journal of Accounting Information Systems* 14, 1 (2013), 1 – 20. <https://doi.org/10.1016/j.accinf.2012.06.015>
- [22] Avinash Kaushik. 2007. *Web analytics: an hour a day*. John Wiley & Sons.
- [23] Elia Kouzari, Lazaros Sotiriadis, and Ioannis Stamelos. 2018. Process Mining for Process Conformance Checking in an OSS Project: An Empirical Research. In *Open Source Systems: Enterprise Software and Solutions*, Ioannis Stamelos, Jesus M. Gonzalez-Barahona, Iraklis Varlamis, and Dimosthenis Anagnostopoulos (Eds.). Springer International Publishing, Cham, 79–89.
- [24] Cong Liu, Shi Wang, Shangce Gao, Feng Zhang, and Jiujun Cheng. 2018. User behavior discovery from low-level software execution log. *IEEJ Transactions on Electrical and Electronic Engineering* 13, 11 (2018), 1624–1632.
- [25] Cong Liu, Jianpeng Zhang, Guangming Li, Shangce Gao, and Qingtian Zeng. 2018. A two-layered framework for the discovery of software behavior: A case study. *IEICE TRANSACTIONS on Information and Systems* 101, 8 (2018), 2005–2014.
- [26] Davide Lorenzoli, Leonardo Mariani, and Mauro Pezzè. 2008. Automatic Generation of Software Behavioral Models. In *Proceedings of the 30th International Conference on Software Engineering (ICSE '08)*. ACM, New York, NY, USA, 501–510. <https://doi.org/10.1145/1368088.1368157>
- [27] Ronny S. Mans, Helen Schonenberg, Minseok Song, Wil M. P. van der Aalst, and Piet J. M. Bakker. 2009. Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital. In *Biomedical Engineering Systems and Technologies*, Ana Fred, Joaquim Filipe, and Hugo Gamboa (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 425–438.
- [28] Bamshad Mobasher, Robert Cooley, and Jaideep Srivastava. 2000. Automatic Personalization Based on Web Usage Mining. *Commun. ACM* 43, 8 (Aug. 2000), 142–151. <https://doi.org/10.1145/345124.345169>
- [29] Laura Mărușter, Niels Faber, René J. Jorna, and Rob Van Haren. 2008. A Process Mining Approach to Analyse User Behaviour. 2 (01 2008), 208–214.
- [30] Jakob Nielsen. 1993. *Usability Engineering*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- [31] Philippe Palanque, Eric Barboni, Célia Martinie, David Navarre, and Marco Winckler. 2011. A Model-based Approach for Supporting Engineering Usability Evaluation of Interaction Techniques. In *Proceedings of the 3rd ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '11)*. ACM, New York, NY, USA, 21–30. <https://doi.org/10.1145/1996461.1996490>
- [32] Sung Park, Arthur D. Fisk, and Wendy A. Rogers. 2010. Human Factors Consideration for the Design of Collaborative Machine Assistants. *Handbook of Ambient Intelligence and Smart Environments* (2010), 961–984. https://doi.org/10.1007/978-0-387-93808-0_36
- [33] Philip Quinn, Andy Cockburn, Géry Casiez, Nicolas Roussel, and Carl Gutwin. 2012. Exposing and understanding scrolling transfer functions. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*. ACM, 341–350.
- [34] Vladimir A. Rubin, Alexey A. Mitsyuk, Irina A. Lomazova, and Wil M. P. van der Aalst. 2014. Process Mining Can Be Applied to Software Tool!. In *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering*

- and Measurement (ESEM '14)*. ACM, New York, NY, USA, Article 57, 8 pages. <https://doi.org/10.1145/2652524.2652583>
- [35] Mahdi Sahlabadi, Ravie Chandren Muniyandi, and Zarina Shukur. 2014. Detecting abnormal behavior in social network websites by using a process mining technique. *Journal of Computer Science* 10, 3 (2014), 393–402. <https://doi.org/10.3844/jcsp.2014.393.402>
- [36] Ben Shneiderman, Catherine Plaisant, Maxine Cohen, and Steven Jacobs. 2009. *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (5th ed.). Addison-Wesley Publishing Company, USA.
- [37] Suriadi Suriadi, Moe T. Wynn, Chun Ouyang, Arthur H.M. ter Hofstede, and Nienke van Dijk. 2013. Understanding process behaviours in a large insurance company in Australia : a case study. In *25th International Conference on Advanced Information Systems Engineering, CAiSE 2013*. Springer, Valencia, Spain, 449–464. https://doi.org/10.1007/978-3-642-38709-8_29
- [38] Harold Thimbleby and Patrick Oladimeji. 2009. Social Network Analysis and Interactive Device Design Analysis. In *Proceedings of the 1st ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '09)*. ACM, New York, NY, USA, 91–100. <https://doi.org/10.1145/1570433.1570453>
- [39] Wil M. P. van der Aalst. 2011. *Process Mining: Discovery, Conformance and Enhancement of Business Processes* (1st ed.). Springer Publishing Company, Incorporated.
- [40] W. M. P. van der Aalst, K. M. van Hee, A. H. M. ter Hofstede, N. Sidorova, H. M. W. Verbeek, M. Voorhoeve, and M. T. Wynn. 2011. Soundness of workflow nets: classification, decidability, and analysis. *Formal Aspects of Computing* 23, 3 (01 May 2011), 333–363. <https://doi.org/10.1007/s00165-010-0161-4>
- [41] Martijn Van Welie. 2001. Task-based user interface design. *SIKS Dissertation Series* 6 (2001).
- [42] Valeria Villani, Lorenzo Sabattini, Julia N Czerniaki, Alexander Mertens, Birgit Vogel-Heuser, and Cesare Fantuzzi. 2017. Towards modern inclusive factories: A methodology for the development of smart adaptive human-machine interfaces. In *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE, 1–7.

Received October 2018; revised February 2019; accepted April 2019