IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 31, 2018, accepted January 10, 2019, date of publication January 22, 2019, date of current version February 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2894116

An Evidence-Based Decision Support Framework for Clinician Medical Scheduling

MINSU CHO[1], MINSEOK SONG[1], SOOYOUNG YOO[2], AND HAJO A. REIJERS[3]
[1]Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang 37673, South Korea
[2]Health ICT Research Center, Seoul National University Bundang Hospital, Bundang-gu 13620, South Korea
[3]Department of Information and Computing Sciences, Utrecht University, 3512 Utrecht, The Netherlands

Corresponding author: Minseok Song (mssong@postech.ac.kr)

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology under Grant 2016R1C1B2016346 and Grant 2014K1A3A7A03073707.

ABSTRACT In healthcare management, waiting time for consultation is an important measure that has strong associations with patient's satisfaction (i.e., the longer patients wait for consultation, the less satisfied they are). To this end, it is required to optimize medical scheduling for clinicians. A typical approach for deriving the optimized schedules is to perform experiments using discrete event simulation. The existing work has developed how to build a simulation model based on process mining techniques. However, applying this method for outpatient processes straightforwardly, in particular medical scheduling, is challenging: 1) the collected data from electronic health record system requires a series of processes to acquire simulation parameters from the raw data; and 2) even if the derived simulation model fully reflects the reality, there is no systematic approach to deriving effective improvements for simulation analysis, i.e., experimental scenarios. To overcome these challenges, this paper proposes a novel decision support framework for a clinician's schedule using simulation analysis. In the proposed framework, a data-driven simulation model is constructed based on process mining analysis, which includes process discovery, patient arrival rate analysis, and service time analysis. Also, a series of steps to derive the optimal improvement method from the simulation analysis is included in the framework. To demonstrate the usefulness of our approach, we present the case study results with real-world data in a hospital.

INDEX TERMS Simulation modeling, process mining, personal clinician schedules, experimental analyses, waiting time for consultation.

I. INTRODUCTION

In outpatient processes, long waiting times for consultation for patients can be considered as critical problems [1], [3]. The longer patients have to wait before their consultation can take place, the less satisfied they are, which may lead to decreasing profits [15]. A confounding factor is that there are significant differences with respect to quality delivery and efficiency among clinicians. In order to handle this problem, it seems worthwhile to consider how the personal appointment schedules of clinicians can be optimized as to improve the overall efficiency of patient management.

A typical approach for optimizing medical schedules for a clinician is *Discrete Event Simulation* (*DES*) [4], [12], [18]. In the healthcare field, various studies have been using DES where clinical activities are considered as the crucial events in outpatient processes [4], [8], [10], [12], [18], [20]. However, it takes generally much time and effort to build an accurate simulation model. This is because the status quo is that simulation models are created by manually recorded data, which may be inaccurate, or interviews, which are time-consuming. To overcome these limitations, Rozinat et al. [18] proposed to combine simulation with process mining as to extract process-related knowledgeable information from so-called event logs [18], [21], [22], [24], [25]. Process mining uses such automatically recorded logs to automatically derive the specific operations in a particular context, which is one of the leading components of a simulation model. The authors explained how to make a *Colored Petri Net* (*CPN*) model using four kinds of analyses [18].

Unfortunately, it still has a couple of challenges to straightforwardly apply this method for outpatient processes, in particular, medical scheduling.

1) The collected data from Electronic Health Record (EHR) system requires a series of processes to acquire simulation

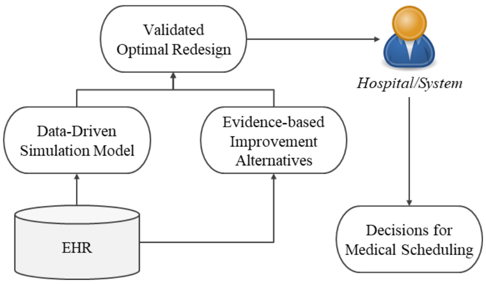

FIGURE 1. Overview of the decision support framework.

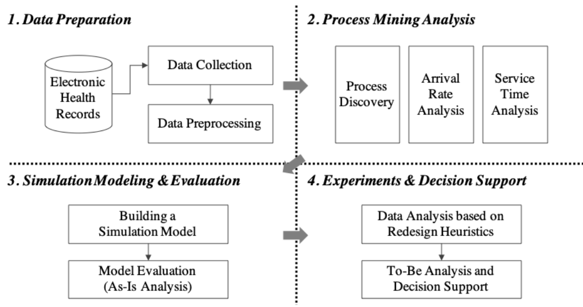

FIGURE 2. The proposed decision support framework for medical scheduling.

parameters from the raw data. Three main elements for building a healthcare simulation model are a process of medical activities, service times, and arrival rates. Out of them, it is hard to find out actual values of the service times and arrival rates from EHR data due to the following reasons:

- Service times: EHR systems, which typically, only record completion time of clinical activities.

- Arrival rates: Patients visit a hospital with a scheduled appointment; thus, the reservation system needs to be considered.

2) Even if the derived simulation model fully reflects the reality, there is no systematic approach to deriving effective improvements, i.e., experimental scenarios. The next step of building a simulation model is to identify all the possible alternatives and determine the best option for the optimal decision making with simulation analysis. However, the existing methods presented in this regard are all heuristic-oriented and unstructured approaches. Therefore, a bridge that connects the simulation model analysis and useful scenarios is still missing.

To overcome these challenges, this paper proposes a novel decision support framework for a clinician's schedule using simulation analysis. Fig. 1 provides the overview of the proposed approach for medical scheduling. It aims at optimizing a clinician's schedule to decrease the waiting time of consultation for patients. To this end, a data-driven simulation model is constructed using process mining analysis, which includes process discovery, patient arrival rate analysis, and service time analysis. Also, improvement alternatives are investigated from data analysis for making experimental scenarios. Then, the validated optimal redesign methods are determined from the simulation analysis.

Using real data, we demonstrate that this simulation-based approach can be practical for experimentally investigating alternative scenarios for improved patient management.

The paper is organized as follows. Section II and Section III present the proposed framework and its application in case studies, respectively. A discussion is provided in Section IV, and conclusions are finally drawn in Section V.

II. METHODS

A. OVERVIEW

The proposed decision support framework for the optimized medical scheduling is composed of four phases: data preparation, process mining analysis, simulation modeling & evaluation, and experiments & decision support. Fig. 2 represents the overview of the proposed framework. First, an appropriate format of data, i.e., an event log, is collected from EHR system's log data of the hospital and preprocessed for effective data analysis in the data preparation phase. After that, three kinds of process mining analysis are performed to derive the inputs for creating a simulation model: process discovery, arrival rate analysis, and service time analysis. Based on these results, a simulation model is constructed, and then the model is evaluated to validate whether the model reflects the behaviors observed from the data (i.e., As-Is analysis). Here, a couple of *Key Performance Indicators* (*KPIs*) are employed. Lastly, in the experiments & decision support phase, improvement alternatives are investigated by performing further data analysis. To this end, best practices for business process redesign [17] are utilized as candidates for process improvement. Then, several scenario-based simulation analyses are performed to identify the optimal redesign method (i.e., To-Be analysis).

B. PHASE 1: DATA PREPARATION

1) DATA COLLECTION

As stated before, we employ process mining approaches for deriving simulation parameters. That is, we need to utilize event logs, which represent the behaviors recorded by an information system and are used in process mining approaches. An event log L is a set of traces T, where a process instance (i.e., a patient in clinical event logs) has one trace. A trace is a finite sequence of events E. An event $e \in E$ includes multiple required attributes AT including the name of the activity (i.e., *act*), the completion time (i.e., *ctime*), the reservation time (i.e., *rtime*), and the name of the resource (i.e., *res*). For the specific attribute, we can get the corresponding value using π function. Here, $\pi : E \rightarrow (AT \nrightarrow V)$ is a function which obtains attribute values recorded for an event. Hence, $\pi_{at}(e) \in AT \nrightarrow V$ signifies to obtain the corresponding value $v \in V$ recorded for attribute $at \in AT$. For instance, $\pi_{act}(e_1) =$ '*Consultation*' represents that the name of the activity of the event e_1 is consultation.

2) DATA PREPROCESSING

After collecting clinical event logs, the data preprocessing step is conducted to improve the accuracy and effectiveness of the data analysis. It includes removing noisy data, identifying outliers, and handling incomplete or error data.

C. PHASE 2: PROCESS MINING ANALYSIS

1) PROCESS DISCOVERY

Process discovery aims at extracting process models from event logs [6], [7], [9], [13], [18], [24], [26]. Through seminal research, many kinds of discovery algorithms have been developed, such as *alpha-mining* [24], *heuristic mining* [26], *genetic mining* [7], *fuzzy mining* [9] *and inductive mining* [13]. In this research, we apply *frequency mining*, which produces a process map (A_L, R_L) based on directly-follows relationships between activities in event logs [6]. Here, A_L is the set of activities in a log L, while R_L is the set of relations between two activities in a log L. A relation $r_{ij} = \{(a_i, a_j)|a_i, a_j \in A_L \wedge a_i > a_j\}$ is an element of R_L, where $a_i > a_j$ represents a notable directly-followed relationship (i.e., a_j is the direct successor of a_i) that has a higher frequency than a pre-determined threshold value.

According to this technique, if there is a relationship from activity A to activity B, then nodes A and B are connected by an arc in the process model. Frequency mining has the powerful advantage that it is able to include all paths in a process model (e.g., with zero threshold value). Therefore, it is a better way than other mining methods to apply in the healthcare domain, because all patient paths in a hospital are relevant.

In the existing approach that creates a simulation model by hand, it is essential to identify the medical activities and the patient flow (i.e., transition) probabilities associated with each activity involved in the outpatient process. However, our approach employs the discovered process model from data. Thus, in such a process, key medical activities or transition probabilities are automatically identified without any human intervention. Therefore, we included only the discovered process model which contains medical activities as one of the simulation parameters.

2) ARRIVAL RATE ANALYSIS

As we stated earlier, patients visit a hospital by a specifically scheduled appointment. To arrive at an accurate simulation model, it is essential to build it based on the characteristics of such schedules, including information on slot capacity and intervals between slots. In this paper, we propose a method to analyze a realistic arrival rate by applying two sorts of information related to the reservation system.

1) The number of appointments for each reservation slot
2) The patients' visiting time (the actual) compared to the reservation time (the planned)

The pseudo-code in Algorithm 1 explains the proposed approach in detail.

By computing how many patients visited the hospital in each slot, we can derive the visiting distribution of patients. After that, we figure out the actual arriving time by applying the second type of information.

Algorithm 1 Deriving Arrival Rate

Input
  Log L
  Time slots for reservation TS(a time slot $ts_k \in TS$)
Output
  The number of appointments for each time slot N
  A collection of patients' visiting time compared to the reservation time D
$N \leftarrow$ an initialized array with size $|TS|$
$D \leftarrow \{\}$
for all traces σ_i in the log L **do**
  $visitT_{\sigma_i} \leftarrow 0$
  for all events e_j in the trace σ_i **do**
    for all time slots $ts_k \in TS$ **do**
      if $\pi_{rtime}(e_j)$ is involved in ts_k **then**
        $N[k] \leftarrow N[k] + 1$
        break;
      end if
    end for
    if $visitT_{\sigma_i} = 0$ **or** $visitT_{\sigma_i} < \pi_{ctime}(e_j)$ **then**
      $visitT_{\sigma_i} \leftarrow \pi_{ctime}(e_j)$
    end if
  end for
  $D \leftarrow D \cup visitT_{\sigma_i}$
end for
return N, D

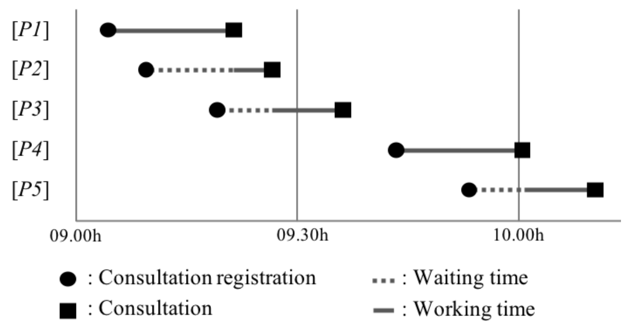

FIGURE 3. Measuring working/waiting time for consultation.

3) SERVICE TIME ANALYSIS

Working time is one of the indispensable components in making a simulation model. However, it is hard to get the accurate working time because most EHR systems in hospitals record only the completed time for each activity [19]. In other words, the duration of activities cannot be divided into waiting and working time because of the absence of the start time of the consultation. For this reason, the status quo is that service time is derived from manual checking [5]. To avoid this laborious and imprecise step, we suggest a new method to estimate the working time for consultation from event logs. The pseudo-code in Algorithm 2 explains the proposed approach in detail.

This method has a realistic assumption that patients visit the consultation room where a doctor works in the

Algorithm 2 Deriving Service and Waiting Time

Input
  Log L
  A resource r_i
Output
  A collection of service time for a resource $(r_i)S_i$
  A collection of waiting time for a resource $(r_i)W_i$
Note that $\pi^{CR}_{ctime}(e_i)$ is the completion time for consultation registration, and$\pi^{C}_{ctime}(e_i)$ is the completion time for consultation.
$S_i \leftarrow \{\}$
$W_i \leftarrow \{\}$
for all events e_j in the log L **do**
  for all resources $r_i \in R$ **do**
    sort by $\pi^{C}_{\mathrm{ctime}}(\mathrm{e_j})$
    $\mathrm{E_i} \leftarrow \{\}$
    if $\pi^{C}_{\mathrm{res}}(\mathrm{e_j}) = \mathrm{r_i}$ **do**
      $\mathrm{E_i} \leftarrow \mathrm{E_i} \cup \mathrm{e_j}$
    end if
    for all events $e_k \in E_i$ **do**
      if e_{k-1} does not exist (i.e., k = 1) or $\pi^{C}_{ctime}(e_{k-1}) < \pi^{CR}_{ctime}(e_k)$ **then**
        $S_i \leftarrow S_i \cup \{\pi^{C}_{ctime}(e_k) - \pi^{CR}_{ctime}(e_k)\}$
      else
        $S_i \leftarrow S_i \cup \{\pi^{C}_{ctime}(e_k) - \pi^{C}_{ctime}(e_{k-1})\}$
        $W_i \leftarrow W_i \cup \{\pi^{C}_{ctime}(e_{k-1}) - \pi^{CR}_{ctime}(e_k)\}$
      end if
    end for
  end for
end for
return S_i, W_i

consecutive order. In other words, the doctor sees patients one at a time, one after another. To explain the principle clearly, we provide a graphical example in Fig. 3. In this figure, for each patient the end times for the consultation registration and consultation are shown. Note that all records are sorted by the end time for consultation. The method we propose to measure the consultation *service time* can be divided into two ways. First, we can distinguish those patients who either get a consultation as the first patient in a time slot (e.g., *P1*) or those patients whose end time for consultation registration is later than the previous patient's end time for the actual consultation (e.g., *P4*). Neither of these types patients has to wait at all since no people are waiting. For these patients, the actual service time of their consultation equals the difference between the registered end times for consultation registration and consultation. By contrast, for the rest of them (e.g., *P2*, *P3*, and *P5*), the end time for the consultation registration of each patient is *earlier* than the end time for consultation of the previous patient. In such case, it is likely that at least one previous patient is waiting at the registration desk or consulting the doctor. Therefore, such patients have to wait before their consultation. The service time for such patients then equals the difference between their own consultation end time and the consultation end time of the previous patient. In absolute terms, a slight error may be introduced in that the preparation time for consultation might be included in the extracted service time. However, we believe that the advantages of such an automated approach outweigh those of manually measuring service and waiting times.

D. PHASE 3: SIMULATION MODELING & EVALUATION

1) BUILDING A SIMULATION MODEL

Based on the results of the three process mining analyses, we can now easily derive a simulation model. This is because we can get every input for the model from the process mining results—the extracted model, the arrival rate, and the service time.

2) MODEL EVALUATION (AS-IS ANALYSIS)

The model is then used to carry out a so-called *as-is analysis*, which allows for a comparison of the results generated by the simulation model with the observed data as present in the actual logs (i.e., records). To evaluate the model thoroughly, we define a couple of KPIs. In this study, we set three measures based on an in-depth discussion with domain experts in the hospital: the waiting time for consultation ($wt(L)$), the controllable waiting time for consultation ($cwt(L)$), and the end time of a clinic session for a specific doctor ($et(L)$). These are given in Eq. (1-3). Note that events are sorted by the execution time for consultation, and KPIs are calculated for each doctor. Also, $\pi^{CR}_{ctime}(e_i)$ and $\pi^{C}_{ctime}(e_i)$is the completion time for consultation registration and consultation, respectively. Moreover, $\pi^{C}_{rtime}(e_i)$ is the reservation time for consultation, and *MAX* is a function to find out the maximum value.

$$wt(L) = \sum_{0 \le e < |c|} \sum_{0 \le i < |e|} \begin{cases} \pi^{C}_{ctime}(e_{i-1}) - \pi^{CR}_{ctime}(e_i) \\ \quad if\ \pi^{C}_{ctime}(e_{i-1}) > \pi^{CR}_{ctime}(e_i) \\ 0 \quad otherwise \end{cases} \tag{1}$$

$$cwt(L) = \sum_{0 \le e < |c|} \sum_{0 \le i < |e|} \begin{cases} \pi^{C}_{ctime}(e_{i-1}) - \pi^{C}_{rtime}(e_i) \\ \quad if\ \pi^{C}_{ctime}(e_{i-1}) > \pi^{CR}_{ctime}(e_i) \\ 0 \quad otherwise \end{cases} \tag{2}$$

$$et(L) = \sum_{0 \le e < |c|} \sum_{0 \le i < |e|} \left\{ MAX\left(\pi^{C}_{ctime}(e_i)\right) \right\} \tag{3}$$

Among them, the second one signifies the difference between the start time of consultation and the reserved time, which excludes the waiting time which is incurred due to patients who registers ahead of the reserved time. Also, the third one represents the timestamp when consultation of the last patient in a session is completed. Based on these KPIs, we evaluate whether the simulation model accurately reflects the real situation or not. This may provide the required confidence to use the simulation model for alternative scenarios. Here, a statistical analysis, e.g., *t-test*, is applied to identify whether the values from the data and simulation model analysis are statistically identical or not [33]. The null and alternative hypothesis for t-test are the means of two groups are equal and not equal, respectively. Besides, we employ evaluation measurements, e.g., *Mean Absolute Percentage Error* (MAPE) [23] for comparing KPIs quantitatively.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|, \tag{4}$$

where A_t is the actual value calculated from the logs, and F_t is the forecast value derived from simulation.

E. PHASE 4: EXPERIMENTS & DECISION SUPPORT

1) DATA ANALYSIS BASED ON REDESIGN HEURISTICS

As mentioned earlier, prior to the scenario-based experimental simulation analysis, it is necessary to conduct preliminary data analysis for identifying a useful scenario which is applicable to improvements of the relevant process. This is because process improvement methods are quite diverse and have a broad range. In this paper, 29 heuristic best practices by Reijers and Mansar [17] are employed as an available set of process improvement alternatives. It covers practical redesign methods such as *activity elimination*, *case types*, and *case assignment*. Based on these best practices, we suggest a series of steps to derive evidence-based simulation scenarios. First, we obtain the applicable best practices with the following conditions.

1) Whether or not a best practice satisfies the goal of the simulation analysis
2) Whether or not a best practice is suitable for the relevant process
3) Whether or not a best practice is already applied to the process
4) Whether or not a best practice is more needed opinions from domain experts than data analysis
5) Whether or not information related to a best practice is stored in the log

After that, we define indicators to identify the availability of each best practice determined through data analysis. The existing work [11] has developed the relevant indicators for each best practice, and they are applied immediately or slightly modified. For example, it is required to measure the number of occurred events within a small unit of time (e.g., 1 min.) for an activity to identify the availability of the *order-based work* best practice, i.e., removing batch-processing and periodic activities in a process. Then, as the measured value exceeds the pre-defined threshold, the relevant best practice is considered as one of the experimental simulation scenarios.

2) TO-BE ANALYSIS & DECISION SUPPORT

The discovered experimental scenarios are tested based on the data-driven simulation model. To this end, we employ the waiting time for consultation and controllable waiting time among the KPIs already presented. Then, the extent of improvements is evaluated for all the experiments. Finally, the optimal scenario-based medical scheduling for a clinician is derived.

TABLE 1. A partial example of clinical event logs.

CaseID	Act.	Comp. Time	Res. Time	Resource	Dept.	Patient Type
P1	CR	May 01 09:00:05	May 01 09:00:00	Mike	Clinic	Follow-up
P1	C	May 01 09:05:10		Peter	Clinic	
P1	CS	May 01 09:06:20		Peter	Dept	
P1	P	May 01 09:30:10		John	PD	
P1	OHPP	May 01 09:35:15		Jason	PresD	
P2	CR	May 01 09:00:20	May 01 09:00:00	Sarah	Clinic	New

CR: Consultation Registration; C: Consultation; CS: Consultation Scheduling; P: Payment; OHPP: outside-hospital prescription printing

III. EXPERIMENTS AND RESULTS

3) MATERIALS

We used the real-world data from EHR system at a fully-digitalized tertiary general university hospital in Korea. Using the clinical event logs, we tried to cooperate with many medical staffs for the effective clinician-specific care scheduling. To this end, we conducted classifying and grouping clinicians by the number of patients and the waiting time for consultation. Among the clusters, we set the target as a group which includes doctors who have a large number of patients and the long waiting time. In the experiment, we selected a doctor and his/her patients who got a consultation in May 2012. The event log contained 15 tasks: consultation registration, consultation, consultation scheduling, test registration, test, test scheduling, payment, sign on selective medical service, referral registration, outside image registration, admission scheduling, outside-hospital prescription printing, in-hospital prescription receiving, treatment, and certificate issuing. Also, it included several attributes such as completion time, resources, departments, patient types, and reservation time. Table 1 shows an example of the partial event log for conducting process mining analyses. After extracting the data, we cleaned it using several preprocessing steps applying existing methods [2]. In summary, the preprocessed logs had about 8,000 events which were performed for about 1,300 patients. To conduct process mining analyses, we applied *ProDiscovery* [16] which was developed by our research group. Furthermore, we used *Automod* [14] to create a simulation model and received the further simulation analysis results using *Autostat* [14]. Also, the present study was approved (IRB No. B-1409/268-107) by the Institutional Review Board of the affiliated institution.

A. PROCESS MINING ANALYSIS RESULTS

We performed the three process mining analyses as described—process discovery, arrival rate analysis, service

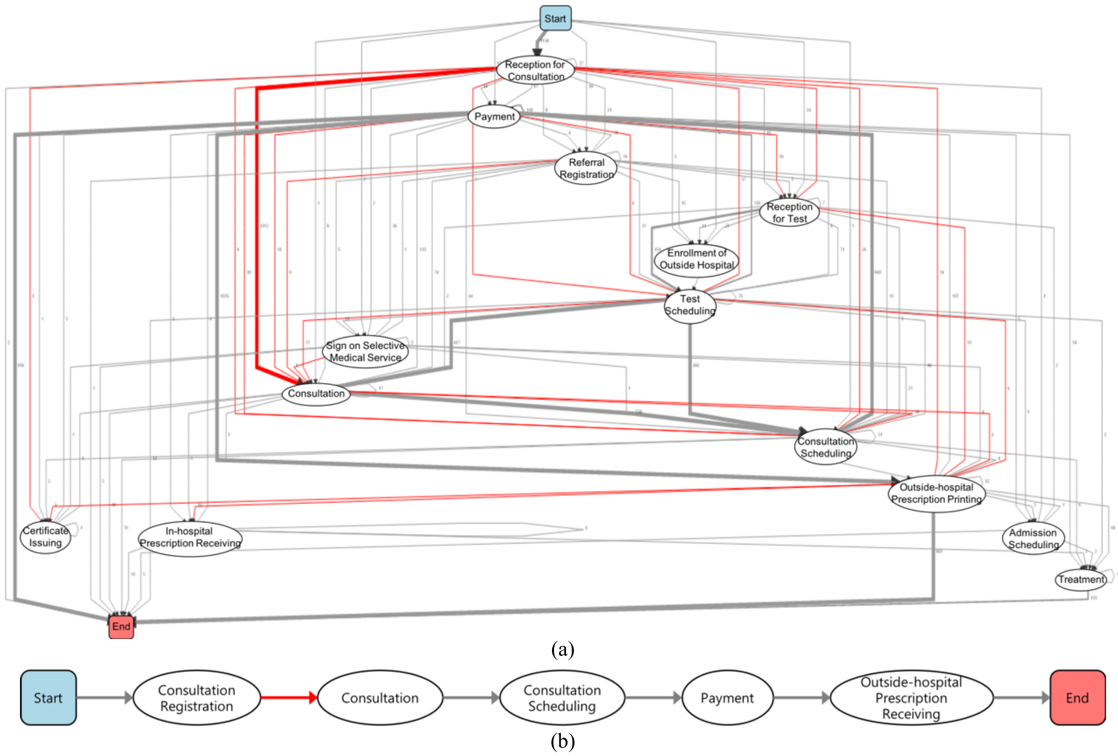

FIGURE 4. **Discovered outpatient processes of doctor A. (a) The whole outpatient process of doctor A. (b) The frequent process pattern of doctor A.**

time analysis. First, we derived a process model using the frequency mining; Fig. 4(a) represents the whole outpatient process from the event log. The derived model was very complicated and resembled a 'spaghetti process'. That is to say, we were able to discover the flows of all outpatients in the hospital using the frequency mining. However, the simulation analysis of this case study aimed to decrease the waiting time for consultation of individual patients. In other words, we had to focus on the major flow which relates to consultation. As a consequence, we tried to find out the major flow by controlling the threshold value to make a simulation model for personal clinician scheduling. Fig. 4(b) describes the frequent process pattern of the discovered outpatient process with a pre-established threshold. As a result of the discovered pattern, most of the outpatients started with the consultation registration, which took 1.11 minutes on average. Then, consultation and scheduling for the next visit were performed, and they averagely required 37.77 and 3.29 minutes, respectively. After getting the consultation scheduling, patients paid for the medical service fees, and the process was finished by receiving the outside-hospital prescription. It was identified that each of the last two activities required about 5.99 and 0.19 minutes.

Second, we calculated the average number of appointments for each reservation slot and the patients' visiting time compared to the reservation time for the arrival rates of patients. In the event log, there were 19 slots in a session which were set up every 10 minutes from 9 a.m. to 12 p.m. Table 2 shows the number of appointments per each slot depending on the patient types: new patients and follow-up patients. The average number of the new and the follow-up patients per slot was 3.71 and 0.55 respectively. In total, about 81 patients visited the hospital to get a consultation from the doctor A in a session on average.

TABLE 2. **The average number of appointments of each reservation slot.**

Reservation Slot	New Patients	Follow-up Patients	Sum
9:00:00	0.25	5	5.25
9:10:00	0.75	4.42	5.17
9:20:00	0.75	4.75	5.5
9:30:00	0.59	0.33	0.92
9:40:00	0.41	4.58	4.99
9:50:00	0.66	4.42	5.08
10:00:00	0.75	4.25	5
10:10:00	0.83	4.42	5.25
10:20:00	0.92	4.25	5.17
10:30:00	0.58	0.5	1.08
10:40:00	0.92	4.17	5.09
10:50:00	0.83	4.5	5.33
11:00:00	0.33	4.75	5.08
11:10:00	0.5	4.33	4.83
11:20:00	0.25	5.08	5.33
11:30:00	0.66	8.08	8.74
11:40:00	0.42	0.75	1.17
11:50:00	0.08	0.75	0.83
12:00:00	0.16	1.08	1.24
Average	0.55	3.71	4.26
Sum	10.64	70.41	81.05

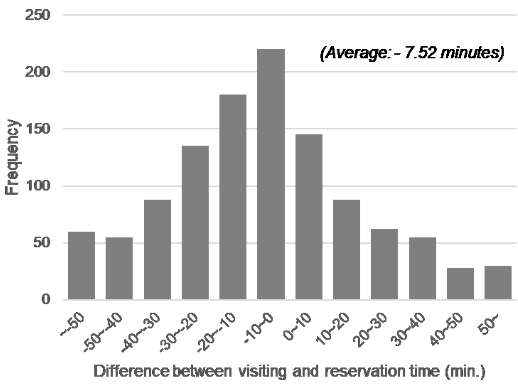

FIGURE 5. The distribution of the difference between visiting time and reservation time.

After that, we calculated the visiting time compared to the reservation time for deriving the arrival rates. Fig. 5 depicts the result that patients visited the hospital 7.52 minutes earlier than the booked time on average. Also, 837 patients (65%) arrived early at the hospital, and 451 patients (35%) were late compared to the reservation time. These two results were applied as the arrival rates in the simulation model.

Lastly, we calculated the service time for consultation using the suggested approach. The average and median of the consultation service time were 3.35 and 2.68 minutes, respectively. More specifically, there was a difference according to the patient type as 3.33 minutes for follow-up patients and 3.56 minutes for new patients on average. To make a precise simulation model, we applied the service time depending on the types of patients.

B. SIMULATION MODELING & EVALUATION RESULTS

Based on the process mining analyses results, the simulation model was created, which covers 19 reservation slots from 9 a.m. to 12 p.m. for each session. To validate the model, we performed the as-is simulation analysis. Table 3 shows the evaluation results between the calculated KPIs from logs and simulation analyses with 500 runs. First, the average of the consultation waiting time (KPI 1) and the controllable waiting time for consultation (KPI 2) from the simulation analysis was 36.41 and 31.36 minutes, respectively. Also, the average of the end time of the clinical session for the doctor (KPI 3) from the simulation analysis was 12:54:28 PM, which displayed a 6-minute time difference with the logs. As far as the statistical testing was concerned, the null hypotheses (i.e., the distribution of the values from the logs are the same with that from simulation analyses) for three KPIs were not rejected as provided in TABLE 3. After that, we conducted a further evaluation using MAPE, and it showed that MAPE values of three KPIs were less than 3.5%, which we take as an indication that our simulation model closely resembles the real patient process. Therefore, it was turned out that the model is suitable to conduct the to-be simulation analyses.

C. EXPERIMENTAL SIMULATION ANALYSIS RESULTS

As a result of the BP-based data analysis, we prepared four scenarios to decrease the waiting time: decreasing the number of appointments per reservation slot, making a break time in the middle of the clinic session, rearranging patients' reservation, and subdividing reservation intervals. A graphical explanation is provided in Fig. 6. Among them, the first two scenarios were relevant with *Extra Resources,* increasing the number of resources in a process, of 29 best practices [17]. Also, the third and the fourth scenario were constructed based on *Case Types* (i.e., distinguishing the process considering a type of cases) and *Order-based Work* (i.e., eliminating batch-processing and periodic activities), respectively. For each scenario, we give the detailed explanation of how it was created.

TABLE 3. The evaluation results between event logs and simulation models using KPIs.

Runs: 500 times (Unit: min.)	KPI1: Waiting time for consultation		KPI2: Controllable waiting time for consultation		KPI3: End time of the clinical session	
	Logs	Simulation	Logs	Simulation	Logs	Simulation
Average	35.09	35.04	31.13	30.08	12:48:52	12:54:28
UCL(95%)	36.47	37.33	32.73	32.66	12:57:35	12:57:44
LCL(95%)	33.71	32.75	29.53	28.09	12:40:09	12:51:11
t-statistic (p-value)	0.013 (0.99)		0.19 (0.85)		1.19 (0.26)	
MAPE	0.14%		3.37%		0.73%	

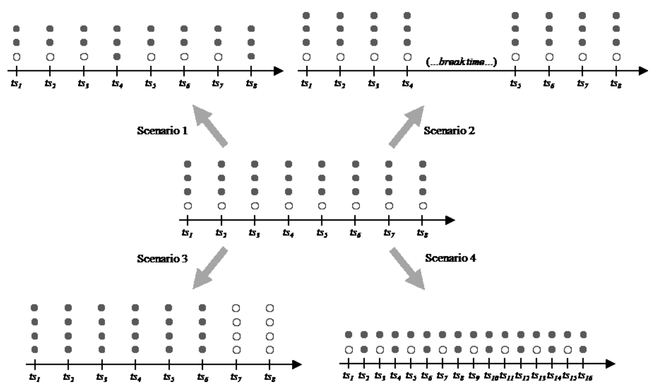

FIGURE 6. Four graphical to-be simulation scenarios.

1) DECREASING THE NUMBER OF APPOINTMENTS PER RESERVATION SLOT

One of the most influenceable factors to waiting time is the number of appointments per slot, that is to say, the number of patients. Assuming that other conditions are the same, it is evident that the fewer patients assigned to a doctor, the fewer time patients have to wait. In the left upper example in Fig. 6, we give a graphical explanation of the first scenario. In scenario 1, we tried to figure out how much waiting time is decreased as the number of patients declines from 5% to 25% at intervals of 5%.

2) MAKING A BREAK TIME IN THE MIDDLE OF CLINIC SESSION

From the event logs, we discovered a trend that the waiting time is on the rise as the time gets closer to the end of the

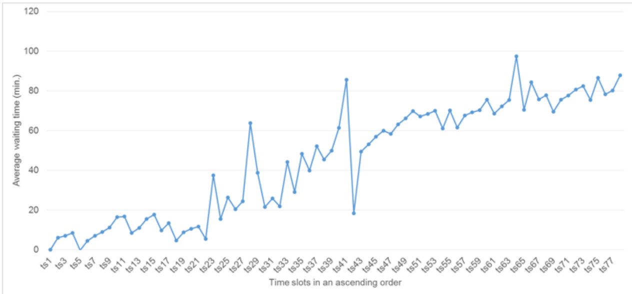

FIGURE 7. An example of average waiting time of each time slot in a clinical session.

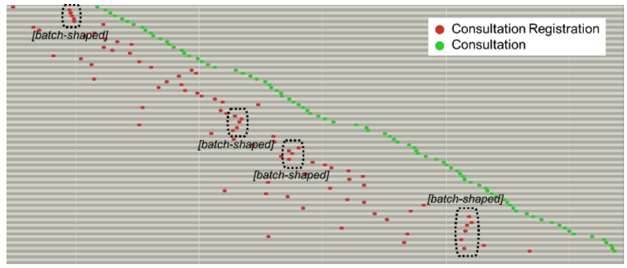

FIGURE 8. Dotted Chart Analysis – The batch shape of consultation registration.

clinic session. Fig. 7 represents an example of an average waiting time of patients who involved in each slot in a clinic session. The figure shows that the average waiting time becomes around 80 minutes at the ending session, while the value is less than 20 minutes within the first 20 time slots. One of the solutions would be to decrease the service time, but it is not realistic because hospitals have to consider patients' satisfaction and there is only so much we can decrease it. In scenario 2, to reduce the number of waiting patients, we implemented an alternative solution which creates a break time in the middle of the clinic session. In the right upper example in Fig. 6, we provide a graphical explanation of the second scenario. Simulation analyses were performed as we inserted a break time of 5 to 25 minutes at intervals of 5 minutes. We tried to figure out how much waiting time is decreased as the break time increases from 5 to 25 minutes at intervals of 5 minutes.

3) REARRANGING PATIENTS' RESERVATION

The third plan is also a solution to cope with the problem of scenario 2, the cumulated waiting patients. From the event logs, we checked out that the consultation service time is depending on the patient type. The patients who visited the hospital for the first time had longer service time than the follow-up patients because the patients should be newly observed with more time. Based on the trend, we rearranged the patients' reservation as the follow-up patients in the beginning and the new patients in the ending of the session. For example, in Fig. 6, suppose that white dots represent the patients who had longer service time. As shown on the left below example (i.e., scenario 3), we can make a scenario to reduce the cumulated waiting time by rearranging as white dots at the ending and gray colors in the beginning side.

4) SUBDIVIDING RESERVATION INTERVALS

The last scenario for decreasing the waiting time is subdividing the number of slots. In the hospital, there was a trend on the batch-shaped consultation registration due to their patient reservation system. Fig. 8 is the dotted chart of two tasks, where red and green dots represent consultation registration and consultation, respectively. In the figure, the y-axis and the x-axis are configured as patients and actual time, respectively, and the rows are sorted by consultation. In the black boxes of the figure, we can identify that multiple registrations on consultation were performed within a few minutes (i.e., batch-processing). As a result, the patients who registered relatively later than others in the same slot had to wait more to get the consultation. To solve this problem, we modified the policy from n patients every 10 minutes to the half of patients every 5 minutes. The detailed graphical explanation for this scenario is presented on the right below example in Fig. 6.

5) TO-BE SIMULATION ANALYSIS

As we explained earlier, we performed the simulation analyses based on four scenarios which decrease the consultation waiting time. To measure the impacts of each scenario, the waiting time for consultation (KPI 1) and the difference between the start time of consultation and the reserved time (KPI 2) were used among three KPIs which were applied to evaluate the simulation model. Table 4 represents the simulation analyses results in each scenario. First, in scenario 1, both KPI 1 and 2 were significantly decreased as the number of appointments per each reservation slot decreased. The reduction of 25% in scenario 1 caused a reduction of about 55% and 65% for KPI 1 and 2, respectively. Second, making a break time in the middle of a clinical session (scenario 2) moderately reduced KPI 1 and 2; a decrease of 25% led to the reduction of about 28% in KPI 1 and 34% in KPI 2. Lastly, in scenario 3 and 4, KPI 1 and 2 slightly decreased due to the adjustments in reserving patient slots and subdividing the reservation intervals.

IV. DISCUSSION

Through the simulation analysis in the experiments, it was identified that the four scheduling strategies presented in this research are effective for deriving a clinician's optimal schedule. The existing problematic scheduling method was to randomly distribute patients in every ten-minute reservation slot manually. Thus, scheduling was performed arbitrarily, and it caused long consultation waiting time for patients. Compared to the existing method, our research provided four improved scheduling approaches that accomplish the reduction of at least 2.84% (scenario 4) and up to 55.20% (scenario 1) in the waiting time for consultation (KPI 1). More in detail, in the case of scenario 1, it becomes evident that the number of patients has a significant impact on the consultation waiting time. Also, the measured values also decreased due

TABLE 4. Scenario-based simulation analysis results.

Scenario 1: Decreasing the number of appointments per reservation slot						
	Current	-5%	-10%	-15%	-20%	-25%
KPI 1 (min.)	35.04	32.15	27.81	22.01	17.44	15.70
(Rate of change (%))	(-)	(-8.24)	(-20.62)	(-37.20)	(-50.23)	(-55.20)
KPI 2 (min.)	30.38	27.11	22.90	17.23	12.83	10.65
(Rate of change (%))	(-)	(-10.77)	(-24.63)	(-43.30)	(-57.78)	(-64.96)
Scenario 2: Making a break time in the middle of clinical session						
	Current	5 min.	10 min.	15 min.	20 min.	25 min.
KPI 1 (min.)	35.04	33.73	32.44	29.43	27.12	25.24
(Rate of change (%))	(-)	(-3.74)	(-7.41)	(-16.01)	(-22.59)	(-27.97)
KPI 2 (min.)	30.38	29.02	27.60	24.73	22.05	20.09
(Rate of change (%))	(-)	(-4.48)	(-9.16)	(-18.59)	(-27.41)	(-33.85)
Scenario 3: Adjusting patients' reservation						
	Current	Adjusted				
KPI 1 (min.)	35.04	34.04				
(Rate of change (%))	(-)	(-2.84)				
KPI 2 (min.)	30.38	29.53				
(Rate of change (%))	(-)	(-2.81)				
Scenario 4: Subdividing reservation intervals						
	Current	Adjusted				
KPI 1 (min.)	35.04	33.40				
(Rate of change (%))	(-)	(-4.69)				
KPI 2 (min.)	30.38	28.93				
(Rate of change (%))	(-)	(-4.78)				

to inserting a break in the middle of the session (scenario 2). That is, the methods in scenario 1 and 2 can be considered as highly substantial improvements of a clinician's appointment schedule when the goal is to decrease patients' waiting time. Furthermore, the other two scenarios, the arrangement of patient groups and subdividing reservation intervals, were also proved as impactful approaches to decrease consultation waiting time.

Interestingly, strategy 1 and 2 potentially negatively affect hospital revenue. After all, the average number of consulted patients per day decreases. Thus, it may require additional expenses as compared to the current system. On the other hand, approach 3 and 4 simply affect the reservation policies without diminishing the number of patients or requiring doctors to spend more time. Therefore, they can be classified as strategies that do not incur any additional expenses. In summary, although all four scheduling scenarios are effective strategies to reduce waiting time, it is necessary to use them as appropriate or mixed strategies depending on multiple factors in the hospital.

After discussing with domain experts in the hospital, we received comments on the results of our simulation analyses. They considered that the methods of case 3 and 4 to change the individual clinicians' schedules are indeed applicable and attractive.

The proposed approach has a significant contribution that provides how to derive parameters required for building a healthcare process simulation using process mining. The process simulation, as it is known, has the main advantage of being able to experiment in advance on existing issues under the simulated environment. In that sense, it considerably simplifies the application of DES in a clinical setting with the data-driven and highly automated approaches. Furthermore, our approach is valuable in that three process mining analyses (e.g., process discovery, arrival rate, and service time analysis) consider the specific characteristics of hospitals and the data they have at their disposal, and that these results are directly reflected in the simulation model. Based on our approach, simulation models can be utilized in diverse healthcare settings to determine improved personal schedules for clinicians.

Also, our approach provides a systematic method that overcomes the limitations of the existing works for scenario-based simulation analysis. This helps to break away from the traditional rule-of-thumb approach and reduces computing time and power with efficient simulation analysis.

Also, our approach has extensive flexibility. In this paper, we focused on how to solve the problem of optimizing personal clinical schedules. However, our approach can handle other processes in the healthcare environment such as clinical test or reception processes. In addition, it can support other service processes such as banking and public office task processes similar to the outpatient process.

To clarify the distinctive traits of our work, we compared our approach with the existing works that adopt the evidence-based aspect. As a result, it was identified that the existing works [29]–[32] consider the domain knowledge or theoretical foundation as the evidence base. Thus, they put the emphasis on the generalized knowledge demonstrated by various sources. Different from these works, the proposed approach determines that the evidence base is the facts discovered from data, i.e., data-driven knowledge. Therefore, this study is differentiated by the approach of a customized and practical evidence-based approach to finding problems with specific data, rather than conventional and generalized problems from the existing literature.

Our work also has several limitations. As far as the proposed framework is concerned, it still needs further automated approaches. In the framework, for example, it is relevant to create a simulation model from process mining analysis results or prepare an improved simulation model that reflects the scenario. In particular, techniques that automatically reflects the improvements based on redesign best practices in the simulation model can maximize the effectiveness of the simulation analysis. Also, this paper covers a single case study to validate our framework. Future research should strive to conduct further case studies. Lastly, we are working to develop the decision support system that supports our framework. It will be helpful for practitioners for effective hospital management.

V. CONCLUSION

In this paper, we suggested a decision support framework for optimizing clinician medical scheduling using discrete event simulation approach, which is constructed based on three process mining analysis including process discovery, arrival rate analysis, and service time analysis. Furthermore, it covered how to derive effective improvement methods to decrease waiting time for consultation. In the case study, we applied the real-world data to the proposed framework. Also, we performed the four scenario-based experiments using the simulation model for the personalized care scheduling. As a result, we showed that not only two cases which need additional costs have a significant effect on the waiting time, but also the changes of reservation systems which do not require more costs decreased the waiting time.

As we stated in section IV, we plan to work on developing a decision support system for medical scheduling. Also, we will improve our framework to make it more automated. More case studies should be performed to validate our approach.

REFERENCES

[1] A. Berhane and F. Enquselassie, "Patients' preferences for attributes related to health care services at hospitals in amhara region, northern ethiopia: A discrete choice experiment," *Patient Preference Adherence*, vol. 9, pp. 1293–1301, Sep. 2015.

[2] R. P. J. C. Bose, "Wanna improve process mining results?: It's high time we consider data quality issues seriously," in *Proc. IEEE Symp. Comput. Intell. Data Mining (CIDM)*, Apr. 2013, pp. 127–134.

[3] W. Cao *et al.*, "A Web-based appointment system to reduce waiting for outpatients: A retrospective study," *BMC Health Services Res.*, vol. 11, no. 1, p. 318, Nov. 2011.

[4] T. J. Carney *et al.*, "Using computational modeling to assess the impact of clinical decision support on cancer screening improvement strategies within the community health centers," *J. Biomed. Informat.*, vol. 51, pp. 200–209, Oct. 2014.

[5] T. Cayirli and E. Veral, "Outpatient scheduling in health care: A review of literature," *Prod. Oper. Manage.*, vol. 12, no. 4, pp. 519–549, Dec. 2003.

[6] M. Cho, M. Song, and S. Yoo, "A systematic methodology for outpatient process analysis based on process mining," *Int. J. Ind. Eng., Theory Appl. Pract.*, vol. 22, no. 4, pp. 480–493, Jul. 2015.

[7] A. K. A. De Medeiros, A. J. M. M. Weijters, and W. M. P. Van Der Aalst, "Genetic process mining: An experimental evaluation," *Data Mining Knowl. Discovery*, vol. 14, no. 2, pp. 245–304, Apr. 2007.

[8] P. K. Sahoo, S. K. Mohapatra, and S.-L. Wu, "Analyzing healthcare big data with prediction for future health condition," *IEEE Access*, vol. 4, pp. 9786–9799, 2016.

[9] C. W. Günther and W. M. P. Van Der Aalst, "Fuzzy mining—Adaptive process simplification based on multi-perspective metrics," *Business Process Management* (Lecture Notes in Computer Science), vol. 4714. pp. 328–343, 2007.

[10] N. R. Hoot *et al.*, "Forecasting emergency department crowding: A discrete event simulation," *Ann. Emergency Med.*, vol. 52, no. 2, pp. 116–125, Aug. 2008.

[11] M. Cho *et al.*, "Evaluating the effect of best practices for business process redesign: An evidence-based approach based on process mining techniques," *Decis. Support Syst.*, vol. 104, pp. 92–103, Dec. 2017.

[12] S. H. Jacobson, S. N. Hall, and J. R. Swisher, "Discrete-event simulation of health care systems,"' in *Patient Flow: Reducing Delay in Healthcare Delivery*, vol. 206, R. Hall, Ed. Boston, MA, USA: Springer, 2013, pp. 273–909.

[13] S. J. J. Leemans, D. Fahland, and W. M. P. Van Der Aalst, "Discovering block-structured process models from event logs—A constructive approach," in *Application and Theory of Petri Nets and Concurrency* (Lecture Notes in Computer Science), vol. 7927. Berlin, Germany: Springer, 2013, pp. 311–329.

[14] D. Müller, "AutoMod: Modeling complex manufacturing, distribution, and logisitics systems for over 30 years," in *Proc. Winter Simulation Conf., Simulation, Making Decis. Complex World*, Dec. 2013, pp. 4037–4051.

[15] C. Nessim, J. Winocour, D. P. Holloway, R. Saskin, and C. M. Holloway, "Wait times for breast cancer surgery: Effect of magnetic resonance imaging and preoperative investigations on the diagnostic pathway," *J. Oncol. Pract.*, vol. 11, no. 2, p. e131-8, Feb. 2015.

[16] P. Analyzer. *Process Analyzer Website*. Accessed: Dec. 28, 2018. [Online]. Available: http://demo.prodiscovery.co.kr

[17] H. A. Reijers and S. L. Mansar, "Best practices in business process redesign: An overview and qualitative evaluation of successful redesign heuristics," *Omega*, vol. 33, no. 4, pp. 283–306, Aug. 2005.

[18] A. Rozinat, R. S. Mans, M. Song, and W. M. P. van der Aalst, "Discovering simulation models," *Inf. Syst.*, vol. 34, no. 3, pp. 305–327, May 2009.

[19] S. Yoo *et al.*, "Assessment of hospital processes using a process mining technique: Outpatient process analysis at a tertiary hospital," *Int. J. Med. Inform.*, vol. 88, pp. 34–43, Apr. 2016.

[20] M. H. Rutberg, S. Wenczel, J. Devaney, E. J. Goldlust, and T. E. Day, "Incorporating discrete event simulation into quality improvement efforts in health care systems," *Amer. J. Med. Qual.*, vol. 30, no. 1, pp. 31–35, Jan./Feb. 2015.

[21] M. Song, H. Yang, S. H. Siadat, and M. Pechenizkiy, "A comparative study of dimensionality reduction techniques to enhance trace clustering performances," *Expert Syst. Appl.*, vol. 40, no. 9, pp. 3722–3737, Jul. 2013.

[22] M. Song and W. M. P. van der Aalst, "Towards comprehensive support for organizational mining," *Decis. Support Syst.*, vol. 46, no. 1, pp. 300–317, Dec. 2008.

[23] L. Deng, Y. Hu, J. P. Y. Cheung, and K. D. K. Luk, "A data-driven decision support system for scoliosis prognosis," *IEEE Access*, vol. 5, pp. 7874–7884, 2017.

[24] W. M. P. Van Der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, vol. 136. Berlin, Germany: Springer, 2011.

[25] W. M. P. van der Aalst, M. H. Schonenberg, and M. Song, "Time prediction based on process mining," *Inf. Syst.*, vol. 36, no. 2, pp. 450–475, Apr. 2011.

[26] A. J. M. M. Weijters, W. M. P. Van Der Aalst, and A. K. A. De Medeiros, "Process mining with the HeuristicsMiner algorithm," Eindhoven Univ. Technol., Eindhoven, The Netherlands, Tech. Rep. WP 166, 2006, pp. 1–34.

[27] H. Zhu, M. Hou, C. Wang, and M. Zhou, "An efficient outpatient scheduling approach," *IEEE Trans. Autom. Sci. Eng.*, vol. 9, no. 4, pp. 701–709, Oct. 2012.

[28] B. Wang, X. Xia, H. Meng, and T. Li, "Bed-scenario-set robust optimization framework with two objectives for uncertain scheduling systems," *IEEE/CAA J. Autom. Sinica*, vol. 4, no. 1, pp. 143–153, Jan. 2017.

[29] K. T. Waxman, "The development of evidence-based clinical simulation scenarios: Guidelines for nurse educators," *J. Nursing Educ.*, vol. 49, no. 1, pp. 29–35, Jan. 2010.

[30] C. M. Clancy and K. Cronin, "Evidence-based decision making: Global evidence, local decisions," *Health Affairs*, vol. 24, no. 1, pp. 151–162, Jan./Feb. 2005.

[31] A. Barratt, "Evidence based medicine and shared decision making: The challenge of getting both evidence and preferences into health care," *Patient Educ. Counseling*, vol. 73, no. 3, pp. 407–412, Dec. 2008.

[32] K. Walshe and T. G. Rundall, "Evidence-based management: From theory to practice in health care," *Milbank Quart.*, vol. 79, no. 3, pp. 429–457, Sep. 2001.

[33] H. Kaur and S. K. Wasan, "Empirical study on applications of data mining techniques in healthcare," *J. Comput. Sci.*, vol. 2, no. 2, pp. 194–200, Feb. 2006.

MINSU CHO received the Ph.D. degree in management engineering from the Ulsan National Institute of Science and Technology, Ulsan, South Korea, in 2018. He is currently a Post-Doctoral Research Associate with the Pohang University of Science and Technology.

His work has been published in *Decision Support Systems* and the *International Journal of Medical Informatics*, and at the BPM Conference. His research interests include business process management, process mining, manufacturing process analysis, clinical pathway analysis, simulation analysis based on process mining, and process mining case studies.

MINSEOK SONG received the Ph.D. degree in industrial and management engineering from the Pohang University of Science and Technology (POSTECH), in 2006, where he is currently an Associate Professor with the Department of Industrial & Management Engineering. Prior to this, he stayed with the Information Systems Department of the Technology Management, Eindhoven University of Technology, as a Post-Doctoral Researcher, from 2006 to 2009. Also, he was an Assistant/Associate Professor with the Ulsan National Institute of Science and Technology.

He has published more than 60 scientific papers in several top-level venues such as *Decision Support Systems*, *Information Systems*, the *Journal of Information Technology*, and the *International Journal of Medical Informatics*. His research interests include business process management, process mining, business analytics, simulation, and social network analysis.

SOOYOUNG YOO received the Ph.D. degree from the Department of Biomedical Engineering, College of Medicine, Seoul National University, Seoul, South Korea, in 2008. She is currently an Assistant Professor with the Office of eHealth Research and Businesses, Seoul National University Bundang Hospital, South Korea, where she is also the Chief of Healthcare ICT Research Center. She has been with the Seoul National University Bundang Hospital, since 2009, and participated in various health IT research projects ranging from electronic health record and information sharing to mobile health and data analytics.

Her major research interests include health IT interoperability standards, clinical decision support systems, information extraction and information retrieval, mobile health, and smart hospital technologies.

HAJO A. REIJERS is currently a Full Professor in business process management and analytics with the Department of Information and Computing Sciences, Utrecht University. Furthermore, he is a part-time Full Professor with the AIS Group, Department of Mathematics and Computer Science, Eindhoven University of Technology (TU/e). His research and teaching focus on business process management, workflow technology, business process improvement, and conceptual modeling. On these and related topics, he published over 150 research papers, book chapters, and professional publications. He is also the Managing Director of the European BPM Round Table, an initiative to connect researchers and practitioners in the area of business process management (≪http://www.bpmroundtable.eu/≫). He is closely cooperating with companies from the services and healthcare domains, and with various international scholars.