Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

WORKSHIFT SCHEDULING USING OPTIMIZATION AND PROCESS MINING TECHNIQUES: AN APPLICATION IN HEALTHCARE

Alberto Guastalla
Emilio Sulis
Roberto Aringhieri

Department of Computer Science
University of Turin
185 Corso Svizzera
10149, Torino, Italy

Stefano Branchi
Chiara Di Francescomarino
Chiara Ghidini

Fondazione Bruno Kessler
Process & Data Intelligence Unit
18 Via Sommarive
38123, Povo, Italy

ABSTRACT

This paper aims at supporting healthcare organizations in automatically generating rostering plans by combining optimization and process mining approaches. Based on event logs from the information system, we propose a decision support system that simulates work schedules. Managing staff workshifts is a complicated issue to solve especially in large and complex organizations such as those in the healthcare sector. A number of different factors can be taken into account, i.e., operative constraints, personal preferences and regulations must be considered in order to produce the best plan. In our approach we exploit the idea for which the patterns included in the realised rostering plans could represent the personal needs and the unspoken habits of the personnel. Based on this remark, we propose a three-step methodological framework – rostering optimization, pattern extraction, pattern adaptation – that it was applied to a real-world scenario.

1 INTRODUCTION

One of the problems that typically affect organizations is managing staff rostering. Workshift scheduling is particularly relevant for service organizations dealing with critical activities, such as healthcare, due to the complexity of managing medical care.

A number of factors need to be taken into account when planning shifts, starting with the operative and contractual constraints, as well as personal needs and the often unspoken habits of staff. All of these factors are usually considered explicitly or implicitly by the staff management who manually prepares shift schedules. In addition to the human effort required, such a schedule may not always be the best plan, considering the high complexity of the problem.

In this paper, we aim to provide support to healthcare organizations in automatically generating rostering plans by combining optimization and process mining approaches, which is the main novelty of this paper. Exploiting the idea for which the patterns included in the realised rostering plans could represent the personal needs and the unspoken habits of the personnel, we propose a methodological framework based on the following three different steps: rostering optimization, pattern extraction, pattern adaptation. We present our decision support method for healthcare management based on real event logs.

In the remainder of the paper we first introduce background in Section 2. Then we describe the three phases of the overall approach: the optimization phase (Section 3), the pattern extraction phase (Section 4), the model adaptation phase (Section 5). Finally, while Section 6 provides insights about the results, Section 7 concludes the paper.

2 BACKGROUND

In this section we discuss first the related work (Section 2.1) and then we report the details of the case study we aim at investigating in the paper (Section 2.2).

2.1 Related work

Simulation and decision-making support systems in healthcare have been extensively applied in different healthcare sectors (Moreira et al. 2019). Hybrid models that combine simulation and optimization are becoming more popular (Lather and Messner 2018). A recent work proposes a robust model that plans bed capacity and optimizes staffing levels (Ordu et al. 2021). Similarly, in (Ahmed and Alkhamis 2009), optimization with integer linear programming integrated with forecasting and DES simulation provides a decision support tool for an hospital emergency department to improve service efficiency.

Staff rostering problems have been widely investigated in the literature. For detailed surveys on this topic, we refer to Ernst et al. (2004), Ernst et al. (2004) and the two volumes by Jiang et al. (2004). The nurse rostering problem in particular has been widely addressed in the literature, as reported in Burke et al. (2004). Real life applications of nurse rostering have been investigated in Kellogg and Walczak (2007), and some applications are available in Bellanti et al. (2004) and Addis et al. (2012).

In the area of business process analysis, the most interesting developments have been precisely in the exploitation of data about the processes collected in information systems (Dumas et al. 2018). The relationship between simulation and information systems has been discussed extensively in recent years (Beese et al. 2019), and some efforts has been done to adopt PM to build a business process simulation model (Martin et al. 2016). Health Information Systems (HIS) can be leveraged to retrieve information on actual event logs. Process mining is a discipline that aims to extract knowledge directly from data collected in the information system of an organization (van der Aalst 2016). Challenges related to healthcare processes, which we discuss in this article, have been systematized recently (Munoz-Gama et al. 2022).

Recent work summarizes the role of knowledge extraction using simulation-based optimization in organizations by performing an application study to provide a flexible pattern mining algorithm in a DSS (Karlsson et al. 2015). Some efforts using healthcare event logs concern the interarrival times (Martin et al. 2015). Applications to the emergency department is reported in Duma and Aringhieri (2020).

A recent work focused on the extraction of significant pathways from a medico-administrative database by adopting a process mining approach (Phan et al. 2019). Authors derived a causal network from healthcare even logs, to be converted into a state-chart model. The corresponding simulation addressed *what-if* scenarios to propose an improved care pathway by focusing on times of occurrence of complications and associated costs.

2.2 Case study

The Presidio Sanitario Cottolengo is a multidisciplinary hospital in Turin, one of the largest cities in Italy. The hospital is able to provide about 150 beds, from General Medicine to Oncology Department. The surgical area includes multiple disciplines as General Surgery, Orthopedics, Urology, Ophthalmology, as well as Rehabilitation, Breast Unit, and Long-term Care. Around 5,452 hospitalizations and more than 350,000 medical services are performed every year, totalling about 40,000 daily hospital stays.

Dataset. Our dataset includes information about personnel on workshifts. In particular, we focus on the most stable positions such as nurses (NUR) and health and social workers (OSS). The dataset includes two types of data, related to 131 employees (NUR and OSS):

- the *workshift plan*, that is a plan related to the planned workshifts for the months of January and February 2021 in different wards: surgical area, long-term care, medical, oncology and rehabilitation ward;

- the *stamping log*, that is data related to the daily entrance and exit stamping carried out by the personnel starting or concluding the workshift during the same months in the same wards.

Table 1: Example of *workshift plan* related to the month of February.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
E_1	3	3	R	R	2	2	2	R	R	3	3	R	R	1	1	2	2	F	R	R	R	1	2	2	2	2	R	3
E_2	2	2	R	R	1	1	1	1	1	R	R	1	1	1	1	R	R	1	1	1	1	R	R	1	1	1	1	R
				...									...									...						
E_n	R	1	2	R	1	1	1	2	2	R	R	1	1	3	3	R	R	2	2	2	R	R	1	3	3	R	R	1

The *workshift plan* contains, for each day of the month, and for each employee the workshift that has been planned, e.g., the morning workshift (indicated with "1"), the afternoon workshift (2), the night workshift (3), the rest workshift (R), the day off (F). For instance, Table 1 shows an example of the planned workshifts related to the month of February. The employee E_1 for instance has a night workshift on February 1 and February 2 and then a rest workshift on February 3 and 4.

Table 2: Excerpt of an example of *stamping log*.

E_1	01-02-2021	21.25	Entrance
E_1	02-02-2021	06.19	Exit
E_2	01-02-2021	13.32	Entrance
E_2	01-02-2021	22.23	Exit
...	...	...	...

The *stamping log* contains, instead, for each day of work and for each employee, the entrance and exit stamping and the corresponding timestamp. For instance, Table 2 shows an example excerpt of a *stamping log*. The log shows that for instance E_1 on February 1 started her workshift at 21.25 and ended it on February 2 at 6.19, i.e., she carried out the night workshift.

3 ROSTERING OPTIMIZATION

In this section, we present a multi-criteria mixed integer linear programming model (MMIP), which models the problem of balancing the monthly working hours of the healthcare personnel in accordance with a list of operative constraints. The following mathematical models handles both NUR and OSS since they have the same operative and contractual contexts.

We introduce the following notation to represent the parameters of the model. Let P, D, and T be respectively the sets of the (i) healthcare personnel, (ii) the days in the time horizon, and (iii) the possible workshifts. Specifically, the set T is composed of four possible workshifts having different durations, that is Morning (7h 44m), Afternoon (7h 44m), Night (8h 14m), Rest (0h 0m). Such durations are denoted with h_t, $t \in T$. Let δ be the total number of working hours required for each employee during the time horizon D. For personal reasons, some employees cannot be considered for the night shift. Let A_{pd} be a 0-1 parameter representing the availability of the employee p to cover a night in $d \in D$. The S_{dt} indicates the minimum number of employees required on day $d \in D$ for the shift $t \in T$.

We introduce the following decision variables. The variable $x^t_{pd} \in \{0,1\}$ is set to 1 when the employee $p \in P$ is assigned to the shift $t \in T$ on day $d \in D$. The variable $y_p \in \mathbb{R}_+$ accounts for the workload assigned to the employee $p \in P$ during the whole time horizon. Let $\alpha = \max_{p \in P} y_p$ be the maximum workload assigned. The variable $\phi_t \in \mathbb{Z}_+$ accounts for the total number of shift $t \in T$ assigned to the personnel.

The constraints (1) model the shift assignment to the personnel.

$$\sum_{t \in T} x^t_{pd} = 1\,, \qquad \forall p \in P, \forall d \in D \tag{1a}$$

$$x^3_{pd} \leq A_{pd}\,, \qquad \forall d \in D, \forall p \in P \tag{1b}$$

$$\sum_{p \in P} x^t_{pd} \geq S_{dt}\,, \qquad \forall d \in D, \forall t \in T \tag{1c}$$

$$\sum_{d \in D} x^t_{pd} \leq \phi_t\,, \qquad \forall p \in P, \forall t \in T \tag{1d}$$

Constraints (1a) guarantee that the assignment of a shift $t \in T$ (including the rest shift) to a employee $p \in P$ on day $d \in D$ while constraints (1b) ensure that night shifts are not assigned to personnel who are not available for this shift. Constraints (1c) guarantee that the minimum number of employees has been assigned to the shift $t \in T$ on day $d \in D$. Finally, constraints (1d) measure the number of employees assigned to each shift $t \in T$ during the time horizon.

$$\sum_{d \in D} \sum_{t \in T} h_t x^t_{pd} = y_p\,, \qquad \forall p \in P \tag{2a}$$

$$y_p \geq \delta\,, \qquad \forall p \in P \tag{2b}$$

$$\alpha \geq y_p\,, \qquad \forall p \in P \tag{2c}$$

Constraints (2) are concerned with the workload assigned to the personnel. Constraints (2a) count the hour workload assigned to each employee $p \in P$ while constraints (2b) impose that such a workload holds the minimum required. Constraints (2c) determine the maximum workload.

$$\sum_{\ell=d}^{d+5} \sum_{t \in T \setminus \{\text{rest}\}} x^t_{p\ell} \geq 6k_{pd}\,, \qquad \forall p \in P, \forall d \in D : d \leq |D| - 5 \tag{3a}$$

$$\sum_{\ell=d}^{d+5} \sum_{t \in T \setminus \{\text{rest}\}} x^t_{p\ell} - 5 \leq 6k_{pd}\,, \qquad \forall p \in P, \forall d \in D : d \leq |D| - 5 \tag{3b}$$

$$\sum_{p \in P} \sum_{d \in D} k_{pd} = 0 \tag{3c}$$

Introducing the dummy 0-1 variable k_{pd} with $p \in P$ and $d \in D$, the set of constraints (3) guarantee that no one works more that 5 days in a row in accordance with the contract rules. More specifically, constraints (3a) and (3b) set the corresponding variable k_{pd} to 1 when 6 shifts in a row are assigned to a employee. But the constraints (3c) avoid this possibility setting to 0 the variable k_{pd}.

$$\sum_{\ell=d}^{d+2} x^3_{p\ell} \geq 3\psi_{pd}\,, \qquad \forall p \in P, \forall d \in D : d \leq |D| - 2 \tag{4a}$$

$$\sum_{\ell=d}^{d+2} x^3_{p\ell} - 2 \leq 3\psi_{pd}\,, \qquad \forall p \in P, \forall d \in D : d \leq |D| - 2 \tag{4b}$$

$$\sum_{\ell=d+3}^{d+4} x^4_{p\ell} \geq 2\psi_{pd}\,, \qquad \forall p \in P, \forall d \in D : d \leq |D| - 4 \tag{4c}$$

$$\psi_{pd} \leq x^4_{p(d+3)}\,, \qquad \forall p \in P, \forall d \in D : d = |D| - 3 \tag{4d}$$

$$x^3_{p(d-3)} + x^3_{p(d-2)} + x^3_{p(d-1)} + x^3_{pd} \leq 3\,, \qquad \forall p \in P, \forall d \in D : d > 3 \tag{4e}$$

Analogously to constraints (3), the constraints (4) model the contract rule for which after three night shifts there will be two rest shifts. This has been done introducing the dummy variable ψ_{pd}, which is equal to 1 when the employee $p \in P$ on day $d \in D$ has the third in a row nigh shift (constraints (4a) and (4b)).

Constraints (4c) impose two rest shifts after three night shifts aside from the case at the end of the planning horizon in which we can assign just one rest shift (constraints (4d)). Constraints (4e) ensure that a employee has no more than three night shift in a row. Finally, constraints similar to (4a) and (4b) to set a dummy variable σ_{pd} equals to 1 when a employee has two night shifts in a row are not reported to save space. Note that the variables σ_{pd} are used in the objective function.

$$\sum_{d}^{D} x_{pd}^{t} \geq w_t - \beta_p^t, \qquad \forall p \in P, \forall t \in T \tag{5}$$

Constraints (5) are introduced to guarantee the shift allowance to the largest number of employees. The allowance is given if a employee did a minimum number (w_t with $t \in T$) of morning, afternoon and night shifts at the end of the time horizon. By consequence, the planning of the shift has to offer this possibility to the personnel. Since this requirement could lead to infeasibility (especially when employees are on holiday), we added the dummy variable β_p^t to account the number of missing shifts t for the employee p during the months. The idea is to use a *soft constraint* for this requirement using the objective functions.

Before introducing the multi-criteria objective function, we would remark that the complete model accounts also for (i) shift pre-assignment, (ii) holidays, and (iii) special work permits. We did not report these constraints as they do not affect the complexity of our three-step approach.

To take into account of different needs discussed with the hospital management, we adopt a multi-criteria objective function as follows.

$$\begin{aligned} \min z = {} & \nu_1 \alpha + \nu_2 \sum_{p \in P} \sum_{t \in T} \beta_p^t - \nu_3 \sum_{d \in D} \left(\lambda_1 \sum_{p \in P} x_{pd}^1 + \lambda_2 \sum_{p \in P} x_{pd}^2 + \lambda_3 \sum_{p \in P} x_{pd}^3 \right) \\ & + \nu_4 \tau - \nu_5 \sum_{p \in P} \sum_{d \in D} \sigma_p^d, \end{aligned} \tag{6}$$

where $\sum_i^5 \nu_i = 1$ and $\tau \geq \phi_t$, $\forall t \in T$. The first term of (6) represents the needs of fairly distribute the workload among the personnel minimizing the maximum workload assigned. The second term of (6) fosters the shift assignment in such a way to make possible the shift allowance to everyone. As soon as the personnel minimum requirement is satisfied by constraints (1c), the third term of (6) *distributes* the exceeding employees (if any) among the three shifts prioritising the morning shift, then the afternoon shift, and finally the night shift (this is done setting the weights $\lambda_1 > \lambda_2 > \lambda_3$ and their sum equals to 1. Finally, the fourth and the fifth terms of (1c) pushes respectively a fair distribution of the type of shifts and of the three nights in a row among the personnel.

We would remark that the objective function (6) is a convex combination of five convex terms, each one representing different criteria. Therefore, the Geoffrion's theorem guarantees to compute a strict supported Pareto solution (Geoffrion 1968).

4 PATTERN EXTRACTION

In order to be able to extract the most frequent shift patterns adopted by the personnel of the Cottolengo Hospital, we extracted a *hybrid workshift log* obtained by combining the information of the planned shifts and the data (employee code and timestamp) related to the actual entrance and exit stamping (*stamping log*) carried out by the personnel when starting or concluding the workshift. To this aim, we used the data related to the actual workshifts for identifying, for each NUR and each OSS, and for each day of the month, whether the employee has worked or not. In the former case, the timestamp of the stamping allows us to understand whether the workshift has been carried out in the morning (1), in the afternoon (2) or in the night (3). In the latter case, if the non-working time is aligned with the information related to the planned workshift, the details about the non-working day can be inferred, e.g., if it is a rest workshift (R), a day off (F), a sick leave (M). When, instead, a disalignment occurs between the planned and the actual workshift, no heuristics can be leveraged and hence a generic no-workshift (X) activity is identified.

The *hybrid workshift log* can be seen as an event log, in which each case is the history of the workshifts of an employee (for the two-month span of the study). This means that all cases in the *hybrid workshift log* have the same length. Each event in a case represents the workshift carried out in a specific day (i.e., the timestamp associated to an event corresponds to the day in which the workshift has been carried out and two consecutive events are carried out in two consecutive days) and is labeled with the "activity" carried out by the employee in the specific day, e.g., workshift in the morning (1), workshift in the night (3), rest workshift (R), sick leave (M) and so on. Besides the workshift types in the list, which are the most frequent ones, other types of workshifts - with a lower frequency - can be found in the *workshift plan*. We omit here the complete list as it is not relevant for the purpose of the paper. Moreover, since we analyzed the data related to different wards, the information related to the ward in which the employee has carried out the workshift is also reported as data attribute for each event.

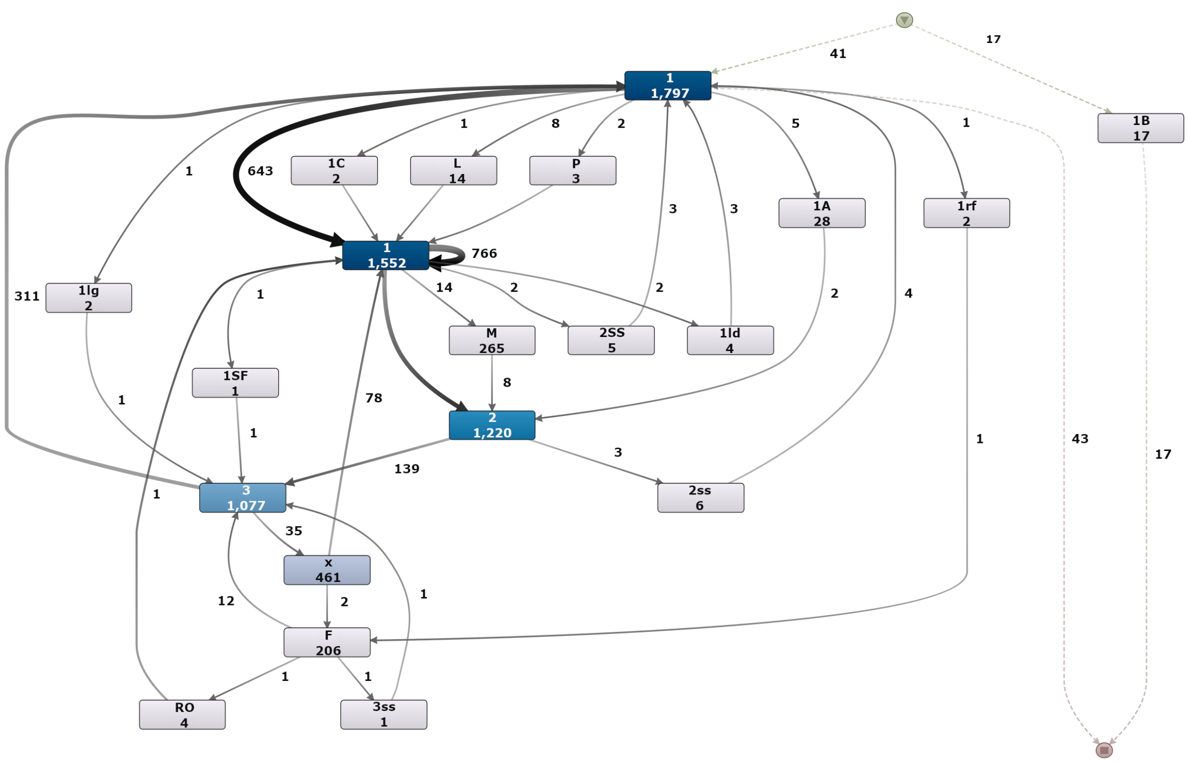

Figure 1: Employee workshift process of the Cottolengo Hospital extracted with Disco.

For instance, Figure 1 provides an overview of the main workshift paths followed by the Cottolengo Hospital employees extracted with Disco (https://www.fluxicon.com/). The model already reveals that the most frequent workshifts are the rest ones, followed by the ones carried out in the morning and by the afternoon ones.

In order to mine frequent contiguous sequential patterns from data, that is to identify constraints that are possibly not explicitly reported as domain knowledge, we resorted to a sequence pattern mining algorithm. We specifically look at the frequency of a pattern in terms of number of employees who followed the pattern at least once. To this aim, we specifically leveraged the implementation of the R CSeqPat library (https://cran.r-project.org/web/packages/CSeqpat/index.html).

Table 3 reports the frequent contiguous patterns that have been extracted from the *hybrid workshift log* related to all the investigated wards. For each pattern, we also report in the table the number of occurrences of the pattern. For instance, two most frequent patterns are mined from the algorithm. The first one is related to one of the explicit constraints already used in the optimization model - that is that three consecutive night workshifts have to be followed by a rest workshift (see constraint 4e). The second pattern states instead that two morning workshifts are usually followed by an afternoon one. The mined patterns have been validated by an employee of the Cottolengo hospital.

Table 3: Patterns extracted from the *hybrid workshift log* through a sequence pattern mining algorithm.

Patterns	**Number of Occurrences**	**Patterns**	**Number of Occurrences**
(3, 3, 3, R)	97	(R, 1, 1, 1)	82
(1, 1, 2)	97	(3, 3, R, 1)	80
(2, R, R)	93	(2, 2, R, R)	78
(1, 2, 2)	90	(2, 3, 3)	77
(1, 1, 1)	90	(3, 3, 3, R, 1)	76
(R, R, 1, 1)	87	(1, 1, R)	75
(R, 1, 1, 2)	82	(2, 2, 2)	74

5 PATTERN ADAPTATION

The last step of our three-step approach concerns the pattern adaptation. The basic idea is to consider the rostering solution provided by the model reported in Section 3 in such a way to adapt the provided workshifts in accordance with the patterns identified in Section 4 while maintaining the feasibility and the optimality of the initial rostering solution.

The algorithm is based on the following adapting mechanism called "rectangular-adaptation". A rectangular-adaptation aims at identifying a rectangle inside the roster in such a way that (i) the north-west shift is equal to the south-east one, and (ii) the north-east shift is equal to the south-west one. An example is reported in Figure 2.

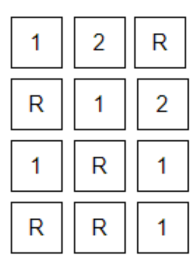

Figure 2: An example of a rectangle inside the rostering in which the shifts in the opposite corners are identical.

Exploiting the particular rectangle structure, it is possible to swap assigned shift while maintaining the feasibility and the optimality of the initial solution, as depicted in Figure 3. The idea is to exploit the rectangular-adaptation within our algorithm, which is based on the Tabu Search methodology (Glover and Laguna 1997): the algorithm starts from the optimal rostering computed by the model; then, applying a sequence of clockwise and counterclockwise rectangular-adaptation, it tries to improve the number, and eventually the goodness, of desired patterns in the final roster while maintaining its feasibility and optimality.

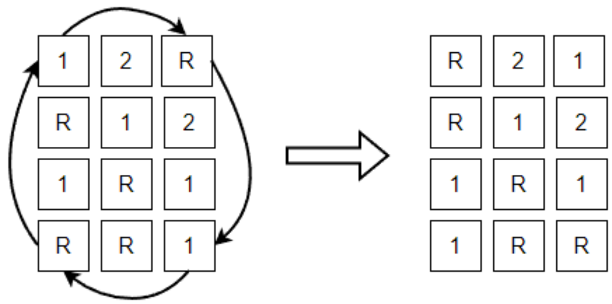

Figure 3: an example of rectangular-adaptation.

A score is associated to the initial solution, which is computed as follows. First of all, we associate to each pattern reported in Table 3 a weight, equal to the number of occurrences reported in the same table. For instance, the weight of "(3, 3, 3, R)" is 97, for "(2, 3, 3)" is 77, and so on.

For each employee roster, we compute the number of occurrences for each pattern in Table 3. The employee score is therefore the sum over all patterns of the number of occurrences multiplied by its weight. For instance, the score of the employee roster "(33R112R1)" is 259 given by the sum of 80 (for "(3, 3, R, 1)"), 82 (for "(R, 1, 1, 2)"), and 97 (for "(1, 1, 2)"). The score of a solution is given by the sum of all employee scores. An outline of the procedure is in Algorithm 1.

Algorithm 1: `Pattern adaptation` procedure.

```
Data: Current Solution S, employees O, Patterns P
Result: Solution S*
OpsScores ← compScores(S, O, P);
TT ← initializeTabuTag(S, O);
k ← 0;
while k < I_ni do
    /* I_ni max number of not-improving iterations */
    Moves ← getAllRectangularAdaptingMoves(S, TT);
    SolScore ← 0;
    for m ∈ Moves do
        S_tmp ← doChanges(S, m);
        if S_tmp is feasible then
            [OpScores_tmp, Sol_tmp] ← compScores(S_tmp, O, P);
            if Sol_tmp > SolScore then
                SolScore ← Sol_tmp;
                BestMove ← m;
                if Sol_tmp > S* then
                    S* ← Sol_tmp;
    S ← doChanges(S, BestMove);
    S ← updateTabuTag(TT, BestMove);
    k ← k + 1;
return S*;
```

At each iteration, the algorithm evaluates all the rectangular-adaptation that can be applied to the current roster to obtain a potential new feasible roster (that is a roster that continues to satisfy all the constraints in the mathematical model reported in Section 3). Among them, the algorithm selects the rectangular-adaptation $\bar{r}$ determining a new roster with the higher overall score. Note that this score could be less than or equal to the score of the current roster. To avoid looping over a number of solutions, the algorithm makes use of a tabu list implemented using tabu tags (Gendreau, Hertz, and Laporte 1994). Therefore the inverse of the rectangular-adaptation $\bar{r}$ is forbidden for a specified number of iterations in order to make the current solution evolve. The algorithms stops after a fixed number iteration is reached.

The pseudocode of our approach in Algorithm 1 makes use of the following procedures:

- `compScores`(S, O, P): compute the score of each employee $o \in O$ from the current solution S using the set of patterns P;
- `getAllRectangularAdaptingMoves`(S): compute every rectangular-adaptation moves from the current solution S;
- `doChanges`(S, m): update the current solution S using the move m;
- `updateTabuTag`(TT, $BestMove$): update the TabuTag data structure TT using the information contained in the move $BestMove$.

6 RESULTS

In this section we report the results of our three-step approach to determine optimal rostering plans but composed of the more frequent patterns, if possible. Our analysis is conducted on a set of real instances obtained from the data provided by the Cottolengo Hospital management, and presented in Section 2.2. For each ward, two instances are extracted from the data, one for nurses and one health and social workers. Since we are considering seven wards, the overall number of instances is 14. The planning horizon is one month. The number of not-improving iterations of the algorithm is set to 100.

Table 4 is organised as follows. The "score" columns and the "pattern" columns report respectively the score value and the number of desired patterns of the initial solution provided by the model, and of the solution computed by our algorithm. The "iterations" columns report the number of improving iterations, that is how many times the algorithms find an improving rectangular adaptation move. The last column reports the running time of the algorithm in seconds. Finally, the last row reports some average results.

Table 4: Results of the pattern adaptation algorithm.

instance	score		pattern		iterations		running
	model	adaptation	model	adaptation	improving	last	time (secs.)
1	222	592	51	72	4	18	113
2	148	518	76	83	3	8	121
3	148	518	76	83	3	8	118
4	148	518	76	83	3	8	109
5	148	444	58	53	2	3	75
6	74	370	45	52	3	5	82
7	222	592	51	72	4	18	104
8	222	518	40	43	2	2	71
9	222	518	40	43	2	2	72
10	148	592	47	56	4	5	57
11	222	592	51	72	4	18	113
12	222	592	51	72	4	18	112
13	296	740	37	56	5	23	43
14	296	740	37	56	5	23	44
average results			53	64	3	11	88

The results reported in the score and pattern columns clearly show the effectiveness of our three-step approach proving its capability to determine an optimal rostering with a number of more frequent patterns. The results reported in the pattern and iterations columns show that a quite small number of improving iterations are enough to determine a consistent number of new desired patterns in the final solution.

7 CONCLUSIONS

Workshift scheduling is particularly relevant for service organizations dealing with critical activities, such as healthcare, due to the complexity of managing medical care.

In this paper we presented a support system that automatically generating rostering plans by combining optimization and process mining approaches. Such integrated approach represents the main novelty of the paper. In particular, we developed and reported a methodological framework based on the three subsequent steps, that is rostering optimization, pattern extraction, pattern adaptation.

The first step consists in a multi-criteria mixed integer linear programming model, which models the problem of determining the monthly rostering balancing the monthly working hours of the healthcare personnel in accordance with a list of operative and contractual constraints. The second step consists in a sequence pattern mining algorithm to mine frequent contiguous sequential patterns from data, that is to identify constraints that are possibly not explicitly reported as domain knowledge. The third and last step consists in the adaptation of the rostering plan provided by the multi-criteria model in such a way to increase the number of the patterns mined in the second step while maintaining the feasibility and the optimality of the initial rostering solution.

We reported our decision support method for healthcare management based on real event logs. The computational results proved the capability of our approach to highly improve the quality of the initial rostering plan.

Ongoing work consist in the validation of our approach involving the main stakeholders, that is the hospital management and the personnel. Future work can consider the integration of the proposed approach with a discrete event simulation model to re-optimize the schedule during its realisation when it diverts from the planned one (eg. absences for COVID).

ACKNOWLEDGMENTS

The authors thank the Cottolengo Hospital management and in particular the Chief Innovation Officer Alessandro Menardi. This research has been partially carried out within the “Circular Health for Industry” project, funded by “Compagnia di San Paolo” under the call “Intelligenza Artificiale, uomo e societ”.

REFERENCES

Addis, B., R. Aringhieri, G. Carello, A. Grosso, and F. Maffioli. 2012. “Workforce Management based on forecasted demand”. In *Advanced Decision Making Methods applied to Health Care*, edited by E. Tanfani and A. Testi, Volume 173 of *International Series in Operations Research & Management Science*, 1–11. Springer.

Ahmed, M. A., and T. M. Alkhamis. 2009. “Simulation optimization for an emergency department healthcare unit in Kuwait”. *European Journal of Operational Research* 198(3):936–942.

Beese, J., K. Haki, S. Aier, and R. Winter. 2019. “Simulation-Based Research in Information Systems - Epistemic Implications and a Review of the Status Quo”. *Business & Information Systems Engineering* 61(4):503–521.

Bellanti, F., G. Carello, F. D. Croce, and R. Tadei. 2004. “A greedy-based neighborhood search approach to a nurse rostering problem”. *European Journal of Operational Research* 153(1):28–40.

Burke, E. K., P. D. Causmaecker, G. V. Berghe, and H. V. Landeghem. 2004. “The state of the art of nurse rostering”. *Journal of Scheduling* 7(1):441–499.

Duma, D., and R. Aringhieri. 2020. “An ad hoc process mining approach to discover patient paths of an Emergency Department”. *Flexible Services and Manufacturing Journal* 32(1):6–34.

Dumas, M., M. L. Rosa, J. Mendling, and H. A. Reijers. 2018. *Fundamentals of Business Process Management*. 2 ed. Springer.

Ernst, A., H. Jiang, M. Krishnamoorthy, B. Owens, and D. Sier. 2004. “An Annotated Bibliography of Personnel Scheduling and Rostering”. *Annals of Operations Research* 127(1):21–144.

Ernst, A., H. Jiang, M. Krishnamoorthy, and D. Sier. 2004. "Staff scheduling and rostering: A review of applications, methods and models". *European Journal of Operational Research* 153(1):3–27.

Gendreau, M., A. Hertz, and G. Laporte. 1994. "A Tabu Search Heuristic for the Vehicle Routing Problem". *Management Science* 40(10):1276–1290.

Geoffrion, A. M. 1968. "Proper efficiency and the theory of vector maximization". *Journal of Mathematical Analysis and Applications* 22(3):618–630.

Glover, F., and M. Laguna. 1997. *Tabu Search*. 1 ed. Springer New York, NY.

Jiang, H., M. Krishnamoorthy, and D. Sier. 2004. "Staff sheduling and rostering: Theory and applications, part I-II". *Annals of Operations Research* 127(1-4):17–20.

Karlsson, I., A. H. Ng, A. Syberfeldt, and S. Bandaru. 2015. "An interactive decision support system using simulation-based optimization and data mining". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 2112–2123: Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kellogg, D. L., and S. Walczak. 2007. "Nurse scheduling: From academia to implementation or not?". *Interfaces* 37(4):355–369.

Lather, J. I., and J. I. Messner. 2018. "Framework for a Hybrid simulation Approach for an Integrated Decision Support System in Healthcare Facilities". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 1406–1417: Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Martin, N., B. Depaire, and A. Caris. 2015. "Using process mining to model interarrival times: investigating the sensitivity of the arpra framework". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 868–879: Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Martin, N., B. Depaire, and A. Caris. 2016. "The Use of Process Mining in Business Process Simulation Model Construction - Structuring the Field". *Business & Information Systems Engineering volume* 58(1):73–87.

Moreira, M. W. L., J. J. P. C. Rodrigues, V. Korotaev, J. Al-Muhtadi, and N. Kumar. 2019. "A Comprehensive Review on Smart Decision Support Systems for Health Care". *IEEE Systems Journal* 13(3):3536–3545.

Munoz-Gama, J. et al. 2022. "Process mining for healthcare: Characteristics and challenges". *Journal of Biomedical Informatics* 127(1):103994.

Ordu, M., E. Demir, C. Tofallis, and M. M. Gunal. 2021. "A novel healthcare resource allocation decision support tool: A forecasting-simulation-optimization approach". *Journal of the Operational Research Society* 72(3):485–500.

Phan, R., V. Augusto, D. Martin, and M. Sarazin. 2019. "Clinical Pathway Analysis Using Process Mining and Discrete-Event Simulation: an Application to Incisional Hernia". In *Proceedings of the 2019 Winter Simulation Conference*, edited by N. Mustafee, K.-H. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 1172–1183: Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

van der Aalst, W. M. P. 2016. *Process Mining - Data Science in Action*. 2 ed. Springer.

AUTHOR BIOGRAPHIES

ALBERTO GUASTALLA is a PhD student at Computer Science Department of University of Turin, Italy. He holds a M.S. in Computer Science from University of Turin. His main research interest focuses on Operational Research, Combinatorial Optimization and Mathematical Programming about transportation and logistics problems with application in healthcare. His email address is alberto.guastalla@unito.it.

EMILIO SULIS is a temporary research assistant in Computer Science at the University of Turin. His research activity focuses on the impact of information technology on society, mainly Artificial Intelligence techniques and organizational processes (process-oriented data science). Recently, he worked on information extraction and representation of organizational processes, mainly in legal and medical field, as well as computational social simulations with agent-based modeling and discrete-event simulation. His email address is emilio.sulis@unito.it.

ROBERTO ARINGHIERI is Associate Professor of Operations Research at the University of Turin, Italy. He holds a M.S. in Computer Science and a Ph.D. in Mathematics for Economic Decisions and Operations Research from the University of Pisa. His main research interest focuses on simulation and optimization algorithms applied to healthcare management. He is co-chair of the EURO Working Group on Operational Research Applied to Health Services. He also serves as Associate Editor for Operations Research for Healthcare. His email address is roberto.aringhieri@unito.it.

STEFANO BRANCHI is a temporary research assistant in Computer Science at Fondazione Bruno Kessler in Process and Data Intelligence (PDI) Unit. He received his Master Degree in Computer Science area Data Science from the University of

Trento. He is currently working in the field of process mining, mainly in predictive and prescriptive process monitoring based on historical execution traces. His email address is sbranchi@fbk.eu.

CHIARA DI FRANCESCOMARINO is a researcher at Fondazione Bruno Kessler (FBK) in the Process and Data Intelligence (PDI) Unit. She received her PhD in Information and Communication Technologies from the University of Trento. She is currently working in the field of process mining, investigating problems related to process monitoring, process discovery, as well as predictive process monitoring based on historical execution traces. She serves as PC member in international conferences in the business process management field and as peer reviewer in international journals. Her email address is dfmchiara@fbk.eu.

CHIARA GHIDINI is the Scientific Coordinator of the digital Health & WellBeing centre at Fondazione Bruno Kessler, Italy, where she also heads the Process & Data Intelligence research unit. She holds a PhD in Computer Science Engineering from Universit La Sapienza of Rome. Her current research interests focus on Artificial Intelligence techniques and their application to Process Mining. She has served as PC/track chair in several conferences on knowledge / process management. Her email address is ghidini@fbk.eu.