Ilmenauer Beiträge zur Wirtschaftsinformatik

Herausgegeben von U. Bankhofer, V. Nissen
D. Stelzer und S. Straßburger

Julian Schulte, Maik Günther, Volker Nissen

Evolutionary Bilevel Approach for Integrated Long-Term Staffing and Scheduling

Paper presented at MISTA 2017, Kuala Lumpur/Malaysia, Dec. 2017

Arbeitsbericht Nr. 2020-02, März 2020

c

Technische Universität Ilmenau
Fakultät für Wirtschaftswissenschaften und Medien
Institut für Wirtschaftsinformatik

Autor: Julian Schulte, Maik Günther, Volker Nissen

Titel: Evolutionary Bilevel Approach for Integrated Long-Term Staffing and Scheduling

First published online in: Gunawan, A.; Kendall, G. (eds.) Proc. of MISTA 2017, 05.-08.12.2017, Kuala Lumpur/Malaysia, pp. 144-157

Ilmenauer Beiträge zur Wirtschaftsinformatik Nr. 2020-02, Technische Universität Ilmenau, März 2020

ISSN 1861-9223

ISBN 978-3-938940-63-1

URN urn:nbn:de:gbv:ilm1-2020200306

Technische Universität Ilmenau, Fakultät für Wirtschaftswissenschaften und Medien, Institut für Wirtschaftsinformatik, PF 100565, D-98684 Ilmenau

http://www.tu-ilmenau.de/wid/forschung/ilmenauer-beitraege-zur-wirtschaftsinformatik/

Inhaltsverzeichnis / Contents

Zusammenfassung / Abstract:

Determining size and structure of a company's workforce is one of the most challenging tasks in human resource planning, especially when considering a long-term planning horizon with varying demand. In this paper an approach for integrated staffing and scheduling in a strategic long-term context is presented by applying evolutionary bilevel optimization. For demonstration, the example of determining the number of employees in different categories over the period of one year in a midsized call center of a utility is used. In doing so, two contrary objectives were optimized simultaneously: reduce the overall workforce costs and retain a high scheduling quality. The results show that the proposed approach could be used to support corporate decision making related to strategic workforce planning, not only for call centers but for any other kind of workforce planning involving personnel scheduling.

Schlüsselwörter: Personnel Scheduling, Staffing, Integrated Staffing and Scheduling, Bilevel Optimization, Metaheuristics

Evolutionary Bilevel Approach for Integrated Long-Term Staffing and Scheduling

Julian Schulte • Maik Günther • Volker Nissen

Abstract Determining size and structure of a company's workforce is one of the most challenging tasks in human resource planning, especially when considering a long-term planning horizon with varying demand. In this paper an approach for integrated staffing and scheduling in a strategic long-term context is presented by applying evolutionary bilevel optimization. For demonstration, the example of determining the number of employees in different categories over the period of one year in a midsized call center of a utility is used. In doing so, two contrary objectives were optimized simultaneously: reduce the overall workforce costs and retain a high scheduling quality. The results show that the proposed approach could be used to support corporate decision making related to strategic workforce planning, not only for call centers but for any other kind of workforce planning involving personnel scheduling.

1 Introduction and Related Work

Companies are challenged by the question of how to organize size and structure of their workforce in order to manage upcoming workload most cost-effectively. This is especially the case when entirely new business units are established, existing units are restructured or current and future demand strongly deviate from each other. Examples of a changing workload are found, among others, in the utility sector. Rising requirements for customer service combined with strong cost pressure require measures for utilities to create a cost-efficient workforce structure. Therefore, in this paper the problem of determining the ideal size and structure of a typical inbound call center of an utility is examined. However, it may be noted that the methodology applied in this paper is not limited to the considered call center, but rather can be applied to problems of other companies and business units.

When considering size and structure (e.g. skill-mix and contract types) of a company's workforce, the purpose of staffing is to determine the adequate future number of employees needed in different categories. As this is already not an easy task with only one type of employee, it gets even more challenging when looking at a heterogeneous workforce due to different types of skills or contracts [7]. Scheduling, as another crucial task in workforce planning, is concerned with getting the right people to the right place at the right time. It is easily recognizable that staffing decisions have direct impact on scheduling quality. Hence, it is reasonable not to look at staffing and scheduling decisions as two consecutive tasks but to employ an integrated planning approach.

Julian Schulte
Technische Universität Ilmenau
E-mail: julian.schulte@tu-ilmenau.de

Maik Günther
Stadtwerke München GmbH
E-mail: guenter.maik@swm.de

Volker Nissen
Technische Universität Ilmenau
E-mail: volker.nissen@tu-ilmenau.de

In the literature, the number of contributions addressing integrated staffing and scheduling is rather limited, especially compared to literature concentrated on staffing or scheduling problems [7, 16, 25]. There are, however, several approaches differing in methodology and considered planning horizon. Avramidis et al. [2] for example provide a simulation-based algorithm that simultaneously optimizes the staffing and scheduling over one day in a multi-skill call center. They focus on how many agents of each type are needed based on the arrival rates and type of calls at a given day. In the model presented by Brunner and Edenharter [8] a column generation based heuristic is applied to identify the weekly demand of physicians with different experience levels. Even though the authors are targeting a long-term planning horizon of one year, they solve each week independently. Beliën and Demeulemeester [3] propose a branch-and-price approach considering a planning horizon of four weeks with the aim of reducing staffing costs by integrating both processes, operation room scheduling, which determines the required nurse staffing level, and nurse scheduling. A branch-and-price methodology with a planning horizon of four weeks was also used in the integrated model developed by Maenhout and Vanhoucke [16], with the purpose of identifying optimal staffing and scheduling policies in a hospital. In a more recent contribution published by Beliën et al. [4], an enumerative MILP algorithm is proposed for optimizing the team sizes of an aircraft maintenance company in order to minimize the overall labor costs for a period of six weeks.

The proposed methods already deliver detailed insight into the needed workforce at a given day, week or month and therefore a necessary basis for further workforce planning. However, the planning horizons considered are short when a strategic perspective is taken. Thus, the approaches so far cannot provide information about the required overall workforce, especially when considering a long-term period, e.g. one year, with varying demand as well as factors like overtime/flextime and holidays of employees. To fill this gap, in this paper an approach for integrated staffing and scheduling in a strategic long-term context using evolutionary bilevel optimization is presented.

Bilevel optimization can be seen as a form of hierarchical optimization problem. More specifically, an upper-level optimization problem has another optimization problem within its constraints and therefore is dependent on the results of the lower-level problem. This hierarchical relationship is closely related to the problem of Stackelberg [22], where a follower (lower-level problem) optimizes his objective based on the given parameters determined by the leader (upper-level problem). The leader, on the other hand, optimizes his own objective under consideration of the follower's possible reactions [10]. In the case of integrated staffing and scheduling, staffing will be treated as upper-level problem with the objective to minimize the overall labor costs, i.e. adjusting number and qualification of employees, but at the same time maximizing the quality of personnel schedules. Scheduling, as the lower-level optimization problem, has the objective to maximize scheduling quality based on the staffing decisions made at the upper-level. The quality in this case is assessed by a fitness function that considers the match of staffing demand and allocation of employees with certain skills at given time intervals as well as employee overtime.

Evolutionary bilevel optimization was successfully applied in various practical applications in fields such as economics, transportation, engineering and management, but, to the best of our knowledge, not yet in workforce planning and scheduling problems (see [20] for a comprehensive review). Evolutionary Algorithms (EA) are a common metaheuristic approach to compute good solutions in an acceptable amount of time, especially when working with real world problems that otherwise cannot be solved to optimality within reasonable computation time [17]. This also applies to workforce planning and scheduling problems, with Genetic Algorithms (GA) as the most often used class of metaheuristics in this domain [7, 25]. Due to its widespread usage and successful application to similar problems, GA were chosen in our case, both to solve the upper-level staffing and the lower-level scheduling problem.

The remainder of the paper is structured as follows: In Section 2 the problem of integrated staffing and scheduling is presented. Section 3 describes the applied evolutionary bilevel approach. In Section 4 the computational results will be discussed. Finally, the conclusions and suggestions for further research are presented in Section 5.

2 Problem Description

In this section, the problem of integrated staffing and scheduling in the environment of a German utility is presented. We consider a strategic context in which the company has to make its overall workforce planning one year in advance to assure that all required employees with the right qualification are available. Due to internal restrictions of the utility, the presented problem is derived and abstracted from a real world problem commonly found in strategic workforce planning.

2.1 Bilevel Optimization

Bilevel optimization problems are proven to be strongly NP-hard [14] and can generally be formulated as follows [10, 20, 24]:

$$
\begin{aligned}
&\min_{x \in X} F(x,y) \\
&\text{subject to} \quad G(x,y) \leq 0 \\
&\qquad\qquad\quad \min_{y \in Y} f(x,y) \\
&\qquad\qquad\quad \text{subject to} \quad g(x,y) \leq 0
\end{aligned}
\tag{1}
$$

where x is the vector of decision variables determined by the upper-level problem and y is the vector determined by the lower-level problem. Besides, $F(x,y)$ and $f(x,y)$ are the objective functions and $\mathrm{G}(x,y)$ and $g(x,y)$ the constraints of the upper- and lower-level problem. For each vector x, y will be the optimization result of the lower-level problem $\min f(x,y)$. Therefore, $\min f(x,y)$ could also be denoted as $y(x)$ [27]. Thus, the result of the upper-level problem is dependent on the result of the lower-level problem, which in turn is dependent on the vector x given by the upper-level problem.

For the considered problem of integrated staffing and scheduling, x will be the staffing decision made at the upper-level determining the number of employees of each type (combination of skill set and contract type). Based on the given workforce structure, the personnel scheduling will be conducted yielding schedules for each day of the planning horizon. Hence, the objective function at the upper-level $\min F(x,y)$ depends on the costs due to staffing decisions as well as the quality $y(x)$ of the created schedules at the lower level.

2.2 Staffing Problem

The here considered call center has a need of three different skill types $s \in S$ with $S = \{\text{agent}, \text{support}, \text{supervisor}\}$. The skills are considered to be categorical, which means they determine the tasks that can be performed by each employee. However, it is possible to cross-train employees so they can perform more than one type of task [7]. The qualification of an employee can therefore be seen as set of different skill combinations $q \subseteq S$. The contract type $t \in T$ of an employee determines his average weekly working time.

Within its staffing decision, the company has to predefine feasible employee types $\bar{E}$ (see Table 1). Each employee type $\bar{e}_{qt} \in \bar{E}$ is defined by its qualification q and contract type t. Furthermore, each employee type $\bar{e}_{qt}$ is linked to costs $c_{\bar{e}qt}$ that arise for employing one employee of this type over the considered planning horizon. Here, the costs of each employee type are represented by a relative factor summing up annual wages, payroll taxes, overhead and training costs. The number of employees of each type is represented by the decision variable $x_{\bar{e}qt}$. The setting of the staffing problem is shown in Table 1.

The objective here is, as part of the upper-level problem, to minimize the overall staffing costs (2a) subject to the output of the lower-level problem (2b).

Table 1 Setting of the staffing problem

Contract type	Qualification	Costs
40 h	agent	1
20 h	agent	0.6
40 h	support	1.1
20 h	support	0.65
40 h	agent - support	1.3
20 h	agent - support	0.75
40 h	supervisor	1.4

Parameters (staffing)

S	set of skills (index s)
q	qualification of an employee ($q \subseteq S$)
T	set of contract types (index t)
$\bar{E}$	set of employee types (index $\bar{e}_{qt}$)
$c_{\bar{e}qt}$	costs for an employee of type $\bar{e}_{qt}$

Decision variable (staffing)

$x_{\bar{e}qt}$	number of employees of type $\bar{e}_{qt}$

Staffing problem (upper level)

$$\min_{x} F\left(\sum_{\bar{e}_{qt}\in\bar{E}} x_{\bar{e}qt}\, c_{\bar{e}qt}\, , y(x)\right) \tag{2a}$$

with

$$x_{\bar{e}qt} \geq 0 \; and \; integer \quad \forall\, \bar{e}_{qt} \in \bar{E}$$
$$q \subseteq S, t \in T$$

subject to

(3a) – (3i) (2b)

2.3 Scheduling Problem

The scheduling problem presented in this paper considers the daily staff scheduling of a call center over a planning horizon of one year. Each week of the planning horizon $w \in W$ is partitioned into periods $p \in P$, representing the operating days of the call center. Moreover, each operating day again is segmented into time intervals $i \in I$. In this practical case, a planning horizon $W = \{1, \dots, 52\}$ with operating days $P = \{1, \dots, 5\}$ and, due to the strategic context, hourly planning intervals $I = \{8, \dots, 17\}$ were chosen, representing the operating times 8 a.m. to 6 p.m.

The set of employees E is determined by the staffing decision at the upper-level with a concrete employee for each $x_{\bar{e}qt}$. It is assumed that the company has a predefined set of possible shift patterns M (see Table 2) with b_{mi} determining whether a shift pattern is covering a specific time interval. In addition, variable n_{es} determines if an employee's qualification contains skill s. It is assumed that each employee has six weeks of holidays each year. Therefore it is possible for employees not to be available at certain periods which is determined by variable a^{e}_{pw}.

Table 2 Possible shift patterns

Shift start (a.m.)	8	8	8	10	10	10	12
Shift duration (h)	4	8	10	4	6	8	4

The assignment of an employee $e \in E$ to a shift m on day p in week w with skill s is controlled by using the binary decision variable y^{es}_{mpw}. An employee can only be assigned if he is available and has the required skill (3b) - (3c). Furthermore, one employee can only be assigned to one shift each day (3d).

For each time interval i on day p in week w and each skill s a certain staffing level d^{s}_{ipw} has to be satisfied. The number of planned employees of each skill at time interval i is determined by variable e^{s}_{ipw} (3e). If a deviation $|e^{s}_{ipw} - d^{s}_{ipw}|$ arises from the staffing target, penalty points are generated by the function P_d (3f). An additional penalty is added if no employees are planned but required or vice versa.

To compensate overtime and minus hours, each employee has a flextime account u_{ew}, which is updated on a weekly basis. Therefore, the deviation of the employee's actual working time l_{ew} (3g) and the average weekly working time h_e is added to his flextime account (3h). However, to provide an equal workload distribution and to ensure that employees are staffed according to their contract types, the penalty function P_u generates penalty points based on how far employees exceeded or fall below their average weekly working time u_{ew}/h_e (3i). The weekly penalty is calculated by multiplying the absolute flextime value times the percentage of deviation.

It has to be noted that both penalty functions have to be carefully balanced, as otherwise employees might fall far beyond their contractual working hours (if over-/understaffing is too expensive) or, on the other side, flextime will not be used at all.

The objective here is to minimize the overall penalty points over the considered planning horizon (3a) subject to given constraints (3b) – (3i) described above.

Parameters (scheduling)

W	set of weeks in planning horizon (index w)
P	set of periods in planning week w (index p)
I	set of time intervals in planning period p (index i)
E	set of employees (index e) determined by the upper-level decision variable $x_{\bar{e}qt}$
S	set of skills (index s)
M	set of shift patterns (index m)
b_{mi}	1 if shift m covering time interval i, 0 otherwise
n_{es}	1 if employee's qualification contains skill s, 0 otherwise
a_{pw}^{e}	1 if employee e is available on day p in week w, 0 otherwise
d_{ipw}^{s}	demand of skill s at time interval i on day p in week w
e_{ipw}^{s}	number of planned employees with skill s at time interval i on day p in week w
P_d	demand penalty function
u_{ew}	flextime account of employee e in week w
l_{ew}	actual working time employee e in week w
h_e	average weekly working time of employee e
P_u	working time penalty function

Binary decision variable (scheduling)

y_{mpw}^{es}	1 if employee e is assigned to a shift m on day p in week w with skill s, 0 otherwise

Scheduling problem (lower level)

$$\min_{y} P_d + P_u \tag{3a}$$

with

$$n_{es}, b_{mi}, a_{pw}^{e}, y_{mpw}^{es} \in \{0, 1\}$$
$$\forall\, e \in E, s \in S, m \in M, p \in P, w \in W$$

subject to

$$y_{mpw}^{es} \leq a_{pw}^{e} \quad \forall\, e \in E, s \in S, m \in M, p \in P, w \in W \tag{3b}$$

$$y_{mpw}^{es} \leq n_{es} \quad \forall\, e \in E, s \in S, m \in M, p \in P, w \in W \tag{3c}$$

$$\sum_{s \in S} \sum_{m \in M} y_{mpw}^{es} \leq 1 \quad \forall\, e \in E, p \in P, w \in W \tag{3d}$$

$$e_{ipw}^{s} = \sum_{e \in E} y_{mpw}^{es}\, b_{mi} \qquad \forall\, s \in S, m \in M, i \in I, p \in P, w \in W \tag{3e}$$

$$P_d = \sum_{s \in S} \sum_{e \in E} \sum_{i \in I} \sum_{p \in P} \sum_{w \in W} \left| d_{ipw}^{s} - e_{ipw}^{s} \right| * \gamma_d\,, \qquad with$$
$$\gamma_d = \begin{cases} 500, & e_{ipw}^{s} > 0 \; and \; d_{ipw}^{s} = 0 \\ 500, & d_{ipw}^{s} > 0 \; and \; e_{ipw}^{s} = 0 \\ 1, & otherwise \end{cases} \tag{3f}$$

$$l_{ew} = \sum_{i \in I} \sum_{p \in P} y_{mpw}^{es}\, b_{mi} \qquad \forall\, e \in E, m \in M, w \in W \tag{3g}$$

$$u_{ew} = u_{e(w-1)} + (l_{ew} - h_{ew}), \qquad \text{with } u_{e0} = 0$$
$$\forall\, s \in S, m \in M, i \in I, p \in P, w \in W \tag{3h}$$

$$P_u = \sum_{e \in E} \sum_{w \in W} \frac{{u_{ew}}^2}{h_e} \tag{3i}$$

3 Evolutionary Bilevel Approach

3.1 Genetic Algorithms

GA are population-based metaheuristics and rely on three basic principles. First, there is a set of solutions (population). Each solution (individual) is evaluated based on its quality (fitness) by applying an objective function (fitness function). Second, variation operators are applied in the process of creating new solutions (reproduction). This can be done by crossover (recombining two or more individuals) and/or mutation (random variation of an individual). Both variation operators are probabilistically applied and exist in many different variants. Finally, individuals with high fitness values are more likely to be selected for reproduction by a selection procedure (see [18, 19, 23] for more detailed information on metaheuristic optimization in general and GA in particular). The GA applied in this paper are based on the basic version shown in Algorithm 1.

The individuals of the here applied GA are represented by matrices, with each row corresponding to an abstract employee type (upper-level algorithm) respectively a concrete employee (lower-level algorithm). The rows at the upper-level are encoded as 4-bit Gray strings, allowing a number between 0 and 15 employees for each type. At the lower-level, 3-bit Gray encoding is used to determine one of seven possible shift patterns for an employee's working day or absence of the employee.

For reproduction, one-point, uniform and two types of n-point crossover are used, each with a probability p=0.25. The first type of n-point crossover randomly selects half of the rows of each matrix and interchanges the entire rows between the two individuals. The second type interchanges one n-bit block of random size per row between the individuals. Moreover, bit flip mutation is used with each bit flipping with the probability of the given mutation rate (see Section 4.1).

3.2 Multi-Objective Optimization

Within the here discussed problem of integrated staffing and scheduling, two objectives have to be optimized. However, both objectives are in conflict with each other, as for example hiring multi-skilled, flexible part-time employees will yield high quality schedules but also increase the staffing costs and on the other hand, reducing the number of employees will reduce labor costs but also the scheduling quality. The resulting multi-objective problem can be solved by using the concept of Pareto efficiency, which will yield a set of Pareto optimal solutions (Pareto front). The final solution to be selected will therefore be a trade-off among the two considered objectives staffing costs and scheduling quality (see [9, 6, 19] for more detailed information on multi-objective optimization).

Algorithm 1 Overview of GA in pseudocode

1: popsize ← desired population size
2: generations ← number of generations to be evaluated
3: P ← build initial population of random individuals with size popsize
4: Best ← select best individual according to fitness of initial population
5: **for** generations times **do**
6: **for** each individual $P_i \in P$ **do**
7: **if** $Fitness(P_i) > Fitness(Best)$
8: $Best \leftarrow P_i$
9: **end if**
10: **end for**
11: $P' \leftarrow \{\}$
12: **for** popsize times **do**
13: Parent $P_a \leftarrow SelectIndividual(P)$
14: Parent $P_b \leftarrow SelectIndividual(P)$
15: Child $C \leftarrow Crossover(P_a, P_b)$
16: $P' \leftarrow P' \cup \{Mutate(C)\}$
17: **end for**
18: $P \coloneqq P'$
19: **end for**
20: **return** Best

3.3 Nested Bilevel Genetic Algorithm

Within the context of integrated (long-term) staffing and scheduling problems, Maenhout and Vanhoucke [16] point out that most researchers (e.g. [1, 15, 11, 26, 3]), including themselves, iteratively alternate between the staffing and the scheduling problem as they are creating and evaluating personnel schedules based on certain staffing decisions. This was also noted by more recent research [12]. This basic procedure also applies for the evolutionary bilevel approach.

Following the taxonomy given by Talbi [24], the here presented procedure can be defined as a nested constructing approach with metaheuristics on both levels. In this type of bilevel model, an upper-level metaheuristic calls a lower-level metaheuristic during its fitness assessment. In doing so, the upper-level heuristic determines the decision variable x (here the number of employees for each type) as input of the lower-level algorithm, which in turn determines the decision variable y. Both variables are subsequently used to solve the bilevel problem at the upper-level. By the existence of a multi-objective optimization problem, non-dominated sorting is used to evaluate the fitness of each individual at the upper-level GA [21]. As a result, a Pareto front will be built of all non-dominated solutions evaluated at the upper-level (see Algorithm 2, line 10 and 11). An overview of the nested bilevel GA applied in this paper is shown in Algorithm 2.

Algorithm 2 Overview of nested bilevel GA in pseudocode

1: **initialization** (see Algorithm 1, lines 1-3)
2: Best ← $\{\}$
3: **for** generations times **do**
4: **for** each individual $P_i \in P$ **do**
5: **call** lower-level GA with P_i as input (see Algorithm 1)
6: **end for**
7: $Fitness(P)$
8: $A \leftarrow ParetoFront(P) \cup Best$
9: $Best \leftarrow ParetoFront(A)$
10: **reproduction** (see Algorithm 1, lines 11-18)
11: **end for**
12: **return** Best

One major issue when applying bilevel optimization are the long computation times. Preliminary experiments showed that after the evaluation of the first four weeks (of 52 in total) at the lower-level GA, it was already roughly possible to determine the quality of the staffing decision. Once the fitness of two individuals showed a deviation of at least 100% in the fourth week, the fitness development of both individuals did not tend to change. This behavior was used to implement termination criteria, which were applied at different points during the fitness assessment of each individual at the lower-level problem. Once one criterion applies, the evaluation of the individual at the lower-level is terminated.

In general, at each termination checkpoint the assessed individual is compared to all individuals included in the current $Best$ Pareto front (upper-level problem). The first two checkpoints are set after the fourth and eighth week. Here, the assessment is terminated if the overall penalty of the assessed individual is higher by a factor of d=2 compared to any of the solutions of the $Best$ Pareto front at the given weeks. However, as there is no point in keeping solutions with lower fitness and higher costs but, on the other hand, solutions with lower fitness and lower costs could be interesting, the termination only applies if the costs of the assessed individual are higher or equal to the compared individual. The last termination checkpoint is set after week twelve. Here, the assessment is terminated if the overall penalty deviates by a factor of d=3 regardless of the solution's costs. To avoid termination due to outliers, two consecutive weeks are checked within the termination checkpoints.

By applying these criteria it is possible to early identify irrelevant staffing decisions (e.g. too many employees or only one employee type) and to concentrate on more promising solutions. The experiments showed that by implementing these three termination checkpoints, the performance already increased significantly. However, to identify subsequent deviations there could also be implemented more checkpoints during the entire fitness assessment.

4 Computational Results and Discussion

4.1 Experimental Setup

The parameters of both GA were set based upon preliminary studies. For the upper-level GA, a population size of 20, a generation number of 40 and n=10 restarts were chosen, with each restart having a random initial population. The lower-level GA was configured with a population size of 50 and a generation number of 80. On both levels the mutation rate was set to $1/v$, with v being the number of bits of the encoded individual. The fitness at the upper-level was evaluated by Eq. (2a), for the fitness evaluation at the lower-level Eq. (3a) was used.

The optimization software was written in Julia [5] and all experiments were executed on Windows 10 machines with Intel Core i5-2400K processors (4 cores, maximum clock rate 3.1 GHz) and 4 GB RAM. By applying the termination criteria the computation time could be reduced by 50%, however, despite parallel computation within the fitness assessment at the upper-level, each restart of the upper-level algorithm took about twelve hours.

For this experimental study, the following scenario is assumed. The call center has an initial staffing level based on the estimated demand for the year 2017 (see Table 3, solution 1). For the coming year, the company expects a 20% increase in demand. The estimated demand for both years (aggregated agent hours) is shown in Fig. 1. Furthermore, Fig. 2 shows the demand fluctuations in hourly resolution for an exemplary day. The data used for the experimental study was created by using a demand generator relying on slightly modified real world data (see [13] for more details on the used demand generator). The demand for the other two skills is calculated based on staffing ratios. In this concrete case, one support for each four and one supervisor for each eight agents is required.

Moreover, it is assumed that each employee has six weeks of holidays. As the generation of employees is done automatically during the optimization procedure, the holidays of each employee are assigned randomly on a weekly basis. However, while doing so it is assured that at maximum 30% of each skill type can be on holidays at the same time, the holidays are equally distributed over the whole year and there are two consecutive weeks of holidays during the summer period (June until September).

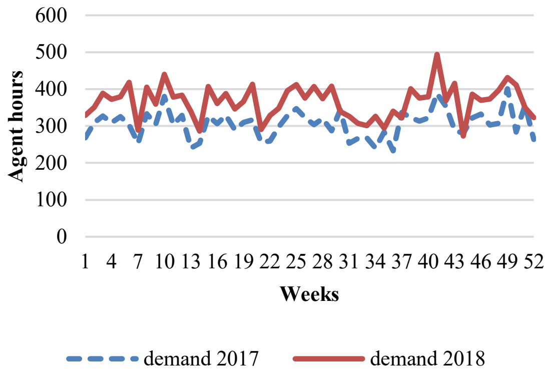

Fig. 1 Weekly demand of agent hours over the planning horizon

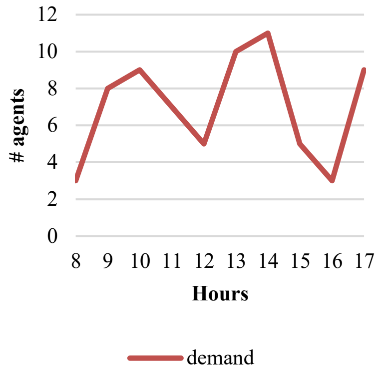

Fig. 2 Exemplary demand of a single day

4.2 Results and Discussion

For the purpose of demonstration and because the Pareto fronts tend to show similar behavior over multiple runs, the upper-level optimization algorithm was executed n=10 times, with each restart yielding a set of Pareto optimal solutions. The combined solutions of all Pareto fronts are shown in Fig. 3. Without the consideration of outliers and solutions that did not pass the termination criteria, a total of 134 possible staffing decisions with different combinations of costs and scheduling penalties were found (based on the demand for 2018). For further discussion, 15 solutions along the new Pareto front were selected. A detailed description of the selected staffing decisions can be found in Table 3.

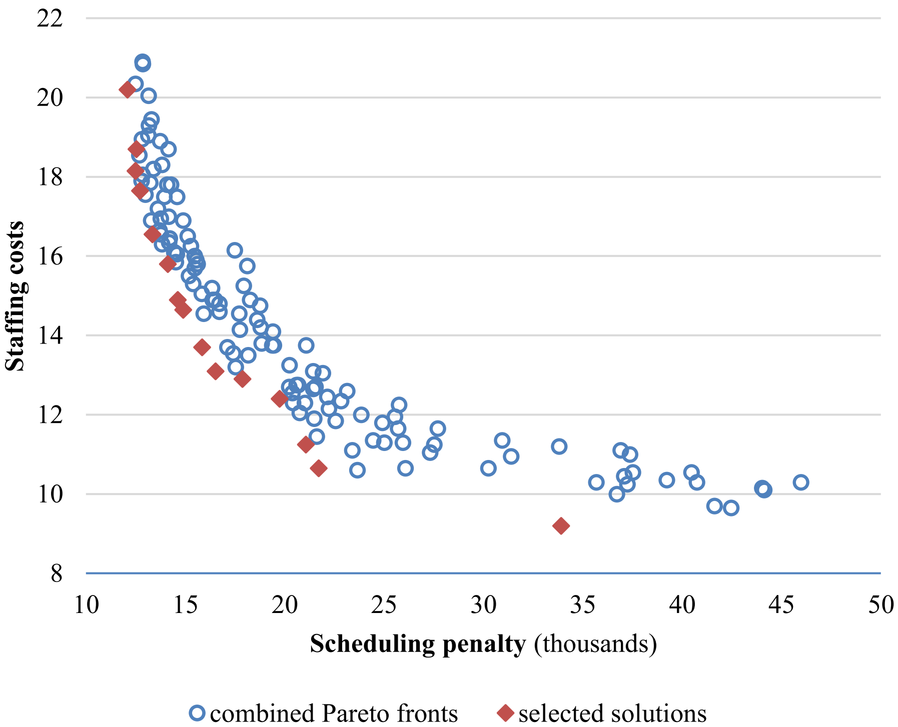

Fig. 3 Possible staffing decisions (n=10 optimization runs)

Table 3 Staffing decisions and corresponding scheduling quality

Solution	Agent		Support		Agent - Support		Supervisor	Penalty			Staffing costs
	40 h	20 h	40 h	20 h	40 h	20 h	40h	Overstaffing	Understaffing	Flextime	
1 (init. 2017)	3	4	1	1	1	1	3	3,5	7,451	4,08	13.4
2 (init. 2018)	3	4	1	1	1	1	3	2,529	10,734	4,942	13.4
3	1	5	0	1	7	3	3	4,991	4,942	2,174	20.2
4	3	4	1	2	4	2	3	4,504	5,387	2,66	18.7
5	2	2	2	0	6	1	3	4,259	6,036	2,212	18.15
6	3	1	0	0	7	1	3	3,903	6,57	2,26	17.65
7	8	0	1	3	1	0	3	4,408	6,231	2,724	16.55
8	6	1	1	0	3	0	3	3,735	7,344	3,056	15.8
9	3	2	1	0	3	2	3	2,9	9,251	2,482	14.9
10	5	2	2	0	1	1	3	3,168	8,647	3,092	14.65
11	3	0	1	0	3	2	3	2,61	10,811	2,44	13.7
12	3	4	2	0	1	0	3	2,58	10,827	3,13	13.1
13	2	5	1	0	2	0	3	2,393	11,505	3,996	12.9
14	2	2	2	0	1	2	3	2,433	12,484	4,842	12.4
15	3	2	2	1	0	0	3	2,542	13,463	5,076	11.25
16	2	2	0	1	2	0	3	2,143	14,918	4,658	10.65
17	2	1	1	2	0	0	3	2,284	16,839	14,804	9 2

The first two solutions shown in Table 3 represent the initial staffing level, optimized with the demand for 2017 and 2018. Solutions 3 to 17 represent the optimized staffing levels based on the demand for 2018. When looking at the results of solution 2, due to the rising demand and therefore an insufficient number of employees, the penalty for understaffing and flextime increased and the overstaffing penalty decreased (compared to solution 1). The resulting scheduling quality of solution 2 now is comparable to the quality of solutions 11 to 13, which furthermore show approximately the same cost level. In case the company wants to retain its service level of 2017 (considering the understaffing penalty of solution 1), the current workforce should be developed towards the staffing level of solution 8. Taking a more general look on the results, it can be noted that no solution below the cost level of 9.2 passed the termination criteria within the fitness assessment. The low costs can simply be explained by the insufficient staffing level, resulting in high understaffing and flextime penalties. Looking at the staffing decisions in Table 3 from bottom to top, it can be seen that the scheduling quality grows with an increasing number of employees and increasing contractual working hours, and, hence, rising staffing costs. Furthermore, when comparing solutions 8 and 7 with 6 to 3, an increased employment of cross-trained and part-time workers can be noticed yielding the highest scheduling quality. However, no solution was found above the cost level of 20.85. This may indicate that there is a point at which scheduling quality cannot be increased under the given staffing and scheduling policies. Thus, measures to further improve the scheduling quality could be, for example, the introduction of new contract types (e.g. contracts with a working time of 30 or 12 hours per week) or more flexible shift patterns to better compensate varying demand over the day (e.g. short 2h shifts).

5 Conclusion and Future Research

In this paper, a model for strategic long-term staffing was presented considering varying demand, different types of employees regarding skills and contractual working times as well as compensation of overtime due to flextime policies. For this purpose, an evolutionary bilevel algorithm with GA on both levels was applied, optimizing the staffing decision at the upper-level and simultaneously evaluating the resulting workforce structure by the creation of personnel schedules over a planning horizon of 52 weeks. This integrated staffing and scheduling approach was demonstrated by the example of the yearly workforce planning of a midsized call center. The computational results indicate that the proposed procedure could be used to support corporate decision making related to strategic workforce planning. Due to the nested structure and independent formulation of the staffing and the scheduling problem, both problems could generically be replaced. Therefore, the model is not limited to the considered call center problem but could be used for any other kind of (strategic) workforce planning involving personnel scheduling.

However, an important limitation arises from the fact that the optimization problem at the lower-level only was executed once, which leads to noisy results. Thus, there is a risk of discarding a possibly good solution due to one "unlucky" optimization run at the lower level. Moreover, it is hard to compare solutions being close to each other (e.g. solutions 4 to 6). These issues could be solved by restarting the lower-level algorithm multiple times, which in turn will lead to a massive increase of computation time. As this is a general challenge when applying bilevel optimization, further research should be aimed at executing the bilevel algorithm more efficiently, e.g. by applying more precise termination criteria, using distributed computation or approximation of the lower-level model. Other opportunities for further research are seen in comparing the here proposed approach to the methods presented in Section 1 (e.g. regarding speed and quality) as well as the consideration of uncertainty and unplanned events during the optimization procedure, such as illness, fluctuation or infra-annual hiring of employees.

As stated in Section 4.2, there may be the need not only to optimize staffing decisions, but also staffing and scheduling policies. The study presented in this paper was limited to optimize the staffing decision as input of the scheduling problem. However, further research should be conducted addressing the possibility to optimize all framework conditions related to personnel scheduling, such as shift types, overtime and break regulations or any other type of adjustable constraints. This could potentially increase the solution quality attainable, but would in turn raise the complexity of the optimization problem considerably.

References

1. Abernathy WJ, Baloff N, Hershey JC, Wandel S, A Three-Stage Manpower Planning and Scheduling Model—A Service-Sector Example, Operations Research 21(3), 693–711 (1973)
2. Avramidis AN, Chan W, Gendreau M, L'Ecuyer P, Pisacane O, Optimizing daily agent scheduling in a multiskill call center, European Journal of Operational Research 200(3), 822–832 (2010)
3. Beliën J, Demeulemeester E, A branch-and-price approach for integrating nurse and surgery scheduling, European Journal of Operational Research 189(3), 652–668 (2008)

4. Beliën J, Demeulemeester E, Bruecker P de, van den Bergh J, Cardoen B, Integrated staffing and scheduling for an aircraft line maintenance problem, Computers & Operations Research 40(4), 1023–1033 (2013)
5. Bezanson J, Edelman A, Karpinski S, Shah VB, Julia. A Fresh Approach to Numerical Computing, SIAM Rev. 59(1), 65–98 (2017)
6. Branke J (ed.), Multiobjective optimization. Interactive and evolutionary approaches. Lecture Notes in Computer Science, vol. 5252. Springer, Berlin (2008)
7. Bruecker P de, van den Bergh J, Beliën J, Demeulemeester E, Workforce planning incorporating skills. State of the art, European Journal of Operational Research 243(1), 1–16 (2015)
8. Brunner JO, Edenharter GM, Long term staff scheduling of physicians with different experience levels in hospitals using column generation, Health care management science 14(2), 189–202 (2011)
9. Coello Coello CA, Lamont GB, van Veldhuizen DA, Evolutionary algorithms for solving multi-objective problems, 2nd edn. Genetic and evolutionary computation series. Springer, New York, NY (2007)
10. Colson B, Marcotte P, Savard G, An overview of bilevel optimization, Ann Oper Res 153(1), 235–256 (2007)
11. Easton FF, Rossin DF, Borders WS, Analysis of Alternative Scheduling Policies for Hospital Nurses, Production and Operations Management 1(2), 159–174 (1992)
12. Erhard M, Schoenfelder J, Fügener A, Brunner JO, State of the art in physician scheduling, European Journal of Operational Research (2017)
13. Günther M, A Generator for Volatile Demand Profiles. A Brief Description of a Tool. GRIN, Munich (2017)
14. Hansen P, Jaumard B, Savard G, New Branch-and-Bound Rules for Linear Bilevel Programming, SIAM J. Sci. and Stat. Comput. 13(5), 1194–1217 (1992)
15. Henderson JC, Krajewski LJ, Showalter MJ, An integrated approach for manpower planning in the service sector, Omega 10(1), 61–73 (1982)
16. Maenhout B, Vanhoucke M, An integrated nurse staffing and scheduling analysis for longer-term nursing staff allocation problems, Omega 41(2), 485–499 (2013)
17. Nissen V, Applications of Evolutionary Algorithms to Management Problems. In: Moutinho, L., Sokele, M. (eds.) Palgrave Handbook of Innovative Research Methods in Management. Palgrave Macmillan, Basingstoke (2017)
18. Rothlauf F, Design of Modern Heuristics. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)
19. Sean Luke, Essentials of Metaheuristics. Lulu (2013)
20. Sinha A, Malo P, Deb K, A Review on Bilevel Optimization. From Classical to Evolutionary Approaches and Applications, IEEE Trans. Evol. Computat., 1 (2017)
21. Srinivas N, Deb K, Muiltiobjective Optimization Using Nondominated Sorting in Genetic Algorithms, Evolutionary computation 2(3), 221–248 (1994)
22. Stackelberg H, The theory of market economy. Oxford University Press, Oxford (1952)
23. Talbi E-G, Metaheuristics. From design to implementation. Wiley Series on Parallel and Distributed Computing, v.74. John Wiley & Sons, Hoboken, NJ (2009)
24. Talbi E-G (ed.), Metaheuristics for Bi-level Optimization. Studies in Computational Intelligence, vol. 482. Springer, Berlin, Heidelberg (2013)
25. van den Bergh J, Beliën J, Bruecker P de, Demeulemeester E, Boeck L de, Personnel scheduling. A literature review, European Journal of Operational Research 226(3), 367–385 (2013)
26. Venkataraman R, Brusco MJ, An integrated analysis of nurse staffing and scheduling policies, Omega 24(1), 57–71 (1996)
27. Yin Y, Genetic-Algorithms-Based Approach for Bilevel Programming Models, Journal of Transportation Engineering 126(2), 115–120 (2000)