A hybrid case-based reasoning approach to detecting the optimal solution in nurse scheduling problem

SVETLANA SIMIĆ*, DRAGANA MILUTINOVIC** AND SLOBODAN SEKULIĆ†, Faculty of Medicine, University of Novi Sad, Hajduk Veljkova 1-9, Novi Sad 21000, Serbia.

DRAGAN SIMIĆ†† AND SVETISLAV D. SIMIĆ§, Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, Novi Sad 21000, Serbia.

JOVANKA ĐORĐEVIĆ§§, Oncology Institute of Vojvodina, Put doktora Goldmana 4, Sremska Kamenica 21208, Serbia.

Abstract
Demand for healthcare is increasing due to a growing and ageing population. Choosing an adequate schedule for medical staff can be a difficult dilemma for managers. The goal of nurse scheduling is to minimize the cost of the staff while maximizing their preferences and the overall benefits for the unit. This paper is focused on a new hybrid strategy based on detecting the optimal solution in nurse scheduling problem. The new proposed hybrid approach is obtained by combining case-based reasoning and general linear empirical model with arbitrary coefficients. The model is tested with original real-world data set obtained from the Oncology Institute of Vojvodina in Serbia.

Keywords: Nurse scheduling problem, case-based reasoning, general linear empirical model, medical staff.

1 Introduction
Demand for healthcare is increasing due to a growing and ageing population, which makes access to medical care more difficult. Medical staff performance represents a significant determinant of public healthcare quality. There is an excessive pressure for cost reduction, which negatively influences work–life balance for a small number of employed physicians and nurses and often results in a decrease of demanded quality of services. Moreover, due to the challenging economic conditions, not only has ever greater number of physicians and nurses from public hospitals moved to live and work abroad, but a lot of them are also employed in private healthcare organizations to earn higher salaries. This tendency has caused the critical issue in medical staff preferences; medical staff satisfaction is a

---

* E-mail: svetlana.simic@mf.uns.ac.rs
** E-mail: dragana.milutinovic@mf.uns.ac.rs
† E-mail: sekulic.slobodan@mf.uns.ac.rs
†† E-mail: dsimic@eunet.rs, dsimic@uns.ac.rs
§ E-mail: simicsvetislav@uns.ac.rs
§§ E-mail: jovankad10@gmail.com

Vol. 28, No. 2, © The Author(s) 2018. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited. Advance Access published 1 October 2018 doi:10.1093/jigpal/jzy047
A hybrid case-based reasoning approach to detecting the optimal solution in nurse scheduling problem

fundamental part in providing the necessary care for patients. There are two aspects to this problem: (i) physician scheduling problem (PSP) and, on the other hand, (ii) nurse scheduling problem (NSP).

This PSP is more complex than the NSP since residents still need an educational praxis to get licensed as physicians. Whereas many physicians generally have individual contracts with their hospital with specific and limited details, it is more challenging for the scheduling process to involve these intricate agreements and these physicians will not be scheduled with other hospital staff.

Choosing an adequate schedule for nursing staff can be a difficult dilemma for nurse managers. They need to obtain balance between the staff’s individual preferences and the overall benefits for the unit and, consequently, the patients. Therefore, traditional approaches to addressing the challenges of clinical staff organization and scheduling are not always effective in modern complex healthcare environment [23]. More state legislatures are mandating specific nurse-staffing levels, and many nurses are dissatisfied with their work schedules. Optimal solutions derived from techniques with high computing times are usually less valuable than the ones based on a flexible algorithm or user-intuitive application [2].

This paper is focused on a new strategy based on hybrid approach to detecting the optimal solution in NSP. The new proposed hybrid approach is obtained by combining case-based reasoning (CBR) and general linear empirical model with arbitrary coefficients. The model is tested with original real-world data set obtained from the Oncology Institute of Vojvodina (OIoV) in Serbia. Also, this paper continues the authors’ previous research in nurse decision-making, scheduling and rostering healthcare organizations which are presented in [12, 15, 17–19].

The rest of the paper is organized in the following way: Section 2 provides an overview of the basic idea in NSP, related work with solution approaches based on general methods, classical heuristics methods and metaheuristics and subsection about CBR reasoning method. Section 3 presents the NSP proposed in this paper, based on hard/soft constraints, CBR representation of empirical data set and the proposed algorithm for NSP. Experimental results and its verification are presented in Section 4. Section 5 provides conclusions and some points for future work.

2 NSP and related work

The NSP is a well-known non-polynomial (NP)-hard scheduling problem that aims to allocate the required workload to the available staff nurses at healthcare organizations to meet the operational requirements and a range of preferences. The NSP is a 2D timetabling problem that deals with the assignment of nursing staff to shifts across a scheduling period subject to certain constraints.

In general, there are two basic types of scheduling used for the NSP: cyclic and non-cyclic scheduling. In cyclic scheduling, each nurse works in a pattern which is repeated in consecutive scheduling periods; whereas, in non-cyclic scheduling, a new schedule is generated for each scheduling period: weekly, fortnightly or monthly. Cyclic scheduling was first used in the early 1970s due to its low computational requirements and the possibility for manual solution [5].

2.1 Related work in NSP

Studies of NSPs date back to the early 1960s. Despite decades of research into automated methods for nurse scheduling and some academic success, it may be noticed that there is no consistency in the knowledge that has been built up over the years and that many healthcare institutions still resort to manual practices. One of the possible reasons for this gap between the nurse scheduling theory and practice is that oftentimes academic community focuses on the development of new techniques rather than developing systems for healthcare institutions [3].
In the past decades, many approaches have been proposed to solve NSP as they are manifested in different models. The three commonly used general methods are mathematical programming, heuristics and artificial intelligence approaches. Many heuristics approaches were straightforward automation of manual practices, which have been widely studied and documented [7, 22].

For combinatorial problems, exact optimization usually requires large computational times to produce optimal solutions. In contrast, metaheuristic approaches can produce satisfactory results in reasonably short times. In recent years, metaheuristics including tabu search algorithm (TS), genetic algorithm (GA) and simulated annealing, have been proven as very efficient in obtaining near-optimal solutions for a variety of hard combinatorial problems including the NSP [4].

Some TS approaches have been proposed to solve the NSP. In TS, hard constraints remained fulfilled, while solutions move in the following way: calculate the best possible move which is not tabu, perform the move and add characteristics of the move to the tabu list. The TS with strategic oscillation used to tackle the NSP in a large hospital is presented in [10].

GA, which is stochastic metaheuristics method, has also been used to solve the NSP. In GA, the basic idea is to find a genetic representation of the problem so that ‘characteristics’ can be inherited. Starting with a population of randomly created solutions, better solutions are more likely to be selected for recombination into novel solutions. In addition, these novel solutions may be formed by mutating or randomly changing the old ones [8].

2.2 CBR

CBR is a technique that has its origins in knowledge-based systems. CBR systems learn from previous situations. The main element of a CBR system is the CASE BASE. It is a structure that stores problems, elements (cases), and their solutions. So, a case base can be visualized as a database that stores a collection of problems with some sort of relationship to solutions to every new problem, which gives the system the ability to generalize to solve any new problem.

The learning capabilities of CBR system rely on their own structures, which consist of four main phases: retrieval, reuse, revision and retain. Figure 1 shows a graphical representation of those four phases. The retrieval phase consists of finding the cases in the CASE BASE that most closely resemble the proposed problem. Once a series of cases have been extracted from the CASE BASE, they must be reused by the system. In the second phase, the selected cases are adapted to fit the current problem. After offering a solution to the problem, it is then revised, to check whether the proposed alternative is in fact a reliable solution to the problem. If the proposal is confirmed, it is retained by the system, modifying some knowledge containers and could eventually serve as a solution for problems in the future.

CBR has been used to solve a variety of problems in health care sciences [24], financial predictions [13, 14, 16], an agent system for detecting Structured Query Language (SQL) injection attacks [11], solving the oil spill problem [9] and everywhere where images can play a key role [6].

3 Modelling the NSP

This research is focused on cyclic scheduling on NSP in planning period in intensive care unit (ICU) at the OIoV. Cyclic scheduling is used here, where each nurse follows a pattern repeated in consecutive scheduling periods.
A hybrid case-based reasoning approach to detecting the optimal solution in nurse scheduling problem

3.1 Hard constraints

Recently, duty rosters are generated manually by head nurse for ICU, which enables the nurses to express their requests and preferences for working/or not working certain shifts, holidays and days off. Nurses in the unit have different skills categories, meaning different qualifications, specialization training, experience and gender, presented in Table 1.

Regular work days are 5 days per week, from Monday to Friday. Regular working hours are 7 hours and 12 minutes. Full-time nurses are defined by: multiple regular work days * regular working hours. When this number is rounded, it represents the total number of shifts allowed per month. Nurses can work in three On-duty shifts: Day (D) (06:30–18:30), Night (N) (18:30–06:30) and Morning Shift (X) (07:30–14:42). There are also Free shifts which include Day off (O), Sick days (B), Sick days on weekend (C), Maternity leave (K), Paid leave (P), Annual leave on weekends (V), Annual leave (Y); the abbreviations for all of these are used in Table 2.

**Hard requests** define a constraint that must be respected in the roster and **Soft requests** define the preferred option expressed by a nurse which is desirable but can be violated in the roster if needed. Some typical values for a few of the constraints are given below:

- Min (max) nurses on shifts: In the OIÖV, three nurses in Day shifts, three nurses in Night shifts;
A hybrid case-based reasoning approach to detecting the optimal solution in nurse scheduling problem

Table 2. Nurse-case duties for 31 January 2014 are starting for calculation nurse schedule for 1 February 2014

| Case no. | Nurse-ID | Date    | Field 1 | Field 2 | Field 3 | Field 4 | Field 5 | Field 6 | Field 7 |
|----------|----------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1        | N-01     | 01.01   | D       | O       | O       | N       | O       | N       | O       |
| 2        | N-01     | 02.01   | O       | O       | N       | O       | N       | O       | O       |
| 25       | N-01     | 25.01   | D       | N       | O       | O       | O       | O       | N       |
| 26       | N-01     | 26.01   | N       | O       | O       | O       | O       | N       |         |
| 27       | N-02     | 01.01   | D       | O       | O       | D       | N       | O       |         |
| 28       | N-02     | 02.01   | O       | O       | D       | N       | O       | O       | D       |
| 51       | N-02     | 25.01   | C       | C       | B       | B       | B       | B       |         |
| 52       | N-02     | 26.01   | C       | B       | B       | B       | B       | B       |         |
| 417      | N-17     | 01.01   | O       | O       | O       | N       | N       | O       | D       |
| 418      | N-17     | 02.01   | O       | O       | N       | N       | O       | D       | N       |
| 441      | N-17     | 25.01   | D       | O       | O       | N       | O       | O       | O       |
| 442      | N-17     | 26.01   | O       | O       | N       | O       | O       | O       |         |
| 495      | N-20     | 01.01   | O       | D       | N       | O       | O       | D       | O       |
| 496      | N-20     | 02.01   | D       | N       | O       | O       | D       | O       | O       |
| 519      | N-20     | 25.01   | D       | N       | O       | O       | O       | O       | N       |
| 520      | N-20     | 26.01   | N       | O       | O       | O       | O       | N       |         |

- It is not desirable to work a *Night shift* followed by a *Day shift*;
- After 5 *Morning shifts*, 2 days off must be assigned;
- After a break of more than 7 days, (annual leave, sick leave) *Day shift* must be assigned;
- Maximum differences between *Day shifts* and *Night shifts* per nurse could be no greater than five;
- At least one of the members of *Shift* must be *shift leader*, which is for every nurse defined in Table 1;
- Max (min) days: Full-time nurses may not work more than predetermined number of days.

The ideal and proposed work shift dynamic is Day–Night–Off–Off–Off (DNOOO). Day–Night, meaning that 2 work shifts and 3 days off in 5 days is ideal shift dynamic. This is recommended by the OIoV management. Also, the DONOO dynamic is allowed, where there are 2 work shifts and 3 days off in 5 days, other combinations of two work shifts and 3 days off in 5 days are allowed as well. But, in the real world, when creating the nurse scheduling, it is impossible to have ideal work shift dynamic. For that reason, a more difficult shift dynamic is allowed, e.g. 3 working days and 2 days off (DDNOO) (DNNOO). Other dynamics of 3 working days and 2 days off are allowed as well. After five *Morning shifts* (XXXXX) 2 days off must be assigned to create (XXXXXOO) dynamic.
3.2 Empirical data set—CBR representation

For this experiment original real-world data set between 1 January and 31 January 2014, from ICU of OloV is used.

The part of experimental data set is presented in Table 2 where the columns are presented: Case No., case number; Nurse-ID = N-01; Date, starting date for the case (for Case No. = 1, Date = 01.01.), Field 1 for date 01.01. the Value of shift = D; Field 2 for date 02.01. the Value of shift = O; Field 3 for date 03.01. the Value of shift = O; Field 4 for date 04.01. the Value of shift = N; Field 5 for date 05.01. the Value of shift = O; Field 6 for date 06.01. the Value of shift = N; Field 7, Case Solution for date 07.01. the Value of shift = O). The end of concise description looks like this: Case No. = 26, Date = 26.01.; Field 1 for date 26.01. the Value of shift = N; Field 2 for date 27.01. the Value of shift = O; Field 3 for date 28.01. the Value of shift = O; Field 4 for date 29.01. the Value of shift = N; Field 5 for date 30.01. the Value of shift = O; Field 6 for date 31.01. the Value of shift = N; Field 7, Case Solution for date 01.02. the Value of shift = is empty and grey. There is no Solution for that case, and it will be calculated when the system calculates schedule for nurse N-01 for 01.02. In CBR basic representation, Case No. = 26 presents NEW CASE, and the hybrid system will try to find Optimal Solution for it. All cases (data set) stored in CASE BASE can be described in the same manner as for the previous nurse. All the cases in CASE BASE which have Solution will be used in reused and revised CBR phases for detecting the best Solution in NSP.

3.3 The algorithm for NSP

The proposed hybrid model is obtained by combining CBR and general linear empirical model with arbitrary coefficients. The basic steps of the proposed hybrid algorithm for NSP are summarized by the pseudo code shown in Algorithm 1 and the most important CBR phases and general linear empirical model with arbitrary coefficients are presented in blue bold colour. Our algorithm is inspired by integration of CBR method, which is discussed in Section 2.2, CASE BASE representation is shown in Section 3.2. The general linear empirical model with arbitrary coefficient defined from (1–4) is in detail presented and discussed in [20].

Equation (1) presents calculation of weighted value \( V_a(t,d) \) Day shift occurrence \( d \) for nurse \( t \), while \( S_7(t,d) \) presents frequency of next letter = \( D \) when it is calculated for nurse \( t \) for pattern. The same logic applies to \( S_6(t,d), S_5(t,d), S_4(t,d), S_3(t,d) \). Also, (2) calculation of weighted value \( V_a(t,n) \) Night shift occurrence \( n \) for nurse \( t \), while \( S_7(t,n) \) presents frequency of next letter = \( N \) when it is calculated for nurse \( t \). The same logic applies to \( S_6(t,n), S_5(t,n), S_4(t,n), S_3(t,n) \). Equation (3) presents calculation of weighted value \( V_a(t,o) \) Day off shift occurrence \( o \) for nurse \( t \), while \( S_7(t,o) \) presents frequency of next letter = \( O \) when it is calculated for nurse \( t \). The same logic applies to \( S_6(t,o), S_5(t,o), S_4(t,o), S_3(t,o) \). The values of \( V_a(t,d) \), \( V_a(t,n) \) and \( V_a(t,o) \) then must be normalized and as such represent probability for shift occurrence for a specific worker for the next day.

\[
V_a(t,d) = S_7(t,d) + 0.8 \times S_6(t,d) + 0.6 \times S_5(t,d) + 0.4 \times S_4(t,d) + 0.2 \times S_3(t,d) \quad (1)
\]

\[
V_a(t,n) = S_7(t,n) + 0.8 \times S_6(t,n) + 0.6 \times S_5(t,n) + 0.4 \times S_4(t,n) + 0.2 \times S_3(t,n) \quad (2)
\]

\[
V_a(t,o) = S_7(t,o) + 0.8 \times S_6(t,o) + 0.6 \times S_5(t,o) + 0.4 \times S_4(t,o) + 0.2 \times S_3(t,o) \quad (3)
\]
Considering that the letters, candidates for the next shift occurring after the pattern string of cases with different lengths, do not have the same significance, the weighted factor for each of them is introduced. Thus, the weighted factor of cases for the frequency of pattern String-7 is 1, the weighted factor cases for the frequency of pattern String-6 is 0.8, the weighted factor cases for the frequency of pattern String-5 is 0.6, the weighted factor cases for the frequency of pattern String-4 is 0.4 and the weighted factor cases for the frequency of pattern String-3 is 0.2, as shown in (1–3).

\[
\arg \max_x V(t,x) := x \in \{d,n,o\}
\]  

(4)

Partially target function, defined by arguments of the maxima (arg max), which occurs in (1–3) is presented in (4), where arg max refers to the inputs, or arguments, at which the function outputs are as large as possible. The arg max are the points of the domain of some function at which the function values are maximized.

4 Experimental results and verification

Table 3 presents calculation for the nurse candidates for the 1 February based on the range from (1–3). For every workday it is necessary to select three nurses for Day shift, and three nurses for Night shift.

4.1 Experimental results

The calculation for the nurse candidates for the 1 February based on the range from (1–3). Looking only at Table 3, it could be concluded that Before revision—the best six candidates for Day shift will consist of the following nurses: N-17 | 73.75%; N-09 | 67.05%; N-06 | 49.29%; N-15 | 39.25%; N-11 | 20.51%; N-12 | 20.51%; and that, Before revision, the best six candidates for Night shift are the following nurses: N-07 | 52.63%; N-06 | 44.36%; N-08 | 42.59%; N-16 | 25.23%; N-03 | 13.95%; N-13 | 12.50%.

Revision process considers additional calculations which are control hard constraints, corrects and modifies calculated data Day and Night shift candidates for Final solutions for the Day and Night shift. Therefore, it is interesting to see Table 3 where, in N-09, there is a great imbalance between Day shift – 9 and Night shift – 4. The system allows the greatest difference between Day shifts and Night shifts to be <= 2 in the same month. Therefore, N-09 cannot work Day Shift, because shifts type imbalance would be even greater. **N-09 becomes Night shift candidate.**

Also, revision, Day shift candidate, will consist of the following nurses in the following order: N-17, N-06, N-15, N-11 and N-12. In the first three Day shift candidates the constraint that at least one member of Shift must be shift leader is satisfied, even two of them N-06 and N-15. Finally, Day shift for 1 February could be completed: N-17, N-15 and shift leader is N-06.

Following the rules and constraints in CBR revise phase, N-09 is Night shift candidate, but on the other side **N-03** has a great imbalance between Night shift—8 and Day shift—4. Therefore, N-11 and N-12 are added from Day shift candidate, also because both of them have higher working percent 16.41 then N-13 with working percent 12.50. And now Night shift candidates are as follows: N-09, N-07, N-08, N-16, N-11, N-12 and N-13. Finally, Night shift for 1 February could be completed: shift leader is N-09, and the members are N-07, N-08. Now, the whole shifts for the 1 February are completed.

In CBR retained phase, now is time to update some cases stored in CASE BASE, which are empty and are shown in grey in Table 2. It is necessary to fill the Field 7 Solution with appropriate Shift
Algorithm 1 The algorithm for NSP

Begin

Step 1: --- Initialization.
- Constraints hard/soft;
- Nurse_number = 20; Nurse skills categories;
- Days = Time period; Ideal shift dynamic = ‘DNOOO’;
- Nurse-day duties = from 1 January to last day for prediction;

Step 2: --- Calculation for time period. Loop over the time period.
for i = 1: Days (all Time period) do
  --- Calculation for Nurses. CBR Retrieval phase
  for j = 1: Nurse_number (all Nurses) do
    New case = String–7
    --- Calculation for case length 7 - String–7
    find pattern String–7 for current nurse
    find frequency S_7(t,d), S_7(t,n), S_7(t,o) for current nurse
    --- Calculation for case length 6 - String–6
    find pattern String–6 for current nurse
    find frequency S_6(t,d), S_6(t,n), S_6(t,o) for current nurse
    --- Calculation for case length 5 - String–5
    find pattern String–5 for current nurse
    find frequency S_5(t,d), S_5(t,n), S_5(t,o) for current nurse
    --- Calculation for case length 4 - String–4
    find pattern String–4 for current nurse
    find frequency S_4(t,d), S_4(t,n), S_4(t,o) for current nurse
    --- Calculation for case length 3 - String–3
    find pattern String–3 for current nurse
    find frequency S_3(t,d), S_3(t,n), S_3(t,o) for current nurse
    --- General linear empirical model
    calculating weighted value ‘D’ ‘N’ ‘O’ for cur. nurse
    Equations (1–3) according (4)
  end for j
  --- End Calculation for Nurses.

Step 3: CBR Reuse phase - Normalizing type of shift for nurses for cur. day
List – (Day/Night) Shift Nurses as candidates for current day
Calculate nurse workload before current day
List – Nurse Workload Most appropriate nurses according hard/soft constraints for current day CBR Revision phase- adaptation process according Day and Night Shift; Nurse Workload, define Shift Leader for current day
end for i
  --- End Calculation for time period. Loop over the time period.

Step 4: CBR Retained phase - Update some cases stored and add New cases in CASEBASE
Post-processing the results and visualization.

End.
A hybrid case-based reasoning approach to detecting the optimal solution in nurse scheduling problem

TABLE 3. Nurse-case duties are starting the calculation for nurse schedule for 1 February, and Final Solution is given for 1 February

| Nurse-ID | Before revision | After revision | Final solution |
|----------|----------------|---------------|----------------|
|          | Day shift | Night shift | Day off | Shift leader | Num. day | Num. night | Dayshift | Night shift |          |
| N-01     | 0        | 8.51        | 91.48   | Yes         | 7        | 6          | -        | -          |          |
| N-02     | 0        | 0           | 100.00  | No          | 3        | 1          | -        | -          |          |
| N-03     | 0        | 13.95       | 86.04   | Yes         | 5        | 8          | -        | 0.00       |          |
| N-04     | 0        | 0           | 100.00  | No          | 0        | 0          | -        | -          |          |
| N-05     | 0        | 0           | 100.00  | Yes         | 6        | 6          | -        | -          |          |
| N-06     | 49.29    | 44.36       | 6.33    | Yes         | 1        | 2          | 69.01    | -          | 69.01 (L)|
| N-07     | 0        | 52.63       | 47.36   | Yes         | 6        | 4          | -        | 52.63      | 52.63    |
| N-08     | 10.18    | 42.59       | 47.22   | No          | 7        | 6          | -        | 42.59      | 42.59    |
| N-09     | 67.05    | 2.62        | 30.32   | Yes         | 9        | 4          | 0.00     | 67.05      | 67.05 (L)|
| N-10     | 5.22     | 6.86        | 87.90   | Yes         | 7        | 6          | -        | -          |          |
| N-11     | 20.51    | 1.70        | 77.77   | Yes         | 5        | 4          | 16.41    | 16.41      |          |
| N-12     | 20.51    | 1.70        | 77.77   | Yes         | 5        | 2          | 16.41    | 16.41      |          |
| N-13     | 10.93    | 12.50       | 76.56   | No          | 2        | 3          | -        | 12.50      |          |
| N-14     | 5.22     | 6.86        | 87.90   | No          | 7        | 7          | -        | -          |          |
| N-15     | 39.25    | 7.47        | 53.27   | Yes         | 6        | 5          | 31.40    | -          | 31.40    |
| N-16     | 16.82    | 25.23       | 57.94   | Yes         | 2        | 2          | -        | 25.23      |          |
| N-17     | 73.75    | 8.51        | 17.73   | No          | 5        | 7          | 73.75    | -          | 73.75    |
| N-18     | 0        | 0           | 100.00  | No          | 6        | 6          | -        | -          |          |
| N-19     | 0        | 0           | 100.00  | No          | 3        | 1          | -        | -          |          |
| N-20     | 0        | 8.51        | 91.48   | No          | 8        | 5          | -        | -          |          |

(Day/Night/Morning) or store some other Free shifts. For example, for case 442, the fields would be ‘OONOOOD’. Also, in this CBR phase it is necessary to add twenty new cases in CASE BASE, e.g. case 537 presents nurse N-17, the beginning date is 27 January 2014 and the six fields would be ‘ONOOOD’ and the Field 7 Solution is empty and grey, and presents new case for solving schedule for 2 February.

The data store in CASE BASE after retained phase is presented in Table 4. The rest of the schedule for the whole period continues as previously described: algorithm, hard constraints, soft constraints and established rules, and in the 2D timetabling is presented in Table 5.

4.2 Verification of experimental results

To verify our research methods, experimental results are compared with real-world data set obtained from the OIoV. A cumulative Workload is calculated for every nurse and summarized for every month, and after that Coefficient of variation (CV) is calculated and compared with original data set and experimental data set. Cumulative workload for February presents workload from the beginning of the year until February. The same logic applies to cumulative workload for March. Coefficient of variation is dimensionless and is defined as the ratio of the standard deviation (SD) to the mean. Hence, coefficient of variation is a useful quantity for comparing the variability in data sets having
A hybrid case-based reasoning approach to detecting the optimal solution in nurse scheduling problem

### TABLE 4. Calculated final solution nurse-case duties schedule for 1 February and the starting calculation for 2 February 2014

| Case no. | Nurse-ID | Date   | Field 1 | Field 2 | Field 3 | Field 4 | Field 5 | Field 6 | Field 7 Solution |
|----------|----------|--------|---------|---------|---------|---------|---------|---------|-----------------|
| 1        | N-01     | 01.01  | D       | O       | O       | N       | O       | N       | O              |
| 2        | N-01     | 02.01  | O       | O       | N       | O       | N       | O       | O              |
|          |          |        |         |         |         |         |         |         |                |
| 25       | N-01     | 25.01  | D       | N       | O       | O       | O       | N       | O              |
| 26       | N-01     | 26.01  | N       | O       | O       | O       | N       | O       | O              |
| 521      | N-01     | 27.01  | O       | O       | O       | O       | N       | O       | O              |
| 27       | N-02     | 01.01  | D       | O       | O       | D       | N       | O       | O              |
| 28       | N-02     | 02.01  | O       | O       | D       | N       | O       | O       | D              |
|          |          |        |         |         |         |         |         |         |                |
| 51       | N-02     | 25.01  | C       | C       | B       | B       | B       | B       | B              |
| 52       | N-02     | 26.01  | C       | B       | B       | B       | B       | B       | V              |
| 522      | N-02     | 27.01  | B       | B       | B       | B       | V       |        |                |
|          |          |        |         |         |         |         |         |         |                |
| 417      | N-17     | 01.01  | O       | O       | O       | N       | N       | O       | D              |
| 418      | N-17     | 02.01  | O       | O       | N       | N       | O       | D       | D              |
|          |          |        |         |         |         |         |         |         |                |
| 441      | N-17     | 25.01  | D       | O       | O       | N       | O       | O       | O              |
| 442      | N-17     | 26.01  | O       | O       | N       | O       | O       | O       | D              |
| 537      | N-17     | 27.01  | O       | N       | O       | O       | O       | D       |                |
|          |          |        |         |         |         |         |         |         |                |
| 495      | N-20     | 01.01  | O       | D       | N       | O       | O       | D       | O              |
| 496      | N-20     | 02.01  | D       | N       | O       | O       | D       | O       | O              |
|          |          |        |         |         |         |         |         |         |                |
| 519      | N-20     | 25.01  | D       | N       | O       | O       | O       | O       | N              |
| 520      | N-20     | 26.01  | N       | O       | O       | O       | O       | N       | O              |
| 540      | N-20     | 27.01  | O       | O       | O       | O       | N       | O       |                |

Different SDs and different means [21]. Then Correlation between Means and statistical Significant Difference F-test between data sets is calculated, and results are presented in Table 6.

Comparison means of real-world data set and experimental data set can be shown as high correlations, and CV is lower in every month in experimental data set as presented in Table 6. Experimental model is, also, verified by Univariate Analysis of Variance, by the Tests of Between-Subjects Effects, and interaction effect between real-world data set, and experimental data set for June was not statistically significant, $F = 0.108$, Sig = 0.744, which is shown in Table 6. This result shows that the differences between real-world and experimental data set do not exist and it could be concluded that experimental results fit quite well when they are compared in Workload aspect.

The nurses’ satisfaction level in original (previous) real-world shift planning and in new nurse scheduling timetable are compared, and on one hand, it can be concluded that nurses are, in general, much more satisfied, but on the other hand a small ward group of nurses are not working together, as in previous timetable. Taken generally, it is much better, for ward work quality, when employees are combined according to different criteria: years of service—experience, specialization training,
TABLE 5. Nurse-day view duties in January and planning period in February

|       | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 10      |
|-------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
|       | January | February| January | February| January | February|
|       |         |         |         |         |         |         |         |         |         |         |
| DQDO  | DQOD    | DOOD    | DQDO    | DQOD    | DQDO    | DQOD    | DQDO    | DQOD    | DQDO    | DQDO    |
| DQOD  | DQOD    | DQDO    | DQDO    | DQDO    | DQDO    | DQDO    | DQDO    | DQDO    | DQDO    | DQDO    |
| DQDO  | DQDO    | DQDO    | DQDO    | DQDO    | DQDO    | DQDO    | DQDO    | DQDO    | DQDO    | DQDO    |
| DQDO  | DQDO    | DQDO    | DQDO    | DQDO    | DQDO    | DQDO    | DQDO    | DQDO    | DQDO    | DQDO    |

TABLE 6. Mean, SD, CV, Correlation between Means, Significant of Univariate Analysis of Variance for Workload by original real-world data set and experimental data set

| Workload  | Real-world data set | Experimental data set | Correlation Means | Univariate Analysis of Variance |
|-----------|---------------------|-----------------------|-------------------|---------------------------------|
|           | Mean SD CV (%)      | Mean SD CV (%)        |                   |                                 |
| January   | 0.387 0.1036 26.77  | 0.387 0.1036 26.74    | 1.000             | -                               |
| February  | 0.408 0.0326 8.00   | 0.411 0.0302 7.34     | 0.746             | -                               |
| March     | 0.401 0.0303 7.55   | 0.401 0.0258 6.43     | 0.813             | -                               |
| April     | 0.401 0.0299 7.45   | 0.405 0.0238 5.87     | 0.836             | -                               |
| May       | 0.404 0.0302 7.48   | 0.406 0.0266 6.55     | 0.801             | -                               |
| June      | 0.399 0.0261 6.55   | 0.402 0.0248 6.16     | 0.817             | F = 0.108                      |

5 Conclusion and future work

The aim of this paper is to propose the new hybrid strategy for detecting the optimal solution in NSP. The new proposed hybrid approach is obtained by combining CBR and general linear empirical
model with arbitrary coefficients. The model is tested with original real-world data set obtained from the OIoV in Serbia.

The data set is represented in CASE BASE as a database structure where problems, elements (cases), and their solutions are stored. All the cases in CASE BASE which have Solution will be used in reused and revised CBR phases for detecting best Solution in NSP, using hard/soft constraints, rules and general linear empirical model with arbitrary coefficients. In CBR retained phase, some cases stored in CASE BASE are updated, and the new cases are added.

Preliminary experimental results encourage further research because data set is stored in database and it is easily manipulated. Our future research will focus on creating new hybrid model combined by intuitive thinking style, to solve problems logically, considering different options until discover the best solution, which will efficiently solve NSP. The new model will be tested with original real-world data set for longer periods, including the year 2017, obtained from the OIoV in Serbia.

References

[1] A. Aamodt and E. Plaza. Case-based reasoning: foundational issues, methodological variations, and system approaches. *AI Communications*, 7, 39–59, 1994.
[2] U. Aickelin and P. White. Building better nurse scheduling algorithms. *Annals of Operations Research*, 128, 159–177, 2004.
[3] G. Beddoe, S. Petrović and J. Li. A hybrid metaheuristic case-based reasoning system for nurse rostering. *Journal of Scheduling*, 12, 99–119, 2009.
[4] B. Cheang, H. Li and B. Rodrigues. Nurse rostering problems—a bibliographic survey. *European Journal of Operational Research*, 151, 447–460, 2003.
[5] K. A. Dowsland. Nurse scheduling with tabu search and strategic oscillation. *European Journal of Operational Research*, 106, 393–407, 1998.
[6] A. Herrero, E. Corchado, M. A. Pellicer and A. Abraham. MOVIHIDS: a mobile-visualization hybrid intrusion detection system. *Neurocomputing*, 72, 2775–2784, 2009.
[7] M. W. Isken and W. M. Hancockm. A heuristic approach to nurse scheduling in hospital units with non-stationary, urgent demand and a fixed staff size. *Journal of the Society for Health Systems*, 2, 24–40, 1991.
[8] K. Leksakul and S. Phetsawat. Nurse scheduling using genetic algorithm. *Mathematical Problems in Engineering*, Article ID 246543, http://dx.doi.org/10.1155/2014/246543, 2014.
[9] A. Mata, E. Corchado and B. Baruque. Solving the oil spill problem using a combination of CBR and a summarization of SOM ensembles, pp. 658–662. Vol. 50 of Advances in Soft Computing, Springer, 2009.
[10] H. Millar and M. Kiragu. Cyclic and non-cyclic scheduling of 12 h shift nurses by network programming. *European Journal of Operational Research*, 104, 582–592, 1998.
[11] C. Pinzón, A. Herrero, J. F. De Paz, E. Corchado and J. Bajo. CBRid4SQL: a CBR intrusion detector for SQL injection attacks, pp. 510–519. Vol. 6077 of LNAI, Springer, 2010.
[12] D. Simić. Nursing logistics activities in massive services. *Journal of Medical Informatics & Technologies*, 18, 77–84, 2011.
[13] D. Simić, V. Kurbalija and Z. Budimac. An application of case-based reasoning in multidimensional database architecture, pp. 66–75. Vol. 2737 of LNCS, Springer, 2003.
[14] D. Simić, V. Kurbalija, Z. Budimac and M. Ivanović. Case-based reasoning for financial prediction, pp. 839–841. Vol. 3533 of LNCS, Springer, Berlin Heidelberg, 2005.
A hybrid case-based reasoning approach to detecting the optimal solution in nurse scheduling problem

[15] D. Simić, D. Milutinović, S. Simić and V. Suknaja. Hybrid patient classification system in nursing logistics activities, pp 421–428. Vol. 6679 of LNAI, Springer, 2011.
[16] D. Simić and S. Simić An approach to efficient business intelligent system for financial prediction. Soft Computing, 11, 1185–1192, 2007.
[17] D. Simić, S. Simić, D. Milutinović and J. Dordević Challenges for nurse rostering problem and opportunities in hospital logistics. Journal of Medical Informatics & Technologies, 23, 2014.
[18] S. Simić, D. Simić, D. Milutinović, J. Dordević and S. D. Simić. A hybrid approach to detecting the best solution in nurse scheduling problem, pp. 710–721. Vol. 10334 of LNAI, Springer, 2017.
[19] S. Simić, D. Simić, D. Milutinović, J. Dordević and S. D. Simić. A fuzzy ordered weighted averaging approach to rerostering in nurse scheduling problem, pp. 79–88. Vol. 649 of Advances in Intelligent Systems and Computing, Springer, 2017.
[20] G. K. Smyth. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology, 3, Article 3, 2004.
[21] B. G. Tabacnick and L. S. Fidell. Using Multivariate Statistics, 6th edn. Pearson, 2013.
[22] M. Warner, B. J. Keller and S. H. Martel. Automated nurse scheduling. Journal of the Society for Health Systems, 2, 66–80, 1990.
[23] W. N. Zelman, M. J. McCue and N. D. Glick. Financial Management of Health Care Organizations: An Introduction to Fundamental Tools, Concepts and Applications. John Wiley & Sons, 2014.
[24] J. M. Corchado, J. Bajo and A. Abraham. GERAmI: improving the delivery of health care. IEEE Intelligent Systems, 3, 19–25, 2008.

Received 25 September 2017