A basic study of how to exchange work shifts using reinforcement learning on a constructive nurse scheduling system

Masato Nagayoshi
Niigata College of Nursing, 240 shinnan-cho
Joetsu, Niigata 943-0147, Japan
E-mail: nagayosi@niigata-cn.ac.jp, elderton@niigata-cn.ac.jp

Hisashi Tamaki
Kobe University, 1-1 Rokkodai-cho, Nada-ku,
Kobe, Hyogo 657-8501, Japan
E-mail:tamaki@al.cs.kobe-u.ac.jp

Abstract

In this paper, we propose a work revision method using reinforcement learning for a constructive nurse scheduling system. The constructive nurse scheduling system has the characteristic of having easy to understand shift schedule creation procedures and rules because the system does not use the evaluation value for the entire shift schedule. We have confirmed the possibility of improving the quality of the shift schedule by the proposed method.

Keywords: nurse scheduling, reinforcement learning, constructive search, work revision method

1. Introduction

Various studies have been conducted on the nurse scheduling problem[1], which is the creation of a shift schedule for nurses. However, for practical use, adjustments including various constraints and evaluation values are required, and the created shift schedule is often not practical as it is, so many head nurses still feel burdened by creating shift schedules[2].

In this paper, we propose a work revision method using reinforcement learning[3] for a constructive nurse scheduling system[4]. The constructive nurse scheduling system has the characteristic of having easy to understand shift schedule creation procedures and rules because the system does not use the evaluation value for the entire shift schedule. We have confirmed the possibility of improving the quality of the shift schedule by the proposed method.

2. Constructive nurse scheduling system

2.1. *Features*

The features of the constructive nurse scheduling system[4] are as follows.

1. The system creates a schedule for each day, starting from the first day.
2. The priority calculation can be extended to take into account detailed conditions.
3. It does not take into account the evaluation value for the entire shift schedule for a month.

2.2. *Work Revisions*

The constructive scheduling system considers only the basic constraints that would be required in a hospital with a large number of nurses, and the possibility exists that a feasible solution that does not satisfy the head nurse is obtained. For this reason, Kurashige et al.[4] describe the following two procedures for the actual modification.

(1) Manual exchange of work shifts.

A work shift of the nurse in the case that does not satisfy the head nurse is manually exchanged with a work shift of another nurse. In this case, it is important that the constraints are satisfied by the exchange. If an exchange is made that does not satisfy the constraints, a warning message is displayed.

(2) Change the shift schedule manually and create it automatically again.

A work shift of the nurse in the case that do not satisfy the head nurse is exchanged to other work shift as designated work shift, and the rescheduling is done. Of course, the next solution displayed is not necessarily a satisfactory solution, but the above procedure is repeated in a timely manner until a satisfactory solution is obtained.

Next, we propose a system that leans this exchange procedure using reinforcement learning.

3. Work Revision Method Using Reinforcement Learning

3.1. *Reinforcement learning*

In this section, we introduce Q-learning (QL)[5] which is one of the most popular RL methods. QL works by calculating the quality of a state-action combination, namely the Q-value, that gives the expected utility of performing a given action in a given state. By performing an action $a \in \boldsymbol{A}_{\mathrm{Q}}$, where $\boldsymbol{A}_{\mathrm{Q}} \subset \boldsymbol{A}$ is the set of available actions in QL and $\boldsymbol{A}$ is the action space of the RL agent, the agent can move from state to state. Each state provides the agent with a reward r. The goal of the agent is to maximize its total reward.

The Q-value is updated according to the following formula, when the agent is provided with the reward:

$$
\begin{aligned}
&Q\big(s(t-1), a(t-1)\big) \\
&\leftarrow Q\big(s(t-1), a(t-1)\big) + \alpha_{\mathrm{Q}}\{r(t-1) \\
&\quad + \gamma \max_{b \in \boldsymbol{A}_{\mathrm{Q}}} Q(s(t), b) - Q(s(t-1), a(t-1))\} \quad (2)
\end{aligned}
$$

where $Q\big(s(t-1), a(t-1)\big)$ is the Q-value for the state and the action at the time step $t-1$, $\alpha_{\mathrm{Q}} \in [0,1]$ is the learning rate of QL, $\gamma \in [0,1]$ is the discount factor.

The agent selects an action according to the stochastic policy $\pi(a|s)$, which is based on the Q-value. $\pi(a|s)$ specifies the probabilities of taking each action a in each state s. Boltzmann selection, which is one of the typical action selection methods, is used in this research. Therefore, the policy $\pi(a|s)$ is calculated as

$$\pi(a|s) = \frac{\exp\left(Q(s,a)/\tau\right)}{\sum_{b \in A_{\mathrm{Q}}} \exp\left(Q(s,b)/\tau\right)} \quad (3)$$

where τ is a positive parameter labeled temperature.

3.2. *Problem Setting for Reinforcement Learning*

The shift schedule created by the constructive nurse scheduling system, which is created in order from the first day, satisfies the shift constraints (such as the number of nurses required for each day). On the other hand, the shift schedule for the entire scheduling period (e.g., one month) is checked, there may be several cases in which the nurse constraints (e.g., such as the limited number of workdays) are not satisfied for each nurse.

Therefore, the number of violations *Vnw* of work shift *w* is calculated as the number of days exceeding *UTnw*, the upper limit of the number of assignments of work shift *w* to each nurse *n*, from the work schedule, and a revision is repeated according to the following:

$$min \ \textstyle\sum_{\mathrm{n}} \sum_{\mathrm{v}} \mathrm{V}_{\mathrm{nw}} \quad (3)$$

The following procedure is to be used for one revision.

(1) Select a work shift w_0 that is the source of the exchange (usually the one with the most violations).

(2) Determine the nurse n_0 with the highest number of violations in the shift w_0.

(3) If the shift w_0 is the night shift, the shift w_0 with the highest number of violations, whether it is the semi-night or the late-night shift, is designated as w_0 for the nurse n_0.

(4) If there is a work shift that is below the lower limit of the number of assignments for the nurse n_0, that work shift w_1 is designated as a destination of the exchange shift. If not, the day shift without the upper and lower limits of the number of assignments is used as the exchange.

(5) Determine the day d_0 with the highest priority among the days when the shift w_0 is exchanged to w_1 for nurse n_0.

(6) Deduce the group $g(j_0)$ in which the nurse n_0 is in charge of a job j_0, which is assigned as the shift w_0.

(7) Determine a nurse n_1 who belongs to group $g(j_0)$ and whose shift on the day d_0 is w_1. If there is more than one nurse, determine the nurse n_1 with the highest priority among the nurses when the shift w_1 is exchanged to w_0 on d_0.

(8) The nurses n_0 and n_1 are exchanged their shifts on the day d_0.

In case there is no corresponding nurses in any of the procedures, the exchange is not valid. In addition, it is also not valid to undo a previous exchange.

Here, minimizing the number of violations is considered to be a very difficult problem, because the number of possible modifications depends on which work shift is being exchanged.

In this study, we propose a work revision method to determine an appropriate exchange procedure using reinforcement learning.

3.3. *RL Agent*

QL is applied to the proposed method to learn an appropriate exchange procedure.

The state space of the RL agent consists of 4 dimensions: the previous exchange days (1 to 30), the total number of violations by all nurses for semi-night, late-night shift, and holiday: V_{nw} (w=1,2,3,4), to be a Markov decision process. The number of possible actions is 4, which is the exchange of semi-night, late-night, holiday, and night shift.

1 step is defined as 1 exchange including unsuccessful cases, 1episode is defined as the time when the shift schedule reaches the target state. Here, the target state is defined as the sum of violations for all nurses and shifts $\sum_n \sum_v V_{nw} = 0$, or when the situation does not improve even after an exchange. The positive reinforcement signal $r_t = 10$ (reward) is given only when the target state is reached and the reinforcement signal $r_t = 0$ at any other steps. At the start of each episode, the shift schedule will be in its initial state before the exchange.

4. Computational Experiments

4.1. *Nurse Scheduling Problem*

The proposed method is applied to a nurse scheduling problem similar to that of Kurashige et al.[4]. First, a three-shift system (day shift, semi-night shift, and late-night shift) is adopted, and the number of nurses is 23, including the head nurse. Furthermore, the number of positions is classified as 3 (head nurse, assistant head nurse, and general), the number of teams is 2 (A and B), and the skill level is 3 (experienced, mid-career, and new). The other constraints are as follows.

- Restrictions on the number of nurses for each shift:
  (1) Required number of day shift on weekdays is greater than or equal to 10.
  (2) Required number of day shift for weekends and holidays is 5.
  (3) Required number of late-night shift is 5.
  (4) Required number of semi-night shift is 5.

Table 1. Evaluation of shift pattern for 2 days.

shift on previous day	shift on the day			
	day	semi-night	late-night	holiday
day	15	1	13	11
semi-night	0	5	0	12
late-night	0	8	5	4
holiday	23	3	0	17

- Constraints on team and skill level for night shifts:
  (5) At least 1 nurse per team should be assigned to each of the semi-night shift and late-night shift.
  (6) At least 3 nurses per team for consecutive semi-night and late-night shifts.
  (7) No more than 2 new nurses may work on the night shifts.
- Restrictions on the position:
  (8) All the work of the head nurse is designated.
  (9) The assistant head nurse works fewer nights.
- Restrictions on shift pattern:
  (10) The interval between holidays is limited to 5 days.
  (11) No more than 4 consecutive days off.
  (12) No more than 2 consecutive days of late-night and semi-night shifts.
  (13) The number of consecutive night shifts is limited to 3 days.

Next, Table 1 shows the evaluation of shift patterns for 2 days with M = 2.

4.2. *RL Agent*

In the state space of the RL agent, the total number of violations is assumed to be [0, 2] and can take 3 states.

The computational experiments have been done with parameters as shown in Table 2. In addition, all initial Q-values are set at 5.0 as the optimistic initial values.

4.3. *Results*

The average of the numbers of steps required to reach the target state and the average of the total number of violations when the target state is reached were observed during learning over 20 simulations, as described in Figs. 1 and 2, respectively.

It can be seen from Figs. 1, 2 that, (1) the total number of violations that can be reached by successive exchanges is 2, (2) the number of violations 3 that can be considered a local solution is sometimes obtained in the early stages of learning, but after more than 100 episodes, the number

Table 2. Parameters for experiments

Parameter	Value
α_Q	0.1
γ	0.9
τ	0.1

of violations 2 that can be considered an optimal solution is obtained in 7 exchanges.

Thus, we confirmed that the proposed method can reduce the number of violations of constraints without using the evaluation value for the entire shift schedule. In addition, we found that 2 nurses had violations of excessive holiday 1 in the modified shift schedule. Since these violations are not concentrated in 1 nurse, it is difficult to think that they lead to a sense of unfairness among general nurses.

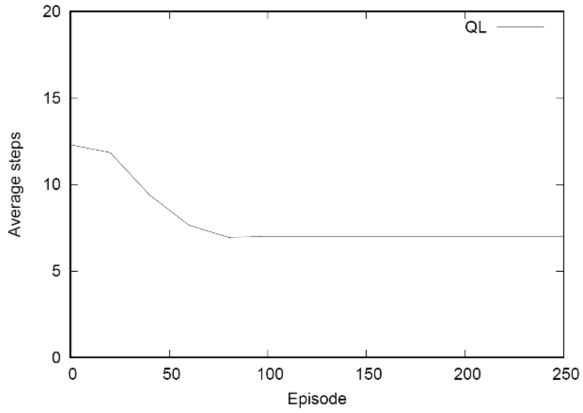

Fig. 1. Required steps to reach the target state.

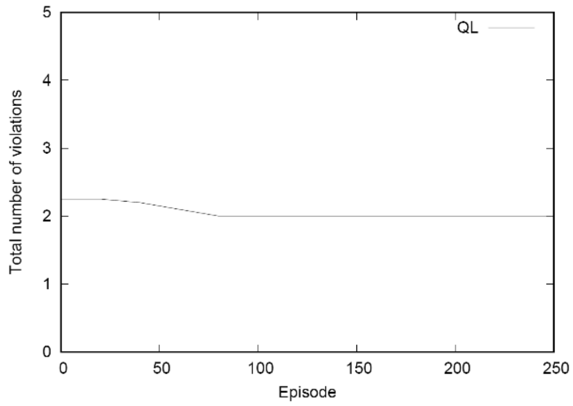

Fig. 2. Total nuber of violations when the target state is reached. Required steps to reach the target state.

5. Conclusion

In this paper, we proposed a work revision method using reinforcement learning for a constructive nurse scheduling system. Through computational experiments, we confirmed the possibility of improving the quality of the shift schedule by the proposed method.

Our future projects include to respond to sudden changes in shift schedule, and to clarify the rules for creating shift schedules, etc.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP19K04906.

References

1. A. Ikegami, A Model for the Nurse Scheduling Problem, IPSJ SIG Notes, **5**, 1-6, 1996. (in Japanese)
2. H. Adachi, S. Nakamura, M. Nagayoshi and N. Okamura, A Survey on The Present Status about Required Time and Recognition of Supports for Managers to Prepare A Work Timetable in The Medical Treatment and Supervision Act Ward, Journal of Japan Academy of Psychiatric and Mental Health Nursing, **30** (1), 59-65, 2021. (in Japanese)
3. R. S. Sutton and A. G. Barto, Reinforcement Learning, A Bradford Book, MIT Press, Cambridge, 1998.
4. K. Kurashige, T. Hashimoto and Y. Kameyama, Nurse Scheduling System in Consideration of Versatility, Journal of Japan Industrial Management Association, **56**(2), 109-120, 2005.
5. C. J. C. H. Watkins and P. Dayan, Technical note: Q-Learning, *Machine Learning* **8**, 279-292, 1992.

Authors Introduction

Dr. Masato Nagayoshi

He is an Associate Professor of Niigata College of Nursing. He graduated from Kobe University in 2002, and received Master of Engineering from Kobe University in 2004 and Doctor of Engineering from Kobe University in 2007.
IEEJ, SICE, ISCIE member.

Dr. Hisashi Tamaki

He is a Professor, Graduate School of Engineering, Kobe University. He graduated from Kyoto university in 1985, and received Master of Engineering from Kyoto University in 1987 and Doctor of Engineering from Kyoto University in 1993. ISCIE, IEEJ, SICE, ISIJ member.