A Hybrid Adaptive Metaheuristic-based Q-Learning Approach for the Home Healthcare Routing and Scheduling Problem

Yun-Chia Liang[1], Angela Hsiang-Ling Chen[2], Sebastian Gunawan[2], and Vanny Minanda[1]

[1] Department of Industrial Engineering and Management, Yuan Ze University, Taoyuan City, Taiwan ROC, 320

[2] Department of Industrial and Systems Engineering, Chung Yuan Christian University, Taoyuan City, Taiwan ROC, 320

ycliang@saturn.yzu.edu.tw

Abstract. This study addresses the Home Healthcare Routing and Scheduling Problem (HHCRSP), a complex extension of the Vehicle Routing Problem involving joint nurse–patient assignment, service scheduling, and route planning under synchronized multi-nurse constraints and strict time windows. To tackle the highly constrained search space, we propose a Hybrid Adaptive Metaheuristic-based Q-Learning (HAMQL). The approach utilizes a survival-selection mechanism to maintain population diversity while leveraging Q-learning as an adaptive memory scheme to guide local search operations. The proposed algorithm was evaluated on benchmark instances of varying scales (up to 100 patients) representing high-density routing scenarios. The results demonstrate promising scalability and high constraint-fulfillment ratios in medium- and large-scale instances, approaching MILP objective function values (or bounds) with faster CPU Time. Finally, we discuss computational trade-offs, particularly regarding convergence rates in large-scale scenarios, and outline paths for future algorithmic refinement.

Keywords: Home Healthcare Routing and Scheduling Problem, Metaheuristic, Q-learning.

1 Introduction

Taiwan's rapid population aging creates intense pressure on healthcare delivery, particularly for rural elderly patients who face severe geographic and infrastructural barriers to conventional medical facilities [1-2]. While home healthcare has emerged as a vital model to improve patient quality of life and optimize resource efficiency, managing it requires solving the Home Healthcare Routing and Scheduling Problem (HHCRSP) [3-4]. As an advanced extension of the Vehicle Routing Problem with Time Windows (VRPTW), HHCRSP introduces highly complex, healthcare-specific constraints. These include strict synchronized multi-caregiver visits (where certain treatments require multiple nurses to arrive simultaneously), localized time windows, and strict caregiver-qualification matching [5].

The combinatorial complexity of these overlapping operational constraints heavily restricts the feasible solution space, causing some state-of-the-art metaheuristics to frequently stall in local optima. To overcome this limitation, this study develops a sophisticated hybrid framework that couples a population-based metaheuristic with Reinforcement Learning (RL). While RL has seen extensive success in sequential decision-making tasks like robotics, its integration into metaheuristic frameworks for logistical optimization remains underutilized [6-7].

The fundamental motivation for developing this hybrid approach is two-fold: (1) Algorithmic Synergy: Population-based selection mechanisms provide robust global exploration capabilities across the discrete solution space, while the adaptive learning memory of Q-learning dynamically tunes neighborhood selection parameters based on search history, mitigating premature convergence; (2) Practical Utility: From a managerial perspective, healthcare providers require an automated scheduling mechanism capable of synchronizing caregiver timetables, minimizing unproductive travel/idle times, and strictly eliminating service-window violations to maximize patient safety and satisfaction. The primary aim of this study is to validate whether embedding adaptive learning memory into a metaheuristic framework can reliably solve high-density, real-world scale routing and scheduling instances where operational constraints are tight.

2 Methodology

This study proposes a Hybrid Adaptive Metaheuristic-based Q-Learning (HAMQL) to solve the highly constrained HHCRSP. Instead of relying on static heuristic parameters, the framework embeds a central Q-table, $Q(d, i, j)$, to act as an adaptive spatial-temporal memory layer that stores the learned expected utility of moving from location i to location j on day d. Over successive iterations, the algorithm maintains a candidate schedule population V, constructing solutions sequentially via an adaptive state transition rule where traditional heuristic visibility is dynamically weighted against historical Q-values. At the end of each iteration, a rank-based survival selection routine filters the population into elite (V_{elite}) and feasible $(V_{feasible})$ tiers based on the objective function of maximizing patient visit fulfillment. High-quality routing trajectories from these surviving schedules are then backpropagated into the central Q-table using temporal-difference learning. This iterative feedback loop structurally guides subsequent iterations toward high-yield routing choices while preserving enough population diversity to prevent the search from stalling in local optima, as formalized in Algorithm 1.

Algorithm 1. Hybrid Adaptive Metaheuristic-based Q-Learning for HHCRSP

1. Initialize $Q(d, i, j) \leftarrow 0 \ \forall d \in D, i, j \in P \cup \{0\}$ and set $K^* \leftarrow \emptyset$
2. For iteration $t = 1$ to T do
3. Compute current exploration rate: $\varepsilon_t = \varepsilon_{start} - \left(\frac{t-1}{T-1}\right) \times (\varepsilon_{start} - \varepsilon_{end})$
4. Initialize an empty candidate solution population $V \leftarrow \emptyset$

5. For candidate index $k = 1$ to M do
6. Initialize schedule $K_k \leftarrow \emptyset$, visit counters, and caregiver spatial-temporal states
7. For each day $d \in D$ do
8. While feasible service decisions remain do
9. Select lead nurse $n \in N$ and define $S = (d, i)$
10. Construct feasible set Ω_k based on demand, qualification, and time window bounds
11. If $\Omega_k = \emptyset$ then break or transition to alternative nurse
12. Sample random variable $r \sim U(0,1)$
13. If $r < \epsilon_t$ then select next patient $j \in \Omega_k$ uniformly at random
14. Else select patient $j \in \Omega_k$ using probability: $P_{ij}(d) = \frac{[Q(d,i,j)]^{\alpha} \cdot \left[\frac{1}{c_{ij}}\right]^{\beta}}{\sum_{m \in \Omega_k} [Q(d,i,j)]^{\alpha} \cdot \left[\frac{1}{c_{ij}}\right]^{\beta}}$
15. Update nurse routing state, synchronize multi-nurse teams if required, and track rewards
16. Record transition tuple (S, j, R, S') into schedule trajectory $\mathcal{T}_k$
17. End While
18. End For
19. Evaluate objective function $F(K_k) = \sum_{p \in P} Y_p$ and append K_k to population V
20. End For
21. Sort population V in descending order of objective value and structural feasibility
22. Partition into elite population V_{elite} ($M_{elite} = [\rho_{elite} M]$) and feasible population $V_{feasible}\left(M_{feasible} = \left[\rho_{feasible} M\right]\right)$
23. Update historical best solution: If $\max(F(K_k)) > F(K^*)$ then $K^* \leftarrow \arg\max(F(K_k))$
24. For each population tier $C \in \{V_{elite}, V_{feasible}\}$ do
25. Assign scaling coefficient ω ($\omega_{elite} > \omega_{feasible}$)
26. For each schedule $K \in C$ and recorded transition $(S, j, R, S') \in \mathcal{T}$ do
27. $Q(d,i,j) \leftarrow Q(d,i,j) + \gamma \cdot \omega \cdot \left[R + \lambda \max_{k \in \Omega'} Q(d,j,k) - Q(d,i,j)\right]$
28. End All for loop
29. Return K^*

3 Results and Discussion

To evaluate the performance and scalability of the proposed HAMQL, computational experiments were conducted across various patient dimensions. The algorithmic performance is validated by analyzing both objective function and CPU Time against an exact MILP solver. The comparative computational results are summarized in Table 1.

Computational experiments across patient dimensions ($|P| \in \{30, 35, 40, 50, 100\}$) validate the performance of the HAMQL algorithm against an exact MILP solver. As summarized in Table 1, HAMQL scales effectively, maintaining high objective function values with stable constraint-fulfillment ratios between 78% and 85% on medium-to-large instances ($|P| \geq 35$). Conversely, while the MILP solver finds the optimal schedule for the small instance ($|P| = 30$) in 64.95 seconds, it experiences combinatorial explosion in larger dimensions due to strict multi-nurse synchronization and time-window constraints, completely stalling at the 36,000-second execution limit. Crucially, the higher CPU runtimes reported for HAMQL represent the algorithm's offline training phase, where exploration overhead is required to mature the dynamic Q-table. Because routing strategies are completely learned during this phase, real-world deployment on operational testing instances is expected to be nearly instantaneous, requiring only localized exploitation of the pre-trained matrix.

Table 1. HAMQL and MILP Results

Dimension	**MILP**		**HAMQL**		
	Optimal Obj. Value	CPU Time	Best Obj. Value Found	Training CPU Time (s)	Fulfillment Ratio (%)
30	17	64.95	18	2,502.3	60.00
35	23*	36,000	29	2,845.0	82.86
40	34*	36,000	34	4,052.3	85.00
50	42	124.04	39	4,906.0	78.00
100	89*	36,000	83	13,992.4	83.00

*We set 36,000 sec. as the stopping criterion for MILP.

4 Conclusion

This study successfully developed an HAMQL that integrates a rank-based survival selection routine with temporal-difference Q-learning to solve the highly constrained HHCRSP. By replacing static heuristic choices with an adaptive spatial-temporal memory layer, the framework effectively guides solution construction through tightly constrained routing spaces. The experimental results validate the practical viability of this approach, demonstrating high and stable patient fulfillment ratios on large-scale instances where exact mathematical solvers experience combinatorial explosion and fail to converge.

While the validation of overlapping operational constraints causes a non-linear CPU runtime expansion during the algorithm's offline training phase, this computational investment builds a robust decision matrix. Future work will focus on transitioning this trained framework from the experimental training phase into real-world operational deployment using live datasets, where schedule generation is expected to be executionally rapid during the testing and inference stage.

Acknowledgement

This research was partially funded by National Science and Technology Council (Taiwan): NSTC 113-2221-E-155-044-MY3.

References

1. Wu, T.-Y., Majeed, A., Kuo, K. N.: An Overview of The Healthcare System in

Taiwan, London J. Prim. Care (Abingdon), 3(2), 115–119 (2010).
2. Cheng, B., Chang, H., Lin, M., Chen, T., Chou, L., Hwang, S.: Rural-Urban Disparities in Family Physician Practice Patterns: A Nationwide Survey in Taiwan, Int. J. Health Plann. Manage., 34(1), e464–e473 (2019).
3. Lin, Y.-Y., Huang, C.-S.: Aging in Taiwan: Building a Society for Active Aging and Aging in Place, Gerontologist, 56(2), 176–183 (2016).
4. Lin, M., Ma, L., Ying, C.: Matching Daily Home Health-Care Demands with Supply in Service-Sharing Platforms, Transp. Res. E Logist. Transp. Rev., 135, 102177 (2021).
5. Bazirha, M., Kadrani, A., Benmansour, R.: Stochastic Home Health Care Routing and Scheduling Problem with Multiple Synchronized Services, Ann. Oper. Res., 320(2), 573–601 (2023).
6. Zhu, M., Wang, Y., Pu, Z., Hu, J., Wang, X., Ke, R.: Safe, Efficient, and Comfortable Velocity Control Based on Reinforcement Learning for Autonomous Driving, Transp. Res. Part C Emerg. Technol., 117, 102662 (2020).
7. Qiu, H., Wang, D., Yin, Y., Cheng, T. C. E., Wang, Y.: An Exact Solution Method for Home Health Care Scheduling with Synchronized Services, Naval Research Logistics, 69(5), 715–733 (2022).
8. Liang, Y.-C., Cuevas Juarez, J. R.: A Novel Metaheuristic for Continuous Optimization Problems: Virus Optimization Algorithm, Engineering Optimization, 48(1), 73–93 (2016).