Computers & Industrial Engineering 182 (2023) 109332

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Home health care routing and scheduling in densely populated communities considering complex human behaviours

Ting Zhang [a], Yang Liu [b,c,*], Xintong Yang [d,*], Jingjing Chen [a], Jiaming Huang [a]

[a] College of Urban Transportation and Logistics, Shenzhen Technology University, Shenzhen, 518118, China
[b] Department of Management and Engineering, Linköping University, SE-581 83 Linköping, Sweden
[c] Industrial Engineering and Management, University of Oulu, 90570 Oulu, Finland
[d] School of Engineering, Cardiff University, Cardiff CF24 3AA, UK

ARTICLE INFO

Keywords:
Home health care
Vehicle routing problem
Stochastic service time
Mixed time windows
Q-learning
Ant-colony optimisation

ABSTRACT

This study focuses on the home health care routing problem (HHCRP) in the scenario of high population density areas where many elders live closely together. This study considers two main objectives. The first is to reduce travel and wait times for nurses or elders. The second concerns socially related objectives in scheduling problems, such as 'quality of life' and empowerment, by considering assumptions related to the acquaintanceship and mutual preferences of nurses and elders. This study models the effects of mutual preferences and acquaintanceship on service time in HHCRP. We use the Markov decision process and chance-constrained programming (CCP) to model the system to conserve the sequential service provision parameters and better represent the influence of stochastic service times. Because traditional deterministic algorithms cannot solve such a model, we apply a model-free reinforcement learning algorithm, Q-learning (QL), as well as the ant colony optimisation (ACO) algorithm. Thus, we tackle this problem by developing a model and algorithm to solve complex, large-scale systems. This study's theoretical and practical contributions are verified by feedback from researchers and practitioners.

Abbreviations

Abbreviations

ACO	ant colony optimisation
BWACO	best-worst ant colony optimisation
CCO	chance constraint of overworking
CCP	chance-constrained programming
CCWT	chance constraint of wait time
HHC	home health care
HHCRP	home health care routing problem
HTW	hard time window
MDP	Markov decision process
MTW	mixed time window
QL	Q-learning
QL-BWACO	Q-learning–best-worst ant colony optimisation
SC	set covering
SHHCRP	stochastic home health care routing problem
STW	soft time window
SVRP	stochastic vehicle routing problem
TSP	travelling salesman problem
VRP	vehicle routing problem

1. Introduction

The problem of societal ageing is becoming increasingly serious, making the home health care routing problem (HHCRP) a critical issue (Lin et al., 2021; Shi et al., 2019). In this study, we focus on high population density countries, specifically on large-scale living quarters in major cities like Beijing, Shanghai, and Guangzhou in China. In these areas, many people live in communities with high population densities, with a density greater than 1,000 persons/km^2 (Statista, 2023). However, different ages gather in different residential areas. Young people move to the suburbs due to the high living cost of the central urban area, while older people remain in the old communities in the central urban area. The city's suburbs attract young talents by developing public transportation, while the central urban area serves the elders by completing medical facilities.

* Corresponding authors at: Department of Management and Engineering, Linköping University, SE-581 83 Linköping, Sweden (Y. Liu). School of Engineering, Cardiff University, Cardiff CF24 3AA, UK (X. Yang).
E-mail addresses: yang.liu@liu.se (Y. Liu), yangx66@cardiff.ac.uk (X. Yang).

https://doi.org/10.1016/j.cie.2023.109332
Received 4 July 2022; Received in revised form 16 May 2023; Accepted 19 May 2023
Available online 1 June 2023

In this context, elders are often clustered in communities, allowing nurses to visit multiple patients in a shorter time (Lenzi et al., 2019). However, each elder may have personalized demands for the duration and schedule of home health care (HHC) services (Low et al., 2011). To address these challenges, we aim to improve the quality and efficiency of HHC service delivery in high-population areas by carefully designing the routing and scheduling of these services. Our goal is to reduce waiting times, workloads, and travel time for nurses while meeting the needs of each elder.

Our research has important implications for addressing the HHCRP in other densely populated areas and the challenges of societal ageing in a rapidly changing world. By proposing a solution for efficient and personalized routing and scheduling of HHC services, we hope to improve the quality of care for elders and reduce the burden on healthcare providers in high population density areas.

One key aspect of our proposed solution is the consideration of human relationships. We recognize that service time can be reduced as nurses become more familiar with elderly patients and their needs, and that factors such as acquaintanceship and individual preferences can significantly impact the quality of life of HHC participants (Lin et al., 2021; Shi et al., 2019; Zhong et al., 2020). Therefore, our study explicitly models the effects of mutual preferences and acquaintanceship on service time in HHCRP, in order to optimize care delivery and enhance the overall quality of services.

In addition to addressing multi-appointment scenarios and skill-demand matching, our proposed solution also takes into account the diverse skill levels of healthcare providers and the specific needs of elderly patients. The selection of caregiver schedules can have a significant impact on both the quality of services and the happiness of elderly patients (Choi, 2020; Lin et al., 2021; Shi et al., 2019; Yang et al., 2018).

To model the system, we use the Markov decision process and chance-constrained programming (CCP) to conserve sequential service provision parameters and better represent the influence of stochastic service times. However, these techniques cannot be solved by traditional deterministic algorithms, especially with a large number of instances, and are computationally expensive to apply using previously published meta-heuristic algorithms (Allahviranloo et al., 2014). Therefore, we utilize a model-free reinforcement learning algorithm, Q-learning (QL), as well as the ant colony optimization (ACO) algorithm, to optimize the proposed solution and address these challenges.

The main contributions of this study are as follows:

- We consider acquaintanceships and individual preferences to improve the sense of satisfaction for participants in HHCRP.
- We consider multi-appointment scenarios and skill–demand matching to increase the empowerment of participants.
- To obtain more realistic results, we consider stochastic service time to simulate the real-world situation in which uncertainty is inevitable.
- We develop a model and an algorithm that can solve complex, large-scale systems with a shorter computational time than in previously used problem-solving algorithms.

The rest of this paper is organised as follows. Section 2 reviews the literature. Section 3 describes and discusses the complexity of the problem. Section 4 proposes a mathematical model for HHCRP, while Section 5 introduces the QL-BWACO-based heuristic algorithm. Section 6 presents the results and analyses of the computational experiments and discusses feedback collected from practitioners and researchers. Section 7 concludes and discusses the practical implications, limitations, and future research.

2. Literature review

Research on HHC has developed over the past two decades. In a detailed review by Fikar & Hirsch (2017), the authors include multiple constraints in their mathematical formulations to simulate the real-world environment.

2.1. Number of appointments per patient per day

This challenge is classified as a single-appointment problem or a multi-appointment problem. If elders require a single appointment per day, it can be considered as a classic travelling salesman problem (TSP) or vehicle routing problem (VRP) (Fikar & Hirsch, 2017). Nickel et al. (2012) describe the acquaintanceship between workers and elders as elder–nurse loyalty. Shi et al. (2019) and Redjem & Marcon (2016) also consider multi-appointments, but they do not consider the effect of acquaintanceship between workers and elders. In our study, therefore, we consider both multi-appointments and acquaintanceships between workers and elders.

2.2. Stochastic service time

Yuan et al. (2015) and Choi (2020) consider stochastic service time in HHCRP and assume that it is normally distributed, optimising it using a branch-and-price algorithm for small-scale experimental cases. Liu et al. (2019) and Shi et al. (2018) assume that service duration is a random variable with known probability distributions while Shi et al. (2019) describe an uncertainty set for service duration. Shi et al. (2018) investigate stochastic HHCRP (SHHCRP) with stochastic travel and service times by applying stochastic programming with resources and reducing it to a deterministic version. Oyola et al. (2016), meanwhile, designate a VRP that considers stochastic service time, demand, or travel time as a stochastic VRP (SVRP) and solve it using CCP and stochastic programming, both of which are modelling methods within stochastic programming theory (Lin et al., 2021). Restrepo et al. (2020) present a two-stage stochastic programming model for employee staffing and scheduling in home healthcare. Du & Zhang (2022) investigate a cross-regional scheduling and routing problem with stochastic service times in the one-day planning horizon. Essentially, SHHCRP is a variant of SVRP. Likewise, the present study uses stochastic programming.

2.3. Time window

In real situations, elders require services and appointments during different periods; these are described mathematically as time windows (Fikar & Hirsch, 2017). They consist of three classes: soft time window (STW), hard time window (HTW), and mixed time window (MTW). Fikar & Hirsch (2015) and Fathollahi-Fard et al. (2020) assume that every service must begin within a period (i.e. HTW), and arrivals outside that time window are not allowed. STW allows nurses to arrive outside the appointment period (Eveborn et al., 2006), which is less realistic since punctuality is considered important in real life. Bertels & Fahle (2006) therefore used both HTW and STW.

HTW prevents waiting or being late, STW allows waiting and being late, MTW penalises waiting but prevents being late, and HTW & STW penalise waiting and being late. The arrows in each sub-plot indicate the arrival timeslots, each of which is associated with an effect. For example, in the HTW case, arrivals before and after the time window are unacceptable, while in the STW case, they are all acceptable but come with either waiting time costs or penalties for being late.

MTW keeps health care providers' visits from occurring after the lower time boundary and results in waiting times when workers arrive before the upper boundary (Yang et al., 2018; Zhang et al., 2018; Zhang et al., 2019). Fig. 1 is a graphical representation of the different time windows. This study uses MTW because it has been shown to be the most reliable approach (Liu et al., 2014).

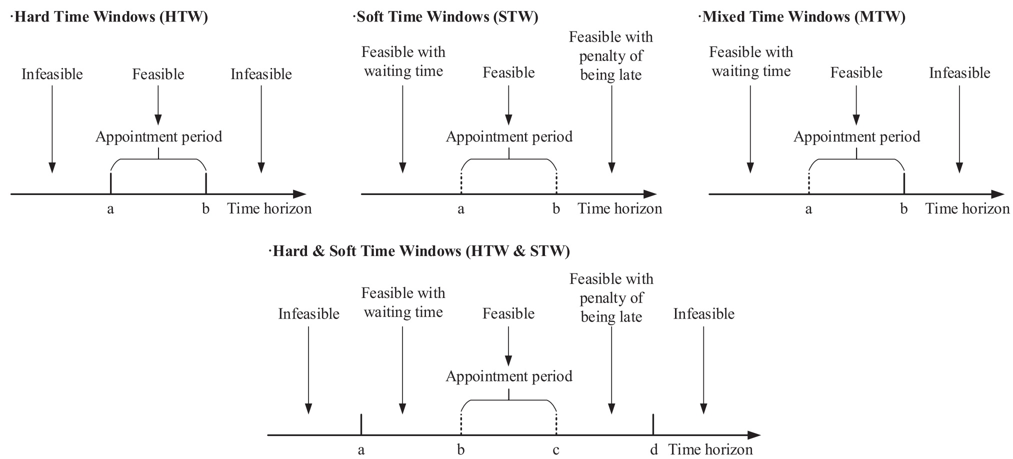

Fig. 1. Graphical representation of different time windows (arrows indicate nurses' arrival slots).

2.4. Skill–demand and preference matching

Skill–demand matching generally represents the availability of elders (or services) for workers with different skills (Fikar & Hirsch, 2017). There are different ways to describe it mathematically, one of which is called 'set covering' (SC), which assumes that only when the skill set of a worker can cover the demand set of elders can the worker serve the elder. Yuan et al. (2015) address the hierarchical assumption that proposes two parameters: 'dummy capacity' for every worker and 'dummy demand' for the elders. Rodriguez et al. (2015) propose an original two-stage approach based on integer linear stochastic programming. Fikar & Hirsch (2015) represent this assumption more directly by assigning each nurse or job a qualification level. Yang et al. (2018) and Zhang et al. (2019), as well as the present study, use the latter representation. A number of studies have investigated matching care demand with supply in the context of traditional institutions (Lin et al., 2021; Shi et al., 2019). Preference matching includes a wide range of personal tendencies (e.g. an elder may need a worker who speaks a certain language). Two recent studies consider a related assumption, 'patient satisfaction', aiming to maximise the privilege of patients to choose specific nurses (Fathollahi-Fard et al., 2020; Lin et al., 2021). Khodabandeh et al. (2021) develop a model to consider the objective of minimizing the difference between the actual and potential skills of the nurses.

2.5. Other constraints

Yang et al. (2018), Zhang et al. (2018), and Zhang et al. (2019) evaluate several simpler SHHCRP scenarios, considering stochastic service time, mixed time window, and skill–demand match. Yang et al. (2021) consider a multi-objective home healthcare routing and scheduling problem (HHRSP) with several conflicting objectives: minimizing routing cost and improving service consistency and workload balance. Mehmet and Çağrı (2022) study a joint multi-depot home health care and dialysis problem of routing and scheduling decisions of health specialists.

Although HHCRP has been increasingly investigated, only a few researchers have studied the stochastic characteristics, and none of them have considered the stochastic scenario in densely populated communities. The present study, therefore, simultaneously considers multi-appointment, mixed time window, skill–demand match, and human relationships (preferences and acquaintanceships) to better reflect real-world scenarios.

3. Problem description

SHHCRP is commonly treated as a variant of VRP, which can be defined on a graph $G = (V, E)$ (Li et al., 2010). The set of nodes, V, contains locations of elders (coordinates). The set of edges, E, contains edges linking each pair of nodes in V. Since workers can only travel to elders via paths, stairs, or elevators (either horizontal or vertical), the shortest distance between any two nodes is assumed to be a constant sum of horizontal and vertical distances traversed by workers. The overall objective of SHHCRP is to allocate demand to nurses to minimise their total waiting time and unfulfilled demands. Fig. 2 shows a graphical illustration of the problem and its assumptions.

Each worker starts from the depot and finally comes back to the depot. Each worker needs to arrive before the end of the time window and wait if he or she arrives before the time window. A worker can only execute a demand when the skill level is higher than the demand level. The more skilled the worker is, the more acquainted the worker is with the elder, and the less the required service time. Each elder may propose multiple demands. Therefore, a worker may visit an elder multiple times.

3.1. Assumptions

More specifically, the problem investigated in this study contains the following assumptions:

- **Single depot.** Each worker must start from and end at the HHC centre (i.e. all routes must start and end at the same node).
- **Multi-appointment.** Each elder may require more than one service per day with different time windows.
- **Acquaintanceship.** Once a worker is assigned to an elder who requires more than one service, the worker is expected to provide the remaining services to the elder unless the worker is not qualified or unable to meet other constraints.
- **Mixed time window.** Each service has an appointed period that forces workers to wait before the lower bound and forbids workers to begin service when arriving after the upper bound.
- **Workload.** Each worker is expected to finish the route within a time horizon.
- **Moving speed.** Each worker moves among elders at a constant speed; thus, the travel cost becomes a value only related to the distance.
- **Skill–demand matching.** Each worker is ranked at a skill qualification level, and each service has a level of required qualifications. It is assumed that the worker can serve an elder only when a service's qualification level is lower than a worker's qualification level. Additionally, the higher the worker's skill, the less the required service time.

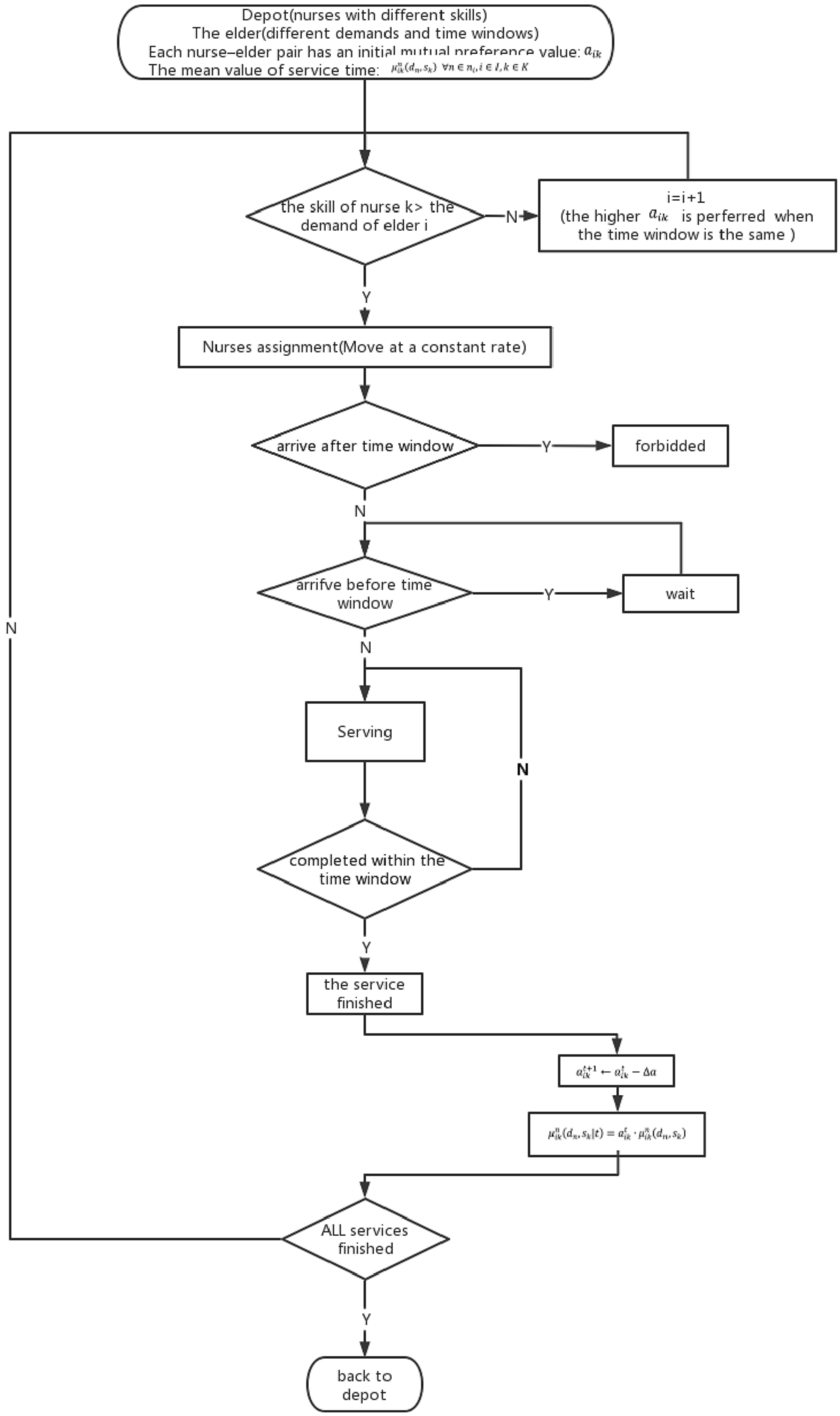

Fig. 2. Problem description.

- **Preference matching.** Each nurse–elder pair has an initial mutual preference value representing how much they mutually prefer each other.
- **Stochastic service time.** Service time is usually distributed and linked to three matching rules: (i) skill–demand match, which is negatively correlated with the mean value (i.e. the better the skills match the service, the less the expected time to fulfil service needs; (ii) preference-match, which is defined as a weight of the mean value (i.e. the better the match, the shorter the expected service time); (iii) acquaintanceship, which is defined as an increment of the preference weight (i.e. as a nurse–elder pair becomes more familiar with each other, the preference weight becomes smaller).

3.2. Symbol definition in the model

3.2.1. Indexes and sets of workers, elders and routes

As shown in the table below, briefly define the parameters that appear in the problem description.

Symbol	Definition
K	the set of workers, which is given by the experimental instances.
I	the set of elders, which is given by the experimental instances
V	the set of coordinates of elders defined as $V = \{(x_i, y_i, z_i) \mid i \in I\}$
N	the set of jobs, which is given by the experimental instances.
R	the set of routes defined as $R = \{R_K \mid R_K = \{r0, r1, r2, \ldots, r_{n_k}, r0,\} \forall k \in K$, and $r_n \in n_i\}$, where r_n is the n-th job of elder i executed in route R_K. Besides, r_0 is the depot where workers start and finish routing
n_i	the set of job that elder i needs, $n_i \in N$. In this paper, it is assumed that an elder may need more than one service or job, therefore $n_i \cap N \neq \{0\} \ \forall i \neq 0$ and $n_i \neq \varnothing$. Note that when and only when $i = 0, n_i = \{0\}$
k	the index of workers defined as $k \in K = \{1,2,3, \ldots k\}$.
i	the index of elders defined as $i \in I = \{0,1,2,3, \ldots i\}$. Note that $i = 0$ represents the depot, which is the start spot and end spot of every worker.
n	the index of jobs defined as $n \in N = \{0,1,2, \ldots, n\}$. Note that $n = 0$ represents the job of $i = 0$.

3.2.2. Skills and preferences match qualities

This article proposes nursing issues based on interpersonal relationships, specifically manifested in the familiarity between nurses and elders. Based on this, both parties make a two-way choice. The following table provides a brief definition of the parameters appearing in the article for this issue.

Symbol	Definition
SQ	the set of skills qualifications. In this paper, we consider three sorts of skills: 1) daily nursing skills, 2) basic medical skills and 3) professional medical skills. They are represented by d_n, b_m, and pm, respectively. Therefore, $SQ = \{d_n, b_m, p_m\}$.
sq_k	the set of skills qualifications that worker k masters, $sq_k \in SQ$
d_n	the set of skills and qualifications required by the job n, $d_n \in S$, $n \in N$.
a_{ik}	a binary variable that represents whether elder i is acquainted with worker k. And it takes the value 0 when they don't know each other, and 1 when they have met each other before. The value is set to 0 for any pair of workers and elders at the beginning of the algorithm and changed to 1 if worker k serves elder i.
Δa	the acquaintanceship decrement

3.2.3. Time-related notations

This article uses a hard time window constraint to define the parameters in the text. At the same time, we link time and cost together and define them.

Symbol	Definition
$(b_i^n, e_i^n]$	the time window of job n of elder i. In particular, the time window of depot, $(b_0^0, e_0^0]$ is defined as the scheduling horizon. We define that $(b_0^0, e_0^0] \rightarrow (0, e_0^0]$, in which e_0^0 is a constant given by experimental instance.
tc_{ij}	the travel cost from elders i to j
dc_{ik}^n	the service time of executing job n of elder i by worker k, which is defined as normally distributed, namely, $dc_{ik}^n \sim N(\mu_{ik}^n (d_n, s_k), \sigma^2)$. As assumed, the mean value, μ_{ik}^n (d_n, s_k) is a function of skill-match-quality.
w_{ik}^n	the waiting time produced when worker k arrives at elder i for job n before b_i^n.
wh_k	the total working hours of worker k
W	the maximum of working hours available for any worker
θ	the cost per time units. In this paper, we consider three sorts of costs: 1) travel cost, 2) service cost and 3) waiting cost. To different extents for different operational purposes, they are differently important, which can be formulated by setting different values of θ.

3.2.4. Decision variables

After the state of a stage is given, a choice that evolves from that state to a certain state in the next stage is called a decision, and the variables that describe the decision are called decision variables.

Symbol	Definition
x_{ijk}^{nm}	a binary variable that takes the value 1 when worker k moves from elder i finishing job n, to elder j for available job m, and 0 otherwise.

4. Mathematical model

Stochasticity requires stochastic programming as a proper representation (Kall et al., 1994). Therefore, we use finite Markov decision process (MDP) and chance-constrained programming (CCP) to hierarchically model the problem, as MDP models sequentiality (Bellman, 1966), and CCP reasonably simulates stochasticity (Charnes & Cooper, 1959). We model the scheduling problem using MDPs and CCP in a nested manner, as depicted in Fig. 3. The MDPs model the outer loop of selecting nurses to be scheduled to capture sequentiality. In contrast, the CCP models the inner loop of allocating elders and their demands to capture stochasticity. At each stage of MDP, an agent (the algorithm) observes the current number of unscheduled nurses and demands and acts (choosing a nurse). Then, a sub-solution (route) restricted to the constraints of CCP is used to calculate a reward regarding routing quality. Then, the system moves to a subsequent state. A complete worksheet is formed when the system transits to an absorbing state (no more nurses or demands).

4.1. Markov decision process

MDP is represented by a tuple $(S, A, \{P_{ss'}\}, \gamma, R)$ (Sutton & Barto, 2018). Note that SHHCRP is assumed to hold the Markov property, which means that as long as the current unfulfilled demands are observed and a nurse is chosen, the solution generated based on CCP is only relevant to the current unfulfilled demands and the chosen nurse and is not influenced by historical system states (Bellman, 1966).

The elements of MDP are formally defined for SHHCRP as follows:

$s \in S = \left\{sr_s^{dd'} \mid d, d' \in D\right\} \cup \left\{s_{absorb}\right\}$ represents the state space, where $d, d' \in D$ represent the levels of demands; $sr_s^{dd'}$ represents the size relationships of demands at levels d and d'; and s_{absorb} represents the absorbing states, which are situations where there are no more nurses or demands.

$a_s = sq_k \in SQ$ represents an action chosen at a state s, which is a skill level of available nurses.

$P_{ss'}(s, a_s)$ is the transition probability of the system transferring to state s' when an action a_s is taken at state s. This is also called the environmental model since, by definition, it describes the dynamics of the environment (Sutton & Barto, 2018). In this study, the environment is modelled by CCP. We assume, however, that the transition probability distribution is unknown since SHHCRP is a combinatorial optimisation problem that suffers from NP-hardness. In principle, transition probability can be computed by solving CCP, but the computational processes are very expensive, especially for problems with a large search space. Therefore, assuming transition probability is unknown and applying a model-free algorithm can be computationally cheaper and more

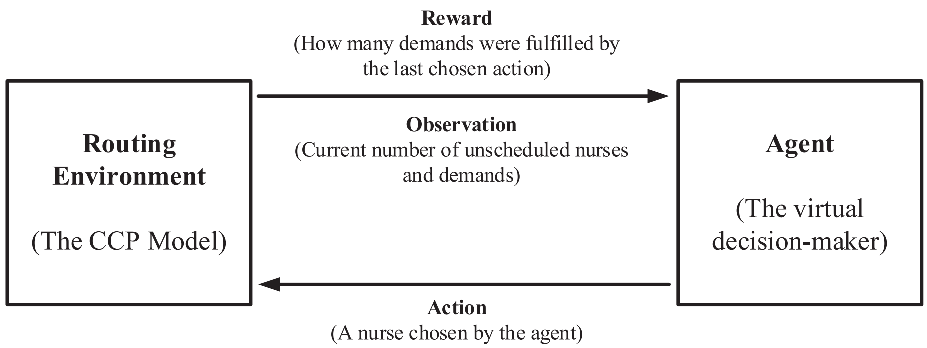

Fig. 3. Finite Markov decision process. At every decision timestep, an agent (the algorithm) observes the number of nurses and demands and selects an action (a nurse in this study). The selected nurse is then assigned a service plan by solving a chance-constrained programming problem. The cost of the solution and the remaining number of nurses and demands are then returned to the agent as a reward and a new observation. This process is repeated until there are no more nurses or demands.

feasible. A model-free algorithm (agent) obtains information about the environment by observing changes in environment dynamics without predicting changes based on a dynamics model (Watkins & Dayan, 1992). For example, in this study, the agent observes the size relation of the current numbers of unscheduled nurses and demands instead of predicting the exact changes in these numbers before choosing an action. The latter is much more computationally costly.

$\gamma \in [0, 1]$ is the discount factor, which represents the importance of future reward.

$R_{hd}(s, a_s) = \lambda_{wh} \bullet wh + \sum_{i\in D}\Delta d_{ss'}^{i}$ is the reward function in which $\Delta d_{ss'}^{i}$ is the amount of demand at level i that is fulfilled by the chosen nurse. Different values of λ_{wh} lead to different degrees of importance of the fulfilled demands and waiting times. When $\lambda_{wh} = 0$, $R_{hd}(\bullet)$ provides a reward that is only related to the number of demands a nurse fulfils, while it trades off between shortening the waiting time and increasing the number of fulfilled demands when $\lambda_{wh} \neq 0$.

The system starts at an arbitrary state s_1. Once a nurse a_1 is chosen, the environment generates a reward $R(s_1, a_1)$ and moves into a new state s_2. This procedure continues until it reaches a terminal state where there are either no more nurses or demands. Then, a worksheet (solution) is completed.

4.2. Chance-constrained programming

The CCP model has an advantage in formulating stochastic decision systems by assuming that an uncertain situation may hold with at least a probability α—namely, a confidence level (Charnes & Cooper, 1959). In a similar study, Li et al. (2010) use an SVRP that assumes that the travel time of vehicles must be adequate so that the arrival time can be within the time windows, and the total workload must be smaller than a given constant within the respective confidence levels. However, since waiting cost is more significant than travel cost in SHHCRP in densely populated areas, we emphasise the waiting time limitation more than the travel time. This assumes a constraint, which should be satisfied at a given confidence level, that the wait time of any nurse for any time window is less than a limited constant. In contrast, the constraint that the total workload is smaller than a given constant should be satisfied at another confidence level.

The mathematical model is proposed as follows. The objective, formalised by Eq. (1), is to minimise the average waiting time. Eq. (2) and Eq. (3) are the equivalents of the chance constraints of waiting time and overworking respectively corresponding to Eq. (A3.3) and Eq. (A3.4) derived in Appendix A3. Eq. (4) requires each demand to only be executed once. Eq. (5) and Eq. (6) require workers to start from and end at the home care centre. Eq. (7) specifies that workers must move on to another demand. Eq. (8) is the constraint of matching skill and demand levels (i.e. a worker can only execute a job if the worker meets the required qualifications). Eq. (9) ensures that workers never arrive after any time window. Eq. (10) is the definitional domain of confidence levels. Eq. (11) is the decision variable that equals 1 if worker k moves from elder i after finishing job n to j for job m to be executed; otherwise, it equals 0. The corresponding mathematical notations, the chance constraints and their deterministic equivalents, and the representations of human relationships are detailed in Appendix A.

$$\text{Min } F\left(x_{ijk}^{nm}\right) = \frac{\sum_{n\in n_i, m\in m_j i, j\in I}\left[x_{ijk}^{nm} \bullet w_{jk}^{m}\right]}{\sum_{n\in n_i, m\in m_j i, j\in I}\left(x_{ijk}^{nm}\right)}, \forall k \in K, \tag{1}$$

s. t.

$$b_{r_{n+1}} - \Phi^{-1}(\alpha)\sqrt{V(dc_{r_n})} - t_{r_n} - tc_{r_n r_{n+1}} - E(dc_{r_n}) \leq C, \forall r_n \in R_k, k \in K, \tag{2}$$

$$\Phi^{-1}(\beta)\sqrt{V(dc_{r_n})} + t_{r_n} + tc_{r_n r_0} + E(dc_{r_n}) + w_{r_0} \leq W, \forall r_n \in R_k, k \in K, \tag{3}$$

$$\sum_{k\in K}\sum_{i\in I}\sum_{j\in I} x_{ijk}^{nm} = 1, \forall n \in n_i, m \in m_j, \tag{4}$$

$$\sum_{j\in I}\sum_{m\in m_j} x_{0jk}^{0m} = 1, \forall k \in K, \tag{5}$$

$$\sum_{i\in I}\sum_{n\in n_i} x_{i0k}^{n0} = 1, \forall k \in K, \tag{6}$$

$$\sum_{i\in I\backslash 0}\sum_{n\in n_i\backslash 0} x_{ijk}^{nm} - \sum_{i\in I}\sum_{n'\in n'_i} x_{jik}^{mn'} = 0, \forall k \in K, j \in I, m \in m_j, \tag{7}$$

$$d_n, d_m \in sq_k, \forall x_{ijk}^{nm} = 1, i, j \in I, n \in n_i, m \in m_j, k \in K, sq_k \in SQ, \tag{8}$$

$$e_{r_n} \geq t_{r_n}, \tag{9}$$

$$\alpha, \beta \in [0, 1], \tag{10}$$

$$x_{ijk}^{nm} = \{0, 1\} \forall i, j \in I, n \in n_i, m \in m_j, k \in K. \tag{11}$$

Symbol	Definition
C	the maximum time within which workers can wait between time windows
r_n	the elder served at the current stage
r_{n+1}	the elder chosen to be served at the next stage
t_{r_n}	the arrival time of the worker k at the current elder r_n
$tc_{r_n r_{n+1}}$	the travel time from r_n to r_{n+1}
dc_{r_n}	the stochastic service time of r_n by k
$w_{r_{n+1}}$	the waiting time of worker k at r_{n+1}

5. Q-learning and best-worst ant colony optimisation (QL-BWACO) heuristic

5.1. Combination of ACO and QL

The heuristic solution is based on the Q-learning (QL) algorithm and the best-worst ant colony optimisation (BWACO) algorithm. Fig. 4 gives an overview of the heuristic solution. At the initial state of the environment, the QL algorithm starts choosing nurses to be scheduled sequentially. Each selected nurse becomes an input for the BWACO algorithm to generate a sub-solution (route). Then, the QL algorithm receives a reward associated with the quality of the route, updates the Q-matrix, and chooses the next nurse based on observing the nest state of the environment. The Q-matrix tabulates the Q values of any pair of action (nurse) and state (numbers of nurses and demands). When the system reaches a terminal state (no more nurses or demands), it finishes one episode and starts a new one, with the system being initialised again. Note that the Q-matrix is only initialised at the beginning. Details about updating the Q-matrix are given in Appendix A4.1.

5.2. Q-learning

Action and value. As shown in Fig. 4, the algorithm will observe the environment at each state and obtain information regarding the current unscheduled nurses and demands. Then it takes an action, i.e., chooses a nurse with a certain skill level and receives a reward. The action is taken either based on the Q-matrix or randomly following the ε-greedy heuristic rule. This means that the action is taken either according to the best-known Q-value with the probability ε or randomly with the probability $1-\varepsilon$ (Rodrigues et al., 2009). As long as a reward is received, the value of that action will be computed. In each state s, according to (Watkins & Dayan, 1992), a Q-value $Q(s, a_s)$ of an action a_s taken at state s is computed by Eq. (12):

$$Q(s, a_s) = R_{hd}(s, a_s) + \gamma \bullet \max_{a_{s'} \in A} Q(s', a_{s'}) \tag{12}$$

where $R_{hd}(s, a_s)$ is the reward of taking action a_s at state s, and $\max_{a_{s'} \in A} Q(s', a_{s'})$ is the best Q-value of the new state that could be obtained by taking the action $a_{s'}$ according to the current Q-matrix. Note that the Q-values are, by definition, the expected returns of taking an action at a state and following a policy thereafter. This formulation means the value of an action chosen at the current state is proportionately (at a given rate γ) associated with the highest value that could be obtained at the new state, in which way the algorithm attaches some degree of importance to future performance.

Q-matrix. Any pair of action and state values are tabulated into a Q-matrix in which columns are actions and rows are states. Once a sub-solution (route) is obtained, the QL agent will receive a reward associated with its quality. One generally needs the state transition probability to compute the expected reward, which is difficult to be achieved directly as explained in Section 3.2.

However, as a model-free reinforcement learning algorithm, Q-learning can find good solutions without full knowledge of the environment, as it only evaluates the action values (Watkins & Dayan, 1992). In general, the goal of QL is to construct a Q-matrix that indicates the values of the actions taken at any state of the MDP, i.e., a Q-matrix that indicates the performance of every nurse given the current unscheduled nurses and demands. The best schedule could be deduced according to the Q-matrix as it converges.

Updating the Q-matrix. In the beginning, the initial value of $Q_0(s, a_s)$ is randomized for each pair of states and actions, forming an initial matrix Q_0. Every time the algorithm receives a reward after taking action at a state, the corresponding element of the Q-matrix will be updated with a learning rate l, by which the agent will proportionally keep the old knowledge and learn the new knowledge. The updating equation is computed by Eq. (13):

$$Q_n(s, a_s) \leftarrow Q_{n-1}(s, a_s) + l \bullet \left[R_{hd}(s, a_s) + \gamma \bullet \max_{a_{s'} \in A} Q_{n\text{-}1}(s', a_{s'}) - Q_{n-1}(s, a_s) \right] \tag{13}$$

In summary, the algorithm will head towards convergence by iteratively interacting with the environment (taking actions and receiving rewards) and updating its action-value matrix (the Q-matrix), as proved by Watkins & Dayan (1992). As long as it converges, the Q-matrix will reflect the exact value of any action taken at any state, and thus the problem could be solved by applying the converged Q-matrix to the environment. However, since the reward defined in Section 3.2 requires sub-solutions of the CCP model, we propose another algorithm to solve the model in Section 4.2 effectively.

5.3. Best-Worst ant-colony optimization

Choosing elders. When the Q-learning procedure provides a nurse with a certain skill level, the BWACO algorithm begins building a route. This sub-procedure was modified according to the meta-heuristic given by (Dorigo & Stützle, 2010).

First, all of the ants start with an empty solution $s_p = \varnothing$. In every construction iteration, each ant completes a current partial solution s_p by sequentially choosing feasible solution elements $C_i^j \in \mathrm{N}(s_p) \subseteq C$, where $\mathrm{N}(s_p)$ is a set of solution elements that satisfies the feasibility. Note that the situation of no feasible solutions during this step is handled differently in different implementations, e.g., as an abandoned or penalized solution. Each ant chooses solution elements probabilistically according to different probability distributions, of which the most widely used one is that defined by the first ACO (Colorni et al., 1991).

That probability transition function is deducted by Eq. (14):

$$P^z_{r_n r_{n+1}}(\mathrm{t}) = \omega^z_{\mathrm{r_n r_{n+1}}}(\mathrm{t}) \bullet \frac{\tau^{\alpha_{aco}}_{r_n r_{n+1}} \bullet \left[\frac{1}{w_{r_{n+1}}(t)}\right]^{\beta_{aco}}}{\sum r_k \in s_z \tau_{r_n r_k} \alpha_{aco} \bullet \left[\frac{1}{w_{r_k}(t)}\right]^{\beta_{aco}}}, \forall r_{n+1} \in s^z_{available'} \tag{14}$$

where $P^z_{r_n r_{n+1}}(\mathrm{t})$ is the probability of ant z choosing edge $[r_n, r_{n+1}]$,

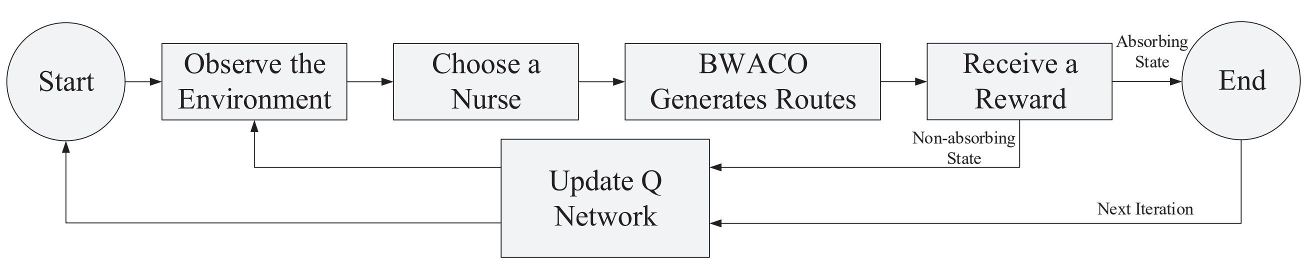

Fig. 4. Overall structure of the hybrid algorithm. At a decision timestep, the agent observes the number of nurses and demands and selects a nurse. This nurse is assigned a service plan by solving the CCP model via the BWACO algorithm. The cost of the planning solution is used to compute a reward for the agent to update the Q-matrix. The agent then receives a new observation, and the process repeats until there are no more nurses or demands. After a certain number of iterations, the Q-matrix will converge to a (sub-)optimum.

$\tau_{r_n r_{n+1}}$ is the pheromone density on edge [r_n, r_{n+1}], and $w_{r_{n+1}}(t)$. is the waiting time at r_{n+1}. Exponents α_{aco} and β_{aco} represent the relative importance of pheromone and heuristic information. $\alpha_{aco} = 0$ means the probabilities only depend on the heuristic information, and $\beta_{aco} = 0$ means the probabilities only depend on the pheromone density. $\omega^z_{r_n r_{n+1}}(t)$ is the function related to the chance constraints defined as Eq. (15).

$$\begin{aligned} ccwt &= b_{r_{n+1}} - \varphi^{-1}(\alpha)\sqrt{V(dc_{r_n})} - t_{r_n} - tc_{r_n}c_{n+1} - E(dc_{r_n}), and \\ cco &= \varphi^{-1}(\beta)\sqrt{V(dc_{r_n})} + t_{r_n} + tc_{r_n r_0} + E(dc_{r_n}), then, \\ \omega^z_{r_n r_{n+1}}(t) &= \left\{ \begin{array}{c} 1, if, ccwt \le C \\ 0, if, ccwt > C, or, cco \le e_{r_{n+1}} \end{array} \right\} \end{aligned} \tag{15}$$

Updating pheromones. In this step, the pheromones are updated to increase the probability of achieving good solutions iteratively. Two mechanisms are used in this process: pheromone deposition and pheromone evaporation. Note that pheromone evaporation protects the algorithm from overly rapid convergence by proportionally decreasing the influence of previous solutions and enhancing the exploration of new possible solutions (Colorni et al., 1991). The best-worst principle was used to update the pheromones according to the best and worst solutions found by ants (Cordon et al., 2000; Wei, 2013).

The pheromone update function is performed according to Eq. (16) and Eq. (17):

$$\tau_{r_n r_{n+1}} = (1-\rho)\tau_{r_n r_{n+1}} + \Delta\tau_{r_n r_{n+1}} \tag{16}$$

$$\Delta\tau_{r_n r_{n+1}} = \left\{ \begin{array}{c} tC_{r_n r_{n+1}}, r_n, r_{n+1} \in s^{z-best}_{suboptimal} \\ -tC_{r_n r_{n+1}}, r_n, r_{n+1} \in s^{z-worst}_{suboptimal} \\ 0, otherwise \end{array} \right\} \tag{17}$$

where $\rho \in [0, 1]$ is the evaporation rate that partially keeps the previous knowledge, and $\Delta\tau_{r_n r_{n+1}}$ is the pheromone increment on edge [r_n,r_{n+1}] involved in the best and the worst solutions.

The algorithm will move towards convergence by iteratively interacting with the environment and updating the Q-matrix, as demonstrated by Watkins & Dayan (1992). If it converges, the Q-matrix will reflect the (approximately) true, expected value of any action taken at any state. Thus, the problem can be solved by applying the converged Q-matrix to the environment.

Since the reward function defined in Section 4.1 requires sub-solutions of the CCP model, we use a BWACO-based heuristic algorithm to solve the CCP model described in Section 4.2. ACO was first introduced in 1992 to address TSP by mimicking ants' behaviour (Colorni et al., 1991). It was recently reviewed in detail and found to be effective for solving many variants of routing problems (Dorigo & Stützle, 2010). In addition to routing problems, ACO is also applied in other scenarios (e.g. network design (Zohal & Soleimani, 2016), flow shop scheduling (Rajendran & Ziegler, 2004), multi-item inventory routing problems (Huang & Lin, 2010), and pipeline routing (Baeza et al., 2017). The BWACO heuristic repeats two procedures: (i) elder-choosing (i.e. choosing elders to build a route for the nurse) and (ii) pheromone-updating (i.e. updating the pheromone density according to the previous results). These two procedures are explained in detail in Appendix A4.2. The heuristic will converge to a (sub-)optimum as ants will greedily choose the route with the highest pheromone density and increase the pheromone density in return.

5.4. Algorithm complexity

Their input quantity usually determines the complexity of algorithms, and the growth rate of complexity varies among different algorithms. This article uses time complexity evaluation to measure algorithm complexity. The time complexity quantitatively describes the algorithm's running time, as shown in the simulation experiment in Table 7. As shown in the figure, the method proposed in this article has made progress based on the combination of the ant colony and q-learning algorithms. Similar instance convergence can be obtained with less computational time, thereby demonstrating its superiority.

Algorithm QL_BWACO

```
1 Input:Instance representations,CCP model parameters,ACO parameters,Q learning
  parameters
2 Output:Variables(*Q_nurses_list).Constants(total workload,total waiting time)
3Initialize variables for output
4 Evaluate solutions
5 Solution recording variables
6 Start iteration
7 for ant in range(ant_num) do
   current_job = depot current_time,waiting,workload = 0
   Bulid routes
   While current_workload <=workload do
   St_mean = service_time_mean[nurse,s][current_job.lv]
   Preference_factor=copy.deepcopy(current_preference[current_job.e][nurse.l])
     If current_time < current_job.twb,then
       Current_waiting +=(current_job.twb - current_time)
       Current_time = copy.deepcopy(current_job.twb)
       Sub_arrival_time.append(copy.deepcopy(current_time))
     Else
       Current_time += (preference_factor*st_mean) + alpha_model_p
8 Search for targets satisfying the time window constraint
9 collect feasible_targets and calculate transition probability
   If (len(feasible_targets))==0 no feasible targets, end routing,then
   Elif (len(feasible_targets)==1 only one feasible target,choose it and update route
   Remove chosen target from visiting list
   For v in range(len(visiting_list))do
     If visiting_list[v].1==feasible_targets[0].1,then
       Visiting_list.remove(visiting_list[v])
     Return feasible_targets[0]
   Else more than 1 feasible targets,calculate transition probabilities
10 update route
   Nurse.tt=copy.deepcopy(shortest_time)
   Nurse.aT=copy.deepcopy(arrival_time_trace)
   Nurse.twt=copy.deepcopy(waiting_time)
   For o in range(len(best_path))do
     Nurse.r.append(best_path[0])
     If best_path[0].lv ==1,then
       Nurse.sd[0]+=1
     eIif best_path[0].lv ==2,then
       Nurse.sd[1]+=1
     elif best_path[0].lv ==3
       Nurse.sd[2]+=1
12 Start Q Learning process
13 Calculate fulfilled and remaining demands
14 Update global solution according to evaluation value
15 If solution_final.ev < sub_solution.ev,then
16 Solution_final = copy.deepcopy(sub_solution)
```

6. Experimental results and discussion

We assume that the stochastic service time is based on normal distributions with different mean values (Table 1). In the following experiments, we assume that skills and demands are limited to three classes: (i) basic care: meal provision, house cleaning, and chatting services, among others, which are considered the most basic and the easiest services; (ii) body checking: physical examination and symptom identification using basic medical devices, among others, which require

Table 1
Initial mean values, $\mu^n_{ik}(d_n, s_k)$, of service time regarding different pairs of skill and demand levels. It is assumed that a nurse with a higher skill level can serve lower-level demands with less service time, and a nurse with a lower skill level cannot serve higher-level demands (denoted '–').

Demand level (d_n)	Skill level (s_k)		
	Primary (Lv. 1)	Senior (Lv. 2)	Professional (Lv. 3)
Basic care (Lv. 1)	25 (min)	20 (min)	18 (min)
Body checking (Lv. 2)	–	30 (min)	20 (min)
Medical guidance (Lv. 3)	–	–	20 (min)

a fundamental knowledge of medicine; (iii) medical guidance: diagnosis, nutrition consulting, and hospital transfer advice, among others, which require professional medical knowledge.

Experimental results with different parameter settings are based on the service time initialisation. We provide results with different parameter settings in the MDP-CCP model and draw managerial implications from the results (Section 6.1). We discuss the advantages of the algorithm compared with other works (Section 6.2) and present feedback collected from practitioners and researchers (Section 6.3). Appendices A5 and A6 present the results of testing the BWACO and QL processes (convergence and hyperparameter experiments). All experiments were performed on a laptop with an Intel Core i5-3230 M 2.60 GHz CPU and an 8 GB DDR3 RAM. The algorithm was coded in Python 3.5.3.

6.1. Parameter settings of the Markov decision process-chance constrained programming (MDP-CCP) model

Table 2 shows the results of testing the parameters associated with the MDP-CCP model, including the waiting time constant, initial preference, preference decrement, and both confidence levels. The experimental results were obtained using the fixed algorithmic parameters listed in Table 3.

6.1.1. Waiting and workload limitations (instances A, B, and C)

We provide the solutions generated by QL-BWACO under different waiting and workload limitations. Based on the results here, managers may increase their confidence in a routing plan with a better balance between flexibility and costs.

As assumed in Section 3, the chance constraint of waiting time (CCWT) limits the maximum time (C) within which workers can wait between time windows. Different values of C can be regarded as the different extent to which institutions allow workers to wait. The flexibility of scheduling is also influenced by waiting limitations. The chance constraint of overworking (CCO) limits the maximum total workload (W) beyond which workers should stop working. The workload is critical for worker health, safety, and fairness.

First, we find that adjusting nurse sequences can keep the average waiting time stable or even reduce it when the waiting limitation increases. Fig. 5 and Fig. 6 show that the average waiting time increases, and the unfulfilled demands decrease as the waiting limitation increases. Although the overall tendency is growing, some values decrease as the limit increases. Such a small difference is attributable to the influence of nurse sequences (corresponding data are boldfaced in Table B1 in Appendix B). We find that the average waiting time with different sequences is stable or even decreases, revealing the algorithm's adaptability. Because nurses with higher skill levels require less service

Table 2
Parameters and instances for testing the Markov decision process-chance constrained programming (MDP-CCP) model.

Parameter values to test MDP-CCP						
C	W	a_{ik}^0	Δa	α	β	
0, 20, 40, 60, 100, 480	400, 480	1, random	0.05, 0.01, 0.2	0.8, 0.9,0.95	0.8, 0.9, 0.95	
Instances						
Index	Elders	Nurses	Jobs	Jobs Lv. 1	Jobs Lv. 2	Jobs Lv. 3
A	30	Lv.1–3 (1, 3, 3)	60	15	37	8
B	30	Lv.1–3 (1, 3, 3)	60	27	19	14
C	30	Lv.1–3 (1, 3, 3)	55	24	28	3
D	80	Given in Section 6.1.4	159	23	107	29
E	150	Given in Section 6.1.4	303	119	140	44

Table 3
Q-learning (QL) and best-worst ant-colony optimisation (BWACO) parameter values for testing the MDP-CCP model.

Parameters		Values
QL	*Learning rate (l)*	*0.9*
	Discount factor (γ)	*0.9*
	Greedy rate (ε)	*0.5*
	Working time importance degree (λ_{wh})	*−0.01*
BWACO	*Initial pheromone (τ_0)*	*20*
	Evaporation rate (ρ)	*0.5*
	Pheromone weight (α_{aco})	*1*
	Heuristic weight (β_{aco})	*1*

time and meet more demands but are more expensive, these results can help decision-makers make trade-offs between more expensive personnel and customer demands that must be satisfied.

Second, we find that nurses can fulfil more demands with a higher waiting limitation. Almost all demands are fulfilled when the waiting limitations are set to 480 min in instances A, B, and C with higher waiting times. When the instances are difficult to schedule, decision-makers might have to accept a higher waiting cost. Therefore, stipulating a proper limitation of waiting time can help fulfil more demands at a smaller cost. When satisfying all demands is a high priority, decision-makers might not be able to fully utilise human resources unless the waiting time limitations are adequately satisfied.

These results approximately reflect the real environment with different demands and provide nearly the best solutions, demonstrating that solutions generated by the method can help decision-makers. Appendix E includes the figures for arrival times and time windows associated with comparative solutions.

6.1.2. Confidence levels of the chance constraints of waiting time and overwork (instance A)

The two confidence levels represent the degree to which decision-makers believe that solutions will satisfy the waiting time and overwork limitations. Setting higher values for the confidence level is equal to assuming a lower probability of breaking the limitations. Since service time is the only random variable, a higher confidence level corresponds to a larger interval of service time estimation. However, the chance constraints are transferred to their deterministic equivalents, which changes the confidence levels to different degrees of time increments. According to Eqs. (2) and (3) and the standard normal distribution, when the confidence levels are higher, the value of the inverse standard normal distribution function becomes smaller, and the estimations of service time increase. We test different values for the confidence levels of the CCWT and CCO. The results are given in Table B2 in Appendix B.

First, we find that when the CCWT confidence level (α) increases, QL-BWACO produces solutions with lower waiting time costs (Fig. 7a). As stated in Section 4, the service time will have different estimated values with different values of α. However, these results show that with higher waiting time limitations, increasing the confidence level of CCWT will decrease the waiting time. This indicates that decision-makers should adjust the confidence level of the CCWT to achieve better scheduling in different environments and that the QL-BWACO algorithm can assist in their decision-making.

Second, the confidence level of CCO is found to only influence the system's performance (Fig. 7b). We test different values of β with a workload limitation of 400 min. We find that as β decreases, the average workload is predicted to be shorter, although the increase rate is low. If decision-makers strictly require nurses to not overwork, the system will predict higher values of working time without breaking the workload limitation (W). Since the solutions are satisfactory, it is unnecessary to reschedule nurse's sequences or routes; thus, the workload is the only value that is changed.

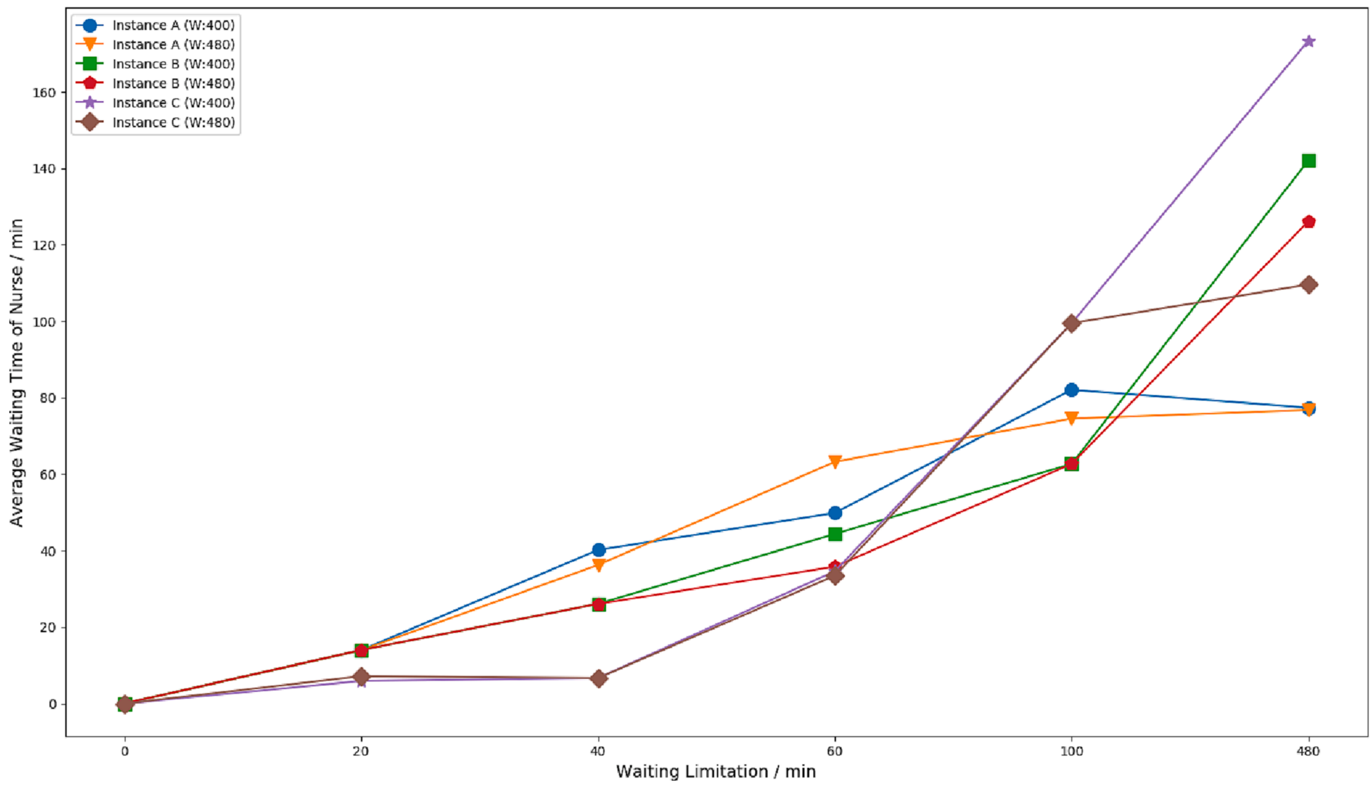

Fig. 5. Average waiting time of nurses. Each line corresponds to an experimental case, and each point corresponds to an estimated average waiting time of *nurses* given a waiting time limitation value.

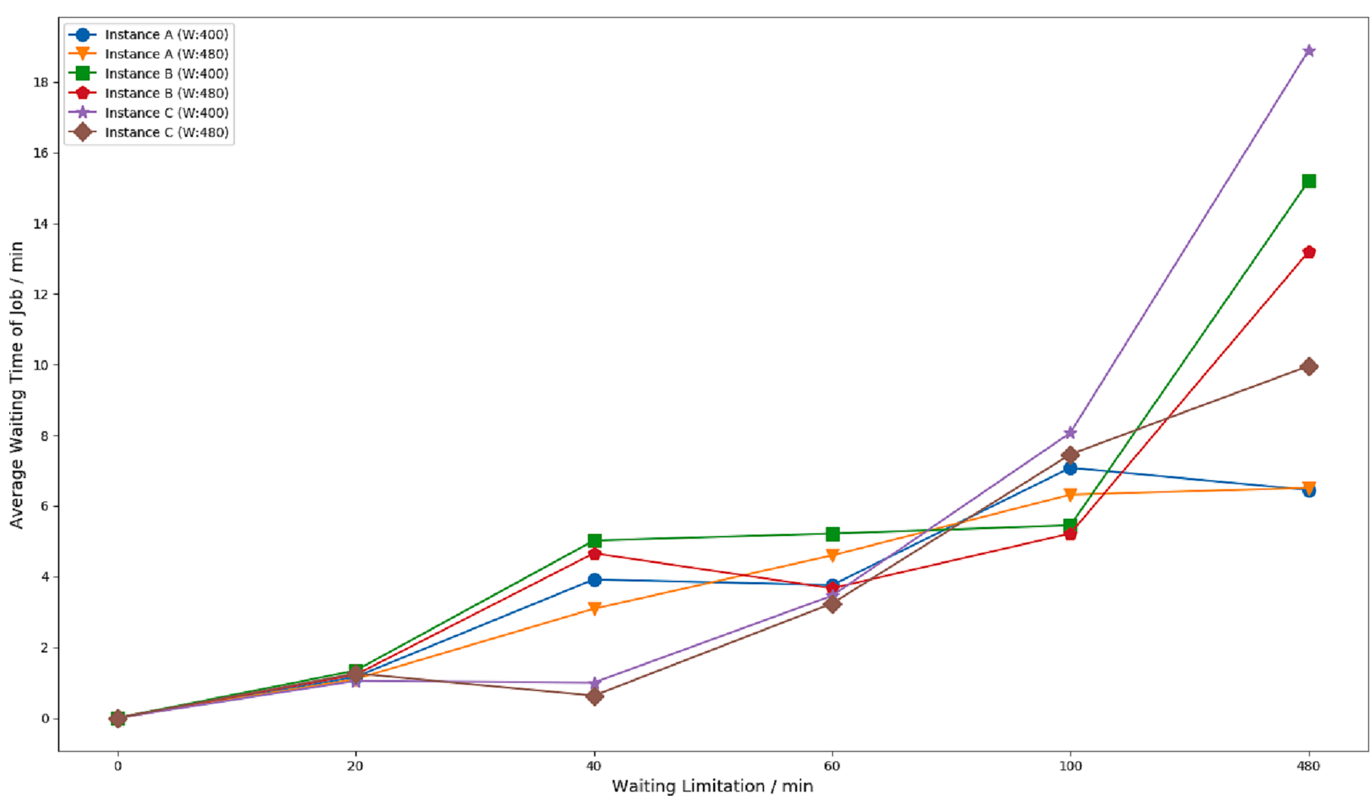

Fig. 6. Average waiting time of jobs. Each line corresponds to an experimental case, and each point corresponds to an estimated average waiting time of *jobs* given a waiting time limitation value.

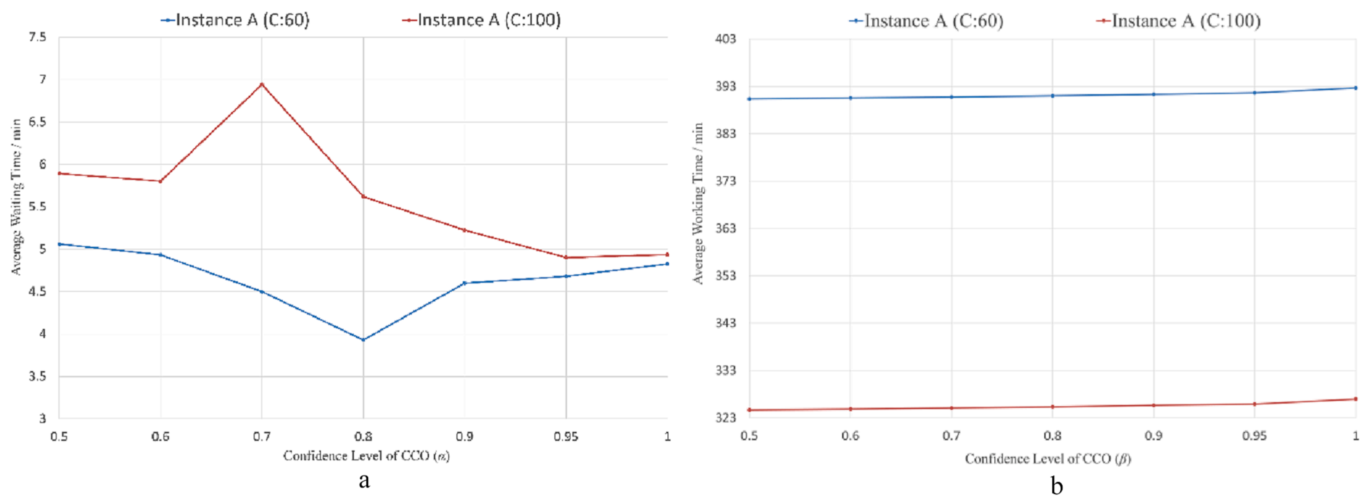

Fig. 7. Sensitivity analysis. a. Changes in the average waiting time for jobs with respect to the CCWT confidence level α. The figure shows that if the decision-maker increases the degree of satisfying the waiting time constraints (the value of α), the algorithm will produce plans with slightly lower waiting costs. b. Changes in the average workload with respect to the CCO confidence level β. The figure shows that if the decision-maker increases the degree of satisfying the workload constraint (the value of β), the algorithm will produce plans with slightly higher predicted workloads.

Adapting the confidence level of the CCWT in different situations is found to be a valuable contributor to making better schedules. Moreover, it can guide QL-BWACO to generate better solutions and provide valuable decision-making advice. Since QL-BWACO produces different solutions when the CCWT confidence level (α) changes, nurses will have different choices for completing their tasks with different confidence levels. In other words, managers can define different flexibilities for their HHC schedules.

Together with the results in Section 6.1.1, we find that HHCRP in densely populated areas increases service rates with shorter travel distances. This means that nurses who provide services in densely populated communities can execute services with higher efficiency and workloads. Hence, in consideration of socioeconomic sustainability, decision-makers should aim to balance the workloads of employees and the quality of care provided to elders with institutional profits under different scenarios.

6.1.3. *Initial preference and decrement (instance A)*

We review the parameters associated with human behaviours, including mutual preferences among workers and elders. We define a preference variable a_{ik}^0 that affects the mean values of service time and a rule to decrease a_{ik}^0 after every service by each worker–elder pair (Δa). Fixed and random initial preference values are simulated for different, predefined relationships between workers and elders, which decision-makers can flexibly change according to different situations.

We find that setting higher decrements (a_{ik}^0) of service time can increase the efficiency and profits of HHC activities. In Table 4 (boldfaced numbers), as a_{ik}^0 increases, service time decreases, and fulfilled demands increase. In other words, nurses can simultaneously save more service time and satisfy more demands with preference matching.

These results generally reveal that considering human relationships in scheduling improves social satisfaction. Assigning nurses to mutually preferred elders makes the organisation of human resources more humane, effective, and efficient, as shown in Table 4. Hence, HHC providers should consider the influence of human relationships. Moreover, the results with constant and random initial preference parameters (Table 4) show a significant difference between considering individual preferences and not doing so. These results might not be the same as those obtained using real-world data. However, they are sufficiently interesting to suggest that managers should humanise schedules and the interactions between nurses and elders to positively contribute to their quality of life.

Further, the QL-BWACO algorithm performs well with both fixed and random initial preference values. The solutions are valuable for decision-makers. The randomly initialised preference matrices are included in Appendix D1, and the increments of elders' preferences after being served are included in Appendix D2.

Table 4
Results with different initial preferences and decrements of instance A. For both constant and random initial preferences, increasing the preference increment factor (Δa) results in increasing demand fulfilment. The different results between constant and random initial preferences also demonstrate the importance of considering individual differences among nurses and elders.

Instance (C, W)	(a_{ik}^0, Δa)	Unfulfilled demands Lv.1–3	Nurses sequences (skills)	Average working time (min)	Average service time of nurse/job (min)	Demands fulfilled rate (%)
A (60, 480)	(1, 0)	(1, 3, 3)	3, 2, 3, 2	405.59	318.88/24.07	88.3
	(1, 0.01)	(1, 3, 0)	3, 2, 3, 3	413.24	313.48/22.39	93.3
	(1, 0.05)	(2, 2, 1)	3, 3, 3, 2	399.40	289.76/21.07	91.7
	(1, 0.1)	(1, 2, 1)	3, 3, 3, 2	386.15	276.35/19.74	93.3
	(1, 0.2)	(0, 0, 0)	3, 3, 3, 2	363.05	251.34/16.76	100
	(random, 0)	(0, 4, 4)	3, 3, 2, 2	400.37	303.82/23.37	86.7
	(random, 0.01)	(2, 4, 2)	2, 3, 3, 2	400.31	308.18/23.71	86.7
	(random, 0.05)	(2, 1, 4)	3, 3, 2, 2	391.84	290.24/21.9	88.3
	(random, 0.1)	(0, 2, 3)	3, 3, 2, 2	389.77	279.61/20.34	91.7
	(random, 0.2)	(1, 1, 1)	3, 3, 2, 3	361.62	241.96/16.98	95.0

6.1.4. Different nurse sources and a larger instance (instance D)

In the results presented in Sections 6.1.1–6.1.3, the solutions for instances A, B, and C do not necessarily occupy all the nurses, which means unavoidable idle time due to the demand distribution. Therefore, we apply a larger instance, designated as instance D (Table 2), with different nurse resources to further evaluate the performance of the model and the algorithm. The parameter settings are fixed as shown in Table 3 and Table 5.

The results in Table 6 show that the algorithm generates a reasonable solution in a large-scale instance with a good nurse source. In practice, decision-making should not depend solely on the predictive results of scheduling and routing. However, those results are an important part of the information that decision-makers need. Furthermore, the results presented in Table 6 illustrate the following characteristics of the SHHCRP:

1) Travelling in general constitutes about 10% of the total work time, which is less than in previous studies that did not consider densely populated communities.
2) Different nurse sequences significantly affect the system's performance.
3) Our results can help decision-makers balance the supply of nurses to fulfil eldercare demands.

6.2. Comparison with related work

To the best of our knowledge, there is currently no benchmark for solving the complex problems considered in this study. We do, however, try to test the practicality of our experimental results by comparing instances and algorithms with those used in the most relevant studies. Nevertheless, our study has the highest problem complexity, as discussed in Section 3.

First, although the assumptions are different in existing studies, our examples are more explicitly provided. As shown in Table 7, ours is the only study that provides complete information about the examples (terms that lack data are marked '-' in Table 7). Therefore, our results have relatively higher reliability and practicality.

Second, the QL-BWACO algorithm described in Section 4 shows comparable or less convergence time than others. Though our method and others are not strictly comparable because of different problem assumptions and hardware support, our method is preferable to others for the following reasons:

1) The problems investigated in this study have the highest complexity (see Section 3 for a detailed comparison of considered constraints).
2) The algorithm used in this study requires fewer iterations for the most similar instances to converge.
3) The algorithm used in this study requires less computation time for the most similar instances to converge.
4) The results in Kergosien et al. (2009) are based on an exact algorithm and a relatively small number of experimental cases, among which the largest required more computational time to converge than was needed for the smallest experimental case tested in our study. Note that the algorithms used in Trautsamwieser & Hirsch (2011) and Nickel et al. (2012) and in our study are searching algorithms. Therefore, the maximum iteration number and CPU time (the third and fourth columns of Table 7, respectively) are not necessarily comparable. Instead, the convergence speeds (the fifth column of Table 7) provide more important information for comparison.

Table 5
Algorithm parameter values of MDP-CCP for larger-instance experiments. MDP-CCP: Markov decision process-chance constrained programming.

Parameter		Value
MDP-CCP	*CCWT constant (C)*	*40* (min)
	CCO constant (W)	*480* (min)
	Initial preference value (a_{ik}^0)	*1*
	Preference decrement (Δa)	*0.05*
	Confidence level of CCWT (α)	*0.9*
	Confidence level of CCO (β)	*0.9*

6.3. Managerial insights

The results of this study have practical implications for the management of home healthcare services in densely populated areas. Our findings suggest that decision-makers in the home care industry should prioritize the importance of human relationships in scheduling, as this can improve nurses' work efficiency and enhance service satisfaction for both caregivers and customers. This may involve strategies such as assigning nurses to work with the same group of elderly patients over time, or allowing nurses to choose their own assignments based on personal preferences and expertise.

Additionally, it is crucial for home care managers to carefully consider the trade-off between workload and profit, particularly in densely populated areas where nurses may be expected to take on heavier workloads. To avoid overworking existing nurses and risking burnout or turnover, it may be more beneficial to hire additional nurses, if the budget allows for it. This could also improve the overall quality of care by enabling nurses to spend more time with each patient and address their needs more effectively.

For policymakers, our findings suggest the need for increased resources to be allocated towards the development of data collection and processing systems for home care services. This would enable the next generation of home caregivers to have access to more effective background knowledge and work efficiently with AI-based technologies. By investing in robust data infrastructure and analytics capabilities, policymakers can better understand the needs and preferences of elderly patients and caregivers and design policies and programs that address these needs more effectively.

7. Conclusions

This study investigates home healthcare resource planning (HHCRP) in a real-world context, specifically in densely populated communities where elderly individuals reside. We analyse crucial constraints such as human relationships, multi-appointments, skill-demand matching, and stochastic service times in order to gain insights that can support HHC decision-making processes. To address the stochastic nature of this sequential decision-making problem, we propose a hybrid mathematical model combining the Markov decision process and chance-constrained programming. The resulting algorithm, called QL-BWACO, is shown to perform reliably and efficiently with medium- and large-scale experimental cases and is empirically documented to be the best among the options used in other studies.

One limitation of this study is the imperfect mathematical model used to represent human relationships and health care provider service times. While the model was built upon expert knowledge-based heuristics and received positive feedback from practitioners, it could be improved with further research and development. Integrating big data and machine learning techniques may offer a promising way to better approximate real-world situations and enhance the accuracy of the human relations model. This will require a well-defined data collection process, as suggested by a healthcare practitioner.

In the future, research on HHCRP in densely populated communities could focus on two main areas: understanding the needs of caregivers and elderly individuals in these communities, including both physical and mental aspects, and building a data collection system for collecting real-world data to monitor, analyze, and predict changes in demand. These data can be used to balance supply and demand and develop and implement more effective models and scheduling algorithms.

Table 6
Results for instances D and E with different nurse resources. Three settings of nurses and two settings of nurses were tested for instances D and E, respectively. For example, the setting '(1, 3, 3)' indicates one nurse for level 1 and three nurses for level 2 and level 3. The resulting solutions and wall-clock runtime document the efficiency and feasibility of the proposed method to handle large-scale problems.

Ins.	Nurses	CPU (min)	Maximum/ converge iteration	Unfulfilled demands Lv.1–3	Nurse sequences (skills)	Idle nurses (skills)	Avg. workload (min)	Avg. waiting (nurse/job)	Avg. travel (min)
D	Lv.1–3 (1, 3, 3)	24.7	400/<100	(4, 39, 16)	2–3-3–2-3–2-1	None	398.73	19.45/1.35	43.13
	Lv.1–3 (1, 5, 3)	28.1		(3, 21, 16)	3–3-2–2-3–2-2–2	1	442.21	34.36/2.29	45.06
	Lv.1–3 (1, 5, 5)	32.1		(3, 18, 7)	3–2-3–3-2–3-3–2	1, 2, 2	446.33	31.48/1.91	48.94
E	Lv.1–3 (3, 5, 7)	75.5	400/<200	(25, 39, 7)	3–2-2–1-3–1-1–2-2–3-3–2-3–3-3	None	459.38	108.08/6.99	46.26
	Lv.1–3 (5, 7, 8)	201.4	800/<500	(5, 16, 1)	2–2-1–2-3–1-1–2-3–2-2–2-2–1-3–3-1–3-3–3	None	435.95	106.63/7.59	49.54

Table 7
Size of experimental cases and performance of algorithms for comparison. Although not strictly comparable because of differences in problem assumptions and hardware settings, this study's method and experimental results are relatively more practical and reliable than those of other studies.

Study	Method	Max. iteration	Max. CPU	Convergence iteration	Convergence CPU	Nurses	Elders	Demands	Nurses/demands level data
(Kergosien et al., 2009)	Cplex	–	0.18 s 1.3 min 1.5 min 3 min	–	0.18 s 1.3 min 1.5 min 3 min	–	–	10 20 30 40	–
(Trautsamwieser & Hirsch, 2011)	Heuristic	10^6	–	$<10^4$ $<10^5$ $<10^5$	–	13 39 75	140 291 420	140 351 512	–
(Nickel et al., 2012)	Heuristic	1000 1000	2 h	<550 <1000	<1.1 h <2 h	11 12	52 95	287 361	–
***This study**	Heuristic	400 400 400 800	6.5 min 28.3 min 1.3 h 3.4 h	<50 <100 <200 <300	<0.7 min <7.1 min <0.7 h <1.3 h	7 7, 9, 11 15 20	30 80 150 150	60 159 303 303	Given in Section 6.1

CRediT authorship contribution statement

Ting Zhang: Conceptualization, Writing – original draft, Methodology. **Yang Liu:** Supervision, Validation, Writing – review & editing. **Xintong Yang:** Methodology, Software. **Jingjing Chen:** Data curation. **Jiaming Huang:** Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data is available in the Appendix.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (71701052), Shenzhen High-Caliber Personnel Research Start-up Project (2020111), Shenzhen Science and Technology Program (JCYJ20210324121203008), Guangdong Key Construction Discipline Research Ability Enhancement Project (2021ZDJS108), Innovation teams of ordinary universities in Guangdong Province (2022KCXTD027). Xintong Yang thanks the China Scholarship Council for financially supporting his PhD studies at Cardiff University (201908440400). The authors would like to thank Prof. Donald Huisingh at the University of Tennessee, Knoxville, USA, for his support and language editing to improve this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cie.2023.109332.

References

Allahviranloo, M., Chow, J. Y. J., & Recker, W. W. (2014). Selective vehicle routing problems under uncertainty without recourse. *Transportation Research Part E: Logistics and Transportation Review, 62*, 68–88. https://doi.org/10.1016/j.tre.2013.12.004

Baeza, D., Ihle, C. F., & Ortiz, J. M. (2017). A comparison between ACO and Dijkstra algorithms for optimal ore concentrate pipeline routing. *Journal of Cleaner Production, 144*, 149–160.

Bellman, R. (1966). Dynamic programming. *Science, 153*(3731), 34–37.

Bertels, S., & Fahle, T. (2006). A hybrid setup for a hybrid scenario: Combining heuristics for the home health care problem. *Computers & Operations Research, 33*(10), 2866–2890.

Charnes, A., & Cooper, W. W. (1959). Chance-constrained programming. *Management Science, 6*(1), 73–79.

Choi, T. M. (2020). Innovative "bring-service-near-your-home" operations under Corona-virus (COVID-19/SARS-CoV-2) outbreak: Can logistics become the messiah? *Transportation Research Part E: Logistics and Transportation Review, 140*, 101961.

Colorni, A., Dorigo, M., & Maniezzo, V. (1991). Distributed optimisation by ant colonies. *Paper presented at the Proceedings of the first European conference on artificial life*.

Cordon, O., de Viana, I.F., Herrera, F., & Moreno, L. (2000). A new ACO model integrating evolutionary computation concepts: The best-worst Ant System. Proc. ANTS 2000, 22-29.

Dorigo, M., & Stützle, T. (2010). Ant colony optimisation: Overview and recent advances. In *Handbook of metaheuristics* (pp. 227–263). Springer.

Du, G., & Zhang, J. (2022). Cross-regional manpower scheduling and routing problem with stochastic service times in home health care. *Computers & Industrial Engineering, 173*, 108668.

Eveborn, P., Flisberg, P., & Rönnqvist, M. (2006). Laps care—An operational system for staff planning of home care. *European Journal of Operational Research, 171*(3), 962–976.

Fathollahi-Fard, A. M., Ahmadi, A., Goodarzian, F., & Cheikhrouhou, N. (2020). A bi-objective home healthcare routing and scheduling problem considering patients' satisfaction in a fuzzy environment. *Applied Soft Computing*, Article 106385.

Fikar, C., & Hirsch, P. (2015). A matheuristic for routing real-world home service transport systems facilitating walking. *Journal of Cleaner Production, 105*, 300–310.

Fikar, C., & Hirsch, P. (2017). Home health care routing and scheduling: A review. *Computers & Operations Research, 77*, 86–95.

Huang, S. H., & Lin, P. C. (2010). A modified ant colony optimization algorithm for multi-item inventory routing problems with demand uncertainty. *Transportation Research Part E: Logistics and Transportation Review, 46*(5), 598–611.

Kall, P., Wallace, S. W., & Kall, P. (1994). *Stochastic programming*. Springer.

Statista (2023). Population density in urban areas of China in 2021, by region (in inhabitants per square kilometer) [Graph]. Retrieved from https://www.statista.com/statistics/279040/population-density-in-urban-areas-of-china-by-region/.

Kergosien, Y., Lenté, C., & Billaut, J.-C. (2009). *Home health care problem: An extended multiple traveling salesman problem.* In Paper presented at the Proceedings of the 4th Multidisciplinary International Scheduling Conference: Theory and Applications (MISTA 2009).

Khodabandeh, P., Kayvanfar, V., Rafiee, M., & Werner, F. (2021). A bi-objective home health care routing and scheduling model with considering nurse downgrading costs. *International Journal of Environmental Research and Public Health, 18*, 900. https://doi.org/10.3390/ijerph18030900

Lenzi, J., et al. (2019). Are degree of urbanisation and travel times to healthcare services associated with the processes of care and outcomes of heart failure? A retrospective cohort study based on administrative data. *PLoS ONE, 14*(10), Article e0223845.

Li, X., Tian, P., & Leung, S. C. (2010). Vehicle routing problems with time windows and stochastic travel and service times: Models and algorithm. *International Journal of Production Economics, 125*(1), 137–145.

Lin, M., Ma, L., & Ying, C. (2021). Matching daily home healthcare demands with supply in service-sharing platforms. *Transp Res E Logist Transp Rev, 145*, Article 102177. https://doi.org/10.1016/j.tre.2020.102177

Liu, R., Xie, X., & Garaix, T. (2014). Hybridisation of tabu search with feasible and infeasible local searches for periodic home health care logistics. *Omega, 47*, 17–32.

Liu, R., Yuan, B., & Jiang, Z. (2019). A branch-and-price algorithm for the home-caregiver scheduling and routing problem with stochastic travel and service times. *Flexible Services and Manufacturing Journal, 31*, 989–1011.

Low, L. F., Yap, M., & Brodaty, H. (2011). A systematic review of different models of home and community care services for older persons. *BMC Health Services Research, 11*, 93. https://doi.org/10.1186/1472-6963-11-93

Mehmet, E., & Çağrı, K. (2022). Home health care and dialysis routing with electric vehicles and private and public charging stations. *Transportation Letters*. https://doi.org/10.1080/19427867.2022.2057899

Nickel, S., Schröder, M., & Steeg, J. (2012). Mid-term and short-term planning support for home health care services. *European Journal of Operational Research, 219*(3), 574–587.

Oyola, J., Arntzen, H., & Woodruff, D. L. (2016). The stochastic vehicle routing problem, a literature review. part ii: Solution methods. *EURO Journal on Transportation and Logistics*, 1–40.

Rajendran, C., & Ziegler, H. (2004). Ant-colony algorithms for permutation flowshop scheduling to minimise makespan/total flowtime of jobs. *European Journal of Operational Research, 155*(2), 426–438.

Redjem, R., & Marcon, E. (2016). Operations management in the home care services: A heuristic for the caregivers' routing problem. *Flexible Services Manufacturing Journal, 28*(1–2), 280–303.

Restrepo, M. I., Rousseau, L.-M., & Vallée, J. (2020). Home healthcare integrated staffing and scheduling. *Omega, 95*, Article 102057. https://doi.org/10.1016/j.omega.2019.03.015

Rodrigues Gomes, E., & Kowalczyk, R. (2009). Dynamic analysis of multiagent Q-learning with ε-greedy exploration. In Proceedings of the 26th annual international conference on machine learning pp. 369-376.

Rodriguez, C., Garaix, T., Xie, X., & Augusto, V. (2015). Staff dimensioning in homecare services with uncertain demands. *International Journal of Production Research, 53*(24), 7396–7410. https://doi.org/10.1080/00207543.2015.1081427

Shi, Y., Boudouh, T., & Grunder, O. (2019). A robust optimisation for a home health care routing and scheduling problem with consideration of uncertain travel and service times. *Transportation Research Part E: Logistics and Transportation Review, 128*, 52–95. https://doi.org/10.1016/j.tre.2019.05.015

Shi, Y., Boudouh, T., Grunder, O., & Wang, D. (2018). Modeling and solving simultaneous delivery and pick-up problem with stochastic travel and service times in home health care. *Expert Systems with Applications, 102*, 218–233.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT Press.

Trautsamwieser, A., & Hirsch, P. (2011). Optimisation of daily scheduling for home health care services. *Journal of Applied Operational Research, 3*(3), 124–136.

Watkins, C. J., & Dayan, P. (1992). Q-learning. *Machine Learning, 8*(3–4), 279–292.

Wei, X. (2013). Improvement and implementation of best-worst ant colony algorithm. *Research Journal of Applied Sciences, Engineering and Technology, 5*(21), 4971–4976.

Yang, M., Ni, Y., & Yang, L. (2021). A multi-objective consistent home healthcare routing and scheduling problem in an uncertain environment. *Computers & Industrial Engineering, 160*, Article 107560.

Yang, X., Zhang, T., Bai, L., Shi, Y., & Chen, W. (2018). Appointment scheduling and routing problem of Community-Home-Health-Care: Based on modified Ant-Colony Algorithm. *Systems Engineering Theory & Practice*.

Yuan, B., Liu, R., & Jiang, Z. (2015). A branch-and-price algorithm for the home health care scheduling and routing problem with stochastic service times and skill requirements. *International Journal of Production Research, 53*(24), 7450–7464.

Zhang, T., Yang, X., Chen, Q., Bai, L., & Chen, W. (2018). Modified ACO for home health care scheduling and routing problem in Chinese communities. *Paper presented at the Networking, Sensing and Control (ICNSC), 2018 IEEE 15th International Conference on*.

Zhang, T., Yang, X., Yu, A., & Wang, X. (2019). Home health care routing problem via off-line learning and neural network. *Procedia CIRP, 83*, 193–197.

Zhong, S., Cheng, R., Jiang, Y., Wang, Z., Larsen, A., & Nielsen, O. A. (2020). Risk-averse optimisation of disaster relief facility location and vehicle routing under stochastic demand. *Transportation Research Part E: Logistics and Transportation Review, 141*. https://doi.org/10.1016/j.tre.2020.102015

Zohal, M., & Soleimani, H. (2016). Developing an ant colony approach for green closed-loop supply chain network design: A case study in gold industry. *Journal of Cleaner Production, 133*, 314–337.