Annals of Operations Research (2020) 291:499–529
https://doi.org/10.1007/s10479-018-3078-6

S.I.: REALCASEOR

A comprehensive multi-objective mixed integer nonlinear programming model for an integrated elderly care service districting problem

Meiyan Lin[1,3] · Kwai Sang Chin[2] · Lijun Ma[1,3] · Kwok Leung Tsui[2]

Published online: 16 October 2018

Abstract
The integrated care service districting (ICSD) problem is an important logistics decision that the elderly care structures (ECS) face when designing service networks to deliver integrated care to the elderly. The ICSD problem, which aims to prepare enhanced care worker recruitment and training plans for all well-designed service districts, is formulated as a multi-objectives mixed integer nonlinear programming (MOMINLP) model. Several criteria are considered, such as balanced workload of care workers among districts, compactness, indivisibility of elderly locations, and the unknown number of districts to be designed. The model considers three objectives simultaneously, including minimizing the total cost of hiring care workers necessary in all service districts, balancing the workload among districts, and achieving as much compactness of district as possible. Results for analysis were obtained by nondominated sorting genetic algorithm II, a well-known multi-objective evolutionary algorithm for continuous multi-objective optimization, which was modified for our MOMINLP model and tested with actual case. Effects of key parameters, including district- and service-related parameters, on these three objectives were analyzed based on different concerns from decision-makers. Furthermore, different correlations among the deviation of service workload and policies for work encouragement were analyzed for ECS. It informs decision-makers about the performance of key factors of the ICSD problem and improves service quality with proper decisions on related parameters.

Keywords Integrated care service districting problem · Multi-objective optimization · NSGA-II · Mixed-integer nonlinear programming · Care worker allocation problem · Integrated elderly care

Abbreviations

ICSD	Integrated care service districting
ECS	Elderly care structure
MOMINLP	Multi-objective mixed integer nonlinear programming

✉ Lijun Ma
ljma@szu.edu.cn

Extended author information available on the last page of the article

MOEA	Multi-objective evolutionary algorithm
NSGA-II	Nondominated sorting genetic algorithm II
TSP	Traveling salesman problem
SBX	Simulated binary crossover
VRP	Vehicle routing problem
PN	Size of population in NSGA-II
GN	Number of generations of NSGA-II

1 Introduction

According to data from world population prospects (United Nations 2017), the global population aged 60 and above is expected to more than double by 2050, rising from 962 million in 2017 to 2.1 billion in 2050. This population is growing faster than all young age groups worldwide. According to a report by Urban Development in China (2015), the situation in China is particularly severe: the population aged 60 and above is expected to increase from 212 million to 483 million by 2050. Specifically, the aging population with disabilities will triple by 2015, increasing from 6.25 million in 2015 to 18.75 million in 2050, and the population aged 80 and above, which stood at approximately 29.68 million in 2015, will increase by 1 million annually over the next 35 years. In addition, China was home to approximately 100 million empty nesters in 2015. With such an aged society, a boom in demand for integrated care from the elderly is expected in the future. Efforts are already being made to meet this demand. The Hong Kong government has increased the value of its annual Elderly Health Care Voucher to HK$2000 per elderly person (aged 65 and above) to encourage the elderly to choose the most suitable integrated care services and establish close relationships with care workers who are familiar with their health conditions (Gov 2018). However, China is far from ready to serve such a large elderly population. As only 0.3 million care workers (of whom only 40,000 are qualified) operate in elderly care structures (ECSs), China lacks 9 million care workers according to global standards (Xinhuanet 2018).

Accordingly, to meet the increasing demand for integrated care (i.e., personal, residential nursing, rehabilitation, and continuous care), ECSs in China are seeking to optimize their activities by, for example, developing recruitment and training programs for care workers (associated with various certificates and skills). Such programs require managers to improve their knowledge of the minimum cost of hiring the care workers needed in service districts. China's ECSs need to solve the care worker allocation problem through service district design. In addition, as a long-term planning strategy, the allocation of care workers and the grouping of the elderly provide the basis for assignment problems (Lin et al. 2016) and scheduling and routing problems (Maya Duque et al. 2015; Pureza et al. 2012). Due to its critical role in guiding long-term recruitment and training planning and as a basis for short-term planning and daily operations, we focus on an integrated care service districting (ICSD) and care worker allocation problem in this paper. Generally, service districts are designed to group elderly customers from different locations into a number of good districts, satisfying geometrical, activity-measured, and cost-related criteria (Jarrah and Bard 2012).

As some elderly customers prefer to receive integrated care at home, whereas some choose to be served in ECSs, China's ECSs adopt two delivery strategies. They either send care workers with matched skills to visit elderly people in their homes or pick up the elderly from designated parking sites in service districts and take them to their care centers to receive services from care workers. Regardless of elderly customers' choice of strategy, service districts must be as compact as possible to guarantee the delivery of care services within 30 min. Compactness is therefore a critical criterion for a good district. A compact district reduces care workers' total travel time and facilitates measurement of the reactivity of the service system, and thereby improves the efficiency of the service system (Benzarti 2012).

Once districts have been designed, the workload, which is composed of the care workload and the travel time in each service district, is determined. Each district is assigned to a team of care workers of several types. A district is well designed and perfectly balanced if the relative deviation of workload in a district from the average workload equals zero. Workload balance, which is related to the need to fairly distribute workload among care worker teams (each in charge of one district), is another important criterion for a good district. This criterion is applied to avoid large discrepancies in workload between districts. Workload imbalance can lead to care worker dissatisfaction and an overall decrease in the morale of ECSs, and thus creates considerable obstacles to retaining the care workforce.

Krause et al. (2005) summarize the compelling evidence that low-wage jobs in hospitality and housekeeping result in a large burden of illness, injury, and disability. Such low-wage jobs are performed by a predominantly female workforce, and are associated with repetitive physical tasks, low job control, low wages, the increasing use of contingency employment, and few opportunities for career advancement. Unfortunately, jobs in ECSs also have these characteristics. A large workload accompanied by low wages can lead to dissatisfaction and high turnover among care workers. To encourage care workers to work hard and stay in their positions, ECSs must develop policies that positively correlate wage cost and workload for care workers. Therefore, hiring care workers at the minimum cost while implementing encouragement policies is another important criterion for establishing good districts.

Despite the importance of these criteria, as explained above, only one or two have been considered as objective to date in the literature. To fill this gap, all three criteria are considered as objectives in the ICSD problem under study here. The aim of the problem is to design good districts that cover all of the care demands of basic units while considering critical criteria such as compactness/compatibility, balanced district workload, and the indivisibility of basic units. The diversity of care workers and the variety of organizational decisions involved in the integrated care delivery process are also considered in the ICSD problem. The main contributions of this study are as follows.

1. To formulate the ICSD problem, a comprehensive multi-objective mixed-integer nonlinear programming (MOMINLP) model is proposed for various ECSs with diverse sizes, aims, and operational criteria. This problem has three objectives: ensuring a rapid rate of response to requests from elderly customers, achieving a good workload balance across districts, and finding the minimum cost of hiring care workers. Trade-off analysis of these three objectives provides ECS managers with visual choices and helps them to make proper decisions.
2. Insights into the effects of key parameters, comprising district- and service-related parameters, on these three objectives. Analysis of district-related parameters serves as a reference helping ECSs to make decisions that ensure the satisfactory performance of the ICSD model (i.e., low cost, a good workload balance, and optimal district compactness). Analysis of service-related parameters offers guidelines for establishing proper correlations between deviations in service workload and compares the effectiveness of policies encouraging care workers to work hard and thereby perform better.
3. The nondominated sorting genetic algorithm-II (NSGA-II) (Deb et al. 2002), a well-known multi-objective evolutionary algorithm (MOEA), is modified to solve the MOMINLP model. The model creates two challenges for NSGA-II: handling nonlin-

ear constraints and handling (mixed) integer optimization. NSGA-II must be modified using a specific technique to handle these challenges.

This paper is organized as follows. Section 2 presents a survey of the related literature. Section 3 describes the ICSD problem and formulates it as an MOMINLP model, followed by model analysis. Section 4 presents the version of NSGA-II modified for the MOMINLP model. Section 5 discusses the effects of key parameters on the objectives. Section 6 draws conclusions, outlines the study's limitations, and suggests directions for future research.

2 Literature review

This section surveys the literature in two categories: operations research literature on resource allocation and districting problems; and literature on methods of solving districting problems, resource allocation problems, and multi-objective optimization problems.

2.1 Districting problems and resource allocation problems

The ICSD problem specified above resembles the service districting problem (Kalcsics 2015), and specifically the healthcare districting problem (Benzarti et al. 2013; Blais et al. 2003; Gutiérrez-Gutiérrez and Vidal 2015; Pezzella et al. 1981), which share some features with the sales territory and distribution districting problem (Benzarti et al. 2013; Konur and Geunes 2016; Lin et al. 2017; Moreno et al. 2017). A sales territory problem is designed to group sales coverage units into districts based on salesmen's responsibility, using criteria such as approximately similar numbers of customers/sizes of workload and the maximum total compactness of districts (Hess and Samuels 1971; Shanker et al. 1975; Zoltners and Sinha 2005). The design of the pick-up and delivery districting problem in the field of logistics considers criteria such as the contiguity and compactness of districts and the degree of workload balance, which consists of the time taken to deliver services to customers and the estimated average travel time within the district and to a centralized depot (Bard and Jarrah 2009; Haugland et al. 2007; Jarrah and Bard 2012). In contrast with measures of activity (number of sales calls, sales potential, customer count, etc.) in the sales territory problem (Zoltners and Sinha 2005), workload is defined in the ICSD problem as consisting of care load and travel load. Unlike the pick-up and delivery service districting problem, which is designed to minimize dissimilarity and the number of districts simultaneously (Bard and Jarrah 2009), the allocation of care workers to districts is integrated into the ICSD problem with the minimization of the cost of hiring care workers, the balancing of workload across districts, and the maximization of district compactness.

The criteria used in the healthcare service districting problem, such as the indivisibility of basic units, respect for borough boundaries, contiguity, compactness, and compatibility, are posed as hard constraints (Benzarti et al. 2013; Blais et al. 2003; Gutiérrez-Gutiérrez and Vidal 2015; Lin et al. 2017). The criteria related to workload equilibrium and personnel mobility are formulated as either constraints or objectives (Benzarti et al. 2013; Blais et al. 2003; Gutiérrez-Gutiérrez and Vidal 2015). Gutiérrez-Gutiérrez and Vidal (2015) consider the diversity of patients and care workers. Lin et al. (2017) focus on one type of service and treat care workers as homogenous. Instead of establishing a fixed number of districts (Benzarti et al. 2013; Blais et al. 2003; Gutiérrez-Gutiérrez and Vidal 2015), Lin et al. (2017) minimize the number of districts created. Unlike Lin et al. (2017), who address only a Meals on Wheels (MOW) service, we formulate an ICSD problem considering diverse care services, various elderly profiles, and several types of care workers. As an extension of the home health care districting problem (Benzarti et al. 2013), which minimizes workload imbalance or total travel distance, the ICSD problem is designed to minimize the cost of hiring care workers across well-designed districts, achieve the best workload balance across districts, and maximize the compactness of districts. In addition, criteria such as compactness, workload equilibrium, and minimum cost are formulated not as constraints or weight-sum objectives, as in related literature, but as multi-objectives. Finally, we analyze the effects of district-related parameters (i.e., the number of districts to design, the threshold for workload imbalance across districts, and the maximum travel duration in any district) on all of the objectives of the ICSD problem. These effects have received little attention in previous works but provide important insights for ECS managers.

Good districts with balanced resource allocation are the basis of good tactical and operational performance. They play a vital role in achieving integrated care objectives by improving delivery efficiency and effectiveness. In a case study, Boldy and Howell (1980) allocate a given amount of home help resources to several geographical areas covered by a County Social Services Department. Guided by the basic principle that with all else equal, everyone should have the same opportunity to receive a given amount of home help regardless of where they live, the authors describe an approach based on the integration of data on clients, population, and the provision of related services.

More specifically, the allocation of care workers can be classified as a human resource allocation problem (Bouajaja and Dridi 2017), which is the process of allocating human resources to various projects or business units to maximize profit or minimize costs (Lin and Gen 2008). The specific personality traits and abilities of human resources are important to cooperative projects. Skill satisfaction level, career path advancement, and motivation concerns were also addressed by Yoshimura et al. (2006). Kang et al. (2011) optimize the scheduling of human resource allocation for a software development project by considering individual constraints (i.e., familiarity with tasks and levels of productivity) and team-level constraints (i.e., team cohesion, communication overhead, and collaboration and management). Lepak and Snell (1999) develop a human resource architecture with four employment modes (internal development, acquisition, contracting, and alliance) for allocating work, which could be expanded to incorporate other capability assessment criteria such as project management, social and personal skills, and workplace preferences (Otero et al. 2009). Unlike previous versions of the human resource allocation problem, the ICSD problem proposed here integrates and describes all of the personality traits and abilities of care workers as individual-level constraints, without considering the team-level constraints addressed by Kang et al. (2011). Skill requirements, employment modes, and capacities are treated as parameters related to services in the proposed model, and the effects of correlations between these parameters on the objectives of the ICSD problem are discussed to provide insights for ECS managers. Such effects have rarely been discussed in the literature.

2.2 Methods of solving multi-objective district design and resource allocation problems

Unlike previous literature, this study builds an integrated MOMINLP model for the ICSD problem that considers the service district design and care worker allocation. In practice, the districting problem is usually modeled as a (variant of the) capacitated clustering problem (Chaves and Nogueira Lorena 2011; Jarrah and Bard 2012; Negreiros and Palhano 2006) and solved by exact methods utilizing mathematical models (Ansari et al. 2015), such as location allocation and set partitioning (Jarrah and Bard 2012), or by (meta)heuristic methods (Ríos-

Mercado and Fernández 2009; Salazar-Aguilar et al. 2012). The human resource allocation problem can also be solved by mathematical methods (Otero et al. 2009) or (meta)heuristic methods (Kang et al. 2011; Lin and Gen 2008). These metaheuristic methods include evolutionary algorithms (e.g., genetic algorithms (GAs), differential evolution algorithms, and harmony search algorithms), swarm intelligence-based algorithms (e.g., particle swarm optimization, ant colony optimization, bee algorithms, firefly algorithms, cuckoo search algorithms, and bat algorithms), and other algorithms (simulated annealing, tabu search, variable neighborhood search, etc.). The main means of dealing with multi-objective optimization problems (MOPs) are classified by Cho et al. (2017) into classical methods based on scalarization techniques [e.g., weighted sum and ε-constraint approaches (Dulebenets 2018); goal programming (Minella et al. 2008)]; (hybrid) metaheuristic methods (MOEAs, multi-objective swarm intelligence based methods, and other metaheuristics); and trust-based algorithms. Using metaheuristics to solve MOPs can overcome many difficulties, such as infeasible areas, local fronts, diversity of solutions, and the isolation of the optimum, and ensure that optimization tasks are completed within a reasonable computational time (Deb 2012).

Clearly, therefore, metaheuristic methods can be used to solve districting problems, resource allocation problems, and multi-objective problems. According to Cho et al. (2017), MOEAs and swarm intelligence based methods are widely regarded as powerful tools and the most appropriate means of solving problems that integrate the above three types of problem, such as this study's ICSD problem with three objectives. The most commonly used MOEAs are NSGA-II (Vanneschi et al. 2017) and its variants (Steiner Neto et al. 2017), MOEAs based on decomposition (MOEA/Ds) (Wang et al. 2017), and multi-objective differential evolution (MODE) (Zhou et al. 2018). Multi-objective swarm intelligence based methods such as multi-objective ant colony optimization (MOACO) (Ariyasingha and Fernando 2015) and multi-objective particle swarm optimization (MOPSO) (Zhang et al. 2018) are also widely used.

NSGA-II is popularly used to solve optimization problems with multiple objectives (Meng et al. 2017), due to its ability to find multiple tradeoff solutions in a single simulation run and its ease and flexibility in focusing on the Pareto-optimal frontier (Jain and Deb 2014). It adopts a crowding distance comparison process to preserve the explicit diversity of Pareto-optimal solutions, and a fast nondominated sorting procedure with $\mathrm{O}\left(\mathrm{M}N^2\right)$ computational complexity to conserve the Pareto-optimal solutions obtained. MOEA/D is shown by Wang et al. (2017) to have excellent features, such as a good search ability, a high level of compatibility with local searches, and a high level of computational efficiency (complexity of $\mathrm{O}(\mathrm{M}NT)$). It decomposes an MOP into a set of single-objective problems that are defined by weighted scalarizing methods (i.e., the weighted sum, penalty-based boundary intersection, and Chebyshev methods) and simultaneously optimized by an evolutionary algorithm. However, these decomposition approaches may perform poorly with some problems. MODE is a relatively novel evolutionary method, and its variants have elitism strategies to improve convergence, such as "best" individual selection, an immediate replacement mechanism, a non-dominated base vector mutation operator, and jumping genes. However, over-exploitation may result in poor performance in solving certain problems. MOACO is becoming increasingly popular as a means of solving dynamic applications, such as finding routes under dynamically changing network topologies [i.e., the traveling salesman problem (Ariyasingha and Fernando 2015)]. The cooperative behavior of ants not only facilitates the discovery of good solutions but also avoids premature convergence to the final solutions. However, convergence to the optimal solutions may be a time-consuming process, even with lower $\mathrm{O}(\mathrm{M}N\log(N))$ computational complexity. MOPSO is extensively applied to MOPs because it produces high-quality solutions in relatively little time (computational complexity of $\mathrm{O}\left(\mathrm{M}N^2\right)$) and displays more stable convergence behavior. However, the personal and global best particles have to be defined properly. These state-of-the-art algorithms have distinct advantages and disadvantages, and all perform well in solving MOPs. Researchers have compared the performance of NSGA-II with that of many MOEAs and swarm intelligence-based methods using test suites or specific real-world problems, and have recognized NSGA-II as one of the most successful tools.

Many MOPs are subject to constraints such as inequality, equality, linearity, and nonlinearity. Constraint handling approaches affect the performance of the state-of-the-art algorithms described above (Asafuddoula et al. 2012; Deb et al. 2002). Studies have shown that using a tournament selection-based algorithm with a population approach performs much better in handling constraints than a number of other existing constraint-handling approaches, such as the rejection of infeasible solutions, penalty functions and their variants, repair methods, decoders, the separate treatment of constraints and objectives, and hybrid methods incorporating knowledge of constraint satisfaction (Brownlee and Wright 2015; Deb 2000; Ray et al. 2001). NSGA-II has been shown to be an excellent method of handling constraints using the definitions of constrained domination and binary tournament selection (Brownlee and Wright 2015). Many MOPs, in which decision variables may have integer as well as real values, require the solution representation (chromosome, particle, etc.) of the above state-of-the-art algorithms to be properly encoded (Yu et al. 2017), and the operators of these algorithms to be properly modified (Subtil et al. 2010). To solve integer and mixed integer constrained optimization problems, a special truncation procedure is incorporated to handle integer restriction on decision variables in a real-coded GA (Deep et al. 2009). The proposed algorithm outperforms others in most of the test cases. NSGA-II with this special truncation procedure is used to solve the ICSD problem proposed in this study due to its excellent performance in solving constrained mixed integer MOPs.

In summary, the ICSD problem is an extension of related problems in literature with some similarity and several differences, and is differently formulated as multi-objective optimization model—MOMINLP model. The effects of parameters provide insights for ECS managers but have received little attention in previous works. NSGA-II is modified to solve the MOMINLP model because of its excellence in solving multi-objective optimization problems and handling constraints, and its flexibility of handling integer and mixed integer variables.

3 Modeling the ICSD problem

ECSs must ensure service quality for the benefit of both the elderly and the care workers. To this end, good districts should first be designed to facilitate the improvement of service quality for the benefit of customers and care workers. The effective grouping of the elderly within a district increases the reactivity of care workers (e.g., in emergencies), which in turn increases customer satisfaction. In addition, districts should be designed such that travel time is reduced, improving the efficiency of the integrated care delivery process. In this section, we propose an MOMINLP model for the ICSD problem.

3.1 Problem definition

Given a certain service territory and set of elderly customers, the ICSD problem is designed to group the locations of the elderly into good districts according to relevant criteria.

3.1.1 Assumptions

Without a loss of generality, we make the following assumptions, which are consistent with existing literature, to simplify the real case.

1. Elderly people with the same profile are homogeneous in terms of care requirements and service demand (i.e., number of visits and average duration of visits).
2. Care workers in the same category are homogeneous in terms of skills, contracts, and workload capacity.
3. All basic units are covered, i.e., every elderly customer admitted must be assigned to a district.

3.1.2 Criteria

The following criteria are used to effectively capture the operational requirements of the districting problem.

(a) Compactness, which can be formulated as a hard constraint by limiting the maximum travel duration between any two basic units in the same district, minimizes the ratio of total travel time to total working time. It is critical to the design of good districts.
(b) The indivisibility of basic units; each unit may be assigned to only one district. This criterion is considered to avoid overlap between care workers' responsibilities and enable them to establish long-term relationships with customers.
(c) The number of care workers in charge of each district.
(d) Workload balance. Having almost the same workload as each other is essential for good districts.

3.1.3 Notations

The parameters and decision variables in the MOMINLP model of the ICSD problem are presented in Table 1.

3.2 The MOMINLP model

Following the notation given above, the model is formulated as follows. It can be used when a manager wants to allocate the minimum number of care workers of each type to each district while considering multiple concerns (i.e., minimum cost, good workload balance, and optimal district compactness).

$$\text{Minimize} \quad O_{min} \tag{1}$$

$$\delta_{max} \tag{2}$$

$$d_{max} \tag{3}$$

$$\text{subject to} \quad m = \sum_{k=1}^{M} y_k \tag{4}$$

$$\mathrm{O}_{min} = \sum_{k=1}^{M} \sum_{s=1}^{S} C_s \gamma_{sk} \tag{5}$$

Table 1 Notation for the ICSD problem

Parameters	
$\boldsymbol{N}$	Total number of basic units in service area
$\boldsymbol{M}$	Upper bound on number of districts to design, $M \leq N$
$\boldsymbol{H}$	Number of elderly profiles considered
$\boldsymbol{S}$	Number of care worker types considered
$\boldsymbol{p_{ih}}$	Number of elderly customer with profile h ($h = 1\ldots H$) in basic unit i ($i = 1\ldots N$)
$\boldsymbol{b_{hs}}$	Number of visits required by elderly person with profile h ($h = 1\ldots$H) and achieved by care worker of type s
$\boldsymbol{t_{hs}}$	Average service duration of visit achieved by care worker of type s ($s = 1\ldots$S) related to profile h ($h = 1\ldots$H)
$\boldsymbol{\tau_i}$	Average estimated travel duration from basic unit i ($i = 1\ldots N$) to next basic unit
$\boldsymbol{\delta_s}$	Admissible percentage deviation in workload of care worker of type s ($s = 1\ldots$S) associated with a given district from average workload across districts, $\boldsymbol{\delta_s} \in [0, 1]$
$\boldsymbol{Q_s}$	Workload (capacity) of care worker of type s ($s = 1\ldots$S), measured in certain time unit (i.e., hour)
$\boldsymbol{Q}$	Baseline of workload (capacity) in contract
$\boldsymbol{C_s}$	Ratio of cost of hiring care worker of type s ($s = 1\ldots$S) to baseline of workload $\boldsymbol{Q}$, such that when $\boldsymbol{Q_s}$ equals $\boldsymbol{Q}$, $\boldsymbol{C_s} = 1$, and increasing $\boldsymbol{Q_s}$ increases $\boldsymbol{C_s}$ according to a certain rule
$\boldsymbol{d^c}$	Maximum travel duration allowed between two basic units in the same district
$\boldsymbol{d_{ij}}$	Travel duration between basic units i and j, assuming that travel between two basic units is achieved by certain transportation mode
$\boldsymbol{d}$	Maximum travel duration between any two basic units, $\boldsymbol{d} = \boldsymbol{max}\{\boldsymbol{d_{ij}} \mid \forall \boldsymbol{i}, \boldsymbol{j} = 1 \ldots \boldsymbol{N}\}$
$\boldsymbol{A}$	Set of pair of basic units (i, j), where $a_{ij} = 0$. $a_{ij} = a_{ji} = 1$ if basic units i and j are compatible; otherwise $a_{ij} = a_{ji} = 0$,where d_{ij} is a large enough value ($d_{ij} > \boldsymbol{d^c}$)
β_{is}	$\beta_{is} = \sum_{h=1}^{H} p_{ih}(t_{hs} + \tau_i)b_{hs}$, the total workload of care workers of type s in basic unit i.
ε	$\varepsilon \ll 10^{-6}$
Decision variables	
$\boldsymbol{x_{ik}}$	$= \begin{cases} 1\, if\, the\, basic\, unit\, i\, (i = 1 \ldots N)\, is\, assigned\, to\, district\, k\, (k = 1 \ldots M), \\ 0\, otherwise \end{cases}$
$\boldsymbol{y_k}$	$= \begin{cases} 1 & if\ district\ k\ (k = 1 \ldots M)\ is\ selected, \\ 0 & otherwise \end{cases}$
$\boldsymbol{m}$	Integer decision variable (number of districts designed)
$\boldsymbol{\gamma_{sk}}$	Integer decision variable (number of care workers of type s in district k)
$\boldsymbol{\delta_{max}}$	Maximum percentage deviation of care workload associated with each district from average care workload across districts
$\boldsymbol{d_{max}}$	Maximum travel duration between two basic units in same district across districts
$\boldsymbol{O_{min}}$	Minimum cost of hiring care workers needed in all districts

$$\sum_{k=1}^{M} x_{ik} = 1, \quad \forall i = 1, 2, \ldots, N \tag{6}$$

$$x_{ik} + x_{jk} \leq y_k, \quad \forall (i, j) \in A, \quad \forall k = 1, 2, \ldots, M \tag{7}$$

$$y_k(1 - \delta_s) \sum_{i=1}^{N} \beta_{is}/m \leq \sum_{i=1}^{N} \beta_{is} x_{ik}, \forall s = 1, 2, \ldots, S; \forall k = 1, 2, \ldots, M \tag{8}$$

$$\sum_{i=1}^{N} \beta_{is} x_{ik} \le y_k (1+\delta_s) \sum_{i=1}^{N} \beta_{is}/m, \quad \forall s = 1, 2, \ldots, S; \quad \forall k = 1, 2, \ldots, M \tag{9}$$

$$y_k (1-\delta\text{max}) \sum_{s=1}^{S} \sum_{i=1}^{N} \beta_{is}/m \le \sum_{s=1}^{S} \sum_{i=1}^{N} \beta_{is} x_{ik}, \quad \forall k = 1, 2, \ldots, M \tag{10}$$

$$\sum_{s=1}^{S} \sum_{i=1}^{N} \beta_{is} x_{ik} \le y_k (1+\delta\text{max}) \sum_{s=1}^{S} \sum_{i=1}^{N} \beta_{is}/m, \quad \forall k = 1, 2, \ldots, M \tag{11}$$

$$d\text{max} \ge d_{ij} \left(x_{ik} + x_{jk} - y_k \right), \quad \forall i, j = 1, 2, \ldots, N, \quad \forall k = 1, 2, \ldots, M \tag{12}$$

$$\gamma_{sk} \le \sum_{i=1}^{N} \beta_{is} x_{ik}/Q_s + 1 - \varepsilon, \quad \forall s = 1, 2, \ldots, S; \quad \forall k = 1, 2, \ldots, M \tag{13}$$

$$\sum_{i=1}^{N} \beta_{is} x_{ik}/Q_s \le \gamma_{sk}, \quad \forall s = 1, 2, \ldots, S; \quad \forall k = 1, 2, \ldots, M \tag{14}$$

$$x_{ik} \le y_k, \quad \forall i = 1, 2, \ldots, N; \quad \forall k = 1, 2, \ldots, M \tag{15}$$

$$x_{ik} \in \{0, 1\}; \quad \forall i = 1, 2, \ldots, N; \quad \forall k = 1, 2, \ldots, M \tag{16}$$

$$y_k \in \{0, 1\}, \quad \forall k = 1, 2, \ldots, M \tag{17}$$

$$m \in [1, \mathrm{M}] \tag{18}$$

$$\gamma_{sk}, \; d\text{max} \in \mathbb{Z}, \; O\text{min}, \; \delta\text{max} \in \mathbb{R} \tag{19}$$

3.2.1 Multiple objectives

The model focuses on determining the minimum cost of hiring care workers, the optimal fairness of workload across the districts designed, and the highest rate of response of service providers in districts with all demands covered and all criteria considered. The first objective function is formulated in Eq. (1), based on Criterion (c), to calculate the cost of hiring the minimum number of care workers. The second objective function minimizes the maximum deviation of the workload in any district from the average workload across districts according to Criterion (d), as indicated in Eq. (2), coupled with Constraints (10) and (11). The third objective function minimizes the maximum travel duration across districts, as indicated in Eq. (3) coupled with Constraint (12). This ensures that districts are compact, based on Criterion (a).

3.2.2 Constraints

Constraint (4) calculates the number of districts created. Constraint (5) derives the total cost of hiring the care workers needed in the districts designed. Constraint (6) ensures the individuality of basic units according to Criterion (b). Constraint (7) guarantees compatibility, as shown in Criterion (a), wherein the duration between two basic units assigned to the same district is bounded by the maximum travel duration d^c. Constraints (8) to (9) explain the minimum and maximum allowable workload of each type within each district, according to Criterion (d). Constraints (13) and (14) calculate the number of care workers of each type needed in each district. Constraint (15) identifies the availability of basic units for districts. Constraints (16) to (19) define the decision variables.

3.3 Model analysis

This section analyzes the model's Pareto-optimal set and key factors, and provides a detailed calculation of computational time.

3.3.1 Pareto-optimal set

The ICSD problem, formulated as an MOMINLP model, can be generalized as a minimization problem of the following kind.

Minimize f($\boldsymbol{x}$) $= (f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), f_3(\boldsymbol{x}))$ subject to constraints $g_i(\boldsymbol{x}) \geq 0, \quad \forall i = 1, \ldots, I;\ h_k(\boldsymbol{x}) = 0, \quad \forall k = 1, \ldots, K, x_i^L \leq x_i \leq x_i^U$, where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ is the vector of decision variables that belong to the parameter space and $f_j, j = 1, 2, 3$ are the three objective functions.

A decision vector $\boldsymbol{x}$ is said to constrain-dominate another decision vector y if any one of the following conditions is satisfied.

1. $\boldsymbol{x}$ is feasible and y is infeasible.
2. Both $\boldsymbol{x}$ and y are infeasible and $\boldsymbol{x}$ has a smaller constraint violation value, CV, where $CV(\boldsymbol{x}) = \sum_{\boldsymbol{i}}^{\boldsymbol{I}} g_i(\boldsymbol{x}) + \sum_{\boldsymbol{k}}^{\boldsymbol{K}} |h_k(\boldsymbol{x})|$ and the bracket operator α returns the negative of α if $\alpha < 0$ and otherwise returns 0.
3. Both $\boldsymbol{x}$ and y are feasible and $\boldsymbol{x}$ dominates y. Vector $\boldsymbol{x}$ dominates y *if and only if* $f_j(\boldsymbol{x}) \leq f_j(\boldsymbol{y}), \quad \forall j = 1, 2, 3, \quad \text{and} \quad \exists j \in \{1, 2, 3\} : f_j(\boldsymbol{x}) < f_j(\boldsymbol{y})$.

A non-dominated decision vector $\boldsymbol{x}$ is considered Pareto-optimal. A Pareto-optimal set is a set of Pareto-optimal decision vectors.

3.3.2 Key model factors

Here, q_s is the annual demand for the s-th type of service from the elderly from all basic units, w_{sk} is the total workload of the s-th type in the k-th district, and w_{total} is the demand for all services from the elderly from all basic units:

$$q_s = \sum_{i=1}^{N} \beta_{is}, \ \ w_{sk} = \sum_{i=1}^{N} \beta_{is} x_{ik}, \ \ w_{total} = \sum_{s=1}^{S} q_s.$$

(a) Model factors

To calculate the bounds, most constraints except those related to the workload limitation, such as Constraints (8), (9), (13), and (14), can be relaxed. Constraints (8) and (9) are combined in Eq. (20). Constraints (13) and (14) are combined in Eq. (20) to give Eq. (21). Therefore, the objective value of the proposed model should be satisfied by Eq. (22), as shown in Eq. (23). Equation (23) includes m, C_s, δ_s, Q_s, and S, which can influence the relaxed bounds of the model. Increasing S (the diversity of care workers) and m (the number of districts created) will increase the number of care workers needed and the related costs. The cost of hiring care workers is also affected by C_s (the ratio of the cost of hiring a care worker of type s to the workload baseline) and Q_s (the contract, related to workload capacity). Thus, heavy workloads should be associated with high salaries (incurring excessive costs) for encouragement. Equation (20) can be deduced to Eq. (24), and Constraints (10) and (11) can be deduced to Eq. (25). Equations (24) and (25) indicate that δ_s may affect workload balance across districts. Constraints (7) and (12) state that d^c may affect district compactness. In sum, the key parameters considered in the model are $S, m, C_s/Q_s, \delta_s,$ and d^c.

$$y_k(1-\delta_s)q_s/m \le w_{sk} \le y_k(1+\delta_s)q_s/m \tag{20}$$

$$y_k(1-\delta_s)q_s/(mQ_s) \le w_{sk}/Q_s \le r_{sk} \le w_{sk}/Q_s + 1 - \varepsilon \le y_k(1+\delta_s)q_s/(mQ_s) + 1 - \varepsilon \tag{21}$$

$$\sum_{k=1}^{m}\sum_{s=1}^{S} y_k C_s(1-\delta_s)q_s/(mQ_s) \le \sum_{k=1}^{m}\sum_{s=1}^{S} C_s r_{sk} \le \sum_{k=1}^{m}\sum_{s=1}^{S} C_s(y_k(1-\delta_s)q_s/(mQ_s) + 1 - \varepsilon) \tag{22}$$

$$\sum_{s=1}^{S}(1-\delta_s)C_s q_s/Q_s \le O_{min} \le \sum_{s=1}^{S}(1+\delta_s)C_s q_s/Q_s + m(1-\varepsilon)\sum_{s=1}^{S} C_s \tag{23}$$

$$\sum_{s=1}^{S}(1-\delta_s)q_s/m \le \sum_{s=1}^{S} w_{sk} \le \sum_{s=1}^{S}(1+\delta_s)q_s/m, \quad \forall y_k = 1 \tag{24}$$

$$(1-\delta_{max})\sum_{s=1}^{S} q_s/m \le \sum_{s=1}^{S} w_{sk} \le (1+\delta_{max})\sum_{s=1}^{S} q_s/m, \quad \forall y_k = 1 \tag{25}$$

(b) Correlation of factors

As mentioned above, C_s/Q_s and δ_s will affect the objectives. This study considers two types of care worker. The correlation between values of δ_s may be independent, positive, or negative. In the real world, ECS managers need the rules for setting δ_s to be as simple as possible. Linear correlation is the most popular and simplest way of describing the positive or negative correlation between δ_s. The correlation between δ_1 and δ_2 is therefore assumed to be linear or independent. When the correlation between δ_1 and δ_2 is linear, two equations are used: $\delta_1 = \theta\delta_2,\ 0 < \theta \le 1$ for positive correlation; and $\delta_1 = \alpha - \delta_2,\ 1 < \alpha \le 2$ for negative correlation. When the correlation is independent, both δ_1 and δ_2 are uniform random values ranging from 0 to 1.

Limits are set on the total workload of each care worker, which should not be greater than twice the baseline or smaller than half of the baseline. Therefore,$Q_s \in [0.5Q,\ 2Q]$. C_s/Q_s can be a constant gradient ϑ, or C_s can be modeled as a stepwise function of Q_s and Q, where s $= 1, 2$. As $C_s = 1$ when $Q_s = Q$, gradient $\vartheta = 1/Q$ when C_s/Q_s is modeled as linear. When C_s is modeled as a stepwise function of Q_s and workload range is divided into 10 pieces, that is, $C_s = (1+\rho)^{(l-4)}$, when $Q_s \in [0.5Q + (l-1) \times 0.15Q, 0.5Q + l \times 0.15Q]$, $\rho \in (0, 0.2]$ is a constant indicating a bonus for hard work and $l \in [1, 10]$. To exam different bonus rules (linear and stepwise), proper levels (pieces) of workload should be set. Too less levels may not illustrate the trends or difference, too many levels will increase the computational time taken for this experiment. Besides, 10 levels (pieces) is easy for workload to be calculated and simple for ECS managers. Therefore, 10 levels of workload are enough to illustrate the trends or tell the difference between different bonus rules. A large ρ means a large bonus for working hard, increasing the total cost of hiring care workers.

As different types of correlation between these factors may affect the objectives, it is necessary to discuss the effects of these correlations to offer insights for decision makers.

3.3.3 Computational time

With the increasing scale of the instance under study, the computational time required to solve the model above increases expectedly. First, the model is subject to $M \times |A|$ compatibility constraints and $M \times S$ or N other constraints. In addition, $M \times N \times N$ constraints are placed on the calculation of the maximum travel duration between any two basic units in any district. The time taken to compute the optimal solution increases with an increase in N, M, and S. Second, the decision variable of district selection y_k has C_M^m alternative optima if the number of districts created is m. In addition, the decision variable of basic unit assignment x_{ik} has at least $(m-1)!$ alternative optima, due to symmetry in branch and bound. The proposed model can be relaxed by properly selecting values of d^c, which may reduce computational time. For instance, when the values of d^c are sufficiently large, computational time will be long due to the absence of pairs in the non-compatibility set ($|A| = 0$). As m is the decision variable in the model and is determined by the selection of districts, it will increase complexity and computational time.

Given the characteristics of the proposed model in terms of objective functions and specific constraints, it is computationally intensive to find the optimal solutions for the model using exact methods such as the branch-and-bound algorithm or branch-and-cut techniques. Therefore, evolutionary optimization approaches based on Pareto optimality are used to solve the model.

4 Modified NSGA-II with constraint handling to solve the MOMINLP model

Commercial optimizers such as Gurobi may be a feasible means of solving the MOMINLP model with blended and hierarchical approaches (Gurobi 2018). However, it tends to be difficult for such optimizers to find the non-dominated set within an acceptable computational time due to the slow convergence rate of exact methods using the branch-and-bound algorithm. Given these challenges, this paper focuses on MOEAs based on Pareto optimality, such as NSGA-II (Deb et al. 2002), which show excellent performance in solving multi-objective optimization problems (Jain and Deb 2014).

In this paper, NSGA-II (Deb et al. 2002) with a truncation procedure (Deep et al. 2009) is used to handle the constraints on and mixed integer decision variables in the MOMINLP model to provide comprehensive solutions. As its overall complexity is $\mathrm{O}(\mathrm{M}N^2)$, NSGA-II can provide good enough computational efficiency. In addition, NSGA-II (Deb et al. 2002) involves fast non-dominated sorting that emphasizes elitism and preserves the best individuals to the current generation, and crowding distance assignment promotes solution diversity. Deb et al. (2002) handle constraints using the definitions of constrained domination and binary tournament selection. The performance of this method is proven to surpass that of other constraint-handling approaches (Deb 2000). The truncation procedure (Deep et al. 2009) used here has been shown to deal effectively with the constraints of integer or mixed integer decision variables. The details are presented in the following subsections.

4.1 Main loop of NSGA-II

The main loop of the modified NSGA-II is run 20 times. Initially, a random parent population P_0 of size PN is created with consideration of box constraints. The truncation procedure described in Sect. 4.1.1 is applied to P_0 to satisfy the integer restrictions. The objective function values of the truncated population are then calculated. At the same time, the constraint violation values are calculated and normalized. Based on the objective values and constraint violation values, the population is sorted based on the principle of constrained domination (described in Sect. 4.1.2). An elitism procedure (outlined in Sect. 4.1.5) is then applied. This procedure also produces *GN* generations of the population.

The *t*th generation of the modified algorithm occurs as follows. Modified tournament selection (described in Sect. 4.1.3), genetic operators (described in Sect. 4.1.4), and a truncation operator are used to create an offspring population O_t of size *PN*. The combined population $R_t = P_t \cup O_t$ is formed. Next,R_t is sorted based on the principle of constrained domination. A population P_{t+1} of size *PN* is formed by applying the elitism procedure to the sorted R_t. This procedure is used for selection, genetic operation, and truncation to create a new population O_{t+1}. At the end of the *GN*-th generation, the non-domination set in the first front is stored. After 20 iterations of the main loop, all non-domination sets stored are combined to form a new population. The new population is sorted based on constrained domination to obtain the non-domination set in the first front (the best one), which is considered the Pareto-optimal set and used in the analysis described in Sect. 5. The key procedures are briefly described in the following subsections, with greater detail provided by Deb et al. (2002).

4.1.1 Truncation procedure for integer restriction

The population is truncated to satisfy the box constraints and the integer restrictions as follows. First, for $\forall i \in [1, N]$, x_i is truncated to x_i^U or x_i^L if x_i is above x_i^U or below x_i^L, respectively. Second, for $\forall i \in [1, N]$, x_i is truncated to integer value $x_i^{'}$ according to the following rules: $x_i^{'} = x_i$ if x_i is an integer, or $x_i^{'}$ is equal to either $[x_i]$ or $[x_i]$ +1, each with probability 0.5 ($[x_i]$ is the integer part of x_i). Deep et al. (2009) use truncation only to handle the integer restrictions, not to satisfy the box constraints. They claim that "the truncation procedure ensures greater randomness in the set of solutions being generated and avoids the possibility of the same integer values being generated" when "a real value lying between the same two consecutive integers is truncated."

4.1.2 Constrained-domination sorting

The population can be sorted into different non-domination levels according to the constraint-domination principle described in Sect. 3.3.1. First, the individuals in the population are separated into two groups according to the constraint violation value CV. In one group, all individuals are feasible (CV=0); in the other, all individuals are infeasible (CV>0). The fast non-dominated sorting approach and the crowding distance method developed by Deb et al. (2002) are applied to the feasible group. All feasible individuals are ranked according to their non-domination level based on the objective function values. Every feasible individual has a higher non-domination rank than every infeasible individual. All infeasible individuals are ranked according to their non-domination level based on the constraint violation values (CVs). Therefore, an individual with a smaller CV has a higher rank. After the various sorting procedures have been applied to these two groups, the sorted feasible individuals and the sorted infeasible individuals are combined to give the sorted population.

4.1.3 Modified tournament selection

Binary tournament selection is modified to favor a feasible individual over an infeasible individual and an individual with a smaller CV over an individual with a larger CV. Two individuals (decision vectors) from P_t are randomly selected, and the modified binary tournament selection procedure is applied to select the better of individuals $\boldsymbol{x}$ and $\boldsymbol{y}$, defined as follows.

1. If the rank of $\boldsymbol{x}$ is not equal to the rank of y, select $\boldsymbol{x}$ if the rank of x is higher than that of y, and select y if the rank of y is higher than that of $\boldsymbol{x}$. Three situations are possible: first, $\boldsymbol{x}$ and y are feasible individuals with different ranks (the feasible individual with the higher rank is selected); second, $\boldsymbol{x}$ and y are infeasible individuals with different CVs (the infeasible individual with the lower CV (higher rank) is selected); or third, one individual is feasible and the other is infeasible (the feasible individual is selected).
2. If the rank of $\boldsymbol{x}$ is equal to the rank of y, select $\boldsymbol{x}$ if the crowding distance of x is better than that of y, and select y if the crowding distance of y is better than that of $\boldsymbol{x}$; otherwise, randomly select x or y. Three situations are possible: first, $\boldsymbol{x}$ and y are feasible individuals with the same rank and different crowding distances (the feasible individual located in a less crowded region is selected); second, $\boldsymbol{x}$ and y are feasible individuals with the same rank and the same crowding distance (an individual is randomly selected); or third, $\boldsymbol{x}$ and y are infeasible individuals (an individual is randomly selected).

4.1.4 Genetic operators

Agrawal et al. (1995) show that real-coded GAs with a simulated binary crossover (SBX) operator and polynomial mutation performed better than binary-coded GAs in solving problems with a continuous search space. Therefore, this SBX operator, polynomial mutation, and the specific truncation operator for integer variables are used with the real-coded NSGA-II reported in this paper. In addition, Agrawal et al. (1995) suggest that GAs that combine binary coding for discrete variables with real coding for continuous variables can efficiently solve mixed integer programming problems. A variable is chosen for crossover. If it is a discrete variable, the single-point crossover or another binary operator can be used for the bits representing that variable only; if it is a continuous variable, the SBX operator proposed by Agrawal et al. (1995) can be used. This suggestion will be adopted in future work.

4.1.5 Elitism procedure

The population of the next generation, P_{t+1}, is formed by appending solutions in each non-dominated set from $\mathcal{F}_1$ to $\mathcal{F}_l$ consecutively until the population size ($|P_{t+1}|$) exceeds the current population size PN. Individuals belonging to the best non-dominated set $\mathcal{F}_1$ are the best individuals in the combined population R_t. If $\mathcal{F}_1$ is smaller than PN, all members (individuals) of the set $\mathcal{F}_1$ are added to the new population P_{t+1}. Members of the set $\mathcal{F}_2$ are added next, followed by members from the set $\mathcal{F}_3$, and so on. Suppose that when adding all members to the non-dominated set $\mathcal{F}_l$, the size of $\bigcup_{i=1}^{l} \mathcal{F}_i \left(\left|\bigcup_{i=1}^{l} \mathcal{F}_i\right|\right)$ exceeds PN for the first time. Then the required number ($PN - \left|\bigcup_{i=1}^{l-1} \mathcal{F}_i\right|$) of the remaining individuals is selected from $\mathcal{F}_l$ based on crowding distance, using the crowded-comparison operator (Deb et al. 2002). For more details, please refer to the original paper (Deb et al. 2002).

4.2 Sketch of modified NSGA-II for the MOMINLP model

As briefly described above, NSGA-II is modified and applied. The use of the modified NSGA-II to solve the MOMINLP model is sketched below.

Step 1. Input a set of choices for values of m (for example, m runs from 2 to 7 due to practical concerns), a set of choices for values of δ, and a set of choices for values of d^c. For each pair (m,δ,d^c), repeat Steps 2 to 4
Step 2. Set proper parameters for NSGA-II, namely the maximum number of generations, population size, running times, distribution indexes for crossover and mutation, and crossover and mutation probability
Step 3. Run the main loop of NSGA-II with fixed times to obtain the non-domination set in the first front
Step 4. Analyze the objective values of the solutions in the non-domination set obtained. The decision maker can decide to terminate the algorithm if the objective values of a certain solution in the current non-domination set suit their aims after analysis
Step 5. Store the selected solution, give the basic units in the same district the same color, and present the solution in a visual user interface

5 Discussion of experimental results

This section models the customer grouping and care worker allocation problem in a real case and analyzes the behavior of the proposed model. Following (Deb et al. 2002), an SBX operator and polynomial mutation with mutation probability $p_m = 1/n$ (where n is the number of decision variables) (Agrawal et al. 1995), along with the specific truncation procedure proposed by Deep et al. (2009), are used to implement NSGA-II. Distribution indexes are used as the crossover and mutation operators, with $\eta_c = 20$ and $\eta_m = 20$, respectively. After 20 runs, the population (size of 92) obtained at the end of 200 generations is used to discuss the effects of the key parameters on the three objectives. The following subsections describe the real case and data and discuss the results.

5.1 Real case and data

The model proposed in this study is developed for the Tai Po integrated care service structure of the Salvation Army (SA-TPIHCSS) in Hong Kong. The SA-TPIHCSS delivers several types of service, including elderly services, rehabilitation services, and education and employment services, to elderly customers living in Tai Po. Based on their residential addresses, the elderly customers are manually clustered into eight groups (named by location), such as Tai Wo Estate, Fu Heng Estate, Kwong Fuk Estate, Wan Tau Tong Estate, and Tai Po Market. The service needs of each group of customers are met by a team consisting of several care workers. However, the work of the ECS is impeded by a shortage of care workers and inefficient delivery procedures. Managers wish to determine the key factors affecting the performance (measured in terms of the cost of hiring care workers, workload balance across teams, and compactness of location) of customer grouping and care worker allocation decisions. Therefore, we build an MOMINLP model to model the customer grouping and care worker allocation problem (integrated as an ICSD problem in this study) and use NSGA-II to solve the model and provide insights for the SA-TIPIHCSS.

On admission to the structure, every elderly customer is classified into one of several profiles. The profile of a customer may be updated at the end of each year. Based on records of admission and discharge in 2012–2014 (for the sample, refer to "Appendix"), elderly customers remain within the structure for a long time, ranging from 1 year to 20 years. Three hundred and twenty elderly customers (cases) had been admitted by 2011. Discharge started in 2012. The absolute difference between the number admitted and the number discharged remained within 10% of 320 (number of cases by 2011) in 2012–2014. Customers may live in the same location, such as the same building/block/street. A location is denoted as a basic unit. Fifty-seven basic units covered the total number of active cases in 2014. As each basic unit is assigned to one district (as per Constraints (6) and (7) of the MOMINLP model), and elderly customers in one basic unit must be assigned to the same group. The distance and duration of travel between any two basic units are obtained from Google Maps. The average estimated travel duration from each basic unit to its next basic unit is calculated using the method described by (Jarrah and Bard 2011). The demand in a basic unit depends on the number of elderly customers, customer profiles, the type(s) of service required, and the frequency of visits needed. The type(s) of service and number of visits required may differ considerably between elderly customers with the same profile in a basic unit. In the model built to solve the ICSD problem, therefore, the calculation of demand is simplified by assuming that elderly customers with the same profile have the same demand in terms of total service time, although they may require different services. The number of customers with each profile in a basic unit is randomly generated considering the demand distribution among basic units. The structure then clusters the basic units into several districts based on the former's demand (as per Constraints (13) and (14), which also calculate the total number of care workers in each team for each group) and the criterion of travel distance/duration (as per Constraint 12).

A team consists of different types (grades) of care worker, such as senior care workers, care workers, care worker assistants, and part-time care worker assistants. Care workers are graded based on their skills, experience, and contracts. All types of care worker can visit elderly customers. Each visit takes 30–60 min (service duration). Care workers' working hours per working week may differ according to their contracts; care workers work for 40 h per week, for example, whereas part-time care worker assistants work for only 20 h per week. Each care worker takes a 1-hr lunch break (from 12:30 to 13:30; see Table 2 in the Appendix for daily work schedules). Care workers have only 1 day off for rest per week, and 7 days of annual vacation. Due to the nature of their job, at least one care worker asks for sick leave per week. Calculating the total available working time for each care worker per year is a complicated process. The available workload of each type of care worker is randomly generated between the minimum and maximum workload defined by the structure or by law. This raises the question of how different requirements for workload balance among care workers affect the performance of the solution to the ICSD problem. This concern is modeled by Constraints (8) to (11), and experiments are conducted with different value settings of δ_s to analyze the effects of δ_s on performance.

Based on the real data and the data generated, two experiments are conducted to analyze the effects of key parameters on three objectives. The first experiment is carried out to analyze the effects of parameters related to district design, namely the number of districts to be designed, m; deviation in workload balance, δ; and the maximum duration of travel between any two basic units in a district, d^c. This experiment considers two types of care worker with the same $\delta_s = \delta(0 < \delta \leq 1)$, i.e., $C_s = 1$ and Q_s is the baseline Q. The second experiment is conducted to analyze the effects of service-related parameters, namely δ_s (percentage deviation in service workload of type s), C_s (the ratio of the cost of hiring a care worker of type s to workload baseline), and the relationship between C_s and Q_s. The number of districts, m, and the maximum duration of travel between two basic units in a district, d^c, are fixed at certain values in the second experiment based on the analysis of the results of the first experiment. Two types of care worker with different values of δ_s and C_s are considered.

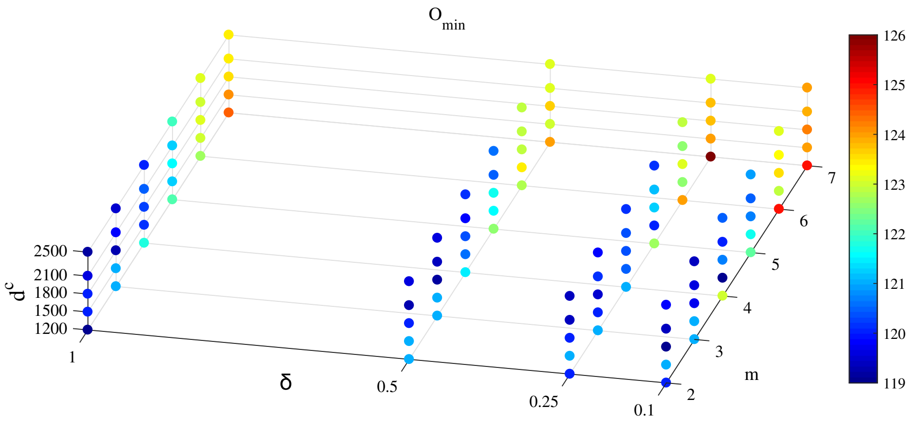

Fig. 1 Mean of O_{min} with all parameter settings

5.2 Discussion of the effects of district-related parameters (m, δ, d^c) on objectives

To analyze the effects of district-related parameters on model performance, we test different sets of key parameters (m, δ, d^c) for the problem. This part of the analysis and the parameter setting are mostly based on real data from the real case. This section analyzes the results obtained using the modified NSGA-II in the first experiment, in which we assume that two types of care worker are needed. They have the same $\delta_s = \delta(0 < \delta \leq 1); C_s = 1$, and Q_s is the baseline Q. To discuss the effect of m on performance, m is increased from 2 to $N/8$ ($N = 57$, $m = 8$ extant districts in the real case) by 1, and m districts are selected ($\mathrm{y_k} = 1, \quad \forall \mathrm{k} = 1 \ldots m$). The percentage deviation δ of service workload belongs to the set $\{100\%,\ 50\%,\ 25\%,\ 10\%\}$. As an excessive value of d^c may lead to a high probability of infeasibility, d^c is not smaller than $d/2$. The maximum travel duration between two basic units d^c in a district ranges from $d/2$ to d; that is, $d^c \in \{d/2, 5d/8, 3d/4, 7d/8, d\}$. When $d^c = d$, the compatibility constraints are relaxed because set A has no pairs of basic units.

5.2.1 Effects of m, δ, and d^c on O_{min} (total cost of hiring care workers)

In this subsection, the effects of key parameters on O_{min} are discussed with the assumptions that $\delta_s = \delta(0 < \delta \leq 1)$, $C_s = 1$, and $Q_s = Q$. O_{min} equals the number of care workers needed when $C_s = 1$. Figure 1 shows the average number of care workers, which is calculated from the objective values of the solutions in the first Pareto-optimal front obtained by NSGA-II. The x, y, and z axes represent m (the number of districts created), δ (the admissible percentage deviation of the workload of care worker of type s), and d^c (the maximum duration of travel between any two basic units in a district) respectively. As the value of m increases (plane of x = 2 to plane of x = 7 from the left to the right side of the x axis in Fig. 1), the color of the dots changes from dark blue to dark red. This change implies that O_{min} (the total number of care workers needed) increases with the number of districts designed, regardless of workload balance limitation or district compatibility. This finding has been tested in the real world in the ECS under study with reference to the MOW service delivery procedure (Lin et al. 2017), and with the Centre local de services communautaires de Côte-des-Neiges in Montreal (Lahrichi et al. 2006).

To analyze the effects of δ and d^c on O_{min}, we focus on the results obtained when the number of districts is fixed. Figure 2 presents the means (presented by dots) and standard deviations (presented by vertical bars) of O_{min} calculated from the non-domination set, where the x axis represents the maximum duration of travel between any two basic units in a district d^c, and lines of different shapes and colors represent the admissible percentage deviation of the workload among services δ. As shown in Fig. 2, the standard deviation of O_{min} is less than 4 for all of the designs. Hiring more care workers at an above average level (high cost) will lead to a good workload balance or greater compatibility between districts (based on the definition of non-domination). As employing fewer care workers is preferable for the ECS, the number of good districts designed should be minimized. Dots unassociated with vertical bars usually present the objective values of infeasible solutions and are called infeasible dots, such as the dots at $d^c \leq 1800,\ m = 2$ and $d^c \leq 1500,\ m = 3$ and the dots at $d^c = 1200,\ m = 4,\ \delta = 0.1$ and $5 < m \leq 7,\ \delta \leq 0.25,\ d^c \leq 1500$, as shown in Fig. 2. These infeasible dots indicate that small values of m, d^c and δ impose strict requirements for district design, district compactness, and workload balance among services, which may lead to infeasibility.

Among the feasible dots at m $\geq 3,\ \delta > 0,\ 25,\ d^c > 1500$, where only blue and red lines count, Fig. 2 also states that for the blue line ($\delta = 1$), O_{min} increases first then drops when $m = 3, 4$, but drops first then increases or remains stable when $m \geq 5$; for the red line ($\delta = 0.5$), O_{min} drops first then increases for all m except $m = 3$. This trend shows that the red and blue lines are close to each other when $d^c = 1800, 2500$. The blue line is above the red line, which means that considerable relaxation of the balance of care workload may or may not help to reduce the number of care workers needed. In the real world, the number of care workers needed mainly depends on the demand distribution across basic units. The ECS hires more care workers when the supply (care workers' total available workload) fails to satisfy demand. When supply satisfies demand, the requirement for workload balance affects only the allocation of demand (basic units) to care workers (districts). Then, when a care worker's total assigned demand nearly equals his/her capacity (i.e., average workload), relaxing workload balance cannot reduce the number of care workers needed; when a care worker's assigned demand nearly equals half of his/her capacity, relaxing workload balance can reduce the number of care workers needed.

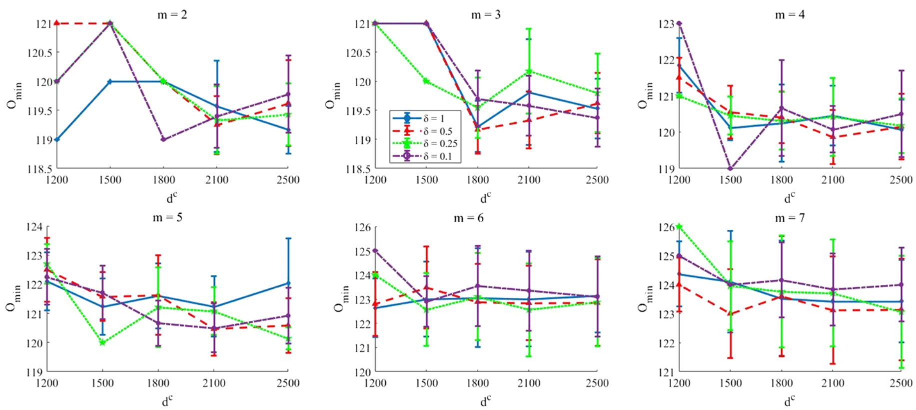

Fig. 2 Means and standard deviations of $O_{\min}$ with different values of m

5.2.2 Effects of m, δ, and d^c on $\delta_{\max}$ (maximum deviation of service workload across districts)

In this subsection, the effects of the key parameters on δ_{max} are discussed. Figure 3 shows that the blue dots indicating low δ_{max} are located in the up zone (m ranges from 2 to 7; d^c ranges from 1800 to 2500). As the value of δ increases (plane of $\delta = 0.1$ to plane of $\delta = 1$ from right to left of y axis), the value of δ_{max} increases slightly. This indicates that a small δ for the workload associated with care services may lead to a good workload balance across districts. This finding is consistent with the analysis of the model stated in Eqs. (24) and (25). $\delta_{\max}$ rises as m increases from 2 to 7 (also illustrated in Fig. 4), which means that it may be more difficult to maintain a workload balance among districts when more service districts are designed. When more districts are designed, achieving the same level of workload balance among districts in the real world requires more constraints to be satisfied and a larger search space, which may lead to more infeasible solutions and greater computational time.

Figure 4 presents the means and standard deviations of δ_{max}, which also reflect the relationships between δ_{max}, m, and d^c. In this figure, the standard deviation of δ_{max} becomes smaller when δ falls from 1 to 0.1, because the error bars shorten in the subfigures from the top left of the figure to the bottom right. δ_{max} is large and fluctuates most when $d^c = 1200$, followed by $d^c = 1500$. The trends in δ_{max} under $d^c > 1500$ remain stable and similar to each other. The green line ($d^c = 1800$) is close to the red line ($d^c = 2100$) when $\delta = 1$; close to the blue line ($d^c = 2500$) when $\delta = 0.5, 0.25$; and close to both red and blue lines ($d^c = 2100, 2500$) when $\delta = 0.1$. This trend indicates that for the same number of designed districts, a good workload balance across districts and a certain level of district compactness can be achieved simultaneously. Specifically, the strict constraint on the fine compatibility of basic units can be satisfied without a deterioration of workload balance across districts. In the real world, given the same requirement of care workers' workload balance and a fixed number of districts, managers can decrease the value of d^c to achieve the same level of workload balance across districts.

Figure 5 illustrates the effects of δ and d^c on δ_{max}. In this figure, feasible dots at $d^c = 1800$ are shown to be closer to each other than other dots at other d^c, which indicates that δ has a limited effect on δ_{max} when $d^c = 1800$. Figure 5 also shows that almost all values

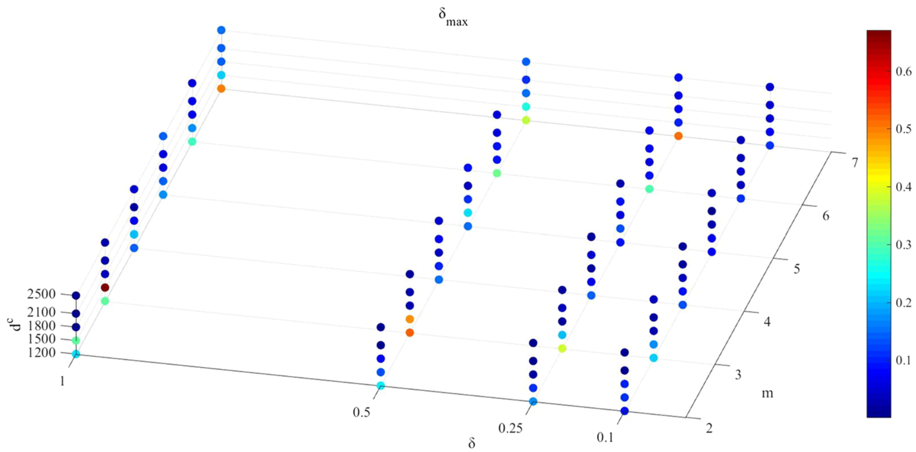

Fig. 3 Mean of $\delta_{\max}$ with all parameter settings

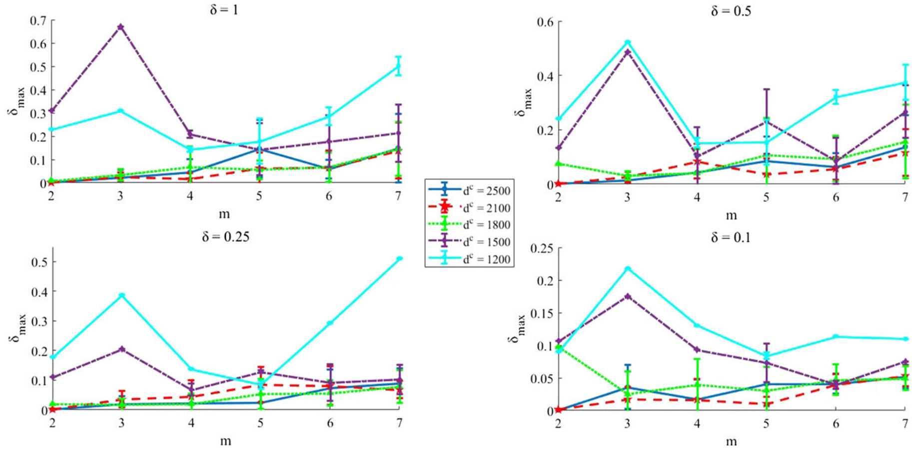

Fig. 4 Means and standard deviations of $\delta_{\max}$ with different values of δ

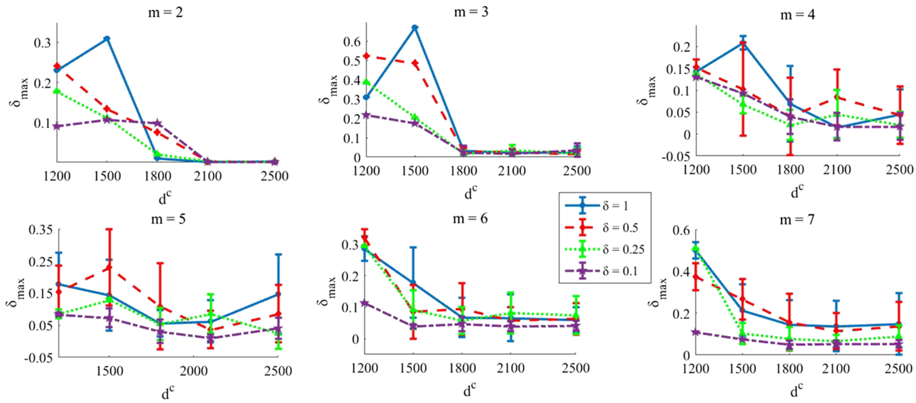

Fig. 5 Means and standard deviations of $\delta_{\max}$ with different values of m

of δ_{max} with any $\delta \leq 1$ when $d^c = 1800$ are less than 0.2. This effect implies that districts' total workload is not smaller than 80% and not greater than 120% of the average workload with mid-level district compactness and few districts.

5.2.3 Effects of m, δ, and d^c on $d_{\max}$ (maximum travel duration between any two basic units in a district)

In this subsection, the effects of the key parameters on d_{max} are discussed. Figure 6 indicates that when $\delta = 0.1$ and $d^c \leq 1200$, the values of dots at $m \leq 4$or$m \geq 7$ exceed the corresponding values of d^c, which means that a small δ, d^c may lead to infeasible solutions when the number of districts planned is excessively small or large. In addition, Fig. 6 indicates that d_{max} decreases with a decreasing d^c(plane of $d^c = 2500$ to plane of $d^c = 1200$ from the top to the bottom of the z axis). In this figure, the strict requirement for the compatibility of basic units is shown to enhance compactness in the districts created, which is consistent with Constraints (7) and (12) in the MOMINLP model.

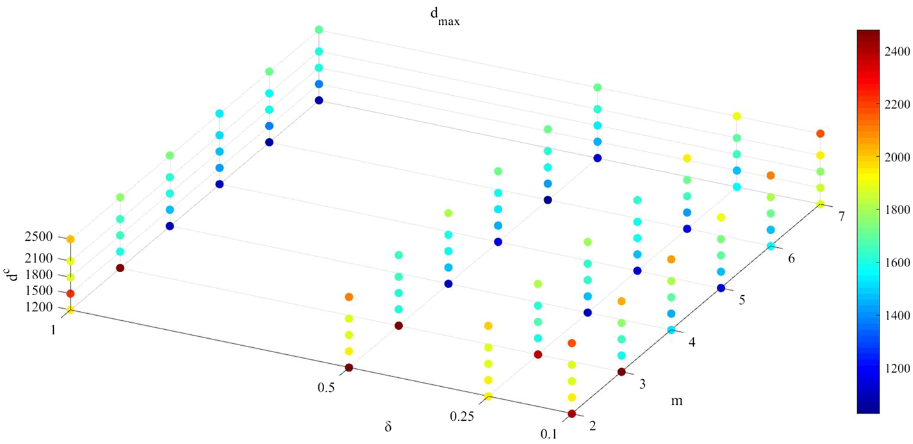

Fig. 6 Means of $d_{\max}$ for all parameter settings

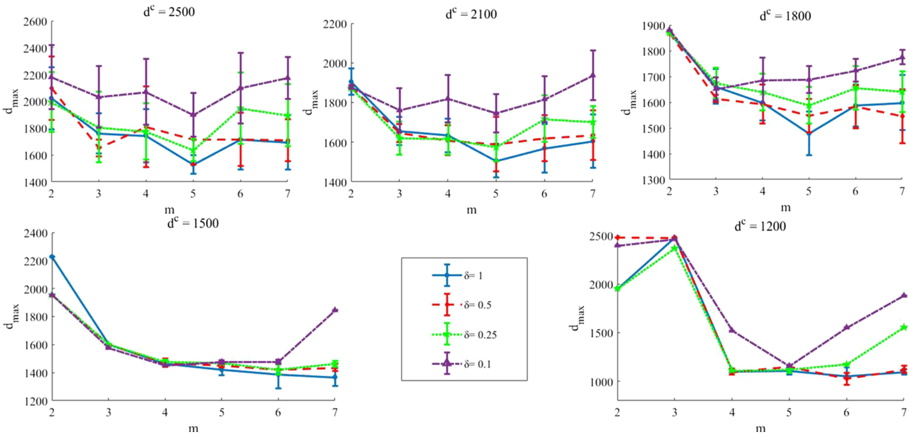

Fig. 7 Means and standard deviations of $d_{\max}$ with different values of d^c

Figure 7 presents the means and standard deviations of d_{max} for different values of d^c. The figure indicates that the standard deviation of d_{max} decreases when d^c is reduced. As most dots at $d^c = 1200, 1500$ present values of d_{max} with infeasible solutions, the discussion focuses on $d^c \geq 1800$. Specifically, the subfigure for $d^c = 1800$ is analyzed due to the similarity between the three subfigures for $d^c \geq 1800$. Figure 7 shows that d_{max} increases as δ decreases (the purple line is above the green line, followed by the red and blue lines). d_{max} fluctuates according to the number of districts designed. $d_{\max}$ is most stable when $\delta = 0.5$, presented as the red line at $d^c = 1800$. Thus, a certain level of compactness can be achieved with a proper admissible percentage deviation on the workload of care workers regardless of how many districts are designed. In the real world, good districts can be designed to achieve compactness by satisfying the given requirements for workload balance (δ) and responsiveness (d^c).

Figure 7 also shows that d_{max} decreases with a decreasing value of d^c and an increasing value of δ when the number of districts is fixed. Relaxing workload balance to some extent increases the compactness of the designed districts. However, due to the trade-off between the cost of hiring care workers, the workload balance across designed districts, and the compactness of districts, a proper δ and d^c should be selected to balance the objectives.

5.2.4 Effect summary

m (the number of districts created), δ (the admissible percentage deviation of the workload of care worker of type s), and d^c (the maximum duration of travel between any two basic units in a district) have important effects on O_{min} (the number of care workers needed),δ_{max} (the maximum deviation of workload across districts), and d_{max} (the maximum duration of travel between any two basic units in any district). Low values of O_{min}, δ_{max}, and d_{max} facilitate the trade-off between these key parameters. Based on the analysis, proper values are selected for key parameters by considering the trade-off between the performance indicators; that is, m equals 4, δ equals [0.5, 1], and d^c equals $3d/4 \approx 1800$. The analysis aids understanding of the relationships between the district-related parameters and the three objectives.

5.3 Discussion of the effects of correlations between service-related parameters on objectives

Based on the conclusions drawn from the analysis in Sect. 5.1, the effects of the correlations between and the functions of the service-related parameters are discussed with the number of districts, m, fixed at 4 and the maximum duration between any two basic units in a district, d^c, fixed at 1800 in the second experiment, wherein two types of care workers are considered with different values of $\delta_s \in [0.5, 1]$ and C_s/Q_s.

5.3.1 Effects of δ_s on O_{min}, δ_{max}, and d_{max}

This section focuses on the effects of the correlations between δ_s on the three objectives based on the analysis above; that is, m $= 4$, $d^c = 1800$, $C_s = 1$, $Q_s = Q_s$. As $\delta_s \in [0.5, 1]$, $\delta_1 = \delta_2$ is assumed for a positive correlation and $\delta_1 = 1.5 - \delta_2$ for a negative correlation; otherwise, δ_1 and δ_2 are uniform random values ranging from 0.5 to 1.

Figure 8 presents the effects of these three types of correlation—that is, independent, negative, and positive—on the objectives. δ_{index} is the index of δ_1, ranging from 0.5 to 1 with 0.05 as an interval when the correlation between δ_1 and δ_2 is linear; for example, $\delta_1 = 0.6$, $\delta_{index} = 3$, or $\delta_1 = 0.8$, $\delta_{index} = 7$; otherwise, each δ_{index} presents an independent random pair (δ_1, δ_2). In Fig. 8, the three subfigures show that the deviations of O_{min}, δ_{max}, and d_{max} are large when the correlation between δ_1 and δ_2 is negative. The blue, red, and green lines fluctuate with different values of (δ_1, δ_2), and show no obvious trends or differences. Moreover, Fig. 8 shows that the blue line is more stable than the other two lines. This result implies that the correlation between δ_1 and δ_2 has a limited effect on the variables associated with the objectives, i.e., O_{min}, δ_{max}, and d_{max}. Therefore, the SA-TPIHCSS can randomly select values of $\delta_s \in [0.5, 1]$ without considering this correlation. This makes sense, because requirements of workload balance influence care workers of the same grade (type) in the real world.

5.3.2 Effects of C_s/Q_s on O_{min}, δ_{max}, and d_{max}

This section discusses the effects of a service-related parameter, the ratio of the cost of hiring a care worker to the workload of the care worker C_s/Q_s, on the three objectives, with the assumptions that $\delta_1 = \delta_2 = 0.5$ and the baseline workload Q is known. Gradient

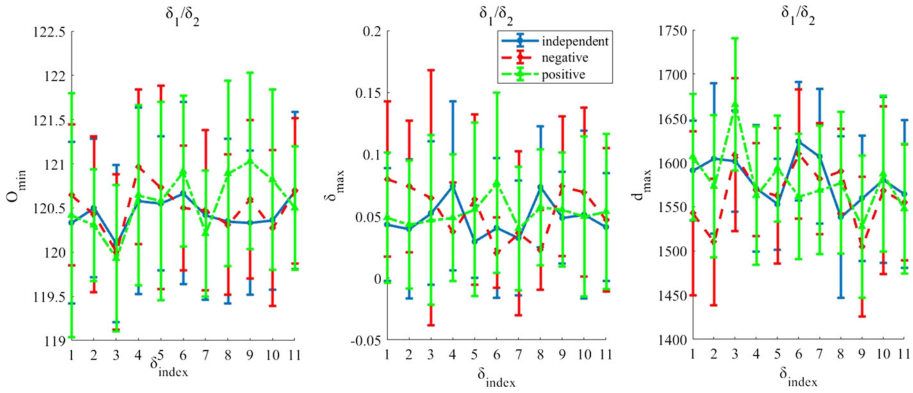

Fig. 8 Means and standard deviations of $O_{\min}$, $\delta_{\max}$, $d_{\max}$ with different correlations of δ_s

$\vartheta = 1/Q$ when C_s is modeled as a linear function of Q_s. When C_s is modeled as a stepwise function of Q_s, the workload range is divided into 10 pieces; that is, $C_s = (1+\rho)^{(l-4)}$ when $Q_s \in [0.5Q + (l-1) \times 0.15Q, 0.5Q + l \times 0.15Q]$, $\rho = 0.2$ is a constant indicating a large bonus for working hard, and $l \in [1, 10]$ is the piece to which Q_s belongs. Q_s ranges from $[0.5Q, 2Q]$, whose bounds refer to the minimum and maximum available workload for each care worker.

Figure 9 presents the means and standard deviations of O_{min}, δ_{max}, and d_{max} for different values of C_s/Q_s. The x axis presents the piece to which Q_s belongs. For each piece, the dots/vertical bars that make up the red and green lines present the means/standard deviations of O_{min}, δ_{max}, and d_{max} in the subfigures from left to right, respectively. In the left-hand subfigure of Fig. 9, the green line displays a slight linear increase with an increase in Q_s, whereas the red line shows an exponential growth with an increase in Q_s. These results match O_{min} with a linear and stepwise correlation between C_s and Q_s, as shown in Eq. (23). In addition, Fig. 9 shows that the vertical bars making up the red line are longer than those making up the green line, which indicates that the standard deviation of $O_{\min}$ is larger when C_s is modeled as a stepwise function of Q_s. This finding also matches the relationship between O_{min} and C_s according to Eq. (5). For the middle subfigure, both the red line and the green line fluctuate with Q_s belonging to different pieces. Determining the difference between the performance of red and of green lines is difficult, which indicates that different formulations of C_s/Q_s have a limited effect on the workload balance across districts. In the right-hand subfigure, most of the red dots are above the green ones, which means that when C_s is a stepwise function of Q_s, the districts designed are less compact than when C_s is a linear function of Q_s. However, most of the green vertical bars are longer than the red ones, which implies that the compactness of districts with a constant C_s/Q_s differs more from the average than that of districts with a stepwise C_s/Q_s. In the real world, providing care workers with a larger bonus for hard work (stepwise C_s/Q_s) requires the ECS to reduce care workers' travel time/cost, which increases the strictness of the requirement for district compactness.

In summary, the correlation between values of δ_s has a limited influence on the three variables associated with the objectives, O_{min}, δ_{max}, and d_{max}, so decision makers in the ECS can randomly choose values of $\delta_s \in [0.5, 1]$. Different relationships between C_s and Q_s mainly affect the cost of hiring care workers, O_{min},; they have a limited influence on the other two objectives, δ_{max} and d_{max}. To save costs, the ECS can choose a linear relationship between C_s and Q_s.

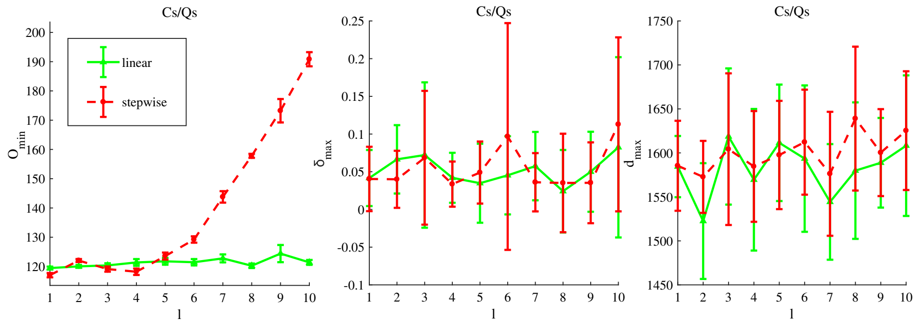

Fig. 9 Means and standard deviations of $O_{\min}$, $\delta_{\max}$, and $d_{\max}$ with different values of C_s/Q_s

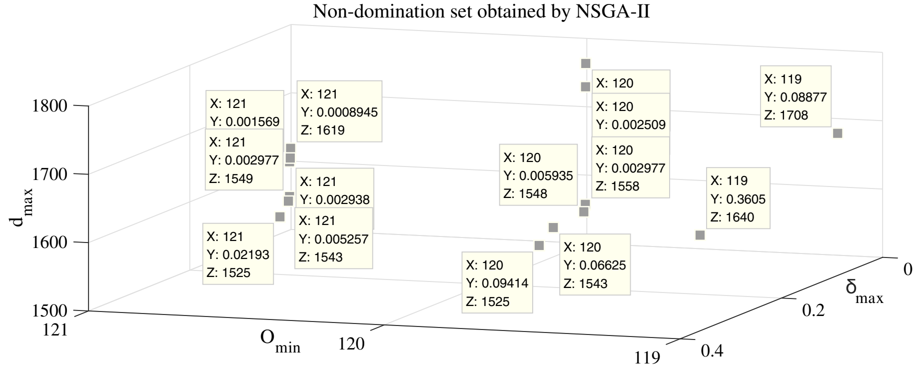

Fig. 10 Non-domination set when $m = 4$, $\delta = 0.5$, and $d^c = 1800$

5.4 Non-domination set obtained by NSGA-II

This section discusses the objective values of the non-domination set obtained by NSGA-II with proper values for the key district-related parameters ($m = 4$, $\delta = 0.5$, and $d^c = 1800$) and service-related parameters ($C_s = 1$, $Q_s = Q$; $\delta_1 = \delta_2 = \delta$). Figure 10 shows the non-dominated solutions, where X presents the cost of hiring care workers, O_{min}; Y presents the maximum deviation of the total workload from the average workload across the districts, δ_{max}; and Z presents the maximum duration of travel between any two basic units in any district, d_{max}.

The trade-off between these three objectives is also shown in Fig. 3: with the same O_{min} (i.e., equaling 120), increasing δ_{max} reduces d_{max} (referring to points at which X: 120); with the same d_{max} (i.e., equaling 1525), increasing O_{min} reduces δ_{max} (referring to points (X: 120, Y: 0.09414, Z: 1525; X: 121, Y: 0.02193, Z: 1525)). For example, when $O_{min} = 119$ (referring to points (X: 119, Y: 0.08877, Z: 1708; X: 119, Y: 0.3605, Z: 1640)), δ_{max} increases from 0.08877 to 0.3605 as d_{max} decreases from 1708 to 1640, which indicates that to achieve more compact districts (by reducing travel time between any two basic units in any district by 68 s), a large sacrifice in workload balance is necessary for the same cost of hiring care workers. In addition, Fig. 3 shows that most values of δ_{max} are below 0.1 except the point (X: 119, Y: 0.3605, Z: 1640). These values mean that in most cases, workload across districts can be balanced well, with a deviation of less than 10% from average.

Figure 3 helps the ECS to make appropriate decisions based on its particular concerns. For instance, to achieve good levels of compactness and workload balance, managers can choose the point (X: 120, Y: 0.005935, Z: 1548), which refers to the district design solution shown in Fig. 11 of the Appendix. The districting results are presented in four colors. Points in the same color indicate the same district, as shown in Fig. 11. Basic units clustered in the same district may not be physically close to each other. However, they can be reached from any basic unit in the same district within 1548 s using the superior transportation mode.

6 Conclusion

In this study, a general MOMINLP model is proposed for an ICSD problem in an integrated care system. The model considers three objectives, namely minimizing the cost of hiring care workers allocated, balancing workload among districts, and maximizing the compactness of districts, and several criteria, including the compactness of districts, the individuality of each basic unit, the compatibility of basic units, the workload balance of care workers, and different elderly profiles. The proposed model is solved using a modified version of NSGA-II and validated using an actual instance collected from the SA-TPIHCSS, with road distance obtained from Google Maps. Numerical analysis was conducted by varying the key parameters of the problem to shed some light on the effects of these key parameters on the objectives of the problem.

The numerical analysis leads to three main conclusions. First, district-related parameters (here the number of districts designed, the balance achieved between types of workload, and the maximum travel duration) have significant effects on the three objectives (i.e., the cost of hiring allocated care workers, workload balance among districts, and district compactness). This finding will help the ECS to select proper key parameters and proper solutions that address their specific concerns. Second, service-related parameters (here the correlations between deviations in service workloads) have limited effects on the three objectives. This means that the ECS can randomly select workload deviation within a certain scope. Third, different policies for awarding bonuses for hard work have important effects on the cost of hiring care workers: when cost is a linear function of workload, the total cost of hiring care workers shows slow linear growth with increasing workload, whereas the total cost of hiring care workers increases exponentially with increasing workload when cost is a stepwise exponential function of workload. However, the relationship between cost and workload has only limited effects on workload balance across districts and district compactness. These findings can help the ECS to select and implement policies that encourage care workers to work hard and thereby improve their satisfaction.

This study is one of very few to explore the effects of key parameters on the objectives of an ICSD problem in the ECS context. The MOMINLP model and modified NSGA-II offer critical insights into ways of ensuring an effective districting system (whose key characteristics are presented in Sect. 4.2) for the ECS. However, the study has a few limitations. First, the instance used to validate the MOMINLP model consists of only two types of care worker. Second, certain criteria, such as the increasing rate of burnout among care workers with increasing workloads and the various skill levels of care workers of each type, are not considered in the ICSD problem.

Future work will take several directions. First, we will focus on the interesting task of developing proper approaches, such as the ε-constraint method and elitist non-dominated sorting approaches (i.e., NSGA-III (Jain and Deb 2014), the non-dominated moth flame optimization algorithm (Savsani and Tawhid 2017)), and the Pareto envelope-based selection algorithm), to solving the tri-objective ICSD problem. The performance of these approaches in solving the ICSD problem will be compared with reference to objective function values and computational time. Second, the MOMINLP model will be extended by considering common operational constraints, such as care workers' burnout and skill levels and the uncertainty of service demand. Third, the MOMINLP model will be validated using different instances of different sizes.

Acknowledgements We thank the guest editors and two anonymous referees for their helpful comments, which have greatly improved the exposition of this paper. This work was partially supported by the National Natural Science Foundation of China (Grant Nos. 71471118, 71871145, 71801158), the Major Project of National Natural Science Foundation of China (Grant No. 71790615), the Key Project of National Natural Science Foundation of China (Grant No. 71431006), the Humanities and Social Science Foundation of Ministry of Education of China (Grant Nos. 14YJC630096, 18YJC630088), Start-Up Funds of Shenzhen University (Grant No. 2018058), the Hong Kong RGC (Grant No. T32-102/14N).

Appendix

See Tables 2, 3 and Fig. 11.

Table 2 Sample records of admission and discharge

Case ID	Profile type	Admission date	Discharge date
300226140005	Invalidism	2014/8/18	
300226140004	Invalidism	2014/6/19	2014/10/14
300226130008	Invalidism	2013/11/16	2013/11/11
300226130007	Invalidism	2013/10/28	
300226130003	Invalidism	2013/4/8	2013/4/8
300226120003	Invalidism	2012/12/27	2013/10/15
300227140073	Ordinary	2014/10/4	
300227140071	Ordinary	2014/9/3	
300227140040	Ordinary	2014/5/22	2014/8/1
300227130140	Ordinary	2013/12/11	2014/2/13
300227130131	Ordinary	2013/11/13	2014/1/27
300227130126	Ordinary	2013/11/8	
300227130073	Ordinary	2013/7/12	
300227130034	Ordinary	2013/4/19	
300227120089	Ordinary	2012/10/27	2014/1/7
300227120072	Ordinary	2012/8/28	
300227120042	Ordinary	300227120042	
300227120002	Ordinary	2012/1/27	2013/10/9

Table 3 The daily schedule of a care worker

Time slots	Work
9:00–10:30	Integrated home care service, cater service, nursing care, etc
10:30–11:00	Prepare the lunch boxes—MOW service
11:00–12:30	Deliver the lunch boxes—MOW service
12:30–13:30	Lunch break
13:30–15:00	Integrated home care service, cater service, nursing care, etc
15:00–16:30	Prepare the lunch boxes—MOW service
16:30–17:30	Deliver the lunch boxes—MOW service

Fig. 11 District design solution with objectives (X: 120, Y: 0.005935, Z: 1548)

References

Agrawal, R. B., Deb, K., & Agrawal, R. B. (1995). Simulated binary crossover for continuous search space. *Complex Systems, 9*(2), 115–148.

Ansari, S., McLay, L. A., & Mayorga, M. E. (2015). A maximum expected covering problem for district design. *Transportation Science, 51*(1), 376–390.

Ariyasingha, I. D. I. D., & Fernando, T. G. I. (2015). Performance analysis of the multi-objective ant colony optimization algorithms for the traveling salesman problem. *Swarm and Evolutionary Computation, 23,* 11–26.

Asafuddoula, M., Ray, T., & Sarker, R., et al. (2012). *An adaptive constraint handling approach embedded MOEA/D.* Paper presented at the 2012 IEEE congress on evolutionary computation.

Bard, J. F., & Jarrah, A. I. (2009). Large-scale constrained clustering for rationalizing pickup and delivery operations. *Transportation Research Part B: Methodological, 43*(5), 542–561.
Benzarti, E. (2012). *Home health care operations management: Applying the districting approach to home health care.* (Doctoral dissertation), Ecole Centrale Paris.
Benzarti, E., Sahin, E., & Dallery, Y. (2013). Operations management applied to home care services: Analysis of the districting problem. *Decision Support Systems, 55,* 587–598.
Blais, M., Lapierre, S. D., & Laporte, G. (2003). Solving a home-care districting problem in an urban setting. *Journal of the Operational Research Society, 54*(11), 1141–1147.
Boldy, D., & Howell, N. (1980). The geographical allocation of community care resources—A case study. *Journal of the Operational Research Society, 31*(2), 123–129.
Bouajaja, S., & Dridi, N. (2017). A survey on human resource allocation problem and its applications. *Operational Research, 17*(2), 339–369.
Brownlee, A. E. I., & Wright, J. A. (2015). Constrained, mixed-integer and multi-objective optimisation of building designs by NSGA-II with fitness approximation. *Applied Soft Computing, 33,* 114–126.
Chaves, A. A., & Nogueira Lorena, L. A. (2011). Hybrid evolutionary algorithm for the capacitated centered clustering problem. *Expert Systems with Applications, 38*(5), 5013–5018.
Cho, J. H., Wang, Y., Chen, I. R., et al. (2017). A survey on modeling and optimizing multi-objective systems. *IEEE Communications Surveys & Tutorials, 19*(3), 1867–1901.
Deb, K. (2000). An efficient constraint handling method for genetic algorithms. *Computer Methods in Applied Mechanics and Engineering, 186*(2), 311–338.
Deb, K. (2012). *Advances in evolutionary multi-objective optimization.* Paper presented at the search based software engineering, Berlin.
Deb, K., Pratap, A., Agarwal, S., et al. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation, 6*(2), 182–197.
Deep, K., Singh, K. P., Kansal, M. L., et al. (2009). A real coded genetic algorithm for solving integer and mixed integer optimization problems. *Applied Mathematics and Computation, 212*(2), 505–518.
Dulebenets, M. A. (2018). A comprehensive multi-objective optimization model for the vessel scheduling problem in liner shipping. *International Journal of Production Economics, 196,* 293–318.
Gov, H. K. (2018). Health care voucher. Retrieved from http://www.hcv.gov.hk/eng/pub_background.htm.
Gurobi. (2018). Gurobi works for multi-objective optimization. Retrieved from http://www.gurobi.com/documentation/7.5/refman/working_with_multiple_obje.html.
Gutiérrez-Gutiérrez, E. V., & Vidal, C. J. (2015). A home health care districting problem in a rapid-growing city. *Ingeniería y Universidad, 19,* 87–113.
Haugland, D., Ho, S. C., & Laporte, G. (2007). Designing delivery districts for the vehicle routing problem with stochastic demands. *European Journal of Operational Research, 180*(3), 997–1010.
Hess, S. W., & Samuels, S. A. (1971). Experiences with a sales districting model: Criteria and implementation. *Management Science, 18*(4), 41–54.
Jain, H., & Deb, K. (2014). An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, Part II: Handling constraints and extending to an adaptive approach. *IEEE Transactions on Evolutionary Computation, 18*(4), 602–622.
Jarrah, A. I., & Bard, J. F. (2011). Pickup and delivery network segmentation using contiguous geographic clustering. *Journal of the Operational Research Society, 62*(10), 1827–1843.
Jarrah, A. I., & Bard, J. F. (2012). Large-scale pickup and delivery work area design. *Computers & Operations Research, 39*(12), 3102–3118.
Kalcsics, J. (2015). Districting problem. In G. Laporte, S. Nickel, & F. da Gama (Eds.), *Location science* (pp. 595–622). Cham: Springer.
Kang, D., Jung, J., & Bae, D. H. (2011). Constraint-based human resource allocation in software projects. *Software: Practice and Experience, 41*(5), 551–577.
Konur, D., & Geunes, J. (2016). Integrated districting, fleet composition, and inventory planning for a multi-retailer distribution system. *Annals of Operations Research*. https://doi.org/10.1007/s10479-016-2338-6.
Krause, N., Scherzer, T., & Rugulies, R. (2005). Physical workload, work intensification, and prevalence of pain in low wage workers: Results from a participatory research project with hotel room cleaners in Las Vegas. *American Journal of Industrial Medicine, 48*(5), 326–337.
Lahrichi, N., Lapierre, S., Hertz, A., et al. (2006). Analysis of a territorial approach to the delivery of nursing home care services based on historical data. *Journal of Medical Systems, 30*(4), 283–291.
Lepak, D. P., & Snell, S. A. (1999). The human resource architecture: Toward a theory of human capital allocation and development. *Academy of Management Review, 24*(1), 31–48.
Lin, M., Chin, K.-S., Fu, C., et al. (2017). An effective greedy method for the Meals-On-Wheels service districting problem. *Computers & Industrial Engineering, 106,* 1–19.

Lin, M., Chin, K. S., Wang, X., & Tsui, K. L. (2016). The therapist assignment problem in home healthcare structures. *Expert Systems with Applications*, *62*, 44–62.

Lin, C.-M., & Gen, M. (2008). Multi-criteria human resource allocation for solving multistage combinatorial optimization problems using multiobjective hybrid genetic algorithm. *Expert Systems with Applications, 34*(4), 2480–2490.

Maya Duque, P. A., Castro, M., Sörensen, K., et al. (2015). Home care service planning. The case of Landelijke Thuiszorg. *European Journal of Operational Research, 243*(1), 292–301.

Meng, K., Lou, P., Peng, X., et al. (2017). Multi-objective optimization decision-making of quality dependent product recovery for sustainability. *International Journal of Production Economics, 188,* 72–85.

Minella, G., Ruiz, R., & Ciavotta, M. (2008). A review and evaluation of multiobjective algorithms for the flowshop scheduling problem. *INFORMS Journal on Computing, 20*(3), 451–471.

Moreno, S., Pereira, J., & Yushimito, W. (2017). A hybrid K-means and integer programming method for commercial territory design: A case study in meat distribution. *Annals of Operations Research*. https://doi.org/10.1007/s10479-017-2742-6.

Negreiros, M., & Palhano, A. (2006). The capacitated centred clustering problem. *Computers & Operations Research, 33*(6), 1639–1663.

Otero, L. D., Centeno, G., Ruiz-Torres, A. J., et al. (2009). A systematic approach for resource allocation in software projects. *Computers & Industrial Engineering, 56*(4), 1333–1339.

Pezzella, F., Bonanno, R., & Nicoletti, B. (1981). A system approach to the optimal health-care districting. *European Journal of Operational Research, 8*(2), 139–146.

Pureza, V., Morabito, R., & Reimann, M. (2012). Vehicle routing with multiple deliverymen: Modeling and heuristic approaches for the VRPTW. *European Journal of Operational Research, 218,* 636–647.

Ray, T., Tai, K., & Seow, K. C. (2001). Multiobjective design optimization by an evolutionary algorithm. *Engineering Optimization, 33*(4), 399–424.

Ríos-Mercado, R. Z., & Fernández, E. (2009). A reactive GRASP for a commercial territory design problem with multiple balancing requirements. *Computers & Operations Research, 36*(3), 755–776.

Salazar-Aguilar, M. A., Ríos-Mercado, R. Z., González-Velarde, J. L., et al. (2012). Multiobjective scatter search for a commercial territory design problem. *Annals of Operations Research, 199*(1), 343–360.

Savsani, V., & Tawhid, M. A. (2017). Non-dominated sorting moth flame optimization (NS-MFO) for multi-objective problems. *Engineering Applications of Artificial Intelligence, 63,* 20–32.

Shanker, R. J., Turner, R. E., & Zoltners, A. A. (1975). Sales territory design: An integrated approach. *Management Science, 22*(3), 309–320.

Steiner Neto, P. J., Datta, D., Arns Steiner, M. T., et al. (2017). A multi-objective genetic algorithm based approach for location of grain silos in Paraná State of Brazil. *Computers & Industrial Engineering, 111,* 381–390.

Subtil, R. F., Carrano, E. G., & Souza, M. J. F., et al. (2010). *Using an enhanced integer NSGA-II for solving the multiobjective generalized assignment problem.* Paper presented at the IEEE congress on evolutionary computation.

United Nations, D. O. E. A. S. A., Population Division. (2017). World population prospects: the 2017 revision. Retrieved from https://esa.un.org/unpd/wpp/Publications/Files/WPP2017_KeyFindings.pdf.

Vanneschi, L., Henriques, R., & Castelli, M. (2017). Multi-objective genetic algorithm with variable neighbourhood search for the electoral redistricting problem. *Swarm and Evolutionary Computation, 36,* 37–51.

Wang, R., Xiong, J., Ishibuchi, H., et al. (2017). On the effect of reference point in MOEA/D for multi-objective optimization. *Applied Soft Computing, 58,* 25–34.

Xinhuanet. (2018). China health care industry development report 2016—talent. Retrieved from http://www.xinhuanet.com/gongyi/yanglao/2017-01/24/c_129457792.htm.

Yoshimura, M., Fujimi, Y., Izui, K., et al. (2006). Decision-making support system for human resource allocation in product development projects. *International Journal of Production Research, 44*(5), 831–848.

Yu, S., Zheng, S., Gao, S., et al. (2017). A multi-objective decision model for investment in energy savings and emission reductions in coal mining. *European Journal of Operational Research, 260*(1), 335–347.

Zhang, X., Zheng, X., Cheng, R., et al. (2018). A competitive mechanism based multi-objective particle swarm optimizer with fast convergence. *Information Sciences, 427,* 63–76.

Zhou, S., Li, X., Du, N., et al. (2018). A multi-objective differential evolution algorithm for parallel batch processing machine scheduling considering electricity consumption cost. *Computers & Operations Research, 96,* 55–68.

Zoltners, A. A., & Sinha, P. (2005). The 2004 ISMS practice prize winner—Sales territory design: thirty years of modeling and implementation. *Marketing Science, 24*(3), 313–331.

Affiliations

Meiyan Lin[1,3] · Kwai Sang Chin[2] · Lijun Ma[1,3] · Kwok Leung Tsui[2]

Meiyan Lin
mylin17@szu.edu.cn

Kwai Sang Chin
mekschin@cityu.edu.hk

Kwok Leung Tsui
kltsui@cityu.edu.hk

1 Department of Management Science, College of Management, Shenzhen University, Shenzhen 518060, Guangdong, People's Republic of China

2 Department of Systems Engineering and Engineering Management and Centre of Systems Informatics Engineering, City University of Hong Kong, Kowloon, Hong Kong, People's Republic of China

3 Institute of Big Data, Intelligent Management and Decision, Shenzhen University, Shenzhen 518060, Guangdong, People's Republic of China