Applied Soft Computing 26 (2015) 350–356

Contents lists available at ScienceDirect

Applied Soft Computing

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

Decision rules for ambulance scheduling decision support systems

Lu Zhen*, Shaopu Sheng, Zhenhua Xie, Kai Wang

School of Management, Shanghai University, 99 Shangda Road, Shanghai 200444, China

ARTICLE INFO

Article history:
Received 25 July 2013
Received in revised form 17 August 2014
Accepted 1 October 2014
Available online 22 October 2014

Keywords:
Decision support systems
Emergency ambulance
Ambulance scheduling

ABSTRACT

This paper studies some decision rules for ambulance scheduling. The scheduling decision rules embedded in the decision support systems for emergency ambulance scheduling consider the criteria on the average response time and the percentage of the ambulance requests that are responded within 15 min, which is usually ignored in traditional scheduling policies. The challenge in designing the decision rules lies in the stochastic and dynamic nature of request arrivals, fulfillment processes, and complex traffic conditions as well as the time-dependent spatial patterns of some parameters complicate the decisions in the problem. To illustrate the proposed decision rules' usage in practice, a simulator is developed for performing some numerical experiments to validate the effectiveness and the efficiency of the proposed decision rules.

1. Introduction

Ambulance scheduling for emergency requests is important to rescue people when their health is in risk of irreparable damage. The cost of medical resources and request volume are growing. Moreover, the traffic conditions in cities are extremely dynamic. All of these issues make the decisions on ambulance scheduling become more and more challenging. The control centers of ambulances are supposed to schedule ambulances from their bases (waiting locations) so that ambulance requesters can be reached in a time efficient manner. In reality, regulatory authorities usually established some performance criteria for the emergency medical resources transportation. In usual practice, the requests of the emergency ambulances have several different levels of priorities. As to each level of priority, the regulatory authorities require an ambulance to arrive at the requester (patient)'s place within a particular response time. A widely used criterion for the ambulance scheduling is that as to the high priority requests, ambulances need arrive at the patients' place in 15 min, which is a golden time and during which the patients should be timely transported to a proper healthcare center where appropriate medical team can give the patients sophisticated medical treatments. The most straightforward but useful decision rule is to allocate an ambulance that is the closest to the requester's place. However, this intuitive scheduling policy cannot guarantee a high percentage of the requests that can be responded within 15 min. The criterion on the percentage of 15 min response is more important than the criterion on the average response time. This study investigate how to design some good decision rules for emergency ambulance scheduling so as to guarantee a high percentage of 15 min response, which is essential for the control centers of ambulance in cities. The challenges for designing highly efficient decision rules of ambulance scheduling lies in the dynamic environment where the spatial distribution of potential requesters varies in the time dimension while and the spatial patterns of traffic situations also vary in the peak hours and off-peak hours. Moreover, the traveling time of ambulances and their stay time at patients' places are stochastic parameter. All of the above mentioned random and dynamic features of the environments further complicate the ambulance scheduling decision. Therefore, this paper makes an explorative study on designing highly efficient decision rules for scheduling ambulances.

2. Related works

The early studies on the resource optimization of the emergency ambulance are mainly related with the minimal covering model [22], which tries to minimize the number of ambulances necessary so as to cover all demand point, and the maximal covering model [6], which tries to maximize the total demand coverage given a feet of fixed size. In the recent years, some scholars concentrate on the dispatching policies. For example, Centrality policy [13], which evolved from the nearest neighbor (NN) policy, is proposed in an effort to reduce the response time in demanding emergency situations. The auction mechanism [14] based on trust is designed for dispatching ambulances for emergency patient transportation. Besides the above studies on ambulance dispatching, some scholars focus on the ambulance redeployment. One stream of the studies on the redeployment models is to apply integer programming

* Corresponding author. Tel.: +86 21 66134237.
E-mail address: lzhen.sh@gmail.com (L. Zhen).

http://dx.doi.org/10.1016/j.asoc.2014.10.001
1568-4946/

methods when an ambulance dispatching decision needs to be made [5,7,8,12,17]. Another stream of the studies is based on applying integer programming methods in a spare time. Dispatchers manage to dispatch so as to keep the ambulance configuration close to the one suggested by the lookup table, which contains the number of available ambulances and the place the ambulances should be dispatched [11]. Besides the studies by using the integer programming, some studies employ the approximate dynamic programming (ADP), which is a useful approach to optimize the ambulance dispatching. Berman [2–4] represents the first papers that provide a dynamic programming approach for the ambulance redeployment problem, and this approach was revisited recently by Zhang et al. [23]. ADP is also used to solve resource allocation problems [9,10,19] and large-scale fleet management [18,21]. However, these papers follow an exact dynamic programming formulation, and as is often the case, this formulation is tractable only in oversimplified versions of the problem with few vehicles and small transportation networks. Salmeroín and Apte [20] develop a two-stage stochastic optimization model to guide the allocation of budget to acquire and position relief assets. Maxwell et al. and Maxwell et al. [15,16] design some optimize algorithm by using the ADP approach in order to make ambulance redeployment decisions in a dynamic setting under uncertainty.

When compared with the above mentioned studies, this paper provides some advantages. In contrast to some DSSs that only consider the average response time, ours captures both the criteria on the average response time and the percentage of the ambulance requests that are responded within 15 min. We seize the random evolution of the system over time, and the stochastic nature of request arrivals, fulfillment processes, and complex traffic conditions as well as the time-dependent spatial patterns of some parameters to establish some formulae with a set of proper parameters which are based on the historical data of the request arrivals during a certain time period. Some experiments are also performed to validate the effectiveness and the efficiency of DSS.

3. Decision rules embedded in decision support systems of ambulance scheduling

3.1. Framework of the decision support systems

The decision support systems (DSS) for emergency ambulance scheduling receive requests that may come from any location at any time in a city. Because the systems need to make a decision timely, they are actually some sort of real-time DSS. The framework of the decision support systems is shown in Fig. 1 as follows.

The framework of the DSS mainly contains five core functions:

(1) Receiver of requests: It receives requests that are sent by patients through call centers, and then transforms the requests into a structured form, which is denoted by ⟨LO, TM, HP, AT, AN⟩. Here, LO is the location of the service requester; TM means an ambulance should arrive at the LO within pre-defined response time; HP denotes the set of hospitals that are suitable for treating the patient; AT denotes the type of the required ambulance, for example, type-A means an advanced vital support vehicle (SVA), and type-B means a basic vital support vehicle (SVB); AN denotes the amount of ambulances required by the service requester.

(2) Sender of instructions: It sends structured instructions ⟨AI, LA, TA, LH, TH⟩ to ambulances. Here, AI is the ambulance (instruction receiver) index; LA is the location of the accident, where the ambulance should arrive first; TA denotes the estimated time when the ambulance should arrive at LA; LH is the location of the hospital, to which the patient is transported by the ambulance; TH denotes the estimated time when ambulance should arrive at LH.

(3) Analyzer of traveling time: It can estimate the traveling time of an ambulance going through a given route. The estimation is based on the road information of the city map. The analyzer is built on two databases. One is the city map database, which stores the route length between two locations; the other is the historical database of trips, which records the estimated traveling speeds in some roads during some periods. When calculating the ambulance travel time between two given locations, the simulator first determines the areas (e.g., A_1, A_2, . . ., A_n) that the journey between the two locations will pass, and finds the velocities (e.g., $v_1, v_2, \ldots, v_n$) that relate with the above obtained areas. Then the travel time between the two locations (denoted by t) is calculated according to the lengths of the n journey segments (e.g., d_1, d_2, . . ., d_n); the formula for calculating t is as follows: $t = d_1/v_1 + d_2/v_2 + \ldots + d_n/v_n$. This formula is used in the estimation of the travel times from the current location of an ambulance to calls' scenes, and then to hospitals; it is also used in the decision process on choosing an ambulance to fulfill the task.

(4) Real-time data of ambulance: It includes the real-time data on the status, location, and undertaking task of each

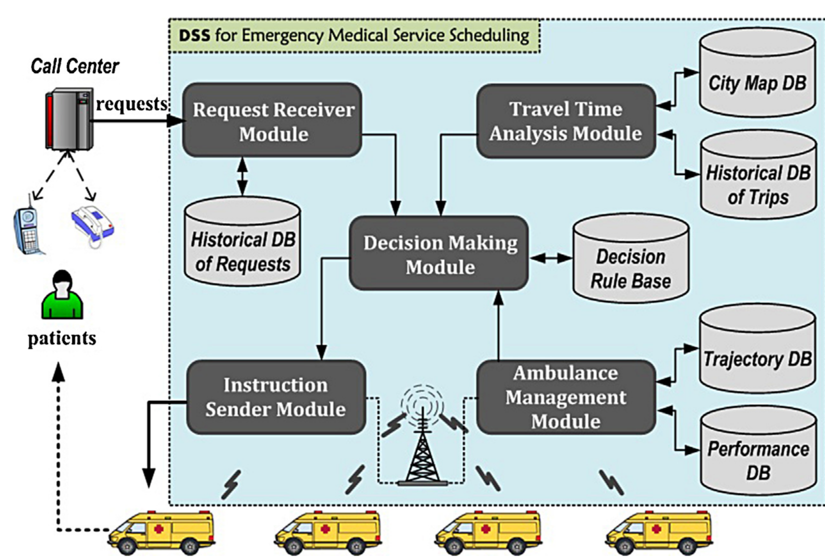

Fig. 1. The framework of the DSS for ambulance scheduling.

ambulance. For example, the ambulance status can be 'in station', 'to accident location', 'at accident location', 'to hospital', 'back to station', and etc. The real-time management of ambulance status is based on two types of databases (DB). One is the trajectory DB that stores the route information of each ambulance. The other is the performance DB that stores performance information of each ambulance.

(5) Decision making module: Its input is the requests that are delivered from the 'receiver of requests'; and its output is the instructions sent to the ambulances. The decision process between the aforementioned input and output is based on the 'analyzer of traveling time' and the 'real-time data of ambulance'. This module can tell the decision makers about which ambulance should be assigned with the task, which hospital should the ambulance transport the patient to. This decision making process is based on some decision rules, which are maintained and managed in a database named by decision rule base, which is elaborated in the next section.

3.2. Region coverage based decision rule

The most straightforward decision rule is to allocate the ambulance that is the closest to the requester. This rule can ensure an ambulance will arrive as quickly as possible. However, this closest distance based decision rule may not perform very well in universal situations. This rule only considers the estimated time of arrival to the patient's location, but ignores that the ambulance must get to a hospital and, after dropping off the patient, return to its base. Thus, while an ambulance is serving a requester, the area surrounding its base is under-covered. Therefore, this subsection proposes a decision rule with considering the region coverage.

The input of the decision rule is a request (i.e., ⟨LO, TM, HP, AT, AN⟩). The main idea for designing the rule is that if the request is very urgent, i.e., TM = 0, the ambulance that can arrive at the location of the patient (i.e., LO) in the shortest time should be assigned with the task. If the request is not very urgent, i.e., TM > 0, there may be several ambulances that can arrive at LO within the time window [0,TM]. When choosing an ambulance from these candidates, we have three criteria: (1) the earlier it can return to its base, the higher priority it is chosen; (2) the more available ambulances are idle in its base, the higher priority it is chosen; (3) the less requests may emerge in the neighborhood of its base, the higher priority it is chosen. The details on the decision rules are shown in Table 1.

In the above decision rules, α and β are two important parameters, the setting of which has influence on the performance of the rules. There is no optimal setting on the α and β parameters for all the situations. When applying the above decision rules in reality, the DSS should determine a proper setting on the α and β parameters according to the historical data of the request arrivals during a certain time period, and the closest distance based decision rule.

3.3. Centrality based decision rule

The closest distance based decision rule prioritizes requesters' calls by closeness. It only minimizes each current response time without considering long-term performances. This study develops another novel ambulance dispatching rule, named by centrality based decision rule. This rule prioritizes requesters' calls on the basis of the so-called centrality in addition to the closeness. The concept of centrality is borrowed and adopted from the studies on complex networks. The centrality reflects the importance of a node in the operational efficiency of the network [1]. Various measures have been designed for the concept of centrality. In the application area of ambulance scheduling, the centrality is some sort of measure that can reflect the efficiency of an ambulance station with respect to the coverage and response for other calls. In this way,

Table 1
The region coverage based decision rule.

Input: a request <LO, TM, HP, AT, AN>
Output: a ambulance u^* is assigned to fulfill the request

Define a set U = all the ambulances whose status is '*idle at its base*' or '*travelling to its base*'.
IF AT = A, **THEN**
  U ← set {u G U; and u.type = A}.
END IF
For all the ambulances u G U, acquire their current locations, i.e., u.location.
Obtain T_e(u.location, LO) by 'travel time analysis module'.
  //T_e is the estimated travel time between u.location and LO.
Obtain $T(u)$ on the basis on T_e by 'Ambulance management module'.
  //$T(u)$ is the estimated time for ambulance u traveling from u.location to LO.
  //$T(u) = T_e + A_u$, A_u is estimated according to u's past performances.
IF TM = 0, **THEN**
$\boldsymbol{u}^* = \text{argmin}_{u \in U} T(u)$.
ELSE
  U ← set {u G U; and $T(u)$ < TM}.
IF $U = \phi$ **THEN**
  $u^* = \text{argmin}_{u \in U} T(u)$
ELSE
IF HP = NULL, **THEN**
  HP ← the hospital that is nearest to LO
END IF
  For $\forall u \in U$, obtain $T'(u)$ and $T''(u)$.
  //$T'(u)$ is the estimated time for ambulance u traveling from LO to HP.
  //$T''(u)$ is the estimated time for ambulance u traveling from HP to its base.
  For $\forall u \in U$, calculate $C_1(u) = T(u) + T'(u) + T''(u)$.
  //$C_1(u)$ is the first criterion, which is the smaller, the better
  For $\forall u \in U$, obtain $C_2 u$), i.e., the number of available ambulances in u's base now.
  //$C_2(u)$ is the second criterion, which is the larger, the better.
  For $\forall u \in U$, obtain $C_3(u)$, i.e., the average rate for a request emerging in the 29 neighborhood of u's base.
  //$C_3(u)$ is the third criterion, which is the smaller, the better.
  //The region is a circle area with its center at u's base and radius equal to a certain value.
$u^* = \arg\min_{u \in U} \{C_1(u) - \alpha \cdot C_2(u) + \beta \cdot C_3(u)\}$.
  //α and β are parameters for the weighted sum of the three criteria.
END IF
END IF

this decision rule can improve the long-term performance of the ambulance scheduling decisions. The details on the decision rules are shown in Table 2.

4. Numerical experiments

Some numerical experiments are conducted to investigate the performance of the proposed decision rules. Fig. 2 shows the interface of a simulator that is developed for conducting the comparative experiments. According to the Poisson distribution, a number of requests are generated in the simulator. Then these generated requests are randomly located within specific locations by following the probability distribution of requests densities among different areas in Shanghai, which is the largest city by population in China. Shanghai has a population of over 23 million and a land area of about 6340 km^2. The ambulance call center in Shanghai usually receives a request and set off an ambulance every 1.2 min on average. Facing so frequent arriving ambulance requests, a good DSS on ambulance scheduling is very necessary for reducing the average response time. For patients, the first few hours are the best time (golden hours) for giving them some proper treatments. Thus the average response time for all the requests reflects the service level of a city's ambulance response DSS. In addition, the percentage of the requests that are responded in 15 min is also used as a criterion in the experiments.

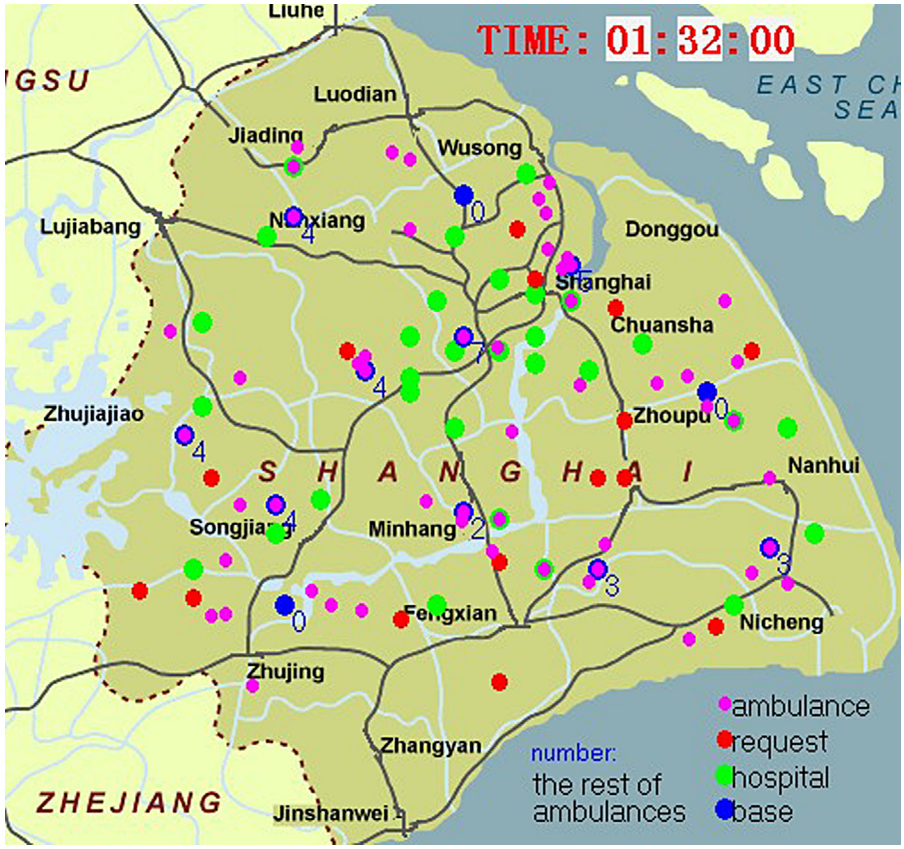

Fig. 2. An interface of the simulator in the experiments.

Before performing some comparative experiments between the proposed rules and other rules, we need to conduct some sensitivity analysis on the parameters α and β. The results are listed in Table 3. We can see that when the parameters α and β are 17 and 22, the performance is the best. Therefore, in the following experiments, the α and β are set as 17 and 22.

For validating the effectiveness of the proposed rules, some comparative experiments are conducted. The experiments are based on some comparisons with two other scheduling strategies, which are introduced as follows.

Strategy 1: For the urgent requests (i.e., TM = 0) and the non-urgent requests (i.e., TM > 0), the ambulance that can arrive at the LO in the shortest time is dispatched.

Strategy 2: For the urgent requests (i.e., TM = 0), the ambulance that can arrive at the LO in the shortest time is dispatched. For the non-urgent requests (i.e., TM > 0), the ambulance that can take patients at the LO and take them to the LH in the shortest time is dispatched.

Three series of comparative experiments are conducted by changing the number of requests, ambulances, and ambulance bases. These experiments can help to investigate the influence of the parameters on the outperformance of the proposed region coverage based rule by comparing with some traditional methods. The comparative experimental results are listed in the following tables.

The results in Table 4 show that the average time ($\bar{T}$) increases and the percentage of requests that are responded in 15 min ($P_{15\mathrm{m}}$) decreases with the number of requests growing for all the scheduling strategies. For the comparison between the proposed region coverage based rule and the two other strategies, Table 4 indicates that the region coverage based rule's $\bar{T}$ is longer than the two strategies, but it outperforms with respect to the criterion on $P_{15\mathrm{m}}$. In addition the outperformance degree of the region coverage based rule on the criterion $P_{15\mathrm{m}}$ becomes more and more evident with the number of requests growing. In reality, the criterion on $P_{15\mathrm{m}}$ is more important than the criterion on $\bar{T}$. Thus the region coverage based rule is more suitable for realistic environments than the two intuitive strategies. For the demo example in the experiments, i.e., Shanghai, the city has 1200 requests every day on average. According to the results in Table 4, it indicates that the region coverage based rule can ensure 80% of all the requests can be responded within 15 min on average.

Similar as the above experiments, the comparisons under different numbers of ambulances and ambulance bases are also performed and the results are shown in Tables 5 and 6, respectively.

The results in Table 5 show that the $\bar{T}$ decreases and the $P_{15\mathrm{m}}$ increases with the number of ambulances growing for all the scheduling strategies. For the comparison between the region coverage based rule and the two other strategies, Table 5 indicates that the region coverage based rule's performance is worse than the two other strategies on the criteria of both the $\bar{T}$ and the $P_{15\mathrm{m}}$ when the number of ambulances is not sufficient. When the number of ambulances exceeds 78, the proposed DSS's $P_{15\mathrm{m}}$ becomes larger than the two other strategies. Another finding from the results in Table 5 is that there exists a certain upper limit on $P_{15\mathrm{m}}$ for the Strategy 1 and the Strategy 2 when increasing the number of ambulances. For example, Strategy 2's $P_{15\mathrm{m}}$ cannot surpass the limit '76%' when adding more ambulances. However, the region coverage based rule's $P_{15\mathrm{m}}$ can reach a level of '85%' easily by adding more ambulances. This phenomenon shows that the region coverage based rule has a good scalability to support the expansion of the ambulance fleets so as to increase the performance of the request responding. Besides the number of ambulances, the number of ambulance bases also has influence on the performance of the strategies. Table 6 shows the comparison between the region coverage based rule and other methods under different numbers of ambulance bases.

Table 2
The centrality based decision rule.

Input: a request ⟨LO, TM, HP, AT, AN⟩
Output: a ambulance u^* is assigned to fulfill the request
Define a set U = all the ambulances whose status is '*idle at its base*' or '*travelling to its base*'.
Define a set RN = the number of requests that appear in each 1 km × 1 km subarea within 5 s
IF AT = A, **THEN**
  U ← set $\{u \in U$; and u.type = A$\}$.
END IF
  For a request r G R, obtain the current locations of the unresponsed requests 5 s before it
  in each l km × 1 km subarea, i.e., r.LO.
  Obtain RN
  IF RN >= 3
   Obtain $W(r)$ by 'Request receiver module',
   //$W(r)$ is the weighted degree of a request, which is the smaller, the better.
   Obtain $D(r)$ by 'Request receiver module',
   //$D(r)$ is the distance centrality of a request, which is the smaller, the better.
   Obtain $B(r)$ by 'Request receiver module',
   //$B(r)$ is the betweenness of a request, which is the smaller, the better.
   Calculate $r^* = \arg\min_{r \in R} \{W(r) + \lambda \cdot D(r) + \mu \cdot B(r)\}$
   //λ and μ are parameters for the weighted sum of the three criteria.
ELSE $r^* = r$
   For all the ambulances u G U, acquire their current locations, i.e., u.location.
   Obtain T_e(u.location, r*.LO) by 'travel time analysis module'.
   //T_e is the estimated travel time between u.location and r*.LO.
   Obtain $T(u)$ on the basis on T_e by 'Ambulance management module'.
   //$T(u)$ is the estimated time for ambulance u traveling from u.location to r*.LO.
   //$T(u) = T_e \pm \Delta_u$, Δ_u is estimated according to u's past performances.
IF TM = 0, **THEN**
   $u^* = \arg\min_{u \in U} T(u)$
ELSE
   $U \leftarrow$ set {u G U; and $T(u) <$ TM}.
IF $\forall u \in U$ **THEN**
   $u^* = \arg\min_{u \in U} T(u)$
ELSE
IF HP = NULL, **THEN**
   HP ← the hospital that is nearest to r*.LO.
END IF
For $\forall u \in U$, obtain $T'(u)$ and $T''(u)$.
   //$T'(u)$ is the estimated time for ambulance u traveling from r*.LO to HP.
   //$T''(u)$ is the estimated time for ambulance u traveling from HP to its base.
   For $\forall u \in U$, calculate $C_1(u) = T(u) + T'(u) + T''(u)$.
   //$C_1(u)$ is the first criterion, which is the smaller, the better
For $\forall u \in U$, obtain $C_2(u)$, i.e., the number of available ambulances in u's base now.
   //$C_2(u)$ is the second criterion, which is the larger, the better.
For $\forall u \in U$, obtain $C_3(u)$, i.e., the average rate for a request emerging in the neighborhood of u's base.
   //$C_3(u)$ is the third criterion, which is the smaller, the better.
    //The region is a circle area with its center at u's base and radius equal to a certain value.
$u^* = \arg\min_{u \in U} \{C_1(u) - \alpha \cdot C_2(u) + \beta \cdot C_3(u)\}$
    //α and β, are parameters for the weighted sum of the three criteria.
   END IF
END IF

The results in Table 6 show that the $\bar{T}$ decreases and the $P_{15\mathrm{m}}$ increases with the number of ambulance bases growing for all the scheduling strategies. For the comparison between the region coverage based rule and the two other strategies, Table 6 indicates that the region coverage based rule's performance is better than the two other strategies on the criteria of both the $\bar{T}$ and the $P_{15\mathrm{m}}$ when the number of ambulance bases is not sufficient. However, when the number of ambulance bases increases to some certain level, the region coverage based rule's $\bar{T}$ becomes worse than the two other strategies. For the criteria on $P_{15\mathrm{m}}$, the region coverage based rule always performs better than the two other strategies. This phenomenon implies that the region coverage based rule can show more evident outperformance in the situations with less ambulance bases. The requests are not distributed evenly among a city because the population densities of the residential areas in the city are different from each other. The uneven degree of the request distribution in the city may have influences on the performance of the ambulance scheduling strategies and the comparison between them. A series of experiments are performed under different uneven degrees of the request distribution. Here the uneven degree is calibrated by the standard deviation of the requests that emerge in all the subareas of the city, which is partitioned into 1 km × 1 km subareas. The results of the experiments are listed in Table 7.

Table 3
Sensitivity analysis on the parameters α and β.

α	β	$\bar{T}$	$P_{15\mathrm{m}}$	α	β	$\bar{T}$	$P_{15\mathrm{m}}$
150	22	21.4	77.3%	17	0	18.9	80.7%
100	22	21.4	77.3%	17	10	18.9	80.7%
50	22	20.3	78.5%	17	15	18.9	80.7%
25	22	19.7	78.5%	17	20	18.9	80.7%
21	22	19.8	78.5%	17	21	18.9	80.7%
20	22	19.8	78.5%	**17**	**22**	**18.8**	**80.7%**
18	22	19.7	78.5%	17	23	19.0	80.7%
17	**22**	**18.8**	**80.7%**	17	24	19.2	80.7%
16	22	18.9	80.7%	17	25	19.5	80.7%
15	22	19.8	78.5%	17	50	19.7	80.7%
10	22	19.8	78.5%	17	100	19.8	78.5%

Notes: $\bar{T}$ denotes the average response time; $P_{15\mathrm{m}}$ denotes the percentage of requests that are responded in 15 min.

Table 4
Comparison between the region coverage based rule and other methods under different numbers of requests.

Number of requests per day	Strategy 1		Strategy 2		Region coverage based rule	
	$\bar{T}$	$P_{15\mathrm{m}}$	$\bar{T}$	$P_{15\mathrm{m}}$	$\bar{T}$	$P_{15\mathrm{m}}$
700	10.9	84.7%	13.7	88.7%	14.5	91.1%
800	11.8	80.9%	14.3	85.4%	15.2	88.2%
900	11.8	79.9%	15.0	85.3%	16.5	88.0%
1000	12.6	76.4%	15.4	83.0%	17.2	86.6%
1100	13.2	72.8%	16.8	80.4%	18.5	84.5%
1200	14.6	70.3%	17.3	75.6%	18.8	80.7%
1300	15.6	65.2%	18.7	72.6%	19.0	75.6%
1400	15.9	64.6%	19.5	66.1%	20.2	70.7%
1500	17.4	56.3%	20.2	57.5%	20.6	62.3%
1600	16.1	50.3%	21.8	52.9%	22.4	58.9%

Notes: $\bar{T}$ denotes the average response time; $P_{15\mathrm{m}}$ denotes the percentage of requests that are responded in 15 min.

Table 5
Comparison between the region coverage based rule and other methods under different numbers of ambulances.

Number of ambulances	Strategy 1		Strategy 2		Region coverage based rule	
	$\bar{T}$	$P_{15\mathrm{m}}$	$\bar{T}$	$P_{15\mathrm{m}}$	$\bar{T}$	$P_{15\mathrm{m}}$
74	16.4	63.4%	17.7	74.0%	22.4	70.4%
75	15.6	67.2%	17.6	74.6%	22.0	72.9%
76	15.2	66.9%	17.5	77.1%	21.4	76.0%
77	15.1	68.8%	17.6	75.8%	21.1	74.2%
78	14.8	69.6%	17.4	76.0%	20.8	76.9%
79	14.6	70.4%	17.4	76.1%	20.5	77.6%
80	14.6	70.3%	17.3	75.6%	18.8	80.7%
81	14.3	71.1%	17.3	76.0%	19.0	81.3%
82	13.7	71.5%	17.3	76.0%	18.7	82.1%
83	13.7	71.7%	17.3	76.0%	18.9	83.6%
84	13.5	72.5%	17.3	76.0%	17.8	85.8%
85	13.4	72.7%	17.3	76.0%	17.8	85.8%

Notes: $\bar{T}$ denotes the average response time; $P_{15\mathrm{m}}$ denotes the percentage of requests that are responded in 15 min.

Table 6
Comparison between the region coverage based rule and other methods under different numbers of ambulance bases.

Number of ambulance bases	Strategy 1		Strategy 2		Region coverage based rule	
	$\bar{T}$	P_{15m}	$\bar{T}$	P_{15m}	$\bar{T}$	P_{15m}
8	24.9	42.8%	21.6	51.1%	19.3	57.8%
9	23.6	45.9%	21.6	53.6%	19.1	58.3%
10	18.6	54.5%	20.0	58.6%	19.0	63.3%
11	19.0	54.3%	18.9	66.5%	18.9	72.8%
12	14.6	70.3%	17.3	75.6%	18.8	80.7%
13	14.4	69.9%	17.2	77.1%	18.8	82.9%
14	13.5	71.7%	17.1	78.8%	18.6	83.4%
15	13.5	73.0%	17.0	78.8%	18.0	86.5%
16	13.5	73.3%	16.9	79.5%	17.9	86.5%

Notes: $\bar{T}$ denotes the average response time; P_{15m} denotes the percentage of requests that are responded in 15 min.

Different from the trends in the data tables of the previous experiments, the results in Table 7 show that the uneven degree does not always influence the $\bar{T}$ and the P_{15m} in a positive nor a negative way under all the three scheduling strategies, i.e., Strategy 1, Strategy 2, and the region coverage based rule. As shown in Table 7, when the uneven degree equals about 24, the $\bar{T}$ values for all the three strategies reach their minimum values, and the P_{15m} values for the three strategies reach their maximum values. For the comparison between the three strategies, the results are similar as the previous experimental results. Some comparative experiments between the proposed two decision rules, i.e., the region coverage based rule and the centrality based rule, are also conducted. Similar with the above performed experiments, the comparisons are also conducted under different numbers of requests, ambulances, and ambulance bases, respectively. The above three series of comparative experiments' results are illustrated in Tables 8–10, respectively.

From the results in Table 8, we can see that none of the two rules can outperform others in all the situations. When the arrival of requests is not very frequent, the region coverage based rule is a bit better than the centrality based rule with respect to the measure P_{15m}. However, when facing a large number of requests each day, the centrality based rule may be more competitive than the other decision rule.

Results in Tables 9 and 10 also show that none of the two rules can outperform others in the universal scenarios. The comparative results are influenced by the amount of resources, i.e., ambulances

Table 7
Comparison between the region coverage based rule and other methods under different uneven degrees of request distribution in the city.

Uneven degree of request distribution	Strategy 1		Strategy 2		Region coverage based rule	
	$\bar{T}$	P_{15m}	$\bar{T}$	P_{15m}	$\bar{T}$	P_{15m}
22.6	16.8	61.4%	20.0	69.5%	20.7	70.9%
23.4	13.8	69.7%	18.3	74.2%	19.3	78.0%
23.4	14.4	66.3%	18.0	73.6%	20.1	75.0%
24.0	14.1	71.6%	17.9	75.4%	19.2	80.1%
24.1	12.3	76.1%	14.1	79.4%	14.9	82.4%
26.3	13.4	70.5%	17.3	78.0%	18.8	80.8%
28.0	13.8	69.5%	17.1	76.3%	18.9	79.9%
31.3	14.6	70.3%	17.3	75.6%	18.8	80.7%
31.9	14.8	65.0%	18.3	72.0%	20.2	76.6%
32.4	16.0	64.0%	18.2	70.9%	20.6	70.7%
36.4	16.2	62.3%	21.2	68.7%	23.1	69.4%

Notes: The uneven degree is calibrated by the standard variation of requests emerging in all the subareas (e.g., 1 km × 1 km square) of the city; $\bar{T}$ denotes the average response time; P_{15m} denotes the percentage of requests that are responded in 15 min.

Table 8
Comparison between the region coverage based rule and the centrality based rule under different numbers of requests.

Number of ambulances	Region coverage based rule		Centrality based rule	
	$\bar{T}$	P_{15m}		$\bar{T}$
400	10.8	97.6%	11.1	95.5%
600	13.2	96.3%	12.7	92.5%
800	15.2	88.2%	15.4	88.3%
1000	17.2	86.6%	17.9	87.2%
1200	18.8	80.7%	18.3	83.1%
1400	20.2	70.7%	18.7	75.2%
1600	22.4	58.9%	21.5	67.9%
1800	26.3	54.8%	23.6	62.3%
2000	28.1	49.2%	26.2	59.6%

Notes: $\bar{T}$ denotes the average response time; P_{15m} denotes the percentage of requests that are responded in 15 min.

Table 9
Comparison between the region coverage based rule and the centrality based rule under different numbers of ambulances.

Number of ambulances	Region coverage based rule		Centrality based rule	
	$\bar{T}$	P_{15m}	$\bar{T}$	P_{15m}
60	25.2	62.6%	26.5	65.2%
65	23.7	65.4%	23.9	68.6%
70	23.1	70.1%	23.0	72.4%
75	22.0	72.9%	20.7	73.0%
80	18.8	80.7%	18.3	83.1%
85	17.8	85.8%	16.3	85.2%
90	17.8	85.8%	15.9	83.5%
95	17.8	85.8%	15.8	83.5%
100	17.8	85.8%	15.8	83.5%

Notes: $\bar{T}$ denotes the average response time; P_{15m} denotes the percentage of requests that are responded in 15 min.

Table 10
Comparison between the region coverage based rule and the centrality based rule under different numbers of ambulance bases.

Number of ambulance bases	Region coverage based rule		Centrality based rule	
	$\bar{T}$	P_{15m}	$\bar{T}$	P_{15m}
8	19.3	57.8%	20.5	63.8%
9	19.1	58.3%	19.6	74.3%
10	19.0	63.3%	19.5	73.9%
11	18.9	72.8%	18.1	74.7%
12	18.8	80.7%	18.3	83.1%
13	18.8	82.9%	17.3	83.5%
14	18.6	83.4%	16.9	85.1%
15	18.0	86.5%	15.3	85.5%
16	17.9	86.5%	15.3	85.5%

Notes: $\bar{T}$ denotes the average response time; P_{15m} denotes the percentage of requests that are responded in 15 min.

and ambulance bases. Generally, when the resources are abundant, the region coverage based rule is a bit better than the centrality based rule. On the contrary, if the resources are limited by some factors such as budgets, the centrality based rule shows its relative merit in scheduling ambulances.

5. Conclusions

This paper designs some decision rules of ambulance scheduling so that the ambulance requesters can be reached in a time efficient manner. Some experiments are also performed to validate the effectiveness and efficiency of the proposed rules. By comparing with the literature on the related topics, the contributions of this paper are mainly listed as follows. Most related studies on decision support systems for emergency medical scheduling only concentrate on the average response time. However, this intuitive scheduling policy cannot guarantee a high percentage of the requests that can be responded within 15 min. In reality, the criterion on the percentage of 15 min response is more important than the criterion on the average response time. Moreover, the proposed decision rules in this paper also consider a dynamic environment where the spatial distribution of potential requesters is changing along the time, and the spatial patterns of traffic situations in the metropolises are also different in peak hours and off-peak hours. The ambulance traveling and serving processes are also in a stochastic environment where the travel time for a certain journey may contain randomness; the service time at the request calls' scenes and hospitals is also uncertain. The above mentioned dynamic and stochastic nature of the request arrivals and ambulance fulfillment processes as well as the environments makes this study different from the existing studies in the related areas. There are also limitations in this study. Some parameters are contained in the decision rules. How to optimize them is an interesting issue for the further researches.

Acknowledgements

This research is supported by Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning, National Natural Science Foundation of China (Grant No. 71422007), Excellent Young Faculty Research Program in Shanghai University under the 085 Project 'Smart City and Metropolis Development'.

References

[1] M. Barthelemy, Betweenness centrality in large complex networks, Eur. Phys. J. B 38 (2) (2004) 163–168.
[2] O. Berman, Dynamic repositioning of indistinguishable service units on transportation networks, Transp. Sci. 15 (2) (1981) 115–136.
[3] O. Berman, Repositioning of distinguishable urban service units on networks, Comp. Oper. Res. 8 (2) (1981) 105–118.
[4] O. Berman, Repositioning of two distinguishable service vehicles on networks, IEEE Trans. Syst. Man Cybern. 11 (3) (1981) 187–193.
[5] L. Brotcorne, G. Laporte, F. Semet, Ambulance location and relocation models, Eur. J. Oper. Res. 147 (3) (2003) 451–463.
[6] R. Church, C. ReVelle, The maximal covering location problem, Pap. Reg. Sci. 32 (1) (1974) 101–118.
[7] M. Gendreau, G. Laporte, S. Semet, A dynamic model and parallel tabu search heuristic or real time ambulance relocation, Parallel Comp. 27 (12) (2001) 1641–1653.
[8] M. Gendreau, G. Laporte, S. Semet, The maximal expected coverage relocation problem for emergency vehicles, J. Oper. Res. Soc. 57 (1) (2006) 22–28.
[9] G.A. Godfrey, W.B. Powell, An adaptive dynamic programming algorithm for dynamic fleet management, I: single period travel times, Transp. Sci. 36 (2002) 21–39.
[10] G.A. Godfrey, W.B. Powell, An adaptive dynamic programming algorithm for dynamic fleet management. II: multiperiod travel times, Transp. Sci. 36 (2002) 40–54.
[11] A. Ingolfsson, The impact of ambulance system status management, in: Presentation, 2006NFORMS Conference, November 7, Pittsburgh, Institute for Operations Research and the Management Sciences, Hanover, MD, 2006.
[12] P. Kolesar, W.E. Walker, An algorithm for the dynamic relocation of fire companies, Oper. Res. 22 (2) (1974) 249–274.
[13] Seokcheon Lee, The role of centrality in ambulance dispatching, Decis. Support Syst. 54 (1) (2012) 282–291.
[14] B. López, B. Innocenti, D. Busquets, A multiagent system for coordinating ambulances for emergency ambulances, IEEE Intell. Syst. 23 (5) (2008) 50–57.
[15] M.S. Maxwell, S.G. Henderson, H. Topaloglu, Ambulance redeployment: an approximate dynamic programming approach, in: M.D. Rossetti, R.R. Hill, B. Johansson, A. Dunkin, R.G. Ingalls (Eds.), Proceedings of the 2009 Winter Simulation Conference, 2009.
[16] M.S. Maxwell, M. Restrepo, S.G. Henderson, H. Topaloglu, Approximate dynamic programming for ambulance redeployment, INFORMS J. Comp. 22 (2) (2010) 266–281.
[17] R. Nair, E. Miller-Hooks, A case study of ambulance location and relocation, in: Presentation, 2006 INFORMS Conference, November 6, Pittsburgh, Institute for Operations Research and the Management Sciences, Hanover, MD, 2006.
[18] W.B. Powell, H. Topaloglu, Fleet management, in: Applications of Stochastic Programming. SIAM Series in Optimization, Math Programming Society, Washinton D.C., 2005, pp. 185–215.
[19] W.B. Powell, J.A. Shapiro, H.P. Simao, A representational paradigm for dynamic resource transformation problems, Ann. Oper. Res. 104 (2001) 231–279.
[20] J. Salmeroín, A. Apte, Stochastic optimization for natural disaster asset prepositioning, Prod. Oper. Manage. 19 (5) (2010) 561–574.
[21] H.P. Simão, J. Day, A.P. George, T. Gifford, J. Nienow, W.B. Powell, An approximate dynamic programming algorithm for large-scale fleet management: a case application, Transp. Sci. 43 (2) (2009) 178–197.
[22] C. Toregas, R. Swain, C. ReVelle, L. Bergman, The location of emergency service facilities, Oper. Res. 19 (6) (1971) 1363–1373.
[23] O. Zhang, A.J. Mason, A.B. Philpott, Simulation and optimisation for ambulance logistics and relocation, in: Presentation, INFORMS 2008 Conference, October 12, Institute for Operations Research and the Management Sciences, Hanover, MD, 2008.