

# 基于机器学习的样本均值近似算法求解应急物资配置问题

□ 胡少龙<sup>1</sup> 范婷睿<sup>2</sup>

(西南交通大学 1. 经济管理学院, 四川 成都 610031; 2. 服务科学与创新四川省重点实验室, 四川 成都 610031)

**[摘要]** 应急物资配置涉及选址和库存优化等决策问题, 可为灾后物资供应提供保障。首先, 考虑应急物资需求的不确定性, 构建两阶段随机规划模型, 以优化设施选址及库存决策。然后, 提出基于机器学习的样本均值近似算法, 整合 k++ 均值聚类 and 分层随机抽样方法改善样本生成效率, 以提高样本均值近似算法求解大规模随机规划模型的效率。最后, 通过仿真实验验证了模型和算法的有效性。结果表明, 当问题规模增大, 所提出的算法能在短时间内求得较好的解。

**[关键词]** 应急物资配置; k 均值聚类; 分层随机抽样; 样本均值近似

**[中图分类号]** C931 **[文献标识码]** A **[文章编号]** 1003-1154(2022)03-0032-07

## 一、引言

洪水、地震、飓风等自然灾害易破坏基础设施, 造成大量灾民无家可归, 需快速向灾民提供应急物资。为了实现这一目标, 需在灾前选择合适的区域储备充足的物资, 以便灾后高效配送<sup>[1]</sup>。应急物资配置通常涉及设施选址和库存两个关键决策, 其中设施选址决策包括物资储备库选址或临时配送中心选址等; 库存决策指不同物资的储备量。学者多应用混合整数规划理论构建数学模型, 优化应急物资配置决策。刘波和李砚<sup>[2]</sup>、王苏生等<sup>[3]</sup>构建了双层规划模型, 优化灾后应急物资分配问题。葛春景等考虑多个受灾点同时且多次需求的情况, 构建了一个多重覆盖选址的混合整数规划模型, 优化物资配置决策<sup>[4]</sup>。于冬梅等从提高服务质量的视角, 考虑设施中断风险, 建立了选址布局网络的多目标混合整数规划模型, 优化设施选择决策<sup>[5]</sup>。

自然灾害难以预测, 受灾范围、发生时间等灾情信息具有不确定性。为保障应急物资配置效率, 需要考虑上述不确定因素。作为研究不确定环境下的最优决策理论, 随机规划是研究上述问题的有效工具。张庆和余森<sup>[1]</sup>、张梦玲等<sup>[6]</sup>、王海军等<sup>[7]</sup>、Wang 等<sup>[8]</sup>构建随机规划模型, 优化灾前物资配置和灾后物资配送等决策。Sanci 和 Daskin<sup>[9]</sup>、Moreno 等<sup>[10]</sup>还考虑了道路受损的情况, 构建随机规划模型, 优化应急物资配置和网络恢复的联合决策。Wang 等基于手机定位数据获取灾害信息构

建情景, 以优化选址和配送等决策<sup>[11]</sup>。

面对大规模问题, 优化软件和精确算法往往难以在可接受时间内求得最优解, 启发式算法具有广泛的应用场景。近年来, 越来越多的学者将机器学习与优化算法结合, 以提高求解效率。李壮年等应用 NSGA-II 遗传算法求解多目标优化问题, 分别提出了基于特征工程和支持向量机等六种机器学习算法的参数优化方法<sup>[12]</sup>。Fu 等应用统计机器学习理论降低随机变量的维数和情景尺度<sup>[13]</sup>; Guevara 等提出了一种机器学习和分布鲁棒优化相结合的方法<sup>[14]</sup>; Ghasemi 等建立了基于机器学习的仿真元模型<sup>[15]</sup>; Zhang 等提出了一个基于极限学习机的集成学习模型和多目标规划的优化框架<sup>[16]</sup>。

样本均值近似算法 (Sample Average Approximation, SAA) 是一种求解大规模随机规划问题的近似算法。Murali 等将禁忌搜索算法与 SAA 结合<sup>[17]</sup>, Aydin 和 Murat 提出了一种基于群体智能的 SAA<sup>[18]</sup>, Li 和 Zhang 等在 SAA 中引入了情景分解算法<sup>[19]</sup>, Bidhandi 和 Patrick 提出了一种基于加速采样的方法改进 SAA<sup>[20]</sup>, Jalali 等提出了一种基于 SAA 的遗传算法<sup>[21]</sup>, Jiang 等提出了一种将 SAA 与牛顿迭代相结合的方法<sup>[22]</sup>, Tao 等将 SAA 与基于种群进化的人工藻类算法相结合<sup>[23]</sup>, 以提高求解大规模随机规划模型的效率。但是, 以上研究难以保障生成最具代表性的样本, 以保证计算效率。如图 1(a) 所示, 当一个样本包含情景 A 和 C, 另一个相比包含情景 B 和 C, 则后者的代表性显然较弱。这是由于情景

**[基金项目]** 国家自然科学基金项目(72101214, 71974122); 四川省科技计划项目(2022NSFSC1115)。

B 和 C 距离较近,所含有的信息比较相似。为解决该挑战,Emelogu 等提出了一种基于机器学习的 SAA,其中机器学习用于生成样本<sup>[24]</sup>。作者利用聚类技术(如  $k$ -means)将相似情景进行聚类,然后随机从每一簇选择一个情景作为该簇的代表性情景,见图 1(b)和(c)。该算法的局限性在于,当每一簇包含的情景数量不同时,只选择一个情景难以具有足够的代表性。因此,本文的贡献是提出整合分层随机抽样解决该挑战。通过标准差决定每个簇的情景选取数量,可以改善样本生成的合理性,以提高 SAA 计算效率。

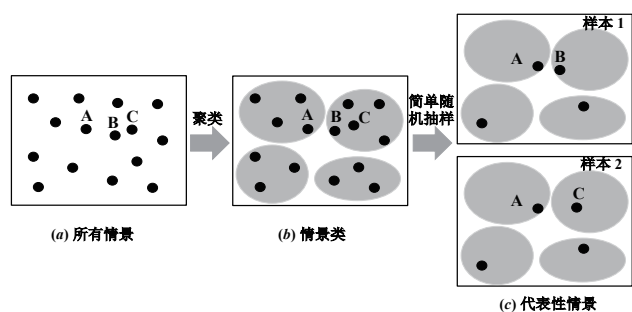


图 1 基于聚类的样本生成过程

综上所述,综合考虑多种应急物资、不同的仓储设施类别和需求的不确定性,以设施的选址、库存和运输成本,以及物资供应不足的惩罚成本最小为目标,设计决策变量分别表示仓储设施选址、库存和灾后配送等决策,构建了基于情景的两阶段随机规划模型,优化应急物资配置决策。此外,应用 SAA 求解模型,并整合了一个机器学习框架以高效地生成样本,进而提高 SAA 的性能。所提出的机器学习框架使用  $k$ -means++ 对全部情景进行聚类,再应用分层随机抽样生成样本。

其余部分内容如下。第二节,构建应急物资配置随机规划模型;第三节,设计基于机器学习的 SAA 算法;第四节,设计数值实验验证算法的有效性;最后,对研究进行总结,并讨论下一步研究方向。

## 二、应急物资配置随机规划模型

### (一) 问题描述

随机规划是一种处理优化模型输入值不确定性的技术之一,模型使用一组离散的情景表示不确定突发事件的影响范围和影响程度。例如,一种情景表示雅安发生 7 级地震,影响了周围市县 120 万人;一种情景表示 11 级台风在宁波登陆,影响了周围市县 50 万人等等。两阶段随机规划是在不确定情况下,做出非预期的第一阶段决策;第二阶段决策是在第一阶段决策和情景已知情况下

进行<sup>[25]</sup>,即如何决定应急物资配置,使得应对各种灾害情景的成本最小。具体地,构建情景集合描述不确定需求,以设施选址、采购、库存、运输和物资供应不足的惩罚成本最小的目标,优化应急物资配置。应急物资配置决策包括:仓储设施的选址和大小,每个设施中储存的各种物资数量,以及应对不同情景时物资的配送、剩余和短缺数量。

### (二) 模型构建

#### 1. 模型假设

考虑饮用水、食物、医疗包三种应急物资;不同设施的储备能力不同;不同物资的单位采购、运输、持有和惩罚成本不同,且不随情景变化。

#### 2. 符号说明

$I$  表示城市集合,由  $i, j$  索引。 $L$  表示设施类型集合,  $l \in L$  分别为大、中、小三种设施类型。 $A$  表示物资类别集合,  $a \in A$  分别为饮用水、食物、医药用品。 $S$  表示自然灾害情景集合,  $s \in S$ 。随情景变化的参数包括情景发生概率  $P_s$  和城市  $j$  对  $a$  类物资的需求  $D_{a,j,s}$ ,令未受灾城市的需求量为 0。储备设施相关参数包括  $l$  类设施的容量  $U_l$ ,  $l$  类设施的固定成本  $CF_l$ 。 $H_{i,j}$  表示城市  $i$  到  $j$  的距离,  $V_a$  表示  $a$  类物资的单位体积,  $CP_a$  表示  $a$  类物资的单位采购成本,  $CT_a$  表示  $a$  类物资的单位运输成本,  $CH_a$  表示  $a$  类物资的单位持有成本,  $G_a$  表示  $a$  类物资的单位惩罚成本。当情景  $s$  发生后,所储备物资被迅速运往各受灾城市,若库存无法满足受灾城市的物资需求,则被视为物资供应不足。对此,设置惩罚成本  $G_a$ ,对未满足物资进行惩罚。

第一阶段决策变量包括:是否在  $i$  城市建造  $l$  型仓储设施  $x_{i,l}$ ,取值分别为 0 和 1;  $i$  城市  $a$  类物资储备量  $y_{a,i}$ 。第二阶段决策变量包括:情景  $s$  下,  $i$  城市运送到  $j$  城市的  $a$  类物资的数量  $q_{a,i,j,s}$ ;情景  $s$  下,  $j$  城市的  $a$  类物资的短缺数量  $w_{a,j,s}$ ;情景  $s$  下,灾害结束后  $i$  城市  $a$  类物资的剩余量  $z_{a,i,s}$ 。

#### 3. 目标与约束函数

$$f = \min fc + pc + \sum_s P_s \cdot (tc_s + hc_s + wc_s) \quad (1)$$

$$fc = \sum_{i,l} CF_l \cdot x_{i,l} \quad (2)$$

$$pc = \sum_{a,i} CP_a \cdot y_{a,i} \quad (3)$$

$$tc_s = \sum_{a,i,j} CT_a \cdot H_{i,j} \cdot q_{a,i,j,s}, \quad \forall s \quad (4)$$

$$hc_s = \sum_{a,i} CH_a \cdot z_{a,i,s}, \quad \forall s \quad (5)$$

$$wc_s = \sum_{a,j} G_a \cdot w_{a,j,s}, \quad \forall s \quad (6)$$

$$\sum_a y_{a,i} \cdot V_a \leq \sum_l U_l \cdot x_{i,l}, \quad \forall i \quad (7)$$

$$\sum_l x_{i,l} \leq 1, \quad \forall i \quad (8)$$

$$z_{a,i,s} = y_{a,i} - \sum_j q_{a,i,j,s}, \quad \forall a, i, s \quad (9)$$

$$w_{a,j,s} = D_{a,j,s} - \sum_l q_{a,j,l,s}, \quad \forall a, j, s \quad (10)$$

$$x_{i,l} = \{0, 1\}, \quad \forall i, l \quad (11)$$

$$y_{a,i}, q_{a,i,j,s}, z_{a,i,s}, w_{a,j,s} \geq 0, \quad \forall a, i, j, s \quad (12)$$

第一阶段使设施固定成本和物资采购成本最小,即  $fc$ 、 $pc$  最小。第二阶段为当需求不确定时,运输、仓储和惩罚成本的期望最小。其中  $tc$  表示运输成本,  $hc$  表示剩余物资的库存成本,  $wc$  表示供应不足的惩罚成本。式(7)限制每个设施的采购总量不超过该设施的储存能力,式(8)限制每个城市最多只能建设一种类别的仓储设施,式(9)计算设施内应急物资的剩余数量,式(10)表示物资短缺量等于需求量减去已分配量。式(11)限定  $x$  只能取 0 或者 1,式(12)为非负约束。

### 三、基于机器学习的样本均值近似算法

样本均值近似(SAA)是一种基于蒙特卡罗模拟的随机离散优化问题的求解方法。这种方法的基本思想是生成一个随机样本,然后用相应的样本平均函数来近似期望值函数。对得到的样本均值近似问题进行求解,该过程重复多次,直到估计目标值与下界的差值低于某一阈值时,停止迭代获取满意解<sup>[26]</sup>。SAA中,更大的样本意味着估计目标值更接近真实值。但是,随着样本的增大,求解的复杂度也随之增长。因而,选择合适的样本量是SAA的关键步骤。本节提出一种机器学习框架选取情景,以生成代表性样本,提高SAA求解效率。首先,应用  $k$ -means++ 聚类方法对灾害情景进行聚类;然后,通过分层随机抽样方法获得样本。下面分别描述  $k$ -means++ 算法、分层随机抽样方法,以及改进后的SAA算法流程。

#### (一) $k$ -means++ 算法

$k$ -means 算法也被称为 Lloyd 算法,由 Stuart Lloyd 在 1957 年提出。 $k$ -means 方法是最流行的无监督学习算法之一,它遵循一种简单和相对有效的方法,将给定的数据集分类成一定数量的簇。首先,令  $K$  为中心集合,随机选择初始的  $|K|$  个中心。然后,数据集中的每个点被分配到最接近它的中心的簇中。最后,计算每个簇的均值作为中心值,并根据新的中心值重复第二步,直到中心值不变则算法收敛。与  $k$ -means 的随机选择初始中心相比,  $k$ -means++ 通过不同数据点之间的距离,迭代的选择  $|K|$  个中心,以确保更快的收敛<sup>[27]</sup>。

$k$ -means++ 算法步骤如下。

第一步:从需求数据集  $D_{a,j,s}$  中随机选取一个需求点作为聚类中心  $C_{a,j,l}$ 。

第二步:通过轮盘赌方法确定下个中心点,分别计算需求点和已选中心的距离,并计算概率。

$$pr_s = \frac{\min \{ \|D_{a,j,s} - C_{a,j,k}\| \}}{\sum_s \min \{ \|D_{a,j,s} - C_{a,j,k}\| \}}, \quad \forall s$$

作为选取依据。其中  $\|D_{a,j,s} - C_{a,j,k}\| = \sqrt{\sum_{a,j} (D_{a,j,s} - C_{a,j,k})^2}$ ;

第三步:重复第 2 步直到  $|K|$  个中心全部被选取;

第四步:针对需求数据集中每个需求点  $D_{a,j,s}$ ,计算它到  $k \in K$  个聚类中心的距离,并将其分到距离最小的聚类中心所对应的簇  $g_k$  中,  $g_1 \cup g_2 \cup \dots \cup g_{|K|} = S$ ;

第五步:对每个簇  $g_k$ ,重新计算聚类中心

$$C_{a,j,k} = \frac{1}{|g_k|} \sum_{s \in g_k} D_{a,j,s};$$

第六步:重复第 4 和 5 步,直到聚类中心位置不再变化。

#### (二) 分层随机抽样算法

由于样本是聚类后从各簇中抽取得到的,抽取方式对于样本的选取也很重要。Emelogu 等应用简单随机抽样,从每簇随机抽取一个情景<sup>[24]</sup>,忽略了不同簇中情景数量差异较大时,选取一个情景难以具有代表性的问题。因此,提出了分层随机抽样算法应对该问题。首先,根据标准差计算各簇应抽取的情景数量。然后,通过随机抽样抽取情景。为防止随机抽取的情景所对应的需求量过大或过小,通过限定其平均值,以确保样本的合理性。例如,若抽取情景所对应的需求量都较大,模型偏好选取大型设施并储备过多物资,会造成成本过大。

分层随机抽样算法步骤如下。

第一步:根据标准差计算各簇应抽取的情景数量<sup>[28]</sup>。

$$N_k = \frac{|g_k| \cdot \sigma_k}{\sum_k |g_k| \cdot \sigma_k} \cdot |N|, \quad k \in K$$

其中,  $\sigma_k$  为簇  $g_k$  的标准差,  $|N|$  表示样本的大小,  $N_k$  表示每个簇抽取的样本数量;

第二步:通过随机抽样抽取情景。令  $g_k$  表示第  $k$  簇中所抽取的情景集合,满足  $|g_k| = N_k$ ,  $\forall k$ , 且  $g_1' \cup g_2' \cup \dots \cup g_{|K|}' = N$ ;

第三步:若

$$\frac{1}{|N|} \sum_{j,s \in N} D_{a,j,s} \geq \frac{1+\gamma}{|S|} \sum_{j,s \in S} D_{a,j,s}, \quad \forall a \text{ 或}$$

$$\frac{1}{|N|} \sum_{j,s \in N} D_{a,j,s} \leq \frac{1-\gamma}{|S|} \sum_{j,s \in S} D_{a,j,s}, \quad \forall a$$

重复第二步,其中  $\gamma \in \{0, 1\}$ ;

第四步:重复第二和第三步,直到生成了  $|M|$  个样本。

#### (三) 算法对比

为对比分析 Emelogu 等提出的算法<sup>[24]</sup>与本文

提出的算法在求解大规模随机规划问题的效率,本节分别描述了两个算法,应用胡少龙等<sup>[29]</sup>提出的SAA框架。

本文提出的算法SKS: SAA,  $k$ -means++, 以及分层随机抽样, 如下。

第一步:应用  $k$ -means++ 算法;

第二步:应用分层随机抽样算法;

第三步:对任意  $m \in M$ , 求解模型

$$\hat{f}^N = \min fc + pc + \frac{1}{|N|} \sum_{s \in N} (tc_s + hc_s + wc_s)$$

约束(2) ~ (12)

其中  $\hat{f}^N$  表示样本问题的目标值。将  $x$  和  $y$  的取值作为候选解, 令其为  $(\hat{x}_1^N, y_1^N), (\hat{x}_2^N, y_2^N), \dots, (\hat{x}_{|M|}^N, y_{|M|}^N)$ ;

第四步:使用全部情景集合  $S$ , 且对任意  $m \in M$ , 求解模型

$$\hat{f}^S = \min fc + pc + \frac{1}{|N|} \sum_{s \in N} (tc_s + hc_s + wc_s)$$

约束(2) ~ (12)

$$(x, y) = (\hat{x}_m^N, y_m^N)$$

其中  $\hat{f}^S$  表示原问题的估计目标值;

第五步:选取满足式  $(x^*, y^*) \in \arg \min \{ \hat{f}^S(x, y) : (x, y) \in \{ (\hat{x}_1^N, y_1^N), (\hat{x}_2^N, y_2^N), \dots, (\hat{x}_{|M|}^N, y_{|M|}^N) \} \}$  的解作为原问题的近似解, 即  $(x^*, y^*)$ ;

第六步:使用式  $gap = \frac{\hat{f}^S(x^*, y^*) - f}{f} \cdot 100\%$  计算估计目标值与最优目标值的差异。

Emelogu 等<sup>[24]</sup>提出的算法SKR: SAA,  $k$ -means++, 以及简单随机抽样, 如下。

第一步:应用  $k$ -means++ 算法;

第二步:应用简单随机抽样, 每簇抽取一个情景;

第三步:同SKS算法的第三步至第六步。

#### 四、算例分析

##### (一) 算例构造

构造飓风灾害算例验证算法的有效性。假设有一飓风多发区域, 各城市受灾情景受飓风登陆点和飓风等级影响, 以下数据均来自已发表文献。飓风通常用萨菲

尔——辛普森级来划分,  $Catg = \{1, 2, 3, 4, 5\}$ , 其中5级最严重。需求和概率两个参数随情景变化, 其中, 需求用受灾人数衡量, 其取决于受灾城市的总人口、登陆城市和飓风等级。飓风的影响程度采用了Dalal和Üster提出的方法表示<sup>[30]</sup>:

$$DF_{i,s} = \begin{cases} 1, & d_{i,s} \leq 20 \text{ miles} \\ 0.8, & 20 \text{ miles} < d_{i,s} \leq 40 \text{ miles} \\ 0.5, & 40 \text{ miles} < d_{i,s} \leq 50 \text{ miles} \\ 0, & d_{i,s} > 50 \text{ miles} \end{cases}$$

距离因素  $DF_{i,s}$  衡量受灾人口比例随距离飓风登陆城市的远近变化。此外, 由于严重的飓风范围广、影响人数多, 飓风等级也应该着重考虑。因此, 受灾人口比为  $v_{i,s} = W_1 \cdot DF_{i,s} + W_2 \cdot \frac{Catg}{5} \Gamma_{(DF_{i,s} > 0)}$  其中  $v_{i,s} \in \{0, 1\}$ ,  $W_1$  和  $W_2$  是权重, 令  $W_1 = W_2 = 0.5$ ;  $\Gamma_{(DF_{i,s} > 0)}$  是基于  $DF_{i,s}$  的指示函数, 即若  $DF_{i,s} > 0$ ,  $\Gamma = 1$ , 否则  $\Gamma = 0$ 。由于没有证据显示飓风的登陆概率, 所以假设每个城市的登陆概率相同。由于重大灾害发生的概率通常较低, 所以假设1-5级飓风发生的概率分别为0.4、0.2、0.2、0.15、0.05。据此随机生成40个城市的人口数量, 以及500个飓风情景。

基于Rawls和Turnquist的研究<sup>[31]</sup>设立如下数据。考虑三种应急物资: 水、食品和医疗包。假设水的单位是1000加仑; 食物为速食产品, 以1000餐为单位; 医疗包为每人一个单位。表1总结了应急物资的关键参数。假设每一种物资未满足的惩罚成本为采购价格的10倍, 持有成本为购买价格的20%。大、中、小三种不同存储容量的设施的固定成本和储存容量, 见表2。

表1 物资的单位采购价格、所占仓储量和运输成本

物资	水	食品	医疗包
单位占有量 (立方米)	144.6	83.33	1.16
单位采购成本 (元)	647.7	5420	140
单位运输成本 (元)	0.3	0.04	0.0006

表2 不同类别设施的固定成本和储存容量

	小型	中型	大型
固定成本 (元)	19600	188400	300000
存储容量 (立方米)	36400	408200	780000

##### (二) 结果分析

表3 三种算法的计算时间和估计目标值

I, S	M, N	计算时间 (秒)			目标值 / Gap		
		Gurobi	SKR	SKS	Gurobi (万)	SKR (%)	SKS (%)
20, 100	5, 5	19	33	33	6070	4.57	1.03



I, S	M, N	计算时间 ( 秒 )			目标值 /Gap		
		Gurobi	SKR	SKS	Gurobi ( 万 )	SKR ( % )	SKS ( % )
20, 100	5, 10	19	36	34	6070	3.61	4.01
	10, 5		56	56		1.16	0.98
	10, 10		62	58		3.63	0.31
20, 200	5, 10	40	57	53	6841	3.70	0.33
	5, 20		58	62		11.15	0.35
	10, 10		99	98		3.04	0.37
	10, 20		110	102		4.01	0.19
20, 500	5, 25	135	120	119	7107	9.58	0.13
	5, 50		143	122		9.20	0.00
	10, 25		224	240		8.33	0.08
	10, 50		278	250		9.21	0.03
40, 100	5, 5	92	106	86	12433	2.25	1.91
	5, 10		152	110		3.90	0.17
	10, 5		433	161		3.96	2.04
	10, 10		310	190		3.87	0.17
40, 200	5, 10	3313	274	212	12092	0.15	0.17
	5, 20		633	449		1.04	0.15
	10, 10		379	319		0.09	0.10
	10, 20		734	800		0.57	0.18
40, 500	5, 25	3600	822	595	19025	4.67	-0.40
	5, 50		1974	1068		-0.57	-0.56
	10, 25		2154	1010		3.77	-0.44
	10, 50		4808	3518		-0.57	-0.50

算法通过 Matlab 进行编程,调用 Gurobi 9.0.1 求解。求解计算机 CPU 为 Intel(R) Core(TM) i9-9900 3.10GHz,内存为 32GB。从计算时间、Gap 值两个角度对比分析 Gurobi、SKR、SKS 的求解效率。令  $Obj_{Gurobi}$  ,  $Obj_{SKR}$  和  $Obj_{SKS}$  分别为应用 Gurobi、SKR 和 SKS 求解模型后得到的目标函数值,则有  $Gap_{SKR} = \frac{Obj_{SKR} - Obj_{Gurobi}}{Obj_{Gurobi}} \cdot 100\%$  和  $Gap_{SKS} = \frac{Obj_{SKS} - Obj_{Gurobi}}{Obj_{Gurobi}} \cdot 100\%$ 。应用 Gurobi、SKR 和 SKS 分别求解城市数量为 20 和 40, 情景数量分别为 100、200、500 时的算例,得到不确定需求下应急物资配置的最小成本和计算时间,见表 3,其中  $I$  为城市集合、 $S$  为情景集合、 $M$  为样本集合、 $N$  为样本情景集合。随着问题规模增大,Gurobi 难以求解,故设置最大计算时间 3600 秒为求解的终止条件。如表 3 所示,

城市数量一定时,随着情景数量增长,Gurobi 计算时间也随之陡增。当城市数量为 40、情景数为 500 时,3600 秒内 Gurobi 无法求出最优解。因此,引入算法求解大规模随机规划问题非常必要。尽管 SKR 和 SKS 计算时间亦呈增长趋势,但增长幅度明显更少。

为便于直观比较,计算不同算例 SKR 和 SKS 的平均求解时间,见图 2。显然,对于小规模算例,Gurobi 具有一定的计算优势。而当数据量增大时(城市数量为 40,且情景数量超过 100),Gurobi 求解时间陡然增大。SKR 和 SKS 的计算时间增幅较小,计算效率明显更高。此外,与 SKR 相比,当城市数量为 20 时,SKS 求解速度较快的优势并不明显。但是,当城市数量增大到 40 时,SKS 的计算时间一直较少,表明其计算效率更高。当  $M$  不变, $N$  值增大意味着情景数

量变多,算法的求解时间主要取决于样本问题的复杂性。例如,当  $I=40$ 、 $S=500$ 、 $M=5$  时,50 个情景的样本问题显然比 25 个情景的更难求解。可通过整合更高效的算法求解样本问题,以提高 SKR 和 SKS 的计算效率。当  $N$  不变,  $M$  值增大意味着样本问题个数增多,求解时间则主要取决于样本问题的计算时间。例如,当  $I=40$ 、 $S=500$ 、 $N=25$  时,10 个样本问题显然要比 5 个求解时间更长。可通过整合并行计算,以提高两种算法的计算效率。

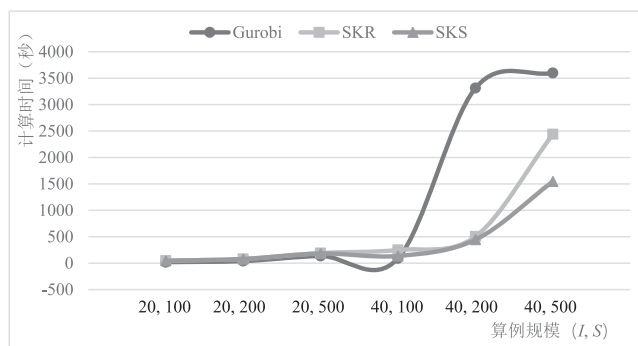


图2 Gurobi、SKR 和 SKS 平均计算时间的对比

以 Gurobi 求解结果为基准,计算 Gap 值并比较 SKR 和 SKS 的计算效率。如表 3 所示,对于算例  $I=40$ 、 $S=500$ , Gurobi 在 3600 秒内无法求得最优解,SKR 和 SKS 可以在较短时间内找到更好的上界。因此,当问题规模增大到一定程度后,SKR 和 SKS 在计算效率上都优于 Gurobi。但与 SKR 不同的是,无论  $M$  和  $N$  值为多少,SKS 都能找到更好的上界。对于不同算例,与 Gurobi 求解结果相比,SKR 的 Gap 在  $-0.57\%$  至  $11.15\%$  之间变化;而 SKS 仅在  $-0.56\%$  至  $4.01\%$  之间变化。为更直观的对比不同算例下 SKR 和 SKS 的 Gap 值的变化,将不同算例下,两个算法的结果求平均值,其结果见图 3。对于所有算例,显然 SKS 所求得的上界都显著优于 SKR。因此,通过对比分析计算时间和 Gap 值发现,SKS 明显优于 SKR。

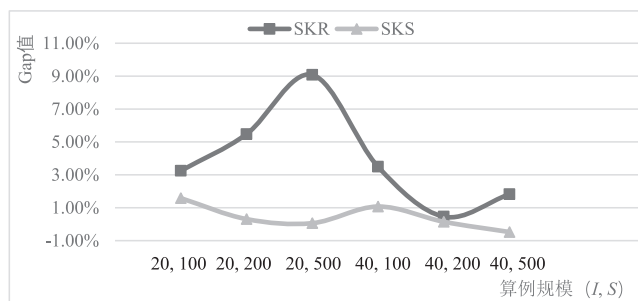


图3 SKR 和 SKS 平均 Gap 值的对比

## 五、结论

合理布局应急物资储备库并存储物资可以在灾害发生后及时做出响应,以缩短物资筹备及运输的时间。通

过最大限度地提高应急物资供给能力,可以保障人民生命财产安全,降低突发事件所造成的损失。

考虑不确定需求,研究了应急物资配置的模型与算法,以优化应急物资储备库选址与库存决策。以设施选址、采购、库存、运输和供应不足的惩罚成本最小为目标,构建了应急物资配置的两阶段随机规划模型。同时,整合  $k$ -means++ 聚类算法和分层随机抽样,设计了一个机器学习框架生成样本,以改进样本均值近似方法,提高其在求解大规模随机规划问题的计算效率。最后,构建了灾害情景对模型和算法进行了仿真验证,通过与 Gurobi 和其他算法的对比分析表明,随着算例规模增大,所提出的算法能在较短时间找到更好的上界。

尽管为求解大规模随机规划问题提供一条新的途径,但研究仍存在不足,可从以下方面进一步完善。首先,仅考虑了单一的灾害后果,未考虑到交通受阻、设施受损等的影响,需在后续的研究中加以讨论。其次,如何有效确定簇的数量。 $k$ -means++ 算法容易理解,聚类效果好,尤其是在处理大数据集的时候,可以保证较好的伸缩性和高效率。但是,初始  $k$  值需要人为设定,该数值选取常依靠经验,可能带来较大误差。仅根据样本数量设定  $k$  值,需要进一步改进。最后,未对比其他聚类算法的效果。下一步研究可以增加不同的聚类算法,通过实验对比分析各聚类方法与 SAA 结合后的优劣。

## [参考文献]

- [1] 张庆,余森.基于两阶段混合整数规划模型的洪涝灾害应急管理研究[J].运筹与管理,2020,29(09):62-69.
- [2] 刘波,李砚.应急物资车辆调度的鲁棒双层优化模型[J].系统工程,2016,34(05):77-81.
- [3] 王苏生,王岩,孙健,等.连续性条件下的多受灾点应急资源配置算法[J].系统管理学报,2011,20(02):143-150.
- [4] 葛春景,王霞,关贤军.重大突发事件应急设施多重覆盖选址模型及算法[J].运筹与管理,2011,20(05):50-56.
- [5] 于冬梅,高雷阜,赵世杰.考虑应急设施中断风险与防御的可靠性选址模型研究[J].运筹与管理,2020,29(09):53-61.
- [6] 张梦玲,王晶,黄钧.不确定需求下考虑供应商参与机制的应急资源配置鲁棒优化研究[J].中国管理科学,2020,28(07):102-111.
- [7] 王海军,刘畅,王婧.应急储备库选址与资源配置随机规划模型研究[J].管理学报,2013,10(10):1507-1511+1519.
- [8] Wang Y, Dong Z S, Hu S. A Stochastic Prepositioning Model for Distribution of Disaster Supplies Considering Lateral Transshipment[J]. Socio-Economic Planning Sci-

ences, 2021, 74: 100930.

[9] Sanci E, Daskin M S. Integrating Location and Network Restoration Decisions in Relief Networks Under Uncertainty[J]. *European Journal of Operational Research*, 2019, 279(02): 335-350.

[10] Moreno A, Alem D, Ferreira D, et al. An Effective Two-stage Stochastic Multi-trip Location-Transportation Model with Social Concerns in Relief Supply Chains[J]. *European Journal of Operational Research*, 2018, 269(03): 1050-1071.

[11] Wang J, Cai J, Yue X, et al. Pre-positioning and Real-time Disaster Response Operations: Optimization with Mobile Phone Location Data[J]. *Transportation Research Part E: Logistics and Transportation Review*, 2021, 150: 102344.

[12] 李壮年, 储满生, 柳政根, 等. 基于机器学习和遗传算法的高炉参数预测与优化[J]. *东北大学学报(自然科学版)*, 2020, 41(09): 1262-1267.

[13] Fu X, Guo Q, Sun H. Statistical Machine Learning Model for Stochastic Optimal Planning of Distribution Networks Considering a Dynamic Correlation and Dimension Reduction[J]. *IEEE Transactions on Smart Grid*, 2020, 11(04): 2904-2917.

[14] Guevara E, Babonneau F, Homem-de-Mello T, et al. A Machine Learning and Distributionally Robust Optimization Framework for Strategic Energy Planning Under Uncertainty[J]. *Applied Energy*, 2020, 271: 115005.

[15] Ghasemi A, Ashoori A, Heavey C. Evolutionary Learning Based Simulation Optimization for Stochastic Job Shop Scheduling Problems[J]. *Applied Soft Computing*, 2021, 106: 107309.

[16] Zhang Y, Xu Y, Dong Z Y, et al. Intelligent Early Warning of Power System Dynamic Insecurity Risk: Toward Optimal Accuracy-earliness Tradeoff[J]. *IEEE Transactions on Industrial Informatics*, 2017, 13(5): 2544-2554.

[17] Murali P, Ordóñez F, Dessouky M M. Facility Location Under Demand Uncertainty: Response to a Large-scale Bio-terror Attack[J]. *Socio-Economic Planning Sciences*, 2012, 46(01): 78-87.

[18] Aydin N, Murat A. A Swarm Intelligence Based Sample Average Approximation Algorithm for the Capacitated Reliable Facility Location Problem[J]. *International Journal of Production Economics*, 2013, 145(01): 173-183.

[19] Li X, Zhang K. A Sample Average Approximation Approach for Supply Chain Network Design with Facili-

ty Disruptions[J]. *Computers & Industrial Engineering*, 2018, 126: 243-251.

[20] Bidhandi H M, Patrick J. Accelerated Sample Average Approximation Method for Two-stage Stochastic Programming with Binary First-stage Variables[J]. *Applied Mathematical Modelling*, 2017, 41: 582-595.

[21] Jalali S, Seifbarghy M, Niaki S T A. A Risk-averse Location-protection Problem Under Intentional Facility Disruptions: A Modified Hybrid Decomposition Algorithm[J]. *Transportation Research Part E: Logistics and Transportation Review*, 2018, 114: 196-219.

[22] Jiang G, Xu C, Fu M C. On Sample Average Approximation Algorithms for Determining the Optimal Importance Sampling Parameters in Pricing Financial Derivatives on Lévy Processes[J]. *Operations Research Letters*, 2016, 44(01): 44-49.

[23] Tao S, Wu C, Sheng Z, et al. Stochastic Project Scheduling with Hierarchical Alternatives[J]. *Applied Mathematical Modelling*, 2018, 58: 181-202.

[24] Emelogu A, Chowdhury S, Marufuzzaman M, et al. An Enhanced Sample Average Approximation Method for Stochastic Optimization[J]. *International Journal of Production Economics*, 2016, 182: 230-252.

[25] Birge J R, Louveaux F. *Introduction to Stochastic Programming*[M]. Springer, New York, NY, 1997.

[26] Kleywegt A J, Shapiro A, Homem-de-Mello T. The Sample Average Approximation Method for Stochastic Discrete Optimization[J]. *SIAM Journal on Optimization*, 2002, 12(02): 479-502.

[27] Arthur, D, Vassilvitskii S. K-meanspp: the Advantages of Careful Seeding. In: *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*[C]. Society for Industrial and Applied Mathematics, Louisiana, New Orleans, 2007: 1027-1035.

[28] Cochran W G. *Sampling Techniques*[M]. New York: John Wiley & Sons, 1977.

[29] 胡少龙, 韩传峰, 孟令鹏, 等. 考虑企业生产能力储备的应急物资配置随机规划模型[J]. *系统工程理论与实践*, 2018, 38(06): 1536-1544.

[30] Dalal J, Üster H. Combining Worst Case and Average Case Considerations in an Integrated Emergency Response Network Design Problem[J]. *Transportation Science*, 2018, 52(01): 171-188.

[31] Rawls C G, Turnquist M A. Pre-positioning of Emergency Supplies for Disaster Response[J]. *Transportation Research Part B: Methodological*, 2010, 44(04): 521-534.