

论著

# 基于离散型麻雀搜索算法的食品抽检路径优化

王建新<sup>1,2</sup>,李腾旭<sup>1,2</sup>,王晔茹<sup>3</sup>

(1. 北京林业大学信息学院,北京 100083; 2. 国家林业草原局林业智能信息处理工程研究中心,北京 100083; 3. 国家食品安全风险评估中心,北京 100022)

**摘要:****目的** 提出一种基于离散型麻雀搜索算法的食品抽检路径高效优化方法。**方法** 通过对抽检点编码,不同编码间路径计算及优化,构建离散型麻雀搜索算法并与其他已知算法进行比较与验证。**结果** 本研究构建的离散型麻雀搜索算法,对于 Burma14、Bays29、Oliver30 和 Att48 等实例,本算法都可以求得已知最优解。对于 Kioa100 和 Ch130 等实例,本算法得到的偏差率,分别是 0.1% 和 1%,优于传统的遗传算法(偏差率分别是 3% 和 4.2%)。**结论** 本研究创建的基于离散型麻雀搜索算法的食品抽检路径优化方法,在求解精度和收敛速度方面有更好的表现,有助于双随机抽检点生成和抽检路径优化的实际工作,为“双随机”抽检信息系统的实用化提供了可行的算法支撑。

**关键词:**食品抽检;双随机;路径优化;旅行商问题;离散型麻雀搜索算法

**中图分类号:**R155.5 **文献标识码:**A **文章编号:**1004-8456(2021)04-0409-06

**DOI:**10.13590/j.cjfh.2021.04.001

## Optimization of food sampling inspection based on discrete sparrow search algorithm

WANG Jianxin<sup>1,2</sup>, LI Tengxu<sup>1,2</sup>, WANG Yeru<sup>3</sup>

(1. School of Information, Beijing Forestry University, Beijing 100083, China;

2. Engineering Research Center for Forestry-oriented Intelligent Information Processing of National Forestry and Grassland Administration, Beijing 100083, China;

3. China National Center for Food Safety Risk Assessment, Beijing 100022, China)

**Abstract:** **Objective** To propose an efficient optimized method for food sampling inspection based on discrete sparrow search algorithm. **Methods** By coding the sampling points, calculating and optimizing the paths between different codes, a discrete sparrow search algorithm was constructed, compared and verified with other known algorithms. **Results** The discrete sparrow search algorithm constructed in this research could obtain the known optimal solution for Burma14, Bays29, Oliver30 and Att48, etc. For examples such as Kioa100 and Ch130, the deviation rates obtained by this algorithm were 0.1% and 1%, respectively, which were better than traditional genetic algorithms (the deviation rates were 3% and 4.2%, respectively). **Conclusion** The food sampling path optimization method based on the discrete sparrow search algorithm created in this research had better performance in terms of solution accuracy and convergence speed, which was helpful to the actual work of double random sampling point generation and sampling path optimization. It provided feasible algorithmic support for the random inspection information system.

**Key words:** Food sampling inspection; double random; path optimization; travelling salesman problem; discrete sparrow search algorithm

民以食为天,食品安全已成为关系国计民生的重大问题<sup>[1]</sup>,而食品安全抽检是保障食品质量安全

的重要手段<sup>[2]</sup>。由于食品安全抽检流程复杂,导致目前食品检测的环节中存在着问题与不足<sup>[3]</sup>。为此,政府希望抽检主体的选择可以更加公正<sup>[4]</sup>,进而提出了双随机标准。双随机标准是指在抽检过程中随机选取抽检对象和抽检人员。实行双随机标准之前,抽检人员通常对那些抽检方便易行的抽检对象进行抽检;但在实行双随机标准后,对随机生成的被检主体集合,需要完全覆盖。因此,如何满足覆盖随机化生成的被检对象的同时,尽量降低

收稿日期:2021-06-14

基金项目:国家重点研发计划(2017YFC1602002,2018YFC1603305);国家

食品安全风险评估中心高层次人才队伍建设 523 项目

作者简介:王建新 男 教授 研究方向为人工智能

E-mail: wangjx@bjfu.edu.cn

通信作者:王晔茹 女 副研究员 研究方向为食品安全和风险评估

E-mail: wangyeru@cfsa.net.cn

人力、物力和财力的浪费,是保证双随机标准执行的可行性和可持续性的一个重要前提。

为了解决上述问题,一个常见的操作是“进一次门,办多件事”。在随机生成的被检对象列表中,如果有重复,例如抽检某个超市中两种或以上的被抽检商品,则需要事先合并,并需要信息系统来安排抽检路线,这样不仅能使抽检过程更加高效,也能对抽检记录实现规范化的留档保存。而系统要实现高效的抽检安排需要在多个抽检点上做路径规划,在计算机领域,这是一个典型的旅行商问题。

旅行商问题 (Travelling sales man problem, TSP)<sup>[5]</sup>是组合优化中典型的 NP-难问题,具有深远的实际意义,在机器人控制<sup>[6]</sup>、电路布线<sup>[7]</sup>等方面都有着广泛的应用。当问题规模较小时,可以求得精确解<sup>[8]</sup>。但是当规模增大时,会带来爆炸式组合增长问题<sup>[9]</sup>。为了解决这一难题,涌现出了一些新兴算法,如遗传算法<sup>[10]</sup>、帝国竞争算法<sup>[11]</sup>等。针对不同的具体问题,这些算法的求解性能有不同的表现,尚未有一个在精度、效率、易实现、易操作等方面都优秀的算法,都在完善和提高的过程中。

麻雀搜索算法是 2020 年由 XUE 和 SHEN<sup>[12]</sup>提出的用于求解连续优化问题的群智能优化算法。本文在上述算法的基础上,提出了一种用于求解离散问题的麻雀搜索算法。经实验表明,该算法在 TSP 上表现优异,相较于传统优化算法具有更高的准确率和更快的收敛速度,并成功在食品抽检路径优化方面进行了应用。

## 1 材料与方法

### 1.1 材料

麻雀搜索算法 (Sparrow search algorithm, SSA) 根据现实中麻雀种群的觅食规律,通过探索食物、追随发现者和反捕食三个过程来使问题不断接近最优解。

在 SSA 中,高适应度的麻雀叫做发现者,低适应度的麻雀为追随者。发现者来为整个种群探索可能存在食物的地方,当种群收到警报时,发现者要改变原来的方向。

追随者根据适应度的高低分为两部分。高适应度的部分跟随发现者飞行。而低适应度的追随者,需要探索其他方向。这个操作使得状态搜索增加了更多的可能性。

反捕食是指当意识到危险的麻雀的适应度较差,会向适应度高的麻雀靠拢。反之,会向低适应

度麻雀靠拢。这种反捕食操作能够增加种群的多样性。

### 1.2 方法

#### 1.2.1 求解 TSP 的离散型麻雀搜索算法

麻雀搜索算法目前用于求解连续型问题,不能直接应用于离散型问题。本文提出的离散型麻雀搜索算法,可用于求解 TSP 类的离散型问题。具体算法流程如图 1 所示。

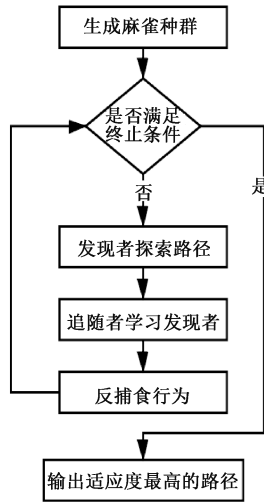


图 1 算法流程

Figure 1 Algorithm flow

#### 1.2.2 表示方式和适应度计算

离散型麻雀搜索算法需要先对抽检点进行编码,能逐个无重复地访问完所有抽检点后形成的编码序列为一只麻雀。多只麻雀即多个编码序列构成了算法的初始种群。可以使用随机方法形成初始种群。

适应度函数能够引导算法逐渐逼近最优解。本文利用路径长度反映适应度的大小,长度越小,适应度越高。因此,适应度函数定为路径长度的倒数,如下:

$$f_i = \frac{1}{p_i} \quad (1)$$

其中 $f_i$ 代表的是第 $i$ 只麻雀的适应度, $p_i$ 代表的是抽检员按第 $i$ 只麻雀所对应路径进行抽检所花费的总成本。

#### 1.2.3 发现者探索食物

区分麻雀群体中发现者和追随者的依据是麻雀个体的适应度值。发现者的任务是探索食物。下面算法是发现者探索食物的算法步骤。事先设定预警值 $w$ 的初始值,在区间 $(0,1)$ 中,随着算法的迭代,这个预警值需要更新。

(i) 首先选择适应度较高的前 $s$ (是一个比例,在 0 和 1 之间)部分作为发现者,其他部分作为追

随者,适应度较高的发现者能够帮助算法尽快收敛。

(ii) 生成一个区间 $[0,1)$ 中的随机数  $r$ ,当  $r$  小于设定的预警值  $w$  时,发现者需要大范围的路径调整;否则转入步骤(iii)。本文提出并使用了两个优化算子辅助调整。

优化算子 1:  $C = \{C_1, C_2, \dots, C_i, C_{i+1}, \dots, C_{j-1}, C_j, \dots, C_n\}$  为 TSP 问题的一条路径。随机选择两个位置  $i$  和  $j$ ,然后将这两个位置之间的路径变为倒序,则新的路径为:  $C_{\text{new}} = \{C_1, C_2, \dots, C_j, C_{j-1}, \dots, C_{i+1}, C_i, \dots, C_n\}$ 。若新路径的适应度大于原路径,则将原路径替换,否则保留原路径。

优化算子 2:为了加快局部收敛速度,同时为了尽可能少的破坏已形成的优化路径,本文使用优化算子 2 进行再优化。首先将路径分为  $k$  段,设  $C = \{C_{h,1}, C_{h,2}, \dots, C_{h,i}, C_{h,i+1}, \dots, C_{h,j-1}, C_{h,j}, C_{h,j+1}, C_{h,j+2}, \dots, C_{h,m}\}$  为其中一段的第  $h$  段路径,在这段路径中随机找到两个位置,设为  $i$  和  $j$ ,交换  $C_{h,i}$ 、 $C_{h,i+1}$  和  $C_{h,j+1}$ 、 $C_{h,j+2}$  的位置,则新的路径为  $C_{\text{new}} = \{C_{h,1}, C_{h,2}, \dots, C_{h,j+1}, C_{h,j+2}, \dots, C_{h,j-1}, C_{h,j}, C_{h,i}, C_{h,i+1}, \dots, C_{h,m}\}$ 。若新的路径适应度大于原来的路径,则替换原路径,否则不替换。

(iii) 当  $r$  大于  $w$  时,发现者只需要进行局部的调整。做法是把一条路径分为  $k$  段,需要保证每段中的抽检点数目大于 2,再对每一段使用优化算子 1 进行优化。当  $k$  值设置过大时会增加算法的运行时间,且使优化算法的视野变窄,导致会破坏已形成的较优路径片段。

(iv) 调整参数  $w$  的值。经过大量实验验证表明,预警值  $w$  的大小对于求解的精度有影响。故这里  $w$  设置为一个随迭代次数变化的值。 $w$  的定义如下:

$$w = \max - \frac{t(\max - \min)}{M}$$

(2)

其中  $\max$ 、 $\min$  代表着预警值  $w$  的最大值、最小值,当迭代次数较少时,发现者的适应度相对较差,进而更加需要大范围的路径调整;随着迭代次数的增加,发现者的适应度会不断提高,同时也为了保护已形成的部分较优路径片段,应该更适合进行局部的路径调整。迭代次数  $t$  所处的区间为  $[1, M]$ 。

1.2.4 追随者学习发现者

由于发现者比追随者解的质量更高,所以在迭代过程中,追随者需要向发现者学习。学习的算法步骤如下。

(i) 根据适应度对麻雀进行排序后,需要再计算出饥饿度的阈值。这里阈值的计算公式如下:

$$T = \frac{p \cdot (1 + s)}{2}$$

(3)

其中  $p$  代表的是种群数量,  $s$  代表的是发现者占种群的比例,  $T$  是经过计算所获得的阈值的大小。

(ii) 当追随者排名位于阈值之后时,需要学习发现者的路径。如果路径中有重复或遗漏,则需要调整至无重复遗漏的形式。具体操作示例如图 1 所示。

(iii) 当追随者位于阈值之前,说明当前路径的适应度较低,故使用 2.1 节中的方法,用随机方式重新生成当前麻雀的序列。

1.2.5 反捕食过程

反捕食行为可以防止路径因为陷入局部最优而过早收敛。在离散型麻雀算法中,反捕食行为需要经过以下几个步骤:

(i) 每一代中都会随机找到  $p \cdot danger$  只麻雀进行反捕食行为,  $danger$  为种群中感知到危险的麻雀所占的比例,是一个输入参数,当参数值过大时,会影响算法收敛的速度,而当参数值过小时,可能会使算法过早收敛。

(ii) 当随机取到的麻雀是发现者时,收到危险预警的发现者需要使用图 1 的算法向追随者学习路径。同时,向追随者学习能够使发现者保留相对高的适应度,不至于延缓算法收敛的进度。

(iii) 当随机到的麻雀是追随者的时候,为了加快收敛,追随者需要向某一只发现者学习路径,如图 2 所示。

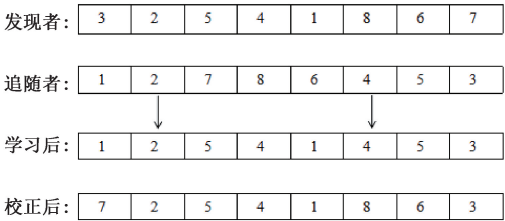


图 2 追随者的学习路径

Figure 2 Followers' learning path

2 结果

为了验证离散型麻雀搜索算法 (Discrete sparrow search algorithm, DSSA) 的有效性,本文选用旅行商问题实例库 (Travelling sales man problem library, TSPLIB) 实例进行实验,并将实验结果同已知的算法进行比较。TSPLIB 是验证 TSP 算法效果的国际通用实例,将城市的位置映射到坐标系中,使得可以通过欧式距离来计算出各个城市间的距离。可以使用 TSPLIB 中的城市来模拟实际抽检工作中的抽检点,这样比随机生成的实验数据具有更好的比较基准。

2.1 算法的实验环境、参数设置和伪代码

实验环境如下:系统为 windows7,内存为 4G, CPU 为 Intel i5-4200U 1.60 GHz, 编程软件为 Eclipse,开发环境为 JDK1.8。

算法参数取值如表 1 所示。

表 1 参数设置

Table 1 Parameter settings

参数	<i>p</i>	<i>s</i>	max	min	<i>M</i>	<i>danger</i>	<i>k</i>
定义	种群数目	发现者的比例	最大预警值	最小预警值	最大迭代次数	反捕食的比例	算子分段数目
值设定	200	0.2	0.8	0.2	500	0.1	2

2.2 对比实验

本文选用遗传算法 (Genetic algorithm,GA) 来进行对比,并选用 2-opt 优化算子<sup>[13]</sup>对遗传算法每一代的最优解进行优化。为了保证两种算法对比的公平性,根据实验和尝试,遗传算法采用最优参数设定:种群数目设为 200,最大迭代次数设为 500,突变率设为 0.05,精英个体数量设为 10。

表 2 的数据集是 TSPLIB(旅行商问题实例),实例名字中的数字对应着参与实验的抽检点数量,例如 Burma14 代表有 14 个抽检点参与抽检;已知最优解为该实例目前所知的最短路径长度;最优解是算法在 20 次运行中得到的最短路径长度。偏差率的计算公式如下:

偏差率 =  $\frac{\text{最优解} - \text{已知最优解}}{\text{已知最优解}} \times 100\%$  (4)

表 2 GA 与 DSSA 的实验结果对比

Table 2 Comparison of experimental results between GA and DSSA

数据集	已知最优解	GA		DSSA	
		最优解	偏差率 /%	最优解	偏差率 /%
Burma14	30.88	30.88	0.00	30.88	0.00
Bays29	9 074.15	9 076.98	0.03	9 074.15	0.00
Oliver30	423.74	423.95	0.05	423.74	0.00
Att48	33 523.71	34 042.18	1.55	33 523.71	0.00
Eil51	426.00	435.08	2.13	428.87	0.67
Berlin52	7 542.00	7 800.67	3.43	7 544.37	0.03
St70	675.00	686.05	1.64	677.10	0.31
Kioa100	21 282.00	21 918.12	2.99	21 307.42	0.12
Ch130	6 110.00	6 363.97	4.16	6 175.61	1.07
A280	2 579.00	2 788.22	8.11	2 689.42	4.28

通过对选取的实例分别使用 GA 和 DSSA 进行 20 次独立实验,由于部分最优解是截尾的整型表示,故所求偏差率比实际稍大。可以看出,对于 Burma14、Bays29、Oliver30 和 Att48 这些实例,DSSA 都可以求得已知最优解。对于 Kioa100 和 Ch130 等实例,在算法若能达到最优解,其偏差率分别是 0.1%和 1%,但优于传统国遗传算法(偏差率分别

是 3%和 4.2%)。在图 3 中可以看出,对于 Att48、Kioa100 和 Ch130,使用 DSSA 得到的结果在最优解、偏差率方面都优于 GA。因此,当抽检点个数增多时,DSSA 具有更高的求解精度和稳定性。另外,表 2 中抽检点的数目从 14 增加到 280,可以模拟实际食品抽检中同一区域下的小规模抽检到跨区域的大规模抽检。

可以通过部分实例的最优路线图以及 DSSA 和 GA 相对比的优化过程图,来进一步说明 DSSA 的优化能力。图 4 至图 7 是 DSSA 的优化结果,可以直观看出其优化结果都是很出色的。图 8 至图 10 是优化过程的比较,可以看出 DSSA 比 GA 的优化效率更高。

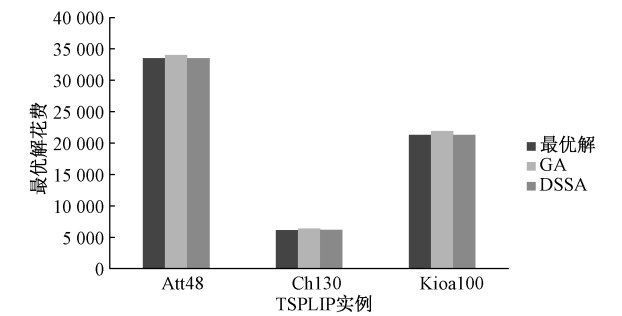


图 3 最优解对比  
Figure 3 Comparison of optimal solutions

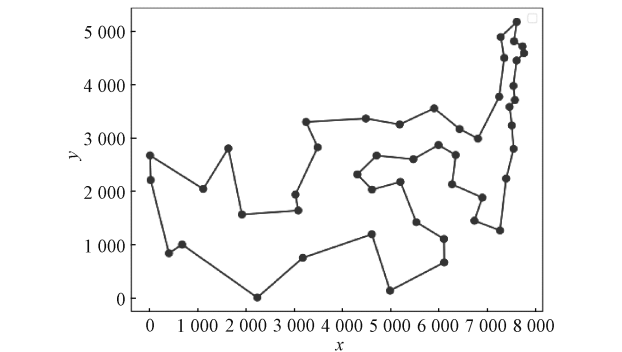


图 4 att48 最优路径  
Figure 4 The optimal path of att48

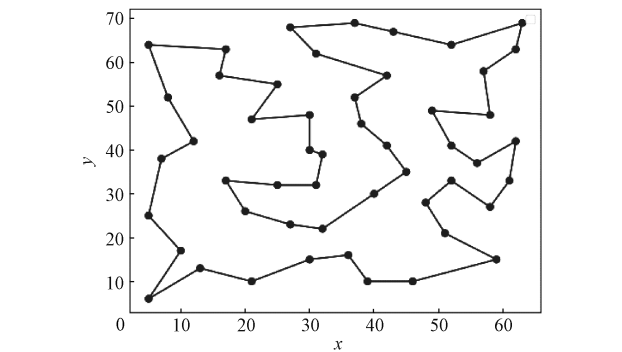


图 5 eil51 最优路线  
Figure 5 The optimal path of eil51

为了增强对比,本文选取了改进的萤火虫算



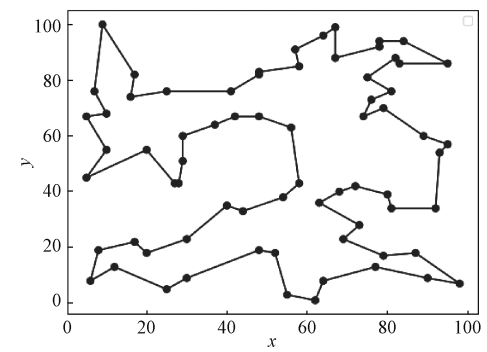


图 6 st70 最优路径图  
Figure 6 The optimal path of st70

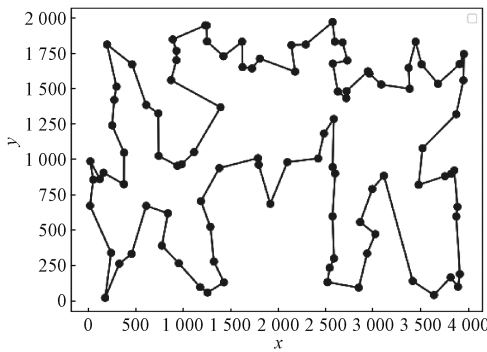


图 7 kioa100 最优路径  
Figure 7 The optimal path of kioa100

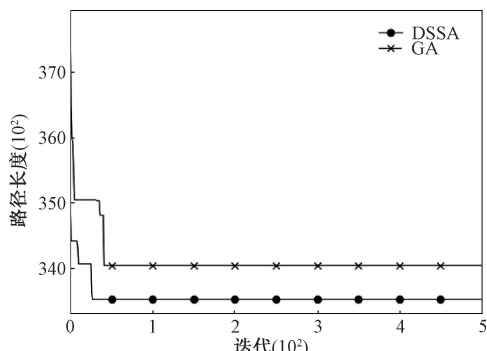


图 8 att48 优化过程  
Figure 8 Optimization process of att48

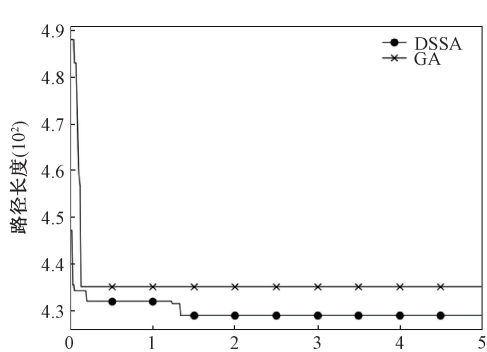


图 9 eill51 优化过程  
Figure 9 Optimization process of eill51

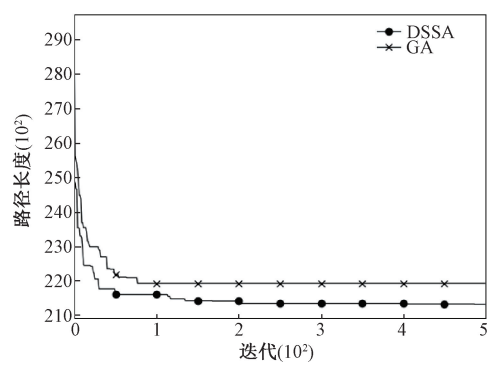


图 10 kioa100 优化过程  
Figure 10 Optimization process of kioa100

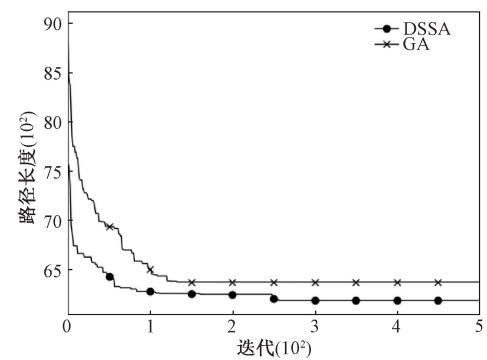


图 11 ch130 优化过程  
Figure 11 Optimization process of ch130

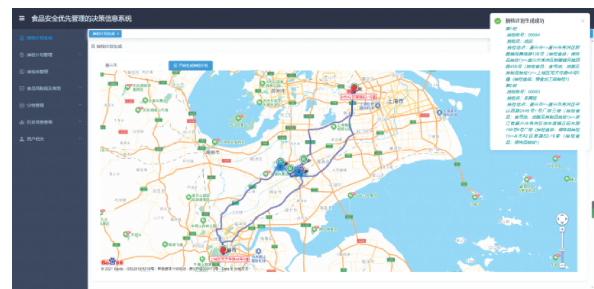


图 12 系统主要运行界面  
Figure 12 The main operating interface of the system

算法 (Improved firefly algorithm, IFA)、蚁群优化算法<sup>[15]</sup> (Ant colony optimization, ACO) 和离散人工萤火虫算法<sup>[16]</sup> (R28) 同 DSSA 进行比较。对比结果如表 3 所示。可以看出, IFA、ACO 和 R28 不仅所得解的精度不够高, 且随着抽检点数目的增加, 偏差率变化的幅度过大。从平均值也可以看出, 这些算法在同一抽检数目的多次试验下, 解的变化范围较大, 稳定性较差。而在实际抽检过程中, 集中派发任务是很常见的情况, 这更要求优化算法具有稳定性以及求解精度。从表中可以看出, DSSA 在不同抽检数目的情况下, 解的平均值接近最优解, 且随着抽检点数目增加, 也具有稳定的求解精度。因此, DSSA 良好的稳定性以及求解精度, 使它更适合

法<sup>[14]</sup> (Improved firefly algorithm, IFA)、蚁器优化

在抽检路径规划中发挥作用。

2.3 实际应用

基于前述的理论与实验,我们开发了实际应用中的食品抽检计划生成及路径优化系统界面如图 12 所示。

表 3 实验结果对比

Table 3 Comparison of experimental results

实例	已知最优解	算法	最优解	平均值	偏差率/%
Burma14	30. 878 5	IFA	30. 878 5	30. 878 5	0. 00
		ACO	44. 468 5	46. 485 4	44. 01
		R28	30. 878 5	30. 878 5	0. 00
		DSSA	30. 878 5	30. 878 5	0. 00
Oliver30	423. 740 6	IFA	424. 697 6	425. 649 6	0. 23
		ACO	468. 700 1	468. 700 1	10. 61
		R28	423. 740 6	438. 860 4	0. 00
		DSSA	423. 740 6	423. 740 6	0. 00
eil51	426. 000 0	IFA	428. 871 8	429. 953 8	0. 67
		ACO	740. 722 6	771. 176 8	73. 88
		R28	429. 484 1	442. 018 7	0. 82
		DSSA	428. 871 8	429. 871 7	0. 67
St70	675. 000 0	IFA	683. 640 4	685. 658 4	1. 28
		ACO	1. 431 9×10 <sup>3</sup>	1. 438 2×10 <sup>3</sup>	112. 13
		R28	679. 454 0	713. 949 6	0. 66
		DSSA	677. 109 6	681. 609 2	0. 31

3 小结

国家大力提倡抽检任务执行中的双随机标准,确保被抽检目标和抽检人员双重随机性。双随机原则使得抽检点对抽检员来说更不可预料、更为发散,这无疑会比以往大幅度加重抽检工作的负荷,让抽检员凭自身的经验规划抽检路径并不现实。本文提出了一种用于求解 TSP 问题的离散型麻雀搜索算法,通过多种优化算子结合来提高算法的收敛速度,通过反捕食行为避免算法过早地陷入局部最优。实验结果表明,该算法是求解 TSP 问题的有效算法。目前该算法已经部署到双随机抽检点生成和抽检路径优化系统中,使得路径优化的时间消耗可以限制在相关人员能够接受的范围之中,促成了系统的实用化。

参考文献

[ 1 ] 徐晓新. 中国食品安全:问题、成因、对策[ J]. 农业经济问题, 2002( 10) :45-48.

[ 2 ] 刘欢. 2018—2019 年全国食品安全监督抽检情况分析[ J]. 食品安全质量检测学报, 2020, 11( 7) :2347-2351.

[ 3 ] 吴琼,宋安东. 食品安全监督抽样工作存在的问题和建议[ J]. 粮食与饲料工业, 2020( 1) :5-7, 15.

[ 4 ] 何平,王煜红,江小明,等. 提高我国食品安全抽检监测有效性的分析和建议[ J]. 中国酿造, 2015, 34( 3) :162-165.

[ 5 ] LEE H P, LIM S P, LEE K H. Solving traveling salesman problems by genetic algorithms[ J]. Progress in Natural Science, 2003( 2) :57-63.

[ 6 ] JIA H W, RUI H X, YUN L M. Research on welding robot path planning using ant colony optimization[ J]. Advanced Materials Research, 2011, 1169:1926-1929.

[ 7 ] 黄训诚. 基于蚁群算法的超大规模集成电路布线研究[ D]. 西安:西安电子科技大学, 2007.

[ 8 ] 王剑文,戴光明,谢柏桥,等. 求解 TSP 问题算法综述[ J]. 计算机工程与科学, 2008( 2) :72-74, 155.

[ 9 ] 马良. 旅行推销员问题的算法综述[ J]. 数学的实践与认识, 2000( 2) :156-165.

[ 10 ] CHEN Y P, KANG L S, YANG H. A gene-pool based genetic algorithm for tsp [ J]. Wuhan University Journal of Natural Sciences, 2003( S1) :217-223.

[ 11 ] 张鑫龙,陈秀万,肖汉,等. 一种求解旅行商问题的新型帝国竞争算法[ J]. 控制与决策, 2016, 31( 4) : 586-592.

[ 12 ] XUE J, SHEN B. A novel swarm intelligence optimization approach: sparrow search algorithm[ J]. Systems ence & Control Engineering An Open Access Journal, 2020, 8( 1) :22-34.

[ 13 ] 公冶小燕,林培光,任威隆. 基于 Grefenstette 编码和 2-opt 优化的遗传算法[ J]. 山东大学学报( 工学版), 2018, 48( 6) : 19-26.

[ 14 ] 王艳,王秋萍,王晓峰. 基于改进萤火虫算法求解旅行商问题[ J]. 计算机系统应用, 2018, 27( 8) : 219-225.

[ 15 ] THIRACHIT S, SUPHAKANT P, CHIDCHANOK L. Combining new fast opposite gradient search with ant colony optimization for solving travelling salesman problem [ J ]. Engineering Applications of Artificial Intelligence, 2014, 35: 324-334.

[ 16 ] 于宏涛,高立群,韩希昌. 求解旅行商问题的离散人工萤火虫算法[ J]. 华南理工大学学报( 自然科学版), 2015, 43( 1) : 126-131, 139.

· 专家寄语 ·

希望《中国食品卫生杂志》继续作为我国营养健康工作重要的交流平台,及时、准确地捕捉营养健康工作和研究中的最新动态,成为政府部门、研究机构、社会团体、营养健康行业的桥梁,促进相互了解、交流合作、深刻思考,成为展示我国营养健康事业发展和对外交流的窗口。

——丁钢强(编委会副主任委员,中国疾病预防控制中心营养与健康所所长)