

DOI:10.19651/j.cnki.emt.2108429

基于 Spark 的并行模拟退火算法求解 TSP*

孙 蓁 刘松佐 武晓晓 巫思敏

(北方民族大学计算机科学与工程学院 银川 750021)

摘要: 模拟退火算法是求解无约束优化问题的有效方法,但求解旅行商问题时存在精度较差、容易陷入局部最优且收敛速度慢等缺点。为了改进上述问题,提出了一种基于 Spark 平台的并行模拟退火算法。修改模拟退火算法的降温函数,构造旅行商问题的解空间,采用大邻域搜索技术和 2-opt 算子增强局部搜索能力,引入 OX 交叉思想增强全局搜索能力,提出交叉协同试验并行策略与 Spark 平台并行实现。选取若干 TSPLIB 数据集进行仿真实验,对求解质量和运行时间两个方面进行测试,与其他 Spark 框架的并行算法进行对比实验。仿真结果表明,该算法求解精度有较大的提高,求解速度上对比其他算法提升 3~10 倍,能够有效求解旅行商问题。

关键词: 并行模拟退火算法;大规模邻域算法;降温策略;旅行商问题;Spark

中图分类号: TP301.6 **文献标识码:** A **国家标准学科分类代码:** 520.3020

Solving TSP based on Spark-based parallel simulated annealing algorithm

Sun Jian Liu Songzuo Wu Xiaoxiao Wu Simin

(College of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China)

Abstract: Simulated annealing algorithm is an effective method for solving unconstrained optimization problems, but it has the disadvantages of poor accuracy, easy to fall into local optimum and slow convergence when solving travel quotient problems. In order to improve the above problems, a parallel simulated annealing algorithm based on Spark platform is proposed. The cooling function of the simulated annealing algorithm is modified to construct the solution space of the travel quotient problem, the local search capability is enhanced by using the large neighborhood search technique and the 2-opt operator, the global search capability is enhanced by introducing the OX crossover idea, and the crossover cooperative experiment parallel strategy is proposed to be implemented in parallel with the Spark platform. A number of TSPLIB datasets are selected for simulation experiments to test both solution quality and running time, and to compare the experiments with other parallel algorithms of Spark framework. The simulation results show that the solution accuracy of this method is greatly improved, and the solution speed is 3~10 times higher than other algorithms, which can effectively solve the traveling salesman problem.

Keywords: parallel simulated annealing algorithm; large neighborhood search algorithm; cooling strategy; traveling salesman problem; Spark

0 引 言

旅行商问题^[1] (traveling salesman problem, TSP) 是较早提出的组合优化问题,本质是有 n 个城市,如何从起始点出发遍历且每个城市遍历一次,回到起始点的路径最小。在车辆路径规划^[2]、物流配送问题^[3] 等领域有重要的现实意义。作为一个 NP 难问题,问题提出初期众多学者使用精确算法求解,后来启发式算法成为求解该问题的主流算

法,如遗传算法、蚁群算法和果蝇算法等。传统的启发式算法求解 TSP 存在陷入局部最优、收敛速度慢和解的质量差等缺点,为克服上述缺点尝试改进传统算法或混合多种算法进行优化求解。

针对模拟退火算法(simulated annealing, SA) 解决 TSP 问题,国内外学者对算法进行优化改进。袁汪凤等^[4] 提出自适应温度并加入回火策略,引入改进最大-最小蚁群算法和 3-OPT,但是参数调节困难。陈科胜等^[5] 提出自适

收稿日期:2021-11-24

* 基金项目:国家自然科学基金(62062002)、北方民族大学中央高校基本科研业务费专项(FWNX09)、北方民族大学校级一般项目(2021XYZJK01)资助

应升温策略和改进 Metropolis 函数,该算法在求解大规模 TSP 问题收敛速度较慢。

为了解决传统 SA 算法的求解质量低和收敛速度慢等问题,采用并行编程用于加速优化。分布式计算平台可提供优秀的并行计算能力,如 Hadoop^[6] 和 Spark^[7] 等分布式平台。Hadoop 依赖于 MapReduce 框架,但计算时需要 CPU 和硬盘频繁进行数据交互,影响计算效率不适合迭代计算。Spark 是可应用于大规模数据处理的开源平台,具有运行速度快,程序编写容易等优点,使 Spark 更适合迭代计算。类似 SA 算法需要大量迭代的算法在 Spark 平台运行优于 Hadoop 等分布式平台,故而使用 Spark 计算框架更适合迭代计算。本文提出基于 Spark 的并行模拟退火算法 (Spark-based parallel simulated annealing algorithm with large neighborhood search, SPSALNS),改进降温函数,加入大规模邻域搜索和 2-opt 增加局部搜索能力,使用并行 OX 变异操作增加全局搜索能力。使用 TSPLIB 中标准数据集作为实验数据,对比 SPSALNS 和其他算法的求解质量和运行时间。

1 基本知识

1.1 模拟退火算法并行策略

模拟退火算法^[8]受固体退火原理的启发,金属固体退温为加温、等温、冷却 3 个过程,算法中一系列操作为“扰动产生新解 计算目标函数差值 判断是否满足接受准则 接受(或舍弃)新解”。根据算法的操作流程并行模拟退火算法可分为如下 3 种方式。

1) 操作并行

将每个操作中需要执行的多个步骤分配给每个 CPU 去完成,由于各个任务是串行的,因此操作并行只能同步进行,并且通信时间耗时长,解的质量和串行无异。

2) 试验并行

(1) 独立试验并行

将初始解分配给每个 CPU 后,由 CPU 独立完成算法并得到各自结果,从中选取最优的为最终结果。此算法依赖用时最长的 CPU,时间复杂度无法估计且每个 CPU 需要单独存储空间。

(2) 协同试验并行

将初始解分配给每个 CPU 后,由每个 CPU 同时产生的新解进行判断,根据接受准则从中选取一个在此基础上所有 CPU 进行下一轮试验。

3) 区域分裂

将组合优化问题的解建立在离散空间上,将此空间分解 N 个子空间,每个 CPU 在一个空间上操作。该方式加速和效率都很高,但要求问题本身可进行区域分裂,可能出现局部混乱的现象。

分析旅行商问题的特性,操作并行有一定加速效果,但对解没有优化,区域分裂会出现路径交叉问题,故本文优化试验并行策略达到优化目的。

1.2 Spark 开发平台

Spark 是一个基于内存的大数据并行计算框架。该平台基于内存计算,提高了大数据环境下数据处理的实时性,保证了高容错性和高扩展性。它允许用户将 Spark 部署在许多普通的硬件上,形成集群;采用将弹性分布式数据集 (resilient distributed dataset, RDD) 的依赖关系构建有向无环图 (directed acyclic graph, DAG), 将 DAG 分解成多个阶段进行调度处理。

Spark 集群分为主节点和工作节点,主节点协调和调度集群中的所有工作节点;在每个工作节点中,都有一些执行器,它们用于管理工作线程。当一个算法在 Spark 上运行时,会在所有工作节点上形成一个由这些工作线程组成的线程池,并对算法的输入进行分区并分配到线程中。最后,Spark 中的主节点在线程池中分配线程数来并行执行算法,每个线程都带着部分输入执行算法。类似遗传算法、蚁群算法和模拟退火算法等存在许多次迭代的算法,使用 Spark 将直接存储在内存中,避免中间结果的存储,大大降低磁盘 I/O 处理,提升迭代效率。

2 解决旅行商问题的并行模拟退火算法

2.1 改进降温函数

模拟退火算法的主要组成要素分别是状态产生函数、状态接受函数、降温函数、热平衡稳定准则、退火结束准则、初始温度的选取。其中降温函数影响算法收搜空间,温度高时,算法进行全局搜索,温度低时,算法进行局域搜索。温度降温较快算法全局搜索时间过短而无法得到全局最优解,降温过慢算法局域搜索能力较弱影响解的质量,两者都会使算法解的质量降低。故本文针对旅行商问题的特性,对算法的降温函数进行改进,改进降温函数如式(1)和(2)所示。

$$T_0 = T_0 \times \alpha \quad (1)$$

$$T = T_0 \times (1 + \cos(t \times \pi / gapIter)) \quad (2)$$

式中: T_0 表示初始温度, T 表示变化后的温度, α 为降温系数, t 迭代次数, $gapIter$ 表示浮动的间隔数。

选择标准降温、快速降温^[9]和本文降温函数进行效果对比,初始条件、降温函数和参数如表 1 所示,降温曲线如图 1 所示。

表 1 降温函数初始化参数

降温标准	降温函数	参数
标准降温	$T_0 = T_0 \times \alpha$	$T_0 = 200, \alpha = 0.99$
快速降温	$T = T_0 \times \exp(-\alpha \times k^{\frac{1}{2}})$	$T_0 = 200, \alpha = 0.99$ k 为迭代次数
本文降温	$T = T_0 \times (1 + \cos(t \times \pi / gapIter))$	$T_0 = 100, \alpha = 0.99$ $gapIter = 100$

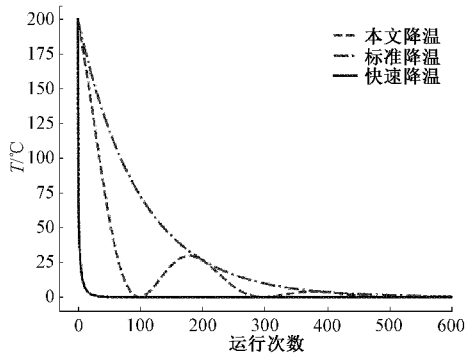


图 1 降温曲线对比图

由图 1 可知,标准降温达到低温状态迭代次数过多,而快速降温速度过快,两者解的质量较差。本文降温结合二者的优点,同时增加了回火过程,使算法全局搜索和局部搜索能力加强,解的精度得到提升。

2.2 大规模邻域搜索

本文产生新解的过程引入大规模邻域搜索算法^[10] (large neighborhood search, LNS)。算法思想为:从当前解中移除多个节点;根据约束条件将移除节点重新插入得到新的解,以此探索当前解的邻域。如图 2 所示,初始解空间为 2-1-8-3-6-7-5-4,随机选取 $N/3$ (N 为城市总数) 破坏移除初始解,图 2 中破坏移除城市为 1、7 和 5,将剩余城市依照初始解依次排序生成破坏的解空间 2-8-3-6-4。对破坏的解空间就行修复,分别将破坏移除的城市插入,图 2 将城市 1 插入后计算其适应度值 $f(\text{path})$,分别计算插入不同位置的适应度值选取值最小的解空间,直到所有被破坏移除城市全部修复完成生成最优解。

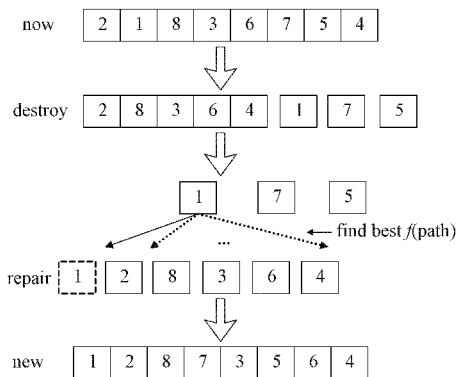


图 2 大规模邻域搜索算法的流程

2.3 局部搜索算子 2-opt

2-opt 是一种随机性算法,基本思想就是随机选择两个元素进行优化,在求解 TSP 问题中得到了广泛的应用。图 3 为 2-opt 示例图,路径为 C_1, C_2, \dots, C_n , 其中 $d(C_i, C_j)$ 表示城市 C_i 和城市 C_j 之间的距离。如果条件 $d(C_i, C_j) + d(C_{i+1}, C_{j+1}) < d(C_i, C_{i+1}) + d(C_j, C_{j+1})$ 满足,则通过去除弧 $(C_i, C_{i+1}), (C_j, C_{j+1})$ 并连接 $(C_i, C_j), (C_{i+1}, C_{j+1})$, 产生更短的路径。

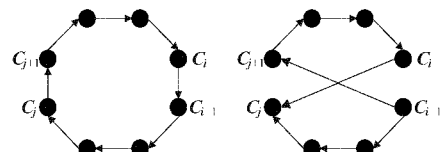


图 3 2-opt 过程图

2.4 OX 交叉策略

为了保证并行模拟退火算法的全局性和优化解的质量,本文加入 OX 交叉策略。使用当前最优路径选取其中某一段路径和当前解的路径交叉重组,生成新路径。如图 4 所示,OX 交叉策略随机选中最优路径起始点 1 和终止点 8,生成一个新路径,并保证新路径中被选中路径的位置与最优路径相同,找出在最优路径选中的城市在当前路径中的位置,再将其余城市按顺序放入生成的新路径中。

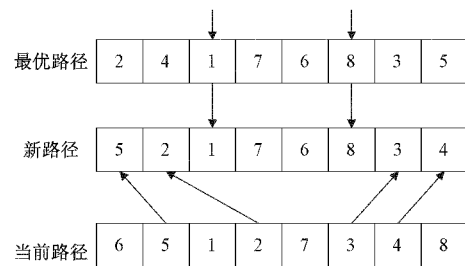


图 4 OX 交叉过程图

2.5 并行模拟退火算法算法流程

现提出一种交叉协同试验并行策略,改善传统并行 SA 寻优精度差和早熟现象,增大 SA 的搜索空间和降低运行时间。该策略既保留独立试验并行的独立求解互不干预的优势,又将协同试验并行选用共享通道通信方式进行信息交换优化解空间结合。将若干种群存入 RDD 中,并行运行扰动产生新解,选取当前最优路径和每个种群进行 OX 交叉,使用 2-opt 进行优化,最后使用 LNS 进行扰动生成新的解。

SPSALNS 步骤如下:

- 1) 根据 TSP 实例贪婪策略初始化种群,存入 RDD 中。
- 2) 计算适应度值,记录最优路径 P_{best} 和最优距离 L_{best} 。
- 3) 使用当前最优路径 P_{new} 和每个种群 OX 交叉、2-opt 优化,使用 LNS 产生新的路径,计算适应度值,选取种群中当前最优路径 P_{new} 。
- 4) 比较两次路径的适应度值之差, $f(P_{new}) \leq f(P_{best})$, 则 $P_{best} = P_{new}$; 否则,按照 Metropolis 准则进行更新;根据式(1)和(2)更新温度。
- 5) 判断是否达到结束条件,如果没有达到执行步骤 3)~5), 否则输出结果。

综上所述,算法流程如图 5 所示。

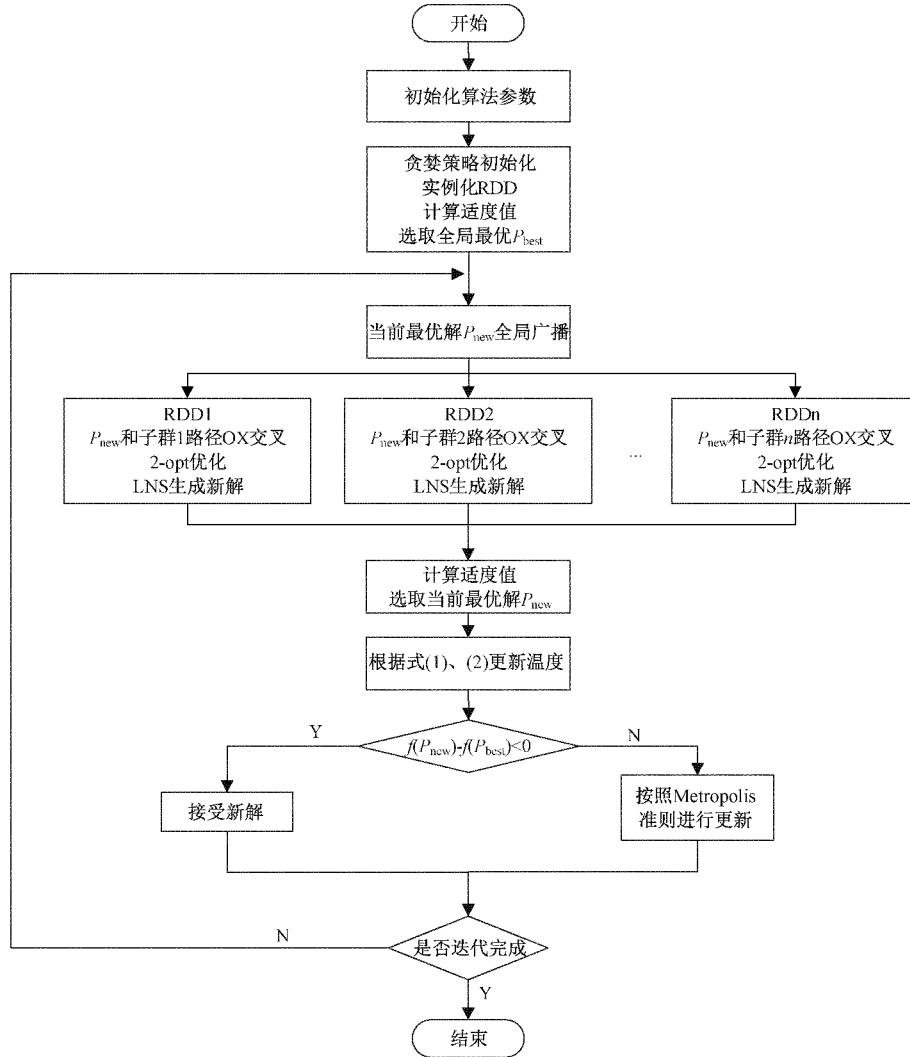


图 5 算法流程

3 数值实验

为了检测大规模邻域搜索的模拟退火算法 (Spark-based parallel simulated annealing algorithm with large neighborhood search, SPSALNS) 的性能, 使用 TSPLIB 中标准数据集作为实验数据进行测试。表中已知最优解、最优解、平均解和平均时间分别用 Hbest, Best、Avg 和 Time(s) 表示。实验环境中节点的属性信息如表 2 所示。

表 2 节点信息表

项目	值
节点个数	6
Spark 版本	2.4.0
Spark 模式	Standalone
JDK 版本	1.8.0
节点 CPU	Intel Core i7-9700 CPU@3.00 GHz * 8
节点内存	8 GB

3.1 求解质量对比

在本节选取 TSPLIB 数据集 Oliver30、Att48、Eil51、Eil76 进行实验对比, 将 SPSALNS 和 SQACS^[11]、SSA^[12] 进行对比。3 种算法分别运行 10 次, 加粗结果为求得已知最优解, 仿真结果如表 3 所示, 时间对比如图 6 所示。

从表 3 和图 6 实验结果中, SPSALNS 在 6 个实例中最优解和平均解都达到已知最优解, SQACS 仅在 Att48 和 Eil51 中最优解和平均解达到已知最优解, 而 SSA 都没有得到已知最优解。在中规模数据集 KroA100 和 Ch130 中, SQACS 和 SSA 未得到已知最优解, 平均解较差。运行时间上本文在 Oliver30 中提升了 10 倍以上, Att48 中提升了 8 倍以上, Eil51 提升了 6 倍以上, Eil76 提升了 4 倍以上, 在中规模数据集 KroA100 和 Ch130 中, 均有 3 倍以上的提升。结果可知, SPSALNS 在求解速度和解的质量上均明显优于其他两种算法。

3.2 Spark 并行算法实验对比

在本节对比使用 Spark 并行的 3 种算法: SFCPGA^[13],

Spark-MMAS^[14] 和 SIMP-ACS^[15], 分别进行 10 次独立实验。

表 3 3 种算法实验结果对比

数据集	算法	Best	Avg	Time(s)
Oliver30	SPSALNS	420	420	2.53
	SQACS	420	421	26.17
	SSA	423	427	29.03
Att48	SPSALNS	33 522	33 522	6.87
	SQACS	33 522	33 583	49.02
	SSA	33 842	34 103	51.50
Eil51	SPSALNS	426	426	9.05
	SQACS	426	427	55.27
	SSA	438	451	60.37
Eil76	SPSALNS	538	538	21.27
	SQACS	538	543	85.16
	SSA	561	582	89.30
KroA100	SPSALNS	21 282	21 282	32.17
	SQACS	21 346	21 540	102.38
	SSA	22 128	22 695	117.75
Ch130	SPSALNS	6 110	6 110	53.60
	SQACS	6 299	6 370	153.32
	SSA	6 379	6 568	167.87

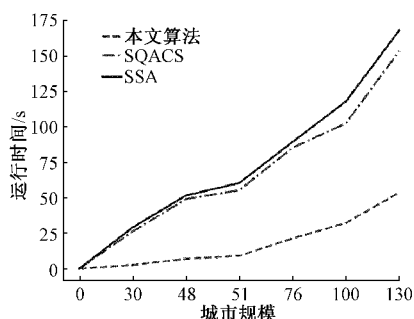


图 6 平均运行时间变化曲线

首先测试中规模的城市群体,使用 Lin105 进行实验,得到的数据如表 4 所示。SPSALNS 最优解和平均解都达到已知最优解,平均运行时间仅 39 s,SIMP-ACS 平均运行时间消耗 134 s,SFCPGA 和 Spark-MMAS 消耗时间过长,平均解均未达到已知最优解。

表 4 Lin105 对比结果

算法	Hbest	Best	Avg	Time(s)
SPSALNS	14 379	14 379	14 379	39
SFCPGA	14 379	14 984	15 549	1 946
Spark-MMAS	14 379	14 379	14 400	2 358
SIMP-ACS	14 379	14 379	14 379	134

为了验证中规模的城市群体有效性,接下来对 Ch150 进行实验,得到的数据如表 5 所示。SPSALNS 最优解达到已知最优解,平均耗时 328 s,平均解优于其他算法,Spark-MMAS 最优解同样达到已知最优解,但平均耗时

3 012 s,消耗时间过长,SFCPGA 最优解和平均解质量差,耗时过长,SIMP-ACS 耗时虽然少于 SPSALNS,但最优解未达到已知最优解,平均解效果质量不理想。

表 5 Ch150 对比结果

算法	Hbest	Best	Avg	Time(s)
SPSALNS	6 528	6 528	6 549	328
SFCPGA	6 528	6 745	6 939	2 843
Spark-MMAS	6 528	6 528	6 579	3 012
SIMP-ACS	6 528	6 543	6 571	134

然后扩大城市规模,选择 KroA200 有 200 个城市进行实验,200 个城市在 TSP 问题中是大规模问题。得到的数据如表 6 所示。SPSALNS 最优解达到已知最优解,平均解优于其他 3 种算法,平均运行时间高于 SIMP-ACS,但远低于 SFCPGA 和 Spark-MMAS。

表 6 KroA200 对比结果

算法	Hbest	Best	Avg	Time(s)
SPSALNS	29 368	29 368	29 391	853
SFCPGA	29 368	30 937	33 639	3 654
Spark-MMAS	29 368	29 426	30 012	3 470
SIMP-ACS	29 368	29 383	29 678	203

选取 Eil51、Lin105、Ch150 和 KroA200 数据集最优解路径图,图中路径清晰无交叉,求得结果均是已知最优解,如图 7~10 所示。

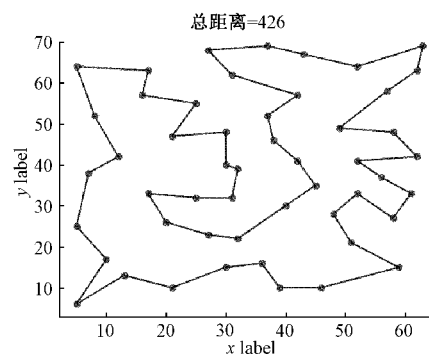


图 7 Eil51 最优路径图

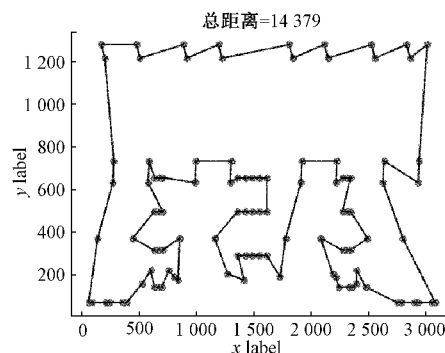


图 8 Lin105 最优路径图

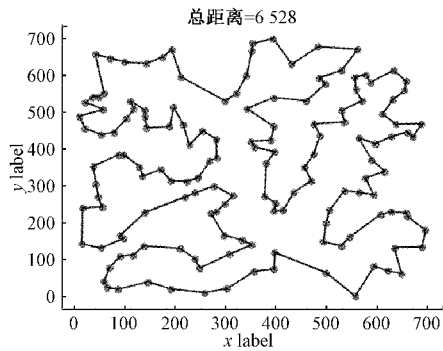


图 9 Ch150 最优路径图

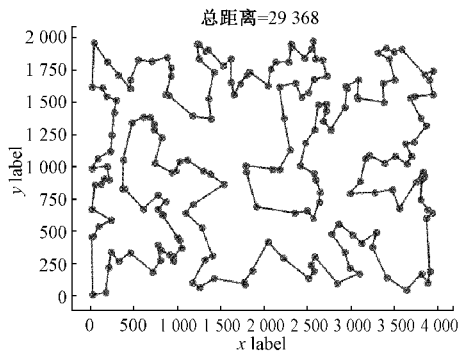


图 10 KroA200 最优路径图

4 结 论

本文采用 Spark 计算框架并行计算,来加快运算速度和优化解的质量。在模拟退火算法的基础上,修改降温函数,提升算法的搜索能力,改进后的降温函数可以很好地跳出局部最优解。在搜索路径中,以 OX 交叉操作提高算法的运算能力和全局搜索能力,结合 2-opt 和 LNS 的局部搜索策略,加强算法局部搜索性能。实验结果表明,该算法同新型算法和其他 Spark 计算框架并行算法进行对比,SPSALNS 优于其他算法,有效地提高求解质量且收敛速度较快。

参考文献

- [1] 吴亚帅,刘新妹,殷俊龄,等.一种求解 TSP 问题的 DSO 路径优化算法[J].国外电子测量技术,2021,40(3):55-61.
- [2] 赵晓敏.基于蚁群算法的 AGV 车辆 End_To_End 最优路径规划研究[J].国外电子测量技术,2020,39(5):90-93.
- [3] 林美,何竹峰.基于云计算的物流车辆路径规划方法[J].电子测量技术,2019,42(13):153-156.
- [4] 袁汪凰,游晓明,刘升,等.求解 TSP 问题的自适应模拟退火蚁群算法[J].计算机应用与软件,2018,35(2):

261-266.

- [5] 陈科胜,鲜思东,郭鹏.求解旅行商问题的自适应升温模拟退火算法[J].控制理论与应用,2021,38(2):245-254.
- [6] 葛苏慧,万泉,白成杰.基于 Hadoop 的高校学生行为预警决策系统研究[J].计算机应用与软件,2021,38(1):6-12.
- [7] 李铭,丁卫平,鞠恒荣,等.基于 Spark 的证据邻域粗糙并行分类高效算法[J].南京理工大学学报,2021,45(4):409-419.
- [8] 钟星宇.基于并行方法的微地震速度模型校正研究[D].荆州:长江大学,2019.
- [9] 王纪程,陈祖斌,江海宇,等.基于极快速模拟退火与网格逐次剖分的微地震定位算法[J].科学技术与工程,2017,17(33):248-252.
- [10] 南丽君,陈彦如,张宗成.改进的自适应大规模邻域搜索算法求解动态需求的混合车辆路径问题[J].计算机应用研究,2021,38(10):2926-2934.
- [11] 徐小平,唐阳丽,王峰.求解旅行商问题的人工协同搜索算法[J/OL].计算机应用,2021:1-7[2021-11-13].
<http://kns.cnki.net/kcms/detail/51.1307.TP.20210804.1442.008.html>.
- [12] XUE J, SHEN B. A novel swarm intelligence optimization approach: sparrow search algorithm[J]. Systems Science and Control Engineering, 2020, 8(1): 22-34.
- [13] 许智宏,赵嘉伟,董永峰,等.基于 Spark 的并行遗传算法在旅行商问题中的应用[J].计算机应用研究,2017,34(7):2080-2083.
- [14] WU Q, ZHU Q. An improved ant colony algorithm based on Spark in TSP[C]. Singapore: International Conference on Future Computer and Communication, WCSE, 2018.
- [15] 吴启帆.基于 Spark 的改进蚁群算法在 TSP 问题中的应用[D].重庆:重庆大学,2018.

作者简介

孙鉴,博士,讲师,主要研究方向为大数据存储与管理等。
E-mail:2014132@nun.edu.cn

刘松佐,硕士研究生,主要研究方向为大数据分析 with 知识工程等。
E-mail:a_bacd@qq.com

武晓晓,硕士研究生,主要研究方向为闪存索引等。
E-mail:343636113@qq.com

巫思敏,硕士研究生,主要研究方向为数据挖掘与知识发现等。
E-mail:624698624@qq.com