

学校编码: 10384

分类号\_\_\_\_\_密级\_\_\_\_\_

学 号: 23020141153211

UDC\_\_\_\_\_

厦 门 大 学

硕 士 学 位 论 文

**基于 Spark 带时间窗约束的车辆路径问题  
蚁群优化算法研究**

**Research On Improved Ant Colony Optimization Algorithm Based On Spark  
For Vehicle Routing Problem With Time Windows**

张 羽 中

指导教师姓名: 张 德 富 教授

专 业 名 称: 计算机技术

论文提交日期: 2017 年 4 月

论文答辩时间: 2017 年 5 月

学位授予日期: 2017 年 月

答辩委员会主席: \_\_\_\_\_

评 阅 人: \_\_\_\_\_

2017 年 月

# 厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为( )课题(组)的研究成果,获得( )课题(组)经费或实验室的资助,在( )实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

# 厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1.经厦门大学保密委员会审查核定的保密学位论文，于  
年 月 日解密，解密后适用上述授权。

2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

厦门大学博硕士学位论文摘要库

## 摘要

作为一个在诸多研究方向都炙手可热的组合优化和整数规划问题，车辆路径问题（Vehicle Routing Problem, VRP）近几十年来在图论、运筹学、应用数学、计算机应用等领域都有着深入而广泛的应用和实践。近十年来，随着我国电子商务的爆发式增长，VRP问题在物流配送、快递运输方面的应用和其带来的经济价值使得它在科学研究领域的重要性不断增强。带时间窗约束的车辆路径问题（Vehicle Routing Problem With Time Windows, VRPTW）是VRP问题的一个分支，具有更高实际应用价值，但相关研究比较少。

随着大数据和云计算技术的发展，大数据并行计算框架Spark和Hadoop在工业界已经取得了不错的成绩，但是在科学研究领域的尝试还较少。

本文的主要工作包括：第一，分别对VRPTW问题、蚁群算法、蚁群算法求解VRPTW问题三个领域进行综述。第二，提出蚁群算法的优化改进策略，改进后的算法融合了“蚁周模型”和最大最小蚂蚁系统（MMAS）的思想，信息素的更新采用全局信息，并将信息素浓度控制在一定范围内使得算法不会过早陷入局部最优；在蚂蚁的状态转移规则中引入时间窗跨度、等待时间和节约值作为影响因子；提出两种信息素的局部蒸发策略；添加信息素挥发率的动态调整策略；最后将改进后的蚁群算法与局部搜索优化策略结合，解的质量有明显的提高。对经典VRPTW数据的实验结果表明，改进后的算法是有效的。第三，将算法的单机实现版本、Spark实现版本和Hadoop实现版本进行实验对比和分析，并验证Spark版本的算法收敛性和扩展性。

本文提出的改进蚁群算法求解带有时间窗约束的VRP问题具有一定的可行性，分布式实现大大减少了算法的求解时间，具有一定的实用价值。

**关键词：**车辆路径；蚁群算法；Spark

## Abstract

As a sought-after combinatorial optimization and integer programming problem in many research directions, vehicle routing problem has in-depth and extensive application and practice in graph theory, operations research, applied mathematics, computer application in recent decades. In nearly a decade, with the explosive growth of electronic commerce in our country, the applications of VRP problem in logistics, express transport and its economic value make it importance in the field of scientific research. Vehicle routing problem with time windows constraint is a branch of VRP problem, has higher application value, but the related research is less.

With the development of big data and cloud computing, Spark and Hadoop as famous big data parallel computing framework have achieved good results in industry, but there is few attempt in the field of scientific research.

In this paper, the main work includes: first of all, three areas of VRPTW problem, ant colony algorithm, ant colony algorithm to solve VRPTW problem are reviewed respectively. Secondly, an improved optimization strategy for the ant colony algorithm is put forward, the improved algorithm combines the ideas of the "Ant-Cycle Model" and the Max-Min Ant System, pheromone is updated by using global information, the pheromone concentration is controlled within a certain range that makes the algorithm avoid getting stuck prematurely into local optimum; time window span, waiting time and saved values are introduced into the transition rules of ant as impact factors; two pheromone evaporation strategy is proposed; a dynamic adjustment strategy of pheromone's volatilization rate is added; finally local search optimization strategy is added to improve the quality of the solution significantly. The experiments of classical VRP data show that the improved algorithm is effective. Thirdly, the standalone version, Spark version and Hadoop version are compared and analyzed in the experiment and the convergence and extensibility of the Spark version is verified.

In this paper, the improved ant colony algorithm to solve the vehicle routing problem with time window constraints has certain feasibility and the distributed implementation greatly reduces the time of the algorithm, and has certain practical value.

**Key words:** Vehicle Routing; Ant Colony Optimization Algorithm; Spark.

厦门大学博硕士学位论文摘要库

# 目 录

第 1 章 绪论 .....	1
1.1 课题背景及研究意义 .....	1
1.2 国内外研究现状 .....	3
1.2.1 VRP 问题研究综述.....	3
1.2.2 蚁群算法综述.....	5
1.2.3 蚁群算法求解 VRP 综述.....	7
1.2.4 蚁群算法并行化综述.....	8
1.3 本文研究工作 .....	9
1.4 本文章节安排 .....	10
第 2 章 基本理论 .....	11
2.1 VRPTW 问题的模型.....	11
2.2 蚁群算法基本原理 .....	13
2.3 基础模型及流程 .....	15
2.4 Spark 框架简介 .....	18
2.4.1 Spark 基本介绍 .....	18
2.4.2 Spark 历史和发展 .....	18
2.4.3 BDAS .....	19
2.4.4 Spark 的优势 .....	21
2.5 Spark 与 Hadoop 的区别 .....	22
第 3 章 改进蚁群算法求解 VRPTW .....	23
3.1 可行解的构建 .....	23
3.2 信息素的更新策略 .....	24
3.3 路径的优化策略 .....	25
3.4 算法的实现步骤 .....	32
第 4 章 基于 spark 和 Hadoop 的并行蚁群优化算法.....	35
4.1 Hadoop 关键技术介绍.....	35



<b>4.2 Hadoop 分布式实现</b> .....	<b>36</b>
4.2.1 算法流程.....	36
4.2.2 主要函数介绍.....	38
<b>4.3 Spark 关键技术介绍</b> .....	<b>39</b>
<b>4.4 Spark 分布式实现</b> .....	<b>40</b>
<b>4.5 实验及分析</b> .....	<b>43</b>
4.5.1 实验环境及参数设置.....	43
4.5.2 实验结果比较.....	46
4.5.3 收敛性分析.....	50
4.5.4 扩展性分析.....	52
<b>第 5 章 结论和进一步工作</b> .....	<b>55</b>
5.1 本文的创新点.....	56
5.2 存在的问题和进一步工作.....	56
<b>参 考 文 献</b> .....	<b>59</b>

# Contents

<b>Chapter 1 Introduction.....</b>	<b>1</b>
<b>1.1 Backgroud of this Subject .....</b>	<b>1</b>
<b>1.2 Related Work.....</b>	<b>3</b>
1.2.1 Overview of VRP .....	3
1.2.2 Overview of Ant Colony Optimization Algorithm .....	5
1.2.3 Overview of ACO For Solving VRP.....	7
1.2.4 Overview Parallel ACO .....	8
<b>1.3 Research Work .....</b>	<b>9</b>
<b>1.4 Chapter Arrangement.....</b>	<b>10</b>
<b>Chapter 2 Basic Theory.....</b>	<b>11</b>
<b>2.1 Model of the VRPTW .....</b>	<b>11</b>
<b>2.2 Basic Theory of ACO .....</b>	<b>13</b>
<b>2.3 Basic Model and Process .....</b>	<b>15</b>
<b>2.4 Introduction of Spark Framework.....</b>	<b>18</b>
2.4.1 Basic Introduction .....	18
2.4.2 History and Development .....	18
2.4.3 BDAS .....	19
2.4.4 Advantages .....	21
<b>2.5 Difference Between Spark and Hadoop.....</b>	<b>22</b>
<b>Chapter 3 Improved ACO for the VRPTW .....</b>	<b>23</b>
<b>3.1 Construction of Feasible Solution.....</b>	<b>23</b>
<b>3.2 Update Strategy of Pheromone .....</b>	<b>24</b>
<b>3.3 Optimization Strategy .....</b>	<b>25</b>
<b>3.4 Implementation Steps .....</b>	<b>32</b>
<b>Chapter 4 Parallel ACO Based on Spark and Hadoop .....</b>	<b>35</b>

<b>4.1 Key Technology of Hadoop .....</b>	<b>35</b>
<b>4.2 Distributed Implementation of Hadoop.....</b>	<b>36</b>
4.2.1 Process .....	36
4.2.2 Main Function.....	38
<b>4.3 Key Technology of Spark.....</b>	<b>39</b>
<b>4.4 Distributed Implementation of Spark.....</b>	<b>40</b>
<b>4.5 Experiment and Analysis.....</b>	<b>43</b>
4.5.1 Environment and Parameter Setting .....	43
4.5.2 Comparison of Results .....	46
4.5.3 Convergence Analysis.....	50
4.5.4 Scalability Analysis.....	52
<b>Chapter 5 Conclusions and Further Work.....</b>	<b>55</b>
<b>5.1 Innovation.....</b>	<b>56</b>
<b>5.2 Problems and Further Work.....</b>	<b>56</b>
<b>References.....</b>	<b>59</b>

厦门大学博硕士学位论文摘要库

## 第1章 绪论

### 1.1 课题背景及研究意义

作为一个在诸多研究方向都炙手可热的组合优化和整数规划问题,车辆路径问题 (Vehicle Routing Problem, VRP) 近几十年来在图论、运筹学、数学、计算机等领域都有着普遍应用。近十年来,随着我国电子商务的爆发式增长,VRP 问题在物流配送、快递运输方面的应用和带来的经济价值使得其在科学研究领域的重要性不断增强。VRP 提出了这样一个问题:假设有一个配送中心及若干停留在配送中心内的配送车辆,配送车辆为了将配送中心内的货物传递给指定客户点集而行驶的最佳配送路径组合是什么。1959年,VRP 问题首次出现在 George Dantzig 和 John Ramserd 的文章中,并应用于汽油运输问题,VRP 的目的是最小化总的配送成本。Clarke 和 Wright 在 1964 年改进了 George Dantzig 和 John Ramserd 的方法,使用一种称为存储算法的有效贪婪方法。

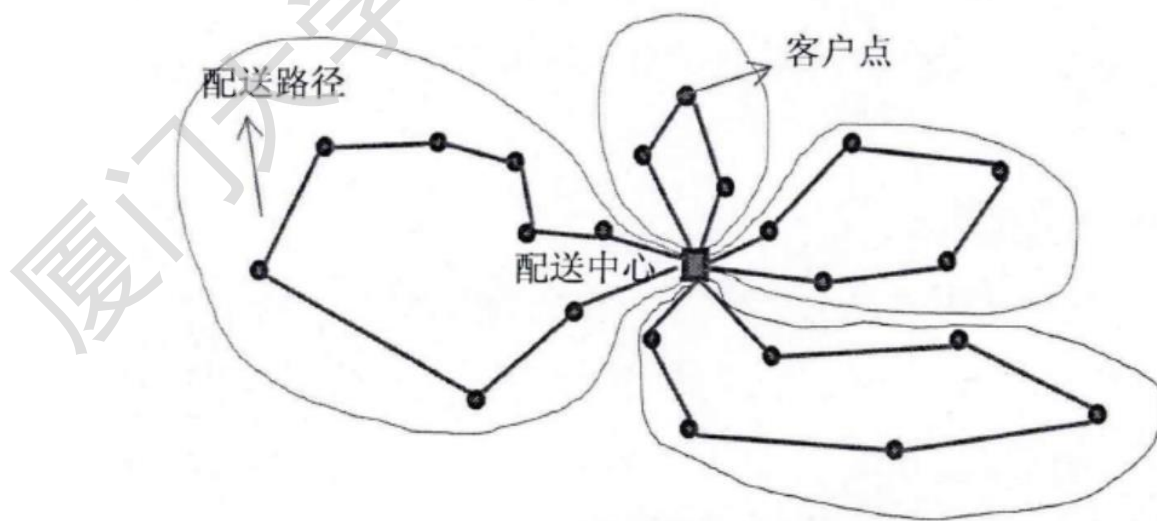


图 1.1: VRP 问题示意图<sup>[39]</sup>

VRP 在工业上的应用也有其重要的经济价值。事实上,使用计算机优化的配

送过程可以给公司带来 5% 的开支节约，因为运输通常是产品成本的重要组成部分，其占到的比例高达 10%。例如，交通部门所带来的贡献占到欧盟 GDP 的 10%。因此，将 VRP 问题应用于工业生产中所带来任何运输成本上的节约，即使低于 5%，都具有重要的经济价值。

VRP 问题存在多个变种，在传统 VRP 问题的基础上，根据限制条件的不同，可以得到不同的变种问题，例如带有重量约束（CVRP）、带有时间窗限制（VRPTW）、多车种（FSVRP）、车辆多次使用（VRPM）等。

寻找问题的最优解是组合优化中的 NP-hard 问题，当问题的规模逐渐增大时，寻找到精确最优解的代价会呈现指数爆炸的增长态势，带来的硬件以及时间上的成本是无法承受的，所以这种方式限制了问题的规模。如何在可接受时间范围内求得问题在可接受误差范围内的解，成为了研究人员研究的重点。因此，考虑现实世界 VRP 问题的规模大小和使用频率，商业应用上倾向于使用启发式算法。

将启发式算法应用在 VRP 中，出现了许多优秀应用，如 Thangiah 等的遗传算法、Ululgu 的模拟退火算法、Taillard 的禁忌搜索算法，但还存在各自的缺点，蚁群算法的出现，为解决 VRP 问题提供了新的思路。

1992 年，Dorig 等<sup>[1]</sup>通过对自然界中蚂蚁群体觅食行为的观察和仿真，提出了蚁群算法，它在旅行商问题（TSP）中最早被使用。在国际上，它是计算智能领域的焦点，具有良好的并行性和鲁棒性，可以很好地应用于大数据并行计算。蚁群算法是通过构造一个人工蚁群来实现的，蚁群中的每一只蚂蚁通过启发信息寻找下一步到达的目的地，直至构造完成整个可行解或者部分可行解，并在经过的路径上释放信息素引导后面蚂蚁的路径搜索，直至收敛到最优路径。蚁群算法已经被成功应用在许多 NP-hard 问题的求解上，然而对于 VRPTW 的相关研究较少。

大数据的数据量是异常巨大的，单一数据集的容量从太字节（TB）到十兆亿字节（PB）不等。传统方式下，对于这些数据的处理在时间上大大超出了人们可以忍受的范围。大数据具有数据量大、数据处理速度快、数据多样性的特点。云计算作为一种新型的并行处理技术，在大数据计算和网络存储方面表现优异。近年来流行的大数据框架很多，如 Spark、Hadoop、Storm、Samza 等。其中以 Spark

和 Hadoop 最为人们所熟知。

Apache Spark 是 Apache 软件基金会下的开源大数据并行计算框架，产生于加州大学伯克利分校 AMPLab。Spark 的运算是基于内存的，相比于 Hadoop MapReduce 在运行完成后将中间数据存放在磁盘，Spark 使用存储器内存运算技术，使得能够在内存中对数据进行分析计算，从而大大提高了计算的速度，运算时间方面相比于 Hadoop MapReduce 缩短了 10 到 100 倍。即便 Spark 在磁盘中运算，运算速度也能比 Hadoop MapReduce 提高 10 倍。Spark 基于弹性分布式数据集 (Resilient Distributed Datasets, RDDs)，RDD 有着较好的容错机制，常用的容错方法有检查点和日志，由于检查点时间一长容易导致数据冗余从而增大网络传输开销，所以 Spark 采用日志更新的方式实现容错。RDD 有着较 Hadoop MapReduce 更为丰富的数据转换操作，为程序的开发提供了便利。因其在大数据处理中运算速度快的特点，使得 Spark 成为了 Apache 软件基金会众多开源项目最为活跃的项目之一。

## 1.2 国内外研究现状

### 1.2.1 VRP 问题研究综述

上世纪 60 年代起，VRP 问题开始引起了越来越多研究人员的注意，研究的方向主要集中在以下几个方面：

#### (1) 精确算法

在研究早期，VRP 涉及的客户节点较少、问题规模较小，众多学者把研究重点放在了精确算法的研究上，主要的精确算法包括以下几种：

(i) 二下标车辆流。Laporte<sup>[5]</sup>针对 CVRP 及 DVRP，在三下标车辆流的基础上，去掉了代表车辆的下标，转而引入了所需车辆数的下界。并在规模为 60 的测试用例上进行了实验。

(ii) 集分割法。Balinski<sup>[6]</sup>针对具有严格约束的 VRP 问题，提出了集分割法。该方法以可行解集为基础进行优化，因而对应 VRP 模型的建立也最为简单。如果 VRP 问题的约束条件比较宽松，则会造成算法的可行解搜寻范围过大，对于

可行解成本下限的确定也较为困难。

## (2) 启发式算法

精确算法的优点在于,对于特定问题,当问题规模较小时,可以求得最优解;缺点是随着问题规模增大,时间复杂度会呈现指数爆炸的增长方式,实际应用中数据量往往较大,因此精确算法在实际应用中存在局限性。最近几年,研究人员的研究重点开始集中于启发式算法。2001年, Breedam<sup>[13]</sup>将启发式算法分为经典启发式算法和元启发式算法。主要的启发式算法如下:

(i)扫描法。1974年 Gillett<sup>[10]</sup>提出基于极坐标的扫描法(Sweep Algorithm, SA)。基本思想是先计算出各个点的极坐标并按由小到大的顺序排序,根据极坐标大小归并到各个子路径中,然后通过 TSP 优化算法进行子路径优化。本质上是通过将极坐标相近的客户节点聚类到相同的子路径中。

(ii)插入法。顾名思义,即按照具体的插入规则,在已知边中迭代插入新节点直至形成一条完整路径,该算法适合用于求解 VRPTW<sup>[12]</sup>。

(iii)节约法。该算法由 Clarke-Wright<sup>[11]</sup>提出,所以又称为 Clarke-Wright 算法,是一种精确的简单启发式算法。方法基本流程如下:1.生成与客户节点数相同的路径;2.合并其中任意两条路径,并计算合并后的收益节约值;3.将收益节约值按照倒叙排序;4.根据约束条件进行路径归并,归并的终止条件是没有更好的解产生。

## (3) 元启发式算法

元启发式算法是一种通用的启发策略,是随机算法和局部搜索优化策略的化合物,它应用领域广泛,但算法的运行效率无法保证。从上世纪 70 年代开始,越来越多的学者开始了对它的研究,主要的研究方向如下:

(i)蚁群算法。蚁群优化算法最初由 Dorigo<sup>[17]</sup>等人于 1992 年提出。该算法模拟自然界蚁群觅食行为而设计。算法的基本思想是:初始化一批人工蚂蚁,最初蚂蚁以随机的方式搜索解空间,蚂蚁搜索完成后更新信息素。信息素渐渐蒸发,较短的路径上信息素浓度更高,从而指导后面的蚂蚁渐渐收敛到全局最短的路径上。

(ii)模拟退火算法。该算法是一种随机概率逐渐降低直至趋向于零的随机算法,



Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to [etd@xmu.edu.cn](mailto:etd@xmu.edu.cn) for delivery details.

厦门大学博硕士学位论文摘要库