

## 摘 要

随着射频识别技术、网络协议和数据存储技术的不断发展，车联网的通信速率得到了大幅提升，车联网对计算资源的需求不断增加，同时对时延的要求也日益严格。时延逐渐成为车联网中日益重要的指标。在这一背景下，云计算技术已经无法单独解决车辆网络中计算资源匮乏的问题。因此车辆边缘计算应运而生。边缘服务器的出现使得计算资源得以下沉至靠近车辆的路侧，车联网中的车辆可以将任务分配给边缘服务器。然而，尽管边缘计算有效地提升了计算效率，但计算资源始终是有限的，难以满足不断增长的需求。最近的研究表明，可以将任务写在给附近的车辆。针对这一问题，实施高效的车联网任务卸载策略来提高资源利用率是一个可行的解决方案。在进行任务卸载时，必然存在一些不可被卸载的任务。本文旨在存在不可卸载任务的情况下，以提高能源利用效率为目标，研究如何利用启发式算法解决车联网中的任务卸载问题。具体而言，本文的研究工作包括以下几个方面：

（1）为了提高资源的利用率，本文提出了基于边缘计算的任务卸载框架，并将问题建模为数学规划的形式。由于该问题属于 NP 难问题，无法在多项式时间内得到精确解，因此我们采用启发式算法即基因算法来解决。实验结果表明，我们提出的调度算法不仅能够提高任务的利用率，还能较好地兼顾公平性。另外，算法的运行时间较短，适应车联网实时任务处理环境。

（2）为了进一步加快算法的运行时间，更好的满足任务对时延的需求，将基于梯度的侏儒猫鼬算法修改为能够解决离散问题的侏儒猫鼬算法。通过重新新定义运算符号，来解决离散问题的更新步骤。实验结果表明，相比较于基因算法，修改后的侏儒猫鼬算法的运行时间更短，能够更好的满足任务对时延的需求。

**关键字：**车联网；边缘计算；任务卸载；启发式算法

# 第一章 绪论

本章主要探讨了车联网中能耗优先调度算法的重要性，介绍了相关研究现状和取得的成果，提出了本文的研究任务，并阐述了主要研究内容，同时介绍了本文的结构组织。

## 1.1 研究背景及研究意义

未来将迎来“万物互联”的时代。随着射频识别技术、网络协议、数据存储以及传感器技术的不断发展，越来越多的物理设备与互联网连接起来，实现了物体之间的信息交互，赋予机器以“智能”。这将深刻改变人们的生产和生活方式，为人们提供更便利、智能和高效的生活和工作体验。<sup>[1]</sup>

车联网是“万物互联”时代的典型应用，汽车的功能已经超出了传统的交通工具范畴，不再只是交通出行的工具，而变成了一个智能互联的计算系统，车辆将承担起对车辆自身的调度与管理的任务，以及满足影音娱乐等用户额外的需求。

车联网是指通过无线通信技术将车辆与互联网连接起来的新一代智能交通系统。它将车辆、道路和互联网三者有机地结合在一起，实现了车辆之间、车辆与基础设施之间以及车辆与云端之间的全面互联互通。

一方面，车联网可以提供丰富的实时信息，帮助驾驶员更好地应对路况变化。通过车载传感器和导航系统，车联网可以收集到车辆位置、速度、加速度等数据，并将这些信息实时上传至云端。同时，车联网也可以获取来自其他车辆和基础设施的交通信息，如拥堵情况、事故警报等。这些信息可以通过智能导航系统反馈给驾驶员，提供最佳的行车路线规划，减少交通拥堵，提升驾驶安全性。车联网还具有智能驾驶辅助功能，为驾驶员提供更加安全和便捷的驾驶体验。通过车载传感器、摄像头和雷达等设备，车联网可以实时监测车辆周围的情况，包括其他车辆、行人、障碍物等，提供预警和辅助驾驶功能。例如，当车辆发生危险驾驶行为或与其他车辆距离过近时，车联网可以及时发出警报，并辅助驾驶员采取相应的应对措施。此外，车联网还可以实现自动泊车、自动驾驶等功能，减轻驾驶员的驾驶负担，提高驾驶的便利性和舒适性。

另一方面，车联网还可以实现车辆与基础设施的互联互通，提高交通效率。通过与交通信号灯、停车场、充电桩等基础设施的连接，车联网可以实现智能交通管理。例如，当车辆接近红绿灯时，车联网可以与信号灯进行通信，优化信号灯的控制，减少等待时间，提高道路通行效率。同时，车联网还可以为驾驶员提供实时

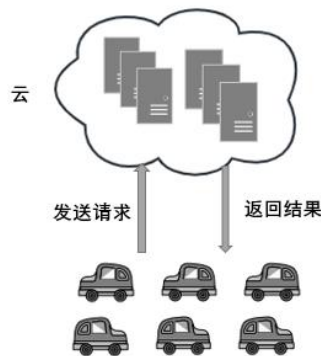


图 1-1 云计算

的停车位信息，方便寻找停车位，并避免因寻找停车位造成的交通堵塞。此外，车联网还可以与充电桩连接，实现电动车的智能充电管理，提升充电效率。

在车联网对车辆自身进行调度与管理的任务中，有些任务对计算资源的要求较高，且对时延敏感，因此需要大量的稳定的计算资源来保证其服务质量。显而易见，车载处理器的计算资源受限于汽车电池<sup>[2]</sup>，计算资源有限，无法满足对互联的计算任务的要求。

已经提出了许多方法来解决计算资源不足问题，例如通过提高能量的利用率，减少能耗侧面的增加计算资源。在文献<sup>[3]</sup>中通过考虑完整的车辆能量管理来减少能耗，在文献<sup>[4]</sup>中通过车辆的滑行来减少能量消耗。

为了解决计算资源受限的问题，基于云的解决方案通过将任务上传到云端（远端、远程云）来进行处理（如图 1-1），云端本身就是一个计算中心，通常具有超大的规模，能提供给用户超强的计算能力，对大规模数据进行聚合、挖掘、分析、优化，可以在短的时间内计算得到结果。

车辆应用产生的计算任务可以通过网络传输到远程云，满足任务对运算和储存资源的需求，在云端进行计算并将结果返回给车辆。由于云端超强的计算能力，因此能够计算和分析复杂的计算任务提供高性能计算资源。

但是因为云端位于网络的远端，在物理位置上往往部署在偏远地区，将车辆计算任务传输到远程云并得到相应的返回结果需要很大的传输时间开销，在数据传输过程中会产生较高的时延。因此对于数据密集型和延迟敏感型的计算任务，即使计算延迟较小（数据量较小），较大的传输延迟仍会使任务总卸载延迟较大。同时，为了将车辆产生的大量数据全部上传到云端，这将会导致网络中的数据量大

幅增加, 占用大量的网络带宽, 使得网络设施负担加重, 从而造成网络性能和稳定性的下降, 这就是云计算中链路过载的问题。

为了解决云计算的时延以及流量问题, 促进数据处理能力向数据源更接近的方向发展, 边缘计算 (edge computing, EC)<sup>[5][6][7]</sup>应运而生。EC 选择了在网络边缘中 (边缘端) 处理数据, 这里将“边缘”定义为数据源和云数据中心之间通道中的任何计算和网络资源。边缘计算。与传统的云计算相比, 边缘计算将服务部署在距离用户更近的地方, 放置到网络的边缘, 可以实现更低的时延和更高的带宽, 减少了核心网络的负载, 解决了云计算中链路过载的问题。同时, 这还有利于数据安全以及隐私保护<sup>[8]</sup>。

边缘计算也可以通过与物联网技术结合, 实现一些与智能车联网的应用, 如自动驾驶、人类行为识别、智慧城市等<sup>[9]</sup>。美国自然基金委和英特尔在 2016 年开始合作讨论建设无线边缘网络。同一年, 中国也兴起了移动边缘计算的研究, 在科研院所、高等院校、企业、行业协会等各方单位的共同努力下, 边缘计算产业联盟成立。

边缘计算并不是云计算的反面, 边缘计算是云计算的延伸, 未来将与云计算协同配合, 为用户提供更高质量的服务。

为了使车联网支持云计算和边缘计算, 科研组织和联盟实现了车到万物通信 (V2X) 技术<sup>[10]</sup>, V2X 主要包括车辆对车辆通信 (V2V)、车辆对基础设施通信 (V2I)。通过 V2X 技术, 车辆可以与其他车辆、交通信号灯、路边停车位、行人信号、路况监测设备等进行数据交换和信息共享。

最近流行的边缘计算技术中, 不仅可以将任务分配给附近的边缘服务器, 还可以分配给附近资源丰富的用户<sup>[11]</sup>。文献<sup>[12]</sup>中关注车载边缘云中的虚拟机资源分配。文献<sup>[13]</sup>中考虑了动态需求和资源约束, 并将所有任务分类为四种待处理列表。然后, 根据它们的特性, 每个列表中的任务将被卸载到不同的节点。

## 1.2 国内外研究历史和现状

### 1.2.1 任务卸载与车辆任务分类

任务卸载<sup>[14]</sup>是指将移动设备 (如智能手机、车载终端等) 上的计算任务分解成若干个子任务, 其中一部分在本地完成, 另一部分则通过网络卸载到云端或者其他设备上计算, 最终将结果返回到本地设备, 从而提高移动设备的计算性能和能耗效率的一种技术。根据卸载的目标, 任务卸载可以分为以下两类:

(1) 计算卸载: 主要目的是将计算密集型任务卸载到云端或其他远程设备上,

减少本地设备的计算负担，提高计算性能和能耗效率；

(2) 通信卸载：主要目的是将数据密集型任务卸载到云端或其他远程设备上进行处理，减少本地设备的数据传输负担，提高通信性能和能耗效率。

在本文中，主要研究的是车联网中的通信卸载。在车联网中，任务卸载主要应用于以下几个方面：

(1) 智能驾驶：智能驾驶需要大量的计算因此需要强大的数据处理能力，通过任务卸载技术，可以将一部分计算和数据处理任务卸载到云端或其他设备上，提高计算性能和能耗效率。

(2) 车辆监控：车辆监控需要实时收集和处理大量的数据，并进行数据分析和决策，通过任务卸载技术，可以将一部分数据处理任务卸载到云端或其他设备上，提高数据处理效率。

(3) 车辆诊断：车辆诊断需要进行大量的数据分析和处理，通过任务卸载技术，可以将一部分数据处理任务卸载到云端或其他设备上，提高诊断效率和精度。

智能车辆的任务并不是都可以卸载的，按关键程度来分，车辆任务可以分为三类：关键任务（Crucial Tasks, CTs）、高优先级任务（High-Priority Tasks, HPTs）和低优先级任务（Low-Priority Tasks, LPTs）。CTs 是车辆安全应用，是保证车辆和乘客安全的关键应用程序，如车辆控制、系统监控和事故预防等。由于 CTs 和安全紧密相关，因此享有除操作系统的系统应用程序外最高的优先级，必须保留充足的计算资源给它，不能因为 HPTs 和 LPTs 的存在而影响 CTs 的正常运行，因此这类任务也不允许卸载，只允许在本地执行，不属于可卸载任务的范畴。该类任务的实例是：车辆控制、车辆预警、红绿灯提醒、道路情况检测（例如道路湿滑程度检测）等。

HPTs 包括与驾驶相关的应用和可选的安全增强应用，这类应用程序对车辆而言是重要但不是必须的，拥有较高的执行优先级，例如实时路径规划和路况提醒等。所以这类应用可以卸载，且这类应用允许出现延迟或卸载失败的情况，但最重要的一点是，该类任务不能影响 CTs。该类任务的实例有：地图路线规划导航、视野增强、车辆传感等。

LPTs 是一类与影音娱乐服务相关的应用，它的优先级较低。例如语音识别，它允许驾驶员发出各种声音命令，通过语音识别命令计算机做一些响应，而不会使驾驶员分心。该类任务的实例是：虚拟现实、语音识别、影视音乐、在线游戏等。

HPTs 和 IPTs 已经被部署到越来越多的车辆上，由于 HPTs 和 LPTs 不会涉及到安全，因此可以将这两类任务进行卸载，来提高车辆资源的利用率。

### 1.2.2 任务卸载的最优化目标

(1) 最小化卸载时延。车辆所产生的任务可以选择在本地执行或者卸载到边缘服务器上执行，前者会产生本地时延和卸载时延，本地时延是指数据在本地运行所消耗的全部时间；卸载时延是指在边缘计算环境中，数据从生成到处理的时间间隔（一般忽略回传时延）。文献<sup>[15]</sup>最优化目标为处理时延。文献<sup>[16][17]</sup>设计了一种基于多臂老虎机理论的自适应学习任务卸载（ALTO）算法，以最小化平均卸载延迟。

(2) 最小能量消耗。通常而言，实际的能耗也是必不可缺的指标。对于车辆而言，可以提高车辆的行驶里程，环节“里程焦虑”。对于边缘服务器而言，可以最大限度地提高网络运营商和服务提供商的收入。<sup>[18][19]</sup>的目标为最小化能量消耗。

### 1.2.3 车联网任务卸载研究现状

车联网中任务卸载的常用算法包括基于强化学习的车联网任务调度、基于博弈论的车联网任务调度和基于启发式算法的车联网任务调度。

(1) 基于强化学习的车辆边缘计算任务调度算法研究。

随着深度强化学习（DRL）算法理论的不完善，DRL 已广泛应用于车联网中智能控制与资源管理与优化等方面<sup>[20]</sup>。DRL 算法可以感知环境的变化并做出决策，能满足车联网场景的需求，被认为是解决车辆边缘计算的有效方案。

DRL 算法主要包括智能体、环境、动作和奖励等部分。一般而言，模型通常把执行决策的主体定义为智能体，并把能够影响智能体进行决策的因素称为环境。智能体与环境进行交互，不断地学习积累经验，直至解决问题。

目前已有众多学者运用 DRL 方法解决车联网中任务卸载问题。如文献<sup>[21]</sup>提出了一种新的层次框架来解决云计算系统中的整体资源分配和电源管理问题，分层包括用于将虚拟机资源分配给服务器的全局层和用于本地服务器的分布式电源管理的本地层，以应对算法的高维问题，实现了服务器集群中实现延迟和能耗之间的最佳权衡。

文献<sup>[22]</sup>提出了一种基于深度强化学习的智能云资源管理架构，使云能够直接从复杂的云环境中自动高效地协商最合适的配置。文献<sup>[23]</sup>提出了一种基于 DRL 的、为云服务商提供资源配置和任务调度的方案，用于最大限度地降低具有大量服务器的云服务提供商的能源成本。文献<sup>[24]</sup>将问题分为两个子优化问题。分别发展了双侧匹配方案和深度强化学习方法，分别用于调度卸载请求和分配网络资源。绩效评估说明了我们构建的系统的有效性和优越性。文献<sup>[25]</sup>将任务卸载问题被分



解为两部分流重定向和卸载决策并提出了一种基于深度强化学习的方案来解决这个优化问题。

## （2）基于博弈论的车辆边缘计算任务调度算法研究。

博弈论是研究决策制定者在冲突或合作情境中进行策略选择的数学理论。当使用博弈论方法对车联网中的任务卸载进行研究时，首先要把目标最优问题转换为博弈论的基础问题，然后对其进行数学建模。建模的目的是利用博弈论来确定均衡状态。均衡状态是指所有的参与者都得到最佳收益的状态，当一个系统的博弈达到均衡时，每个参与者都无法仅通过改变自己的策略而在系统中得到更好的结果。博弈的本质就是多个相互影响的参与者在选择各自的决策时最大化自己的收益。

博弈论的方法主要包括以下几种：

（1）纳什均衡方法。纳什均衡是博弈论最重要的概念之一。它指的是在一个博弈中，每个玩家都采取了最优的策略，而且没有人想改变自己的策略。这种情况下，所有玩家都处于理性选择的状态，即达到了稳定状态。纳什均衡在很多实际问题中都有广泛的应用。

（2）最小化最大损失方法。最小化最大损失是一种双方博弈的决策方法，也被称为“最小保证赢法”。这种方法的基本思想是：在不确定对手的行动时，尽量减少自己可能遭受的最大损失。这种策略对于不确定性很大的问题非常有用。

（3）博弈树方法。博弈树是博弈论中常用的一种分析方法。它是一种图形，用于显示博弈的各种可能情况和策略选择，以及每种情况下玩家的收益或损失。通过分析博弈树，可以找到纳什均衡点和最优策略。

文献<sup>[26]</sup>将任务卸载子问题建模为精确势博弈，并提出了一种多智能体分布式深度确定性策略梯度来实现纳什均衡。将资源分配子问题划分为两个独立的凸优化问题，并利用基于梯度的迭代方法和 KKT 条件提出了最优解。

文献<sup>[27]</sup>将问题表述为潜在的博弈并证明它存在纳什均衡，设计了一种去中心化算法，用于在博弈中找到纳什均衡。

文献<sup>[28]</sup>将多用户卸载决策问题建模为非合作博弈，证明该问题存在纳什均衡，并提出了一种基于机器学习技术的全分布式计算卸载算法来寻找该纳什均衡点。

## （3）基于启发式算法的车辆边缘计算任务调度算法研究

启发式算法是一种解决复杂问题的计算方法，它通过利用经验或启示来指导搜索过程，以找到可能的解决方案。与精确算法不同，启发式算法可能无法保证找到最优解，但通常能够在合理的时间内找到接近最优解的解决方案<sup>[29]</sup>。启发式

表 1-1 启发式算法的主要思想

算法名称	主要思想
粒子群算法	鸟群或鱼群等生物群体的行为
模拟退火算法	固体材料退火过程中的原子结构调整
遗传算法	生物遗传和进化理论
蚁群算法	蚂蚁搜寻食物时的行为

算法是基于直观自然现象或者结构自然背后的经验构造的算法（如表1-1），在可接受的时空开销内给出问题的一个可行解（次优解）。贪婪算法是一种直观的启发式算法，它每一步选择当前最好的选择（局部最优），而不考虑长远影响（整体最优）。贪婪算法通常易于实现和理解，但可能会得到次优解。模拟退火算法受到固体退火过程的启发，通过接受一定概率的劣解，以避免陷入局部最优解，并逐渐趋向全局最优解。模拟退火算法适用于求解复杂问题，但需要合适的参数设置和调整<sup>[30]</sup>。

文献<sup>[31]</sup>通过考虑通用型任务计算时间分布来分析排队延迟，通过考虑给定概率满足延迟约束，以最小化受延迟和计算资源约束的 ITS 算子费用提出了一个非线性优化问题。针对该问题，该文提出一种线性化方法，并设计了一种基于整数线性规划的算法同时引入了一种称为任务卸载的成本效益启发式算法的启发式算法来解决该问题。

文献<sup>[32]</sup>研究了一个由多用户、一个边缘节点以及一个云服务器组成的移动云计算系统，提出了一个优化目标为用户之间的能耗、计算和最大延迟的总体成本的非凸二次约束二次规划问题，并提出了一种高效的启发式算法来解决这个问题。

Huynh 等人<sup>[33]</sup>开发了一种基于粒子群的启发式算法，用来解决异构网络中多用户多服务器的两层计算卸载策略问题，该问题的优化目标为总计算开销，包括完成时间和能耗。

文献<sup>[34]</sup>提出了一种用于共享交通流信息的联网车辆之间的通信框架，将联网车辆视为人工蚂蚁，能够根据交通流量的动态进行自我计算以做出自适应决策，使车辆能计算出到达目的地的最佳路径。

#### 1.2.4 车辆边缘计算面临的挑战

目前，基于车联网的车辆边缘计算面临着诸多挑战。

（1）实时性问题，车辆位置、速度、路况等方面的影响会导致车辆任务的实时变化，进而导致车辆任务卸载环境的不稳定性，上一时刻的最优分配方案可能下



一时刻就是最劣方案，导致任务卸载方案的可靠性降低。

(2) 资源动态变化问题，车辆位置、速度等方面的影响，车辆的计算等资源会发生相应改变。目前，大多数车联网任务卸载的研究都假设车联网中计算资源在较长时间内保持相对稳定。而在现实情况下，各种因素会导致资源状态发生明显的变化，导致车联网任务卸载算法的适应性降低。

(3) 卸载环境复杂问题，在车联网中，由于任务关键程度的不同，车联网任务卸载环境变得更加复杂。不是所有的任务都是可卸载的，对于那些不适合在边缘设备上进行处理的任务，应当优先本地来完成。这样可以保证车辆的行车安全。

根据以上的分析，车联网中任务卸载面临的问题对求解精确解的传统优化算法提出了挑战，传统优化算法智能求解特定的问题，以及高复杂度的特点，使得难以解决车辆边缘计算任务卸载。通过将任务卸载给附近的车辆是一种可能的解决方案，但是又带来了求解的方案难以兼顾公平的问题。

因此，在本文中，车联网存在不可卸载的任务情况下，以提高能源利用效率为目标，通过将最优化目标设置为能耗的平方来兼顾公平性，通过启发式算法来生成任务卸载方案。

### 1.3 论文的研究内容和组织结构

本文主要研究了车联网任务卸载算法，在车辆任务不能完全卸载情况下，即有部分任务必须车辆在本地执行的情况下，如何利用能耗与 CPU 频率之间的非线性关系进行任务卸载用来提高整体的能量利用率，使得整体能耗下降。

论文共分为五章，论文的具体内容安排如下：第 1 章对全文内容进行概述和章节安排，首先介绍了车联网和移动边缘计算研究背景和意义，并简述了边缘计算中任务卸载的国内外研究现状，对全文内容进行概述，最后给出论文组织章节安排。

第 2 章介绍了相关的知识背景，如数学规划的定义以及与整数规划形式和 0-1 线性规划形式两种不同的数学规划形式，其次综述了三种车联网任务卸载现有的相关算法研究。

第 3 章首先将问题进行建模为数学规划的形式，该问题分为任务可分情况和任务不可分情况，因为该问题是一个非线性整数规划问题，没有多项式时间范围内的解决方法，因此提出使用遗传算法解决该问题，并通过仿真实验验证了算法的性能实验。

第 4 章首先说明了贪心法调整的策略，然后重新定义了运算符，重新定义的

运算符有双目加法运算符、双目减法运算符、双目乘法运算符、单目调整运算符。通过重新定义运算符的方式，将侏儒猫鼬算法修改为解决离散问题的算法，解决本文提出的离散形式的问题。

第 5 章全文总结与展望。总结全文的研究内容和主要贡献，以及需要后续深入研究的问题。



## 第二章 任务卸载技术研究

本章介绍了相关的知识背景，如数学规划的定义以及与本文相关的数学规划问题的不同描述形式（整数规划形式和 0-1 线性规划形式），然后介绍了本文中使用的两种算法，基因算法和侏儒猫鼬算法，最后进行了章节小结。

### 2.1 相关知识背景

本节介绍了想要将算法修改为离散算法的相关知识。如数学规划的定义、数学规划问题的不同描述形式（整数规划和 0-1 线性规划）。

#### 2.1.1 数学规划

数学规划，又叫优化问题，是一类数学模型，用于在给定约束条件下寻找最优解决方案。它在现代科学、工程和经济领域中具有广泛的应用。常见的数学规划包括线性规划、整数规划、非线性规划和混合整数规划等类型。

线性规划（Linear Programming, LP）：线性规划是数学规划中最常见的类型。线性规划是在给定线性约束条件下最大化或最小化一个线性目标函数。线性规划的目标函数和约束条件都是线性的，这使得它具有许多良好的数学特性，例如可解析性、单调性和凸性。线性规划可以用单纯形算法、内点算法<sup>[35]</sup>等方法来求解。下面是一个线性规划的实例：

$$\begin{aligned} \max \quad & z = 40x_1 + 45x_2 + 24x_3 \\ \text{s.t.} \quad & 2x_1 + 3x_2 + x_3 \leq 100 \\ & 3x_1 + 3x_2 + 2x_3 \leq 120 \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0 \end{aligned}$$

整数规划（Integer Programming, IP）：整数规划是线性规划的扩展，其决策变量被限制为整数。这种类型的规划适用于需要做出离散决策的问题，比如资源分配、项目排程等。常用的方法有分支定界法<sup>[36]</sup>、割平面法<sup>[37][38]</sup>等。下面是一个整形规划的实例：

$$\begin{aligned} \max \quad & f = 5x_1 + 8x_2 \\ \text{s.t.} \quad & x_1 + x_2 \leq 6 \\ & 5x_1 + 9x_2 \leq 45 \\ & x_1, x_2 \in \mathbb{Z} \end{aligned}$$

非线性规划（Nonlinear Programming, NLP）：非线性规划是指目标函数或约束条件中存在非线性项的优化问题。非线性规划可以包括凸优化、非凸优化等问题，通常需要使用不同的算法进行求解。常用的方法包括梯度下降法、牛顿法、拟牛顿法等。

混合整数规划（Mixed Integer Programming, MIP）：混合整数规划是整数规划的进一步扩展，允许部分决策变量为连续变量，部分为整数变量。混合整数规划在实际问题中很常见，例如生产调度、设施选址等。求解方法包括分支定界法、启发式算法等。

数学规划是一种强大的工具，可以帮助我们在复杂的环境下寻找最优解决方案。通过使用数学规划，我们可以优化资源分配、提高效率、降低成本，从为我们的生活和经济发展做出更大的贡献。

### 2.1.2 数学规划问题描述形式

本论文的问题和并行机器调度问题（parallel machines scheduling problem）有同质性，并行机器调度问题是指在一些并行处理的机器上调度执行一组作业的问题。该问题与日常生活紧密联系，通常出现在工业车间生产、物流快递配送、计算机任务分配等领域，可以帮助企业有效地安排生产计划、降低成本、提高效率，也可以在计算机系统中动态地调度任务，提高系统总体性能。

该问题的形式化描述如下：有一个工件（任务）集合，其中，任务共有  $n$  个。还有一个机器的集合即，共有  $m$  个机器，所有的机器都是一样的，并且每台机器一次只能处理一个任务。每一项任务应该必须在其中一台机器上进行，在所有机器上运行的时间不会改变。任何一台机器处理任务  $i$  所需的时间由给出。调度中分配给机器的作业的子集用表示。作业一旦开始处理，就必须在不中断作业的情况下完成，即任务一旦开始做那就必须做完。

此外，作业之间没有优先关系，也就是说，在任何时间，作业可以不受限制地随意分配到任意一台机器上。研究了以最小化最后一个工件离开系统的时间为目标的最优工件分配问题，即最大完工时间准则。

最小最大完工时间的混合整数规划公式如下：

$$\min \max_{1 \leq j \leq m} \sum_{i=1}^n p_i x_{ij} \quad (2.1)$$

1	1	2	2	1	2
---	---	---	---	---	---

图 2-1 解的表示

$$\text{s.t. } \sum_{j=1}^m x_{ij} = 1 \quad i = 1, \dots, n \quad (2.2)$$

$$x_{ij} \in \{0, 1\} \quad (2.3)$$

公式2.1表明，本论文应该耗时最久的那台机器尽可能地小，因为所有的作业完成才算是完成了任务。公式2.2表明每个作业必须分配到一台机器上。如果  $x_{ij} = 1$ ，意味着作业  $i$  被分配到  $j$  的机器上。

然而，工件问题这个形式的空间复杂度是  $O(mn)$ ，而且因为每个变量都是二进制变量，所以时间复杂度是  $O(2^{mn})$ 。本论文将并行机器的调度表述为另一种形式：

$$\min \max_{1 \leq j \leq m} \sum_{1 \leq i \leq m, x_i=j} p_i \quad (2.4)$$

$$\text{s.t. } x_i \in \{0, 1, \dots, m\}, \quad i = 1, \dots, n \quad (2.5)$$

其中，工件问题该形式的空间复杂度为  $O(n)$ ，时间复杂度为  $O(m^n)$ 。

工件问题每个变量之间的取值有相关性，如几个变量的值交换，很大的概率就会导致最优解变成次优解甚至是劣解。

后文中，称呼每一个  $x_i$  为变量，所有变量的解集合即  $X = (x_1, x_2, \dots, x_n)$  称为该问题的一个解。

在后文设计离散形式的算法时，解的表示是一个关键的步骤，解具有与问题域相关的必要的信息。例如上文中使用基因算法解决问题，基因算法为了表示解，将解重新进行了编码，用二进制字符串的形式表示解。为了建立算法和问题之间的联系，对于有  $n$  个任务的问题，对应的解的维度也是  $n$ 。因此，在后文进行离散算法表示时，并行机器调度问题的解是用一个长度等于作业数量的数组表示的。

如图 2-1 所示，一共有六个任务，第 1, 2, 5 号任务放在第一台机器上处理，第 3, 4, 6 号任务放在编号为 2 的机器上进行处理。



## 2.2 基因算法

因为本文的问题是一个 NP 难（NP-Hard）问题，多项式时间内求解不出该问题的精确解，直接计算法求解是不现实的，因此采用了基因算法求解该问题。

基因算法是一种典型的启发式智能优化算法。其鲁棒性来源主要有三点：首先来源于个体多样性，通过引进交叉、变异等操作，使得不同个体之间进行信息的交流，有助于保持种群的多样性，从而不易陷入问题的局部最优解。其次来源于选择策略。基因算法中的选择策略来自于达尔文自然选择学说中的“适者生存”原则，通过对适应度高的个体进行选择 and 繁殖，逐步提高整个种群的适应性，这提高了算法的鲁棒性。最后来源于参数设置。基因算法中的参数设置，例如交叉的概率、变异的概率，这对算法的表现具有较大影响。通过对参数的合理设置，可以提高算法的鲁棒性，并且使其更容易达到全局最优解。

基因算法已被应用于各个领域。在文献<sup>[39]</sup>中，Sun 等人使用基因算法设计 CNN 架构。张等人<sup>[40]</sup>使用基因算法来最小化能量消耗。Yoon 等人<sup>[41]</sup>提出了一种高效的基因算法，使用新颖的归一化方法来解决无线传感器网络中的最大覆盖部署问题。在复杂的三维环境中，Roberge 等人<sup>[42]</sup>使用它计算固定翼无人机的可行和准最优轨迹。

以下是基因算法的实现步骤：

(1) 使用初始化参数来初始化种群：设置种群的大小，每个个体的基因编码方式和以及初值范围，设置循环最大次数为，然后根据这些参数生成初始种群。同时设置目标函数以及约束函数族。

(2) 评估适应度：将每个个体的基因编码字符串解码成相应的解，然后计算其适应度值。适应度值是个体在解空间中执行任务的优劣程度的量化指标。

基因算法二进制解码公式如下：

$$\delta = \frac{U_x - L_x}{2^l - 1} \quad (2.6)$$

$$x = L_x + \delta \cdot \sum_{i=1}^l A_i 2^{i-1} \quad (2.7)$$

其中， $\delta$  是偏移量， $l$  是二进制码的长度， $A_i$  是第  $i$  位二进制位的值， $x$  是解码后的值。例如，给定  $U_x = 5$ ， $L_x = -5$  以及  $l = 10$ ，那么  $\delta = \frac{5 - (-5)}{2^{10} - 1} \approx 0.009775$ 。

基因算法二进制编码二进制长度公式如下：

$$l = \left\lceil \log_2 \left( \frac{U_x - L_x}{\text{SearchAccuracy}} \right) \right\rceil \quad (2.8)$$

其中， $U_x$  和  $L_x$  分别是解的最大值和最小值， $\text{SearchAccuracy}$  是求解的精度。例如，给定  $U_x = 5$  和  $L_x = -5$ ，精度是 0.01，那么  $l = \left\lceil \log_2 \left( \frac{5 - (-5)}{0.01} \right) \right\rceil = 10$ ，也就是说，需要十个二进制位来表示解。

适应度计算公式如下：

$$fit_i = e^{-\frac{y_i}{\text{mean}(y)}} \quad (2.9)$$

其中， $y_i = f(x_i)$ ， $\text{mean}(y)$  是种群中所有的个体的平均值。

(3) 选择操作：按照轮盘赌选择的规则，从种群中选择一部分较好的个体进行进化操作，以期得到更加优秀的后代个体。在进行轮盘赌选择时，标准化处理的公式为：

$$p_i = \frac{\hat{fit}_i}{\sum_{i=1}^n \hat{fit}_i} \quad (2.10)$$

其中， $n$  是种群中的个体数量。

(4) 进行遗传操作：包括交叉操作和变异操作。交叉操作是将两个个体的染色体进行配对，然后按照一定的概率产生新的后代染色体；变异操作是随机改变个体染色体的一个或多个基因，以增加种群的多样性。

(5) 更新种群：这是“适者生存”法则在基因算法中的应用。将新生成的后代个体插入到种群中，也就是按照一定的比例，某些适应度高的个体替换掉适应度低的个体。

(6) 判断遗传算法是否满足停止准则：停止的准则可以设置最大迭代次数或者在种群中出现一定数量的相似个体等条件，决定是否终止遗传算法。如果不满足条件则跳转到第（2）步。

## 2.3 侏儒猫鼬算法

### 2.3.1 算法简介

和粒子群、基因算法类似，侏儒猫鼬算法（Dwarf Mongoose Optimization Algorithm, DMOA）也是一种基于动物行为的启发式优化算法，能够用来解决线性以及非线性的优化问题。该算法是由 Jeffrey O. Agushaka<sup>[43]</sup>等科学家于 2022 年提出的，算法的灵感源于侏儒猫鼬的捕猎行为，通过模拟侏儒猫鼬在觅食和捕猎时的行为特征<sup>[44]</sup>来进行对问题的求解。



图 2-2 侏儒猫鼬

DMOA 在理论上比文献中的一些算法更能找到不同优化问题的全局最优解，因为它具有以下独特的属性：DMOA 特殊的探索和开发能力，这模仿了侏儒猫鼬的半游牧行为和补偿适应。DMOA 只有一个可调参数，可以提供更高的效率和可靠性。简化调优过程：开发人员只需关注一个参数的调整，而无需考虑其他参数的影响。提高效率：更容易地找到最佳参数配置。提高可靠性：减少了参数之间的相互影响，使算法更加稳定和可靠。当参数之间存在复杂的交互作用时，调优过程可能会变得复杂且困难，同时也增加了算法出现错误或不稳定性的风险。

侏儒猫鼬算法被应用于机器学习，建模等方面。文献<sup>[45]</sup>提出了一种改进的侏儒猫鼬优化器的智能元启发式形式，用于太阳能光伏系统的最佳建模和参数提取。文献<sup>[46]</sup>提出了一种二元侏儒猫鼬优化算法，以解决机器学习模型在数据挖掘种高维特征选择问题。文献<sup>[47]</sup>使用经过侏儒猫鼬优化算法（DMOA）调整的 SqueezeNet 轻量级的卷积神经网络来预测心脏病。文献<sup>[48]</sup>提出了一种基于矮猫鼬优化的多跳路由方案（DMOSC-MHRS）的安全集群来生成到达目的地的最佳路线。

### 2.3.2 侏儒猫鼬

侏儒猫鼬（英文名：Dwarf mongoose）是一种小型的食肉动物，属于猫鼬科。它们主要分布在非洲撒哈拉以南地区的草原、稀树草原和山地森林中。

侏儒猫鼬的体型相对较小（如图 2-2），平均身長约为 25-30 厘米，尾长约为 15-20 厘米。它们的体重通常在 200-350 克之间。其身体呈灰褐色，背部有黑色纵纹，腹部呈浅黄色。侏儒猫鼬是杂食性动物，除了昆虫和小型动物外，它们还会食

用水果、种子和其他植物物质。它们的食性多样性使它们能够适应不同的环境条件。它们主要在白天活动，它们具有机敏的行动能力和快速适应环境的能力。

侏儒猫鼬通常生活在母系社会的家庭群体中，由一对终身结合的阿尔法配偶领导。在家庭群体中，雌性猫鼬的地位高于雄性<sup>[49]</sup>。根据年龄和性别，不同的猫鼬扮演着守卫、保姆等不同的角色<sup>[50]</sup>。

### 2.3.3 侏儒猫鼬算法种群分组

为了模仿它们的觅食行为，优化过程建立了三个社会结构：阿尔法小组，侦察兵小组，保姆小组。

(1) 阿尔法小组。阿尔法小组是又进行出发去觅食（探索新的可行解）的行为的种群，食物的位置由下面的公式给出：

$$X_i^{t+1} = X_i^t + \text{phi} \times \text{peep} \times (X_i^t - X_{\text{peep}}) \quad (2.11)$$

其中， $X_i^{t+1}$  是新产生的解， $\text{phi}$  服从于在  $[-1,1]$  之间的均匀分布。 $\text{peep}$  是雌性首领的数量， $X_{\text{peep}}$  是选出的雌性首领的位置。

与侏儒猫鼬的行为不同，雌性首领在阿尔法小组中产生，每个个体都有可能成为雌性首领，适应度高的个体成为雌性首领的可能性更大，适应度差的个体成为雌性首领的可能性小，每个个体成为雌性首领的概率计算方式如下：

$$\alpha = \frac{\text{fit}_i}{\sum_{i=1}^n \text{fit}_i} \quad (2.12)$$

其中， $\text{fit}_i$  是第  $i$  个个体的适应度， $n$  是种群数量。

适应度的计算方式为：

$$\text{fit} = e^{-\frac{y}{y}} \quad (2.13)$$

其中， $y$  为目标函数的值。

计算适应度使用公式2.13的好处有：首先，考虑了每个个体的表现与平均表现的比值，因此能够根据整体表现水平动态地调整适应度值。这意味着在不同阶段，个体的适应度能够更好地反映其在当前群体中的相对表现。其次，数函数的性质能够有效地抑制过大的适应度值，使得在优化过程中，个别极端值对整体优化的影响被降低，从而增强了算法的鲁棒性。最后，体现相对优势：通过将个体的表现与整体平均水平进行比较，适应度算法能够更好地体现个体的相对优势，从而促进种群中优秀个体的选择和保存。

(2) 保姆小组。保姆通常是附属的群体成员，和幼仔呆在一起，定期轮换，让首领带领其他成员进行日常觅食。保姆的数量取决于种群规模，种群越大，保姆数量越多。保姆通常在中午和晚上回来给幼崽喂奶。也就是说，保姆小组的成员不进行觅食行为。

在算法中，当一个个体很久没有进行觅食（更新）时，就说明不是陷入了局部最优，就是这个解太差了，更新的可能比较小，将其放入保姆小组中，并重新进行初始化。

(3) 侦察兵小组。侦察兵寻找下一个睡觉的土堆（睡眠丘），因为猫鼬不会回到先前睡觉的土堆，这保证了探险的有效性。在该算法中，侦察兵小组和阿尔法小组是一个小组，首先会进行阿尔法小组的更新行为，然后再进行侦察兵小组的更新行为。

睡眠丘的计算公式如下：

$$sm_i = \frac{fit_{i+1} - fit_i}{\max \{|fit_{i+1}, fit_i|\}} \quad (2.14)$$

睡眠丘的平均值反映了该轮迭代的移动范围， $\varphi$  计算公式如下：

$$\varphi = \frac{\sum_{i=1}^n sm_i}{n} \quad (2.15)$$

CF 是表示侦察兵小组移动能力的参数，其计算公式如下， $iter$  为当前迭代次数， $T$  为最大迭代次数。

$$CF = \left(1 - \frac{iter}{T}\right)^{\left(2\frac{iter}{T}\right)} \quad (2.16)$$

在图 2-3 中， $x = \frac{iter}{T}$ ， $x \in [0, 1]$ ，当迭代次数增加时，其值会减小，也就是说，随着迭代次数增加，侦察兵小组移动能力会减小。

侦察兵小组的更新方式如下：

$$X_{i+1} = \begin{cases} X_i - CF * \text{phi} * \text{rand} * [X_i - \bar{\mathbf{M}}] & \text{if } \varphi_{i+1} > \varphi_i \\ X_i + CF * \text{phi} * \text{rand} * [X_i - \bar{\mathbf{M}}] & \text{else} \end{cases} \quad (2.17)$$

其中， $\text{phi}$  服从于  $[0, 1]$  之间的均匀分布， $\bar{\mathbf{M}} = \sum_{i=1}^n \frac{X_i \times sm_i}{X_i}$

### 2.3.4 算法流程

优化从阿尔法小组出发（探索空间）觅食开始，将保姆小组留在巢穴中。一旦找到觅食地点，阿尔法小组就一直觅食到中午（保留更新值）。

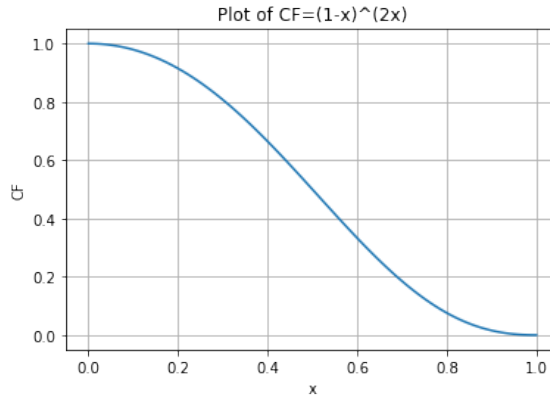


图 2-3 CF 变化图

当他们返回交换保姆时，按照保姆交换标准，很久没觅食的成员会成为保姆。一旦保姆被交换，他们就不会回到先前觅食的地方，以避免过度放牧。

侦察兵会发现一个新的觅食地点，并通知雌性首领，带领家族到达新的地点。

可以看出，侏儒猫鼬算法中种群的主体部分（阿尔法小组或者侦察兵小组）一次迭代中会进行两次更新行为。

显而易见，侏儒猫鼬算法在进行每一步的更新时，使用了类似于梯度（如公式2.18）的更新方式，因此该算法只能解决连续形式的问题，不能解决离散形式的问题，需要本论文将其改造为像基因算法一样的离散型算法。

$$X_i^{t+1} = X_i^t + \text{phi} \times \text{peep} \times (X_i^t - X_{\text{peep}}) \quad (2.18)$$



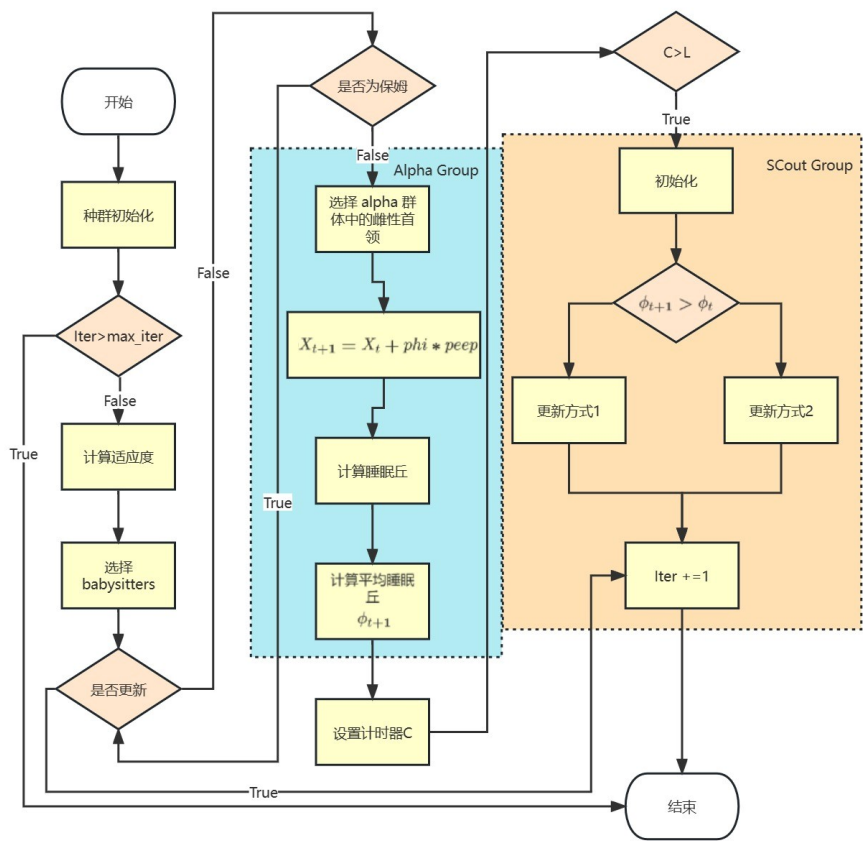


图 2-4 侏儒猫鼬算法流程图

2.4 本章小结

本章详细介绍了将算法修改为离散形式的相关知识，包括数学规划和数学规划描述形式本章所介绍的内容将为后续离散算法的提出提供借鉴的基础。

### 第三章 基于基因算法的车辆任务卸载算法

本章首先将问题进行建模为数学规划的形式，该问题又分为任务可分情况和任务不可分情况，因为该问题是一个非线性整数规划问题，没有多项式时间范围内的解决方法，因此提出使用遗传算法解决该问题，并通过仿真实验验证了算法的性能实验。实验结果表明，算法可以达到节能的目的的同时很好的兼顾公平性。

#### 3.1 系统模型

本节主要介绍了系统模型，分为三个部分，系统场景，车辆计算能力模型，以及车辆能耗模型。

##### 3.1.1 系统场景

在本文中，如图 3-1所示，层次结构由车辆云、边缘云和中心云组成。中心云负责全局调度，它执行的任务包括执行复杂的计算任务和进行全局的决策。边缘云是车辆云的控制器，负责创建、维护和删除车辆云。车辆云由一定范围内的、参与共享其计算资源的智能车辆组成。基站、边缘服务器和远程的中心云服务器通过有线来连接，基站和边缘服务器物理上距离十分接近，因此可以看作一体，后文中以边缘云来称呼这两者的总和。

每辆车都可以访问边缘云并与其进行通讯。车辆可以通过将任务卸载到其他车辆的方式来节省能源。这样可以提高整体资源利用率。

任务卸载过程如下：首先，实时流量信息被传输到云服务器进行分析。实时信息包括目的地、当前位置、时间、车速等。云服务器在汇总数据后，进行大数据分析以获取道路拥堵情况，然后推断未来某个时间段内车辆的位置信息，并将结果返回到车辆附近的边缘云。

这些车辆在未来一段时间内将会参与资源共享。然后，边缘云根据这些信息分配任务，并将车辆分为提供者和请求者，他们在  $T$  时间段共享资源。

##### 3.1.2 车辆计算能力模型

如何定义车辆的计算资源是建立起问题描述非常重要的一步，在此，我使用计算每秒执行的指令条数来衡量车辆的计算资源，本论文将资源共享时间定义为  $T = \{1, \dots, T\}$ 。车辆数量定义为  $N$ 。

本使用元组  $J_{it} = \{r_{it}, r'_{it}, d_{it}\}$  来表示时间为  $t$  时车辆  $i$  的任务大小， $d_{it}$  是任务的数据大小。

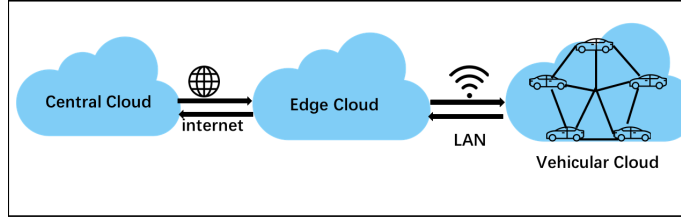


图 3-1 系统场景图

任务分为必须在本地执行的任务，以及可以分配出去的任务。这一点并不抽象，例如有些涉及隐私的任务，或者是对实时性要求高的任务，就必须在本地执行。

在时间  $t$  时， $r_{it}$  是车辆  $i$  必须要在本地执行的任务， $r'_{it}$  是车辆  $i$  能够分配给别人的任务，并且它们的值是由需要的指令条数来表示的。

为了更好的描述这个模型，本论文定义  $\mathbf{X}_t = \{x_{ij}\} \in \{0, 1\}^{N \times N}$  为分配矩阵，如果  $x_{ij} = 1$ ，代表车辆  $i$  执行车辆  $j$  所卸载的任务。注意，如果  $x_{ii} = 1$ ，车辆  $i$  所有的任务都在本地执行。

本论文将车辆  $i$  在时间为  $t$  时的容量定义为  $C_{it}$ 。在时间  $t$ ，当车辆被选为提供者 (P)，所有需求者需要的资源需要少于这辆车的容量。公式化表达如下：

$$\sum_{j=1}^N x_{ij}^t \cdot r'_{jt} \leq C_{it}, i = 1, \dots, N \quad (3.1)$$

一个关键的步骤是怎样衡量车辆的计算资源，在本篇文章中，计算资源定义在每秒能够执行的指令条数上。公式化表达如下：

$$C_{it} = M_{it} \cdot \Delta T - r_{it} \quad (3.2)$$

其中， $\Delta T$  是资源共享的持续时间，并且它是一个比建立车辆边缘网络的更小的时间粒度，在  $\Delta T$  秒后，分配矩阵会发生变化。

对于一个单核的 CPU，每秒能够执行的指令条数 ( $m_{it}$ ) 和 CPU 的频率 ( $f_{it}$ ) 有如下的关系：

$$M_{it} = v_i \cdot f_{it} + \theta_i \quad (3.3)$$

其中， $v_i$  和  $\theta_i$  是待估计的参数。

最后，公式3.2中的 CPU 的容量  $C_{it}$  在能被下面的公式计算出来：

$$C_{it} = (v_i \cdot f_{it} + \theta_i) \times \Delta T - r_{it} \quad (3.4)$$

### 3.1.3 车辆能耗模型

当本论文考虑车辆的能源消耗时，本论文将它分为两个部分（1）计算所需要的能量以及（2）传输所需要的能量。

根据<sup>[51][52][53][54]</sup>，计算所需要的能耗能够被下列公式计算：

$$E = \lambda_i \cdot f_{it}^3 \cdot \Delta T \quad (3.5)$$

其中， $f_{it}$  是 CPU 在  $t$  时刻的频率。如果一辆车被选为了需求者（R），这辆车的频率会下降为  $f'_{it}$ 。因为任务被分配出去了，所以他消耗的能量会减少。消耗的能量能够被以下公式计算：

$$E_i^{save} = (\lambda_i \cdot f_{it}^3 - \lambda_i \cdot f'^3_{it}) \times \Delta T, \forall i \in R \quad (3.6)$$

传输能耗和传输的时间有线性关系，传输的时间取决于数据大小和传输速率的比值 ( $b_{ij}$ ):

$$E = P_0 \cdot \frac{d_{it}}{b_{ij}} \quad (3.7)$$

其中， $P_0$  是传输率。尽管功率可以变化<sup>[55][56][57]</sup>，但为了简单起见，本论文在本文中使用固定值。

最大的传输速率 ( $b_{ij}$ ) 可以通过香农公式计算：

$$b_{ij} = W \log(1 + \text{SNR}) \quad (3.8)$$

其中，SNR 是信噪比， $W$  是频道的带宽。因为它和每一个场景相关，本论文将它考虑为一个常数。

对于每一辆车，接收信息所消耗的能量是：

$$E_i^{rec} = \sum_{j=1, j \neq i}^N x_{ij}^t \cdot P_0 \cdot \frac{d_{jt}}{W \log(1 + \text{SNR})}, \quad i = 1, \dots, N \quad (3.9)$$

对于每一辆车，发送信息所需要的能量是：

$$E_i^{send} = \sum_{i=1, i \neq j}^N x_{ij}^t \cdot P_0 \cdot \frac{d_{it}}{W \log(1 + \text{SNR})}, \quad j = 1, \dots, N \quad (3.10)$$

定义  $E_i^{blnc}$  是车辆  $i$  在时刻所消耗的能量总和，它可以被这么计算：

$$\begin{aligned}
 E_{it}^{blnc} = & \sum_{j=1, j \neq i}^N x_{ij}^t \cdot P_0 \cdot \frac{d_{jt}}{W \log(1 + \text{SNR})} \\
 & + \sum_{j=1, j \neq i}^N x_{ji}^t \cdot P_0 \cdot \frac{d_{it}}{W \log(1 + \text{SNR})} \\
 & + \lambda_i \cdot f_{it}^3 \cdot \Delta T, \quad i = 1, \dots, N
 \end{aligned} \tag{3.11}$$

本论文算法的设计目标是：最小化所有车辆能量消耗的平方。平方和和普通的和相比，平方和更容易受到极端值的影响，因为平方项会使得较大或较小的值对平方和的贡献更加显著。

这个目标既能够考虑最小化能量消耗，又能考虑公平，也就是说平衡了能量消耗和公平。公式表示如下：

$$\min \sum_{t=1}^T \sum_{i=1}^N (E_{it}^{blnc})^2 \tag{3.12}$$

#### 3.1.4 分配矩阵例子说明

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$\mathbf{X}$  是一个分配矩阵的例子。车辆集合为  $\{A, B, C, D, E\}$ ，车辆 A、B 和 D 没有分配给其他车辆任务，因此它们不会产生传输能量消耗。而车辆 C 执行车辆 E 的卸载任务。每列的总和大于或等于 1。

表 3-1 本文符号说明

参数	解释
$T$	资源共享时间
$C_{it}$	CPU 容量
$r_i$	车辆 $i$ 的工作负载
$r'_{jt}$	资源请求者 $j$ 需要的资源
$M_{it}$	每秒百万次指令数 (MIPS)
$\theta_i$	估计参数
$E_{it}^{blnc}$	车辆 $i$ 消耗的能量
$E_i^{rec}, E_i^{send}$	车辆 $i$ 接收或发送的能量

## 3.2 问题建模

在本节中，本论文将问题分为两种情况来建模，一种是任务可分情况（3.3.1 节），另一种是任务不可分情况（3.3.2 节）。当任务可分的情况下，问题能够直接求解数学规划的方式解决；当任务不可分时，不能通过求解数学规划的方式来解决。

### 3.2.1 任务可分情况

当车辆在进行某些同质任务时，可以将大任务任意的拆成一些小任务，因为任务的粒度太小，因此可以近似看成任何一个任务可以随意拆分然后分配到任意的车辆上，也就是任务可分的情况。

在任务可分时，根据本地留存的高优先级任务的大小又分为：高于平均任务（大任务），和低于平均任务大小（小任务）这两种情况。

#### （1）小任务情况

当本地任务低于平均任务大小（小任务）时，该问题较为简单，直接将所有任务平均分配到车辆上就能解决。可以看出，该问题等价于：首先将固定值  $C$  切制成  $n$  个数字，使得的他们的三次方之和最小。设切分的数分别为  $a_1, a_2, \dots, a_n$ ，根据均值不等式： $\sqrt[n]{\frac{\sum_{i=1}^n a_i^3}{n}} \geq \frac{\sum_{i=1}^n a_i}{n}$ ，可得： $\frac{\sum_{i=1}^n a_i^3}{n} \geq \left(\frac{\sum_{i=1}^n a_i}{n}\right)^3$ ，其中当  $a_1 = a_2 = \dots = a_n = \frac{C}{n}$  时，等号成立。

因此，要使  $\sum_{i=1}^n a_i^3$  最小，需要让  $\sum_{i=1}^n a_i$  的值尽可能平均分配给  $n$  个数，即令  $a_1 = a_2 = \dots = a_n = \frac{C}{n}$ 。

当本地任务是小任务时，能够保证参与资源共享的车辆，任务的大小可以卸载为平均值。

#### （2）大任务情况



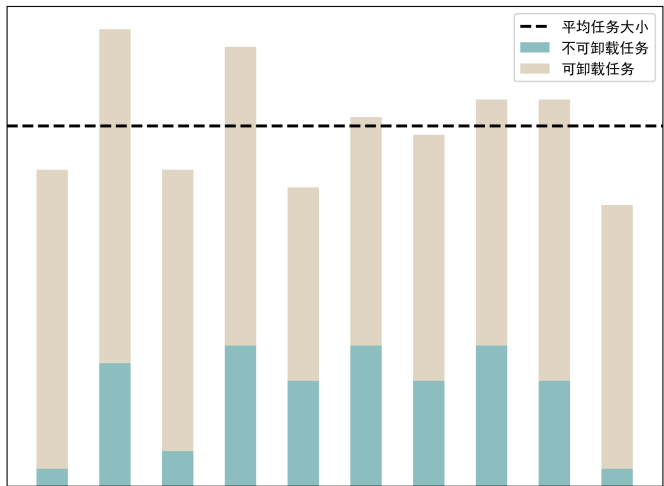


图 3-2 小任务情况

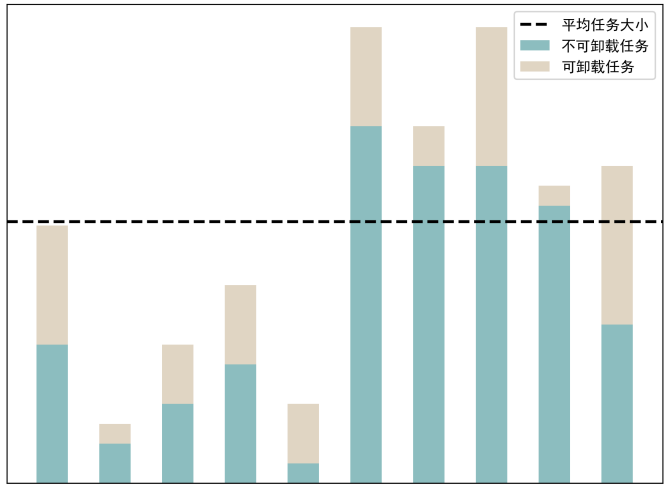


图 3-3 大任务情况

但是当任务高于平均任务大小（大任务）时，该问题需要进行分类的讨论，是否还需要将再给有大任务的汽车分配任务。

该问题等价于：有一个固定值  $C$ ，将其切割为  $n$  个数，要求这  $n$  个数的三次方最小，已知有  $m$  个数必须大于  $\frac{c}{n}$ ，如何切割。

在有限制条件的情况下，该问题可以使用拉格朗日乘数法来求解。设切分后的数为  $a_1, a_2, \dots, a_n$ ，且有  $m$  个数大于  $\frac{c}{n}$ 。

有以下约束条件：

$$\begin{cases} \sum_{i=1}^n a_i = C \\ x_i - a_i \leq 0, i = 1, \dots, m \end{cases} \quad (3.13)$$

目标函数为:

$$\min \sum_{i=1}^n a_i^3 \quad (3.14)$$

根据拉格朗日乘数法, 构造带拉格朗日乘数  $\lambda_1, \lambda_2, \dots, \mu$  的拉格朗日函数:

$$\begin{aligned} L(a_1, a_2, \dots, a_n, \lambda_1, \lambda_2, \dots, \lambda_m, \mu) \\ = \sum_{i=1}^n a_i^3 + \sum_{i=1}^m \lambda_i (x_i - a_i) + \mu \left( \sum_{i=1}^n a_i - C \right) \end{aligned} \quad (3.15)$$

其中,  $\lambda_i, \mu$  是拉格朗日乘子。

如公式3.16所示, 分别对  $L$  关于  $a_i$ 、 $\lambda_i$ 、 $\mu$  求偏导, 并令其等于零。

$$\begin{cases} \frac{\partial L}{\partial a_i} = 0 \\ \frac{\partial L}{\partial \lambda_i} = 0, i = 1, \dots, m \\ \frac{\partial L}{\partial \mu} = 0 \end{cases} \quad (3.16)$$

得到以下结果:

$$\begin{cases} \frac{\partial L}{\partial a_i} = 3a_i^2 - \lambda_i + \mu = 0, & i = 1, \dots, m \\ \frac{\partial L}{\partial a_i} = 3a_i^2 + \mu = 0, & i = m+1, \dots, n \\ \frac{\partial L}{\partial \lambda_i} = a_i - x_i = 0, & i = 1, \dots, m \\ \frac{\partial L}{\partial \mu} = \sum_{i=1}^n a_i - C = 0 \end{cases} \quad (3.17)$$

根据  $\frac{\partial L}{\partial \lambda_i} = 0$  可得对于所有必须大于平均值的  $a_i$ ,  $a_i = x_i$ 。根据  $\frac{\partial L}{\partial a_i} = 0$ , 可得  $a_i = \sqrt{\frac{-\mu}{3}}$ , 即对于没有限制的  $a_i$ , 其值都相等。

可得最优的分配方案是: (1) 对于本地任务是大任务的车辆, 将他们所有可分得任务分出去。(2) 其余车辆再均分剩余的任务。

### 3.2.2 任务不可分情况

在任务不可分时, 该问题没有多项式时间内的解决方法, 因此, 在下一节中, 本文打算通过智能算法来获得一个近似解, 来满足靠近最优解和卸载方案计算时

间的平衡。该问题的优化目标是通过调度各个车辆的任务，最小化所有能耗的平方，该问题最终被建模为如下形式：

$$\min \sum_{t=1}^T \sum_{i=1}^N (E_{it}^{blnc})^2 \quad (3.18)$$

$$\begin{aligned} \text{s. t. } E_{it}^{blnc} = & \sum_{j=1, j \neq i}^N x_{ij}^t \cdot P_0 \cdot \frac{d_{jt}}{W \log(1 + \text{SNR})} \\ & + \sum_{j=1, j \neq i}^N x_{ji}^t \cdot P_0 \cdot \frac{d_{it}}{W \log(1 + \text{SNR})} \\ & + \lambda_i \cdot f_{it}^3 \cdot \Delta T, \quad i = 1, \dots, N \end{aligned} \quad (3.19)$$

$$\sum_{j=1}^N x_{ij}^t \cdot r'_{jt} \leq C_{it} \quad (3.20)$$

$$\sum_{i=1}^N x_{ij}^t \geq 1 \quad (3.21)$$

$$f_{it} \geq 0 \quad (3.22)$$

$$x_{it}^t \in \{0, 1\} \quad (3.23)$$

正如前文所提到的，公式 (3.20) 说明了对于每一辆车辆  $i$ ，需求的总和不能超过他的容量。公式 (3.21) 说明了车辆  $i$  的任务，必须要被执行，无论是本地执行还是分配给其他车辆。

### 3.3 仿真实验

#### 3.3.1 实验环境与仿真参数设计

本章实验环境为：处理器为：Intel(R) Core(TM) i7-7700HQ；显卡的型号为：NVIDIA GeForce RTX 1050；机带 RAM 为 8GB；操作系统为 Windows10 专业版；Python 版本为 3.8，Numpy 版本为 1.19.2，matplotlib 版本为 3.3.2。

本文参数设置如下，车辆 CPU 使用 Cortex-A57，Cortex-A57 处理器是 ARM 性能最高的处理器，旨在推动移动和企业计算应用。Cortex-A57 的频率从 700 MHz 到 1900 MHz。频率集为 700, 800, 900, ..., 1900，因此，只有 13 个频率级别。当车辆在 700MHz 处理不了任务时，频率升为 800MHz 来处理任务。

在表3-2中， $U[x, y]$  是  $x$  和  $y$  之间的均匀分布。 $N(x, y)$  是正态分布，均值为  $x$ ，方差为  $y$ 。

表 3-2 仿真参数

参数	值/分布	参数	值/分布
$f_i$	[700, 1900]	$r_{it}$	$U[820, 10000]$
$v_i$	7.683	$r'_{it}$	$U[200, r_{it}]$
$\theta$	-4558.52	$P_0$	0.2
$\lambda$	0.00125	$W$	10 MHz
$\Delta T$	1	SNR	677

由于我们的空间有限，我们假设车辆共享资源的时间为  $T = 10$  秒和  $\Delta T = 1$  秒。

如公式 (3.3) 所示，指令数为  $M_{it} = v_i \cdot f_{it} + \theta_i$ 。并且公式 (3.4) 中的 CPU 容量为  $C_{it} = (v_i \cdot f_{it} + \theta_i) \times \Delta T - r_{it}$ 。基于<sup>[58]</sup>中提供的分析，CPU 容量参数  $v_i, \theta_i$  其值分别为 7.683 和 -4558.52。

经过计算， $M$  的最大值和最小值分别为 819 和 10039。所有任务执行工作负载所需的计算资源  $r_{it}$  服从于  $U[820, 10000]$  的均匀分布，并且假设车辆  $i$  可以卸载给其他车辆的任务计算资源值  $r'_{it}$  服从于  $U[200, r_{it}]$ ，因为某些子任务必须在本地执行。数据的大小从 1MB 到 10MB 不等。我们设定带宽  $d$  的值为 27Mb/s。

本文选择了 7 种种类的参与资源共享的车辆数目，分别是 10, 15, 20, 25, 30, 35, 40。在算法中，种群数量被设置为 40。而当车辆数量为 35 和 40 时，种群数量被设置为 150。因为当车辆数量为 35 和 40 时，种群的规模太小，算法不会收敛或者不能很好的收敛到最优值附近。

本文将变异概率设置为 0.01，参数用于控制每个个体发生变异的概率。该概率越高，个体发生变异的可能性就越大。一般情况下，变异概率会设置为较大的值，以保持种群的多样性，并防止算法陷入局部最优解。

### 3.3.2 评价标准

本论文设置了个指标来衡量本论文的实验，分别是，性能指标，公平指标，种群离散程度。在实验中，本论文的性能指标是指节能的百分比。节能的百分比，其定义如下：

$$P = \min \left( 100 \cdot \left( 1 - \frac{\sum_{i=1}^N E_{it}^{blne}}{E_{lc}} \right) \right), t = 1, \dots, T \quad (3.24)$$

其中， $E_{loc}$  是本地执行任务所需要消耗的能量。为了反映本论文任务分配的公平性，本论文在标准差的基础上定义了公平系数 (fairness coefficient, FC)。与标准

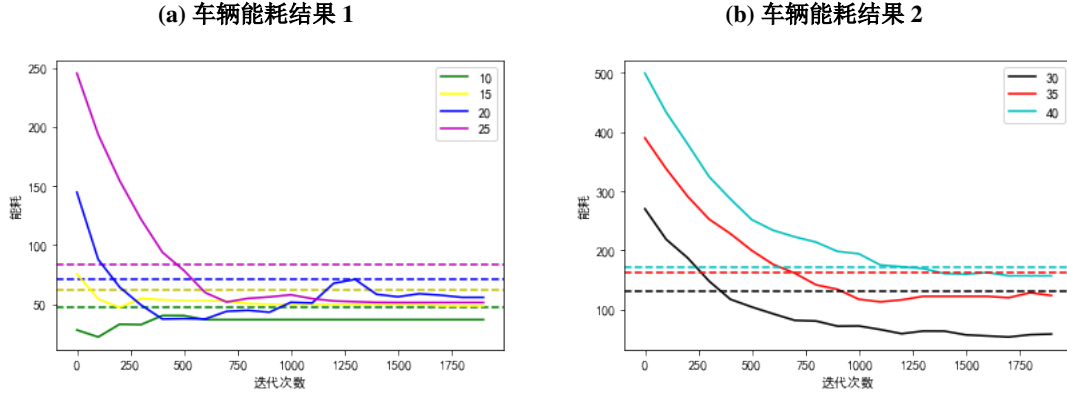


图 3-4 车辆能耗结果

差一样，数值越小，公平性越高。标准差定义如下：

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}} \quad (3.25)$$

FC 的定义如下：

$$FC = \max \sqrt{\frac{1}{N} \sum_{i=1}^N (E_{it}^{blnc} - \bar{E})^2}, t = 1, \dots, T \quad (3.26)$$

其中， $t$  是迭代的次数， $\bar{E}$  是所有车辆能耗的平均值。

### 3.3.3 实验结果

在图 3-4a 和 3-4b 中，虚线是任务没有负载出去而在本地执行所需要消耗的能量。本论文可以看到，当车辆数为 10 时，该算法迭代几十轮就可以得到很好的结果。当车辆数为 20 时，需要 350 次迭代才能得到一个较好的结果。而当车辆数量为 30 时，能耗在 750 次迭代之前变化较大，而在 750 次迭代之后变化就会明显变小。也就是说，能源消耗的减少率在 750 次迭代前较快，而在 750 次迭代后，能源消耗的减少率就会降低。当车辆数为 40 时，几乎没有节能效果。

可以得出结论，随着车辆数量的增加，迭代的次数也在增加。因此，在边缘服务器上运行该算法时，需要合理地调整迭代次数，因为车辆数量的变化会显著的影响算法的运行时间。

在图 3-5 中，经过 100 次迭代以后，所有数目运行时间和迭代次数呈线性关系。然而，当迭代次数固定时，算法的运行时间和车辆之间的关系不是线性的，而是非线性的，参与资源共享的车辆的数量越多，潜在的分配方案越多，算法的运行时间增加的越多。

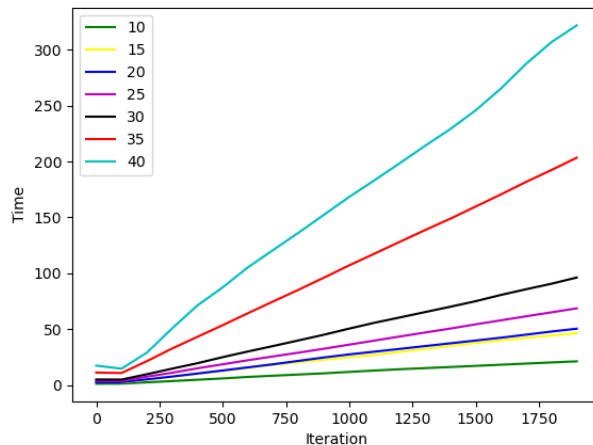


图 3-5 时间图

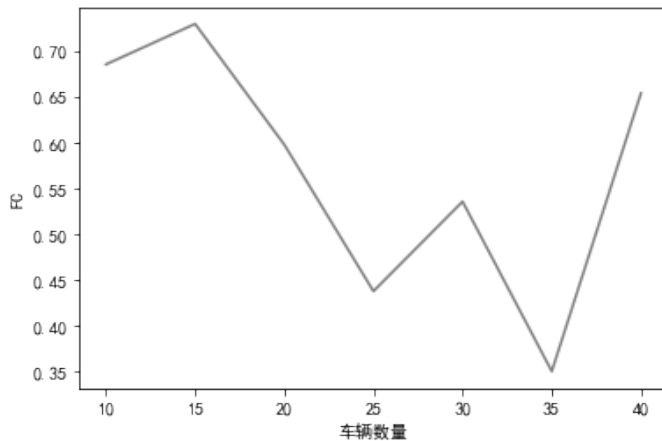


图 3-6 平衡因子 FC 值

如图 3-6 所示，随着车辆的增加，车辆能量消耗的平衡首先增加，然后减少。当车辆数目为 30 和 40 时，能耗的平衡系数反而增加。

如图 3-7 所示，当车辆数量在 10 到 25 辆之间时，本论文的系统可以节省 20% 以上的能量消耗，但是当车辆数量为 30 辆时，只有节省 17% 的能量消耗。但当车辆数为 30 时，只能节省 17%，而当车辆数为 40 时，几乎可以节省 20%。当车辆数量为 40 辆时，几乎没有节省能量。

综合考虑性能测量、算法运行时间以及平衡因子，选择 25 辆汽车参与资源共享是非常合适的。



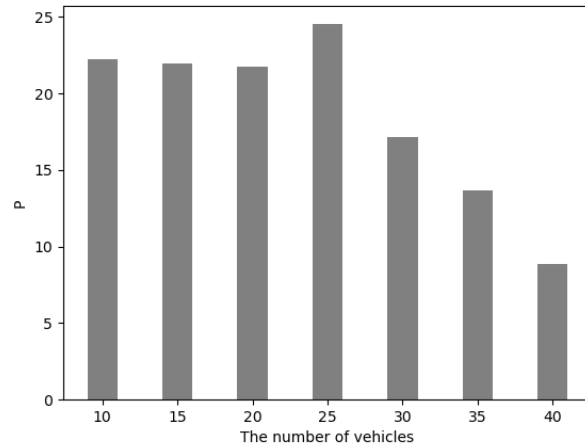


图 3-7 性能度量 P 值图

### 3.4 本章小结

本章所研究的是车联网领域的任务卸载和分配的方案，通过能耗优先的原则设计出一套卸载算法。首先，本章详细介绍了车联网中任务卸载的系统模型，并对其中的要素进行了说明，包括系统场景、车辆计算模型及车辆能耗模型。接着，作者将问题建模为数学规划形式，提出了一种基于粒子群算法的卸载与分配策略，然后设计了相关评价标准，并对实验结果进行了分析。评价标准表明，该算法能够很好地提高资源利用率，同时也兼顾了公平性。

然而，在研究过程中也发现了一些缺陷，比如，基因算法存在着编码和解码的过程，需要进行交叉和变异操作，这导致了计算量太大和计算时间过长等问题。因此，在下一步的研究中，本论文将引入新的智能算法，并将其改进为适用于离散问题的算法，以加快算法的运行时间。

## 第四章 基于贪心法调整的离散侏儒猫鼬算法的任务卸载算法

通过上一章的研究，本论文完成了对车联网中对任务卸载算法研究的任务，但也指出了解决方法存在计算量太大、计算时间过长等问题。本章首先说明了贪心法调整的策略，然后重新定义了运算符，重新定义的运算符有双目加法运算符、双目减法运算符、双目乘法运算符、单目调整运算符。通过重新定义运算符的方式，将侏儒猫鼬算法修改为解决离散问题的算法，解决本文提出的离散形式的问题。

### 4.1 贪心法调整策略

为了克服变量之间存在的相关性，基因算法有着解码和编码两个过程，来配合交叉操作，使得交叉操作更具有随机性，克服了变量之间的相关性。

本论文不使用基因算法，而是想将侏儒猫鼬算法改造为能够求解离散形式问题的原因是，基因算法的解码和编码两个过程十分浪费时间，所以，为了加快侏儒猫鼬算法的运行时间，本论文在此使用基于贪心法的调整策略。

本论文采用的贪心法是处理时间最短法（Shortest Processing Time, SPT），该算法是将作业调整到当前剩余处理时间最短的机器上。SPT 算法的执行步骤如下：首先，将所有待处理的任务按照所需处理时间从小到大排序。其次，依次将任务分配给选择当前剩余处理时间最短的机器。直到所有任务都被分配并完成。

下面是 SPT 的一个实例：假设有 6 个任务需要调度，它们的处理时间分别为 3, 4, 5, 4, 3 和 7，有两台机器可用。首先，按照所需处理时间从小到大对任务进行排序：time=[3, 3, 4, 4, 5, 7]。

接下来，依次将任务分配给可用的处理器，使得处理时间最短的任务尽量早地被处理。第一批任务的处理时间分别为 3、3，将这两个任务分配给两台处理器中处理时间较短的那台。此时，两台处理器的处理时间分别为 3 和 3。现在剩余的任务列表为 [4, 4, 5, 7]。

接着，在两台处理器中选择处理时间最短的处理器（即第一个处理器），并为其分配下一个处理时间最短的任务，即处理时间为 4 的任务。第一个处理器的处理时间增加到了 7。现在剩余的任务列表为 [4, 5, 7]。

然后，选择处理时间最短的处理器（即第二个处理器），并为其分配下一个处理时间最短的任务，即处理时间为 4 的任务。第二个处理器的处理时间增加到了 7。现在剩余的任务列表为 [5, 7]。

继续选择处理时间最短的处理器（即第一个处理器），并为其分配下一个处理

时间最短的任务，即处理时间为 5 的任务。第一个处理器的处理时间增加到了 12。现在剩余的任务列表为 [7]。

最后，选择处理时间最短的处理器（即第二个处理器），并为其分配最后一个任务，即处理时间为 7 的任务。第二个处理器的处理时间增加到了 14，此时所有任务都被分配并完成。

因此，使用 SPT 算法进行调度后，两台处理器的处理时间分别为 12 和 14

## 4.2 重定义运算符

### 4.2.1 双目加法运算符

双目加法运算符 ( $\oplus$ )。操作为：对于两个变量，进行交叉操作。

交叉操作是基因算法中一个解更新行为，是指在两个父代个体（解）之间交换染色体上的一部分基因信息（变量）并生成新的子代个体（解）。作为进化算法的重要步骤之一，用于产生更好的后代解，并增加群体的多样性。

交叉操作通常需要满足如下两个条件：首先，确定交叉方式，即交叉方式有多种形式，例如单点交叉、多点交叉、均匀交叉和线性交叉等。本文选择了双点交叉的方式，双点交叉可以在更大范围内交换基因片段，有助于生成更多多样化的后代个体，有利于种群的多样性维持和进化。其次，保证基因信息的完整性和正确性，即交叉操作需要保证新生成的解是有效的解，因此需要保证每个基因在子代中都不缺失或重复。

在双点交叉（Two-point Crossover）中，需要随机选择两个交叉点的位置，然后将两个父代个体的这两个交叉点之间的基因序列进行互换，从而产生新的子代个体。下面以一个长度为 6 的染色体为例，说明双点交叉的具体操作过程：假设有两个父代个体分别是：

$$S_1 = [0, 1, 0, 1, 0, 1]$$

$$S_2 = [1, 0, 1, 0, 1, 0]$$

本论文随机选择两个交叉点位置  $k_1=2$  和  $k_2=5$ ，然后将两个父代个体在这两个位置之间的基因信息进行互换，得到新的子代个体  $S_3$  和  $S_4$ ：

$$S_3 = [0, 0, 1, 0, 1, 1]$$

$$S_4 = [1, 1, 0, 1, 0, 0]$$

子代个体中来自父代个体的基因信息（变量）在交叉点之前或之后是完全一致的，但在交叉点之间则发生了互换。

### 4.2.2 双目减法运算符

双目减法运算符 ( $\ominus$ )。如图 4-1 所示, 解中每一个变量相应元素的值是否不同。如果是, 则获得其值, 如果不是, 则置为 0。也就是说, 获得两个解中相同的部分, 不同的部分置为 0。

1	2	2	1	1	1	2	2	1	1
1	2	2	2	2	2	1	2	2	1
1	2	2	0	0	0	1	2	2	1

图 4-1 减法运算符

### 4.2.3 双目乘法运算符

双目乘法运算符 ( $\otimes$ )。如图 4-2 所示, 每个变量的取值分别相乘, 第一个变量的取值是  $n$  重伯努利分布, 其值只有 0 和 1。

1	1	1	1	1	1	0	0	0	0	0
1	2	2	2	1	1	1	2	2	1	1
1	2	2	2	1	1	0	0	0	0	0

图 4-2 乘法运算符

### 4.2.4 单目调整运算符

单目调整运算符  $\odot X_1$ 。如图 4-3 所示, 对  $X_1$  中所有为 0 的元素进行 SPT 调整, 调整步骤如下:

一共两台机器, 编号分别为 1 和 2,  $time = [3, 4, 5, 4, 3, 7]$ ,  $X_1 = [0, 1, 2, 2, 0, 0]$ , 对值为 0 的元素进行 SPT 调整, 值为  $k(k \neq 0)$  的元素的意思: 该任务的调整到第  $k$  台机器上。调整步骤如下:

(1) 首先, 将两个数组打包, 得到

$[3, 4, 5, 4, 3, 7]$

$[0, 1, 2, 2, 0, 0]$

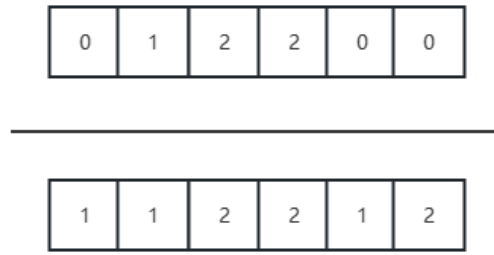


图 4-3 单目运算符

(2) 将任务按照所需处理时间从小到大进行排序，得到。

[3, 3, 4, 4, 5, 7]

[0, 0, 1, 2, 2, 0]

(3) 1 号机器的处理时间为 4，2 号机器的处理时间为 9，将所需处理时间最短的且对应分配机器值为 0 的任务，即任务处理时间为 3 的任务调度到 1 号机器上，1 号机器处理时间变成 7。

(4) 1 号机器的处理时间为 7，2 号机器的处理时间为 9，将所需处理时间最短的且对应分配机器值为 0 任务，即任务处理时间为 3 的任务调度到 1 号机器上，1 号机器处理时间变成 10。

(5) 1 号机器的处理时间为 10，2 号机器的处理时间为 9，将所需处理时间最短的且对应分配机器值为 0 任务，即任务处理时间为 7 的任务调度到 2 号机器上，1 号机器处理时间变成 16。

(6) 将分配的方案映射回到没有排序的数组中得到 [1,1,2,2,1,2]。即  $\odot X_1 = [1, 1, 2, 2, 1, 2]$ 。

### 4.3 离散侏儒猫鼬算法伪代码

图 4-4 是离散侏儒猫鼬算法的伪代码。

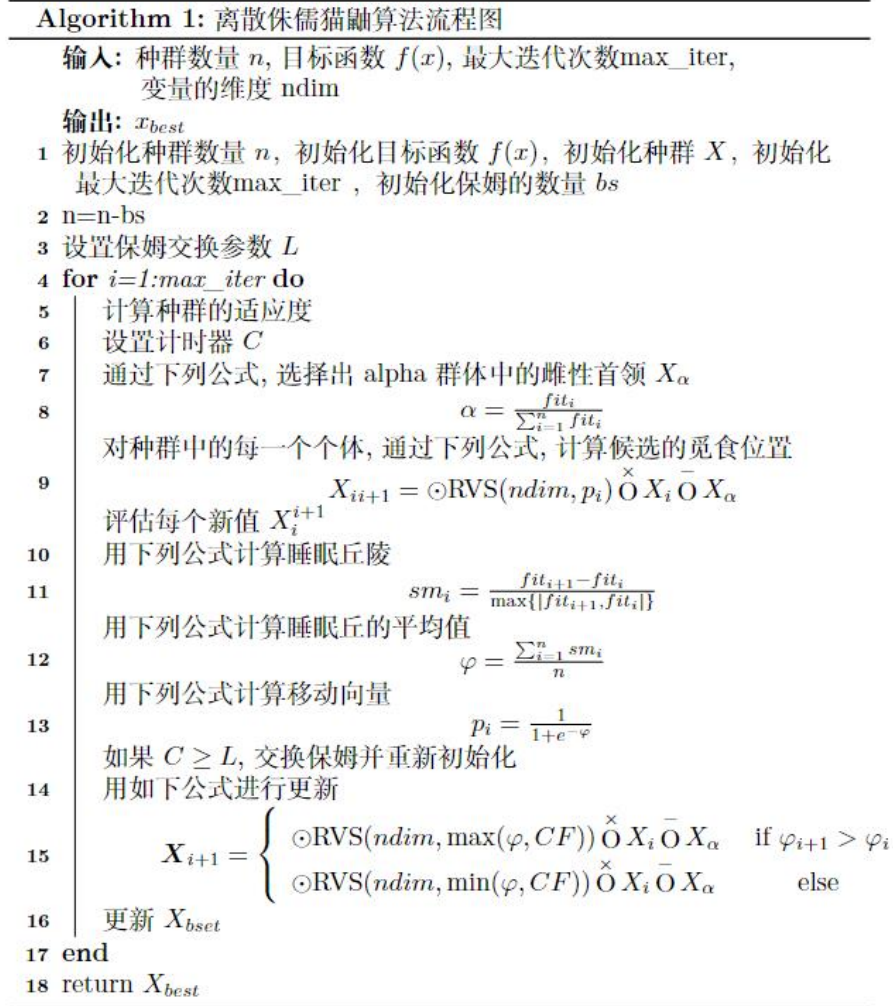


图 4-4 离散侏儒猫鼬算法伪代码

离散侏儒猫鼬算法流程如下:

- (1) 进行参数初始化。
- (2) 对于每一轮迭代进行更新:
  - (2-1) 计算适应度, 选出雌性首领。
  - (2-2) 通过更新公式, 进行更新。其中,  $RVS$  是维度为变量维度, 其值只有 0 或者 1 的向量。
  - (2-3) 计算睡眠丘和睡眠丘平均值。
  - (2-4) 计算出概率  $p_i$ 。
  - (2-5) 如果  $C \geq L$ , 交换保姆并重新初始化。

(2-2) 根据睡眠丘平均值在进行一轮更新。

(3) 输出最佳的种群个体。

## 4.4 仿真实验

### 4.4.1 消融实验

通过两个实验来验证各个变量之间的相关性，从而验证调整策略的重要性。实验一是验证解中某些变量的取值发生改变对目标函数的影响，实验二是验证两个解进行交叉行为对目标函数的影响。第二个实验实际是第一个实验的更进阶的版本，第一个实验交换的是同一个解中的某些变量，第二个实验交换的是不同解中的连续的变量。

这两个实验都说明了解变量之间的相关性，为了消除变量之间的相关性，本论文生成了解以后，还需要对解的变量进行调整。

实验一：验证解中某些变量的取值发生改变对目标函数的影响。

(1) 随机生成初始解，且保证解的合法性，计算目标函数。

(2) 对每个解的某些位置的变量，交换他们的取值。

(3) 计算完成操作后的解所对应的目标函数，生成图像。

实验二：两个解进行交叉行为对目标函数的影响。(1) 随机生成初始解，且保证解的合法性，计算目标函数。

(2) 解两两配对，进行交叉操作。

(3) 计算完成操作后的解所对应的目标函数，生成图像。

### 4.4.2 实验环境

如表4-1所示,本章实验环境如下:处理器为: Intel(R) Core(TM) i7-7700HQ; 显卡的型号为: NVIDIA GeForce RTX 1050; 机带 RAM 为 8GB; 操作系统为 Windows10 专业版; Python 版本为 3.8, Numpy 版本为 1.19.2, matplotlib 版本为 3.3.2。

本文参数设置如下, 车辆 CPU 的频率集为 700, 800, 900, ..., 1900, 只有 13 个频率级别。当车辆在 700MHz 处理不了任务时, 频率升为 800MHz 来处理任务。

和上一章实验相同, 车辆共享资源的时间为  $T = 10$  秒和  $\Delta T = 1$  秒。CPU 容量参数  $v_i, \theta_i$  其值分别为 7.683 和 -4558.52。M 的最大值和最小值分别为 819 和 10039。所有任务执行工作负载所需的计算资源  $r_{it}$  服从于  $U[820, 10000]$  的均匀分布, 并且假设车辆  $i$  可以卸载给其他车辆的任务计算资源值  $r'_{it}$  服从于  $U[200, r_{it}]$ , 因为某些子任务必须在本地执行。数据的大小从 1MB 到 10MB 不等。我们设定带

表 4-1 实验环境

名称	型号或版本	名称	型号或版本
处理器	Intel(R) Core(TM) i7-7700HQ	Python	3.8
显卡	NVIDIA GeForce RTX 1050	Numpy	1.19
RAM	8.0GB	Matplotlib	3.3
操作系统	Windows10 专业版		

宽  $d$  的值为 27Mb/s。参与资源共享的车辆数目分别是 10, 15, 20, 25。在算法中, 为了和基因算法进行比较, 种群数量被设置为 40。 $p$  值被设置为 0.4。

#### 4.4.3 评价标准

(1) 使用分配方案后的能耗, 包括基因算法求得的解所对应的能耗, 以及离散侏儒猫鼬算法求得的解所对应的能耗。

(2) 算法运行时间。算法的运行时间可以直接反映出算法的性能, 较短的运行时间通常代表着更高效的算法。通过比较不同算法的运行时间, 可以客观评估它们的性能优劣。对于需要实时响应的应用, 算法的运行时间直接影响用户体验。短时间内得到结果可以提高用户满意度和使用体验。在本实验中, 算法的运行时间时每一轮迭代 (100 次迭代) 使用的时间长短。

(3) 和第三章中的车辆平衡系数 (FC) 相同, FC 的定义如下:

$$FC = \max \sqrt{\frac{1}{N} \sum_{i=1}^N (E_{it}^{blnc} - \bar{E})^2}, t = 1, \dots, T \quad (4.1)$$

其中,  $t$  是迭代的次数,  $\bar{E}$  是所有车辆能耗的平均值。

(4) 种群的多样性: 种群多样性无疑是启发式算法性能的一个关键问题, 主要有两种方法能够衡量种群的多样性。

第一类方法是基于距离的度量方法 (distance-based measurement, DMB), 通过计算个体之间的欧氏距离来衡量种群间的多样性。DBMs 有几种不同的形式, 如个体间的平均空间距离、个体与种群中心位置之间的距离的平均值、个体间最大距离等<sup>[59][60]</sup>。但是相同的是, 距离越远代表着种群多样性越丰富, 反之, 距离越小则代表着的种群多样性越差。

第二种方法是基于频率的测量方法 (frequency-based measurement, FBM)。在 FBM 中, 通过统计具有代表性、有优势的个体的频率<sup>[61]</sup>或者平均基因频率<sup>[62]</sup>来评价种群多样性。此外, 熵度量是属于 FBM 的另一种方法, 因为熵 (公式 4.2) 使



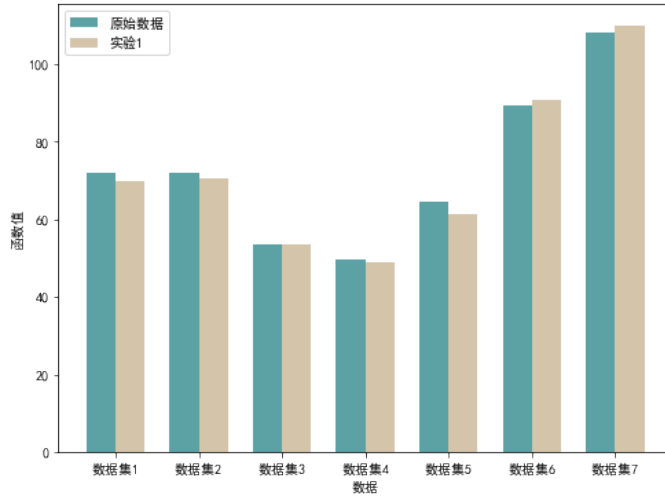


图 4-5 实验一实验结果

用了基因出现的概率/适应度。

$$S = - \sum_{i=1}^n p_i \log_2(p_i) \quad (4.2)$$

其中,  $(S)$  表示熵,  $(n)$  表示系统中可能的状态数,  $(p_i)$  表示第  $(i)$  个状态发生的概率。

在本文中, 本论文使用了一种度量方法, 即将种群中每个解的各个维度的方差之和的二分之一次方作为多样性  $D$  的衡量标准。下面是计算多样性  $D$  的公式:

$$D = \sqrt{\sum_{j=1}^m D_j^2}, \quad D_j = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \quad (4.3)$$

其中,  $m$  是解的维度,  $n$  是种群的数量,  $x_{ij}$  是第  $i$  个粒子的第  $j$  个维数。

#### 4.4.4 实验结果及分析

在图 4-5 中, 绿色柱状图表示未更新解所对应的函数值, 而黄色柱状图表示更新后所对应的变量值。从图中可以看出, 仅仅交换两个变量值并不会显著改变更新和未更新之间的区别。这表明我们不能仅仅通过交换两个变量的位置来进行更新。

在图 4-6 中, 绿色柱状图代表未经更新的函数值, 而黄色柱状图则表示经过交叉操作后所得到的变量值。尽管相较于实验一有所改善, 但这种交叉更新策略的效果并不十分显著。同时, 这几次实验之间的差异较大, 进一步表明了变量之间存在相关性。这说明我们不能仅仅依靠交换连续多个变量的位置来进行更新。

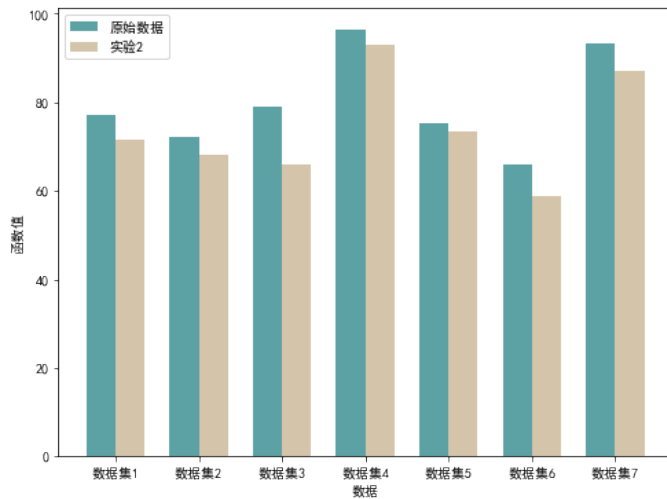


图 4-6 实验二实验结果

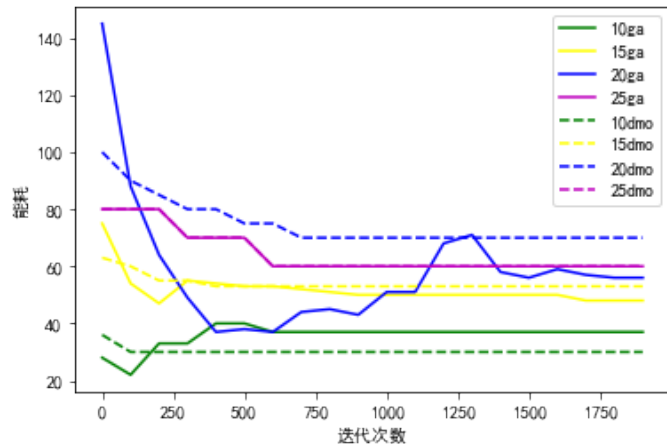


图 4-7 能耗图

图 4-7 中可以看出，虚线代表基因算法求解问题的能耗，而实线则是侏儒猫鼬算法求解的能耗。标签前面的数字表示车辆的数量，而同颜色的线代表着相同数量的车辆。

值得注意的是，尽管侏儒猫鼬算法在最终收敛时可能得到比基因算法差的解，但由于调整策略的优势，它只需进行少数迭代即可获得较好的结果，即其收敛性较好。也就是说，侏儒猫鼬算法具有较强的探索能力（广泛的搜索范围），但在寻找局部最优解方面不如基因算法。这一特点与侏儒猫鼬不会过度放牧的特性相对应。

通过观察图 4-8，可以发现基因算法种群差异度一直比较小，侏儒猫鼬算法种群差异度的变化幅度较大。根据种群差异度的定义，这表明，在算法运行过程中

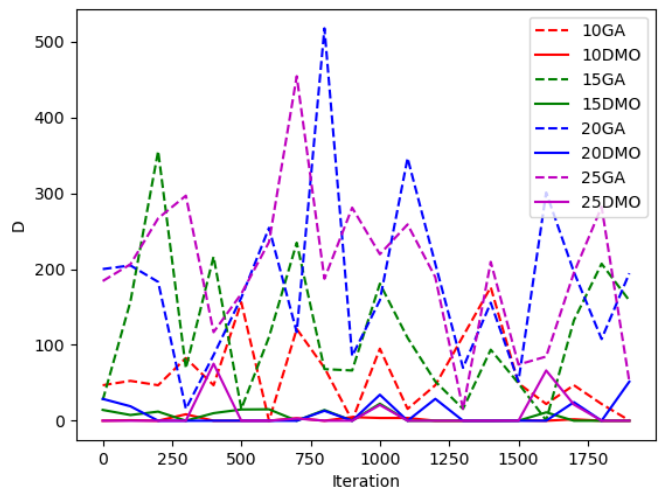


图 4-8 侏儒猫鼬算法种群差异度图

仍在产生新的解，并且新产生的解种群差异度较大。这样一来，侏儒猫鼬算法就能够避免陷入劣解，有可能产生更优质的新解。

这能和前面能耗图的结果相对应，种群差异度大表示，其探索的范围广，但是因为种群差异度一直很大，所以难以在已经寻找到的局部最优附近进行下一步的优化。

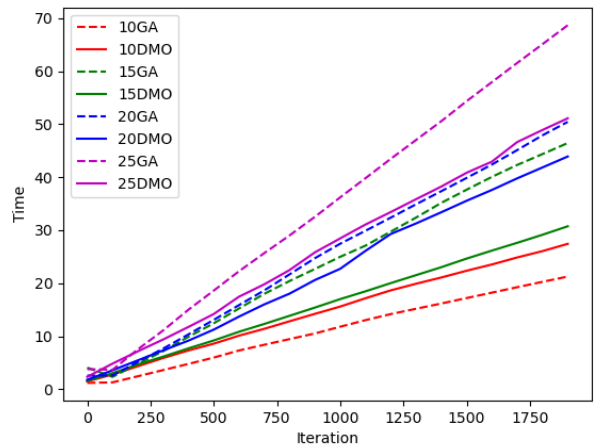


图 4-9 算法运行时间图

图 4-9 中可以看出，虚线代表基因算法求解问题的时间，而实线则是侏儒猫鼬算法的运行时间。标签前面的数字表示车辆的数量，而同颜色的线代表着相同数量的车辆。值得注意的是，侏儒猫鼬算法的运行时间明显比基因算法的运行时间要短。

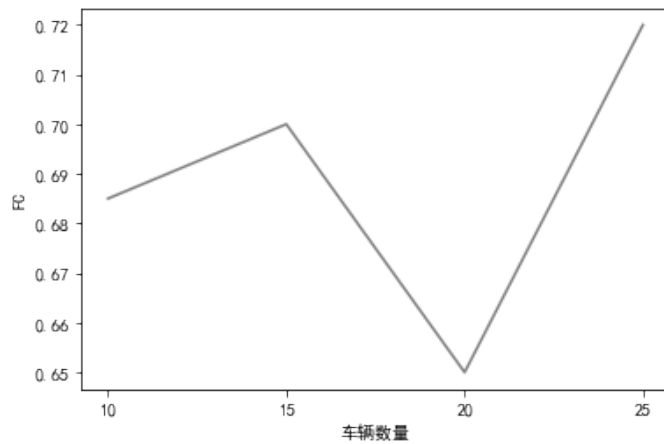


图 4-10 平衡因子 FC 值

当车辆数量为 15 和 20 时，侏儒猫鼬算法的运行时间只有原算法的三分之二；而当车辆数量为 10 和 25 时，则能将运行时间节省约四分之一。因此，从图中可以清晰地看出侏儒猫鼬算法在解决问题时具有更高的效率和优势。

如图 4-10 所示，随着车辆数量的增加，车辆能量消耗的平衡并没有明显的规律。这与我们使用基因算法求解结果的趋势不相同。然而，这些数值都在 0.6 至 0.7 之间浮动，因此，车辆任务卸载的分配方案相对较为公平。在平衡方面，侏儒猫鼬算法表现良好。

## 4.5 本章小结

本章在上一章的基础上探索了新的方法来解决第三章提出的问题。本论文解决了基因算法中编码解码过程耗时较长的问题，并通过重新定义运算符和增加贪心法调整的策略，对侏儒猫鼬算法进行了改进，将其转变为离散侏儒猫鼬算法。

实验结果表明，与基因算法相比，侏儒猫鼬算法在运行时间上有显著的缩短，但是在求解的分配方案不如基因算法求解的分配方案。

这表明，该算法也存在一些问题，例如，由于提高了搜寻能力导致搜索局部最优解附近的能力相对于基因算法较弱。



## 第五章 总结与展望

### 5.1 论文的主要工作及贡献

电动汽车越来越受到人们的追捧，其中很重要的一个原因就是可玩的智能系统。虽然电动汽车的电池续航里程越来越大，但是对比人们对于里程的渴望而言是远远不够的，如何在有限的容量中让汽车的行驶里程变得更大，仍然需要解决，而任务卸载是其中一个有希望能够解决该方法的方法，将任务卸载到边缘端或者是其他的汽车上，通过合理的调度来降低所有车辆的能耗和。这就需要我们设计合理的算法来进行调度。但是该问题的模型复杂，且并不能通过凸优化的方法来解决，因此需要其他的方法来解决。智能算法是作为求解复杂问题的佼佼者，我们通过智能算法来求解该问题。本文采用基因算法以及基于贪心法调整的侏儒猫鼬算法来解决本文的复杂问题。本文的总结如下：

（1）研究了车联网中任务卸载的问题模型，将能耗模型分为两部分来描述，将其描述为数学规划的形式。

（2）本文使用了基因算法来求解该问题，但是从本质上来讲，并没有对基因算法进行创新。

（3）本文在侏儒猫鼬算法的基础上提出了基于贪心法调整的侏儒猫鼬算法，核心理念是通过贪心法（处理时间最短法）以及重定义运算符来实现对种群进行更新。

### 5.2 未来的工作

针对车联网中能耗优先的任务调度问题，本文考虑的目标是最小化能耗的平方之和，这能有效兼顾分配方案的公平性。但是该场景下依旧有许多问题需要解决，未来仍需要进行探索。

（1）本文第三章使用了基因算法来解决该问题，但是基因算法存在编码解码过程导致计算速率偏慢。因此，下一步的研究可以考虑加入一些其他的近似算法来加快这个过程。

（2）本文第四章使用了离散侏儒猫鼬算法来解决该建模问题，虽然离散侏儒猫鼬算法能够比基因算法更快速的解决问题，但是其算法的不确定性也比基因算法更大，没有一个比较稳定的迭代次数。因此，下一步的研究可以考虑后期动态的不确定性，迭代前期加大不确定性，迭代后期减小不确定性，进一步提高问题的解决效率。

（3）劣解并不是没有用的，可以考虑吸收模拟退火算法的思想，有概率保留劣解，从而提高搜索能力。

## 参考文献

- [1] 余文科, 程媛, 李芳, 等. 物联网技术发展分析与建议[J]. 物联网学报, 2020, 4(4): 105-109.
- [2] CANO Z P, BANHAM D, YE S, et al. Batteries and fuel cells for emerging electric vehicle markets[J]. Nature Energy, 2018, 3(4): 279-289.
- [3] ROMIJN T C J, DONKERS M C F, KESSELS J T B A, et al. A distributed optimization approach for complete vehicle energy management[J/OL]. IEEE Transactions on Control Systems Technology, 2019, 27(3): 964-980. DOI: [10.1109/TCST.2018.2789464](https://doi.org/10.1109/TCST.2018.2789464).
- [4] BARTA D, MRUZEK M, LABUDA R, et al. Possibility of increasing vehicle energy balance using coasting[J]. Advances in Science and Technology. Research Journal, 2018, 12(1).
- [5] 施巍松, 张星洲, 王一帆, 等. 边缘计算: 现状与展望[J]. 计算机研究与发展, 2019, 56(1): 69-89.
- [6] SHI W, CAO J, ZHANG Q, et al. Edge computing: Vision and challenges[J]. IEEE internet of things journal, 2016, 3(5): 637-646.
- [7] KHAN W Z, AHMED E, HAKAK S, et al. Edge computing: A survey[J]. Future Generation Computer Systems, 2019, 97: 219-235.
- [8] CAO K, LIU Y, MENG G, et al. An overview on edge computing research[J]. IEEE access, 2020, 8: 85714-85728.
- [9] HE Y, SI P B, SUN E C, et al. Joint task management in connected vehicle networks by software-defined networking, computing and caching[C]//2017 International Conference on Network and Information Systems for Computers (ICNISC). IEEE, 2017: 47-50.
- [10] GARCIA M H C, MOLINA-GALAN A, BOBAN M, et al. A tutorial on 5g nr v2x communications[J]. IEEE Communications Surveys & Tutorials, 2021, 23(3): 1972-2026.
- [11] 李智勇, 王琦, 陈一凡, 等. 车辆边缘计算环境下任务卸载研究综述[J]. 计算机学报, 2021, 44(5): 963-982.
- [12] YU R, ZHANG Y, GJESSING S, et al. Toward cloud-based vehicular networks with efficient resource management[J]. Ieee Network, 2013, 27(5): 48-55.
- [13] LIU C, LIU K, GUO S, et al. Adaptive offloading for time-critical tasks in heterogeneous internet of vehicles[J/OL]. IEEE Internet of Things Journal, 2020, 7(9): 7999-8011. DOI: [10.1109/IIOT.2020.2997720](https://doi.org/10.1109/IIOT.2020.2997720).
- [14] WANG B, WANG C, HUANG W, et al. A survey and taxonomy on task offloading for edge-cloud computing[J]. IEEE Access, 2020, 8: 186080-186101.
- [15] ZHENG K, MENG H, CHATZIMISIOS P, et al. An smdp-based resource allocation in vehicular cloud computing systems[J]. IEEE Transactions on Industrial Electronics, 2015, 62(12): 7920-7928.



- [16] SUN Y, GUO X, SONG J, et al. Adaptive learning-based task offloading for vehicular edge computing systems[J]. IEEE Transactions on vehicular technology, 2019, 68(4): 3061-3074.
- [17] SUN Y, GUO X, ZHOU S, et al. Learning-based task offloading for vehicular cloud computing systems[C]//2018 IEEE International Conference on Communications (ICC). IEEE, 2018: 1-7.
- [18] KLAIMI J, SENOUCI S M, MESSOUS M A. Theoretical game approach for mobile users resource management in a vehicular fog computing environment[C]//2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC). IEEE, 2018: 452-457.
- [19] LIN C C, DENG D J, YAO C C. Resource allocation in vehicular cloud computing systems with heterogeneous vehicles and roadside units[J]. IEEE Internet of Things Journal, 2017, 5(5): 3692-3700.
- [20] LIU X Y, DING Z, BORST S, et al. Deep reinforcement learning for intelligent transportation systems[A]. 2018.
- [21] LIU N, LI Z, XU J, et al. A hierarchical framework of cloud resource allocation and power management using deep reinforcement learning[C]//2017 IEEE 37th international conference on distributed computing systems (ICDCS). IEEE, 2017: 372-382.
- [22] ZHANG Y, YAO J, GUAN H. Intelligent cloud resource management with deep reinforcement learning[J]. IEEE Cloud Computing, 2017, 4(6): 60-69.
- [23] CHENG M, LI J, NAZARIAN S. Drl-cloud: Deep reinforcement learning-based resource provisioning and task scheduling for cloud service providers[C/OL]//2018 23rd Asia and South Pacific Design Automation Conference (ASP-DAC). 2018: 129-134. DOI: [10.1109/ASPDAC.2018.8297294](https://doi.org/10.1109/ASPDAC.2018.8297294).
- [24] NING Z, DONG P, WANG X, et al. Deep reinforcement learning for vehicular edge computing: An intelligent offloading system[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2019, 10(6): 1-24.
- [25] NING Z, DONG P, WANG X, et al. Deep reinforcement learning for intelligent internet of vehicles: An energy-efficient computational offloading scheme[J]. IEEE Transactions on Cognitive Communications and Networking, 2019, 5(4): 1060-1072.
- [26] XU X, LIU K, DAI P, et al. Joint task offloading and resource optimization in noma-based vehicular edge computing: A game-theoretic drl approach[J]. Journal of Systems Architecture, 2023, 134: 102780.
- [27] HE Q, CUI G, ZHANG X, et al. A game-theoretical approach for user allocation in edge computing environment[J]. IEEE Transactions on Parallel and Distributed Systems, 2019, 31(3): 515-529.
- [28] CAO H, CAI J. Distributed multiuser computation offloading for cloudlet-based mobile cloud computing: A game-theoretic machine learning approach[J]. IEEE Transactions on Vehicular Technology, 2017, 67(1): 752-764.

- [29] GAD A G. Particle swarm optimization algorithm and its applications: a systematic review[J]. Archives of computational methods in engineering, 2022, 29(5): 2531-2561.
- [30] 谢云. 模拟退火算法综述[J]. 计算机应用研究, 1998, 15(5): 7-9.
- [31] BELOGAEV A, ELOKHIN A, KRASILOV A, et al. Cost-effective v2x task offloading in mec-assisted intelligent transportation systems[J]. IEEE access, 2020, 8: 169010-169023.
- [32] CHEN M H, DONG M, LIANG B. Resource sharing of a computing access point for multi-user mobile cloud offloading with delay constraints[J]. IEEE Transactions on Mobile Computing, 2018, 17(12): 2868-2881.
- [33] HUYNH L N, PHAM Q V, PHAM X Q, et al. Efficient computation offloading in multi-tier multi-access edge computing systems: A particle swarm optimization approach[J]. Applied Sciences, 2019, 10(1): 203.
- [34] BUI K H N, JUNG J J. Aco-based dynamic decision making for connected vehicles in iot system [J]. IEEE Transactions on Industrial Informatics, 2019, 15(10): 5648-5655.
- [35] NESTEROV Y. A method for unconstrained convex minimization problem with the rate of convergence  $O(1/k^2)$ [C]//Dokl. Akad. Nauk. SSSR: Vol. 269. 1983: 543.
- [36] LITTLE J D, MURTY K G, SWEENEY D W, et al. An algorithm for the traveling salesman problem[J]. Operations research, 1963, 11(6): 972-989.
- [37] DANTZIG G B. Maximization of a linear function of variables subject to linear inequalities[J]. Activity analysis of production and allocation, 1951, 13: 339-347.
- [38] GOMORY R E. On the relation between integer and noninteger solutions to linear programs[J]. Proceedings of the National Academy of Sciences, 1965, 53(2): 260-265.
- [39] SUN Y, XUE B, ZHANG M, et al. Automatically designing cnn architectures using the genetic algorithm for image classification[J]. IEEE transactions on cybernetics, 2020, 50(9): 3840-3854.
- [40] ZHANG X, LIU X, LI W, et al. Trade-off between energy consumption and makespan in the mapreduce resource allocation problem[C]//International Conference on Artificial Intelligence and Security. Springer, 2019: 239-250.
- [41] YOON Y, KIM Y H. An efficient genetic algorithm for maximum coverage deployment in wireless sensor networks[J/OL]. IEEE Transactions on Cybernetics, 2013, 43(5): 1473-1483. DOI: [10.1109/TCYB.2013.2250955](https://doi.org/10.1109/TCYB.2013.2250955).
- [42] ROBERGE V, TARBOUCHI M, LABONTÉ G. Comparison of parallel genetic algorithm and particle swarm optimization for real-time uav path planning[J]. IEEE Transactions on industrial informatics, 2012, 9(1): 132-141.
- [43] AGUSHAKA J O, EZUGWU A E, ABUALIGAH L. Dwarf mongoose optimization algorithm [J]. Computer methods in applied mechanics and engineering, 2022, 391: 114570.
- [44] RASA O A E. The dwarf mongoose: a study of behavior and social structure in relation to ecology in a small, social carnivore[M]//Advances in the Study of Behavior: Vol. 17. Elsevier,

- 1987: 121-163.
- [45] MOUSTAFA G, SMAILI I H, ALMALAWI D R, et al. Dwarf mongoose optimizer for optimal modeling of solar pv systems and parameter extraction[J]. Electronics, 2023, 12(24): 4990.
  - [46] AKINOLA O A, AGUSHAKA J O, EZUGWU A E. Binary dwarf mongoose optimizer for solving high-dimensional feature selection problems[J]. Plos one, 2022, 17(10): e0274850.
  - [47] BALASUBRAMANIAM S, SATHEESH KUMAR K, KAVITHA V, et al. Feature selection and dwarf mongoose optimization enabled deep learning for heart disease detection[J]. Computational intelligence and neuroscience, 2022, 2022.
  - [48] ALRAYES F S, ALZHRANI J S, ALISSA K A, et al. Dwarf mongoose optimization-based secure clustering with routing technique in internet of drones[J/OL]. Drones, 2022, 6(9). <https://www.mdpi.com/2504-446X/6/9/247>. DOI: 10.3390/drones6090247.
  - [49] RASA O A E. The ethology and sociology of the dwarf mongoose (*helogale undulata rufula*)[J]. Zeitschrift für Tierpsychologie, 1977, 43(4): 337-406.
  - [50] RASA O. Ecological factors and their relationship to group size, mortality and behaviour in the dwarf mongoose *helogale undulata* (peters, 1852)[J]. Cimbebasia, 1986, 8(3): 15-21.
  - [51] GE R, FENG X, CAMERON K. Performance-constrained distributed dvs scheduling for scientific applications on power-aware clusters[C/OL]//SC '05: Proceedings of the 2005 ACM/IEEE Conference on Supercomputing. 2005: 34-34. DOI: 10.1109/SC.2005.57.
  - [52] ELNOZAHY E N M, KISTLER M, RAJAMONY R. Energy-efficient server clusters[C]//FALSAFI B, VIJAYKUMAR T N. Power-Aware Computer Systems. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003: 179-197.
  - [53] LI J, MARTINEZ J. Dynamic power-performance adaptation of parallel computation on chip multiprocessors[C/OL]//The Twelfth International Symposium on High-Performance Computer Architecture, 2006. 2006: 77-87. DOI: 10.1109/HPCA.2006.1598114.
  - [54] PIGUET C, SCHUSTER C, NAGEL J L. Optimizing architecture activity and logic depth for static and dynamic power reduction[C/OL]//The 2nd Annual IEEE Northeast Workshop on Circuits and Systems, 2004. NEWCAS 2004. 2004: 41-44. DOI: 10.1109/NEWCAS.2004.1359011.
  - [55] LIU X, LI W, XIE R. A primal-dual approximation algorithm for the k-prize-collecting minimum power cover problem[J]. Optimization Letters, 2021: 1-13.
  - [56] LIU X, LI W, DAI H. Approximation algorithms for the minimum power cover problem with submodular/linear penalties[J]. Theoretical Computer Science, 2022.
  - [57] DAI H, DENG B, LI W, et al. A note on the minimum power partial cover problem on the plane [J]. Journal of Combinatorial Optimization, 2022, 44(2): 970-978.
  - [58] BAHREINI T, BROCANELLI M, GROSU D. Vecman: A framework for energy-aware resource management in vehicular edge computing systems[J/OL]. IEEE Transactions on Mobile Computing, 2021: 1-1. DOI: 10.1109/TMC.2021.3089338.

- [59] URSEMR K. Diversity-guided evolutionary algorithms[C]//International Conference on Parallel Problem Solving from Nature. Springer, 2002: 462-471.
- [60] MORRISON R W, DE JONG K A. Measurement of population diversity[C]//International conference on artificial evolution (evolution artificielle). Springer, 2001: 31-41.
- [61] GOUVEA M M, ARAUJO A F. Diversity control based on population heterozygosity dynamics [C]//2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence). IEEE, 2008: 3671-3678.
- [62] COLLINS R J, JEFFERSON D R. Selection in massively parallel genetic algorithms[M]. University of California (Los Angeles). Computer Science Department, 1991.