

基于负载预测的多目标优化任务卸载策略

彭世明, 林士飏, 贾硕, 杨苗会

(山东理工大学交通与车辆工程学院, 山东 淄博 255000)

摘要: 在车联网中随着智能网联汽车的升级, 新兴车载应用对计算能力提出更高要求, 车载单元本身的计算能力已远远不够, 移动边缘计算(MEC)的出现可以为车辆提供更加可靠的服务。为解决车联网边缘计算中的任务卸载问题, 提出基于负载预测的多目标优化卸载策略算法, 降低任务时延并实现边缘服务器负载均衡。通过基于自适应优化神经网络的负载预测算法预测 MEC 服务器的负载, 提前感知 MEC 服务器的负载变化, 解决任务卸载滞后问题。以最小化时延及 MEC 服务器负载均衡为目标, 综合考虑通信环境、计算资源、任务量等因素构建多目标优化模型。通过非支配排序遗传算法-III (NSGA-III) 获得最优任务卸载策略。仿真实验结果表明, 该算法能对 MEC 服务器的负载进行较精确的预测。相比 MTUOA、NSGA2、QTD 和 AOS, 该算法的任务时延分别降低 1.7%、7.3%、12.4%、17.5%, 并在 MEC 服务器负载均衡中取得显著优势, 解决 MEC 服务器负载不均衡的问题。此外, 该算法可以根据不同通信小区的通信环境及车辆数等制定差异化的任务卸载方案。

关键词: 车联网; 移动边缘计算; 任务卸载; 负载预测; 负载均衡; 时延

中图分类号: TP391

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0066766

Multi-Objective Optimization Task Offloading Strategy Based on Load Prediction

PENG Shiming, LIN Shiyang, JIA Shuo, YANG Miaohui

(School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo 255000, Shandong, China)

[Abstract] With the upgrading of intelligent connected vehicles in the Internet of Vehicles (IoV), new on-board applications introduce higher requirements for computing power. The computing power of on-board units is far from sufficient. The emergence of Mobile Edge Computing (MEC) can provide more reliable services for vehicles. Aiming at the task offloading problem in edge computing for the IoV, a multi-objective optimization offloading strategy algorithm based on load prediction is proposed to reduce the task delay and realize load balance among edge servers. By using a load prediction algorithm based on adaptive optimization neural networks to predict the load of MEC servers, the load changes of MEC servers are sensed in advance, solving the problem of task offloading lag. A multi-objective optimization model is constructed with the goal of minimizing latency and balancing MEC server load, considering factors such as communication environment, computing resources, and task volume. The optimal task offloading strategy is obtained through the Non-dominated Sorting Genetic Algorithm (NSGA) - III (NSGA-III). The simulation results show that this algorithm can accurately predict the load of MEC servers. Compared with the MTUOA, NSGA2, QTD, and AOS algorithms, NSGA-III reduced task latency by 1.7%, 7.3%, 12.4%, and 17.5%, respectively, and achieved significant advantages in MEC server load balancing, solving the problem of MEC server load imbalance. In addition, the proposed algorithm can develop differentiated task offloading plans based on factors such as the communication environment and number of vehicles in different communication cells.

[Key words] Internet of Vehicles (IoV); Mobile Edge Computing (MEC); task offloading; load prediction; load balancing; time delay

0 引言

车联网(IoV)作为物联网技术的一种具体应用, 集成了人工智能、5G以及大数据等前沿技术, 是未来智能交通系统的重要组成部分^[1-2]。车辆借助其装备的车载电脑及通信技术, 获得更准确、全面的交通信息, 实现智能导航、无人驾驶、安全预警等功能^[3-4]。随着智能交通系统的升级, 车辆对网络质

量、时延等要求越来越高。车辆装备的车载电脑难以高效且及时地处理任务, 因此解决车辆计算能力受限问题已刻不容缓^[5]。移动边缘计算(MEC)的出现为汽车行业打破现有技术瓶颈^[6]。MEC服务器通常部署在距离车辆较近的位置, 将车辆的计算任务卸载到边缘网络^[7-8], 从而拓展车辆的计算能力。因此, 在车联网中基于MEC设计有效的计算卸载方案, 实现资源的优化分配是值得研究的问题。

收稿日期: 2023-01-16 修回日期: 2023-04-04

基金项目: 产学研合作协同育人项目(202102473012)。

通信作者 E-mail: shiyang@sdu.edu.cn

低时延对于智能交通系统而言至关重要,直接决定了无人驾驶、安全预警等功能能否实现应用。边缘计算任务卸载策略以时延作为性能评估的重要指标。

LIU等^[9]以最小化时延为目标,提出一种基于马尔可夫链决策的任务卸载策略,分析卸载策略的时延及能耗,寻找最优计算卸载策略。SONG等^[10]提出一种定期分发任务的有效方法,该分配策略在满足任务时延及服务质量要求的基础上增加了边缘计算网络的任务处理数量。与基准方案相比,该方法在时延、边缘网络承载的任务数量及负载均衡方面都有更优的表现。REN等^[11]在MEC系统中提出最优联合通信与计算资源分配算法,首先推导本地与MEC服务器的最小系统延迟与最优资源分配模型,然后提出一种针对部分卸载的分段凸优化方法,该方法与先前的研究相结合,使用次梯度优化方法解决执行延迟问题。ZHANG等^[12]在基于分层云的车辆边缘计算卸载框架内对MEC服务器计算资源受限问题进行研究,采用Stackelberg博弈论方法制定最优多级卸载方案,该方案可以有效地降低网络延迟,并且在满足任务延迟限制的同时,最大化系统的收益。SALEEM等^[13]提出一种支持部分计算卸载的MEC框架,考虑所需的能量消耗、部分卸载和资源分配约束,将总时延最小化制定为混合整数非线性规划问题,提出一种联合部分卸载和资源分配方案,该方案在降低时延的同时限制了用户设备的本地能耗。

上述研究以时延为重要性能评估指标制定任务卸载策略,但是任务分配不均衡可能会使某些服务器负载过高,从而造成服务器资源浪费和服务质量下降^[14]。ZHANG等^[15]提出一种基于软件定义网络(SDN)的任务卸载框架,基于该框架设计负载均衡任务卸载方案,得到任务时延最小化的任务卸载策略。YAO等^[16]将任务卸载问题转化为整数线性规划问题,通过改进的遗传算法(GA)搜索最优解,在负载均衡的条件下降低任务时延。DAI等^[17]在车辆边缘计算网络中将负载均衡与卸载相结合,联合优化选择决策、卸载率和计算资源提出一种低复杂度算法,该算法在优化时延与负载均衡的同时具有快速收敛性。LUO等^[18]在基于边缘计算的5G车载自组织网络中,将内容预取至路侧设施和车辆单元,采用基于图论的方法解决资源分配问题,最终通过一种平衡贪婪算法实现负载均衡。林峰等^[19]在蜂窝车联网与MEC融合应用场景中提出动态负载均衡算法,该方法充分考虑边缘服务器自身运行状态,动态调整指标权重并合理分配计算任务,从而提升集群负载均衡度。余翔等^[20]在MEC车联网架构内,考虑MEC服务器负载不均以及多优先级车辆任务的情况,提出基于遗传算法的任务卸载策略,对比分析发现该策略在提高任务完成率的同时可实现负载均衡。

上述研究综合考虑了任务时延及服务器的负载

均衡,制定任务卸载策略。然而,由于车联网中车辆单元在短时间内聚集和消散的特点,大量的车辆任务卸载使得MEC服务器的负载快速变化并持续波动,因此仅依据当前的计算负载制定任务卸载策略,无法适应未来一段时间的负载变化,继而导致任务时延增大以及卸载滞后。因此,建立可靠的方法预测MEC服务器的负载尤为重要。

本文提出一种基于负载预测的多目标优化卸载策略算法MTUOA-LP。在车联网场景中通过基于自适应优化神经网络的负载预测算法预测MEC服务器的计算资源量,为提高计算资源预测的准确率,引入自适应遗传算法优化长短期记忆(LSTM)网络的超参数,预测MEC服务器的计算资源。设计基于负载预测的多目标优化卸载策略算法,以最小化时延及MEC服务器负载均衡为目标,综合考虑通信环境、计算资源及任务量等因素构建多目标优化模型,采用非支配排序遗传算法-III(NSGA-III)求解多目标优化模型,获得理想的任务卸载策略。

1 系统模型与优化问题制定

1.1 场景模型

MEC网络系统架构如图1所示,其分为集中控制层、分布式MEC层和车辆层3层。分布式MEC层中的路侧单元(RSU)部署于路侧,且均配有单一MEC服务器,形成独立的通信小区。在MEC服务器之间通过光纤构成通信链路,可实现资源共享。车辆层中的车辆单元与MEC服务器通过RSU建立信号连接,通过本地RSU将任务上传至分布式MEC层。MEC服务器周期性采集车辆节点广播信息并获取周边MEC服务器信息,为车辆单元提供边缘计算服务。集中控制层配备中心控制器,中心控制器与MEC服务器间通过光纤构成通信链路,可获取网络系统的状态信息,制定任务卸载策略。

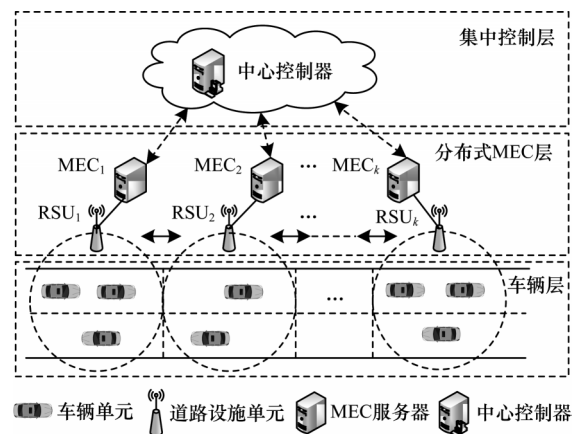


图1 MEC网络系统架构

Fig.1 Architecture of MEC network system

本文将车辆单元集合表示 $V = \{v_1, v_2, \dots, v_h\}$, 每个车辆单元都有1个或多个任务,其中所有任务都可以被卸载但不能进一步划分,所有任务都相互独

立。任务集合用 $A = \{A_{i,j} | i \in H, j \in N\}$ 表示, 其中 $H = \{1, 2, \dots, h\}$, $N = \{1, 2, \dots, n\}$, $A_{i,j}$ 表示第 i 个车辆单元的第 j 个任务。任务属性表示 $A_{i,j} = \{\beta, b, T_{\max}\}$, 其中, β 表示每 bit 计算需要的 CPU 周期数, b 表示任务数据大小, T_{\max} 表示任务最大可容忍时延。MEC 服务器集合表示 $M_{\text{MEC}} = \{M_k | k \in G\}$, 其中, $G = \{1, 2, \dots, m\}$ 。不同通信小区的车辆单元信息表示 $B_k = \{v_i | i \in H\}$, 其中, $\bigcap_{k=1}^m B_k = \emptyset, \bigcup_{k=1}^m B_k = V$ 。

在 MEC 网络系统中, 车辆单元已使用的计算资源集合表示 $Q = \{f_{\text{loc},i} | i \in H\}$, 其中, $f_{\text{loc},i} \in [0, 100]$ 。MEC 服务器已使用的计算资源集合表示 $P = \{f_{\text{mec},k} | k \in G\}$, $f_{\text{mec},k} \in [0, 100]$ 。

1.2 通信模型

在 MEC 网络系统中, 车辆单元的卸载任务需要通过无线信道从车载端传输至 RSU, 再上传至本地 MEC 服务器。RSU 通常位于 MEC 服务器附近且两者之间通过光纤建立通信连接, 因此忽略了 RSU 到 MEC 服务器的通信时延。假设车辆到 RSU 上行链路信道是频率平坦型块衰落的瑞利信道^[21,22], 那么车辆 v_i 与 B_k 之间的上行传输速率表达式如下:

$$R_{i,k} = \frac{W}{m} \log \left(1 + \frac{P_i d_i^{-\gamma} h^2}{N_0} \right) \quad (1)$$

其中: $d_i^{-\gamma}$ 表示车辆 v_i 与 B_k 之间的路径损耗; γ 表示路径损耗指数; d 表示车辆到 R_k 之间的距离; W 表示信道带宽; P_i 表示车辆 v_i 的发射功率; h 为常数, 表示上行信道衰落系数; N_0 表示白高斯噪声功率; m 表示通信小区 k 内的车辆单元数量。

由于本文考虑的任务类型为计算密集型任务, 输出相较于输入非常小, 因此不考虑从 MEC 服务器返回车辆单元的下行传输时延^[23]。

1.3 时延模型

在 MEC 网络系统中, 车辆单元任务可在车载端计算或卸载至 MEC 服务器计算。本文任务时延模型主要考虑 2 部分, 分别是任务处理时延(车载端处理时延和 MEC 服务器处理时延)和上行传输时延。

1) 车载端处理时延

当任务 $A_{i,j}$ 在车载端处理时, 为 $A_{i,j}$ 分配计算资源 $f_{\text{loc},i}$ 。CPU 频率表示计算资源, $f_{\text{loc},i}$ 越小则车载端处理能力越弱, 任务需要卸载到 MEC 服务器进行处理的程度越大。

车辆采用先到先服务的规则处理任务, 当本地端处于满载状态时, 后到的任务进入等待队列。任务 $A_{i,j}$ 的本地端处理时延 T_{loc} 表达式如下:

$$T_{\text{loc}} = \frac{\beta \times b}{f_{\text{loc},i}} + T_w \quad (2)$$

其中: T_w 表示等待时间。

2) 上行传输时延

当任务要上传至 MEC 服务器处理时, 车辆单元将任务上传至本地 MEC 服务器, 若车辆单元与目标

MEC 服务器未建立直接通信连接, 可通过 MEC 之间的光纤连接由本地 MEC 服务器实现任务上传。由于车辆任务在 MEC 之间的传输时延非常小, 因此本文不考虑 MEC 之间的传输时延。

根据任务大小及上行传输速率 $R_{i,k}$ 计算车辆单元 i 的任务上行传输时延, 表达式如下:

$$T_{\text{up}} = \frac{\sum_{j=1}^n b_j \times x_{i,j}}{R_{i,k}} \quad (3)$$

其中: $x_{i,j} = \{0, 1\}$, $x_{i,j}$ 取值为 0, 表示任务在车载端计算, $x_{i,j}$ 取值为 1, 表示任务卸载至 MEC 服务器计算。

3) MEC 服务器处理时延

当任务 $A_{i,j}$ 卸载至 MEC 服务器 k 处理时, 根据 M_k 分配的计算资源计算处理时延。

通过本文设计的负载预测算法 $A_{i,j}$ 预测 M_k 的计算资源使用情况 $f_{\text{pre},k}$, MEC 服务器给每个任务分配的计算资源 $f_{\text{mec},\text{alloc}}$ 的表达式如下:

$$f_{\text{mec},\text{alloc}} = \frac{\sum_{k=1}^m (f_{\text{max}} - f_{\text{pre},k})}{h \times n} \quad (4)$$

其中: f_{max} 表示设置的 MEC 服务器最大负载。

MEC 服务器采用先到先服务的规则处理任务, 当服务器处于满载状态时, 后到的任务进入等待队列。因此, 任务 $A_{i,j}$ 在 MEC 服务器 k 的处理时延 T_{proc} 表达式如下:

$$T_{\text{proc}} = \frac{\beta \times b}{f_{\text{mec},\text{alloc}}} + T_w \quad (5)$$

其中: T_w 表示等待时间。

任务 $T_{i,j}$ 的总时延表达式如下:

$$T_{i,j} = \begin{cases} T_{\text{loc}}, & x_{i,j} = 0 \\ T_{\text{up}} + T_{\text{proc}}, & x_{i,j} = 1 \end{cases} \quad (6)$$

1.4 MEC 服务器负载均衡模型

在 MEC 服务器的负载均衡中, 服务器的负载分布情况是 1 个关键指标, 标准差可以用来评价服务器负载集合的离散程度。如果标准差较小, 说明服务器的负载分布较均衡, 有助于实现车辆任务的合理分配及计算资源的优化利用, 防止 MEC 服务器负载过高; 如果标准差较大, 说明服务器的负载分布较不均衡, 有些服务器负载较重, 而有些服务器负载较轻, 需要对负载较重的服务器进行任务迁移或资源调度等操作来实现负载均衡。因此, 标准差可以作为有效评价 MEC 服务器负载均衡的指标。

本文采用负载均衡标准差^[20,24,25]对 MEC 服务器的负载进行定量分析, 负载均衡标准差越小, MEC 服务器的负载越均衡。负载均衡标准差表达式如下:

$$\sigma = \sqrt{\frac{\sum_{k=1}^m (f_{\text{mec},k} - f_{\text{mec,avg}})^2}{m}} \quad (7)$$

其中: $f_{mcc,avg}$ 表示 MEC 服务器负载的平均值。

1.5 优化问题制定

任务时延是制定车辆任务卸载策略的关键指标,快速处理对于计算型任务尤为重要。MEC 服务器负载均衡也是任务卸载策略的关键指标,MEC 服务器处于负载均衡状态,可缓解个别服务器负载压力,提高系统计算资源利用率。本文任务卸载模型以最小化时延及最小化 MEC 服务器负载均衡标准差为目标,将任务分配问题视为多目标优化问题。多目标优化问题定义如下:

$$\left\{ \begin{array}{l} \min T = \sum_{i=1}^h \sum_{j=1}^n T_{i,j} \\ \min \sigma = \sqrt{\frac{\sum_{i=1}^m (f_{mcc,k} - f_{mcc,avg})^2}{m}} \\ \text{s.t.} \\ \text{C1: } T_{i,j} \leq T_{\max}, \forall i \in H, \forall j \in N \\ \text{C2: } f_{loc,i} \leq f_{\max}, \forall i \in H \\ \text{C3: } f_{mcc,k} \leq f_{\max}, \forall k \in M \\ \text{C4: } x_{i,j} \in \{0, 1\}, \forall i \in H, \forall j \in N \end{array} \right. \quad (8)$$

其中:C1表示任务的最大容忍时延约束,保证任务处理时延不超过 T_{\max} ;C2、C3表示车辆单元及 MEC 服务器的最大计算资源约束;C4表示任务只能选择一种任务处理方式。

2 任务卸载策略

在车联网中,因车辆单元在空间和时间上分布的不均衡性,导致 MEC 服务器负载不均衡,并且车辆任务具有低时延的要求,部分 MEC 服务器会在短时间内达到很高的负荷甚至过载,这往往会增加任务时延。然而,有的 MEC 服务器处于低负荷状态,降低计算资源使用效率。因此,本文以时延及 MEC 服务器负载均衡为目标制定任务卸载策略。

由于车联网中车辆单元的特性,MEC 服务器的负载具有变化快和持续波动的特点,因此仅依据当前的计算负载制定任务卸载策略,无法适应未来一段时间内的负载变化,继而导致任务时延增大以及卸载滞后。因此,本文制定负载预测方法,为任务卸载策略提供可靠的计算资源状态变化信息。

2.1 负载预测模型

2.1.1 LSTM 预测模型

长短期记忆网络是一种常用于处理序列数据的深度学习模型,由多个门控单元组成,可以自适应地选择要记忆或遗忘的信息,并在处理长序列数据时能够有效地保持梯度稳定性。相比传统的递归神经网络(RNN),长短期记忆网络具有更强的记忆能力和更好的长期依赖建模能力^[26]。LSTM 可以完美地

模拟多个输入变量的问题,适用于时间序列预测^[27]。因此,引入 LSTM 对 MEC 服务器进行负载预测具有重要意义。

在 MEC 服务器中主要依靠 CPU 为任务计算提供算力支持,因此本文用 CPU 使用率表示负载。由于 MEC 服务器中内存使用率、磁盘读操作数、磁盘写操作数与 CPU 使用率之间具有关联性,因此选取以上 4 个特征共同作为 LSTM 网络的输入特征信息。在预测方式上,选取 i 个连续时间序列数据为输入信息获得预测值,预测模型表达式如下:

$$Y_{t+1} = f(Y_t, X_t, Z_t, C_t, \dots, Y_{t-(i-1)}, X_{t-(i-1)}, Z_{t-(i-1)}, C_{t-(i-1)}) \quad (9)$$

其中: Y_t 表示 t 时刻的 CPU 使用率; X 表示内存使用率; Z 表示磁盘读操作数; C 表示磁盘写操作数。

LSTM 预测模型采用均方根误差(RMSE)评估预测结果, RMSE 反映预测值与真实值之间的偏差程度,值越小结果越优,其计算式如下:

$$R_{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_c(i) - y_0(i))^2}{n}} \quad (10)$$

其中: $y_c(i)$ 为预测值; $y_0(i)$ 为实测值; n 为样本个数。

2.1.2 自适应优化神经网络模型

LSTM 模型的非线性建模性能与 4 种超参数密切相关:学习率、隐含层节点数、训练次数以及批量大小。LSTM 超参数组成的解空间规模大,若采用传统方法寻找最优超参数将消耗巨大的算力及时间。本节采用自适应优化神经网络算法 AGA-LSTM 寻找最优网络模型。

1) 自适应遗传算法是一种模拟自然选择和遗传机制的随机化搜索算法。该算法利用进化论的思想搜索最优解,通过编码、交叉、变异等操作不断进化,从而逐步接近全局最优解。

衡量遗传算法性能的 2 个主要指标是全局寻优能力和收敛速度。遗传算法中交叉概率和变异概率对算法寻优能力及收敛速度起着重要作用。交叉概率增大可以提高种群个体的进化速度,但会破坏适应度较高的个体,交叉概率降低容易导致个体进化速度缓慢。变异概率增大,使得种群进化方向多变,不利于保留优势个体,变异概率降低则影响局部搜索能力。

在传统遗传算法中,交叉和变异概率固定不变,可能导致早熟收敛、收敛速度慢。针对上述问题,本文采用一种自适应遗传算法(AGA)对 LSTM 参数进行寻优。AGA 的基本思想是使交叉概率和变异概率随个体适应度值自动改变,在避免算法早熟和局部收敛的同时,提高全局寻优搜索能力及效率。

交叉概率和变异概率计算表达式如下:

$$P_c = \begin{cases} \frac{k_1(f_{\max} - \hat{f})}{f_{\max} - f_{\text{avg}}}, & \hat{f} > f_{\text{avg}} \\ k_3, & \hat{f} \leq f_{\text{avg}} \end{cases} \quad (11)$$

$$P_m = \begin{cases} \frac{k_2(f_{\max} - f)}{f_{\max} - f_{\text{avg}}}, & f > f_{\text{avg}} \\ k_4, & f \leq f_{\text{avg}} \end{cases} \quad (12)$$

其中: P_c 、 P_m 分别为交叉概率和变异概率; k_1 、 k_2 、 k_3 、 k_4 为设定值; f_{\max} 表示种群中最大适应度; f_{avg} 表示种群中平均适应度; \hat{f} 表示配对个体中较大的适应度; f 表示个体的适应度。

2) 适应度函数。本文通过 AGA 确定 LSTM 模型 4 种超参数最佳值, AGA 根据 LSTM 预测算法的评价指标 RMSE 计算适应度。适应度函数表达式如下:

$$f = \frac{1}{R_{\text{RMSE}}} \quad (13)$$

2.2 基于自适应优化神经网络的负载预测算法

在车联网中 MEC 服务器的负载变化快, 为了在任务卸载策略中实现对负载变化的及时响应, 需要建立可靠的负载预测算法, 根据过去负载的变化规律预测未来 MEC 服务器的负载, 为制定任务卸载策略提供可靠的计算资源状态变化信息。

由第 2.1 节可知, AGA 算法优化 LSTM 网络模型, 可以有效提高预测的准确性。本节以该方法为基础建立基于自适应优化神经网络的负载预测算法 LP-AGA-LSTM, 算法主要分为数据处理、预测模型构建和预测结果计算。

首先, 不同特征之间的量纲不一致, 为了消除特征之间的差异性需要对数据特征进行归一化处理, 本节采用 Max-Min 对数据集中所用到的特征进行归一化处理, 表达式如下:

$$\hat{y} = \frac{y - y_{\min}}{y_{\max} - y_{\min}} \quad (14)$$

其中: \hat{y} 表示归一化后的值; y 表示原始值; y_{\max} 表示样本内每个特征的最大值; y_{\min} 表示样本内每个特征的最小值。

通过 AGA-LSTM 确定 LSTM 最佳超参数值, 建立网络模型。由于训练后的 LSTM 模型通过数据学习负载的内在变化规律, 因此相同 MEC 服务器不需要重复训练, 根据算法的设定输入所需的历史数据, 便可快速得到负载预测值。

算法 1 基于自适应优化神经网络的负载预测算法

输入 最大进化代数 D , 种群大小 N , 数据样本集合 DATA, 自适应设定值 k_1 、 k_2 、 k_3 、 k_4 , 超参数解空间 ξ

输出 最优超参数值 gbest, 预测值 fore

1. 对 DATA 进行归一化

2. 初始化种群 pop

3. for $i=1$ to D do

4. 种群解码 values

5. for $j=1$ to N do

6. $\text{RMSE}_j = \text{LSTM}(\text{values}[j], \text{DATA})$

7. $\text{fit}[j] = 1/\text{RMSE}_j$ // 计算适应度值

8. end for

9. 更新种群最优个体 pbest

10. 更新历史最优个体 gbest

11. 种群复制

12. 根据式(11)计算 P_c 进行种群交叉

13. 根据式(12)计算 P_m 进行种群变异

14. end for

15. fore = fore_LSTM(gbest, DATA)

// 用最优参数建模预测

2.3 基于负载预测的多目标优化任务卸载策略

在前期建立相关模型的基础上, 本节提出基于负载预测的多目标优化任务卸载策略, 通过预测 MEC 服务器的负载, 分析 MEC 网络系统中计算资源的分布情况, 合理分配任务, 优化 MEC 服务器负载及任务时延。

非支配排序遗传算法-III 具有高效、均衡、可扩展、多样性等优点, 适用于多目标优化问题的求解^[28]。本文通过 NSGA-III 求解第 1 节提出的多目标优化模型, 迭代更新获得理想的任务卸载策略。

2.3.1 非支配排序遗传算法-III 模型

1) 编码在 MEC 网络系统中, 车辆单元产生的任务可在车载端执行或卸载至 MEC 服务器, 将任务所在的车辆编码为 0, MEC 服务器分别编码为 1, 2, ..., m 。染色体表示所有任务的卸载策略, 基因表示各个任务的卸载策略。图 2 所示为任务卸载策略的编码实例, 每辆车有 n 个任务, 染色体长度为 $n \times h$ 。

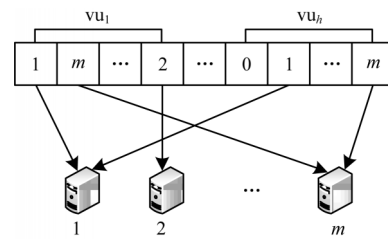


图 2 任务卸载策略编码实例

Fig.2 Coding example of task offloading strategy

2) 评价标准。本文采用非支配排序及拥挤度距离来评估 NSGA-III 所产生任务卸载策略的优劣。

根据支配关系将任务卸载策略划分到不同的前沿等级中, 优先选择前沿等级更小的任务卸载策略。当任务卸载策略的前沿等级相同时, 为了选择更好的解决方案, 进而考虑每个策略的拥挤度距离, 即该策略周围解的密度, 密度越大说明策略的优势越大。优先选择拥挤度距离更大的任务卸载策略。拥挤度距离计算式如下:

$$\begin{cases} D(i,j) = \frac{f_j(i+1) - f_j(i-1)}{f_{\max,j} - f_{\min,j}} \\ D(i) = \sum_{j=1}^n D(i,j) \end{cases} \quad (15)$$

其中: $D(i,j)$ 表示个体 i 在目标 j 上的距离; $f_{\max,j}$ 、 $f_{\min,j}$ 分别表示目标 j 的最大值与最小值; $f_j(i+1)$ 、 $f_j(i-1)$ 表示在目标 j 中个体 i 的 2 个相邻值; $D(i)$ 表示个体 i 的拥挤距离。

2.3.2 基于负载预测的多目标优化任务卸载算法

任务卸载策略综合考虑最小化任务时延和 MEC 服务器负载均衡 2 个目标优化, 卸载策略的单一目标最优并不能保证另一目标也最优, 需要同时权衡 2 个目标制定任务卸载策略。为此, 本文构建 NSGA-III 模型, 对第 1 节建立的多目标优化模型寻求最优解。

为实现任务卸载策略对负载变化的及时响应, 在第 2.2 节建立负载预测算法, 给任务卸载策略提供可靠的计算资源状态变化信息。本文的任务类型为计算密集型任务, 综合考虑通信环境、任务大小及传输时延, 为任务卸载策略提供下一时隙的 MEC 服务器负载信息而不进行多步预测。

基于负载预测的多目标优化任务卸载算法的执行过程如算法 2 所示。

算法 2 基于负载预测的多目标优化任务卸载算法

输入 车辆单元集合 vu 、任务集合 G ; MEC 服务器集合 MEC、不同通信小区车辆信息 RSU、带宽 W 路径损耗指数 γ 、车辆发射功率 P 、白高斯噪声功率 N_0 、上行信道衰落系数 h

输出 最优任务卸载策略 g_{best}

1. 调用算法 1 获得各 MEC 服务器负载预测值
2. 编码与初始化种群 pop
3. for $i=1$ to D do
4. 根据式(1)~式(6)计算卸载方案的总时延 T
5. 根据式(7)计算卸载方案的负载均衡标准差 σ
6. 交叉、变异
7. 对种群进行非支配排序及基于参考点选择
8. 生成新种群 pop_{i+1}
9. end for
10. 根据评价标准选择最优个体 g_{best}

基于负载预测的多目标优化任务卸载算法需要结合之前的基础工作来实现。首先, 调用算法 1 获得 MEC 网络系统中各服务器的计算资源预测值; 其次, 调用 NSGA-III 算法完成任务卸载策略优化; 最后根据评价标准选择最优个体作为任务卸载策略。

3 仿真

3.1 实验环境与数据集

本文仿真实验在内存 16 GB、处理器为 Intel® Core™ i7-8650U、频率 1.90 GHz 的 Windows 10 操作系统下进行, 使用 Python 3.8 对本文所提算法进行试

验仿真, 实验的具体参数配置如表 1 所示。

表 1 参数配置信息

Table 1 Parameter configuration information		
名称	值	含义
k	5	MEC 服务器数量
RSU_k	6、7、8、9、10	各通信小区车辆数
f_{mec}/GHz	15	MEC 服务器的 CPU 频率
f_{vu}/GHz	1.9	车载端 CPU 频率
n	[10,15]	单车辆卸载任务数
b/MB	0.6	卸载任务的数据量
T_{\max}/s	3	任务最大容忍时延
β	30	单位 bit 所需 CPU 周期数
W/MHz	100	带宽
d/m	70	车辆到 RSU 之间的距离
γ	2	路径损耗指数
P/W	1.3	车辆发射功率
N_0	3×10^{-13}	白高斯噪声功率
h	4	上行信道衰落系数

本文在内存 128 GB、处理器为 Intel® Xeon® Gold 5118 CPU, 频率 2.30 GHz, Windows 7 操作系统下采集 10 000 组数据作为实验数据集, 每组数据间隔时间为 0.1 s。数据集中的特征有时间、CPU 使用率、内存占用率、磁盘读操作数以及磁盘写操作数。

3.2 计算资源预测算法性能

本文将 LP-AGA-LSTM 算法与 LP-GA-LSTM 及 PSO-LSTM 算法进行对比, 对比方法分别采用传统遗传算法及粒子群算法优化 LSTM 超参数, 对负载进行预测。

本文实验按照 6:2:2 的比例将数据集划分为训练集、验证集和测试集, LSTM 超参数取值范围如表 2 所示。

表 2 LSTM 网络超参数取值

Table 2 Hyperparameter values of LSTM network	
超参数	取值范围
第 1 个隐藏层节点数	16, 32, 64, 128, 256
第 2 个隐藏层节点数	16, 32, 64, 128, 256
学习率	[0.001, 0.01]
训练次数	[50, 250]
批量大小	32, 64, 128, 256, 512

本文设定 3 种算法的最大进化代数为 30 次, 种群规模为 10, LSTM 预测时间步长 $i=10$, LP-AGA-LSTM 算法中 $k_1=k_3=0.8$, $k_2=k_4=0.3$, LP-GA-LSTM 算法中变异概率及交叉概率分别为 0.3 与 0.8, PSO-LSTM 算法中权重系数为 0.8, 加速因子为 0.5。本文实验以 RSME 作为评价指标, 对上述 3 种算法进行对比分析, 实验结果如图 3 所示。

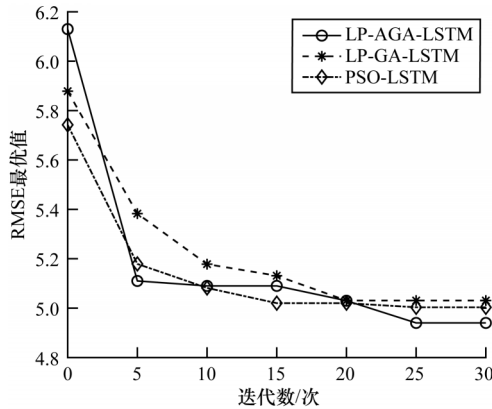


图3 迭代次数与RMSE最优值的关系

Fig.3 The relationship between the number of iterations and the optimal value of RMSE

从图3可以看出,随着迭代次数的增加,算法的RMSE最优值收敛速度逐渐降低,直至达到收敛状态。LP-AGA-LSTM迭代25次的RMSE收敛至4.93,LP-GA-LSTM迭代20次RMSE收敛至5.03,PSO-LSTM迭代15次RMSE收敛至5.02。

在实验中,虽然PSO-LSTM与LP-GA-LSTM先达到收敛状态,但它们陷入局部最优状态,在接下来的迭代中并没有搜寻到更优的RMSE值。LP-AGA-LSTM在迭代数为0~5次时能够以较快的速度收敛,在10~15次时跳出局部最优继续在空间中搜寻,直至收敛到最优值。本文算法在应对局部最优状态时有更优的表现。这是因为本文在LP-AGA-LSTM算法中引入了自适应的变异及交叉算子,自适应调整个体交叉及变异概率,对表现较差的个体赋予更大的交叉变异概率,提高局部搜索能力。迭代次数与RMSE均值对比如图4所示。

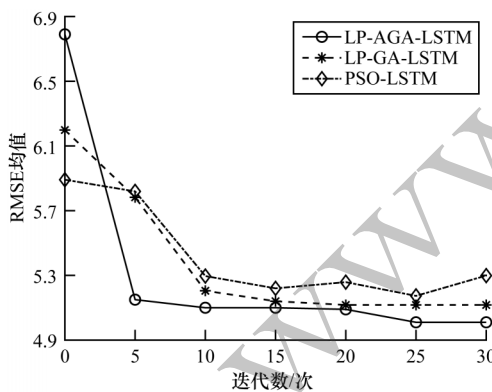


图4 迭代次数与RMSE均值的关系

Fig.4 The relationship between the number of iterations and the RMSE mean

从图4可以看出,随着迭代次数的增加,各代种群的RMSE均值呈现收敛趋势。LP-GA-LSTM与LP-AGA-LSTM分别于20次、25次达到收敛状态,而PSO-LSTM算法在30次内没有收敛,甚至还有小幅度的震荡。

从全局看LP-AGA-LSTM算法具有更优的局部搜索能力和全局搜索能力,收敛速度也得到提高。因此,LP-AGA-LSTM具有更优的寻优能力,提高负载预测的精度。经LP-AGA-LSTM算法寻找到的最优LSTM模型参数为第1个隐藏层与第2个隐藏层节点数256个、训练次数115、批量大小128、学习率0.009。

3.3 卸载策略性能

为验证MTUOA-LP算法的性能,本文引入以下4种卸载方案进行对比分析。

1)多目标优化任务卸载策略MTUOA,制定任务卸载策略方面与MTUOA-LP算法相同,以时延及MEC服务器负载均衡标准差为优化目标得到任务卸载策略,算法未考虑负载预测。

2)基于Qos的任务卸载策略QTD^[10]。为有效挖掘边缘计算网络的算力,QTD算法在时延及Qos等指标约束下,增加边缘网络承载任务数量。

3)基于NSGA2^[29]的车联网边缘计算任务卸载方案,在车联网边缘计算体系下,建立卸载模型,通过NSGA2算法得到卸载策略。

4)全卸载策略(AOS)将车辆的所有任务都随机卸载到MEC网络系统中的MEC服务器。

3.3.1 总时延与负载均衡标准差比较

图5所示为各卸载策略单辆车任务数与任务平均时延的关系。MTUOA-LP的平均时延均最优,与MTUOA、NSGA2、QTD、AOS对比,总体平均时延分别降低1.7%、7.3%、12.4%和17.5%。QTD算法希望增加边缘网络承载的任务数,但是该策略让边缘网络时刻接近过载状态,并且因更多的任务卸载至边缘端,导致车载端的计算资源没有得到充分利用,增大了时延。NSGA2算法因考虑能耗方面也增加卸载至边缘端的任务数量,导致时延增加。MTUOA算法未考虑负载预测,导致服务器过载,增加任务的排队时延。

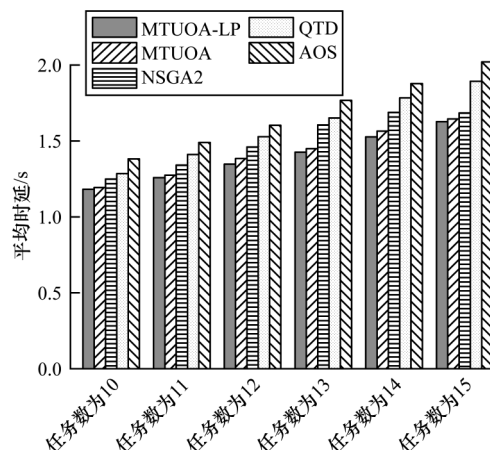


图5 任务数与平均时延的关系

Fig.5 The relationship between the number of tasks and the average time delay

图6所示为各卸载策略单辆车任务数与MEC服务器负载均衡标准差的关系。MTUOA-LP算法具有

显著优势。QTD 算法的负载均衡标准差优于其他 3 个算法,该算法让服务器负载接近满载值,但未考虑负载预测,导致任务卸载后 MEC 服务器的实际负载与理想状态有差异。NSGA2、QTD、AOS 算法均没有直接以负载均衡为目标制定卸载策略,因此它们的负载均衡标准差呈现不稳定的特点。MTUOA 算法虽然考虑负载均衡因素,但是未考虑负载预测,存在信息滞后问题,负载标准差是 MTUOA-LP 的 7.75 倍。

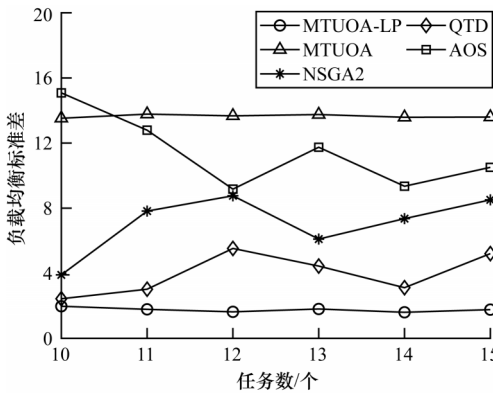


图 6 任务数与负载均衡标准差关系

Fig.6 The relationship between the number of tasks and the standard deviation of load balancing

与其他卸载策略相比,MTUOA-LP 算法能优化时延,并且在负载均衡方面取得显著优势。

3.3.2 通信小区卸载率比较

在 MEC 网络系统中,由于不同通信小区通信环境及车辆数等具有差异,因此不同通信小区的任务卸载率及平均传输时延存在差异。

图 7 所示为在 MTUOA-LP 算法中任务数与不同通信小区卸载率的关系,主要表现 2 个特点:1)随着任务数的增加,车辆单元计算能力不足,需要卸载更多任务至 MEC 服务器,导致卸载率增大;2)不同通信小区的车辆数不同,车辆数多的通信小区通信压力大,因此在任务数相同的情况下,随着车辆数增加卸载率降低。

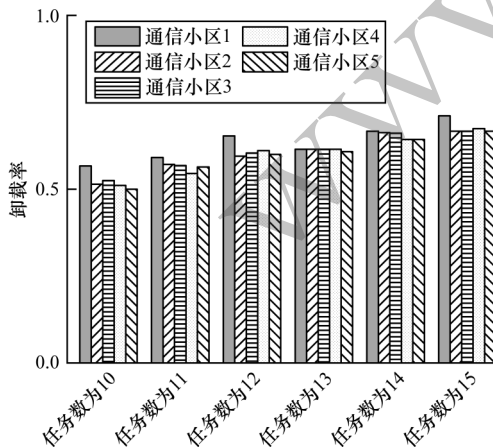


图 7 任务数与不同通信小区卸载率关系

Fig.7 The relationship between the number of tasks and the offloading rate among different communication cells

3.3.3 通信小区传输时延比较

图 8 所示为在 MTUOA-LP 算法中任务数量与不同通信小区平均传输时延的关系,主要表现 2 个特点:1)随着任务数的增加,车辆单元计算能力不足,需要卸载更多任务至 MEC 服务器,因通信资源是恒定的,随着传输任务数量增加,平均传输时延增大;2)不同通信小区的车辆数不同,车辆数多的通信小区通信压力大,平均传输时延增加。

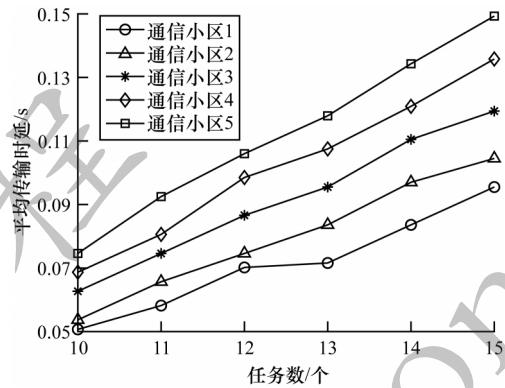


图 8 任务数与不同通信小区传输时延关系

Fig.8 The relationship between the number of tasks and the transmission time delay among different communication cells

通过比较不同通信小区的任务卸载率及平均传输时延可以看出,MTUOA-LP 算法综合考虑车辆数及通信环境等因素,针对不同通信小区制定差异化的任务卸载方案。

4 结束语

本文针对 MEC 网络系统架构中的任务卸载问题,提出一种基于负载预测的多目标优化任务卸载策略 MTUOA-LP。通过 LP-AGA-LSTM 算法对 MEC 服务器计算资源进行预测,以时延及负载均衡为目标,综合考虑通信环境、计算资源及任务量等因素给出最优的任务卸载策略。实验结果表明,LP-AGA-LSTM 算法预测精准率以及 AGA 的局部搜索能力、收敛速度都得到显著提高。MTUOA-LP 算法能优化任务时延,在负载均衡方面取得显著优势,能有效解决 MEC 服务器负载不均衡的问题。此外,MTUOA-LP 算法可以针对各个通信小区的环境、车辆数等因素,制定差异化的任务卸载策略。下一步将考虑更加复杂的通信环境,面对差异化的任务,添加循环预测机制,从而完善任务卸载策略。

参考文献

[1] VEGNI A M, LOSCRI V. A survey on vehicular social networks[J]. Communications Surveys & Tutorials, 2015, 17(4): 2397-2419.
 [2] ZHAI Y L, SUN W X, WU J Q, et al. An energy aware

- offloading scheme for interdependent applications in software-defined IoV with fog computing architecture[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 22(6):3813-3823.
- [3] MU S D, XIONG Z X, TIAN Y X. Intelligent traffic control system based on cloud computing and big data mining[J]. *IEEE Transactions on Industrial Informatics*, 2019, 15(12): 6583-6592.
- [4] GAO Q, LIN S Y, ZHU G. Joint vehicular and static users multiplexing transmission with hierarchical modulation for throughput maximization in vehicular networks[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 21(9):3835-3847.
- [5] LUONG H P, PANDA M, VU H L, et al. Beacon rate optimization for vehicular safety applications in highway scenarios[J]. *IEEE Transactions on Vehicular Technology*, 2018, 67(1):524-536.
- [6] GUO H Z, ZHANG J, LIU J J. FiWi-enhanced vehicular edge computing networks: collaborative task offloading[J]. *IEEE Vehicular Technology Magazine*, 2019, 14(1):45-53.
- [7] 谢人超, 廉晓飞, 贾庆民, 等. 移动边缘计算卸载技术综述[J]. *通信学报*, 2018, 39(11):138-155.
XIE R C, LIAN X F, JIA Q M, et al. Survey on computation offloading in mobile edge computing [J]. *Journal on Communications*, 2018, 39(11): 138-155. (in Chinese)
- [8] 杨天, 杨军. 移动边缘计算中的卸载决策与资源分配策略[J]. *计算机工程*, 2021, 47(2):19-25.
YANG T, YANG J. Offloading decision and resource allocation strategy in mobile edge computing[J]. *Computer Engineering*, 2021, 47(2):19-25. (in Chinese)
- [9] LIU J, MAO Y Y, ZHANG J, et al. Delay-optimal computation task scheduling for mobile-edge computing systems[C]//*Proceedings of International Symposium on Information Theory*. Washington D. C. , USA; IEEE Press, 2016:1451-1455.
- [10] SONG Y Z, YAU S S, YU R Z, et al. An approach to QoS-based task distribution in edge computing networks for IoT applications[C]//*Proceedings of International Conference on Edge Computing*. Washington D. C. , USA; IEEE Press, 2017:32-39.
- [11] REN J K, YU G D, CAI Y L, et al. Latency optimization for resource allocation in mobile-edge computation offloading[J]. *IEEE Transactions on Wireless Communications*, 2018, 17(8):5506-5519.
- [12] ZHANG K, MAO Y M, LENG S P, et al. Optimal delay constrained offloading for vehicular edge computing networks[C]//*Proceedings of International Conference on Communications*. Washington D. C. , USA; IEEE Press, 2017:1-6.
- [13] SALEEM U, LIU Y, JANGSHER S, et al. Latency minimization for D2D-enabled partial computation offloading in mobile edge computing [J]. *IEEE Transactions on Vehicular Technology*, 2020, 69(4):4472-4486.
- [14] XU X L, GU R H, DAI F, et al. Multi-objective computation offloading for Internet of Vehicles in cloud-edge computing[J]. *Wireless Networks*, 2020, 26(3):1611-1629.
- [15] ZHANG J, GUO H Z, LIU J J, et al. Task offloading in vehicular edge computing networks: a load-balancing solution[J]. *IEEE Transactions on Vehicular Technology*, 2020, 69(2):2092-2104.
- [16] YAO D H, GUI L, HOU F, et al. Load balancing oriented computation offloading in mobile cloudlet [C]//*Proceedings of the 86th Vehicular Technology Conference*. Washington D. C. , USA; IEEE Press, 2018:1-6.
- [17] DAI Y Y, XU D, MAHARJAN S, et al. Joint load balancing and offloading in vehicular edge computing and networks[J]. *IEEE Internet of Things Journal*, 2019, 6(3):4377-4387.
- [18] LUO G Y, YUAN Q, ZHOU H B, et al. Cooperative vehicular content distribution in edge computing assisted 5G-VANET[J]. *China Communications*, 2018, 15(7):1-17.
- [19] 林峰, 段建岚, 李传伟, 等. C-V2X 边缘服务器的动态负载均衡算法研究[J]. *计算机工程*, 2020, 46(12):201-206, 221.
LIN F, DUAN J L, LI C W, et al. Research on dynamic load balancing algorithm for C-V2X edge server[J]. *Computer Engineering*, 2020, 46(12):201-206, 221. (in Chinese)
- [20] 余翔, 刘一勋, 石雪琴, 等. 车联网场景下的移动边缘计算卸载策略[J]. *计算机工程*, 2020, 46(11):29-34, 41.
YU X, LIU Y X, SHI X Q, et al. Mobile edge computing offloading strategy under Internet of vehicles scenario[J]. *Computer Engineering*, 2020, 46(11): 29-34, 41. (in Chinese)
- [21] WANG Y T, SHENG M, WANG X J, et al. Mobile-edge computing: partial computation offloading using dynamic voltage scaling[J]. *IEEE Transactions on Communications*, 2016, 64(10):4268-4282.
- [22] LIANG L P, CHENG W C, ZHANG W, et al. Orthogonal frequency and mode division multiplexing for wireless communications [C]//*Proceedings of Global Communications Conference*. Washington D. C. , USA; IEEE Press, 2019:1-7.
- [23] 赵海涛, 朱银阳, 丁仪, 等. 车联网中基于移动边缘计算的内容感知分类卸载算法研究[J]. *电子与信息学报*, 2020, 42(1):20-27.
ZHAO H T, ZHU Y Y, DING Y, et al. Research on content-aware classification offloading algorithm based on mobile edge calculation in the Internet of vehicles[J]. *Journal of Electronics & Information Technology*, 2020, 42(1):20-27. (in Chinese)
- [24] WU Q, XU X L, ZHAO Q Z, et al. Tasks offloading for connected autonomous vehicles in edge computing [J]. *Mobile Networks and Applications*, 2022, 27(6):2295-2304.
- [25] 邹卯荣, 傅明, 熊兵. 基于时延与负载的 SDN 控制器部署模型[J]. *计算机工程*, 2019, 45(4):30-35.
ZOU M R, FU M, XIONG B. Deployment model of SDN

- controllers based on time delay and load[J]. Computer Engineering, 2019, 45(4):30-35. (in Chinese)
- [26] SHAH D, CAMPBELL W, ZULKERNINE F H. A comparative study of LSTM and DNN for stock market forecasting[C]//Proceedings of International Conference on Big Data. Washington D. C. , USA; IEEE Press, 2019; 4148-4155.
- [27] GERS F A. Learning to forget; continual prediction with LSTM[C]//Proceedings of the 9th International Conference on Artificial Neural Networks. Washington D. C. , USA; IEEE Press, 1999; 850-855.
- [28] DEB K, JAIN H. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints [J]. IEEE Transactions on Evolutionary Computation, 2014, 18(4): 577-601.
- [29] 刘建华, 罗荣鑫, 刘佳嘉, 等. 基于NSGA2的车联网边缘计算任务卸载方案[J/OL]. 西安理工大学学报: 1-11 [2023-03-04]. <http://kns.cnki.net/kcms/detail/61.1294.n.20230130.1410.002.html>.
- LIU J H, LUO R X, LIU J J, et al. Task offloading scheme of Internet of Vehicles edge computing based on NSGA2 [J/OL]. Journal of Xi'an University of Technology: 1-11 [2023-03-04]. <http://kns.cnki.net/kcms/detail/61.1294.n.20230130.1410.002.html>. (in Chinese)

编辑 薛晋栋

计算机工程
www.ecice06.com