

# 基于遗传神经网络的边缘集群大规模数据流调度研究

张中超

(广东电网有限责任公司河源供电局, 广东 河源 517000)

**摘要:** 物联网技术和互联网技术的快速发展, 促使边缘设备从单一的数据消费角色转换为生产者。随着边缘设备的智能化发展, 其在执行数据任务时仍存在调度缺陷, 基于此, 研究基于遗传神经网络的边缘集群大规模数据流调度方法。以约束条件对应边缘设备与数据中心, 在给定的时间范围内建立虚拟队列, 确定边缘集群中的大规模数据流调度请求; 通过实际的调度请求划分数据类型, 在散度化理论下设定数据流所属类型, 按照双重约束情况表征边缘集群大数据流向量; 通过遗传神经网络中的适应度函数确定调度权值, 按照重要程度进行调度排序, 以激发函数对应目标实现大规模数据流调度。以四组不同类型的传感器作为边缘设备, 并以传感器在不同日期内的传输数据作为测试样本, 以数据调度完整度和调度效率作为测试指标, 在新方法应用过程中可以保证数据调度的完整性和效率, 具有应用价值。

**关键词:** 遗传神经网络; 边缘集群; 大规模数据流; 调度方法

中图分类号: TP183 文献标识码: A 文章编号: 1003-7241(2025)07-0110-05

## Research on Large Scale Data Flow Scheduling of Edge Clusters Based on Genetic Neural Networks

ZHANG Zhongchao

(Guangdong Power Grid Co., Ltd., Heyuan Power Supply Bureau, Heyuan 517000, China)

**Abstract:** The rapid development of IoT technology and internet technology prompts edge devices to transition from a single data consumer role to a producer. With the intelligent development of edge devices, there are still scheduling defects when executing data tasks. Based on this, a large-scale data flow scheduling method for edge clusters based on genetic neural networks is studied. Corresponding to edge devices and data centers with constraints, it establishes a virtual queue within a given time range to determine large-scale data flow scheduling requests in the edge cluster, it divides data types based on actual scheduling requests, sets the type of data flow under divergence theory, and characterizes the edge cluster big data flow vector according to dual constraints. It determines scheduling weights through fitness functions in genetic neural networks, sorts scheduling according to importance, and achieves large-scale data flow scheduling by stimulating function corresponding goals. Using four different types of sensors as edge devices, and the transmission data of sensors on different dates as test samples, with data scheduling integrity and scheduling efficiency as test indicators, can ensure the integrity and efficiency of data scheduling in the application process of the new method, which has application value.

**Keywords:** genetic neural network; edge cluster; large scale data flow; scheduling methods

## 0 引言

近年来, 各个科研领域的焦点逐渐聚焦于智能化与自动化技术中, 我国更是受到多种先进技术的启发, 在多个领域开始研发智能化技术。无论从信息管理手段还是从信息调度中, 均已经开始了智能化和自动化的研究, 尤其在信息调度中更偏向于边缘数据的分析和处理。由于数据调度过程中存在多种因素的干扰, 尤其在边缘设备的快速发展中, 大量的数据开始在边缘设备中自动流转, 为此需要应用有效的调度方法对数据进行处理, 以此保

证数据的有效调度。文献[1]中采用自动加权的方式设计了分簇调度方法, 主要针对大规模数据建立调度数学模型, 并针对边缘设备中的数据进行多目标分类, 以此完成分类模块和调度模块的设计, 实现大规模数据集的有效调度, 在一定程度上能够保证数据的完整性, 但由于分类过程存在多个目标的划分, 在调度效率上存在一定的不足之处。文献[2]中选择改进灰狼优化算法设计调度方法, 在考虑调度目标的基础上重新设计标准数据集, 在不断收敛过程中判断数据的调度标准和类型, 以此构建一个标准化的调度模型, 可以在一定程度上保证数据调度的效率, 但会存在数据流丢失的问题。为此, 为实现边缘集群设备中大规模数据流的有效调度, 此次采用遗

\*基金项目: 广东电网有限责任公司科技项目(GDKJXM20230471)

收稿日期: 2023-12-20

神经网络算法进行研究,设计一个新的调度方法,为保证大规模数据流的调度提供技术支持。

## 1 约束条件确定大规模数据流调度请求

大规模数据的调度研究以数据中心为参考,在边缘设备集群效应中,不同的服务请求会产生不同的调度目标,因此,选择约束性条件对大规模的数据量调取请求进行分析,实现具体且标准的请求处理<sup>[3]</sup>。设定在一个服务中心存在有多个边缘设备,每组设备对应有不同的用户,在规定时间内用户对数据的实例请求,如下:

$$\begin{cases} h(n,u) = \begin{cases} 1, \rightarrow u \\ 0, \text{else} \end{cases} \\ j(m,n,u) = \begin{cases} 1, \text{if } h(n,u) \neq 0, (m,u) \rightarrow u \\ 0, \text{else} \end{cases} \end{cases} \quad (1)$$

式中,式(1)主要用于验证用户是否发出了数据调取请求,式(2)表示数据中心是否接收到用户的数据调度请求。 $n$ 为用户<sup>[4]</sup>; $u$ 为数据实例的调度请求; $m$ 为数据中心; $h(n,u)$ 为请求发出状态函数。 $j(m,n,u)$ 为请求接收函数。由于不同的用户会产生多样化的数据请求,因此在确定用户的数据请求时,需要采用约束条件进行分析,具体为

$$l(m,n,u) = \begin{bmatrix} \lambda \\ \varphi \\ \phi \\ \kappa \end{bmatrix} (u'') + \frac{\varepsilon_{u,n}^2 k(m) + \delta_{u,n}^2 k(m)}{2} \quad (2)$$

式中, $l(m,n,u)$ 为用户访问数据中心后发出实例请求所需的时间; $u''$ 为数据请求的虚拟队列; $\lambda, \varphi, \phi, \kappa$ 为约束条件,可形成约束矩阵 $\varepsilon_{u,n}^2, \delta_{u,n}^2$ 分别为请求的时间变异系数<sup>[5]</sup>; $k(m)$ 为数据中心历史存在的数据请求副本。按照历史副本衡量出现在数据中心的具体调度请求,计算方式为

$$i(m) = \frac{\sum_{n,u} h(n,u) o(u) j(m,n,u)}{k(m)} \quad (3)$$

式中, $i(m)$ 为衡量用户在数据中心发出调度请求的概率。 $o(u)$ 为实际发出的数据请求<sup>[6]</sup>。

通过请求所需的时间建立虚拟队列,以此对照历史请求副本进行分析,实现用户具体请求的确定,在此基础上采用散度化理论转换数据流向量。

## 2 散度化理论表征边缘集群大数据流向量

在数据调度过程中若只参考数据请求,会存在数据输出错误的情况,因此,通过散度化理论能够对边缘集群中的数据流进行向量分析,按照不同类型的数据进行监督学习,以此为后续的数据调度提供支持<sup>[7]</sup>。分别对数据流进行标签设定,将大数据流请求划分为不同的类型:

$$p_1 = \frac{Q \frac{w_1}{E}}{\sum_T Q \frac{w_1}{E}} \quad (4)$$

$$p_2 = \frac{Q \frac{w_2}{E}}{\sum_T Q \frac{w_2}{E}} \quad (5)$$

$$p_3 = \frac{Q \frac{w_3}{E}}{\sum_T Q \frac{w_3}{E}} \quad (6)$$

$$p_4 = \frac{Q \frac{w_4}{E}}{\sum_T Q \frac{w_4}{E}} \quad (7)$$

式中, $p$ 为数据流类型的软标签<sup>[8]</sup>; $T$ 为数据类型总数,此次 $T=1,2,3,4$ 。为简化此次调度流程直接将数据流划分为四组软标签,分别用 $p_1, p_2, p_3, p_4$ 来表示<sup>[9]</sup>; $W_1, W_2, W_3, W_4$ 为对应数据类型划分的参考变量; $E$ 为各类型数据划分所需时间。

分别在各自数据类型中对其调度过程进行向量转换,能够针对具体的数据类型进行具体输出,也可以保证数据的结构在输出过程中不受到改变,在散度化过程中以最小化损失函数进行双重约束,具体为

$$Y(p_1, p_2, p_3, p_4) = \text{soft}_{\min} \frac{\exp(U_{p_1, p_2, p_3, p_4})}{\sum_U \exp(U_{p_1, p_2, p_3, p_4})} \quad (8)$$

式中, $Y(p_1, p_2, p_3, p_4)$ 为数据类型的向量转换结果; $U_{p_1, p_2, p_3, p_4}$ 为对应数据类型的输出向量; $\text{soft}_{\min}(\cdot)$ 为最小化损失函数; $U$ 为标准向量表示形式。

通过对数据类型的划分,在对应类型中转换其向量形式,以此针对调度请求建立调度目标,实现边缘集群的大数据流调度。

## 3 基于遗传神经网络调度大规模数据流

遗传神经网络将遗传算法与神经网络相结合,是一种具有更加复杂机制的神经网络,能够解决多维度的复杂问题,并且具有更好的表达方式<sup>[10]</sup>。此次以遗传神经网络实现大规模数据流调度,能够实现更加紧凑和简洁的调度表达。遗传算法自带有适应度函数,能够对需求目标进行调整,可以将其应用在不同类型数据流的调度过程中,如下:

$$q = \frac{1}{a_1 z_1 w_1 + a_2 z_2 w_2 + a_3 z_3 w_3 + a_4 z_4 w_4} \quad (9)$$

式中: $w_1, w_2, w_3, w_4$ 为设定的调度目标<sup>[11]</sup>; $a_1, a_2, a_3, a_4$ 为适应度权值,可以表示数据流调度的重要程度,此次将所有调度目标的权值总和设定为1,即 $a_1 + a_2 + a_3 + a_4 = 1$ <sup>[12]</sup>; $z_1, z_2, z_3, z_4$ 分别为调度参数; $q$ 为适应度函数。分别对数据

流的调度目标  $w_1, w_2, w_3, w_4$  进行表示:

$$w_1 = \min \left( \max_{1 \leq x \leq e} (s_x) \right) \quad (10)$$

$$w_2 = \min \left( \max_{1 \leq x \leq e} \sum_{r=1}^v \sum_{d=1}^{d_x} c_{rx d} f_{rx d} \right) \quad (11)$$

$$w_3 = \min \left( \sum_{x=1}^e \sum_{r=1}^v \sum_{d=1}^{d_x} c_{rx d} f_{rx d} \right) \quad (12)$$

$$w_4 = \min \left( \sum_{x=1}^e \sum_{r=1}^v \sum_{d=1}^{d_x} c_{rx d} f_{rx d} t_{rx d} \right) \quad (13)$$

式中:  $w_1$  为调度完成量目标;  $w_2$  为调度最小负载目标;  $w_3$  为总负载目标;  $w_4$  为调度利用率目标;  $e$  为边缘集群中数据流集合<sup>[13]</sup>;  $v$  为待调度的数据量总量;  $r$  为第  $r$  种调度数据类型;  $x$  为第  $x$  道调度顺序;  $s_x$  为完成时间;  $d_x$  为初始时间;  $c_{rx d}$  为所需时间;  $f_{rx d}$  为调度所需总工序;  $t_{rx d}$  为数据流调度量<sup>[14]</sup>。

将设定的调度目标融合至神经网络之中, 实现边缘集群大规模数据流的调度:

$$t_g = z(a_1 w_1 + a_2 w_2 + a_3 w_3 + a_4 w_4) + b_y \quad (14)$$

式中:  $t_g$  为大数据流在遗传网络中的激发输出调度结果;  $b_y$  为激发函数<sup>[15]</sup>。

当边缘集群数据输入至神经网络时, 通过网络中的学习机制判断具体调度目标, 并依据目标进行更替学习, 实现最终的大数据流规则调度。至此, 本文基于遗传神经网络完成大规模数据流的调度方法设计。

#### 4 实验测试分析

通过遗传神经网络对大数据的调度方法进行重新研究, 针对于边缘集群的大规模数据调度设计了新的调度方法, 为验证该方法具有实际应用价值, 采用对比测试的方式进行论证。

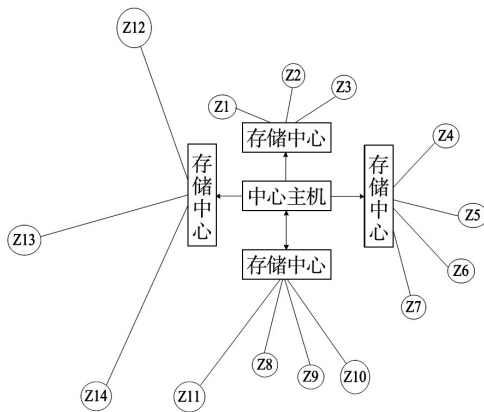


图1 测试环境

分别采用基于自适应加权的调度方法和改进灰狼优化算法的调度方法作为对照组, 与新方法进行对比, 验证不同调度方法的有效性。由于大规模数据在调度过程中容易产生数据丢失的问题, 且当大规模数据处于网络边

缘环境中, 其调度的过程容易受环境影响, 造成调度效率低下的问题。基于此, 以某个传感器网络作为测试环境, 对其连接的传感器结构进行展示, 具体见图1所示。

如图1所示, 此次以多组传感器网络结构作为测试环境, 在该组成结构中能够看出, 不同传感器所处的位置与中心位置具有一定的距离, 并且传输距离具有差异性。当传感器通路距离中心越近时, 数据调度的效率越高, 且不容易受到外界攻击和内部的信号干扰, 则数据的调度完整度也就越高。从图1中可以看出, 在14组传感器中 Z1, Z2, Z3 距离中心较近, 而 Z11, Z12, Z13, Z14 位于整个网络结构的边缘, 距离中心较远。将距离较远的4组传感器作为测试对象, 以近半个月的传输数据作为统计对象, 对每组传感器内产生的数据进行统计, 见表1。

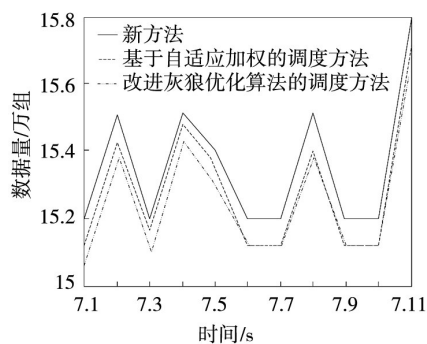
表1 数据量统计结果/万组

日期	传感器			
	Z11	Z12	Z13	Z14
7.1	15.2	14.5	12.2	10.3
7.2	15.5	14.2	12.6	10.5
7.3	15.2	14.5	12.4	10.4
7.4	15.5	14.5	12.5	10.6
7.5	15.4	14.5	12.5	10.6
7.6	15.2	14.5	12.6	10.6
7.7	15.2	14.5	12.6	10.8
7.8	15.5	14.5	12.6	10.2
7.9	15.2	14.8	12.4	10.2
7.10	15.2	14.5	12.2	10.2
7.11	15.8	14.6	12.4	10.5

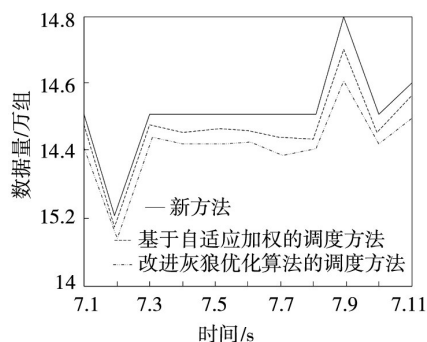
根据表1中内容所示, 对每组传感器在不同日期内总计传输数据进行统计, 能够看出不同传感器中产生的数据具有差异性, 但每组通路中产生的数据量均超过10万组, 符合大数据集群数据规模, 可以作为此次测试数据。按照上述各组传感器中产生的数据作为后续调度测试内容, 以两个阶段完成测试, 在第一阶段验证各组方法的调度完整度, 即是否能够在一次调度中完成全部数据的调度, 在第二阶段验证各组方法的调度效率, 即是否可以快速完成数据调度。将所有数据上传至 MATLAB 测试平台中, 并分别连接三组调度方法进行测试, 结果见图2。

如图2所示, 以每日的传输数据作为调度样本, 在新方法应用下能够实现各组传感器数据的调度, 且调度后的数据结果与实际情况相一致, 没有出现丢失现象, 具有调度完整度。两组传统方法在调度过程中出现了丢失问题, 尤其在 Z10 传感器中丢失的数据较多, 主要是该传感器距离中心较远, 在调度中无法保证数据的完整度。综上所述, 在不同距离的传感器位置设定下, 新方法在调度中不会受到影响, 能够实现全部数据的有效调度, 具有应

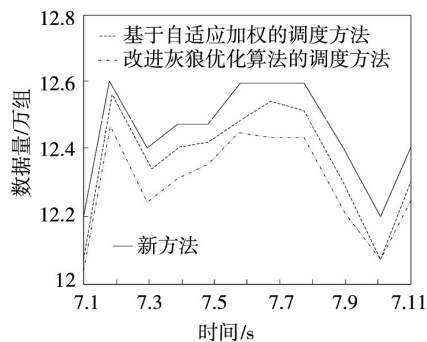
用价值。



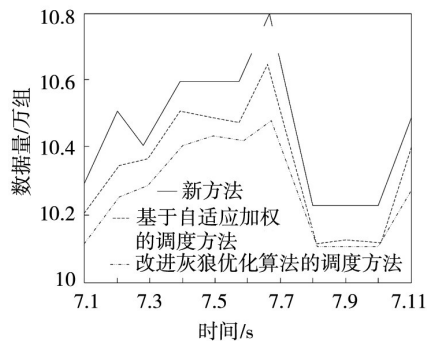
(a) Z11



(b) Z12



(c) Z13



(d) Z14

图2 各组方法调度完整性测试结果

第一阶段验证了不同方法的调度完整性,能够说明新方法更具有应用价值,为进一步分析新方法的应用效果,在此基础上,验证不同方法的调度效率。由于传感器网络为运行共同体,选择的4组传感器在调度中应为同步

状态,则在调度效率测试中以4组传感器同时调度为验证前提,并需要保证数据调度的完整度。按照每组传感器的总计传输数据作为测试数据,分别在测试平台中通过三组调度方法进行验证,结果见图3。

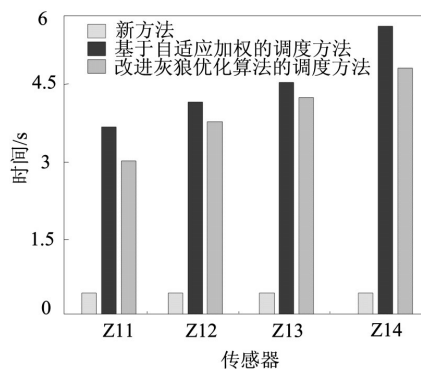


图3 不同方法调度效率对比结果

根据图3中内容所示,在同时进行4组传感器数据调度时,新方法能够保证每组传感器的调度时间在1.25 s之内,各组传感器的调度时间较为均匀,说明具有一定的调度效率。而两组传统方法在应用过程中受无法全部调取影响,需要在数据传输过程中进行多次调度,则对于每组传感器的调度效率不能保持在同一水平,距离较远的传感器所需的调度时间较多,距离较近的传感器调度时间较少,因此,在同步调度过程中更容易受到其他干扰。综上所述,在遗传神经网络算法的基础上设计了新的调度方法,不仅具有较高的调度完整度,可以真实地保存调度数据,也可以实现较快的调度,具有应用价值。

## 5 结束语

基于神经网络设计新的数据流调度方法,并通过实验测试论证了新方法的有效性,不仅可以保证大规模数据流调度的完整性,又可以保证大数据流调度的效率,具有较好的应用价值。但由于时间有限,在研究过程中仍存在少许不足之处,如对不同的数据类型没有进行具体划分,直接以边缘设备作为测试对象进行测试分析,后续研究中会针对此类问题进一步完成论证,验证调度方法的有效性,以此保证大规模数据流调度的应用效果。

## 参考文献:

- [1] 陈宇, 张勇, 陈实. 大规模卫星集群网络自适应加权分簇算法[J]. 北京理工大学学报, 2021, 41(11): 1188-1192.
- [2] 李开言, 王亚刚, 张涛. 基于改进灰狼优化算法的边缘计算任务调度方法[J]. 信息与控制, 2022, 51(4): 489-497, 512.
- [3] 郑守建, 彭晓晖, 王一帆, 等. 一种基于综合匹配度的边缘计算系统任务调度方法[J]. 计算机学报, 2022, 45(3): 485-499.
- [4] 陆克中, 陈超凡, 蔡桓, 等. 面向概念漂移和类不平衡数据流的在线分类算法[J]. 电子学报, 2022, 50(3): 585-597.

[5] 张翔, 廖海君, 周振宇, 等. 含规模化5G基站租赁共享储能  
的配电网混合博弈优化调度[J]. 电测与仪表, 2023, 60(5): 23-32.

[6] 张翀宇, 陈彦明, 李炜. 边缘计算中面向数据流的实时任  
务调度算法[J]. 计算机科学, 2022, 49(7): 263-270.

[7] 杨桂华, 卫嘉乐. 基于免疫蚁群优化算法的仓储任务调度  
[J]. 组合机床与自动化加工技术, 2023(1): 179-183.

[8] 林秋珍, 赵越, 孙忠阁. 大规模光纤网络数据流的自动分  
类[J]. 激光杂志, 2021, 42(9): 80-84.

[9] 杨茸, 牛保宁. OIQ-tree: 一种支持大规模空间文本数据流  
上连续k近邻查询的索引[J]. 计算机学报, 2021, 44(8): 1732-1750.

[10] 杨华, 孙欣伊, 贾宗星, 等. 基于博弈论的命名数据网  
络拥塞控制策略[J]. 东华大学学报(自然科学版), 2021, 47(4):  
49-54, 61.

[11] 查满霞, 祝永晋, 朱霖, 等. 面向实时流数据处理的边缘  
计算资源调度算法[J]. 计算机应用, 2021, 41(S1): 142-148.

[12] 吴明杰, 陈庆奎. 面向边缘集群内AI数据流的双平面调

度模型[J]. 小型微型计算机系统, 2021, 42(6): 1332-1339.

[13] 侯晓双, 张俊. 图数据流上时间尊重图模式匹配算法研究  
[J]. 计算机应用研究, 2021, 38(7): 1988-1992.

[14] 赵全, 汤小春, 朱紫钰, 等. 大规模短时间任务的低延迟  
集群调度框架[J]. 计算机应用, 2021, 41(8): 2396-2405.

[15] 吴任博, 刘淑琴. 考虑市场交易的新型电力系统调度辅助  
决策方法研究[J]. 电测与仪表, 2023, 60(11): 81-89.

作者简介: 张中超(1987—), 男, 硕士, 工程师, 研究方向: 电网运  
行, 信息安全、电力系统数据分析、调度自动化系统等。

(上接第12页)

通过对PID、Simth、改进Simth三种控制方法构建的  
控制系统进行仿真实验, 对比分析实验结果得出无扰动  
情况下的改进Simth预估器控制系统具有最好的阻尼效  
果和最快的响应速度。引入不同的干扰信号后, 这种改  
进Simth预估器控制系统相比其余两种方法, 仍具有良好  
的动态性能和稳定性能, 实现了较好的控制效果优化。

### 参考文献:

[1] JIANBO J, AIMIN M, PENG L, et al. Design of mixed ex-  
perimental platform for fuzzy Smith time-delay compensation control  
system of double-capacity water tank[J]. Experimental Technology  
and Management, 2017, 34(1): 84-89.

[2] 高宏岩, 郭春江. 基于Matlab和双容水箱的现代控制理论  
实验设计[J]. 实验技术与管理, 2019, 36(5): 52-56, 71.

[3] LUYBEN W L. Liquid level control: Simplicity and com-  
plexity[J]. Journal of Process Control, 2020(86): 57-64.

[4] 郝泽军. 双容水箱液位控制系统设计[J]. 电子技术, 2018,  
47(12): 73-77.

[5] 冯慧山. Simth串级PID控制在锅炉主蒸汽温度控制中的  
应用[J]. 石油化工自动化, 2016, 52(3): 41-45.

[6] AHMAD S, ALI S, TABASHA R. The design and imple-  
mentation of a fuzzy gain-scheduled PID control ler for the Festo MPS  
PA compact workstation liquid level control[J]. Engineering ence and  
Technology, an International Journal, 2020, 23(2): 307-315.

[7] 常瑞丽, 韩军. 双容水箱液位控制系统的开发与研究[J].  
机床与液压, 2014, 42(10): 166-168, 171.

[8] 夏百花, 王雪洁. 纯滞后系统控制算法的比较分析[J]. 无  
线互联科技, 2018, 15(24): 98-99.

[9] 寇发荣, 李冬, 许家楠, 等. 电动静液压主动悬架的内模-Smith  
时滞补偿控制[J]. 液压与气动, 2018(12): 48-53.

[10] NIE YONGHUI, ZHANG PENGYU, CAI GUOWEI, et  
al. Unified Smith predictor compensation and optimal damping control  
for time-delay power system [J]. International Journal of Electrical

Power&Energy Systems, 2020(117): 105670.

[11] 张蒙. 新型Smith预估控制器在大延迟系统中的应用[J].  
电力自动化设备, 2011, 31(4): 136-138.

[12] DENG J, HAO C. The smith-pid control of three-tank-sys-  
tem based on fuzzy theory[J]. Journal of Computers, 2011, 6(3):  
514-523.

[13] 高锦, 章家岩, 冯旭刚. 基于Smith预估补偿的啤酒发酵  
温度控制策略[J]. 食品与机械, 2019, 35(8): 115-120.

[14] 王永帅, 陈增强, 孙明玮, 等. 一阶惯性大时滞系统Smith  
预估自抗扰控制[J]. 智能系统学报, 2018, 13(4): 500-508.

[15] 李冬辉, 高峰. 基于扰动观测器的压缩式制冷系统改进  
Smith预估解耦控制[J]. 上海交通大学学报, 2019, 53(5): 593-599.

[16] 王永祥, 张建平, 张锐琳, 等. 船舶中央冷却水多段比值  
控制系统设计[J]. 船舶工程, 2018, 40(5): 72-76.

[17] JOSE MARIO ARAUJO, TITO L. M. SANTOS. Control  
of second-order asymmetric systems with time delay: smith predictor  
approach[J]. Mechanical Systems and Signal Processing, 2020(137):  
106355.

[18] 杨辰飞, 牛晨旭, 吕蒙. Smith预估控制器在大延迟系  
统中的应用[J]. 工业控制计算机, 2019, 32(11): 106-107.

[19] 平玉环, 管志敏, 李宗耀. 基于Smith预估的模糊自适应  
主汽温控制系统[J]. 中国电力, 2018, 51(11): 9-14.

[20] 时靖, 马柳, 陈广华. 一种改进型Smith预估控制器的设  
计与仿真[J]. 国外电子测量技术, 2012, 31(7): 56-58.

作者简介: 高甜甜(1998—), 女, 硕士研究生, 研究方向: 智能控制。  
通信作者: 王沁(1981—), 男, 博士, 副教授, 研究方向: 智能控制、  
机器人设计与仿真。