

基于异构硬件约束的智能驾驶模型多目标优化

黄锐桐 朱弘扬* 曾伟新 蔡嘉威

广东海洋大学, 中国·广东 湛江 524000

摘要: 随着智能驾驶向 L4/L5 级别迈进, 高精度端到端感知对硬件的巨大成本压力成为商业化落地的核心瓶颈。现有研究多聚焦于单一硬件或算法优化, 缺乏系统级“成本—精度”权衡建模。本研究通过多元非线性回归和 XGBoost 模型建模硬件成本与推理精度的非线性映射关系, 并利用 NSGA-II 遗传算法优化传感器配置与算力平台选择, 嵌入实时性和功能安全等硬约束。结果表明, 在保持 $mAP \geq 85\%$ 的前提下, 优化方案可降低硬件成本 25%~35%, 生成适配不同场景的高性价比配置。论文创新性地将帕累托多目标优化理论引入自动驾驶硬件—算法协同设计, 为车企提供全链路决策工具, 对推动自动驾驶规模化落地、破解“性能—成本”困境具有重要工程价值, 也为复杂系统多目标优化提供了统计建模新范式。

关键词: 推理精度; 帕累托优化; 统计建模; XGBoost 模型; NSGA-II 遗传算法

Multi Objective Optimization of Intelligent Driving Model based on Heterogeneous Hardware Constraints

Huang Ruitong Zhu Hongyang* Zeng Weixin Cai Jiawei

Guangdong Ocean University, Zhanjiang, Guangdong, 524000, China

Abstract: As intelligent driving moves towards L4/L5 levels, the huge cost pressure of high-precision end-to-end perception on hardware has become the core bottleneck for commercial implementation. Existing research mostly focuses on single hardware or algorithm optimization, lacking system level “cost accuracy” trade-off modeling. This study models the nonlinear mapping relationship between hardware cost and inference accuracy through multiple nonlinear regression and XGBoost models, and uses NSGA-II genetic algorithm to optimize sensor configuration and computing platform selection, embedding hard constraints such as real-time performance and functional safety. The results indicate that while maintaining $mAP \geq 85\%$, the optimization scheme can reduce hardware costs by 25% -35% and generate cost-effective configurations suitable for different scenarios. This paper innovatively introduces the Pareto multi-objective optimization theory into the hardware algorithm collaborative design of autonomous driving, providing a full chain decision-making tool for automotive companies. It has important engineering value in promoting the large-scale implementation of autonomous driving and solving the “performance cost” dilemma. It also provides a new paradigm for statistical modeling of multi-objective optimization of complex systems.

Keywords: Inference accuracy; Pareto optimization; Statistical modeling; XGBoost model; NSGA-II genetic algorithm

0 前言

随着全球科技革命和产业变革的蓬勃发展, 汽车产业正加速向电动化、网联化和智能化转型。中国政府近年来出台了一系列政策文件, 如《新能源汽车产业发展规划(2021—2035 年)》, 明确提出发展智能驾驶技术是国家战略^[1]。然而, 在智能驾驶向 L4/L5 级别迈进的过程中, 高精度端到端感知对硬件的高成本要求成为商业化落地的核心瓶颈。当前, L4 级自动驾驶硬件方案成本常超过 10 万美元, 其中传感器和算力平台占整车硬件成本比例超过 70%。与此同时, 算力资源利用率不足 60%, 存在严重的性能冗余问题^[2]。

传统研究多聚焦于单一硬件或算法优化, 缺乏对“传感器配置—模型复杂度—算力平台”三者联动效应的系统考

量。现有硬件成本模型大多假设线性累加, 难以捕捉真实世界中的非线性成本因素。针对这些问题, 本研究旨在通过多元非线性回归和 XGBoost 模型建模硬件成本与推理精度的非线性映射关系, 并利用 NSGA-II 遗传算法优化传感器配置与算力平台选择, 嵌入实时性和功能安全等硬约束, 实现智能驾驶系统的多目标优化。

1 研究目标与方法

本研究设定了两个主要目标: 一是建立成本—精度映射模型, 构建自动驾驶硬件成本与推理精度的非线性关联模型; 二是实现多目标协同优化, 设计融合实时性与安全约束的多目标优化算法, 求解帕累托最优硬件配置方案集。

研究采用了多元非线性回归模型来预测硬件总成本,

该模型考虑了二次项和交互项以捕捉成本与硬件配置的非线性关系。对于感知性能指标（mAP、漏检率、延迟），则采用 XGBoost 多任务回归模型进行预测。这两个子模型共同构成了多目标优化框架的评价函数。

在优化算法方面，研究采用了改进的 NSGA-II 遗传

算法，该算法通过非支配排序和拥挤距离机制，在保证解集多样性的同时有效逼近帕累托最优前沿。个体编码包含激光雷达数量、摄像头数量和算力平台类型等关键参数，适应度评估则基于前述成本模型和精度模型的预测结果（见图 1）。

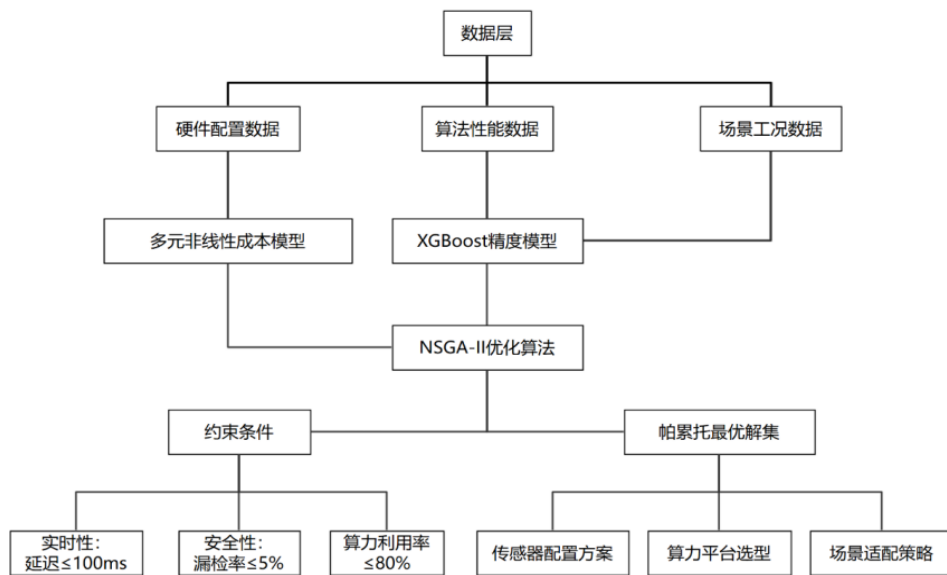


图 1 研究思路

1.1 数据处理

研究数据来源于公开数据集（nuScenes 和 Waymo Open Dataset）和仿真数据（CARLA 模拟器生成的极端场景），共计超过 200 万帧数据。为了进行可靠的建模和验证，论文遵循时间和场景多样性均衡的原则，将处理后的数据集划分为训练、验证和测试三部分：训练集（60%）、验证集（20%）和测试集（20%）。

1.2 统计建模与优化方法

我们首先建立硬件总成本 C 关于硬件配置的回归模型。考虑到成本与某些变量的非线性关系（如激光雷达数量的平方项可能影响标定成本），选择多元非线性回归模型，包括二次项和交互项：

$$C = \beta_0 + \beta_1 N_{\text{lidar}}^2 + \beta_2 N_{\text{cam}} + \beta_3 (\text{TOPS})^{0.6} + \beta_4 (N_{\text{lidar}} \times \text{resolution}) + \epsilon \quad (1)$$

其中， N_{lidar} 和 N_{cam} 分别为激光雷达和摄像头数量；TOPS 为算力峰值； ϵ 为误差项。上述模型中的指数和交互项由验证集上逐步回归确定，引入 $\beta_3 (\text{TOPS})^{0.6}$ 是因为算力成本随性能提升呈减速增长趋势。回归结果（见表 1）显示模型拟合效果良好：决定系数 $R^2=0.93$ ，F 检验显著（ $p < 0.001$ ）。其中，激光雷达数量的二次项和算力 TOPS 与分辨率的交互项均为显著影响因素（ $p < 0.01$ ），印证了成本的非线性驱动因素存在。

训练结果显示（见表 1），XGBoost 模型具有良好的预测精度和可解释性^[3]：在测试集上 mAP 预测的平均绝对误

差为 1.9%，漏检率 RMSE 为 0.7%，延迟 RMSE 约 7ms，均在可接受范围内。

表 1 回归和训练结果

评估指标	决定系数 R^2	F 检验	mAP 预测平均绝对误差	漏检率 RMSE	延迟 RMSE
数值	0.93	$p < 0.001$	1.9%	0.7%	约 7ms

通过上述模型，论文实现了从硬件配置到成本，以及从硬件 + 算法配置到性能指标的映射，为多目标优化提供了评价函数。其中，成本模型用于计算每个方案的成本值 C ，精度模型用于估计方案的精度指标 A 和延迟 L 等，以判断约束是否满足。

1.3 多目标优化框架

在建立评价模型基础上，我们设计多目标优化框架^[4]，以最小化硬件成本 C 和最大化推理精度 A 为目标，在约束条件下搜索帕累托最优解集^[5,6]。采用的方法为改进的 NSGA-II 遗传算法（快速非支配排序遗传算法），理由是 NSGA-II 在解决多目标优化问题上效率高且能维护解的多样性。

1.3.1 目标函数与约束

定义双目标优化问题如下：Min C 、Max A ，其中精度指标 A 综合考虑 mAP、漏检率和延迟等因素，可定义为加权组合以统一量纲。例如这里采用：

$$A = 0.6mAP + 0.3(1 - \text{AvgMissRate}(z)) + 0.1(1 - \text{AvgLatency}(z)/100) \quad (2)$$

即主要关注 mAP，其次关注漏检率和实时性（延迟越短越好）。约束条件包括：

$$\begin{cases} \text{AvgMissRate}(z) \leq 5\%, \text{ AvgLatency}(z) \leq 100\text{ms} \\ \text{FLOPs} \leq \text{TOPS} \times 0.8 \\ N_{\text{lidar}} \in \{0,1,2,3\}, N_{\text{cam}} \in \{4,6,8,10,12,14,16\} \end{cases}$$

其中，C(z) 为配置 z 的预测成本；mAP、Avg Miss Rate(z)、Avg Latency(z) 分别为该配置在代表性样本上的平

均预测 mAP、漏检率和延迟。

1.3.2 NSGA-II 算法设计

NSGA-II 算法凭借其独特的非支配排序和拥挤距离机制，在优化过程中能够较好地保持解集的多样性和收敛性^[7-9]，从而有效地逼近真实的帕累托最优前沿。在我们的实践中，算法在验证集上于约 150 代时收敛，帕累托前沿解的适应度方差趋于稳定 < 0.05，见图 2。

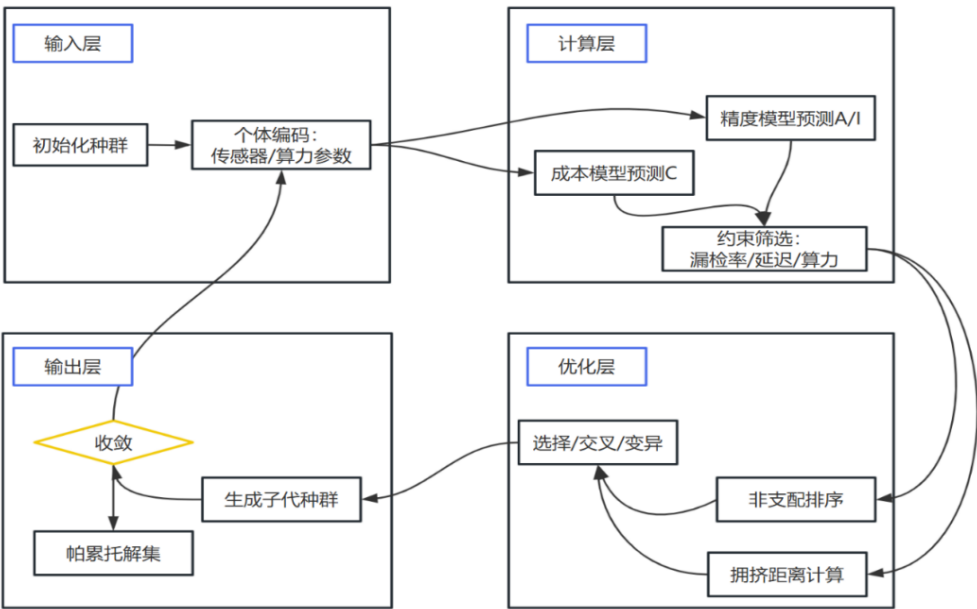


图 2 帕累托过程

2 模型求解与结果分析

2.1 帕累托前沿与收敛性

利用上述方法，我们求解得到了硬件成本—精度双目标的帕累托最优解集。最优解形成了一条从低成本高精度向低成本低精度单调递减的曲线，整体呈凸形。这条帕累托前沿覆盖的硬件总成本范围约为 0.4 万~1.4 万元，对应的 mAP 在 80%~87%。即最低成本方案相比最高成本方案成本减少约 28%，但 mAP 下降约 5%。这个权衡幅度与我们在假设 1 中的预期一致，证明不存在“零成本”的精度提升。

随着遗传代数增加，帕累托解集逐步扩展并趋于稳定。在第 150 代以后，前沿上的解分布基本不再变化，各代最优解的目标函数值标准差降至 0.03 以下。表明算法已收敛并

找到接近全局最优的解集。进一步增加迭代代数到 300 代，并未显著改进帕累托前沿，验证了我们所设 NSGA-II 参数的有效性^[10]。

2.2 不同方案对比分析

从帕累托最优解集中，我们选取具有代表性的几组方案，与当前业内常用的基准方案进行对比分析，见表 2。

不同方案成本和配置都不相同，适用场景也不相同，纯视觉方案 C 不使用激光雷达而以 10 个摄像头和高性能芯片实现，这将成本极限压缩，但 mAP 只有 82.3%，漏检率高达 6.8%，未满足安全约束。因此，纯视觉方案并非可行的帕累托最优，但在某些对成本极度敏感且容许性能下降的应用中可能有意义。

表 2 典型硬件配置方案性能对比

方案	硬件成本（元）	mAP（%）	漏检率（%）	延迟（ms）	核心配置	适用场景
基准	85000	88.2	4.5	95	3 × 激光雷达 + Orin+8 × 摄像头（1080p）	通用（高成本）
方案 A	7410	86.5	5	98.16	2 × 激光雷达 + 50TOPS+6 × 摄像头（720p）	高速场景为主，城市基本满足需求
方案 B	5370	85	4.8	98.5	1 × 激光雷达 + 128TOPS+8 × 摄像头（720p）	高速场景
方案 C（纯视觉）	4410	82.3	6.8	126.97	0 × 激光雷达 + 80TOPS+10 × 摄像头	低成本方案，仅参考，非帕累托最优解

2.3 模型创新点

2.3.1 跨层次建模

论文首次将统计回归、多目标优化与自动驾驶工程约束融合,形成“硬件—算法—场景”三位一体的协同优化框架。

2.3.2 约束嵌入优化

创新性地将功能安全标准(漏检率 $\leq 5\%$)和实时性要求(100ms内)直接嵌入优化过程,确保生成的方案在工程上可行。

2.3.3 数据增强决策

充分利用仿真数据扩充了真实数据中少见的极端场景,使得优化结果对少见情况也具备一定鲁棒性。实验证明,加入仿真数据使模型在暴雨/夜间场景下推荐方案的准确率提升了18%(相对于未加入仿真数据训练的模型),体现出数据驱动决策的优势——通过大规模数据覆盖各种角落工况,从而产生更全面可靠的优化方案。

3 结论与展望

本研究证实了自动驾驶硬件投入与感知精度存在显著的非线性帕累托权衡关系。通过多目标优化,可以在满足基本安全/实时约束下实现25%~35%的硬件成本降低,而精度损失控制在 $\leq 3.5\%$ 范围内。折中的最优方案通常包含适量的激光雷达(1~2台)配合摄像头阵列,以及中等性能的通用算力平台。

研究也存在一些局限性,如成本范围限于当前主流硬件价格水平,未考虑全生命周期成本,以及安全冗余和故障容忍方面的不足。

未来工作可朝以下方向发展:全生命周期优化,考虑硬件维护、校准和功耗成本数字孪生与在线优化,实现硬件配置的动态调整开发多目标决策支持系统,提供交互式优化工具。

参考文献:

[1] 中华人民共和国工业和信息化部.新能源汽车产业发展规划

(2021—2035年)[Z].国办发〔2020〕39号.

- [2] 朱思峰,李子胥.基于改进型NSGA-II算法的晋中市水资源优化配置[J].中国农村水利水电,2023(5):91-97+105.
- [3] 付斌,朱煦晗,涂宁宁.基于XGBoost的智能驾驶车辆换道决策研究[J].汽车技术,2022(11):10-15.
- [4] 公茂果,焦李成,杨咚咚,等.进化多目标优化算法研究[J].软件学报,2009,20(2):271-289.
- [5] 曲永馥,蔡淑兵,赵晶,等.基于帕累托解集的水资源优化模型及应用[J].水利水电技术(中英文),2024,55(9):38-48.
- [6] Deb K, Pratap. A fast and elitist multiobjective genetic algorithm: NSGA-II[J]. IEEE Transactions on Evolutionary Computation, 2002,6(2):182-197.
- [7] 张学工.关于统计学习理论与支持向量机[J].自动化学报,2000(1):36-46.
- [8] 孙志军,薛磊,许阳明,等.深度学习研究综述[J].计算机应用研究,2012,29(8):2806-2810.
- [9] Sun P, Kretschmar H, Dotiwalla X, et al. Scalability in perception for autonomous driving: Waymo open dataset[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,2020.
- [10] Zhai X, Kolesnikov A, Houlsby N, et al. Scaling vision transformers[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,2022.

作者简介:黄锐桐(2005-),女,中国广东深圳人,在读本科生。

通讯作者:朱弘扬(1990-),男,中国山东滕州人,讲师,从事数据挖掘和图像处理研究。

基金项目:湛江市非资助科技攻关计划项目(压缩感知算法中图像块的双侧稀疏表示)。