

DOI:10.13718/j.cnki.xsxb.2022.12.002

基于机器学习对串联排队系统等待时间的预测^①

卫安妮, 赵宁, 张志坚

昆明理工大学 理学院, 昆明 650500

摘要: 串联排队系统是排队网络的基本结构, 研究串联排队系统对分析排队网络具有重要意义. 串联排队系统中站与站之间存在关联性, 上游站的输出过程是下游站的输入过程, 对于不满足马尔可夫性的排队系统, 下游站的到达过程很难用解析的方法分析. 对于一般的串联排队系统, 本文基于机器学习对串联排队系统的平均等待时间进行预测, 通过数值实验比较机器学习中线性回归模型和非线性回归模型的预测效果. 实验结果表明, 非线性回归模型优于线性回归模型, XGBoost 算法对串联排队系统的平均等待时间的预测准确度较高. 此外, 本文将 XGBoost 算法与传统的近似分析方法进行比较, 发现 XGBoost 算法的预测效果优于传统的近似分析方法.

关键词: 串联排队系统; 机器学习; 仿真模拟; 平均等待时间; XGBoost 算法

中图分类号: TP181

文献标志码: A

文章编号: 1000-5471(2022)12-0011-11

Prediction of Waiting Time in Tandem Queueing Systems Based on Machine Learning

WEI Annie, ZHAO Ning, ZHANG Zhijian

College of Science, Kunming University of Science and Technology, Kunming 650500, China

Abstract: Tandem queueing system is the basic structure of queueing network, and it is important to study the tandem queueing system to analyze the queueing network. There are correlations between stations in a tandem queueing system, and the output process of the upstream station is the input process of the downstream station. For queueing systems that do not satisfy Markovianity, the arrival process of the downstream station is difficult to analyze by the analytical method. For the general tandem queueing system, this paper proposes to predict the mean waiting time of the system based on machine learning, and compare the prediction effect of linear regression model and nonlinear regression model in machine learning through numerical experiments. The experimental results show that the nonlinear regression model outperforms the linear regression model, and the XGBoost algorithm has a higher accuracy in predicting the mean waiting time of the tandem queueing system. In addition, this paper compares the XGBoost algorithm with the traditional approximate analysis method and finds that the prediction effect of the XGBoost algorithm

① 收稿日期: 2022-08-04

基金项目: 2021 年度工业控制技术国家重点实验室开放课题(ICT2021B51).

作者简介: 卫安妮, 硕士研究生, 主要从事排队论研究.

通信作者: 赵宁, 副教授, 博士研究生导师.

is better than that of the traditional approximate analysis method.

Key words: tandem queueing system; machine learning; simulation; mean waiting time; XGBoost algorithm

排队网络模型在流水生产线、交通运输、计算机通信等领域应用十分广泛,吸引了众多学者的关注.串联排队系统是排队网络的基本结构^[1],顾客在一个站接受服务后按照一定的规则接受下一个站的服务,研究该系统对深入分析复杂的排队网络具有重要意义.

串联排队系统的研究最早可追溯到 20 世纪 50 年代,文献[2-5]研究了具有马尔可夫性的串联排队系统的平均等待时间等性能指标.随后,关于满足马尔可夫性的串联排队系统得到了广泛研究.然而,实际生活中排队系统一般不满足马尔可夫性,这导致串联排队系统的性能很难用解析的方法来求解,通常使用近似方法进行分析.文献[6]提出了排队网络分析方法(queueing network analysis, QNA)研究不满足马尔可夫性的串联排队系统.文献[7]利用到达过程和服务时间的一阶矩和二阶矩的近似提出了排队网络方法(queueing network, QNET)估计顾客的平均逗留时间.文献[8]基于分解的方法使用联合矩对 MAP/MAP/1 排队网络进行了分析.文献[9]同样基于分解算法提出鲁棒排队网络分析器算法(robust queueing network analyzer, RQNA)近似开排队网络的稳态性能.文献[10]使用固有比的方法近似串联排队系统的平均排队时间.文献[11]采用指标比研究 M/G/1-G/1 串联排队系统的平均等待时间.文献[12]提出三阶近似的方法分析 GI/G/1-G/1 串联排队系统的平均等待时间.文献[13]基于泛函重对数律和重对数律极限的方法,分析 GI/G/1-G/1 串联排队系统的性能指标的波动程度.

近年来,基于机器学习分析排队系统引起一些学者的关注.文献[14]利用支持向量机(support vector machine, SVM)对排队系统中到达和服务时间的概率密度函数进行分类和识别,并通过支持向量回归(support vector regression, SVR)解决概率密度函数回归的问题.文献[15]使用机器学习的方法对患者的治疗数据进行预测,根据预测的治疗时间推断其等待时间,结果表明随机森林模型为每日治疗时间提供了最佳的预测.文献[16]使用分位数、普通最小二乘(ordinary least square, OLS)回归以及机器学习算法对某医院患者的平均等待时间进行预测,结果表明套索回归(lasso regression, Lasso)和分位数回归方法的准确率更高.文献[17]使用交通模拟器对神经网络进行训练,得到一个自适应交通系统.文献[18]使用神经网络方法对银行排队的等待时间进行预测,证明机器学习是预测排队等待时间的一种可行方法.文献[19-20]使用高斯过程回归预测单服务器和多服务器排队网络的平均逗留时间.

在日常生活中,串联排队系统广泛存在于生产系统等领域.串联排队系统中站与站之间存在关联性,上游站的输出过程是下游站的输入过程,对于不满足马尔可夫性的排队系统,下游站的到达过程很难用解析的方法分析.本文考虑具有两个站的串联排队系统,其到达过程和服务时间均服从一般分布,通过模拟串联排队系统的平均等待时间生成训练集,使用机器学习预测一般串联排队系统的平均等待时间,并与近似方法进行比较.

本文结构如下:第 1 节描述了两个站的串联排队模型;第 2 节介绍了常见的线性和非线性机器学习回归算法;第 3 节利用机器学习的方法预测串联排队系统的平均等待时间;第 4 节将机器学习中的 XGBoost 算法与其他近似方法进行比较;第 5 节为结论.

1 模型描述

本文研究图 1 所示的串联排队系统,该系统由两个站串联而成,每个站有一个服务器,并且服务器前的缓冲区无限大.顾客的到达过程为更新过程,顾客到达系统后依次在每个站接受服务,服务完成后离开系统.系统的服务规则为先到先服务(first come first served, FCFS),每个服务器的服务时间服从一般分布.

假设系统的到达率为 λ_a ,相邻顾客到达时间间隔为 X , X 服从一般分布.第 1 个站和第 2 个站的服务器的服务时间分别记为 S_1 和 S_2 . X , S_1 和 S_2 的平方变异系数分别为 C_a^2 , $C_{S_1}^2$ 和 $C_{S_2}^2$, 且

$$C_a^2 = \frac{D(X)}{E^2(X)}$$

$$C_{S_1}^2 = \frac{D(S_1)}{E^2(S_1)}$$

$$C_{S_2}^2 = \frac{D(S_2)}{E^2(S_2)}$$

其中 $E(X)$ 和 $D(X)$ 分别表示到达时间间隔的期望和方差, $E(S_i)$ 和 $D(S_i)$, $i=1,2$ 分别表示第 1 个站和第 2 个站的服务器的服务时间的期望和方差. 串联排队系统中第 1 个站和第 2 个站的服务强度分别记为 ρ_1 和 ρ_2 , 且

$$\rho_1 = \lambda_a E(S_1)$$

$$\rho_2 = \lambda_a E(S_2)$$

令 $\rho = \max\{\rho_1, \rho_2\}$. 串联排队系统中第 1 个站和第 2 个站的平均排队时间分别记为 W_1 和 W_2 .

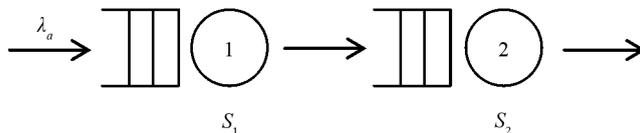


图 1 两个站的串联排队系统

2 机器学习回归算法

近年来, 机器学习快速发展, 广泛应用于数据挖掘、人工智能、医疗保健、排队等领域. 与传统回归方法相比, 机器学习能够分析和挖掘数据中的规律, 并对新的样本进行预测, 适合处理复杂的回归问题. 下面介绍机器学习中常见的回归算法.

2.1 线性回归模型

机器学习中常见的线性回归模型为多元线性回归 (multiple linear regression, MLR)、岭回归 (ridge regression, Ridge) 以及套索回归 (lasso regression, Lasso). 线性回归模型属于一种监督学习算法, 研究两个随机变量之间的线性关系. 该模型可表示为

$$Y = X\beta + \epsilon$$

其中: X 表示线性回归模型的自变量集合, Y 表示线性回归模型的因变量, β 表示偏回归系数, ϵ 表示模型拟合后每一个样本的误差项.

为了求解线性回归模型的参数, 将该模型的目标函数表示为^[16]

$$J(\beta) = \sum (Y - X\beta)^2 + \lambda[\alpha \|\beta\|_1 + (1 - \alpha) \|\beta\|_2^2]$$

其中: λ 为目标函数的惩罚系数, $\|\beta\|_1$ 和 $\|\beta\|_2^2$ 分别表示回归系数 β 的 l_1 和 l_2 正则项. 该目标函数由最小二乘误差平方和与正则项组成. 当 $\lambda=0$ 时, 该目标函数退化为 MLR 的目标函数; 当 $\lambda \neq 0, \alpha=0$ 时, 该目标函数退化为 Ridge 的目标函数; 当 $\lambda \neq 0, \alpha=1$ 时, 该目标函数退化为 Lasso 的目标函数. 对于 Ridge 和 Lasso 回归而言, 寻找合理的 λ 值是平衡模型方差和偏差的关键. MLR 的目标函数通过最小二乘误差平方和的方法获得; Ridge 是一种替代最小二乘的压缩估计拟合方法, 通过正则化减少方差, 能够将系数向零的方向进行压缩, 在 MLR 中实现变量重要性的筛选; Lasso 也是通过正则化减少方差, 与 Ridge 的差异在于 Lasso 可以将系数压缩至零, 能够实现变量的筛选, 得到输出参数较少的稀疏模型^[21].

2.2 非线性回归模型

对于比较复杂的非线性回归模型, 需要在因变量和多个自变量之间构建复杂的非线性关系. 机器学习的非线性回归算法主要包括 K 近邻 (k-nearest neighbor, KNN)、支持向量机 (support vector machine, SVM)、决策树 (decision tree, DT)、随机森林 (random forest, RF)、梯度提升树 (gradient boosting decision tree, GBDT) 以及极端梯度提升 (extreme gradient boosting, XGBoost) 算法. 本文将以上非线性回归

算法分为 3 类: 递归划分方法、黑箱方法和集成学习方法^[22].

递归划分方法主要包括决策树(DT)算法. 该算法按照一定的规则持续拆分数据, 每次将数据划分为两个相对一致的子集, 直到达到目标, 从而形成树状结构, 直观反映变量的重要性, 但该算法结构不稳定, 容易产生过拟合的现象.

黑箱方法包括 K 近邻算法(KNN)以及支持向量机(SVM)算法. 这类算法的输入到输出过程是通过一个模糊的“箱子”进行处理. KNN 通过比较已知样本和预测样本的相似度, 寻找最相似的 k 个样本作为未知样本的预测. 采用多重交叉验证法选取最佳 k 值. SVM 利用某些支持向量构成的“超平面”, 将不同类别的样本点进行划分, SVM 算法与其他单一的算法相比, 能够将低维不可分的空间转化为高维的线性可分空间, 具有较高的预测准确性, 但其最大的缺点是容易受共线性影响, 运算成本高. 这类方法对数据缺失较敏感, 处理大规模数据的效率较低.

集成学习方法通过选择某种结合策略将若干弱学习器集合起来, 以得到一个预测效果较好的强学习器. 随机森林(RF)、梯度提升树(GBDT)以及极端梯度提升(XGBoost)算法是一类以决策树(DT)为基学习器的集成学习算法. RF 采用多棵决策树的投票机制, 即将多棵树的回归结果进行平均, 最终得到样本的预测值. 类似的, GBDT 也是通过对多棵树的结果进行综合, 不同的是每棵树是从之前所有树的残差中学习的, 并以新树每个叶子的信息增益来进行最后的全局预测. XGBoost 采用了随机森林的思想, 作为升级版的 GBDT 算法, XGBoost 使用损失函数的一阶导和二阶导作为残差的近似值, 而 GBDT 仅利用损失函数的一阶导作为残差的近似值. 集成学习方法通常优于单一的回归方法, 但预测速度明显下降, 随着学习器数目的增加, 所需的存储空间也急剧增加^[23].

通常采用线性回归模型以及非线性回归模型进行预测时, 需要将不同模型的运行时间成本和准确率进行对比分析, 从中选择合理的模型进行预测. 本文将准确率作为衡量标准, 选择较优的模型对串联排队系统的平均等待时间进行预测.

3 基于机器学习预测平均等待时间

3.1 训练集数据

为了生成机器学习所需的训练集数据, 首先对串联排队系统进行模拟, 得到不同参数下串联排队系统的平均等待时间.

将参数 $C_a^2, E(S_1), E(S_2), C_{S_1}^2, C_{S_2}^2, \rho$ 作为模型的输入参数, 使用 MATLAB 对图 1 所示的串联排队系统进行模拟, 其中参数取值如表 1 所示. 为了生成机器学习所需的训练集数据, 使用表 1 所列的参数范围, 模拟了 11 494 种不同参数组合下的串联排队系统, 其中相邻顾客的到达时间间隔和每个站的服务时间均服从伽马分布.

表 1 训练集的参数

参数	范围
C_a^2	{0.3, 0.5, 0.8, 1, 1.5, 2, 3, 5, 10}
$C_{S_1}^2$	{0.1, 0.2, ..., 1} ∪ {1.5, 1.6, ..., 3} ∪ {8}
$C_{S_2}^2$	{0.5, 0.9, 1, 1.5, 3}
$E(S_1)$	{5, 6, ..., 15} ∪ {20, 25, 26, 30}
$E(S_2)$	{5, 6, ..., 15} ∪ {20, 25, 27, 30}
ρ	{0.1, 0.2, ..., 0.9}

对于每个串联排队系统, 模拟运行 30 个样本, 每个样本取第 400 001 个顾客到 600 000 个顾客在第 1 个站和第 2 个站的等待时间的平均值分别作为第 1 个站和第 2 个站的平均等待时间, 其中平均等待时间的置信水平大于 95%, 保证了模拟数据的可靠性.

3.2 数据预处理

在对训练集数据进行建模分析之前, 需要对数据进行预处理以满足模型的要求. 对于等待时间非常小的数据, 即使预测值只有微小的偏差, 其相对误差也会非常大, 造成整体的平均误差偏大, 影响预测效果. 因此, 本文选取第 1 个站和第 2 个站的平均等待时间均大于 0.1 的数据进行分析. 从图 2 和图 3 的分布来看, 模拟得到的串联排队系统的第 1 个站和第 2 个站的平均等待时间均呈现严重的右偏现象, 为了满足机器学习模型的数据要求, 提高模型精度和训练效率, 本文对第 1 个站的平均等待时间 W_1 和第 2 个站的平均等待时间 W_2 分别进行对数处理, 令

$$W'_i = \log(W_i + 1), i = 1, 2$$

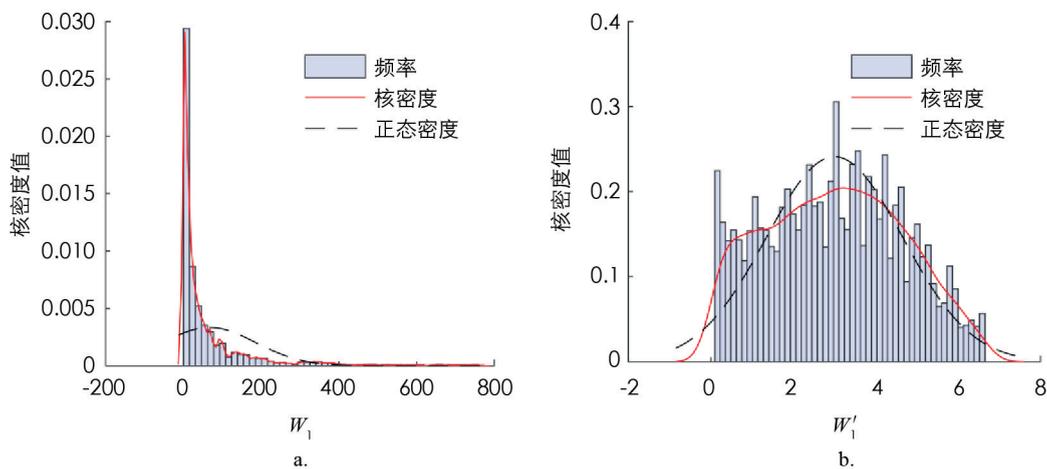


图 2 第 1 个站的平均等待时间 W_1 及其对数变换 W'_1

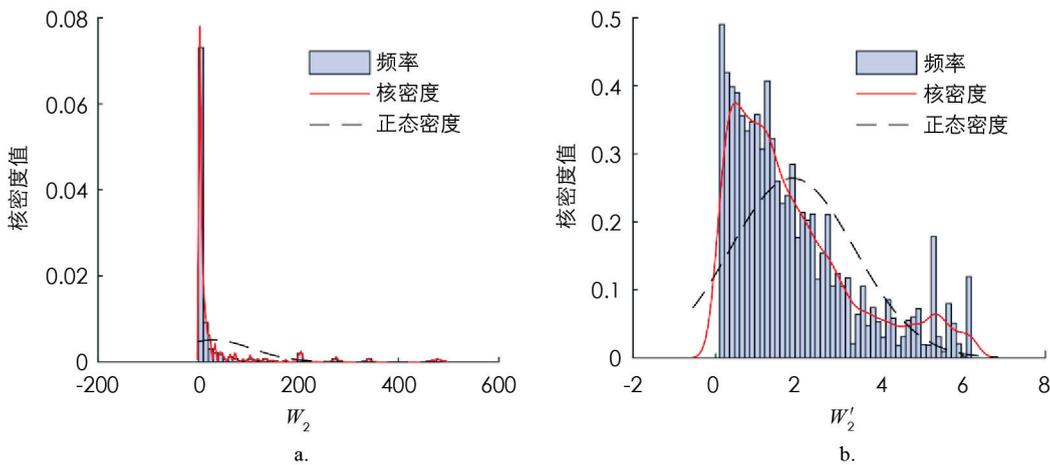


图 3 第 2 个站的平均等待时间 W_2 及其对数变换 W'_2

计算模拟得到的数值 W_i 和 $W'_i (i=1, 2)$ 的核密度如图 2, 3 所示. 从图中可知经过对数处理后, W'_i 的分布近似服从正态分布, W'_2 的右偏现象明显有所缓解, $W'_i, i=1, 2$ 的值域均缩小到 $[-2, 8]$ 之间, 采用 W'_i 加快了梯度下降求最优解的速度, 即机器学习训练的速度.

3.3 基于线性回归模型预测平均等待时间

本文采用 python 的 sklearn 程序包构建模型, 将数据集按照 85% 和 15% 的比例随机分为训练集和测试集. 将具有两个站的串联排队系统的到达时间的平方变异系数 C_a^2 、平均服务时间 $E(S_i)$ 、服务时间的平方变异系数 C_s^2 以及瓶颈站的服务强度 ρ 作为特征, 分别预测第 1 个站的平均等待时间 W_1 和第 2 个站的平均等待时间 W_2 .

对于线性回归模型, 即多元线性回归(MLR)、岭回归(Ridge)以及套索回归(Lasso), 利用 10 重交叉验证的方法确定最佳的 λ 值, 如表 2 所示. 基于最佳 λ 值使用训练集学习相应的线性回归模型, 并对测试集进

行预测. 基于线性回归模型预测的第 1 个站和第 2 个站的平均等待时间以及其模拟值如图 4,5 所示, 其中 W_i , $i=1,2$ 表示第 i 个站平均等待时间的模拟值, W_i^* , $i=1,2$ 表示第 i 个站平均等待时间的预测值.

表 2 线性回归模型的最佳 λ 值

线性回归模型	λ 值		
	MLR	Ridge	Lasso
W_1	0	0.000 14	0.000 01
W_2	0	0.000 06	0.000 01

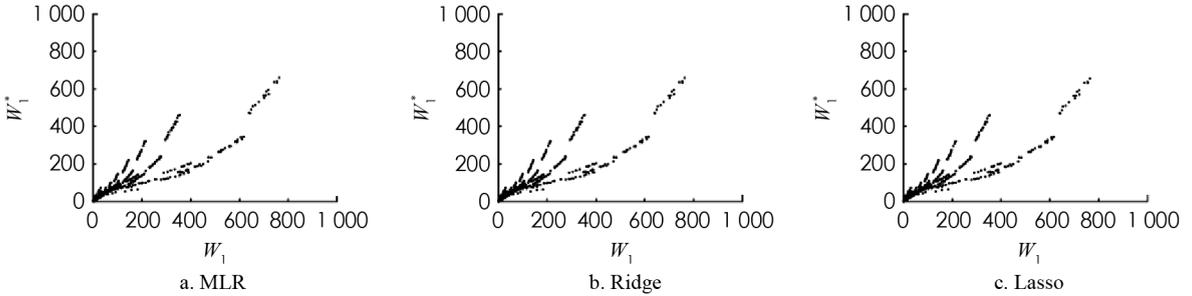


图 4 基于线性回归模型对第 1 个站平均等待时间的预测

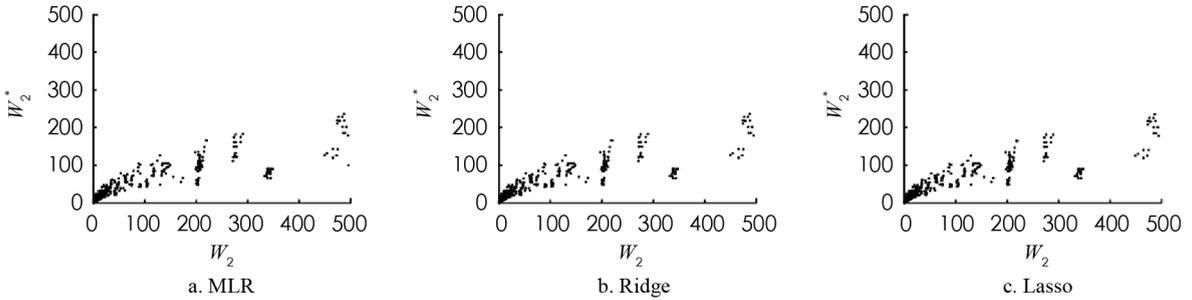


图 5 基于线性回归模型对第 2 个站平均等待时间的预测

由图 4,5 可知, MLR, Ridge 和 Lasso 回归在 W_i , $i=1,2$ 较小时的预测较准确, 但对于平均等待时间较长的情况预测不准确. 为了定量的评估模型, 使用平均相对误差 $r = \frac{|W_2 - W_2^*|}{W_2}$ 来衡量模型的准确性,

MLR, Ridge 和 Lasso 回归对于第 1 个站的平均等待时间预测的平均相对误差分别为 65.02%, 65.03%, 65.04%; 对于第 2 个站的平均等待时间预测的平均相对误差分别为 80.57%, 80.57%, 80.51%. 由此可知, 3 种线性回归模型对于串联排队系统的平均等待时间预测误差较大. 由于 MLR, Ridge 和 Lasso 回归均为线性回归模型, 串联排队系统中顾客的平均等待时间与顾客的到达时间分布、服务器的服务时间分布等因素有着复杂的非线性关系, 因此, 使用线性回归对平均等待时间预测的效果不佳.

3.4 基于非线性回归模型预测平均等待时间

下面使用非线性回归模型对串联排队系统的平均等待时间进行预测, 其中包括 K 近邻(KNN)、支持向量机(SVM)、决策树(DT)、随机森林(RF)、梯度提升树(GBDT)以及极端梯度提升(XGBoost)算法. 大多数机器学习算法都需要对参数进行设置, 参数设置不同, 学习得到的模型性能往往有显著的差异. 因此, 使用非线性回归模型对平均等待时间进行预测时, 需要对参数调优. 为了提高参数优化的效率, 使用 10 重交叉验证的方法调整参数, 并通过网格搜索找到最佳的参数组合. 各模型参数的范围和最优值如表 3 所示, 其中参数最优值 (x_1, x_2) 中 x_1 和 x_2 分别表示预测 W_1 和 W_2 所设定的模型参数.

对于上述非线性回归模型, 使用最优的参数组合对训练集数据进行学习, 并对测试集进行预测. 非线性回归模型预测的第 1 个站和第 2 个站的平均等待时间及其模拟值如图 6,7 所示. DT, RF, GBDT 以及 XGBoost 算法的预测效果明显优于 KNN 和 SVM 算法, 这是由于 RF, GBDT 和 XGBoost 是基于 DT 的集成学习算法, 结合了对所有弱学习器的预测, 优于单一的学习器.

表 3 模型的参数范围和最优值

非线性回归模型	参数	范围	最优值(x_1, x_2)
KNN	K	{1, 2, ..., 13}	(2, 1)
	Kernel	{'rbf', 'linear', 'poly', 'sigmoid'}	('rbf', 'rbf')
SVM	epsilon	{0.1, 0.3, 0.5, 0.7, 1.1, 1.3, 1.5}	(0.1, 0.1)
	C	{100, 300, 500, 700, 900, 1000}	(900, 900)
	gamma	{0.001, 0.002, ..., 0.01}	(0.009, 0.009)
	max_depth	{19, 20, 21, 22, 23, 25, 27, 29}	(22, 23)
DT	min_samples_split	{1, 2, 3, 4, 5, 6, 8, 10}	(1, 4)
	min_samples_leaf	{1, 2, 3, 4, 5, 8, 10, 12}	(2, 2)
RF	n_estimators	{100, 150, 200, ..., 500} ∪ {1000}	(350, 200)
	learning_rate	{'ls', 'lad', 'huber'}	('ls', 'ls')
GBDT	loss	{0.01, ..., 0.05} ∪ {0.1, 0.15, 0.2, ..., 0.7}	(0.4, 0.5)
	n_estimators	{100, 200, ..., 1800}	(1500, 800)
	learning_rate	{0.01, ..., 0.05} ∪ {0.1, 0.15, 0.2, ..., 0.9}	(0.1, 0.2)
XGBoost	max_depth	{1, 2, ..., 10}	(5, 5)
	n_estimators	{100, 200, ..., 1000}	(1000, 900)

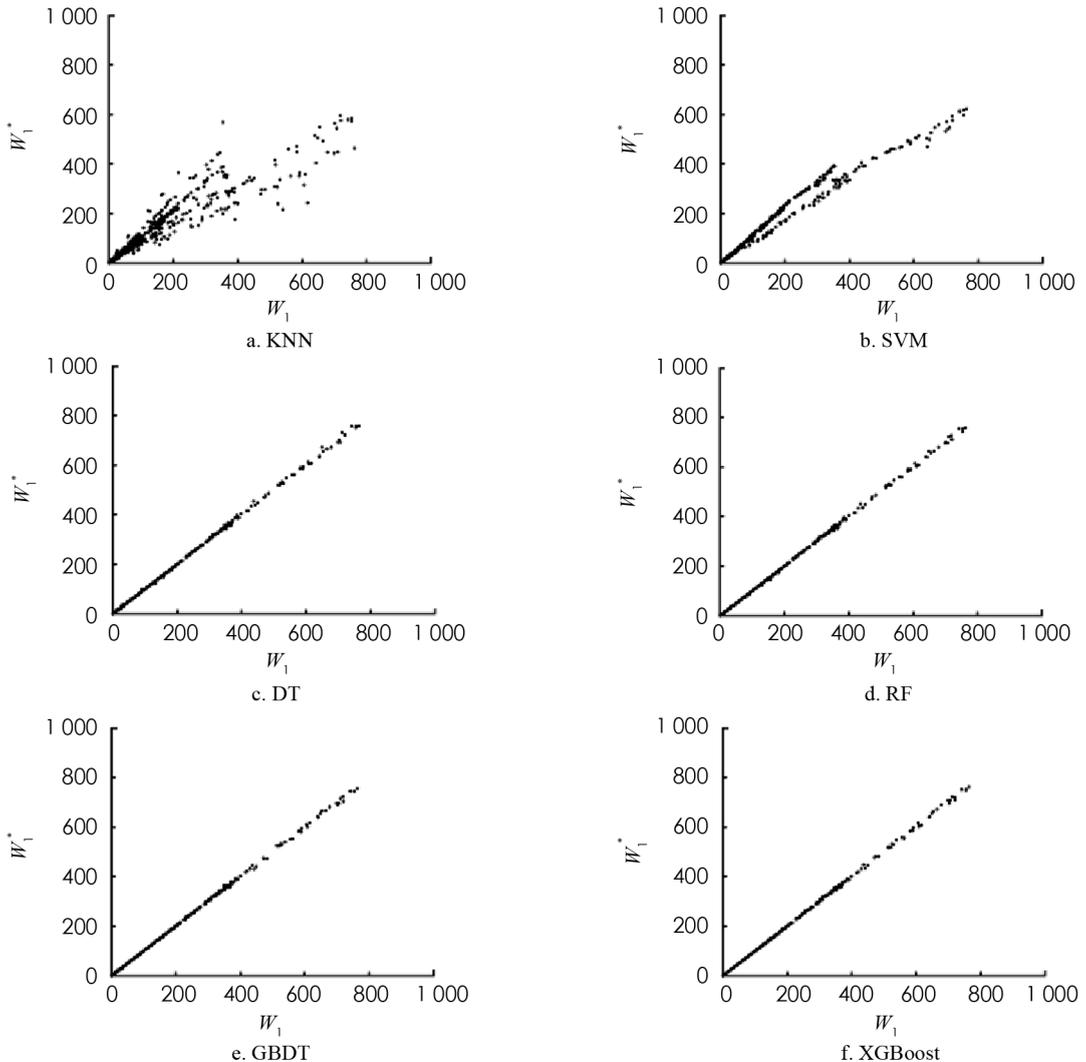


图 6 基于非线性回归模型对第 1 个站平均等待时间的预测

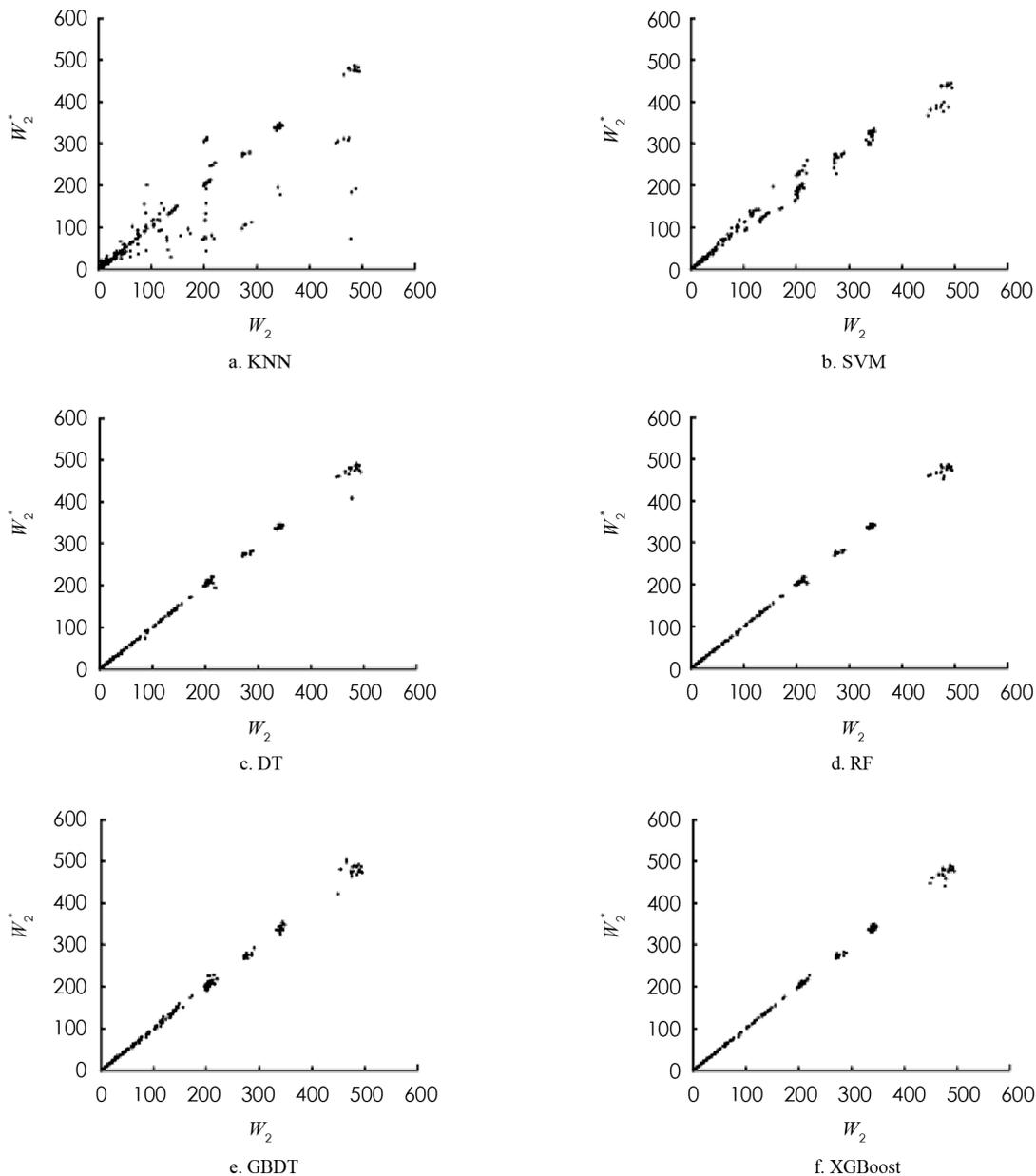


图 7 基于非线性回归模型对第 2 个站平均等待时间的预测

为了比较 DT, RF, GBDT 以及 XGBoost 算法的预测效果, 将平均相对误差 r 作为模型的评价标准, 4 种模型的第 1 个站的平均等待时间的平均相对误差分别为 3.42%, 2.33%, 2.79%, 1.60%, 第 2 个站的平均等待时间的平均相对误差分别为 8.34%, 4.13%, 3.34%, 1.86%。实验结果表明, DT 算法对于第 2 个站的平均等待时间的预测效果较差, XGBoost 算法对第 1 个站和第 2 个站的平均等待时间的预测准确率分别为 98.40% 和 98.14%, 对串联排队系统的平均等待时间的预测较优。

4 机器学习与近似方法的比较

目前, 对于到达过程为更新过程, 服务时间服从一般分布的排队系统的平均等待时间的研究均采用近似方法. 文献[24]研究了 GI/G/1 排队系统的平均等待时间. 文献[6]基于文献[24]的方法使用排队网络分析方法(QNA)研究了具有非马尔可夫性的串联排队系统的平均等待时间. 基于布朗运动, 文献[7]利用一阶矩和二阶矩的近似方法提出了使用 QNET 方法估计串联排队系统中顾客的平均逗留时间。

为了验证本文方法的有效性, 下面分别对 GI/G/1-G/1 系统以及 M/G/1-G/1 系统的平均等待时间的误差进行比较.

系统 1: 在 GI/G/1-G/1 串联排队系统中, 假设相邻顾客到达时间间隔服从伽马分布, $X \sim Gamma\left(\frac{1}{5}, \frac{150}{\rho}\right)$, 第 1 个站和第 2 个站的服务时间分别为 $S_1 \sim Gamma\left(\frac{1}{8}, 200\right)$, $S_2 \sim Gamma(2, 15)$, 令 $\rho = \{0.1, 0.2, \dots, 0.9\}$.

系统 2: 在 M/G/1-G/1 串联排队系统中, 假设到达过程为泊松过程, 相邻顾客到达时间间隔服从指数分布 $X \sim \exp\left(\frac{\rho}{30}\right)$, 第 1 个站和第 2 个站的服务时间分别为 $S_1 \sim Gamma(2, 10)$, $S_2 \sim Gamma\left(\frac{10}{9}, 27\right)$, 令 $\rho = \{0.1, 0.2, \dots, 0.9\}$.

对于上述串联排队系统, 分别采用 Kingman 方法以及本文提出的 XGBoost 方法对第 1 个站的平均等待时间进行预测; 使用 QNA, QNET 以及本文提出的 XGBoost 方法对第 2 个站的平均等待时间进行预测, 各种方法预测的相对误差如表 4,5 所示.

表 4 系统 1 中两个站的平均等待时间的相对误差比较

ρ	W_1		W_2		
	Kingman	XGBoost	QNA	QNET	XGBoost
0.1	63.08%	0.39%	67.70%	57.10%	3.03%
0.2	47.83%	0.79%	53.38%	38.91%	0.90%
0.3	37.47%	0.80%	43.25%	27.21%	0.13%
0.4	29.67%	0.09%	35.16%	19.23%	0.19%
0.5	23.58%	0.24%	27.84%	13.35%	0.04%
0.6	18.56%	0.52%	21.01%	9.27%	0.50%
0.7	14.41%	0.81%	13.97%	6.27%	0.25%
0.8	10.36%	0.05%	5.14%	3.05%	0.00%
0.9	7.27%	0.17%	6.92%	0.42%	0.00%
平均误差	28.02%	0.43%	30.49%	19.42%	0.56%

表 5 系统 2 中第 2 个站的平均等待时间的相对误差比较

ρ	QNA 方法	QNET 方法	XGBoost 方法
0.1	7.01%	4.42%	1.28%
0.2	6.78%	4.00%	3.01%
0.3	5.96%	3.83%	1.07%
0.4	4.80%	3.71%	0.43%
0.5	3.51%	3.41%	0.18%
0.6	1.79%	3.20%	0.41%
0.7	0.59%	3.25%	0.65%
0.8	3.61%	3.33%	0.36%
0.9	7.20%	2.75%	0.11%
平均误差	4.58%	3.55%	0.83%

由表 4 可知, 在不同的繁忙程度下, 对于第 1 个站的平均等待时间, XGBoost 方法、Kingman 方法的

平均相对误差分别为 0.43%, 28.02%。Kingman 方法是对平均等待时间上限的近似分析, 其预测效果比 XGBoost 方法差。

对于第 2 个站的平均排队时间, XGBoost 方法、QNA 方法以及 QNET 方法的平均相对误差分别为 0.56%, 30.49% 以及 19.42%。相比于其他方法, XGBoost 方法的相对误差最小且平均误差均小于 1%。在繁忙程度 ρ 较小时, 顾客的平均等待时间较短, 相对误差较大。由此可知, 本文提出的 XGBoost 方法明显优于其他方法, 并且在繁忙程度 ρ 较大时, 预测效果最佳。

在 M/G/1-G/1 排队系统中, 第 1 个站的平均等待时间存在精确解析表达式, 因此, 仅对第 2 个站的平均等待时间的相对误差进行比较。QNA 方法、QNET 方法均通过考虑离去过程的一阶矩和二阶矩来刻画串联排队系统中第 1 个站对第 2 个站的影响。虽然这些方法很容易计算平均等待时间的近似值, 但是其计算的精确度不高。由表 5 可知, 当 $\rho = \{0.1, 0.2, 0.3, 0.5, 0.6, 0.8, 0.9\}$ 时, 本文提出的 XGBoost 方法相对误差最小, XGBoost 方法、QNA 以及 QNET 方法的平均相对误差分别为 0.83%, 4.58% 以及 3.55%。XGBoost 方法的平均相对误差最小, QNET 方法优于 QNA 方法, 这是由于 QNA 方法在平方变异系数较大的情况下, 参数分解方法的性能下降导致预测效果不佳^[7]。综上所述, XGBoost 方法优于其他方法, 近似效果较好, 能够比较准确地计算串联排队系统的平均等待时间。

5 结论

本文采用机器学习中的线性回归算法和非线性回归算法预测串联排队系统的平均等待时间。将仿真的通用性与机器学习的计算效率相结合, 提高了平均等待时间预测的准确性。大量的数值实验表明, XGBoost 方法对平均等待时间的预测效果较好。

本文主要研究了两个站的串联排队系统, 未来可以使用该方法对其他排队系统进行深入研究, 例如具有多个服务站的串联排队系统、具有有限缓冲区的串联排队系统、具有批量服务的串联排队系统以及复杂的排队网络等。

参考文献:

- [1] KLIMENOK V, DUDIN A, VISHNEVSKY V. Tandem Queueing System with Correlated Input and Cross-Traffic [M]// Computer Networks. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013: 416-425.
- [2] O'BRIEN G G. The Solution of some Queueing Problems [J]. Journal of the Society for Industrial and Applied Mathematics, 1954, 2(3): 133-142.
- [3] JACKSON R R P. Queueing Systems with Phase Type Service [J]. Journal of the Operational Research Society, 1954, 5(4): 109-120.
- [4] JACKSON R R P. Random Queueing Processes with Phase-Type Service [J]. Journal of the Royal Statistical Society: Series B (Methodological), 1956, 18(1): 129-132.
- [5] BURKE P J. The Output of a Queueing System [J]. Operations Research, 1956, 4(6): 699-704.
- [6] WHITT W. Performance of the Queueing Network Analyzer [J]. Bell System Technical Journal, 1983, 62(9): 2817-2843.
- [7] HARRISON J M, NGUYEN V. The QNET Method for Two-Moment Analysis of Open Queueing Networks [J]. Queueing Systems, 1990, 6(1): 1-32.
- [8] HORVÁTH A. A Joint Moments Based Analysis of Networks of MAP/MAP/1 Queues [J]. Performance Evaluation, 2010, 67(9): 759-778.
- [9] WHITT W, YOU W. A Robust Queueing Network Analyzer Based on Indices of Dispersion [J]. Naval Research Logistics (NRL), 2022, 69(1): 36-56.
- [10] WU K, MCGINNIS L. Interpolation Approximations for Queues in Series [J]. IIE Transactions, 2013, 45(3): 273-290.
- [11] 吴登磊, 赵宁, 刘文奇. 基于指标比对串联排队系统平均排队时间的近似方法 [J]. 南京航空航天大学学报, 2020,

52(4): 644-649.

- [12] 侯佳辰, 赵宁, 刘文奇, 等. 串联排队系统平均等待时间的近似分析 [J]. 山西大学学报(自然科学版), 2022, 45(1): 41-49.
- [13] GUO Y J. Asymptotic Variability Analysis for a Two-Stage Tandem Queue, Part I: The Functional Law of the Iterated Logarithm [J]. *Journal of Mathematical Analysis and Applications*, 2017, 450(2): 1479-1509.
- [14] HU G S, DENG F Q. The Analysis of Queuing System Based on Support Vector Machine [C]//ICARCV 2004 8th Control, Automation, Robotics and Vision Conference. New York: IEEE Press, 2004: 2320-2325.
- [15] JOSEPH A, HERRERA D, HIJAL T, et al. Sci-Fri AM: Quality, Safety, and Professional Issues 04: Predicting Waiting Times in Radiation Oncology Using Machine Learning [J]. *Medical Physics*, 2016, 43(8): 4953.
- [16] PAK A. Predicting Waiting Time to Treatment for Emergency Department Patients [J]. *International Journal of Medical Informatics*, 2021, 145: 104303.
- [17] NATAFGI M B, OSMAN M, HAIDAR A S, et al. Smart Traffic Light System Using Machine Learning [C]//2018 IEEE International Multidisciplinary Conference on Engineering Technology. New York: IEEE Press, 2018: 1-6.
- [18] KYRITSIS A I, DERIAZ M. A Machine Learning Approach to Waiting Time Prediction in Queueing Scenarios [C]//2019 Second International Conference on Artificial Intelligence for Industries (AI4I). New York: IEEE Press, 2019.
- [19] TAN B, KHAYYATI S. Supervised Learning-Based Approximation Method for Single-Server Open Queueing Networks with Correlated Interarrival and Service Times [J]. *International Journal of Production Research*, 2022, 60(22): 6822-6847.
- [20] KHAYYATI S, TAN B. Supervised-Learning-Based Approximation Method for Multi-Server Queueing Networks under Different Service Disciplines with Correlated Interarrival and Service Times [J]. *International Journal of Production Research*, 2022, 60(17): 5176-5200.
- [21] 王岩, 苏子怡, 李晓锋, 等. 基于数据驱动的地下车站能耗预测模型对比研究 [J]. 都市快轨交通, 2022, 35(3): 135-140.
- [22] 雷相东. 机器学习算法在森林生长收获预估中的应用 [J]. 北京林业大学学报, 2019, 41(12): 23-36.
- [23] 杨剑锋, 乔佩蕊, 李永梅, 等. 机器学习分类问题及算法研究综述 [J]. 统计与决策, 2019, 35(6): 36-40.
- [24] KINGMAN J F C. Some Inequalities for the Queue GI/G/1 [J]. *Biometrika*, 1962, 49(3-4): 315-324.

责任编辑 张构