

带有 N 策略的可修重试排队系统的双目标优化

何柳青

燕山大学理学院, 河北 秦皇岛

收稿日期: 2023年3月14日; 录用日期: 2023年4月4日; 发布日期: 2023年4月19日

摘要

本文考虑了一个基于 N 策略下带有预留时间和启动时间的 $M/M/1$ 可修重试排队系统。通过概率母函数法求得系统的稳态概率, 并给出了一些性能指标。考虑了双目标优化问题, 旨在同时最小化成本和期望等待时间, 借助NSGA-II算法来寻找Pareto最优解集。建立二者之间的回归模型, 检验从帕累托最优解集获得的最小成本和期望等待时间的关系。

关键词

重试排队, N 策略, 预留时间, 双目标优化, 回归分析

Bi-Objective Optimization of the Repairable Retrial Queue with N -Policy

Liuqing He

School of Science, Yanshan University, Qinhuangdao Hebei

Received: Mar. 14th, 2023; accepted: Apr. 4th, 2023; published: Apr. 19th, 2023

Abstract

In this paper, we consider the $M/M/1$ retrial queue with a repairable server as well as reserved time and setup times under the N -policy. The stationary probabilities of the system are obtained by the generating function method, and some performance measures are given. Bi-objective optimization problem is considered to minimize the cost and expected waiting time at the same time, and NSGA-II algorithm is used to find the Pareto optimal solution set. The regression model between them is then constructed and the relationship between the minimum cost and the expected waiting time obtained from the Pareto optimal solution set is tested.

Keywords

Retrial Queue, N -Policy, Reserved Time, Bi-Objective Optimization, Regression Analysis

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

若顾客到达时, 服务台处于忙期, 那么他将加入重试轨道等待再次请求服务, 这样的行为被称为重试。20 世纪 50 年代 Cohen [1]首次发表了关于重试排队的文章, 自此以后, 关于重试的研究不断涌现, 有关重试排队的分析可参阅 Falin [2]和 Artalejo [3]等人的著作。

实际生活中, 经常会发生服务台无法工作的情况, 需要维修后才能继续服务顾客, 研究可修排队有着重要的现实意义。Shogan [4]和 Neuts [5]等人分别研究了单服务台和多服务台的可修排队系统。Wang [6]等人将其引入到重试排队系统中, 考虑了 M/M/1 可修重试排队。

在经典的排队系统中, 服务台总是能够提供服务, 而休假排队允许服务台在某些时刻去休假。N 策略就是众多休假策略的一种情况。当系统为空时, 服务台将处于休假状态。只有当队列长度达到 N 时, 服务台才再次开启工作。关于 N 策略休假的研究最早可以追溯到 Kulkarni [7]的文章。有关休假排队的分析还可以参阅田乃硕[8]的著作。

大多数关于排队模型优化的研究都集中在单目标问题上, 然而现实世界中许多优化问题需要同时考虑多个目标函数。关于在排队中考虑多目标优化的问题, 可以参阅 Wu [9], Khodemani-Yazdi [10]和 Hajipour [11]等人的文献。

本文考虑了一个基于 N 策略下带有预留时间和启动时间的 M/M/1 可修重试排队系统, 并且考虑了双目标优化的问题, 其实用性能在诸多领域为系统管理者提供指导。

2. 模型描述

考虑一个基于 N 策略下带有预留时间和启动时间的 M/M/1 可修重试排队系统。顾客到达服从参数为 λ 的泊松流。若服务台空闲, 新到达的顾客立即接受服务。否则, 进入重试轨道等待。排在重试轨道前面的顾客会重新尝试, 直到获得服务为止。服务时间和重试时间分别服从参数为 μ 和 δ 的指数分布。当系统中最后一个服务完成时, 服务台不会立即关闭, 而是会预留一段时间。在此期间, 如果有新的顾客到达, 服务台将工作。否则, 将被关闭。当系统中的顾客数达到 N 时, 服务台将重新启动。预留时间和启动时间分别服从参数为 σ 和 θ 的指数分布。服务台工作时, 可能会发生故障。损坏的服务台立即被送去修复。服务台故障时, 新到达的顾客不会加入系统。服务台寿命和维修时间分别服从参数为 β 和 α 的指数分布。假设该模型所涉及的时间变量都是相互独立的, 如到达时间、维修时间等。

定义 $(N(t), I(t))$ 为时刻 t 系统的状态, 其中 $N(t)$ 表示轨道中的顾客数, $I(t)$ 表示服务台的状态(0: 空闲; 1: 繁忙; 2: 休假; 3: 故障)。到达顾客对系统的状态一无所知, 假设顾客都以概率 q 进入, 那么实际到达率为 $\Lambda = \lambda q$ 。

3. 稳态分析

服务台各状态下的稳态概率为

$$\begin{aligned}
P_0(1) &= \left[\frac{\delta}{\lambda + \delta} - \frac{\mu}{\lambda + \delta} \frac{\lambda\delta + (\lambda + \delta)\sigma N}{(\lambda + \delta)\Lambda - \delta\mu} \right] P_{0,0}, \\
P_1(1) &= \frac{\lambda\delta + (\lambda + \delta)\sigma N}{\delta\mu - (\lambda + \delta)\Lambda} P_{0,0}, \\
P_2(1) &= \sigma \left(\frac{N}{\Lambda} + \frac{1}{\theta} \right) P_{0,0}, \\
P_3(1) &= \frac{\beta}{\alpha} \frac{\lambda\delta + (\lambda + \delta)\sigma N}{\delta\mu - (\lambda + \delta)\Lambda} P_{0,0},
\end{aligned}$$

且

$$\begin{aligned}
P'_0(1) &= \frac{1}{\lambda + \delta} \left\{ \frac{2\mu\sigma N\lambda + \mu N(N-1)\sigma(\lambda + \delta)}{2[\delta\mu - (\lambda + \delta)\Lambda]} + \frac{\mu\Lambda(\lambda + \delta)[\lambda\delta + \sigma N(\lambda + \delta)]}{[\delta\mu - (\lambda + \delta)\Lambda]^2} + \sigma N \right\} P_{0,0}, \\
P'_1(1) &= \left\{ \frac{2\sigma N\lambda + N(N-1)\sigma(\lambda + \delta)}{2[\delta\mu - (\lambda + \delta)\Lambda]} + \frac{\Lambda(\lambda + \delta)[\lambda\delta + \sigma N(\lambda + \delta)]}{[\delta\mu - (\lambda + \delta)\Lambda]^2} \right\} P_{0,0}, \\
P'_2(1) &= \sigma \left[\frac{N(N-1)}{2} \frac{1}{\Lambda} + \frac{N}{\theta} \right] P_{0,0}, \\
P'_3(1) &= \frac{\beta}{\alpha} \left\{ \frac{2\sigma N\lambda + N(N-1)\sigma(\lambda + \delta)}{2[\delta\mu - (\lambda + \delta)\Lambda]} + \frac{\Lambda(\lambda + \delta)[\lambda\delta + \sigma N(\lambda + \delta)]}{[\delta\mu - (\lambda + \delta)\Lambda]^2} \right\} P_{0,0},
\end{aligned}$$

其中

$$P_{0,0} = \left[\left(\frac{\mu}{\lambda + \delta} + 1 + \frac{\beta}{\alpha} \right) \frac{\lambda\delta + (\lambda + \delta)\sigma N}{\delta\mu - (\lambda + \delta)\Lambda} + \frac{\delta}{\lambda + \delta} + \frac{\sigma N}{\Lambda} + \frac{\sigma}{\theta} \right]^{-1}.$$

一名顾客到达并选择进入系统的期望等待时间为

$$W = \frac{P'_0(1) + P'_1(1) + P'_2(1) + P'_3(1)}{\Lambda [P_1(1) + P_2(1)]},$$

轨道中的平均顾客数为

$$E(N) = P'_0(1) + P'_1(1) + P'_2(1) + P'_3(1).$$

4. 双目标优化

本节建立了一个双目标优化模型，旨在同时最小化成本和期望等待时间。期望等待时间是决定服务质量的重要因素，然而在排队系统中，服务成本与服务质量常常发生冲突，因此本文将二者结合起来，寻找到使得它们同时达到最小的解集，并考虑它们之间存在的回归关系。

本文所构建的模型中，阈值 N 和服务率 μ 是重要参数，显著影响整个系统，因此本文构建了期望等待时间函数 $W(N, \mu)$ 和成本函数 $F(N, \mu)$ 。下面给出成本函数的具体表达式

$$F(N, \mu) = C_n E(N) + C_s \mu + C_1 P_1(1) + C_2 P_2(1) + C_3 P_3(1)$$

其中

C_n 代表每名顾客在系统内的单位时间成本。

C_s 代表服务台提供服务的单位时间成本。

C_1 、 C_2 和 C_3 分别代表了系统处于忙期、假期和故障期的单位时间成本。

该双目标优化问题陈述如下

$$\min_{\alpha>0, \mu>0} [F(N, \mu), W(N, \mu)]。$$

借助 NSGA-II 算法来寻找 Pareto 最优解集 (F_p, W_p) 。首先, 随机产生初始种群, 对其进行非支配快速排序, 通过选择、交叉和变异操作得到第一代子代种群。然后, 合并父代种群和子代种群, 进行快速非支配排序, 计算个体拥挤度, 选取合适的个体组成新的父代种群。最后, 通过选择、交叉和变异操作产生新的子代种群。依此类推, 直到满足最大迭代次数, 结束运行。

表 1 给出了 λ 取不同值时的 Pareto 最优解。实际上每个 λ 对应的最优解集中都有 50 组数据, 在这里只截取了部分数据做展示。

Table 1. Pareto optimal solutions when λ takes different values

表 1. λ 取不同值时的 Pareto 最优解

λ	N^*	μ^*	F_p	W_p
1	3	6.912023	3.349037	3.103350
	3	6.700880	3.360424	3.041215
	3	6.583578	3.367124	3.006738
	3	6.407625	3.377720	2.955084
	3	6.137830	3.395382	2.876039

1.5	3	6.958944	3.430634	3.127595
	3	6.777126	3.443041	3.074415
	3	6.478006	3.470367	2.988106
	3	6.331378	3.482200	2.945427
	3	6.143695	3.493223	2.889942

2	3	6.894428	3.522581	3.119519
	3	6.70088	3.539369	3.063344
	3	6.577713	3.550676	3.027669
	3	6.419355	3.565989	2.981891
	3	6.390029	3.579238	2.975362

图 1 展现了不同 λ 取值下的 F_p 关于 W_p 的走势。从中可以看出, F_p 和 λ 成正比关系。 F_p 和 W_p 之间成反比关系, 近似于逆函数、负指数分布或对数函数。然而在回归过程中, 发现建立对数函数拟合的效果是最优的。

观察图 2，可发现残差均值接近于 0，标准差接近于 1，数据呈标准正态分布。从图 3 中可看出图上的点近似地在一条直线附近，这意味着残差正态性条件是达到的，可建立各个变量间的回归关系。因此，建立多元回归模型如下

$$F_p = \beta_0 + \beta_1 \lambda + \beta_2 W_p + \beta_3 \ln W_p$$

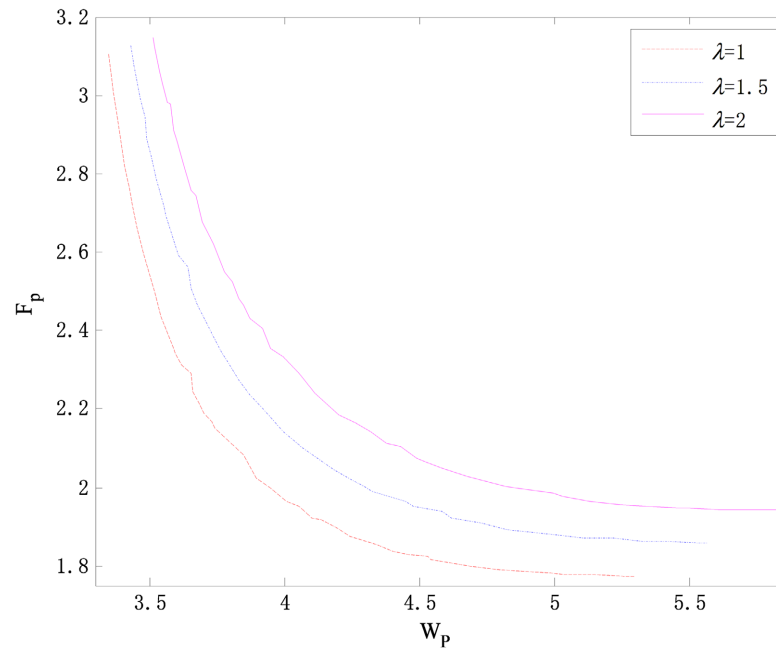


Figure 1. The trend chart of Pareto optimal solutions

图 1. Pareto 最优解走势图

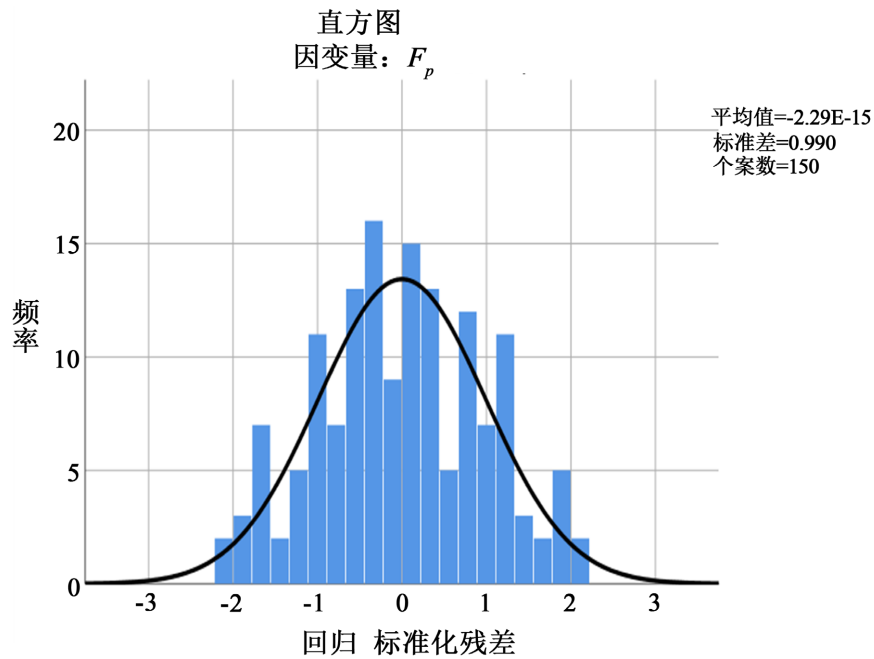


Figure 2. Histogram for the regression residuals

图 2. 回归残差项直方图

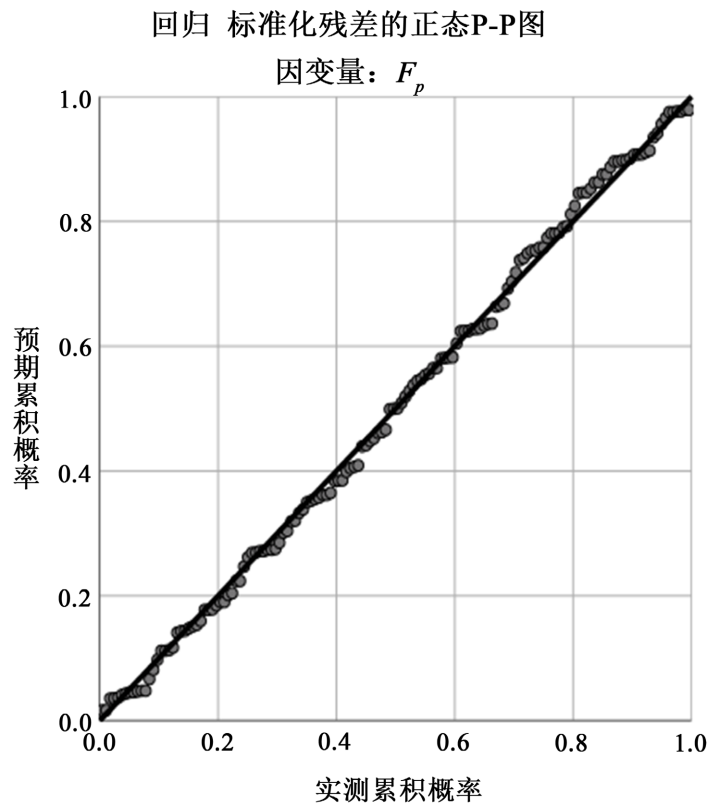


Figure 3. K-S test for the regression residuals
图 3. 回归残差项的 K-S 检验

基于表 1 中的数据, 使用加权最小二乘估计法(WLS)得到参数的估计量, 结果见表 2。

Table 2. Regression results
表 2. 回归结果

变量	数值结果
λ	0.313 ^{***} (0.021)
W_p	3.498 ^{***} (0.177)
$\ln W_p$	-17.551 ^{***} (0.771)
截距项	12.086 ^{***} (0.360)
F 统计量	775.499 ^{***}
调整后的 R^2	0.940

注: ***表示系数在 1%的水平下显著, 括号内为稳健的标准差。

从表 2 中可以看出, 在 1%的显著性水平下, 各个自变量对因变量的影响和回归方程都有显著的统计学意义。调整后的拟合优度为 $R^2 = 94\%$, 本文所选取的自变量一共可以解释因变量 94%的变化, 这说明模型拟合效果良好。

因此得到回归方程

$$F_p = 12.086 + 0.313\lambda + 3.498W_p - 17.551\ln W_p$$

从回归方程中,可以看出: λ 每变动一个单位, F_p 平均变动 0.313 个单位; W_p 每变动一个单位, F_p 平均变动 3.498 个单位; $\ln W_p$ 每变动 1%, F_p 平均变动 17.551 个单位。其中, λ 对 F_p 的影响是正向的,这是因为当 λ 变大时,轨道中的顾客数增加,显然会导致成本的增加。 W_p 的参数估计值虽然为正,但 $\ln W_p$ 的参数估计值为负,并且 $\ln W_p$ 对 F_p 的影响更大,这是因为顾客在系统中的期望逗留时间增加,导致系统变得拥堵,轨道中的顾客数增加,因此成本增加。

5. 结论

本文考虑了基于 N 策略下带有预留时间和启动时间的 M/M/1 可修重试排队系统。利用概率母函数的方法,得到了系统的稳态概率,并给出了系统的关键性能指标。制定了一个基于性能指标和成本要素的成本函数,考虑了双目标优化问题,借助 NSGA-II 算法最小化成本函数和期望等待时间,得到了多组 Pareto 最优解集。随后建立了一个回归模型,检验了从帕累托最优解集获得的最小成本和期望等待时间的关系。

参考文献

- [1] Cohen, J.W. (1957) Basic Problems of Telephone Traffic Theory and the Influence of Repeated Calls. *Philips Telecommunication Review*, **18**, 49-100.
- [2] Falin, G.I. and Templeton, J. (1997) *Retrial Queues*. Chapman and Hall, London.
- [3] Artalejo, J. and Gomez-Corral, A. (2008) *Retrial Queueing Systems: A Computational Approach*. Springer, Berlin.
- [4] Shogan, A.W. (1979) A Single Server Queue with Arrival Rate Dependent on Server Breakdowns. *Naval Research Logistics Quarterly*, **26**, 487-496. <https://doi.org/10.1002/nav.3800260310>
- [5] Neuts, M.F. and Lucantoni, D.M. (1979) A Markovian Queue with N Servers Subject to Breakdowns and Repairs. *Management Science*, **25**, 849-861. <https://doi.org/10.1287/mnsc.25.9.849>
- [6] Wang, J., Cao, J. and Li, Q. (2001) Reliability Analysis of the Retrial Queue with Server Breakdowns and Repairs. *Queueing Systems*, **38**, 363-380.
- [7] Kulkarni, V.G. and Choi, B.D. (1990) Retrial Queues with Server Subject to Breakdowns and Repairs. *Queueing Systems*, **7**, 191-208. <https://doi.org/10.1007/BF01158474>
- [8] 田乃硕. 休假随机服务系统[M]. 北京: 北京大学出版社, 2001.
- [9] Wu, C.H. and Yang, D.Y. (2021) Bi-Objective Optimization of a Queueing Model with Two-Phase Heterogeneous Service. *Computers and Operations Research*, **130**, Article No. 105230. <https://doi.org/10.1016/j.cor.2021.105230>
- [10] Khodemani-Yazdi, M., Tavakkoli-Moghaddam, R. and Bashiri, M., Rahimi, Y. (2019) Solving a New Bi-Objective Hierarchical Hub Location Problem with an M/M/C Queueing Framework. *Engineering Applications of Artificial Intelligence*, **78**, 53-70. <https://doi.org/10.1016/j.engappai.2018.10.004>
- [11] Hajipour, V., Farahani, R.Z. and Fattahi, P. (2016) Bi-Objective Vibration Damping Optimization for Congested Location—Pricing Problem. *Computers & Operations Research*, **70**, 87-100. <https://doi.org/10.1016/j.cor.2016.01.001>