

具有不同故障和延迟维修的重试队列的优化分析

韩云娜

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

本文研究具有两种类型故障和延迟维修的M/M/1重试队列。在稳态条件下, 我们得到了服务台状态和轨道顾客数的概率母函数的表达式。建立了以期望成本最小化和顾客期望等待时间最小化为目标的双目标优化模型。拟议的分析办法使我们能够在成本费用和服务质量之间取得适当的平衡。

关键词

重试排队, 负顾客, 被动故障, 延迟维修, 双目标优化

Optimization Analysis of Retrial Queues with Different Breakdowns and Delayed Repairs

Yunna Han

School of Science, Yanshan University, Qinhuangdao Hebei

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

Abstract

In this paper we study M/M/1 retrial queues with two types of breakdowns and delayed repairs. Under steady-state conditions, we obtain expressions for the probability generating functions of the server state and the number of orbital customers. A bi-objective optimization model with the objectives of minimizing the expected cost and minimizing the expected waiting time of the customers is developed. The proposed analytical approach allows us to achieve an appropriate balance between operational costs and service quality.

文章引用: 韩云娜. 具有不同故障和延迟维修的重试队列的优化分析[J]. 统计学与应用, 2023, 12(2): 332-338.
DOI: 10.12677/sa.2023.122035

Keywords

Retrial Queueing, Negative Customers, Passive Breakdown, Delayed Repairs, Bi-Objective Optimization

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

重试排队的通用模型、方法、结果，示例和应用可以在 Artalejo 和 Corral [1] 以及 Tian 和 Zhang [2] 中找到。由于现实中一些不可预知的因素，如服务器寿命有限、外部干扰、服务器故障、启动失败等，在空闲或忙碌期间，服务器可能会出现故障并需要维修。Kulkarni 和 Choi [3] 介绍了服务器受到故障和维修的重试队列，并得出了系统的稳定条件和稳态分布。此后，对故障和可修的重试排队从稳定性和可靠性的角度进行广泛研究。Zhang 等 [4] 研究了具有两类故障和延迟修复的 M/G/1 重试队列分析，通过补充变量法计算出一些重要的性能指标和可靠性指标。Krishna 等 [5] 研究了具有两种故障和维修的马尔科夫重试队列。

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。Tavakkoli 等 [6] 提出了一种新的多目标模型，用于解决具有拥堵和定价策略的设施位置问题。该模型考虑了不移动服务设施因 M/M/m/k 队列之后的随机需求而拥堵的情况。Wu 和 Yang [7] 研究了具有两阶段异构服务排队模型的双目标优化问题。Khodemani 等 [8] 提出了一个双目标分层枢纽选址问题，旨在同时使总成本和总行程时间最小化。他们分别使用 M/M/C 和 M/M/1 排队系统来模拟中心和本地枢纽的队列。Hajipour [9] 研究具有固定服务器和随机需求为 M/M/M/K 的拥塞系统队列的双目标优化问题。更多关于多目标优化的问题，可以参阅文献 [10] [11] [12]。

2. 模型描述

在具有负顾客、被动故障和延迟维修的重试队列中。顾客根据速率 λ 的泊松过程到达系统。服务台的服务速率服从参数为 μ 的指数分布。到达顾客发现服务台不可用就会以概率 q 加入轨道或以互补概率 $1-q$ 离开系统。重试时间服从参数为 ν 的指数分布。服务台可能会遇到两种类型的故障，即被动故障和主动故障。被动故障意味着当服务台处于空闲状态时系统根据速率为 η 的泊松过程发生故障。被动故障的修复时间服从参数为 θ 的指数分布，由于在空闲期间缺乏对服务台的监控，当发生被动故障时服务台无法立即获得修复，而处于延迟修复状态。延迟维修服从参数为 δ 的指数分布。主动故障意味着当服务台繁忙时，负面顾客的到来会影响系统，导致服务台故障，同时迫使正在服务的顾客离开系统。负顾客根据速率为 ϕ 的泊松过程到达。主动修复时间服从参数为 β 指数分布。

最后，假设到达、服务、故障、修复、延迟维修和重试过程相互独立。

令 $N(t)$ 表示时刻 t 轨道中的顾客数， $I(t)$ 表示时刻 t 系统的状态(0: 服务台空闲，1: 服务台繁忙，2: 服务台处于被动故障状态，3: 服务台处于主动故障状态，4: 服务台处于延迟修复状态)。

3. 稳态分析

$\pi(j, i)$ 表示系统处于状态的加入概率。

$$\pi(j, i) = \lim_{t \rightarrow \infty} P\{N(t) = j, I(t) = i\}, (j, i) \in \Omega.$$

为了获得系统稳态分布, 我们定义概率母函数

$$G_i(z) = \sum_{j=0}^{\infty} \pi(j, i) z^j, \quad |z| < 1, \quad i \in \{0, 1, 2, 3, 4\}.$$

服务台处于空闲状态的概率为

$$P_0 = G_0(1) = \frac{\mu\nu\theta\delta\kappa_1(\lambda q + \theta)(\lambda q + \delta) + \beta\varphi\theta\delta\mu\kappa_2}{\kappa_2[\theta\delta\kappa_1(\lambda + \nu) - \beta\mu\lambda\kappa_3]} \pi(0, 1) \quad (1)$$

服务台繁忙的概率为

$$P_1 = G_1(1) = \frac{\beta\mu\lambda\nu\kappa_3(\lambda q + \theta)(\lambda q + \delta) + \beta\varphi\theta\delta\mu\kappa_2(\lambda + \nu)}{\kappa_2[\theta\delta\kappa_1(\lambda + \nu) - \beta\mu\lambda\kappa_3]} \pi(0, 1) \quad (2)$$

服务台处于被动修复状态的概率为

$$P_2 = G_2(1) = \frac{\mu\nu\delta\eta\kappa_1(\lambda q + \theta)(\lambda q + \delta) + \beta\varphi\delta\mu\eta\kappa_2}{\kappa_2[\theta\delta\kappa_1(\lambda + \nu) - \beta\mu\lambda\kappa_3]} \pi(0, 1) \quad (3)$$

服务台处于主动修复状态的概率为

$$P_3 = G_3(1) = \frac{\varphi\mu\lambda\nu\kappa_3(\lambda q + \theta)(\lambda q + \delta) + \varphi^2\theta\delta\kappa_2(\lambda + \nu)}{\kappa_2[\theta\delta\kappa_1(\lambda + \nu) - \beta\mu\lambda\kappa_3]} \pi(0, 1) \quad (4)$$

服务台处于延迟修复状态的概率为

$$P_4 = G_4(1) = \frac{\mu\nu\theta\eta\kappa_1(\lambda q + \theta)(\lambda q + \delta) + \beta\varphi\theta\eta\mu\kappa_2}{\kappa_2[\theta\delta\kappa_1(\lambda + \nu) - \beta\mu\lambda\kappa_3]} \pi(0, 1) \quad (5)$$

其中

$$\begin{aligned} \kappa_1 &= \beta(\varphi + \mu) - \lambda q(\beta + \varphi) \\ \kappa_2 &= \lambda[\lambda q(\lambda q + \eta q + \theta + \delta) + \eta q(\theta + \delta) + \theta\delta] \\ \kappa_3 &= \eta q(\theta + \delta) + \theta\delta \\ A_1 &= \nu\kappa_1(\lambda q + \theta)(\lambda q + \delta) + \beta\varphi\kappa_2 \\ A_2 &= \lambda\mu\nu\kappa_3(\lambda q + \theta)(\lambda q + \delta) + \varphi\theta\delta\kappa_2(\lambda + \nu) \\ \pi(0, 1) &= \frac{\kappa_2[\theta\delta\kappa_1(\lambda + \nu) - \beta\mu\lambda\kappa_3]}{\mu[\theta\delta + \eta(\theta + \delta)]A_1 + (\beta + \varphi)A_2} \end{aligned} \quad (6)$$

服务台空闲、繁忙、被动修复、主动修复、延迟修复时的平均轨道大小 N_i 分别为

$$N_0 = G'_0(1) = \left(\sum_{j=0}^{\infty} \pi(j, 0) z^j \right)' \quad (7)$$

$$N_1 = G'_1(1) = \left(\sum_{j=0}^{\infty} \pi(j, 1) z^j \right)' \quad (8)$$

$$N_2 = G'_2(1) = \left(\sum_{j=0}^{\infty} \pi(j, 2) z^j \right)' \quad (9)$$

$$N_3 = G'_3(1) = \left(\sum_{j=0}^{\infty} \pi(j, 3) z^j \right)' \quad (10)$$

$$N_4 = G'_4(1) = \left(\sum_{j=0}^{\infty} \pi(j, 4) z^j \right)' \quad (11)$$

设 $E(N)$ 表示轨道中顾客的平均数量。因此,

$$E(N) = N_0 + N_1 + N_2 + N_3 + N_4 \quad (12)$$

假设被标记顾客在到达时发现服务台不可用, 并决定加入重试轨道, 他在轨道上的预期等待时间为

$$W = \frac{E(N)}{\lambda_{ret}} = \frac{E(N)}{\lambda q [G_1(1) + G_2(1) + G_3(1) + G_4(1)]} \quad (13)$$

4. 双目标优化

大多数关于排队模型优化设计的研究都集中在单目标问题上, 其中成本或利润函数是优化目标。然而, 现实世界中许多优化问题需要同时优化多个目标函数。多方面考虑是决策所必需的。预期等待时间是决定顾客对服务质量满意度的最重要因素。在排队系统中, 成本往往与服务质量相冲突。另一方面, 系统对成本和服务质量的要求是不同的。通过分析我们观察到服务率 μ 和延迟修复率 δ 对预期等待时间 W 的影响较大, 同时也会影响系统的预期成本。在本节中, 我们借助遗传算法寻找同时满足最小成本和最少等待时间的 Pareto 最优解集。制定了一个双目标优化模型, 旨在同时最小化期望成本 F_p 和预期等待时间 W_p 。

单位时间的期望成本函数为

$$F(\mu, \delta) = C_h E(N) + C_s \mu + C_e \delta + C_o P_1 + C_f P_2 + C_a P_3 + C_i P_4 \quad (14)$$

其中 C_h 是系统中每个顾客的持有成本, C_s 是单位时间内提供服务的成本, C_e 是单位时间延迟修复的成本, C_o 是保持系统运行的单位时间成本, C_f 是服务台处于被动故障状态的单位时间成本, C_a 是服务由于负顾客而处于故障状态的单位时间成本, C_i 是服务台处于延迟修复状态的单位时间成本。

双目标优化问题表述为

$$\min [F_p, W_p] \quad (15)$$

多目标遗传算法是用来分析和解决多目标优化问题的一种进化算法, 是以遗传算法为基础并基于 Pareto 最优概念得到的。其核心就是协调各个目标函数之间的关系, 找出使得各个目标函数都尽可能达到比较小的(或者比较大的)函数值的最优解集。我们选择以下系统参数 $q = 0.4$, $\varphi = 0.5$, $\beta = 1.4$, $\eta = 0.2$, $\theta = 1.5$, $\nu = 1.8$, $C_h = 0.5$, $C_s = 2$, $C_e = 2.5$, $C_o = 3$, $C_f = 4$, $C_a = 4$, $C_e = 4.5$ 。图 1 给出了 $\lambda = 2.2, 2.4, 2.6$ 时使用多目标遗传算法得到的非支配解, 各种 λ 值下的 Pareto 最优解汇总在表 1 中。这些结果为决策者提供了有用的指导方针。由图 1 和表 1 可以看出:

1) 这三条曲线表明, 成本的增加导致顾客等待时间的减少。当 W_p 固定时, F_p 随着平均到达率 λ 的增加而增加。

2) 当 W_p 趋近无穷大时, 最小期望成本 F_p 收敛于某一值, 该值可视为最小成本。

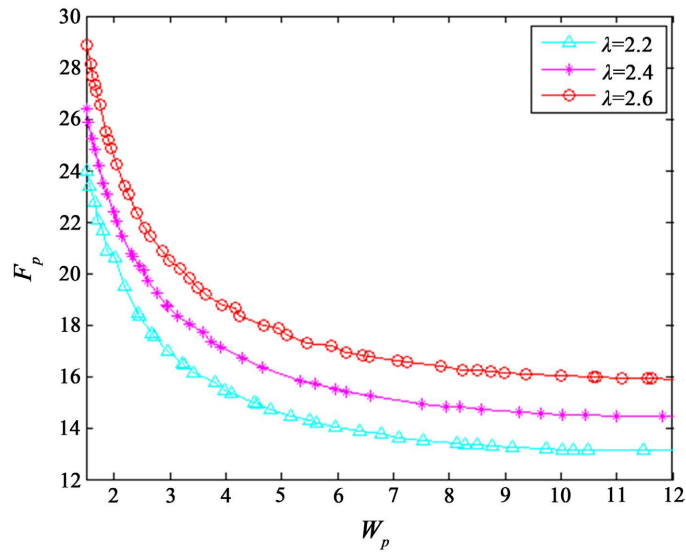


Figure 1. Pareto front solution found by genetic algorithm
图 1. 遗传算法找到的 Pareto 前沿解

Table 1. The Pareto optimal solutions for various values of λ
表 1. 不同参数 λ 下的 Pareto 最优解

λ	δ^*	μ^*	W_p	F_p
2.2	1.013685239	3.047898338	10.47939618	13.1248014
	1.013685239	3.472140762	7.531527811	13.51567069
	1.265884653	4.156402737	4.548649438	14.91442655
	1.563049853	5.43597263	2.68099876	17.65367789
	2.738025415	7.427174976	1.50336452	23.98950303
	...	...	...	...
2.4	1.013685239	3.376344086	12.77151603	14.41235041
	1.095796676	3.786901271	8.566595525	14.73423676
	1.430107527	4.813294233	4.291037585	16.69683244
	1.449657869	6.831867058	2.535452456	20.15018117
	2.174975562	8.474095797	1.661489684	24.79362586
	...	...	...	...
2.6	1.001955034	3.978494624	12.65447072	15.88314227
	1.078201369	4.395894428	8.751180879	16.16272361
	1.258064516	5.107526882	5.465581314	17.29136247
	1.926686217	6.044965787	3.183612616	20.15862665
	2.251221896	8.467253177	1.890007053	25.18239912
	...	...	...	...

从表 2 中可以看出，单个自变量 t 检验的 P 值都小于 0.01，回归方程 F 检验的 P 值小于 0.01，因此，

在 0.01 的显著水平下, 各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 $R^2 = 0.996$, 模型拟合效果良好。

Table 2. Regression results
表 2. 回归结果

变量	结果
λ	8.846*** (0.164)
W_p^{-1}	13.694*** (0.636)
e^{-W_p}	18.712*** (1.528)
截距项	-8.039*** (0.411)
F 统计量	11125.899***

注: ***表示系数在 1%的水平下显著, 括号内为标准误差。

从图 1 中可以看出 F_p 与 W_p 之间的关系是呈反比关系, 近似为一个负指数的指数函数, F_p 与 λ 成正比关系。从图 2 回归残差的直方图中可以看出左右两侧对称, 从 K-S 正态性检验图中可以看出, 散点基本靠近直线。明预期累计概率与实测累计概率十分吻合, 说明服从正态分布, 满足回归的假设条件, 可以建立回归关系。

因此, 建立多元线性回归模型如下

$$F_p = -8.039 + 8.846\lambda + 13.694W_p^{-1} + 18.712e^{-W_p} \quad (16)$$

这个回归模型将有助于管理人员寻求在给定的 λ 估计值下达到满意的服务水平所需的最小成本。

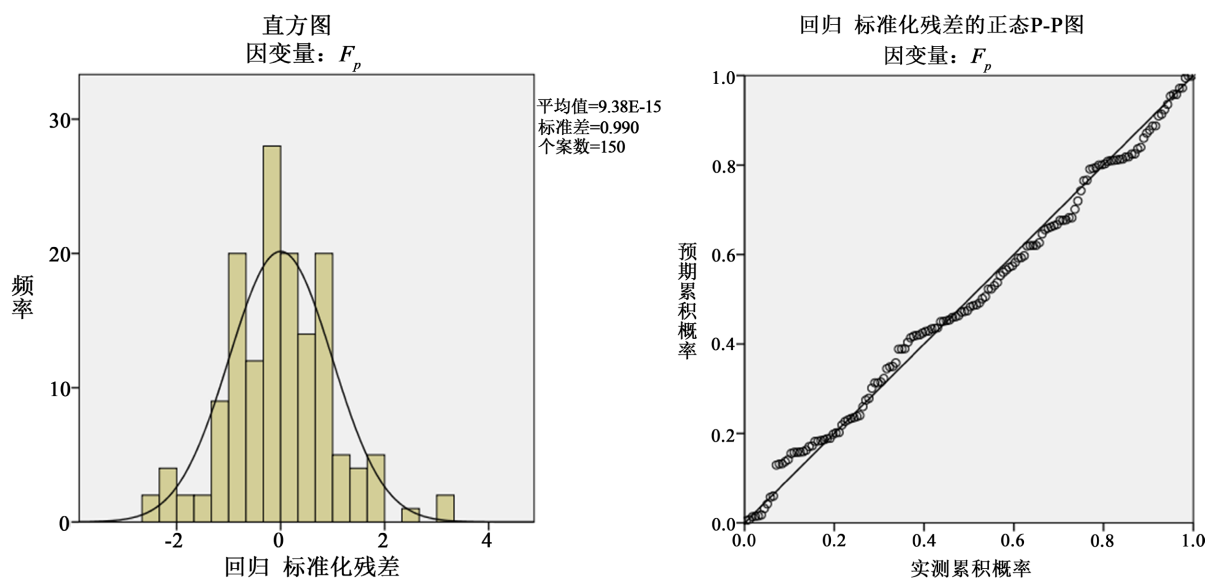


Figure 2. Histogram and K-S normality test for the regression residuals

图 2. 回归残差的直方图和 K-S 正态性检验

5. 结论

我们研究了 M/M/1 重试排队系统, 该系统具有延迟维修的被动故障和繁忙时期由负顾客引起的主动故障。利用概率母函数, 我们得到了重要的性能指标。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, 使用遗传算法最小化期望成本函数和预期等待时间, 最后得到的 Pareto 最优解集。表明需要在运营成本和顾客等待时间之间进行权衡。

参考文献

- [1] Artalejo, J.R. and Gómez-Corral, A. (1999) Retrial Queueing Systems. *Mathematical and Computer Modelling*, **30**, 13-15. [https://doi.org/10.1016/S0895-7177\(99\)00127-2](https://doi.org/10.1016/S0895-7177(99)00127-2)
- [2] Tian, N.S., Li, J.H. and Zhang, Z.G. (2009) Matrix Analytic Method and Working Vacation Queues—A Survey. *International Journal of Information and Management Sciences*, **20**, 603-633.
- [3] 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>
- [4] Gao, H., Zhang, J. and Wang, X. (2020) Analysis of a Retrial Queue with Two-Type Breakdowns and Delayed Repairs. *IEEE Access*, **8**, 172428-172442. <https://doi.org/10.1109/ACCESS.2020.3023191>
- [5] Krishna, K.B., Rukmani, R., Thanikachalam, A., et al. (2016) Performance Analysis of Retrial Queue with Server Subject to Two Types of Breakdowns and Repairs. *Operational Research*, **18**, 521-559. <https://doi.org/10.1007/s12351-016-0275-4>
- [6] Tavakkoli-Moghaddam, R., Vazifeh-Noshafagh, S., Taleizadeh, A.A., et al. (2017) Pricing and Location Decisions in Multi-Objective Facility Location Problem with M/M/m/k Queueing Systems. *Engineering Optimization*, **49**, 136-160. <https://doi.org/10.1080/0305215X.2016.1163630>
- [7] Wu, C.H. and Yang, D.Y. (2021) Bi-Objective Optimization of a Queueing Model with Two-Phase Heterogeneous Service. *Computers & Operations Research*, **130**, Article No. 105230. <https://doi.org/10.1016/j.cor.2021.105230>
- [8] Khodemani-Yazdi, M., Tavakkoli-Moghaddam, R., Bashiri, M., et al. (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>
- [9] 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>
- [10] Vishkaei, B.M., Fathi, M., Khakifirooz, M. and De Giovanni, P. (2021) Bi-Objective Optimization for Customers' Satisfaction Improvement in a Public Bicycle Sharing System. *Computers & Industrial Engineering*, **161**, Article No. 107587. <https://doi.org/10.1016/j.cie.2021.107587>
- [11] Wu, C.H., Yang, D.Y. and Yong, C.R. (2023) Performance Evaluation and Bi-Objective Optimization for F-Policy Queue with Alternating Service Rates. *Journal of Industrial and Management Optimization*, **19**, 3819-3839. <https://doi.org/10.3934/jimo.2022111>
- [12] Pouraliakbari-Mamaghani, M., Mohammadi, M., Arshadi-Khamseh, A., et al. (2021) A Bi-Objective Robust Possibilistic Programming Model for Blood Supply Chain Design in the Mass Casualty Event Response Phase: a M/M/1/K Queueing Model with Real World Application. *International Journal of Operational Research*, **42**, 229-275. <https://doi.org/10.1504/IJOR.2021.118976>