

基于随机完全区组设计的嵌入式系统性能分析

楼俊钢¹, 江建慧¹, 赵时旻², 孔亮亮¹

(1. 同济大学 计算机科学与技术系, 上海 201804; 2. 上海申通轨道交通研究咨询有限公司, 上海 201103)

摘要: 提出了基于随机完全区组设计的嵌入式系统性能分析方法. 以城市轨道交通自动售检票系统的自动检票机的性能测试为例, 说明了运用该方法的过程. 比较了 ReWorks, Linux 和 Windows XPE 等 3 种不同操作系统下自动检票机的性能, 测评结果表明该方法是有用的.

关键词: 嵌入式系统; 性能分析; 随机完全区组设计; 性能基准程序; 假设检验

中图分类号: TP 302.7

文献标识码: A

Analysis of Performance of Embedded Systems Based on Randomized Complete Blocks Designs

LOU Jungang¹, JIANG Jianhui¹, ZHAO Shimin², KONG Liangliang¹

(1. Department of Computer Science and Technology, Tongji University, Shanghai 201804, China; 2. Shanghai Shentong Rail Transit Research & Consultancy Co. Ltd., Shanghai 201103, China)

Abstract: This paper presents a performance analysis approach of embedded systems based on the randomized complete block designs. A case study is made of this approach in performance testing of automatic gate machines used in the automatic fare collection system of Shanghai Metro. Embedded systems based on ReWorks, Linux or Windows XPE are used in automatic gate machines. The assessment results show that the approach is helpful and effective.

Key words: embedded system; performance analysis; randomized complete block design; performance benchmarks; hypothesis testing

针对特定应用, 比较在不同硬件平台或体系结构下的计算机系统的性能, 是性能评估研究的一项重要内容^[1]. 一般情况下, 比较过程主要考虑以下 2

点: ①确定在哪种硬件平台和操作系统下可以达到最好的性能; ②减少误差对比较过程以及结果的影响, 使得比较的结论更加可信.

通用计算机已经有了一些成功的性能评测技术^[2-3], 如解析法^[4]、模拟法^[5-6]、测量法和基准程序法等. 这些评测方法各有特点, 分别适用于处于生命周期不同阶段的目标系统. 其中, 基准程序法因具有直接、简单、费用低等优点而被广泛地用来评估系统的处理能力.

嵌入式系统的性能评测原理与高性能计算机相似, 嵌入式系统根据应用有其自身的特点, 对于处理器和硬件系统, 现有的性能评测基准程序可以经过适当的修改进行测试. 而由于面向微处理器的基准程序和面向操作系统的基准程序极大地依赖于应用领域, 面向专门应用的基准程序相对来讲用得比较少.

基于测试进行性能分析是常用的性能评估方法之一, 任何一组测试数据, 都可以看作所研究的各种条件下的一种模型或者排列. 由于任何度量都带有一定程度的随机误差, 因此必须综合很多测试数据的结果, 才能使所得的结论具有足够的精度. 统计学的重要作用之一, 是为确定测试的重复次数提供合理的依据. 而在性能比较方面, 现有的方法一般是先确定一组比较的依据、准则, 并通过测试或者其他方法得到系统的指标值, 从而确定性能的优劣. 在比较简单或者指标值相差较大时, 这种方法简单实用; 但对于比较复杂的过程或者指标值较为接近时, 采用这种方法往往会得出错误的或者不恰当的结论. 本文的目的是为了解决这种情况下进行比较所遇到的问题, 用统计学的观点来分析得到的测试数据, 使结果更加客观. 另外, 在对影响性能各因素进行研究

收稿日期: 2008-06-21

基金项目: 上海市科委重点科技攻关计划资助项目(04D15032, 06D15003)

作者简介: 楼俊钢(1982—), 男, 博士生, 主要研究方向为软件可靠性工程、性能评估等. E-mail: loujungang0210@hotmail.com

江建慧(1964—), 男, 教授, 博士生导师, 工学博士, 主要研究方向为容错计算、软件可靠性、微处理器体系结构、系统性能评估等. E-mail: jhjiang@tongji.edu.cn

时,固定其他因素而使所研究因素作某些变动是经常用到的方法,但由于随机误差与测试误差交织在一起,使数据处理变得烦琐.重复测试可以得出随机误差,但若进行全面测试,无疑会加大测试工作量.通过采用随机化完全区组设计,在相同条件下每次只做一组测试,而组与组之间的相应条件允许有所不同,则可使测试工作量大大降低.因此,本文提出基于随机完全区组设计的嵌入式系统性能分析和比较方法.本文方法能分辨测试中误差对实际性能差距的影响,特别适用于 2 种以上不同产品的比较;而通过使用完全区组设计,可以有效减少测试的重复次数,使测试更加有效率;最后把它们应用于上海城市轨道交通自动售检票系统终端设备(自动检票机)的嵌入式系统的性能测试中,获得了比较理想的结果.

1 基于随机完全区组设计的性能分析与比较

所谓测试的统计设计,就是设计测试的过程,使得收集的数据适用于统计方法分析,得出有效和客观的结论.因此,测试包括 2 个方面:测试过程的设计和测试数据的统计分析.它们是紧密相关的,因为分析方法直接依赖于所用的设计.测试的统计设计的基本原理如下:

(1) 重复.就是测试的重复进行,重复测试有 2 条重要的性质.第一,允许测试者得到测试误差的估计量.第二,如果样本均值用作测试中一个因素效应的估计量,则重复允许测试者求得这一效应的更为精确的估计量.

(2) 随机.它是使用统计方法的基石,测试的分配和实施的次序都随机确定.

(3) 随机化完全区组设计.它是用来提高测试精确度的方法,一个区组就是整个测试的一个部分,相比测试数据全体,它们本身的性质更为类似.随机完全区组设计牵涉到每个区组内部对感兴趣的测试内容进行比较.利用随机化区组设计,执行各组处理时的条件差异,能够与处理之间的差异分开,并且与实验误差分开,这种设计相当于把测试条件细分为具有相对相同条件的区组.如果在测试过程中,以杂乱无章的方式取得结果,那么最终的排列也是不平衡和不规则的.若在测试之前进行适当的设计,则用较少的工作量也能取得等量的效果.

1.1 随机化完全区组设计

随机区组设计(randomized complete blocks design)是根据局部控制和随机排列原理进行的.将测试按不同性质划分为等于重复次数的区组,使区组内环境差异最小而区组间环境允许存在差异,每个区组即为 1 次完整的重复,区组内各处理都独立地随机排列,如表 1 所示.这是随机排列设计中最常用、最基本的设计.随机区组试验设计把测试过程中各处理随机排列在 1 个区组中,区组内条件基本一致,区组间可以有适当的差异.随机区组试验由于引进了局部控制原理,可以从测试的误差方差中分解出区组变异的方差,从而减少试验误差,提高 F 检验和多重比较的灵敏度和精确度.

随机化完全区组设计的优点是:①设计简单,容易掌握;②富于伸缩性,单因素、复因素以及综合测试等都可应用;③能提供无偏的误差估计,有效降低非处理因素等测试条件下的单向差异,降低误差,提高测试效率.缺点是:这种设计方法不允许处理数太多.因为处理数多,区组必然增大,局部控制的效率降低,所以,处理数一般不要超过 20 个,最好在 10 个左右.

1.2 总体框架

图 1 为一般性能测试的流程图.本文提出采用区组设计思想来指导测试设计的方法,具体测试设计过程如图 2 所示.区组的划分依赖于因素水平和响应变量的选择等,通过采用随机区组设计,可以有效去除误差的影响,并有效减少测试重复次数.

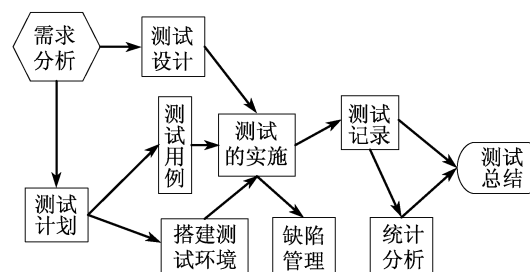


图 1 一般性能测试的流程图

Fig.1 Framework of general performance testing

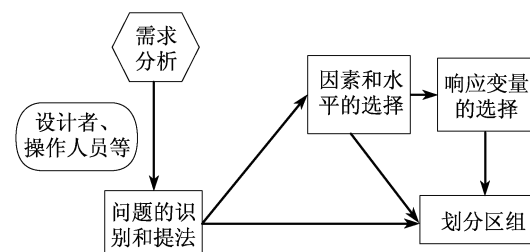


图 2 区组划分过程

Fig.2 Step of block partition

问题的识别和提法:明确测试的目的,需要吸引所有相关人员的参与,包括设计者、测试人员、用户、操作人员等,通常他们会有很多想法,却常常容易被忽略.

因素和水平的选择:选择测试中准备改变的因素、这些因素变化的范围,以及在进行测试时对这些因素规定的水平.还必须考虑将这些因素控制在期望范围内,以及测量这些数值.仔细研究分析所有可能的重要因素,特别是在测试的早期阶段.

响应变量的选择:在选择响应变量时,应该确信一个变量会为所研究的过程提供有用的信息.一般可以取测量特性的平均值或标准差,或两者一起作为响应变量.

划分区组:主要依据因素水平、响应变量的选择来进行.

测试的实施:在测试时,监视测试的过程,确保每一过程按计划进行.先行制定测试计划是成功的关键.在测试过程中牢记测试的目的,如在有些测试中,一开始就知道有些因素水平会使响应得出不同的数值,比较 2 个系统 A 和 B,A 是标准的,B 成本较低或者具有其他优势,测试者会有兴趣弄清它们性能是否会有差异.

数据分析:分析数据采用统计方法,使得结论和结果都是客观的.统计方法不能证明 1 个因素(或几个因素)有特殊的效应,它们仅对测试结果的可靠性和有效性提供准则.

1.3 统计分析

一般说来,如果要比较 a 种处理并且有 b 个区组,随机完全区组设计如表 1 所示.

表 1 随机完全区组设计

Tab.1 Randomized complete blocks design

区组 1	区组 2	...	区组 b
y_{11}	y_{12}	...	y_{1b}
y_{21}	y_{22}	...	y_{2b}
\vdots	\vdots		\vdots
y_{a1}	y_{a2}	...	y_{ab}

在每个区组内,每种处理有一个测试值,对这一设计取统计模型为

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \end{cases}$$

式中: μ 是总均值; τ_i 是第 i 种处理的效应; β_j 是第 j 个区组的效应; ϵ_{ij} 是通常的随机误差项.处理与区

组是固定因素,它们的效应定义为对总均值的偏差.笔者的目的是检验处理均值是否相等.于是,可以假设

$$\begin{aligned} H_0: \mu_1 &= \mu_2 = \mu_3 = \dots = \mu_a \\ H_1: &\text{至少有一个 } i \text{ 使得 } \tau_i \neq 0 \end{aligned}$$

以 y_{ib} 表示在处理 i 下的所有测试值的总和, y_{aj} 表示在区组 j 内的所有测试值的总和, y_{ab} 表示全体测试值的总和,则 $N = ab$ 是测试值的总个数,用数学式表示为

$$\begin{aligned} y_{ib} &= \sum_{j=1}^b y_{ij}, \quad i = 1, 2, \dots, a \\ y_{aj} &= \sum_{i=1}^a y_{ij}, \quad i = 1, 2, \dots, b \\ y_{ab} &= \sum_{i=1}^a \sum_{j=1}^b y_{ij} = \sum_{i=1}^a y_{ib} = \sum_{j=1}^b y_{aj} \end{aligned}$$

用 S_T, S_r, S_t, S_E 分别表示总校正平方和、区组平方和、处理平方和以及试验误差平方和,则

$$S_T = \sum_{i=1}^a \sum_{j=1}^b \left(y_{ij} - \frac{y_{ab}^2}{N} \right)^2 = \sum_{i=1}^a \sum_{j=1}^b y_{ij}^2 - \frac{y_{ab}^2}{N}$$

同理可得

$$S_r = \sum_{i=1}^a \frac{y_i^2}{b} - \frac{y_{ab}^2}{N}, \quad S_t = \sum_{j=1}^b \frac{y_{aj}^2}{a} - \frac{y_{ab}^2}{N}$$

同时可以计算得到

$$S_E = S_T - S_r - S_t$$

进而,统计量

$$M_r = S_r / (a - 1), \quad M_t = S_t / (b - 1),$$

$$M_E = \frac{S_E}{(a - 1)(b - 1)}$$

式中: M_r, M_t, M_E 分别为区组均方、处理均方以及试验误差均方.因为有 N 个测试值, S_T 有 $N - 1$ 个自由度.有 a 种处理和 b 个区组,所以 S_r 和 S_t 分别有 $a - 1$ 和 $b - 1$ 个自由度.而 S_E 有 $ab - 1 - (a - 1) - (b - 1) = (a - 1)(b - 1)$ 个自由度.由此得

$$S_T = S_r + S_t + S_E$$

左右两边自由度相等,用通常的关于误差的正态性假定,由 CACHRAN 定理^[7]可以证明 $S_r/\sigma^2, S_t/\sigma^2, S_E/\sigma^2$ 是独立分布的卡方随机变量.可以证明均方的期望值是

$$E(M_r) = \sigma^2 + b \sum_{i=1}^a \tau_i^2 / (a - 1)$$

$$E(M_t) = \sigma^2 + a \sum_{i=1}^b \beta_i^2 / (b - 1)$$

$$E(M_E) = \sigma^2, \sigma^2 \text{ 为总方差}$$

所以,检验处理均值的等式,可以用检验统计量 $F_0 =$

M_r/M_E .

当零假设为真时,它的分布是 $F_{a-1, (a-1)(b-1)}$, 否定区域是 F 分布的上尾部, 当 $F_0 > F_{a-1, (a-1)(b-1)}$ 时(α 称为检验的显著性水平), 否定 H_0 . 所以, 可以得到随机完全区组设计如表 2 所示的方差分析表.

表 2 随机完全区组设计的方差分析表
Tab.2 Variance analysis of randomized complete blocks design

变差来源	平方和	均方	F_0
处理	$\sum_{i=1}^a \frac{y_{ib}^2}{b} - \frac{y_{ab}^2}{N}$	$\frac{S_r}{a-1}$	
区组	$\sum_{j=1}^b \frac{y_{aj}^2}{a} - \frac{y_{ab}^2}{N}$	$\frac{S_t}{b-1}$	$\frac{M_r}{M_E}$
误差	S_E	$M_E = \frac{S_E}{(a-1)(b-1)}$	
总和	$\sum_{i=1}^a \sum_{j=1}^b y_{ij}^2 - \frac{y_{ab}^2}{N}$		

1.4 相对效率

上文说明了随机完全区组设计的误差减少性质. 注意到, 在不计算处理时的总平方和中(表 2), 由区组间差异所得的误差可能会占较多比例. 这样, 如果不采用随机完全区组设计的测试, M_E 的值可能会太大, 没有随机完全区组设计那样灵敏.

估计随机完全区组设计对一般测试的相对效率是有帮助的. 定义这一相对效率的一种方法是^[8]

$$R = \frac{(f_b + 1)(f_r + 3) \frac{\sigma_r^2}{\sigma_b^2}}{(f_b + 3)(f_r + 1) \frac{\sigma_r^2}{\sigma_b^2}}$$

式中: σ_r^2 和 σ_b^2 分别表示完全随机设计与一般测试的测试误差; f_r 和 f_b 分别表示相应误差的自由度. 该统计量可以看作一般测试对与之有相同灵敏度的随机完全区组设计在重复次数上的增加倍数.

为了计算相对效率, 必须有 σ_r^2 和 σ_b^2 的估计值, 随机完全区组设计的 M_E 是 σ_b^2 的无偏估计量, 并且可以证明 σ_r^2 误差方差的无偏估计量为^[8]

$$\tilde{\sigma}_r^2 = \frac{(b-1)M_r + b(a-1)M_E}{ab-1} \quad (1)$$

2 自动检票机性能测试

2.1 结构与性能指标

城市轨道交通自动检票机位于车站的付费区和非付费区之间, 检票机能接受轨道交通专用车票和公共交通卡, 并满足乘客持票快速通过的需求.

对于自动检票机系统而言, 性能指标是否满足业务需求关系到它的成败. 因此, 性能测试对于自动检票机系统尤其重要. 笔者关注的主要性能指标是交易处理时间, 包括单程票和交通卡的进站交易处理时间、单程票和交通卡的出站交易处理时间. 下面以单程票交易进站处理时间为例, 具体说明这些指标的含义及其计算方法.

图 3 为单程票交易进站处理时间的流程图. 单程票进站交易处理时间是指乘客使用单程票进站时, 从车票靠近天线到通道放行这段时间. 具体包括: 车票费用计算时间, 密钥安全性、票种合法性、状态、使用地点、余值、有效期、进/出站次序、更新信息等检查时间, 读写卡时间. 具体测试时, 使用模拟器通过与自动检票机主控单元之间报文交互来记录时间值, 因此, 实际测试得到的时间值中未包含读写卡时间.

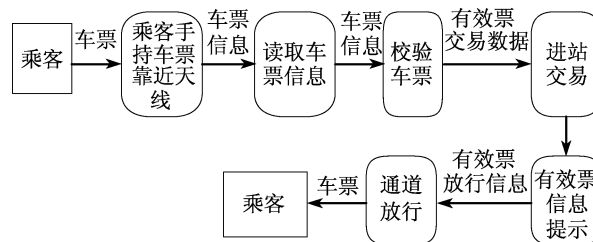


图 3 单程票进站处理时间流程图

Fig.3 Flow chart of transaction time

2.2 基于不同操作系统的自动检票机的性能测试

测试中要求对基于 ReWorks, Linux 和 Windows XPE 3 个操作系统的自动检票机系统的性能进行比较测试, 目的是确定基于这 3 种不同操作系统的自动检票机在 4 组测试指标下是否具有相同的性能, 或者验证它们性能不同是由于误差引起的可能性. 图 4 给出了测试 1 000 次后基于 ReWorks 操作系统性能测试时间值分布百分比情况. 可以看出, 它们基本上符合正态分布.

把所有测试结果按不同操作系统、不同测试指标, 随机地安排 $3 \times 4 = 12$ 个测试单元, 并记录相应的测试数据. 把 12 个测试单元按照单程票进站交易处理时间、交通卡进站交易处理时间、单程票出站交易处理时间、交通卡出站交易处理时间分为 4 个区组. 不同操作系统的自动检票机当作不同的处理, 它的一个完全区组设计如表 3 所示. 表中的数值为测试平均值. 按照表 2 公式可以计算得到如表 4 所示的方差分析.

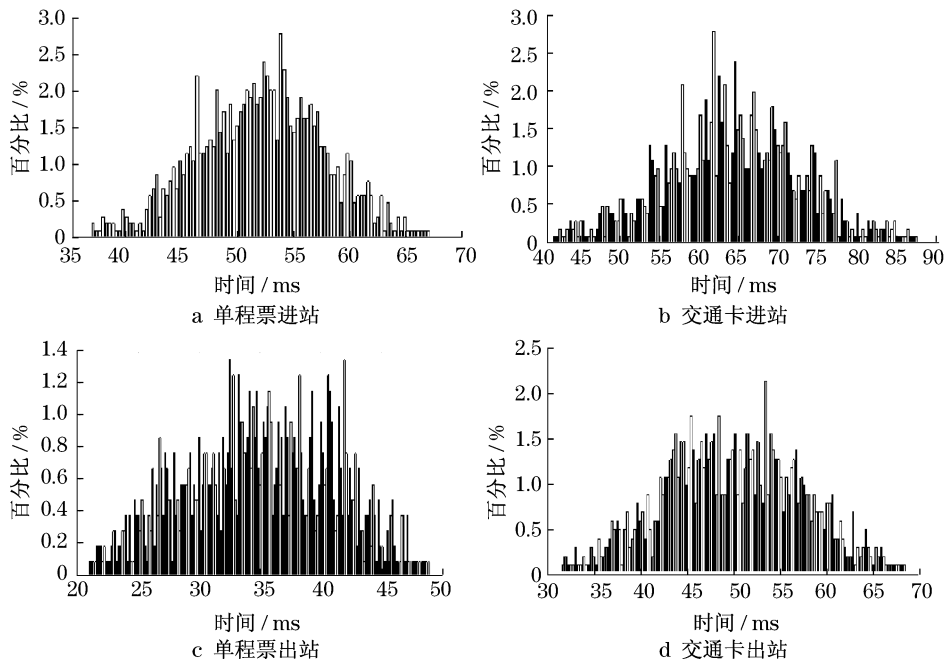


图 4 处理时间分布百分比
Fig.4 Percent distribution of transaction time

表 3 完全区组设计
Tab.3 Randomized complete blocks design ms

操作系统	区组			
	单程票进站	单程票出站	交通卡进站	交通卡出站
ReWorks	52.5	63.4	35.9	48.8
Linux	237.0	178.0	239.0	97.0
Windows XPE	52.0	60.0	36.5	45.1

表 4 方差分析
Tab.4 Variance analysis

变差来源	平方和	均方	F_0
处理(操作系统类型)	1245.46	622.73	
区组(指标类型)	824.48	412.24	38.04
误差	98.20	16.37	
总和	2168.14		

取 $\alpha = 0.05$, F 的临界值 $F_{\alpha, a-1, (a-1)(b-1)} = F_{0.05, 2, 6} = 3.74$, 因为 $38.44 > 3.74$. 由此可以得出结论:各操作系统下的自动检票机的性能有显著性差异, 由于误差造成的可能性很小. 另外, 根据式(1)计算得 $\sigma_e^2 = 124.33$, 采用随机完全区组设计可以减少的重复倍数为: $R = \frac{(6+1)(8+3)}{(6+3)(8+1)} \cdot \frac{124.33}{16.37} = 7.21$. 这意味着, 采用了区组化设计之后, 测试重复次数可以大大减少, 但仍然能够保持同样的误差灵敏度.

3 结论

在以性能为标准之一选择产品的组件时, 使用随机化区组设计主要有以下优点:

(1) 统计测试设计理论论述了在一组测试中各项数据的数目及其相互关系, 利用数学理论, 可获得测试方案所提供资料的定量度量. 统计方法允许度量结论中可能出现的误差, 或者对一个命题附加上置信水平, 其优点是它对做出判断的过程加进了客观性, 从而可以使测试结论更加准确.

(2) 利用随机化区组设计, 执行各组处理时的条件之间的差异, 能够与处理之间的差异分离开来, 并且能与实验误差分离开来, 这种设计相当于把测试条件细分为具有相对相同条件的区组, 由此有效减少了测试需要的重复次数, 使测试更加有效, 较优的测试设计与恰到好处的测试重复数目的综合效果, 可以节省测试工作量, 这种经济上的节省, 可大大补偿为设计测试实验所花费的额外时间和脑力劳动.

(3) 能分辨误差对实际性能差距的影响, 特别适用于 2 种以上不同产品的比较.

同时, 对测试中的非统计学知识的利用也是一个需要考虑的方面, 这类非统计学知识在选择因素、确定因素水平、决定进行多少次重复、解释分析的结果等方面是非常重要的.

(下转第 1678 页)