

# 一种基于贝叶斯算法的页面置换评测系统

毛布<sup>1</sup>, 谢汶<sup>2</sup>

(1. 四川自贡广播电视大学, 四川 自贡 643000; 2 四川大学, 成都 610000)

**摘要:** 页面调度算法会影响到操作系统的性能, 文章针对操作系统的效率提高性评估原则, 提出了一种基于贝叶斯算法的页面置换策略评测系统。通过这种评测系统促进页面置换策略的改进, 从而提高操作系统的性能。

**关键词:** 贝叶斯算法; 页面置换; 评测系统

**中图分类号:** TP181

**文献标识码:** A

## 引言

评价一个算法的优劣, 可通过在一个特定的存储访问序列(页面走向)上运行它, 并计算缺页数量来实现。常见的页面置换算法有先进先出 FFO 置换算法、最久未使用算法(LRU)、最优置换算法(OPT)<sup>[1]</sup>。

最佳置换算法(OPT)是由 Belady 于 1966 年提出的一种理论上的算法。<sup>[2]</sup>其所选择的被淘汰页面, 将是以后永不使用的, 或许是在最长(未来)时间内不再被访问的页面。采用最佳置换算法, 通常可保证获得最低的缺页率。但由于人们目前还无法预知一个进程在内存的若干个页面中, 哪一个页面是未来最长时间内不再被访问的, 因而该算法是无法实现的, 但可以利用该算法去评价其它算法的优劣。

我们可通过贝叶斯算法计算当前的置换算法是否接近理想中的 OPT 算法, 从而达到改进当前所采用的置换算法来提高操作系统性能。

## 1 贝叶斯算法模型

贝叶斯理论提供了一种计算假设概率的方法, 基于假设的先验概率、给定假设下观察到不同数据的概率以及观察到的数据本身。用  $P(h)$  表示在没有训练数据前假设  $h$  拥有的初始概率。 $P(h)$  被称为  $h$  的先验概率。先验概率反映了关于  $h$  是一正确假设的机会的背景知识, 如果没有这一先验知识, 可以简单地将每一候选假

设赋予相同的先验概率。类似地,  $P(D)$  表示训练数据  $D$  的先验概率,  $P(D|h)$  表示假设  $h$  成立时  $D$  的概率, 机器学习中, 我们关心的是  $P(h|D)$ , 即给定  $D$  时  $h$  的成立的概率, 称为  $h$  的后验概率。<sup>[3]</sup>

贝叶斯公式提供了从先验概率  $P(h)$ 、 $P(D)$  和  $P(D|h)$  计算后验概率  $P(h|D)$  的方法, 公式如下:

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)} \quad (1)$$

$P(h|D)$  随着  $P(h)$  和  $P(D|h)$  的增长而增长, 随着  $P(D)$  的增长而减少, 即如果  $D$  独立于  $h$  时被观察到的可能性越大, 那么  $D$  对  $h$  的支持度越小。

## 2 利用贝叶斯算法评估页面置换算法

假定系统为某进程分配了三个物理块, 并考虑有以下页面号引用串: 7, 0, 1, 2, 0, 3, 0, 4, 2, 3, 0, 3, 2, 1, 2, 0, 1, 7, 0, 1。当进程运行时, 先将 7, 0, 1 三个页面装入内存。以后, 当进程要访问页面 2 时, 将会产生缺页中断。此时操作系统根据最佳置换算法, 将选择页面 7 予以淘汰。这是因为页面 0 将作为第 5 个被访问的页面, 页面 1 是第 14 个被访问的页面, 而页面 7 则要在第 18 次页面访问时才需调入。下次访问页面 0 时, 因它已在内存而不必产生缺页中断。当进程访问页面 3 时, 又将引起页面 1 被淘汰; 因为, 它在现有的 1, 2, 0 三个页面中, 将是以后最晚才被访问的。图 1 示出了采用最佳置换算法时的置换图。由图可看出, 采用最佳置换算法发

生了 6 次页面置换。

这里,我们仍用上面的例子,但采用 FIFO 算法进行页面置换,如图 2。当进程第一次访问页面 2 时,将把第 7 页换出,因为它是最先被调入内存的;在第一次访问页面 3 时,又将把第 0 页换出,因为它在现有的 2,0,1 三个页面中是最老的页。由图 2 可以看出,利用 FIFO 算法时进行了 12 次页面置换,比最佳置换算法正好多一倍。我们又采用了 LRU 算法,从图 3 中,我们可以看出,共发生了 9 次页面置换。

从上面三种算法中明显可以看出,OPT 优于其它两种算法。但因为 OPT 算法只是一种理想算法,实践中我们只能是尽量接近这种理想值,现在我们通过贝叶斯算法来评估系统当时采用的置换算法。

贝叶斯法则为计算给定训练数据下任一假设的后验概率提供了原则性方法,因此可以直接将其作为一个基本的学习方法:计算每个假设的概率,再输出其中概率最大的。OPT 算法是一种理想算法,我们可通过贝叶斯算法来判定当前的算法是否接近 OPT,由此来判定当前算法的适用价值<sup>[4]</sup>。

针对 OPT 算法中的页面,人为规定用正 + 表示被置换,用负 - 表示不会被置换。我们通过 LINUX 平台中抽样得出先验知识:在所页面中,根据算法被置换的概率是 0.008 通过 OPT 算法后,理想状态下应被置换的概率为 98%。通过 OPT 算法计算后,理想状态下不应该被置换的概率为 97%,根据贝叶斯表达方法总结如下:

$$P(\text{exchange}) = 0.008 \quad P(\bar{\text{exchange}}) = 0.992 \quad (2)$$

$$P(+|\text{exchange}) = 0.98 \quad P(-|\text{exchange}) = 0.02 \quad (3)$$

$$P(+|\bar{\text{exchange}}) = 0.03 \quad P(-|\bar{\text{exchange}}) = 0.97 \quad (4)$$

我们提出问题:假定发生了一次页面置换,这次页面置换是否是理想状态下的置换呢?让我们求后验概率  $P(\text{exchange}|+)$  和  $P(\bar{\text{exchange}}|+)$ 。利用公式 (1) 至 (4) 找到极大后验假设:

$$P(+|\text{exchange})P(\text{exchange}) = 0.0078 \quad (5)$$

$$P(+|\bar{\text{exchange}})P(\bar{\text{exchange}}) = 0.0298 \quad (6)$$

$$\text{hMAP} = (\text{exchange}) \quad (7)$$

确切的后验概率可将上面的结果归一化以使它们的和为 1,这时得出:

$$P(\text{exchange}|+) = 0.0078 / (0.0078 + 0.0298) = 0.21 \quad (8)$$

$$P(\bar{\text{exchange}}|-) = 0.79 \quad (9)$$

该步骤的根据在于:贝叶斯公式说明后验概率就是上面的量除以  $P(+)$ ,虽然  $P(+)$  没有作为问题陈述的一部分直接给出,但因已知  $P(\text{exchange}|+)$  和  $P(\bar{\text{exchange}}|-)$

$|-)$  的和必为 1(即,该页面要么被置换,要不不被置换),因此可以进行归一化。

该结果说明,当某个页面被置换时,该页面在理想状态下应该被置换的值为 0.21;当某个页面被置换时,该页面在理想状态下不应被置换的值为 0.79。

注意,虽然被置换的页面的后验概率比先验概率要大,但最大可能的假设仍为此页面不被置换。

根据得出的后验概率等数据,我们可以评测当前所采用的页面置换算法的优劣,以便促进对页面算法的改进。

### 3 评测模型

根据以上的评估页面置换算法,我们选取了三个不同的页面转换算法案例。借助 Netica 软件包建立评测模型<sup>[2]</sup>。

#### 3.1 建立变量集、变量域和贝叶斯网络节点<sup>[5]</sup>

在本实验中,我们参照 OPT 中得出的有效数据,它们分别代表以下含义:

$P(\text{exchange})$ : 在所有页面中,应该被置换的概率。

$P(+|\text{exchange})$ : 对确实应该置换的页面的准确率。

$P(-|\bar{\text{exchange}})$ : 对确实不应被置换的页面的准确率。利用图 1 至 3 的数据,并根据公式 (1) 至 (9) 计算后得出下表:

表 1 各算法对应数据表

	FIFO	LRU	OPT
被置换的概率	0.6	0.45	0.3
命中次数	5	8	11
$P(\text{exchange})$	0.3	0.3	0.3
$P(+ \text{exchange})$	0.5	0.67	1
$P(- \bar{\text{exchange}})$	0.5	0.34	1
$P(\text{exchange} +)$	0.3	0.47	无意义
$P(\bar{\text{exchange}} -)$	0.7	0.53	无意义

这样,我们通过列表的形式,由先验概率计算出了后验概率  $P(\text{exchange}|+)$ 、 $P(\bar{\text{exchange}}|-)$ 。

$P(\text{exchange}|+)$  意味着是当页面交换发生时,这次页面交换应该发生的概率。 $P(\bar{\text{exchange}}|-)$  意味着是当页面交换没有发生时,这次页面交换不应该发生的概率。

#### 3.2 建立评测模型

我们可以同过抽样举例,例出更多的数据,通过计算得出后验数据。这样我们就可以建立一个评测模型<sup>[6]</sup>。

本系统采用了简单直观的评价模型,只考虑了在某

次页面置换下 FIFO、LRU 与 OPT 的比较。当在理想状态时, 我们求后验概率无实际意义。但通过表 1 我们得到了一个比较不同页面置换算法的数据表。从表 1 中可以看出, LRU 算法明显在此例页面置换中优于 FIFO 算法。我们又通过表 1 归纳得出了评测模型表 2。但更精确的评价模型还需要引入概率学、机器学习等相关理论, 需要作为一个相对独立的课题继续展开研究。

表 2 评测模型表

初始数据	$P(\text{exchange}) P(+ \text{exchange}) P(- \text{exchange})$
输出数据	$P(\text{exchange} +) P(\text{exchange} -)$
比较	通过 $P(\text{exchange} +) P(- \text{exchange} -)$ 的值来判断在某例页面置换中, 算法的优劣性。 $P(\text{exchange} +)$ 越趋于 1 表明越理想。

#### 4 结束语

本文研究表明, 基于贝叶斯算法的页面置换评测系统可借助 Netica 软件包建立评测模型。传统评页面置换算法的优劣都是通过计算“命中次数”来完成, 而本

文所提出的基于贝叶斯的推算方法是一种基于先验数据的算法, 最后的结果值很大程度上也依赖先验数据, 那么先验概率  $P(\text{exchange})$  就是一个比较关键的数据了, 那么我们就需要通过在操作系统平台上, 实际运行后采样数据, 尽可能的得到更为真实的数据。

#### 参考文献:

- [1] 汤子赢. 操作系统. 3 版 [M]. 西安: 西安电子科技大学出版社, 2007.
- [2] Jensen F V. An introduction to Bayesian networks [M]. New York: Springer Veriage
- [3] Tom M Mitchell 机器学习. 1 版 [M]. 北京: 机械工业出版社, 2003
- [4] 孟宪福, 陈 莉. 基于贝叶斯理论的协同过滤推荐算法计算机应用, 2009, 29(10): 2733-2735.
- [5] 曹卫东, 贺国光. 连续航班延误与涉及的贝叶斯网络分析 [J]. 计算机应用, 2009, (2): 606-610
- [6] 曹 晖. 基于动态贝叶斯博弈的攻击预测模型 [J]. 计算机应用, 2007, 27(6): 1545-1547

## A Page Replacement Evaluation System Based on Bayesian Theory

MAO Bu<sup>1</sup>, XIE Wen<sup>2</sup>

(1. Sichuan Television and Broadcast College, Zigong 643000, China; 2. Sichuan University, Chengdu 610000, China)

**Abstract** Since paging calculation will influence the performance of operating system, this paper focuses on the evaluation principles of upgrading operating system efficiency and proposes an evaluation system of page-replacement strategy that is based on Bayesian Optimization Algorithm. Which aims to improve the replacement strategy and enhance the performance of operating system.

**Key words** Bayesian theory; page replacement; evaluation system