MTA Számitástechnikai és Automatizálási Kutató Intézete, Közlemények 18/1977.

A BAYESIAN APPROACH TO THE PROBLEM OF OPTIMAL PROGRAM–PAGING STRATEGIES

A. Benczur – A. Krámli

Several works in the recent literature on computer sciences are devoted to the optimal program-paging problem (see e.g. Franaszek and Wagner [1], Denning [2]. Starting from the standpoint of these works M. Arató [3] proposed a statistical approach to this problem using Bayesian sequential decision theory developed by Bellman (see. e.g. [4]).

We give the most important special case the exact proof of the optimality of the "least frequently used" strategy

We repeat briefly the model used by Arató, to make more understandable this note.

The program consists of n pages and the computer has a twolevel hierachical memory. We suppose that k pages can be stored in the high speed memory (central memory), and $n-k$ pages (in fact the whole program) is stored on a slower access memory device (disk).

The sequence of references to the different pages forms a sequence $\{\eta_t\}$ of i.d.d. random variables with unknown probability distribution

$$P_i = P(\eta_t = i) \qquad i = 1,2, \ldots, n.$$

In Arató's model the pages can be rearranged before each reference without extra cost. (Such a situation is realized if the pages are subroutines of a main program, and between two consecutive calls there is a relatively great time interval.) But if the main program calls a routin, which is absent from the central memory the cost increases with 1 unity.

As in the case of known distribution $\{P_i\}$ the optimal strategy depends only on the order of probabilities P_i (before every call we must store the first k most probable pages in the central memory) we suppose, that the unkown parameter of the distribution is the permutation w of n elements giving the right order of P_i-s in the following way:

If $\{P_1, \ldots, P_n\}$ is a decreasing sequence of probabilities $(P_1 + \ldots + P_n = 1)$, then $P_i = p_{w(i)}$, where $w(i)$ denotes the one to one mapping realized by the permutation w.

By the Bayesian principle we suppose that the parameter w is also a random variable. If we have no previous information about its distribution, we can consider every permutation equally like, i.e. the apriori distribution ξ of n is the uniform distribution on $n!$ elements. In Arató's work is treated the case of arbitrary apriori distribution. Let us denote by D_N the set of sequential decision procedures $\{d_0, \ldots, d_{N-1}\}$ for a program consisting of N step. By Arató's model d_t is a subset of the labels of the pages, wich are absent from the central memory after the observation of references

$$\eta_1, \ldots, \eta_t.$$

Before the program is started we define d_0 on the basis of the a priori distribution ξ; by the symmetriry of ξ, d_0 can be choosen arbitrarily.
Let us introduce the random process $x_t^{d_{t-1}}$

$$X_t^{d_{t-1}} = \begin{cases} 0 & \text{if} \quad \eta_t \notin d_{t-1} \\ 1 & \text{otherwise} \end{cases}$$

By this terms the utility function has the form:

$$V(\xi,N) = \max_{\{d_1, \ldots, d_{N-1}\} \in D_N} E_\xi\left(\sum_{t=1}^{N} X_t^{d_{t-1}}\right) \tag{1}$$

E_ξ means the expectation taken on the basis of the uniform a priori distribution ξ. Our aim is to determine the class $D'_N \subset D_N$ of sequential decision procedures for which the maximum in (1) is reached. The solution of this problem can be obtained by solving recursively the Bellman equation

$$V(\xi(\eta_1, \ldots, \eta_{t-1}),\ N - t + 1) = \max_{d_{t-1}} E_{\xi(\eta_1, \ldots, \eta_{t-1})}(X_t^{d_{t-1}}) + \tag{2}$$

$$+ V(\xi(\eta_1, \ldots \eta_t), N - t),$$

where $\xi(\eta_1, \ldots, \eta_t)$ means the aposteriori distribution of the parameter w after having observed the references $\eta_1, \ldots, \eta_t$.

As the aposteriori distribution does not depend on the decision procedure the right hand side of (2) has maximum for such a d_{t-1} which maximize $E_{\xi(\eta_1, \ldots, \eta_{t-1})}(X_t^{d_{t-1}})$.

We shall prove the following

Lemma 1. *The expectation* $E_{\xi(\eta_1, \ldots, \eta_{t-1})}(X_t^{d_{t-1}})$ *reaches its maximum for the decision* d_{t-1} *consisting of the* $n - k$ *least frequently used pages.*

If there are pages of equal frequencies their changing by each other has no influence on the expectation

$$E_{\xi(\eta_1, \ldots, \eta_{t-1})}(X_t^{d_{t-1}})$$

The proof of the Lemma can be carried out by direct calculation and comparison of the aposteriori probabilities.

First we calculation the aposteriori probabilities of a fixed permutation w. If we denote by W the set of all permutations of natural numbers $1 \ldots n$, and by $\{k_1, \ldots, k_n\}$ the frequencies of pages $(k_1 + \ldots + k_n = t - 1)$, then

(3)

$$P(w_0|\eta_1, \ldots, \eta_{t-1}) = \frac{P(\eta_1, \ldots \eta_{t-1} | w_0)}{\sum_{w \in W} P(\eta_1, \ldots, \eta_t | w)} = C \cdot \prod_{i=1}^{n} P_{w_0(i)}$$

It is sufficient to prove that for every pair of frequencies

$$k_r \leq k_s$$

$$P(\eta_t = r|\eta_1, \ldots \eta_{t-1}) = \sum_{w \in W} \prod_{i=1} p_i^{k_{w(i)} + \delta_{w(i),r}}$$

(4)

$$P(\eta_t = \Delta|\eta_1, \ldots, \eta_{t-1}) = \sum_{w \in W} \prod_{i=1} p_i^{k_{w(i)} + \delta_{w(i),s}}$$

where $\delta_{i,j}$ is the Kronecker's symbol. Inequality (4) follows by comparing the terms on the left and right hand side, which belong to pairs (w,w^*) of permutations with the following properties:

There on $i < j \leq n$ are two fixed natural numbers for which

$$w(i) = r \qquad w(j) = s$$
$$w^*(i) = s \qquad w^*(j) = r$$

and $w(k) = w^*(k)$ for every $k \neq i,j$.

The second assertion of Lemma follows from the symmetry of inequality (4). Lemma 1 and the Bellman equation written in the form (2) gives.

THEOREM 1. For the uniform a priori distribution ξ of the parameter w, and for every N the optimal Bayesian sequential decision procedure is the "least frequently used" stategy.

References

[1] P. A. Franaszek and T. J. Wagner, Some distribution-free aspects of paging algorithm performance Journ. ACM. 21 (1974), 1, 31-39.

[2] P. J. Denning, Virtual memory, Computing Surveys, Vol. 2, N^{0}3, (1970), 153-89.

[3] M. Arató, On optimal performance of page storage hierarchies, to appear

[4] M.H. DeGroot, Optimal Statistical Decisions Mc Graw-Hill, (1970).

Ö s s z e f o g l a l ó

Optimális program lapolási eljárások Bayes-féle tárgyalása

Benczúr A. – Krámli A.

A dolgozat a Bayers-féle szekvenciális döntéselmélet segitségével igazolja, hogy bizonyos feltételek mellett a "legritkábban használt lap" stratégiája minimalizálja a lapolási hibák várható számát.

Р Е З Ю М Е

БАЙЕСОВСКИЙ ПОДХОД К ПРОБЛЕМЕ ОПТИМАЛЬНОГО ПОСТРАНИЧНОГО ПРОГРАММИРОВАНИЯ

А. Бенцур - А. Крамли

В работе, используя Байесовскую теорию последовательного решения, доказывается, что при некоторых условиях стратегия "наименее часто использованной страницы" минимизирует ожидаемое число страничного сбоя.