A NEW CRITERION FOR COMPARISON OF SELF ORGANIZING SCHEMES

Pratiksha Bhawalkar[1], Dr. Meenakshi Sagdeo[2]

***Abstract:** Researchers have devised different schemes for quick and less costly retrieval of records stored in various data structures. In linear array, as the records are arranged sequentially, the cost to locate the requested record is directly proportional to its position in the array. To minimize the search cost different Self Organizing Schemes (SOS) have been developed wherein records are arranged almost according to their request probabilities in the long run. Two well known self organizing schemes are Transposition Scheme (TR) and Move-to-Front Scheme (MTF). Many researchers have studied different SOS and compared their performances on the basis of Average Search Cost (ASC). It has been shown, that TR scheme always has lesser ASC than that of MTF scheme [2]. In this paper we consider a modified Move-to-Front scheme (MMTF), which has ASC between that of TR and MTF. Further we develop a new criterion to compare various self organizing schemes. We present the output of the computer programs for substantiating the theoretical results. We also discuss the relation between ASC and the new criterion under study.*

***Keywords:** Linear search, Move-to-Front scheme, Optimal ordering, Self organizing schemes, Transposition scheme.*

1. INTRODUCTION

The concept of self organizing scheme was introduced by McCabe [1] in 1965. In his pioneering work he introduced the Move-to-Front scheme and Transposition scheme. Move to Front (MTF) and Transposition (TR) are two well known self organizing schemes.

In this paper we consider fixed number of records stored in a linear array. In MTF scheme, requested record is placed in the first position after use by moving other records backward to make a room for it. In TR scheme the requested record interchanges its position with the preceding record. If the requested record is at the first position it retains its position in MTF as well as TR scheme.

The MTF and TR schemes were devised under the following assumptions and we have also devised a new self- organizing-scheme (Modified Move- to- Front scheme) under the same assumptions.

Assumptions

- There are a n records, where n is a fixed positive integer.
- The records are requested independently of all other requests.

- Only one record is requested at any instance of time.
- The requests probabilities of various records are unknown and non- zero but remains constant throughout the search.
- No record of previous request is kept.

In section 2, we give notations and definitions needed for analysis. In section 3, we give the general method of analysis of any SOS. In section 4, we describe our new criterion for comparison, in sections 5, 6 and 7 we derive the results regarding the new criterion for MTF, TR and MMTF schemes. In section 8, we give comparison of TR, MMTF and MTF schemes on the basis of the new comparison criterion. We also present the results of related computer programs. Finally in section 9, we give conclusions.

2. NOTATIONS AND DEFINITIONS

Notations:

n: Total number of records.
R_i: Record number i (i = 1…n)
p_i: Request probability for R_i
$\underline{p}$: Request probability vector of order 1 x n
$\underline{p} = [p_1 \ \ p_2 \ \dots \ p_n]_{1xn}$ where $p_i > 0$ and $\sum_{i=1}^{n}(p_i) = 1$
[E(X)]: Expected number of records moved including the requested record.
S_n: Sum of squares of first n natural numbers = n.(n+1).(2n+1)/6

Definitions:

1. Matrix Δ:

We define the matrix Δ of order n x n as follows:
$\Delta = \{\partial\ (i, j)\}_{nxn}$
Where $\partial(i, j)$ is the stationary probability that i^{th} record (R_i) is in j^{th} position (i, j=1,2….n)
Every record will occupy one of the positions from 1 to n and every position will be occupied by one of the records. Hence, $\sum_{i=1}^{n}(\delta(i, j)) = 1$ for j = 1…n and $\sum_{j=1}^{n}(\delta(i, j)) = 1$ for i = 1…n Thus, Δ is a doubly stochastic matrix.

2. Average Search Cost (ASC):

Average search cost of a record is its position in the list in the long run.

Let C_i denote the expected position or search cost of i^{th} record (R_i) in sequential search.

Therefore, $C_i = \sum_{j=1}^{n}(j\ \delta(i, j))$

Where $\delta(i, j)$ = stationary probability that i^{th} record (R_i) is in the j^{th} position (i, j = 1…n)
Average search cost for any scheme is defined as
ASC = $\sum_{i=1}^{n}(p_i C_i)$ (2.1)

3. Optimal Ordering:

When the records are arranged according to decreasing order of their request probabilities, they are said to be in optimal ordering. However in real life, this optimal ordering is never achieved.

Now we introduce a new self organizing scheme and call it Modified Move- to-Front Scheme

4. Modified Move- to- Front scheme (MMTF)

In Modified Move- to-Front scheme (MMTF) we divide the set of 'n' records into two parts and apply MTF separately to these parts. Consider a position 'k' such that $1 \leq k \leq [n/2]$, where [n/2] is the integer part of (n/2). The MMTF scheme operates in the following way.

(i) If the record is requested from the first position, then it will retain its position.

(ii) If the record is requested from the position 'j' (j= 2 to k) it will be placed at position '1'by shifting the records originally in the positions 1 to (j - 1) one position backward to make room for it.

(iii) If the record is requested from the position 'j' (j= k+1 to n) it will be placed at position 'k' by shifting the records originally in the positions k to (j - 1) one position backward to make room for it.

3. ANALYSIS OF A SELF ORGANIZING SCHEME

In this paper we consider a special request probability distribution, where only one of the 'n' records is having relatively very high request probability say, 'p' and the remaining (n - 1) records have the same request probability say 'r'. Let 'R_1' be the record with request probability 'p'.

Thus,$\underline{p}=[p\ r\ r.\ r]_{1xn}$ such that, $p + (n - 1)r=1$ and $p>r>1$ (3.1)

Let the i^{th} state of the Markov Chain be defined as the arrangement of records where, the record R_1 occupies the i^{th} position (i= 1… n)

The TPM P in this case will be of order n x n. That is $P = \{p_{jk}\}_{nxn}$ j, k = 1 …n where p_{jk} is the probability that record R_1 goes from position j to position k. Let the stationary probability vector for 'R_1' be denoted by $\underline{\pi} = [\pi_1, \pi_2 \ldots \pi_n]_{1x\,n}$ where π_j is the stationary probability that R_1 will be in the position j and $\sum_{j=1}^{n}(\pi_j) = 1$. The vector $\underline{\pi}$ is obtained by solving the system of 'n' equations given by $\underline{\pi} = \underline{\pi}$ P. This is equivalent to solving $\pi_k = \sum_{j=1}^{n} (\pi_j)\, p_{jk}$ such that $\pi_k \geq 0$ and $\sum_{k=1}^{n} (\pi_k) = 1$.

Using $\underline{\pi}$ we can find out the ASC.

$$ASC = \sum_{i=1}^{n} (p_i C_i) = \sum_{i=1}^{n} p_i \left(\sum_{j=1}^{n} j\, \delta(i,j) \right)$$

$$= p \sum_{j=1}^{n} (j\, \pi_j) + (n-1)\, r \sum_{j=1}^{n} (j\, (1- \pi_j) / (n-1))$$

On simplification we get,

$$ASC = (p - r) \sum_{j=1}^{n} j\, \pi_j + \frac{rn(n+1)}{2} \tag{3.2}$$

We can compare the different SOS on the basis of ASC.

4. NEW COMPARISON CRITERION:

Motivation:

Different self organizing schemes were compared on the basis of their average search cost (ASC) till now. We thought of devising new comparison criteria which can be used to compare different SOS. We find out the expected number of records moved in the long run. If more records are displaced under a given scheme, it means more cost will be involved in the long run. Thus, we can compare various SOS on the basis of this new criterion.

We define a new variable [E(X)] which represents the expected number of records moved backward, in the scheme under consideration. Let $[E(X)]_j$ represent the expected number of records moved backward when record 'R_1' is in position 'j', where j = 1, 2, …, n

$$\text{Therefore, } [E(X)] = \sum_{j=1}^{n} \pi_j\, [E(X)]_j \tag{4.1}$$

Here $[E(X)]_j$ will be obtained according to the scheme under consideration.

We will find out [E(X)] for different SOS. We may attach suffix to [E(X)] to denote the SOS under consideration. In the next section we find out [E(X)] for MTF Scheme.

5. NEW CRITERION FOR MTF SCHEME:

[E(X)] for MTF Scheme is given by,

$$[E(X)]_{MTF} = \sum_{j=1}^{n} \pi_j\, [E(X)]_j \tag{5.1}$$

First we find $[E(X)]_j$ for j = 1, …, n under MTF scheme.

Method to find $[E(X)]_j$ for different values of j under MTF scheme:

Now, $[E(X)]_1$ represents the expected number of records moved when 'R_1' is in the first position.

$$[E(X)]_1 = 0.p + 1.r + 2.r + \ldots, + (n-1).r = \left[\frac{n(n-1)}{2}\right] r$$

In general, $[E(X)]_j$ represents the expected number of records moved when record 'R_1' is in the j^{th} position (j = 1, …, n) is given by

$$[E(X)]_j = (j - 1).p + \left[\frac{n(n-1)}{2} - (j - 1)\right] r \qquad (5.2)$$

Now we present the formula for $[E(X)]_{MTF}$ as a theorem.

Theorem 5.1:

[E(X)] for MTF scheme is given by

$$[E(X)]_{MTF} = (p - r) \sum_{j=1}^{n} \pi_j (j - 1) + \frac{nr(n-1)}{2}$$

Proof :

Equation (5.1.2) for MTF Scheme is given by

$$[E(X)]_j = (j - 1).p + \left[\frac{n(n-1)}{2} - (j - 1)\right] r, \quad j = 1, \ldots, n.$$

The value of [E(X)] under the MTF scheme is given by $[E(X)]_{MTF}=\sum_{j=1}^{n} \pi_j$

$$[E(X)]_j=\sum_{j=1}^{n} \pi_j \left[(j - 1).p + \left[\frac{n(n-1)}{2} - (j - 1)\right] r\right]$$

$$=(p-r)\sum_{j=1}^{n} \pi_j(j-1)+\frac{nr(n-1)}{2} \qquad (5.3)$$

This proves the theorem.

6. NEW COMPARISON CRITERION FOR TR SCHEME

[E(X)] for TR Scheme is given by,

$$[E(X)]_{TR} = \sum_{j=1}^{n} \pi_j [E(X)]_j \qquad (6.1)$$

First we find $[E(X)]_j$ for j = 1, …, n under TR scheme.

Method to find $[E(X)]_j$ for different values of j under TR scheme:

Now, $[E(X)]_1$ represents the expected number of records moved when 'R_1' is in the first position.

$$[E(X)]_1 = 0.p + 1.r + 1.r + \ldots, + 1.r = (n - 1) r$$

$[E(X)]_2$ represents the expected number of records moved when record 'R_1' is in the second position.

$$[E(X)]_2 = 0.r + 1.p + 1.r + \ldots, + 1.r = 1.p + (n - 2)r$$

ISBN - 978-93-81583-55-5

In general, $[E(X)]_j$ represent the expected number of records moved when record 'R_1' is in the j^{th} position.(j = 1,…, n) we get,

$[E(X)]_j$ =0. r+1.r+1.r, .+ 1.p …,+1.r = 1.p +(n–2)r (6.2)

In the next section we present the formula for $[E(X)]_{TR}$ as a theorem.

Theorem 6.1:

[E(X)] for TR scheme is given by,$[E(X)]_{TR} = (nr–1)\pi_1 + (1- r)$

Proof:

Recall that $[E(X)]_j$ for TR Scheme is given by

$[E(X)]_j = 1. p + (n – 2) r$, for j = 2, …, n.

The value of [E(X)] under TR scheme is given by

$[E(X)]_{TR}=\sum_{j=1}^{n} \pi_j [E(X)]_j = (n - 1) r \pi_1 + \sum_{j=2}^{n}[p + (n – 2) r] \pi_j$

However,$\sum_{j=1}^{n} \pi_j =1$ Therefore, $\sum_{j=2}^{n} \pi_j =1–\pi_1$ and p+ (n – 1) r = 1 Therefore, p + (n – 2) r = 1 – r

Thus,$[E(X)]_{TR} =(n–1)r\pi_1 +(1 – r)(1 - \pi_1)$

$= (nr – 1) \pi_1 + (1 – r)$ 6.3

This proves the theorem.

7. NEW COMPARISON CRITERIA FOR MMTF SCHEME

[E(X)] for MMTF Scheme is given by,

$[E(X)]_{MMTF} = \sum_{j=1}^{n} \pi_j [E(X)]_j$

$= \sum_{j=1}^{k} \pi_j [E(X)]_j + \sum_{j=k+1}^{n} \pi_j [E(X)]_j$ (7.1)

First we find $[E(X)]_j$ for j = 1, …, n under MMTF scheme.

Method to find $[E(X)]_j$ for different values of j under MMTF scheme:

Let $[E(X)]_1$ represent the expected number of records moved backward when 'R_1' is in the first position.

$[E(X)]_1$ =0.p+1.r+2.r + …, +(k-1)r + 1.r + 2.r + …, + (n - k). r

$= \left[\frac{k(k-1)}{2}\right] r + \left[\frac{(n-k)(n-k+1)}{2}\right] r$

In general, for j= 1, …, k.

$[E(X)]_j = (j-1).p + \left[\frac{k(k-1)}{2} - (j-1)\right] r + \left[\frac{(n-k)(n-k+1)}{2}\right].r$ (7.2)

$[E(X)]_{k+1}$=0.r +1.r+ 2.r + …, +(k–1).r+[1.p+ 2.r +.,+(n – k).r]

In general, for j > k, where j= k+1, …, n

$[E(X)]_j=\frac{k(k-1)}{2}.r+[(j–k).p+(\frac{(n-k)(n-k+1)}{2}.– (j– k)).r]$ (7.3)

Now we present the formula for $[E(X)]_{MMTF}$ as a theorem.

Theorem 7.1:

[E(X)] for MMTF scheme is given by

$$[E(X)]_{MMTF}=ASC_{MMTF}-(n-k+1)kr-(p-r)[(k-1)\sum_{j=k+1}^{n}\pi_j-1]$$

Proof:

The value of [E(X)] under the MMTF scheme is given by

$$[E(X)]_{MMTF}=\sum_{j=1}^{n}\pi_j\,[E(X)]_j=\sum_{j=1}^{k}\pi_j\,[E(X)]_j+\sum_{j=k+1}^{n}\pi_j\,[E(X)]_j$$

From equation (7.2) and (7.3) we get,

$$[E(X)]_{MMTF}=\sum_{j=1}^{k}\pi_j\{(j-1).p+\left[\frac{k(k-1)}{2}-(j-1)\right]r+\left[\frac{(n-k)(n-k+1)}{2}\right].r\}$$

$$+\sum_{j=k+1}^{n}\pi_j\{\frac{k(k-1)}{2}.r+[(j-k).p+(\frac{(n-k)(n-k+1)}{2}.-(j-k)).r]\}$$

$$=\sum_{j=1}^{n}\pi_j\{[\frac{k(k-1)}{2}+\frac{(n-k)(n-k+1)}{2}]r\}+\sum_{j=1}^{k}\pi_j\{(j-k)(p-r)\}+\sum_{j=k+1}^{n}\pi_j\{(j-k)(p-r)\}$$

$$=k(k-1-n)r+\frac{n(n+1)}{2}r+(p-r)[\sum_{j=1}^{n}j\pi_j-(k-1)\sum_{j=k+1}^{n}\pi_j-1]$$

$$=ASC_{MMTF}-(n-k+1)kr-(p-r)[(k-1)\sum_{j=k+1}^{n}\pi_j-1]$$

This proves the theorem.

In the next section we compare different SOS on [E(X)].

8. COMPARISON OF DIFFERENT SELF ORGANIZING SCHEMES ON THE BASIS OF NEW CRITERION [E(X)]:

Now we write a C++ program for all the formulae obtained using new comparison criteria E(X) for different SOS. We then compare different SOS by using these results.

Table I
Comparison of E(X) and ASC for different SOS
(n = 10 and k=5)

	p = 0.4		p = 0.6		p = 0.7	
	E(X)	ASC	E(X)	ASC	E(X)	ASC
TR	0.655555	4.06667	0.441152	3.044444	0.331746	2.533334
MMTF	1.85846	4.197063	1.264444	3.153637	0.954558	2.621275
MTF	3.428571	4.428571	2.344828	3.344828	1.772727	2.772727

ISBN - 978-93-81583-55-5

Observations:

From the Table I it is observed that

1. As 'p' increases the value of E(X) decreases for all self organizing schemes. Thus, if request probability of record 'R_1' increases then the rate of convergence also increases

2. Also for the fixed value of 'n' and 'p' for any SOS we have, $[E(X)]_{TR} < [E(X)]_{MMTF} < [E(X)]_{MTF}$

3. As 'p' increases, the average search cost decreases for all SOS.

Table II
Comparison of E(X) and ASC for different SOS
(p = 0.7 and k = 5)

	n = 10		n = 15		n = 20		n = 25	
	E(X)	ASC	E(X)	ASC	E(X)	ASC	E(X)	ASC
TR	0.331746	2.533334	0.320773	3.271429	0.315433	4.01579	0.312277	4.76250
MMTF	0.954558	2.621275	1.473483	3.330639	2.113004	4.060378	2.798246	4.798248
MTF	1.772727	2.772727	2.532178	3.532178	3.286765	4.286765	4.039474	5.039474

Observations:

From the table II it is observed that

1. As the number of records increases for the TR scheme, the value of [E(X)] decreases and for the MMTF and MTF scheme, the value of [E(X)] increases. Therefore for TR scheme there will be less number of records moved as number of records increases compared to other two schemes, which shows that TR scheme is slower in reaching optimal ordering.

2. Also notice that, for fixed number of records and fixed 'p',

3. $[E(X)]_{TR} < [E(X)]_{MMTF} < [E(X)]_{MTF}$

Table III
Comparing different values of $[E(X)]_{MMTF}$ and ASC_{MMTF}
(p=0.7)

	n = 10		n = 15		n = 20		n = 25		n = 30	
	E(X)	ASC	E(X)	ASC	E(X)	ASC	E(X)	ASC	E(X)	ASC
k=5	0.95455	2.62127	1.47348	3.33063	2.11300	4.06037	2.79824	4.79824	3.50569	5.54017
k=7	-	-	1.34236	3.37093	1.85897	4.09055	2.47236	4.82236	3.13267	5.56026
k=10	-	-	-	-	1.71478	4.13583	2.17105	4.85855	2.72832	5.59039
k=12	-	-	-	-	-	-	2.09517	4.88267	2.56220	5.61047

Observations:

From the table III it is observed that

1. As 'n' increases the corresponding value of [E(X)] and ASC increases for MMTF scheme. Therefore if we increase number of record it results in more displacement of record, which will lead to increase in the cost, but will be faster in reaching optimal ordering.

2. Also it is observed that as 'k' increases for MMTF scheme, the value of [E(X)] decreases.

9. CONCLUSION

The expected number of records moved in the long run can thus be used to compare various self organizing schemes. Further, it has been proved by the researchers that TR is the slowest in attaining optimal ordering amongst all the SOS keeping no record of previous request. The proposed new criterion also throws light on this property of TR. In case of TR as 'n' increases, ASC also increases. This means that as 'n' increases, the expected position of 'R_1' in the long run increases. At the same time [E(X)] that is the expected number of records moved decreases. Therefore, TR scheme becomes slower in reaching optimal ordering than the other schemes when 'n' increases.

Thus, the new comparison criterion can be used to compare various SOS as it is related to the average search cost. At the same time the relative changes in [E(X)] and ASC reflect the speed of convergence to the optimal ordering.

10. REFERENCES

1. Mc Cabe, J. (1965) On Serial Files with Relocatable Records. *Operations Research,12,p 601-618.*
2. HENRICKS W.J.(1972) The Stationary Distribution of an Interesting Markov Chain.*J Appli. prob., p 231-233.*
3. Rivest R. (1976) On Self Organizing Sequential Search Heuristics.*Comm.Of the A. C. M., p 63 -67*
4. Tenenbaum A. (1978) Simulation of Dynamic Sequential Search Algorithms. *Comm. Of the A. C. M., 21, p 790-791.*
5. Bitner J. R. (1979) Heuristics That Dynamically Organize Data Structures.*SIAM J. COMPUT. Vol. 8, p 82 – 110.*
6. Phatarfod R.M. (1994) On the Transition Probabilities of the Move - To - Front Scheme. *J Appli. Prob., 31, p 570 – 574.*
7. Sagdeo M. M, (2006), Ph. D. Thesis "Some Contributions to the Theory of Probabilistic Aspects of Linear Search in Dynamically Organized Linear Arrays." (RTM Nagpur University)

****

1.Pratiksha Bhawalkar , Assistant Professor, Thakur Institute of Management Studies & Research, Kandivali, Mumbai.
2.Dr. Meenakshi Sagdeo, Associate Professor, Govt. I. Y. College, Jogeshwari, Mumbai.