Some Results on Optimal Search in Discrete Spaces*

ZHU Qing-xin[1], ZHOU Ming-tian[2], John Oommen[3]

[1](State Key Laboratory of Broadband Optical Fiber Transmissions and Communication Networks, University of University of Electronic Science and Technology of China, Chengdu 610054, China);

[2](School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China);

[3](School of Computer Science, Carleton University, Ottawa, Ontario K1S 5B61, Canada)

E-mail: qxzhu@263.net

http://www.uestc.edu.cn

Received October 12, 2000; accepted May 9, 2001

Abstract: In this paper the searching problem for an object in a set of N locations is considered. The goal is to allocate the resources so as to maximize the probability of locating the object. By using Lagrangian operator method the problem of optimal search with the unknown target distribution is studied. Some selection criteria and error estimate results are derived.

Key words: optimal search; detection function; target distribution; Lagrangian operator; error estimate

Optimal search problem involves computing how to allocate resources (for example, searching time) so as to maximize the probability of detection, or to minimize the expected consumption (cost) of resources used for detecting the target[1]. One example is searching for a record X in a distributed database located at one of N storage sites, denoted by $\{C_1, \ldots, C_N\}$. The probability that we will find X when searching C_i given that X is indeed in C_i is typically called the detection function, denoted by $b(i,t)$. Let K denote the total time that is assigned for searching, the optimal search problem involves allocating time $k_i(\Sigma k_i = K)$ for each site C_i such that either the probability of detecting target X is maximum, or the expected cost (time consumed) to find target X is minimum. This allocation of time is called the optimal search strategy. The optimal search theory has many applications, including developing military and strategic policy[2,3].

The study of the optimal search for a stationary target was pioneered by B.O. Koopman[4]. Since then many authors[1~3] have generalized the Koopman's work, but all of these results were obtained under the assumption that the target distribution is known in advance, obviously it is not true in many cases. In Refs.[5,6] we investigated the problem of optimal search when the target distribution is unknown, to our knowledge these were the first available results in this aspect. In Ref.[7] we considered an optimal search problem with non-regular detection

* Supported by the Visiting Scholar Foundation of Kay Laboratory in University (高等学校重点实验室访问学者基金); the Returned Overseas Student Research Fundation of the Ministry of Education of China (归国留学人员基金)

ZHU Qing-xin was born in 1954. He is a professor of the University of Electronic Science and Technology of China. His research interests are control theory, computer networks and communications, and the network information securities. **ZHOU Ming-tian** was born in 1939. He is a professor and doctoral supervisor of the University of Electronic Science and Technology of China. His research interests are computer networks, distributed and parallel processing, distributed object technology, and the network information securities. **John Oommen** was born in 1953. He is a professor and doctoral supervisor of the Carleton University, Canada. His research interests are learning system, pattern recognition, and adaptive data structure.

function. In this paper we shall further derive the selection criteria and error estimate results for the optimal search problem with the unknown target distribution. Although these results are obtained for discrete systems, they can be easily extended to the continuous case.

1 Upper and Lower Bounds of Optimal Search Strategies

First we derive the upper and lower bounds of optimal strategies when the detection function is exponential.

Theorem 1. Suppose an object is located in one of N cells with an arbitrary probability distribution $p(i)$, $(i=1,2,\ldots,N)$. Assume that the detection function is exponential:

$$b(i,z)=1-e^{-z},\ (1\leqslant i\leqslant N).$$

Then the detection probability $P[f^*]$ of the optimal search strategy f^* satisfies the following inequality:

$$1-e^{-K/N}\leqslant P[f^*]\leqslant 1-C_N e^{-K/N},$$

where $C_N=N[p(1)\ldots p(N)]^{1/N}$ and $K>0$ is the total cost constraint.

Proof. Without loss of generality we may assume that

$$p(i)>0,\ \sum_{i=1}^{N}p(i)=1.$$

Consider the Lagrangian

$$l(i,\lambda,z)=p(i)b(i,z)-\lambda(z).$$

Solve $\partial l/\partial z=0$ to yield the optimal search plan:

$$z_i=f_\lambda^*(i)=\ln\frac{p(i)}{\lambda},\ \sum_{i=1}^{N}z_i\leqslant K.$$

That is

$$\sum_{i=1}^{N}\ln\frac{p(i)}{\lambda}\leqslant K,\ \lambda\geqslant[p(1)\ldots p(N)]^{1/N}e^{-K/N}.$$

The equality is achieved in the boundary. Thus we can easily derive the upper and lower bounds for $P[f^*]$. This completes the proof of the theorem. □

Similarly we can generalize the above result to the case of general regular detection function $b(i,z)$. That is, $b(i,z)$ is continuously differentiable and the derivative $b'(i,\cdot)$ is a decreasing function with $b'(i,0)>0$ and $b'(i,\infty)=0$.

Theorem 2 (Bounds of optimal strategies for regular detection functions). Suppose an object is located in one of N cells with an arbitrary probability distribution $p(i)$, $(i=1,2,\ldots,N)$. Assume that the detection function is regular with $b'(i,\cdot)$ decreasing and satisfying $b'(i,0)>0$ and $b'(i,\infty)=0$. Then the detection probability $P[f^*]$ of the optimal search strategy f^* satisfies the following inequality:

$$b(i,0)+\lambda K\leqslant P[f^*]\leqslant b(i,0)+b'(i,0)K.$$

2 Selection Criteria and Error Estimates

Suppose an object is hidden in one of N locations, but the probability distribution is unknown. The first issue to be considered in designing a search strategy is to make a good guess for the target distribution $q(i)$, $(i=1,2,\ldots,N)$, which will maximize the detection probability.

Suppose that the target distribution is $(a_1,a_2,\ldots,a_N)$ but the searcher chooses $(b_1,b_2,\ldots,b_N)$ as the approximation of the target distribution. Intuitively we would think that the optimal search plan which is based on the true distribution should have a superior detection probability than that of the search plan which is based on the guessed one. This intuition is justified below.

Theorem 3. In searching N cells with an exponential detection function, suppose that the true target distribution is $a=(a_1,a_2,...,a_N)$ and that the searcher assumes the target distribution is $b=(b_1,b_2,...,b_N)$. Let $P[f_a^*]$ denote the detection probability of the optimal search plan f_a^* based on the distribution a and $P[f_b^*]$ be the detection probability of the optimal search plan f_b^* based on the distribution b, then $P[f_b^*]\leqslant P[f_a^*]$.

We omit the proof of Theorem 3 here. The interested readers may contact the first author for a strict and complete proof of this theorem.

From the proof of Theorem 3 we have the following corollaries:

Corollary 4 (First criterion for choosing target distributions). In the search of N cells with the exponential detection function, one criteria for choosing the target distribution $b=(b_1,b_2,...,b_N)$ is to minimize the following expression

$$D(a,b)=e^{-K/N}\left(\prod_{i=1}^{N}b_i\right)^{1/N}\sum_{i=1}^{N}\frac{a_i}{b_i}$$

for all possible values of $a=(a_1,a_2,...,a_N)$.

Corollary 5 (Error estimates).

In the search of N cells with the exponential detection function, if the real target distribution is $a=(a_1,a_2,...,a_N)$ and the chosen distribution is $b=(b_1,b_2,...,b_N)$, then the error in the detection probability of the optimal search plan is given by

$$E(a,b)=e^{-K/N}\sum_{i=1}^{N}\left[\left(\prod_{i=1}^{N}b_i\right)^{1/N}a_i-\left(\prod_{i=1}^{N}a_i\right)^{1/N}b_i\right]/b_i.$$

The following two examples give some numerical results.

Example 1. In a two-cell search problem, suppose that the true target distribution is $a=(2/3,1/3)$, and the detection function is $b(t)=1-e^{-t}$, and the total search time $K=4$ minutes. Then the optimal detection probability is $P[f^*]=0.873$.

Suppose that the target distribution is unknown to the searcher, and the searcher chooses $(b,1-b)$, $(0<b<1)$. Then the detection probability

$$P[f_b^*]=1-(0.667-0.333b)e^{-2}[b(1-b)]^{-1/2}.$$

From this expression we can draw the graph of detection probability which illustrates the result of Theorem 3.

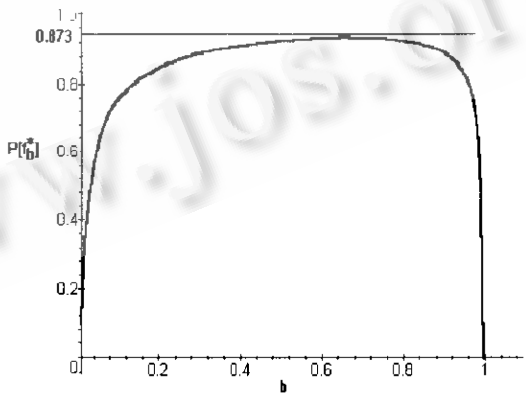

Fig. 1 Graph of the detection probability

Example 2. In the case of two cells and the exponential detection function, assume that the target distribution is again a random variable with the distributions:

$$P[p_1=a_i]=P[p_2=1-a_i]=q_i,\ (i=1,2,...,m).$$

Let $z_k(k=1,2)$ denote the time spent to search cell k in the optimal search plan, and $P[f^*]$ denote the detection probability of the optimal search plan. Suppose the cost is bounded by $K=4$. We can tabulate the value of

$P[f^*]$ as follows:

q	p_1/p_2	z_1/z_2	$P[f^*]$
0.1	0.5/0.5	2.0/2.0	0.86
0.15	0.6/0.4	2.2/1.8	0.87
0.20	0.7/0.3	2.42/1.58	0.88
0.35	0.8/0.2	2.69/1.31	0.89
0.15	0.9/0.1	3.1/0.9	0.92
0.05	1.0/0.0	4.0/0.0	0.99

From the above table we see that $E[p_1]=0.745$ but $E[P(f^*)]=0.892$. The detection probability corresponding to $E[p_1]$ is only 0.8779, which is less than $E[P(f^*)]$. In fact, $P[f^*]$ is independent of q_1, and thus the mean value of $E[p_1]$ is not related to $E[P(f^*)]$ too.

References:

[1] Stone, Lowrence D. Theory of Optimal Search. Mathematics in Science and Engineering, New York: Academic Press, 1975.

[2] Benkoski, S.J., Monticino, M.G., Weisinger, J.R. A survey of the search theory literature. Naval Research Logistics, 1991,38:469~494.

[3] Castannon, D.A. Optimal search strategies in dynamic hypothesis testing. Journal of IEEE Transactions on System, Man and Cybernetics, 1995,25(7):1130~1138.

[4] Koopman, B.O. Searching and screening. OEG Report, no. 56, Center for Naval Analysis, Rosslyn, Va., USA, 1946.

[5] Zhu, Qing-xin. Optimal search problems in discrete and continuous spaces. SCS Dissertations, Carleton University Press, 1996.

[6] Zhu, Qing-xin, Oommen, J. On the optimal search problem: the case when the target distribution is unknown. In: Proceedings of the 17th International Conference of the Chilean Computer Science Society, Valparaiso Chile, 1997. 268~277.

[7] Zhu, Qing-xin, Oommen, J. Optimal search with non-regular detection functions. Journal of UESTC, 1999,28(9):125~128 (in Chinese).

附中文参考文献:

[7] 朱清新,Oommen,J. 离散空间中获取函数为非正则情况下的最优探索. 电子科技大学学报,1999,28(9):125~128.

关于离散空间中最优搜索策略的一些结果

朱清新[1], 扈明天[2], John Oommen[3]

[1](电子科技大学 光纤通信国家重点实验室,四川 成都 610054);

[2](电子科技大学 计算机科学与工程学院,四川 成都 610054);

[3](加拿大卡尔顿大学 计算机学院,渥太华 K1S 5B61,加拿大)

摘要:研究关于 N 个位置的最优搜索问题. 最优搜索问题是研究如何将用于搜索的资源(如时间等)分配到 N 个位置使得发现目标的概率为最大. 以往人们在研究最优搜索问题时总是假设目标的分布函数是已知的,但实际情况往往不是这样. 用拉格朗日算子理论来研究目标的分布函数是未知的情况下的最优搜索问题,得出了一系列新的结果,包括分布函数的近似方法和误差估计公式. 最后给出了两个例子.

关键词:最优搜索;判决函数;目标分布;拉格朗日算子;误差估计

中图法分类号:O229 **文献标识码**:A