MANAGEMENT SCIENCE
Vol. 19, No. 6, February, 1973
Printed in U.S.A.

THE JOURNAL SELECTION PROBLEM IN A UNIVERSITY LIBRARY SYSTEM*†

D. H. KRAFT‡ AND T. W. HILL, JR.§

The problem of selecting which journals to acquire in order to best satisfy library objectives is examined and modeled as a zero-one linear programming problem. This is done using an objective function based on expected usage as a measure of journal worth and on cost constraints which account for the scarcity of capital. A dynamic programming algorithm is used to break down the larger problem into smaller subproblems and to generate a feasible solution. Special cases where the solution is optimal are presented and discussed in terms of their implications for the library. An example problem is presented to illustrate the algorithm.

Introduction

In this paper we discuss the problem of deciding which technical journals should be acquired, given a particular demand pattern exhibited by the users of a library and a set of limited resources available to the library. The resulting model takes the form of a zero-one linear programming problem.

The importance of the problem, from the "information deluge" point of view, has been demonstrated by Leimkuhler and Neville [14] and Meier [16]. At the same time, Hacker [5] has pointed out that financing to handle such growth is limited, with outside demands for these monies increasing, and Metcalf has said that growth is the underlying cause for nearly all of the financial problems of a library [17]. While Downs [3] states that complete records of all human activities would be ideal even if impractical, Morse [19] disagrees and feels that only the most important and most useful items should remain in a library collection. Since secondary storage facilities and microforms are starting to come into use, space is not as critical as it has been in the past. However, costs are still involved.

Thus, a choice must be made and it is crucial that the choice be made properly. The importance of the proper choice of journals has been established by Wulfekoetter [27] and by Straus, Strieby and Brown [24]. Yet, nowhere in the literature [23], [24], [27] has much attention been paid to the problem of how to decide which journals to acquire.

The Objective Function

Let the decision variables, $Y_{j,q-l,l}$, take on the value of one if the library has acquired the issues of journal j published in period l, as of period q, and zero if the library has not acquired these issues. Both q and l denote discrete periods of time so that they must cover a specified planning horizon consisting of r such periods plus a zero or initial point. Thus, $q = 0, 1, 2, \cdots, r - 1, r$ and $l = 0, 1, 2, \cdots, q - 1, q$. Assume that each period is of equal length and that all variables relate to an action as of the end of a specific period. There is a set of s journals ($j = 1, 2, \cdots, s - 1, s$) which includes all journals that have even the remotest chance of being selected.

* Received March 1971; revised July 1972.

† This research effort was conducted under the auspices of the National Science Foundation Office of Science Information Services, Grant GN-759

‡ University of Maryland.

§ Purdue University.

Next define $G_{j,q-l,l}$ as the worth of acquiring the issues of journal j published in period l as of period q; and $H_{j,q-l,l}$ as the worth of not having acquired these issues. This leads to the objective function

$$(1) \qquad Z' = \sum_{j=1}^{s} \sum_{q=0}^{r} \sum_{l=0}^{q} [Y_{j,q-l,l}G_{j,q-l,l} + (1 - Y_{j,q-l,l})H_{j,q-l,l}]$$

which is to be maximized to yield as much worth as possible.

In order to simplify the objective function, consider an inverse, linear relationship between $G_{j,q-l,l}$ and $H_{j,q-l,l}$, i.e.,

$$(2) \qquad H_{j,q-l,l} = -\rho G_{j,q-l,l} \qquad \forall j,\ q, \quad \text{and} \quad l \quad [8].$$

Rearranging terms, our objective function now becomes

$$(3) \quad Z' = (1 + \rho) \sum_{j=1}^{s} \sum_{q=0}^{r} \sum_{l=0}^{q} Y_{j,q-l,l}G_{j,q-l,l} - \rho \sum_{j=1}^{s} \sum_{q=0}^{r} \sum_{l=0}^{q} G_{j,q-l,l}.$$

If $\rho > -1$, we can achieve the same result by maximizing

$$(4) \qquad Z = \sum_{j=1}^{s} \sum_{q=0}^{r} \sum_{l=0}^{q} Y_{j,q-l,l}G_{j,q-l,l}.$$

This implies that (4) is a suitable objective function for the journal selection model as long as the worth of acquiring an item is greater than the worth of not acquiring it.

The library has been viewed as an information retrieval system [2], [7] and as a service organization [1]. It can be shown that this objective function incorporates the relevant factors involved with the performance evaluation for such systems [8]. As a journal becomes more pertinent to some set of users, its worth should go up and so the worth if acquired also increases. Moreover, the worth if not acquired should decrease and may even become a deficit since a pertinent item has been overlooked. In fact, when related to the constraints, the worth of acquiring a journal that is only slightly pertinent is quite small, since scarce resources that might be put to better use have been consumed.

This function allows us another advantage over past studies of library systems. Lister [15], when considering the use of compact storage for the less active items in a library collection, noted that the optimal storage policy was most sensitive to variations in the cost of the users not having an item immediately available. We can avoid the analogous problem of determining a specific cost of not being able to provide service directly, in terms of an acquired journal, by merely specifying a boundary condition for ρ. Thus, we do not need to concern ourselves with finding a specific value for ρ, as long as $\rho > -1$.

The Constraints

By definition

$$(5) \qquad Y_{j,q-l,l} = 0 \quad \text{or} \quad 1 \qquad \forall j,\ q, \quad \text{and} \quad l.$$

Since we are considering acquisition of a journal issue as of period q, and once an item is acquired it must remain in the library system and will not be considered for acquisition again, we have

$$(6) \qquad Y_{j,q-l,l} \geqq Y_{j,q-l-1,l} \qquad \forall j,\ q, \quad \text{and} \quad l.$$

All the issues of a journal for each period are considered as an inseparable unit analogous to a monograph or book, to be considered independently of *all* other items, or units.

Finally, using the cost model discussed by Leimkuhler [12] and by Williams et al. [26], we have

$$\begin{aligned} \sum_{j=1}^{s} \sum_{l=0}^{q} Y_{j,q-l,l}(C_1 + C_2 + C_3\lambda_{j,q-l,l} + K_{j,q-l,l}) \\ - \sum_{j=1}^{s} \sum_{l=0}^{q-1} Y_{j,q-l-1,l}(C_1 + K_{j,q-l,l}) \leqq b_q \qquad \forall q, \end{aligned} \tag{7}$$

where

b_q = the budget for period q for use in the acquisition of new journals and maintenance of the collection of journals,

C_1 = the cost of initially adding the issues of a journal to the collection, excluding subscription cost,

C_2 = the periodic storage and recurring costs, including some allotment for expected replacement of lost, stolen, or defaced items,

C_3 = the cost per expected use of an item for circulation,

$K_{j,q-l,l}$ = the subscription cost to acquire the issues of journal j published in period l if acquired in period q,

$\lambda_{j,q-l,l}$ = the expected usage in period q of the issues of journal j published in period l.

Williams et al. estimate C_1, C_2, and C_3, based on studies of university library costs, to be \$19.80, \$0.194, and \$1.48, respectively. Subscription costs range generally from \$5.00 to \$30.00 annually. The cost estimates for selection decisions should be based on the additional or incremental cost incurred due to the acquisition of an additional item [13]. While Williams et al. largely ignore the problem of estimation of such incremental costs, their estimates can show the relative size of the cost figures.

The total model is shown in Figure 1, using one potential measure of journal worth, expected usage, as discussed below.

$$\text{Max } E(Z) = \sum_{j=1}^{s} \sum_{q=0}^{r} \sum_{l=0}^{q} Y_{j,q-l,l}\, \lambda_{j,q-l,l}\,,$$

$$\text{s.t. } \sum_{j=1}^{s} \sum_{l=0}^{q} Y_{j,q-l,l}\ (C_1 + C_2 + C_3\lambda_{j,q-l,l} + K_{j,q-l,l}) - \sum_{j=1}^{s} \sum_{l=0}^{q-1} Y_{j,q-l-1}\ (C_1 + K_{j,q-l,l}) \leqq b_q \qquad \forall q$$

$$Y_{j,q-l,l} \geqq Y_{j,q-l-1,l} \qquad \forall j,\ q, \text{ and } l$$

$$Y_{j,q-l,l} = 0 \text{ or } 1 \qquad \forall j,\ q, \text{ and } l$$

FIGURE 1. The total model

Journal Usage as a Criterion for Measuring Journal Worth

In considering what criterion to use when measuring the worth of acquiring a specific journal, one factor that consistently comes to mind is the usage of, or demand for, that journal. Letting $G_{j,q-l,l}$ represent the number of demands for the issues of journal j published in time period l made in period q has several interesting implications. First, this means that the objective function now represents the number of demands that the library has the potential to directly satisfy. The variable $H_{j,q-l,l}$, as expressed in equation (2), takes the role of a corrective factor; a portion of the unsatisfied demand is deducted from the potentially satisfied demand in the objective function to compensate for bias. Since usage is nonnegative for all items, in terms of maximizing expected usage it is always better to acquire an item than to not acquire it. Thus $\rho > -1$, and the objective function of (4) is valid.

Second, it is to be expected that relevance or pertinence in terms of a contingency table analysis [25] and usage are directly related. As demand decreases, one would expect relevance to decrease so that a decrease in $G_{j,q-l,l}$ becomes representative of the smaller worth of acquiring a nonpertinent journal. On the other hand, as usage decreases, $H_{j,q-l,l}$ should increase and become representative of the increased worth of not acquiring a nonpertinent journal.

In order to estimate journal usage over time, consider a Markovian approach. Adopting a modified form of Morse and Elston's [20] model of probablilistic obsolescence [9], we have

$$(8)\quad \lambda_{j,q-l,l} = E(G_{j,q-l,l}) = (\lambda_{j,0,l} - ac/(c-b))b^{q-l} + (ac/(c-b))c^{q-l} \qquad \forall j, q, \text{and } l,$$

where a, b, and c are nonnegative parameters, b and c are no greater than unity, b and c are unequal, and $\lambda_{j,0,l}$ is the initial expected estimate of usage in period l of the issues of journal j published in period l. By specifying a, b, c, and $\lambda_{j,0,l}$ for all j and l, we can use (8) to estimate usage over time. Morse and Elston found that for various fields of scientific endeavor b ranges from 0.20 to 0.60; while a ranges from 0.32 to 0.68 and c ranges from 0.91 to 0.975 [9]. Thus from (8) the objective function must now be considered as an expected value, so

$$(9)\qquad E(Z) = \sum_{j=1}^{s} \sum_{q=0}^{r} \sum_{l=0}^{q} Y_{j,q-l,l}\lambda_{j,q-l,l}\,.$$

A Dynamic Programming Solution Algorithm

Conventional integer programming methods [22] will not handle problems of the size generated by this model (for s journals and r time periods we have $s(r+3)r/2$ decision variables, so for $s = 50$ and $r = 10$ we have 3,250 variables). Consider a dynamic programming approach to break down the larger problem into smaller ones of the knapsack variety.

The state variable and equations are defined by

$$X_q = b_q - \sum_{j=1}^{s} \sum_{l=0}^{q-1} Y_{j,q-l,l}(C_2 + C_3\lambda_{j,q-l,l}) \qquad \forall q$$

(10) = the amount of money left over in period q for acquiring new journals after paying for the storage and circulation of former acquisitions in period q;

and the recursive relationship of the objective function over time can be stated as

$$F_q^*(X_q) = \mathrm{Max}_{\{Y_{j,q-l,l}\}} \left[\sum_{j=1}^{s} \sum_{l=0}^{q} Y_{j,q-l,l}\lambda_{j,q-l,l} + F_{q+1}^*(X_{q+1})\right]$$

(10a) = the optimal return at stage or period q as a function of the state variable X_q and the decision variables $Y_{j,q-l,l}$.

The constraints are (5), (6), and (7), with (7) being rewritten as

$$(11)\quad \sum_{j=1}^{s} \left[\sum_{l=0}^{q} Y_{j,q-l,l}(C_1 + C_2 + C_3\lambda_{j,q-l,l} + K_{j,q-l,l}) - \sum_{l=0}^{q-1} Y_{j,q-l-1,l}(C_1 + C_2 + C_3\lambda_{j,q-l,l} + K_{j,q-l,l})\right] \leqq X_q \qquad \forall q$$

and

$$(12)\quad X_{q+t} = b_{q+t} - \sum_{j=1}^{s} \sum_{l=0}^{q+t-1} Y_{j,q+t-l,l}(C_2 + C_3\lambda_{j,q+t-l,l}) \geqq 0 \qquad \forall q, \text{ and } t = 1, 2, \cdots, r-q.$$

Note that the decisions made in period q affect returns and decisions made in period $q+1$ and all future periods, so that a recursive relationship is needed to take these effects into account. Moreover, (12) insures the feasibility of future budget constraints by not allowing us to acquire more items in period q than can be stored and circulated

(1) The journal selection problem must be set up as outlined in Figure 1. This means specifying the costs (C_1, C_2, C_3, $K_{j,q-l,l}$), expected journal usage ($\lambda_{j,q-l,l}$), budgets ($b_q \geqq 0$), and the setting of $Y_{j,0,0}$ to one for all journals purchased prior to the planning horizon with the setting of all other $Y_{j,q-l,l}$ values to zero. This also means one must define a planning horizon and divide it into r equal time periods; and one must define a complete set of s journals and label them with a number j, $j = 1, 2, \cdots, s$. One must also estimate any parameters used to depict expected usage or any other variables over time. One must check to see if a feasible solution exists. If

$$b_q \geqq \sum_{j=1}^{s} Y_{j,0,0}(C_2 + C_3\lambda_{j,q,0}) \quad \text{for all} \quad q = 1, \cdots, r,$$

a feasible solution does exist and we can proceed to Step (2). If at least one q exists such that the condition is violated, no feasible solution exists and we must terminate our algorithm

(2) Set $q = 1$. Let $Y_{j,1,0} = 1$ if $Y_{j,0,0} = 1$ for all $j = 1, 2, \cdots, s$.

(3) Calculate the cost functions

$$\begin{aligned} C_{j,q-l,l} &= C_1 + C_2 + C_3\lambda_{j,0,q} + K_{j,0,q} \quad \text{if} \quad q = l, \\ &= (1 - Y_{j,q-l-1,l})(C_1 + K_{j,q-l,l}) + C_2 + C_3\lambda_{j,q-l,l} \quad \text{if} \quad q > l, \end{aligned} \tag{14}$$

for all $j = 1, 2, \cdots, s$ and for all $l = 0, 1, \cdots, q$. Also define

$$X_q = b_q - \sum_{j=1}^{s} \sum_{l=0}^{q-1} Y_{j,q-l-1,l}(C_2 + C_3\lambda_{j,q-l,l}).$$

(4) Calculate the ratio $R_{j,q-l,l}$ for all $j = 1, 2, \cdots, s$ and $l = 0, 1, \cdots, q$ where

$$\begin{aligned} R_{j,q-l,l} &= \lambda_{j,0,q}/C_{j,0,q} \quad \text{if} \quad l = q, \\ &= \lambda_{j,q-l,l}(1 - Y_{j,q-l-1,l})/C_{j,q-l,l} \quad \text{if} \quad l < q. \end{aligned}$$

(5) Rank the ratios $R_{j,q-l,l}$ in decreasing order Define the variable $\theta_{j,q-l,l}$ as the rank of the item having the ratio $R_{j,q-l,l}$. Thus the item with the largest ratio will have a θ value of 1

(6) Define $C(\theta_{j,q-l,l}) = C_{j,q-l,l}$ $\forall$ j, q, and l and p_q as the largest integer such that

$$\sum_{\theta_{j,q-l,l}=1}^{p_q} C(\theta_{j,q-l,l}) \leqq X_q . \tag{15}$$

Thus, we sum over the ranks of the items, in order to select the highest ranked items and to acquire as many as possible under the constraint limitations. Going down the list of ranked items, set the first p_q $Y_{j,q-l,l}$ variables equal to one.

(7) If (15) is an equality, we have optimized the return at stage q. Go to Step (9)

(8) Here (15) is a strict inequality, so we must use branch and bound techniques, such as those outlined in [4] to achieve some desired level of near-optimality if not optimality at stage q.

(9) Calculate

$$X_{q+t} = b_{q+t} - \sum_{j=1}^{s} \sum_{l=0}^{q+t-1} Y_{j,q+t-l-1,l}(C_2 + C_3\lambda_{j,q+t-l,l})$$

for $t = 1, 2, \cdots, r - q$.

(10) If $X_{q+t} \geqq 0$ for all t, we are feasible for all future time periods in terms of the policy for period q. Go to Step (12).

(11) If $X_{q+t} < 0$ for at least one $t = 1, 2, \cdots, r - q$ we have violated the budget constraint for period $q + t$ through the storage and circulation of items accumulated before period $q + 1$. To restore feasibility, begin to set $Y_{j,q-l,l} = 0$ for enough variables to satisfy the violated constraint. This must be done by beginning at the lowest value of $R_{j,q-l,l}$ for which $Y_{j,q-l,l} = 1$, and $Y_{j,q-l-1,l} = 0$ if $q > l$, and going up the scale as the ratio increases and setting each $Y_{j,q-l,l} = 0$ as needed.

(12) Set q equal to the value $q + 1$.

(13) If $q \leqq r$, we must re-cycle to continue the process for new time periods. We must first update our decision variables by letting $Y_{j,q-l,l} = 1$ when $Y_{j,q-l-1,l} = 1$ for all $j = 1, \cdots, s$ and $l = 0, 1, \cdots, q - 1$. Now go to Step (3).

(14) Here $q > r$, so we have completed the solution and must terminate. Calculate

$$\sum_{j=1}^{s} \sum_{q=0}^{r} \sum_{l=0}^{q} Y_{j,q-l,l}\lambda_{j,q-l,l}$$

for the expected usage from our selection policy.

FIGURE 2. The dynamic programming algorithm.

in all future periods. We assume that all of the budgets, as well as the cost and usage parameters, are nonnegative and are known at the time of planning.

A result from the theory of optimal control of stochastic systems shows that under certain conditions the optimal policy is independent of the state of the system [18]. Since our model satisfies these conditions, we can generate a series of independent knapsack problems provided we ignore the recursiveness of the problem.

An algorithm embodying these ideas is given in Figure 2 below. Specifically the algorithm determines a suboptimal feasible policy, if one exists, by solving the knapsack problem: determine

$$F_q^*(X_q) = \mathrm{Max}_{\{Y_{j,q-l,l}\}} \sum_{j=1}^{s} \sum_{l=0}^{q} Y_{j,q-l,l}\lambda_{j,q-l,l} \tag{13}$$

subject to (5), (6), (11), and (12), for each q.

The conditions for the existence of a feasible solution (Step (1)) are that the budgets for each period can cover the cost of storing and circulating the items bought previously to the planning horizon as described by the given initial selection variables $Y_{j,0,0}$. The satisfaction of this condition implies that at the very least the policy of acquiring no new journals is feasible. If this is the case, we seek better solutions; if this is not the case, no solution is feasible, terminate the solution algorithm.

The algorithm is iterative in nature; one starts with $q = 1$ and proceeds by increasing q by 1 each time until all r problems have been solved (Steps (2), (12), (13) and (14)). Further, by starting with $q = 1$ and working up to $q = r$, we can maintain the relationship $Y_{j,q-l,l} \geqq Y_{j,q-l-1,l}$ (Steps (2) and (13)), and can properly calculate the increased cost due to acquisition in period q (Step (3)) as well as the increased usage (Step (4)). In the calculation of the ratios (Step (4)), the original benefit-over-cost function is multiplied by $(1 - Y_{j,q-l-1,l})$ to set that ratio to zero and virtually eliminate the ratio from consideration, if $Y_{j,q-l-1,l} = 1$ which means that acquisition of j will not occur in period q but has already taken place. The calculation of these ratios, the ranking of these ratios in descending order (Step (5)), the selection of those items with the highest ratios to be acquired (Step (6)), the branch-and-bounding to check for better solutions if the budget figure is not met exactly (Step (8)), and the termination of the qth problem if the budget is met exactly (Step (7)) are all part of the knapsack problem and can be solved accordingly [4]. While the knapsack problem solution technique of Step (8) could be terminated at a suboptimal point, generating another method of deviation from optimality, this need not be the case. Moreover, the main cause of suboptimality is not branch-and-bounding methods of Step (8) but the removal of the linkages between time periods needed to form the set of independent knapsack problems. To keep the solution feasible for all future time periods, the routine (Steps (9)–(11)) is used to insure that X_{t+q} is nonnegative for all $t = 1, 2, \cdots, r - q$.

When Is Suboptimal Optimal?

The algorithm produces an obviously suboptimal solution in that each stage or period is optimized separately and independently with no regard to future effects of a current policy decision. This is exemplified by (12) which shows that items currently selected have circulation and storage costs associated with them in all future periods which can directly influence the future decisions made in those periods. In essence, the recursiveness of the dynamic programming formulation has been dropped to yield a feasible, suboptimal solution. Moreover, the worth of this solution in terms of a global optimal answer is in general unknown.

However there are some special cases where optimizing at each stage can lead to the global optimum solution. These cases and their implications are discussed below.

PROPOSITION. *The dynamic programming algorithm of Figure 2, optimizing each stage separately, leads to a global optimal solution when a feasible solution exists, if any of the following cases holds:*

(i) $C_2 = C_3 = 0$.

(ii) $C_3 = 0$; $K_{j,q-l,l} = K_q \; \forall \; j, q$ *and* l; *and* $C_1 + X_2 + K_q > 0$ *and*

$$\langle X_q/(C_1 + C_2 + K_q)\rangle = \langle b_q/(C_1 + C_2 + K_q)\rangle \; \forall \; q.$$

(iii) $\lambda_{j,q-l,l} = \lambda_q$, $K_{j,q-l,l} = K_q \; \forall \; j, q,$ *and* l; *and* $C_1 + C_2 + C_3\lambda_q + K_q > 0$ *and*

$$\langle X_q/(C_1 + C_2 + C_3\lambda_q + K_q)\rangle = \langle b_q/(C_1 + C_2 + C_3\lambda_q + K_q)\rangle \; \forall \; q.$$

Here $\langle a \rangle =$ *the largest integer* $\leqq a$.

PROOF.

Case (i). In this case, we can see that the state equation (10) becomes $X_q = b_q \; \forall \; q$, and that the budget constraints (11) become

$$\sum_{j=1}^{s} \sum_{l=0}^{q} Y_{j,q-l,l}(C_1 + K_{j,q-l,l}) - \sum_{j=1}^{s} \sum_{l=0}^{q-1} Y_{j,q-l-1,l}(C_1 + K_{j,q-l,l}) \leqq X_q \qquad \forall \; q.$$

All other relationships in the model remain the same. With the link between each stage or time period having been broken in terms of the future budgets and policies being unaffected by the storage and circulation costs of current selections, the computational algorithm of Figure 2 yields a global solution. Thus, the journal selection model decomposes into a series of knapsack problems that can be solved independently for the optimal solution to the total selection model.

Case (ii). This case implies that (10) now becomes

$$X_q = b_q - C_2 \sum_{j=1}^{s} \sum_{l=0}^{q-1} Y_{j,q-l-1,l} \qquad \forall \; q$$

and that (11) now is

$$(C_1 + C_2 + K_q)[\sum_{j=1}^{s} Y_{j,0,q} + \sum_{j=1}^{s} \sum_{l=0}^{q-1} (Y_{j,q-l,l} - Y_{j,q-l-1,l})] \leqq X_q \qquad \forall \; q.$$

Again, all other relationships in the model remain unchanged. While we have not yet eliminated the dynamic relationship between stages, the process which established the computational algorithm as one yielding the optimal solution such as in Case (i), we can form conditions which will accomplish this end.

Since

$$\sum_{j=1}^{s} Y_{j,0,q} + \sum_{j=1}^{s} \sum_{l=0}^{q-1} (Y_{j,q-l,l} - Y_{j,q-l-1,l})$$

is a series of additions and subtractions of zero-one variables, it must be an integer. Thus, with $C_1 + C_2 + K_q > 0$, we can rewrite the cost constraints as

$$\sum_{j=1}^{s} Y_{j,0,q} + \sum_{j=1}^{s} \sum_{l=0}^{q-1} (Y_{j,q-l,l} - Y_{j,q-l-1}) \leqq \left\langle \frac{X_q}{C_1 + C_2 + K_q} \right\rangle = \left\langle \frac{b_q}{C_1 + C_2 + K_q} - \frac{C_2 \sum_{j=1}^{s} \sum_{l=0}^{q-1} Y_{j,q-l-1,l}}{C_1 + C_2 + K_q} \right\rangle.$$

Since

$$\langle X_q/(C_1 + C_2 + K_q)\rangle = \langle b_q/(C_1 + C_2 + K_q)\rangle \qquad \forall\, q,$$

we can eliminate the dynamic relationship between cost constraints. To obtain this, note that $\langle a - b\rangle = \langle a\rangle$ implies that $\langle a\rangle \leqq a - b$. Thus, using (10), we have

$$\langle b_q/(C_1 + C_2 + K_q)\rangle \leqq b_q/(C_1 + C_2 + K_q) - C_2\sum_{j=1}^{s} \sum_{l=0}^{q-1} Y_{j,q-l-1,l}/(C_1 + C_2 + K_q)$$

or

$$\sum_{j=1}^{s} \sum_{l=0}^{q-1} Y_{j,q-l-1,l} \leqq b_q/C_2 - (C_1 + C_2 + K_q)\langle b_q/(C_1 + C_2 + K_q)\rangle/C_2 \qquad \forall\, q.$$

This occurs by maintaining a large enough budget to account and compensate for all past acquisitions in terms of current storage and circulation costs. Of course, we still have the feasibility requirements that

$$X_{q+t} = b_{q+t} - C_2\sum_{j=1}^{s} \sum_{l=0}^{q+t-1} Y_{j,q+t-l-1,l} \geqq 0 \quad \text{for all} \quad t = 1, \cdots, r - q$$

which implies that

$$\sum_{j=1}^{s} \sum_{l=0}^{q+t-1} Y_{j,q+t-l-1,l} \leqq \langle b_{q+t}/C_2\rangle \quad \text{for all} \quad t = 1, \cdots, r - q.$$

Thus, the computational algorithm yields an optimal solution to the journal selection model when all the conditions have been met.

Case (iii). In a situation parallel to that of Case (ii), we have

$$X_q = b_q - (C_2 + C_3\lambda_q)\sum_{j=1}^{s} \sum_{l=0}^{q-1} Y_{j,q-l-1,l}$$

and

$$(C_1 + C_2 + C_3\lambda_q + K_q)\left[\sum_{j=1}^{s} Y_{j,0,q} + \sum_{j=1}^{s} \sum_{l=0}^{q-1} (Y_{j,q-l,l} - Y_{j,q-l-1,l})\right] \leqq X_q \qquad \forall\, q,$$

with all other relationships in the model remaining constant. The objective function is now to maximize

$$\sum_{q=0}^{r} \lambda_q \sum_{j=1}^{s} \sum_{l=0}^{q} Y_{j,q-l,l}\,.$$

As before, since

$$\sum_{j=1}^{s} Y_{j,0,q} + \sum_{j=1}^{s} \sum_{l=0}^{q-1} (Y_{j,q-l,l} - Y_{j,q-l-1,l})$$

is an integer and $C_1 + C_2 + C_3\lambda_q + K_q$ is positive, we have

$$\sum_{j=1}^{s} Y_{j,0,q} + \sum_{j=1}^{s} \sum_{l=0}^{q-1} (Y_{j,q-l,l} - Y_{j,q-l-1,l}) \leqq \left\langle \frac{x_q}{(C_1 + C_2 + C_3\lambda_q + K_q)} \right\rangle = \left\langle \frac{[b_q - (C_2 + C_3\lambda_q)\sum_{j=1}^{s} \sum_{l=0}^{q-1} Y_{j,q-l-1,l}]}{(C_1 + C_2 + C_3\lambda_q + K_q)} \right\rangle.$$

As before, the conditions for global optimality are

$$\langle x_q/(C_1 + C_2 + C_3\lambda_q + K_q)\rangle = \langle b_q/(C_1 + C_2 + C_3\lambda_q + K_q)\rangle$$

or

$$\sum_{j=1}^{s} \sum_{l=0}^{q-1} Y_{j,q-l-1,l} \leqq \frac{b_q}{(C_2 + C_3\lambda_q)} - \frac{(C_1 + C_2 + C_3\lambda_q + K_q)}{(C_1 + C_2\lambda_q)} \left\langle \frac{b_q}{(C_1 + C_2 + C_3\lambda_q + K_q)} \right\rangle.$$

Feasibility further requires that

$$X_{q+t} = b_{q+t} - (C_2 + C_3\lambda_{q+t}) \sum_{j=1}^{s} \sum_{l=0}^{q+t-1} Y_{j,q+t-l-1,l} \geqq 0$$

or

$$\sum_{j=1}^{s} \sum_{l=0}^{q+t-1} Y_{j,q+t-l-1,l} \leqq \langle b_{q+t}/(C_2 + C_3\lambda_{q+t}) \rangle \quad \text{for all} \quad t = 1, \cdots, r - q,$$

and the proof is complete.

As discussed earlier, the cost coefficients must represent the actual amount of additional expenditure and effort required by the selection of a new journal. With journals being selected in an environment where circulation and storage of library materials are considered and budgeted separately relative to the acquisitioning functions, it is logical to let $C_2 = C_3 = 0$. The relevant costs for selection decisions are subscription and ordering costs only. In fact, in the Purdue University Libraries [8], one could also let $C_1 = 0$ since only the actual subscription costs are used in making journal selection decisions, with ordering costs handled separately.

Case (ii) also has some very interesting implications for the librarian. First, the additional restrictions on the number of items bought previous to period q and on the total cumulative number of acquisitions made as of all future periods in terms of the budgets for period q and future budgets, respectively, lead to an optimal solution via the computational algorithm.

It seems quite reasonable to find situations where one would desire restrictions on the number of items selected. This idea seems reinforced when assuming that certain costs are zero in terms of selection decisions; for certain functions such as ordering, storing, or circulating journals the library may have enough resources to handle only a fixed amount of material. If one does not exceed the limits on selections, no additional costs are incurred by new selections in terms of these functions.

Also, if one can increase the budget enough in each period so that these additional requirements are satisfied in all periods, we can guarantee an optimal solution. Since $\sum_{j=1}^{s} \sum_{l=0}^{q-1} Y_{j,q-l-1,l}$ and $\sum_{j=1}^{s} \sum_{l=0}^{q+t-1} Y_{j,q+t-l-1,l}$ are nondecreasing functions of q, b_q must also be a nondecreasing function of q to satisfy all conditions. One difficulty with this case is the assumption that $K_{j,q-l,l} = K_q$, which implies that in period q all items have a constant subscription cost regardless of age or of which journal or specific volume is being selected. Another problem is that there is no a priori way to determine a sufficient increase in the values of the budgets over time.

The same comments made in Case (ii) regarding the restrictions on the number of items selected can be made for Case (iii). The solution obtained from the computational algorithm is optimal if the conditions are met. The implications of budgets needing to be increased enough again indicate the relevance of this result to people who are interested in understanding the consequences of setting budgets. One may now be able to define how much money is adequate to maintain a certain collection and to be able to update a library by acquiring recent volumes of ongoing journals so as to retain continuity.

It is not unreasonable to consider a new library as being in a situation indicated by these latter cases. With circulation, acquisition, and storage departments not already established, each new journal may mean incremental costs. Further, one may be able to purchase several items at a relatively constant subscription price from dealers or from gifts and exchange departments of other libraries. Finally, with no past usage data, one may assume a relatively constant usage for several items purchased in a specific period. However, it is the assumptions of constant subscription price and usage for all items selected in a specific time period that could be the great weakness of this case in general situations. This, of course, depends on the length of the planning horizon and the prevailing economic conditions. If the horizon length is short, this assumption may not be "too far" wrong.

The relevance of these cases stems from the fact that they shed some light on the relationships that exist between collection maintenance and budget limitations. Hopefully they will lead to some more general insights into acquisition policies that will be of assistance to the library decision maker, especially in terms of the number of items selected.

It is also important to note that the magnitude of the costs C_2 and C_3 compared to the other costs can give some idea of the error made in disconnecting the periods as was done here. If we take the cost estimates of Williams et al. as previously given, we see that C_2 and C_3 are an order of magnitude smaller than the other costs.

An Example Problem

In order to illustrate the algorithm, an example is presented below [8]. The size of the problem will be kept small to maintain an understanding of the computations.

Suppose we consider the purchasing of a few journals to add breadth and depth to an ongoing collection. This means that past usage data are easily accessible, costs and facilities are fixed and relatively stable, and the decision in question is the selection of some subset of a small number of available journals over a small time horizon. To be specific, let us consider the purchase of $s = 4$ journals related to some field of scientific endeavor, such as operations research, over an $r = 5$ period planning horizon. Assume that we are interested in a yearly subscription cycle for each year so that each time period represents a one-year period. The initial input data are presented in Table 1.

The parameters a, b, and c are used to calculate expected usage, $\lambda_{j,q-l,l}$, according to the revised Markovian model described by (8). This model uses the initial usage estimates, $\lambda_{j,0,q}$, which were assigned to the journals randomly but as nondecreasing functions of publishing date. The cost figures for C_1, C_2, and C_3 are taken from estimates in the study by Williams et al. [26]. The budgets, increasing over time, were selected in order to allow for the existence of at least one feasible solution and to allow for a solution that, a priori, seemed reasonable in terms of continuity of ongoing journals. The initial subscription costs, assigned randomly to the journals, were set up as increasing functions of an item's age. In this example, it was assumed that subscription costs were not a function of an item's date of publication. The final item in the initial data set is the specification of journals already being purchased as of the initial period, i.e., initial acquisitions.

It is not to be implied that this example problem is at all representative of any large general class of problems, but it will be seen that this example does illustrate the mechanics of the algorithm. It should be noted that for even a small problem such as this, much input data are required.

TABLE 1

Initial Values and Parameters

Usage Parameters					Costs	
a	b	c		c_1	c_2	c_3
0.60	0.50	0.93		19.80	0.194	1.48
Budgets						
b_q	$q = 1$	2		3	4	5
	115	125		130	140	150
Initial Usage						
$\lambda_{j,0,q}$	$q = 0$	1	2	3	4	5
$j = 1$	1.50	2 00	2.00	2.50	2.50	3.00
2	3.00	4.00	5.00	6.00	7.50	8.00
3	1.00	1 25	1.50	1.75	2.00	2.25
4	2 00	2.00	2.00	2.00	2.00	2.00
Subscription Costs						
$K_{j,t,0}$	$t = 0$	1	2	3	4	5
$j = 1$	10	11	12	13	14	15
2	15	15	15	20	20	25
3	5	5	10	10	10	10
4	10	11	12	13	14	15
Initial Acquisitions						
$Y_{j,0,0}$	$j = 1$	1	2	3	4	
		0	1	1	0	

The solution generated by the algorithm (solution 2) is presented in Table 2 along with some other relevant solutions for this specific problem. The table lists the solution with its objective function, the budgets for the problem (b_q), expenditures in each time period (g_q), and $t_{j,q}$, the period in which the issues of journal j published in period q are acquired. If $t_{j,q} = r + 1 = 6$, the item is never acquired in the planning period. Also, each solution is feasible or infeasible only in terms of the cost constraints. The constraints on continuity (6) are maintained through the solution procedures of the various algorithms, as are the constraints (5) regarding the values which the decision variable can take, zero or one.

The first answer, illustrated in Table 2 as solution 1, is the one formed by considering an "empty" policy of buying no new journal issues at all. The feasibility of this answer establishes the existence of at least one feasible solution to the example problem.

Solution 3 is a "full" policy, consisting of buying all items as soon as possible. This policy, having the largest value of the objective function of any set of zero-one decision variables for this example, can be used as a comparison for the other solutions. The "full" policy yielded a value of 153.7011 as the maximal upper limit to the objective function, but it is not a totally valid comparison since this policy is infeasible. The optimal solution must have a value of its objective function that lies between the value for the "empty" policy, 16.3095, and the "full" policy. However, these limits are obviously superseded by tighter bounds discussed below.

Only in the "full" policy did the algorithm allow older items to be selected over new material. Thus, no $Y_{j,t,q} = 1$ for $t > 0$ unless $Y_{j,t-1,q} = 1$ for all but the latter solution. Another point is that there is some loss of continuity in the best feasible solution, number 2. The initial conditions specify that journals 2 and 3 are being acquired. In the first two time periods, we select current issues of journals 1, 2, and 4; while in the last three time periods, we select current journals 1, 2, and 3. In other words, journal 3 is dropped for the first two time periods in favor of journal 4. Then, the situation reverses itself and stays so throughout the remaining periods. If continuity is desired,

TABLE 2

Solutions Generated by the Dynamic Programming Algorithm

Solution 1			Objective Function = 16.3093			
	$q = 0$	1	2	3	4	5
b_q	—	115 0	125.0	130.0	140.0	150.0
g_q	—	5.0352	4.3144	3.8739	3.5775	3.3570
$t_{j,q}$						
$j = 1$	6	6	6	6	6	6
2	0	6	6	6	6	6
3	0	6	6	6	6	6
4	6	6	6	6	6	6
Solution 2			Objective Function = 120.5742			
	$q = 0$	1	2	3	4	5
b_q	—	115.0	125.0	130.0	140 0	150.0
g_q	—	111.8372	121.6492	126.0104	135 7799	144.9031
$t_{j,q}$						
$j = 1$	6	1	2	3	4	5
2	0	1	2	3	4	5
3	0	6	6	3	4	5
4	6	1	2	6	6	6
Solution 3			Objective Function = 153.7011			
	$q = 0$	1	2	3	4	5
b_q	—	115.0	125 0	130 0	140.0	150.0
g_q	—	204.9664	154.9553	166.7731	178.5332	189.4297
$t_{j,q}$						
$j = 1$	1	1	2	3	4	5
2	0	1	2	3	4	5
3	0	1	2	3	4	5
4	1	1	2	3	4	5
Solution 4			Objective Function = 118.1835			
	$q = 0$	1	2	3	4	5
b_q	—	115.0	125.0	130.0	140.0	150.0
g_q	—	105.7472	115.3543	125.3628	135.4561	144.7412
$t_{j,q}$						
$j = 1$	6	1	2	3	4	5
2	0	1	2	3	4	5
3	0	1	2	3	4	5
4	6	6	6	6	6	6
Solution 5			Objective Function = 120 7617			
	$q = 0$	1	2	3	4	5
b_q	—	115 0	125.0	130.0	140.0	150.0
g_q	—	111.8572	121.6492	131.3803	140.9649	149.6187
$t_{j,q}$						
$j = 1$	6	1	2	3	4	5
2	0	1	2	3	4	5
3	0	6	6	6	6	6
4	6	1	2	3	4	5

a librarian might examine this solution and decide to purchase only journals 1, 2, and 3 throughout the planning horizon. This revised plan is illustrated in solution 4. Solution number 5 consists of a policy of acquiring journals 1, 2, and 4 throughout the horizon and provides a contrast to the previous answer. Solution 4 is feasible while solution 5 is not, but solution 2 is the best feasible answer we have generated.

Using a Lagrangian formulation of the problem we have developed an upper bound on the optimal value of the objective function, in this case the bound is 123.0564 [10]. The optimal solution lies within the interval bounded by the value of the objective function of the dynamic programming solution and this upper bound.

Summary and Conclusions

In this paper we have developed a model of the journal selection problem in the form of a zero-one linear programming problem. The objective function was shown to reflect many of the various means of evaluating the performance of information systems. The constraints were generated primarily from budget and cost considerations. We also developed an algorithm based upon a dynamic programming formulation, which generates a feasible solution if one exists.

The special cases where the algorithm generates optimal solutions have several interesting implications for the library system. If the unit costs associated with acquisition, storage, and circulation are zero, an optimal solution is generated. This may be true for several libraries which set up separate budgets for the journal selection function based solely on the subscription costs. The acquisition, storage, and circulation functions become ongoing operations that are independent of the number or type of journal items selected. If the subscription cost and/or the expected usage of each item is dependent solely on the period in which the decision is being made, we have the potential for an optimal solution to be generated from our dynamic programming algorithm. To allow it to be optimal, we must set limits on the solution in terms of the number of journal items acquired in each period. This limit reinforces the notion of no increase in the cost of the acquisition, storage, and circulation functions, due to the acquisition of new journal items up to a fixed limit on the amount of material in the collection. While these latter cases may not be realistic for all libraries, they do serve to illustrate the point and may approximate reality for some libraries.

The data requirements for the selection model are severe, though this is hardly unique to this model. The quantity of data needed is rather large, but with the exception of the demand parameter estimates, and the incremental costs, the data can be easily obtained from standard sources such as accounting figures. Usage seems to be the largest problem, since there is a good deal of data required for this item and since unbiased estimates may be very difficult to obtain. Since usage will determine journal worth and hence policy, a decision maker may bias estimates in favor of those items he decides a priori are good for one reason or another. Research must be done to determine methods of obtaining good, unbiased estimates that can be had with relative ease and accuracy, and they must be reasonably inexpensive to obtain.

One possible way to reduce the amount of data required is to establish simple parametric forms for $\lambda_{j,0,q}$ (expected usage in period q of the issues of journal j published in period q) and for $K_{j,q,t}$ (subscription cost of the issues of journal j published in period t if acquired in period $t + q$) versus the variable q. Of course, this will then add to the problem of finding good estimates for all of the parameters used in the model.

It is likely that the cost and difficulty of implementing this model would be too high for current practice. However, the corresponding real life problem it represents must be solved. Thus, the importance of the model does not lie in its immediate applicability, but rather in what it has told us about the complex decision problem it portrays and the directions it suggests for future studies.

References

1. BAKER, N. R., "An Analytical Framework for Service Organizations," XIV International Meeting of TIMS, Mexico City, August 1967.
2. BECKER, J. AND HAYES, R., *Information Storage and Retrieval: Tools, Elements, Theories*, Wiley, N. Y., 1963.
3. DOWNS, R. B., "The Implementation of Book Selection Policy in University and Research

Libraries", in Goldhor, H. (Ed.), *Selection and Acquisition Procedures in Medium-Sized and Large Libraries*, Illinois-Union Bookstore, Champaign, Illinois, 1963.
4. GREENBERG, H. AND HEGERICH, R. L., "A Branch Search Algorithm for the Knapsack Problem," *Management Science*, Vol. 16, No. 5 (January 1970), pp. 327–332.
5. HACKER, H. L., "Financial Problems of the Large Library," *The Library Quarterly*, Vol. 38, No. 1 (January 1968), pp. 41–57.
6 JAIN, A. K., "A Statistical Study of Book Usage," Ph.D. Dissertation, Purdue University, School of Industrial Engineering, Lafayette, Indiana, 1967.
7. KORFHAGE, R AND DELUTIS, T., "A Basis for Time and Cost Evaluations of Information Systems," Research Memorandum Series 69–6, School of Industrial Engineering, Purdue University, Lafayette, Indiana, June 1969
8. KRAFT, D. H., "The Journal Selection Problem in a University Library System," Unpublished Ph.D. Dissertation, School of Industrial Engineering, Purdue University, Lafayette, Indiana, January 1971.
9. ——, "A Comment on the Morse-Elston Model of Probabilistic Obsolescence," *Operations Research*, Vol. 18 (November-December 1970), pp 1228–1233.
10. —— AND HILL, T. W., JR., "A Langrangian Formulation of the Journal Selection Model," A paper presented at the 39th National Operations Research Society of America, Dallas, May 1971.
11. —— AND —— "A Journal Selection Model and Its Implications for a Library System," accepted by *Information Storage and Retrieval*, 1972.
12. LEIMKUHLER, F. F., "Storage Policies for Information Systems," Research Memorandum Series 69-8, School of Industrial Engineering, Purdue University, Lafayette, Indiana, June 1969.
13. —— AND COOPER, M., "Cost Accounting and Analysis for University Libraries," University of California, Office of the Vice-President, Report P-2, Berkeley, California, January 1970.
14. —— AND NEVILLE, A., "The Uncertain Future of the Library," *The Johns Hopkins University Magazine* (Fall 1967).
15. LISTER, W. C., "Least Cost Decision Rules for the Selection of Library Materials for Compact Storage," Ph.D. Dissertation, School of Industrial Engineering, Purdue University, Lafayette, Indiana, 1967.
16. MEIER, R., "Overload and Growth in Communications-Oriented Institutions," in Massarick, F. and Ratoosh, P. (Eds), *Mathematical Explorations in Behavior Science*, Irwin and Dorsey Press, Homewood, Illinois, 1965.
17 METCALF, K. P., "Why We Need to be Investigated," *College and Research Libraries* (October 1954).
18. MIDLER, J. L., "Optimal Control of a Discrete Time Stochastic System Linear in the State," *Journal of Mathematical Analysis and Applications*, Vol 25, No. 1 (January 1969), pp. 114–120.
19. MORSE, P., *Library Effectiveness: A Systems Approach*, MIT Press, Cambridge, Massachusetts, 1968.
20. —— AND ELSTON, C., "A Probabilistic Model for Obsolescence," *Operations Research*, Vol. 17, No 1 (1969).
21. O'NEILL, E. T., "Journal Usage Patterns and Their Implications in the Planning of Library Systems," Ph.D. Dissertation, School of Industrial Engineering, Purdue University, Lafayette, Indiana, January 1970.
22. SAATY, T. L., *Optimization in Integers and Related Extremal Problems*, McGraw-Hill, New York, 1970.
23. STRABLE, E G. (Ed.), *Special Libraries: A Guide for Management*, Special Libraries Association, N. Y., 1966. (A project of the Illinois Chapter.)
24. STRAUS, L. J., STRIEBY, I. AND BROWN, A. L., *Scientific and Technical Libraries: Their Organization and Administration*, Interscience Publishers, Wiley, N. Y., 1964.
25. SWETS, J. A. "Information Retrieval Systems," in Kochen, M. (Ed.), *The Growth of Knowledge*, Wiley, N. Y., 1967.
26. WILLIAMS, G., BRYANT, E. C., WIEDERKEHR, R. AND PALMOUR, V. E., *Library Cost Models: Owning Versus Borrowing Serial Publications*, Center for Research Libraries, Chicago, 1968.
27. WULFEKOETTER, G., *Acquisition Work: Processes Involved in Building Library Collections*, University of Washington Press, Seattle, Washington, 1961.