Recall-Precision Trade-Off: A Derivation

Michael Gordon* and Manfred Kochen
*Computer and Information Systems, Graduate School of Business, The University of Michigan,
Ann Arbor, MI 48109*

The inexact nature of document retrieval gives rise to a fundamental recall precision trade-off: generally, recall improves at the expense of precision, or precision improves at the expense of recall. This trade-off is borne out empirically and has qualitatively intuitive explanations. In this article, we explore this relationship mathematically to explain it further. We see that the recall-precision trade-off hinges on a deceleration in the proportion of relevant documents which are retrieved, successively, over time. Further, we examine several mathematical functions sharing this property and conclude that the equation that best models recall as a function of time is the logarithm of a quadratic function. Our conclusion meets the following requirements: the function we derive predicts non-decreasing recall over time until the last relevant document is retrieved (regardless of the density of relevant documents in the collection) without imposing any artificial restrictions on either what percentage of the collection would need to be examined to achieve perfect recall or what the level of precision would be at that time. Other models examined fail to meet one or more of these criteria.

I. Introduction

It has been widely observed that document retrieval systems produce a trade-off between furnishing relevant documents and failing to furnish irrelevant documents. At one extreme, a literature searcher can ensure that he retrieves all relevant documents by scanning the entire collection (and thereby examining a vast number of irrelevant documents). At the other, he may wish to avoid all irrelevant documents and may attempt to do so by examining only a very few likely relevant documents (thereby missing many relevant ones). In practice, retrieval usually falls between these two extremes.

Similar trade-offs have been discussed in signal detection theory [1, 2], medical diagnostics and decision theory

[3], and pattern recognition studies [4]. In all cases, we have objects that may possess a certain characteristic and also judgments about whether they do. For instance, a document may or may not be relevant, and a retrieval system may or may not retrieve it based on its matching rule, which is a prediction of relevance. Similarly, a person may be diseased or disease-free, a determination a physician diagnosing this individual attempts to make. Consequently, a condition which exists may be detected (which we call a "hit" or "true positive") or not ("miss" or "false negative"). Alternatively, a non-existing condition may be erroneously detected ("false positive" or "false alarm") or correctly judged as absent ("true negative").

Two variables are useful to characterize the performance of a document retrieval system: (1) Recall (or hit rate, h) which we define as the conditional probability of retrieving a document given it is relevant; and (2) Precision (or acceptance, a) which is the conditional probability that a document is relevant given it is retrieved. We may also interpret h as sensitivity and a as accuracy, the terms most commonly used in medical diagnosis [5]. A third variable, Fallout (which we denote b), the conditional probability of retrieving a document given it is irrelevant, is sometimes reported instead of precision. Figure 1 gives an example of estimates for these variables for a single query posed for a collection of 1,000 documents. Recall (hit rate, h) is estimated to be $1/3$ (10 out of 30 relevant documents retrieved); precision (acceptance, a) $1/10$ (10 of 100 retrieved documents are relevant); and fallout (b) $9/97$ (90 of 970 irrelevant documents retrieved). Each of the measures h, a, and b defines a probability measure derived from cumulative retrieval behavior up to a given point in time (here, the retrieval of the first 100 documents). Later, we explicitly include parameters with these measures (i.e., we consider $h(t)$, $a(t)$, and $b(t)$ — t representing time) to study the h vs. a trade-off as more and more documents are retrieved.

Despite many attempts to explicate "relevance" [6-8], there is still no completely satisfactory conceptualization, either operationally or theoretically. Nonetheless, h and a have been estimated and relations between them have been postulated [9], and empirically plotted [10-14]. The intu-

	Retrieved (Signal Noted) (Positive Diagnosis)	Not Retrieved (No Signal Noted) (Negative Diagnosis)	
Relevant (Signal Present) (Condition Exists)	HIT 10	MISS 20	30
Not Relevant (Signal Absent) (Condition Absent)	FALSE ALARM 90	TRUE NEGATIVE 880	970
Total	100	900	1,000

FIG. 1.

itively plausible trade-off between a (accuracy) and h (sensitivity) is typified by Figure 2 (adapted from ref. 15., p. 168). The first mathematical derivation of this inverse relation [16] was based on an assumption of retrieval based on the presence of query terms in the full text of a document and used rank-frequency laws for those query terms. Decision-theoretic based evaluations of document retrieval systems must also be related to precision and recall [17, 18]. Relations between recall and precision have also been analyzed by Yu and Salton [19].

In what follows, we present a mathematical model in an effort to explain the origin of the recall-precision (h vs. a) trade-off.

II. General Considerations

We first establish a mathematical relationship between h (recall, i.e., hit rate) and a (precision, i.e., acceptance):

Lemma

$$a = hp/r \qquad (1)$$

where

$p =$ the probability that a randomly chosen document is relevant to a selected query, i.e., P (relevant)

$b = P$ (document is retrieved | it is irrelevant)

and

$r = hp + b(1 - p)$, i.e., the probability an arbitrary document has been retrieved. $\qquad (2)$

Proof:

$h = P$ (document is retrieved | it is relevant)

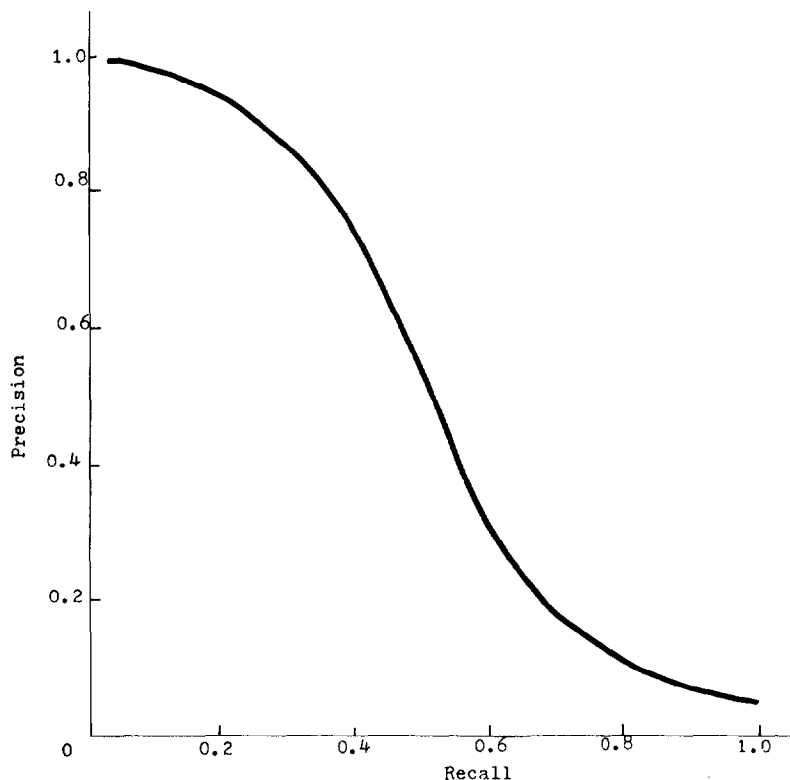$= P$ (retrieved and relevant)/ P (relevant), so that

$hp = P$ (retrieved and relevant).

FIG. 2. Typical recall-precision trade-off.

Similarly,

$$b(1 - p) = P \text{ (retrieved and irrelevant)}.$$

Thus,

$$r = P \text{ (retrieved and relevant)}$$
$$+ P \text{ (retrieved and irrelevant)}$$
$$= P \text{ (retrieved)}.$$

Thus, $hp/r = P \text{ (retrieved and relevant)}/$
$$P \text{ (retrieved)}$$
$$= P \text{ (relevant | retrieved)}$$
$$= a \text{ (acceptance or precision)}$$

as we wished to show. (This lemma is well known; it follows immediately from the definition of conditional probability by Bayes' rule.)

If we assume, as an idealization that does not diminish the value of the argument, that a is a continuous and twice differentiable function of h, we obtain an expression for the derivative, $\partial a/\partial h$. This idealization can be considered to be a curve best fitted to the aggregate retrieval data arising from a system's entire user population, each of whose queries produces an a vs. h scatter plot of points as each successive document is retrieved. Such an interpretation gives rise to functions such as $h(t)$, $a(t)$, and $r(t)$ which, respectively, consider hit-rate, acceptance, and retrieval-rate as continuous functions of time. In all the arguments that follow, it should be kept in mind that a, h, and r are probabilities, not frequencies.

We first show the conditions under which $\partial a/\partial h$ must be negative, that is, the conditions under which precision (a) decreases monotonically as recall (h) increases, by proving the following:

Theorem 1

$$\partial a/\partial h < 0 \quad \text{iff} \quad \partial r/\partial h > r/h$$

Proof: Differentiate equation (1) with respect to h to obtain:

$$\partial a/\partial h = p/r - (hp)/r^2 \, \partial r/\partial h$$
$$= p/r \, [1 - (h/r) \, \partial r/\partial h] \tag{3}$$

Since $p/r > 0$, $\partial a/\partial h < 0$ iff $1 - (h/r) \, \partial r/\partial h < 0$, or   (4)
equivalently, iff $\partial r/\partial h > r/h$.   Q.E.D.

In words, we have shown that precision and recall are inversely related only when $\partial r/\partial h > r/h$. We show next that the condition $\partial r/\partial h > r/h$ is a characteristic of operational retrieval systems. We note that

$\partial r/\partial h > r/h$
iff $\partial(\text{num-ret}/D)/\partial(\text{num-}r\text{-ret}/pD) >$
   $(\text{num-ret}/D)/(\text{num-}r\text{-ret}/pD)$
iff $\partial(\text{num-ret})/\partial(\text{num-}r\text{-ret}) > \text{num-ret}/\text{num-}r\text{-ret}$ (5)

where num-ret stands for number of documents retrieved, num-r-ret stands for number of relevant documents retrieved, and D is the number of documents in the collection.

Again, p refers to the probability of an arbitrary document being judged relevant to a given query under consideration.

Intuitively, we argue that, for any number of relevant documents that a searcher has already retrieved, m, he will have to examine more documents to find the next relevant one than he had to, on average, in retrieving those relevant documents he has found already. Pictorially, Figure 3 illus-

trates this point, the slope of the hypotenuse of the triangle with base $[0, m]$ always being less than the tangent to the curve at m, for all values of m. This observation about the marginal retrieval performance informally validates (5).

More formally, we examine the relationship between $\partial r/\partial h$ and r/h by considering the hit rate as a continuous function of time, $h(t)$, and computing its McLaurin expansion to prove the following:

Theorem 2

$$\partial r/\partial h > r/h \quad \text{iff} \quad h''(0) < 0,$$

where $h(t)$ measures hit rate as a function of time and $h'(t)$ and $h''(t)$ are its first and second derivatives, dh/dt and d^2h/dt^2, respectively.

We assume that $h(t)$ is sufficiently smooth to be represented by its McLaurin series. This is a mathematically strong assumption but not very restrictive of real retrieval systems.

Proof: Express $h(t)$ and $h'(t)$ by their McLaurin series

$$h(t) = h(0) + th'(0) + (t^2/2!)h''(0)$$
$$+ (t^3/3!)h'''(0) + \cdots \tag{6}$$
$$h'(t) = h'(0) + th''(0) + (t^2/2!)h'''(0)$$
$$+ (t^3/3!)h^{(4)}(0) + \cdots \tag{7}$$

Let us suppose it takes 1 time unit to search the entire document collection and examine every document for relevance and that the rate at which documents are searched (retrieved) is constant. Thus

$$r(t) = \begin{cases} t & 0 \le t \le 1 \\ 1 & t > 1 \end{cases} \tag{8}$$

As long as we continue to retrieve, therefore,

$$dr/dt = 1 \tag{9}$$

Additionally,

$$h(0) = 0. \tag{10}$$

As an approximation to (6) and (7) we assume $h^{(k)}(0)$ is negligible for $k \ge 3$.

It follows from these approximations and equations (7) and (9) that, until retrieval ceases:

$$\partial r/\partial h = (dr/dt)/(dh/dt)$$
$$= (dr/dt)/h'(t)$$
$$= 1/(h'(0) + th''(0)) \tag{11}$$

Also, by equations (6), (8), (10) and our approximations:

$$r/h = r(t)/h(t)$$
$$= t/(h(0) + th'(0) + (t^2/2!)h''(0))$$
$$= t/(th'(0) + (t^2/2!)h''(0))$$
$$= 1/(h'(0) + (t/2)h''(0)) \tag{12}$$

We see, therefore, that for any positive time, t, (11) > (12) iff $h''(0) < 0$. Thus we have shown $dr/dh > r/h$ iff $h''(0) < 0$.   Q.E.D.
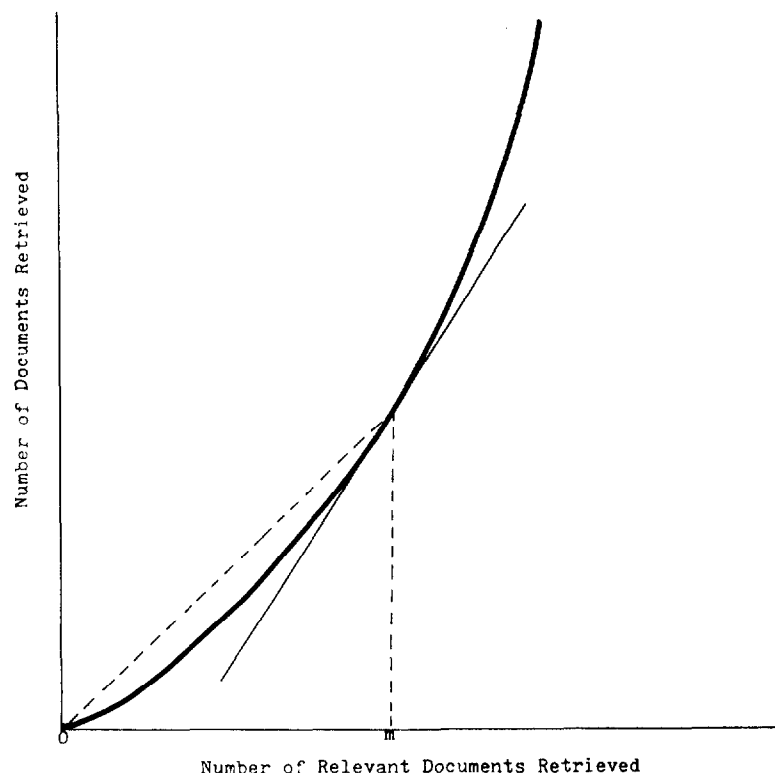
FIG. 3. Retrieved vs. relevant-retrieved documents.

Together Theorems 1 and 2 reveal that precision (a) falls with increasing recall (h) whenever $h''(0) < 0$. The condition $h''(0) < 0$ indicates that the retrieval of relevant documents is decelerating at $t = 0$, when retrieval begins. More generally, we assume that $h''(t)$ is negative for all values of t. This implies that $h'(t)$ is monotonic decreasing. In other words, the rate of retrieving relevant documents (dh/dt) is maximum at $t = 0$. This condition is perfectly plausible. For a recall precision curve representing aggregate retrieval over several queries, it means that the earlier a document is presented, the more likely it is to be relevant. A similar interpretation holds for a recall-precision curve derived from a single query. These interpretations agree with Robertson's "probability ranking principle" [20].

III. Examination of Hit-Rate as a Function of Time

We have seen, logically, that precision (acceptance, a) will decline with increasing recall (hit rate, h) iff $h''(0) < 0$. Further, we have plausibly postulated that $h''(t) < 0$ for all t. In this section, we examine some further constraints on $h(t)$ and discuss them in relationship to various functional forms that might describe $h(t)$.

It is straightforward to provide a best and worst case analysis for recall (h), fallout (b), and precision (a). For a collection with p*100% relevant documents requiring 1 time unit to be searched in its entirety, optimal retrieval occurs if all relevant documents are selected before any non-relevant document is selected. Thus,

$$h_{opt}(t) = \begin{cases} (1/p)t & \text{for } t \le p \\ 1 & \text{for } t > p \end{cases}$$

Equation (2) shows that, for optimal retrieval, fallout, b, is 0 through the retrieval of the last relevant document at $t = p$. That is,

$$b_{opt}(t) = (r(t) - p*h_{opt}(t))/(1 - p)$$
$$= (t - p*(1/p)t)/(1 - p)$$
$$= 0 \quad \text{through } t = p.$$

Additionally, optimal precision is shown by equation (1) to be

$$a_{opt} = h_{opt}p/r$$
$$= (1/p)t*p/t$$
$$= 1 \quad \text{through } t = p.$$

Under worst case retrieval, we assume the retrieval system performs exactly as if it were randomly selecting documents. Thus $h_{worst}(t) = r(t)$ (the hit rate is the same as the retrieval rate). Again using equation (2), we obtain

$$b_{worst} = (r(t) - ph_{worst}(t))/(1 - p)$$
$$= (r(t) - p*r(t))/(1 - p)$$
$$= r(t) \quad \text{through } t = p.$$

For such retrieval, equation (1) shows that

$$a_{\text{worst}}(t) = h_{\text{worst}}p/r$$
$$= r(t)p/r(t)$$
$$= p \quad \text{through } t = p.$$

All of these results are as expected.

We have postulated already that $h(0) = 0$. To continue our analysis, we also assume for now that $h(t)$ is coincident with the graph of best retrieval, $h_{\text{opt}}(t)$ at $t = 0$. (Intuitively we are suggesting that the first document retrieved is relevant.) Since $h_{\text{opt}}(t) = (1/p)t$ for $t \leq p$ we have

$$h'(0) = h'_{\text{opt}}(t) = 1/p. \quad (13)$$

We consider several functional forms of $h(t)$ consistent with these constraints in an effort to model hit-rate (recall, h) as a function of time.

We first consider quadratic functions, that is, those of form:

$$h(t) = \alpha t^2 + \beta t + \gamma \quad (14)$$

and try to determine values for parameters, α, β, and γ. We see that $h'(t) = 2\alpha t + \beta$ and, thus, $h'(0) = \beta = 1/p$. Similarly, $h(0) = \gamma = 0$.

To obtain a value for the coefficient α, we specify the time at which the hit rate, $h(t)$, first becomes 1.00 to be k, $p \leq k \leq 1$.

$$\text{Putting } h(k) = \alpha k^2 + (1/p)k = 1 \quad (15)$$

and solving, we get $\alpha = (p - k)/(k^2 p)$.

Substituting this into (14), we see that, under these assumptions,

$$h(t) = \begin{cases} ((p - k)/(k^2 p))t^2 + (1/p)t & t \leq k \\ 1 & t > k \end{cases} \quad (16)$$

and, thus,

$$h'(t) = 2((p - k)/(k^2 p))t + (1/p) \quad (17)$$

until the last relevant document is retrieved at $t = k$.

We note that hit-rate must be a non-decreasing function from $t = 0$ until $t = k$. We see, in fact, that $h'(t) > 0$ only when $2((p - k)/(k^2 p))t + (1/p) > 0$. That is, since $k > p$, $h'(t) > 0$ only when $t < k^2/(2(k - p))$.

In other words, hit rate increases until $t = k^2/(2(k - p))$. For this to be a realistic model of retrieval, $k^2/(2(k - p))$ must be at least as big as k (time when hit-rate first attains its maximum). Solving $k^2/2(k - p) \geq k$, we see that $k \leq 2p$.

In words, this means that, under this model of retrieval, the last relevant document must be retrieved by twice the time required to retrieve all (and only) the relevant documents. Said differently, at most 50% of the retrieved documents can be non-relevant by the time of retrieving the last relevant one. That is, precision is always at least 50%. Since this behavior is contradicted by ordinary retrieval experience, we reject the possibility that $h(t)$ is quadratic under our assumptions.

We next consider the possibility that $h(t)$ is a logarithmic function of t, that is, $h(t) = \log(f(t))$. This is plausible because the time to find a record in a file is known to be proportional to the logarithm of the file size for many good search algorithms.

Since $h(0) = 0$, we see that $\log(f(0)) = 0$, or, equivalently that $f(0) = 1$. (18)
Additionally,
$h'(0) = 1/p$ means $f'(0)/f(0) = 1/p$, which implies $f'(0) = 1/p$. (19)

We use these two constraints in examining hit-rates which are the logarithms of linear functions, that is, those in which $f(t) = \alpha t + \beta$. From equation (18), we get $\beta = 1$. Using $f'(t) = \alpha$ — together with equation (19) — shows $\alpha = 1/p$. Thus, under our assumptions, if $h(t)$ is the logarithm of a linear function of t, we must have

$$h(t) = \log((1/p)t + 1) \quad (20)$$

This equation increases with t (for all values of p) until the final relevant document is retrieved at $t = t_f$. Since all relevant documents are retrieved when $h(t) = 1$, we have $\log((1/p)t_f + 1) = 1$, or

$$t_f = p*(e - 1) \quad (21)$$

where e is the transcendental number $2.718\ldots$

Since this time, t_f, must be less than the time at which we retrieve the last document in the collection (at $t = 1.00$), we must ensure that

$$p*(e - 1) \leq 1.00 \quad \text{or}$$
$$p \leq .582 \text{ (approximately)} \quad (22)$$

This constraint, that no more than 58.2% of the documents be relevant to a query, does not restrict the class of queries in any significant way. However, by equation (1), if all relevant documents are retrieved at $t_f = p*(e - 1) \cong p*1.718$, then, at that time, when the hit-rate is perfect (100%), precision, a, is $1/1.718 \cong 58.2\%$. Unfortunately, precision is normally far below this level even for levels of recall considerably less than 100%. Thus, we reject the possibility that hit-rate is modelled by equation (20).

The last candidate formula for hit-rate that we will consider is

$$h(t) = \log(\alpha t^2 + \beta t + \gamma) \quad (23)$$

We consider two sets of constraints in determining values for the parameters in this equation. First, using $h(0) = 0$, we obtain $\gamma = 1$. Similarly, $h'(0) = 1/p$ implies that $\beta = 1/p$. The value of α in the equation for $h(t)$ above may be made to depend on the time at which all relevant documents become retrieved, which we denote by k, $p \leq k \leq 1$. Solving $\log(\alpha k^2 + (1/p)k + 1) = 1$ for α, we see that

$$h(t) = \log(((e - (k/p + 1))/k^2)t^2 + (1/p)t + 1) \quad (24)$$

which is the first form of equation (23) we consider.

For equation (24) to be suitable, it must be an increasing (or nondecreasing) function in the interval $[0, k]$. Differentiating (24), we get

$$h'(t) = ((2/pk^2)(p(e - 1) - k)t$$
$$+ (1/p))/((1/pk^2)(p(e - 1) - k)t^2$$
$$+ (t/p) + 1) \qquad (25)$$

We set equation (25) equal to 0 and solve for t to see that $h(t)$ has a local extremum at

$$t_e = k^2/(2(k - p(e - 1))). \qquad (26)$$

We see that $t_e > 0$ when

$$k/p > e - 1. \qquad (27)$$

Also, $t_e < k$ when

$$k/p > 2(e - 1). \qquad (28)$$

Thus, by equations (27) and (28), $h(t)$ has a local extremum between $t = 0$ (when retrieval begins) and $t = k$ (when hit-rate reaches 1.00) for $k/p > 2(e - 1) \cong 3.436$. Notice that p/k reveals what the level of precision is when recall reaches 100%. The condition $k/p > 3.436$, therefore, is equivalent to the realistic condition that precision be less than 29.1% when recall reaches 100%. Since $h(t)$ will not be monotonic under these conditions, we must reject equation (24) as a candidate function for $h(t)$ for general values of p and k.

Instead, we consider $h(t) = \log(\alpha t^2 + \beta t + \gamma)$ with an alternative set of hypotheses to derive a function consistent with the retrieval behavior we expect. In particular, we maintain the hypotheses that $h(0) = 0$ and $h(k) = 1.00$. But we replace the hypothesis that $h'(t) = 1/p$ by $h'(k) = 0$. The intuition behind this assumption is that the hit-rate ceases to increase when the last relevant document is retrieved. As before, $h(0) = 0$ implies $\gamma = 1$. Further by stipulating that $h(k) = 1$ we see

$$\log(\alpha k^2 + \beta k + 1) = 1 \quad \text{or}$$
$$\alpha k^2 + \beta k + 1 = e \qquad (29)$$

From $h'(k) = 0$ we obtain

$$(2\alpha k + \beta)/(\alpha k^2 + \beta k + 1) = 0 \qquad (30)$$

Solving (29) and (30) for α and β, we see that

$$h(t) = \log(((1 - e)/k^2)t^2 + (2(e - 1)/k)t + 1) \qquad (31)$$

Note that differentiating $h(t) = \log(\alpha t^2 + \beta t + 1)$ gives $h'(t) = (2\alpha t + \beta)/(\alpha t^2 + \beta t + 1)$ which has only one local extremum — at $t = -\beta/(2\alpha)$ — for any values of α or β. By hypothesis, (31) contains an extremum at $t = k$. Thus, equation (23) has no extrema in the interval $(0, k)$ and is, therefore, monotonic increasing during this interval as we desire.

By differentiating equation (31) and evaluating at $t = 0$, we see that

$$h'(0) = 2(e - 1)/k \qquad (32)$$

Since the instantaneous hit-rate at $t = 0$ must not exceed the rate of optimal retrieval [equation (13)] we get $2(e - 1)/k \le 1/p$, or, equivalently,

$$k/p \ge 2(e - 1) = \text{(approximately)} \; 3.436. \qquad (33)$$

This restriction in the values k can take on in relation to p indicates that, under this model, recall will fall short of 100% at least until one has examined 3.43 times as many documents as ultimately prove to be relevant. That is, 29% (approximately) is the best precision possible for those seeking perfect recall.

IV. Conclusion

A mathematical model of the recall-precision (a vs. h) trade-off is useful in several ways. One, it can shed light on the limits of effective retrieval: What precision must one tolerate to achieve a specified level of recall? Two, a theoretical formulation of this trade-off can help in the building of document retrieval simulations from which operational systems can be derived.

In developing such a model, we have shown that a recall-precision trade-off derives from recall, $h(t)$, decelerating as more documents are retrieved. We have examined several such mathematical functions modeling $h(t)$ to see which seems most consistent with expected retrieval behavior.

We have concluded that the best form is $h(t) = \log(((1 - e)/k^2)t^2 + (2(e - 1)/k)t + 1)$, primarily since it is monotonic until $h(t)$ reaches 1.00, predicts realistic levels of precision, and imposes no unrealistic restrictions on when hit-rate reaches 1.00. (In this model, 1 time unit is required to search the entire database which contains $p\%$ relevant documents, and the last relevant document is retrieved at $t = k$, $p \le k \le 1$.)

Curves showing the recall-precision behavior predicted by this model are shown in Figure 4. Note that equation (31) defines a family of curves parameterized by k. (k can be regarded as the time at which the last relevant document is retrieved or, equivalently, the fraction of the entire collection that has been searched when the last relevant document has been retrieved.) Since equation (31) is monotonic from $t = 0$ through $t = k$, knowing what percentage of a real collection must be searched to achieve any specified level of recall determines a particular value for the parameter k.

Future research should be directed at comparing how well this theoretical model predicts actual retrieval and at estimating its parameter, k, for particular retrieval methods. Justification of the model, together with a determination that different values of k are appropriate for different retrieval methods, presents an analogy to coding methods in Shannon and Weaver's mathematical theory of communication [21]. Like channel capacity, our research may offer a bound on retrieval effectiveness for various retrieval methods.
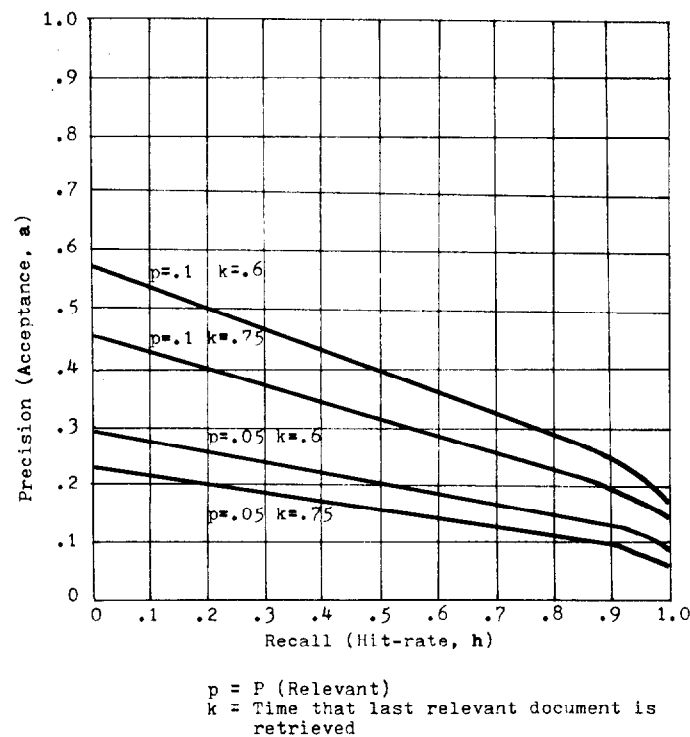
p = P (Relevant)
k = Time that last relevant document is
    retrieved

FIG. 4. Derived recall-precision relations.

References

1. Peterson, W. W.; Burdrall, T. G.; Fox, W. C. *IRE Trans. Info. Th.* 4:171; 1954.
2. Van Meter, D.; Middleton, D. *IRE Trans. Info. Th.* 4:119; 1954.
3. Wald, A. *Statistical Decision Functions.* New York: Wiley; 1950.
4. Tanner, W. P. Jr.; Swets, J. A. *Psychol. Rev.* 61:401; 1954.
5. Swets, J. A. "Information Retrieval Systems," *Science* 141:245–250; July 19, 1963.
6. Goffman, W. "On Relevance as a Measure," *Info. Storage and Retrieval,* 2(3):201–203; December 1964.
7. Saracevic, T. "Relevance: A Review of and a Framework for the Thinking on the Notion in Information Science," *Journal of the American Society for Information Science,* 26(6):321–343; 1975.
8. Cooper, W. S. "A Definition of Relevance for Information Retrieval," *Info. Storage and Retrieval,* 7(1):19–37; 1971.
9. Cleverdon, C. W. "On the Inverse Relationship of Recall and Precision," *J. Doc.,* 28:195–201; 1972.
10. Cleverdon, C. W. "Progress in Documentation: Evaluation Tests of Information Retrieval Systems," *J. Doc.,* 26(1):55–67; 1970.
11. Salton, G. *The Smart Retrieval System-Experiments in Automatic Document Processing,* Englewood Cliffs, N.J.: Prentice-Hall; 1971.
12. Lancaster, F. W. *Evaluation of the Medlars Demand Search Service,*

National Library of Medicine, Bethesda, January 1968.
13. Lancaster, F. W. *Information Retrieval Systems-Characteristics, Testing and Evaluation,* 2nd Ed., New York: Wiley, 1979.
14. Blair, D.; Maron M. E. "An Evaluation of Retrieval Effectiveness for a Full-Text Document Retrieval System," *COMM ACM,* 28(3):289–299; 1985.
15. Salton, G.; McGill, M. J. *Introduction to Modern Information Retrieval,* New York: McGraw-Hill, 1983.
16. Kochen, M. *Principles of Information Retrieval,* Los Angeles: Wiley-Becker and Hayes/Melville, 1974, p. 156–163.
17. Kraft, D.; Bookstein, A. "Evaluation of Information Retrieval Systems: A Decision Theoretic Approach," *Journal of the American Society for Information Science,* 29(1):31–40; 1978.
18. Bookstein, A.; Swanson, D. R. "A Decision Theoretic Foundation for Indexing," *Journal of the American Society for Information Science,* 26(1):45–50; 1975.
19. Yu, C. T.; Salton, G. "Precision Weighing — An Effective Automatic Indexing Method," *Journal of the American Society for Information Science,* 23(1):76–88; 1976.
20. Robertson, S. E. "The Probability Ranking Principle in IR," *J. Doc.* 33:294–304; 1977.
21. Shannon, C.; Weaver, W. *The Mathematical Theory of Communication,* University of Illinois Press, Urbana, 1964.