Brief Communication

A Model for the Stopping Behavior of Users of Online Systems*

Paul B. Kantor
Tantalus, Inc., 3257 Ormond Road, Cleveland, Ohio 44118

We examine a model in which the user of an online system continually updates his/her estimated probability of success, and quits or continues according to the expected utility of each action. The prior distribution of the unknown probability is a beta distribution, with mean determined by the *a priori* expectation of success, and variance determined by the confidence with which the user has that prior expectation. The stopping criterion depends upon the accumulated number of positive and negative reinforcements, and is a straight line in a suitable coordinate system.

The user of an information-retrieval system reasons at two different levels. At the "working" level, she decides what steps to take next in carrying out the search at hand. At the "monitor" level, she simultaneously decides whether to continue the search at all or to terminate it. The same is true of a researcher in a library, an auditor reviewing financial records, or a detective of any kind. In this article we propose a very general model for the "monitor" process and apply it to the information-retrieval situation. We show that there is a very simple cutoff criterion which determines whether the search will be terminated.

The resilience of the searcher to repeated failure is found to depend in a natural way on both the *a priori* estimate of success and the believed precision of that *a priori* estimate of success. In other words, a person who expects a 50% chance of success *a priori* will be more resistant to failure than one who expects 20% chance of success, and a person whose estimate of 50% is marked by a high degree of certainty will be even more resistant to failure.

The entire analysis is based on a Bayesian formulation in which we use the most natural, i.e., the beta distribution, model for the prior uncertainty in the probability of success. This model has two parameters [1] which can be converted naturally into the estimated probability of success and the precision of that estimate. This model is "natural" in the sense that its form remains invariant as new information is added. The model also incorporates the notion of expected utility and takes account of the utility, generally negative, of continuing a search and the utility of various possible results for the outcomes of taking the next step.

The Bayesian Model

The Bayesian model is a powerful technique for estimating an unknown probability. When a probability is unknown we can only say that it may assume any of several values. Since a probability must lie between 0 and 1, the natural starting point is to say that it is somehow distributed between 0 and 1. For example, it might have a unform distribution. In the Bayesian method, an assumed distribution for the unknown parameter is revised ("updated") as new information is received. This information comes in the form of events whose precise nature is governed by the probability to be estimated. In our particular example the two possible events are that the next item retrieved is good or that it is bad. If the probability that an item is good is "p," then finding a good item should cause us to update our assumption about the value of p in a way that makes it larger than it was before. Mathematically, the Bayesian updating procedure takes the initial probability distribution $f(p)$, multiplies it by the probability of the observed occurrence, whatever the value of p might be, and then readjusts the normalization so that the new function is again a probability distribution. In the example given, the updating due to a good document replaces $f(p)$ by $pf(p)$, and renormalizes it by dividing by the original expected value of p, $<p>$, giving $pf(p)/<p>$.

This is a powerful and popular technique for dealing with unknown probabilities. To summarize, the un-

*Supported in part by the National Science Foundation under Grant No. IST 83-18630.

Received October 10, 1983; accepted April 18, 1986

known probability (p) is replaced by a distribution, which represents our present state of ignorance about that probability. This is called the prior distribution. The distribution is then updated according to a rigorous mathematical procedure. Objections to the Bayesian method, and the objectors are numerous, point out that all of this mathematical rigor is built on a foundation of sand, since our original assumptions about the distribution of the parameter may be entirely wrong.

In this case, the fact that the assumptions may be wrong are not a grave impediment. We are using the Bayesian model to describe whether, and under what circumstances, a searcher holding a certain view about the prior distribution will give up on a search. Thus, whether his original view is right or wrong, the model ought to be applicable. In what follows, we consider only two possible events, the finding of a good item, which introduces a factor p, and the finding of a bad item, which, correspondingly introduces a factor $(1 - p)$.

The Decision Model

When a person has a choice C vs C' and there are two possible outcomes R vs R', the decision C or C' is, in a "rational Von Neuman-Morgenstern scheme," made by evaluating the expected utility of each of the two alternatives and selecting the one with the larger utility. It is perfectly clear, by introspection alone, that people do not actually perform this calculation moment by moment during their working days. However, those who are more successful, and who get more utility out of their working days, probably subconsciously do something close to it. That is, they are continually aware of new information which comes to them, and they revise their estimates of the value of continuing with a given course of action. Thus the model we consider is not fully prescriptive of human behavior, but it is very likely to be descriptive of the broad features of this behavior. The expected utilities are given in terms of the conditional probabilities $p(R \mid C)$, $p(R' \mid C)$, $p(R \mid C')$, and $p(R' \mid C')$, as Eqs. (1) and (2). We use the fact that $p(R \mid C) + p(R' \mid C) = 1$ and $P(R \mid C') + P(R' \mid C') = 1$ to simplify:

$$v(C) = p(R \mid C)v(R) + p(R' \mid C)v(R') + v_0(C)$$
$$= p(R \mid c)[v(R') - v(R')] + v(R') + v_0(C), \quad (1)$$

$$v(C') = p(R \mid C')v(R) + p(R' \mid C')v(R') + v_0(C')$$
$$= p(R \mid C')[v(R) - v(R')] + v(R') + v_0(C'). \quad (2)$$

Here, v_0 is an intrinsic value (or a cost, if $v_0 < 0$) of the indicated choice. The conditional probabilities are themselves unknown. They will be estimated using the Bayesian model. In our particular case R stands for "relevant information retrieved," or, from a behavioral viewpoint, "positive reward." Thus, R' stands for "nonrelevant information retrieval," or "negative reward." C represents the choice to continue, so that C' represents the choice not to continue. The conditional probability of retrieving either relevant or nonrelevant documents is 0 if the searcher quits, so that $v(C') = v_0(C')$.

The fact that v_0 may depend on how much searching has already occurred and on the history of the search so far introduces some computational complexity, but no conceptual difficulties, and will not be discussed here (see [2]). Letting $p = p(R \mid C)$, we can write $v(c) = v(C \mid p) = v_0(C) + [v(R) - v(R')]p + v(R')$. The parameter p is assumed to be governed by the beta distribution

$$P(p = t) = (\text{constant}) * t^{a-1} * (1 - t)^{b-1} = g(t;a,b), \quad 0 \le t \le 1, \quad a,b \text{ integers}. \quad (3)$$

In this distribution the expected value of p is $E(p) = a/(a + b)$. Given this uncertainty about p, the expected value of $v(c)$ is given by

$$\int dt\, v(C,t)g(t;a,b) = v_0(C) + v(R') + [a/(a + b)][v(R) - v(R')]. \quad (4)$$

With repeated trials, the shape of the distribution g changes. At any step, if the prior distribution is represented by g, then $P(p = t \mid R)$ is proportional to $tg(t)$. Similarly, $P(p = t \mid R')$ is proportional to $g(t)(1 - t)$. Hence with an accumulation of S successes and F failures, the initial distribution g will be multiplied by t or $1 - t$, according to Bayes's rule, S and F times, respectively.

$$g(t;a,b \mid S,F) = \text{constant} * t^{a+S-1}(1 - t)^{b+F-1}. \quad (5)$$

The expected value of the choice to continue is then given by

$$v' = v_0(C) + v(R') + (a + S)/(a + b + S + F)[v(R) - v(R')]. \quad (6)$$

We assume that $v(R) > v(R')$. Hence the condition to continue is that $(a + S)/(a + b + S + F)$ remain large enough, i.e.,

$$(a + S)/(a + b + S + F) > [v_0(C') - v_0(C)]/[v(R) - v(R')] \equiv \pi. \quad (7)$$

Thus, the condition for continuing is that the slope of the line joining the point (F,S) to the point $-(b,a)$ exceed the value $\pi/(1 - \pi)$. This critical value depends in a natural way on the relative value of the retrieved information and the cost of continuing. If the cost is high, the line is steep. If the value of retrieved documents is high, the slope is small, and the searcher can tolerate a large number of failures before quitting.

Examples are shown in Figure 1, where the region below the line is the region in which one should quit. The region above the line is the region in which one should continue. It is easy, although somewhat cumbersome, to reexpress the parameters a and b in terms of the *a priori* estimate of the probability $a/(a + b)$ and an estimate of the precision of this probability. With reference to standard tables we see [3, p. 930, row 26.1.33] that the variance of this *a priori* estimate is given by Eq. (8).

$$\text{variance} = (a + b)/[(a + b)^2 (a + b + 1)]. \quad (8)$$

The standard error of the *a priori* estimate is given by the square root of this variance. If both the parameters a and b are multiplied by a positive number Q^2, the result is that the estimate of the prior possibility remains unchanged, but the estimate of the standard error decreases by a factor of Q. Hence larger values of a and b correspond to more precise estimates of the prior probability. The corresponding distribution function is more sharply peaked. Similarly, relatively low values of a and b correspond to a broad distribution representing the fact that the prior belief specifies a certain mean but does not do so with great precision or, in the psychological sense, confidence. This effect is shown by the various lines in Figure 1.

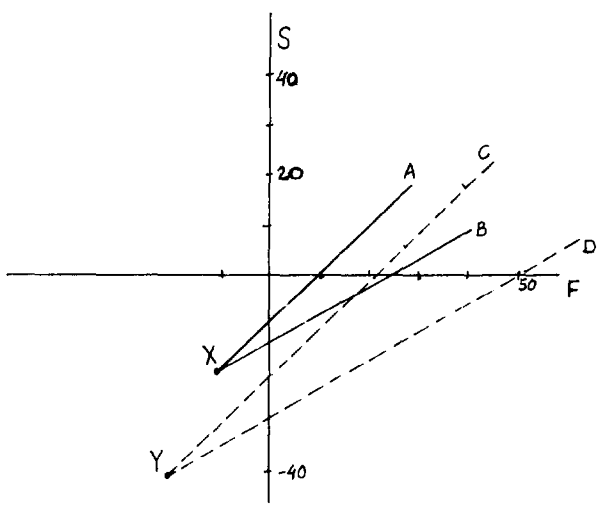

FIG. 1. The quitting boundaries. A user begins at the origin ($S = F = 0$). He will tolerate failure, without success, until reaching the stopping boundary. The boundaries are shown for several values of $p \equiv \pi/(1 - \pi)$.

A: Prior 2/3	$a = 20$; $b = 10$	$r = 1/2$
B: Prior 2/3	$a = 20$; $b = 10$	$r = 4/11$
C: Prior 2/3	$a = 20$; $b = 20$	$r = 1/2$
or Prior 3/4	$a = 30$; $b = 10$	$r = 1/2$
D: Prior 2/3	$a = 40$; $b = 20$	$r = 4/11$

Note that *every* line has more than one interpretation. For example, line C corresponds to a prior estimate of 3/4, or an estimate of 2/3, with a narrower uncertainty. Generally, the higher the prior (or, the more firmly it is "believed"), the more resistant the patron will be to discouragement.

Note that each stopping boundary corresponds to infinitely many prior estimates of the probability of success, with different estimates of the variance. The origination points X and Y both refer to a prior of 2/3. For X ($a = 10$, $b = 20$), the patron can tolerate ten successive failures (if π, the value ratio is 0.5.); that is $(a + 0)/(a + 0 + 10) = 0.5$. But with ($a = 20$, $b = 40$) the line originates at y, and the patron can tolerate 20 successive failures.

Realistic experiments [5] suggest that the density of relevant items, even in a well-chosen retrieved set, may be as low as 0.04. In this case an initial string of ten failures has a probability of $(0.96)^{10} = 0.6648$. So a searcher with a totally unrealistic estimate of $p = 2/3$ and high confidence in that estimate (var $= (10 \times 20)/(31 \times 30 \times 30)) = 0.007$ has a 66% chance of giving up before getting a good document. If this original confidence is low, he will almost surely quit. For example, with $a = 2$, $b = 4$ he will quit at the second or third failure, with probability 92% or 88%, respectively.

Discussion

Does this research have implications for real search strategy? It suggests that if items are examined serially: (1) it is important that the end user have either a realistically low estimate of the chance of success or an unreasonably firm confidence in an unrealistically high estimate of the chance of success, and (2) it is vital that the system be as effective as it can in putting the relevant documents to the head of the list. There is an interesting dualism between this argument, stimulated by some observations of Albert and Kraft [4] and the recent note by Bookstein [2]. Bookstein has suggested that a retrieval system obtain relevance statements from the user and continually update its estimate of some probability distribution. Although his particular example involved the notion of term independence, which is a difficult one to defend [6,7], the principle is, apart from that, a sound one. As the present article suggests, the system user interface may very well represent a dialogue between two Bayesian systems which are, or should be, updating their estimates of probabilities of success as a search progresses.

It would indeed be interesting to attempt an experimental test of whether the user of the information system does follow this rational Bayesian model. There are a number of difficulties—obtaining prior estimates of the probability of success and the precision of this estimate (in order to determine the parameters a and b) as well as identifying whether an individual step in the process is regarded as a success, a failure, or merely neutral. Detailed study of transaction logs [8] shows that there are many events of a housekeeping nature which are to be regarded as neither successes or failures in a model of this type.

This very simple model suggests that there is practical value in learning more about actual rates of success, and about end-user quitting behavior.

References

1. Lindley, D. V. *Introduction to Probability and Statistics. Part 2: Inference.* London: Cambridge University Press; 1965. See especially pages 143-144.
2. Bookstein, Abraham. "Information Retrieval: A Sequential Learning Process." *Journal of the American Society for Information Science.* 34(5):331-342; 1983.
3. Abramowitz, M.; Stegun I. *Handbook of Mathematical Functions.* New York, NY: Dover Press; 1965.
4. Albert, Danny; Kraft, Donald H. "A Dynamic Search Stopping Rule for an Information Storage and Retrieval System." In P. P. Cooper and J. W. Hall, Eds.: *Management of Information Systems.* Proceedings of the Mid-Year Meeting of the American Society for Information Science Houston, TX; May 21–24, 1978. Dallas, TX: Xerox Corporation; 1978. Fiche 4.
5. Chamis, A., Saracevic, T., Trivison, D. "Research on Information Seeking and Retrieving: A Progress Report." In: *Proceedings of the National Online Meeting, New York, 1986.* Melford, NJ: Learned Information, Inc.; 1986.
6. Yu, C. T.; Buckley, C.; Lam, K.; Salton, G. "A Generalized Term Dependence Model in Information Retrieval." *Report No. TR83-543.* Ithaca, NY: Cornell University Department of Computer Science; 1983.
7. Kantor, Paul B. "Maximum Entropy and the Optimal Design of Automated Information Retrieval Systems." *Information Technology.* 3(2):88–94; 1984.
8. Chapman, Janet. "A State Transition Analysis of Online Information-Seeking Behavior." *Journal of the American Society for Information Science.* 32(5):325–333; 1981.