Maximum entropy and the optimal design of automated information retrieval systems

Paul B Kantor

The application of the maximum entropy principle is extended to problems of information storage and retrieval. The extension includes continuous or 'fuzzy' relevance valuations, fuzzy descriptors, and prior or feedback constraints. A decomposition property of the entropy function is used to express the total entropy in terms of the entropy of nonoverlapping components. Each component is described by a richness parameter which is determined by a set of coupled constraint equations given in closed form. A method is outlined for solving those equations in real time, and possible grounds for applying the maximum entropy principle are explored. The relation to term weighting, and the possibility of constructing rigorous relations between information and effort, are also discussed.

Keywords: information retrieval, information storage, entropy, richness parameters

Cooper[1] and Huizinga[2] have suggested that the maximum entropy principle (MEP) will correctly determine the retrieval value of the various components of an information reservoir[3]. According to this principle, the 'value' (or 'relevance') is distributed within the reservoir in such a way that all known constraints on structure and content are met and the entropy function is maximized.

The notion of entropy maximization has great importance in the study of physical ensembles[4] where its power traces to the fact that the microstate of maximum entropy is the most probable microstate and nearly all allowed microstates are quite close to it. The notion of entropy maximization plays an important role in the statistical theory of contingency tables[5] where it prescribes a 'null hypothesis' in the presence of any number of complicated marginal conditions. In this role it is not subject to 'experimental verification', but it has clearly had a great impact. In essence, the MEP makes precise the notion of 'at random' in complicated constrained situations.

In the area of information retrieval, the MEP serves as a design principle which is amenable to experimental test. That is, given some constraints on the distribution of value, the MEP leads to a unique determination of the value of any Boolean atomic component of the reservoir. For examples given four terms A, B, C, D and estimates of the expected (average) value per item in A, in C, in $B \cup D$ and in $A \cup C \cup B$, the MEP leads to a unique estimate of the value in any atomic component such as $A\bar{B}C\bar{D}$, etc.

In a companion paper[6], this problem has been analysed completely, using the calculus of variations, for several important situations. The present paper aims to provide a less technical summary of the key results. To fix notation, suppose that there are only three descriptors, defining subsets A, B, C of a reservoir R.

$$A = \{r \varepsilon R \colon A(r) = 1\}$$

The case of fuzzy descriptors adds no essential features[6]. However, it is realistic and important to recognize that value may be neither 0 nor 1 but somewhere in between. Maron, Robertson and Cooper[7, 8] have described such values as probabilities of relevance, while Salton[9] has described them, in objective terms, by inner products in a suitable vector space.

The present analysis makes use of the entropy function, defined on any nonnegative distribution $p_1, \ldots, p_m \geq 0$; $p_1 + p_2 + \ldots + p_m = 1$.

Tantalus Inc., Suite 218, 2140 Lee Road, Cleveland, Ohio 44118, UK
Received 28 June 1983, revised 10 January 1984

$$S(p) = -p_1 \ln p_1 - p_2 \ln p_2 - \ldots - p_n \ln p_n \qquad (1)$$

$\ln p$ is the natural logarithm of p. If $x = 0$, $x \ln x = 0$

Also, the decomposition property of the entropy function is used. If

$$n(\alpha) \geq 0 \quad \text{all } \alpha \varepsilon Q \qquad \sum_{\alpha \varepsilon Q} n(\alpha) = 1 \qquad (2a)$$

and for each choice of

$$p(\alpha, v) \geq 0 \; v \varepsilon V \qquad \sum_{v \varepsilon V} p(\alpha, v) = 1 \qquad (2b)$$

and

$$q(\alpha, v) \equiv n(\alpha) p(\alpha, v) \qquad (2c)$$

so that

$$q(\alpha, v) \geq 0, \; \sum_{\alpha, v} q(\alpha, v) = 1 \qquad (2d)$$

then

$$S(q) = S(n) + \sum_{\alpha \varepsilon Q} n(\alpha) S(p(\alpha)) \qquad (3)$$

All the possible values of α and v are represented as the axes of a table

α	Structure	v
0 . . .		
α	$n(\alpha)$	$\ldots \; q(\alpha, v) = p(\alpha, v) \cdot n(\alpha)$
1		

The decomposition property (which may be proved algebraically) means that the entropy of the entire distribution is the entropy of the structure column (the $n(\alpha)$) plus the weighted sum of the entropy of the rows, each row having weight $n(\alpha)$.

It is assumed[6] that, within each row, the value of any item may lie anywhere between 0 and 1. Thus $q(\alpha, v)$ is precisely that fraction of the entire reservoir which has 'description α and value v'.

By application of the Lagrangian method for constrained optimization, it has been shown that every row α may be characterized by a richness parameter ρ, and that, by the MEP:

$$p(\alpha, v) = \frac{\rho(\alpha) e^{\rho(\alpha) v}}{e^{\rho(\alpha)} - 1} \geq 0 \qquad (4)$$

$$\int_0^1 dv \, p(\alpha, v) = 1 \qquad (5)$$

It has been shown[6] that the richness parameters are determined by constraints on the value content of the rows. The results of that analysis are reviewed briefly using simple examples. Suppose that a given problem has 8 atoms and 3 constraints, whose effect is indicated as:

α
1
2
3 } Constraint 1
Constraint 2 { 4
5
6 } Constraint 3
7
8

(Constraint 1 brackets atoms 1–5; Constraint 2 brackets atoms 3–6; Constraint 3 brackets atoms 4–7.)

The richness parameters[6] $\rho_1, \ldots, \rho_8$ are given by

$$\begin{aligned} \rho_1 = \rho_2 &= \lambda_1 \\ \rho_3 &= \lambda_1 + \lambda_2 \\ \rho_4 = \rho_5 &= \lambda_1 + \lambda_2 + \lambda_3 \\ \rho_6 &= \lambda_2 + \lambda_3 \\ \rho_7 &= \lambda_3 \\ \rho_8 & \quad \text{undetermined} \end{aligned} \qquad (6a\text{–}f)$$

The Lagrange multipliers $\lambda_1, \lambda_2, \lambda_3$ are in turn determined by 3 content equations; giving the 'total value' believed to be in groups 1, 2 and 3.

$$\begin{aligned} V_1 &= n(1)v(\lambda_1) + n(2)v(\lambda_1) + n(3)v(\lambda_1 + \lambda_2) \\ &\quad + n(4)v(\lambda_1 + \lambda_2 + \lambda_3) + n(5)v(\lambda_1 + \lambda_2 + \lambda_3) \\ V_2 &= n(3)v(\lambda_1 + \lambda_2) + n(4)v(\lambda_1 + \lambda_2 + \lambda_3) + \\ &\quad n(5)v(\lambda_1 + \lambda_2 + \lambda_3) + n(6)v(\lambda_2 + \lambda_3) \\ V_3 &= n(4)v(\lambda_1 + \lambda_2 + \lambda_3) + n(5)v(\lambda_1 + \lambda_2 + \lambda_3) \\ &\quad + n(6)v(\lambda_2 + \lambda_3) + n(7)v(\lambda_3) \end{aligned} \qquad (7a\text{–}c)$$

Since $v(\rho) = \frac{1}{2} + \frac{1}{2}(\coth \rho/2 - 2/\rho)$ these equations can only be solved numerically.

To describe the general situation, another table may be formed, the rows of which correspond to the atomic components and the columns of which correspond to the content constraints. A check at the intersection of a row and column indicates that the constraint applies to the corresponding component.

The present case would be represented as:

α	1	2	3
1	√		
2	√		
3	√	√	
4	√	√	√
5	√	√	√
6		√	√
7			√
8			

The set of all atomic components is denoted by A and the set of constraints by P. The constraints involving α are represented by

$$P(\alpha) = \{Q \varepsilon P \; Q \text{ constrains row } \alpha\} \qquad (8)$$

and the affected components by

$$A(P) = \{\alpha \varepsilon A \; P \text{ constrains row } \alpha\} \qquad (9)$$

The results of the analysis are that the richnesses ρ are determined by

$$\rho(\alpha) = \sum_{Q \varepsilon P(\alpha)} \lambda_Q \qquad (10)$$

The Lagrange multipliers λ are determined from the content constraints

$$V_P = \sum_{\alpha \varepsilon A(P)} v(\rho(\alpha))n(\alpha) \tag{11}$$

or,

$$V_P = \sum_{\alpha \varepsilon A(P)} v\left(\sum_{P \varepsilon P(\alpha)} \lambda_P\right) \cdot n(\alpha) \tag{12}$$

THE COOPER–HUIZINGA ENHANCEMENT PHENOMENON

Without solving the equations that determine the Lagrange multipliers (and thus the effective temperatures), we may see, qualitatively, the enhancement phenomenon noted by Cooper and Huizinga. We have the basic relation

$$\rho_\alpha = \sum_{P \varepsilon P(\alpha)} \lambda_P \tag{13}$$

The situation is represented by a table with rows labelled by α and columns labelled by the conditions imposed. An entry appears whenever the condition applies to the row. For example, with 3 'terms' A, B, C we might have

α	V_A	V_B	V_C	(C-H) V_R	(MC) V_{R^*}
$\Phi = \overline{ABC}$				√	
$\overline{AB}C$			√	√	√
$\overline{A}B\overline{C}$		√		√	√
$\overline{A}BC$		√	√	√	√
$A\overline{BC}$	√			√	√
$A\overline{B}C$	√		√	√	√
$AB\overline{C}$	√	√		√	√
$\Pi = ABC$	√	√	√	√	√
Lagrange multiplier	λ_A	λ_B	λ_C	λ_D	λ_{R^*}

Let us compare $\overline{A}BC$ with $\overline{AB}C$, $AB\overline{C}$ using the Cooper–Huizinga assumption

$$\begin{aligned} \rho(\overline{A}BC) &= \lambda_B + \lambda_C + \lambda_P \\ \rho(\overline{AB}C) &= \lambda_C + \lambda_R \\ \rho(\overline{A}B\overline{C}) &= \lambda_B + \lambda_R \\ \rho(\overline{A}BC) &= \rho(\overline{AB}C) + \{\rho(\overline{A}B\overline{C}) - \lambda_R\} \end{aligned} \tag{14a–d}$$

However, λ_R determines the $\rho(\overline{ABC})$ and since this atom is surely poorer than $\overline{AB}\overline{C}$, we must have $\lambda_R > \rho(\overline{ABC})$; therefore

$$\rho(\overline{A}BC) > \rho(\overline{AB}C) \tag{15}$$

and so the intersection of B and C is enriched relative to $\overline{B}C$, independent of A.

In the minimum constraint formulation, λ_R is replaced by λ_R* which corresponds to the richness of the set determined by the (Boolean 'OR') union of the descriptors. In this case, it is seen that

$$\rho(\overline{A}BC) > \rho(\overline{AB}C) \tag{16}$$

if and only if $\rho(\overline{A}B\overline{C}) > \lambda_R$* — that is, roughly speaking, if $\overline{A}B\overline{C}$ is 'richer than average' within the union. In the minimal constraint case, we must recognize that this need not always be true. Thus the Cooper–Huizinga enhancement of intersections may be a consequence of the use of the constraint, and may be sometimes violated. Clearly, it need not be violated, since we might have

$$\rho(\overline{ABC}) < \rho(\overline{A}B\overline{C}) < \rho(A\overline{BC}) < \rho(AB\overline{C}) < \ldots < \rho(ABC) \tag{17}$$

SUMMARY

To summarize, application of the MEP has been reduced to the problem of computationally determining the Lagrange multipliers subject to conditions such as (7). It has also been found that the richness parameter ρ is related to an 'effective temperature' T by the equation $T = -1/\rho$. The value $\rho = 0$, which corresponds to 'average richness' corresponds also to $T = \pm\infty$. Positive values of ρ correspond to negative temperatures and vice versa.

The very important function v, which expresses the mean value of items in a row in terms of the richness of that row, is:

$$v(\rho) = \frac{1}{2} + \frac{1}{2}(\coth \rho/2 - 2/\rho) \tag{18}$$

The hyperbolic cotangent is given by

$$\coth x = \frac{e^x + e^{-x}}{e^x - e^{-x}} \tag{19}$$

The apparently infinite term, $2/\rho$ (for $\rho \simeq 0$) serves to remove the corresponding infinity of the hyperbolic cotangent. In fact, as ρ varies from $-\infty$ to $+\infty$, $v(\rho)$ increases monotonically from 0 to 1 (see Figure 1).

This reformulation of the maximum entropy procedure has several very important features. All of them relate to the question of how the MEP can be implemented in real time, so that it may be built into a retrieval system and, ultimately, tested.

The key features are:

- Reduction of the dimensionality of the problem. For N terms the number of variables is reduced from 2^N atomic components to $\simeq N$ constraint equations.
- Freedom from unjustified or superfluous assumptions. The formalism presented here is a minimal constraint version of the MEP, which does not require any assumed constraints (such as $\rho(R)$).
- Clear analytic structure. The function v is an entire function in the complex ρ-plane and homotopic solution methods are expected to work well.
- Conceptual insight into the nature of 'richness', the relation between terms and their combinations, and the overall mathematical structure of the maximum entropy procedure.

The ultimate test of the MEP must be in the field, using large, real databases and real inquirers (or 'end-users' as they are unfortunately called).

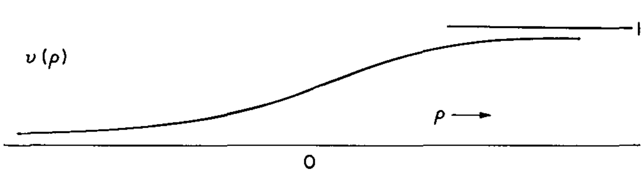

Figure 1. The constraint function $v(\rho)$

A particularly promising approach to real time implementation has been outlined[6], which involves homotopic optimization[10]. This procedure uses the analyticity of the problem to trace the solution from a trivial simplification, back to the solution of the full problem.

SPECULATION

While we await the development of the real time implementation, and the experimental testing of the MEP, it may be appropriate to speculate on the validity of the MEP. A distinction is made between aesthetic and conceptual speculation.

Aesthetic

It has been known for some time that the entropy function is the unique continuous function satisfying certain very reasonable constraints — constraints which should apply to a measure of uncertainty[11]. Thus we are fairly certain that the MEP captures the distribution, subject to given constraints, having the most residual uncertainty. This seems to mean that we are 'assuming as little as possible' and that is, in turn, aesthetically attractive.

Furthermore, from the viewpoint of mathematical aesthetics, the picture we have reached is compelling in its clarity and simplicity. However, that simplicity is due to the exact correspondence between the present problem and the elegant machinery of statistical mechanics — machinery that has been developed by many investigators over scores of years. Thus simplicity and elegance do not necessarily mean that *we* are on the right track.

Conceptual

There is a potential conceptual objection. When statistical mechanics is applied to, for example, the physics of gases, we are all somewhat prepared to accept the notion that the molecules rush about as if at random. There are difficulties — as in the paradox of the H-theorem. We know that the motion of the molecules is not random, but is governed by the physical laws of motion. This has led to the development of a rich field of mathematical analysis — originally ergodic theory — later the study of the mixing properties of dynamical systems.

However, no one has ever interviewed a molecule, nor do we ascribe to molecules volition, intelligence and other entropy-decreasing virtues. For information storage and retrieval, the situation is quite different. The items, whether culled from this journal or the *National Enquirer*, represent the products of volition, intelligence and other entropy-decreasing virtues. Furthermore, their organization (that is, the selection of descriptors, and the assignment of items to the rows of the table) represents the further application of intelligence (arguably, in some cases, a higher order of intelligence).

In the face of all this, it seems almost perverse to argue that the key feature of what remains is its randomness. This is a disturbing objection, and not clearly wrong. For example, is it not true that the ideal descriptors would exactly correspond to the questions asked — so that every row would have $\rho = \pm\infty$ and $\nu = 1$ or 0? In the face of this argument we can point to three factors. First, as in mathematical sociology, it does appear that the voluntary acts of individual human intelligences can exhibit stochastic regularity: the classic example is the near constancy of national suicide rates. The suicides of 1983 are surely different persons from the suicides of 1982. Why is their total number nearly the same? Second, and more specific to the case of information retrieval, the descriptors are relatively fixed (one of the key virtues of retrieval by relatedness is to weaken this restriction), while queries are constantly shifting. Thus, although descriptors may have been originally defined in ways that decrease the randomness (by imposing correlations among descriptors) evolution in meaning of terms and areas of interest may quickly increase the randomness. (Thus, the term 'electronic' once implied, for a vast literature, the term 'vacuum tube', while the term 'solid state' occurred only in the technical literature of physics.)

A third reason is the remarkable result obtained by VanCampenhout and Cover[12]. They show that if the average of many i.i.d. random variables is constrained, the conditional distribution of any one of them is given asymptotically by the product of the prior distribution and the maximum entropy form. This suggests that if value does 'behave randomly' within each row then the maximum entropy principle is not only the least constraining assumption; it is asymptotically rigorous (as the rows become large).

At present, it is the author's opinion that the applicability of the MEP to retrieval from large data reservoirs will soon be subject to experimental test even in complex situations. The question of whether it 'ought to be right' may therefore be profitably postponed.

IMPLICATIONS AND OPEN QUESTIONS

In the companion paper[6], certain mathematical problems have been specified for solution; there are also questions that go beyond direct calculation. Two particularly interesting areas are: the validity of term weighting and the 'infodynamics' of information reservoirs.

Term weighting

The concept of term weighting may be abstracted as follows: given a query Q, is there a term weighting function $W(Q)$ that assigns to every descriptor A a real number $W(Q, A)$ such that:

- $W(Q, AB) = W(Q, A) + W(Q, B)$
  $W(Q, ABC) = W(Q, AB) + W(Q, C)$
  etc.

and

- $W(Q, \alpha) \geq W(Q, \beta)$ if and only if the average value (in response to Q) of row α is at least as high as the average value of row β (precision measure).
- $W(Q, \cdot)$ is some help in achieving the (occasionally desirable) goal of recall — complete extraction of the value from the reservoir.

Clearly, the latter two conditions together define a multiple criterion problem, and the rigorous elaboration of the last condition involves a careful analysis of the interplay between the information retrieved and the effort required. Rows should be studied in order of decreasing richness to obtain as much value as possible, but, at some point, the inquirer will probably abandon the goal of 'total recall'. The problem of term weighting has been discussed extensively by Sparck Jones and Robertson[13].

If we consider only the first two conditions, we have a well defined mathematical question: if the maximum entropy principle is correct, does a weighting function exist? We suspect that the answer is, in general, negative.

The detailed relation between maximum entropy and the current approaches to term weighting will be explored elsewhere. The relation is more nearly one of conflict than of harmony, as several examples will show. To make contact with the usual discussions, the value v must be restricted to 0 or 1. Also, the rows are labelled by vectors called document vectors $(x_1, x_2, \ldots)$. In a particular row, $x_j = 0$ if the jth term does not apply to documents on the corresponding row. If it does apply, x_j may be either a Boolean indicator $(x_j = 1)$ or some sort of 'count indicator' (x_j = number of times term j appears in each document of the corresponding row). The latter case is a discrete version of 'fuzzy descriptors'. Let us concentrate on the Boolean or binary version. If A is term 1, etc. the table now becomes

Com-ponent	Document vector	n	$q(\alpha, v=0)$	$q(\alpha, v=1)$
$\overline{ABC}$	000	$n(000)$	x	x'
$A\overline{BC}$	100	$n(100)$	a	a'
$\bar{A}B\bar{C}$	010	$n(010)$	b	b'
...	001	$n(001)$	c	c'
	110	$n(110)$	ab/P	$a'b'/Q$
	101	$n(101)$	ac/P	$a'c'/Q$
	011	$n(011)$	bc/P	$b'c'/Q$
	111	$n(111)$	abc/P^2	$a'b'c'/Q^2$
			P	Q

The table has been filled in using the pattern resulting from the 'binary independence hypothesis'. At once, it is clear that this hypothesis requires that six parameters $a, \ldots, C'$ explain seven numerical properties of the reservoir — that is

$$n(011) = (bc/P) + (b'c'/Q)$$

etc. The problem is not apparent when terms are considered only two at a time, since the table

0	0	$n(00)$	x	y
1	0	$n(10)$	a	a'
0	1	$n(01)$	b	b'
1	1	$n(11)$	ab/P	$a'b'/a$

has four parameters with which to describe three variables. If tempted to argue that one overconstraint is something 'we could live with', one should recall that the degree of overconstraint is

$$\text{overconstraint} = 2^n - 1 - 2n$$

Thus for four terms, the problem is $16 - 1 - 8 = 7$ times overconstrained. (The fact of overconstraint has been noted by van Rijsbergen — in unpublished remarks at the ACM-SIGIR conference, 6–8 June 1983.) It certainly means that the binary independence hypothesis has drastic implications for the structure of large reservoirs. One might, of course, turn the argument around and assert that there are certain 'good terms' for which this structure does prevail.

The more general question 'do term weights exist?' is now considered. A term weight is a vector such as (w_1, w_2, w_3) such that

$$w \cdot x(\alpha) > w \cdot x(\beta) \text{ if } \rho(\alpha) > \rho(\beta)$$

Our constraint tables provide a good tool for exploring this question. We note that, for binary value (0 or 1) the function v becomes

$$v(\rho) = e^{\rho}/(1 + e^{\rho}) = 0.5(1 + \tanh \rho/2)$$

but this explicit form is not needed for what follows.

Let us consider several possible constraint tables

Comp	Constraints A	B	C
000			
100	x		
010		x	
001			x
110	x	x	
101	x		x
011		x	x
111	x	x	x

Table *(ABC)*

Comp	Constraints A	B	C	R
000				x
100	x			x
010		x		x
001			x	x
110	x	x		x
101	x		x	x
011		x	x	x
111	x	x	x	x

Table *(ABCR)*

Comp	Constraints A	B	R*
000			
100	x		x
010		x	x
001			x
110	x	x	x
101	x		x
011		x	x
111	x	x	x

Table *(ABR*)*

Comp	Constraints *A* or *B*	*C*
000		
100	x	
010	x	
001		x
110	x	
101	x	x
011	x	x
111	x	x

Table*(A* or *B, C)*

Table (ABC) supports term weighting. For example

$$\rho(110) = \lambda_A + \lambda_B = \rho(100) + \rho(010)$$

and it is easy to see that $w = (\lambda_A, \lambda_B, \lambda_C)$. However, this is a rather special case. Table $(ABCR)$ does not have this property, as

$$\rho(110) = \rho(100) + \rho(010) - \lambda_R$$

Table (ABR^*) has the same problem, and so does the innocent looking $(A \text{ or } B,\ C)$ for which

$$\rho(111) = \lambda_{A \text{ or } B} + \lambda_C = \rho(100) + \rho(011) - \lambda_{A \text{ or } B}$$

What we have is a picture in which the richness parameter serves as a term weight in one case, but fails in three. This does not, of course, prove that term weights do not exist, but it causes grave doubts about the compatibility of term weights and the maximum entropy principle.

Consider now the conditional probabilities $Pr\{(x_1, x_2, x_3)/v = 0\}$ and $Pr\{(x_1, x_2, x_3)/v = 1\}$. The ratio

$$Z(x_1, x_2, x_3) = \frac{Pr\{(x_1, x_2, x_3)/v = 1\}}{Pr\{(x_1, x_2, x_3)/v = 0\}}$$

can serve to rank the rows of the table[14]. Specifically

$$Z(x_1, x_2, x_3) = e^{\rho(x_1, x_2, x_3)} \cdot \frac{Pr\{v = 0\}}{Pr\{v = 1\}} = \frac{P}{Q} e^{\rho}$$

The ratio $\frac{P}{Q}$ has nothing to do with the document vector (x_1, x_2, x_3) and so ranking by Z is the same as ranking by ρ or by e^{ρ}. However, binary independence suggests using $\log Z$ as a weight. That is

$$w = (\log Z(100), \log Z(010), \log Z(001))$$

Writing $e^y = P/Q$ we see that

$$\log Z = \rho + y$$

Hence, even in the most favourable case, Table (ABC), the values of $\log Z$ are not additive. In other words, if the maximum entropy principle is correct, the binary independence principle is not, as long as constraints of the form in Table ABC are allowed.

Effort and Information

The concept of 'infodynamics' is still rather vague. It is used here in a very technical sense, akin to the discipline of thermodynamics. In thermodynamics, one of the fundamental relations (the First Law) may be expressed as

$$dU = TdS - pdV \tag{20}$$

where dU represents a small change in the internal energy of a thermodynamic system; T represents the temperature; dS represents a small increase in the entropy of the system; pdV represents the pressure of the system multiplied by a small change in the volume which it occupies. This product is the work done by the system, on its environment.

This law states that during a reversible change (one for which the equation of state applies throughout), the heat added to the system (TdS) produces an increase in internal energy (dU) or external work (ρdV).

In the present analysis, there is an analogy between the value and the internal energy U. In fact, it has been found (Ref 6, Appendix II) that the derivatives $S'(\rho)$ and $v'(\rho)$ are related by

$$S'(\rho)/v'(\rho) = -\rho = 1/T \tag{21}$$

This follows from the first law, if there is no external work. In other words, this discussion relates the average value of a component to the disorder of the component, but sheds no light on the effort required to change either of them. At present, the literature on the rigorous study of the relation between information and mechanics (the study of mechanical energv) is very sparse, consisting of a single monograph[15] which is more fundamental in outlook. It is the author's belief that some progress can be made towards 'infodynamics' by combining the maximum entropy analysis and detailed study of the effort required to maintain or examine the several components of a reservoir. The link, if one can be found, will be in the relations between the cost functions for a component:

- $M(\alpha; N)$ = cost of maintaining component α at size N,
- $C(n; \alpha)$ = cost of examining n elements of component α

and the 'success functions':

- $N(k; n; \alpha; v^*)$ = Probability that the examination of n elements from α will produce exactly k whose value exceeds v^*,
- $S(n; \alpha; V^*)$ = Probability that the sum of the values of elements drawn from α will first exceed V^* at the nth drawing.

In the companion paper[6], the success functions have been expressed in terms of the richness, or effective temperature, of the components. The problem of 'infodynamics' is to expand the analysis to include M and C in a rigorous fashion.

ACKNOWLEDGEMENTS

It is a pleasure to acknowledge several conversations with Professor William Cooper and Dr Frederick Kantor. This work was supported by NSF Grant No IST–8110510.

REFERENCES

1 **Cooper, W S** 'Exploiting the maximum entropy principle to increase retrieval effectiveness' *J. Am. Soc. Inf. Sci.* Vol 34 No 1 (1983) pp 31–39

2 **Cooper, W S and Huizinga, P** 'The maximum entropy principle and its application to the design of probabilistic retrieval systems' *Inf. Technol: Res. & Dev.* Vol 1 No 2 (1982) pp 99–112
3 **Kantor, P B** 'Evaluation of the feedback in information storage and retrieval systems' in **M E Williams (Ed.)** *Annu. Rev. Inf. Sci. & Technol.* Vol 17 (1982) pp 99–120
4 **Kittel, C** *Elementary statistical physics* Wiley & Sons, USA (1958)
5 **Good, I J** 'Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables' *Ann. Math. Stat.* Vol 34 (1963) pp 911–932
6 **Kantor, P B** 'Maximum entropy, continuous value, effective temperature and homotopy theory in information retrieval' to be submitted to *IEEE Trans. Inf.*
7 **Cooper, W S and Maron, M E** 'Foundations of probabilistic and utility-theoretic indexing' *J. Assoc. Comput. Mach.* Vol 25, No 1 (1978) pp 67–80
8 **Robertson, S E, Maron, M E and Cooper, W S** 'Probability of relevance: a unification of two competing models of document retrieval' *Inf. Technol.: Res. & Dev.* Vol 1 No 1 (1982) pp 1–21
9 **Salton, G** *Automatic information organization and retrieval* McGraw-Hill, USA (1968)
10 **Zangwill, W I and Garcia, C B** *Pathways to solutions and equilibria* Prentice-Hall, USA (1981)
11 **Khinchin, A I** *Foundations of statistical mechanics* Dover, USA (1967)
12 **VanCampenhout, J M and Cover, T M** 'Maximum entropy and conditional probability' *IEEE Inf. Theory* Vol 27 No 4 (1981) pp 483–489
13 **Robertson, S E and Sparck Jones, K** 'Relevance weighting of search items' *J. Am. Soc. Inf. Sci.* Vol 27 (1967) pp 129–146
14 **Chow, D and Yu, C T** 'Constructing feedback queries' *J. Assoc. Comput. Mach.* Vol 29 No 1 (1982) pp 127–151
15 **Kantor, F W** *Information mechanics* Wiley & Sons, USA (1976)

APPENDIX

In the companion paper[6], the following mathematical results are established. The variable β is defined as

$$\beta = -\rho$$

The entropy of a row with richness β is given by

$$S(\beta) = 1 + \ln 2 + \ln(2\beta^{-1} \sinh \beta/2) - 2\beta^{-1} \coth(\beta/2)$$

The expected or average value of v in a row is given by

$$v(\beta) = 0.5(1 - (\coth \beta/2 - 2/\beta))$$

The β-values of rows are determined by maximizing:

$$\sum_{\alpha} n(\alpha)S(\beta(\alpha)) + \sum_{P} \lambda_P(n(\alpha)v(\beta(\alpha)) - V_P)$$

Differentiating with respect to any particular $\beta(\alpha)$ yields

$$n(\alpha)S'(\beta) + \sum_{P} \lambda_P n(\alpha)v'(\beta) = 0$$

where the sum extends over all conditions which constrain row α. It is further shown that

$$S'(\beta)/v'(\beta) = \beta = -\rho$$

Thus, for each row α

$$\rho(\alpha) = \sum_{P\varepsilon P(\alpha)} \lambda_P$$

and

$$V_P = \sum_{\alpha\varepsilon A(P)} n(\alpha)v\left(\sum_{Q\varepsilon P(\alpha)} \lambda_Q\right)$$

where $A(P)$ is the set of rows constrained by the Pth condition and $P(\alpha)$ is the set of conditions constraining the αth row.