A LOGIC OF INFORMATION SYSTEMS

Norman Dalkey

September 1987
CSD-870057

Proceedings, 2nd AAAI Workshop on Uncertainty in Artificial Intelligence, Philadelphia, PA, August 1986, pp. 63-70.

A LOGIC OF INFORMATION SYSTEMS

N.C. Dalkey
UCLA Cognitive Systems Laboratory
Los Angeles, CA. 90024-1596

Abstract. A logic can be formulated with information systems as elements. The calculus of this logic is similar to, but not identical with, Boolean algebra. The logic is inductive--conclusions have more information than premises. Inferences have a strong justification; they are valid for all proper scoring rules.

DOMINANCE.

Information systems (*IS*) are well-known constructs in the knowledge sciences. Examples are: experiments, communication coding schemes, signal systems, pattern recognition techniques, surveillance systems, medical diagnosis, many expert systems, etc. Despite the wide variety of applications *IS* have a common underlying structure:

1. A set of events E (hypotheses, events of interest, target events, states of the world,...)

2. A set of events I (observations, data, signals, messages,...)

3. A joint probability distribution $P(E.I)$ on hypotheses and observations (the period in $P(E.I)$ denotes the logical conjunction "and".)

IS have a significant property from the standpoint of creating a logic, they allow *dominance* --one information system can have a higher expected value than another *for all payoff functions*. This property contrasts sharply with probability distributions. If P and Q are any two non-identical probability distributions, then there is a payoff function (decision matrix) that engenders a higher expected value for P, and another payoff function that engenders a higher expected value for Q.

Representation of expected value is simplified by the notion of *proper scoring rule*. Let P be a probability distribution on the partition of events $E = (e_1, ..., e_n)$ and let $S(P, e)$ be a function which assigns a score (rating, reward, payoff) to P given that the event e occurs. S is called *proper* (admissible, reproducing, honesty promoting) if it fulfills the condition

$$\sum_E P(e)S(P,e) \geq \sum_E P(e)S(Q,e) \tag{1}$$

That is, a score rule S is proper if the expectation is a maximum when the score is determined by the same distribution as that determining the expectation.

There is an infinite family of functions that fulfill (1). Among them is the logarithmic score $S(P,e) = \log P(e)$, and the set of decisional scores. For the latter, let $U(a,e)$ be the payoff if action a is taken and the event e occurs, and let $a^*(P)$ be the optimal action if P is the probability distribution on E. $S(P,e) = U(a^*(P), e)$ is a proper score. The expectation of the logarithmic score is the negative of the Shannon entropy of P (often called the information in P) and links the theory of proper scores to information theory. Decisional scores tie the theory of proper scores to decision theory.

Abbreviate $\sum_E P(e)S(P,e)$ by $G(P)$ and $\sum_E P(e)S(Q,e)$ by $G(P,Q)$. The expected score of an IS is given by

$$H(P) = \sum_I P(i)G(P(E \mid i))$$

where $P(i) = \sum_E P(e.i)$ is the initial probability of the observation i. $H(P)$ is thus the average over the potential observations of the expected score of the posteriors. The expected relative score is defined analogously,

$$H(P,Q) = \sum_I P(i)G(P(E \mid i), Q(E \mid i))$$

It is readily verified that H fulfills the analogue of (1), that is

$$H(P) \geq H(P,Q) \tag{2}$$

It is also straightforward to demonstrate that $H(P)$ is convex, and $H(P,Q)$ is linear in P. (Dalkey 1987).

An IS P is said to dominate an IS Q, in symbols $P \geq Q$, if $H(P) \geq H(Q)$ for all proper score rules S. It is clear that $\geq$ is a partial order, i.e., it fulfills:

1. Transitivity. $P \geq Q$ and $Q \geq R \rightarrow P \geq R$
2. Reflexivity: $P \geq P$.

3. Antisymmetry: $P \geq Q$ and $Q \geq P \rightarrow P = Q$

In addition, for a given set of events E and a given prior distribution (E), $\geq$ has an absolute upper bound P^* where for each e, $P^*(e \mid i) = 1$ for some i, and 0 otherwise. P^* is often called "perfect information" in decision theory, or more jocularly, "the clairvoyant." $\geq$ also has an absolute lower bound P^0, where $P^0(e.i) = P(e)P(i)$. P^0, in effect, consists in implementing the prior distribution. It is readily verified that

$$P^* \geq P \geq P^0 \tag{3}$$

The second inequality $P \geq P^0$ is often called the positive value of information principle (PVI), any *IS* is at least as valuable as the prior *IS* (assuming that information is free.) (3) is a well-known illustration of the fact that $\geq$ is not an empty relation. (LaValle 1978).

LATTICE STRUCTURE

To proceed further in using the dominance relation as the basis for a logic, it is pertinent to examine the lattice properties of the relation. Lattices have received extensive attention as foundations for logics. (Birkhoff 1940).

A partial order such as $\geq$ is called a lattice if for each pair of elements P, Q there is least upper bound (l.u.b) w.r.t. $\geq$ and a greatest lower bound. Examples can be found of pairs of *IS* that do not have a l.u.b., and thus $\geq$ is not in general a lattice. However, for an important subclass of *IS*, namely, those with binary hypotheses, $\geq$ is a lattice.

Theorem 1. For the set of *IS* with binary hypotheses, $\geq$ is a lattice.

Proof: Theorems 8 and 8′ in (Dalkey 1980).

IS with binary hypotheses are those which address a yes-no question: Does the patient have AIDS? Is there life on Mars? Will a Republican be elected president of the U.S. in 1988? Is the crystal structure of the substance octahedral? In practice, binary hypotheses are part of the stock-in-trade of the analyst. In addition, although the typical *IS* is not binary, decisional problems often "boil down to" a binary question.

Denote the l.u.b. of P and Q by $P+Q$, and the g.l.b. by $P \cdot Q$. $P \cdot Q$ expresses the information that P and Q have in common. $P+Q$ expresses the "sum" of the information in the two.

THE CALCULUS

Given the operations + and $\cdot$, a calculus can be formulated. Listed below are a set of postulates for the calculus. They are listed in parallel, one set for + and the analogous set for $\cdot$. Although they are listed as postulates, they can be verified by the construction methods described below. The calculus differs from Boolean Algebra in that it does not have complements (negation) and is not distributive. The lack of a negation is partially compensated for by the duality rule described below.

P1.	$P+P \geq P$	$P \cdot P \leq P$
P2.	$P+Q \geq Q+P$	$P \cdot Q \leq Q \cdot P$
P3.	$P+Q \geq P$	$P \cdot Q \leq P$
P4.	$P \geq Q$ and $P \geq R \rightarrow P \geq Q+R$	$P \leq Q$ and $P \leq R \rightarrow P \leq Q \cdot R$
P5.	$P+Q = P+R \rightarrow P+Q = P+Q \cdot R$	$P \cdot Q = P \cdot R \rightarrow P \cdot Q = P \cdot (Q+R)$

It is noteworthy that the first four postulates for + are homologous to the basic four postulates for the propositional calculus; however, they are not as powerful because of the lack of a negation. The first four are true of any lattice. P5 expresses a property that does not hold for lattices in general and distinguishes *IS* logic. I do not have a proof that P1-5 are complete. The existence of a model--the canonical representation described below--shows that the postulates are consistent.

The basic inference rules for the calculus are the transitivity of $\geq$ and the rule of replacement--in any statement, if $P = Q$, then P can be replaced by Q in any position. In addition, the usual rules of substitution for variables and the inference rules for two-valued logic (e.g., modus ponens for $\rightarrow$) are assumed. A derived rule, the duality principle, is particularly useful. It states that any postulate or theorem remains true if + is replaced by $\cdot$ throughout, and $\geq$ is replaced by $\leq$. Note that with the duality principle, only the + versions of P1-5 are needed. The $\cdot$ versions can be derived immediately with the duality principle.

From P1-5 a variety of theorems can be generated. Among the more familiar:

T1. Idempotence: $P+P = P$, $P \cdot P = P$.

T2. Associativity: $(P+Q)+R = P+(Q+R)$, $P \cdot (Q \cdot R) = (P \cdot Q) \cdot R$.

T3. Consistency: $P \geq Q, P+Q = P, P \cdot Q = Q$ are mutually equivalent.

T5. Absorption: $P = P+P \cdot Q = P \cdot (P+Q)$.

T6. Semi-distributivity: $P \cdot (Q+R) \geq P \cdot Q + P \cdot R, (P+Q) \cdot (P+R) \geq P+Q \cdot R$.

The next three theorems are provided with proofs as an example of using the calculus. Let $P \mid Q$ mean that P does not dominate Q and Q does not dominate P.

T7. $P \mid Q \rightarrow P+Q > P$ and $P+Q > Q$, i.e., $P \mid Q$ implies that $P+Q$ strictly dominates both P and Q.

Proof: Suppose $P \geq P+Q$. From P3 , $P+Q \geq Q$, and thus by transitivity of $\geq$, $P \geq Q$ contrary to the hypothesis.

T8. $P = Q \rightarrow P+R = Q+R$.

Proof: From $P1$, $P+R = P+R$. Whence, by replacement, $P+R = Q+R$.

T9. $P+Q = P+R \rightarrow P+Q = P+Q+R$.

Proof: From the hypothesis and $T8$, $P+Q+Q = P+R+Q$. From $T1$, $Q+Q = Q$, and thus by replacement, $P+Q+Q = P+Q = P+Q+R$.

T9 is of special interest with regard to the design of information systems. It states that even though P, Q, and R are mutually non- dominating--i.e., each contains information neither of the other two contain--if $P+Q = P+R$, then either Q or R is eliminable. This contrasts with T7, which states that if $P \mid Q$, then the sum is strictly more informative than either alone.

T10. $P+Q = P+R \rightarrow P+Q+R = P+Q \cdot R$.
Proof: Immediate from P5 and T9.

T10 states that if the sum of P and Q is the same as the sum of P and R, then the sum of all three is just the sum of P and the common part of Q and R.

INFERENCE

One mode of application of *IS* logic stems from elaborating the set of theorems derivable in the calculus. This body of results appears promising in the design of information systems, e.g., design of experiments. In a sense, this mode is deductive, determining the consequences of the postulates.

A somewhat different mode stems from applying the logic to the problem of combining evidence. This mode is inductive. The basic inference rule in this mode is: If P and Q are known, but the dependencies (correlation) of P and Q are not known, assume $P+Q$. The justification for this rule requires some preliminaries.

Let the observation set for P be I and the observation set for Q be J; i.e., P is the joint distribution $P(E.I)$ and Q is the joint distribution $Q(E.J)$ Let IJ denote the cartesian product of I and J. The composition of P and Q, denoted by $P.Q$ is a joint distribution $R(E.I.J)$. Knowing P and Q is not sufficient to determine R. All that is known is that R must be compatible with both P and Q, i.e.:

$$P(E.I) = \sum_J R(E.I.J) \tag{4}$$

$$Q(E.J) = \sum_I R(E.I.J)$$

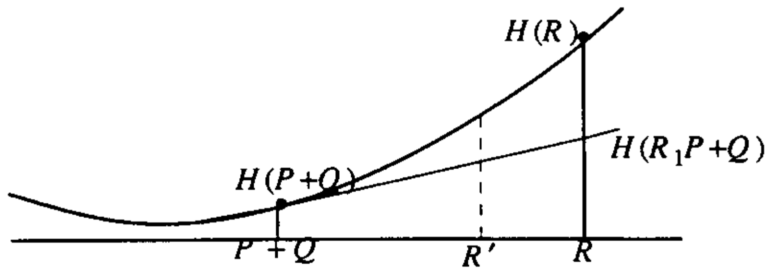
Let $K(P,Q)$ be the set of R which fulfills (4). K is the set of compositions of P and Q. It is an immediate consequence of PVI that for any R in K, $R \geq P$ and $R \geq Q$. Since (4) is a set of linear constraints, K is convex and closed. K does not contain all *IS* which dominate both P and Q; however, for any R which dominates P and Q, there is an R' in K such that $R \geq R'$. (Dalkey 1987).

$P+Q$ is in K, since $P+Q$ dominates both P and Q, and if $P+Q$ were not in K, there is an R' in K, $P+Q \geq R'$. If $P+Q \neq R'$, $P+Q$ would not be the l.u.b. of P and Q. Thus, $P+Q$ is the g.l.b. of K.

The justification for assuming $P+Q$ when P and Q are known is based on the following theorem:

Theorem 2. If R is in $K(P,Q)$ then $H(R, P+Q) \geq H(P+Q) \geq \max [H(P), H(Q)]$ for every proper score.

Proof. Since K is convex, $R' = aR + (1-a)(P+Q), 0 \leq a \leq 1$, is in K, and since $P+Q$ is the g.l.b. of K, $H(R') \geq H(P+Q)$; thus, $H(R')$ is monotonically decreasing (with decreasing a) between R and $P+Q$. $H(R', P+Q)$ is the line tangent to $H(R')$ at $P+Q$. Thus, since $H(R')$ is convex, $H(R, P+Q) \geq H(P+Q)$ (cf. Figure 1).



$H(P+Q) \geq \max [H(P), H(Q)]$ from P3.

We can assume that the actual composition R of P and Q (i.e., the R that would be observed in a sufficient set of observations of E.I.J) is in K. The theorem states that whatever R may be, the actual relative expectation of $P+Q$ $H(R, P+Q)$ is greater than the apparent expectation of either P or Q. In other words, the expectation of $P+Q$ is guaranteed and guaranteed to be greater than that of P or Q, no matter what the payoff function of the user. It is this guarantee which justifies the use of the term *logic*.

Note that the inference from P and Q to $P+Q$ is *inductive*. We cannot derive $P+Q$ from P and Q by deductive reasoning. We could derive the actual composition R from P and Q if we knew the dependencies between P and Q, e.g., if we knew they were independent. Without knowing the dependencies, however, we can recommend accepting $P+Q$ on the basis of the strong guarantee.

In the interesting case where P and Q are mutually non-dominating (if P dominates Q for example, then from consistency, T4, $P+Q = P$) T7 assures that $P+Q$ strictly dominates P and Q.

COMPUTATION

For the case of *IS* with binary hypotheses, the computation of $P+Q$ and $P \cdot Q$ is particularly simple. Let $t(i)$ denote the vector $(P(i \mid e), P(i \mid \bar{e}))$, the supra-bar indicating negation or "non-e." Order these vectors in the decreasing order of the ratio $P(i \mid e)/P(i \mid \bar{e})$, and reindex the observations numerically in the new order. Define $T(i) = \sum_{j \leq i} T(j)$. $T(0) = (0, 0)$. Since $\sum_{I} P(i \mid e) = 1$, if there are m observations in I, $T(m) = (1, 1)$.

The vectors $T(i)$ can be plotted in the plane, and joining them with straight lines generates a concave, piece-wise linear curve in the unit square lying above the diagonal, as in Figure 2. The convex closure $C(P)$ of this curve--i.e., the points between the curve and the diagonal and including the curve and the diagonal--can be called a canonical representation of the *IS* P. It can be shown that $P \geq Q$ if and only if $C(Q) \subset C(P)$.

(Dalkey 1980).

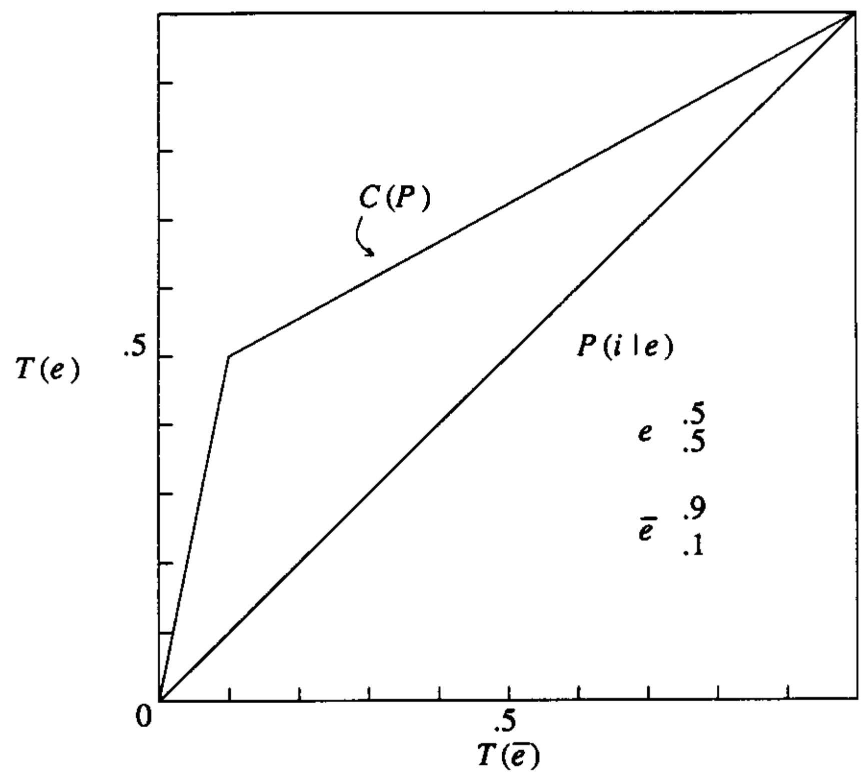


Fig. 2

If two *IS* P and Q are plotted, then $C(P+Q) = \langle C(P), C(Q)\rangle$ the convex closure of $C(P)$ and $C(Q)$. $P \cdot Q = C(P)$. $C(Q)$, the intersection of the two representations. This construction is illustrated in Figure 3. For small *IS*--those with a relatively few observations-- the construction is readily made by hand. For larger *IS*, the construction is easily programmable for a computer.

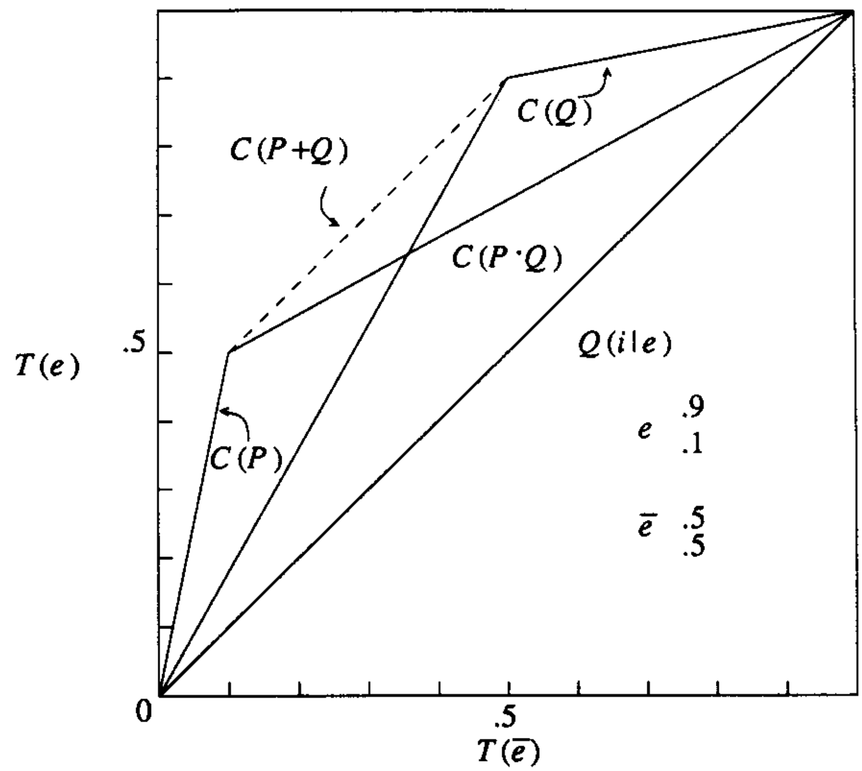


Fig. 3

EXAMPLE

Suppose you are worried about AIDS. You know there are two tests, each developed by a different organization. You know that each has been studied by its developer, and relatively good statistics exist concerning its diagnosticity. However, statistics are insufficient concerning the joint diagnosticity of the two tests taken together. You would prefer the (potentially) greater information of both tests, but you have no way of interpreting the results if the two tests give conflicting results. Let A mean "The patient has AIDS," and + mean "The test result is positive." Suppose the likelihoods of test results are: (Any similarity between the numbers in the table and those for actual tests is a miracle. I made the numbers up.)

center box tab(;); c c | c | c c | c | c | c. ;;T_1;T_2 _ A ;+;.5;.9 ; ;-;.5;.1 _ $\bar{A}$;+;.1;.5 ; ;-;.9;.5

The diagnosticity of T_1, if the patient has AIDS, is low, but it is high if the patient doesn't have AIDS. These characteristics are reversed for T_2. For the minimal composition T_1+T_2 of the two tests, there are four possible outcomes: both positive, both negative, or the two mixed cases.

center box tab(;); c c c | c c | c c | n. ;T_1;T_2;$P(\cdot \mid T_1 . T_2)$ _ A;+;+;.5 ;+;-;.0 ;-;+;.4 ;-;-;.1 _ $\bar{A}$;+;+;.1 ;+;-;.0 ;-;+;.4 ;-;0;.5

The O's at + – are a mathematical artifact. In practice, the likelihoods for these cases would be determined by a separate computation.

The table gives the likelihoods of test results given the disease state. To determine the posterior probabilities of the disease states given test results, it is necessary to know the prior probabilities of the disease states for a relevant population. The table below lists the posteriors for prior probabilities of AIDS of .5 and .2.

Posterior Probability

center box tab(;); c | c |c | c. T_1;T_2;$P(A)$=.5;$P(A)$=.2 +;+;.83;.56 +;-;?;? -;+;.5;.2 -;-;.17;.05

Note that the double negative is a good deal more reassuring if the prior probability is low. The values for this example are those used in Figures 2 and 3. The joint likelihoods were read from the graph in Figure 3.

IS WITH NON-BINARY HYPOTHESES

Although *IS* with binary hypotheses are of practical importance, most *IS* arising in practice have multiple hypotheses. (I recently read of a computerized medical diagnostic service with a list of over 1900 disease states and over 4000 symptoms and test results. It was called "Hypochondriac Heaven.") Examples can be devised with as few as three hypotheses where a pair of *IS* do not have a l.u.b. w.r.t. $\geq$. Thus $\geq$ is not a lattice for non-binary *IS*. On the other hand examples are easily devised for pairs of non-binary *IS* which do have a l.u.b. It should be clear from the preceding analysis that for pairs of *IS* with non-binary hypotheses, if $P+Q$ exists, then the inference from P and Q to $P+Q$ is just as solid as it is for binary *IS*.

At present there is no algorithm for determining whether a pair of non-binary hypothesis *IS* have a l.u.b. This is clearly an area inviting research.

There is a non-trivial fall-back possibility. If a specific payoff function--e.g., the logarithmic score--is considered "adequate," then the H function imposes a complete order on *IS*, and $P+Q$ and $P \cdot Q$ w.r.t. that score always exist. For this case, $P+Q = \underset{K(P,Q)}{arg\ \min} H(R)$. Define the set $L(P,Q) = \{R \mid P \geq R, Q \geq R\}$. $P \cdot Q = \underset{L(P,Q)}{arg\ \max} H(R)$. The guaranteed expectation for $H(P+Q)$ holds for the given score rule. This formalism was the basis for a weaker inductive logic proposed earlier.

(Dalkey 1985).

REFERENCES

Birkhoff, G. (1940). Lattice Theory. American Mathematical Society (Colloquium Publications, *25*), New York.

Dalkey, N.C. (1980). The Aggregation of Probability Estimates. UCLA-ENG-CSL-8025.

Dalkey, N.C. (1985). Inductive Logic and the Maximum Entropy Principle. *In* Maximum Entropy and Bayesian Methods in Inverse Problems, eds. C. Ray Smith & W.T. Grundy. Boston: D. Reidel Publishing Co.

Dalkey, N.C. (1987). Information Systems. UCLA-ENG-CSL Report, in preparation.

LaValle, . (1978). Fundamentals of Decision Analysis. New York: Rhinehart & Winston.