Reflections on Information and Information Value

Thomas B. Sheridan

***Abstract*—*Information*, as explicated by C. E. Shannon (1949), refers to the reduction in uncertainty about the state of an event after a message has been sent relative to the uncertainty about the state of the event before the message was sent. *Information value*, as explicated by R. A. Howard (1966) and others, refers to the difference in what one can gain by action taken knowing the state of the event relative to what one can gain by action taken without such knowledge. This paper considers the relation between the two concepts, information and information value, to the authors knowledge not heretofore discussed. It is concluded that the two concepts are independent, characterizing quite different aspects of information seeking and using. In application, the cost of the first term (information) subtracts from the second term (information value), and/or diminishes the latter in relation to information access time.**

I. Introduction

The well established notions of information and information value are essentially as follows:

Information: Assume that an event x can occur in one of many mutually exclusive and collectively exhaustive states i, and $p(x_i)$ is the probability that the event is in state i. The accepted Shannon definition of information H in a message specifying which event has occurred is

$$H = \Sigma_i p(x_i) \log_2[1/p(x_i)] = -\Sigma_i p(x_i) \log_2 p(x_i).$$

The Shannon measure indicates the average uncertainty about the state i (or the entropy) before the message i was sent. Thus the message is defined by its uncertainty reducing capability. When all event states have equal probability, $-\log_2[p(x_i)] = \log_2 N$. This term corresponds to the number of steps in a split-half procedure for narrowing down from the initial N equally probable states to the final certainty of which state is the true one. The H defined above for (in general) unequally probable states can also be called H_{avg} because it is a weighted average over log transformations on component possibilities.

Q(a,s)

Information value: Let $V(u_j \mid x_i)$ be the gain or reward for taking action u_j when an event x is in state i. If x_i is known exactly, then a rational decision-maker adjusts u_j (selects j) to maximize V for each occurrence of x_i, in each instance yielding $\max_j[V(u_j \mid x_i)]$. In this case the average reward over a set of x_i is

$$V_{\text{avg}} = \Sigma_i p(x_i)\{\max_j[V(u_j \mid x_i)]\}.$$

If x_i is known only as a probability density, $p(x_i)$, then the best a rational decision-maker can do is to adjust u_j once, to be the best in consideration of the whole density function $p(x_i)$. In this case the average reward over a set of x_i is

$$V'_{\text{avg}} = \max_j\{\Sigma_i p(x_i) V(u_j \mid x_i)\}.$$

Information value, then, is the difference between the gain in taking the best action given each specific x_i as it occurs, and the gain in taking the best action in ignorance of each specific x_i, i.e., knowing only $p(x_i)$. This difference is

$$V^*_{\text{avg}} = V_{\text{avg}} - V'_{\text{avg}}$$

Manuscript received June 11, 1993; revised February 11, 1994.

The author is with the Human-Machine Systems Laboratory, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

IEEE Log Number 9404974.

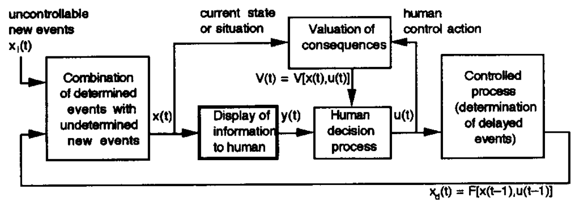

Fig. 1. Human-machine system in which roles of information and information value may be considered.

These two concepts and their application in human-machine systems may be considered in conjuction with Fig. 1.

State $x(t)$ is determined by both uncontrolled independent externalities and feedback from a controlled process (which typically has a delay in it). Selected aspects of $x(t)$ are measured and displayed to the human operator as information $y(t)$. Both $x(t)$ and human control action $u(t)$ determine the value consequences

$$V(t) = V[x_i(t), u_j(t)] = V(u_j \mid x_i)p(x_i).$$

The information display is highlighted to emphasize that it is the system design variable: how much information $y(t)$ to provide in order to maximize V?

II. How Do Information and Information Value Relate?

Curiously, the literature, at least that with which the author is familiar, seems devoid of discussion regarding the relation between these two straightforward concepts, which from their names and even their domains of application might seem to have a natural tie. In the field of human-machine systems, it seems that the modelers and analysts using Shannon information measures, e.g., Hick's law (1952) and Fitts' law (1954), have not concerned themselves with economic value, while operations researchers and economists have done the reverse. A model of human sampling behavior was proposed by the author (Sheridan, 1970) based on information value theory, but the relation to information theory was not mentioned.

Clearly the Shannon measure of information (uncertainty reduction) is a measure of the *effort or other cost* required to *discover* the truth from an initially uncertain state. It can be applied to specification of x_i by a machine or by cognitive effort, e.g., in the former case by providing the information in a display, in the latter case by retrieving information from memory, or by finding information by visual search. It says nothing about what one can gain by having the information, i.e., by taking action based on the information.

Just as clearly the idea of information value as stated above regards only what can be gained by *taking action* based on the information as compared to taking action without it. It can also include the *cost of action*. However it says nothing about the effort required or cost to get the information.

Thus, it seems, we have in these two ideas the fundamental and complementary ingredients to state what one pays to acquire

0018-9472/95$04.00 © 1995 IEEE

information together with what one earns by taking action based on it. However the two quantities defined above are not in the same units. H_{avg} is itself dimensionless, whereas V^*_{avg} has units of value.

By adding a cost coefficient C per bit we get

$$H^*_{\text{avg}} = -C\Sigma_i p(x_i)\log_2[p(x_i)]$$

now in value units. Then we have (1) the cost of information seeking or uncertainty reduction (e.g., time away from other things) or the dollar cost of providing the information (e.g., the cost of hardware, software or manpower), and (2) the gain for action or control (minus the costs of action itself), both in commensurable units of worth.

The difference, putting action gain first,

$$\begin{aligned} V^*_{\text{net}} &= V^*_{\text{avg}} - H^*_{\text{avg}} \\ &= \Sigma_i p(x_i)\{\max_j[V(u_j \mid x_i)]\} \\ &\quad - \max_j\{\Sigma_i p(x_i)V(u_j \mid x_i)\} + C\Sigma_i p(x_i)\log_2[p(x_i)] \end{aligned}$$

is the key quantity to be evaluated. (Because $\log_2[p(x_i)]$ in the last term is negative, the latter two terms essentially subtract from the first.) If $V^*_{\text{net}} > 0$, it is wise to seek (and use for control action) the information. Otherwise, the opportunity is not worthwhile. This, then, becomes a normative model for information seeking/using behavior by a human operator.

Sometimes the cost of accessing information in real time is regarded primarily as the degradation of control due to delay before the information is used, i.e., the information gets out of date quickly, and the more careful and full the information accessing activity, the longer the delay. This is not a cost for time away from other tasks, but rather for time away from some other way to get information about the same task. For example, when an aircraft pilot goes head down to read instruments or to call up information from the computerized flight management system his attention is necessarily distracted from observing the changing terrain, weather or positions of other aircraft out the window (Patrick, 1992). In this case the information cost is not an independent subtractive term but rather a multiplicative function which monotonically diminishes the information benefit with increased time t. In this case

$$\begin{aligned} V^*_{\text{net}} &= [\Sigma_i p(x_i)\{\max_j[V(u_j \mid x_i)]\} \\ &\quad - \max_j\{\Sigma_i p(x_i)V(u_j \mid x_i)\}][1 - C'(t)] \end{aligned}$$

where $C'(t)$ increases monotonically as time increases. After $C'(t) > 1$, using the information source is worse than not using it.

III. Use of the V^*_{net} Measure

The V^*_{net} measure can be a basis for deciding, for example:

- what information to put into a data base, given the probable effort of a user to find it in the data base and the probable value of its use when found;
- whether a procedure should require an operator to check a certain instrument, given the probable effort to do the checking and the probable gain of knowing what the instrument says versus not knowing it;
- whether to do some research, considering the probable cost of the research and the probable advantage of getting useful answers from the research.

To use the V^*_{net} measure, one must have: (1) an explicit value function $V(u_j \mid x_i)$; (2) the probability density $p(x_i)$; and (3) the relative information accessing cost-per-bit C or information degradation factor C'. In real life the set of states x_i has a very large number of dimensions and may be continuous or at least have a large number of levels along each dimension, suggesting that the modelling problem is intractable. I believe that useful results may be obtained from considering only a few variables with but a few levels of each, i.e, where an equivalent one-dimensional array of x_i has less than 100 elements.

requirements

Assessing $V(u_j \mid x_i)$ is probably the most difficult aspect of applying this theory. For simple cases $V(u_j \mid x_i)$ can be derived from control analysis based on some model of what states are good or safe, and, given current state x_i, how any particular control action u_j makes that state better or worse. It is common for continuous control systems to assume some quadratic negative weighting of error from a nominal course, time away from nominal course, and magnitude of control action (e.g., which might translate into fuel used).

It is tempting to reflect on how $V(u_j \mid x_i)$ relates to the sensibilities of people for changes in their environmental state (or what is sometimes called the surrounding "situation"). Well established by psychophysical methods are three relevant characteristics of people's preferences:

(1) People value change for the better more than they value static existence in a "good" state. Some philosopher said it as "Pursuit is sweeter than possession." -> who?

(2) People are sensitive to physical changes in state in proportion to what exists. The just-noticeable difference in stimulus magnitude ΔI is proportional to the initial magnitude I (known to psychologists as "Weber's law"). This also means that sensation is a logarithmic function of I (known to psychologists as "Fechner's law").

(3) People tend to value changes (for the good) better if they occur faster, up to some point, then dislike change that is too rapid. In other words the marginal utility for rate of change tends to be single-peaked.

These empirical facts about human preference might provide heuristics to help model $V(u_j \mid x_i)$.

The density $p(x_i)$ can be estimated subjectively. In situations where there is some experience, subjective probability estimates are commonly used and reasonably reliable.

One would like the cost variable C to scale information-provision in terms commensurable to $V(u_j \mid x_i)$. To infer this one might determine from simulator experiments which information opportunities are sought and which are ignored by a trained subject. Then the condition for which $V^*_{\text{avg}} = H^*_{\text{avg}}$ can be established in selected cases.

The corresponding variable C' is based on the risks of delay and inattention. Again, in selected situations, and given certain information on how better to do a task, one might be able to establish experimentally how long one would be willing to delay before the information were regarded as useless.

IV. Conclusion

The concepts of information as uncertainty reduction and information value for control are differentiated in the context of a theoretical framework for optimizing use of information in man-machine systems.

It may be that many opportunities are being missed for making available information which can be easily accessed, which can be used to take profitable actions, and for which the actions themselves are not too costly. On the other hand, many expert systems and/or human operator procedures or practices may be overly wasteful or otherwise inappropriate because some or all of the effort to access the information, the relative gain in taking action based on the information, and the cost of the action are too large or too small in relation to one another.

REFERENCES

[1] C. E. Shannon, "Communication in the presence of noise," *Proceedings of IRE*, vol. 37, pp. 10–22, 1949.
[2] H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory*, Cambridge, MA: Harvard University Graduate School of Business Administration, 1961.
[3] R. A. Howard, "Information value theory," *IEEE Trans. Systems Science and Cybernetics*, vol. SSC-2, pp. 2–26, 1966.
[4] W. E. Hick, "On the rate of gain of information," *Quart. J. Experimental Psychology*, vol. 4, pp. 11–26, 1952.
[5] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *J. Experimental Psychology*, vol. 47, pp. 381–391, 1954.
[6] T. B. Sheridan, "On how often the supervisor should sample," *IEEE Trans. Systems Science and Cybernetics*, vol. SSC-6, pp. 140–145, 1970.
[7] N. J. M. Patrick, *Ph.D. Thesis* in progress, Dept. of Mechanical Engineering, MIT, 1994.

Fuzzy Hypergraph and Fuzzy Partition

Hyung Lee-Kwang and Keon-Myung Lee

Abstract— **In this paper, the concept of hypergraph is extended to the fuzzy hypergraph. In the fuzzy hypergraph, the concepts of α-cut hypergraph, strength of edge and dual fuzzy hypergraph are developed. It is shown that the fuzzy hypergraph and α-cut hypergraph are useful to represent a fuzzy partition. An application example also shows that the strength of edge can be used to decompose the data set in a clustering problem.**

I. INTRODUCTION

The hypergraph was introduced by Berge [1] and has been considered as a userful tool to analyze the structure of a system and to represent a partition, covering and clustering [2]–[4]. The notion of hypergraph has been extended in the fuzzy theory and the concept of fuzzy hypergraph was provided by Kaufmann [5], [6]. However, it has been pointed out that the Kaufmann's definition of fuzzy hypergraph is not appropriate to represent various systems such as fuzzy partition.

In this paper, we generalize the concept of fuzzy hypergraph and redefine it to be useful for the analysis and fuzzy partition of a system. Some useful concepts will be developed such as the strength of edge (class), strong class, α-cut hypergraph and dual fuzzy hypergraph. We will see that the proposed fuzzy hypergraph is useful to describe visually a fuzzy partition or covering. Furthermore, the strength of edge (class) allows us to select strong classes in a partition, and a discrimination of the strong classes from the other parts can reduce the remaining data to handle. The proposed concepts can be used in system analysis, circuit clustering and pattern recognition, etc. [13]–[15].

In section II, brief definitions and terminologies of hypergraph are introduced. In section III, we develop the concepts of fuzzy hypergraph, strength of edge, α-cut hypergraph and dual fuzzy hypergraph. Section IB shows a possible application of the proposed fuzzy hypergraph in fuzzy partition. We will see that by using the proposed concepts, a decomposition or reduction of data in a clustering is possible.

In the literature [11], [12], when a *fuzzy set* $\tilde{A}$ is given in a universal set X, the fuzzy set $\tilde{A}$ is represented by

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) \mid \mu_{\tilde{A}}(x) > 0\}$$

where $\mu_{\tilde{A}}(x)$ is the membership function representing the membership degree of element x in the fuzzy set $\tilde{A}$. The *support of the fuzzy set* $\tilde{A}$, supp($\tilde{A}$), is defined as:

$$\text{supp}(\tilde{A}) = \{x \mid \mu_{\tilde{A}}(x) > 0\}$$

The support of a fuzzy set is a crisp set. We can cut a fuzzy set $\tilde{A}$ at level α and have *α-cut* A_α of $\tilde{A}$ such as:

$$A_\alpha = \{x \mid \mu_{\tilde{A}}(x) \geq \alpha\}$$

When two fuzzy sets $\tilde{A}$ and $\tilde{B}$ are given, we can obtain the union $\tilde{A} \cup \tilde{B}$ and the intersection $\tilde{A} \cap \tilde{B}$. The sets $\tilde{A} \cup \tilde{B}$ and $\tilde{A} \cap \tilde{B}$ are defined by their membership function $\mu_{\tilde{A}\cup\tilde{B}}(x)$ and $\mu_{\tilde{A}\cap\tilde{B}}(x)$ respectively: $\mu_{\tilde{A}\cup\tilde{B}}(x) = \max[\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)]$, $\mu_{\tilde{A}\cap\tilde{B}}(x) = \min[\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)]$.

In general, a family $(A_1, A_2, ..., A_m)$ of subsets $(\forall i, A_i \neq \phi, A_i \in X)$ in a set X is called a *partition* of X if the following conditions are satisfied [8], [9].

$$\cup_i A_i = X, \quad i = 1, \ldots, m$$
$$A_i \cup A_j = \phi, \quad i, j = 1, \ldots, m (i \neq j)$$

If the family $(A_1, A_2, \ldots, A_m)$ allows $A_i \cap A_j \neq \phi$ for $i \neq j$, it is called a *covering* (or *cover*) of X. A *fuzzy partition* of set X is a family $(\tilde{A}_1, \tilde{A}_2, \ldots, \tilde{A}_m)$ of fuzzy sets $(\forall i, \tilde{A}_i \neq \phi, \tilde{A}_i \neq X)$ such that

$$\cup_i \text{ supp}(\tilde{A}_i) = X \tag{1}$$

$$\sum_{i=1}^{m} \mu_{\tilde{A}_i}(x) = 1, \forall x \in X \tag{2}$$

In the clustering, the fuzzy set $\tilde{A}_i$ is called a *fuzzy class*. We can also develop the concept of fuzzy covering. Let's call a family $(\tilde{A}_1, \tilde{A}_2, \ldots, \tilde{A}_m)$ a *fuzzy covering* of X if it verifies only the above condition (1) but (2).

II. HYPERGRAPH

The *hypergraph* $H = (V, \mathcal{E})$ was proposed by Berge [1] and is defined as follows:

$$\begin{aligned} &H = (V, \mathcal{E}) \text{ where} \\ &V = \{x_1, x_2, \ldots, x_n\}\text{: a finite set of vertices} \\ &\mathcal{E} = \{E_1, E_2, \ldots, E_m\}\text{: a family of subsets of } V \\ &E_j \neq \phi, j = 1, \ldots, m \\ &\cup_j E_j = V \end{aligned}$$

The set V is called the set of vertices and $\mathcal{E}$ is the set of edges (or hyperedges). In the diagram, the edge E_j is represented by a solid

Manuscript received March 23, 1993; revised February 12, 1994.
The authors are with the Department of Computer Science, Korea Advanced Institute of Science and Technology (KAIST), Taejon, 305-701, Seoul Korea.
IEEE Log Number 9404973.

0018-9472/95$04.00 © 1995 IEEE