THE DESIGN AND EVALUATION OF SITUATION ASSESSMENT STRATEGIES

Andrew BORDEN

Mind operates on sensations to create information for its own use.
Franklin, Oyama[1]

Introduction and Basic Concepts

The purpose of this section is to establish the framework for the analysis that follows and to introduce some of the basic terms and definitions in an informal way. The concepts of "Entropy" and "Information" will be introduced here, but defined more formally (and quantitatively) later and tools to measure them will be described.

The core concept in this paper is "Situation Assessment". "Situation" in the context of warfare consists of the composition, readiness, location and status of adversary systems and forces. "Assessment" takes place within a "Frame of Discernment". A frame of discernment is a set of distinguishable possibilities, one of which is the *actual* situation. The objective of situation assessment is to make the distribution of probabilities on the frame of discernment asymmetrical so that one possible situation is determined to be highly probable and the others less probable. Typically, we want the high probability to exceed some threshold before we accept it as a successful assessment. That is, we begin the process at some level of uncertainty (entropy) and we reduce the uncertainty until we have sufficient confidence to make the assessment. The reduction of uncertainty *is* information. For emphasis, the reduction of uncertainty is not *produced by* information, it *is* information. The "assessment" is accomplished by the systematic generation and use of information, usually according to some efficient strategy.

Information can only be produced in the presence of Knowledge. In the absence of knowledge, all messages from the environment are noise. Moreover, Knowledge, in the form of a data base for example, is not enough. There must be a procedure for *using the knowledge to reduce uncertainty* when Messages from the environment are received.

INFORMATION & SECURITY. An International Journal 1, 1 (Summer 1998), 63-74.

It could be that the information-producer has the option to select types of messages from the environment. Each type of message could make a different information contribution, depending on the data base, the current uncertainty and how noisy the message communications channel is. Moreover, each type of message could have a different cost, perhaps time. It is clearly necessary then, to develop a strategy for selecting messages to reduce the entropy most efficiently.

If all this seems abstract, let the reader take heart. We will illustrate these ideas first with a familiar example, then with a generalization of the example complete with mathematical formulas and figures.

Example of the Basic Concepts

The example we will use is that of a radar parametric data base being developed and used to generate a strategy to identify radars quickly and with high confidence. The output of this process would be the Radar Classification Algorithm (RCA) in a Radar Warning Receiver (RWR). This is a straightforward illustration of situation assessment in which an accurate and relatively complete data base exists.

The RCA selects parameters to be measured depending on what is already known, how much information the new parameter will contribute and how much the parameter measurement will cost. The idea is to reduce entropy as quickly as possible until the probability of one radar classification exceeds the specified threshold. Then, the aircrew is notified so that appropriate defensive actions can be initiated.

The procedure for iteratively selecting parameters is the RCA strategy. Unfortunately, finding the *optimal* strategy is a Non-deterministic, Polynomial Time (Complete) (NP-Complete) problem. This means that we will almost always have to settle for a less-than-optimal solution, but it must be demonstrably good.[2] (More about the "demonstrable" part later). The impact of NP-completeness is that, to ensure good run-time performance, *very hard work* is required during the algorithm design phase.[3] How we accomplish this very hard work quickly and reliably is the key point in this paper.

Roadmap to the Analysis

The remainder of this paper is in five parts, followed by conclusions. The first part contains a brief introduction to the mathematics of information, based on Shannon's classic work from the 1940's.[4] The second part shows how knowledge is obtained and used to construct a data base. The third part contains an explanation of how the data base becomes an active memory which is the background for entropy reduction (information production). The fourth part contains an explanation of how the active memory and other factors support the development of an efficient situation assessment strategy. Finally, a tool for designing and/or evaluating strategies is described and its use is illustrated with a quantitative example. The tool can be used to measure the utility of attack and protect measures in Information Warfare.[5] The RCA example will be the unifying thread in this analysis.

We will follow Shannon in calling data from the environment, for example parametric measurements, “messages,” not “information.”

Fundamental Definitions in the Theory of Information

Entropy and Information

Definition 1: *Information* is the degree to which uncertainty (entropy) is reduced.

We define "information" in terms of "uncertainty" because, when performing any type of situation assessment, we begin with relative uncertainty and attempt to replace it with certainty. Therefore, uncertainty is the starting point for all information theoretic definitions. Another name for uncertainty is "entropy." We can compute entropy whenever we have a probability distribution. Given a probability distribution $\{P(i)\}$, $i = 1,2,3,...,N$, where each $P(i) > 0$, the entropy is[6]:

$$H = - \sum_i P(i) * \mathrm{Log}_2\, P(i)$$

Entropy is measured in bits. For example, if “Land” and “Sea” were equally probable, Paul Revere’s initial entropy was exactly one bit. The signal from the Old North Church (one lantern if by land, two if by sea) contained exactly one bit of information, reducing the entropy to zero. For another example, if there are four, equally probable possibilities (with probability 0.25), our initial entropy is:

$$H = - (0.25*(-2) + 0.25*(-2) + 0.25*(-2) + 0.25*(-2)) = 2 \text{ bits.}$$
$$(\text{because } \mathrm{Log}_2\, 0.25 = -2)$$

If in any given situation, there is a current set of distinguishable possibilities and a current assessment of probabilities, the entropy is computed by the equation. As you might expect, entropy is low when one of the possibilities has a very high probability and all the others have low probabilities. The converse is true. Entropy is high when all possibilities are almost equally probable. Our objective in situation assessment is to reduce entropy, i.e. to sharpen the probability distribution so that we can select one of the possibilities with a specified level of confidence.

Mutual Information

A message from an environment is the result of a measurement of one or more parameters of something in the environment. The "information" contained in a message is the reduction in entropy produced by the message. So, if we receive a new message from the environment, we may revise our assessment of the current probabilities and re-compute the entropy. The difference between the new and old entropies is the information (in bits) contributed by the message (Definition 1).

If there are several parameters which could be measured, we want to choose the one which is likely to provide the most information (entropy reduction). The expected amount of information to be derived from a new parameter measurement is called the "mutual information" between the new information source and our current knowledge. La France provides a formula to compute mutual information.[7] Mutual information is measured in bits.

> *Definition 2*: *Mutual information* between a new message source and our current knowledge is the expected value of the information (entropy reduction) to be obtained by evaluating a message from the new source.

For example, suppose that, as the result of making a parametric observation of a situation, we have new probabilities for the situation in the above example: 0.125, 0.125, 0.25, 0.5. The new entropy is:

$$H = -(0.125*(-3) + 0.125*(-3) + 0.25*(-2) + 0.5*(-1)) = 1.75 \text{ bits.}$$

So, this parameter measurement has reduced our entropy by 0.25 bits. The weighted average entropy reduction over *all possible values* of the new parameter is the mutual information between the new parameter and what we already know. In other words, the mutual information of a candidate information source is the amount to which we can expect uncertainty (entropy) to be reduced by the source. (Definition 2)

Information Payoff

If each candidate data source has a cost associated with it, we would, of course, compute the ratio of mutual information to cost. If the cost is time, we would come up with a rate of information production from the new source in bits per second.

> *Definition 3*: The *information payoff* from a candidate message source is the ratio of mutual information to the cost of using the candidate source.

Suppose we have a number of intelligence resources available to assess a situation. If the knowledge base includes statistical descriptions of each possible situation, we can compute the expected value of the information payoff from each resource. In this case, the cost might not be only in time. It might be in lack of covertness, fuel, risk or some other commodity. We would select the intelligence resources which have the highest information payoffs, based on what we already know. This is done iteratively until the classification of the situation is successful or until all the intelligence resources have been used without success.

Knowledge Acquisition and Data Base Development

Figure 1 illustrates how passive knowledge (a data base) is developed, then becomes an active knowledge base or associative memory. Without some supervision or

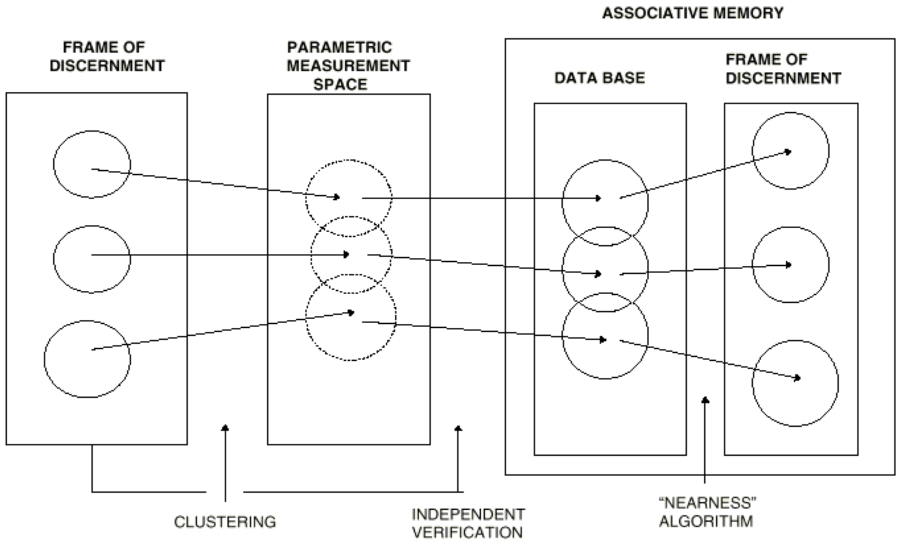


Figure 1: Knowledge Development

independent verification of the meaning of messages (parametric measurements), little real knowledge develops. It is possible to do some clustering within the parameter measurement space and to hypothesize that certain values of different parameters appear to occur together and may be related. Visualize the Lieutenant in Electronic Warfare School looking at his equipment for the first time. He or she sees a plethora of messages which seem to be very noisy indeed. After a few hours, aural pitch, location on the CRT and scan frequency may assemble themselves into rough patterns, but real knowledge remains very sparse until certain parameter clusters are related to objects (Radars) that actually exist in the real world.

With independent verification, rough clusters can be refined and related to distinguishable objects (radars) in the frame of discernment. Now, we have the beginnings of a data base!

Two things are now needed to develop a relatively complete, usable data base. The first is a dedicated, purposeful collection effort, guided at every step by independent verification. The second is the design of a sound data base format, based on the intended use of the information. If the intended use is to facilitate smart, structured queries and sorts, the best format is highly relational, with every table having perhaps only two fields - a key and a single parameter. It is well-known that this dedicated collection effort exists and that it continually supports and updates the resulting data base.

Note that, in Figure 1, there is overlap between the clusters of parameter measurements. For one or more reasons, it is impossible to distinguish absolutely between objects in the frame of discernment by using the data base. The probable reason is that not enough parameters are available to make a classification with zero uncertainty. This is almost certainly the case when only one parameter measurement is available to evaluate. The overlap can be regarded as noise in the channel through which the parameter was sent. Thinking about ambiguity in terms of noise is natural in the context of Shannon's information theory.

If we have one or more parameter measurements for an object (radar), how do we use the data base to reduce entropy? The answer is that we use some sort of "Nearness" function to identify the best candidate classification. There are a number of ways to specify nearness, but the one that is consistent with Shannon's mathematics is the Bayesian one. That is, we compute the conditional probabilities for each candidate classification, given the parameter that we have measured. The candidate with the highest conditional probability is the nearest, so is the best guess based on what we currently know. When we get near enough to one radar (when the probability gets high enough), we commit and take appropriate action.

Development of an Active Memory

In Figure 2, the parameter messages coming from the environment are shown as dashed lines to indicate that one or more, but not all, of the parameters may be selected. We are using our nearness function only with the tables in the data base which are relevant to the selected parameters.

The data base with the nearness function is an active, associative memory. An associative memory does precisely what we have described. It takes a partial description of an object and finds the nearest classification - the one having the highest probability of being the correct one. Recall that, as the probability distribution sharpens (as one object gets very near to our parameters), the entropy becomes less, so information is being produced.

The Figure shows that parameter measurements are selected iteratively according to the situation assessment strategy until we are near enough to some known radar in the environment to make the classification.

Designing a Strategy for Situation Assessment

In an RWR, it is necessary to deinterleave pulse trains in order to isolate a radar signal which requires classification. Since the average Pulse Recurrence Interval (PRI) is used to accomplish deinterleaving, this parameter is automatically available

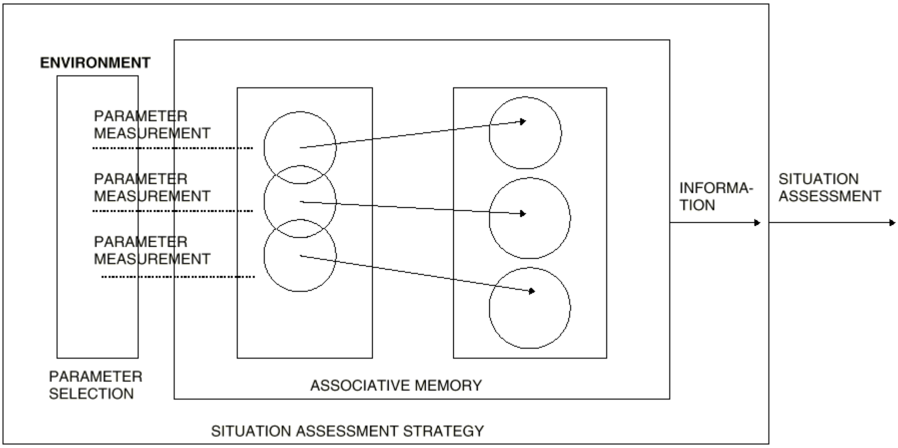


Figure 2: Applying the Situation Assessment Strategy

without additional cost in time. Therefore, in designing a good RCA, the first step is to compute the conditional probabilities for candidate radars, given PRI. If PRI has given us enough information (reduced entropy enough), we make the classification decision based on the nearest match (highest probability). If not, we have the very interesting question regarding which parameter to measure next. This decision has the very practical and important result of ensuring that valuable resources are used in the most efficient way.

The next section contains a description of a tool (the situation assessment evaluation tool) which explains how efficient parameter selection can be based on Bayesian probabilities and Shannon's information theory. The original motivation for developing this tool was precisely to enable the design of a good RCA for use in RWR's. It is generally usable in any situation assessment task however.

The Situation Assessment Tool (SAET)

Description of the Tool

Using the principles described above, we have developed a program which will build a nearly optimal situation assessment strategy, given a statistical description of possible outcomes, confidence requirements and the capabilities of the assessment system to perform parameter measurements..

The tool computes the initial entropy based on an *a priori* assessment of the probabilities of the objects in the frame of discernment and selects the parameter measurement with the best information payoff. (In the case of the RCA, it is PRI). For each possible PRI value, the parameter measurement with the next highest payoff is selected. The process continues recursively until a decision tree is complete with each node labelled with a “Classification” (radar type), “Failure” (to reach the specified confidence level), or “Unknown”, (meaning that the combination of parameters does not match anything in the data base). At each node in the decision tree, statistics are available on the current conditional probabilities, the current entropy, current cost and other statistics which permit evaluation of the Classification strategy.

At each node in the decision tree, the list of conditional probabilities serves as a nearness function. Of course, the determination of conditional probabilities and entropies is very computation intensive. This is a mixed blessing. When the strategy is completed, we know everything about its throughput, entropy reduction and expected response time. We can assign it a report card and/or edit it to make any needed tradeoffs. For example, we might be willing to accept a slightly lower confidence level to replace a “Failure” result with a Classification. The maintenance of performance statistics is the “demonstrable” part of “demonstrably good” as discussed previously in the second section.

Figure 3 is a flow chart for the program. There is also a turnkey Data Base - to - Knowledge Base Module which discretizes parameter measurements into windows and computes all the probabilities which are needed to start the program. The user also has the option to input a partial decision tree and let the program complete the design work.

This tool can be used in IW evaluation as follows:

Using best *a priori* information, a situation assessment strategy is developed for a given Situation. Statistics on efficiency (bits of information and time required) are recorded.

An information attack and/or an information protect measure is applied. A revised strategy is developed and efficiency is measured again. The difference in efficiency is a measure of how the difficulty of situation assessment has been changed by the information attack/protect measure.

Numerical Example

Consider the situation in which an adversary is designing an RCA. If he knows the *a priori* distribution of radars in our Air Defense System, his initial uncertainty is much less and the work he has to do to design a good classification algorithm is much easier. For the purposes of this illustration, we will assume that the adversary is using a "good" method to design his algorithm - like our design program mentioned above.

FIGURE 3: SITUATION ASSESSMENT EVALUATION TOOL FLOWCHART (1)

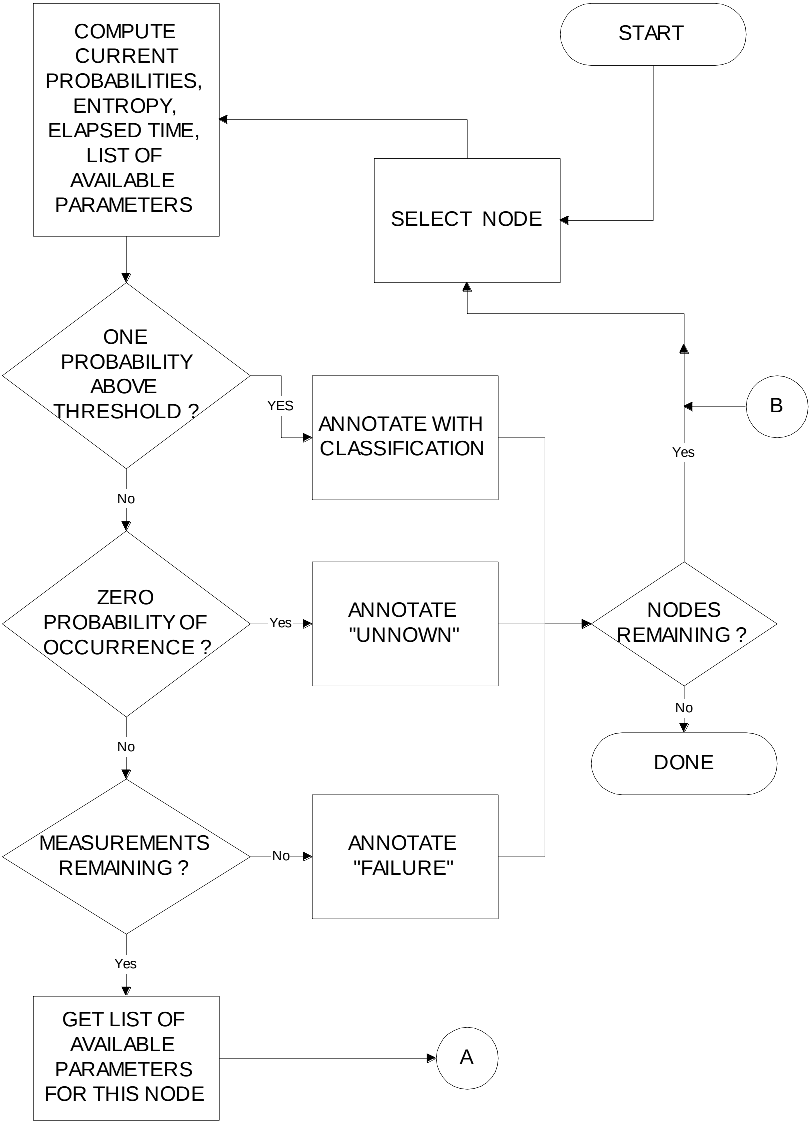
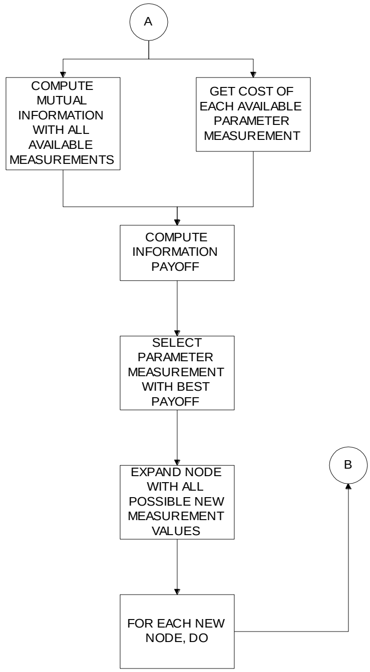


FIGURE 3: SITUATION ASSESSMENT EVALUATION TOOL FLOWCHART (2)



We are postulating an environment in which there are five types of radars and five possible parameter measurements, each requiring a different amount of time to accomplish. The objective of the classification is to use sequences of measurements which get us to 90 % confidence as quickly as possible. The strategy for utilizing parameter measurements is embedded in a decision tree.

Table 1 shows the real world probability of occurrence of radars in the environment. If the adversary does not know the real-world probabilities, he is forced to use a default assumption that all the radars are equally probable.

RADAR TYPE	**A**	**B**	**C**	**D**	**E**
REAL-WORLD PROBABILITIES OF OCCURRENCE	0.1	0.1	0.1	0.2	0.5
DEFAULT (ASSUMED) PROBABILI-TIES	0.2	0.2	0.2	0.2	0.2

Table 1 : Real World and default probabilities of radars

	PROBABILITIES KNOWN	PROBABILITIES ASSUMED
INITIAL ENTROPY (bits)	1.96	2.32
FINAL ENTROPY (bits)	0.35	0.32
INFORMATION MARGIN (bits) (INITIAL - FINAL)	1.61	2.00
MEAN TIME FOR CLASSIFICATION (sec)	0.87	1.14

Table 2: Entropy change (information margin) and time required for classification. (90 % confidence required).

We now design classification algorithms using the two different assumptions about the *a priori* probability of occurrence of radars. Table 4 shows how the resulting classification algorithms perform. The initial entropy is much greater if the *a priori* probabilities are unknown. The result is that the information margin (the amount of work that must be done to make a classification) is much greater (2.00 versus 1.61 bits). The cost in seconds is also greater.

The conclusion to be drawn from this example is that the work required in doing this situation assessment (information margin of the friendly situation as assessed by the adversary) can be increased by 25 % if we successfully apply certain information protect measures, i.e. operational security (OPSEC).[8]

For each of the situations in Table 2 (known and unknown *a priori* probabilities), the program took about a minute to design a classification algorithm from scratch, provide performance numbers for every possible combination of parameters and compute expected values for elapsed time and entropy reduction. The computation was done using a Pentium 100.

Conclusions

- Information is created or produced when parametric messages from the environment are processed by an active memory.
- The active (associative) memory consists of a data base and an algorithm which computes a “nearness” function.
- The nearness computation reduces uncertainty and produces information

- Developing a good situation assessment strategy means selecting parameters whose values will produce the most information (reduce uncertainty) at the least cost. This process can be automated.

The situation assessment design and evaluation tool can be used to quantify the utility of certain attack and protect measures in Information Warfare.

---

1 Stan Franklin, *Artificial Minds* (Cambridge, MA.: The MIT Press, 1995); Susan Oyama, *The Ontogeny of Information* (Cambridge, MA.: Cambridge University Press, 1985).

2 Andrew Borden, "Salvaging the ARC," *The Journal of Electronic Defense*, Publication of the Association of Old Crows (October 1988).

3 Nils J. Nilsson, *Principles of Artificial Intelligence* (Palo Alto, CA.: Tioga Publishing Co., 1980).

4 C.E. Shannon, "A Mathematical Theory of Communications," *Bell Syst. Tech. J.* 27, (1948): 379-423 and 623-656.

5 Andrew Borden, "An Information Warfare Roadmap," in *Information Aspects of Security and Development of Modern Societies*, Proceedings of the AFCEA Europe Seminar, ed. Velizar Shalamanov and Todor Tagarev (Sofia: AFCEA Sofia, 1996), 5-16.

6 Shannon, "A Mathematical Theory of Communications," and Pierre La France, *Fundamental Concepts in Communications* (Prentice Hall International Editions, 1990).

7 La France, *Fundamental Concepts in Communications*.

8 Borden, "An Information Warfare Roadmap," contains a discussion of Information Margin and other measures of effectiveness in the context of Information Warfare.

Andrew BORDEN is a computer scientist with long experience in Electronic Warfare. He has published many papers on the subject of decision-making systems. Mr. Borden is a retired USAF officer. His last assignment was Deputy Chief of Staff for Intelligence in what is now the USAF Air Intelligence Agency. He has advanced degrees in mathematics from Kansas State and Ohio State Universities and advanced graduate work in computer science at the University of Texas. Currently, he works for DRH Consulting, San Antonio, Texas. Address for correspondence: 1210 Scenic Knoll, San Antonio, TX 78258, USA.