AI Magazine Volume 5 Number 3 (1984) (© AAAI)

Probability Concepts For An Expert System Used For Data Fusion

Herbert E. Rauch

Lockheed 92-20/205
Palo Alto Research Laboratory
3251 Hanover Street
Palo Alto, CA 94304

Abstract

Probability concepts for rule-based expert systems are developed that are compatible with probability used in data fusion of imprecise information Procedures for treating probabilistic evidence are presented, which include the effects of statistical dependence. Confidence limits are defined as being proportional to root-mean-square errors in estimates, and a method is outlined that allows the confidence limits in the probability estimate of the hypothesis to be expressed in terms of the confidence limits in the estimate of the evidence. Procedures are outlined for weighting and combining multiple reports that pertain to the same item of evidence. The illustrative examples apply to tactical data fusion, but the same probability procedures can be applied to other expert systems

KNOWLEDGE-BASED EXPERT SYSTEMS are a class of computer programs intended to serve as consultants for decision making. These programs use a collection of facts, rules of thumb, and other knowledge about a limited field to help make inferences in the field. They differ substantially from conventional computer programs in that their goals may have no algorithmic solution, and they must make inferences based on incomplete or uncertain information. They are called expert systems because they address problems normally thought to require human specialists for solution, and knowledge-based because researchers have found that amassing a large amount of knowledge, rather than sophisticated reasoning techniques, is responsible for the success of the approach.

Advantages of an expert system are that is can be designed to supply one or more hypotheses to the user, request additional information from the user, explain to the user the reasons for the hypotheses or for the requests for additional information, and allow the addition or deletion of knowledge and rules without extensive reprogramming. A recent survey article by Gevarter (1983) discusses expert system applications in areas such as medical diagnosis, geology, and computer configuration analysis and evaluates the limitations of current systems. Limitations exist on the current use of expert systems because formalizing the knowledge of experts is a difficult task; building the system is laborious and time-consuming; operation is effective only in a relatively limited field; and degradation is not always graceful when problems are outside of the limited field. Newer expert systems are being developed that find ways around these limitations, and future use and growth of expert systems should increase. Duda and Shortliffe (1983) discuss current research in expert systems, while the book edited by Barr and Feigenbaum (1982) presents more detailed material.

This article is concerned with the data fusion aspect of expert systems. The correlation and fusion of information from sensor systems is becoming increasingly important for

This research was supported by the Lockheed Independent Research Program An early version of this paper was presented at the Seventeenth Annual Asilomar Conference on Circuits, Systems and Computers, Pacific Grove, California, October 31-November 2, 1983.

military command and control and for technical intelligence analysis. High data rates from increasingly sophisticated sensors are overwhelming the existing methods of processing this data. The high volume of data makes timely interpretation difficult and very demanding of human resources. It would be beneficial to have a system that processes routine information automatically so as to free the human analyst to concentrate on non-routine tasks. Rauch, Firschein, Perkins, and Pecora (1982), Brown and Goodman (1983), Pecora (1984), and Rauch (1984) outline how expert systems might be applied as automated decision aids for tactical data fusion.

In traditional approaches to data fusion the quality of signal data is characterized by quantities such as root mean square error (which will be called standard deviation) and correlation. Data from various sources can be combined using weightings based on the standard deviation and correlation in measurement errors. When information from human experts is used, the experts make estimates and indicate the quality of the estimates.

For the purpose of data fusion, it is necessary that the signal information and the human estimates be expressed in similar forms. When a rule-based expert system is used to aid in this data fusion, it is necessary that the values and weighting of the signal information and the human estimates be expressed in a form that can be interpreted by rules. The rules must include a procedure for weighting conflicting or time-varying reports, for weighting correlated data, and for propagating confidence limits through a hierarchy of production rules.

Methods for using expert systems to combine uncertain information from various sources have been developed in the literature. For example, Shortliffe and Buchanan (1975) use a probability model based on assumptions of independent evidence as discussed by Adams (1976). Duda, Hart, and Nilsson (1976) use a Bayesian approach; Ben-Bassat (1980) uses a classification method; and Garvey, Lowrance and Fischler (1981) and Dillard (1983) use the Dempster-Shafer theory. Martin-Clovaire and Prade (1983) discuss some of these ideas in more detail. However, many of these approaches do not treat correlated data, multiple reports, or propagation of confidence limits through a hierarchy of production rules. In this article, probability concepts for rule-based expert systems are developed which have these properties. The illustrative examples used throughout this paper are for tactical data fusion, but the probability concepts can be applied to any rule-based expert system required to combine uncertain data from multiple sources.

Probabilistic Rules and Evidence

A typical deterministic rule for an expert system is in the form IF EVIDENCE E IS TRUE, THEN HYPOTHESIS H IS TRUE. When a probabilistic indication of likelihood is introduced, the rule becomes IF EVIDENCE E IS TRUE, THEN HYPOTHESIS H IS TRUE WITH PROBABILITY P_1. On the other hand, another version of the rule is IF EVIDENCE E IS NOT TRUE, THEN HYPOTHESIS H IS TRUE WITH PROBABILITY P_0. When a probabilistic indication of evidence is introduced, the information available becomes EVIDENCE E IS TRUE WITH PROBABILITY P_E. The probability of the hypothesis being true P_H can be calculated given the probability of the evidence being true P_E and the assumptions about probabilities P_0 and P_1.

Statistical Dependence	AND operation: $PROB(A \text{ and } B)$	OR operation: $PROB(A \text{ or B})$
Independence	$P_A P_B$	$P_A + P_B - P_A P_B$
Maximum Dependence	MINIMUM (P_A, P_B)	MAXIMUM (P_A, P_B)
Minimum Dependence	MAXIMUM $(P_A + P_B - 1, 0)$	MINIMUM $(P_A + P_B, 1)$

NOTE: P_A and P_B are the robabilities that A and B are true.

Consequences of Logical AND operation and logical OR operation can be calculated with independence, maximum dependence, and minimum dependence for two items of evidence.

Table 1.

$$P_H = P_1 P_E + P_O(1 - P_E)$$
$$= (P_1 - P_O)P_E + P_O$$

The relation between the evidence probability P_E and the hypothesis probability P_H is linear. Hence, if σ_E and σ_H are the standard deviations (root-mean-square) in the errors in the estimates of the probabilities of evidence P_E and hypothesis P_H, respectively, the two standard deviations will have the same linear relation.

$$\sigma_H = (P_1 - P_0)\sigma_E$$

For the examples here, the individual items of evidence are designated A and B, and the form of the rule is IF A IS TRUE, AND IF B IS TRUE, THEN H IS TRUE The hypothesis H might be that a SAM (surface-to-air missile) launcher is operational, and the evidence A might be ELINT (electronic intelligence) that a class of radar is transmitting, while the evidence B might be two-dimensional image data that show extensive preparation for installation of a missile launcher or artillery. Let E be the total evidence from A and B so that P_E is the probability that both A AND B ARE TRUE Let the probability that A is true (P_A) be equal to 0.5 and the probability that B is true (P_B) be equal to 0.5. In order to calculate the joint probability that A IS TRUE AND B IS TRUE, it is necessary to make an assumption about the statistical dependence between the item of evidence A and B.

	$P_B = 0.5$ TRUE	$1 - P_B = 0.5$ NOT TRUE
$P_A = 0.5$ TRUE	$P_A P_B$	$P_A(1 - P_B)$
$1 - P_A = 0.5$ NOT TRUE	$(1 - P_A)P_B$	$(1 - P_A)(1 - P_B)$

Note: The sum of the four joint probabilities in matrix is equal to unity. Columns sum to probability at top. Rows sum to probability at left.

Procedure of calculating matrix of joint probabilities involves probablility products when two items A and B are independent.

Table 2.

	$P_B = 0\ 5$ TRUE	$1 - P_B = 0\ 5$ NOT TRUE
$P_A = 0.5$ TRUE	$\mathrm{MAX}(P_A, P_B)$	$P_A - \mathrm{MAX}(P_A, P_B)$
$1 - P_A = 0.5$ NOT TRUE	$P_B - \mathrm{MAX}(P_A, P_B)$	$\mathrm{MAX}(1 - P_A, 1 - P_B)$

Note: $1 - \mathrm{MIN}(P_A, P_B)$ is equal to $\mathrm{MAX}(1 - P_A, 1 - P_B)$

Procedure for calculating matric of joint probabilities involves MAXimum and minimum operations when two items of evidence A and B have MAXimum dependence.

Table 3.

	$P_B = 0.5$ TRUE	$1 - P_B = 0\ 5$ NOT TRUE
$P_A = 0.5$ TRUE	$P_A - \mathrm{MAX}\ (P_A, 1 - P_B)$	$\mathrm{MAX}\ (P_A, 1 - P_B)$
$1 - P_A = 0.5$ NOT TRUE	$\mathrm{MAX}(1 - P_A, P_B)$	$1 - P_A - \mathrm{MAX}(1 - P_A, P_B)$

Procedure for calculating matrix of joint probabilities when two items of evidence A and B have minimum dependence is the same as when A has MAXimum dependence with NOT B.

Table 4.

Dependence of Evidence

Three possibilities for the statistical dependence are: the two items of evidence are statistically independent; the two items of evidence have maximum dependence; and the two items of evidence have minimum dependence. In the terminology used here, when A has maximum dependence with NOT B, it is stated that A has minimum dependence with B. The consequences of the logical AND operation and the logical OR operation are presented in Table 1 for the three possibilities as a function of P_A and P_B. A fourth, more general, possibility is introduced here that includes the three previous possibilities as special cases. The fourth possibility requires the introduction of the dependence parameter D (with D taking on values between -1 and +1). When D equals zero, the two items of evidence are independent; when D equals one, the two items have maximum dependence, and when D equals minus one, the two items have minimum dependence. The same procedure holds for more than two items of evidence, as long as all operations have the same dependence parameter D.

To return to the example, first consider the case where it is assumed that the two items of evidence are statistically independent. This means the occurrence of radar transmitting (item A) seems to be completely independent of the location of extensive preparation (item B). When two items of evidence are statistically independent, the joint probability is equal to the product of the two individual probabilities. For the example, this is (0.5) times (0.5,) which equals (0.25), as illustrated in Table 2.

Second, consider the case where it is assumed that the two items of evidence have maximum dependence, so whenever a radar of a certain class is transmitting, the preparation seems to be extensive and vice versa. This means that the joint probability that A is true and B is true (simultaneously) is maximized subject to the given probabilities P_A and P_B. For the example, with P_A equal (0.5) and P_B equal (0.5), maximum dependence gives the result that both A and B are true (simultaneously) with probability equal (0.5) as illustrated in Table 3.

Third, consider the case where it is assumed that the two items of evidence have minimum dependence so whenever a radar of a certain class is transmitting, the preparation usually is not extensive, and vice versa. This means that the joint probability of A and NOT B have maximum dependence and the joint probability that A is true and B is true (simultaneously) is minimized subject to the given probabilities P_A and P_B. For the example, with $P_A = 0.5$ and $P_B = 0.5$, minimum dependence means that both A and B are simultaneously true with probability zero as illustrated in Table 4.

Notice that for this example, varying the assumptions about statistical dependence has allowed the resulting probability to vary between 0 and 0.5 even though the probabilities of the items of evidence are held constant. Introducing the dependence parameter D allows the resulting probability to vary continuously between the extreme assumptions.

When there are three or more items of evidence, it is desirable that the truth table for the AND/OR relations have the property that the order in which the variables are considered (in multiple AND relations or in multiple OR relations) does not change the resulting probability. To ensure this property, the procedure that will be followed is first calculate the probability under the assumption that the items of evidence are independent (the probability from these calculations will be designated C_1). When the dependence

Statistical Dependence	AND operation: PROB (A AND B)	OR operation: PROB (A OR B)
Independence	$P_A^2\sigma_B^2 + P_B^2\sigma_A^2 + \sigma_A^2\sigma_B^2$	$(1-P_A)^2\sigma_B^2 + (1-P_B)^2\sigma_A^2 + \sigma_A^2\sigma_B^2$
Maximum or Minimum Dependence	$\frac{1}{2}(\sigma_A^2 + \sigma_B^2) + Q$	$\frac{1}{2}(\sigma_A^2 + \sigma_B^2) - Q$

Maximal Dependence:
$Q = \frac{1}{2}(\sigma_A^2 - \sigma_B^2)\text{Minimum}[\frac{\delta V}{L}, 1].$

$\delta V = \frac{P_B - P_A}{(\sigma_A^2 + \sigma_B^2)^{\frac{1}{2}}}$

Minimal Dependence:
$Q = \frac{1}{2}(\sigma_A^2 + \sigma_B^2)\text{Minimum}[\frac{\delta W}{L}, 1]$ if $\delta W \geq 0$,

$Q = \frac{1}{2}(\sigma_A^2 + \sigma_B^2)\text{Maximum}[\frac{\delta W}{L}, -1]$ if $\delta W < 0$,

$\delta W = \frac{P_B + P_A - 1}{(\sigma_A^2 + \sigma_B^2)^{\frac{1}{2}}}$

Note: P_A and P_B are the probabilities that A and B are true (with $P_B > P_A$) and σ_A and σ_B are the standard deviations in these probabilities and L is an *ad hoc* constant ($L = 2$ is reasonable).

Square of Standard Deviation (confidence limits) after Logical AND Operation and Logical OR Operation can be calculated with independence, maximum dependence, and minimum dependence with two items of evidence.

Table 5.

parameter D is positive, the second calculation is of maximum dependence (designated C_2). When the dependence parameter D is negative, the two calculations are probabilities under the assumption of independence (C_1) and under the assumption of minimum dependence (C_3). The resulting probability is a linear combination of the two appropriate calculations.

$$P = \begin{cases} DC_2 + (1-D)C_2, & \text{for } 0 \leq D \leq 1; \\ |D|C_3 + (1-|D|)C_1, & \text{for } -1 \leq D \leq 0. \end{cases}$$

Propagation of Confidence Limits

The standard deviation (root-mean-square error) in a probability estimate is a convenient way to express the confidence limits in the estimate. When each item of evidence that makes up a rule has an associated standard deviation as well as a probability, it is useful to be able to calculate both the probability the hypothesis is true and the standard deviation of hypothesis probability. In particular, rule-based expert systems have multiple levels of hypotheses so that hypotheses at lower levels can become evidence for hypotheses at higher levels. Hence, in addition to determining the probability that hypotheses are true, it may be necessary to determine the standard deviation of the hypothesis probability estimate.

The procedure to calculate the standard deviation is based on that used to calculate the probability of a hypothesis. It will be assumed that the errors in the probability estimate are independent, but the resulting standard deviation will be calculated under three assumptions about statistical dependence: the items of evidence are statistically independent, the items of evidence have maximum dependence, and the items of evidence have minimum dependence. The dependence parameter D (with D taking on values between -1 and 1) will be used to interpolate linearly between the square of the two appropriate standard deviations. Other methods of interpolation are possible, but linear interpolation will be used in this example.

It is not difficult to calculate the standard deviation when the two items of evidence are statistically independent. Let P_A, P_B, and P_E be the probabilities, let δ_A, δ_B and δ_E be the error in the probability estimate, and let σ_A,σ_B and σ_E be the associated standard deviations. The procedure for determining the standard deviation δ_E for the logical AND operation starts with the original probability calculations $P_E = P_A P_B$.

The original probability calculations are modified by adding the errors in probability estimates.

$$P_E + \delta P_E = (P_A + \delta P_A)(P_B + \delta P_B)$$

Subtracting the original from the modified gives the error equation.

$$\delta P_E = (\delta P_A)P_B + (\delta P_B)P_A + (\delta P_A)(\delta P_B)$$

Squaring the error equation and taking the mean value gives the mean square error where it has been assumed the errors δ_A and δP_B are independent and the mean square errors are equal to δ_A^2 and δ_B^2 respectively.

$$\delta_E^2 = \delta_A^2 P_B^2 + \delta_B^2 P_A^2 + \delta_A^2 \delta_B^2$$

The same procedure is used to determine the standard deviation for the logical OR operation when the two items of evidence are independent.

The calculations for the standard deviation are more difficult when the items of evidence have maximum dependence or minimum dependence because the operations MAXIMUM and MINIMUM can be nonlinear. If the probabilities are such that the calculations occur in a region that is far from the boundary of the MAXIMUM or MINIMUM, the nonlinear aspect of these operations can be ignored. In that case, the calculation of the standard deviation is straightforward for both maximum and minimum dependence.

When the nonlinearities are important, it is necessary to develop some *ad hoc* procedure. The *ad hoc* procedure presented here is based on calculating the standard deviation when the probabilities P_A and P_B are exactly on the boundary of MAXIMUM or MINIMUM, and then linearly interpolating to the case far from the boundary. On the boundary the square of the standard deviation is proportional to $\delta_A^2 + \delta_B^2$. Normalized variables δV and δW are defined (for when the items of evidence have minimum and maximum statistical dependence, respectively) that determine how many σ the calculations are from the boundary.

$$\text{minimum dependence:} \quad \delta V = \frac{P_B - P_A}{\left(\sigma_A^2 + \sigma_B^2\right)^{\frac{1}{2}}}$$

$$\text{maximum dependence:} \quad \delta W = \frac{P_B + P_A - 1}{\left(\sigma_A^2 + \sigma_B^2\right)^{\frac{1}{2}}}$$

If the normalized variables δV and δW are more than "$L\sigma$" from the boundary (with $L = 2$ reasonable for this example), then the nonlinearities of the boundary are neglected. The consequences of the logical AND operation and the logical OR operation on the square of the standard deviations are presented in Table 5 for the three possible assumptions about statistical dependence. The *ad hoc* quantity Q is intended to compensate for the nonlinearites.

When P_A and P_B are equal and dependence is maximum, the *ad hoc* quantity is zero, and the exact solution is obtained for the standard deviation. When the sum of P_A and P_B is equal to one and there is dependence is minimum, the *ad hoc* quantity Q is zero, and the exact solution is obtained for the standard deviation. The exact solution is also obtained when the quantities of interest are far from the nonlinearity. The particular *ad hoc* procedure presented here is not the only one possible, but it leads to a reasonable practical solution.

Multiple Reports

In the example thus far it has been assumed there is one report or message for each item of evidence. A more complicated situation arises when there are multiple reports about the same item. For this example, reports might relate to extensive preparation for installation of a missile launcher or artillery. The first report might be daytime photographs from an airplane, which show extensive preparation for installation of a missile launcher or artillery. The second report might be from Synthetic Aperture Radar [SAR] during a night overflight. The third report might be technical intelligence from long-range visual observation. All three of these reports pertain to the same item—extensive preparation for installation of a missile launcher or artillery—but the confidence in each of the reports might not be the same.

The procedure proposed here for combining multiple reports concerning the same item of evidence gives the resulting probability P as a linear combination of the probabilities P_i of the individual reports where the weighting W_i is a function of the standard deviation in the errors in the reports.

$$P = \frac{\left(\sum_i W_i P_i\right)}{W},$$
$$W = \sum_i W_i.$$

If the errors in the reports are independent, the weighting W_i equals the inverse of the square of the standard deviation σ_i,

$$W_i = \frac{1}{\sigma_i^2}.$$

In general, the errors in the individual reports will not be independent, but will be correlated. Assume there are N reports, and the convariance of the error between the i and j reports is given by R_{ij}. To obtain the desired weighting, first form the $N \times N$ square matrix $[R]$ with elements R_{ij}. Next find the $N \times N$ square matrix $[S]$ with elements S_{ij}, which is the inverse of the matrix $[R]$. Using standard least-squares estimation, it can be shown that the estimate of the resulting probability P, which has the minimum standard deviation σ, is given by the following weighting:

$$W = \sum W_i = \frac{1}{\sigma^2},$$
$$W_i = \sum_j S_{ij},$$
$$[S] = [R]^{-1}.$$

When the errors in the reports are uncorrelated, the matrix $[R]$ is diagonal, and the weighting W_i is equal to the inverse of the variance R_{ii}, which is the square of the standard deviation σ_i. Sometimes it is not convenient to keep all previous reports, so a sequential version of the weighting procedure is useful. Assume all reports up to now have been combined to form a probability P_1 with a standard deviation σ_1. The new report has a probability P_2 with a standard deviation σ_2. The errors in the combination of the previous reports and the new report have a correlation of ρ. The weighting for the sequential version is the same as the weighting when two reports are combined.

$$P = \frac{(W_1 P_1 + W_2 P_2)}{W},$$
$$W = W_1 + W_2 = \frac{1}{\sigma^2},$$
$$W_i = \frac{\left(1 - \frac{\rho\sigma_i}{\sigma_j}\right)\left(1-\rho^2\right)^{-1}}{\sigma_i^2}.$$

Once again, when the errors are uncorrelated so that ρ is zero, the weighting reduces to the inverse of the square of the standard deviation. Notice that the weighting procedure proposed for multiple reports is similar to the standard procedure for combining signal data when there are multiple measurements, so the signal information and the human estimates can be expressed in similar forms and combined in similar ways.

Conclusion

The procedures developed here for rule-based expert systems are compatible with probability used in data fusion of signals. Probabilities for evidence and hypotheses are treated, and a method is presented that allows propagation of probability through production rules when items of evidence are not independent, but have some statistical dependence. Procedures are outlined for weighting conflicting or time-varying reports, for weighting correlated data, and for propagating confidence limits through a hierarchy of production rules.

References

Adams, J. Barclay (1976) A Probability Model of Medical Reasoning and the MYCIN Model. *Mathematical Biosciences* 32: 177-186.

Barr, Avron and Edward A Feigenbaum (Eds.) (1982) *The Handbook of Artificial Intelligence.* William Kaufman, Inc.: Los Altos, California.

Ben-Bassat, Moshe (1980) Multimembership and Multiperspective Classification: Introduction, Applicationsa, and a Bayesian Model *IEEE Transactions on Systems, Man, and Cybernetics.* Vol SMC-10, No 6, 331-336.

Brown, David A and Goodman, Harvey S. (1983) Artificial Intelligence Applied to C_3I *Signal,* Vol. 37, No. 7, 27-32.

Dillard, R. A. (1983) Tactical Inferencing with the Dempster-Shafer Theory of Evidence *Proceedings of Seventeenth Asilomar Conference on Circuits, Systems, and Computers*

Duda, Richard O ,Peter E. Hart, and Nils J. Nilsson (1976) Subjective Bayesian Methods for Rule-Based Inference Systems. *Proceedings of National Computer Conference,* 1075-1082.

Duda, Richard O. and Edward H. Shortliffe (1983) Expert Systems Research. *Science,* Vol. 220, No. 4594, 261-268.

Garvey, Thomas D., Lowrance, John D., and Fischler, Martin A. (1981) An Inference Technique for Integrating Knowledge from Disparate Sources. *Proceedings of IJCAI-7,* Vancouver, B. C., Canada, 1319-325.

Gevarter, William B. (1983) Expert Systems: Limited but Powerful. *IEEE Spectrum,* Vol. 71, No. 8, 39-45

Martin-Clouaire, Roger and Henri Prade (1983) On the Problems of Representation and Propagation of Uncertainty *Second NAFIP Workship* Schenectady, N. Y., June 29-July 1, 1983.

Pecora, Vincent J , Jr. (in press, 1984) EXPRS-A Prototype Expert System using Prolog for Data Fusion, *AI Magazine,* Vol. 5 No. 2.

Rauch, Herbert E., Oscar Firschein, Walton A. Perkins, and Vicent J. Pecora (1982) An Expert System for Tactical Data Fusion. *Proceedings of Sixteenth Asilomar Conference on Circuits, Systems, and Computers,* 235-238.

Rauch, Herbert E. (1984) Expert Systems as Automated Decision Aides. *Proceedings of IEEE 1984 International Conference on Acoustics, Speech, and Signal Processing,* San Diego, California, 39B.2.1 to 39B.2.4.

Shortliffe, Edward H. and Bruce G. Buchanan (1975) A Model of Inexact Reasoning in Medicine. *Mathematical Biosciences* 23, 351-379.