FOREST C. DEAL, JR.

USING CONDITIONAL ENTROPY TO EVALUATE A CORRELATOR/TRACKER

Shipborne correlator/trackers accept reports about nearby ships and aircraft, decide how many such platforms must be present to have caused the reports, and determine where they are located. Conditional entropy is shown to be a useful concept for evaluating correlator/tracker performance.

INTRODUCTION

When a Navy ship is at sea, its captain would like to know the identities and positions of all platforms (ships and aircraft) in the vicinity. Since platforms within radar range are easy to follow, only those platforms outside radar range are considered. For each of those platforms, the captain may receive reports (from sources located elsewhere) at various times, each containing a position estimate with the parameters of a 90% confidence ellipse, and possibly information such as ship class or even ship name. Unfortunately, such a report seldom contains both specific identifying information and precise positional information, so it can be difficult to determine exactly how many platforms are present. For example, if the six reports shown in Figure 1 refer to simultaneous observations, it is difficult to tell whether four, five, or six platforms are present. With six reports, the chance that all of the associated platforms lie within their 90% confidence ellipses is only 53%. Even assuming that they do, it is possible, for example, that reports 3, 4, and 6 correspond to only one platform since all their ellipses overlap, or that they correspond to two or three distinct platforms.

Because manual correlation can be difficult, reports are fed into an automated correlator/tracker. It correlates the reports by partitioning them into sets, each of which is purported to be associated with exactly one nearby platform. It also tracks the platforms by estimating their positions at any given time, usually the present. If correlation has already been performed, tracking can proceed independently, and therefore a tracker can be evaluated as an isolated algorithm. A correlator, however, usually relies on a tracker for feedback about which partitions are the most reasonable. Therefore, a correlator cannot, in general, be evaluated independently; instead, the entire correlator/tracker must be evaluated as a unit.

The simplest and most direct way to evaluate a correlator/tracker is to test it on simulated data and calculate measures of performance (MOP's) that are functions of the correlation matrix of the result. This article describes the correlation matrix and determines how well some existing MOP's evaluate certain extreme correlation matrices. It demonstrates how conditional entropy (a concept from information theory that is a measure of the average extra information obtained by learning an additional fact about a random variable) can be used in MOP's, and compares MOP's that use this concept with the other MOP's.

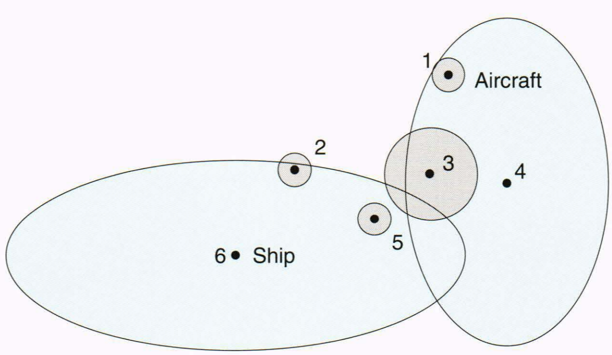

Figure 1. Confusing reports. Two or more of these reports may refer to the same platform. For example, reports 3, 4, and 6 may correspond to only one platform since all their ellipses overlap, or they may correspond to two or three distinct platforms.

CORRELATION MATRIX

The job of correlation is to place the received reports into groups that can be associated with individual platforms, usually resulting in a partition of the reports that is not exactly the same as the true partition. One particular representation of the difference between the two partitions is the correlation matrix (sometimes called the confusion matrix). An element E_{ij} of this matrix is the number of reports describing the platform associated with column j that were determined to relate to the hypothetical platform associated with row i. This is not the usual correlation matrix that is square, symmetric, and positive definite, and that quantifies statistical correlation between pairs of random variables. Instead, the columns of this correlation matrix correspond to actual platforms, and the rows correspond to hypothetical platforms; therefore, the number of rows does not necessarily equal the number of columns. Since we can ignore platforms on which no reports were made, and we would not generate a hypothetical platform having no reports associated with it, each row of the matrix contains at least one nonzero element, as does each column. Although the matrix resembles a contingency table, we are not interested in testing for statistical independence, but would like to test how far the rows deviate from being identified with the columns. Simulation is usually necessary to obtain such a matrix because for real data the true partition might itself be unknown, and the correlation matrix would therefore be incalculable.

Several definitions will prove useful. These are given in terms of the matrix elements E_{ij} and the normalized matrix elements P_{ij}, each of which represents the probability that a randomly selected report (with each report equally likely) is associated with a particular row and a particular column:

$$E_{++} = \sum_i \sum_j E_{ij} ,$$

$$E_{i+} = \sum_j E_{ij} ,$$

$$E_{+j} = \sum_i E_{ij} ,$$

$$E_{i*} = \max_j E_{ij} ,$$

$$E_{*j} = \max_i E_{ij} ,$$

$$P_{ij} = E_{ij}/E_{++} ,$$

$$P_{i+} = \sum_j P_{ij} ,$$

$$P_{+j} = \sum_i P_{ij} ,$$

$$P_{i*} = \max_j P_{ij} ,$$

$$P_{*j} = \max_i P_{ij} .$$

All these quantities are nonnegative, and only the E_{ij} and P_{ij} are possibly 0; the rest must be positive.

EXTREME CORRELATIONS AND TWO STANDARD MEASURES OF PERFORMANCE

No matter how many platforms are present or how many reports are associated with each platform, at least three extreme correlations are possible: perfection, compression, and extension. Perfection occurs when all reports are perfectly correlated; the correlation matrix is square, and each of its rows and each of its columns contain exactly one nonzero element. Compression occurs when all reports are correlated with one hypothetical platform, and therefore the correlation matrix consists of exactly one row. Extension occurs when each report is correlated with a separate hypothetical platform, and therefore the number of rows in the correlation matrix equals the number of reports. For the special circumstance where the same number of reports has been generated for each platform, confusion is also possible. In that situation, every entry of the correlation matrix has a value of 1, and the number of rows in the matrix equals the number of reports generated per platform. Figure 2 shows examples of correlation matrices corresponding to perfection, compression, extension, and confusion for the situation where three reports have been received for each of three platforms.

Two standard MOP's called the track purity (TP) and the track continuity (TC) have the following definitions:

$$TP = \sum_i P_{i*} ,$$

$$TC = \sum_j P_{*j} .$$

The idea for these MOP's comes from the concept of perfection in correlation. Since a row in the correlation matrix should have only one nonzero element, a measure of how well this requirement is met is provided by dividing the maximum element of the row by the row sum. An extension of this measure to the entire matrix is the sum of the row maxima divided by the sum of all the elements, which is simply an alternative definition for the track purity. The track purity varies between 0 and 1 and attains the value 1 if and only if each row of the correlation matrix contains exactly one nonzero element. The track continuity is obtained in the same manner by working with the columns of the matrix instead of the rows.

Table 1 shows the track purity, the track continuity, and their geometric mean for four extreme correlations where there are N platforms and R reports per platform. Whereas TP and TC individually may reward imperfect correlation with the maximum score of 1.0, their geometric mean reaches this value if and only if each row and each column contain exactly one nonzero element, thus indi-

$$\boxed{\begin{matrix} 0 & 0 & 3 \\ 3 & 0 & 0 \\ 0 & 3 & 0 \end{matrix}}$$

Perfection

$$\boxed{\begin{matrix} 3 & 3 & 3 \end{matrix}}$$

Compression

$$\boxed{\begin{matrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{matrix}}$$

Extension

$$\boxed{\begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix}}$$

Confusion

Figure 2. Correlation matrices showing perfection, compression, extension, and confusion for the situation where three reports have been received for each of three platforms. Each column corresponds to an actual platform, and each row corresponds to a hypothetical platform.

Table 1. Measures of performance for four extreme correlations.

Measure	Perfection	Compression	Extension	Confusion
TP	1	$1/N$	1	$1/N$
TC	1	1	$1/R$	$1/R$
$\sqrt{(TP)(TC)}$	1	$1/\sqrt{N}$	$1/\sqrt{R}$	$1/\sqrt{NR}$

Note: TP = track purity, TC = track continuity, N = number of platforms, R = reports per platform.

cating the existence of a one-to-one identification between rows and columns, and therefore signifying that the correlation is indeed perfect.

CONDITIONAL ENTROPY

The concepts of entropy, joint entropy, and conditional entropy (all from information theory) have been in use for more than forty years.[1] We can apply them in evaluating a correlator/tracker by selecting a particular correlation matrix and considering a random variable, *Report*, that takes on the identity of one of the reports considered in the correlation matrix, with equal probabilities. Then, by using two functions, row( ) and column( ), we create two additional random variables, *Row* and *Column*:

$$Row = \mathrm{row}(Report) \;,$$
$$Column = \mathrm{column}(Report) \;.$$

Now we calculate their individual, joint, and conditional entropies. The entropy, H, of a random variable that takes on only a finite number, n, of possible values with probabilities $P_1, P_2, \ldots, P_n$ is given by

$$H = \sum_{k=1}^{n} (-P_k \ln P_k) \,,$$

where 0 ln 0 is taken by convention to be 0. Therefore, the individual and joint entropies of *Row* and *Column* are given by

$$H(Row) = \sum_i (-P_{i+} \ln P_{i+}) \;,$$
$$H(Column) = \sum_j (-P_{+j} \ln P_{+j}) \;,$$
$$H(Column,\, Row) = \sum_{ij} (-P_{ij} \ln P_{ij}) \;.$$

Each conditional entropy is simply the difference between a joint entropy and an individual entropy:

$$\begin{aligned} H(Row \mid Column) &= H(Column,\, Row) - H(Column) \\ &= \sum_{ij} (-P_{ij} \ln P_{ij}) \\ &\quad - \sum_j (-P_{+j} \ln P_{+j}) \;, \end{aligned}$$

$$\begin{aligned} H(Column \mid Row) &= H(Column,\, Row) - H(Row) \\ &= \sum_{ij} (-P_{ij} \ln P_{ij}) \\ &\quad - \sum_i (-P_{i+} \ln P_{i+}) \;. \end{aligned}$$

Finally, the average of the conditional entropies is called the average conditional entropy (ACE):

$$\begin{aligned} ACE(Row,\, Column) = {}& \tfrac{1}{2}[H(Row \mid Column) \\ & + H(Column \mid Row)] \;. \end{aligned}$$

How are these conditional entropies physically realizable? Suppose that someone randomly (with equal probabilities) and repeatedly selects reports, and we must send messages to someone else, indicating the row and column numbers of each report. Being clever, we encode the messages judiciously on the basis of the various probabilities, and minimize the average length of message required to describe each report. The average message length measures the joint entropy of *Row* and *Column*. Similarly, if we only needed to report the column, we would judiciously select a different code, and its average message length would measure the entropy of *Column*. The difference between these average message lengths is the conditional entropy of *Row* given *Column*, a measure of how much extra information is provided by the identity of the row.

CONDITIONAL ENTROPY MEASURES OF PERFORMANCE

The track purity and track continuity MOP's mentioned previously measure goodness of correlation and range between 0 and 1, where 1 is the most favorable value. The three different types of entropy measures given previously measure error in correlation and are nonnegative, where 0 is the most favorable value. To convert the entropy measures into the more standard form, we simply take the exponential of their negative; that is, we define the informational purity (IP), the informational continuity (IC), and the fidelity (F) as follows:

$$IP = \exp[-H(Column \mid Row)] \;,$$
$$IC = \exp[-H(Row \mid Column)] \;,$$
$$F = \exp[-ACE(Row,\, Column)] \;.$$

The following inequalities are proved in the boxed insert:

$$IP \le TP \;,$$
$$IC \le TC \;.$$

Since $F = \sqrt{(IP)(IC)}$, it follows that

$$F \le \sqrt{(TP)(TC)} \;.$$

To see that equality may hold in the previous three equations, one can simply look at the four extreme correlations given earlier.

PROOF THAT $IP \le TP$ AND $IC \le TC$

The proof of the theorem is simplified by the following lemma:

Lemma

If for $i = 1, \ldots, n$ we have $0 \le x_i$, and $0 \le c_i \le 1$ along with $\sum_{i=1}^{n} c_i = 1$, and we define $0^0 = 1$, then

$$\prod_{i=1}^{n} x_i^{c_i} \le \sum_{i=1}^{n} c_i x_i \ .$$

Proof

Without loss of generality, we may assume that all the $x_i > 0$, for if $x_j = 0$ and $c_j > 0$, the left side of the inequality is zero and the result is trivial; whereas if $x_j = 0$ and $c_j = 0$, the jth factor on the left is 1 and the jth term on the right is 0, and we need only to prove the result for the remaining $n - 1$ pairs of c and x. Similarly, we may assume that all $c_i > 0$, for if some $c_j = 0$, again the jth factor on the left is 1 and the jth term on the right is 0, and we need only to prove the result for the remaining $n - 1$ pairs of c and x. When all values are positive, the result follows from Jensen's inequality.[2] That theorem states that any particular convex function of the expected value of a random variable cannot be greater than the expected value of that particular convex function of the random variable. If we choose the exponential function as the convex function and let the random variable take on the values $y_1, \ldots, y_n$ with positive probabilities $c_1, \ldots, c_n$, then Jensen's inequality indicates that

$$\sum_{i=1}^{n} \exp(c_i y_i) \le \sum_{i=1}^{n} c_i \exp(y_i) \ .$$

The lemma follows by identifying x_i with $\exp(y_i)$ for $i = 1, \ldots, n$.

Theorem

$$IP \le TP \quad \text{and} \quad IC \le TC \ .$$

Proof

$$\begin{aligned} IP &= \exp\left\{ -\left[\sum_{ij} (-P_{ij} \ln P_{ij}) - \sum_{i} (-P_{i+} \ln P_{i+}) \right] \right\} \\ &= \left(\prod_{ij} P_{ij}{}^{P_{ij}} \right) \left(\prod_{i} P_{i+}{}^{P_{i+}} \right)^{-1} \\ &\le \left(\prod_{ij} P_{i*}{}^{P_{ij}} \right) \left(\prod_{i} P_{i+}{}^{P_{i+}} \right)^{-1} \\ &\le \left(\prod_{i} P_{i*}{}^{P_{i+}} \right) \left(\prod_{i} P_{i+}{}^{P_{i+}} \right)^{-1} \\ &\le \prod_{i} \left(\frac{P_{i*}}{P_{i+}} \right)^{P_{i+}} \\ &\le \sum_{i} P_{i*} \\ &= TP \ , \end{aligned}$$

where the last inequality is due to the lemma. The proof that $IC \le TC$ is similar.

RATIONALE FOR CONDITIONAL ENTROPY MEASURES OF PERFORMANCE

Why use the conditional entropy MOP's, when they are similar to the corresponding standard MOP's and yet more complicated? First, the standard MOP's concern themselves only with the row and column maxima, and are not functions of smaller values that may be related to the vast majority of the reports. For each conditional entropy MOP, every value is used, and minor changes in the correlation matrix can affect the value of the MOP. Second, evaluating the correlation is equivalent to quantifying the similarity between two different ways of partitioning the same objects (i.e., partitioning reports into platforms). Therefore, we should be able to determine how similar two different correlator/trackers are in their correlation of a batch of reports by using an MOP on a matrix constructed in a manner similar to the correlation matrix. If their partitions are close together, resulting in an MOP value near 1.0, we would expect that the MOP values obtained by comparing the two partitions with the true partition should be nearly the same. This similarity does not necessarily occur when the geometric mean of TP and TC is used, but it always occurs if the fidelity is used. If the partitions are close together, the fidelity is near 1.0 and ACE is near 0. Because ACE has been shown to be a metric, which by definition obeys the triangle inequality,[3] the ACE values obtained by comparing the two partitions with the true partition must be nearly the same. Since the mapping from ACE to the fidelity is continuous and monotonic and has a bounded derivative, the fidelity values must also be nearly the same.

COMPUTATIONAL FORMS

For computational ease, the following forms can be used:

$$H(Column \mid Row) = \frac{\sum_{i} (E_{i+} \ln E_{i+}) - \sum_{ij} (E_{ij} \ln E_{ij})}{E_{++}} \ ,$$

$$H(Row \mid Column) = \frac{\sum_{j} (E_{+j} \ln E_{+j}) - \sum_{ij} (E_{ij} \ln E_{ij})}{E_{++}} \ ,$$

$$ACE(Row,\ Column) = \frac{H(Column \mid Row) + H(Row \mid Column)}{2} \ .$$

PRACTICAL CONSIDERATIONS

The preceding discussion assumes that each incoming report is associated with a specific platform, and that each is processed by the correlator/tracker and associated with a particular hypothetical platform. This assumption is not always true. For example, a sensor can produce a false alarm and report the presence of a platform when nothing is there. Since we are working with a simulation, however, this condition is easily avoided by asserting that the input data of the simulation will contain no false alarms.

Another option is to allow reports that are false alarms, but to consider each one as corresponding to its own unique platform. In this way, each report is the only report in its column of the correlation matrix. Other problems are not so easily handled. Although a correlator/tracker could receive a report and completely disregard it, we simply ignore this possibility. Instead of disregarding a report, a correlator/tracker could indicate that it cannot determine to which hypothetical platform the report belongs, thereby declaring it an ambiguity.

If ambiguities are allowed, then no good way to modify the MOP of fidelity is apparent; further research in this area would be useful. One way to handle ambiguities is to associate all of them with one additional hypothetical platform row, evaluating the fidelity as before. This method makes sense from a point of view based on game theory, because if the punishment for creating ambiguities were too harsh, the correlator/tracker itself could be programmed to append this new row and obtain a better MOP value. This option, however, makes the triangle inequality cease to hold for ACE; thus, any measure of how close two correlator/trackers came to each other would have no meaning. Another option is to calculate the fidelity for the correlation matrix without the reports that led to ambiguities, and then to multiply the result by the fraction of reports that did not become ambiguities. This method, however, produces a fidelity of 0 if every report becomes an ambiguity, a result worse than that for any other possible correlation. In particular, the correlator/tracker would do better to declare all the ambiguities to belong to one hypothetical platform. Other options for modifying the fidelity have been investigated, but none seems acceptable in all situations.

WEIGHTING OF REPORTS

In all the preceding results, the reports were weighted equally. If we give weights to the reports (perhaps because the platforms associated with them are of different interest), then we can redefine E_{ij} to be the sum of the weights of the associated reports instead of simply the number of associated reports. The rest of the results remain valid as stated.

CONCLUSION

To evaluate a correlator/tracker by means of a correlation matrix, we have produced a measure of performance, the fidelity, that is the exponential of the negative of a metric. It can be used to compare the results of correlator/trackers with each other, as well as with perfection. It incorporates the concept of conditional entropy, is sensitive to all elements of the correlation matrix, and is bounded above by a function of standard measures of performance.

REFERENCES

[1] Shannon, C. E., "A Mathematical Theory of Communication," *Bell Sys. Tech. J.* **27**, 379–423 (Part I) and 623–656 (Part II) (1948).

[2] Rudin, W., *Real and Complex Analysis*, 2nd Ed., McGraw–Hill, New York, p. 64 (1974).

[3] Denker, M., Grillenberger, C., and Sigmund, K., *Ergodic Theory on Compact Spaces*, Springer–Verlag, New York, p. 63 (1976).

THE AUTHOR

FOREST C. DEAL, Jr., received a B.S. degree in mathematics and physics from Rollins College. After a year studying applied mathematics at Brown University, he enlisted in the U.S. Navy and became an officer. He obtained M.A. and Ph.D. degrees in mathematical sciences from The Johns Hopkins University and joined APL in 1977, where he worked in the Strategic Systems Department on multivariate estimation, ensemble statistical analysis, observability and aliasing, sample sizing, and reliability. Since 1985, Dr. Deal has worked in the Fleet Systems Department on sensor and communication systems modeling, algorithm development, and measures of performance. He is a member of the Institute of Mathematical Statistics and the American Statistical Association.