The application of Dempster-Shafer theory of evidence to intelligence analysis

S G Coulson & R May

RED Scientific Limited, 1 Oriel Court, Omega Park,

Alton, Hampshire GU34 2YT

Abstract

The Dempster-Shafer theory of evidence has been used since the 1970s to model decision making under uncertainty. Recent applications of the Dempster-Shafer theory and its extensions have focused on Artificial Intelligence and target identification. Here we examine the application of Dempster-Shafer theory to intelligence analysis.

Dempster-Shafer theory offers a number of advantages over traditional statistical methods of intelligence analysis; in particular, by assigning a quantitative measure of uncertainty to overlapping subsets of propositions or hypothesis it avoids the Bayesian dilemma of having to assume prior probabilities. Rather than consider the probability of a hypothesis being correct, the Dempster-Shafer framework utilises a belief *(Bel)* function to measure the confidence that an event lies within a given region and not in any proper subset of that region.

In this paper we develop a matrix methodology to calculate belief functions to analyse intelligence reports. We compare the use of the Dempster-Shafer framework with a Bayesian approach to describe an intelligence analyst's prior knowledge. We consider two special cases, first where there is complete information and secondly where we only have sufficient information to assign a probability to one hypothesis, under these circumstances the Dempster-Shafer framework reduces to a traditional probability problem.

1. Introduction

Intelligence analysis, understanding what has happened and intelligence assessment, making predictions on what will happen both deal in uncertainty. The role of an intelligence analysis is to develop hypotheses and assign probabilities to them (Fisk 1972). What makes the task of an intelligence analyst different from that of classical decision analysis is that the information presented to the analyst is often incomplete and may be deliberately misleading. Any model of intelligence analysis must take account of these challenges.

Some degree of error or uncertainty is inherent in both intelligence analysis and assessment. Broadly speaking, there are two main types of uncertainty, errors from misleading or inaccurate information and unknown outcomes. The first of these affects intelligence analysis and the second intelligence assessment. Here we are interested in both types of uncertainty and will use the term intelligence analysis to refer to analysis (understanding) and assessment (prediction) processes.

Within intelligence analysis it is rare that probabilities can be rigorously assigned to a hypothesis using an objective measure of event space. Even the most simple of possible outcomes can be complicated by unexpected actions occurring. As an example, a common question for strategic intelligence is to ask what are the chances that a given country will go to war against its neighbours, traditionally an analyst would assign a probability to the hypothesis that war will occur and the complement of that probability to the hypothesis that war does not occur. This approach can be negated if the bellicose country takes an unexpected course of action, such as hybrid warfare or war by proxy fighters, which enables it to deploy military force without committing an overt act of war.

The alternative to assigning probabilities to a given event space is to adopt a frequentist approach, where the probability that an outcome A occurs $P(A)$ is determined by observing the number of times $N(A)$ that A occurs during a number n of events. Formally, $P(A) = \lim_{n \to \infty} \frac{N(A)}{n}$, this actuarial approach to defining probabilities relies on there being a sufficiently large number of events to determine the convergence of $\frac{N(A)}{n}$. In practice, there are often very few repeatable events to enable frequentist probabilities to be assigned in intelligence analysis.

Within intelligence analysis, the assignment of probabilities to hypotheses is usually a subjective process based on the judgement of either a single analyst or a team of analysts. The probability of an outcome is usually expressed qualitatively rather than quantitatively, for example an outcome may be described as likely or highly unlikely. The risk of individual and organisational bias influencing subjective probabilities has long been recognised (Kam 1988). Uncertainty Yardsticks are often used within the intelligence community to provide a consistent methodology for equating qualitative judgements with quantitative probabilities (Omand 2010).

The use of subjective probabilities in intelligence can also cause problems when trying to combine judgment-based probabilities with quantitative analysis based on frequentist or objective probabilities, such as forensic techniques e.g. DNA fingerprinting. There is a risk that intelligence derived from such sources may unduly influence the mind of an intelligence analyst.

The need for a rigorous methodology to combine information from different intelligence reports in a consistent manner is of particular interest to the intelligence community. Since the 1960s, the intelligence analysts have considered the use of Bayesian techniques to provide a framework for intelligence analysis (CIA 1967). Recently, the US intelligence community has been mandated to develop rigorous techniques to improve intelligence analysis1including the use of Bayesian techniques.

The Bayesian approach is described formally in Section 3, but for now it is sufficient to say that it relies on the assignment of a prior probability that a given event may occur. This has attracted criticism when Bayesian techniques are applied to subjective probabilities on the grounds that it conveys an impression of precision to a highly uncertain probability (Walley 1987). As discussed above, any application of Bayesian techniques to intelligence analysis will often involve highly subjective measures of probability; however, one advantage of a framework that requires an analyst to provide a prior probability of an event is that this provides a record of the analyst's initial assumptions or biases.

A generalisation of the Bayesian approach is provided by the Dempster Shafer theory. In a series of papers Dempster (1966, 1967, 1969) expanded the Bayesian approach to treat unknown parameters and his work was further developed by Shafer (1976). The Dempster-Shafer theory has found numerous applications in Artificial Intelligence (AI) (Laskey & Leehner 1989), target identification (Buede & Girardi1997) and decision making (Beynon, Curry & Morgan 2000).

In this paper we develop a simple model for intelligence analysis, in Section 2,and define some basic mathematical properties that a practical intelligence analysis model must obey. In Section 3 we review the Bayesian approach to intelligence analysis in terms of our model and demonstrate that it is consistent with our defined properties.

In Section 4 we formally introduce the Dempster-Shafer theory using a similar representation to Beynon et al. (2000) and apply the approach to our intelligence model. This is the first time, we are aware of, that the Dempster-Shafer theory has been applied to intelligence analysis. In Section 4 we develop a matrix approach to simplify Dempster-Schafer calculations and to demonstrate that the Dempster-Shafer theory satisfies our intelligence model requirements.

[1] U.S., Intelligence Reform and Terrorism Prevention Act of 2004, Section 1017 a, http://www.nctc.gov/docs/pl108_458.pdf

2. An intelligence analysis model

We adopt a simple model for intelligence analysis whereby an analyst starts with an initial hypothesis H that a given assertion is true. The analyst receives a series of $n \in \mathbb{N}$ intelligence reports E = {E1, E2, ……, En}, assumed to be distinct and independent.

The analyst assigns a probability that *H* is true, $\mathrm{P}(\mathrm{H}|\mathrm{E}_0), \mathrm{P} \in [0,1]$, that is in the absence of any evidence, the prior probability that the hypothesis is correct. This initial probability can be thought of a measure of the analyst's preconception that the hypothesis is correct.

Each intelligence report received allows the analyst to revise the probability associated with *H*. So after analysing the first intelligence report E_1, the probability that *H* is true can be written as $\mathrm{P}(\mathrm{H}|\mathrm{E}_0, \mathrm{E}_1)$. After analysing the second intelligence report in the series, the probability changes to $\mathrm{P}(\mathrm{H}|\mathrm{E}_0, \mathrm{E}_1, \mathrm{E}_2)$; after analysing the i-th intelligence report the probability becomes $\mathrm{P}(\mathrm{H}|\mathrm{E}_0, \mathrm{E}_1, \mathrm{E}_2, \cdots, \mathrm{E}_\mathrm{i})$.

Here we are concerned with the operation required to update the probabilities given each subsequent intelligence report

$$\mathrm{P}(\mathrm{H}|\mathrm{E}_0, \mathrm{E}_1, \cdots, \mathrm{E}_\mathrm{i}) \rightarrow \mathrm{P}(\mathrm{H}|\mathrm{E}_0, \mathrm{E}_1, \cdots, \mathrm{E}_\mathrm{i}, \mathrm{E}_{\mathrm{i}+1}).$$

Before we look at specific approaches we can make some general remarks on the properties that we require:

Remark 2.1. *The order in which intelligence reports are received should not affect the final probability*

$$P\big(H|E_{0,} \cdots, E_i, E_j\big) = P\big(H|E_{0,} \cdots, E_j, E_i\big), \qquad i \neq j,$$

Remark 2.2. *Similarly, a binary operator ∘ combining subsets of E should be commutative*

$$P\big(H|E_{0,} \cdots, E_i\big) \circ P\big(H|E_{0,} \cdots, E_j\big) = P\big(H|E_{0,} \cdots, E_j\big) \circ P\big(H|E_{0,} \cdots, E_i\big)$$

If the event space $\mathcal{E}$ containing the hypothesis *H* can be completely partitioned by a number of alternative hypotheses $\mathcal{E} = \{H, H^1, H^2, \cdots, H^m\}$, then, by the law of total probability, the probability that the total set of intelligence reports *E* received is true is given by

$$P(E) = \sum_i P(E|H^i)P(H^i)$$

where $P(H^i)$ is the probability that the i-th hypothesis, H^i is true.

3. Bayesian Intelligence Analysis

Using the intelligence model developed above, the probability that a given hypothesis *H* is correct by application of Bayes' theorem is given by

$$P\big(H|E_{0,} \cdots, E_i\big) = \frac{P\big(E_{0,} \cdots, E_i|H\big)P(H)}{\sum_j P\big(E_{0,} \cdots, E_i|H^j\big)P(H^j)} \tag{3.1}$$

where the H_js partition the set of all possible hypotheses from the previous section. This is a generalisation of the form presented in most discussions on applying Bayesian theory to intelligence analysis which assume that the hypothesis set $H = \{H, H^c\}$, i.e. the hypothesis and its complement.

If an additional intelligence report E_{i+1} is received, equation 3.1 can be updated to give a revised probability $P\big(H|E_{0,} \cdots, E_i, E_{i+1}\big)$ that the hypothesis is true.

$$P(H|E_{0,}\cdots,E_i,E_{i+1}) = \frac{P(E_{0,}\cdots,E_i,E_{i+1}|H)P(H|E_{0,}\cdots,E_i)}{\sum_j P(E_{0,}\cdots,E_i,E_{i+1}|H^j)P(H^j|E_{0,}\cdots,E_i)} \tag{3.2}$$

As an example of the application of Bayesian analysis in intelligence assessment, consider the following problem from operational intelligence. Blueland must defend land consisting of three valleys (A, B & C) lying parallel to each other running north and south, with valley A at the far east of Blueland, B in the middle and C to the west. The high ground between each valley cannot be traversed and the length of the valleys is such that any troops located within one valley would be unable to deploy rapidly into one of the other valleys. Blueland's vital ground lies at the southern end of the three valleys and they must prevent an incursion by enemy forces from Redland located to the north of the valleys. Blueland have insufficient forces to adequately defend all three valleys, the Blueland commander directs his intelligence staff to assess which valleys are Redland most likely to attack through.

On analysis of the terrain and topography of each valley, Blueland intelligence conclude that there is an equal probability that Redland could attack down A, B and C. So that the probability associated with the hypothesis $H \equiv Redland\ attack\ vally\ A$ is $P(H) = \frac{1}{3}$.

Suppose that Blueland receive imagery intelligence (IMINT) E_1 indicating that Redland forces are manoeuvring to attack valley A. Based on previous reporting, Blueland assess that there is an 4/5 probability that this intelligence is correct; i.e. the conditional probability $P(E_1|H) = \frac{4}{5}$.

Using the law of total probabilities, the probability that E_1 is correct is

$$P(E_1) = P(E_1|H)P(H) + P(E_1|H^*)\mathrm{P}(H^*) = \left(\frac{4}{5}.\frac{1}{3}\right) + \left(\frac{1}{5}.\frac{2}{3}\right) = \frac{6}{15}.$$

where H^* is the complement of H, i.e. Redland attack down valley B or C.

From 3.2, the conditional probability that H is correct given E_1 is $P(H|E_1) = \frac{2}{3}$.

So despite the reasonably high accuracy of the intelligence source, the probability that H is correct has only increased by a factor of two.

Now, suppose that the second intelligence report received, E_2, is an agent report (HUMINT) that, once the reliability of the source is taken into account, suggests that there is only a probability of 1/10 that Redland will attack down valley A. Using 3 to combine this new intelligence, we find that $P(H|E_1, E_2) = \frac{2}{11}$, that is the latest reporting has deceased Blueland's assessment that valley A will be the route of Redland's attack. The decrease in the probability is lower than the prior probability of 1/3.

It is straightforward to verify that this Bayesian model satisfies the condition $P(H|E_1, E_2) = P(H|E_2, E_1)$ required by Remarks 2.1 and 2.2, a formal proof is given in the Appendix, which also gives a justification for the revised probability formula in Equation 3.2. In reality, the order in which E_1 and E_2, are received may have an undue influence on the minds of the analysts, this will be explored later on.

In the above example, assigning a prior probability that valley A would be attacked was relatively straightforward, in real-life possible enemy courses of action are usually assessed in terms of *most likely* and *most dangerous*, each with a usually highly subjective associated probability.

Both of the intelligence reports received, E_1 and E_2, referred unambiguously to the subject of the hypothesis, in the majority of cases, the relationship between an intelligence report and a hypothesis is not so explicit but has to be inferred by the analyst. Suppose that report E_1 instead of indicating that Redland were preparing to attack valley A, now indicates that Redland forces are moving west of valley C. The immediate reaction is to conclude that the probability that H is correct is now $P(H) = \frac{1}{2}$; however, noting that if H is given, then the probability that E_1 is correct, $P(E_1|H) = \frac{1}{2}$, since there is an equal chance that valley B could be the direction of Redland's attack.

Using the law of total probabilities, the probability that E_1 is correct is

$$P(E_1) = P(E_1|H)P(H) + P(E_1|H^*)P(H^*) = \left(\frac{1}{2}.\frac{1}{3}\right) + \left(\frac{1}{2}.\frac{2}{3}\right) = \frac{1}{2}$$

where H^* is the complement of H, i.e. Redland attack down valley B or C.

Applying Bayes theorem, the conditional probability that H is correct given E_1 is $P(H|E_1) = \frac{1}{3}$, i.e. receipt of report E_1 has left the probability that valley A will be attacked unchanged from the prior probability.[2]
One of the most difficult challenges for intelligence analysts using a Bayesian approach is to assign prior probabilities to their hypotheses, the Dempster-Shafer approach provides an alternative framework that avoids the need to assign priors.

4. Dempster-Shafer intelligence analysis

The Dempster-Shafer approach considers a *frame of discernment*, a finite set of hypotheses $\mathcal{H} = \{h^1, h^2, \cdots h^n\}$ that completely partition the space of all possible hypotheses $\mathcal{H}$. Such that $h^i \cap h^j = \emptyset, i \neq j$ and $\bigcup_j h^j = \mathcal{H}$. Unlike the Bayesian approach which calculates the probability that a set of evidence $E = \{E_1, E_2, \ldots\ldots, E_n\}$ supports a single hypothesis $h^i \epsilon\ H$, the Dempster-Shafer approach calculates the degree to which the evidence supports a subset of outcomes $\mathcal{H}' \subseteq \mathcal{H}$.

For a finite set of hypothesis $\mathcal{H} = \{h^1, h^2, \cdots h^n\}$, the number of subsets that completely partition $\mathcal{H}$ is 2^n including the empty set $\{\emptyset\}$. The Dempster-Schafer Theory defines a *basic probability assignment (bpa)*, where $m: 2^n \to [0,1]$, whenever $m(\emptyset) = 0$ and $\sum_a (m(a)|a \in 2^n) = 1$.

The *bpa* $m(H')$ is a measure of the confidence that the true hypothesis lies within H' and not in any proper subset of H'. The empty set corresponds to zero confidence, that is the rejection of all hypotheses within the subset $\mathcal{H}' \subset \mathcal{H}$.

Derived from the *bpa* are two measures of confidence belief (*Bel*) and plausibility (*Pl*), defined by

$$Bel: 2^n \to [0,1]$$

$$Bel(A) \equiv \sum_{B \subseteq A} m(B), \qquad \forall A \subseteq \mathcal{H} \qquad (4.1)$$

and,

$$Pl: 2^n \to [0,1]$$

$$Pl(A) \equiv \sum_{B \cap A \neq \emptyset} m(B), \qquad \forall A \subseteq \mathcal{H} \qquad (4.2)$$

Where A and B are two non-empty subsets of $\mathcal{H}$, such that $B \subset A \subset \mathcal{H}$.

The belief function is a measure of the confidence that a hypothesis lies within a set A or any subset of A; while, the plausibility measures the confidence that a given hypothesis is not disbelieved.

If we define the complement or *doubt* in A, A^*, to consist of the set of all hypotheses $h^i \notin A, h^i \in \mathcal{H}$. Then it can be shown that:

$$Bel(A) = 1 - Pl(A^*)$$

and,

$$Pl(A) = 1 - Bel(A^*).$$

[2] This is similar (but subtly different) to the *Monty Hall* problem (Selvin 1975)

The Dempster-Shafer theory allows intelligence reports to be combined to update values of *Bel* and *Pl*, in a manner analogous to equation 3.2 in the Bayesian approach.

Consider two independent intelligence reports E_1 and E_2 with *bpa* m_1 and m_2, respectively. Then the combined *bpa* is defined as:

$$[m_1 \oplus m_2](a) = \begin{cases} \dfrac{\sum_{A\cap B=a} m_1(A)m_2(B)}{1-\sum_{A\cap B=\emptyset} m_1(A)m_2(B)}, & a \neq \emptyset \\ 0, & a = \emptyset \end{cases} \tag{4.3}$$

In contrast to Bayes' theorem, the Dempster-Shafer combination rule is not uniquely determined from the axioms of probability (Denoeux1995), this implies that other combination rules are possible. In particular, the normalisation factor $1-\sum_{A\cap B=\emptyset} m_1(A)m_2(B)$ may lead to unexpected results if

$$1-\sum_{A\cap B=\emptyset} m_1(A)m_2(B) \approx 1$$

(Dubois & Prade 1988).

Applying the Dempster-Shafer combination rule to our earlier example of the three-valley problem immediately illustrates some of the key differences between Dempster-Shafer and the Bayes approach to intelligence analysis. The first intelligence report received, E_1 indicating that Redland forces are manoeuvring to attack valley A. As before, Blueland assess that there is a 4/5 probability that this intelligence is correct; using the Dempster-Shafer approach, this implies that we can assign the following *bpa*s based on E_1

$$m_1\{A\} = \frac{4}{5}, \qquad m_1\{A,B,C\} = \frac{1}{5}$$

where the set $\{A,B,C\}$ is the set off all other possible hypothesis including *{A}* itself, as opposed to the Bayesian approach where the complement *{A*}* would be assigned the remaining probability to *{B,C}*. Under the Dempster-Shafer approach we avoid being forced to make assumptions concerning unavailable probabilities and having to assign prior probabilities.

As before, a second intelligence report, E_2, is received suggesting that there is only a probability of 1/10 that Redland will attack down valley A. That is

$$m_2\{A\} = \frac{1}{10}, \qquad m_2\{A,B,C\} = \frac{9}{10}$$

Using the combination rule given by 4.3 we can assign combined *bpa* $m_3 = m_1 \oplus m_2$ by first noting that $\sum_{A\cap B=\emptyset} m_1(A)m_2(B) = 0$, since the only subsets of $\mathcal{H} = \{A,B,C\}$ that have non-zero *bpas* are $\{A\}$ and $\{A,B,C\}$. Hence

	$m_2\{A\} = \frac{1}{10}$	$m_2\{A,B,C\} = \frac{9}{10}$
$m_1\{A\} = \frac{4}{5}$	$m_3\{A\} = \frac{2}{25}$	$m_3\{A\} = \frac{18}{25}$
$m_1\{A,B,C\} = \frac{1}{5}$	$m_3\{A\} = \frac{1}{50}$	$m_3\{A,B,C\} = \frac{9}{50}$

From the definitions in 4.1 and 4.2 we can evaluate the belief function $Bel(A) = \frac{41}{50}$ and the plausibility function $PL(A) = \frac{41}{50}$; similarly, $Bel(A, B, C) = \frac{9}{50} = Pl(A, B, C)$. We note that when the non-zero subsets of hypothesis are singletons, then $Bel(a) = Pl(a)$ and the *bpa* are the equivalent of probability distributions.

In this example, the belief that valley A is the target of Redland's attack from combining the intelligence in reports E_1 and E_2 is $\frac{41}{50}$, numerically this is much greater than the $\frac{2}{11}$ from Bayesian analysis; however, it should be remembered that the *bpa* assigned in this example are not the same as in the Bayesian example. If we assume that the first intelligence report received E_0 is the same as our prior probability from the Bayesian example, then

$$m_0\{A\} = \frac{1}{3}, m_0\{B, C\} = \frac{2}{3}, m_1\{A\} = \frac{4}{5}, m_1\{B, C\} = \frac{1}{5}$$

using the combination rule we obtain the result $Bel(m_0 \oplus m_1\{A\}) = Pl(m_0 \oplus m_1\{A\}) = P(H|E_1) = \frac{2}{3}$.

One of the drawbacks in applying the Dempster-Shafter theorem to solving problems in decision making is that the combination rule can be difficult to calculate, in the next section we develop a matrix approach to evaluate the combined *bpas*.

5. A matrix approach to Dempster-Shafer theory

Consider a non-empty *frame of discernment*, $\mathcal{H} = \{h^1, h^2, \cdots h^n\}$, if we ignore the empty set $\{\emptyset\}$, $\mathcal{H}$ has $2^n - 1$ partitions. Further, consider a function Γ that assigns a $bpa \in [0,1]$ to each hypothesis $h^i \in \mathcal{H}$, we define **M** as the $2^n - 1$ dimensional vector of Γ acting on $\mathcal{H}$. Such that

$$\boldsymbol{M} = \begin{pmatrix} m(h^1) \\ m(h^2) \\ \vdots \\ \vdots \\ m(h^n) \end{pmatrix}$$

For a given $\mathcal{H}$, suppose there exist two *bpa* assigning functions Γ_1 and Γ_2 that generate the vectors $\boldsymbol{M_1}$ and $\boldsymbol{M_2}$ respectively. We define $\boldsymbol{M_3}$ as the $(2^n - 1, 2^n - 1)$ matrix obtained from the product $\boldsymbol{M}_1^T \boldsymbol{M}_2$**:**

$$\boldsymbol{M_3} = \boldsymbol{M}_1^T \boldsymbol{M}_2 = \begin{pmatrix} m_1(h^1)m_2(h^1) & \cdots & m_1(h^1)m_2(h^n) \\ m_1(h^2)m_2(h^1) & \cdots & m_1(h^2)m_2(h^n) \\ \vdots & & \vdots \\ & \ddots & \\ m_1(h^n)m_2(h^1) & \cdots & m_1(h^n)m_2(h^n) \end{pmatrix}$$

We can decompose $\boldsymbol{M_3}$ into the sum of two $(2^n - 1, 2^n - 1)$ matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$, where the elements of $\boldsymbol{X}, x_{ij}$ are defined by

$$x_{ij} \equiv \begin{cases} m_1(h^i)m_2(h^j), & h^i \cap h^j = \emptyset \\ 0, & otherwise \end{cases}$$

and the elements of $\boldsymbol{Y}, y_{ij}$ are defined by

$$y_{ij} \equiv \begin{cases} m_1(h^i)m_2(h^j), & h^i \cap h^j \neq \emptyset \\ 0, & otherwise \end{cases}$$

COROLLARY 1. *For an n dimensional hypothesis set the number of non-zero elements of* $\boldsymbol{X}$ *is*

$$\sum_{r=1}^{n} \binom{n}{r} \left(2^{(n-r)} - 1\right)$$

where $\binom{n}{r}$ *denotes the binomial coefficient.*

COROLLARY 2. *For an n dimensional hypothesis set the number of non-zero elements of* $\boldsymbol{Y}$ *is*

$$(2^n - 1)^2 - \sum_{r=1}^{n} \binom{n}{r} \left(2^{(n-r)} - 1\right)$$

The denominator in the Dempster-Shafer combination rule can be written in terms of the elements of $\boldsymbol{X}$

$$1 - \sum_{h^i \cap h^j = \emptyset} m_1(h^i) m_2(h^j) = 1 - \sum_{j=1}^{N} \sum_{i=1}^{N} x_{ij} \tag{5.1}$$

for $N = 2^n - 1$

Similarly, the combined bpa, m_3 is given by

$$[m_1 \oplus m_2](a) = \frac{\sum_{h^i \cap h^j = a}^{N} y_{ij}}{1 - \sum_{j=1}^{N} \sum_{i=1}^{N} x_{ij}} \tag{5.2}$$

THEOREM 1. *The Dempster-Shafer combination rule satisfies REMARK 2.2*

$$[m_1 \oplus m_2](a) = [m_2 \oplus m_1](a)$$

PROOF

Let $\boldsymbol{M}_3 = \boldsymbol{M}_1^T \boldsymbol{M}_2$ and $\boldsymbol{M}_3' = \boldsymbol{M}_2^T \boldsymbol{M}_1$. Applying the well-known properties of the matrix transpose:

$$\begin{aligned} \boldsymbol{M}_3 &= \boldsymbol{M}_1^T \boldsymbol{M}_2 \\ &= (\boldsymbol{M}_2^T \boldsymbol{M}_1)^T \\ &= (\boldsymbol{M}_3')^T \end{aligned}$$

and if $\boldsymbol{M}_3 = \boldsymbol{X} + \boldsymbol{Y}$, and $\boldsymbol{M}_3' = \boldsymbol{X}' + \boldsymbol{Y}'$, then

$$\boldsymbol{M}_3^T = \boldsymbol{X}^T + \boldsymbol{Y}^T$$

Since $m_1(h^i) m_2(h^j) = m_2(h^j) m_1(h^i)$

$$\sum_{i}^{N} \sum_{j}^{N} x_{ij} = \sum_{j}^{N} \sum_{i}^{N} x_{ji}$$

Hence $\boldsymbol{X}^T = \boldsymbol{X}'$ and by similar arguments $\boldsymbol{Y}^T = \boldsymbol{Y}'$

In the matrix form of the Dempster-Shafer combination rule

$$\frac{\sum_{h^i \cap h^j = a}^{N} y_{ij}}{1 - \sum_{j=1}^{N} \sum_{i=1}^{N} x_{ij}} = \frac{\sum_{h^i \cap h^j = a}^{N} y_{ji}}{1 - \sum_{j=1}^{N} \sum_{i=1}^{N} x_{ji}}$$

$$= [m_2 \oplus m_1](a) \blacksquare$$

Hence the Dempster-Shafer combination rule satisfies the required commutative properties of our intelligence model. Using the matrix forms in equations 5.1 and 5.2 it is possible to derive some results for some special cases, we consider two of these in the next section.

Case 1: Complete Information

Consider the case when an intelligence report E_1 enables an analyst to assign a non-zero *pba* to each hypothesis $h^i \in \mathcal{H}$. If a second intelligence report, E_2, is received that also assign a non-zero *pba* to each hypothesis, then we define the Dempster-Shafer combination matrix $\boldsymbol{M_C}\left(= \boldsymbol{M_1^T M_2}\right)$ as *complete*. $\boldsymbol{M_C}$ has the following properties.

THEOREM 2. *Except for the trivial case (one hypothesis), the determinant of a complete combination matrix* $|M_3|$ *is zero.*

PROOF

The vector $\boldsymbol{M_2}$ forms a basis for each row of $\boldsymbol{M_3}$, hence the rows are linearly dependent. This implies that $|M_3| = 0 \blacksquare$

Decomposing the complete matrix into $\boldsymbol{M_C} = \boldsymbol{X_C} + \boldsymbol{Y_C}$, where as before, the elements of $\boldsymbol{X_C}, x_{ij}$ are defined by

$$x_{ij} \equiv \begin{cases} m_1(h^i)m_2(h^j), & h^i \cap h^j = \emptyset \\ 0, & otherwise \end{cases}$$

and the elements of matrix $\boldsymbol{Y_C}$ are similarly defined as before. If $m_1(h^i)m_2(h^j) = m_1(h^j)m_2(h^i)$, then $\boldsymbol{X_C}$ is *symmetric* and the value of $1 - \sum_{j=1}^{N}\sum_{i=1}^{N} x_{ji}$ in the denominator of the combination rule can be easily evaluated by taking the sum of the absolute values of the $2^n - 1$ eigenvalues of $\boldsymbol{X_C}, \sum|\lambda_i|$.

Case 2: Only one hypothesis has a defined bpa

If two intelligence reports only enable a *bpa* to be assigned to one of the hypothesis in $\mathcal{H}$, h^A, say such that from the first intelligence report $m_1(h^A) = a_1$and from the second intelligence report $m_2(h^A) = a_2$. Then

$$m_3(h^A) = a_1 + a_2 - a_1 a_2$$

and

$$m_3(h^A, \dots \dots, h) = 1 - m_3(h^A)$$

regardless of the size of $\mathcal{H}$.

On receiving the first intelligence report, the value of $el(h^A) = a_1$, after combining the information in the second intelligence report, $Bel(h^A)$increases by $a_2(1 - a_1)$. The Plausibility of $h^A, Pl(h^A) = Bel(h^A)$, i.e. the situation is the same as for normal probability theory. This offers an advantage to an intelligence analyst, that if not all the elements of the hypothesis set $\mathcal{H}$ can be defined, then they can be included within the partition of $\mathcal{H}$that contains all its elements.

6. Discussion and Conclusion

We have developed a simple model for intelligence analysis and used it to review the application of Bayes' theorem to intelligence analysis. Our model relied on the intelligence analyst declaring a prior probability that a given hypothesis was correct. In the example we considered, it was straightforward to assign prior probabilities based on the possible courses of action. In reality, assigning prior probabilities is not such a simple task, by slightly varying the information contained within an intelligence report so that it reduced the possibility of one of the hypotheses, we showed that the report did enable the analyst to update the assessment that a given course of action would be taken. In short, the intelligence report did not improve the analyst's understanding of the situation.

While we showed that the Bayes' formulism was independent of the order that intelligence reports were received, in reality individual reports can often influence analysts out of proportion to their ability to update prior probabilities. This has been recognised as a real concern within studies of intelligence (Kam 1988) and will be investigated in future work.

The application of the Dempster-Shafer theory to intelligence analysis altered the structure of our intelligence model, so that the analyst was no longer required to assign prior probabilities to each hypothesis. There is still a requirement to define a hypothesis set $\mathcal{H}$; however, by extending the argument used in the case where a specific *bpa* is only defined for one hypothesis, all unknown hypotheses with $\mathcal{H}$ can be included within the partition that contains all elements of $\mathcal{H}$.

A criticism of the application of Bayes' theorem to intelligence analysis is that computation and interpretation of the results may be difficult and time consuming. While this could also be a drawback to practical implementation of the Dempster-Schafer theory, the matrix formulisation developed here should simplify some of the calculations required, so that the process can become automated to enable faster evaluation by intelligence analysis.

Appendix

THEOREM. *The final probability a hypothesis is correct given by the Bayesian intelligence model is independent of the order the intelligence reports are received, i.e.*

$$P\left(H|E_{0,}\cdots,E_i,E_j\right) = P\left(H|E_{0,}\cdots,E_j,E_i\right), \qquad i \neq j,$$

PROOF
From Bayes' Theorem

$$P(H|E_0) = \frac{P(E_0|H)P(H)}{P(E_0)}$$

Making the transformations $H \longrightarrow H|E_0$ and $E_0 \longrightarrow E_1$, then

$$P(H|E_0,E_1) = \frac{P(E_1|(E_0|H))P(H|E_0)}{P(E_1)}$$

using the definition of conditional probability

$$P(E_1|H|E_0) = \frac{P(H \cap E_0 \cap E_1)}{P(H \cap E_0)}$$

Hence

$$P(H|E_0,E_1) = \frac{P(H \cap E_0 \cap E_1)}{P(H \cap E_0)}\frac{P(H|E_0)}{P(E_1)}$$

$$P(H|E_0,E_1) = \frac{P(H \cap E_0 \cap E_1)}{P(H \cap E_0)}\frac{P(E_0 \cap H)}{P(E_1)P(E_0)}$$

e.g.

$$P(H|E_0,E_1) = \frac{P(H \cap E_0 \cap E_1)}{P(E_1)P(E_0)}$$

Similarly, if intelligence report E_1 is received before E_0, then

$$P(H|E_1,E_0) = \frac{P(H \cap E_1 \cap E_0)}{P(E_0)P(E_1)} = P(H|E_0,E_1)$$

By induction on $i \geq 1$, and replacing the order of the indices, result follows■

References

A mathematical model for intelligence warning, *CIA Intelligence report* 1967, No. 1396/67

BEYNON, MALCOLM, CURRY, BRUCE, MORGAN, PETER, 2000 *Omega* 28 37-50

BUEDE D. M., GIRARDI P. A., 1997 *IEEE Transaction on Systems, Man and Cybernetics-Part A: Systems and Humans*, 27(5) 569-577

DEMPSTER A. P., 1966 *Ann Math Stat* 37, 355-374

DEMPSTER A. P., 1967 *Biometrika*, 54, 515-528

DEMPSTER A. P., 1969 *Ann Math Stat* 40, 953-969

DENOEUX, T., 1995 *IEEE Transactions on Systems, Man and Cybernetics*; 27(5) 569-577

DUBOIS, D., PRADE, H., 1988 *Computational Intelligence*, 4, 244-264

FISK, CHARLES E., 1972 The Sino-Soviet Border Dispute: A comparison of the conventional and Bayesian methods for intelligence warning. *Studies in Intelligence*, 16(2), 53-62

KAM, EPHRAIM, *Surprise Attack*, 1988 Harvard University Press, Cambridge, Massachusetts

LASKEY, M. B., LEEHNER, P. E. 1989 *Artificial Intelligence* 41 66-77

OMAND D., *Science advice and the management of risk in government and business*, http://www.foundation.org.uk/events/pdf/20101110_omand.pdf, 2010

SELVIN, S. 1975 "A problem in probability (letter to the editor)". *American Statistician* 29 (1): 67

SHAFER, G., 1976 *A mathematical theory of evidence.* Princeton University Press, Princeton, New Jersey

WALLEY, P., 1987 *Ann Stat*, 10, 741-761