Fusion of data from sources with different levels of trust

David A. Nevell
QinetiQ
Malvern, UK
dnevell@QinetiQ.com

Simon R. Maskell
QinetiQ
Malvern, UK
srmaskell@QinetiQ.com

Paul R. Horridge
QinetiQ
Malvern, UK
prhorridge@QinetiQ.com

Hayleigh L. Barnett
QinetiQ
Malvern, UK
hlbarnett@QinetiQ.com

Abstract – In the context of this paper, trust is defined to be "a measure of to what degree an information source is believed to be capable of producing information that conforms to fact." No standard method has been adopted by the intelligence community for fusing data from sources with different levels of trust. This paper proposes an approach that extends the standard application of Bayesian inference to allow for the fact that any piece of intelligence data may be less than fully trustworthy. Based on a prototypical intelligence scenario from which synthetic data was generated, results indicate that trust models produce results which are closer to the ground truth than those for a model containing no trust variables, exhibit less variability and which provide a better basis for making correct decisions.

Keywords: Tracking, data association, Kalman filtering, estimation.

1 Introduction

For an intelligence analyst dealing with data that contain elements of subjectivity, the concept of data trustworthiness is an important one. The analyst may be faced with thousands of pieces of information about a disparate set of subjects from a wide range of intelligence sources. These sources may be highly subjective as in the case of Human Intelligence (HUMINT) or have subjectivity added to them through the interpretation of sensor outputs as in the case of Measurement and Signal Intelligence (MASINT). The decisions that depend upon an interpretation may be highly sensitive to how well this intelligence is trusted. In addition, there may be little time for a lengthy or manual analysis. How should this data be fused and interpreted?

Although a great deal of research has been undertaken worldwide into the concept of trust (e.g. [1], [2]) this does not yet appear to have manifested itself in the form of a standard approach adopted by the intelligence community. One reason for this may be the fact that the term *trust* means different things to different people. There are a number of similar terms that, rightly or wrongly, tend to get used interchangeably or confused. For example, confidence, belief, credibility, repeatability, reliability, accuracy, precision and uncertainty are all words which can occur in the intelligence domain. To make things worse, a definition of one of these terms often involves use of another and there is a danger that everything is confused by a self-referential semantic muddle. It is clear then, that any work that seeks to analyse and manipulate trust needs to be clear exactly what is meant by it.

In this particular military context it is also important that existing ways of expressing trust-like concepts are taken into account. Table 1 contains the structure for the classification of intelligence data taken from standard military guidelines.

Reliability of Source	Credibility of information
A Completely reliable	1 Confirmed by other sources
B Usually reliable	2 Probably true
C Fairly reliable	3 Possibly true
D Not usually reliable	4 Doubtful
E Unreliable	5 Improbable
F Cannot be judged	6 Truth cannot be judged

Table 1: Classification of intelligence data

In addition, cognisance needs to be taken of the way that confidence levels are defined, which uses the Confirmed (>95%), Probable (>75%), and Possible (>50%) probability scale.

2 Methodology

2.1 A definition of trust

Taking into account the context explained above, it is proposed that there are three broad concepts that need to be captured. The third of these represents an extension to standard intelligence analysis and for which the term trust will be used. These concepts are.

1. A measure of how well information conforms to a datum. The datum is a measurable frame of reference, often taken to be the mean or a specified target (but not the ground truth, which is unknown). To this concept the terms uncertainty or variance can be broadly associated.
2. A measure of how much belief is attached to individual elements of information, based on how sure one can be that they conform to, or are relevant to, fact. To this concept the term confidence can be broadly associated.
3. A measure of to what degree an information source is believed to be capable of producing information that conforms to fact. If this capability is called

trustworthiness, then the measure of it, derived from analysing intelligence data, is called trust.

The key difference between the way that (2) and (3) are defined here is that whilst the former is information oriented, the latter is information source oriented. For example, if an analyst has a low level of confidence about a piece of information it does not mean that he is untrustworthy – it may just mean that he has an very good self-perception of how much weight should be placed on his observation.

Trust is defined on a [0, 1] scale and can be applied to the both the information source itself, and (by inference) the information it produces. A trust level of 1 represents a belief that the information conforms exactly to fact, and a trust level of 0 represents a belief that the information is of no merit whatsoever. It should be noted that a trust level of 0 does not imply that the information never conforms to fact; rather it represents a position of neutrality. A source of information that never conforms to fact, if identified, could be construed as contributing useful, if negative, information.

2.2 The trust model

A mathematical framework was required in order to fulfil the aims of the project. Since the main objective was to show the advantages of a trust model over a model without trust, it could be argued that the actual type of model into which to add a trust variable was of secondary importance. Many years of work in the "belief" domain has lead to the development of competing candidate methodologies such as Bayesian probability theory, fuzzy logic and Dempster-Shafer. Since it has now been shown that one of the paradigms for representing uncertainty in the context of information fusion is that of Bayesian probability theory, which can articulate belief through the assignment of probabilities to mutually exclusive hypotheses [3], this was the route chosen.

The adopted approach then, was to extend the standard application of Bayesian inference to allow for the fact that any piece of intelligence data might not be fully trustworthy. This extension allows a range of sources of untrustworthiness to be captured and quantified; examples range from the probability that a specific expert may have made a mistake to the probability that a generic data source has a historic level of reliability.

The approach is intended to be as simple as possible. Each individual element of intelligence, which will be referred to as an observation, can be associated with one or more sources, which are represented as elements in a hierarchy. This hierarchy is referred to as the trust hierarchy. The trust hierarchy has within it sources that are both intrinsically subjective (e.g. an image analyst) and intrinsically objective (e.g. a sensor producing an image). An observation that is made as a result of an analyst interpreting the output of a sensor can be thought of as having two parents, the analyst and the sensor. In general, the structure of the trust hierarchy dictates to what extent an observation can be trusted, given knowledge of which its parents are and how trustworthy they are. A generic overview of the model is shown in Figure 1.

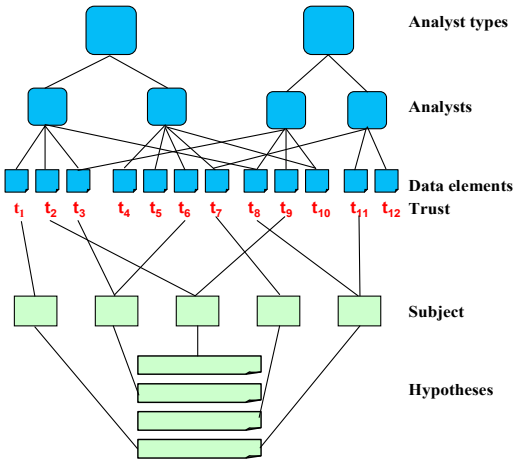

Figure 1: Generic model structure

Information elements (observations) are shown across the centre of the figure. Each has an associated level of trust in the interval [0,1] which is derived from the trustworthiness of its parents in the trust hierarchy above. The term analyst is used in a very general sense; it can refer to anyone from a trained expert to a village elder. Not all observations necessarily have both analysts and sensors as parents. Observations are also linked to subjects or objects in the hypothesis hierarchy below. The lowest part of that hierarchy contains hypotheses which most closely relate to operational decision making, e.g. factory A is manufacturing vehicle type B. The precise structure of the hypothesis hierarchy will depend upon the operational scenario in question; there may be many more levels than shown here

The individual information sources, information elements and hypotheses form the nodes of the network. Prior estimates for trust levels (for the trust hierarchy) and probabilities (for the hypotheses) are required, – although these can be set at neutral values if prior knowledge is low. Loopy belief propagation [4] is then used to jointly estimate the posterior trustworthiness of each element in the trust hierarchy and the probabilities of the hypotheses. These take into account any untrustworthiness that has been detected; there is no need to re-run the model with untrustworthy sources removed, this has already been accounted for. Further details of how the concept of trust is used in the model are described in Section 2.4.

It was important that the Bayesian model provided an opportunity for Intelligence Report levels of confidence, credibility and reliability that were listed in Table 1, to be incorporated. The alphabetic part of the code relating to the reliability of a source can be used to inform the prior estimates of information source trust levels. It is also possible to use the numeric part of the code relating to the credibility of a single information

element to inform that element's prior trust in the model (normally set at neutral). The overall levels of confidence (High, Medium, Low) are already standard inputs to the models.

2.3 Methodology test bed

A context and data relating to that context were needed with which to demonstrate the trust hierarchy and hypothesis hierarchy. The chosen context was what was considered to be a prototypical intelligence scenario relating to vehicle manufacture in a fictional town. The intelligence requirement was to ascertain, from information gathered from delivery vehicles, which factories within the town were manufacturing certain types of vehicle. Although not originally set up with the issue of trust in mind, this pre-existing scenario provided an ideal testing ground for the methodology.

A high level view of the scenario is given in Figure 2. The town has 10 factories, each capable of making up to 5 different types of vehicle. Delivery of components to enable manufacture of vehicles is on a just-in-time basis, so it is possible to assess which vehicles are being made in which factory on the basis of observing those deliveries.

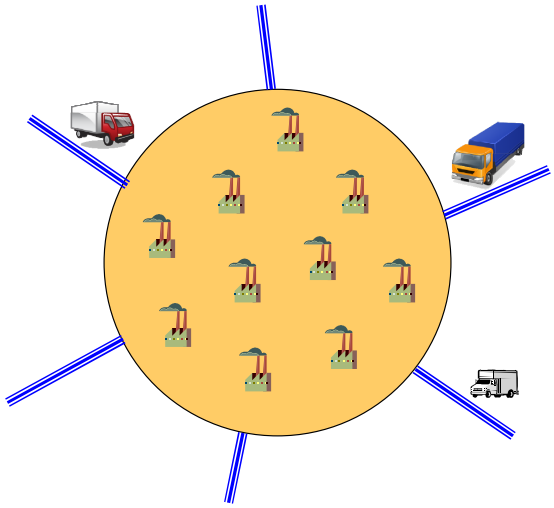

Figure 2: High level view of the intelligence scenario

There are 6 roads leading into the town; delivery vehicles may use any of them but there is no known correlation between entry road and destination factory. Each vehicle is observed on camera twice, once on one of the 6 roads leading into the town and later at one of the ten factories.

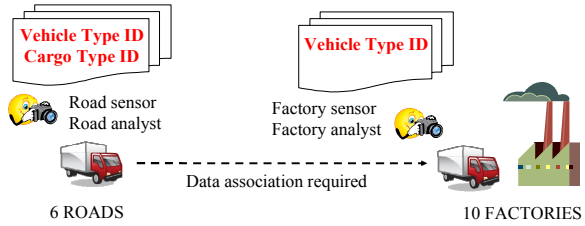

Figure 3: Intelligence scenario – data structure

For every camera, an analyst, subject to error, reports on what he has seen (Figure 3). The cameras at the factories are of a lower quality than those of the roads and only the delivery vehicle type can be observed there whilst earlier, the road analyst may have an opportunity to spot the single cargo type too (some of the cargo is covered). A correctly identified cargo type provides stronger evidence about possible manufacturing activity than a correctly identified delivery vehicle so there is a need to associate the road and factory observations. This can only be done on the basis of timings and the (error-prone) observations.

Data association, a crucial part of the intelligence data analysis process, was not considered to be core to the initial requirements for assessing the trust methodology. To that end, at this stage of the analysis an assumption was made that the correct associations could be successfully made; this was built into the data generation process. The basic information element then, referred to as an observation, was based on each sighting of a vehicle as it entered a factory and consisted of the following:

- Time (of delivery at factory).
- Delivery vehicle ID (1 of 3 types).
- Confidence in vehicle ID (High, Medium, Low).
- Cargo ID (1 of 10) – obtained via the data association process.
- Confidence in cargo ID (High, Medium, Low).

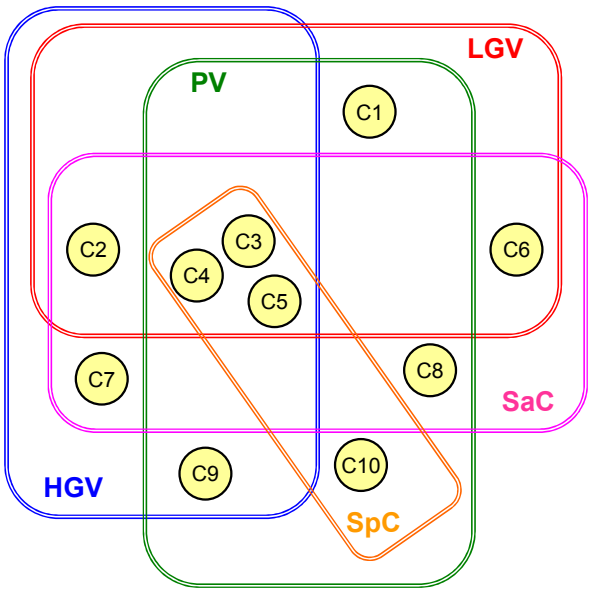

Figure 4: Components required for vehicle manufacture

The analysts have prior knowledge of which components are used to make which vehicle, and of which delivery vehicles can carry each component. The linkages are far from straightforward though. Figure 4 shows the considerable degree of commonality between the required components for each of the five vehicles (shown by HGV (Heavy Goods Vehicle), LGV (Light Goods Vehicle), PV (Panel Van), SaC (Saloon Car) and SpC (Sports Car)).

When this commonality, along with the degree of error introduced by the quality of the sensors and the shortcomings of the analysts themselves is taken into account, it is clear that the question about which factories are making which vehicles can only be answered through the accumulation of a large number of observations. Conclusions made about the manufacturing profiles are potentially very sensitive to systematic sources of untrustworthiness.

2.4 Detailed modeling of trust

It is worth looking in more detail how the concept of trust is utilised in this particular example. Figure 5 shows a simplication of the trust hierarchy and the interface with the top of the hypothesis hierarchy.

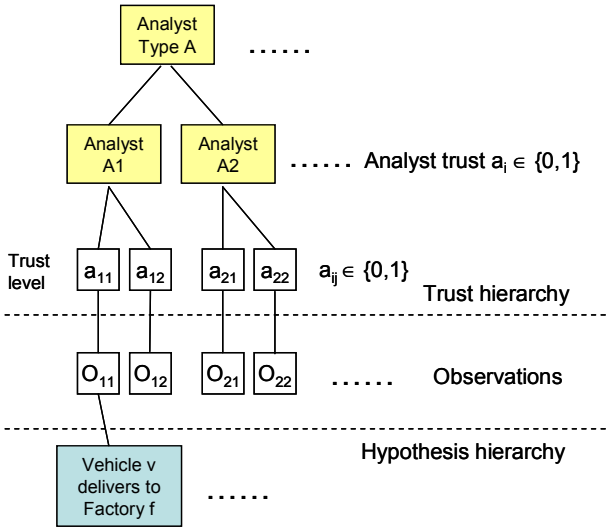

Figure 5: Part of the detailed trust hierarchy with linkage to the hypothesis hierarchy

In this case each observation has one parent, a single analyst. The relationship between analyst trustworthiness and observation trustworthiness was modelled in the way shown in Table 2. A default value for *q* of 0.90 was utilised throughout the analysis. A similar relationship can also be used between other levels in the trust hierarchy (e.g. between Analyst Type and Analyst).

		Analyst trustworthiness	
		0	1
Observation trustworthiness	0	q	1-q
	1	1-q	q

Table 2: Relationship between analyst and observation trustworthiness

Moving downwards to the first level of the hypothesis hierarchy, the probability that a vehicle of a certain type has conducted a delivery depends upon the trustworthiness of the observation, as well the level of confidence associated with the observation. There are two possibilities, as shown in Table 3.

Observation cannot be trusted

		P(Actual delivery vehicle type)		
		HGV	LGV	PV
Observed delivery vehicle type	HGV	1/3	1/3	1/3
	LGV	1/3	1/3	1/3
	PV	1/3	1/3	1/3

Observation can be trusted

		P(Actual delivery vehicle type)		
		HGV	LGV	PV
Observed delivery vehicle type	HGV	P	(1-P)/2	(1-P)/2
	LGV	(1-P)/2	P	(1-P)/2
	PV	(1-P)/2	(1-P)/2	P

Table 3: Relationship between observed vehicle type and probabilities associated with actual vehicle type

When the observation cannot be trusted (top half of the table), regardless of which of the three delivery vehicles has been logged, a flat profile of probabilities is used for the actual vehicle type. This assumes a position of complete neutrality such that the untrustworthy observation is effectively ignored. If the observation can be trusted then it is given a probability weighting of P, the confidence associated with the observation, and a flat weighting is applied to the remaining vehicle types. In this analysis P values of .95, .8 and .6 were used to represent High, Medium and Low confidence levels. This then reflects the Confirmed, Probable and Possible intelligence data classification described in Section 1.

Further calculations in the hypothesis hierarchy involve probabilities only, culminating in the generic bottom line probability that Factory *f* is manufacturing Vehicle *v*. None of these further calculations involve the explicit concept of trust.

2.5 Data generation

To provide a means of assessing the proposed methodology, a data generator was developed that produced observation data and also had the capacity to inject known issues of untrustworthiness into chosen information sources, i.e. analysts and/or sensors.

With the data generator it was then possible to set up a structured set of data sets upon which the comparison between the trust model and the no-trust model could be formally analysed. To allow for random variation, multiple repeat runs were set up for each configuration to be tested.

2.6 Experimental Approach

The runs were aimed at satisfying this main objective: *To demonstrate the extent to which the introduction of a trust factor model facilitates "better" solutions to intelligence*

fusion problems than those produced by using a model without a trust factor.

The concept of *"better"* has been captured in the following three ways;

1. *It is possible to correctly identify sources of untrustworthiness.* This was expressed in terms of the mean posterior level of trust associated with those information sources that are known to be untrustworthy.
2. *It is possible, taking into account identified untrustworthiness, to more closely identify the underlying ground truth.* This was quantified by the following log-likelihood statistic *F*:

$$F = \log_{10}\left(\prod_{i,j}\left(1 - abs(g_{i,j} - e_{i,j})\right)\right) \qquad (1)$$

   $g_{i,j}$ is the ground truth relating to the manufacture of Vehicle *j* at Factory *i* (taking the value 0 or 1), and $e_{i,j}$ is the probabilistic estimate of that ground truth. *F* takes a maximum value of 0 if the estimate is perfect. This captures the probabilistic difference between the ground truth and the posterior probabilities relating to every combination of factories manufacturing vehicles. The mean and variance of F are calculated for a set of repeated runs.
3. *It is possible to then make better executive decisions based on the fused data.* This was measured by considering what the analyst would conclude was the statistically most likely combination of vehicles that each factory was producing (obtainable from their joint distribution) and then counting and classifying the errors made. The classification errors were defined as follows:

   Misclassification type 1: The most likely estimate is for a greater number of vehicle types being manufactured than is actually true.

   Misclassification type 2: The most likely estimate is for the correct number of vehicle types being manufactured but the wrong types.

   Misclassification type 3: The most likely estimate is for a lower number of vehicle types being manufactured than is actually true.

With these metrics, the experimental approach was focused on directly comparing the result of using the Bayesian model with a trust variable for each element in the trust hierarchy against using it without the capability of measuring trust. There was a particular need to ensure that the trust model did not introduce Type 1 (false positive) errors. This is expressed in Figure 6.

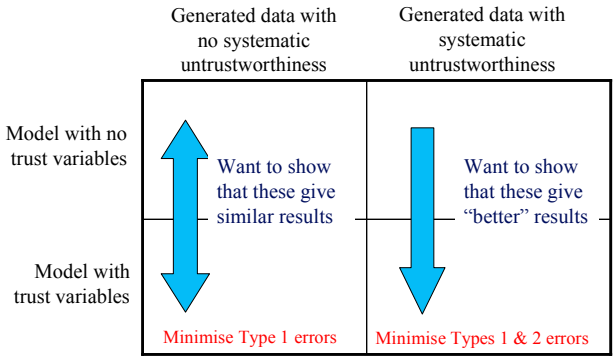

Figure 6: Experimental approach

3 Results

The following results are for a baseline where:

- No factory makes more than two different vehicle types.
- Where manufactured, all vehicles are manufactured at the same rate.
- Analysts have a relatively high degree of average confidence about their observations, both for vehicle types and for cargo types.

2000 observations were randomly generated for each run, representing about three weeks of data in the scenario. The introduction of systematic untrustworthiness was limited to analysts only. Untrustworthy analysts were set to produce what were approximately random observations (trust level=0.25). Trustworthy analysts were set to an underlying trust level of 0.90. To reduce unwanted variability between comparative runs, common data were used wherever possible.

3.1 Identification of untrustworthy information sources

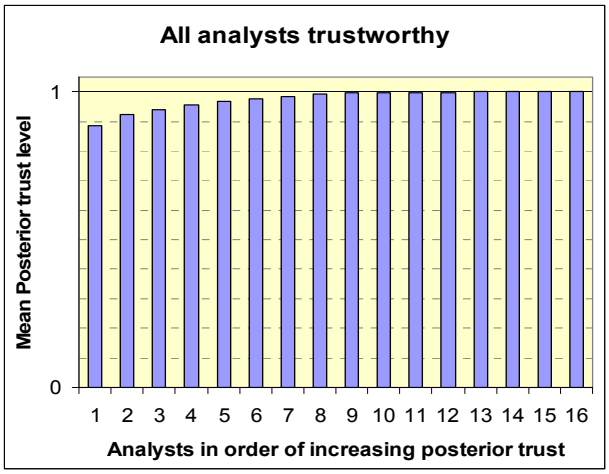

Figure 7: Mean posterior trust (no untrustworthy analysts)

For the baseline case where the observation data contains no untrustworthy information sources, Figure 7 shows that the trust model consistently reaches the correct conclusion that all analysts can be trusted, with the mean

least trustworthy analyst being assigned a trust factor of about 0.9, the same as the prior. However, when one analyst is set as untrustworthy on each run (a different one each time), the trust model always correctly identifies him and assigns a mean posterior trust level of close to 0.1 (Figure 8). All other analysts are deemed to be close to 100% trustworthy.

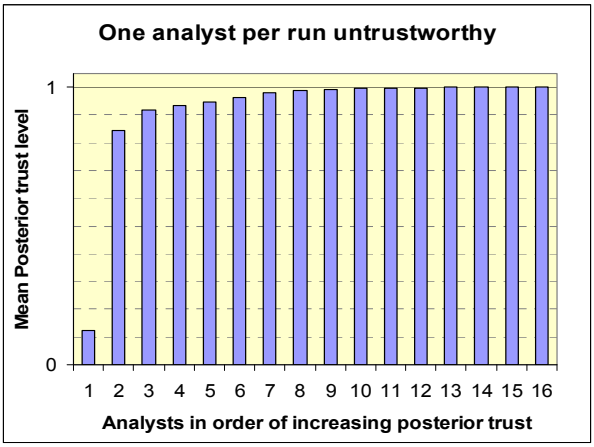

Figure 8: Mean posterior trust (1 untrustworthy analyst)

As the number of untrustworthy analysts is further increased to two and then three, the model successfully identifies them and discounts their information, almost completely in every case. Figure 9 shows the mean posterior trust values for the case where there are three untrustworthy analysts per run.

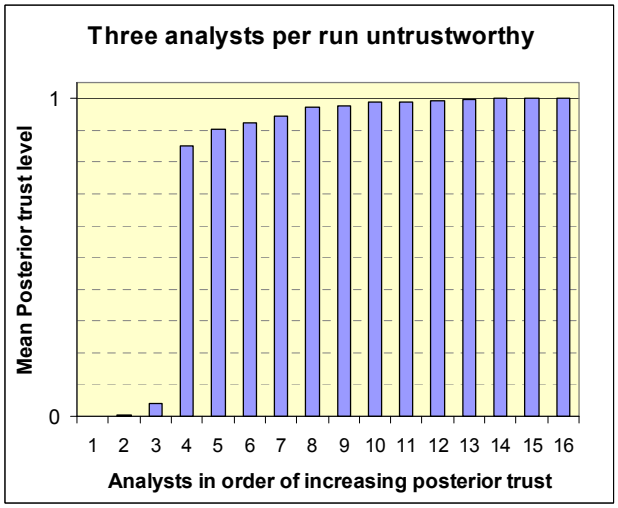

Figure 9: Mean posterior trust (3 untrustworthy analysts)

3.2 Match with ground truth

The results for the baseline case are shown in Figure 10.

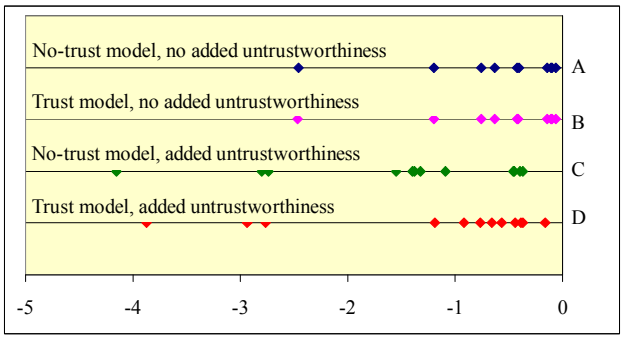

Figure 10: Log-likelihood statistic F with 1 untrustworthy analyst

The results for run sets A and B are almost identical, indicating that the introduction of the trust factor in B has a stable effect when no untrustworthiness is present. The results for D are slightly better than those for C, indicating that the extra trust terms in the model are capable of reducing the effect of added untrustworthiness. The difference between C and D is made more significant by the fact that there is a strong pair wise improvement; 8 of the 10 corresponding runs show a reduction in F. This difference however becomes much more pronounced as further untrustworthy analysts are introduced. Figure 11 shows the results for the resulting cases C and D when there are two untrustworthy analysts present with trust levels set at 0.25.

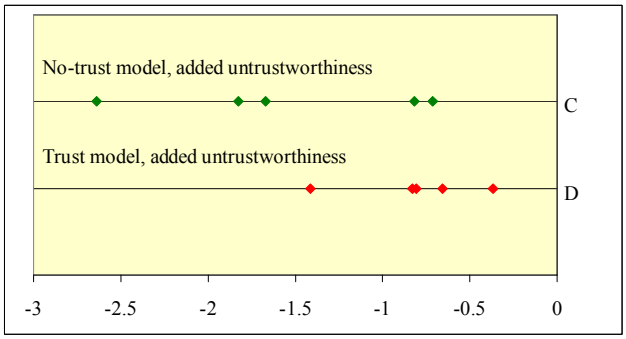

Figure 11: Log-likelihood statistic F with 2 untrustworthy analysts

C: Mean = -1.53 Variance = 0.63
D: Mean = -0.81 Variance = 0.15
Here a significant reduction is seen in both the mean and the variance. The difference between C and D is even greater as the number of untrustworthy analysts is increased to three per run as Figure 12 shows.

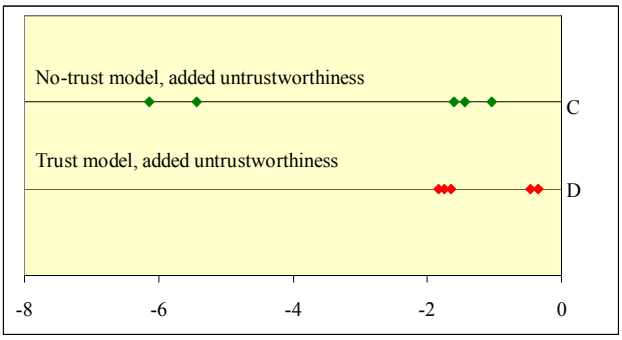

Figure 12: Log-likelihood statistic F with 3 untrustworthy analysts

C: Mean = -3.13 Variance = 6.00
D: Mean = -1.20 Variance = 0.54

3.3 Identification of most likely manufacturing profiles

Table 4 shows the results classification for the baseline run and with increasing numbers of untrustworthy analysts.

			Misclassification type		
Number of untrustworthy analysts	Model	Correct classification rate	1	2	3
0	No-trust	96%	3	0	1
0	Trust	96%	3	0	1
1	No-trust	88%	8	1	5
1	Trust	92%	5	0	5
2	No-trust	88%	5	1	0
2	Trust	98%	1	0	0
3	No-Trust	70%	11	0	4
3	Trust	94%	1	0	2

Table 4: Summary of manufacturing classifications

In each of the 3 cases (1 to 3 untrustworthy analysts), the trust model returns fewer misclassifications.

3.4 Sensitivity runs

Many other sensitivity runs were performed and for a large proportion of them the results reflected what has already been shown previously in Section 3. This included changing the prior estimates of the trust levels for elements in the trust hierarchy, the number of observations, adding untrustworthy sensors, as well as a number of features of the hypothesis hierarchy which affected the overall ability of the models to identify the ground truth. However, these changes did not alter the fundamental relationship between the two types of model; i.e. the addition of a trust term always provided some degree of improvement in the defined key metrics.

4 Conclusions

- The introduction of a trust variable has been successfully implemented into a Bayesian network model using Belief Propagation as a means to estimating the Bayesian model parameters. It has been shown, over a wide variety of generated data, to produce results which are closer to the ground truth than those for a model containing no trust variables, exhibit less variability and which provide a better basis for making correct decisions.

- When systematic untrustworthiness was not present in the data the addition of trust terms to the model did not result in data sources being falsely identified as untrustworthy and when systematic untrustworthiness was present in the data the addition of trust terms to the model did allow untrustworthy data sources to be clearly identified.

- The model has the ability to discount other data from sources that have been identified as producing untrustworthy observations – an innovation that results in improved performance.

- The method can run in only a few minutes on large amounts of data.

- The methodology can be implemented without the need for new types of input from intelligence analysts and can use existing data assessment codes to inform inputs into the trust model.

References

[1] S. Young & J. Palmer, *Pedigree and Confidence: Issues in Data Credibility and Reliability,* Overwatch Systems, Austin, TX, U.S.A.

[2] Ion Matei, John S. Baras & Tao Jiang, *A Composite Trust Model and its Application to Collaborative Distributed Information Fusion,* 12th International Conference on Information Fusion, Seattle, WA, USA, July 6-9, 2009

[3] S. Maskell, *A Bayesian Approach to Fusing Uncertain, Imprecise and Conflicting Information*, Information Fusion Journal. 9(2):259-277. April 2008

[4] J. S. Yedidia, W.T. Freeman, and Y. Weiss, *Understanding belief propagation and its generalizations*, Mitsubishi Electric Research Laboratories, TR-2001-22, January 2002