

Information Evaluation: a formalisation of operational recommendations

Vincent Nimier

ONERA Centre de Châtillon

BP 72, 29 avenue de la Division Leclerc, 92322 Châtillon, France

nimier@onera.fr

Abstract *An important step in the intelligence gathering process is the fusion of information provided by several sources. The objective of this process is to build an up-to-date and correct picture of the current situation with the overall available information in order to make adequate decisions. In the framework of intelligence, in opposition to air defense domain for example, there is no real automatic process and the fusion process is made by human operators. The stanag 2022 give a methodology for the evaluation of information in the framework of intelligence and for human processing. However, given the increase in the amount of information that the human operator need to process, some automatic processing is being considered. Therefore there is a need for a formal methodology for the evaluation of information. In this paper, we present some consideration about the correspondence between the STANAG 2022 information evaluation methodology and a formal and mathematical evaluation approach gave by an automatic process.*

1. Introduction

An important step in the intelligence gathering process is the fusion of information provided by several sources. The objective of this process is to build an up-to-date and correct view of the current situation with the overall available information in order to make adequate decisions. To succeed in this process, it is important to associate, with each available information, some attributes like the number of the sources that support it, their reliability, the degree of truth of the information etc. The Standardization Agreements (STANAG) 2022 of North Atlantic Treaty Organization (NATO) gives some elements of a framework for such common definitions. However, actually, in the operational intelligence process, these attributes are managed by the human operator when he fuses information provided by the different sources. Indeed, there is no real methodology to do this in a formal manner. When relaying this fusion process to a machine, we need to develop formal definitions and algorithms to manage these attributes in addition to fusing information. Furthermore, in a context of interoperability where different systems exchange information, common definitions of these attributes have to be shared.

The purpose of this paper is first to analyze the STANAG 2022 recommendations about information evaluation and then to set a first step in the definition of a formal and non ambiguous system for evaluating information. Indeed, as it will be shown, the present recommendations, written in natural language are rather ambiguous and imprecise and are open to discussion. Secondly, knowing that in general an automatic fusion process is based on some mathematical theories like: probability, possibility, Dempster Shafer, we propose some correspondence between the result that this automatic process may provide and an operational evaluation in accordance to the evaluation process describe in the STANAG 2022.

This paper is organized as follows. Section 2 presents a review of STANAG 2022 recommendations and points out to the main notions that underline these recommendations. Section 3 analyzes the different assumptions underlying the evaluation proposed in the STANAG. Section 4, presents the link between the notions described in the STANAG and the results of automatic processing based on the use of mathematical theory. A model for a data fusion process is presented in section 5. It takes into account the notion of reliability. The conclusion to this paper is presented in section 6.

2. Review of STANAG 2022 recommendations

The Annex to STANAG 2022, Edition 8 ([1]) explicitly mentions that the aim of information evaluation is to indicate the degree of confidence that may be placed in any item of information which has been obtained for intelligence. This is achieved by adopting an alphanumeric system of rating. This system combines a measurement of the reliability of the source of information with a measurement of the credibility of that information when examined in the light of existing knowledge.

Examining the whole text leads us to point out that the two main concepts in this evaluation system are the reliability of the sources and the credibility of the

information. These concepts are defined in the STANAG 2022 recommendations, as follows.

The reliability of the source is designated by a letter between A and F signifying various degrees of confidence as indicated below.

- a source is evaluated A if it is completely reliable. It refers to a tried and trusted source which can be depend upon with confidence
- a source is evaluated B if it is usually reliable. It refers to a source which has been successfully used in the past but for which there is still some element of doubt in particular cases.
- a source is evaluated C if it is fairly reliable. It refers to a source which has occasionally been used in the past and upon which some degree of confidence can be based.
- a source is evaluated D if it is not usually reliable. It refers to a source which has been used in the past but has proved more often than not unreliable.
- a source is evaluated E if it is unreliable. It refers to a source which has been used in the past and has proved unworthy of any confidence.
- a source is evaluated F if its reliability cannot be judged. It refers to a source which has not been used in the past.

The credibility of information is designated by a number between 1 and 6 signifying varying degrees of confidence as indicated below.

- If it can be stated with certainty that the reported information originates from another source than the already existing information on the same subject, then it is classified as "confirmed by other sources" and rated 1.
- If the independence of the source of any item of information cannot be guaranteed, but if, from the quantity and quality of previous reports, its likelihood is nevertheless regarded as sufficiently established, then the information should be classified as "probably true" and given a rating of 2.
- If, despite there being insufficient confirmation to establish any higher degree of likelihood, a freshly reported item of information does not conflict with the previously reported behavior pattern of the target, the item may be classified as "possibly true" and given a rating of 3.
- An item of information which tends to conflict with the previously reported or established behavior pattern of an intelligence target should be classified as "doubtful" and given a rating of 4
- An item of information which positively contradicts previously reported information or conflicts with the established behavior pattern of an intelligence target in a marked degree should be classified as "improbable" and given a rating of 5
- An item of information is given a rating of 6 if its truth cannot be judged.

The previous definitions of information evaluation can be criticized. Indeed, since they are given in natural language, they are quite imprecise and ambiguous and a lot of points are open to discussion. Furthermore, this evaluation is given in a discrete scale while generally automatic processing gives evaluation results in a continuous scale between 0 and 1.

However, even if the previous recommendations are a little bit vague, they still present three basic concepts that form a cornerstone in the evaluation system. These concepts are the following:

- the reliability of a source
- the number of independent sources that support an information
- the fact that the information tends to conflict with some available information.

It is clear that these three concepts are independent and that they have to be included in all quotation systems, whether these systems are automatic or not.

3. Analysis of the different evaluation concepts

3.1. Basic concepts

The concept used in the STANAG was introduced in an operational framework with human operator working on the fusion process. If some tasks of the fusion process have to be automatic, we need to find some correspondence between operator reasoning and the result of automatic process using mathematical theory. For this we try in this section to give a more comprehensive and formal definitions of reliability and credibility based on some mathematical theory.

3.2. Reliability

In mathematical logic, the reliability of a source is defined in a binary way as follows [10]: *an information source is (totally) reliable if and only if the information it delivers is true in the real world.* For instance, a sensor which measures the temperature is totally reliable if and only if the temperature it indicates is the correct one; a human expert is totally reliable if and only if any information (opinion, conjecture etc) he gives is true.

According to the recommendations of STANAG 2022, the reliability of a source is not a binary notion but a scale one and is defined in reference to its use in the past. It can be measured for example, as the ratio of the number of times the source gave true information to the number of times it gave information. However, this definition does not take into account the actual environment of use of the information source. For

instance, even if it is known to be reliable, an infra-red sensor loose reliability when it rains. So we have to take into account the condition in which the source is used.

It must also be noticed that there is no consensus yet on a formal definition of reliability. For instance, we can read in the APJ 2.1 “*every piece of information produced by an impeccable source is not necessarily correct*”. If “impeccable” intends to mean “reliable”, this sentence is contradictory with the definition given previously. Here, it implies that the reliability of a source is not defined by its ability to deliver truth and that even a reliable source can be wrong. A source can be wrong not because it is not sincere but because its model may be miss-adapted to the situation.

To give some mathematical models to the above considerations, we note C the event “the source is used in favorable condition”. That means that it is used such that its performances are well known and they are the best as possible. In other terms, it gives true information when possible. So the reliability can be modeled by the probability $P(C)$ which is equal to 1 if it is totally reliable and equal to 0 if it is unreliable. We will see further, thanks to this model, how the sentence: “*every piece of information produced by a **totally reliable** source is not necessarily **true***” can be modeled.

3.3. Credibility

The credibility of an information is a more natural concept for a mathematical information fuser. Indeed it can be considered as the likelihood function of the information. This credibility is evaluated based on the following two basic concepts:

3.3.1. Conflict

This paragraph addresses the concept of conflict with a focus on its use in probability. First we define the concordance function thanks to a Kronecker product. The conflict function is a straightforward derivation of this last function. Secondly, we expand this function to take into account the notion of uncertainty in the decision process.

Let a set of hypotheses $D=\{H_1, \dots, H_n\}$. After observation, a declaration is given by a source. Let $d=i$ with $i \in \{1, \dots, n\}$ this declaration. Then the conflict between each hypothesis and the declaration can be defined in a simple way by the relation:

The declaration $d=i$ is in conflict with an hypothesis H_j if $i \neq j$. If $i=j$ we say the declaration concurs with the hypothesis.

From a strict mathematical point of view we can define the concordance by the mathematical function δ_{ij} which is equal to 1 if $i=j$ and equal to 0 if not. With this

definition the conflict can be described by the function $1 - \delta_{ij}$.

These functions are binary. Moreover, we can give some shade to the concordance or the conflict between the declaration and an hypothesis by defining the probability $P(d=i/H_j)$. In numerous cases the concordance function is maximum when the declaration agrees ($i=j$) and at the limit $P(d=i/H_i)=1$, this implies that $P(d=i/H_j)=0$ if $i \neq j$.

3.3.2. Independence of sources

In this paper we consider that two sources are independent if they are two physical entities. Two informations are not independent if they come from the same source at different time periods.

4. Definition of a mathematical function for evaluation

As we have seen in section 3.2 the concept of reliability can be directly modeled by a probability. The credibility is more difficult to define because it encompasses the notion of conflict in addition to the number of sources.

Let us now consider a simple model for a “closed world” assumption and an exclusive decision process. The “closed world” means that we know and it is possible to establish all the classes of object and this set of classes is exhaustive. We denote $D=\{H_1, \dots, H_n\}$ this set of classes. An exclusive decision process means that all the information given by a source corresponds to a decision about one elementary hypothesis and only about one.

In the following we consider the general idea:

For the closed world assumption, we will not consider the rating of 6 for the credibility. The reason is straightforward. If an hypothesis belongs to the frame of discernment that means this hypothesis is possible, if not, we do not need to put it on this frame. Thus in the following, we will consider that if no information is available on one hypothesis, the credibility of this hypothesis is equal to 3. In the same spirit if one information confirms one hypothesis then the credibility of this hypothesis equals to 2 and if two informations confirm one hypothesis the credibility of this hypothesis will be equal to 1. Then the scale given to the credibility will be between 1 and 5.

4.1. Mono hypothesis approach

In this section we suppose that an automatic process fuses information based on one of the following mathematical theory approaches. The objective is that this process provides its results according to the recommendations such that an operator can manage the results with other information given by human intelligence.

4.1.1. Heuristic approach

At the beginning of the process and without any information from any of the sources, the only information available is a priori information. In this case, we can suppose that the a priori information is provided by a fictitious source with reliability equal to F . Given that all the hypotheses which make up the frame of discernment are possible, we assume that the credibility of each hypothesis is equal to 3.

If an information $d=H_j$ given by a source reaches the fusion process, the hypothesis H_j will have a credibility of 2 “probably true” and the other hypotheses H_i , for $i \in \{1, \dots, n\}$, with $i \neq j$ will have the credibility of 4 “doubtful”. If a second information $d=H_j$ arrives to the fusion process, two cases are then possible. The second information confirm the first one and then H_j will have the credibility of 1 “confirmed” in this case the others hypothesis will have the credibility of 5 “improbable”. In the case where the two informations are different, both informations conserve the credibility of 2 and the others remain with a credibility of 4. In the case of a third or more informations arriving to the fusion process, we can therefore derive a general procedure.

Let $Q(H_i)$ the number of informations $d = H_i$ that confirm the hypothesis H_i and given by a set of sources. Let us suppose that p sources give one information. Then

If $Q(H_i) = p$ the credibility of this hypothesis is 1 and the credibility of the others is 5

If $Q(H_i) \in \{1, \dots, p-1\}$ the credibility of this hypothesis is 2 and the credibility of the hypothesis with $Q(H_i)=0$ is equal to 4.

If no information is available about any of the hypothesis, $Q(H_i)=0$ for all of the hypotheses, the credibility is equal to 3.

4.1.2. probabilistic approach

The procedure in the previous section is not totally satisfactory and becomes ill suited when the number n of delivered informations is large. The main reason is that the evaluation of the credibility of a information can not to be restrained to a scale of five. In this section we propose the use of the probability theory which allows a continuous scale from 0 to 1. Let $P(H_i) = q/p$ the number of informations that confirm a hypothesis over the number of sources that give informations. The previous procedure can be rewritten as follows:

- If $P(H_i) = 1$ the credibility of this hypothesis is 1 and the credibility of the others is 5
- If $P(H_i) \neq \{0,1\}$ the credibility of this hypothesis is 2 and the credibility of the others hypotheses with $P(H_i)=0$ is equal to 4.
- If no information is available about any of the hypothesis the credibility is equal to 3.

In the framework of probability theory, an additional extension of the first procedure can be made by considering that the source not only provides an information of the form $d=H_i$ but also it provides an

evaluation of the credibility based on its own sources of information and knowledge. In this case, a source k provides information of the form $P(d_k = H_i/H_j)$ with $j \in \{1, \dots, n\}$. This information is not a unique decision but a vector of decisions with their probabilities. If many sources are available a fusion center can fuse the information given by the sources and calculate a probability of the form: $P(H_i/ d_1, \dots, d_n)$. Then the credibility can be given by:

- If $P(H_i/ d_1, \dots, d_n) = 1$ the credibility of this hypothesis is 1 and the credibility of the others is 5
- If $P(H_i/ d_1, \dots, d_n) \neq \{0,1\}$ the credibility of this hypothesis is 2 and the credibility of the hypotheses with $P(H_i/ d_1, \dots, d_n) = 0$ is equal to 4.
- If no information is available about any of the hypothesis therefore the credibility is equal to 3.

In practical situation the first and the third cases nearly never occur. We can adopt some heuristic approach to determine a threshold s so that the procedure become:

- If $P(H_i/ d_1, \dots, d_n) > s$ the credibility of this hypothesis is 1 and the credibility of the other hypothesis is 5
- If $P(H_i) \neq \{1-s, s\}$ the credibility of this hypothesis is 2 and the credibility of the hypothesis with $P(H_i) < 1-s$ is equal to 4.
- If no information is available about any of the hypotheses the credibility is equal to 3.

4.1.3. possibility approach

Thanks to the possibility theory framework, we can note that the definition of the credibility of one hypothesis is closely related to the credibility given to the other hypotheses in the same frame of discernment. For example in the first procedure if a hypothesis has a credibility of 1 the others have the credibility of 5. This consideration allows to rewrite the procedure in terms of possibility and necessity as follows:

- If $\Pi(H_i) = 1$ and $N(H_i) = 1$ then the credibility is equal to 1 “confirmed”
- If $\Pi(H_i) = 1$ and $0 < N(H_i) < 1$ then the credibility is equal to 2 “probable”
- If $\Pi(H_i) = 1$ and $N(H_i) = 0$ then the credibility is equal to 3 “possible”
- If $0 < \Pi(H_i) < 1$ and $N(H_i) = 0$ then the credibility is equal to 4 “doubtful”
- $\Pi(H_i) = 0$ and $N(H_i) = 0$ then the credibility is equal to 5 “improbable”

Where $\Pi(H_i)$ and $N(H_i)$ are the possibility and the necessity of the hypothesis H_i . The evaluation of these two terms is application dependent. A possible way to be coherent with the examples given above is the following:

- If no information is available then $\Pi(H_i) = 1$ and $N(H_i) = 0$
- If m is the number of independent sources and $p \leq m$ the number of sources affected to the hypothesis H_i then $\Pi(H_i) = 1$ and $N(H_i) = p/m$. for declared

hypothesis, and $\Pi(H_j) = 1 - N(\bar{H}_j)$ with $i \neq j$ for the hypothesis not declared by any sources.

4.2. Multi hypothesis approach Dempster-shafer theory

In the previous section the model adopted is a very simple one. We suppose the frame of discernment is closed and the decisions given by the sources have a one to one correspondence with each hypothesis of this frame. In the following, we adopt a model which is a more realistic one. We suppose the decision can take its value in a subset of hypotheses. For example the frame of discernment D represents all the terrestrial vehicles. A source may give an information of the form “the object is an armored vehicle” which represent an subset of D composed by the elementary hypothesis {AMX10, Char Leclerc, ...}. For this application framework we deal with the dempster-shafer theory. Let D the frame of discernment and d_i the decision given by a source which takes his value in 2^D . Then the operational evaluation is given by the following relations:

- If $PI(d_i) = 1$ $Cr(d_i) = 1$ then the operational credibility is equal to 1 “confirmed”
- $PI(d_i) = 1$ $0 < Cr(d_i) < 1$ then the operational credibility is equal to 2 “probable”
- $PI(d_i) = 1$ $Cr(d_i) = 0$ then the operational credibility is equal to 3 “possible”
- $0 < PI(d_i) < 1$ $Cr(d_i) = 0$ then the operational credibility is equal to 4 “doubtful”
- $PI(d_i) = 0$ $Cr(d_i) = 0$ then the operational credibility is equal to 5 “improbable”

In the relation above we make the distinction between the credibility given by the theory of Dempster Shafer and noted Cr and the operational credibility given by Stanag 2022.

This approach given by the theory of Dempster Shafer enable to consider the problem given by two sources working with different granularity of classification. For example one source may send an information of the form “this object is an AMX 10” and an other source may give an information of the form “this object is an armored vehicle” the tow source referencing the same object. In that case what information we have to consider and what operational credibility have we to give to this information. If we know that an AMX 10 is an armored vehicle the we can consider that the information “armored vehicle” is confirmed by the tow source an then have the rating of 1 and the information “AMX 10” have only a rating of 2. This phenomenon can be taken into account using the theory of Dempster Shafer. Some further work has to be made at this level.

4.3. The open world approach

Just now the rating of 6 was not use. The reason for using such rating is that the information given by a source is not referenced as an information corresponding

to our a priori knowledge. In this case the evaluation of the credibility of this information can not be judged. This case is foreseeing in the theory of Dempster Shafer by affecting a masse different of zero to the empty set.

5. General model for data fusion

5.1. Introduction

In the preceding section we have define in a more manner the notion of evaluation in the framework of the probability theory, the possibility theory and the Dempster Shafer theory. Let now study the use of this evaluation rating in a fusion process in the framework of probability theory. For a complete revue of the use of reliability in a fusion process see [9].

5.2. Probability model

Now we consider the architecture figure 1, where two sensors C_o and C_r are linked to a fusion center FC.

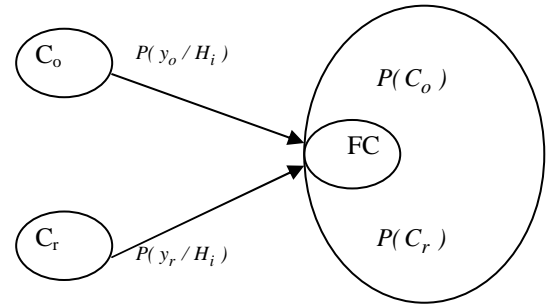


Fig 1. General architecture

Each sensor has its own model. Then it is possible for each sensor to send its credibility under the form, for example, of probability if the fusion center is an automatic one (figure 1), or under the form described by the STANAG 2022 (given the conversion in section 4.1) if the fusion center is an operator.

In this part, we suppose that the fusion center is an automatic machine that knows the models used for each of the sensor. Therefore it knows the limit of each of them and can deliver a reliability in the form of a probability $P(C_o)$ or $P(C_r)$. It follows that the fusion formula including the use of reliability is given by:

$$\begin{aligned}
 P(H_i / y_o, y_r) = & P(H_i / y_o, y_r, C_o, C_r) P(C_o \cap C_r) \\
 & + P(H_i / y_o, y_r, \bar{C}_o, C_r) P(\bar{C}_o \cap C_r) \\
 & + P(H_i / y_o, y_r, C_o, \bar{C}_r) P(C_o \cap \bar{C}_r) \\
 & + P(H_i / y_o, y_r, \bar{C}_o, \bar{C}_r) P(\bar{C}_o \cap \bar{C}_r)
 \end{aligned}$$

where the probability of each hypothesis is given by a formula of the form

$$P(H_i / y_o, y_r, \overline{C_o}, C_r) = \frac{P(y_o / H_i, \overline{C_o})P(y_r / H_i, C_r)P(H_i)}{P(y_o, y_r / \overline{C_o}, C_r)}$$

for example for the second term. The reliability coefficient of the sensors $P(\overline{C_o} \cap C_r)$ can be decomposed into individual reliability of the sensors thanks to the relation:

$$P(\overline{C_o} \cap C_r) = (1 - P(C_o))P(C_r)$$

For the likelihood of each hypothesis there is two term: The first one $P(y_o / H_i, C_o)$ is given by the sensor

thanks to his internal model. The second one $P(y_o / H_i, \overline{C_o})$ is the likelihood of the hypothesis H_i

knowing that the sensor is out of the range of its validity domain. This function is defined by the fusion center and depends of the strategy adopted. For more details of this fusion process in the framework of tracking see [7].

5.3. Discussion

The above given model tries to follow the principal guidelines used by operators in the framework of intelligence (STANAG 2022 and AJP 2.1) and tries to give some adjustments for an automatic fusion process. In the example given in section 5.2, we suppose that the credibility (equivalent to a probability in this example) is directly given by the source. The fusion center has to determine the reliability of the source and to establish an alternate strategy (minimum commitment, worst case, ...etc) to ensure credibility in the case the source is not reliable.

To make the idea more clear, we consider a very simple example with one source and one fusion center. In this case, after making an observation, the source can give a probability $P(H_i / y, C)$ that depends on its internal

model. The fusion center has to give the reliability $P(C)$ and the alternate probability $P(H_i / y, \overline{C})$. The result of

the fusion center is:

$$P(H_i / y) = P(H_i / y, C)P(C) + P(H_i / y, \overline{C})P(\overline{C})$$

In this case the minimum commitment strategy for the alternate probability is $P(H_i / y, \overline{C}) = 1/n$ where n is

the number of hypotheses in the discernment frame. In fact, if a source is unreliable $P(C)=0$ and whatever the provided credibility is, the result of the fusion process will be: $P(H_i/y)=1/n$. Which is the minimum commitment strategy in the framework of probability. Therefore, given the previous model, we can see that a the sentence "every piece of information produced by a

totally reliable source is not necessary true" can be modeled by the following value in the probability framework:

- $P(H_i/y,C)=0$
- $P(C)=1$

After the level of the fusion process, the result of this information is:

- $P(H_i/y,C)=1/n$

This means that false information given by a totally reliable source gives no information

6. Conclusion

The purpose of this paper is to formalize some "informal" STANAG recommendations on information evaluation in the fusion process. Some of the informal concepts underlying the STANAG recommendations have been given some formal interpretations. However, as it has been shown, no consensus exists yet on these definitions and more work is needed.

7. References

- [1] Annex to STANAG 2022 (Edition 8). North Atlantic Treaty Organization (NATO) *Information Handling Services*, DODSID, Issue DW9705
- [2] S. Konieczny and R. Pino-Pérez. *On the logic of merging*. In Principles of Knowledge Representation and Reasoning (KR'98), pages 488–498, 1998.
- [3] S. Konieczny and R. Pino-Pérez. *Merging with Integrity Constraints* In Proceedings of ESCQARU'99
- [4] C. Baral and S. Kraus and J. Minker and V.S. Subrahmanian. *Combining multiple knowledge bases*. IEEE Trans. on Knowledge and Data Engineering, 3(2), 1991.
- [5] L. Cholvy. Proving theorems in a multi-sources environment. Proceedings of IJCAI, 1993.
- [6] C. Liau. A conservative Approach to Distributed Belief Fusion. Proceedings of FUSION, 2000.
- [7] V. Nimier *Supervised Multisensor tracking algorithm* FUSION-98, International Conference on Multisource-Multisensor Information Fusion, Las Vegas, USA, July 6-9, 1998, pp 149-156.
- [8] V. Nimier, Introducing Contextual Information in Multisensor tracking Algorithms Advances in intelligent Computing, Springer, 1994.
- [9] Galina L. Rogova, V. Nimier. *Reliability in information fusion* Fusion 2004, Stockholm, june 28-july 1.
- [10] L. Cholvy, R. Demolombe, and A. Jones. *Reasoning about the safety of information : from logical formalization to operational definition*. In Lecture notes in Artificial Intelligence, number 869. Springer-Verlag, 1994