A model for an imperfect knowledge base for high-level information fusion experiments

A-L. Jousselme*, P. C. Costa‡, A. Akli*, G. Arcieri†

* CS Group, La Garde, France, anne-laure.jousselme@csgroup.eu
‡ George Mason University, Fairfax, VA, USA, pcosta@gmu.edu
† Centre for Maritime Research and Experimentation, Italy, gianfranco.arcieri@cmre.nato.int

***Abstract*—Evaluation of fusion algorithms constitutes a pivotal phase in the life-cycle of designing solutions for fusion problems. Benchmark datasets together with standardized evaluation criteria and metrics are thus needed to run proper experiments. When it comes to higher-level tasks such as situation assessment, datasets are conspicuously absent. Among the reasons are the lack of formal characterization of situation assessment, and the lack of structure to guide the collection of data at an appropriate level of semantics and granularity. In this paper, we propose a model for an imperfect knowledge base (IKB) to support high-level information fusion experiments. The model includes characterization of imperfect information from partially reliable sources. The concept of *infon* is an elementary piece of information which can bear on both concrete and abstract objects. The corresponding knowledge base encodes infons about physical objects of interest in a situation (*e.g.*, a vessel), about sources providing infons, and about other infons. The imperfection dimensions of information are captured and connected to reliability dimensions of the source. An implementation within a graph database is proposed, exemplified with data of a maritime scenario. An example of comparison of artificial and human agents executing a fusion task is shown. Future work are finally briefly discussed.**

***Index Terms*—Uncertainty; Situation assessment; Source quality; Information quality.**

I. Introduction

Evaluation of fusion algorithms is a critical step in the life-cycle of the design of solutions to fusion problems. Benchmarks and reference datasets are classically used to compare different solutions, analyze their sensitivity to some parameters, and properly predict errors. While the machine learning community relies for decades on the UCI[1] data repository for instance, the fusion community is less equipped in this respect. Still, relevant datasets for exercising multiple target tracking algorithms exist (*e.g.*, [1]). But when it comes to higher-level tasks such as situation assessment, datasets are critically missing, or they often exhibit features involving lower fusion tasks. For instance, the SYNCOIN dataset produced within the MURI project on hard and soft fusion[2] and made available to the Evaluation of Technologies for Uncertainty Reasoning (ETUR) Working Group[3] gathers data from heterogeneous sources (*e.g.*, images, radar contacts, intelligence reports, tweets, etc). It contains low level data (such as objects precise positions, raw images) which require going through all the chain from sensors' output to the focused task of situation assessment. Similarly, the VAST[4] challenge datasets include highly heterogeneous data with some data requiring low level process.

A benchmark dataset for high-level information fusion experiments would need specific features to enable focusing on the targeted problem. In particular, the underlying model should have the ability to handle and capture meta-information about uncertainty, about the sources and possible links, about links between objects, etc.The purpose of this work is to define a model for an imperfect knowledge base (IKB) to store and retrieve data suitable for high-level fusion experiments. It should (minimally) meet the following requirements:

- capture meta-information about source quality;
- distinguish between source and information quality dimensions;
- establish formal connections between information and the source providing it;
- capture multi-source estimations about attribute values;
- capture expressions of uncertainty (non-specificity, confidence, randomness, etc);
- be agnostic to the type of source (humans, sensors, classifiers, etc);
- handle different domains and types scales (nominal, numerical, etc).

The ultimate goal is to provide a structured way for storing (and generating) data for high-level fusion experiments, at the desired granularity and semantic levels, capturing the required ingredients for such tasks in terms of uncertainty expressiveness and situation description.

In Section II, we provide some background about higher-levels of information fusion tasks and situational awareness, highlighting their specific features, compared to lower-level tasks. We also present the challenge in evaluating the impact of uncertainty representation and handling in this context. We propose in Section III a model for imperfect knowledge suitable to capture multi-source information and its semantics for higher-level fusion tasks. This model is exemplified in Section IV on a maritime use case with implementation in

[1]`archive.ics.uci.edu`
[2]`www.eng.buffalo.edu/ nagi/MURI/MURI/Welcome.html`
[3]`eturwg.c4i.gmu.edu`
[4]Visual Analytics Benchmark Repository

a graph database. Conclusions are drawn in Section V along sketches for future work.

II. Background

Before identifying relevant features of high-level information fusion, we first introduce the JDL fusion levels and the concepts pertaining to uncertainty representation and reasoning.

A. *High-level information fusion and situation awareness*

The semantic growth from detection or tracking tasks to identification or threat assessment tasks, is naturally reflected in the Joint Directors of Laboratories (JDL) functional model of Data Fusion which drives the research and discussions in the field since its inception in 1987 (see [2] for a survey on the evolution of the model).

That semantic growth is reflected in the so-called levels of fusion: As the tasks get closer to the decision maker, data to be processed get farther from signals to include more meaningful features. The techniques to support processes at different levels may not meet the same expectations: More reasoning, interpretation and explanation is expected for higher level tasks than for lower level tasks.

The shift of focus from individual (physical) objects (Level 1 of the JDL) to the more abstract object of situation (Level 2) renders difficult the formalization of that task. Some clarification of the abstract concept of "situation assessment" was eventually found in Mica Endsley's model of situation awareness (SAW) [3], where SAW involves "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future". These three cognitive tasks leading to situation awareness are embedded, denoting some constant re-perception and re-comprehension to get to the projection on future states. Situation assessment is the process by which the decision maker gains situation awareness, *i.e.* a mental state, the state of knowledge required to achieve a given task, or mission.

Compared to Levels 0 and 1 of the JDL, for which expectations in terms of target characterization are usually captured by the Tactical Picture (kinematic features, categories of targets), Levels 2 and 3 do not have the same level of formalization and characterization. The main contributions to formalization of higher levels of fusion include notably Kokar *et al.*'s work on the definition of the core SAW ontology [4], which provides precise mathematical relationships between situation elements, defining the principal classes of entities of Situation, Situation object, Goal, Relation, Event, Attribute and Value. The Situation Theory Ontology (STO) [5] provides formal grounds to address the situation assessment problem, defining concepts such as Situation, Attribute, Value, Individual, Infons. Although it enables reasoning about situations, the STO does not capture uncertainty concepts.

B. *Representation and reasoning under uncertainty*

Handling uncertainty in fusion solutions is indeed a major issue which generates many questions such as what *uncertainty* means, where it comes from, on what it bears, how to interpret the associated numerical values or measures, how to distinguish between its different nature, etc. Uncertainty in Levels 0 and 1 tends to be aleatory in nature (*i.e.*, intrinsic randomness), while in Levels 2 and above it is mostly epistemic (*i.e.*, insufficient, imperfect knowledge). Thus, the apparatus used to represent and quantify uncertainty in Levels 0 and 1 is mostly aimed at aleatory uncertainty and ill-equipped to capture the key uncertainty aspects that exist in Levels 2 and above. The URREF ontology [6] was designed to address this issue by formalizing the concepts needed to represent uncertainty in higher JDL Levels. For instance, it defines: Uncertainty Nature, Uncertainty Derivation, Uncertainty Type, Uncertainty Theory. Discriminating between the different facets of uncertainty has the advantage of (1) avoiding untimely uses of definitions and models of uncertainty, (2) clarifying links with the already well developed logics of knowledge and belief, and (3) providing guidelines for the selection of the appropriate mathematical model to process uncertainty-based information [7].

Figure 1 displays the top-level concepts and relationships of the representation and reasoning under uncertainty domain, some of them captured in the URREF ontology. An Uncertainty Theory has two main components: Uncertainty Function and Calculus. An Uncertainty Function bears on Uncertain Variables which are of different Uncertainty Nature, either aleatoric or epistemic.Examples of Uncertainty Functions are probability functions, belief or plausibility functions, possibility or necessity functions, fuzzy sets, imprecise probabilities or random sets just to name a few. Uncertainty Functions captures different Uncertainty Types, typically non-specificity, fuzziness or discord, referring for instance to Klir and Yuan's typology [8]. Calculus of the Uncertainty Theory encompasses all inference rules and operators to be possibly applied on Uncertainty Functions.

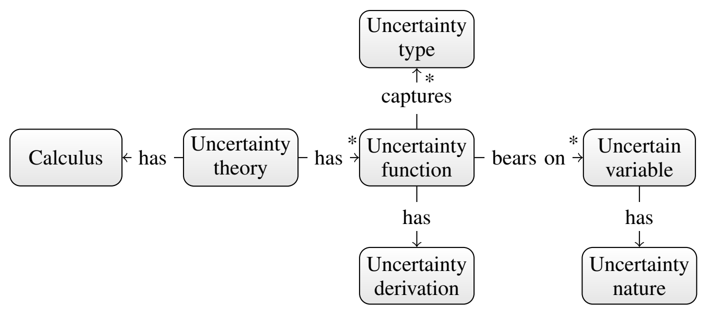

Fig. 1. Main concepts of the Representing and reasoning under uncertainty domain [6], [7]. One-to-many relationships are tagged with *

Although this characterization of uncertainty is important at any level of processing, the diversity is less present at lower levels while it is critical at higher levels. To be able to exercise fusion solutions, benchmark datasets should capture these different facets of uncertainty.

C. Features of high-level information fusion

To sum-up, we provide below a (non-exhaustive) list of features which characterize high-level fusion tasks, and that should be reflected in a benchmark dataset. Compared to Level 1, for a Level 2 fusion problem:

- relevant individual objects are already detected, identified and monitored (*i.e.*, tracked);
- relevant attributes of individual objects are identified and estimated (by several sources);
- association issues could be overlooked (or at least they are less a focus than for Level 1);
- uncertainty is diverse, in nature (both aleatoric and epistemic uncertainty need to be handled), in interpretation (both objective and subjective uncertainty can be expressed by sources), in type (imprecision, vagueness or randomness);
- heterogeneity of sources: sources can be sensors and humans, with highly different meanings for reliability assessment (a radar may provide poor detection but does not have any intent to spoof, compared to a human);
- semantic upgrade from low-level acts on three aspects:
  - n-ary: attributes can bear on several individual objects;
  - meaning: attributes are meaningful to humans (*e.g.*, Radar-Cross-Section translated into size notion);
  - granularity: attribute domains are coarse-grained (*e.g.*, finite number of levels for speed rather a continuous values), as is the timescale (generally useless to get positional information at each millisecond).

In the next section, we propose a model which enables capturing data suitable for high-level information fusion tasks.

III. A MODEL FOR IMPERFECT KNOWLEDGE

We first identify the main quality dimensions for sources and information quality. We then propose a model for an imperfect knowledge base which enables capturing these dimensions.

A. From source to information and vice-versa

By definition, a source of information provides information and thus, should be assessed primarily according to its capacity in providing information [9]. The quality of a source (possibly along different dimensions) describes "how good" the source is expected to perform, or "how much" we can consider its output as valid, true, good. While it seems relevant to evaluate source and information quality independently from each other, both aspects are necessarily closely linked, as it is expected that a source of "good" quality will deliver information of "good" quality. Moreover, in case of direct access to information from the source, some information quality dimensions can be transferred to the source. Hence, during a training phase, the quality of the information output by a source is assessed on the long run of a series of trials, and a synthesis of it is further *transferred* to the source as a more or less perennial quality dimension (see Fig. 2).

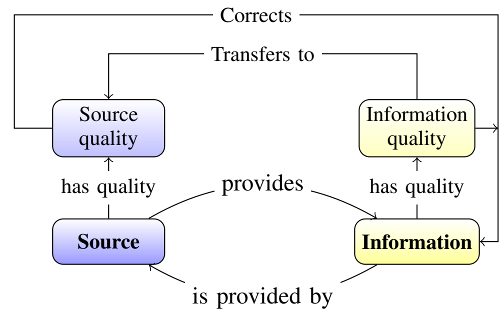

Fig. 2. Information quality and source quality relationships in case of learning and subsequent correction processes

At the correction step, and hence before further use of the information item in the fusion process, the source quality (either learnt or provided as meta-information by other sources) is used to *correct* the information item. Besides, quality factors can be used to simply qualify the output information or as a predictive parameter of the estimation produced.

B. Source and information quality dimensions

In accordance to standard information evaluation procedures (*e.g.*, NATO AJP 2.1), source quality and information quality dimensions are assessed independently, acknowledging that a "good" source can provide "bad" information at a specific instant in time. In [10], the dimensions are named *reliability* for the source quality and *credibility* for the information quality. The reliability refers to past tries of the source, where the correctness of the results or answers by the source have been validated. The credibility of information refers to a comparison with other pieces of information provided by independent sources, without referring to any notion of correctness or truth. The evaluation of source quality is instrumental to a sound fusion process, as sources are in general only partially reliable. For instance, if the source is highly reliable, its information may be reinforced, while if it is not reliable its information can be discounted and even discarded by some correction mechanisms (*e.g.*, [11]). In information fusion correction procedures, some meta-information about the source quality is either assumed to be known beforehand, or estimated based on training sets with some ground truth information. Intelligence analysts implement intuitive similar mechanisms based on the previously rated credibility of information and reliability of the source.

Figure 3 displays a top-level model for source and information quality dimensions, with three types of relationships: *impacts* (*is impacted by*), *provides* (*is provided by*) and *has quality* (*is quality of*). This selection includes the main dimensions typically considered but other dimensions (criteria) can be considered. The blue part of Fig. 3 corresponds to source quality dimensions while the yellow part displays information quality dimensions of an information item provided by the source at a given instant in time.

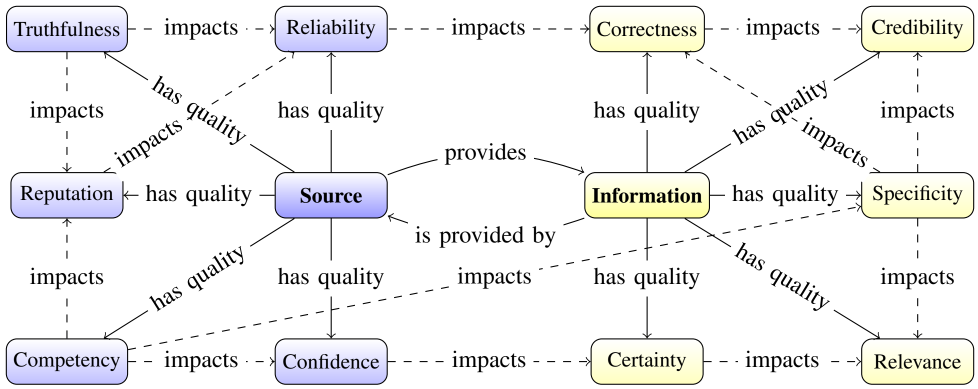

Fig. 3. Source quality dimensions (in blue) for a source s and information quality dimensions (in yellow) provided by s. Three types of relationships are displayed

Information quality - The quality of information (along different dimensions) describes "how good" information is to support the task at hand. While the assessment can or not consider its origin (*i.e.*, the source), the concepts below describe information quality without access to its origin. They include unary concepts (Specificity, Certainty) and n-ary concepts in relation with other information items (Correctness, Credibility, Relevance).

- CORRECTNESS is the "quality of being in agreement with the true facts or with what is generally accepted"[5].
- SPECIFICITY can be defined as "the degree of uniqueness that a function shows towards an element, and an element towards a function, in discourse" [12]. It assumes some granularity of reference and essentially relates to the cardinality or interval length of the set bearing information (*e.g.*, speed between 0 and 5 knots, type is either a fishing vessel or a service ship).
- CERTAINTY represents how much the agent thinks the statement is true (correct). For instance, "I am certain to a degree of 0.8 that the type of the vessel is a cargo", or the probability of the vessel being a cargo is 0.97 (as an output of a classifier).
- CREDIBILITY is "the fact that someone [or something] can be believed or trusted"[6]. Although credibility can thus be assigned to a source, we will keep this criterion for information in accordance to the AJP2.1, but we admit it is just a matter of convention. An information item's credibility is thus assessed relatively to other items (including possible background knowledge), considering some agreement between them.
- RELEVANCE is "the degree to which something is related or useful to what is happening or being talked about"[7]. This thus refers to the impact of the information item on our current state of knowledge, constrained by our needs.

Some relationships between these information quality dimensions exits. For instance:

- SPECIFICITY impacts CREDIBILITY (*i.e.*, the less specific an item, the more likely to be credible) but also correctness (the less specific an item, the more likely to be correct.
- CERTAINTY impacts RELEVANCE in the sense that the less certain an item, the less relevant (useful) the item.
- CORRECTNESS impacts CREDIBILITY as a correct information item is more likely to be credible than an incorrect one, assuming that it is compared with enough items from independent sources.

From the standpoint of information quality, SPECIFICITY and CERTAINTY are both outbound only concepts, which are not impacted by other information quality dimensions, but that impact others. Certainty and specificity form *informativeness*: An information item is highly informative if it bears on a non-empty single (maximally certain) singleton (maximally specific). Instead, CREDIBILITY and RELEVANCE are inbound only concepts not impacting any others. CORRECTNESS however both impacts and is impacted by others.

Source quality - Source quality can be assessed either by direct access to information provided by the source, or by means of external factors without direct access to that information. In the first case, the assessment can be for instance based on past experiments with the source, where its output has been compared to some ground truth and validated. This is the case for instance for sensors characterized by Receiver Operating Characteristics (ROC) curves, for classifiers or detectors characterized by confusion matrices but also for humans who gained trust of their interviewers, after a series of validated testimonies.

The source quality can also be assessed based on contextual factors, without direct access to information. For instance, a witness under influence may be expected to lie, a radar under bad weather is expected to under-perform. In this case, the quality of the source would not be assessed directly by the information provided, but could rather be, based on similar sources (*i.e.*, of the same type) used under similar conditions. Regardless the context, a source can also be assessed based on its reputation, such as for instance some governmental

[5] https://dictionary.cambridge.org/dictionary/english/correctness
[6] https://dictionary.cambridge.org/dictionary/english/credibility
[7] https://dictionary.cambridge.org/dictionary/learner-english/relevance

websites. This is a case of second-order assessment, where we did not have access to data ourselves but rather rely on others' assessment. Finally, a source can also be assessed based on its disposition to provide good information, such as its truthfulness (*e.g.*, a witness who may not be free to reveal some facts). It can be contextual or not.

- CONFIDENCE of the source is defined as "the quality of being certain of your abilities or of having trust in people, plans, or the future"[8], and "certainty" could be seen as a synonym. In case of a source, (self-)confidence refers thus to its quality of being certain of its statement. It is a level to which the source is confident that its statement is correct. It can be expressed as a degree of belief. Although it is particularly suited for human sources who can express some forms of belief about their statements ("I think that ..."), self-confidence can be assigned to some artificial sources (*e.g.*, a classifier) based, for instance, on some possible internal probabilistic assessment (or scores).
- COMPETENCY is "an important skill that is needed to do a job"[9] or "the ability to do something well" (competence). For a source, competency refers to its *a priori* general ability to provide a specific type of information. For instance, an engineer may not be competent to provide a medical diagnosis. As reputation, competency is a (rather) perennial factor but is relative to the question asked. A source may be competent for X and not for Y.
- TRUTHFULNESS is "the quality of being honest and not containing or telling any lies"[10]. It is a synonym of sincerity. It is typically a human feature, as humans are able to lie or deliberately spoof.
- REPUTATION is defined as "the general opinion that people have about someone or something"[11]. In case of a source, reputation is thus a general assessment about the ability of a source to provide "good" (or "bad") information. Reputation is an external assessment not directly related to the information provided by the source. It could be based on its type for instance (*e.g.*, artificial intelligence vs human), origin (country, institution, brand), on global characteristics, on observed past behavior, on contextual use.
- RELIABILITY is defined as "the quality of being able to be trusted or believed because of working or behaving well"[12]. It explicitly expresses a link with a third party, the trustor or believer. The reliability of the source refers to its ability to be relied on. The reliability concept is very close to trust, and is actually the dimension establishing the link between the source (the trustee) and the agent (the trustor).

As for information quality dimensions, some relationships can be established. For instance:

- REPUTATION can be impacted by the TRUTHFULNESS of the source and its COMPETENCY. It conveys the ideas of accumulation of evidence toward some behavior, either directly or reported by others. REPUTATION impacts thus directly RELIABILITY.
- COMPETENCY may impact the CONFIDENCE of the source in its statement (if it is aware of its competency). A lack of COMPETENCY would rather lead to uncertain statement such as "I am not sure that ...". COMPETENCY also affects REPUTATION.
- TRUTHFULNESS impacts both REPUTATION (when assessed on the long run) but also directly impacts RELIABILITY even if REPUTATION is not affected yet. TRUTHFULNESS can be affected by external factors we can gather under "context" in which sources may be under pressure to not share some information.
- The way RELIABILITY is assessed may depend on several factors, such as the ability of the source to provide consistent correct results under specific conditions, which would imply access to training data. The RELIABILITY may also be estimated based on the REPUTATION of the source, its TRUTHFULNESS or the contextual conditions of operations.

Finally, as expressed earlier, information and source quality are linked. For instance, assessing correctness (if ground truth is available) or credibility (if information from other sources is instead available) of information items provided by a source would provide some assessment of the source's reputation, either through its truthfulness or its competency. Hence:

- CORRECTNESS $\rightarrow$ REPUTATION – On the long run, a source providing correct information will build a better reputation than one which makes mistakes or has been proved to lie;
- CREDIBILITY $\rightarrow$ REPUTATION – When ground truth is not accessible, a source providing non credible information will impact negatively its reputation.

On the other hand:

- RELIABILITY $\rightarrow$ CORRECTNESS – A reliable source is likely to provide more correct information than an unreliable one. Of course, this is not guaranteed and should not prevent any cautious approach to be applied. Essentially, and by definition, reliability is the main factor connecting source and information, and acts a kind of "mediator".
- COMPETENCY $\rightarrow$ SPECIFICITY – A competent source is more inclined to provide specific answers.

The information correction mechanisms are generally agnostic to the type of source and simply rely on a quality factor known as "reliability", while many behavior and dispositions to provide information of "good" or "bad" quality can be encoded [13]. The characterization of the quality dimensions above may help defining specific sources behaviors in light of available meta-information.

[8] https://dictionary.cambridge.org/dictionary/english/confidence
[9] https://dictionary.cambridge.org/dictionary/english/competency
[10] https://dictionary.cambridge.org/dictionary/english/truthfulness
[11] https://dictionary.cambridge.org/dictionary/english/reputation
[12] https://dictionary.cambridge.org/dictionary/english/reliability

C. A data model for infons

We can now derive our model for capturing imperfect knowledge in an high-level information fusion context. The model is articulated around the five main concepts of Object, Infon, Attribute, Value and Domain. Figure 4 displays the top-level model. Let us assume we observe a scene

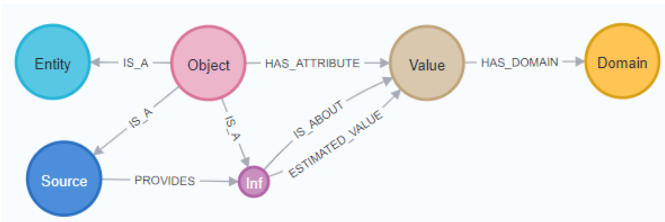

Fig. 4. Imperfect Knowledge Base top level model

(described by a volume of space and time) in which some entities (concrete objects) evolve and about which some sources capture and provide information. Our goal is to understand what is gong on, hence to assess the situation. Objects can of different nature: Entity objects are concrete objects, actors in the scene and are the focus of the estimation process (we want to understand what they do, who they are, etc). Source objects provide information about the situation and about entities evolving in it. Objects can have possibly the two roles. An Infon is an elementary abstract object carrying informational content. An Object is characterized by a set of attributes (or features), which can be evaluated (estimated, measured), represented here by a series of relationships HAS_ATTRIBUTE.

Objects which provide Infons are Sources (of information). An Infon is a tuple $\phi = (o, a, v, t)$ of an object o, an attribute a, a value v and a timestamp t. In natural language this reads: Infon ϕ = "the Attribute a **of** Object o **is** (has Value) v at time t". Values for objects' attributes belong to some Domain. Depending on the source providing the infon, the same attribute value can be expressed on different Domains.

Infons can bear on attributes of both physical and abstract objects with either numerical, nominal or fuzzy values. For instance, "ϕ_1: the Location **of** Vessel #4328 **is** $(52.3, 24.1)$", "ϕ_2: the Reliability **of** Analyst Anna **is** B". Infons carry thus the result of an evaluation and represent elementary constructs to be further communicated, and processed. Possibly, o is a set of objects, enabling thus infons to carry information about relationships between objects.

An infon is itself an Object, characterized by several attributes, as defined in Section III-B. Essentially, infons have basic attributes such as Correctness, Specificity, Certainty, while other dimensions can be estimated such as Credibility or Relevance. Source quality can be expressed along the dimensions introduced in Section III-B, such as Reliability, Truthfulness, Competence. The list can be enlarged as needed.

As an infon is also an object, it can bear on an infon attribute value. Nested infons allow in particular to express some uncertainty notions about other infons (*e.g.*, probability of Vessel A type being a Trawler), following the same model than the one used to evaluate physical or abstract quantities. In particular, we can express some certainty about an infon. For instance "ϕ_3: the Certainty **of** ϕ_1 **is** 1", as provided by Source 3:

$$\phi_3 = (\phi_2, a_3, v_3, t) = (((o_1, a_1, v_1, t), a_2, v_2, t), a_3, v_3, t)$$

Furthermore, second-order uncertainty can also be expressed. For example: "The Specificity (a_3) **of** the probability (a_2) **of** "Position **of** Vessel A **is** (F3, G3)" (ϕ_1) **is** 0.9 (v_2), **is** [0.8,1] (v_3)", that could be translated into an imprecise probability. Equivalently, the model allows to capture quality assessments about sources (as illustrated in Section IV, in Fig. 7).

IV. Illustration

In order to illustrate some aspects of the high-level information fusion imperfect knowledge base (IKB), we provide an implementation with the Neo4j[13] graph database, exemplified on the maritime scenario described in [14].

The situation to be assessed is the following one: *The contact with a fishing vessel from RightLand has been lost and the Watch Officer (WO) from LeftLand now is facing with two unidentified tracks, called Vessel A (in LeftLand) and Vessel B (in RightLand), as the only two possible locations for the missing vessel. The goal of the WO is to estimate the intent of the vessel: If it crossed the boarder, it is assume it has a bad intent. Several heterogeneous sources of information combine cooperative and non-cooperative sensors, which output is processed either by algorithms or by humans, plus purely human sources.*

A. Importing data

We first import data from the Risk Game [14], as a series of 208 basic infons[14] bearing on the two vessels' attributes. Each infon is tagged with its correctness, specificity, and the certainty expressed by the source's confidence in its statement. A total of 1932 nodes and 3314 relationships, with 443 infons, have been created for this illustrative example.

TABLE I
Risk Game IKB features

Nodes	1932
Relationships	3314
Labels	32
Relationship Types	24
Property Keys	6

Figure 5 illustrates an excerpt of the IKB with two vessels, observed by three sources, providing 4 infons about the type of the two vessels. Attributes of the two vessels describe ground truth information: Vessel A HAS_TYPE Trawler, HAS_LOCATION F4, HAS_LENGTH 30 (meters), HAS_FLAG RightLand, etc. Moreover, Vessel A HAS_INTENT

[13] `neo4j.com`

[14] Infons correspond to "pieces of information" in [14].

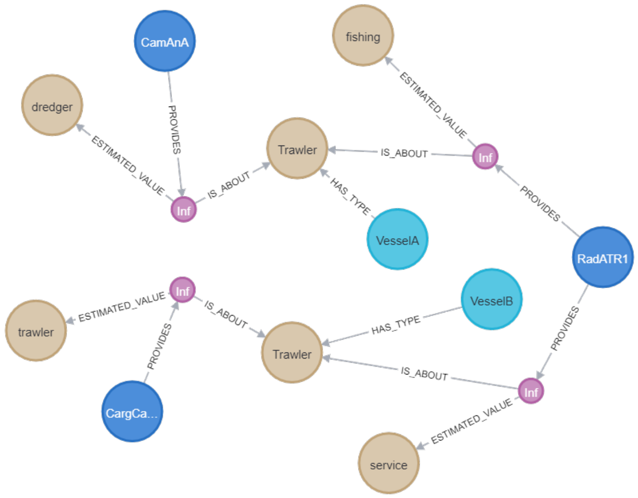

Fig. 5. Excerpt of the database for two vessels observed by three sources

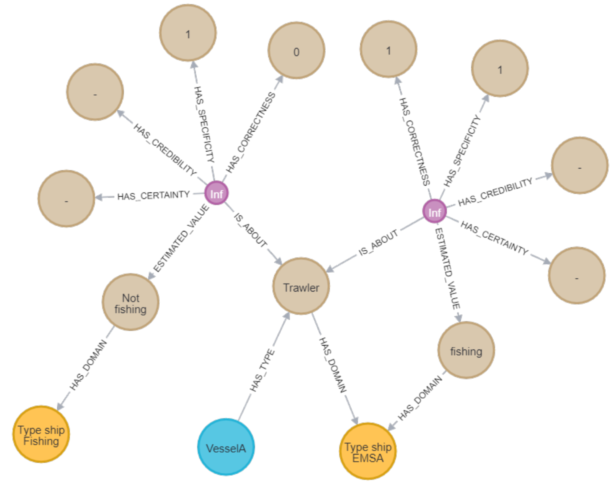

Fig. 6. Two infons defined on different domains with different quality attributes

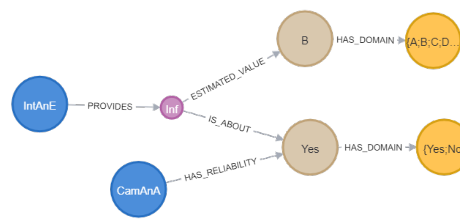

Fig. 7. One source providing estimation about the reliability of another source

Bad, HAS_OPPORTUNITY High, HAS_CAPACITY Moderate, HAS_THREAT Yes, meaning that in our scenario, this vessel is actually a potential threat. Vessel B has similar attributes but, importantly, has no bad intent and HAS_LENGTH 20 (meters) which distinguishes it from Vessel A. Only property HAS_TYPE is displayed in the graph. Coverage (which is a feature of Competency for the sources - see Section III-B) can be expressed in terms of the vessels seen by the sources: A and B for the radar, only A for the camera and only B for the cargo. Source 1 (a radar with associated tracker), named RadATR1, provides INFONS about the type of the two vessels. A camera analyst provides INFONS about Vessel A type, and a cargo captain about Vessel B type. INFONS have all ESTIMATED_VALUES about the type of one of the two vessels.

In Figure 6, two infons about the type of Vessel A are displayed: They differ by the domain on which they report their value (EMSA[15] type for the ground truth and one infon, and a binary domain (Fishing; Not fishing) for the other infon. Quality dimensions are displayed. One infon is correct (ESTIMATED_VALUE = fishing, HAS_CORRECTNESS = 1) and the other one is false (ESTIMATED_VALUE = Not fishing, HAS_CORRECTNESS = 0). They are both specific relatively to their domain (HAS_SPECIFICITY = 1). And they have no certainty nor credibility value yet.

Figure 7 shows an example of sources' quality dimensions, estimated by another source. In this case, the reliability of the camera analyst, which is truly reliable, is estimated as reliable of level B (according to the standard AJP2.1 scale) by an intelligence analyst. We provided here only examples of the expressiveness of the IKB model, illustrating features fitting to high-level information experiments. We sketch below an example of use of the IKB for exercising fusion solutions.

[15]European Maritime Safety Agency

B. Data processing

Data from the IKB can be read and processed with different fusion solutions. The encoding of basic infons can be understood by both machines and humans. We present here such an example, where infons are processed by either a human, or an artificial FUSION METHOD implemented within evidence theory [15] as the UNCERTAINTY THEORY. The Risk Game [14] aims at capturing the impact of information quality on human judgement. To this aim, players are briefed on a scenario of maritime surveillance (see beginning of Section IV), and on their role as Watch Officer to find out which among Vessel A and Vessel B corresponds to the missing vessel. Players are presented a series of physical cards with information about the two vessels as provided by several sources, mixing sensors and humans. After selecting the vessel and attribute of interest (*e.g.*, type of Vessel B), the quality is randomized by a dice roll. Each information card corresponds to an infon (*e.g.*, type of Vessel B is Trawler) with quality dimensions of CERTAINTY, SPECIFICITY and CORRECTNESS. Only certainty and specificity are perceptible by the player, the correctness is hidden. The player processes then a series of infons and assesses her/his belief toward the two possible events on a scale between 0 and 1 with 0.2 steps (see the circle-shape curves in Fig. 8). The artificial fusion solution (artificial agent) is implemented as an evidential network [16], [17] with a dedicated source model [13] and uncertain variables focusing on the two entities' attributes (position, speed, heading, type, size). Let X be an UNCERTAIN VARIABLE of interest which true state in denoted by x^*. Infons are encoded by simple mass functions as UNCERTAINTY FUNCTIONS over the corresponding UNCERTAIN VARIABLES, such that:

$$m(E) = \alpha, \;\; m(\mathcal{X}) = 1 - \alpha, \;\text{ with } E \subset \mathcal{X} \text{ and } \alpha \in (0, 1)$$

and $\sum_{E \subseteq \mathcal{X}} m(A) = 1$, with α being the degree of belief (or certainty) assigned to the fact that x^* belongs to E, and read from the certainty value of the corresponding infon. At each step, mass functions are combined conjunctively and the global conflict between infons is computed as the resulting mass on the empty set, $m(\emptyset)$.

Fig. 8 displays comparative human and artificial assessments toward the two hypotheses A (Vessel A is the missing vessel) in red and B (Vessel B is the missing vessel) in blue. BelA and BelB are the player's judgements displayed with circle shape, while the pair $(\text{Bel}(E); \text{Pl}(E))$, $E \in \{A, B\}$, of unnormalized belief and plausibility values are displayed with square and diamond shapes respectively.

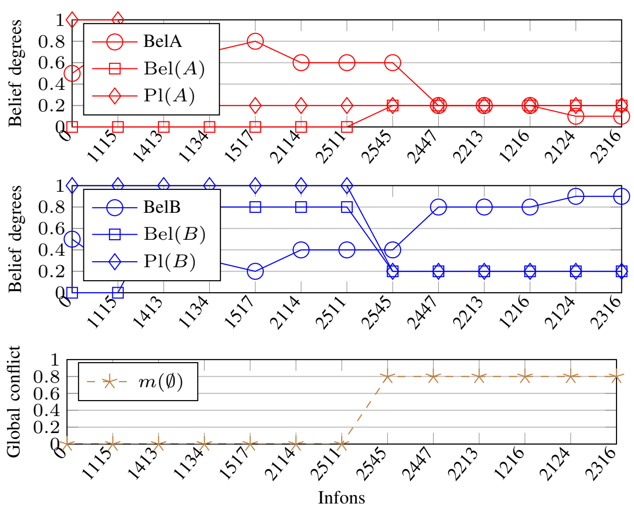

Fig. 8. Artificial versus human reasoning: Case of false information leading to false final belief; the conflict computed by the artificial solution would have raised an alert

Comparing the human and artificial agent, we observe that the player reaches a final state of certainty (0.9 for B and 0.1 for A) while likely the information received was wrong, as shown by the global conflict displayed in the bottom graph in brown. The artificial agent itself reaches a state of high epistemic uncertainty. The two agents are made OBJECTS in the IKB, so that they become sources providing infons themselves, available for further comparison or processing.

V. CONCLUSIONS

We described in this paper a model for an imperfect knowledge base suitable to high-level information experiments. We built an instance of a corresponding graph database, capturing data about a maritime surveillance use case. Such a model is well suited to the complexity, diversity and required expressiveness of a high-level fusion problem. We illustrated elements of the model around the different concepts of OBJECT, ENTITY, SOURCE, INFON, DOMAIN, VALUE and a series of attributes relationships capturing quality about the sources, infons and entities. The INFON conveys the basic elements of information to be encoded by uncertainty functions in fusion solutions, but as well as by human agents. The same encoding was used to feed ChatGPT prompts in [18]. In our ongoing work, we are using this model to cast relevant datasets such as the ones already made available by the ETUR working group. The intent is then to publish both the corresponding databases with dedicated queries and already extracted data, so others can replicate our results, and enable benchmark experiments for high-level fusion to be conducted. Other research will include refining the model and analyzing its complexity.

REFERENCES

[1] B. Le Saux, N. Yokoya, R. Hänsch, and M. Brown, "Data fusion contest 2019 (dfc2019)," 2019. [Online]. Available: https://dx.doi.org/10.21227/c6tm-vw12

[2] A. N. Steinberg, C. L. Bowman, F. E. White, E. P. Blasch, and J. Llinas, "Evolution of the JDL model," *Perspectives on Information Fusion magazine*, vol. 6, pp. 36–40, jun 2023.

[3] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human Factors*, vol. 37, no. 1, pp. 65–84, 1995.

[4] C. J. Matheus, M. M. Kokar, and K. Baclawski, "A core ontology for situation awareness," in *Proceedings of the 6th Annual Conference on Information Fusion*, Cairns, Qc, Australia, July 2003, pp. 545–552.

[5] M. M. Kokar, C. J. Matheus, and K. Baclawski, "Ontology-based situation awareness," *Information Fusion*, vol. 10, no. 1, pp. 83–98, 2009, sI on High-level Information Fusion and Situation Awareness.

[6] P. Costa, A.-L. Jousselme, K. Laskey, E. Blasch, V. Dragos, J. Ziegler, P. de Villiers, and G. Pavlin, "URREF: Uncertainty representation and reasoning evaluation framework for information fusion," *Journal of Advances in Information Fusion*, vol. 13, no. 2, pp. 137–157, Dec. 2018.

[7] A.-L. Jousselme, P. Maupin, and E. Bossé, "Uncertainty in a situation analysis perspective," in *6th Int. Conf. on Information Fusion*, Cairns, Australia, July 2003, pp. 1207–1214.

[8] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Upper Saddle River, NJ: Prentice Hall International, 1995.

[9] A.-L. Jousselme, A.-C. Boury-Brisset, B. Debaque, and D. Prévost, "Characterization of hard and soft sources of information: A practical illustration," in *Int. Conf. of Information Fusion*, Spain, 2014.

[10] STANAG 2191, "Allied Joint Doctrine for Intelligence Procedures - AJP-2.1," North Atlantic Treaty Organization, Brussels, Belgium, 2016.

[11] F. Pichon, D. Mercier, E. Lefèvre, and F. Delmotte, "Proposition and learning of some belief function contextual correction mechanisms," *International Journal of Approximate Reasoning*, no. 72, pp. 4–42, 2016.

[12] J. R. Trillo, "A mathematical model for the analysis of variation in discourse," vol. 37, no. 3, pp. 527–550.

[13] F. Pichon, D. Dubois, and T. Denœux, "Relevance and truthfulness in information correction and fusion," *International Journal of Approximate Reasoning*, vol. 53, pp. 159–175, 2012.

[14] A.-L. Jousselme, G. Pallotta, and J. Locke, "Risk game: Capturing impact of information quality on human belief assessment and decision making," *Int. Journal of Serious Games*, vol. 5, no. 4, pp. 23–44, 2018.

[15] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[16] P. P. Shenoy, "Valuation-Based Systems: A framework for managing uncertainty in expert systems," in *Fuzzy Logic for the Management of Uncertainty*, L. A. Zadeh and J. Kacprzyk, Eds. John Wiley and Sons, 1992, pp. 83–104.

[17] A.-L. Jousselme, N. Ben Abdallah, and F. Pichon, "A Risk Game formalisation in support to Maritime Situation Awareness automation: Analysis of reasoning profiles," NATO STO Centre for Maritime Research and Experimentation, Tech. Rep. CMRE-FR-2016-011, June 2016.

[18] A.-L. Jousselme, P. de Villiers, A. de Freitas, E. Blasch, V. Dragos, G. Pavlin, P. Costa, K. Laskey, and C. Laudy, "Uncertain about ChatGPT: enabling the uncertainty evaluation of large language models," in *Proc. of the Int. conference of Information Fusion*, Charleston, South Carolina, Jun. 2023.