Ontology Alignment using Relative Entropy for Semantic Uncertainty Analysis

Erik P. Blasch
Defence R&D Canada-Valcartier
2459 Pie-XI Blvd. North
Québec City, QC G3J 1X5
erik.blasch@drdc-rddc.gc.ca

Éric Dorion
Defence R&D Canada-Valcartier
2459 Pie-XI Blvd. North
Québec City, QC G3J 1X5
eric.dorion@drdc-rddc.gc.ca

Pierre Valin, Eloi Bossé
Defence R&D Canada-Valcartier
2459 Pie-XI Blvd. North
Québec City, QC G3J 1X5
eloi.bosse@drdc-rddc.gc.ca

Abstract – The development and use of many diverse ontologies to support the representational needs of different sources and different contexts is common and necessary. However, the increased sharing of databases implementing heterogeneous ontologies pose the problem of ontological alignment. Ontology alignment typically consists of manual operations from users with different experiences and understandings and limited reporting is conducted in the quality of mappings. To assist the International Organization for Standards (ISO) in standards development for information and data quality assessment, we propose an approach using relative entropy for semantic uncertainty analysis. Information theory has widely been adopted and provides uncertainty assessment for quality of service (QOS) analysis. Quality of information (QOI) or Information Quality (IQ) definitions for semantic assessment can be used to bridge the gap between ontology (semantic) uncertainty alignment and information theory (symbolic) analysis. Pragmatically aiding users of the shared ontologies requires assessments of the cognitive mental models, recognition of semantic classifications, and action over timeliness, throughput, confidence, and accuracy of the translations. In this paper, we explore issues of ontology uncertainty alignment utilizing the elements of information theory (KL divergence or relative entropy). A maritime domain situational awareness example with ship semantic labels is shown to demonstrate ontology alignment uncertainty assessment for data quality standards to assist users for pragmatic surveillance.

Keywords: Information Fusion, Situational Assessment, ontology alignment, uncertainty analysis, information theory

1 Introduction

In this paper, we explore the requirements for an ontology alignment uncertainty analysis through information quality and information theory. Ontology development includes syntactic, semantic, and pragmatic issues that require uncertainty assessment. To assess these issues is the Ontology Alignment Evaluation Initiative (http://oaei.ontologymatching.org/) with the goals of:

- assessing strength and weakness of alignment/matching systems;
- comparing performance of techniques;
- increasing communication among algorithm developers;
- developing evaluation techniques; and
- improving the work on ontology alignment/matching.

To meet these goals, controlled experimental evaluation of the techniques and performances are conducted. To further develop the evaluation techniques, we need to address the uncertainty issues in ontological alignment.

Our goal is to prioritize the uncertainty information so that a user could determine which information needs better translation (semantic alignment), or which data lacks uncertainty. Currently, the user does not list uncertainty in an ontology mapping. During the matching of ontologies, the user lists the information without reporting the capability of the quality of the mapping.

1.1 Ontological Alignment and Merging

Gruber [1, 2] succinctly defines an **ontology** as *a specification of a conceptualization*. Ontologies have been designed for knowledge management, use, and sharing. Since an ontology is a theory of a consistent (but not complete) formal vocabulary, agents operate on the information which can be uncertain. Classes, relations, functions, or other objects in normal human discourse and readable text describing what the names mean enable communication; however, if agents do not share the exact vocabulary, there can be mismatches or inexact meanings in the understanding of knowledge.

To design decision support information fusion systems [3], there is a need to have a defined vocabulary of information such as object identifications. Definitions of semantic content is important for situation awareness [4, 5] and metric analysis for fusing different ontologies through sensor management [6]. When designing decision-support systems, not only is the vocabulary important, but how the user interprets the information for threat prediction and situation awareness [7] which can be affected by experience and emotions [8].

Like federating databases, merging two ontologies includes specifications of conceptualizations which are interlocking systems of institutional facts [9, 10]. Colomb establishes that to merge ontologies, the argument requires

a contextual definition of what it means for systems of facts from operational organizations to be interlocking.

To be interoperable, information fusion systems need a shared conceptualization of the application domain [11]. Translation uncertainty is a natural consequence of the ontology alignment process. In order to support information fusion and decision making, it is necessary to qualify and quantify such uncertainty. The goal is to quality the uncertainty associated with alignment of two ontological information fusion systems [12] for decision making [13]. Information theory has long been regarded as a method of assessing vocabulary content, however, there is need to also evaluate the uncertainty in meaning for the successful merging of databases. Issues in semantic content for a sensor exploitation system include (a) *ontology standardization* [14] - multiple semantic labels can be attributed to the same information, (b) *ontology alignment* [15] - when there is not a shared vocabulary a linking is required, and (c) *ontology analysis* [16] – same, incomplete, or imprecise labels for different information.

1.2 Information Fusion Ontology Research

The use of ontologies in information fusion includes situational awareness (SA) [17, 18], command and control [19], decision support, and situation management [20, 21]. These systems are designed for a specific application. Examples of such issues are agent communication net-centric warfare [22]. Current developments in other communities include biomedical research [23] and web services [24]. These communities have a plethora of articles describing examples of interoperability topics that could be leveraged by the information fusion community.

As the user works with labels, it is important to incorporate the user with machine fusion systems through ontology representations (i.e. semantic content labels) as shown in **Figure 1**. The synergy between the ontology community using data fusion [25, 26] and the information fusion looking at ontology issues in high-level information fusion [27, 28], threat ontologies based on the target intent, capability, and opportunity [29]. Other examples include probabilistic methods [30], data sources [31] and scene understanding [32].

Examples include knowledge-based mission analysis [33], semi-automated ontology development [34], and semantic-oriented services architectures [35], and soft messaging [36]. With various applications researching ontology developments, there is a need to understand the alignment process and match the reporting results with users needs in high-level information fusion [37, 38, 39].

Duckham and Worboys [16] list many categories to be considered in ontology research:

Semantic alignment: whether or not the labeling correspond to meaningful categories of alignment.
Inaccuracy of information: the segmentation of the geographical boundaries
Imprecision: the representation of the segmented areas

Vagueness: uncertainty in the boundaries
Uncertainty: the stochastic randomness in the collection boundaries.

1.3 Data Quality Measures in Ontology

The *ISO 211 Data Quality Information* [40] standard is one example of numerous efforts to standardize information and data quality. The *ISO 19115 Metadata* standard includes non-quantitative and quantitative quality information; however, the ISO 19114 Quality Evaluation report includes on quantitative reporting from data quality attributes of scope, measure, procedure, results, and type as well as the time of reporting. From the reporting of quality measures, such as ontology alignment uncertainty, there is a need to assess the non-quantitative information in the quantitative reporting.
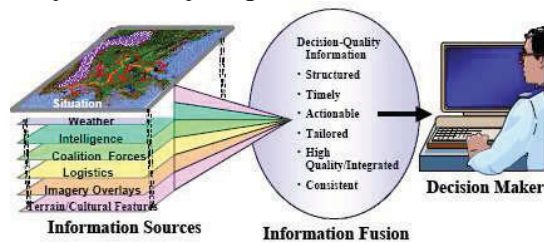

Figure 1. Evaluating and reporting Data quality measures

The rest of the paper is as follows. Section 2 overviews the uncertainty analysis as it relates to metrics of data quality. Section 3 establishes the notion of a semantic distance and Section 4 overviews information theory. Section 5 provides a maritime-domain awareness model with a conclusion in Section 6.

2 Uncertainty Analysis

Uncertainty has many definitions and recent efforts to understand the information content the World Wide Web has fostered research in quality of information or information quality analysis. Klir, established a taxonomy of uncertainty categories shown in **Figure 2**. [41]
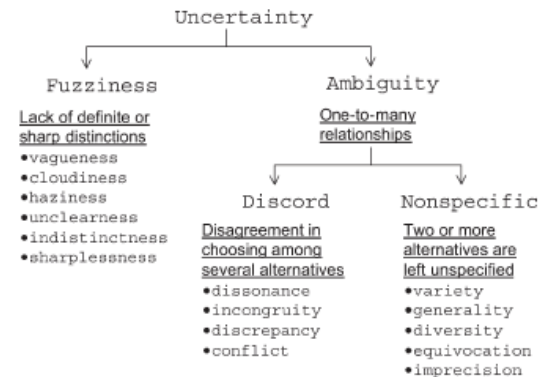

Figure 2. Klir and Yun's basic types of uncertainty [42].

2.1 Accuracy and Precision

Statistical methods of uncertainty analysis from measurement systems include accuracy and precision, shown in Figure 2. The use of distance metrics (accuracy) and precision metrics (standard deviations) help to analyze whether the measurement is calibrated and repeatable. We would desire the same analysis for semantic content with precise meanings and consistent understanding.
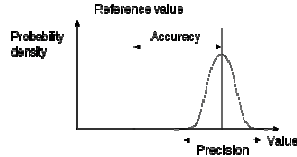

Figure 3. Uncertainty defined as precision and accuracy.

Ontologies must be able to account for the uncertainty between databases and be able to model it qualitatively, semantically, and quantitatively. Using the quality of service (QoS) and quality or information (IQ) we can better realize a modeling capability between ontologies and uncertainty analysis.

2.2 Information Quality and QoS

Quality of Service metrics include accuracy, probability of *error*, *timeliness*, *throughput*, and *cost*. Blasch, [6] mapped these into fusion metrics to include confidence as lack of error as shown in the left column of Figure 4. Information quality has also a coalescing of terms and relates to the QoS metrics. The goal then is to see how these quality definitions can be used for uncertainty analysis (shown in Figure 3 right column – from Klir's definitions) for ontology alignment. One concern is that *reliability* has yet to be defined in the fusion literature, but is well known in industry [43, 44]. From **Figure 4**, these definitions can be loosely correlated in related categories as shown in the rows in Figure 4. For example, QoS accuracy maps to IQ accuracy. The key for ontology research is to determine the mapping of representational IQ into the discord ambiguity and fuzziness areas moving from sound mathematical principles in information theory to translation uncertainty in ontology alignment.
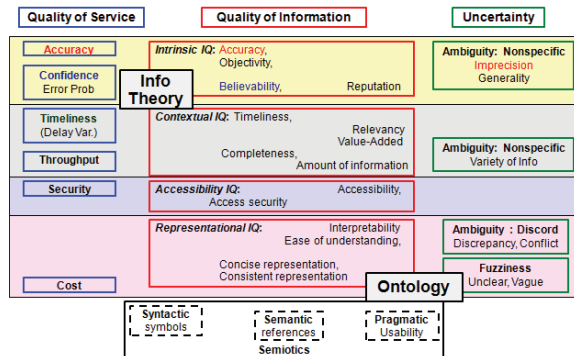

Figure 4. Mapping between QOS, QOI, and Uncertainty.

Data quality is important to net-centric fusion [45]. While many of the definitions are not consistent with the information quality definitions, it does validate the need to clearly define a vocabulary of uncertainty [46]. Intrinsic IQ accuracy and believability relates to nonspecific ambiguity imprecision and generality. Contextual IQ completeness and timeliness correspond with non-specific ambiguity of variety. Accessibility IQ does not specifically relate to an uncertainty category. Definitions of data with semantic uncertainty in Klir's model include vagueness, completeness, fuzziness, interpretability and conflict are mapped to representational IQ. To utilize the IQ taxonomy, two overlaps arise with completeness and consistency.

Figure 4 is a mapping between the *Quality of Service* definitions, the *Information Quality* definitions, and categories of *uncertainty*. When we look at ontological alignment issues, it could utilize a mapping of uncertainty (fuzziness and discord ambiguity) and representation IQ for pragmatic and semantic analysis; while information theory has been applied to QOS and QOI information over syntactic symbols. The requirements of information quality demonstrate the need for linking the formalism of information theory with the knowledge representation from ontology research.

A key measure of information in the theory is known as *entropy*, *H*, which is usually expressed by the average number of bits needed for storage or communication. Intuitively, entropy quantifies the average uncertainty of an information source represented by a random variable.

$$H(X) = \mathbb{E}_X [\, I(x) \,] = - \sum_{x \in X} p(x) \operatorname{Log} p(x) \qquad (1)$$

To determine *accuracy* of translations between ontologies, we need to establish the concept of a ***semantic distance***.

3 The Concept of Semantic Distance

When two symbols denote concepts that are close but are not synonyms or completely congruent, we say that there is some sort of related similarity (or distance) between these terms. In trying to align distinct ontologies together, we establish correspondence relationships between symbols where semantic distance is an inherent part of the resulting matching solution. In practice, we often discard such considerations, but this may result in semantic loss.

We would like some sort of mechanism to alleviate the uncertainty on the exactness of a translation. To do so, we need some sort of a method to measure that distance, but even before that we are confronted with the very nature of that distance, what it represents and how it is defined. Of course, there are several views on this from many research efforts. We will assume that *semantic distance* and *semantic similarity* relate to the measurement construct.

3.1 Semantic Similarity

Rodriguez [47] gives an account of four classes of models for semantic similarity assessment and how it is defined within those contexts:

- *Feature-based* models define the semantic similarity by comparison of concept characteristics such as properties, roles, rules, etc.
- Models based on *semantic relations* make use of semantic networks and how far two nodes (concepts) are from one another. This model is not foreign to a certain practice motivated by structuralism.
- Models based on *information content* establish semantic similarity between concepts by comparison of how much information there is to support the concept and how specific it is. Its underlying assumption is that abstract concepts present less information and influence some function occurrence within a corpus of text. This model is less dependent on structure but very sensitive to aspects related to its assumption. It might not be the right model for semantic interoperability since two distinct ontologies might not present the same characteristics of abstractedness.
- *Context-based models* consider syntactic, semantic, pragmatic and stylistic factors to establish how words can be substituted within a sentence. The higher the degree of substitution indicates the higher the semantic similarity.

Rodriguez proposes the matching-distance model for a single ontology, which is limited when dealing with large-scale semantic interoperability standards.

Determining the most suitable model for source-to-fusion node ontological alignment is still an active research area. Thus, we ask ourselves how semantic distance measures could be exploited in refining the fusion process. [48] However, we are quite confident in stating that disregarding semantic distance may lead to erroneous handling of information and thus impede the overall fusion process. To avoid this, we suggest that this aspect be explicitly reflected in a semantic distance measurement.

Semantic distance refers to the disparity between the user's intentions and the meaning in the interface language, while articulatory distance refers to the distance between the physical form of the interface language and its meaning. [49] On the execution side, semantic distance of the interface is reduced if the user's intentions can be expressed in the command language in a concise manner [50]. The goal should be to match the level of description required by the interface language to the level at which the person thinks of the task. On the evaluation side, semantic distance is reduced if the displayed objects represent the higher level concepts that people naturally adopt when reasoning in the problem domain. In this case, the goal should be to provide a powerful, productive way of thinking about the domain. [51] Figure 5 presents a model of how a user would interact with an automated ontology alignment decision support aid.
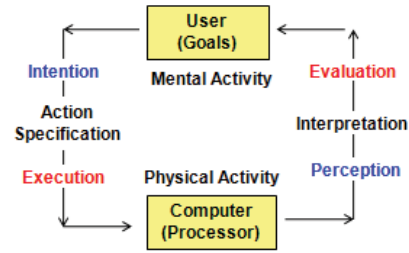

Figure 5. Execution & evaluation (*Based on* [52]).

Modeling the human user typically includes three interacting subsystems: (1) *the perceptual system,* (2) the *motor system,* and (3) the *cognitive system*. The cognitive system is fundamentally parallel (with selective interrupt handling) in its recognition phase and fundamentally serial in its action phase. Since a user conducts the ontology mapping, there are semantic issues of cognitive bias and syntactic issues in symbol or word associations. Finally, there are pragmatic issues surrounding ones experiences, familiarity, and use of the various ontologies.

3.2 Cognitive Bias

Judgment or decision-making errors include various user biases in ontology mappings. Most ontology translations are done by hand to link meanings as a computer algorithms have yet to match the plethora of established ontologies. Over the years, there have been many biases listed in the literature [53] categorized into behavioral, decision-making, probability and belief, social, memory, and cognitive groups. Bias clearly relates to ontology uncertainties that have to be accounted for in ontology alignment. Table 1 lists a few of these (as examples) that can be important for presentation of information decision quality in user induced errors.

Table 1. User Factors that Influence Decision Quality

	Description
Availability Bias (Tversky, 1973)	Recall of specific incidents affects judgments of frequency
Attentional Bias	Tendency to neglect relevant data when making correlation/ association decisions
Belief Bias	Likelihood of the conclusion affects logical strength argument evaluation
Confirmation Bias (Wason, 1960)	Seeks data consistent with own opinion and ignore disconfirming evidence
Concrete Info. Bias (Nisbett 1980)	Experienced data tends to dominate abstract statistical summary information
Frequency Bias (Einhorn, 1978)	Focus on absolute frequency of events rather than their relative frequency for predictive relationship evaluation
Miscommunication (Cannon-Bowers,93)	Smooth communication and coordination depend on team shared mental models.
Primacy Effect (Anderson, 1981)	Information presented first is weighted more heavily than data presented later

Some of the ontological issues for evaluation uncertainty and bias mitigation include:

Availability Bias [54]: Utilize context sensitive cues to determine the incident likelihood to aid in categorization spanning the incident history to reduce the likelihood that prior incidents will be more heavily weighted when making a decision

CONFIRMATION BIAS [55]: Provide knowledge agents that alert users when the value for a particular cue falls outside of a pre-specified range.

CONCRETE INFORMATION BIAS [56]: Afford displays that provide context dependent cues to help focus attention on the situational awareness and not current perceptions.

FREQUENCY BIAS [57]: Instrument dynamic processes to represent functional relationships between variables rather than the value of individual variables.

MISCOMMUNICATION DUE TO LACK OF SHARED MENTAL MODELS [58]: Foster a shared view of the problem space for common understanding of the current situation to facilitate effective communication.

PRIMACY EFFECT [59]: Manipulate salience to ensure equal visibility of all information being utilized to make a decision.

Along with the cognitive aspects of ontological semantic mapping is the syntactic (symbolic) and pragmatic goals.

3.3 Recognition and Action

Symbolic manipulation has been well established in the ordering and processing of word. A user's working memory, consist of symbols, called **chunks**, that may themselves be organized into larger units [60, 61]. However, successful retrieval of a chunk depends on whether associations to it can be found. Retrieval error results from (1) lack of effective associations, or (2) similar associations to several chunks interfere with the retrieval of the target chunk. Linking a user's ontological knowledge with domain knowledge requires **semantic coding** of information and actions.

One of the leading developments included hand-eye coordination for spatial analysis, which can be utilized for ontological semantic analysis. Utilizing *Fitt's Law* [62] to determine, we can determine the timeliness of action. The time to move the hand to a target depends only on the relative *precision* required, that is, the ratio between the target's distance and its size. The time T_{POS} to move the hand to a target of size S that lies a distance D away is given by:

$$T_{POS} = I_M \, \mathrm{Log}_2 \, [D/S + 0.5] \tag{2}$$
$$\text{where } I_M = 100 \; [70\text{-}120] \; \text{msec/bit.}$$

Using the information theory principle, decision time T increases with uncertainty about the judgment or decision to be made:

$$T = I_C \, H \tag{3}$$

where H is the information-theoretic entropy of the decision and $I_C = 150$ [0-157] msec/bit. For n equally probable alternatives (called *Hick's law* [63]),

$$H = \mathrm{Log}_2 \, (n+1) \tag{4}$$

For n alternatives with different probabilities, p_i of occurrence,

$$H = \Sigma_i \, \mathrm{Log}_2 \, (1 / p_i + 1). \tag{5}$$

Utilizing the information-theoretic approaches for target recognition, Card applied these methods to word recognition based on the frequency of letters in the semantic dictionary. For example, recognition rates of words fell between 17 to 25 words per minute.

While word recognition does not solve the problems of word ontology alignment; it does demonstrate that timelines, confidence, and throughput can be addressed; while accuracy, fuzziness, and ambiguity in meaning have yet to be solved.

To address the ontology alignment problem, and utilizing the elements of information quality, we can establish link between uncertainty analysis and ontology mapping through information theory.

4 Information Theory

We will utilize the relative entropy analysis to assess the information content. [64, 65]

4.1 KL Divergence

In information theory, the **Kullback–Leibler divergence** [66], (also *information divergence*, *information gain*, or *relative entropy*) is a non-symmetric measure of the difference between two probability distributions P and Q. KL measures the expected number of extra bits required to code samples from P when using a code based on Q, rather than using a code based on P. Typically P represents the "true" distribution of data, observations, or a precise calculated theoretical distribution. The measure Q typically represents a theory, model, description, or approximation of P. Like Discrimination gain, there is minimal discrimination information (MDI) that can be utilized for sensor placement.

For probability distributions P and Q of a discrete random variable their K–L divergence is defined to be

$$D_{KL} \, (P \| Q) \; = \; \sum_i \; P(i) \, \mathrm{Log} \, \frac{P(i)}{Q(i)} \tag{6}$$

The K-L divergence is only defined when $P > 0$ and $Q > 0$ for all values of i, and when P and Q both sum to 1.

One might be tempted to call it a "distance metric" on the space of probability distributions, but this would not be correct as the Kullback–Leibler divergence is not symmetric – that is, $D_{KL} \, (P \| Q) \neq D_{KL} \, (Q \| P)$ – nor does it satisfy the triangle inequality. Still, being a pre-metric, it

generates a topology on the space of generalized probability distributions, of which probability distributions proper are a special case.

Following Renyi [67], the term is sometimes also called the **information gain** about X achieved if P can be used instead of Q. It is also called the **relative entropy**, for using Q instead of P. KL is closely related to information theory concepts of mutual information and relative information.

The **mutual information**,

$$
\begin{aligned}
I(X;\, Y) &= D_{\text{KL}}\ (P(X,\, Y) \,\|\, P(X)\, P(Y)) \\
&= \mathbb{E}_X\ \{\, D_{\text{KL}}\ (P(Y\,/\,X) \,\|\, P(Y))\, \} \\
&= \mathbb{E}_Y\ \{\, D_{\text{KL}}\ (P(X\,/\,Y) \,\|\, P(X))\, \}
\end{aligned}
\tag{7}
$$

is the KL divergence of the product $P(X)P(Y)$ of the two marginal probability distributions from the joint probability distribution $P(X,Y)$ — i.e. the expected number of extra bits that must be transmitted to identify X and Y if they are coded using only their marginal distributions instead of the joint distribution. Equivalently, if the joint probability $P(X,Y)$ *is* known, it is the expected number of extra bits that must on average be sent to identify Y if the value of X is not already known to the receiver.

The **conditional entropy**,

$$
\begin{aligned}
H(X\,|\,Y) &= \text{Log}\, N - D_{\text{KL}}\ (P(X,\, Y) \,\|\, P_{\text{U}}\,(X)\, P(Y)) \\
&= H(X)\ -\ I(X;\, Y)
\end{aligned}
\tag{8}
$$

is the number of bits which would have to be transmitted to identify X from N equally likely possibilities, *less* the KL divergence of the product distribution $P_{\text{U}}(X)\ P(Y)$ from the true joint distribution $P(X,Y)$ — i.e. *less* the expected number of bits saved which would have had to be sent if the value of X were coded according to the uniform distribution $P_{\text{U}}(X)$ rather than the conditional distribution $P(X/Y)$ of X given Y.

4.2   Symmetrized divergence

Kullback and Leibler defined the divergence as:

$$
D_{\text{KL}}\,(P\,\|\,Q) + D_{\text{KL}}\,(Q\,\|\,P)
\tag{9}
$$

which is symmetric and nonnegative. This quantity has sometimes been used for feature selection in classification problems, where P and Q are the conditional pdfs of a feature under two different classes.

An alternative is given via the **λ divergence**,

$$
\begin{aligned}
D_\lambda\,(P\,\|\,Q) = {}&\lambda\ D_{\text{KL}}\,(P\,\|\,\lambda P + [1-\lambda]\, Q) \\
&+ (1-\lambda)\ D_{\text{KL}}\,(Q\,\|\,\lambda P + [1-\lambda]\, Q)
\end{aligned}
\tag{10}
$$

which can be interpreted as the *expected information gain* about X from discovering which probability distribution X is drawn from, P or Q, if they currently have probabilities λ and $(1-\lambda)$ respectively.

The value $\lambda = 0.5$ gives the ***Jensen–Shannon divergence***, defined by

$$
D_{\text{JS}} = (\tfrac{1}{2})\ D_{\text{KL}}\,(P\,\|\,M) + (\tfrac{1}{2})\ D_{\text{KL}}\,(Q\,\|\,M)
\tag{11}
$$

where M is the average of the two distributions,

$$
M = (\tfrac{1}{2})\ (P + Q)
\tag{12}
$$

D_{JS} can also be interpreted as the capacity of a noisy information channel with two inputs giving the output distributions p and q. The Jensen–Shannon divergence is the square of a metric that is equivalent to the Hellinger metric, and the Jensen–Shannon divergence is also equal to one-half the so-called *Jeffrey's divergence* [68 ,69].

4.3   Definition of Mutual Information

Formally, the mutual information of two discrete random variables X and Y can be defined as: [70]

$$
I\,(X;\, Y) = \sum_{y\,\in\,Y} \sum_{x\,\in\,X} p(x,\, y)\ \text{Log}\left(\frac{p(x,\, y)}{p_1(x)\, p_2(y)}\right)
\tag{13}
$$

where $p(x,y)$ is the joint probability distribution function of X and Y, and $p_1(x)$ and $p_2(y)$ are the marginal probability distribution functions of X and Y respectively.

4.4   Normalized Information variants

Normalized variants of the mutual information are provided by the *coefficients of constraint* [71] or *uncertainty coefficient* [72]

$$
C_{\text{XY}} = \frac{I\,(X;\, Y)}{H(Y)}\ \text{ and }\ C_{\text{YX}} = \frac{I\,(X;\, Y)}{H(X)}
\tag{14}
$$

The two coefficients are not necessarily equal. A more useful and symmetric scaled information measure is the *redundancy*

$$
R = \frac{I\,(X;\, Y)}{H(X)\ + H(Y)}
\tag{15}
$$

which attains a minimum of zero when the variables are independent and a maximum value of

$$
R_{\text{max}} = \frac{\min\,(\,H(X),\, H(Y))}{H(X)\ + H(Y)}
\tag{16}
$$

when one variable becomes completely redundant with the knowledge of the other. Another symmetrical measure is the *symmetric uncertainty* [73], given by

$$
U(X,\, Y) = 2\,R\ = \ 2\,\frac{I\,(X;\, Y)}{H(X)\ + H(Y)}
\tag{17}
$$

which represents a weighted average of the two uncertainty coefficients.

We are interested in using the information theoretic measures for performance analysis [74] and ontological alignment evaluation in deducing uncertainty.

5 Example

Our example comes from a need to protect coastal areas [75], support operator needs, and [76] provide valuable performance metrics [77].

In this example, we are interested in the tracking and identification of ships [78]. Since the ships utilize a reporting system of their information and position, it can be ontologically used for semantic uncertainty analysis. Issues arise in prediction of future locations [79], but we would investigate this at a later time. We use the Automatic Information System (AIS) database of ship identification [80]. While there are 99 designations of ship types, not all are used. For purposes of the analysis, we use a subset of the information to determine the semantic uncertainty. Recent efforts include decision support using the ship designations [81].

We assume that we have two systems that are collecting information on the maritime scenario. Both systems collect and label their products differently. We need to align (or match) the ontology representations of the physical objects to the situation for coordination with the image displays of the coastal areas.

5.1   Use of the Classifications

For the ship designation [82], the classifications are meant to separate the meanings of the different types of ships. The general categories include:

[20-29] Wing in ground (WIG) with different hazardous categories
[30-39] pleasure and military ops
[40-49] High speed craft (HSC) with various hazardous categories
[50-59] port law enforcement
[60-69] passenger with various hazardous categories
[70-79] cargo, with various hazardous categories
[80-89] tanker, with various hazardous categories
[90-99] other with various hazardous categories

We use a subset of the categories for the simulation.

21 Wing in ground Hazardous category A
41 Hi speed craft Hazardous category A
50 Pilot Vessel
52 Search and Rescue Vessel
56 Spare Local Vessel
61 Passenger Hazardous category A
71 Cargo Hazardous category A
81 Tanker Hazardous category A

From the designations, there are semantic issues associated with the labeling. Before looking at a situation in which the operators are communicating; it is noted that the number designation and the ship size could confer physical and quantifiable data; however, if a ship is not communicating with the port authorities, the human observers would try to identify the ship and most likely place it in one of the categories.

5.2   Case 1: Similar meaning

21 Wing in ground Hazardous category A
41 Hi speed craft Hazardous category A

We calculate the KL divergence, KL-Symmetric Divergence and symmetric uncertainty [83]. Also, in Figure 6 we plot the entropy, conditional entropy, and mutual information of the classifications between category 21 and 41.
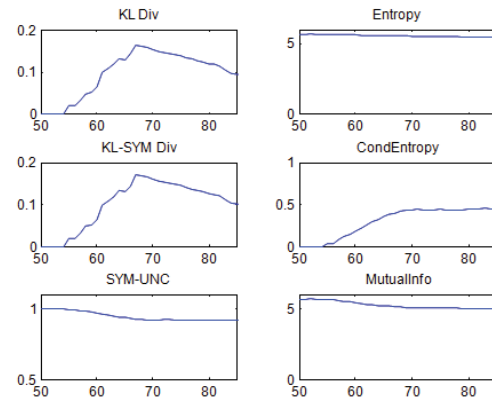

Figure 6. Information based ontology alignment for the case of similar meanings.

While the analysis includes semantic classification information, some things to note are that the divergence (or semantic distance) decreases when the same words are used to identify the ship. An example might be that the operator indicates a ship type and then clarifies the hazardous instance. However, the operator most likely remembers the most recent information that was discussed and might have a bias in the analysis.

5.3   Case 2: Different Meaning

21 Wing in ground Hazardous category A
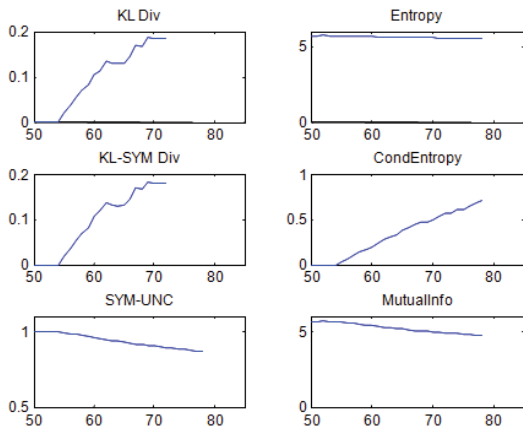52 Search and Rescue Vessel

Figure 7. Information based ontology alignment for the case of different meanings.

From Figure 7, it is easy to see that with no overlap, the divergence in semantic distance is larger when different categories are used to designate the ship type. Problems may arise when an undesignated ship is conducted unauthorized activities. During the excitement, there are issues associated with the cultural factors affected the decision making [**84**], which could lead to ontology discord.

Semantic uncertainty can be determined from various sources of information and we utilize the KL metrics to determine the semantic relationships of ontology mapping. When the ontology is aligned, the KL divergence decreases, whereas when the ontologies are not aligned, the KL divergence increases. It is important to note that the KL value does not necessarily correspond with the ontology alignment quality, but rather indicates a possible uncertainty with the ontology alignment.

6 Conclusions

We explored the issues in ontology alignment through methods of information quality and information theory. For ontological alignment, many open questions still remain such as (1) how to deal with mismatch and mapping uncertainty in ontologies, (2) coordination between different standards for interoperability, and (3) how to evaluate ontology alignment using uncertainty measurements. We presented a simple case for maritime domain awareness for ship classifications and used the information-theoretic KL divergence as a measure to capture the dissimilarity in ontologies.

The ISO quality standards community is actively seeking methods to determine the quantitative and qualitative uncertainties in ontology alignment. Together, the integration of the hard-soft ontology alignment would aid in command and control, synergistic collection of surveillance information, and decision support systems to provide reasoning and understanding over the situation.

7 References

Ontology
[1] T. R. Gruber, "A translation approach to portable ontologies," *Knowledge Acquisition*, 5(2):199-220, 1993.
[2] T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, Vol. 43, Issues 4-5, pp. 907-928, November 1995.
Situational Awareness
[3] E. P. Blasch "Assembling an Information-fused Human-Computer Cognitive Decision Making Tool," *IEEE Aerospace and Electronic Systems Magazine,* June 2000*,* pp. 11-17.
[4] E. Blasch, I. Kadar, K. Hintz, J. Biermann, C. Chong, and S. Das, "Resource Management Coordination with Level 2/3 Fusion", *IEEE AES Magazine ,* Mar. 2008.
[5] E. Blasch, I. Kadar, J. Salerno, M. M. Kokar, S. Das, G. M. Powell, D. D. Corkill, and E. H. Ruspini, "Issues and challenges of knowledge representation and reasoning methods in situation assessment (Level 2 Fusion)", *J. of Adv. in Information Fusion,* Dec. 2006.
[6] E. Blasch, M. Pribilski, B. Roscoe, *et. al.*, "Fusion Metrics for Dynamic Situation Analysis," *Proc SPIE 5429*, 2004.
[7] G. Chen, D. Shen, C. Kwan, J. Cruz, M. Kruger, E. Blasch, "Game Theoretic Approach to Threat Prediction and Situation Awareness," *Advances in Journal for Information Fusion,* July 2008.
[8] M. Wei, G. Chen, J Cruz, L. Haynes, M. Kruger, & E. Blasch "Game-Theoretic Modeling and Control of Military Operations with Partially Emotional Civilian Players", *Decision Support Systems*, Vol. 44, No. 3, Elsevier, Feb, 2008.
[9] R. M. Colomb and M. N. Ahmad, "Merging Ontologies requires interlocking institutional worlds," *App. Ontology*, IOS Press, 2007.
[10] R. M. Colomb and M. N. Ahmad, "A perdurant ontology for interoperating information systems based on interlocking institutional worlds," *App. Ontology*, Vol. 5, IOS Press, 2010.
[11] B. Drabble, T. Black, C. Kinzig, and G. Whitted, "Ontology based dependency analysis: Understanding the impacts of decisions in a collaborative environment," *Int. Symp. on Collaborative Technologies and Systems*, 2009.
[12] E. Blasch, É. Dorion, P. Valin, E. Bossé, and J. Roy, "Ontology Alignment in Geographical Hard-Soft Information Fusion Systems," *Fusion2010*, 2010.
[13] D. L. Hall, J. Llinas, M. McNeese, and T. Mullen, "A Framework for Dynamic Hard/Soft Fusion," *Fusion08*, 2008.
[14] M. Schorlemmer and Y. Kalfoglou, "Institutionalizing Ontology-based semantic integration," *App. Ontology*, Vol. 3, IOS Press, 2008.
[15] S. Mencke, M. Kunz, D. Rud, R. Dumke, "Quality-Driven, Semantic Information System Integration: The QuaD Framewok," Ch. 7 in *Cases on Semantic Interoperability for Information Systems Integration: Practices and Applications*, Ed. Y. Kalfoglou, 2010.
[16] M. Duckham and M. Worboys, "Automated Geographical Information Fusion and Ontology Alignment", Ch 6 in *Spatial Data on the Web* (eds.) A. Belussi, B. Catania, E. Clementini, and E. Ferrari, Springer, 2007.
[17] P. Smart, "Knowledge-Intensive Fusion for Situational Awareness," DTC/WP270, Univ. Southampton, UK, 2005.
[18] C. J. Matheus, M. M. Kokar, K. Baclawski, J. A. Letkowski, C. Hall, M. L. Hinman, and D. M. Boulware, "SAWA: an assistant for higher-level fusion and situation awareness," *Proc SPIE 5813*, 2005.
[19] K. Wallenius, "Support for Situation Awareness in Command and Control," *Fusion07*, 2007.
[20] N. Baumgartner and W. Retschitzegger, "Towards a Situation Awareness Framework Based on Primitive Relations," *IEEE Conf. Info. Decision & Control*, 2007.
[21] N. Kartha and A. Novstup, "Ontology and rule-based knowledge representation and decision support," *Proc. SPIE 7352*, 2009.

[22] H. J. Lee, *Ontology-based Data Fusion within a Net-centric Information Exchange Framework*, PhD Dissertation, Univ. Arizona, 2009.

[23] L. T. Vizenor, O. Bodenreider, and A. T. McCray, "Auditing associative relations across two knowledge sources," *J. Biomed. Informatics*, 42, 2009.

[24] A. Nikolov, V. Uren, and E. Motta, "Towards Data Fusion in a multi-ontology Environment," *Linked Data on the Web (LDOW09)*, Spain, 2009.

[25] A.-C. Boury-Brisset, "Ontology-based approach for information Fusion," Fusion03, 2003.

[26] D. L. McGuinness, "Ontologies for Information Fusion," Presentation at *Fusion03*, 2003.

[27] E. P. Blasch and S. Plano, "Ontological Issues in Higher levels of Fusion: User Refinement in the Fusion Process," *Fusion03*, 2003.

[28] C. Nowak, "On ontologies for high-level information fusion," *Fusion03*, 2003.

[29] A.-C. Boury-Brisset, "Ontological Engineering for Threat Evaluation and Weapon Assignment: a Goal-driven approach," Fusion07, 2007.

[30] K. B. Laskey, P. C. G. da Costa, E. Wright and K. Laskey, "Probabilistic Ontology for Net-Centric Fusion," *Fusion07*, 2007.

[31] M. Gagnon, "Ontology-Based Integration of Data Sources," *Fusion07*, 2007.

[32] M. Zandipour, B. J. Rhodes, and N. A. Bomberger, "COALESCE : A Probabilistic Ontology-Based Scene Understanding Approach," *Fusion08*, 2008.

[33] J. Roy and A. Auger, "Knowledge and Ontological Engineering Techniques for Use in Developing Knowledge-Based Situation Analysis Support Systems," *Defence R&D Canada – Valcartier Tech. Mem. 2006-757*, 2008.

[34] A. P. Stirtzinger and C. S. Anken, "Semi-automated ontology generation and evolution," *Proc. SPIE 7347*, 2009.

[35] R. Hull, K. Bimson, R. Hyle, and R. Thiebauth, "Semantic service-oriented architecture for range operation: evolving the role of semantics in the enterprise," *Proc. SPIE 7350*, 2009.

[36] K. Sambhoos, J. Llinas, and E. Little, "Graphical Methods for Real-Time Fusion and Estimation with Soft Message Data," *Fusion08*, 2008.

[37] E. Bosse, J. Roy, and S. Wark, *Concepts, Models, and Tools for Information Fusio*n, Artech House, 2007.

[38] D. A. Lambert, "Tradeoffs in the Design of Higher-level Fusion Systems," *Fusion07*, 2007.

[39] S. Das, *High-Level Data Fusion*, Artech House, 2008.

[40] *Standards Guide, ISO/TC 211 Geographic Information/ Geomatics*, June 2009, available at http://www.isotc211.org/

Uncertainty

[41] C. Liu, D. Grenier, A-L. Jousselmen, and E. Bosse, "Reducing Algorithm Complexity for Computing an Aggregate Uncertainty Measure," *IEEE Trans. on Sys Man and Cybernetics - A*, Vol. 37, No. 5, Sept 2007.

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

[43] E. Blasch, "Derivation of a Reliability Metric for Fused Data Decision Making," *IEEE NAECON Conf.*, July 2008.

[44] L. Bai and E. Blasch, "Two-Way Handshaking Circular Sequential k-out-of-n Congestion System," *IEEE Journal of Reliability*, Mar 2008.

[45] I. Kadar, E. Blasch, and C. Yang, "Network and Service Management Effects on Distributed Net-Centric Fusion Data Quality," July, *Fusion08*.

[46] E. P. Blasch and Jim Schmitz, "Uncertainty Issues in Data Fusion", *NSSDF 02,* August 2002.

Semantic Distance

[47] M. A. Rodriguez. *Assessing semantic similarity among entity classes*. PhD. Thesis, University of Maine, 2000.

[48] É. Dorion and S. Fortin, "Multi-Source Semantic Integration – Revisiting the Theory of Signs and Ontology Alignment Principles," *Fusion07*, 2007.

[49] S. K. Card, T. P. Moran, and A. Newell. *The psychology of human-computer interaction.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1983.

[50] J. Mackinlay, S. K. Car, and G. G. Robertson, "A Semantic Analysis of the Design Space of Input Devices," *Human Computer Interaction*, Vol. 5, Erlbaum, 1990.

[51] J. Rasmussen, A. M. Pejtersen, and L. P. Goodstein, *Cognitive Systems Engineering*, John Wiley and Sons, 1994.

[52] D. A. Norman & S. Draper, (Eds.), *User Centered System Design: New Perspectives on Human-Computer Interaction.* Hillsdale, NJ: Lawrence Erlbaum, 1986.

Cognitive Bias

[53] http://en.wikipedia.org/wiki/List_of_cognitive_biases

[54] A. Tversky and D. Kahneman, "Availability: a heuristic for judging frequency and probability," *Cognitive Psychology* **5**, 207-232, 1973.

[55] P. C. Wason, Peter C. "On the failure to eliminate hypotheses in a conceptual task", *Quarterly Journal of Experimental Psychology* (Psychology Press) **12** (3): 129–140, 1960.

[56] R. Nisbett and L. Ross, *Human inference: Strategies and shortcomings of social judgment*. Prentice-Hall, NJ, 1980.

[57] H. J., Einhorn and R. M. Hogarth, "Confidence in judgment: Persistence of the illusion of validity," *Psychological Review*, **85**, 396-416, 1978.

[58] J. E. Cannon-Bowers, E. Salas, and S. Converse, S. "Shared Mental Models in Expert Team Decision-Making,. in J. Castellan (ed.), *Individual and Group Decision-Making: Current Issues*. Mahwah, NJ: Lawrence Earlbaum and Associates, Inc., 1993.

[59] N. H. Anderson, *Foundations of information integration theory*. New York: Academic Press, 1981.

[60] G. A. Miller, "The magical number seven plus or minus two: Some limits on our capacity for processing information." *Psychological Review,* 63,81-97, 1956.

[61] H. A. Simon, "How big is a chunk?," *Science, 183,* 482-488, 1974.

[62] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *Journal of Experimental Psychology*, Vol. 47, No. 6, June 1954.

[63] W. E. Hick, "On the rate of gain of information," *Quarterly Journal of Experimental Psychology, ,*4, 11-26, 1952.

Information Theory

[64] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2 ed., New York, Wiley, 2006.

[65] http://en.wikipedia.org/wiki/Kullback-Leibler_divergence

[66] S. Kullback, *Information theory and statistics*, John Wiley and Sons, NY, 1959.

[67] A. Rényi , *Probability Theory*. New York: Elsevier, 1970.

[68] Y. Rubner, C. Tomasi, L. J. and Guibas, "The Earth Mover's distance as a metric for image retrieval," *International Journal of Computer Vision*, **40**(2): 99–121, 2000.

[69] H. Jeffreys *Theory of Probability.* 3rd ed. Oxford University Press; Oxford, U.K.: 1961.

[70] http://en.wikipedia.org/wiki/Mutual_information

[71] C. H. Coombs, R. M. Dawes, & A. Tversky, *Mathematical Psychology: An Elementary Introduction*, Prentice-Hall, Englewood Cliffs, NJ, 1970.

[72] W. H, Press, B. P. Flannery, S.A. Teukolsky, S. A. & W. T, Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge, p. 634, 1988.

[73] I. Witten and E. Frank, Data Mining: *Practical machine learning tools and techniques*, Morgan Kaufmann, San Francisco, 2005.

[74] H. Chen, G. Chen, E. Blasch, and P. Douville, "Information Theoretic Measures for Performance Evaluation and Comparison ," *ISIF Conf. Fusion09*, 2009.

AIS

[75] Z. Li, H. Leung, P. Valin, and H. When, "High Level Data Fusion System for CanCoastWatch", *Fusion07*, 2007.

[76] M. Nilsson, J. van Laere, T. Ziemke and J. Edlund "Extracting rules from expert operators to support situation awareness in maritime surveillance", *Fusion08*.

[77] D. Waller, M. R. MacLeod, and T. McCallum, "Measures of Effectiveness and Performance for the Northern Watch Station," *DRDC CORA TM 2008-053*, July 2009.

[78] C. Yang and E. Blasch, "Kalman Filtering with Nonlinear State Constraints", *IEEE Transactions AES*, Vol. 45, No. 1, Jan. 2009.

[79] C. Yang, and E. Blasch, W. Garber, and R. Mitchell, "A Track-ID Maneuver Indicator from Target's Range-Doppler Image, *J. Advances in Information Fusion*, July, 2007.

[80] D. Husjord and E. Pedersen, "On Guidance and Decision-Support Displays in Ship-to-Ship Lightering Operations," *IEEE SMC eNewsletter*, Issue 31, June, 2010.

[81] http://en.wikipedia.org/wiki/Automatic_Identification_System

[82] Ship Type: http://www.bosunsmate.org/ais/message5.php

[83] J. Goñi, and I. Martincorena. Information Theory Toolbox v1.0. Univ. of Navarra - Dpt. Physics & App. Math. Pamplona, Spain, 2007.

[84] M. Wei, G. Chen, J Cruz, L. Haynes, M. Kruger, & E. Blasch "Game-Theoretic Modeling and Control of Military Operations with Partially Emotional Civilian Players", *Decision Support Systems*, Vol. 44, No. 3, Elsevier, Feb, 2008.