ISIF Working Group Report

Paulo C. G. Costa
George Mason University
Fairfax, VA, USA
pcosta@gmu.edu

Anne-Laure Jousselme
CS Group - France
La Garde, France
Anne-Laure.Jousselme@csgroup.eu

On the occasion of the 25th anniversary of ISIF, this paper presents the origin, challenges and key developments of the Evaluation of Techniques for Uncertainty Representation Working Group (ETUR WG), sponsored by ISIF since 2012.

UNCERTAINTY REPRESENTATION AND REASONING EVALUATION FRAMEWORK

Although in recent times chat generative pretrained transformer, commonly known as ChatGPT, has become the epitome of artificial intelligence ubiquity in human lives, the truth is that we have long been subjected to increasingly pervasive sensors, wide availability of large volumes of heterogeneous data, easily accessible machine learning frameworks, and other aspects that enable such systems to exist. Furthermore, the seamless fashion in which such systems pervade our daily lives usually disguises the complexity of the interactions that happen among the various information systems so that data can be properly accessed. Uncertainty management is a key aspect of these interactions and is a critical component to ensure sound results when using multiple data sources. This is especially true when the underlying sources of uncertainty are also heterogeneous, such as in systems that operate above level 2 of the Joint Directors of Laboratories (JDL) framework, a.k.a. high-level information fusion (HLIF) systems. Not surprisingly, the problem of uncertainty representation and reasoning in HLIF systems has attracted interest that extends beyond the information fusion (IF) community.

Even in modern times, fusing hard and soft information from diverse sensor or source types (human-as-a-sensor included) and the associated uncertainty is a task that still relies heavily on human intervention, creating a scalability conundrum that current technologies are incapable of solving. Despite the widespread acknowledgment that HLIF systems must support automated knowledge representation and reasoning in the presence of uncertainty, there is no consensus on the appropriate approach to adopt (which theory, uncertainty function or model, fusion rule, etc.), on the performance criteria that should guide the design of an HLIF system in terms of uncertainty handling, or on how to assess such criteria.

IF applications typically must deal with information that is incomplete, imprecise, inconsistent, and otherwise in need of a sound methodology for representing and managing uncertainty. Complex and dynamic use cases make such tasks even more difficult, because for the same input conditions, apparently minor differences in how uncertainty is handled may drastically affect the output of the IF process. Evaluation of information fusion systems (IFSs) presents intrinsic challenges due to their complexity and the sheer number of variables influencing their performance. Low-level IF tasks generally address random phenomena for which numerical data are collected. The impact of uncertainty representation is well understood and generally quantifiable. However, higher levels of IF tasks need to handle uncertainty not only due to the variability of data (aleatory uncertainty) but also due to lack of knowledge (epistemic uncertainty). The approach chosen for representing uncertainty has an overall impact on system performance that is hard to quantify or even to assess from a qualitative viewpoint.

The evaluation of how uncertainty is addressed within a given IFS is distinct from, although closely related to, the evaluation of the overall performance of the system. Metrics for evaluating the overall performance of an IFS are more encompassing in scope than those focused on the uncertainty handling within the system. The metrics for the overall system capture not only the effects of the uncertainty representation but also the effects of other aspects that can affect the performance of the system (e.g., the implementation).

In 2010, an Uncertainty Forum was organized by Simon Maskell and John Lavery from the U.S. Army Research Office as part of the International Conference on Information Fusion (FUSION) held in Edinburgh, United Kingdom, to discuss some different ways of representing and dealing with uncertainty using a common and single scenario as a reference point. As the organizers mentioned, "The goal of the Uncertainty Forum is not to come to specific conclusions about a linear or other ranking of approaches for representing uncertainty but rather to widen the spectrum of available options and link these options with situations in which they perform well".[1] Prior to this forum, a vehicle-borne improvised explosive device (V-IED) scenario was submitted to five scientists, with each expert "defending" one approach or framework: the Bayesian method (Simon Godsill), Dempster–Shafer theory (Arnaud Martin), transferable belief model (David Mercier), Dezert–Smarandache theory (Jean Dezert), and human intelligence/processing (Peter Gill). The analysis conducted in [1] revealed that beyond the mathematical framework selected, personal choices of modeling impact the solution provided and the results obtained.

Hence, to help fusion scientists better sail among the different approaches dealing with uncertainty, the International Society of Information Fusion (ISIF) charted the Evaluation of Technologies for Uncertainty Representation (ETUR) working group,[2] which has been discussing this topic since FUSION 2012 in Singapore. The goal of this group is to provide a forum to address the problem of the assessment and evaluation of the different uncertainty representation approaches developed so far. The ultimate objective would be to provide objective cri-

[1] http://isif.org/fusion/proceedings/fusion2010/plenary-speakers.htm
[2] https://eturwg.c4i.gmu.edu

teria to assess specifically how uncertainty is handled in fusion systems and define basic concepts to be eventually accepted and standardized.

The main outcome is the uncertainty representation and reasoning evaluation framework (URREF), which includes an ontology, evaluation procedures and associated datasets, applications, and other components that aim at providing the foundational theory, mechanisms, and standardization artifacts required to evaluate the impact of uncertainty in information fusion systems. The framework provides a means for relating evaluation criteria specifically focused on uncertainty handling, with other information quality aspects such as the nature of uncertainty (aleatory vs. epistemic), the derivation of uncertainty (objective vs. subjective), the type of uncertainty (imprecision vs. uncertainty), and the uncertainty theory (belief functions, probability, fuzzy sets, possibility, etc.) [2], [3].

Within the URREF, a major task is to formally identify the concepts that are pertinent to the evaluation of uncertainty of an IFS, which is a seen as a means to ensure that all evaluations follow the same semantic constraints and abide by the same principles of mathematical soundness. This is enabled by the ontology reference model developed for the framework, known as the URREF ontology. The first stone to formal representation of the uncertain reasoning domain was put by the Uncertainty Reasoning for the World Wide Web Incubator Group of the World Wide Web Consortium, which published in March 2008 an uncertainty ontology "to demonstrate some basic functionality of exchanging uncertain information".[3] This effort was then pursued through the ISIF ETUR working group.

The URREF and its ontology component were developed through an iterative process, an essential part of which was to apply the framework to a set of use cases. The use cases not only serve as benchmarks but also reflect a range of considerations relevant to evaluation of uncertainty representation within the context of an overall fusion application. The focus is on high-level fusion tasks, which require a closer human interaction (human as a source or as a decision-maker). Applying the framework to use cases and their associated datasets grounds the ideas in concrete application areas and helps to uncover requirements that emerge as the framework is applied to a concrete problem.

As such, the work on developing these use cases has been generating new insights and requirements for the URREF (e.g., [4], [5], [6]). Among the different use cases proposed through the years, the three that have been mostly consistent throughout the discussions are *maritime surveillance,* where a harbor area is monitored by a set of sources mixing sensors and humans [6]; *rhino poaching,* which involves a decision support system that directs the patrol effort of the rangers to the areas with elevated risk of poaching [7]; and *cyber threat,* which comprises an expert model for cyber threat analysis [8].

The URREF is not a system or software application that can be "directly applied" to a use case. Yet the use cases are essential for the group to achieve an understanding of all the nuances and idiosyncratic aspects of the process of evaluating techniques that are fundamentally different in their assumptions and views of the world. They provide the grounding for establishing the URREF concepts and mechanisms needed to mitigate the effects that the underlying assumptions of each theory have in biasing the design of evaluations—each usually geared toward the strengths of one technique at the expense of the others. The URREF does not completely remove the subjectivity and biases involved in evaluating uncertainty representation techniques, but it is a strong step in that direction.

We offer a final thought about the ETUR working group, the URREF, and its unconventional nature and contributions to the IF society. On the one hand, the problem being addressed is a fundamental issue that requires a deep understanding of the many aspects of IF in general, of HLIF in particular, and of uncertainty theories. It is no coincidence that the group has such a wide background among its members, mixing expertise across the whole spectrum of the JDL model. On the other hand, its results can be assessed in terms of knowledge shared and formally captured about a difficult problem. This is when the group's work and contributions shine. Since 2012, it met in all FUSION conferences, with roughly 150 biweekly meetings at the time of this writing and 12 ETUR special sessions. In addition, a *Journal of Advances in Information Fusion* special issue, more than 70 articles in FUSION conferences, and tutorials, panels, and other events have brought a more throughout understanding of its topic and its importance to our community.

[3] https://www.w3.org/2005/Incubator/urw3/group/draftReport.html

REFERENCES

1. Jousselme, A.-L., and Maupin, P. Uncertainty representations for a vehicle-borne IED surveillance problem. In *Proc. Int. Conf. Information Fusion,* Singapore, July 2012.
2. Costa, P. C., Laskey, K. B., Blasch, E., and Jousselme, A.-L. Towards unbiased evaluation of uncertainty reasoning: The URREF ontology. In *Proc. Int. Conf. Information Fusion,* Singapore, July 2012.
3. Costa, P. C. G., Jousselme, A.-L., Laskey, K., Blasch, E., Dragos, V., Ziegler, J., et al. URREF: Uncertainty representation and reasoning evaluation framework for information fusion. *Journal of Advances in Information Fusion,* Vol. 13, 2 (Dec. 2018), 137–157.
4. Dragos, V., Ziegler, J., and de Villiers, J. P. Application of URREF criteria to assess knowledge representation in cyber threat models. In *Proc. 21st Int. Conf. Information Fusion,* Cambridge, United Kingdom, 2018.
5. Pavlin, G., Jousselme, A.-L., De Villiers, J. P., Costa, P. C. G., and de Oude, P. Towards the rational development and evaluation of complex fusion systems: a URREF-driven approach. In *Proc. Int. Conf. Information Fusion,* Cambridge, United Kingdom, July 2018.
6. Jousselme, A.-L. Semantic criteria for the assessment of uncertainty handling fusion models. In *Proc. 19th Int. Conf. Information Fusion,* Heidelberg, Germany, July 2016.
7. Koen, H., De Villiers, J. P., Pavlin, G., de Waal, A., de Oude, P., and Mignet, F. A framework for inferring predictive distributions of rhino poaching events through causal modelling. In *Proc. Int. Conf. Information Fusion,* Salamanca, Spain, 2014.
8. Kiesling, T., Krempel, M., Niederl, J., and Ziegler, J. A model-based approach for aviation cyber security risk assessment. In *Proc. 11th IEEE Int. Conf. Availability, Reliability and Security,* Salzburg, Austria, 2016, 517–525.