18th International Conference on Information Fusion
Washington, DC - July 6-9, 2015

A critical assessment of two methods for heterogeneous information fusion

Valentina Dragos
ONERA–The French Aerospace Lab
Chemin de la Hunière
Palaiseau, France
valentina.dragos@onera.fr

Xavier Lerouvreur
Airbus Defence And Space
Rue des Cosmonautes
Toulouse, France
xavier.lerouvreur@airbus.com

Sylvain Gatepaille
Airbus Defence And Space
Parc d'affaires des Portes
Val de Reuil, France
sylvain.gatepaille@airbus.com

Abstract* – Data fusion in heterogeneous environments plays a major role in assisting end users by providing them with an increased situational awareness so that decisions can be made about events in the field. Heterogeneous fusion involves combining different types of soft and hard data such that the situation or the resulting output is more precise, accurate, complete or easy to comprehend by decision makers. If soft data conveys more sophisticated information that is difficult to measure and hard data can be described with specificity, the question of how to take advantage of their complementarities is attracting considerable attention from data fusion community. This paper presents two methods for heterogeneous fusion, differing in procedures used to combine information items. We propose two methods that enrich a situation by adding supplementary attributes to entities, so that entities have a better characterisation. A domain ontology and reasoning capacities support both methods, although they implement different enrichment solutions. First, a picture of entities and relationships is created by using only hard data provided by sensors and then this picture is enriched thanks to soft data, in the form of succinct or more complex observation reports. The enrichment allows the situation to be understood and processed in a meaningful way by end users; however uncertainty arises as various items are matched. The paper also discusses underlying uncertainties induced by both methods along criteria of the current URREF framework.

Keywords: *heterogeneous fusion, soft data, ontology, uncertainty, URREF*

1 Introduction

Information fusion is at the core of many solutions developed to perform situation assessment for critical tasks, such as border monitoring, surveillance of areas or entity tracking. In complex environments, users have to fuse data provided by multiple sources in order to obtain reliable information that allows building an accurate picture of the situation. For example, an acoustic sensor alone cannot help a user to identify a vehicle.

There are three primary differences between sensor data and soft data which make heterogeneous fusion a challenging task. First, sensors typically deliver data in streams: data is produced continuously, often at well-defined time intervals, without having been explicitly asked for. Those streams need to be processed in near real-time, as data arrives, because sensor streams can refer to real-world events, like traffic accidents, which need to be responded to. Besides, saving raw sensor streams to disk can be expensive. In contrast, soft data arrives sporadically, late after the event occurred. Second, sensors are blind producers of data and thus they are unable to change data delivered. Soft data is the result of various modes of perception, affect, skills and knowledge which all fall under the umbrella of subjective assessment and include social and cultural formations that shape any individual. In addition, humans also have the ability to provide a personal view, attitude or appraisal, in the form of a judgment of belief when reporting facts and events. Finally, sensors are typically connected together in ad-hoc networks that cover a geographic area, such that receiving data from arbitrary nodes allows the user to take into account the network when performing the analysis. Unlike sensors, human sources can be part of hidden networks, whose undisclosed ties can lead to unreliable conclusions. For example, dependent information pieces can be analyzed under the false assumption of there being several independent sources, with impact on the result accuracy. This paper presents two methods to fuse heterogeneous information for entity tracking and identification. A general architecture was developed, which creates a situation thanks to sensor data and then enrich this situation by adding soft data items. Both methods are built upon the same architecture, but they provide different enrichment mechanisms. The analysis of underlying uncertainties is carried out by using the URREF ontology [8] as a common basis, along various criteria applying to soft data and to data handling and assignment.

The paper is structured as follows: section 2 discusses the application context, an entity tracking and identification scenario. Related approaches are presented in section 3. Fusion methods are introduced in section 4 while analysis of uncertainty is carried out in section 5. Conclusion and perspectives for future work end this paper.

2 Application context

2.1 Entity tracking and identification

Identifying a mobile entity, knowing exactly where the entity is and monitoring its trajectory in real-time has already attracted a lot of interests from both academia and industrial communities, due to the large number of applications it enables. The work described in this paper is part of a research project aiming at developing incremental capabilities for information fusion to be further tested in realistic operational environments. As described in more details in [22] this project provides a comprehensive development framework and execution environment of algorithmic software components for information fusion, which can be then linked into processing chains. The obtained chains can implement fully capable and effective fusion architectures according to their design. A more specific goal of this project is to develop solutions able to combine sensor and human-based data adapted to surveillance–related queries. Although a basic scenario for entity tracking and identification was adopted for illustration purposes, the problem tackled by this paper is the combination of soft and hard data, thus tracking algorithms based exclusively on sensor data are not described in detail.

2.2 Definition of a situation

Data provided by various sensors along with human reports or brief messages are fused to identify and track several entities in order to monitor and protect a zone of interest. The outcome of the fusion is a situation, to be provided to men in the field involved in operations or to commanders in tactical and operational headquarters. Each entity is described as a vector of features, which, according to the sensor data used in the fusion process, provides the position and kinematics of the entity, its type (environment domain and nature) and also relations to contextual information such as geographical features (roads, airways or navigable waterways to name a few) or to other entities in the situation. More precisely, an entity is described as a set of states, representing the knowledge of this entity at a point in time. An entity state gathers all the estimated features mentioned earlier as well as more technical information related to traceability and information assessment, such as state likelihood, for instance. Hence, a situation of *n* entities can be defined as the union of the set of entities $\{E_p\}_{p\in\{1,\dots,n\}}$ and the set of $q+p$ collected observations $\{O_i^{Sensor}, O_j^{Soft}\}_{i\in\{1,\dots,p\}, j\in\{1,\dots,q\}}$ some of which could be false alarms or disinformation. One can add the set of contextual information (geophysical but also a priori knowledge) which is also an input of the fusion process. Each entity E_i has a set of states $\{ES_k\}$ with $ES_k=\{t_k, K_k, Tr_k, A_k\}$, a time stamped vector of features, composed of the knowledge K_k (kinematics, nature and additional properties), the traceability to observations used to produce K_k, and the assessment of K_k,represented as a probability or a likelihood or even as a simple score. Entity states can be built upon sensor-based data and soft observation reports: this only depends on the ability of the algorithms to associate these observations to a given entity. The data model used in the framework to represent the situation or picture of it at a given moment is relational, and so, implemented in XML but also designed to embed more generic knowledge representations such as generic properties in the form of hierarchical Key/Value pairs or ontologies in XML compatible syntax, typically RDF-S [30] or OWL [10].

2.3 A general architecture for heterogeneous fusion

The general architecture developed to implement heterogeneous fusion is illustrated in fig. 1. It should be noted that in this architecture we focus only on the general cycle allowing to take into consideration both sensor and human –based data, although technical details of the joint processing are provided in next sections.

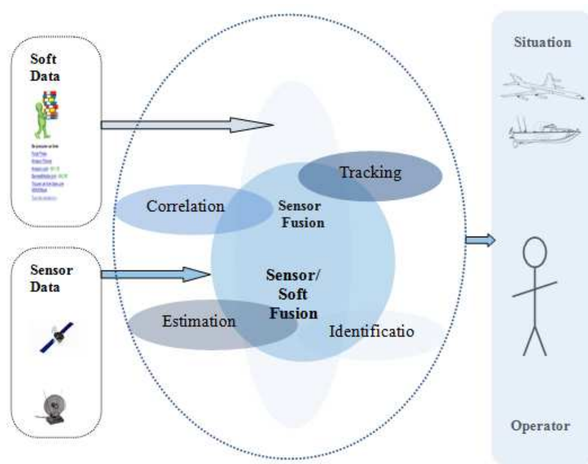

Figure 1: General architecture for heterogeneous fusion

Entity tracking and identification are carried out by using several sources that generate information items that can be classified as either hard or soft. Sensors create items by measuring physical properties of entities with quantitative values, errors and biases. For this work, the types of sources available include ground moving target indicator (GMTI) infrared and visible light imaging (IMINT) and signals intelligence (SIGINT) sensors. All of these sensing modalities generate elements that can be described by mathematical and numerical or symbolic representations (e.g. using a universe of discourse), and serve as inputs to automated processing procedures. Sensor measurements result in observations of objects for which they provide information about properties like location, speed or signal characterization when these objects are electromagnetic emitters. Soft information derived from human or open source is fundamentally different in that their content is often more qualitative and requires additional context elements for complete human interpretation. By definition, soft information requires a human observer somewhere along the intelligence processing chain, be that directly in the production of reports or messages, or indirectly in the gathering of human data. Besides, human sources are often the only observers having the ability to identify or to infer complex entities, such as cells of insurgent activities or relationships of persons of interest. Subjective content is often introduced in soft data, as humans use inherently their own knowledge to represent and analyse the world they perceive, and because they communicate observations by using languages. As the fig. 1 illustrates, heterogeneous fusion is carried out thanks to two information fusion cycles, designed to take into account characteristics of sensor and human sources. The core of our architecture is a sensor-based kernel fusion that provides several processes for entity correlation and tracking along with estimation of their states. The kernel implements a short-time classical tracking algorithm, as data are provided by sensors on a regular frequency. The outcome of this process is a situation, whose entities are described by their spatio-temporal coordinates and their kinematics. At this stage, the type of entities is also estimated but only thanks to sensor-based data. The second layer of this architecture enriches this previously created situation by integrating soft data elements on a stream and irregular basis, as they become available. The enrichment aims at refining the states of entities (more precise location, more accurate type description) or at adding supplementary attributes to entities (allegiance, military or civilian nature, etc.). Thus, heterogeneous fusion can be considered as a long-time fusion cycle, triggering specific processes as soft data observations arrive. Those processes first provide matching mechanisms to assign soft data observations to entities of the situation and then perform fusion strategies in order to combine elements of entity states with items extracted from soft data.

2.4 Information fusion and user analysis

This general architecture is intended to ease the development of applications based on heterogeneous information fusion, from analysis of sources to end user interpretation. However, there are several issues to consider when performing hard and soft data fusion by using the architecture and its embedded fusion cycles. Thus, association approaches try to assign real observations to entities of the situation, although some of those entities do not necessarily correspond to real-world objects. In contrast, soft data can provide elements about objects that are not part of the situation, as it is the case with complex entities such as convoys that sensors fail to identify which may result in creating an incorrect picture of the situation. Traceability is another important aspect to be considered, as the final situation is composed of entities whose states are estimated by using several sources. Therefore, a user analysing the situation must be able to retrieve the observations used to infer a specific entity feature. And finally, several types of uncertainties arise depending on the solution adopted for the overall heterogeneous fusion approach, which is to say both matching of soft sources and enrichment of the situation. Uncertainty criteria capture the imperfections of various approaches and the level of this uncertainty must be accounted for as data fusion progresses.

3 Related work

Heterogeneous information fusion is an emerging topic within the information fusion community, addressing both theoretical and applied aspects relevant for higher levels of the JDL model [23]. Various research efforts have addressed the fusion or analysis of soft data, by providing solutions to structure free messages [3] or to classify reports [6]. Although based on natural language processing techniques [19], several approaches aim at integrating semantic analysis [17], [18] when processing soft data in order to overcome limitations of key-word spotting [16] and have ambitious goals, such as threat recognition [26]. However, the challenge of heterogeneous fusion is to combine both sensor and human based information items.

Generic mathematical frameworks for heterogeneous fusion can be approached from either a parametric or a nonparametric modeling perspective. The first ones specify parametric models of the state space based on domain knowledge, whereas in the non-parametric approach these models are empirically determined from data. A common parametric approach to tackle heterogeneous fusion uses probabilistic graphical models to build a unified representation of data, such as Hidden Markov Models or richer Bayesian Networks [9]. In practice, such models may be difficult to construct due to limited domain knowledge. Random Finite Set theory (RFS) is another theory suitable for the fusion of disparate information [21]. RFS uses the random finite set as a common representation and data in the form of qualitative statements or quantitative values are translated into this representation. RFS can also encode the disparate forms of uncertainty inherent in the data.

In [28] the foundation of an emerging framework for hard-soft information fusion based on Dempster-Shafer (DS) theory is described. The solution takes into account inherent data and source uncertainties such as reliability and credibility, as introduced by NATO standards [25] and uses the conditional approach, an extension of DS frame to handle soft data.

A graph-based representation is proposed in [13] to process uncertainties of soft data for the purpose of situation assessment. An attributed data graph is created to represent intelligence information, where nodes represent entities and the arcs correspond to relationships. A template-graph is used to specify a situation of interest,

and uncertainties correspond to an inexact matching of those structures.
From a different perspective, various approaches propose solutions for hard-soft data fusion by using Controlled Natural Languages in the form of subsets of natural languages, obtained by restricting the grammar and vocabulary in order to reduce or eliminate ambiguity and complexity. In the military field, BML (Battle Management Language) [4] was developed based upon the Joint Consultation, Command and Control Information Exchange Data Model (JC3IEDM), [24]. It provides a standardized representation to communicate orders, request and reports, and it is sufficiently expressive to formulate both military and non-military data exchanges for a variety of tasks.
An analytical review of recent developments for multi-sensor information fusion is presented by Khaleghi and colleagues in [20] while trends and pitfalls of soft data integration are discussed in [12].

4 Methods for heterogeneous fusion

This section presents two methods for heterogeneous information fusion, combining sensor data and human observations. Observations are messages in the form of XML files providing, in addition to soft information, several meta-data such as: geographical area of reported events, time of information delivery or its security level.
While both methods perform the assignment of soft observations to entities and the refinement of their states according to incoming items, they differ in the way those procedures are carried-out, as described hereafter.
A domain ontology [15] supports both methods, providing a standard model of entities along with relationships. The ontology is represented in OWL [10], a description–logic based formalism [2], allowing logical inferences.

4.1 Method 1: Assign then fuse

The overall solution adopted by this method is summarised in fig. 2 and consists of the following two steps: assignment of observations to entities of the situation and then fusion of new elements of observations to attributes of entities.
Description of data**:** Soft data used for this method are brief annotations of sensor-based reports. Each annotation is a short message, conveying information about the type of entities. The type, as stated by those messages, will not be considered for tracking purposes, instead it will help human-operators to have a better description of the situation. Besides, the type can be iteratively refined, as more observations arrive.
Processing steps: Assignment of observations to entities is carried out in the light of spatio-temporal correlation. As observations are associated with a timestamp and have specific locations, this method first estimates a correlation coefficient to describe the probability of an observation to be assigned to an entity of the situation.
The current states of the entity along with its previous states are taken into account for this estimation, as soft observations are not necessarily synchronised with the current situation.

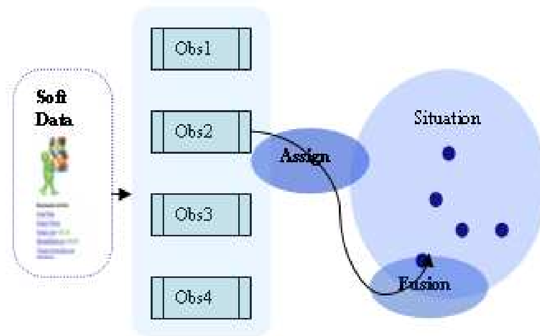

Figure 2: Method 1: assign then fuse

Results of this estimation are then ranked and the observation maximizing the value is selected as input to the fusion step. Fusion is the second step of this method and it aims at refining the type of entities by taking into account their initial type, as provided by tracking and identification algorithms, and the type, as provided by soft observations.
For this method both labels describing types are previously labelled by a domain ontology and the fusion consists in using reasoning mechanisms to combine them. More specifically, as operators are interested in having a precise description of entities, the fusion algorithm identifies the most specific concept of both labels, which is then used to describe the entity, as illustrated in fig. 3, where the final state of entity highlights the type "bus", as a more specific and informative concept than "vehicle".
The outcome of this method is the refinement of entity types which allows a more accurate description of entities and improves the overall situation.

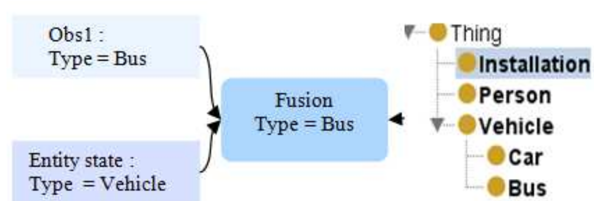

Figure 3: Refinement of type by fusion

Remarks: This method is suitable for situation assessment in dynamic environments, when brief observations are constantly arriving and there is a need to constantly update entity states. Those changes are not subject to traceability requirements, given their frequency.

4.2 Method 2: Fuse then assign

Fig. 4 illustrates this solution that consists on first a fusion of information extracted from soft data in the form of features of entities and then the assignation of those attributes to entities of the situation.

Description of data**:** Data used for this method is composed of free text messages, whose size can range from a few phrases to more important volume. The content is also heterogeneous and messages can refer to different aspects such as entity location, evolution, but also, in a more general context, knowledge about political and social environments. For this method, soft data will be used to retrieve additional attributes of entities. More particularly, attributes describe the type (vehicle, bus, person, etc.), allegiance (foe, friend, neutral) and nature (civilian, military, insurgent, etc.) of entities.

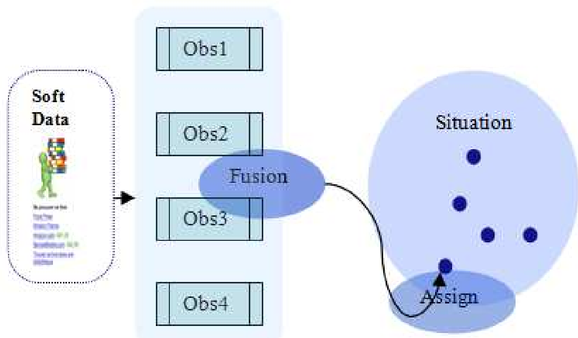

Figure 4: Method 2: fuse then assign

Processing steps: The first step aims at exploiting free text messages in order to extract features of entities. We adopted an approach based on collocation identification. Collocations are associations of words which co-occur frequently within the same sentence, whether because the meanings of words are related to each other (i. e. vehicle-road, car-driver) or because the two words make up a compound noun (car stop, subway station). We focus on collocations composed of two words (also named by-grams) as often entities of interest are named by short lexical units.

Traditional methods for collocation extraction are based on the evaluation of a statistical score to estimate the relevance of word pairs. However, data used for this work are short messages whose volume is not appropriate for statistical validation. For this reason, we developed a complementary approach able to normalize the meaning of collocations thanks to an existing ontology. First, collocations are generated by using a window placed over a sentence, such that two words are analysed at a time and moving the window from the first to the last word of the sentence. Then the meaning of the collocation is analyzed by using an ontology, whose concepts are used to label the overall collocation or its words, individually. Lexical similarities are used to label collocation by ontological concepts. Lexical similarity associates a real number to a pair of words and it is a measure of the degree to which the words are similar. Several measures were proposed to estimate lexical similarity, a selection of which is presented by Cohen and colleagues in [7]. At the end of this phase, ontological entities are assigned to collocations, as shown in tab. 1.

Collocation	Ontology concept
Unknown vehicle	Vehicle
Bus moving	Bus

Table 1: Concepts –collocation matching

For this work we use an ontology of the ISR[1] field, providing us with a standard model of various domain entities and relationship. This ontology was created from scratch, thanks to support of domain experts and highlights the main categories of domain entities while provides categories to classify them, as illustrated in fig. 5. Having this specific domain model allows us to improve the semantic description of collocations, by adding the class of concepts, as shown in fig. 5.

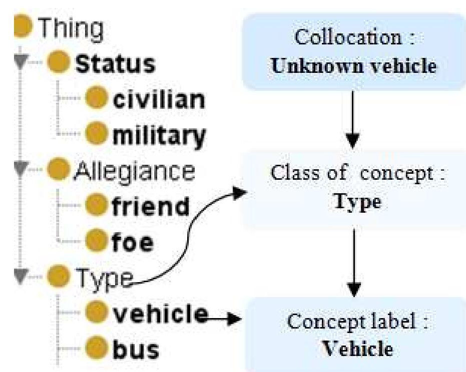

Figure 5: Ontology and semantic labelling of collocations

The processing chain previously described identifies at most three features of entities, according to both the content of messages and the classes of concepts as modelled by the ontology. Those features are used to create additional properties, assigned afterwards to entities in the light of spatio-temporal correlation, by taking into account the current and previous states of entities and the time stamp of the observation report out of which features were extracted. The outcome of this method is the enrichment of entity states, by adding supplementary properties which cannot be inferred from sensor data, and the construction of a refined and more comprehensible situation.

Remarks: This method is suitable when situation assessment involves a rather long-time analysis of a dynamic evolution and triggers modifications of the situation that have the ability to impact the overall picture; therefore there is a need to keep track of the chain of sources providing various elements.

[1]Intelligence, Surveillance & Reconnaissance

5 Assessment of uncertainty

When performing heterogeneous fusion by following the aforementioned methods, uncertainty is introduced into the final outcome via: soft data itself, the more or less precise assignment of items to entities of the situation, and the fusion process. In this section we analyse those uncertainties under URREF [8], the uncertainty representation and reasoning evaluation framework. URREF is a generic framework developed to facilitate the analysis of uncertainty for high level information fusion, and it is supported by an ontology modelling criteria along with related elements. The framework was already used for various applications such as the evaluation of fusion approaches based on Bayesian networks [27] or the analysis of wide-area motion imagery solutions [5]. Efforts to develop a generic methodology for URREF use cases are discussed by Ziegler and colleagues in [29]. For this work, the analysis is carried out without taking into account the quality of the initial situation. The process under analysis is the enrichment of the situation, and a critical assessment is performed in order to compare uncertainties underlying both methods.

5.1 Analysis of soft data uncertainties

The set of criteria discussed in this section refer to the characterisation of soft data reports as they are provided by information sources, and before they enter any processing. For the first method, as reports are very short messages that can be rather considered as annotations, criteria relevant for uncertainty analysis are: relevance to problem, accuracy and precision. Relevance to problem highlights a piece of information related to the application field, such as *Vehicle*, a main concept of the ontology. Accuracy can be used to capture the extent to which the information piece agrees with the standard model, as it is the case when various sources use different words to name entities (i.e. *light-duty vehicle* instead of *vehicle*). Accuracy is then related to the use of non-ambiguous lexical expressions. Precision captures the ability for an information item to relate to more specific concepts of the ontology. For instance, *school bus* is more specific that *bus*. As data used for the second method are more complex reports, they can be characterised by credibility. Under URREF, credibility is specified by three criteria (Observational sensitivity, self-confidence and Objectivity) intended to capture the ability of an item to be believable or worthy of trust. For our second fusion method, Self-confidence is relevant to analyse the quality of soft data reports, and it can be assessed by superficial natural language processing [11]. It should be noted that for the first method, uncertainty criteria provide a characterisation of the intrinsic quality of data, while for the second method the characterisation is done in a relative manner.

5.2 Analysis of data–handling uncertainties

Traceability and interpretation are two criteria provided by URREF to analyse uncertainty related to data handling. For this work, interpretation can be used as related to the number and quality of attributes each method is able to add to each entity of the situation. Thus, intuitively, the second method will provide a more comprehensive picture to end users, as it is able to specify the type, the allegiance and the nature of entities while the first method has the capacity to improve only their types. The second data-handling criterion, traceability captures the ability of keeping track of the evolution of entities, as additional properties are added, and therefore to chronologically interrelate various entity states in a way that is verifiable. Traceability is relevant only for the second method, as it implements a long-time fusion cycle and attributes provided can be crucial for the course of action, as it is the case with entity allegiance. We do not consider traceability for the first method as sources deliver fast-changing items having a limited influence on the final outcome.

5.3 Analysis of assignment uncertainties

For both methods, spatial and temporal coordinates are used to identify entities of the situations to which soft data items can be assigned.

Two mechanisms induce uncertainties when performing this assignment; first, the spatio-temporal correlation is not based on a perfect match of values, but rather on the analysis of various entity positions at different times; second, even in the case of a strong correlation, features provided by soft data should still be relevant to entities under analysis. Under URREF ontology, correctness and consistency are criteria able to capture those uncertainties. Thus, correctness can be estimated in the form of a distance highlighting how close a set of observations and a given entity are according to space and time coordinates. Consistency is intended to capture how coherent the entity states and incoming observations are. The criterion can be assessed thanks to a similarity measure, by implementing specific estimation procedures for each method. While for the first method, correctness is a function of the similarity between the annotation, as provided by observation reports and the type of entity, as assigned by classification algorithms, for the second method, correctness will take into account only the two types of entity, as they are provided by the fusion and the classification algorithms, respectively.

Correctness and consistency are independent criteria, however, from a practical standpoint, one or other criteria can be estimated first and low values of it can be sufficient to point out high uncertainties.

5.4 Analysis of fusion uncertainty

Fusion creates for both methods a new situation according to soft data input. Four criteria can be used to capture

induced uncertainties, all falling under the umbrella of uncertain evidence, as modeled by URREF.
Fig. 6 illustrates a comparison of uncertainty evaluation for both fusion methods along the same set of URREF criteria. Besides showing criteria specific to each method according to input data and processing chains, the picture also shows the way various criteria are to be considered in time.

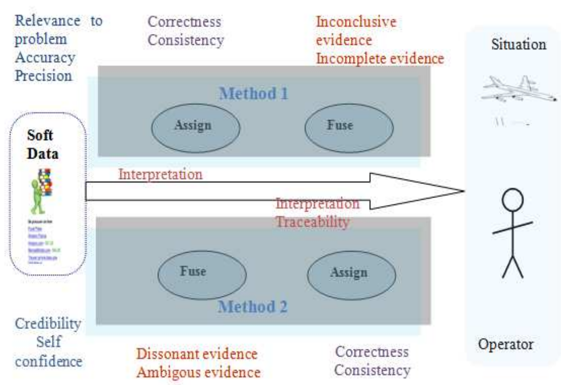

Figure 6: Appraisal of uncertainty assessment

The first method updates entity types by following inferences based on ontological descriptions of the type. First, incomplete evidence can highlight soft data that seem to confirm a particular feature, while ignoring significant related features data that may contradict the overall state. This can be the case with sources used by the first method, who specify the type of entities (bus, vehicle, or person) without realizing that they are part of a group (convoy). Although they provide reliable attributes, sources ignore the overall picture and thus the evidence is not complete. Such incomplete statements are inherent when natural language is used to convey data, as it was already shown by Auger and colleagues in [1].
Dissonant evidence is a criterion relevant for the second method, when various soft reports are gathered before performing the identification of features and their integration into the situation. This criterion implies a clash between features identified within reports and it can be captured in the form of violations of ontological assertions (i.e. pedestrian and vehicle as entity types, although it is obvious that an entity can have a type or another).
Ambiguous evidence is a criterion intended to capture the ability for an information item to equally refer to several different entities. Ambiguous items can be provided by the second method, namely when the only attribute identified is the nature of entity and it proves to be insufficient to identify a unique entity whose state will be updated. Finally, inconclusive evidence is used to describe situations when the input of the fusion process in composed of items that don't allow inferring a conclusive and definitive description. This criterion is relevant for the first method, to point out cases when it is not possible to infer a new type by combining the entity type as provided by the tracking and identification algorithms and soft data annotations. Without facing a contradiction such as military vs. civilian and while being consistent with the ontological model (both types are ontological concepts), the inferences are unable to find a common concept as a synthetic description of inputs.
Those cases should be triggered as inconclusive to users, who can then provide an expert solution to solve the inconclusiveness (keep the initial entity type, for instance). By using URREF, uncertainty related to both methods is described by a set of common criteria. Nevertheless, as shown in fig .3, those criteria are to be estimated at different processing steps for each method. The use of URREF criteria offers a common ground to assess uncertainties, once we identify at which points they enter into different processes, how data flow from one step to another and how it is transformed within the system for each method.

6 Conclusion and future work

This paper proposes two methods for heterogeneous information fusion and discusses their conceptualizations and challenging aspects, along with an analysis of underlying uncertainties.
Heterogeneous fusion is performed by building a situation thanks to sensor data and refining this situation by using soft data observations. While having the same architecture, methods differ in soft data processing. URREF criteria are used for uncertainty analysis, as a common basis for automatic detection of irrelevant data and assessment of uncertainty throughout.
Directions for future work are threefold. First, we can improve the overall fusion approach, by allowing a better integration of soft data observations. Thus, entities can also be created based on soft observations.
Second, the implementation of uncertainty criteria will allow us to have numerical estimations to characterise the process. This implementation can be based on ontological semantics, as the solution developed in a previous work by Gurevych and colleagues to assess coherence values [14] to natural language statements.
And third, a joint protocol is needed explaining how uncertainty criteria can be considered in order to avoid the processing of irrelevant data or the propagation of unreliable results.
This protocol requires human intervention, since there are still roles for domain experts and analysts like experience and intuition based decision making which are extremely difficult to integrate in the form of automatic estimation of uncertainty criteria.

References

[1] Auger, A. and Roy, J., *Expression of uncertainty in Linguistic Data*, Proceedings of the 11th International Conference on Information Fusion, Köln, Germany, 2008.

[2] Baader, F., Calvanese, D., McGuiness, D., Nardi, D., and Patel-Schneider, P.-F. Editors, The description logic Handbook: Theory, Implementation and Applications, Cambridge University Press, 2003.
[3] Biermann, J. et all., From Unstructured to Structured Information in Military Intelligence: Some Steps to Improve Information Fusion, SCI Panel Symposium, London, United Kingdom, 2004.
[4] Biermann, J., Nimier, V. Garcia, J. Rein, K., Krenc, K., Snidaro, L.: Multi-level fusion of hard and soft information, Proc. of the 17th International Conference on Information Fusion, Salamanca, Spain, 2014.
[5] Blash, E., Costa, P. C. G., Laskey, K. , Ling, H., and Chen, G., The URREF Ontology for Semantic Wide Area Motion Imagery Exploitation. In Proc. of the 7th International Conference on Semantic Technologies for Intelligence, Defense, and Security (STIDS 2012). Costa, P.; Laskey, K. (eds.), pp. 88-95. Fairfax, USA, 2012.
[6] Carr, O., and Estival, D., *Document classification in structured military messages,* Proceedings of the Australasian Language technology Workshop, 2003.
[7] Cohen, W., Ravikumar, P., Fienberg, S.: A comparison of string distance metrics for name-matching tasks. In proceedings of IJCAI-2003, Workshop on Information Integration on the Web pages, 2003.
[8] Costa, P.C.G., Blackmond Laskey, K., Blasch, E., Jousselme, A.L., Towards unbiased evaluation of uncertainty reasoning: The URREF ontology., Proc. of the 15th Conference on Information Fusion, Singapore, 2012.
[9] Das, S., High-level Data Fusion, Artech House 2008.
[10] Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P.F., Rudolph, S.: OWL 2 Web Ontology Language Primer, W3C Recommendation (2009).
[11]Dragos, V., Assessment of uncertainty in soft data: a case study, In Proc. of the 17th International Conference on Information Fusion, Salamanca, Spain, 2014.
[12]Dragos, V. and Rein, K. Integration of soft data for information fusion: pitfalls, challenges and trends, In Proc. of the 17th International Conference on Information Fusion, Salamanca, Spain, 2014.
[13] Gross, G., Nagi, R. and Sambhoos, K., Soft information, dirty graphs and uncertainty representation/ processing for situation understanding, Proc. of the 13th Int. Conf. on Information Fusion, Edinburgh, UK, 2010.
[14]Gurevych, I., Malaka, R., Porzel, R., Zorn, H.P., *Semantic coherences scoring using an ontology*, Proc. of the North American Chapter of the Association for Computational Linguistics on Human Language technology, Stroudsburg, USA 2003.
[15] Gruber, T. A translation approach for Portable Ontology Specification, *Knowledge Acquisition*, 5(2), 199-220, 1993.
[16]Haarmann, B., Sikorski, L. and Schade, U., *Text analysis beyond keyword spotting*. In Proceedings of Military Communication & Information Systems Conference (MCC) Amsterdam, Nederland, 2011.
[17]Hecking, M., *Content analysis of HUMINT reports*, Proceedings of the 11th Command and Control Research and Technology Symposium, June 2006, San Diego, USA.
[18]Hecking, M., *System ZENON .Semantic analysis of intelligence reports,* Proceedings of LangTech forum, Rome, Italy, 2008.
[19] Jenge, C., Kawaletz, S., Schade, U., *Combining different NLP methods for HUMINT reports analysis*. NATO RTO IST Panel Symposium, Stockholm, Sweden.
[20] Khaleghi, B. , Khamis, A., Karray, F.O., Razavi, N., S., Multisensor data fusion: A review of the state-of-the-art, Journal of Information Fusion, Volume 14 Issue 1, January, 2013 pages 28- 44.
[21] Khaleghi, B., Khamis, A. and Karray, F. Random finite set theoretic based soft/hard data fusion with application for target tracking, in Multisensor Fusion and Integration for Intelligent Systems (MFI), 2010, IEEE Conference on, pp. 50–55, Salt Lake City, USA, 2010.
[22]Lerouvreur, X., Dragos V., Dambreville F., Principles of a Unified Framework for Heterogeneous Information Fusion, In Military Communications and Information Systems Conference MCC'2013, St Malo, France, 2013.
[23]Llinas, J., Bowman, C., Rogova, G., Steinberg A., Waltz, E., and White, F. *Revisiting the JDL data fusion model*, Proc. of the 7th International Conference on Information Fusion, Stockholm, Sweden, 2004.
[24]MIP JC3IEDM Overview, Greding, Germany, 2005,http://mipsite.org/publicsite/Baseline_3.0/JC3IEDM-Joint_C3_Information_Exchange_Data_Model/
[25] NATO STANAG 2511 Intelligence Reports, 2003
[26] Sikorski, L., Haarmann, B., Schade, U., *Computational linguistic tools exploited for automatic threat recognition,* NATO RTO IST Panel Symposium, Madrid, Spain 2011.
[27] De Villiers, JP., Pavlin, G., Costa, P., Laskey, K., Jousselme, A.L. , A URREF interpretation of Bayesian network information fusion, Proc. of the 17th Int. Conf. on Information Fusion, Salamanca, Spain, 2014.
[28]Wickramarathney, T. L., Premaratney, K., Murthiy, M.N., Scheutzz, M., Kubler, S., and Praviao, M., Belief theoretic methods for soft and hard data fusion, In Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, Czech Republic, 2011.
[29] Ziegler, J. Detje, F. Application of empirical methodology to evaluate information fusion approaches, Proc. of the 16th International Conference on Information Fusion, Istanbul, Turkey, 2013.
[30]W3C, RDF Schema 1.1, W3C Recommendation, 25/02/2014, Available online, http://www.w3.org/TR/rdf-schema, Accessed 16/02/2015.