Tracking uncertainty propagation from model to formalization: illustration on trust assessment

Valentina Dragos,* Jean Dezert, * and Kellyn Rein, †
* ONERA-The French Aerospace Lab, Palaiseau, France
valentina.dragos@onera.fr, jean.dezert@onera.fr
† ITF, Fraunhofer, FKIE, Wachtberg, Germany
kellyn.rein@fraunhofer.fkie.de

***Abstract*—This paper investigates the use of the URREF ontology to characterize and track uncertainties arising within modeling and formalization phases. The estimation of trust in reported information, a real-world problem of interest to practitioners in the field of security, was adopted for illustration purposes. A functional model was developed to describe the analysis of reported information, and it was implemented with belief functions. When assessing trust in reported information, the uncertainty arises not only from the quality of sources or information content, but also due to the inability of models to capture the complex chain of interactions leading to the final outcome and to constraints imposed by formalisms. A primary goal of this work is to separate known approximations, imperfections and inaccuracies form potential errors, while explicitly tracking the uncertainty from modeling to formalization phases. A secondary goal is to illustrate how criteria of the URREF ontology can offer a basis to analyze performances of fusion systems at early stages, ahead of implementation. Ideally, as uncertainty analysis runs dynamically, it can use the existence or absence of observed states and processes inducing uncertainty to adjust the tradeoff between precision and performance of systems on-the-fly.**

***Index Terms*—uncertainity, reported information, trust, belief functions, information fusion, DSmT, URREF ontology**

I. Introduction

A key element when designing an information fusion system is the way the system designer isolates and analyses real world phenomena. A model is abstracted into simpler representation, in which components, modules, interactions, relationships and data flows are easier to express. The decisions made when defining a model and choosing a formalization solution can affect the way a fusion system could be implemented to meet practical requirements with acceptable quality of results. Analysis of sources and types of uncertainly impacting the results is then a critical precursor to system implementation. It aims at developing a meaningful assessment of whether the solution is able to deliver an outcome whose quality fulfills the operational need.

Uncertainty tracking provides guidance on highlighting common challenges when abstracting a relevant but still simpler world description, selecting the right level of detail in building the model, detecting the approximations induced by formalization and providing a checklist to help ensure that all uncertainty sources and factors have been identified and considered ahead of system implementation. Done in early stages of system implementation, uncertainty tracking can be a powerful basis for developing analysis-based projections of performance.

This paper illustrates the use of the evaluation criteria defined by the uncertainty representation and reasoning framework (URREF) ontology to identify and assess uncertainties arising during the modeling and formalization phases of an information fusion system intended to estimate the trust in reported information.

Trust in reported information is a real-world problem grounded in many applications relying heavily on reported items, with different persons observing and then reporting on objects, individuals, actions or events. For such contexts, using inaccurate, incomplete or distorted items can result in unfortunate consequences and analysts need to ensure the consistency of reported information by collecting multiple items from several sources.

From the perspective of an information analyst, trust can be analyzed along two dimensions: the subjective evaluation of items reported by the source itself, called self-confidence and source's evaluation by the analyst, in the form of reliability. While self-confidence encompasses sincerity or intention, the reliability of sources is rather related to the quality of previously reported items, the competence of the source for specific topics and misleading intent. Trust estimation aims at capturing into an aggregated value the combined effects of self-confidence and reliability on the perceived quality of information. The model is represented with belief functions, a formalism which offers sound mathematical basis to implement specific fusion operators capable to estimate trust by combining self-confidence and attributes of source's reliability.

Trust estimation offers a good application to illustrate uncertainty tracking: the phenomenon is complex, so that any model adopted is a simplification of the real world interactions. Uncertainties can be made explicit for static elements of the model, such as sources or items, but also for the dynamic processes of combining items with one another. Moreover, there are several formalisms suitable to describe the fusion, and adopting belief functions will have an impact on the way the information system could be implemented and the uncertainty of its results.

The remainder of this paper is divided into 8 sections: section II discuses related approaches developed to tackle trust modeling and assessment. The problem addressed in this paper in presented in section III. Section IV describes the model developed to analyze reported information and its implementation with belied functions is presented in section V. The analysis of uncertainty during modeling and formalization phases is discussed in VI while examples and several scenarios for trust assessment are discussed in section VII. Conclusions of the paper are presented in section VIII.

II. Background and related work

The subject of trust modeling and assessment is not a new research topic, spanning areas as diverse as agent systems Granatyr et al. [2015], logical modeling and argumentation Paglieri et al. [2014], service provision on the Internet Josang et al. [2007], decision making under uncertainty, social networks analysis Sherchan et al. [2013], or crowdsourcing applications Venanzi et al. [2013]. Applications of trust analysis in are also of interest in the military field where techniques were developed in order to identify clues of veracity in interview statements Twitchell et al. [2006].

The concept of trust in these communities varies in how it is represented, computed and used. Although having an obvious social dimension, trust is not only understood and addressed towards other humans, but also towards information pieces Venanzi et al. [2013],information sources Koster et al. [2017], Internet sites Dong et al. [2015],data and knowledge fusion algorithms Dong et al. [2014], intelligent agents Granatyr et al. [2015], services in Internet of things Guo and Chen [2015].

As pointed out in Grandison and Sloman [2000], trust is tied up with the relationships between individuals and is directly related to their actions . While encompassing aspects of human interactions, the concept of trust can have specific scopes such as: provision,is a trust dimension occurring between the user of a service or resource, and the provider of that resource; access trust links the owner of a resource and users accessing the resource; delegation connects the individual who delegates responsibility for some action or decision and the individual to whom that action or decision is delegated; identity trust is about being certain that the identity if an entity is identical to the identity claimed by the entity itself.

While definitions of trust vary from one domain to another, there are clearly some common elements. The first commonality for all research areas cited above is to consider trust as a user-centric notion that needs to be addressed in integrated human-machine environments which rely heavily on information collected by human even if further processing can be executed automatically. Second, all definition associate some degree of uncertainty with trust, which is then captured by concepts such as subjective certainty or agent's beliefs Falcone and Castelfranchi [2001], subjective probability Castelfranchi and Falcone [2000] or even the feeling of security McKnight and Chervany [1996].

An interesting analysis intended to capture the lexical dimension of trust as a concept is described in [20]. The authors used collocation analysis exploiting mutual information to identify words that occur most frequently with trust. Mutual information is a method characterizing the strength of word associations in the form of a statistical association of words. Results obtained in the on the basis of COCA (Corpus of Contemporary American English) corpus highlighted integrity, honesty, confidence, competence and character as nouns closely related to trust in lexico-semantic spaces.

Trust goes hand in hand with veracity and deception concepts and several research effort aimed at identifying features of veracity (truthfulness) and deception in natural language statements.

In the field of natural language processing, veracity in related to the identification and quantification of features of certainty/ uncertainty. In the context of textual data analysis, Lukoianova and Rubin [2014] addresses veracity along the dimensions of truthfulness /deception, objectivity / subjectivity and credibility / implausibility. The authors developed a veracity index ranging from true/objective/credible to untrustworthy/subjective/ implausible to characterize texts in the context of big data analysis. While a comprehensive description of veracity in textual data, along with methods developed to identify veracity and misinformation are presented in Berti-Équille and Borge-Holthoefer [2015], the topic is also considered as part of a larger research direction tackling opinions detection.

Opinion detection is part of subjectivity analysis, which focuses on the automatic identification of private states, such as emotions, speculations, sentiments, evaluations, or beliefs, in natural language statements. Automatic procedures to detect subjective/ objective texts stem from the idea that humans express various degrees of subjectivity Saurí and Pustejovsky [2012], Saurí and Pustejovsky [2009] that are marked linguistically and can be identified with automatic procedures, Rubin et al. [2006], Rubin [2007].

Various research efforts addressed features of subjectivity in the field of natural processing: dimensions of subjectivity are discussed in Chen [2008] while Riloff and Wiebe [2003] proposes a method to learn patterns of subjective expressions; Yu and Hatzivassiloglou [2003] discusses the methods for separating facts from opinions, while sentiment analysis is undertaken in Taboada et al. [2011]. Several approaches investigate the way human various levels of certainty or uncertainty impact results of information retrieval Rubin [2010].

The techniques developed so far classify sentences as objective or subjective, by using either rule-based classifiers Wiebe and Riloff [2005], shallow lexical analysis, data-driven methods that rely on the availability of annotated datasetWiebe et al. [2005] or lexical resources such as SentiWordNet Esuli and Sebastiani [2006], a resource for opinion mining built on top of WordNet.

In many applications, subjectivity analysis is rather used as a filtering phase to increase the quality of data or to generate more viable data sets. Results provide a basis to summarize the initial collection Carenini et al. [2008] or to undertake a deeper semantic analysis Wiebe and Riloff [2005].

A complementary direction to investigate trust is the analysis of deception, notion that is commonly defined as a message knowingly transmitted with the intent to foster false beliefs or conclusions. The topic is emerging at the intersection of natural language processing and machine learning fields, with a solid background of studies on deception, from areas such as interpersonal psychology and communication Buller and Burgoon [1996], Hancock et al. [2007].

The main concern for addressing automation in deception identification is to avoid biases of human judgment, and to provide a more robust method to perform statement analyses. Most of the time, the automation of deception identification is based on the linguistic cues derived from the word classes from the Linguistic Inquiry and Word Count (LWIC) a gold standard in computerized text analysis explaining how the words used on a everyday basis are able to reveal thoughts, feelings, personality, and motivations Pennebaker et al. [2001]. The resource was extensively employed to study deception detection, Mihalcea and Strapparava [2009]. However, those efforts are subject to inherent limitations, as it is difficult to identify linguistic differences between truth-tellers and liars for certain types of deception strategies. Currently little is known about the adaptation of automated deception detection tools for statement analysis in various domains, although some authors argue that some tools for deception identification in textual data provide conflicting findings Hauch et al. [2012].

Most contributions on trust estimation adopts models keeping the distinction between source and information provided Paglieri et al. [2014]; some focus on building trust by using argumentation Tang et al. [2011], beliefs fusion or judgment aggregation Everaere et al. [2015] or investigate their joint integration Pigozzi [2015].

The main contribution of this paper is the development of a trust estimation model combining source's reliability and the self-confidence of reported statements and the analysis of different uncertainty types associated to the model and its formalization with belief functions.

III. Human sources and reported information

A. Assertions, opinions and reported information

Human sources are constantly used to supply observations, hypotheses, subjective beliefs and opinions about what they see or learn. From an analyst's point of view, decisions have to be made upon indirect reporting, thus the analysis of reported items is a critical step. This analysis is a multilevel process, relying heavily on analys's ability to understand the content of messages and assess their quality from additional clues provided. Practical cases described below highlight levels of indirection occurring when collecting information, with impact on trust construction.

Case 1 : Let X be an analyst collecting evidence in order to decide whether or not an individual is involved in terrorist activities. In particular, he takes into account reports submitted by Y, a human source. Those reports usually consist on a mixed set of assertions (e.g., descriptions of events or states observed by Y) and opinions (i.e., judgments, assessments, or inferences) expressed by Y about assertion which give the analyst an insight into how strongly Y commits to the assertion, see fig. 1.

In the statement contained in fig. 1, the source Y (Mary) lets us know that she does not commit her full belief to the assertion that John is a terrorist, otherwise the reporter would have used phrasing such as I am completely convinced or it is without doubt or simply reported John is a terrorist as an unadorned statement.

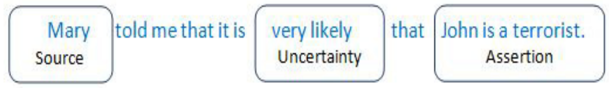

Fig. 1. Assertions and opinions in human messages.

In this case, the analyst knows that Y is not completely certain about her own statements so the analyst must make a judgment about the veracity of John being a terrorist based upon factors such as previous experience with Y's assessments in the past, or, perhaps, on the fact that other sources are relating the same information.

Case 2: Again, let X be an analyst collecting evidence in order to decide whether or not an individual is involved in terrorist activities. In this case, he takes into account reports submitted by Y, a human source, who is relating information obtained from a secondary source (Mary), see fig. 2.

Fig. 2. Hearsay, assertions and opinions in human messages.

The reporter does not just report on her direct observation or her deductions or beliefs, but also conveys information which the reporter has received from yet another source, e.g., Mary in the statement in fig. 2.

In this report, we have more levels of complexity to deal with. First, Y informs us that the assertion is Mary's opinion. The information that the assertion comes from Mary is a further distancing mechanism on the part of Y, as (unlike in Figure 2), he is neither claiming the opinion nor the assertion.

This case introduces yet more uncertainty. How sure can we be that the reporter has accurately repeated what Mary said? For example, did Mary really say it is likely or did the reporter insert (intentionally or unintentionally) this based upon the reporters assessment of the reliability of Mary as a source of information. Furthermore, perhaps Mary made this statement within a context which would strengthen or weaken this statement, but this context has not been passed on by the reporter.

The goal of the analyst is to take this primary information into account, but also to encode his own belief about the quality of the source further in the analysis. All these different attitudes have to be evaluated by the analyst, who has maybe additional background information or prior evaluation of the source that have to be considered.

In both cases discussed above, the outcome of the analyst is the primary information, augmented with a coefficient that help measure and track the different levels of trust of items, for their future exploitation. In the following, we call this quality the *trustworthy of the information* and highlight the main cases for trust construction with respect to the considered scenario.

The context, see Jenkins et al. [2015], describing the conditions and circumstances of information reporting is another feature to be considered when analyzing reported pieces. For instance, a witness may produce different information at different moments of time, i. e. in the aftermath of an attack or long after. Along the same lines, the information will differ on content and possible in quality if the source liberally decides to provide it on its own, or if it is interviewed. And finally, the dynamics of context make trust a temporal attribute, as trustworthy items can become untrustworthy in the light of a new context.

B. Concepts and notions for trust assessment

In the following we introduce several concepts that are relevant for trust analysis and often have different meanings in various research fields.

For the purpose of this paper, the notion of trustworthiness is considered as the quality of a source that is willing and able to report information in an accurate way with respect to the ground truth, and who can be relied upon to share sincerely and clearly his beliefs on the uncertainty level of reported information. The item provided by such a source is then trusted by analysts.

When reporting, the explicit and implicit uncertainty level assigned to reported assertions by the source is captured by self-confidence. Sources may include explicit indicators of uncertainty through the use of various lexical clues such as possibly, probably, might be, it is unlikely, undoubtedly, etc., all of which signal the writers confidence in the veracity of the information being conveyed.

Furthermore, the reporter may include formulations that indicate that the information in the sentence is not the result of direction observation by the reporter, and thus are implicit sources of uncertainty (lack of confidence) about the information being conveyed. For example, the reporter may be passing on information which originated with another source (hearsay); for example, according to, we were told, participants reported and similar constructions indicate that the information has originated from a source other than the reporter. Additionally, use of believe, assume, imply, suspect, or doubt all express opinions or inferences (dubbed by Bednarek in Bednarek [2006] as mindsay) by the reporter which creates some uncertainty about the credibility of the information.

In addition to self-confidence, how strong the analyst is willing to accept items from a source is another dimension to be considered when analysing reported items. This dimension is captured by the reliability of sources, generally understood as the degree to which prior historical reports have been consistent with fact. As an overall characterization, reliability is used for this work to rate sources that can be trusted with respect to their reputation, competence and supposed intentions.

The reputation of a source is an attribute built upon its own history of successes and failures in delivering accurate, good quality information, while the sources competence is related to the possession of the skills, knowledge and capabilities to report on various topics.

Intentions are specific to human sources as only humans have the capacity to deliberately provide false or misleading information. Sensors may provide erroneous data due to a number of factors such as device failure or environmental conditions, but never due to intention. Humans, however, may relay information which is partially or completely false because of internal factors such as ideology or maliciousness.

In the light of those notions, integrating reported information into further processing requires mechanisms that are able to analyze the quality of the item (see Rogova and Bossé Rogova and Bosse [2010]) and flexible enough to compensate for insufficient quality on one or several dimensions.

IV. A FUNCTIONAL MODEL OF TRUST

This section introduces the model developed to construct trust in reported information by taking into account the reliability of the source and its own characterization of reported items. The advantage of this distinction is to better dissociate the impact of both sources' beliefs and analyst's opinions on the source on the information provided.

The description of various attributes of sources and information we refer to when developing the model of trust is related to notions introduced by the STANAG 2511, NATO [2003] but some of those notions are addressed differently: reliability of sources is understood in terms of competence, reputation and intentions, while credibility is restrained to features of self-confidence.

Even if the primary function of a source is to provide information, we keep the distinction between the source and the information by considering separate dimensions for each element and combining them in order to have an overall characterization of the outcome in the form of trust. The rationale behind this is the observation that even reliable sources can provide inaccurate or imprecise information from one report to another, which is even more plausible in the case of human sources.

By focusing on the global characterization of reported items, the model aims at providing a better understanding of how trust is to be constructed from various dimensions during a process having humans as a centric element in both the production and the analysis of information.

The model consists of several elements, related by functions, as described in fig. 3.

In the following, we present the model by adopting a granularity that is detailed enough to describe different elements of the model, but still rough enough to avoid the adoption of a specific formalism for their representation.

A. Elements of the trust model

The model is composed of two elements: an information source and reported items. The analyst is considered outside the model, although it has multiple interactions with its elements.

Definition of information source: an information source is an agent providing an information item while being able

Fig. 3. Model for trust analysis.

to characterize its level of uncertainty. Source is a relative notion, depending on the perspective of analysis. As a general setting, information is propagated within a chain relating the real world to some decision maker, and agents along the path can be both trained observers, supposed to provide reports, but also witnesses or lay observers, adding items without being primarily considered as information sources, but opportunistic ones.

The notion of source is central in many information fusion applications and several research efforts aimed at modeling their properties. A general analysis of sources is undertaken by Hall and Jordan Hall and Jordan [2010], who identify three main classes: S-Space, composed of physical sensors, H-Space for human observers and I-Space for open and archived data on the Internet. In Jousselme et al. [2014], a unified characterization of hard and soft sources is described, along with a detailed description of their qualities and processing capabilities.

Information reported by humans is unstructured, vague, ambiguous and subjective and is often contrasted with information coming from physical sensors, described as structured, quantitative and objective. While humans can deliberately change the information or even lie, sensors are prone to errors and therefore hard items are not always accurate.

Processing hard sensor information is widely covered Khaleghi et al. [2013] in the research community, while the integration of human sources comes with new challenges. The model addresses human sources, and reported items can refer to actions, events, persons or locations of interest.

For human agents, the source is part of the real world, (a community, a scene, an event) and can be either directly involved in the events reported, or just uttering as a witness.

Definition of reported information: Reported information is a couple $(I, \chi(I))$, where I is an item of information and $\chi(I)$ the confidence level as assigned by the source. Each item I has assertive i_a and subjective i_s components conveying factual and subjective contents respectively. Extracted from direct speech, reported information is used by analysts to represent assertions reported by other persons.

The analysis of reported information is a topic to be addressed as the fusion of information from soft sources receives increasing attentions in recent years. If some authors developed logic-based approached to model distortions of items exchanged between agents having intentions and the ability to deceive Cholvy [2011], there are more challenges arising when the information is analyzed in its textual form.

Features of uncertainty, as expressed in natural language statements are analyzed in Auger and Roy [2008] while Dragos and Rein [2014] provides a broader discussion of pitfalls and challenges related to soft data integration.

B. *Functions of the trust model*

The model introduces several functions estimating features of reliability, self-confidence and trust, as described hereafter.

Definition of a reliability function: a reliability function is a mapping associating a real value to an information source.

The real value is a quantitative characterization of the source, inferred with respect to its previous failures or its reputation and the relevance of its competence and skills for specific domains.

For this model, reliability of human sources combines features of competence, reputation, and intentions.

Competence defines to what extent a human source can understand the events it reports on, has the ability to accurately describe them and is able to follow the logic of processes producing the information. Thus, the quality of information reported by a source depends on the level of training and expertise, and can be high for a task poor for others, out of source's depth.

Reputation is the overall quality of a source, as estimated from the history of its previous failures and intentions refer to attitudes or purposes, often defined with respect to a hidden purpose or plan to achieve.

Reliability is a complex concept and, from a practical standpoint it is difficult to have information about the global reliability of a source. Thus, the model developed describes reliability along three attributes: the competence of a source, its reputation and intentions and seeks to estimate each attribute. This solution allows us to compensate for insufficient information on one or several aspects of reliability and to conduct, in some cases, the analysis of source's reliability based on a single attribute.

Evaluation of reliability Assessing reliability is of real interest when opportunistic sources are considered and analyst has no clue about the way the source can behave and lacks the ability to monitor or control the human providing information, or the environment in which the source operates

Various methods can be developed to estimate competence, reputation and intentions of the source. Thus, competence is closely related to the level of training of an observer or can be defined by domain knowledge. Values can be expressed either in a linguistic form (*bad, good, fair, unknown*) or by a number.

Reputation is an attribute to be constructed upon previous failures of the source but also by considering its level of conflict with other sources, and can be expressed by numeric or symbolic values.

While reputation and competence can be, at least in some cases, estimated from prior knowledge, characterizing intentions of a source is subject to human perception and analysis. Judgment of human experts is needed as usually there is no *a priori* characterization of the source with respect to its intentions and also because it is important to assess those aspects from the subjective point of view of an expert in the form of binary values only.

From a practical standpoint, it is suitable to provide to an expert the description of competence reputation and intentions as assessed independently. This way, experts can have the opportunity to develop different strategies of using reliability: they can decide to assign different importance to those attributes under different contexts or can use their own hierarchy of attributes. For instance, an expert may consider irrelevant the information provides by a source whose competences is lower than a thresholds or if he suspects the source of having malicious intentions.

Definition of a self-confidence function: a self-confidence function is a mapping linking a real value and an information item. The real value is a measure of the information credibility as evaluated by the sensor itself and is of particular interest for human sources, as often such sources provide appreciations of the information conveyed.

For this work, we adopt the definition of self-confidence as introduced by the homonym concept in the URREF (Uncertainty Representation and Reasoning Evaluation Framework) ontology developed by the Evaluation of Technologies for Uncertainty Representation Working Group (ETURWG), see Costa et al. [2012].

Identifying features of self-confidence requires methods related to a research topic in the field of natural language processing: the identification of assertions and opinions in texts. In this field, the commonly adopted separation of those notions considers assertions as statements that can be proven or not, while opinions are hypotheses, assumptions and theories based on someone's thought and feelings and cannot be proven.

Evaluation of self-confidence: Estimation of self-confidence aims at assessing a numerical value, able to capture the degree how strong the author stands behind assertions in the statement, on the basis of lexical clues he included when uttering. More generally, markers of authors commitment are in the form of hedges, modal verbs and forms of passive/ active language. A hedge is a mitigating word that modifies the commitment to the truth of propositions, i.e., certainly, possibly. Their impact can be increased by a booster (highly likely) or decreased by a downtowner (rather certain).

Modal verbs indicate if something is plausible, possible, or certain (*John could be a terrorist, you might be wrong)*. Moreover, in some domains sentences making use of the passive voice are considered as an indicator of uncertainty, in that sense that author puts emphasis on assertions in the statement and seeks to distance himself from the items reported. In contrast, less uncertainty is related to active voice sentence that clearly identifies the action and who is performing that action.

Quantifying self-confidence is a topic of particular interest for intelligence analysis, and it was early addressed by Kent in 1962, Kent [1964] who created a standardized list of words of estimative probability. This list was intended to become a common basis to be used by analysts to produce uncertainty assessments.

Kesselman describes in Kesselman [2008] a study conducted to analyze the way the list was used by analysts over the past, identifies new trends to convey estimations and proposes a new list having the verb as a central element addressed.

Given the variety of linguistic markers for uncertainty, the estimation of a numerical value based on their combination seems unrealistic, as often the same sentence contains not just one but multiple expressions of uncertainty. Besides, assigning numerical values to lexical expressions is not an intuitive task, and Rein shows that there are no universal values to be associated in a unique manner to hedges or other uncertainty markers, see Rein [2016]. As the author argues further, it is rather possible to order those expressions and use this universal ordering as a more robust way to compare combinations of uncertainty expression, and thus highlight different levels of uncertainty in natural language statements.

Using the model for trust analysis: The model proposed in this work allows assessing levels of trust in reported information by combining various attributes of the source, introduced under the hat of reliability, and self-confidence, capturing the credibility of information as stated by the human.

The model is source-centric and considers aspects of reliability as main factors having the ability to correct, alter, or qualify the information reported by the source. If several rules to rank, prioritize, or combine attributes introduced by the model can be drafted empirically, the estimation of a trust value requires a formal representation of the model.

A possible solution to estimate a unified value for trust is to consider reliability and self-confidence within the framework of an uncertainty theory and to rely on the set of combination rules the theory defines, for example those developed in probability theory, in possibility theory, or in belief functions theory. All these theories provide various operators to combine reliability and self-confidence in order to estimate trust.In the following we describe the model by using belief functions and illustrate trust estimation based on them for various practical scenarios. In this work, we consider only belief functions because they offer a very general theoretical framework to deal with various aspects of uncertainty. In fact, probabilities, possibilities are very special cases of belief functions and Subjective Logic sometimes used in trust analysis by some authors is also based on a simplified form of belief functions.

V. Trust formalization with belief functions

A. *Basic Belief Assignment*

Belief Functions (BF) have been introduced by Shafer in his his mathematical theory of evidence Shafer et al. [1976], also called Dempster-Shafer Theory (DST), to model epistemic uncertainty. The frame of discernment (FoD) of the decision problem under consideration, denoted Θ, is a finite set of exhaustive and mutually exclusive elements. The powerset of Θ denoted 2^Θ is the set of all subsets of Θ, empty set included. A body of evidence is a source of information characterized by Basic Belief Assignment (BBA), or a mass function, which is a mapping $m(.) : 2^\Theta \rightarrow [0,1]$ that satisfies $m(\emptyset) = 0$, and the normalization condition $\sum_{A\in 2^\Theta} m(A) = 1$. The belief (a.k.a credibility) Bel(.) and plausibility Pl(.) functions usually interpreted as lower and upper bounds of unknown (subjective) probability measure $P(.)$ are defined from $m(.)$ respectively by

$$\text{Bel}(A) = \sum_{B\subseteq A | B\in 2^\Theta} m(B) \tag{1}$$

$$\text{Pl}(A) = \sum_{B\cap A\neq\emptyset | B\in 2^\Theta} m(B) \tag{2}$$

An element $A \in 2^\Theta$ is called a focal element of the BBA $m(.)$, if and only if $m(A) > 0$. The set of all focal elements of $m(.)$ is called the core of $m(.)$ and is denoted $\mathcal{K}(m)$. This formalism allows to model a full ignorant source by taking $m(\Theta) = 1$. The Belief Interval (BI) of any element A of 2^Θ is defined by

$$BI(A) \triangleq [Bel(A), Pl(A)] \tag{3}$$

The width of belief interval of A, denoted $U(A) = Pl(A) - Bel(A)$ characterizes the degree of imprecision of the unknown probability $P(A)$, often called the uncertainty of A. We define the uncertainty (or imprecision) index by

$$U(m) \triangleq \sum_{A\in\Theta} U(A) \tag{4}$$

to characterize the overall imprecision of the subjective (unknown) probabilities committed to elements of the FoD bounded by the belief intervals computed with the BBA $m(.)$.

Shafer did propose Dempster's rule of combination to combine multiple independent sources of evidence Shafer et al. [1976], which is the normalized conjunctive fusion rule. This rule has been strongly disputed in BF community since Zadeh's first criticism in 1979, and since the 1990's many rules have been proposed to combine (more or less efficiently) BBAs, see discussions in Smarandache and Dezert [2015], in particular the proportional conflict redistribution rule no 6 (PCR6).

Instead of working with quantitative (numerical) BBA, it is also possible to work with qualitative BBA expressed by labels with the linear algebra of refined labels proposed in Dezert-Smarandache Theory (DSmT), Smarandache and Dezert [2015] (Vol. 2 & 3).

B. Trust formalization model

Because belief (also known as credibility) are well defined mathematical concepts in the theory of belief functions, we prefer to use self-confidence terminology to represent the confidence declared by a source Y on its own assertion A (considered as a piece of information for the analyst X). Let's denote by A the assertion given by the source, for instance $A =$ *John is a terrorist*. The valuation $v(A)$ made my the source Y about the assertion A can be done either quantitatively (by a probability or a BBA) or qualitatively (by a label associated to a linguistic form). Let's concentrate on the quantitative representation of $v(A)$ for simplicity[1].

The basic piece of information provided by a source Y consists of A (the assertion), and $v(A)$ (its valuation). To be as general as possible, we suppose that $v(A)$ is a basic belief mass assignment defined with respect to the very basic frame of discernment $\Theta_A \triangleq \{A, \bar{A}\}$ where $\bar{A}$ denotes the complement of A in Θ_A, that is $v(A) = (m(A), m(\bar{A}), m(A\cup\bar{A}))$. Note that only two values of the triplet are really necessary to define $v(A)$ because the third one is automatically derived from the normalization condition $m(A) + m(\bar{A}) + m(A\cup\bar{A}) = 1$. So one could also have chosen equivalently $v(A) = [Bel(A), Pl(A)]$ instead the BBA. In a probabilistic context, one will take $m(A\cup\bar{A}) = 0$ and so $v(A) = P(A)$ because $Bel(A) = Pl(A) = P(A)$ in such case.

The self-confidence of the source Y is an extra factor $\alpha_Y \in [0,1]$ which characterizes the self-estimation of the quality of the piece of information $(A, v(A))$ provided by the source itself. $\alpha_Y = 1$ means that the source Y is 100% confident in his valuation $v(A)$ about assertion A, and $\alpha_Y = 0$ means that the source Y is not confident at all in his valuation $v(A)$. In the theory of belief functions, this factor is often referred as the discounting factor of the source because this factor is usually used to discount the original piece of information $(A, v(A))$ into a discounted one $(A, v'(A))$ as follows Shafer et al. [1976]:

$$m'(A) = \alpha_Y \cdot m(A) \tag{5}$$

$$m'(\bar{A}) = \alpha_Y \cdot m(\bar{A}) \tag{6}$$

$$m'(A\cup\bar{A}) = \alpha_Y \cdot m(A\cup\bar{A}) + (1-\alpha_Y) \tag{7}$$

The idea of this Shafer's discounting technique is to diminish the belief mass of all focal elements with the factor α_Y and redistribute the missing discounted mass $(1-\alpha_Y)$ to the whole ignorance $A\cup\bar{A}$. Note that the valuation of discounted piece of information is always degraded because its uncertainty index is always greater than the original one, that is $U(m') > U(m)$, which is normal.

The reliability factor r estimated by the analyst X on the piece of information $(A, v(A))$ provided by the source Y must take into account both the competence C_Y, the reputation R_Y and the intention I_Y of the source Y. A simple model to establish the reliability factor r is to consider that C_Y, R_Y and I_Y factors are represented by numbers $[0,1]$ associated to chosen subjective probabilities, that is $C_Y = P(Y \text{ is competent})$, $R_Y = P(Y \text{ has a good reputation})$ and $R_Y = P(Y \text{ has a good intention (i.e. is fair)})$. If each of this factor has equal weight, then one could use $r = C_Y \times R_Y \times I_Y$ as simple product of probabilities. However in practice, such simple modeling does not fit well with what the analyst really needs for taking into account epistemic uncertainties in

[1] In fact and without loss of generality one can always map a qualitative representation to a quantitative one by a proper choice of scaling and normalization (if necessary).

Competence, Reputation and Intention. In fact each of these factors can be viewed as a specific criterion influencing the level of the global reliability factor r. This is a multi-criteria valuation problem. Here we propose a method to solve it.

We consider the three criteria C_Y, R_Y and I_Y with some associated importance weights w_C, w_R, w_I in $[0,1]$ with $w_C + w_R + w_I = 1$. We consider the frame of discernment $\Theta_r = \{r, \bar{r}\}$ about the reliability of the source Y, where r means that the source Y is reliable, and $\bar{r}$ means that the source Y is definitely not reliable. Each criteria provides a valuation on r expressed by a corresponding BBA. Hence, for the competence criteria C_Y, one has $(m_C(r), m_C(\bar{r}), m_C(r \cup \bar{r}))$, for the reputation criteria R_Y one has $(m_R(r), m_R(\bar{r}), m_R(r \cup \bar{r}))$ and for the intention criteria I_Y, one has $(m_I(r), m_I(\bar{r}), m_I(r \cup \bar{r}))$.

To get the final valuation of reliability r of the source Y one needs to fuse efficiently the three BBAs $m_C(.)$, $m_R(.)$ and $m_I(.)$ taking into account their importance weights w_C, w_R, and w_I. This fusion problem can be solved in applying the importance discounting approach combined with PCR6 fusion rule of DSmT Smarandache et al. [2010] to finally get the valuation $v(r) = (m_{PCR6}(r), m_{PCR6}(\bar{r}), m_{PCR6}(r \cup \bar{r}))$ from which either the decision (r, or $\bar{r}$) can be drawn (using BI distance for instance). If a firm decision is not required, an approximate probability $P(r)$ can also be inferred with some lossy transformations of BBA to probability measure Smarandache and Dezert [2015]. Note that Dempster's rule of combination cannot be used here because it is not responding to the importance discounting, as explained in Smarandache et al. [2010].

The trust model consists in using both the piece of information $(A, v(A))$ and self-confidence factor α_Y provided by the source Y, and the reliability valuation $v(r)$ expressed by the BBA $(m(r), m(\bar{r}), m(r \cup \bar{r}))$ to infer the trust valuation about the assertion A. For this, one proposes to use the mass $m(r)$ of reliability hypothesis r of the source Y as a new discounting factor of the BBA $m'(.)$ reported by the source Y taking into account its self-confidence α_Y. Hence the trust valuation $v_t(A) = (m_t(A), m_t(\bar{A}), m_t(A \cup \bar{A}))$ of assertion A for the analyst X is defined by

$$m_t(A) = m(r) \cdot m'(A) \tag{8}$$

$$m_t(\bar{A}) = m(r) \cdot m'(\bar{A}) \tag{9}$$

$$m_t(A \cup \bar{A}) = m(r) \cdot m'(A \cup \bar{A}) + (1 - m(r)) \tag{10}$$

or equivalently by

$$m_t(A) = m(r)\alpha_Y \cdot m(A) \tag{11}$$

$$m_t(\bar{A}) = m(r)\alpha_Y \cdot m(\bar{A}) \tag{12}$$

$$m_t(A \cup \bar{A}) = m(r)\alpha_Y \cdot m(A \cup \bar{A}) + (1 - m(r)\alpha_Y) \tag{13}$$

Of course, some strategies based on the level of $m(r)$ can be developed in order to not apply automatically this discounting technique depending on the problem or assertion under concern.

VI. Uncertainty analysis for trust estimation under URREF criteria

Tracking uncertainties from problem description to model construction and formalization is done under the URREF criteria. The goal is to make explicit the uncertainty arising when the problem is abstracted by the model and the model is then simplified in order to fulfill constraints of specific formalism, fig. 6.

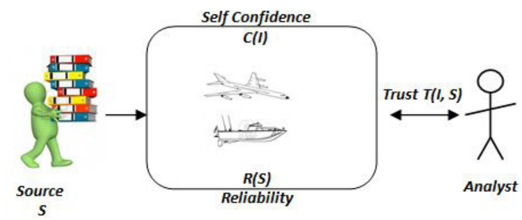

Fig. 4. Trust estimation from source to analyst

Besides listing uncertainty criteria relevant for trust estimation, the analysis also discusses the mapping of URREF criteria to attributes of the model and sheds a light on overlapping and imperfects matchings. This mapping also offers a basis to identify the limitations of the URREF ontology, by stressing out elements whose characterization in terms of uncertainty is out of its reach or its intended scope.

A. Uncertainties form problem definition to model construction

Let M be the model for trust estimation, with elements introduced in paragraph IV: the source Y, the reported item I with its assertive i_a and subjective i_s parts and $\chi(I)$ the confidence level assigned by the source Y to I.

Static elements of the model are impacted by forms of uncertainty arising when items are being produced by the source, gathered and analyzed by the analyst before entering the fusion system. From an information fusion process standpoint, the input elements of the model are the source, a human reporting, and the item provided and they can be evaluated with the following URREF criteria *Objectivity, ObservationalSensitivity and SelfConfidence* and all subclasses of the more general concept *InputCriteria*.

Objectivity is an attribute of the source, related to its ability to provide factual, unbiased items, without adding their own points of view or opinions. For a source Y providing information item i, having i_s and i_a as subjective and factual parts respectively, objectivity can be expressed as:

$$Objectivity(Y, I) = \psi_o(i_s, i_a) \tag{14}$$

where $\psi_o(i_s, i_a)$ represents the mathematical quantified expression of the subjective over the factual content content of i.

ObservationalSensitivity is a source's criterion capturing its ability to provide accurate reports. For the considered model, the criterion is an aggregation of competences C and reputation R, two attributes of the model.

$$ObservationalSensitivity(Y, i) = \psi_{os}(C, R) \quad (15)$$

where $\psi_{os}(C, R)$ is a function aggregating values of competence and reputation.

Items entering the system are described by *SelfConfidence* and *RelevanceToProblem*.

SelfConfidence is the expression how strong the source supports his own assertions. Again, considering i_s and i_a as subjective and factual items conveyed by I, the criterion can be expressed as:

$$SelfConfidence(I) = \psi_{sc}(i_s) \quad (16)$$

with $\psi_{sc}(i_s)$ a function quantifying the subjective content of item I.

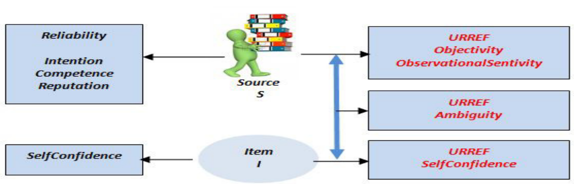

Fig. 5. Mapping of model attributes to URREF criteria

Fig. 6 shows the mapping between teh set of source's attributes and information items, as identified during the modeling stage, and the set of relevant URRER uncertainty criteria. The mapping shows a perfect match between SelfConfidence as introduced by the model and the homonym URREF criterion and several imperfect matches described hereafter.

At source level, the URREF criteria are not able to capture in a distinct manner features of competence, reputation and intentions, the main attributes of the sources added by the model under Reliability. To some extent, Competence and Reputation can be related to *ObservationlSentitivity*, but intentions clearly remains out of reach for URREF criteria, and having a too general description of sources is clearly a main limitation of the URREF ontology.

At information level, using the URREF criteria provides an additional attribute for items characterization as *RelevanceToProblem* aims at analyzing the item in the light of the specific problem to be solved. Obviously, a description of the problem is needed to establish correlations.

B. Uncertainties from model construction to formal representation

The model for trust estimation was built by adopting a granularity that is detailed enough to identify its elements, but still rough enough to avoid the adoption of a specific formalism to represent those elements and to implement an estimation of trust accordingly.

The goal of having formal models as approximations for reality is for the purpose of making inferences. The formalism provides equations for relating the information to parameters and thus estimates the trust according to the model. The inference algorithm is tightly coupled with how uncertainty is represented within the formalism.

However any representation formalisms comes with additional constraints and approximations. Let F be the DST formalization of the trust estimation model, with parameters introduced in paragraph V. The formalism induces two types of uncertainty related to its capacity to handle incomplete, ambiguous or contradictory evidence and knowledge.

The uncertainty of evidence handling is captured by *Ambiguity*, *Dissonance*, *Conclusiveness* and *Completeness*.

Ambiguity measures to which extent the formalism can handle data sets which support different conclusions.

$$Ambiguity(F) = \phi_a(\alpha_Y, R_Y) \quad (17)$$

where the function $\phi_a(\alpha_Y, R_Y)$ considers the self-confidence factor α_Y provided by the source Y and the reliability of Y provided by the analyst R_Y to estimate the degree of ambiguity. The measures is of interest when items having high values of selfconfidence are provided by unreliable sources.

Dissonance captures the ability of the formalism to represent inconsistent evidence. For BBA representations, dissonance can be related to the capacity of the formalism to assign belief mass to an element and its negation and can therefore be assessed for every BBA representation build for the model. For example, the dissonances for sources competence can be in the form:

$$Dissonance(F) = \phi_d(m_C(r), m_C(\bar{r})) \quad (18)$$

where $\phi_d(m_C(r), m_C(\bar{r}))$ is a function combining the belief mass assigned to the fact that the source is considered as competent or incompetent, respectively.

Dissonance is of interest to highlight practical situations when there are significant differences of belief masses assigned at attribute level, as it is the case when a source is considered incompetent (low $m_C(r)$, high$m_C(\bar{r})$) but has a good reputation (high $m_R(r)$, low $m_R(\bar{r})$).

Conclusiveness is a measure expressing how strong the evidence supports a conclusion or unique hypothesis.

$$Conclusiveness(F) = \phi_{cc}(m_t(A), m_t(\bar{A}), m_t(A \cup \bar{A})) \quad (19)$$

where $\phi_{cc}(m_t(A), m_t(\bar{A}), m_t(A \cup \bar{A}))$ is a function combining the belief masses estimated for trustful, untrustful or unknown qualifications of assertion A respectively.

The measure indicates to which extent the result of inferences can support a conclusion, in this case whether the hypothesis that the assertion under analysis is trustful or not. It can be used during the inference process, to show how taking into account additional elements such as the competence of the source, its reputation or intentions impact the partial estimations of trust.

Completeness is a measures of the range of the available evidence, and captures the ability of formalism to take into account how much is unknown. The measures is somehow similar to *Dissonance*, as is can be assessed for every BBA representation build for the model. Hereafter completeness of source's reliability is described as:

$$Completeness(F) = \phi_{cp}(m_(r \cup \bar{r})) \tag{20}$$

where $\phi_{cp}(m_(r \cup \bar{r}))$ is a function depending of the belief mass assigned to unknown.

The measure is useful to be estimated and analyzed before entering the fusion process, in order to have a picture of how complete is the evidence describing the various elements of the model, and to avoid performing fusion on highly incomplete data sets.

The main form of epistemic uncertainty in the formalization F can be evaluated using *Adaptability*, a subcass of *Expressiveness* and *KoowlegdeHandling*. *Adaptability* encompasses the ability of the formal representation to allow different configurations of the model and is achieved thanks to importance coefficients. Coefficients are introduced in order to cover several practical situations, such as analyst ignoring the reputation of a source or its intentions. The measure is defined at global level for the formal representation F:

$$Adaptability(F) = \phi_{ad}(N_{as}, N_p) \tag{21}$$

where $\phi_{ad}(N_{as}, N_p)$ is a function depending of both N_a the number of attributes in the formal representation, and N_p the number of parameters allowing building different configurations.

Adapatbility can be used as a criterion to choose the most suitable formal representation when several are available to represent the model.

Both *EvidenceHandling* and *KnowledgeHandling* are subclasses of *RepresentationCriteria*.

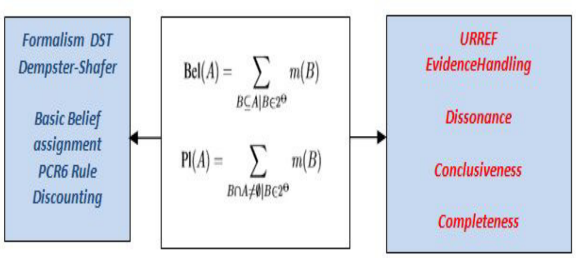

Fig. 6. Mapping of formalism uncertainties to URREF criteria

This section analyses the nature of uncertainties arising when going from defining the problem of trust estimation for reported items definition to model construction and further to model formalization with belief functions. The newt section show how uncertainties can be highlighted for particular scenarios of trust analysis.

VII. Use cases of uncertainty analysis for trust estimation

A. Running example and method for uncertainty tracking

As a running example let's consider an assertion A and its valuation $v(A)$ provided by the source Y as follows: $m(A) = 0.7$, $m(\bar{A}) = 0.1$ and $m(A \cup \bar{A}) = 0.2$. Its self-confidence factor is $\alpha_Y = 0.75$. Hence the discounted BBA $m'(.)$ is given by

$$m'(A) = 0.75 \cdot 0.7 = 0.525$$
$$m'(\bar{A}) = 0.75 \cdot 0.1 = 0.075$$
$$m'(A \cup \bar{A}) = 1 - m'(A) - m'(\bar{A}) = 0.4$$

Let's assume that the BBAs about the reliability of the source based on Competence, Reputation and Intention criteria are given as follows:

$$m_C(r) = 0.8, m_C(\bar{r}) = 0.1, m_C(r \cup \bar{r}) = 0.1$$
$$m_R(r) = 0.7, m_R(\bar{r}) = 0.1, m_R(r \cup \bar{r}) = 0.2$$
$$m_I(r) = 0.6, m_I(\bar{r}) = 0.3, m_I(r \cup \bar{r}) = 0.1$$

with importance weights $w_I = 0.6$, $w_R = 0.2$ and $w_C = 0.2$.

After applying the importance discounting technique presented in Smarandache et al. [2010] which consists to discount the BBAs with the importance factor and redistribute the missing mass onto the empty set, and combining the discounted BBAs with PCR6 fusion rule, we finally get after normalization the following BBA

$$m(r) = 0.9335$$
$$m(\bar{r}) = 0.0415$$
$$m(r \cup \bar{r}) = 1 - m(r) - m(\bar{r}) = 0.025$$

The final trust valuation of assertion A reported by the source Y taking into account its self-confidence $\alpha_Y = 0.75$ and the reliability factor $m(r) =$ is therefore given by Eqs. (11)–(13) and one obtains

$$m_t(A) = 0.4901$$
$$m_t(\bar{A}) = 0.0700$$
$$m_t(A \cup \bar{A}) = 1 - m_t(A) - m_t(\bar{A}) = 0.4399$$

Note that if we $m_C(r) = m_R(r) = m_I(r) = 1$, then we will always get $m(r) = 1$ whatever is the choice of weightings factors, which is normal. If one has a total conflict between valuations of reliability based on Competence, Reputation and Intention criteria then Dempster's rule cannot be applied to get global reliability factor $m(r)$ because of 0/0 indeterminacy in formula of Dempster's rule. For instance, if one has $m_C(r) = m_R(r) = 1$ and $m_I(\bar{r}) = 1$, then $m(r)$ is indeterminate with Dempster's rule of combination, whereas it corresponds to the average value $m(r) = 2/3$ using PCR6 fusion rule (assuming equal importance weights $w_C = w_R = w_I = 1/3$), which makes more sense.

In the next subsections we explore several typical scenarios for trust assessment corresponding to different situations of BBAs distributions and track the uncertainty according to

URREF criteria. Each scenario illustrates specific instances of the model developed for trust estimation.

The method adopted to track uncertainty defines the following measures to estimate URREF criteria:

$$SelfConfidence = \alpha_Y$$
$$Ambiguity = |\alpha_Y - m(r)|$$
$$Objectivity = m_I(r)$$
$$ObservationalSensitivity = min(m_C(r), m_R(r))$$

As shown in previous formulas, UREF criteria are estimated based on features of the BBA formalization and are assigned to static elements of the model, respectively the source and information item. While o$Objectivity$ and $ObservationalSensitivity$ captures imperfections of observations, $SelfConfidence$ and $Ambiguity$ show inaccuracies of reporting information to analysts.

Those criteria are assessed before entering the fusion phase, and describe the initial uncertainty present in the system before inferences.

In addition, $Dissonance$, $Conclusiveness$ and $Completeness$ will be estimated at scenario level by adopting the following formulas:

$$Dissonance = 1 - |m_t(A) - m_t(\bar{A})|$$
$$Conclusiveness = |m_t(A) - m_t(\bar{A})|$$
$$Completeness = 1 - m(A \cup \bar{A})$$

Moreover, at model level, $Adaptability$ will not be assessed as it is based on the number and values of adjustment coefficients which are constant variables for all scenarios.

Those criteria will be assessed for elements inferred thanks to fusion: the reliability of the source, the updated BBAs of the initial assertion and estimated trust. Having values of URREF criteria provides a richer description of results and helps their interpretation.

Hereafter scenarios for trust estimation and their associated uncertainty are illustrated.

B. Scenarios for trust assessment and uncertainty analysis

Scenarios introduced bellows provide examples of trust construction using various operators and highlights uncertainty assigned to elements of the model and its propagation during the fusion process.

Scenario 1 - Consensus: Suppose that Y provides the assertion A, while stating that A certainly holds and that X considers Y as a reliable source.

In this case, the trust will be constructed on the basis of two consensual opinions: the analyst X that considers Y as a reliable source, and the source's conviction that the information provided is certain. In this case, $m(A) = 1$, $\alpha_Y = 1$ and $m(r) = 1$, so that $m'(A) = 1$ and $m_t(A) = m(r) \cdot m'(A) = 1$. The result will be in the form $(A, v(A))$ initially provided by the source.

URREF criteria	Objectivity	ObservationalSensitivity
Values	1	1

TABLE I
CONSENSUS: OBSERVATION UNCERTAINTY

URREF criteria	SelfConfidence	Ambiguity
Values	1	0

TABLE II
CONSENSUS: REPORTING UNCERTAINTY

	Dissonance	Conclusiveness	Completeness
Updated BBAs	0	1	1
Reliability	0	1	1
Trust	0	1	1

TABLE III
CONSENSUS: DYNAMIC UNCERTAINTY ASSERTION

This scenario illustrates an ideal situation for trust assessment, where the source is trustful, well known by the analyst and observations are carried out and reported in perfect conditions. As shown in tab. I and II, there is no uncertainty induced by the source, and once fusion is performed the items impacted show high values for conclusiveness and completeness, while dissonance is 0 for the updates BBAs values, source's reliability and estimated trust, as shown in tab. III.

Scenario 2 - Uncertain uttering: Y is considered as a reliable source by X and reports the assertion A, while showing a low level of certainty $v(A)$ about the veracity of A. This example is relevant for reliable sources providing (possibly) inaccurate descriptions of events due to, say, bad conditions for observation. This scenario corresponds by example to the following case for inputs: $\alpha_Y = 0.6$

$$m(A) = 0.8, m(\bar{A}) = 0.1, m(A \cup \bar{A}) = 0.1$$

$$m_C(r) = 0.9, m_C(\bar{r}) = 0, m_C(r \cup \bar{r}) = 0.1$$

$$m_R(r) = 0.9, m_R(\bar{r}) = 0, m_R(r \cup \bar{r}) = 0.1$$

$$m_I(r) = 0.3, m_I(\bar{r}) = 0.3, m_I(r \cup \bar{r}) = 0.6$$

and $w_C = 0.5$, $w_R = 0.5$ and $w_I = 0$.
Hence, one gets

$$m'(A) = 0.48, m'(\bar{A}) = 0.06, m'(A \cup \bar{A}) = 0.46$$

and

$$m(r) = 0.9846, m(\bar{r}) = 0, m(r \cup \bar{r}) = 0.0154$$

Therefore, one finally obtains as trust valuation

$$m_t(A) = 0.4726, m_t(\bar{A}) = 0.0591, m_t(A \cup \bar{A}) = 0.4683$$

This case shows that self-confidence has an important impact on values of the discounted BBA, as $m'(A)$ is decreased from 0.8 to 0.48 and thus the remaining mass is redistributed on $m'(A \cup \bar{A})$.

The combination of competence, reliability and intention are in line with the assumption of the scenario, which states that Y is a reliable source. After normalization, values for trust assessment clearly highlight the impact of uncertain uttering, as the BBA shows a mass transfer from $m_t(A)$ to $m_t(A \cup \bar{A})$. Still, values of trust are close to BBA integrating the self-confidence, which confirms the intuition that when the analyst X considers Y as a reliable source, the assertion A is accepted

URREF criteria	Objectivity	ObservationalSensitivity
Values	0.3	0.9

TABLE IV
UNCERTAIN UTTERING: OBSERVATION UNCERTAINTY

URREF criteria	SelfConfidence	Ambiguity
Values	0.6	0.38

TABLE V
UNCERTAIN UTTERING: REPORTING UNCERTAINTY

	Dissonance	Conclusiveness	Completeness
Updates BBAs	0.3	0.7	0.9
Reliability	0.02	0.98	0.985
Trust	0.59	0.41	0.54

TABLE VI
UNCERTAIN UTTERING: FUSION UNCERTAINTY

with an overall trust level almost equal to the the certainty level stated by the source.

This scenario illustrates uncertainty induced by observations failures, as $Objectivity$, ans $SelfConficende$ are low, see IV, V.

While the quality of the source is highlighted by high values of $Conclusiveness$ and $Completeness$, showing the analyst's confidence in reports analyzed, the impact of imperfect observation is shown on the overall estimation of trust, with a combination of $Dissonance$, $Conclusiveness$ and $Completeness$ having values close to 0.5, VI.

Scenario 3 - Reputation: Suppose that Y provides A and $v(A)$ and X has no global description of Y in terms of reliability. As reliability of Y is not available, the reputation will be used instead, as derived from historical data and previous failures. This scenario corresponds by example to the following case for inputs: $\alpha_Y = 1$

$$m(A) = 0.8, m(\bar{A}) = 0.1, m(A \cup \bar{A}) = 0.1$$

$$m_C(r) = 0.1, m_C(\bar{r}) = 0.1, m_C(r \cup \bar{r}) = 0.8$$

$$m_R(r) = 0.9, m_R(\bar{r}) = 0.1, m_R(r \cup \bar{r}) = 0$$

$$m_I(r) = 0.1, m_I(\bar{r}) = 0.1, m_I(r \cup \bar{r}) = 0.8$$

and $w_C = 0.1$, $w_R = 0.8$ and $w_I = 0.1$.
Hence, one gets

$$m'(A) = 0.8, m'(\bar{A}) = 0.1, m'(A \cup \bar{A}) = 0.1$$

and

$$m(r) = 0.9449, m(\bar{r}) = 0.0196, m(r \cup \bar{r}) = 0.0355$$

Therefore, one finally obtains as trust valuation

$$m_t(A) = 0.7559, m_t(\bar{A}) = 0.0945, m_t(A \cup \bar{A}) = 0.1496$$

For this scenario the source is confident about its own assertions, and therefore

$$m(A) = 0.8, m(\bar{A}) = 0.1, m(A \cup \bar{A}) = 0.1$$

and

$$m'(A) = 0.8, m'(\bar{A}) = 0.1, m'(A \cup \bar{A}) = 0.1$$

have identical BBA distributions. Reliability of the source is build namely on its reputation, as there are clues about the competence and intentions of the source. Hence, the overall BBA

$$m(r) = 0.9449, m(\bar{r}) = 0.0196, m(r \cup \bar{r}) = 0.0355$$

is close to the initial reputation distribution

$$m_R(r) = 0.9, m_R(\bar{r}) = 0.1, m_R(r \cup \bar{r}) = 0$$

Values of trust shows the impact of using incompletely reliable sources, which decreased the certainty level of the initial BBA

$$m'(A) = 0.8, m'(\bar{A}) = 0.1, m'(A \cup \bar{A}) = 0.1$$

to

$$m_t(A) = 0.7559, m_t(\bar{A}) = 0.0945, m_t(A \cup \bar{A}) = 0.1496$$

They also support the intution that the trust assigned by the analyst to A will have an upper limit equal to the reputation of the source.

URREF criteria	Objectivity	ObservationalSensitivity
Values	0.1	0.1

TABLE VII
REPUTATION: OBSERVATION UNCERTAINTY

URREF criteria	SelfConfidence	Ambiguity
Values	1	0.6

TABLE VIII
REPUTATION: REPORTING UNCERTAINTY

	Dissonance	Conclusiveness	Completeness
Updated BBAs	0.3	0.7	0.9
Reliability	0.07	0.93	0.955
Trust	0.34	0.66	0.84

TABLE IX
REPUTATION: FUSION UNCERTAINTY

This scenarios is similar the previous one as in both cases there are incomplete descriptions of the source. For this particular case, a historical recording of sources failures offers a basis to overcome the missing pieces and, in spite of low values for $Objectivity$ and $ObservationalSensitivity$, see tab. VII the final trust evaluation is improved with respect to the previous scenario, and shows a better combination of $Dissonance$, $Conclusivenes$ and $Completeness$, as shown in tab. IX.

Scenario 4 - Misleading report: In this case, Y provides the assertion A, while stating that it certainly holds and X considers Y as a completely unreliable source. For this case, the analyst knows that the report is somehow inaccurate, for example, it cannot be corroborated or it contradicts fully or in part information from other (reliable) sources. The analyst suspects the source of having misleading intentions, and can therefore assign a maximal uncertainty level to the information reported. This scenario corresponds by example to the following case for inputs: $\alpha_Y = 1$

$$m(A) = 1, m(\bar{A}) = 0, m(A \cup \bar{A}) = 0$$

$$m_C(r) = 0.1, m_C(\bar{r}) = 0.1, m_C(r \cup \bar{r}) = 0.8$$

$$m_R(r) = 0.1, m_R(\bar{r}) = 0.1, m_R(r \cup \bar{r}) = 0.8$$

$$m_I(r) = 0.1, m_I(\bar{r}) = 0.8, m_I(r \cup \bar{r}) = 0.1$$

and $w_C = 0.1$, $w_R = 0.1$ and $w_I = 0.8$. Hence, one gets

$$m'(A) = 1, m'(\bar{A}) = 0, m'(A \cup \bar{A}) = 0$$

and

$$m(r) = 0.0237, m(\bar{r}) = 0.9109, m(r \cup \bar{r}) = 0.0654$$

Therefore, one finally obtains as trust valuation

$$m_t(A) = 0.0237, m_t(\bar{A}) = 0, m_t(A \cup \bar{A}) = 0.9763$$

Values for this scenario show high self-confidence of the source and high accuracy of the assertion provided; therefore the initial BBA is unchanged after fusion with self-confidence. Nevertheless, the impact of having misleading intention is visible first on the mass distribution assigned to reliability and then on the overall values of trust. With respect to the initial values

$$m(A) = 1, m(\bar{A}) = 0, m(A \cup \bar{A}) = 0$$

and the partially fused ones

$$m'(A) = 1, m'(\bar{A}) = 0, m'(A \cup \bar{A}) = 0$$

, the integration of a misleading source transfers the mass assignation almost exclusively to $m_t(A \cup \bar{A})$.
Intuitively, the assertion A wile be ignored, as the reliability if the sources is drastically decreased by a high mass assignment on misleading intentions.

URREF criteria	Objectivity	ObservationalSensitivity
Values	0.1	0.1

TABLE X
MISLEADING REPORT: OBSERVATION UNCERTAINTY

URREF criteria	SelfConfidence	Ambiguity
Values	1	0.977

TABLE XI
MISLEADING REPORT: REPORTING UNCERTAINTY

	Dissonance	Conclusiveness	Completeness
Assertion	0	1	1
Source	0.11	0.89	0.935
Trust	0.763	0.237	0.03

TABLE XII
MISLEADING: FUSION UNCERTAINTY

This scenario illustrates the impact of misleading sources on trust estimation. Hence, the use case has very good values for reporting induced uncertainty, with high $SelfConfidence$ and low $Ambiguity$ see tab. XI, but the overall trust characterization shows strong $Dissonance$, to be corroborated with low $Conclusiveness$ and almost zero $Completeness$, as shown in tab. XII.

Scenario 5 - Ambiguous report: The source Y provides A and $v(A)$, the uncertainty level. Suppose that $v(A)$ has a low value, as the source is not very sure about the events reported, and that X considers Y to be unreliable. This scenario corresponds by example to the following case for inputs: $\alpha_Y = 0.3$

$$m(A) = 0.6, m(\bar{A}) = 0.2, m(A \cup \bar{A}) = 0.2$$

$$m_C(r) = 0.1, m_C(\bar{r}) = 0.8, m_C(r \cup \bar{r}) = 0.1$$

$$m_R(r) = 0.1, m_R(\bar{r}) = 0.8, m_R(r \cup \bar{r}) = 0.1$$

$$m_I(r) = 0.1, m_I(\bar{r}) = 0.1, m_I(r \cup \bar{r}) = 0.8$$

and $w_C = 0.2$, $w_R = 0.4$ and $w_I = 0.4$.
Hence, one gets

$$m'(A) = 0.18, m'(\bar{A}) = 0.06, m'(A \cup \bar{A}) = 0.76$$

and

$$m(r) = 0.0223, m(\bar{r}) = 0.4398, m(r \cup \bar{r}) = 0.5379$$

Therefore, one finally obtains as trust valuation

$$m_t(A) = 0.0040, m_t(\bar{A}) = 0.0013, m_t(A \cup \bar{A}) = 0.9946$$

This scenario is an illustration for the worst practical case and is relevant when the analyst receives a report provided by a source that lacks skills or competence to provide accurate descriptions of events. In this case, the reports is incomplete, ambiguous, or even irrelevant. In addition to low competence and reliability, the source is also unsure about the statement.

The first modification of BBA shows the strong impact of self confidence, which changes drastically the BBA of the initial assertions, from

$$m(A) = 0.6, m(\bar{A}) = 0.2, m(A \cup \bar{A}) = 0.2$$

to

$$m'(A) = 0.18, m'(\bar{A}) = 0.06, m'(A \cup \bar{A}) = 0.76$$

. Unsurprisingly, the overall reliability is low:

$$m(r) = 0.0223, m(\bar{r}) = 0.4398, m(r \cup \bar{r}) = 0.5379$$

and the results of the final combination show an important mass assigned to $m_t(A \cup \bar{A}) = 0.9946$. Intuitively, the information provided is useless, and considered an highly uncertain.

URREF criteria	Objectivity	ObservationalSensitivity
Values	0.1	0.1

TABLE XIII
AMBIGUOUS REPORT: OBSERVATION UNCERTAINTY

URREF criteria	SelfConfidence	Ambiguity
Values	0.3	0.27

TABLE XIV
AMBIGUOUS REPORT: REPORTING UNCERTAINTY

	Dissonance	Conclusiveness	Completeness
Assertion	0.6	0.4	0.8
Source	0.583	0.417	0.47
Trust	0.973	0.027	0.006

TABLE XV
AMBIGUOUS REPORT: FUSION UNCERTAINTY

This scenario shows the combined effects of uncertain reporting and incomplete sources description for trust estimation. First, the outcome is affected by high values of uncertainty induced during observation, tab. XIII and reporting passes, tab XIV. Then, fusion leads to a trust estimation having high values of $Dissonance$, and very low values of $Conclusiveness$ and $Completeness$. The same criteria estimated for reliability show the main difference with respect to the previous case which was also based on unreliable sources. While for scenario 4 the source still has important $Completeness$, the measure is drastically decreased for this scenario, see tab. XV.

As highlighted by the examples above, there is a need for a more detailed investigation of uncertainties occurring when modeling the problem of estimating trust in reported information and adopting a specific formalism for its representation. Using URREF criteria offers a unified basis to analyze inaccuracies affecting the estimation process during different phases (observation, reporting, and fusion) and facilitates the tracking of uncertainty propagation by highlighting when uncertainties are added into the system and which partial results and affected. The uncertainty analysis provides additional characterization of different results and finally helps their interpretation.

VIII. CONCLUSION

This paper presents a model to estimate trust in reported information and illustrated the use of URREF criteria to track uncertainty affecting the results, from model development to its formalization with belief functions.

First, a model for trust estimation is developed combining several attributes of the sources and its own appreciation of items reported. The model is implemented using belief functions, and takes advantage of its mathematical background to define fusion operators for trust assessment. Several scenarios are adopted to illustrate uncertainty analysis, highlighting when uncertainty occurs and how it affects partial results for different practical applications.

Tracking uncertainty is suitable for fusion systems, when various human sources send observations of questionable quality and there is a need to constantly update the trust associated to reports to be analyzed and URREF ontology offers a unified framework to choose appropriate criteria to capture uncertainty. From a practical standpoint, uncertainty analysis offers additional description of results and improves user interpretation.

REFERENCES

Alain Auger and Jean Roy. Expression of uncertainty in linguistic data. In *Information Fusion, 2008 11th International Conference on*, pages 1–8. IEEE, 2008.

M Bednarek. Evaluation in media discourse: Analysis of a newspaper corpus continuum, 2006.

Laure Berti-Équille and Javier Borge-Holthoefer. Veracity of data: From truth discovery computation algorithms to models of misinformation dynamics. *Synthesis Lectures on Data Management*, 7(3):1–155, 2015.

David B Buller and Judee K Burgoon. Interpersonal deception theory. *Communication theory*, 6(3):203–242, 1996.

Giuseppe Carenini, Raymond T Ng, and Xiaodong Zhou. Summarizing emails with conversational cohesion and subjectivity. In *ACL*, volume 8, pages 353–361, 2008.

Cristiano Castelfranchi and Rino Falcone. Trust is much more than subjective probability: Mental components and sources of trust. In *System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on*, pages 10–pp. IEEE, 2000.

Wei Chen. Dimensions of subjectivity in natural language. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, pages 13–16. Association for Computational Linguistics, 2008.

Laurence Cholvy. How strong can an agent believe reported information? In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 386–397. Springer, 2011.

Paulo CG Costa, Kathryn B Laskey, Erik Blasch, and Anne-Laure Jousselme. Towards unbiased evaluation of uncertainty reasoning: The urref ontology. In *Information Fusion (FUSION), 2012 15th International Conference on*, pages 2301–2308. IEEE, 2012.

Xin Luna Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Kevin Murphy, Shaohua Sun, and Wei Zhang. From data fusion to knowledge fusion. *Proceedings of the VLDB Endowment*, 7(10):881–892, 2014.

Xin Luna Dong, Evgeniy Gabrilovich, Kevin Murphy, Van Dang, Wilko Horn, Camillo Lugaresi, Shaohua Sun, and Wei Zhang. Knowledge-based trust: Estimating the trustworthiness of web sources. *Proceedings of the VLDB Endowment*, 8(9):938–949, 2015.

Valentina Dragos and Kellyn Rein. Integration of soft data for information fusion: Pitfalls, challenges and trends. In *Information Fusion (FUSION), 2014 17th International Conference on*, pages 1–8. IEEE, 2014.

Andrea Esuli and Fabrizio Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, volume 6, pages 417–422. Citeseer, 2006.

Patricia Everaere, Sébastien Konieczny, and Pierre Marquis. Belief merging versus judgment aggregation. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pages 999–1007. International Foundation for Autonomous Agents and Multiagent Systems, 2015.

Rino Falcone and Cristiano Castelfranchi. Social trust: A cognitive approach. In *Trust and deception in virtual societies*, pages 55–90. Springer, 2001.

Jones Granatyr, Vanderson Botelho, Otto Robert Lessing, Edson Emílio Scalabrin, Jean-Paul Barthès, and Fabrício Enembreck. Trust and reputation models for multiagent systems. *ACM Computing Surveys (CSUR)*, 48(2):27, 2015.

Tyrone Grandison and Morris Sloman. A survey of trust in internet applications. *IEEE Communications Surveys & Tutorials*, 3(4):2–16, 2000.

Jia Guo and Ray Chen. A classification of trust computation models for service-oriented internet of things systems. In *Services Computing (SCC), 2015 IEEE International Conference on*, pages 324–331. IEEE, 2015.

David Lee Hall and John M Jordan. *Human-centered information fusion*. Artech House, 2010.

Jeffrey T Hancock, Lauren E Curry, Saurabh Goorha, and Michael Woodworth. On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, 45(1):1–23, 2007.

Valerie Hauch, Jaume Masip, Iris Blandon-Gitlin, and Siegfried L Sporer. Linguistic cues to deception assessed by computer programs: a meta-analysis. In *Proceedings of the workshop on computational approaches to deception detection*, pages 1–4. Association for Computational Linguistics, 2012.

Michael P Jenkins, Geoff A Gross, Ann M Bisantz, and Rakesh Nagi. Towards context aware data fusion: Modeling and integration of situationally qualified human observations to manage uncertainty in a hard+ soft fusion process. *Information Fusion*, 21:130–144, 2015.

Audun Josang, Roslan Ismail, and Colin Boyd. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618–644, 2007.

Anne-Laure Jousselme, Anne-Claire Boury-Brisset, Benoit Debaque, and Donald Prevost. Characterization of hard and soft sources of information: A practical illustration. In *Information Fusion (FUSION), 2014 17th International Conference on*, pages 1–8. IEEE, 2014.

Sherman Kent. Words of estimative probability. *Studies in Intelligence*, 8(4):49–65, 1964.

Rachel F Kesselman. *Verbal probability expressions in national intelligence estimates: a comprehensive analysis of trends from the fifties through post 9/11*. PhD thesis, MERCYHURST COLLEGE, 2008.

Bahador Khaleghi, Alaa Khamis, Fakhreddine O Karray, and Saiedeh N Razavi. Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*, 14(1):28–44, 2013.

Andrew Koster, Ana LC Bazzan, and Marcelo de Souza. Liar liar, pants on fire; or how to use subjective logic and argumentation to evaluate information from untrustworthy sources. *Artificial Intelligence Review*, 48(2):219–235, 2017.

Tatiana Lukoianova and Victoria L Rubin. Veracity roadmap: Is big data objective, truthful and credible? *Advances in Classification Research Online*, 24(1):4–15, 2014.

D Harrison McKnight and Norman L Chervany. The meanings of trust. 1996.

Rada Mihalcea and Carlo Strapparava. The lie detector: Explorations in the automatic recognition of deceptive language. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 309–312. Association for Computational Linguistics, 2009.

NATO. STANAG 2011. Technical report, Intelligence reports, 2003.

Fabio Paglieri, Cristiano Castelfranchi, Célia da Costa Pereira, Rino Falcone, Andrea Tettamanzi, and Serena Villata. Trusting the messenger because of the message: feedback dynamics from information quality to source evaluation. *Computational and Mathematical Organization Theory*, 20 (2):176–194, 2014.

James W Pennebaker, Martha E Francis, and Roger J Booth. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71:2001, 2001.

Gabriella Pigozzi. Belief merging and judgment aggregation. *Stanford Encyclopedia of Philosophy*, 2015.

Kellyn Rein. *I think it is likely that it might be soExploiting Lexical Clues for the Automatic Generation of evidentiality Weights for Information Extracted from English Text*. Dissertation, University of Bonn, 2016.

Ellen Riloff and Janyce Wiebe. Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 105–112. Association for Computational Linguistics, 2003.

Galina L Rogova and Eloi Bosse. Information quality in information fusion. In *Information Fusion (FUSION), 2010 13th Conference on*, pages 1–8. IEEE, 2010.

Victoria L Rubin. Stating with certainty or stating with doubt: Intercoder reliability results for manual annotation of epistemically modalized statements. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers*, pages 141–144. Association for Computational Linguistics, 2007.

Victoria L Rubin. Epistemic modality: From uncertainty to certainty in the context of information seeking as interactions with texts. *Information Processing & Management*, 46(5):533–540, 2010.

Victoria L Rubin, Elizabeth D Liddy, and Noriko Kando. Certainty identification in texts: Categorization model and manual tagging results. In *Computing attitude and affect in text: Theory and applications*, pages 61–76. Springer, 2006.

Roser Saurí and James Pustejovsky. Factbank: A corpus annotated with event factuality. *Language resources and evaluation*, 43(3):227–268, 2009.

Roser Saurí and James Pustejovsky. Are you sure that this happened? assessing the factuality degree of events in text. *Computational Linguistics*, 38(2):261–299, 2012.

Glenn Shafer et al. *A mathematical theory of evidence*, volume 1. Princeton university press Princeton, 1976.

Wanita Sherchan, Surya Nepal, and Cecile Paris. A survey of trust in social networks. *ACM Computing Surveys (CSUR)*, 45(4):47, 2013.

Florentin Smarandache and Jean Dezert. *Advances and applications of DSmT for information fusion*, volume 1-4.

American Research Press (ARP), 2015.

Florentin Smarandache, Jean Dezert, and J-M Tacnet. Fusion of sources of evidence with different importances and reliabilities. In *Information Fusion (FUSION), 2010 13th Conference on*, pages 1–8. IEEE, 2010.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307, 2011.

Yuqing Tang, Kai Cai, Peter McBurney, Elizabeth Sklar, and Simon Parsons. Using argumentation to reason about trust and belief. *Journal of Logic and Computation*, 22(5):979–1018, 2011.

Douglas P Twitchell, David P Biros, Mark Adkins, Nicole Forsgren, Judee K Burgoon, and Jay F Nunamaker. Automated determination of the veracity of interview statements from people of interest to an operational security force. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, volume 1, pages 17a–17a. IEEE, 2006.

Matteo Venanzi, Alex Rogers, and Nicholas R Jennings. Trust-based fusion of untrustworthy information in crowdsourcing applications. In *Proceedings of the 2013 international conference on autonomous agents and multi-agent systems*, pages 829–836. International Foundation for Autonomous Agents and Multiagent Systems, 2013.

Janyce Wiebe and Ellen Riloff. Creating subjective and objective sentence classifiers from unannotated texts. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 486–497. Springer, 2005.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2-3):165–210, 2005.

Hong Yu and Vasileios Hatzivassiloglou. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 129–136. Association for Computational Linguistics, 2003.