

# Reasoning with Uncertain Information and Trust

Murat Sensoy<sup>a,e</sup>, Geeth de Mel<sup>b,c</sup>, Achille Fokoue<sup>b</sup>, Timothy J. Norman<sup>e</sup>, Jeff Z. Pan<sup>e</sup>, Yuqing Tang<sup>d</sup>,  
Nir Oren<sup>e</sup>, Katia Sycara<sup>d</sup>, Lance Kaplan<sup>c</sup>, Tien Pham<sup>c</sup>

<sup>a</sup>Computer Science, Ozyegin University, Istanbul, Turkey

<sup>b</sup>IBM T. J. Watson Research Center, NY, USA

<sup>c</sup>US Army Research Laboratory, Maryland, USA

<sup>d</sup>Carnegie Mellon University, Pittsburgh, USA

<sup>e</sup>Computing Science, University of Aberdeen, UK

## ABSTRACT

A limitation of standard Description Logics is its inability to reason with uncertain and vague knowledge. Although probabilistic and fuzzy extensions of DLs exist, which provide an explicit representation of uncertainty, they do not provide an explicit means for reasoning about second order uncertainty. Dempster-Shafer theory of evidence (DST) overcomes this weakness and provides means to fuse and reason about uncertain information. In this paper, we combine DL-Lite with DST to allow scalable reasoning over uncertain semantic knowledge bases. Furthermore, our formalism allows for the detection of conflicts between the fused information and domain constraints. Finally, we propose methods to resolve such conflicts through trust revision by exploiting evidence regarding the information sources. The effectiveness of the proposed approaches is shown through simulations under various settings.

**Keywords:** Information Fusion, Trust, Uncertainty, Description Logics

## 1. INTRODUCTION

Effective and efficient decision making plays a crucial role in success of any operation. Key to successful decision making is the effective interpretation of the available data about the given domain – i.e., Data-to-Decisions (D2D). This is especially true for coalition operations where the operations are critical and data-centric. It is important to note that the data gathered by sources may represent supporting or negating evidence about a particular phenomenon in the domain. For example, an acoustic array may have recorded a series of engine signatures and have deduced that it is of a *military truck* – i.e., its interpretation of the gathered evidence. Thus, the engine signatures can be taken as the *evidence* that support a proposition such as *military vehicle is in motion*. This can be taken as the *opinion* of the sensor about a particular phenomena based on its current observations. Assume that a seismic sensor has also picked-up a set of vibration signatures and deduced that it is of a heavy vehicle moving from north – i.e., its opinion about the environment based on its current observations. These opinions may support a global proposition such as *military vehicle moving from north*.

However, in data heavy environments such as coalition operations utilizing data to make informed decisions is not straightforward. This is because data that has to be consumed in order to make decisions are from multiple parties with different granularities and confidence levels. For example, a coalition partner may obfuscate data in order to hide a specific information or may reduce the resolution of data if it has less trust on the sharing party. Therefore, these data will have an inherent uncertainty that has to be considered while fusing to make decisions. There are many approaches studied in literature to address this issue and evidential theory<sup>1</sup> is probably the best known. However, it has been shown that the approaches based on the evidential theory suffer from evidence compatibility issues when presented with conflicting opinions. In environments – such as the ones this work is based on – such conflicts are common, thus, we need a new approach to reason about uncertainty in the face of conflicting opinions.

The importance of the D2D problem is well understood in the military; D2D is recognized as one of the top seven challenges to be addressed by the Department of Defense (DoD) Quadrennial Defense Review (QDR)\*. As an entity of the DoD, the research division of the United States Army – i.e., US Army Research Laboratory (ARL) – has invested a lot

---

Further author information ⇒E-mail: murat.sensoy@ozyegin.edu.tr, Telephone: +90-216-564-9493

\*<http://www.acq.osd.mil/chieftechnologist/areas/>

of time and effort in realizing this challenge. ARL provides assistance in conducting basic and applied research in D2D through a number of collaborative efforts and International Technology Alliance (ITA)<sup>†</sup> is one such transatlantic effort. The ITA program is a research program to address issues related to mobile ad-hoc networks for military coalitions. The research is aimed at fundamental advances in information and network sciences that will enhance decision making for coalition operations.

Rest of the paper is structured as follows. In Section 2 we set out the preliminaries for our work. It discusses a formal model for knowledge representation and highlights the need for reasoning based on uncertainty. It also provides clues to mechanisms that can assist in performing effective reasoning in uncertain environments. Section 3 introduces a scenario which highlights the need to have a mechanism to make use of opinions belonging to multiple parties in order to make informed decisions in the face of lack of trust. In Section 4 we introduce the syntax and semantics of our formalism to represent uncertainty in knowledge and in Section 5 we show how that formalism can be used to perform trust revisions. We evaluate our approach in Section 6 and conclude the document in Section 7 with a discussion of related work.

## 2. BACKGROUND

In order to intelligently reason in uncertain domains – such as the ones discussed in Section 1 – we need a language to capture the domain effectively and efficiently. In this work, we shall use Description Logics (DLs) for this purpose. Providing a full overview of DLs is out-of-the scope of this paper. However, we refer the reader to Baader *et al.*<sup>2</sup> for an overview of DLs.

### 2.1 DL-based Knowledge Representation

Even for the smallest propositionally closed DL,  $\mathcal{ALC}$  (which only provides class constructors  $\neg C, C \sqcap D, C \sqcup D, \exists R.C$  and  $\forall R.C$ ), the complexity of logical entailment is EXPTIME – a class of decision problems that can be solved by a deterministic Turing machine. Recently, Calvanese *et al.*<sup>3</sup> proposed DL-Lite, which can express most features in UML class diagrams with a low reasoning overhead (with data complexity  $AC_0$ ). It is for this reason that we base our model on DL-Lite<sub>core</sub> (referred to here as DL-Lite, although there are extensions<sup>4</sup>), and hence provide a brief formalisation to ground the subsequent presentation of our model.

A DL-Lite knowledge base  $\mathcal{K} = (\mathcal{T}, \mathcal{A})$  consists of a TBox  $\mathcal{T}$  and an ABox  $\mathcal{A}$ . Axioms of the following forms compose  $\mathcal{K}$ :

1. class inclusion axioms:  $B \sqsubseteq C \in \mathcal{T}$  where  $B$  is a basic class  $B := A \mid \exists R \mid \exists R^-$  and  $C$  is a general class  $C := B \mid \neg B \mid C_1 \sqcap C_2$  (where  $A$  denotes an named class,  $R$  denotes a named property, and  $R^-$  is the inverse of  $R$ )  
E.g., a car subsumes a vehicle (i.e.,  $\text{Car} \sqsubseteq \text{Vehicle}$ )
2. individual axioms:  $B(a), R(a, b) \in \mathcal{A}$  where  $a$  and  $b$  are named individuals. E.g., a jeep is a type of a vehicle – i.e.,  $\text{Vehicle}(\text{Jeep})$  – and jeep can travel is rough terrain – i.e.,  $\text{canTravel}(\text{Jeep}, \text{RoughTerrain})$ .

Description Logics have a well-defined model-theoretic semantics, which are provided in terms of interpretations. An interpretation  $\mathcal{I}$  is a pair  $(\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ , where  $\Delta^{\mathcal{I}}$  is a non-empty set of objects and  $\cdot^{\mathcal{I}}$  is an interpretation function, which maps each class  $C$  to a subset  $C^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$  and each property  $R$  to a subset  $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ . Using a trivial normalisation, it is possible to convert class inclusion axioms of the form  $B_1 \sqsubseteq C_1 \sqcap C_2$  into a set of simpler class inclusions of the form  $B_1 \sqsubseteq B_i$  or  $B_1 \sqsubseteq \neg B_j$ , where  $B_1, B_i$ , and  $B_j$  are basic concepts.<sup>3</sup> For instance, during normalisation,  $B_1 \sqsubseteq B_2 \sqcap \neg B_3$  is replaced with  $B_1 \sqsubseteq B_2$  and  $B_1 \sqsubseteq \neg B_3$ .

Though the statements in knowledge bases created base on the above formalism is supposed to contain facts, it is in-fact important to note that those statements may be probabilistic in nature. For example, “detection of a moving vehicle” is in fact can only be stated with a 90% accuracy. Thus, we need a mechanisms to reason about such uncertain statements. Dempster-Shafer theory of evidence (DST) provides an explicit framework to reason about such knowledge bases and in the next section we briefly discuss its variations.

---

<sup>†</sup><https://www.usukita.org/>

## 2.2 Subjective Opinions

Dempster-Shafer Theory (DST) offers means to characterise an agent's view of the state of world by assigning *basic probability masses* to subsets of truth assignments of propositions in the logic. Jøsang<sup>5</sup> proposed Subjective Logic (SL), which can be considered as an interpretation and extension of DST with logical operators (e.g., conjunction, deduction, abduction and so on). Jøsang<sup>5</sup> coined the term *subjective opinions* to refer to uncertain statements. In SL, all the operators are grounded on probability theory – as oppose to DST – thus, allowing one to consider the mathematical properties of the fusion easily. In this work, we take Jøsang's view of DST to represent and reason about uncertain statements.

A *binomial opinion* about a proposition  $x$  is represented by a triple  $w_x = (b_x, d_x, u_x)$  which is derived from the basic probability masses assigned to subsets of truth assignments of the language. In the opinion  $w_x$ ,  $b_x$ , also denoted by  $b(w_x)$ , is the belief about  $x$  — the summation of the probability masses that entail  $x$ ;  $d_x$ , also denoted by  $d(w_x)$ , is the disbelief about  $x$  — the summation of the probability masses that entail  $\neg x$ ; and  $u_x$ , also denoted by  $u(w_x)$ , is the uncertainty about  $x$  — the summation of the probability masses that neither entail  $x$  nor entail  $\neg x$ . The constraints over the probability mass assignment function require that  $b_x + d_x + u_x = 1$  and  $b_x, d_x, u_x \in [0, 1]$ . When a more concise notation is necessary, we use  $(b_x, d_x)$  instead of  $(b_x, d_x, u_x)$ , since  $u_x = 1 - b_x - d_x$ . The negation over an opinion  $w_x$  is defined as  $\neg(b_x, d_x, u_x) = (d_x, b_x, u_x) = (b_{\neg x}, d_{\neg x}, u_{\neg x})$ .<sup>5</sup>

DEFINITION 1. Let  $w_1 = (b_1, d_1, u_1)$  and  $w_2 = (b_2, d_2, u_2)$  be two opinions about the same proposition. We call  $w_1$  a *specialisation* of  $w_2$  ( $w_1 \preceq w_2$ ) iff  $b_2 \leq b_1$  and  $d_2 \leq d_1$  (implies  $u_1 \leq u_2$ ). Similarly, we call  $w_1$  a *generalisation* of  $w_2$  ( $w_2 \preceq w_1$ ) iff  $b_1 \leq b_2$  and  $d_1 \leq d_2$  (implies  $u_2 \leq u_1$ ).

An agent  $i$ 's opinion about a proposition  $x$  is denoted by  $w_x^i = (b_x^i, d_x^i, u_x^i)$ . This opinion  $w_x^i$  may not be *directly* used by another agent  $j$ . Agent  $j$  could have a view of the reliability or competence of  $i$  with respect to  $x$ . Shafer<sup>1</sup> proposed a discounting operator  $\otimes$  to normalise the belief and disbelief in  $w_x^j$  based on the degree of trust  $j$  has of  $i$  with respect to  $x$ :  $t_i^j$ . The discounted opinion,  $w_x^j$ , is computed as  $(b_x^i \times t_i^j, d_x^i \times t_i^j)$ . The trustworthiness of information sources can be modelled using Beta probability density functions.<sup>6</sup> A Beta distribution has two parameters  $(r + 1, s + 1)$ , where  $r$  is the amount of positive evidence and  $s$  is the amount of negative evidence for the trustworthiness agent  $i$  agent has for agent  $j$ . The degree of trust  $t_j^i$  is then computed as the expectation value of the Beta distribution:  $t_j^i = (r + 1)/(r + s + 2)$ .

In the next section, we introduce a coalition-based scenario in which opinions generated from multiple heterogeneous information sources are used to make informed decisions about critical situations by revising the trust associated with the information.

## 3. MOTIVATIONAL SCENARIO

A coalition operating in a mountainous area has planned for a high-value-target (HVT) extraction. The coalition consists of trusted partners  $P_1, P_2$  and the local partners  $P_{loc}$ .  $P_1$  is executing the HVT extraction and the command and control (C2) receives information from a *local informant* – i.e.,  $P_i$  – about suspicious activity on a road leading to the location where HVT resides. However, sensor resources belonging to  $P_1$  deployed in the area have not picked-up any recent activity. An observation post owned by  $P_{loc}$  in the north region also reports vehicle movement along the road; the observation is obtained by interpreting the evidences gathered by using long-range observation devices. However, the trust  $C_2$  has on local informants/militia are very limited due to their past experiences. Meanwhile, the trusted coalition partner  $P_2$  is executing a reconnaissance operation over the same area using a sensing resource  $S$  owned by  $P_1$ .  $P_2$ , too, observes some activity on the road based on the aerial images of  $S$ . Note that  $P_2$  has a good trust value on  $P_i$  on events such as reconnaissance, thus,  $P_2$  can vouch for  $P_i$  in this context.  $C_2$  now 1) increases its trust on the local informants and revises the trust assessments it has on similar tasks with the local informants 2) start making plans for the eventualities associated with the current task.

In Table 1, we provide a snapshot of the information sources C2 has access to with respect to their trustworthiness. Assume that  $P_i$  reports the observation of the vehicular movement along the road with an opinion of  $(0.9, 0, 0.1)$ . C2 interprets this opinion based on the trust assessment it has on  $P_i$  given in Table 1. Thus, the discounted opinion of the local informant is  $(0.412, 0, 0.588)$  which has a higher uncertainty as oppose to the original report. However,  $P_2$  has a better trust value on  $P_i$  in such scenarios; assume that  $P_2$ 's trust in  $P_i$  is 95%. Thus, it can be shown that  $P_2$ 's opinion of what  $P_i$  reported as  $(0.855, 0, 0.145)$ . Since C2 trust  $P_2$  better, it can be shown that C2 indeed can interpret the opinion expressed by  $P_i$  as  $(0.838, 0, 0.162)$  in this context, which has a greater confidence level than the original discount based on C2's own experience about  $P_i$

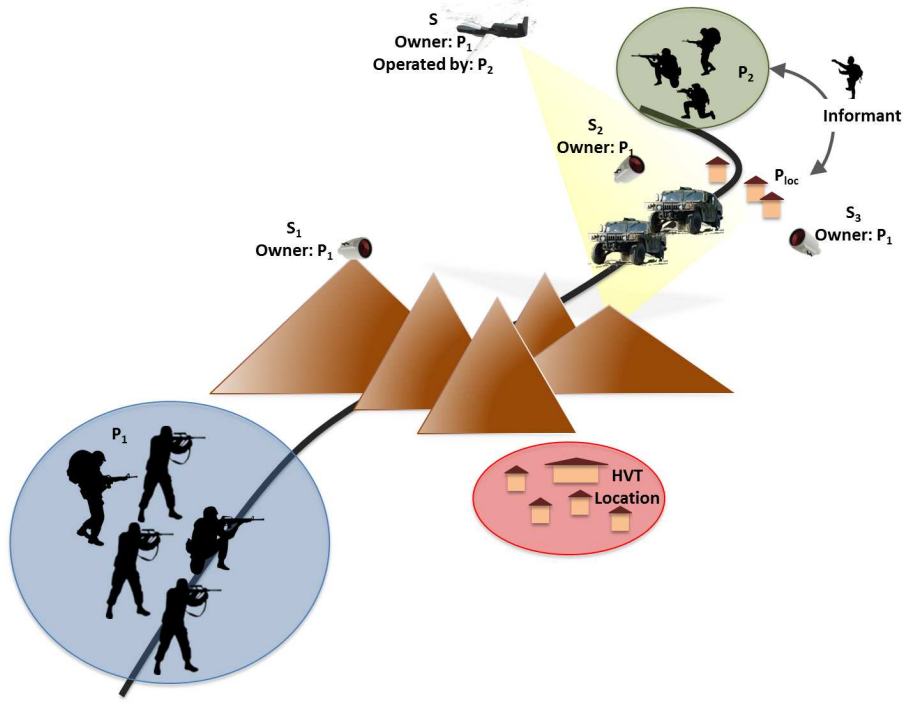


Figure 1. Scenario: high value target extraction

Table 1. Information sources and their trustworthiness

Source	Definition	Evidence	Degree of Trust
$P_i$	Informant	(10, 12)	0.458
$P_{loc}$	Coalition's Local partner	(10, 3)	0.786
$S_{1...3}$	Acoustic array of $P_1$	(1000, 0)	0.999
$P_2$	Coalition partner of $P_1$	(50, 0)	0.981

An important property to note in the above scenario is the fact that how  $C_2$  can revise its trust on local informants when evidence from trusted partners and resources come to light. This is because, though the prior trust of local informants is limited to  $C_2$ 's past experience, with the added information from trusted partners, the less trustworthy source can indeed increase its trust, at least in some contexts. Such properties can also be used to create trust matrices for future collaborations. Having provided an overview of DLs, DST, and subjective opinions, in the next section we provide a formalisation of subjective DLs so that uncertain statements could be captured and reasoned in DL efficiently.

#### 4. SDL-Lite

We propose *Subjective DL-Lite* (or *SDL-Lite* for short), which extends *DL-Lite<sub>core</sub>* with subjective opinion assertions of the form  $B:w$ , where  $w$  is an opinion and  $B$  is an ABox axiom (i.e., assertion). Each ABox axiom is associated with one opinion. ABox axioms have the form  $B(a)$  or  $R(a, b)$ , where  $B$  is basic class,  $R$  is a property, and  $a$  and  $b$  are individuals.

##### 4.1 SDL-Lite Semantics

In common with *DL-Lite* ontologies, the semantics of an ontology in *SDL-Lite* is defined in terms of *subjective interpretations*. Let  $\mathcal{W}$  be the set of all possible subjective binary opinions. A subjective interpretation is a pair  $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$  where the domain  $\Delta^{\mathcal{I}}$  is a non-empty set of objects and  $\cdot^{\mathcal{I}}$  is a subjective interpretation function, which maps:

Syntax	Semantics
$\top$	$\top^{\mathcal{I}}(o) = (1, 0, 0)$
$\perp$	$\perp^{\mathcal{I}}(o) = (0, 1, 0)$
$\exists R$	$b((\exists R)^{\mathcal{I}}(o_1)) \geq \max_{\cup_{o_2}} \{b(R^{\mathcal{I}}(o_1, o_2))\}$ and $d((\exists R)^{\mathcal{I}}(o_1)) \leq \min_{\cup_{o_2}} \{d(R^{\mathcal{I}}(o_1, o_2))\}$
$\neg B$	$(\neg B)^{\mathcal{I}}(o) = \neg B^{\mathcal{I}}(o)$
$R^-$	$(R^-)^{\mathcal{I}}(o_2, o_1) = R^{\mathcal{I}}(o_1, o_2)$
$B_1 \sqsubseteq B_2$	$\forall o \in \Delta^{\mathcal{I}}, b(B_1^{\mathcal{I}}(o)) \leq b(B_2^{\mathcal{I}}(o))$ and $d(B_2^{\mathcal{I}}(o)) \leq d(B_1^{\mathcal{I}}(o))$
$B_1 \sqsubseteq \neg B_2$	$\forall o \in \Delta^{\mathcal{I}}, b(B_1^{\mathcal{I}}(o)) \leq d(B_2^{\mathcal{I}}(o))$ and $b(B_2^{\mathcal{I}}(o)) \leq d(B_1^{\mathcal{I}}(o))$
$B(\mathbf{a}):w$	$b(w) \leq b(B^{\mathcal{I}}(\mathbf{a}^{\mathcal{I}}))$ and $d(w) \leq d(B^{\mathcal{I}}(\mathbf{a}^{\mathcal{I}}))$
$R(\mathbf{a}, \mathbf{b}):w$	$b(w) \leq b(R^{\mathcal{I}}(\mathbf{a}^{\mathcal{I}}, \mathbf{b}^{\mathcal{I}}))$ and $d(w) \leq d(R^{\mathcal{I}}(\mathbf{a}^{\mathcal{I}}, \mathbf{b}^{\mathcal{I}}))$

Table 2. Semantics of Subjective DL-Lite

- an individual  $\mathbf{a}$  to an element of  $\mathbf{a}^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ ,
- a named class  $A$  to a function  $A^{\mathcal{I}} : \Delta^{\mathcal{I}} \rightarrow \mathcal{W}$ ,
- a named property  $R$  to a function  $R^{\mathcal{I}} : \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \rightarrow \mathcal{W}$ .

To provide a semantics for  $\mathcal{SDL}$ -Lite, we extend interpretations of  $\mathcal{DL}$ -Lite class and property descriptions, and of axioms under unique name assumption. The semantics are presented in Table 2. The semantics of  $\exists R$  is derived from the rule  $R(\mathbf{a}^{\mathcal{I}}, \mathbf{b}^{\mathcal{I}}) \rightarrow \exists R(\mathbf{a}^{\mathcal{I}}), \forall \mathbf{b}^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ . This rule constrains the minimum belief and the maximum disbelief that  $\exists R(\mathbf{a}^{\mathcal{I}})$  can have. For any individuals  $\mathbf{a}$  and  $\mathbf{b}$ , the belief in  $\mathbf{a}$  having a property  $R$  (i.e.,  $\exists R(\mathbf{a})$ ), is not less than belief in  $\mathbf{a}$  having the property  $R$  with  $\mathbf{b}$  (i.e.,  $R(\mathbf{a}, \mathbf{b})$ ), and disbelief in  $\exists R(\mathbf{a})$  is not more than disbelief in  $R(\mathbf{a}, \mathbf{b})$ . An ontology provides us with domain constraints in the form of TBox axioms. For instance, the axiom  $B_1 \sqsubseteq B_2$  means that every instance of class  $B_1$  is also an instance of class  $B_2$ . This trivially implies  $\neg B_2 \sqsubseteq \neg B_1$ , i.e., an individual that is not an instance of  $B_2$  cannot be an instance of  $B_1$ . Therefore, given an individual  $\mathbf{a}$ , the axiom  $B_1 \sqsubseteq B_2$  implies that our belief in  $B_2(\mathbf{a})$  cannot be less than our belief in  $B_1(\mathbf{a})$  and our disbelief in  $B_2(\mathbf{a})$  cannot be more than our disbelief in  $B_1(\mathbf{a})$ . That is,  $b(B_1^{\mathcal{I}}(\mathbf{a}^{\mathcal{I}})) \leq b(B_2^{\mathcal{I}}(\mathbf{a}^{\mathcal{I}}))$  and  $d(B_2^{\mathcal{I}}(\mathbf{a}^{\mathcal{I}})) \leq d(B_1^{\mathcal{I}}(\mathbf{a}^{\mathcal{I}}))$  must hold. Similar constraints also exist in Table 2 for  $B_1 \sqsubseteq \neg B_2$ .

**DEFINITION 2.** An  $\mathcal{SDL}$ -Lite knowledge base  $\mathcal{K} = (\mathcal{T}, \mathcal{A})$  is consistent if and only if it has a model. A model of  $\mathcal{K}$  is an interpretation of  $\mathcal{K}$  that satisfies the constraints in Table 2.

If  $\mathcal{K}$  is consistent, it can have many models, but one of them is the most general model with respect to the partial ordering on opinions by Definition 1. Providing a detailed description on how to detect consistency, and how to compute the most general model of a consistent  $\mathcal{SDL}$ -Lite knowledge base is out-of-the-scope of this paper; we refer the reader to Murat *et al.*<sup>7</sup> for the details. In the rest of the paper, we assume that the opinion about a specific ABox assertions is provided by a single source. When there is more than one source for an assertion, only one of them is chosen (e.g. based on their trustworthiness). This will be relaxed in future. Having described  $\mathcal{SDL}$ -Lite we now examine a novel application of the system, describing how evidence from multiple sources can be reasoned about based on the trust placed in these sources.

## 5. TRUST-BASED EVIDENCE ANALYSIS

Here we get to the crux of the problem being addressed in this paper: how can we draw reliable conclusions regarding the state of the world, given evidence acquired from disparate sources (agents), about whom we have variable trust? We refer to this process as trust-based evidence analysis. Our aim is not to offer a new mechanism for assessing the trustworthiness of information sources; in fact, we exploit a widely-studied model<sup>6</sup> for this purpose based on Beta distributions as described in Section 2.2. The novelty of this work lies in the use of such models to guide evidence analysis.

### 5.1 Handling Inconsistencies

SDL-Lite presented in the previous section provides a tractable means to capture and interpret evidence acquired from other agents. The fact that we have evidence from multiple agents, however, means that there are likely to be inconsistencies in the evidence received. Thus, given evidence (i.e., opinions) from various sources, our knowledge-base may not be consistent. This is despite the use of *discounting* through DST. Discounting provides us with a “best-guess” of the reliability of agents based on an aggregation of our prior experiences with, and other knowledge of them as evidence sources. As with any computational model of trust, the trust assessments that drive discounting are vulnerable to: lack of evidence about other agents and the effects of whitewashing;<sup>8</sup> a conflation of the probability of malicious behaviour and lack competence/expertise in the evidence-provider; strategic liars; and collusion among evidence-providers. In our running example, for instance, local police and civilian sources have relatively low trustworthiness, not because of any perceived malicious intent but due to a belief that they lack experience in providing precise information. With more evidence, trustworthiness of information sources may be modelled more accurately, but our challenge is to support the analysis of evidence given the status quo.

To illustrate this challenge, consider our example scenario in which  $P_i$  reports of a vehicular movement along the road. Based on the trustworthiness values given in Table 1, C2’s discounted opinion of  $P_i$ ’s observation is  $(0.412, 0, 0.588)$ . However, the discounted opinion C2 obtained from observations of  $P_1$ ’s acoustic array (i.e.,  $S_{1..3}$ ) is  $(0.099, 0.799, 0.102)$ . This clearly represents a conflict since  $0.412 + 0.799 > 1$  and would result in an inconsistent knowledge-base. Let  $w_1 = (0.412, 0, 0.588)$  and  $w_2 = (0.099, 0.799, 0.102)$ . The *conflicting portions* of  $w_1$  and  $w_2$  are  $c_{12} = 0.412$  and  $c_{21} = 0.799$ . Let us refer to the trustworthiness of the sources of  $w_1$  and  $w_2$  as  $t_1$  and  $t_2$  respectively. In our example, from Table 1,  $t_1 = 0.458$  and  $t_2 = 0.999$ . In order for us to transform our inconsistent knowledge-base into a *consistent* knowledge-base, from which we can draw valid conclusions given our semantics, we need to determine additional discounting factors  $x_1$  and  $x_2$  for *opinions*  $w_1$  and  $w_2$  such that  $0 \leq c_{12}.x_1 + c_{21}.x_2 \leq 1$ .

In this paper, we specify this problem as that of finding *additional* discounting factors for the belief-mass distributions of pieces of evidence to make our knowledge-base consistent. In general, our conflict resolution problem is a tuple  $\langle \mathcal{C}, \mathcal{X} \rangle$  where  $\mathcal{C}$  is the set of conflicting portions that appear in the extended knowledge base, and  $\mathcal{X}$  is a set of additional discounting factors corresponding to  $\mathcal{C}$ . We require that, in  $\langle \mathcal{C}, \mathcal{X} \rangle$ ,  $\forall c_{ij} \in \mathcal{C}$ ,  $\exists c_{ji} \in \mathcal{C}$  and  $\exists x_i, x_j \in \mathcal{X}$ . Then, a solution to this problem is an assignment of values to each  $x_i \in \mathcal{X}$  such that

$$\forall c_{ij}, c_{ji} \in \mathcal{C}, \forall x_i, x_j \in \mathcal{X} \quad 0 \leq c_{ij}.x_i + c_{ji}.x_j \leq 1$$

There are many heuristic approaches to solving this problem, among them being to consider only consistent knowledge to draw conclusions from the evidence received; i.e.  $\forall x_i \in \mathcal{X}, x_i = 0$ . This, however, could lead to a significant loss of evidence. Here, we explore a number of increasingly refined approaches that guarantee the generation of a consistent knowledge-base: *trust-based deleting*, *trust-based discounting* and *evidence-based discounting*.

### 5.2 Trust-based deleting

If two opinions  $w_1$  and  $w_2$  are in conflict, the opinion from the less trustworthy source is deleted, and if both sources are equally trustworthy both opinions are deleted. Thus, if the trust we have in the source of opinion  $w_1$  is greater than that of the source of  $w_2$  ( $t_1 > t_2$ ) then  $x_2 = 0$  and  $x_1 = 1$ , and in the event that  $t_1 = t_2$  we assign  $x_1 = x_2 = 0$ .

### 5.3 Trust-based discounting

If two opinions  $w_1$  and  $w_2$  are in conflict, they are discounted in proportion to the trustworthiness of their sources. That is, the additional discounting factor for  $w_1$  and  $w_2$  is computed using  $t_1/(c_{12}t_1 + c_{21}t_2)$  and  $t_2/(c_{12}t_1 + c_{21}t_2)$ , respectively, where  $t_1$  and  $t_2$  are the trustworthiness of the sources of the opinions. In our example, an additional discount factor for  $P_i$ 's opinion is 0.386 and that of  $S_{1...3}$  is 0.842, since the trustworthiness of  $P_i$  and  $S_{1...3}$  are 0.458 and 0.999, respectively. Therefore, to resolve the conflict, the original opinion of  $P_i$  is discounted by  $0.458 \times 0.386 = 0.177$  and that of  $S_{1...3}$  is discounted by  $0.999 \times 0.842 = 0.841$ . However, this approach neglects the amount of evidence used to calculate trust in sources.

### 5.4 Evidence-based discounting

Within the evidence analysis domain, the information that we have to work with relates to past experiences with a specific agent (i.e., information source)  $\varrho_k$  where information received has proven reliable or unreliable according to some criteria (as would be captured in any trust assessment model). In other words, the amount of positive evidence we have for agent  $\varrho_k$ , namely  $r_k$ , and the amount of negative evidence for that agent, namely  $s_k$ . From this evidence, we calculate trustworthiness of  $\varrho_k$ , denoted as  $t_k$  described in Section 2.2. When we receive opinion  $w_i^k$  from  $\varrho_k$ , we discount it by  $t_k$  and add the resulting opinion  $w_i$  to our knowledge base. However, as explained before, additional discounting by factor  $x_i$  is required when  $w_i$  is in conflict with another opinion in the knowledge base. Discounting  $w_i$  by  $x_i$  implies discounting the original opinion  $w_i^k$  by  $t_k \cdot x_i$ . This corresponds to revising the trustworthiness of  $w_i^k$  as  $t_k \cdot x_i$  by speculating about the trustworthiness of  $\varrho_k$  regarding this single opinion. That is, even though the trustworthiness of  $\varrho_k$  is  $t_k$  based on the existing evidence  $(r_k, s_k)$ , it becomes  $t_k \cdot x_i$  for this specific opinion  $w_i^k$ ; so,  $t_k \cdot x_i$  effectively becomes the trust in  $w_i^k$ . Here, we create a metric to measure how much we speculate about the trustworthiness of  $\varrho_k$  regarding  $w_i^k$ .

First, to decrease trust from  $t_k$  to  $t_k \cdot x_i$ , we need additional negative evidence, which is called *speculative evidence* and denoted as  $\rho_i$ . Our intuition is that it is less likely for a trustworthy agent to present additional negative *speculative evidence* than it is for an untrustworthy agent, and thus the receipt of such evidence should be tempered by  $(\bar{t}_k)^\kappa$ . Here,  $\bar{t}_k$  represents the *distrust* we have in agent  $\varrho_k$ ; i.e. the likelihood that we will receive additional negative evidence given our experiences with the source. The calibration constant  $\kappa \geq 0$  enables us to vary the influence that prior experience has on our prediction that an individual will present negative evidence in the future. If  $\kappa = 0$ , for example, we assume that all sources are equally likely to provide negative evidence. Now, using the Beta distribution formula for trust, we obtain:

$$\begin{aligned} t_k \cdot x_i &= \frac{r_k + 1}{r_k + s_k + 2} \cdot x_i = \frac{r_k + 1}{s_k + r_k + 2 + \rho_i \cdot (\bar{t}_k)^\kappa} \\ &= \frac{r_k + 1}{s_k + r_k + 2 + \rho_i \cdot \left(\frac{s_k + 1}{r_k + s_k + 2}\right)^\kappa} \end{aligned}$$

Rearranging this for  $\rho_i$  yields:

$$\rho_i = \frac{\nu_i(1 - x_i)}{x_i} \quad \text{where} \quad \nu_i = \frac{(r_k + s_k + 2)^{\kappa+1}}{(s_k + 1)^\kappa} \quad (1)$$

Given two conflicting subjective opinions  $w_i$  and  $w_j$ , there can be different additional discounting factors that can be used to resolve the conflict. Let us assume that  $x_i$  and  $x_j$  are additional discounting factors used to resolve the conflict. The cost of this resolution in terms of the total amount of speculative evidence can be computed as

$$\frac{\nu_i(1 - x_i)}{x_i} + \frac{\nu_j(1 - x_j)}{x_j}$$

where  $\nu_i$  and  $\nu_j$  are constants that are calculated as described in Equation 1. When we have multiple conflicts, they may interact in such a way that resolving one may also affect the resolution of another. The overall amount of *speculative evidence* necessary to resolve all of these relevant conflicts can easily be formulated as a function of additional discount factors. Once, we have this function, we can find the discounting factors to have a solution with the minimum total speculative evidence.

Assume we have a set of conflicting opinions  $\{\langle w_i, w_j \rangle, \dots, \langle w_m, w_n \rangle\}$  and, derived from trust evidence about agents, coefficients  $\{\nu_i, \nu_j, \dots, \nu_m, \nu_n\}$ . To determine the optimum discounting factors  $\{x_i, x_j, \dots, x_m, x_n\}$  for these opinions, we construct the following optimisation problem with a multivariate non-linear objective function and linear constraints.

$$\begin{aligned}
& \arg \min_{\vec{x}} f(\vec{x}) \quad \text{where} \\
f(\langle x_i, x_j, \dots, x_m, x_n \rangle) &= \frac{\nu_i(1-x_i)}{x_i} + \frac{\nu_j(1-x_j)}{x_j} + \dots \\
& \quad \frac{\nu_m(1-x_m)}{x_m} + \frac{\nu_n(1-x_n)}{x_n} \\
& \text{such that} \quad 0 \leq x_i \leq 1, 0 \leq x_j \leq 1, \dots \\
& \text{and} \quad 0 \leq c_{ij}x_i + c_{ji}x_j \leq 1, \dots
\end{aligned} \tag{2}$$

Existing constrained non-linear programming methods can be used to solve this problem in order to estimate the best discounting factors. There are various techniques that may be used including *Interior-Point* and *Active-Set* algorithms. In this work, we use *Interior-Point* approximation. Details of these methods are out of the scope of this paper and can be found elsewhere.<sup>9</sup>

In this section we have formalised the problem of computing additional discounting factors for *opinions* received about the world from other agent so that we may formulate a consistent *SDL-Lite* knowledge-base from which we can draw reliable conclusions. We have presented a number of approaches to the resolutions of inconsistencies between opinions including an optimisation-based approach, evidence-based discounting. Next, we evaluate these approaches with respect to their robustness in the face of liars.

## 6. EVALUATION

We have evaluated our approach through a set of simulations. In each simulation, we define the domain by randomly generating an *SDL-Lite* TBox that contains 100 concepts and roles, as well as axioms over those, e.g.,  $B_1 \sqsubseteq B_2$  and  $B_2 \sqsubseteq \neg \exists R_3$ . For each role or concept, there is one information source that provides opinions about its instances, e.g.,  $B_1(a):(0.8, 0, 0.2)$  and  $R_3(a, b):(0.5, 0.1, 0.4)$ . There are 10 information sources in total, each is an expert on 10 concepts and roles, and provides its opinions about those.

In our simulations, we assume there is one information consumer that uses the information from sources to make decisions. Each simulation is composed of 10 iterations. At each iteration  $t$ , the consumer needs to gather information about an individual  $a$ . We generate ground truth about  $a$ , which is composed of one assertion about  $a$  for each concept and role with an associated opinion. Each information source knows the ground truth only about the concepts and roles of their expertise. However, they may not provide the ground truth to the consumer when it is requested. Behaviours of the information sources are determined by their behavioural type, which are summarised as follows.

- **Honest:** Most of the time, this type of sources provide the ground truth about the assertion of their expertise with small Gaussian noise  $N(0, 0.01)$ . With probability  $P_b$ , honest sources behave like malicious ones and provide bogus information.
- **Malicious:** This type of sources aim at misleading the information consumer by providing bogus opinions. More specifically, given  $(b, d, \_)$  is the ground truth about an assertion, a malicious source provides the opinion  $(abs(\epsilon_1), 0.9 + \epsilon_2, \_)$  if  $b \approx d$ ; otherwise it provides the opinion  $(d + \epsilon_1, b + \epsilon_2, \_)$ , where  $\epsilon_1, \epsilon_2 \in [-0.05, 0.05]$ . There are two types of malicious sources, which are defined as follows:
  - i. **Simple liars:** they always provide bogus opinions.
  - ii. **Strategic liars:** they behave like honest sources to build trust and then provides bogus information exploiting the built trust. After providing misleading information to the consumer, they change their identity to avoid negative evidence against them.



After collecting opinions about different assertions from information sources, the information consumer uses its trust in these sources to discount these opinions and uses the proposed reasoning mechanisms for *SDL-Lite* to compute interpretations. Ideally, these interpretations should be close to the ground truth if all sources are accurate and their trustworthiness is modelled correctly. If there are some malicious sources, there may be conflicts in the collected information. In the case of conflicts, the consumer resolve the conflicts using *Naive Deleting* (NDL), *Trust-based Deleting* (TDL), *Trust-based Discounting* (TDC), or *Evidence-based Discounting* (EDC) with  $\kappa = 1$ . In NDL, all conflicting opinions are deleted from the knowledge base to resolve the conflicts. The consumer computes the interpretations for concept and role assertions related to  $\mathbf{a}$ , after resolving the conflicts if any. Then, we measure the performance as the *mean absolute error* in the computed interpretations. Let  $(b, d, u)$  be the ground truth and  $(b', d', u')$  be the computed interpretation for assertion  $B(\mathbf{a})$ , then the *absolute error* in the interpretation is computed as  $err_{B(\mathbf{a})} = abs(\delta_b) + abs(\delta_d)$ , where  $\delta_b = b' - b$  and  $\delta_d = d' - d$ . For instance, if the ground truth about  $B(\mathbf{a})$  is  $(0.9, 0.05, 0.05)$ , but the computed interpretation is  $(0.05, 0.9, 0.05)$ , then the error would be 1.7.

At the end of each iteration, the consumer learns the ground truth and updates the trustworthiness of the information sources with new evidence  $(r^t, s^t)$  computed as in Equation 3, which is based on the intuition that the information is still useful if it has a small amount of noise or is slightly discounted.

$$(r^t, s^t) = \begin{cases} (0, 1), & \text{if } \delta_b > 0.1 \text{ or } \delta_d > 0.1 \\ (1, 0), & \text{if } -0.1 \leq \delta_b \leq 0.01 \text{ and } -0.1 \leq \delta_d \leq 0.01 \\ (0, 0), & \text{otherwise.} \end{cases} \quad (3)$$

Each of our simulations are repeated 10 times and our results are significant based on *t-test* with a confidence interval of 0.95.

Without any evidence, the trustworthiness of sources is computed as 0.5. Thus, there are is no conflict in the beginning of our simulations. If all sources have deterministic behaviours, i.e., malicious sources are simple liars and  $P_b = 0$ , then trustworthiness of sources are easily modelled over time and the opinions from liars are significantly discounted. In such settings, conflicts are totally avoided and information consumers using either of the four proposed methods have the same level of success. Figure 2 shows an example of this setting where honest sources always provides the truth ( $P_b = 0$ ) and malicious sources are simple liars. Here, the *ratio of liars* ( $R_{liar}$ ) is 0.5, i.e., half of the sources are malicious.

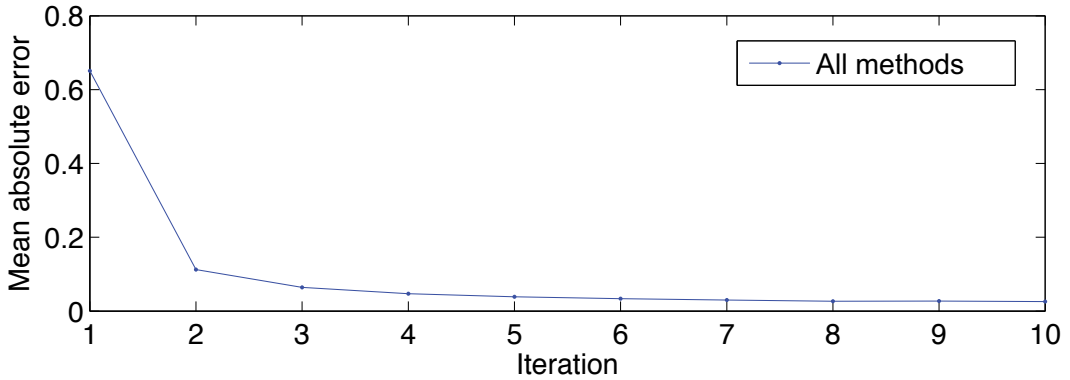


Figure 2. Simple liars ( $R_{liar} = 0.5$  and  $P_b = 0$ )

When honest sources provide bogus information occasionally, the conflicts may arise in the knowledge base of the consumer, because the information from these sources are not significantly discounted. Figure 3 shows our results for  $R_{liar} = 0.5$  and  $P_b = 0.1$ , where all malicious sources are simple liars. In this setting, NDL leads to significant errors in the computed interpretations. While TDL does much better than NDL, it is outperformed by discounting based approaches TDC and EDC. Both of these approaches have similarly good performance though TDC does slightly better.

Simple liars may not be enough to model malicious sources in real life. That is why we change the type of malicious sources to strategic liars and repeat our simulations. Figure 4 shows our results for  $R_{liar} = 0.5$  and  $P_b = 0.1$ . In this settings, trust evaluations become misleading, since strategic liars build trust, make their impact and then change their identity to avoid any negative evidence. As a result, as shown in the figure, TDC fails significantly more than EDC after a

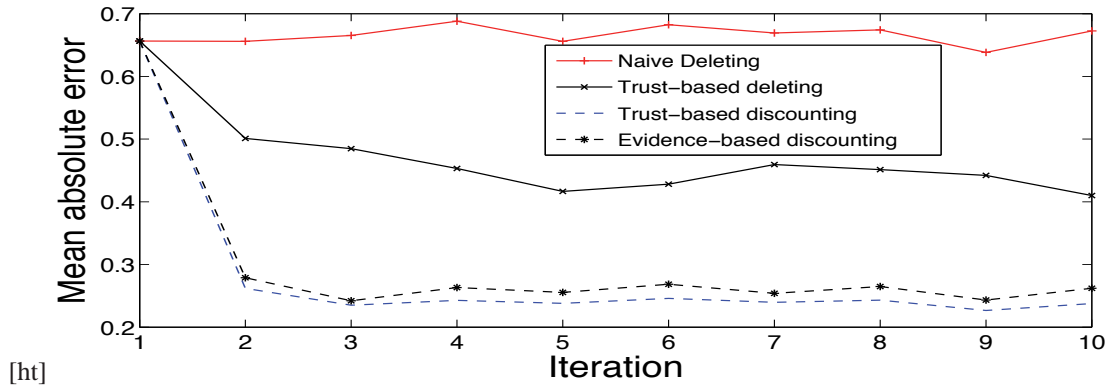


Figure 3. Simple liars ( $R_{liar} = 0.5$  and  $P_b = 0.1$ )

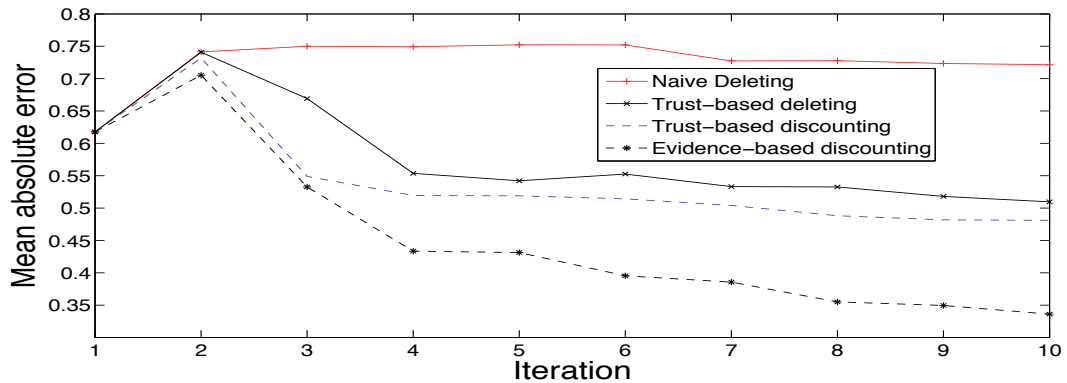


Figure 4. Strategic liars ( $R_{liar} = 0.5$  and  $P_b = 0.1$ )

few iterations. We repeat the simulations with strategic liars for different  $R_{liar}$  values; our results are shown in Figure 5. Our results indicate that evidence-based discounting is much more robust in the presence of realistic malicious behaviour than trust-based discounting or deletion.

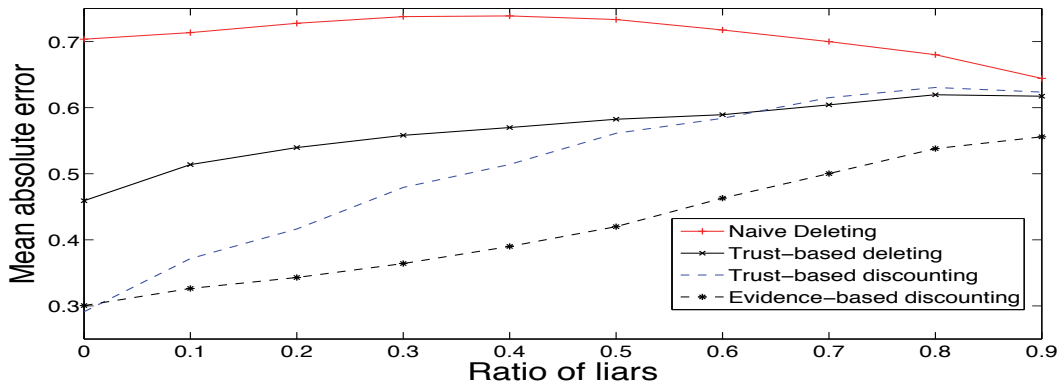


Figure 5. Strategic liars with varying  $R_{liar}$  ( $P_b = 0.1$ )

## 7. DISCUSSION

DL-Lite is a tractable subset of DLs with a large number of application areas.<sup>10</sup> Its scalability makes it very useful especially for the settings where large amount of data should be queried. However, in a network of heterogeneous sources, any information provided by the sources could be uncertain, incomplete, and even conflicting. DL-Lite cannot accommodate such information. Pan et al.<sup>11</sup> proposed a framework of tractable query answering algorithms for a family of fuzzy query

languages over large fuzzy DL-Lite<sup>12</sup> ontologies. On the other hand, DST and its extensions such as Subjective Logic explicitly takes into account *uncertainty* and *belief ownership*.<sup>5</sup>

Gobeck and Halaschek<sup>13</sup> present a belief revision algorithm for OWL-DL, which is based on trust degrees to remove conflicting statements from a knowledge base. However, as the authors point out, the proposed algorithm is not guaranteed to be optimal. In our work, we embed statement retraction implicitly into the opinion revision procedure with a global optimal criteria which is grounded on a Beta distribution formalisation of trust.

Fact-finding algorithms aim to identify the *truth* given conflicting claims. Pasternack and Roth<sup>14</sup> propose to translate these claims to a linear program, which is solved to obtain belief scores over claims. For example, with TruthFinder,<sup>15</sup> the belief scores obtained can be interpreted as the result of simultaneously minimising the frustration coming from the sources against the claims. These approaches do not consider semantics while reasoning about belief and trustworthiness as we do here.

There are several models for computing trust and reputation in multiagent systems. In these models, direct evidence is combined with indirect evidence to model trust in agents. Direct evidence is based on personal observations, while indirect evidence is received from other agents that serve as information sources. Jøsang and Ismail proposed the beta reputation system (BRS).<sup>6</sup> It estimates the likelihood of proposition “Agent *i* is trustworthy” – i.e., trustworthiness of the agent *i* – using beta probability density functions. For this purpose, aggregation of direct evidence and indirect evidence (i.e., ratings) from information sources are used as the parameters of beta distributions. Evidence shared by sources are equivalent to binary opinions in Subjective Logic.<sup>5</sup> Whitby *et al.* extended BRS to handle misleading opinions from malicious sources using a majority-based algorithm.<sup>16</sup> Teacy *et al.* proposed TRAVOS,<sup>17</sup> which is similar to BRS, but it uses personal observations about information sources to estimate their trustworthiness as we do in this paper.

In this paper, we describe conflicts between binomial opinions and propose an approach to resolve conflicts before performing fusion. Conflicts in knowledge lead to inconsistencies that hamper the reasoning over the knowledge. Therefore, before using such knowledge bases, their conflicts should be resolved. Gobeck and Halaschek<sup>13</sup> present a belief revision algorithm for ontologies, which is based on trust degrees of information sources to remove conflicting statements from a knowledge base. However, as the authors point out, the proposed algorithm is not guaranteed to be optimal. Dong *et al.*<sup>18</sup> propose to resolve conflicts in information from multiple sources by a voting mechanism. Double counting in votes is avoided by considering the information dependencies among sources. The dependences are derived from Bayesian analysis.

## ACKNOWLEDGMENTS

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defense or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute for Government purposes notwithstanding any copyright notation hereon.

## REFERENCES

1. G. Shafer, *A mathematical theory of evidence*, Princeton university press, 1976.
2. F. Baader, D. L. McGuinness, D. Nardi, and P. Patel-Schneider, eds., *Description Logic Handbook: Theory, implementation and applications*, Cambridge University Press, 2002.
3. D. Calvanese, G. D. Giacomo, M. Lenzerini, R. Rosati, and G. Vetere, “DL-Lite: Practical Reasoning for Rich DLs,” in *Proc. of the DL2004 Workshop*, 2004.
4. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati, “Data complexity of query answering in description logics,” in *Proc. of KR 2006*, pp. 260–270, 2006.
5. A. Jøsang, *Subjective Logic*, Book Draft, 2011.
6. A. Jøsang and R. Ismail, “The beta reputation system,” in *Proc. of the 15th Bled Electronic Commerce Conference e-Reality: Constructing the e-Economy*, pp. 48–64, 2002.
7. M. Sensoy, A. Fokoue, J. Z. Pan, and T. J. Norman, “Reasoning about uncertain information and conflict resolution through trust revision,” in *Proceedings of 12th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, p. to appear, 2013.

8. C. Burnett, T. J. Norman, and K. Sycara, "Bootstrapping trust evaluations through stereotypes," in *Proceedings of the Ninth International Conference on Autonomous Agents and Multiagent Systems*, pp. 241–248, 2010.
9. A. Ruszczyński, *Nonlinear optimization*, vol. 13, Princeton university press, 2011.
10. D. Calvanese, G. Giacomo, D. Lembo, M. Lenzerini, and R. Rosati, "Tractable reasoning and efficient query answering in description logics: The dl-lite family," *J. Autom. Reason.* **39**(3), pp. 385–429, 2007.
11. J. Z. Pan, G. Stamou, G. Stoilos, S. Taylor, and E. Thomas, "Scalable Querying Services over Fuzzy Ontologies," in *the Proc. of the 17th International World Wide Web Conference (WWW2008)*, 2008.
12. U. Straccia, "Answering vague queries in fuzzy DL-Lite," in *Proc. of the 11th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pp. 2238–2245, 2006.
13. J. Golbeck and C. Halaschek-Wiener, "Trust-based revision for expressive web syndication," *Journal of Logic and Computation* **19**, pp. 771–790, Oct. 2009.
14. J. Pasternak and D. Roth, "Knowing what to believe (when you already know something)," in *Proc. of the 23rd International Conference on Computational Linguistics*, (Beijing, China), 2010.
15. X. Yin, J. Han, and P. S. Yu, "Truth discovery with multiple conflicting information providers on the web," in *Proceedings of the Conference on Knowledge and Data Discovery*, 2007.
16. A. Whitby, A. Jøsang, and J. Indulska, "Filtering out unfair ratings in Bayesian reputation systems," *The Icfain Journal of Management Research* **4**(2), pp. 48–64, 2005.
17. W. Teacy, J. Patel, N. Jennings, and M. Luck, "TRAVOS: Trust and reputation in the context of inaccurate information sources," *Autonomous Agents and Multi-Agent Systems* **12**(2), pp. 183–198, 2006.
18. X. L. Dong, L. Berti-Equille, and D. Srivastava, "Integrating conflicting data: The role of source dependence," in *Proc. of the 35th International Conference on Very Large Databases*, (Lyon, France), August 2009.