Bayesian Knowledge-driven Ontologies

Intuitive Uncertainty Reasoning for Semantic Networks

Eugene Santos Jr., Jacob C. Jurmain
Thayer School of Engineering
Dartmouth College, Hanover, NH 03755
Email: {Eugene.Santos.Jr, Jacob.C.Jurmain}@Dartmouth.edu

***Abstract*—Uncertainty handling for semantic networks is a difficult problem which has slowed the effort to fully develop a semantic web. Uncertainty handling becomes particularly challenging when incompleteness is present in a domain, as it frequently is when modeling real-world complexity. To date, work on uncertainty frameworks for semantic networks has not intuitively captured a useful notion of uncertainty, for reasons including weaknesses in underlying uncertainty theories and assumption conflicts with semantic networks.**

We propose a framework which is a synthesis of semantic networks and Bayesian Knowledge Bases, which are a generalization of Bayesian Networks to accommodate incompleteness. This synthesis represents knowledge as "if-then" conditional probability rules between description logic assertions. We define simple methods for reasoning about semantic information under uncertainty and about uncertainty itself. Our results show potential to remove some obstructions in the path to a semantic web.

Keywords – Bayesian Knowledge-driven Ontologies; semantic networks; probability theory

I. INTRODUCTION

Semantic networks, also known as ontologies, are the very foundation of the semantic web, and are widely used in machine knowledge representation. They facilitate the definition and conceptual characterization of classes, individuals, and their relationships within a domain. Semantic networks can effectively define a networked encyclopedia of conceptual information in a domain, and their deductive reasoning mechanism facilitates algorithmic revelation of implicit, inferable knowledge in the model according to the rules of description logic [1] which is a decidable subset of predicate calculus.

Unfortunately, semantic networks do not conveniently represent uncertainty. Uncertainty often involves the existence of multiple conflicting possible instantiations (i.e. states) of a domain. Where uncertainty exists, we would typically like to have some understanding of the relative likelihoods of the possible instantiations, so that we can reduce or prioritize the set of instantiations we must consider. Such information is usually not known ahead of time, but can often be derived through some form of reasoning based on knowledge of the interactions between the assertions which differentiate the instantiations. But in order to reason about uncertainty, we must first represent it. Conventional semantic networks can only represent uncertainty ad-hocly by creating parallel models describing each possible instantiation individually. Of course, this approach does not capture interactions between the instantiations' assertions, so no formal reasoning can be performed. Furthermore, the number of parallel ontologies needed to completely define a domain is exponential with respect to the number of sets of differentiating assertions in the domain. So maintaining parallel ontologies is usually so labor-intensive as to be impracticable. What is needed is a systematic way to represent multiple variations on a domain.

Several theories of uncertainty exist which can introduce strong uncertainty semantics into description logic. Two prominent theories which have enjoyed success are fuzzy logic and possibility theory. These have been applied in frameworks such as Fuzzy OWL [17] and possibilistic description logic [10]. However, in both theories, some deep interactions between variables are lost during inferencing. The lost information is unnecessary for modeling the notions of fuzzy set membership and possibility, but as we will detail in Section II, we would prefer to capture a more complex notion of uncertainty which supports chains of "if-then" interactions between variables. One uncertainty theory which has strong semantics and fully captures these variable interactions is probability theory. Unfortunately, to the best of our knowledge all of the representation frameworks for semantic networks which are rooted in probability exhibit lossy reasoning or have unintuitive restrictions on their flexibility. The probabilistic description logics defined by Lukasiewicz [6] and based on Nilsson's probabilistic logic [8] experience decay in relative precision during reasoning due to their expression of probabilities as intervals. Approaches using Bayesian Networks [9], such as BayesOWL [4], MEBN/PR-OWL [3], and P-CLASSIC [5], contain an unintuitive conflict in their assumptions: Bayesian Networks require complete specification of the domain's probability distribution with no incompleteness, but semantic networks have a finer granularity which allows for incompleteness. Some domains with incompletely defined relationships can only be represented in BN-based frameworks by overdefining them. We address these issues in more detail in Section II.

There exists one probabilistic knowledge representation framework which may show better results when unified with semantic networks. Bayesian Knowledge Bases [12], or BKBs, are a generalization of Bayesian Networks designed specifically to handle incompleteness, and they do not experience reasoning decay like probabilistic logic. BKBs represent domain knowledge as sets of "if-then" conditional probability rules between propositional variable instantiations. They use those conditional probabilities to compute marginal probabilities of the domain's instantiations. BKBs also

Sponsors: ONR MURI Program, AFOSR, DTRA, and DHS

facilitate asking "what if" questions of the model by allowing the user to restrict the set of instantiations computed to ones where certain variable instantiations are set as evidence. BKBs represent knowledge with the same granularity as semantic networks, but they are not an immediate substitute for them because they only reason about propositional knowledge, not predicated knowledge like semantic networks do. A synthesis of BKBs and semantic networks which preserves the capabilities of both is desirable, but to our knowledge this paper is the first attempt.

We therefore propose the knowledge representation and reasoning framework called Bayesian Knowledge-driven Ontologies (BKOs). BKOs unite the predicate reasoning capabilities of semantic networks with the probabilistic reasoning capabilities of BKBs. They represent knowledge as description logic assertions like semantic networks, but also represent conditional probability rules between those assertions like BKBs. We will show that BKOs can validly reason about both types of knowledge without disrupting the other, based on two insights: First, that generalizing the rule of universal instantiation to its probabilistic analog allows description logic to validly handle uncertainty in its reasoning process. Second, that a "fully reasoned" semantic network, i.e. one in which all implicit propositional knowledge has been made explicit, naturally conforms to the semantics of BKBs. We will show that BKOs validly model uncertainty and incompleteness in both their representation and reasoning, possess strong probabilistic semantics that fully capture variable interactions, and do not vitiate the flexibility of the semantic network. We will show how BKOs may be used to answer the *probabilistic membership query*

$$P(a_{j_1} \in C_{j_1}, a_{j_2} \in C_{j_2}, \dots a_{j_m} \in C_{j_m} \mid a_{i_1} \in C_{i_1}, a_{i_2} \in C_{i_2}, \dots a_{i_n} \in C_{i_n})$$

the intuitive interpretation of which is: In the context of a given ontology, and given the knowledge that a set of individuals $\{a_{i_{1\dots n}}\}$ are each known to be members of respective classes $\{C_{i_{1\dots n}}\}$, what is the joint probability of the set of individuals $\{a_{j_{1\dots m}}\}$ each being members of respective classes $\{C_{j_{1\dots m}}\}$?

We begin in Section II with a brief survey of representative approaches to augmenting semantic networks with uncertainty reasoning. Next, Sections III and IV provide background on description logic theory and BKB theory. Section V develops BKOs' method of knowledge representation. In Section VI, we define the method of probabilistic predicate reasoning using the probabilistic rule of universal instantiation. We follow with Section VII which defines the method of computing marginal probabilities of domain instantiations. Finally, in Section VIII, we provide our concluding remarks and a look at future directions and potential applications for our theory.

II. Related Work

The simplest approach to introducing uncertainty to semantic networks is to list probabilities as attributes on classes. Some applications will use upper ontologies to define special predicates for storing probabilities. These methods are essentially annotation with probabilities. They provide no real probabilistic reasoning functionality, and to the best of our knowledge there is little to no published work on the subject.

Straccia [18] introduces fuzzy logic to semantic networks. Fuzzy logic is an uncertainty theory designed to represent the notion of ambiguity using partial set membership, such as the classic half-empty or half-full glass. Fuzzy logic's axioms are identical to probability theory, except that fuzzy logic lacks the axiom that the union of all outcomes sums to one. The absence of that axiom means that fuzzy logic's reasoning is a more coarse treatment of information interaction, using min and max functions in place of the arithmetic functions that probability theory would use. Consider the following example. (Notation: for an individual or class a, a class C, and $p \in [0,1]$, a∈C:p states that a has membership in C with degree p.) Given the assertions a∈C:0.7, a∈D:0.4, C∈E:0.2, and D∈E:0.6, what is the membership of a in E? In fuzzy set theory, this is simply max(min(0.7, 0.2), min(0.4,0.6)) = 0.4. Note that most of the information contained in this reasoning chain had no effect on the outcome except inasmuch as it was greater or less than 0.4. A change in the degree of membership of D in E would only alter the result if it dropped below 0.4, and a change in the degree of membership of a in C would not alter the result at all. This can be unintuitive when we consider modeling any notion of causality, since we typically think that a change in a root variable should somehow affect the result. Fuzzy logic is therefore more suited to its intended purpose of comparing entity descriptions than it is to capturing variable interactions.

Qi et al. [10] introduces possibility theory to semantic networks. Possibility theory models the notion of uncertainty of events, but like fuzzy logic it does not fully capture causal interactions. Possibility theory models the uncertainty of a single event with two numbers from the range [0,1]: the event's possibility, which is the degree to which the event could be expected to happen, and the event's necessity, which is the degree to which the event must happen. These numbers are related in that the necessity of an event is equal to one minus the possibility of the event's complement. Despite possibility theory's sophisticated uncertainty representation capability, its reasoning mechanism still does not intuitively capture causality. Consider the following example and note the parallels to the example we used for fuzzy logic: (Notation – for events A and C, and $p, q \in [0,1]$, p>q, C|A:(p,q) states that the possibility of C given A is p and the necessity of C given A is q.) Given the assertions C|A:(0.7,0.5), D|A:(0.4,0.3), E|C:(0.2,0.1), and E|D:(0.6,0.55), what is the possibility and necessity of E given A? The answer is simply that the possibility is max(min(0.7, 0.2), min(0.4,0.6)) = 0.4 and the necessity is max(min(0.5, 0.1), min(0.3,0.55)) = 0.3. As we discussed for fuzzy logic, this is a coarse treatment of causality.

Probability theory is ideally formulated to represent complex causal interactions. We assume that the reader is familiar with the formulation and reasoning mechanics of probability theory, such as the notions of sample spaces, probability distributions, and conditional probabilities. Let us now narrow our focus to frameworks founded in probability theory. Two groups of frameworks have been widely studied: those founded in Nilsson's probabilistic logic [8], and those founded in Bayesian Networks [9].

Regarding Nilsson's probabilistic logic-based frameworks, such as [6], we see the difficulty they encounter in the following example. Recall that assertions in probabilistic DL are made probabilistic not by assigning them a probability, but by declaring an interval in which that probability is said to be found. This interval-based definition causes erosion of relative precision (measured as the width of the probability's interval divided by its average) with every calculation. Suppose we have two probabilistic axioms, "Tweety is-a Bird" with probability between 0.70 and 0.80 (relative precision 0.13), and "Birds can Fly" with probability between 0.90 and 0.99 (relative precision 0.10). We wish to find the marginal probability that "Tweety can Fly". Since the probabilities are only known as intervals, we must multiply their bounds to get the extreme cases of the marginal probability. The lowest possible probability is 0.9 * 0.7 = 0.63 and the highest possible probability is 0.8 * 0.99 = 0.79, so the marginal probability on "Tweety can Fly" is within the interval [0.63, 0.79]. Notice that this interval has a relative precision of 0.23, wider than either of the relative precisions on the original axioms. The representation of probabilities as intervals is an artifact of probabilistic DL's foundation in Nilsson's probabilistic logic [8], which is subject to the same decay in precision.

Regarding Bayesian Network-based approaches, such as PR-OWL [3], BayesOWL [4], and P-CLASSIC [5], consider the notion of incompleteness in a domain. Incompleteness is the under-definition of a domain's uncertainty, i.e., the domain's probability distribution could match one of a number of possible probability mass functions. Recall that BNs assume completeness by assuming that all variables whose joint distributions are not completely known are independent. Semantic networks do not share this completeness assumption, so there are incomplete domains which can be represented with conventional semantic networks (sans uncertainty handling) but which can only be expressed with BN-based frameworks by making up unsupported and potentially inaccurate constraints. Furthermore, we find notions which can be represented in semantic networks that are unintuitive when we try to express them in BNs even with complete information. For example, if we wanted to describe the probability distribution between the variable "airplane model" and a discretized "gas mileage" variable, the distribution would become unintuitive when we tried to define probabilities for the gas mileage of an engineless glider model. Even the notion of context-specific independence [2] does not avoid this problem because it would still require the "gas mileage" variable to have some distribution given a "glider model" value, but any distribution, even independence, is unintuitive. Disregarding uncertainty, a semantic network would have no trouble expressing this domain's concepts , because it could simply omit the glider's gas mileage property from any consideration.

III. BACKGROUND – DESCRIPTION LOGIC

We will briefly introduce a simple description logic with definitions and notation based on set theory. The definitions are conceptually equivalent to formal description logic as presented by Baader et al. (2007) [1], but simplified for accessibility and more closely founded in set theory to simplify our derivations in Sections V-VII. Our simplification is to ignore the possibility of mapping ontologies to multiple interpretations, and instead to just consider classes and individuals as simple sets under a single interpretation. We believe that generalizing to multiple interpretations is possible and we will explore further in future work.

The fundamental concept of description logic is the *class* (also *concept*), which is a set. An *individual* is an element of a class. A *role* is a binary operator acting from one individual (the owner) to another individual (the filler). Classes, individuals, and roles generally have real world interpretations as categories, objects, and relationships between objects.

While the words "class" and "concept" are for the most part interchangeable in description logic, "class" generally refers to a more set-theoretic notion of classes/concepts as groups of individuals, while "concept" is used in the context of the descriptive nature of classes/concepts, i.e. that they characterize the nature of the individuals in them. We will mostly use "class" to emphasize the set-theoretic foundation of our theory.

Atomic classes are irreducible. They may be used in expressions called *constructors* to inductively define new classes, called *constructed classes*. The permitted expressiveness of constructors is specific to the particular description logic being used. Simple construction operators are complement, union, intersection, role existential quantification, and role value restriction; additional operators are defined in more expressive description logics. In general, the more expressive a description logic is, the longer its reasoning takes and the greater the risk of it being able to express undecidable problems. Ensuring decidability while achieving maximum expressivity is a pervasive issue in description logic research.

Description logic makes the *open world assumption*: that the absence of a particular statement within a description of a domain does not imply that statement's falsehood (or truth). The implication is that any description is incomplete, because we can always add new individuals, classes and rules to it. Here lies an important and subtle distinction: the open world assumption does not imply that every domain is *necessarily* infinite, but does imply that every domain is *possibly* infinite, i.e. cannot be proven finite. For practical purposes we will assume than any description of a domain is finite, but admit the possibility that the domain which it describes is infinite.

A. Notation

$\top$ (down tack character, not the letter) is the universal class, i.e., the class that includes all individuals. Because $\top$ contains all individuals, it necessarily contains all nonempty classes. $\bot$ is the empty class, i.e., the class that contains no individuals.

The complement of a class C, $\neg\mathrm{C}$, is $\top - \mathrm{C}$.

B. Asserting Knowledge

In description logic, knowledge is expressed through assertional axioms and terminological axioms. Assertional axioms are propositional: they characterize a single individual's membership in classes. Terminological axioms are predicated: they define general rules applying to any and all qualified individuals. The set of assertional and terminological axioms in an ontology are often referred to as the Abox and the Tbox, respectively.

Definition 3.1: An *assertional axiom* is a class assertion or a role assertion.

- A *class assertion* declares that $a \in C$ for a class expression C and an individual a. Description logic commonly uses the notation $\mathrm{C}(\mathrm{a})$.

- A *role assertion* declares that $b\ R\ c$ for a role expression R and individuals b and c. b R c states that c is a filler of the role R for an owner b. Description logic commonly uses the notation R(b,c).

Definition 3.2: $b\,R\,c \Leftrightarrow b \in R_c \Leftrightarrow c \in R_b^{-1}$.

Intuitively, the role assertion b R c and the concept assertions $b \in R_c$ (i.e. b is a member of the class defined by having the role R with filler c) and $c \in R_b^{-1}$ (i.e. c is a member of the class defined by being the filler of role R with owner c) all imply each other.

Definition 3.3: A *terminological axiom* is a statement asserting a relation between two classes. Standard forms in description logic are subsumption, equivalence, and disjointness axioms.

- A subsumption axiom is of the form $\mathrm{C} \sqsubseteq \mathrm{D}$, for any classes or constructors C and D. $\mathrm{C} \sqsubseteq \mathrm{D}$ states that all elements of C are also elements of D.
- An equivalence axiom is of the form $\mathrm{C} = \mathrm{D}$, for any classes or constructors C and D. $\mathrm{C} = \mathrm{D}$ states that all elements of C are elements of D, and all elements of D are elements of C, i.e. $\mathrm{C} = \mathrm{D} \Leftrightarrow \mathrm{C} \sqsubseteq \mathrm{D}$ and $\mathrm{D} \sqsubseteq \mathrm{C}$.
- A disjointness axiom is of the form $\mathrm{C} \sqcap \mathrm{D} = \bot$, for any arbitrary concept expressions C and D. $\mathrm{C} \sqcap \mathrm{D} = \bot$ states that there are no elements which belong to both C and D.

In some ontology languages, such as the variants of OWL, knowledge can be presented and used in the form of property characteristics [7], which define specific inference rules for instantiations of properties such as functionality, transitivity, and symmetricality. This expressive capability is fairly ad-hoc. In this paper we will only consider formal description logics, and therefore only property characteristics which can be directly expressed in them. In the future, property characteristics' analogs in BKO theory would probably be implemented as informal construction tools in an editor rather than formally in BKO theory, but we intend to investigate their formal inclusion in the theory as custom inference rules.

C. Reasoning

Terminological axioms are expressed as predicated statements, but used in arguments, they can be used to derive new assertional axioms. These new assertional axioms can then be used in new arguments, revealing yet more axioms. Long chains of reasoning can form in this way. These arguments hinge on the rule of universal instantiation, which simply states if something is true in general for all individuals in a class, it is true for each specific individual in that class. For our purposes we express the rule of universal instantiation as the basic property of set theory: if $\mathrm{C} \sqsubseteq \mathrm{D}$ and $\mathrm{a} \in \mathrm{C}$, infer $\mathrm{a} \in \mathrm{D}$. If $\mathrm{C} \sqcap \mathrm{D} = \bot$ and $\mathrm{a} \in \mathrm{C}$, infer $\mathrm{a} \notin \mathrm{D}$.

IV. Background – Bayesian Knowledge Bases

Bayesian Knowledge Bases [12] (abbrev. BKBs) are a generalization of Bayesian Networks to admit incompleteness in the probability distribution. BKBs model probabilistic knowledge in an intuitive "if-then" rule structure which quantifies dependencies between states of random variables. Reasoning with BKBs is performed as belief updating, which computes the posterior probability of a target variable state, belief revision, which computes the posterior probabilities of domain instantiations, and partial belief revision, which computes the posterior probabilities of sets of target variable states. BKBs excel at modeling causal and correlative information because they inherently provide backtrackable explanations of simulation outcomes [15]. They see typical use on problems of human intent modeling, such as war gaming [11], predicting outcomes of strategic actions [14], and explanatory analysis of complex events [15].

There are two equivalent formulations of BKB theory. One, presented in Santos et al. (2003) [13], is founded on the notion of a BKB as a set of conditional probability rules (abbrev. CPRs) and the other, presented in Santos et al. (1999) [12], on the notion of a BKB as a directed acyclic graph. We present a condensed version of the CPR-based formulation from Santos et al. (2003) [13]. The notation is slightly modified but expresses equivalent concepts.

Definition 4.1 [Santos et al. (2003) Def. 2.1]: Let $A_1 \ldots A_n$ be a collection of finite discrete random variables (abbrev. rvs) where $r(A_i)$ denotes the set of possible values for A_i. A conditional probability rule (CPR) is a statement of the form

$$\mathrm{P}\big(\mathrm{A}_{\mathrm{i}_\mathrm{n}} = \mathrm{a}_{\mathrm{i}_\mathrm{n}} \,|\, \mathrm{A}_{\mathrm{i}_1} = \mathrm{a}_{\mathrm{i}_1} \wedge \mathrm{A}_{\mathrm{i}_2} = \mathrm{a}_{\mathrm{i}_2} \wedge \ldots \wedge \mathrm{A}_{\mathrm{i}_{\mathrm{n}-1}} = \mathrm{a}_{\mathrm{i}_{\mathrm{n}-1}}\big) = \mathrm{p}$$

for some positive integer n where $\mathrm{a}_{\mathrm{i}_\mathrm{j}} \in \mathrm{r}(\mathrm{A}_{\mathrm{i}_\mathrm{j}})$ such that $i_j \neq i_k$ for all $j \neq k$ and p is a weight between [0,1].

A CPR R's antecedent, denoted ant(R), is the conjunction of rv assignments to the right of the vertical bar. R's consequent, denoted con(R), is the rv assignment to the left of the vertical bar. R states that given the consequent, the antecedent holds with probability p. Each rv assignment in the antecedent is called an *immediate ancestor* of the consequent, and the consequent is called an *immediate descendant* of the rv assignments in the antecedent. We can define this recursively over a set of CPRs for *ancestor* and *descendant*.

Definition 4.2 [Santos et al. (2003) Defs. 2.2 and 2.6]: Given two CPRs

$$\mathrm{R}_1\colon \mathrm{P}\big(\mathrm{A}_{\mathrm{i}_\mathrm{n}} = \mathrm{a}_{\mathrm{i}_\mathrm{n}} \,|\, \mathrm{A}_{\mathrm{i}_1} = \mathrm{a}_{\mathrm{i}_1} \wedge \ldots \wedge \mathrm{A}_{\mathrm{i}_{\mathrm{n}-1}} = \mathrm{a}_{\mathrm{i}_{\mathrm{n}-1}}\big) = \mathrm{p}_1,$$
$$\mathrm{R}_2\colon \mathrm{P}\big(\mathrm{A}_{\mathrm{j}_\mathrm{m}} = \mathrm{a'}_{\mathrm{j}_\mathrm{m}} \,|\, \mathrm{A}_{\mathrm{j}_1} = \mathrm{a'}_{\mathrm{j}_1} \wedge \ldots \wedge \mathrm{A}_{\mathrm{j}_{\mathrm{m}-1}} = \mathrm{a'}_{\mathrm{j}_{\mathrm{m}-1}}\big) = \mathrm{p}_2,$$

we say that R_1 and R_2 are *mutually exclusive* if there exists some $1 \le k < n$ and $1 \le l < m$ such that $i_k = j_l$ and $\mathrm{a}_{\mathrm{i}_\mathrm{k}} \neq \mathrm{a'}_{\mathrm{j}_\mathrm{l}}$. Otherwise, we say they are *compatible*.

Intuitively, Definition 4.2 states that mutually exclusive CPRs' antecedents are not simultaneously satisfiable because they are conditioned on different values of the same rv.

Definition 4.3 [Santos et al. (2003) Def. 2.3]: R_1 and R_2 are consequent-bound if (1) for all $k < n$ and $l < m$, $\mathrm{a}_{\mathrm{i}_\mathrm{k}} = \mathrm{a'}_{\mathrm{j}_\mathrm{l}}$ whenever $i_k = j_l$, and (2) $i_n = j_m$ but $\mathrm{a}_{\mathrm{i}_\mathrm{n}} \neq \mathrm{a'}_{\mathrm{j}_\mathrm{m}}$.

Intuitively, Definition 4.3 states that consequent-bound CPRs only conflict in their consequent. Their antecedents are simultaneously satisfiable and their consequents assign different values to the same rv.

Definition 4.4 [Santos et al. (2003) Def. 2.4]: A Bayesian Knowledge Base B is a finite set of CPRs such that

- for any distinct R_1 and R_2 in B, either (1) R_1 is mutually exclusive with R_2 or (2) $\mathrm{con}(R_1) \neq \mathrm{con}(R_2)$, and
- for any subset S of mutually consequent-bound CPRs of B, $\sum_{R \in S} P(R) \leq 1$.

Definitions 4.5-7 establish the concept of inferences, which are the basis of BKBs' expression of probability distributions.

Definition 4.5 [Santos et al. (2003) Def. 2.5]: A subset S of B is said to be a *deductive set* if for each CPR R in S where

$$R: P(A_{i_n} = a_{i_n} \mid A_{i_1} = a_{i_1} \wedge \ldots \wedge A_{i_{n-1}} = a_{i_{n-1}}) = p$$

the following two conditions hold:

- For each k = 1,…, n-1 there exists a CPR R_k in S such that $\mathrm{con}(R_k) = \{A_{i_k} = a_{i_k}\}$.
- There does not exist some R' ∈ S where R' ≠ R and con(R') = con(R).

Intuitively, the first condition establishes that each CPR's antecedents are supported by the consequents of other CPRs. The second condition requires that each rv assignment be supported by a unique chain.

Definition 4.6 [Santos et al. (2003) Def. 2.7]: A deductive set I is said to be an *inference* over B if I consists of mutually compatible CPRs and no rv assignment is an ancestor of itself in I. The set of rv assignments induced by I is denoted V(I). The probability of I is defined to be $P(I) = \prod_{R \in I} P(R)$.

Definition 4.7 [Santos et al. (2003) Def. 2.8]: Two inferences are compatible if all their CPRs are mutually compatible.

The following theorems establish that inferences can define a partial joint probability distribution.

Theorem 4.1 [Santos et al. (2003) Theorem 2.2]: For each set of rv assignments V, there exists at most one inference I over B such that V = V(I).

Theorem 4.2 [Santos et al. (2003) Theorem 2.3]: For any set of mutually incompatible inferences Y in B, $\sum_{I \in Y} P(I) \leq 1$.

Theorem 4.3 [Santos et al. (2003) Theorem 2.4]: Let I_0 be some inference. For any set of mutually incompatible inferences $Y(I_0)$ such that for all $I \in Y(I_0)$, $I_0 \subseteq I$, $\sum_{I \in Y(I_0)} P(I) \leq P(I_0)$

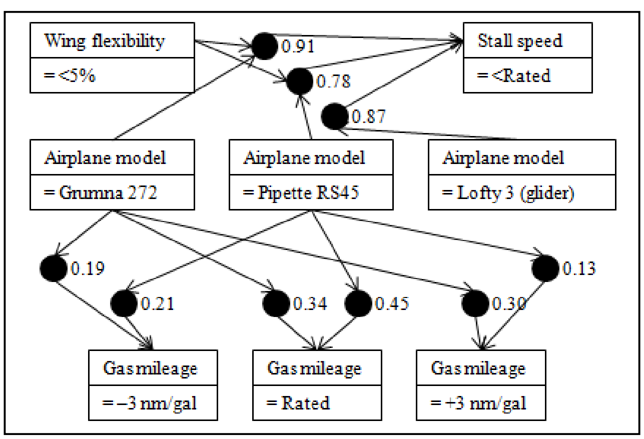


Figure 1. A hypothetical BKB fragment represented as a directed graph.

Fig. 1 above presents a graphical representation of an example BKB. Each black node (S-node) represents the weight of a CPR. Its parent white nodes (I-nodes) represent that CPR's antecedent, and its child node represents that CPR's consequent.

V. Bayesian Knowledge-driven Ontologies

For a domain description consisting of all of that domain's known individuals and classes, an instantiation of that domain is an assignment of the known individuals to the known classes. An individual may be assigned to one or more than one class, and a class may be assigned any number of individuals. A BKO models a probability distribution over all of a domain's possible instantiations. BKO theory uses if-then rules over the distribution's event space to restrict the distribution's probability mass function (abbrev. pmf). We describe these rules as "restrictions" on the distribution's pmf. BKO theory supports incompleteness, i.e. it does not require complete definition of the pmf, so a valid BKO may be compatible with more than one pmf.

To formulate this theory, we first define the nature of the distribution and its variables. We then define expressions used to restrict probability mass functions over that sample space. Finally, we define the nature of a BKO as a knowledge base.

A. Domain State Distributions

Definition 5.1: For a domain description consisting of a finite set of individuals $\{a_1 \ldots a_m\}$ and a finite set of distinct atomic classes $\{C_1 \ldots C_n\}$ and their complements $\{\neg C_1 \ldots \neg C_n\}$, define the domain's *state distribution* as a discrete multivariate probability distribution over the variables (referred to as *atomic variables*) V_{ij} whose states are $r(V_{ij}) = \{a_j \in C_i,\ a_j \in \neg C_i\}$. The sample space of the state distribution is defined as the following cross-product:

$$\Omega = \prod_{j=1}^{m} \prod_{i=1}^{n} r(V_{ij}) = \prod_{j=1}^{m} \prod_{i=1}^{n} \{a_j \in C_i, a_j \in \neg C_i\}$$

Intuitively, the sample space consists of all possible combinations of complete descriptions of each individual's memberships in classes. In the context of domain instantiations as discussed in the introduction, each outcome of the sample space is one possible domain instantiation.

Atomic classes and variables are insufficient for most reasoning tasks. We can define new constructed variables over the sample space which can make use of constructed classes or recombine atomic classes.

Notation: A set of classes $\{C_1 \ldots C_n\}$ is said to *span* a class D if $\cup \{C_1 \ldots C_n\} = D$. $\{C_1 \ldots C_n\}$ is said to be *world-spanning* if $\cup \{C_1 \ldots C_n\} = \top$.

Definition 5.2: Let Q be a domain containing an individual, a, and a world-spanning set of constructed or atomic classes $\{C_1 \ldots C_n\}$, and let the domain have a state distribution with sample space Ω. Then the set $\{C_1 \ldots C_n\}$'s *constructed variable* is a variable V over Ω such that $r(V) = \{a \in C_1, \ldots, a \in C_n\}$.

Example 5.1: Consider a domain describing a monochromatic ball. The ball can have one of three colors: red, blue, or green.

These three colors and their complements are atomic classes, so the state distribution is multivariate over three variables with instantiations {"the ball is red", "the ball is not red"}, {"the ball is blue", "the ball is not blue"}, {"the ball is green", "the ball is not green"}. We know the ball is monochromatic, so the set of classes {red, blue green} is world-spanning and we can define a new variable with states {"the ball is red", "the ball is blue", "the ball is green"}.

B. Restricting the Probability Mass Function

In BKO theory, the pmf is restricted using probabilistic assertional axioms and probabilistic terminological axioms. Probabilistic assertional axioms are probabilistically propositional, i.e. they characterize a single individual's conditional probability of membership in a class. Probabilistic terminological axioms are probabilistically predicated, i.e. they define general rules which characterize the class membership distributions of any and all qualified individuals. We will refer to both probabilistic and classical (i.e. non-probabilistic) axioms throughout the paper.

Definition 5.3: Let Ω be the sample space of a domain's state distribution. A *probabilistic assertional axiom* (PAA) is a conditional probability rule of the form

$$\mathrm{P}\big(\mathrm{V}_{\mathrm{i}_n} = \{\mathrm{a}_{\mathrm{i}_n} \in \mathrm{B}_{\mathrm{i}_n}\} \,|\, \mathrm{V}_{\mathrm{i}_1} = \{\mathrm{a}_{\mathrm{i}_1} \in \mathrm{B}_{\mathrm{i}_1}\} \wedge \ldots \wedge \mathrm{V}_{\mathrm{i}_{n-1}} = \{\mathrm{a}_{\mathrm{i}_{n-1}} \in \mathrm{B}_{\mathrm{i}_{n-1}}\}\big) = \mathrm{p}$$

where each V_k is a variable over Ω and $\{\mathrm{a}_{\mathrm{i}_k} \in \mathrm{B}_{\mathrm{i}_k}\} \in \mathrm{r}(\mathrm{V}_{\mathrm{i}_k})$.

Notation: The PAA in Definition 5.3 above can be expressed equivalently using the following shorthand:

$$\mathrm{P}\big(\mathrm{a}_{\mathrm{i}_n} \in \mathrm{B}_{\mathrm{i}_n} \,|\, \mathrm{a}_{\mathrm{i}_1} \in \mathrm{B}_{\mathrm{i}_1} \wedge \ldots \wedge \mathrm{a}_{\mathrm{i}_{n-1}} \in \mathrm{B}_{\mathrm{i}_{n-1}}\big) = \mathrm{p}.$$

This notation works because which variable each value belongs to is implicit in the value's name. Variables' names carry no real meaning in a BKO. In effect they exist merely to partition axioms, and that partitioning is derived directly from the axioms. The exception is when a constructed variable shares a value with another variable, such as in Example 5.1, but this notation still holds because the values held in common between the variables are mutually implicative, i.e. they must either be assigned to both variables or neither.

Definition 5.4: A *probabilistic terminological axiom* (PTA), T, is a statement of the form $\mathrm{P}(\mathrm{x} \in \mathrm{D} \,|\, \text{any } \mathrm{x} \in \mathrm{C}) = \mathrm{p}$ for any classes C and D where $\mathrm{p} \in [0,1]$ is the probability that any given individual x known to be a member of class C is also a member of class D.

T's antecedent, denoted ant(T), is the condition on x to the right of the vertical bar. T's consequent, denoted con(T), is the class membership assignment to the left of the vertical bar. In this case ant(T) = any $\mathrm{x} \in \mathrm{C}$ and con(T) = $\mathrm{x} \in \mathrm{D}$.

Definition 5.5: The instantiation of a PTA $\mathrm{T}\colon \mathrm{P}(\mathrm{x} \in \mathrm{D} \,|\, \text{any } \mathrm{x} \in \mathrm{C}) = \mathrm{p}$ for an individual a is defined as $\mathrm{T}|_{\mathrm{a}}\colon \mathrm{P}(\mathrm{a} \in \mathrm{D} \,|\, \mathrm{a} \in \mathrm{C}) = \mathrm{p}$.

PTAs are not conditional probability rules as defined in BKB theory. PTAs do describe a restriction on the pmf, but that restriction is not immediately applicable. Rather, PTAs are predicated rules which may be used to infer additional propositions (see Section 6) and those inferred propositions are then applied to directly restrict the pmf. Note also that all classical axioms can be represented as PAAs or PTAs:

Definition 5.6: A classical assertional axiom Z is equivalent to the PAA P(Z) = 1. A classical subsumption axiom $\mathrm{C} \sqsubseteq \mathrm{D}$ is equivalent to the PTA $\mathrm{P}(\mathrm{x} \in \mathrm{D} \,|\, \text{any } \mathrm{x} \in \mathrm{C}) = 1$. A classical equivalence axiom $\mathrm{C} = \mathrm{D}$ is equivalent to the set of PTAs $\{\mathrm{P}(\mathrm{x} \in \mathrm{D} \,|\, \text{any } \mathrm{x} \in \mathrm{C}) = 1, \mathrm{P}(\mathrm{x} \in \mathrm{C} \,|\, \text{any } \mathrm{x} \in \mathrm{D}) = 1\}$. A classical disjointness axiom $\mathrm{C} \cap \mathrm{D} = \perp$ is equivalent to the PTAs $\mathrm{P}(\mathrm{x} \in \mathrm{D} \,|\, \text{any } \mathrm{x} \in \mathrm{C}) = 0$ or $\mathrm{P}(\mathrm{x} \in \mathrm{C} \,|\, \text{any } \mathrm{x} \in \mathrm{D}) = 0$.

Proposition 5.1:

$$\mathrm{P}(\mathrm{x} \in \mathrm{D} \,|\, \text{any } \mathrm{x} \in \mathrm{C}) = 0 \Leftrightarrow \mathrm{P}(\mathrm{x} \in \mathrm{C} \,|\, \text{any } \mathrm{x} \in \mathrm{D}) = 0$$

C. Definition of a Bayesian Knowledge-driven Ontology

To define a BKO as a knowledge base, we first extend the definitions of mutual exclusivity (Definition 4.2) and consequent-boundedness (Definition 4.3) to PTAs. We make PTAs relatable to PAAs by considering PTAs as the set of all their possible instantiations and requiring mutual exclusivity or consequent-boundedness of all of those instantiations.

Let $\mathrm{T}_1\colon \mathrm{P}(\mathrm{x} \in \mathrm{D}_\mathrm{i} \,|\, \text{any } \mathrm{x} \in \mathrm{C}_\mathrm{i}) = \mathrm{p}_1$

and $\mathrm{T}_2\colon \mathrm{P}\big(\mathrm{x} \in \mathrm{D}_\mathrm{j} \,\big|\, \text{any } \mathrm{x} \in \mathrm{C}_\mathrm{j}\big) = \mathrm{p}_2$

be two PTAs.

Definition 5.7: T_1 and T_2 are mutually exclusive if $\mathrm{C}_\mathrm{i} \cap \mathrm{C}_\mathrm{j} = \perp$

Intuitively, Definition 5.8 is analogous to Definition 4.2. Mutually exclusive PTAs will never apply to the same individual in the same domain instantiation because that individual cannot be in both C_i and C_j.

Definition 5.8: T_1 and T_2 are consequent-bound if $\mathrm{C}_\mathrm{i} \cap \mathrm{C}_\mathrm{j} \neq \perp$ and $\mathrm{D}_\mathrm{i} \cap \mathrm{D}_\mathrm{j} = \perp$.

Intuitively, Definition 5.9 is analogous to Definition 4.3. Consequent-bound PTAs only conflict in their consequent. Their antecedents are simultaneously satisfiable and their consequents assign different values to the same rvs.

Let $\mathrm{R}\colon \mathrm{P}\big(\mathrm{a}_{\mathrm{i}_n} \in \mathrm{B}_{\mathrm{i}_n} \,|\, \mathrm{a}_{\mathrm{i}_1} \in \mathrm{B}_{\mathrm{i}_1} \wedge \ldots \wedge \mathrm{a}_{\mathrm{i}_{n-1}} \in \mathrm{B}_{\mathrm{i}_{n-1}}\big) = \mathrm{p}$ be some PAA and $\mathrm{T}\colon \mathrm{P}(\mathrm{x} \in \mathrm{D} \,|\, \text{any } \mathrm{x} \in \mathrm{C}) = \mathrm{p}$ be some PTA.

Definition 5.9: R and $\mathrm{T}|_{\mathrm{a}_{\mathrm{i}_n}}$ are mutually exclusive if there exists some $1 \leq k < n$ such that $\mathrm{a}_{\mathrm{i}_k} = \mathrm{a}_{\mathrm{i}_n}$ and $\mathrm{B}_{\mathrm{i}_k} \cap \mathrm{C} = \perp$.

Intuitively, Definition 5.9 is also analogous to Definition 4.2. R and T cannot be mutually exclusive unless con(R) is conditional on every individual in the domain not being in C, which is only possible if $\mathrm{C} = \perp$. For R and $\mathrm{T}|_{\mathrm{a}_{\mathrm{i}_n}}$ to be mutually exclusive, ant(R) must contain a condition that $\mathrm{a}_{\mathrm{i}_n}$ not be in C.

Definition 5.10: R and T are consequent-bound if $\mathrm{B}_{\mathrm{i}_k} \cap \mathrm{C} \neq \perp$ for all $1 \leq k < n$ and $\mathrm{B}_{\mathrm{i}_n} \cap \mathrm{D} = \perp$.

Intuitively, Definition 5.11 is also analogous to Definition 4.3. For R and T to be consequent-bound, ant(R) must be simultaneously satisfiable with all of T's possible instantiations' antecedents, and con(R) must assign a different value to the same rv as $\mathrm{con}(\mathrm{T}|_{\mathrm{a}_{\mathrm{i}_n}})$.

Definition 5.11: A Bayesian Knowledge-driven Ontology is a finite set of PAAs and PTAs such that

- for any distinct PTAs T_1 and T_2 in B, either (1) T_1 is mutually exclusive with T_2 or (2) $con(T_1) \neq con(T_2)$, and
- for any distinct PAAs R_1 and R_2 in B, either (1) R_1 is mutually exclusive with R_2 or (2) $con(R_1) \neq con(R_2)$, and
- for any PAA R and PTA T in B such that $con(R) = a \in C$, either (1) R is mutually exclusive with $T|_a$ or (2) $con(R) \neq con(T|_a)$, and
- for any subset S of mutually consequent-bound PAAs and/or PTAs of B, $\sum_{R \in S} P(R) \leq 1$.

Compare Definition 5.11 to Definition 4.4, the definition of a BKB. Intuitively, the definition of a BKO ensures that all PAAs and all PTA instantiations, taken as a set of CPRs, could form a BKB. The mapping is implemented in Section VII.

Proposition 5.2: Any subset of a BKO is also a BKO.

Example 5.2: (This is a running example that continues in Sections VI – VII.) Here we introduce a simple BKO about two fish, "Tuna" and "Herring". Normal text indicates the entity is an individual or class, *italics* indicates the entity is a relational operator, and **bold** indicates the entity is a DL function.

(1) Tuna, Herring ∈ Fish
(2) P(Tuna *ate* Herring) = 0.99
(3) P(x ∈ Predator | any x ∈ $ate_{(\mathbf{some}\ \mathrm{Fish})}$) = 0.9
(4) P(x ∈ Has_Parasites | any x ∈ Fish ∩ Predator) = 0.3

VI. The Predicate Reasoning Method for BKOs

PTAs can be used to infer new PAAs analogously to classical terminological inferencing through the probabilistic rule of universal instantiation. We prove below that once all such inferences have been made and incorporated into the BKO, the BKO is equivalent to a BKB and is ready to be reasoned over using BKB theory's methods.

Definition 6.1: The *probabilistic rule of universal instantiation* states the following: For a domain containing an individual a and a probabilistic terminological axiom $T: P(x \in D \mid \text{any } x \in C) = p$, infer $T|_a$.

BKO predicate reasoning is simply the inferring of all possible PTA instantiations for a domain using the probabilistic rule of universal instantiation. Any PTA can be instantiated for any individual, but in practice, instantiating all PTAs for every individual will create many PAAs with incomplete support chains. Some of these can be pruned from the knowledge base.

Definition 6.2 [equivalent to Santos et al. (1999) Def. 3.1]: A PAA R is said to be *well-supported* if there is some set of PAAs $\{S_1 \ldots S_n\}$, $n \geq 0$, and some set of PTAs $\{T_1, \ldots, T_m\}$, $m \geq 0$, and some set of individuals a, such that

$$con(S_1) \wedge \ldots \wedge con(S_n) \wedge con\left(T_1|_{a_{i_1} \in I}\right) \wedge \ldots \wedge con\left(T_m|_{a_{i_m} \in I}\right) = ant(R).$$

If R is not well-supported, then it is said to be *unsupported*. $S_{1\ldots n}$ and $T_{1\ldots m}|_{a_{i_{1\ldots m}}}$ are said to support R.

Intuitively, a well-supported PAA is either unconditional, or the pmf is defined over its antecedents. An unsupported PAA is one for which we know the pmf will be unconstrained and therefore unknown given its antecedents.

We will now show that PAAs whose support chains are not fully well-supported may be pruned from the knowledge base without affecting its results.

Definition 6.3: A PAA R is said to be grounded if it is unconditional or if its supporting rules are all grounded. If R is not grounded, then it is said to be ungrounded.

For any BKO B, let I(B) be the set of all inferences over B.

Theorem 6.1: Let B be a BKO containing a set, Q, of ungrounded PAAs. Then B – Q is a BKO and I(B) = I(B – Q).

Proof: By Proposition 5.2, B – Q is a BKO. By Definition 6.3, an ungrounded PAA does not satisfy Definition 4.5 and therefore will never be part of a deductive set. Therefore, no element of Q will ever be in an inference, so I(B) = I(B – Q).

Definition 6.4: A BKO B is *fully-reasoned* when for any individual a and PTA T, either $T|_a \in B$ or $T|_a$ is ungrounded.

Definition 6.5: A BKO B is *fully-grounded* when all PAAs in B are grounded.

A fully-reasoned BKO can be transformed into an equivalent BKB (see Section VII), but may still have rules in it that the BKB will not use. By ensuring the BKO is also fully-grounded we will avoid working with unnecessary rules.

Example 6.1: To render the BKO from Example 5.2 fully-reasoned, we would infer the following PTA instantiations:

(5) P(Tuna ∈ Predator | Tuna $ate_{(\mathbf{some}\ \mathrm{Fish})}$) = 0.9
(6) P(Tuna ∈ Has_Parasites | Tuna ∈ Fish ∩ Predator) = 0.3
(7) P(Herring ∈ Predator | Herring $ate_{(\mathbf{some}\ \mathrm{Fish})}$) = 0.9
(8) P(Herring ∈ Has_Parasites | Herring ∈ Fish ∩ Predator) = 0.3

Note that the antecedents are unsupported, but the information needed to make (5), (6), and (8) supported is clearly present in the BKO. Generate PAAs (9), (10), and (12) to express description logic axioms that "bridge" these terms. The lack of a solution for (11) shows where the reasoning chain for Herring breaks and becomes unsupported.

(9) P(Tuna ∈ $ate_{(\mathbf{some}\ \mathrm{Fish})}$ | Tuna *ate* Herring ∧ Herring ∈ Fish) = 1
(10) P(Tuna ∈ Fish ∩ Predator | Tuna ∈ Fish ∧ Tuna ∈ Predator) = 1
(11) P(Herring ∈ $ate_{(\mathbf{some}\ \mathrm{Fish})}$ | Herring *ate* ? ∧ ? ∈ Fish) = 1
(12) P(Herring ∈ Fish ∩ Predator | Herring ∈ Fish ∧ Herring ∈ Predator) = 1

To make this BKO fully-grounded, we follow the support chain from (11). This leads us to prune (7), (12), and (8). The BKO made by (1-6) and (9-10) is fully-reasoned and fully-grounded.

VII. Uncertainty Reasoning for BKOs

We now define the means of transforming a BKO into an equivalent BKB.

Lemma 7.1: Given a BKO B, the subset S containing all PAAs in B comprises a BKB.

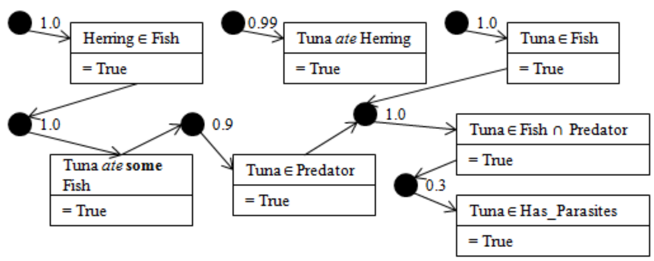
Proof: By Proposition 5.2, S is a BKO and therefore obeys the constraints of Definition 5.12. Since S consists entirely of PAAs, which by Definition 5.3 are CPRs, these constraints reduce to those of Definition 4.4, the definition of a BKB.

Theorem 7.1: Given a fully-reasoned BKO B, the subset S containing all PAAs in B comprises a BKB with the same set of possible pmfs as B.

Proof: By Lemma 6.1, S is a BKB. By Definition 6.5, all pmf restrictions due to the PTAs in B are present in S, so B and S both describe the same set of possible pmfs.

Once a BKO is fully-reasoned, Theorem 7.1 means that its PAAs are a BKB, and can be reasoned over using BKB theory's methods. These are belief updating, which computes posterior probabilities of target variable states, belief revision, which computes posterior probabilities of domain instantiations, and partial belief revision, which computes posterior probabilities of sets of target variable states. These can be conducted on a BKO-derived BKB just as for any other BKB as in [12]. The probabilistic membership query from the introduction is simply a partial belief revision problem.

Example 7.1: BKB form of the fully-reasoned, fully-grounded BKO generated at the end of Example 6.1.



VIII. Conclusion

A new framework was presented for formal uncertainty and incompleteness handling in semantic networks. The framework, called Bayesian Knowledge-driven Ontologies, is founded on a synthesis of description logic and Bayesian Knowledge Bases. These knowledge representations' theories are founded on compatible assumptions, so the synthesis is intuitive and powerful. We presented a formal method of reasoning both logically and probabilistically on BKOs.

Our future work will define BKO theory fully based on Baader et al.'s (2007) description logic [1], including its more abstracted notion of classes and roles as formal concepts in models subject to interpretations. We intend to codify reasoning algorithms and research their efficiency. We will test implementations of BKOs in knowledge engineering problems. We also intend to apply BKOs to our ongoing research in human intent modeling, and we are conducting investigations into using BKOs to enable more complex robot behaviors and machine reasoning. We believe that BKOs are capable of subsuming semantic networks, so we intend to create a BKO extension of OWL which subsumes its current implementation. On the subject of the semantic web, we believe that BKOs are capable of providing a formal method of solving several roadblocks in semantic web research. Most immediately, BKOs enable the formal representation of uncertain knowledge, a roadblock which, as we show in Section II, had not been unreservedly solved. Additionally, the formalization of ontology alignment and merging appears to be low-hanging fruit, thanks to existing work from BKB theory on the subject of merging knowledge from different sources [16]. This would be a major step toward enabling fully autonomous ontology merging, which would make disparate knowledge in the semantic web much more machine-accessible.

Acknowledgment

This work was supported in part by grants from ONR MURI Program, AFOSR, DTRA, and DHS. Also, special thanks to help and support from Securboration, Inc.

References

[1] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi and P. F. Patel-Schneider. *The Description Logic Handbook*. Cambridge: Cambridge University Press, 2007.

[2] C. Boutilier, N. Friedman, M. Goldszmidt, D. Koller. "Context-specific independence in Bayesian Networks." In Proc. UAI, 1996, pp. 115-123.

[3] P. C. G. Costa and K. B. Laskey, "PR-OWL: A framework for probabilistic ontologies." in *Proc. FOIS*, 2006, pp. 237-249.

[4] Z. Ding, Y. Peng, R. Pan, Z. Ding, Y. Peng and R. Pan, "BayesOWL: Uncertainty modeling in semantic web ontologies," presented at Soft Computing in Ontologies and Semantic Web, 2005.

[5] D. Koller, A. Levy and A. Pfeffer, "P-CLASSIC: A tractable probabilistic description logic," presented at AAAI, 1997.

[6] T. Lukasiewicz, "Expressive probabilistic description logics," *Artif. Intell.* vol. 172(6-7), pp. 852-883, 2008.

[7] D. L. McGuinness and F. v. Harmelen. "OWL web ontology language overview: W3C recommendation." Internet: http://www.w3.org/TR/owl-features/, Feb. 10 2004 [Apr. 1 2011].

[8] N. J. Nilsson. "Probabilistic logic." *Artif. Intell.* vol. 28(1), pp. 71-87, 1986.

[9] J. Pearl. "Bayesian networks: A model of self-activated memory for evidential reasoning." (UCLA Technical Report CSD-850017). In *Proc. of the 7th Conf. of the Cognitive Science Society*, 1985, pp. 329–334.

[10] G. Qi, J. Z. Pan and Q. Ji. "A possibilistic extension of description logics." In *Proc. DL*, 2007 .

[11] E. Santos Jr., B. McQueary and L. and Krause, "Modeling adversarial intent for interactive simulation and gaming: The fused intent system," (refereed abstract), in *Proc. SPIE: Defense & Security Symposium*, Vol. 6965, Orlando, FL, 2008.

[12] E. Santos, Jr. and E. S. Santos. "A framework for building knowledge-bases under uncertainty." *J. Exp. Theor. Artif. Intell.*, vol. 11, pp. 265-286, 1999.

[13] E. Santos, Jr., E. S. Santos and S. E. Shimony. "Implicitly preserving semantics during incremental knowledge base acquisition under uncertainty." *International Journal of Approximate Reasoning*, vol 33(1), pp. 71-94, 2003.

[14] E. E. Santos, E. Santos, Jr., J. T. Wilkinson, J. Korah, K. Kim, D. Li and F. Yu. "Modeling Complex Social Scenarios using Culturally Infused Social Networks." Submitted to IEEE SMC 2011.

[15] E. E. Santos, E. Santos, Jr., J. T. Wilkinson and H. Xia. "On a framework for the prediction and explanation of changing opinions." In *Proc. IEEE SMC*, 2009, 1446-1452.

[16] E. Santos, Jr., J.T. Wilkinson, and E. E. Santos. "Bayesian knowledge fusion," in *Proc. 22nd International FLAIRS Conf.* AAAI Press, 2009.

[17] G. Stoilos, G. Stamou, V. Tzouvaras, J. Z. Pan and I. Horrocks. "Fuzzy owl: Uncertainty and the semantic web." in *Proceedings of the International Workshop on OWL: Experiences and Directions*, 2005.

[18] U. Straccia. "Reasoning within fuzzy description logics." *Journal of Artificial Intelligence Research*, vol. 14, 2001.