Uncertainty in Knowledge Provenance

Jingwei Huang and Mark S. Fox

Enterprise Integration Laboratory, University of Toronto
40 St. George Street, Toronto, ON M5S 3G8, Canada
{jingwei,msf}@eil.utoronto.ca

Appeared in: *Proceedings of the 1st European Semantic Web Symposium*, Heraklion, Greece, May 2004, Springer Lecture Notes in Computer Science.

Abstract. Knowledge Provenance is an approach to determining the origin and validity of knowledge/information on the web by means of modeling and maintaining information sources and dependencies, as well as trust structures. This paper constructs an uncertainty-oriented Knowledge Provenance model to address the provenance problem with uncertain truth values and uncertain trust relations by using information theory and probability theory. This proposed model could be used for both people and web applications to determine the validity of web information in a world where information is uncertain.

1 Introduction

With the widespread use of the World Wide Web and telecommunications making information globally accessible, comes a problem: anyone is able to produce and distribute information on the web; however, the information may be true or false, current or outdated, or even outright lies. The concerns regarding how to determine the validity of web information are receiving more and more attention. Interest in addressing the issue of web information trustworthiness has appeared under the umbrella of the "Web of Trust" which is identified as the top layer of the Semantic Web and is still in its infant stage of development (see [2] slides 26&27).

Knowledge Provenance (hereafter, referred to as **KP**) is proposed in [6] to create an approach to determining the origin and validity of web information by means of modeling and maintaining information sources and dependencies, as well as trust structures. The major questions KP attempts to answer include: Can this information be believed to be true? Who created it? Can its creator be trusted? What does it depend on? Can the information it depends on be believed to be true? This proposed approach could be used to help people and web software agents to determine the validity of web information.

Four levels of KP have been identified, as follows:

- Level 1 (**Static KP**) focuses on provenance of static and certain information;
- Level 2 (**Dynamic KP**) considers how the validity of information may change over time;
- Level 3 (**Uncertainty-oriented KP**) considers information whose validity is inherently uncertain;

- Level 4 (**Judgment-based KP**) focuses on societal processes necessary to support provenance.

Static KP and Dynamic KP have been studied in [6] and [10] respectively. This paper focuses on uncertainty-oriented KP.

In Levels 1 and 2 of KP, an information creator is either trusted or distrusted, and a proposition is trusted by a provenance requester to have a truth value of "True", "False", or "Unknown". However, it is common to find that a person may trust an information creator to a certain degree rather than completely trust or completely distrust it. Furthermore, a proposition created by the information creator may also be believed to be true to an extent rather than absolutely "True" or "False". The questions here are how to define these types of uncertainty and how to use uncertain values to infer the validity of a proposition.

Level 3, or uncertainty-oriented KP, addresses this type of provenance problem in a world where information is uncertain. This paper focuses on the basic and the most important aspects of uncertainty in provenance, that is, uncertain trust relationships and uncertain truth values. "Trust Degree" (subjective probability) is introduced to represent uncertain trust relationships; "Degree of Certainty", the probability of a proposition to be true, is used to represent uncertain truth values; and an uncertainty-oriented KP model is constructed to infer the degrees of certainty for different types of propositions by applying information theory and probability theory. This uncertainty-oriented KP model can be used to determine the validity of web information with uncertain trust relationships and uncertain truth values.

The content of this paper is organized as follows. Section 2 introduces the related research; section 3 introduces the basic concepts of knowledge provenance; section 4 provides a motivating scenario for developing an uncertainty-oriented KP model; section 5 constructs an uncertainty-oriented KP model by applying probability theory and information theory; section 6 provides an example to use uncertainty-oriented KP for provenance reasoning; and section 7 provides a summary and future research.

2 Related Research

The issue of web information trustworthiness has appeared under the umbrella of the "Web of Trust" that is identified as the top layer of the Semantic Web [2].

No doubt, digital signature and digital certification [18] play important roles in the "Web of Trust". However, they only provide an approach to certify an individual's identification and information integrity, but they do not determine whether this individual can be trusted. Trustworthiness of the individual is supposed to be evaluated by each web application. For the purpose of secure web access control, Blaze et al [4] first introduced "decentralized trust management" to separate trust management from applications. Since then, trust management has grown from web access control to more general trust concerns in various web applications. PolicyMaker [4] introduced the fundamental concepts of policy, credential, and trust relationship. REFEREE [5] introduced trust protocol; Kinateder and Rothermal [14] developed a distributed reputation system with a trust building model; Herrmann [9] used Jøsang's subjective logic [12] to evaluate the trust values of software components. Twigg [19] applied Jøsang's subjective logic based trust model to support routing decision for P2P and ad hoc networks. Golbeck et al [8] and Richardson et al [16] developed the models of trust propagation in social networks.

Trust management attempts to answer the question of whether an individual is trusted to do a specific action to a specific resource [13]. However, KP needs to answer whether the information created by an individual in a specific field can be believed to be true. Even though KP may be regarded as a specific form of trust management in which the action is understood as telling true information, KP still needs to handle certain problems beyond the current range of trust management. In the context of KP, trust management only considers trust relationships between information users and information creators; however, it does not consider the dependencies among web information. KP needs to consider both of them.

Regarding uncertainty in trust management, uncertainty logics provide various methods for representing and updating uncertainty/belief [3]. Jøsang [12] proposed subjective logic to represent uncertain trust values with an opinion triangle in which an opinion is represented as a triple (b, d, u) where b, d, u denote the degrees of belief, disbelief, and uncertainty respectively, and the sum of them equals to 1. This method can discern the difference between "unknown" and "disbelief", but it requires a degree of uncertainty in addition to degree of belief or disbelief, thus possibly causing some difficulties to users. Gil & Ratnakar [7], as well as Golbeck et al [8] represent uncertain trust relationships by grading with discrete numbers corresponding to a set of linguistic descriptions. The advantages of this method are simple and easy to use. The disadvantages are that users usually have different understandings on the linguistic descriptions, thereby resulting inconsistent in defining and understanding the descriptions of trust relationship. Fuzzy logic has the similar difficulties. Probability is a more direct solution adopted by many researchers to represent uncertain trust relationships. This paper also uses probability to represent both uncertain trust relationships and uncertain truth values.

3 What is Knowledge Provenance?

Knowledge Provenance is an approach to determining the origin and validity of knowledge/information on the web by means of modeling and maintaining information sources and interdependence, as well as trust relations. This section introduces the basic concepts of KP.

The basic unit of web information to be considered in KP is a "proposition". A proposition, as defined in Proposional Logic, is a declarative sentence that is either true or false. A proposition is the smallest piece of information to which provenance-related attributes may be ascribed. An information creator may define a phrase, a sentence, a paragraph, even a whole document as a proposition. Not only text but also an xml element could be defined as a proposition. The taxonomy of the propositions in KP is illustrated in figure 1. KP_prop is the most general class of propositions; An Asserted_prop is an assertion that is not dependent on any other propositions; A Dependent_prop is a proposition which truth is dependent on other propositions; An Equivalent_prop is a copy and its truth value is the same as the proposition it depends on; A Derived_prop is a derived conclusion based on some premises; A Composite_prop could be the “and”/ “or” / “negation” of other proposition(s).

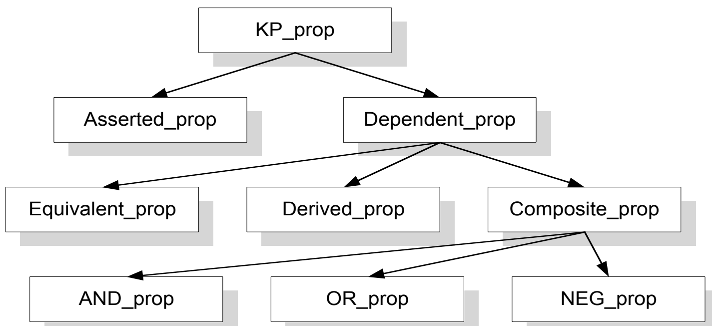

Figure 1. Taxonomy of Propositions in Knowledge Provenance

To use KP, information creators need to annotate web documents with KP metadata to describe the provenance-related attributes, such as who is proposition creator and what is the premise proposition on which this proposition depends. A web browser “plugin” is expected to assist information creators to annotate their web documents; information users (provenance requesters) need to define their personalized trust relations to tell whom they trust; an online KP software agent (a KP reasoner) will trace KP tags (KP metadata) in web documents across web pages, combining information sources and dependencies, as well as trust relationships, to deduce the origin and validity of tagged information.

In the context of KP, “trust” means that one party believes the information created by another party in a specific field to be true. A trust relation in KP is defined as a of triple (*a, c, f*) where the provenance requester *a* "trusts" (directly or indirectly) information creator *c* in a specific knowledge field *f*, here, "trust" means that *a* believes any proposition created by *c* in field *f* to be true; “indirect trust” means that *a* do not directly know *c* but trusts *c* by the media of some references who trust *c*.

3.1 Static Knowledge Provenance

As mentioned earlier, there are 4 levels of KP. Level 1, or static KP, is the basis for other levels. So, we give a brief introduction to static KP first. A detailed description of static KP can be found in [6].

Static KP focuses on static and certain information. Static KP needs to answer the following competency questions that define the requirements for static KP.

– Is this proposition true, false, or unknown?
– Who created this proposition?
– Is the creator of the proposition trusted?

– Does the truth of this proposition depend on any other propositions? If so, what?
– What is the digital signature verification status of this proposition?
– Which knowledge fields does this proposition belong to?
– In these fields, can the information creator be trusted?

Static KP Axioms

The axioms for static KP are summarized as follows. The formal specification of these axioms in First Order Logic can be found in [6].

- A KP-prop is "trusted", if the creator or publisher of the proposition is "trusted" in a field that covers* one of the fields of the proposition, and the digital signature verification status is "Verified".
- For an asserted, or derived, or equivalent KP-prop that has no creator specified, the creator of the document is the default creator of the KP-prop.
- If a proposition does not have a creator, then the digital signature verification status of the KP-prop is determined by the digital signature verification status of the document.
- The default assigned truth value of a KP-prop is "True". That is, if a proposition creator does not give the truth value of a proposition, the creator implicitly declare the truth value is "True".
- The trusted truth value of an asserted-prop is the same as its assigned truth value, if the asserted-prop is trusted by the provenance requester; otherwise the trusted truth value is "Unknown".
- The trusted truth value of an equivalent-prop is the same as the trusted truth value of the proposition it depends on, if this equivalent-prop is trusted; otherwise the trusted truth value is "Unknown".
- The trusted truth value of a derived-prop is the same as its assigned truth value, if the derived-prop is trusted and the KP-prop it depends on is "True"; otherwise the trusted truth value is "Unknown". Note that it is unnecessary to include everything used to derive the truth value in the dependencies.
- The trusted truth value of a negative-prop is the negation of the trusted truth value of the KP-prop it depends on, if the negative-prop is trusted by the provenance requester; otherwise the trusted truth value is "Unknown".
- The trusted truth value of an And-prop is "True" if all the KP-props it depends on are "True"; The trusted truth value of an And-prop is "False" if at least one of the KP-props it depends on is "False"; and the trusted truth value of an And-prop is "Unknown" if at least one of the KP-props it depends on is "Unknown" and none of them is "False".
- The trusted truth value of an Or-prop is "True" if at least one of the KP-props it depends on is "True"; The trusted truth value of an Or-prop is "False" if all of the KP-props it depends on

* The relations among different knowledge fields could be very complex, which is beyond our topic on KP. We assume that a common recognized taxonomy of knowledge fields is used.

are "False"; and the trusted truth value of an Or-prop is "Unknown" if at least one of the KP-props it depends on is "Unknown" and none of them is "True".

4. Motivating Scenario of Uncertainty-oriented KP

The following two cases provide a clue for constructing uncertainty-oriented Knowledge Provenance model.

Case 1: Uncertain Truth Values

Consider the proposition found on a web page that “Acupuncture on pain-relief points cuts blood flow to key areas of the brain within seconds” discovered by a scientist in Harvard Medical School. Instead of giving truth value as True (1) or False (0), the proposition creator may assign a numeric truth value between 0 and 1 to the proposition. This numeric truth value represents the degree of confidence (subjective probability) that the creator believes the proposition to be true. When a reader reads this proposition from the web, what numeric truth value does the reader give to this proposition? And how to calculate it? Intuitively, the numeric truth value given by the reader will depend on how much the reader trust the proposition creator and whether this proposition is dependent on other propositions.

Case 2: Uncertain Trust in Information Creators

Further consider the trust relationship between a reader and the proposition creator in the above example. A reader may trust the creator in the field of “Neuroscience” to a certain degree rather than completely "trust" or completely "distrust" it. Here, "trust" means to believe any proposition created by the creator on the topic of "Neuroscience". The degree of trust could be represented with a number in interval [0,1.0] where 1.0 is corresponding to complete trust and 0 is corresponding to complete distrust. For example, the reader trusts the creator to a degree of 0.9, that should be understood as any proposition about “Neuroscience” created by the creator is believed to be true by the reader with a subjective probability of 0.9.

These two cases reveal the following points for building uncertainty-oriented KP:

- The truth value of a proposition may be uncertain. The degree of confidence (subjective probability) for a proposition to be true could be introduced to extend a binary truth value to a numeric truth value.
- A proposition creator may assign a numeric truth value to a proposition, and a numeric trusted truth value of a proposition may be calculated according to how much the provenance requester trusts this proposition and the trusted truth value of the proposition that this proposition depends on.
- Trust relationships may be uncertain. The degree of belief (subjective probability) could be introduced as trust degree to represent uncertain trust relationships.

5. Uncertainty-oriented KP Model

This section aims to construct an uncertainty-oriented KP model by applying probability theory and information theory. The following terms defined in static KP need to be used. (Note: in this paper, "KP agent" represents "provenance requester").

assigned_truth_value(x, y): proposition *x* has truth value of *y* assigned by its creator.

trusted_truth_value(a,x,y): KP agent *a* trusts that proposition *x* has truth value *y*.

trusted(x,a): proposition *x* is trusted by agent *a*.

Several notations and definitions used in this paper are introduced as follows:

Pr(Y) denotes the probability of event *Y*;

"TTV_x" denotes *trusted_truth_value(a, x, "True")*, that is, the trusted truth value of proposition *x* (trusted by KP agent *a*) is "True". In our discussion, only one provenance requester (agent *a*) is involved, so, "*a*" does not appear in "TTV_x". Other notations below are similar.

"ATV_x" denotes assigned*_truth_value(x, "True")*, i.e., the truth value of proposition *x* assigned by proposition creator is "True";

"$Trusted_x$" denotes *trusted(x, a)*, that is, KP agent *a* trusts proposition *x*.

When only one proposition is involved, the footnote representing the proposition can be omitted, e.g., "TTV_x" is written as "*TTV*".

Consider that a proposition has only two possible determined truth values: "True" or "False", therefore, "$\neg ATV_x$" represents *assigned_truth_value(x, "False")*; and similarly "$\neg Trusted_x$" represents that agent *a* distrust proposition *x*. Note that as a simple method to handle uncertainty, "Unknown" was used to represent a status in which truth value cannot be determined in static KP. In this paper, we will introduce a method to represent uncertain truth value. So, "Unknown" will no longer be used as a truth value.

From the motivating scenario in the last section, we know that proposition creator may assign a numeric truth value to a proposition. This numeric truth value assigned by proposition creator is called "**assigned certainty degree**" and is used to represent uncertain assigned truth value. It is defined as follows.

Definition 1: the assigned certainty degree (denoted as *acd*) of a proposition given by the proposition creator is defined as the degree of confidence (subjective probability) of the proposition creator to assign the truth value of "True" to the proposition.

$$acd = Pr(ATV) \qquad (5\text{-}1)$$

Similar to static KP where a proposition has a trusted truth value (trusted by a provenance requester), a proposition may have a numeric trusted truth value. This numeric truth value is called "**certainty degree** (or degree of certainty)" and is used to represent uncertain trusted truth value. It is defined as follows.

Definition 2: the certainty degree (denoted as *cd*) of a proposition is defined as the probability in which provenance requester believe the proposition being "True", that is, the probability of the trusted truth value to be "True".

$$cd = Pr(TTV) \qquad (5\text{-}2)$$

Finally, "**trust degree**" is defined to represent uncertain trust relations.

Definition 3: the trust degree (denoted as *td*) of a proposition is defined as the degree of belief (subjective probability) for the provenance requester to trust this proposition.

$$td = Pr(Trusted) \qquad (5\text{-}3)$$

In the following, first, knowledge provenance model for asserted propositions is constructed, and then the same approach is applied to other types of propositions including "Derived", "Equivalent", "AND", "OR", as well as "NEG".

5.1 Uncertain model of Asserted Propositions

When an asserted proposition has an assigned certainty degree that represents uncertain assigned truth value given by the proposition creator, what is the certainty degree (uncertain trusted truth value) of this proposition? According to Axiom 1 of static KP:

for-all (a,x,v)
((type(x, "asserted_prop") ^ trusted(x, a) (5-1-0)
^ assigned_truth_value[(1)]*(x, v))*
->trusted_truth_value(a, x, v))

the certainty degree of an asserted proposition should depend on (1) the assigned certainty degree given by the proposition creator; (2) the trust degree of the proposition. From the axiom, it is easy to understand, when the trust degree is 1.0, the certainty degree will be the same as the assigned certainty degree, as shown in figure 2(a). But when trust degree is 0 (corresponding to "unknown"), what is the value the certainty degree? Furthermore, when trust degree is less than 1.0 and greater than 0, what is the relation among certainty degree, assigned certainty degree, and trust degree?

First, let us consider the case of trust degree being zero. According to information theory ([17] [15]), "entropy" is used to measure uncertainty of information, and the entropy of a variable *x* which has *n* possible outcomes $v_1, \ldots, v_n$ is defined as follows.

$$H(x) = - \sum_{1,\ldots,n} p_i \log p_i \qquad (5\text{-}1\text{-}1)$$

where, p_i is the probability for the variable to have outcome v_i; for a given *n*, when all the p_i are equal to (1/n), which is corresponding to the most uncertain situation, entropy *H(x)* is maximal and equals to *log n*; entropy *H(x)* is minimal and equals to 0 if and only if one p_i is 1.0 and all others are 0, which is corresponding to the most certain situation. In the case of the variable having only two possible outcomes, the entropy becomes:

$$H(x) = - (p \log p + (1 - p) \log (1 - p)) \qquad (5\text{-}1\text{-}$$

---

(1) Predicate assigned_truth_value(…) is used to replace the predicate truth_value(…) defined in [fox&huang2003]. They have the same definition.

2)

And the entropy has maximal value if and only if $p=0.5$.

In our context of uncertainty-oriented KP, if the trust degree of an asserted proposition is 0, then no matter what value the assigned certainty degree is, there is no information for determining the certainty degree of the proposition, which is corresponding to the most uncertain situation where the entropy should be maximal. As a proposition has only two determined values "True" and "False", in this case, the probability of this proposition being "True" should be 0.5, that is, the certainty degree of this proposition should be 0.5. Therefore, based on information theory, we assign 0.5 to the certainty degree of this asserted proposition when trust degree is 0, as shown in figure 2(b). As a matter of fact, this situation of asserted proposition can be extended to other types of propositions. When a proposition is distrusted, no matter what type the proposition is, there is no information available to determine its certainty degree, so according to information theory the certainty degree of the proposition should be 0.5. For this reason, we have the following axiom.

Axiom 5-1:

for-all (a,x) ((type(x, "KP_prop") ^ ¬ trusted(x, a))
->certainty_degree(a, x, 0.5)).

Now consider the general situation when trust degree is any real value that ranges from 0 to1.0. Recall axiom 1 of Static KP (formula 5-1-0). We know that the trusted truth value of an asserted-prop is dependent on (1) whether the asserted-prop is trusted by the provenance requester; (2) the assigned truth value given by the proposition creator. By using the sum rule and conditional probability of Probability theory, the probability of the trusted truth value of an asserted proposition being "True" is calculated with the following formula:

$$Pr(TTV) = Pr(TTV \mid Trusted, ATV) * Pr(Trusted, ATV) + Pr(TTV \mid Trusted, \neg ATV) * Pr(Trusted, \neg ATV) + Pr(TTV \mid \neg Trusted) * Pr(\neg Trusted) \quad (5\text{-}1\text{-}3)$$

Because whether a proposition is trusted by the provenance requester and what is the assigned truth value of the proposition given by its creator are independent to each other, according to the product rule of probability theory, we have

$$Pr(Trusted, ATV) = Pr(Trusted) * Pr(ATV) \quad (5\text{-}1\text{-}4)$$
$$Pr(Trusted, \neg ATV) = Pr(Trusted) * Pr(\neg ATV)$$

Apply (5-1-4) and $Pr(\neg Y) = 1 - Pr(Y)$ to (5-1-3),

$$Pr(TTV) = Pr(TTV \mid Trusted, ATV) * Pr(ATV) * Pr(Trusted) + Pr(TTV \mid Trusted, \neg ATV) * Pr(Trusted) * (1 - Pr(ATV)) + Pr(TTV \mid \neg Trusted) * (1 - Pr(Trusted)) \quad (5\text{-}1\text{-}5)$$

The conditional probabilities in the above formula can be determined as follows. According to axiom 1 of static KP, when the assigned truth value of a proposition is assigned as “True”, and the proposition is trusted, the trusted truth value is “True”, that is, the probability in which the trusted truth value is “True” is 1.0, i.e.

$$Pr(TTV|\ Trusted, ATV) = 1.0 \qquad (5\text{-}1\text{-}6)$$

Similarly, when the assigned truth value of a proposition is assigned as “False”, and the proposition is trusted, the trusted truth value is “False”, that is, the probability in which the trusted truth value is “True” is 0, i.e.

$$Pr(TTV|\ Trusted, \neg ATV) = 0 \qquad (5\text{-}1\text{-}7)$$

According to information theory and our discussion earlier in this section, when a proposition is distrusted, no matter what the assigned certainty degree given by the proposition creator is, there is no information to determine the certainty degree, which is corresponding to the most uncertain situation and the “entropy” has maximal value, so the certainty degree of this proposition should be 0.5, i.e.

$$Pr(TTV|\ \neg Trusted) = 0.5 \qquad (5\text{-}1\text{-}8)$$

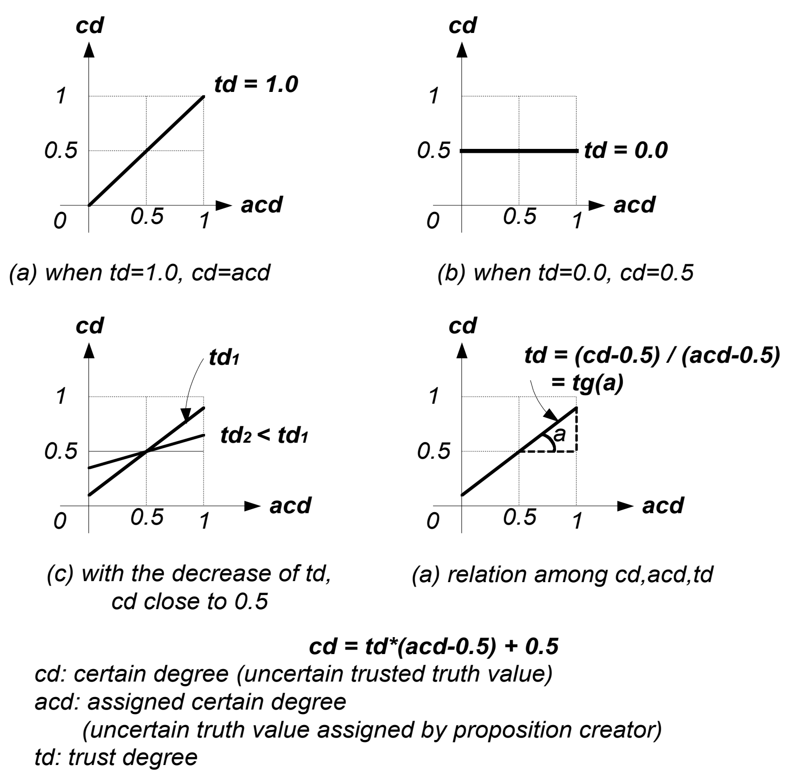

Figure 2. Relation among certain degree, assigned certain degree, and trust degree

Applying (5-1-6) (5-1-7) (5-1-8) and definitions in (5-1) (5-2) (5-3) to formula (5-1-5), we have

$$cd = td*(acd - 0.5) + 0.5 \qquad (5\text{-}1\text{-}9)$$

The relation among the certainty degree, assigned certainty degree, and trust degree of a proposition, revealed by formula (5-1-9), can be illustrated in figure 2. When trust degree is 1.0 (completely trust), the certainty degree is the same as the assigned certainty degree given by the proposition creator (see figure 2 (a)); with the decrease of trust degree, the certainty degree is close to 0.5 ("unknown") (see figure 2 (c)); when the trust degree is 0 (completely distrust), the certainty degree should be 0.5 ("unknown") (see figure 2 (b)); if assigned certainty degree $acd = 0.5$, then certainty degree $cd = 0.5$, no matter what the trust degree is (see figure 2 (a)(b)(c)).

Theorem 5-1: The certainty degree of an asserted proposition is dependent on the trust degree of the proposition and the assigned certainty degree given by the proposition creator. The relation among them satisfies:

$$cd = td*(acd - 0.5) + 0.5 \qquad (5\text{-}1\text{-}9)$$

The derivation of formula (5-1-9) gives the proof of this theorem.

In the following subsections, the approach used above is applied to set up uncertainty-oriented KP model for other types of propositions including "Equivalent", "Derived", "AND", "OR", and "NEG".

5.2 Uncertain Model of Equivalent Propositions

Theorem 5-2: The certainty degree of an equivalent proposition x is dependent on the trust degree of x and the certainty degree of the proposition y that this equivalent proposition depends on. The relation among them satisfies:

$$cd_x = td_x * (cd_y - 0.5) + 0.5 \qquad (5\text{-}2\text{-}1)$$

The proof of this theorem can be found in [Huang &Fox 2003B].

5.3 Uncertain Model of Derived Propositions

Theorem 5-3: The certainty degree of derived proposition x is dependent on the trust degree of x and the assigned certainty degree given by the proposition creator as well as the certainty degree of proposition y that x depends on. The relation among them is:

$$cd_x = td_x*cd_y*(acd_x - 0.5) + 0.5 \qquad (5\text{-}3\text{-}1)$$

The proof of this theorem is similar to theorem 5-1 and can be found in [11].

This model has the similar properties of the uncertainty model for asserted propositions (formula (5-1-9). When trust degree is 1.0 (completely trusted) and the certainty degree of premise y is 1.0 (“True”), the certainty degree of derived proposition *x* is the same as the assigned certainty degree given by its creator; if trust degree is 0 (completely distrusted) or the certainty degree of premise *y* is 0 (“False”) or the assigned certainty degree of derived proposition x given by its creator is 0.5 (“Unknown”), the certainty degree of proposition *x* will be 0.5 (“Unknown”); with the decrease of trust degree of *x* and certainty degree of *y*, the certainty degree of derived proposition *x* is close to 0.5.

5.4 Uncertain Model of Composite Propositions

As the premise of a derived proposition may be a composite (“AND”/ “OR”/ “NEG”) proposition, uncertainty-oriented KP needs to answer how to calculate the certainty degree of a composite proposition.

“AND” Propositions

Consider “AND” proposition $z = (x \wedge y)$. According to product rule of probability theory:

$$Pr(A \wedge B) = Pr(A|B)*Pr(B),$$

or

$$Pr(A \wedge B) = Pr(B|A)*Pr(A),$$

and if A is conditionally independent to B (i.e., $Pr(B|A)= Pr(B)$), then

$$Pr(A \wedge B) = Pr(B)*Pr(A)$$

In order to calculate $Pr(x \wedge y)$, the relation between *x* and *y*, either the statement of *x* and *y* being conditional independent or the conditional probability $Pr(x|y)$ (or $Pr(y|x)$) needs to be provided by the proposition creator. In the context of KP, this claimed relation between *x* and *y* needs to be trusted by provenance requester. So, in KP, that “AND” proposition $z = (x \wedge y)$ is trusted should be understood as the relation between *x* and *y* (conditional probability) is trusted. The certainty degree of *z*, TTV_z, is calculated as follows.

$$\begin{aligned} Pr(TTV_z) = {} & Pr(TTV_z|Trusted_z, (x \wedge y))*Pr(Trusted_z, (x \wedge y)) \\ & + Pr(TTV_z|Trusted_z, \neg(x \wedge y))*Pr(Trusted_z, \neg(x \wedge y)) \\ & + Pr(TTV_z|\neg Trusted_z)*Pr(\neg Trusted_z) \end{aligned} \quad (5\text{-}4\text{-}1)$$

It is easy to understand that if proposition *z* is trusted and $x \wedge y$ is true, then *z* is true; if the proposition *z* is trusted but $x \wedge y$ is false, then *z* is false; and if proposition *z* is distrusted, then there is no information to determine the truth of *z*, that is, if the conditional probability used to calculate $Pr(x \wedge y)$ is distrusted, then the correctness of the computing result of $Pr(x \wedge y)$ is unknown. So we have:

$$\begin{aligned} & Pr(TTV_z|Trusted_z, (x \wedge y)) = 1.0 \\ & Pr(TTV_z|Trusted_z, \neg(x \wedge y)) = 0 \\ & Pr(TTV_z|\neg Trusted_z) = 0.5 \end{aligned} \quad (5\text{-}4\text{-}2)$$

In addition, whether proposition z is trusted is conditionally independent to whether $x \wedge y$ is true. Therefore,

$$Pr(Trusted_z, (x \wedge y)) = Pr(Trusted_z)* Pr(x \wedge y) \qquad (5\text{-}4\text{-}3)$$

Furthermore, $cd_y = Pr(y)$, and $cd_x = Pr(x)$, so, we have

$$Pr(x \wedge y) = Pr(x|y)* Pr(y) = Pr(x|y)* cd_y \qquad (5\text{-}4\text{-}4)$$

or

$$Pr(x \wedge y) = Pr(y|x)* Pr(x) = Pr(y|x)* cd_x$$

Applying (5-4-2) to (5-4-4) and definition (5-2) (5-3) to (5-4-1), we have the formula to calculate the certainty degree of "AND" proposition, $z = x \wedge y$, as follows.

Axiom 5-3: if $z = (x \wedge y)$, then

$$\boldsymbol{cd_z = td_z*(Pr(x|y)* cd_y - 0.5) + 0.5} \qquad (5\text{-}4\text{-}5)$$

or

$$\boldsymbol{cd_z = td_z*(Pr(y|x)*cd_x - 0.5) + 0.5}$$

"OR" Propositions

Consider "OR" proposition $z = (x \vee y)$. Because

$$Pr(x \vee y) = Pr(x) + Pr(y) - Pr(x \wedge y) \qquad (5\text{-}4\text{-}6)$$

and $Pr(x \wedge y)$ appears in $Pr(x \vee y)$, the relation (conditional probability) between proposition x and y need to be specified and need to be trusted also.

Similar to uncertainty-oriented KP model of "AND" propositions, the certainty degree of "OR" proposition z is calculated as follows, the proof is omitted.

Axiom 5-4: if $z = (x \vee y)$, then

$$\boldsymbol{cd_z = td_z*(cd_x + cd_y - Pr(x|y)*cd_y - 0.5) + 0.5} \qquad (5\text{-}4\text{-}10)$$

or

$$\boldsymbol{cd_z = td_z*(cd_x + cd_y - Pr(y|x)* cd_x - 0.5) + 0.5}$$

"NEG" Propositions

Uncertainty-oriented KP model of "NEG" proposition is very simple. Consider "NEG" proposition $x = \neg y$. According to probability theory,

$$Pr(\neg y) = 1 - Pr(y)$$

So, we have

Axiom 5-5: if $x = \neg y$, then

$$\boldsymbol{cd_x = 1 - cd_y} \qquad (5\text{-}4\text{-}11)$$

6. Example

An example to illustrate how to use uncertainty-oriented KP model is given as follows. Some basic concepts of KP involved can be found in section 3.

A reader find a web page containing the following propositions: an asserted proposition (Asserted_prop: “New finding”) that “Acupuncture on pain-relief points cuts blood flow to key areas of the brain within seconds”, an equivalent proposition (Equivalent_prop: “Brain areas”) that “The specific brain areas affected are involved in mood, pain and cravings”, and a derived proposition (Derived_prop: “Implications of finding”) that “This finding could help explain why some studies have found acupuncture helpful in treating depression, eating problems, addictions and pain.” In the example, the proposition on which the equivalent proposition is dependent is in another web document. Assume that this web page is annotated with kp metadata.

The following is an example of annotating one proposition. Other propositions could be annotated in similar way. An example of annotating a whole web document could be found in [6](section 5).

```
<kp:Derived_prop rdf:id="Implications_of_finding"
      is_dependent_on="#Conditions"
      creator ="Bruce Rosen"
      certainty_degree = 0.9
      in_field ="Neuroscience"
>
This finding could help explain why some studies have found acupuncture helpful in treating depression, eating problems, addictions and pain.
</kp:Derived_prop>
```

Figure 3 illustrates the major kp metadata associated with each proposition, the dependencies of the propositions, and the provenance reasoning process using uncertainty-oriented KP model. To use KP, the reader needs define his/her trust relations (shown as “trust_degree” boxes in figure 3). Certainly, a KP agent (KP reasoner) can provide a set of default trust relations to certain common used information sources. A KP agent will conduct provenance reasoning as requested from the reader. According to theorem 5-3, in order to calculate the certainty degree of derived proposition “Implications of finding”, KP agent needs to calculate the trust degree to this proposition, according to the assigned certainty degree of the proposition, and the certainty degree of its premise, the AND_prop “condition1”. The latter leads to calculating the certainty degrees of Equivalent_prop “Brain areas” and Asserted_prop “New finding”. And the calculation of the certainty degree of Equivalent_prop “Brain areas” leads to the calculating of the certainty degree of Asserted_prop “Brain_regions”. So, the calculation process can be outlined in 5 steps as shown in figure 3 in which each step is represented with a box marked by step number and the formula used. For example, step (1) calculating the certainty degree of asserted proposition “New finding” by using formula (5-1-9).

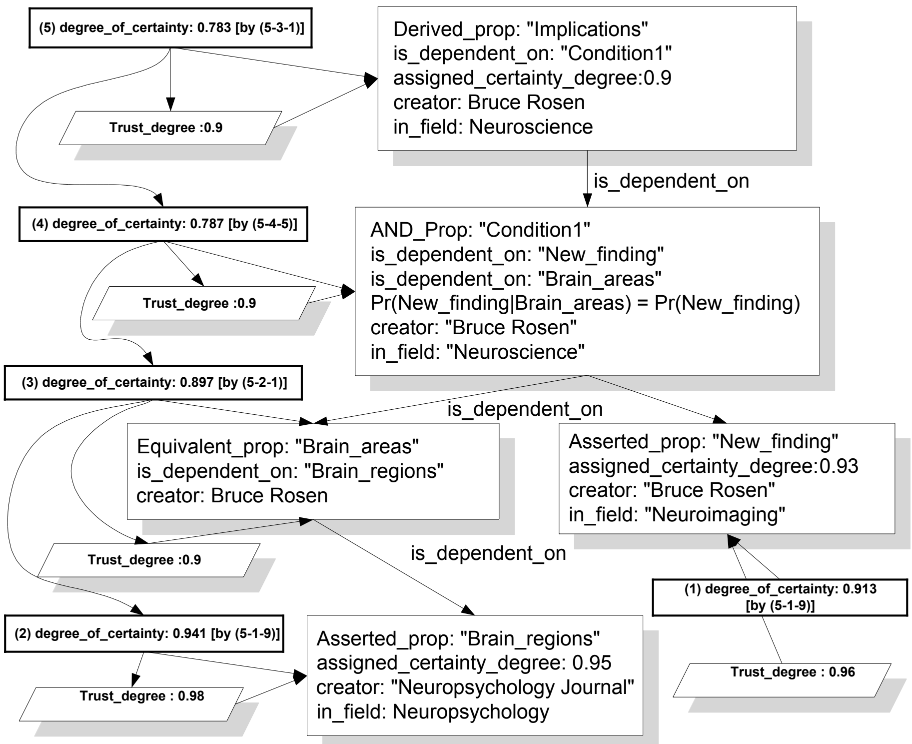

Figure 3. Example of provenance reasoning using Uncertainty-oriented KP

7. Summary

This paper proposes an uncertainty-oriented KP model to address the provenance problem with uncertain trust relationships and uncertain truth values. "Trust Degree" (subjective probability) is introduced to represent uncertain trust relationships; "Certainty Degree", the probability of a proposition to be true, is used to represent uncertain truth value; and an uncertainty-oriented KP model is constructed to infer the certainty degree for different types of propositions by using information theory and probability theory. This uncertainty-oriented KP model could be used to determine the validity of web information in a world where information is uncertain.

As mentioned in introduction section, Knowledge Provenance comprises of four levels: Static, Dynamic, Uncertain, and Judgment-based KP. To continue our work, we will develop judgment-based KP that focuses on societal processes necessary to support knowledge provenance in which indirect judgment of information trustworthiness in social processes will be considered.

Acknowledgements
This research was supported, in part, by Bell University Laboratory.

References

1. Berners-Lee, T., Hendler, J., and Lassila, O., (2001), "The Semantic Web", *Scientific American*, May 2001.
2. Berners-Lee, T., (2003), Semantic Web Status and Direction, *Int. Semantic Web Conf. 2003*, keynote. http://www.w3.org/2003/Talks/1023-iswc-tbl/
3. Bhatnager, R.K., and Kanal, R., (1986), Handling Uncertain Information: A Review of Numeric and Non-numeric Methods, in *Uncertainty in Artificial Intelligence*, edited by L. Kanal and J. F. Lemmer, Elsevier Science Publishers.
4. Blaze, M., Feigenbaum, J. and Lacy, J., (1996), Decentralized Trust Management, *Proceedings of IEEE Conference on Security and Privacy*, May, 1996.
5. Chu, Y., (1997), Trust Management for the World Wide Web, Master Thesis, MIT.
6. Fox, M. S., and Huang, J., (2003), "Knowledge Provenance: An Approach to Modeling and Maintaining the Evolution and Validity of Knowledge", EIL Technical Report, Uni.of Toronto, May 2003, http://www.eil.utoronto.ca/km/papers/fox-kp1.pdf
7. Gil, Y. and Ratnakar, V., (2002), "Trusting Information Sources One Citizen at a Time", *Proceedings of Int. Semantic Web Conf.2002*.
8. Golbeck, J., Hendler, J., and Parsia, B., (2002), Trust Networks on the Semantic Web, University of Maryland, College Park.
9. Herrmann P., (2003), Trust-Based Protection of Software Component Users and Designers, *Proceedings of 1st Int. Conf. On Trust Management*, *LNCS 2692*, Springer, PP.75-90.
10. Huang, J. and Fox, M. S., (2003), " Dynamic Knowledge Provenance ", EIL Technical Report, University of Toronto, June 2003. http://www.eil.utoronto.ca/km/papers/kp2-TR03.pdf
11. Huang, J., and Fox, M.S., (2003B), "Uncertainty-oriented Knowledge Provenance", EIL Technical Report, University of Toronto, September 2003.
12. Jøsang, A., (2001), A Logic for Uncertain Probabilities, *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, V.9, N.3, 2001, PP.279-311.
13. Khare, R., and Rifkin, A., (1997), "Weaving and Web of Trust", *World Wide Web Journal*, Vol. 2, No. 3, pp. 77-112.
14. Kinateder, M. and Rothermel K., (2003), Architecture and Algorithms for a Distributed Reputation System, *Proceedings of 1st Int. Conf. On Trust Management, LNCS 2692*, Springer, PP.1-16.
15. MacKay, D.J.C. (2003), *Information Theory, Inference, and Learning Algorithm*, Cambridge University Press, 2003.
16. Richardson, M., Agrawal, R., and Domingos, P., (2003), Trust Management for the Semantic Web, *Proc. of Int. Semantic Web Conf. 2003*, PP.351-368.
17. Shannon, C.E. (1948), A Mathematical Theory of Communication, *The Bell System Technical Journal*, Vol.27, pp379-423, 623-656, October, 1948.

18. Simon, E., Madsen, P., Adams, C., (2001), An Introduction to XML Digital Signatures, http://www.xml.com/pub/a/2001/08/08/xmldsig.html
19. Twigg A., (2003), A Subjective Approach to Routing in P2P and Ad Hoc Networks, *Proc. of 1st Int. Conf. On Trust Management, LNCS 2692*, Springer, PP.225-238.