Measuring the Ignorance and Degree of Satisfaction for Answering Queries in Imprecise Probabilistic Logic Programs

Anbu Yue[1], Weiru Liu[1], and Anthony Hunter[2]

[1] School of Electronics, Electrical Engineering and Computer Science,
Queen's University Belfast, Belfast BT7 1NN, UK
{a.yue, w.liu}@qub.ac.uk

[2] Department of Computer Science, University College London,
Gower Street, London WC1E 6BT, UK
a.hunter@cs.ucl.ac.uk

Abstract. In probabilistic logic programming, given a query, either a probability interval or a precise probability obtained by using the maximum entropy principle is returned for the query. The former can be noninformative (e.g., interval $[0, 1]$) and the reliability of the latter is questionable when the priori knowledge is imprecise. To address this problem, in this paper, we propose some methods to quantitatively measure if a probability interval or a single probability is sufficient for answering a query. We first propose an approach to measuring the ignorance of a probabilistic logic program with respect to a query. The measure of ignorance (w.r.t. a query) reflects how reliable a precise probability for the query can be and a high value of ignorance suggests that a single probability is not suitable for the query. We then propose a method to measure the probability that the exact probability of a query falls in a given interval, e.g., a second order probability. We call it the degree of satisfaction. If the degree of satisfaction is high enough w.r.t. the query, then the given interval can be accepted as the answer to the query. We also provide properties of the two measures and use an example to demonstrate the significance of the measures.

1 INTRODUCTION

Probabilistic logic programming is a framework to represent and reason with imprecise (conditional) probabilistic knowledge. An agent's knowledge is represented by a *probabilistic logic program* (PLP) which is a set of (conditional) logical formulae with probability intervals. The impreciseness of the agent's knowledge is explicitly represented by assigning a probability interval to every logical formula (representing a conditional event) indicating that the probability of a formula shall be in the given interval.

Given a PLP and a query against the PLP, traditionally, a probability interval is returned as the answer. This interval implies that the true probability of the query shall be within the given interval. However, when this interval is too wide, it provides no useful information. For instance, if a PLP contains knowledge $\{(fly(X)|bird(X)[0.98, 1]$, $(bird(X)|magpie(X))[1, 1]\}$, then the answer to the query *Can a magpie fly?* (i.e., $?(fly(t)|magpie(t)))$ is a trivial bound $[0, 1]$.

One way to enhance the reasoning power of a PLP is to apply the maximum entropy principle [1]. Based on this principle, a single probability distribution is selected and it is assumed to be the most acceptable one for the query among all possible probability distributions. As a consequence, a precise probability is given for a query even when the agent's original knowledge is imprecise. In the above example, by applying the maximum entropy principle, 0.98 is returned as the answer for the query. Intuitively, accepting a precise probability from (a prior) imprecise knowledge can be risky. When an agent's knowledge is rich enough then a single probability could be reliable, however, when an agent's knowledge is (very) imprecise, an interval is more appropriate than a single probability.

Therefore, in probabilistic logic programming as well as other conditional probabilistic logics, there is a question that has not been fully investigated, that is, how useful is a probabilistic logic program (PLP) to answering a given query? This question is important in two ways: first, it helps to analyze if a PLP is adequate to answer a query and second, if a PLP is sufficiently relevant to a query, then shall a single probability be obtained or shall a probability interval be more suitable? If it is an interval that is more suitable, then how can we get a more meaningful interval (which is satisfactory to certain extent), rather then a loose bound?

To answer the above questions, in this paper, we propose two concepts, *the measure of ignorance* and *the measure of the degree of satisfaction*, w.r.t. a PLP and a query. The former analyzes the impreciseness of the PLP w.r.t. the query, and the latter measures which (tighter) interval is sufficiently reliable to answer the query.

The main contributions of this paper are as follows. First, we propose a general framework which formally defines the measure of ignorance and the measure of the degree of satisfaction, and the postulates for these two measures. We also provide several consequence relations based on the degree of satisfaction. Second, by using the divergence of probabilistic distribution, we instantiate our framework, and show that the measure of ignorance and the measure of the degree of satisfaction have many desirable properties and provide much useful information about a PLP w.r.t. a query. Third, we prove that our framework is an extension of both reasoning on probabilistic logic program and reasoning under the maximum entropy principle. Fourth, we prove that these measures can be viewed as a second-order probability. More specifically, a high level of ignorance means a high probability about the given PLP (the agent's knowledge) is towards total absence of knowledge. The degree of satisfaction is the second-order probability about the actual probability for a conditional event given in the query falls in the given interval (provided in the query).

This paper is organized as follows. A brief review of probabilistic logic programming is given in Section 2. In Section 3, we formally analyze probabilistic logic programming and the maximum entropy principle, and provide our general framework. In Section 4, we give instantiations of the framework. We then use an example to demonstrate the significance of the measures in Section 5. Finally, we compare our approach with related work and conclude the paper in Section 6.

2 PROBABILISTIC LOGIC PROGRAMMING

We briefly review conditional probabilistic logic programming here [2, 3].

We use Φ to denote the finite set of predicate symbols and constants symbols, $\mathcal{V}$ to denote the set of *object variables*, and $\mathcal{B}$ to denote the set of *bound constants* which describe the bound of probabilities and bound constants are in [0,1]. We use $a, b, \ldots$ to denote constants from Φ and $X, Y \ldots$ to denote object variables from $\mathcal{V}$. An *object term* t is a constant from Φ or an object variable from $\mathcal{V}$. An *atom* is of the form $p(t_1, \ldots, t_k)$, where p is a predicate symbol and $t_1, \ldots, t_k$ are object terms. We use Greek letters $\phi, \varphi, \psi, \ldots$ to denote *events* (or *formulae*) which are obtained from atoms by logic connectives $\wedge, \vee, \neg$ as usual. A *conditional event* is of the form $(\psi|\phi)$ where ψ and ϕ are events, and ϕ is called the *antecedent* and ψ is called the *consequent*. A *probabilistic formula*, denoted as $(\psi|\varphi)[l, u]$, means that the probability of conditional event $\psi|\varphi$ is between l and u, where l, u are bound constants. A set of probabilistic formulae is called a *conditional probabilistic logic program (PLP)*, a PLP is denoted as P in the rest of the paper.

A *ground* term, (resp. event, conditional event, probabilistic formula, or PLP) is a term, (resp. event, conditional event, probabilistic formula, or PLP) that does not contain any object variables in $\mathcal{V}$.

All the constants in Φ form the Herbrand universe, denoted as HU_Φ, and the Herbrand base, denoted as HB_Φ, is the finite nonempty set of all atoms constructed from the predicate symbols in Φ and constants in HU_Φ. A subset I of HB_Φ is called a *possible world* and $\mathcal{I}_\Phi$ is used to denote the set of all possible worlds over Φ. A function σ that maps each object variable to a constant is called an *assignment*. It is extended to object terms by $\sigma(c) = c$ for all constant symbols from Φ. An event φ satisfied by I under σ, denoted by $I \models_\sigma \varphi$, is defined inductively as:

- $I \models_\sigma p(t_1, \ldots, t_n)$ iff $p(\sigma(t_1), \ldots, \sigma(t_n)) \in I$;
- $I \models_\sigma \phi_1 \wedge \phi_2$ iff $I \models_\sigma \phi_1$ and $I \models_\sigma \phi_2$;
- $I \models_\sigma \phi_1 \vee \phi_2$ iff $I \models_\sigma \phi_1$ or $I \models_\sigma \phi_2$;
- $I \models_\sigma \neg\phi$ iff $I \not\models_\sigma \phi$

An event φ is satisfied by a possible world I, denoted by $I \models_{cl} \varphi$, iff $I \models_\sigma \varphi$ for all assignments σ. An event φ is a *logical consequence* of event ϕ, denoted as $\phi \models_{cl} \varphi$, iff all possible worlds that satisfy ϕ also satisfy φ.

In this paper, we use $\top$ to represent (ground) tautology, and we have that $I \models_{cl} \top$ for all I and all assignments σ. And we use $\perp$ to denote $\neg\top$.

If Pr is a function (or distribution) on $\mathcal{I}_\Phi$ (i.e., as $\mathcal{I}_\Phi$ is finite, Pr is a mapping from $\mathcal{I}_\Phi$ to the unit interval [0,1] such that $\sum_{I \in \mathcal{I}_\Phi} Pr(I) = 1$), then Pr is called a *probabilistic interpretation*. For an assignment σ, the probability assigned to an event φ by Pr, is denoted as $Pr_\sigma(\varphi)$ where $Pr_\sigma(\varphi) = \sum_{I \in \mathcal{I}_\Phi, I \models_\sigma \varphi} Pr(I)$. When φ is ground, we simply write it as $Pr(\varphi)$. When $Pr_\sigma(\phi) > 0$, the conditional probability, $Pr_\sigma(\psi|\phi)$, is defined as $Pr_\sigma(\psi|\phi) = Pr_\sigma(\psi \wedge \phi)/Pr_\sigma(\phi)$. When $Pr_\sigma(\phi) = 0$, $Pr_\sigma(\psi|\phi)$ is undefined. Also, when $(\psi|\phi)$ is ground, we simply written as $Pr(\psi|\phi)$.

A probabilistic interpretation Pr satisfies or is a *probabilistic model* of a probabilistic formula $(\psi|\phi)[l, u]$ under assignment σ, denoted as $Pr \models_\sigma (\psi|\phi)[l, u]$, iff $u \geq Pr_\sigma(\psi|\phi) \geq l$ or $Pr_\sigma(\phi) = 0$. A probabilistic interpretation Pr satisfies or is a *probabilistic model* of a probabilistic formula $(\psi|\phi)[l, u]$ iff Pr satisfies $(\psi|\phi)[l, u]$ under all assignments. A probabilistic interpretation Pr satisfies or is a *probabilistic model* of a PLP P iff for all assignment σ, $\forall(\psi|\phi)[l, u] \in P, Pr \models_\sigma (\psi|\phi)[l, u]$. A probabilistic formula $(\psi|\varphi)[l, u]$ is a *consequence* of the PLP P, denoted by $P \models (\psi|\varphi)[l, u]$, iff all probabilistic models of P satisfy $(\psi|\varphi)[l,u]$. A probabilistic formula $(\psi|\varphi)[l,u]$ is a *tight consequence* of P, denoted by $P \models_{tight} (\psi|\varphi)[l,u]$, iff $P \models (\psi|\varphi)[l,u]$, $P \not\models (\psi|\varphi)[l,u']$, $P \not\models (\psi|\varphi)[l',u]$ for all $l' > l$ and $u' < u$ ($l', u' \in [0,1]$). It is worth noting that if $P \models (\phi|\top)[0,0]$ then $P \models (\psi|\phi)[1,0]$ where $[1,0]$ stand for the empty set.

A query is of the form $?(\psi|\phi)$ or $?(\psi|\phi)[l,u]$, where ψ and ϕ are ground events and $l,u \in [0,1]$. For query $?(\psi|\phi)$, by the tight consequence relation, a bound $[l,u]$ is given as the answer, such that $P \models_{tight} (\psi|\phi)[l,u]$. For query $?(\psi|\phi)[l,u]$, a bound $[l,u]$ is given by the user. A PLP returns *True (or Yes)* if $P \models (\psi|\phi)[l,u]$ and *False (or No)* if $P \not\models (\psi|\phi)[l,u]$ [3].

The principle of maximum entropy is a well known techniques to represent probabilistic knowledge. Entropy quantifies the indeterminateness inherent to a distribution Pr by $H(Pr) = -\Sigma_{I \in \mathcal{I}_\Phi} Pr(I) \log Pr(I)$. Given a logic program P, the *principle of maximum entropy model* (or *me-model*), denoted by $me[P]$, is defined as: $H(me[P]) = \max H(Pr) = \max_{Pr \models P} -\Sigma_{I \in \mathcal{I}_\Phi} Pr(I) \log Pr(I)$

$me[P]$ is the unique probabilistic interpretation Pr that is a probabilistic model of P and that has the greatest entropy among all the probabilistic models of P.

Let P be a ground PLP, we say that $(\psi|\varphi)[l,u]$ is a *me-consequent* of P, denoted by $P \models^{me} (\psi|\varphi)[l,u]$, iff P is unsatisfiable, or $me[P] \models (\psi|\varphi)[l,u]$.

We say that $(\psi|\varphi)[l,u]$ is a *tight me-consequent* of P, denoted by $P \models^{me}_{tight} (\psi|\varphi)[l,u]$, iff either P is unsatisfiable, $l = 1$, $u = 0$, or $P \models \bot \leftarrow \varphi$, $l = 1$, $u = 0$, or $me[P](\varphi) > 0$ and $me[P](\psi|\varphi) = l = u$.

3 GENERAL FRAMEWORK

Example 1. Let P be a PLP:

$$P = \left\{ \begin{array}{l} (fly(X)|bird(X))[0.9,1], (bird(X)|magpie(X))[1,1] \\ (sickMagpie(X)|magpie(X))[0,0.1], (magpie(X)|sickMagpie(X))[1,1] \end{array} \right\}$$

From P, we can infer that $P \models_{tight} (fly(t)|magpie(t))[0,1]$,
$P \models_{tight} (fly(t)|sickmagpie(t))[0,1]$, $P \models^{me}_{tight} (fly(t)|magpie(t))[0.9,0.9]$, and $P \models^{me}_{tight} (fly(t)|sickMagpie(t))[0.9,0.9]$.

In the above example, we get the same answers for queries on the proportions that magpies and sick magpies can fly. Since the proportion of sick magpies in birds is smaller than the proportion of magpies in birds, the knowledge about birds can fly should be more cautiously applied to sick magpies than magpies. In other words, the statement that more than 90% birds can fly is more about magpies than sick magpies. Therefore, to accept that 90% magpies can fly is more rational than to accept 90% sick magpies can fly. However, this analysis can not be obtained directly from comparing the bounds inferred from P.

In this section, we provide a framework to measure the ignorance of a PLP w.r.t. a conditional event and the degree of satisfaction for a conditional event with a user-given bound under a PLP.

Definition 1 (Ignorance). *Let $\mathcal{PL}$ be the set of all PLPs and $\mathcal{E}$ be a set of conditional events. Function* $\mathsf{IG} : \mathcal{PL} \times \mathcal{E} \mapsto [0, 1]$ *is called a* measure [3] of ignorance*, iff for any PLP P and conditional event $(\psi|\phi)$ it satisfies the following postulates*

[Bounded] $\mathsf{IG}(P, \psi|\phi) \in [0, 1]$.
[Preciseness] $\mathsf{IG}(P, \psi|\phi) = 0$ *iff* $P \models_{tight} (\psi|\phi)[u, u]$ *for some* $u \in [0, 1]$ *or* $P \models \bot \leftarrow \phi$.
[Totally Ignorance] $\mathsf{IG}(\emptyset, \psi|\phi) = 1$*, if* $\not\models_{cl} \phi \rightarrow \psi$ *and* $\not\models_{cl} \phi \rightarrow \neg\psi$.
[Sound] *If* $\mathsf{IG}(P, \psi|\phi) = 1$ *then* $P \models (\psi|\phi)[l, u]$ *iff* $\emptyset \models (\psi|\phi)[l, u]$.
[Irrelevance] *If P and another PLP P' do not contain common syntaxes, i.e.$\Phi \cap \Phi' = \emptyset$, then* $\mathsf{IG}(P, \psi|\phi) = \mathsf{IG}(P \cup P', \psi|\phi)$*, where P and P' are defined over Φ and Φ' respectively.*

For simplicity, we use $\mathsf{IG}_P(\psi|\phi)$ *to denote* $\mathsf{IG}(P, \psi|\phi)$ *for a given PLP P and $(\psi|\phi)$. Value* $\mathsf{IG}_P(\psi|\phi)$ *defines the level of* ignorance *about $(\psi|\phi)$ from P.*

If $P = \emptyset$, only tautologies can be inferred from P. Therefore, from any PLP P, $\mathsf{IG}_P(\psi|\phi) \leq \mathsf{IG}_\emptyset(\psi|\phi)$, which means that an empty PLP has the biggest ignorance value for any conditional event. When $\mathsf{IG}_P(\psi|\phi) = 0$, event $(\psi|\phi)$ can be inferred precisely from P, since a single precise probability for $(\psi|\phi)$ can be obtained from P.

Definition 2 (Degree of Satisfaction). *Let $\mathcal{PL}$ be the set of all PLPs and $\mathcal{F}$ be a set of probabilistic formulae. Function* $\mathsf{SAT} : \mathcal{PL} \times \mathcal{F} \mapsto [0, 1]$ *is called a measure of* degree of satisfaction *iff for any PLP P and ground probabilistic formula $\mu = (\psi|\phi)[l, u]$, it satisfies the following postulates:*

[Reflexive] $\mathsf{SAT}(P, \mu) = 1$*, iff* $P \models \mu$.
[Rational] $\mathsf{SAT}(P, \mu) = 0$ *if* $P \cup \{\mu\}$ *is unsatisfiable.*
[Monotonicity] $\mathsf{SAT}(P, \mu) \geq \mathsf{SAT}(P, (\psi|\phi)[l', u'])$*, if* $[l', u'] \subseteq [l, u]$.
$\mathsf{SAT}(P, \mu) > \mathsf{SAT}(P, (\psi|\phi)[l', u'])$*, if* $[l', u'] \subset [l, u]$ *and* $\mathsf{SAT}(P, (\psi|\phi)[l', u']) < 1$.
[Cautious Monotonicity] *Let* $P' = P \cup \{(\psi|\phi)[l', u']\}$*, where* $P \models^{me} (\psi|\phi)[l', u']$
Then $\mathsf{SAT}(P', \mu) \geq \mathsf{SAT}(P, \mu)$.

For simplicity, we use $\mathsf{SAT}_P(\mu)$ to denote $\mathsf{SAT}(P, \mu)$.

The reflexive property says that every consequence is totally satisfied. Rational says that 0 is given as the degree of satisfaction of an unsatisfiable probabilistic formula. Monotonicity says that if we expect a more precise interval for a query, then the chance that the exact probability of the query is *not* in the interval is getting bigger. Cautious monotonicity says that, if P and P' are equivalent except for the bound of $(\psi|\phi)$, and P' contains more knowledge about $(\psi|\phi)$, then the degree of satisfaction of $(\psi|\phi)[l, u]$ under P' should be bigger than that of $(\psi|\phi)[l, u]$ under P.

Proposition 1. *Function* SAT *is* consistent with the maximum entropy principle*, that is, it satisfies the following conditions for any PLP P and any conditional event $(\psi|\phi)$ with $P \not\models \bot \leftarrow \phi$ and $l, u \in [0, 1]$.*

$$\mathsf{SAT}_P((\psi|\phi)[l, u]) \begin{cases} = 0 \text{ if } P \models^{me} (\psi|\phi)[l', l'], l' \notin [l, u] \\ > 0 \text{ if } P \models^{me} (\psi|\phi)[l', l'], l' \in [l, u] \end{cases}$$

---

[3] In mathematical analysis, a measure $m : 2^S \mapsto [0, \infty]$ is a function, such that
1) $m(E_1) \geq 0$ for any $E \subseteq S$
2) $m(\emptyset) = 0$
3) If $E_1, E_2, E_3, \ldots$ is a countable sequence of pairwise disjoint subsets of S, the measure of the union of all the E_i's is equal to the sum of the measures of each E_i, that is,
$m(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} m(E_i)$

For a query $?(\psi|\phi)[l,u]$, when $\mathsf{SAT}_P((\psi|\phi)[l,u]) < 1$ it means that the exact probability of $(\psi|\phi)$ in $[l,u]$ could be wrong based on the knowledge in P.

In our framework, given a PLP P, a conditional event $(\psi|\phi)$, and a probabilistic formula $(\psi|\phi)[l,u]$, the ignorance value $\mathsf{IG}_P(\psi|\phi)$ and the degree of satisfaction $\mathsf{SAT}_P(\mu)$ reveal different aspects of the impreciseness of the knowledge in P w.r.t. $(\psi|\phi)$ and $(\psi|\phi)[l,u]$. The former says how much this P can tell about $(\psi|\phi)$ and the latter says to what degree a user can be satisfied with the bound $[l,u]$ with $(\psi|\phi)$.

Proposition 2. *Let P be a PLP and $(\psi|\phi)$ be a conditional event. If $\mathsf{IG}_P(\psi|\phi) = 0$ then $\mathsf{SAT}_P((\psi|\phi)[l,l]) = 1$ for some $l \in [0,1]$.*

Definition 3. *Let $\mathsf{SAT}_P(\mu)$ be the degree of satisfaction for a PLP P and $\mu = (\psi|\phi)[l,u]$ be a probabilistic formula. We define two consequence relations as*
$P \models^{SAT \geq w} \mu$ *iff* $\mathsf{SAT}_P(\mu) \geq w$,
$P \models^{SAT \geq w}_{tight} \mu$ *iff* $P \models^{SAT \geq w} \mu$ *and* $P \not\models^{SAT \geq w} (\psi|\phi)[l',u']$ *for every* $[l',u'] \subset [l,u]$.

Proposition 3. *Let $\mathsf{SAT}_P(\mu)$ be the degree of satisfaction for a PLP P and a probabilistic formula $\mu = (\psi|\phi)[l,u]$, then*

$$P \models \mu \textit{ iff } P \models^{SAT=1} \mu$$
$$P \models_{tight} \mu \textit{ iff } P \models^{SAT=1}_{tight} \mu$$

If SAT *is also consistent with the maximum entropy principle, then*
$P \models^{me}_{tight} \mu$ *iff* $\lim_{\epsilon \to 0+} P \models^{SAT \geq \epsilon}_{tight} \mu$

In this proposition, we use $\mathsf{SAT}_P(\mu) = 1$ instead of $\mathsf{SAT}_P(\mu) \geq 1$, since the degree of satisfaction cannot be bigger than 1.

The above proposition says that our framework is a generalization of PLP under its original semantics as well as under the maximum entropy principle. That is, the classical consequence relations $\models$ and $\models_{tight}$ are too cautious - they are equivalent to requiring the degree of satisfaction of μ w.r.t P to be 1, which means that the true probability of $(\varphi|\phi)$ must fall in the bound $[l,u]$. On the other hand, the reasoning under the maximum principle ($\models^{me}_{tight}$) is credulous – it excludes all the other possible probability distributions except for the most possible one.

Given a query $?(\varphi|\phi)[l,u]$ against a PLP P, the degree of satisfaction $\mathsf{SAT}_P(\mu)$ tells the probability that $p(\varphi|\phi) \in [l,u]$. For a query $?(\varphi|\phi)$, the bound $[l,u]$ returned by $P \models_{tight} (\psi|\phi)[l,u]$ may be noninformative as discussed above. In our framework, we provide three ways to generate a more informative interval $[l',u']$ with $\mathsf{SAT}_P((\varphi|\phi)[l',u']) \geq a$, where a is threshold given by the user. First, a user may want to know the highest acceptable lower bound, so the lower bound is increased from 0 to l' until $\mathsf{SAT}_P((\varphi|\phi)[l',u]) \geq a$ holds. Second, a user may want to know the lowest upper bound, so u is decreased to be u' until $\mathsf{SAT}_P((\varphi|\phi)[l,u']) \geq a$ is true. Third, a user may want to create an interval $[l',u']$ around $me[P]$, the precise probability given by the maximum entropy principle, where $\mathsf{SAT}_P((\varphi|\phi)[l',u']) \geq a$ holds. To formalize these three scenarios, we define three consequence relations $\models^{SAT \geq a}_{maxLow}$, $\models^{SAT \geq a}_{minUp}$ and $\models^{SAT \geq a}_{aroundMe}$ for them respectively as

– $P \models^{SAT \geq a}_{maxLow} (\psi|\phi)[l',u]$ iff $P \models^{SAT \geq a} (\psi|\phi)[l',u]$ where $P \models_{tight} [l,u]$, and $l' > l$

- $P \models^{SAT \geq a}_{minUp} (\psi|\phi)[l, u']$ iff $P \models^{SAT \geq a}_{tight} (\psi|\phi)[l, u']$, where $P \models_{tight} [l, u]$ and $u > u'$
- $P \models^{SAT \geq a}_{aroundMe} (\psi|\phi)[l', u']$ iff $P \models^{SAT \geq a}_{tight} (\psi|\phi)[l', u']$ where $P \models_{tight} (\psi|\phi)[l, u]$, and $\exists b \geq 0, P \models^{me}_{tight} [m, m], l' = \max\{l, m - b\}, u' = \min\{u, m + b\}$

Example 2. Let $P = \{(fly(t)|bird(t))[0.90, 1], (bird(t)|magpie(t))[1, 1]\}$ be a PLP. From P, we can only infer that $P \models_{tight} (fly(t)|magpie(t))[0, 1]$, and $P \models^{me}_{tight} (fly(t)|magpie(t))[0.9, 0.9]$. As discussed above, the bound $[0, 1]$ is meaningless and there is not enough knowledge to infer that exactly 90% magpies can fly. In reality, taking $[0.9, 0.9]$ as the answer for this query is too risky, and there is no need to get a precise probability for the query. A more informative interval $[l, u]$ then $[0, 1]$ would be useful. Assume that a user is happy when there is a 80% (i.e. $a = 0.8$) chance that the actual probability of the query is in $[l, u]$, then we are able to use the above three consequence relations to get the following

$$
\begin{aligned}
P &\models^{SAT \geq 0.8}_{maxLow} (fly(t)|magpie(t))[0.7, 1] \\
P &\models^{SAT \geq 0.8}_{minUp} (fly(t)|magpie(t))[0, 0.96] \\
P &\models^{SAT \geq 0.8}_{aroundMe} (fly(t)|magpie(t))[0.7, 1]
\end{aligned}
$$

From the highest lower bound 0.7, the user can assume that a magpie very likely can fly. The user should not think that all magpies can fly either, since the lowest upper bound 0.96 is less than 1. The bound $[0.7, 1]$ gives an estimate for the probability of a magpie can fly.

In our framework, the user can calculate the degree of satisfaction for a query with a user-given bound, and the user can also calculate the tightest bound for a query s.t. the degree of satisfaction w.r.t. this bound is greater than a user-given threshold.

4 INSTANTIATING THE FRAMEWORK

In this section, we provide an instantiation of our framework by defining a specific ignorance function and a satisfaction function. But first, we define a quasi-distance between probability distributions based on Kullback-Leibler divergence (KL-divergence) [4].

One of the most common measures of distance between probability distributions is the KL-divergence.

Definition 4. *Let Pr and Pr' be two probability distributions over the same set of interpretations $\mathcal{I}_\Phi$. The KL-divergence between Pr and Pr' is defined as:*

$$KL(Pr \| Pr') = -\Sigma_{I \in \mathcal{I}_\Phi} Pr(I) \log \frac{Pr'(I)}{Pr(I)}$$

KL-divergence is asymmetric and is also called *relative entropy*. It is worth noting that $KL(Pr, Pr')$ is undefined if $Pr'(I) = 0$ and $Pr(I) \neq 0$. This means that Pr has to be absolutely continuous w.r.t. Pr' for $KL(Pr \| Pr')$ to be defined.

4.1 MEASURABLE SPACE

In this subsection, we first define a measurable space, in which we can measure how *wide* a set of probability distributions is.

Let $\mathbf{Pr}_{\Phi}$ be the set of all probability distributions on the set of interpretations $\mathcal{I}_{\Phi}$. Let $\mathbf{Pr}_1$ and $\mathbf{Pr}_2$ be two subsets of $\mathbf{Pr}_{\Phi}$, $\mathbf{Pr}_1$ and $\mathbf{Pr}_2$ are *separated* if each is disjoint from the other's *closure* [4]. A subset $\mathbf{Pr}$ of $\mathbf{Pr}_{\Phi}$ is called *inseparable* if it cannot be partitioned into two separated subsets. Empty set $\emptyset$ is defined as inseparable. For example, the intervals $[0, 0.3], [0.4, 1]$ are separated and each of them is inseparable in the set of real numbers $\mathcal{R}$. Obviously, any subset $\mathbf{Pr}$ can be partitioned into a set of inseparable sets. Formally, there exists $\mathbf{Pr}_1, \mathbf{Pr}_2, \ldots$, such that every $\mathbf{Pr}_i$ is inseparable, $\mathbf{Pr}_i \cap \mathbf{Pr}_j = \emptyset$ $(i \neq j)$, and $\mathbf{Pr} = \bigcup_i \mathbf{Pr}_i$.

It is worth noting that, the set of all probabilistic models for a PLP is a convex set, which is an *inseparable set*. So, we only need to define a measurable space over all inseparable sets.

Definition 5. *Let $(\psi|\phi)$ be a conditional event and $\mathbf{Pr}$ be a subset of $\mathbf{Pr}_{\Phi}$. Suppose that $\mathbf{Pr}$ is inseparable, $l = \inf_{Pr \in \mathbf{Pr}} Pr(\psi|\phi)$, and $u = \sup_{Pr \in \mathbf{Pr}} Pr(\psi|\phi)$. We define $\delta^{ub} : 2^{\mathbf{Pr}_{\Phi}} \times \mathcal{F} \rightarrow [0, 1]$ and $\delta^{lb} : 2^{\mathbf{Pr}_{\Phi}} \times \mathcal{F} \rightarrow [0, 1]$ as*

$$\delta^{ub}(\mathbf{Pr}, (\psi|\phi)) = \min_{Pr \in \mathbf{Pr}\ Pr \models (\psi|\phi)[u, u]} KL(Pr || Pr_{unif})$$

$$\delta^{lb}(\mathbf{Pr}, (\psi|\phi)) = \min_{Pr \in \mathbf{Pr}\ Pr \models (\psi|\phi)[l, l]} KL(Pr || Pr_{unif})$$

where Pr_{unif} is the uniform distribution on $\mathcal{I}_{\Phi}$.

If $Pr(\phi) = 0$ for all $Pr \in \mathbf{Pr}$, we define $\delta^{ub}(\mathbf{Pr}, (\psi|\phi)) = \delta^{lb}(\mathbf{Pr}, (\psi|\phi)) = 0$.

For simplicity, we use $\delta^{ub}_{\mathbf{Pr}}(\psi|\phi)$ to denote $\delta^{ub}(\mathbf{Pr}, (\psi|\phi))$ and use $\delta^{lb}_{\mathbf{Pr}}(\psi|\phi)$ to denote $\delta^{lb}(\mathbf{Pr}, (\psi|\phi))$.

Value $\delta^{ub}_{\mathbf{Pr}}(\psi|\phi)$ (resp. $\delta^{lb}_{\mathbf{Pr}}(\psi|\phi)$) measures how much additional information needs to be added to the uniform distribution in order to infer the upper (resp. lower) bound of the conditional event $(\psi|\phi)$ given subset $\mathbf{Pr}$.

Definition 6. *Let $\mathbf{Pr}$ be an inseparable subset of $\mathbf{Pr}_{\Phi}$ and $(\psi|\phi)$ be a conditional event defined on Φ. Let $\mathbf{Pr}_{IS}$ contains all the inseparable subsets of $\mathbf{Pr}_{\Phi}$. We define $\vartheta_{\psi|\phi} : \mathbf{Pr}_{IS} \mapsto [0, 1]$ as $\vartheta_{\psi|\phi}(\mathbf{Pr}) = \mathrm{sign}(p - u) * \delta^{ub}_{\mathbf{Pr}}(\psi|\phi) - \mathrm{sign}(p - l) * \delta^{lb}_{\mathbf{Pr}}(\psi|\phi)$, where $p = Pr_{unif}(\psi|\phi)$, $l = \min_{Pr \in \mathbf{Pr}} Pr(\psi|\phi)$ and $u = \max_{Pr \in \mathbf{Pr}} Pr(\psi|\phi)$. Here, $\mathrm{sign} : R \mapsto R$ is defined as $\mathrm{sign}(x) = 1$ if $x \geq 0$ and $\mathrm{sign}(x) = -1$ otherwise.*

In the above definition, if $Pr(\phi) = 0$ for all $Pr \in \mathbf{Pr}$, then we canonically define $\vartheta_{\psi|\phi}(\mathbf{Pr}) = 0$ since $\delta^{ub}_{\mathbf{Pr}}(\psi|\phi) = \delta^{lb}_{\mathbf{Pr}}(\psi|\phi) = 0$.

Let σ_{Φ} denote the smallest collection such that σ_{Φ} contains all the inseparable subsets of $\mathbf{Pr}_{\Phi}$ and it is closed under complement and countable unions of its members. Then, $\langle \mathbf{Pr}_{\Phi}, \sigma_{\Phi} \rangle$ is a *measurable space* over the set $\mathbf{Pr}_{\Phi}$. Obviously, $\mathbf{Pr}_{\Phi} \in \sigma_{\Phi}$, and if $\mathbf{Pr} = \{Pr \mid Pr \models P\}$ for any PLP P, then $\mathbf{Pr} \in \sigma_{\Phi}$.

We extend function $\vartheta_{\psi|\phi}$ to the members of σ_{Φ}.

Definition 7. *Let $\mathbf{Pr}$ be a member of σ_{Φ} and $(\psi|\phi)$ be a conditional event. Define $\vartheta_{\psi|\phi} : \sigma_{\Phi} \mapsto [0, 1]$ as $\vartheta_{\psi|\phi}(\mathbf{Pr}) = \sum_{\mathbf{Pr_i} \in \mathfrak{P}} \vartheta_{(\psi|\phi)} \mathbf{Pr_i}$ where $\mathfrak{P}$ is a partition of $\mathbf{Pr}$ such that each element of $\mathfrak{P}$ is inseparable.*

[4] The closure of a set S is the smallest *closed* set containing S.

Informally, value $\vartheta_{(\psi|\phi)}(\mathbf{Pr})$ measures how *wide* the probability distributions in $\mathbf{Pr}$ is when inferring ψ given ϕ. For example, when all the distributions in $\mathbf{Pr}$ assign the same probability for the conditional event $(\psi|\phi)$, then the set $\mathbf{Pr}$ is acting like a single distribution when inferring ψ given ϕ, and $\mathbf{Pr}$ has *width* 0 for inferring ψ given ϕ.

From the definition, we know that function $\vartheta_{(\psi|\phi)}$ is a *measure*. Since it is a measure, we can define a probability distribution based on it, and we show that this probability distribution can be used as an instantiation of ignorance in the next subsection.

4.2 INSTANTIATION OF IGNORANCE

Definition 8. *Let P be a PLP and $(\psi|\phi)$ be a conditional event. Then a KL-divergence based ignorance, denoted $\mathsf{IG}_P^{KL}(\psi|\phi)$, is defined as $\mathsf{IG}_P^{KL}(\psi|\phi) = \frac{\vartheta_{(\psi|\phi)}(\mathbf{Pr})}{\vartheta_{(\psi|\phi)}(\mathbf{Pr}_\Phi)}$, when $\vartheta_{(\psi|\phi)}(\mathbf{Pr}_\Phi) > 0$, where $\mathbf{Pr} = \{Pr \mid Pr \models P\}$. And we define $\mathsf{IG}_P^{KL}(\psi|\phi) = 0$ when $\vartheta_{(\psi|\phi)}(\mathbf{Pr}_\Phi) = 0$.*

Since $\vartheta_{\psi|\phi}$ is a measure, IG_P^{KL} is an uniform probability distribution. Therefore, $\mathsf{IG}_P^{KL}(\psi|\phi)$ is the probability that a randomly selected probability distribution from set $\mathbf{Pr}_\Phi$ assigns $\psi|\phi$ a probability value that is in the interval $[l,u]$, where $P \models_{tight} (\psi|\phi)[l,u]$. If this probability is close to 1, then reasoning on P is similar to reasoning on an empty PLP; when it is close to 0, it indicates that a tighter bound for $(\psi|\phi)$ can be inferred from P.

Proposition 4. *Let P be a PLP and $(\psi|\phi)$ be a conditional event. Suppose that $P \models_{tight} (\psi|\phi)[l,u]$ and $pm = me[P](\psi|\phi)$. Then $IG_P^{KL}(\psi|\phi) = IG_{P_1}^{KL}(\psi|\phi) + IG_{P_2}^{KL}(\psi|\phi)$, where $P_1 = P \cup \{(\psi|\phi)[pm,u]\}, P_2 = P \cup \{(\psi|\phi)[l,pm]\}$.*

This proposition says that the ignorance of a PLP about a conditional event is the sum of the ignorance of lacking knowledge supporting probability distributions above and below the maximum entropy probability. The ignorance can also be calculated according to maximum entropy as below.

Proposition 5. *Let P be a PLP and $(\psi|\phi)$ be a conditional event. Suppose that $P \models_{tight} (\psi|\phi)[l,u]$, $\emptyset \models_{tight}^{me} (\psi|\phi)[p_{me},p_{me}]$, and $\mathbf{Pr} = \{Pr \mid Pr \models P\}$, then $\vartheta_{(\psi|\phi)}(\mathbf{Pr}) = \mathrm{sign}(u-p_{me}) * \max_{Pr\models P^u} H(Pr) - \mathrm{sign}(l-p_{me}) * \max_{Pr\models P^l} H(Pr)$ where $P^u = P \cup \{(\psi|\phi)[u,u]\}$ and $P^l = P \cup \{(\psi|\phi)[l,l]\}$.*

4.3 INSTANTIATION OF SATISFACTION FUNCTION

Given a PLP P, a set of probability distributions can be induced such that $\mathbf{Pr} = \{Pr \mid Pr \models P\}$ and a unique probability distribution $me[P]$ in the set that has maximum entropy can be determined. In $\mathbf{Pr}$, some distribution is likely to be the actual probability distribution. Based on the maximum entropy principle, $me[P]$ is the most likely one, and the probability $me[P](\psi|\phi)$ is the most likely probability for the event $(\psi|\phi)$. Intuitively, the probability value that is closer to $me[P](\psi|\phi)$ is more likely to be the actual probability of $(\psi|\phi)$. Based on this, an interval that contains values closer to $me[P](\psi|\phi)$ are more likely to contain the actual probability of $(\psi|\phi)$. Of course, a loose interval is always more likely to contain the actual probability of $(\psi|\phi)$ than a tight interval.

From the KL-divergence, we can define how close a value is to $me[P]$ as:

$$\begin{aligned}
\nu^{pos}_{P,(\psi|\phi)}(v) &= \min_{Pr\models P, Pr(\psi|\phi)=v} KL(Pr||me), \text{ where } v \geq me[P] \\
\nu^{neg}_{P,(\psi|\phi)}(v) &= \min_{Pr\models P, Pr(\psi|\phi)=v} KL(Pr||me), \text{ where } v \leq me[P] \\
dis^{pos}_{P,(\psi|\phi)}(u,v) &= |\nu^{pos}_{P,(\psi|\phi)}(u) - \nu^{pos}_{P,(\psi|\phi)}(v)| \\
dis^{neg}_{P,(\psi|\phi)}(u,v) &= |\nu^{neg}_{P,(\psi|\phi)}(u) - \nu^{neg}_{P,(\psi|\phi)}(v)|
\end{aligned}$$

Let dis be $dis^{pos}_{P,(\psi|\phi)}$ (resp. $dis^{neg}_{P,(\psi|\phi)}$). It is easy to see that dis is a distance function on $\mathcal{R}^{[pme,u]}$ (resp. $\mathcal{R}^{[l,pme]}$), where $P \models_{tight} (\psi|\phi)[l,u]$, $pme = me[P](\psi|\phi)$ and $\mathcal{R}^{[a,b]} = \{x \mid x \in [a,b], x \in \mathcal{R}\}$, i.e. dis satisfies the following:

- $dis(u,v) \geq 0$
- $dis(u,v) = 0$ iff $u = v$
- $dis(u,v) = dis(v,u)$
- $dis(u,v) \leq dis(u,x) + dis(x,v)$

Again, from the distance functions $dis^{pos}_{P,(\psi|\phi)}$ and $dis^{neg}_{P,(\psi|\phi)}$, a probability distribution can be defined. So, by KL-divergence, the possible probabilities of a conditional event $(\psi|\phi)$ are measurable. Consider every probability is equally possible, then the (second order) probability that the actual (first order) probability of $(\psi|\phi)$ falls in an interval $[a,b]$ is the *length* of $[a,b]$ divided by the *length* of $[l,u]$, where $P \models_{tight} (\psi|\phi)[l,u]$, according to the distance function $dis^{pos}_{P,(\psi|\phi)}$ and $dis^{neg}_{P,(\psi|\phi)}$. Formally, we define the degree of satisfaction as this second order probability:

Definition 9. *Let P be a PLP and $(\psi|\phi)$ be a conditional event. Suppose that $P \models_{tight} (\psi|\phi)[l,u]$ and $P \models^{me}_{tight} (\psi|\phi)[p_{me}, p_{me}]$, then we have that:*

$$\mathsf{SAT}^{KL}_P((\psi|\phi)[a,b]) = \begin{cases} 0.5\left(\frac{dis^{pos}_{P,(\psi|\phi)}(p_{me},\min(u,b))}{dis^{pos}_{P,(\psi|\phi)}(p_{me},u)} + \frac{dis^{neg}_{P,(\psi|\phi)}(p_{me},\max(a,l))}{dis^{neg}_{P,(\psi|\phi)}(p_{me},l)}\right), & \textit{if } p_{me} \in [a,b] \\ 0, & \textit{otherwise} \end{cases}$$

Proposition 6. *Let P be a PLP, then the function SAT^{KL}_P defined in Definition 9 satisfies all the postulates in Definition 2, and it is consistent with the maximum entropy principle, that is, it satisfies the conditions in Proposition 1.*

5 EXAMPLES

We illustrate the usefulness of our framework with two examples.

Example 3. Let P be a PLP as given in Example 1. In our framework, we calculate the KL-ignorance and KL-satisfaction for our queries. We have $\mathsf{IG}^{KL}_{(fly(t)|magpie(t))}(P) = 0.11$ and $\mathsf{IG}^{KL}_{(fly(t)|sickMagpie(t))}(P) = 0.0283$. This indicates that P is more useful to infer the proportion of magpies that can fly than to infer the proportion of sick magpies that can fly. We also have that $\mathsf{SAT}_P^{KL}((fly(t)|magpie(t))[0.8,1]) = 0.58$, $\mathsf{SAT}_P^{KL}((fly(t)|sickMagpie(t))[0.8,1]) = 0.53$. By comparing these KL degrees of satisfaction, we know that magpies are more likely to fly than sick magpies.

Example 4 (Route planning). [1]. Assume that John wants to pick up Mary after she stopped working. To do so, he must drive from his home to her office. Now, John has the following knowledge at hand: Given a road (ro) from R to S, the probability that he can reach (re) S from R without running into a traffic jam is greater than 0.7. Given a road in the south (so) of the town, this probability is even greater than 0.9. A friend just called him and gave him advice (ad) about some roads without any significant traffic. Clearly, if he can reach S from T and T from R, both without running into a traffic jam, then he can also reach S from R without running into a traffic jam. Furthermore, John has some concrete knowledge about the roads, the roads in the south of the town, and the roads that his friend was talking about. For example, he knows that there is a road from his home (h) to the university (u), from the university to the airport (a), and from the airport to Mary's office (o). Moreover, John believes that his friend was talking about the road from the university to the airport with a probability between 0.8 and 0.9 (he is not completely sure about it, though). The above and some other probabilistic knowledge is expressed by the following PLP P:

$$P = \left\{ \begin{array}{lll} ro(h,u)[1,1], & ro(u,a)[1,1], & ro(a,o)[1,1], \\ ad(h,u)[1,1], & ad(u,a)[0.8,0.9], & so(a,o)[1,1], \\ (re(R,S)|ro(R,S))[0.7,1], & \multicolumn{2}{l}{(re(R,S)|ro(R,S) \wedge so(R,S))[0.9,1],} \\ \multicolumn{3}{l}{(re(R,S)|ro(R,S) \wedge ad(R,S))[1,1],} \\ \multicolumn{3}{l}{(re(R,S)|re(R,T) \wedge re(T,S))[1,1]} \end{array} \right\}$$

John wants to know the probability of him running into a traffic jam, which can be expressed by the query: $Q_0 = ?(re(h,o)|\top)$.

In [1], Q_0 can be answered by $P \models_{tight} (re(h,o)|\top)[0.7,1]$, and by $P \models_{tight}^{me} (re(h,o)|\top)[0.93,0.93]$. The user can either accept a noninformative bound $[0.7,1]$ or accept a unreliable precise probability 0.93, and no further reasoning can be done.

Using our method, we can get that $\mathsf{IG}_P^{KL}(re(h,o)|\top) = 0.066$. The ignorance value $\mathsf{IG}_P^{KL}(re(h,o)|\top)$ indicates that the knowledge is reliable about $(re(h,o)|\top)$. However, the actual probability of $(re(h,o)|\top)$ may be still different from 0.93, since $\mathsf{IG}_P^{KL}(re(h,o)|\top) > 0$.

John is wondering whether he can reach Mary's office from his home, such that the probability of him running into a traffic jam is smaller than 0.10. This can be expressed by the following probabilistic query: $Q_1 = ?(re(h,o)|\top)[0.90,1]$. John is also wondering whether the probability of him running into a traffic jam is smaller than 0.10, if his friend was really talking about the road from the university to the airport. This can be expressed as a probabilistic query: $Q_2 = ?(re(h,o)|ad(u,a))[0.90,1]$.

In [1], in the traditional probabilistic logic programming both Q_1 and Q_2 are given the answer "No"; by applying the maximum entropy principle Q_1 is given the answer "No" and Q_2 is given the answer "Yes". For Q_1 John will accept the answer "No", however, for Q_2, John may be confused and does not know which answer he should trust.

Using our method, we can calculate the degree of satisfaction of these two queries. For Q_1, $\mathsf{SAT}_P^{KL}(Q_1) = 0$, which means the bound $[0.9,1]$ does not contain the probability given by applying the maximum entropy principle, and thus John has no confidence that he can reach Mary's office on time. For Q_2, $\mathsf{SAT}_P^{KL}(Q_2) = 0.724$, the relative high value "0.724" can help John to decide whether he should set off to pick up Mary.

Using our method, John can get an estimation of the probability that he can reach Mary's office from his home without running into a traffic jam. If it is a special day for him and Mary, he hopes that his estimation be more accurate, otherwise, he can tolerate a less accurate estimation. Formally, he needs to decide the threshold a for $\models_{maxLow}^{SAT \geq a}$. For example, for Q_2, he may set $a_N = 0.6$ for a normal day, and $a_I = 0.75$ for an important day. Therefore, he can infer that $P \models_{maxLow}^{SAT \geq 0.6} (re(h,o)|ad(u,a))[0.922, 1]$ and $P \models_{maxLow}^{SAT \geq 0.75} (re(h,o)|ad(u,a))[0.897, 1]$. If it is an ordinary day and the lowest probability is bigger than 0.90, then he can set off. On an important day, he will need to investigate more about the traffic (to decrease the ignorance of $(re(h,o)|ad(u,a))$) or he has to revise his plan, since $0.897 < 0.9$.

On the another hand, we can also analyze the usefulness of the advice from his friend. By analyzing his friend's knowledge, we have $\mathsf{IG}_P^{KL}(re(h,o)|ad(u,a)) = 0.0184$. This means that his friend's advice is indeed useful, since this ignorance value is significantly smaller than $\mathsf{IG}_P^{KL}(re(h,o)|\top)$. So, John needs to call his friend to make sure that his friend is really talking about the road from the university to the airport.

The degrees of satisfaction for various intervals are given in Table 1. From the table, we can see that, the degree of satisfaction decreases as the interval becomes tighter. This means that the second order probability that the actual probability of $(\psi|\phi)$ falls in $[l, u]$ is getting smaller.

Table 1. Degrees of satisfaction for queries Q_1 and Q_2

Bound	$(re(h,o)\|\top)$	Bound	$(re(h,o)\|ad(u,a))$
$[0, 1]$	1	$[0, 1]$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$[0.70, 1]$	1	$[0.88, 1]$	1
$[0.75, 1]$	0.785	$[0.897, 1]$	0.75
$[0.80, 1]$	0.658	$[0.922, 1]$	0.60
$[0.86, 1]$	0.500	$[0.94, 1]$	0.50
$[0.90, 1]$	0.000		

6 RELATED WORK AND CONCLUSION

Related work. In recent years there have been a lot of research on integrating logical programming with probability theory. These probabilistic logic programs have been studied from different views and have different syntactic forms and semantics, including *conditional probabilistic logic programming* [5, 3], *causal probabilistic logic programming* [6–8], *success probabilistic logic programming* [9, 10], and some others [11].

In causal probabilistic logic programming [6, 7], a rule $pr(a|\phi) = v$ is intuitively interpreted as *a is caused by factors determined by ϕ with probability v*. A causal probability statement implicitly represents a set of conditional independence assumptions:

given its cause C, an effect E is probabilistically independent of all factors except the (direct or indirect) effects of E (see [6] for detail). Formally, if $pr(\psi|\phi_1) = y_1 \in P$ and $pr(\psi|\phi_2) = y_2 \in P$ where $y_1 \neq y_2$, then no possible world of P satisfies $\phi_1 \wedge \phi_2$.

In [9, 10], the real number attached to a rule represents the probability that this rule is alliable (or satisfiable). In another word, a PLP in this view represents a set of (classical) logic programs, and the probability of each logic program is decided by all probabilities of all the rules. Then for any query, the answer is the probability of choosing a classical logic program from the set that can successfully infer the query. In this formalization, we can only query about the probability of ψ and cannot query about the probability of $(\psi|\phi)$, since $(\psi|\phi)$ is meaningless in classical logic programs.

In [11], the probabilities are attached to atoms, such as: $b[0.6, 0.7] \leftarrow a[0.2, 0.3]$, which means that if the probability of a is in between 0.2 and 0.3 then the probability of b is in between 0.6 and 0.7. Intuitively, the interpretation of rules is more close to casuality than conditioning. As a consequence, if we have another rule: $b[0.2, 0.3] \leftarrow c[0.5, 0.6]$, then $Pr(a) \in [0.2, 0.3]$ and $Pr(c) \in [0.5, 0.6]$ cannot be both true,

In this paper, we focus on the framework of conditional probabilistic logic programming for representing conditional events.

Because of its weakness in reasoning, subclasses cannot inherit the properties of its superclass in the basic semantics of PLP. For instance, subclass magpie can not inherit the attribute "can fly" from its superclass bird in Example 1, since $P \models_{tight} (fly(t)|magpie(t))[0, 1]$. In [12–14], Lukasiewicz provided another method to enhance the reasoning power mainly on the issue of inheritance. In this setting, *logic entailment strength* λ is introduced. With strength 1, subclasses can completely inherit the attributes of its superclass; with strength 0 subclasses cannot inherit the attributes of its superclass; with a strength between 0 and 1, subclasses can partially inherit the attributes of its superclass. Value strength appears to be similar to the degree of satisfaction in our framework, but they are totally different. First, λ is not a measurement for a query, but is given by a user to control the reasoning procedure, in other words, we cannot know beforehand the strength in order to infer a conclusion. Second, even if we can use a strength as a measurement, i.e. even if we can obtain the required strength to infer an expected conclusion, it is not an instance of degree of satisfaction, because the cautious monotonicity postulate in Definition 2 is not satisfied. Given a PLP P, assume that we can infer both $(\psi|\phi)[l_1, u_1]$ by strength $\lambda = \lambda_1$ and $(\psi|\phi)[l_2, u_2]$ by strength $\lambda = \lambda_2$. Now assume that $(\psi|\phi)[l_1, u_1]$ is added to P, however, in order to infer $(\psi|\phi)[l_2, u_2]$, we still need to have the strength $\lambda = \lambda_2$ given. That is, adding additional information to P does not avoid requiring the strength λ_2 if $(\psi|\phi)[l_2, u_2]$ is to be inferred. In contrast, if we have $(\psi|\phi)[l_1, u_1]$ added in the PLP, then the degree of satisfaction of $(\psi|\phi)[l_2, u_2]$ will increase.

In [15, 16], the authors provided a second order uncertainty to measure the reliability of accepting the precise probability obtained by applying maximum entropy principle as the answer to a query in propositional probabilistic logic. The second order uncertainty for $(\psi|\phi)$ and PLP P is defined as $(-\log l - \log u)$ where $P \models_{tight} (\psi|\phi)[l, u]$. Similarly, we provided ignorance function to measure the usefulness of a PLP to answering a query. If a precise probability for a query is inferred from a PLP P then P contains full information about the query, and therefore accepting the probability is totally reliable. More precisely, their second order uncertainty is directly computed from the probability interval of the query inferred from P. In contrast, our ignorance is computed from the PLP, which provides more information than an interval. Therefore, our measure of ignorance is more accurate in reflecting the knowledge in a PLP.

Conclusion. In this paper, we investigated the issues surrounding how much we can *trust* a result for a query given a PLP with imprecise knowledge. We proposed a framework to measure both ignorance and the degree of satisfaction of an answer to a query under a given PLP. Using the consequence relations provided in this paper, we can get an informative and reliable interval as the answer for a query or alternatively we know how much we can trust a single probability. The proofs that our framework is an extension of both traditional probabilistic logic programming and the maximum entropy principle (in terms of consequence relations) show that our framework is theoretically sound.

References

1. Kern-Isberner, G., Lukasiewicz, T.: Combining probabilistic logic programming with the power of maximum entropy. Artificial Intelligence **157**(1-2) (2004) 139–202
2. Lukasiewicz, T.: Probabilistic logic programming. In: ECAI. (1998) 388–392
3. Lukasiewicz, T.: Probabilistic logic programming with conditional constraints. ACM Trans. Comput. Log. **2**(3) (2001) 289–339
4. Gray, R.M.: Entropy and information theory. Springer-Verlag New York, Inc., New York, NY, USA (1990)
5. Costa, V.S., Page, D., Qazi, M., Cussens, J.: CLP($\mathcal{BN}$): Constraint logic programming for probabilistic knowledge. In: UAI. (2003) 517–524
6. Baral, C., Gelfond, M., Rushton, J.N.: Probabilistic reasoning with answer sets. In: LPNMR. (2004) 21–33
7. Baral, C., Hunsaker, M.: Using the probabilistic logic programming language p-log for causal and counterfactual reasoning and non-naive conditioning. In: IJCAI. (2007) 243–249
8. Saad, E.: Qualitative and quantitative reasoning in hybrid probabilistic logic programs. In: ISIPTA'07 - Fifth International Symposium On Imprecise Probability: Theories And Applications. (2007)
9. Raedt, L.D., Kimmig, A., Toivonen, H.: Problog: A probabilistic prolog and its application in link discovery. In: IJCAI. (2007) 2462–2467
10. Fuhr, N.: Probabilistic datalog: Implementing logical information retrieval for advanced applications. JASIS **51**(2) (2000) 95–110
11. Dekhtyar, A., Dekhtyar, M.I.: Possible worlds semantics for probabilistic logic programs. In: ICLP. (2004) 137–148
12. Lukasiewicz, T.: Probabilistic logic programming under inheritance with overriding. In: UAI. (2001) 329–336
13. Lukasiewicz, T.: Weak nonmonotonic probabilistic logics. Artif. Intell. **168**(1-2) (2005) 119–161
14. Lukasiewicz, T.: Nonmonotonic probabilistic logics under variable-strength inheritance with overriding: Complexity, algorithms, and implementation. Int. J. Approx. Reasoning **44**(3) (2007) 301–321
15. Rödder, W., Kern-Isberner, G.: From information to probability: An axiomatic approach - inference isinformation processing. Int. J. Intell. Syst. **18**(4) (2003) 383–403
16. Rödder, W.: On the measurability of knowledge acquisition, query processing. Int. J. Approx. Reasoning **33**(2) (2003) 203–218