Evaluation and Comparison Criteria for Approaches to Probabilistic Relational Knowledge Representation

Christoph Beierle[1], Marc Finthammer[1],
Gabriele Kern-Isberner[2], Matthias Thimm[2]

[1]Dept. of Computer Science, FernUniversität in Hagen, 58084 Hagen, Germany
[2]Dept. of Computer Science, TU Dortmund, 44221 Dortmund, Germany

Abstract. In the past ten years, the areas of *probabilistic inductive logic programming* and *statistical relational learning* put forth a large collection of approaches to combine relational representations of knowledge with probabilistic reasoning. Here, we develop a series of evaluation and comparison criteria for those approaches and focus on the point of view of knowledge representation and reasoning. These criteria address abstract demands such as language aspects, the relationships to propositional probabilistic and first-order logic, and their treatment of information on individuals. We discuss and illustrate the criteria thoroughly by applying them to several approaches to probabilistic relational knowledge representation, in particular, Bayesian logic programs, Markov logic networks, and three approaches based on the principle of maximum entropy.

1 Introduction

Originally, probabilistic logic was based on propositional logic, using conditionals of the form $(B \mid A)[x]$ to express that *if A, then B with probability x*. In order to exploit its more expressive power, various approaches to combine probabilistic logic with first-order logic have been proposed (see [3, 7]) like Baysian logic programs (BLP) [7, Ch. 10], Markov logic networks (MLN) [5], or relational Bayesian networks [8]. The principle of maximum entropy [14] is used to define the probabilistic relational approaches in [13, 12, 6].

There are different motivations and objectives for choosing a particular representation for probabilistic relational knowledge. Suppose we want to model situations in a zoo (this scenario is adapted from [4]). There are elephants and keepers, and we want to say something about whether elephants like their keepers. Thus, we want to formalize generic statements like *Generally, elephants like their keepers* or *Elephants like their keepers with a probability of 0.9*. Furthermore, we might want to state information about individuals, e. g., that Fred is an elephant keeper the elephants do not like very much; this might be expressed by *Elephants like Fred only with a probability of 0.3*. There are also situations

This research was partially supported by the DFG (BE 1700/7-2 and KE 1413/2-2).

where it is useful to list all individual elephants and keepers that are in the zoo. Given a knowledge base representing our zoo model, we would like to be able to use inference methods to answer questions about individuals occurring in the model and relationships among them.

Example 1. To give a concrete example, consider the knowledge base KB with

c_1: $(\mathit{likes}(X,Y) \mid \mathit{elephant}(X),\ \mathit{keeper}(Y))[0.9]$

c_2: $(\mathit{likes}(X,\mathit{fred}) \mid \mathit{elephant}(X))[0.3]$ $\qquad$ c_3: $(\mathit{likes}(\mathit{clyde},\mathit{fred}))[1.0]$

where X and Y are variables. There is a general statement (c_1) that represents the probability of elephants liking their keepers, and two more specific statements (c_2 resp. c_3) that model the relationships for some individuals Clyde and Fred.

Note that naive grounding of KB using all possible instantiations yields contradictory information since, for instance, we get both $\mathit{likes}(\mathit{clyde},\mathit{fred}))[0.3]$ and $\mathit{likes}(\mathit{clyde},\mathit{fred}))[1]$. Nonetheless, KB makes perfect sense from a commonsense point of view as, for instance, rule c_2 could be treated as an exception to c_1, inhibiting the instantiation of Y with the constant *fred* in c_1.

Despite the variety of different approaches to probabilistic relational knowledge representation, inference, and learning, not much work has been done on systematically comparing them. In [2], a comparison between several statistical relational learning systems is done, with an emphasis on the learning aspects. In [9], a schema for expressivity analysis is proposed and used to show that relational Bayesian networks [8] are at least as expressive as MLNs. While providing access to different modeling and learning approaches, the focus of the software suite ProbCog [10] is its practical use and integration in technical systems. A software platform providing a common interface to a series of probabilistic relational systems and supporting their evaluation and comparison is presented in [18]. In [18], also some meta-level criteria for the evaluation and comparison of different approaches are given. In this paper, we extend the discussion of these criteria, focusing on knowledge representation aspects, apply them to further approaches, and develop a series of new criteria especially with respect to the role of individuals, prototypical elements, and universes. It has to be noted that our investigation of these criteria stays on an abstract level since the objective is not to take specific formalizations into account, but to address desirable properties and interesting features from a general and commonsense perspective. A technical comparison with respect to default reasoning properties of the three approaches employing the principle of maximum entropy can be found in [12].

After briefly recalling the notions of BLP, MLN, and three approaches based on maximum entropy in Sec. 2, our comparison and evaluation criteria are presented along several dimensions, dealing with language aspects (Sec. 3), the relationship to strict and propositional knowledge (Sec. 4), and individuals and universes (Sec. 5). In Sec. 6, we conclude and point out further work.

2 Background: Probabilistic Relational Approaches

Bayesian logic programming combines logic programming and Bayesian networks [7, Ch. 10]. The basic structure for knowledge representation in Bayesian logic programs are *Bayesian clauses* which model probabilistic dependencies between *Bayesian atoms* as in the following BLP corresponding to Ex. 1:

$c_1:\ (\mathit{likes}(X,Y)\,|\,\mathit{elephant}(X),\ \mathit{keeper}(Y))$

$c_2:\ (\mathit{likes}(X,\mathit{fred})\,|\,\mathit{elephant}(X))$ $\qquad c_3:\ \mathit{likes}(\mathit{clyde},\mathit{fred})$

While in Ex. 1, a probability for each clause is given, expressing a constraint on a satisfying distribution, for each Bayesian clause c, a function cpd_c must be defined, expressing the conditional probability distribution $P(\mathsf{head}(c)\,|\,\mathsf{body}(c))$ and thus partially describing an underlying probability distribution P. For instance, $\mathsf{cpd}_{c_1}(\mathsf{true},\mathsf{true},\mathsf{true}) = 0.9$ would express our subjective belief that $\mathit{likes}(X,Y)$ is true with probability 0.9 if $\mathit{elephant}(X)$ and $\mathit{keeper}(Y)$ are true. In order to aggregate probabilities that arise from applications of different Bayesian clauses with the same head, BLPs make use of *combining rules*. Semantics are given to Bayesian logic programs via transformation into propositional forms, i. e. into Bayesian networks [15] (see [7, Ch. 10] for details).

Markov logic [5] establishes a framework which combines Markov networks [15] with first-order logic to handle a broad area of statistical relational learning tasks. The Markov logic syntax complies with first-order logic where each formula is quantified by an additional weight value, e.g.

$(\mathit{elephant}(X) \wedge \mathit{keeper}(Y) \Rightarrow \mathit{likes}(X,Y),\quad 2.2)$

$(\mathit{elephant}(X) \Rightarrow \mathit{likes}(X,\mathit{fred}),\quad -0.8)$ $\qquad (\mathit{likes}(\mathit{clyde},\mathit{fred}),\ \infty)$

Semantics are given to sets of Markov logic formulas by a probability distribution over propositional possible worlds that is calculated as a log-linear model over weighted ground formulas. The fundamental idea in Markov logic is that first-order formulas are not handled as hard constraints (which are indicated by weight ∞), but each formula is more or less softened depending on its weight. A *Markov logic network (MLN)* L is a set of weighted first-order logic formulas $(F_i,\ w_i)$ together with a set of constants C. The semantics of L is given by a ground Markov network $M_{L,C}$ constructed from F_i and C [7, Ch. 12]. The standard semantics of Markov networks [15] is used for reasoning, e.g. to determine the consequences of L (see [7, Ch. 12] for details).

The syntax of *relational probabilistic conditional logic* (RPCL) [19] has already been used in the representation of Ex. 1 and employs conditionals of the form $(B\,|\,A)[x]$ with first-order formulas A, B and $x \in [0,1]$. A conditional $(B\,|\,A)[x]$ represents a constraint on a probability distribution $P : \Omega \rightarrow [0,1]$ on the set of possible worlds Ω and states that the conditional probability of B given A is x. In order to interpret conditionals containing free variables several relational semantics have been proposed, see [19, 13]. The *grounding semantics* [13] uses a *grounding operator* $\mathcal{G}$, e. g. universal instantiation, that translates a set $\mathcal{R}$ of conditionals with free variables into a set of ground conditionals. Then, a probability distribution P *$\mathcal{G}$-satisfies* $\mathcal{R}$, denoted by $P \models_{\mathcal{G}} \mathcal{R}$, iff $P(B'\,|\,A') = x$ for every ground $(B'\,|\,A')[x] \in \mathcal{G}(\mathcal{R})$. Both *averaging* and *aggregating semantics* [12, 19] do not require a grounding operator but interpret the intended probability x of a conditional with free variables only as a guideline for the probabilities of its instances and the actual probabilities may differ from x. More precisely, a probability distribution P *$\varnothing$-satisfies* $\mathcal{R}$, denoted by $P \models_{\varnothing} \mathcal{R}$, iff for every $(B \mid A)[x] \in \mathcal{R}$ it holds that $P(B_1 \mid A_1) + \ldots + P(B_n \mid A_n) = nx$ where $(B_1 \mid A_1), \ldots, (B_n \mid A_n)$ are the ground instances of $(B \mid A)$. A probability function P $\odot$-*satisfies* $\mathcal{R}$, denoted by $P \models_{\odot} \mathcal{R}$, iff for every $(B \mid A)[x] \in \mathcal{R}$ it holds that $P(B_1 \wedge A_1) + \ldots + P(B_n \wedge A_n) = x(P(A_1) + \ldots + P(A_n))$ where $(B_1 \mid A_1), \ldots, (B_n \mid A_n)$ are the ground instances of $(B \mid A)$. Note that these three semantics are extensions of classical probabilistic semantics for propositional probabilistic conditional logic [11]. Based on any of these semantical notions the *principle of maximum entropy* [14, 11] can be used for reasoning. The *entropy* H is an information-theoretic measure on probability distributions and is defined as a weighted sum on the information encoded in every possible world $\omega \in \Omega$: $H(P) = -\sum_{\omega \in \Omega} P(\omega) \log P(\omega)$. By employing the *principle of maximum entropy* one can determine the unique probability distribution that is the optimal model for a consistent knowledge base $\mathcal{R}$ in an information-theoretic sense via

$$P_{\mathcal{R}}^{ME_{\circ}} = \arg \max_{P \models_{\circ} \mathcal{R}} \mathcal{H}(P) \tag{1}$$

with $\circ$ being one of $\mathcal{G}$, $\varnothing$, or $\odot$. We abbreviate the approaches of reasoning based on the principle of maximum entropy with grounding, averaging, and aggregating semantics with $\mathrm{ME}_{\mathcal{G}}$, $\mathrm{ME}_{\varnothing}$, and $\mathrm{ME}_{\odot}$, respectively. We say that a formula $(B \mid A)[x]$ is $\circ$-inferred from $\mathcal{R}$ iff $P_{\mathcal{R}}^{ME_{\circ}} \models_{\circ} (B \mid A)[x]$ with $\circ$ being one of $\mathcal{G}$, $\varnothing$, or $\odot$.

3 Language Aspects

We start with discussing properties concerning the language of an approach to probabilistic relational knowledge representation, i. e. aspects relating to syntax and semantics. Firstly, the semantics of the components of the knowledge representation language should be as declarative as possible. In particular, it should be possible to express basic concepts directly and to have an intuitive meaning for all language constructs, inference and learning results:

(L-1) Direct expression of probabilities in the form of "A holds with a probability of x".

(L-2) Direct expression of conditional probabilities as in "Provided that A holds, then the probability of B is x".

Since an RPCL knowledge base supports representing formulas of the form $(B \mid A)[x]$ which constrain the conditional probability of B given A to x in any model, $\mathrm{ME}_{\mathcal{G}}$, $\mathrm{ME}_{\varnothing}$, $\mathrm{ME}_{\odot}$ obviously fulfill (L-1) and (L-2). The same holds for BLPs when taking into account the conditional probability distribution functions cpd_c which must be defined for any Bayesian clause c. Since in an MLN there is no obvious correspondence between the weight of a formula and its corresponding probability and because conditionals are not supported, (L-1) and (L-2) do not apply to MLNs.

(L-3) Qualitative statements like "A is more probable than B" or "A is very probable".

Such qualitative statements can be expressed within none of the five approaches.

(L-4) Commonsense meaning: Probabilities are used for expressing uncertainty, and for each basic construct of the knowledge representation language there should be a clear intuitive or commonsense meaning. Examples of such meanings are a statistical interpretation of expressions (with respect to a population), or a subjective degree of belief (with respect to the set of possible worlds).

The difference between statistical and subjective interpretations can be illustrated in the elephant-keeper-example 1 by contrasting "Most elephants like their keepers" (statistical, as it refers to a whole population) vs. "Mostly, an elephant likes its keeper" (subjective, as it refers to situations, i.e., possible worlds).

A Bayesian clause c in a BLP expresses qualitative information about the conditional probability of the clause's head given the body of the clause; the actual conditional probability is given by cpd_c which is applied for each instance. Thus, these informations (together with the combining rules) yield subjective conditional probabilities as a commonsense meaning of a BLP.

Although it can be observed that the greater the weight w of an MLN clause F the more impact F will have on the probability distribution in the resulting ground Markov network $M_{L,C}$, a more precise intuitive meaning of (F, w) is not evident. Besides this general negative statement, the probabilities resulting from an MLN can be classified as subjective, as the MLN semantics is based on possible worlds.

For each ground conditional $(B \mid A)[x]$ in an RPCL knowledge base, its commonsense meaning is given by the conditional probability of B given A being x for grounding, averaging, and aggregating semantics. However, the commonsense interpretation of conditionals with free variables is substantially different in these three semantics. For grounding semantics, a relational conditional is understood as a universal quantification of the subjective conditional probability of each ground instance within the range of the grounding operator. For averaging and aggregating semantics, the commonsense meaning is a mixture of statistical interpretation and degrees of belief. The averaging semantics yields a statistics of subjective conditional beliefs. For instance, the conditional c_1 in Ex. 1 with an interpretation via $\mathrm{ME}_{\varnothing}$ reads as "Considering *a random* elephant-keeper-pair, the average subjective probability that the elephant likes its keeper is 0.9." The aggregating semantics exchanges the role of statistical (or population based) and subjective view by providing kind of a subjectively weighted statistics. Here, c_1 is understood as "Considering *all* elephant-keeper-pairs, the expected subjective probability that elephants like their keepers is 0.9." In contrast to the averaging semantics, the aggregating semantics gives more weight to individuals (or tuples of individuals) that are more likely to fulfill the premise of a conditional.

By taking both statistical and subjective aspects into account, both averaging and aggregating semantics allow a more realistic approach to commonsense reasoning in a relational probabilistic context. When entering a zoo (or considering the vague population of all elephants and keepers in the world) and uttering conditional c_1 of Ex. 1, human beings are very likely to express something like

"In (about) 90 % of all situations that involve an elephant and its keeper, I will notice that the elephant likes the keeper." This statement takes both beliefs about possible worlds and about the population into account, and it is exactly this perspective that averaging and aggregating semantics aim to represent. For a further discussion and evaluation of these semantics, see [12].

(L-5) Closure properties: The results obtained by inference should be expressible in the knowledge representation language, thus enabling, e. g., the integration of inferred knowledge into a knowledge base. Another closure aspect refers to the query language: Can any formula being allowed in a knowledge base be used in a query?

Given a (ground) query Q for a BLP, BLP inference can be used for computing a cpd_Q for Q by generating all possible combinations of evidence for Q, allowing one to add this information as a BLP clause. Since with MLNs also probabilities are computed, MLN inference results can not be used directly in an MLN knowledge base where a weight is required for a formula. On the other hand, ME inference results can be directly integrated into an RPCL knowledge base (independently of the actual semantics).

In all approaches, queries must be ground, and taking a logic formula F from a corresponding knowledge base, every ground instance of F can be used in a query. For example, given the body of the BLP clause

$$(likes(clyde, jim) \mid elephant(clyde),\ keeper(jim))$$

as evidence, the BLP inference mechanism will determine the conditional probability of $likes(clyde, jim)$ given the evidence. Consequently, open queries are not allowed in any of the approaches; if a support system offers posing queries with free variables (as it is allowed e.g. in Alchemy [5]), then such a query is being treated as an abbreviation for posing a sequence of all possible ground instantiations of that query.

(L-6) Semantical invariance: A general requirement for logic-based representations applies also here: The semantics should be the same if the same knowledge is expressed by syntactic variants.

Let KB be a knowledge base in any of the three relational approaches. Since for any variable renaming σ, the respective semantics of KB and $\sigma(KB)$ coincide, semantical equivalence with respect to variable renaming holds for BLPs, MLNs, and $\mathrm{ME}_\circ$.

Another form of syntactic variants arises from propositionally equivalent formulas, e. g. A and $A \wedge A$. In places where such formulas are allowed, they do not give rise to a different semantics in any of the five approaches. However, it should be noted that this case has to be distinguished carefully from the case of adding a syntactic variant of a knowledge base element to that knowledge base: If $F \in KB$ and σ is a variable renaming replacing some variable in F with a new variable not occurring in KB, then in general $KB \cup \{\sigma(F)\}$ has a different semantics both for BLPs and for MLNs. For instance, when using *noisy-or* as the combining function, the probability expressed by F—and thus also by $\sigma(F)$—will typically increase when adding $\sigma(F)$ to KB.

Example 2. Consider a BLP consisting of the single clause $c = (A(\mathsf{X}) \,|\, B(\mathsf{X}))$ with $\mathsf{cpd}_c(\mathsf{true}, \mathsf{true}) = 0.9$, $\mathsf{cpd}_c(\mathsf{true}, \mathsf{false}) = 0.5$ and with *noisy-or* being the combining rule for predicate A. Then querying this BLP with $(A(d) \,|\, B(d))$ results (obviously) in the probability 0.9 for $A(d)$ being true given $B(d)$ is true. However, adding the clause c' with $c' = (A(\mathsf{Y}) \,|\, B(\mathsf{Y}))$ (with $\mathsf{cpd}_c = \mathsf{cpd}_{c'}$) which is a syntactical variant of c results in a probability of $1-(1-0.9)\cdot(1-0.9) = 0.99$ as both c and c' determine a probability of 0.9 and these probabilities have to be combined by the corresponding combining function (*noisy-or* in this case) to obtain the final answer to the given query.

Example 3. Similarly, consider an MLN consisting of the single formula $(B(\mathsf{X}) \Rightarrow A(\mathsf{X}),\ 1)$. Querying this MLN with $(A(d) \,|\, B(d))$ results in the (approximated) probability 0.764974 for $A(d)$ being true given $B(d)$ is true. However, adding the syntactic variant $(B(\mathsf{Y}) \Rightarrow A(\mathsf{Y}),\ 1)$ results in an (approximated) probability of 0.861964 (these probabilities have been computed with the Alchemy system).

As inference in RPCL is defined on well-defined semantics, syntactical variants do not influence the outcome of inference (for grounding, averaging, and aggregating semantics).

(L-7) Explanation capabilities for inference: It is desirable to have explanation capabilities of inference results. Which elements of the knowledge base are responsible (to what degree) for an inferred result? Or which elements of the knowledge base did not affect a result in any way? Can every result (or at least some results) be derived (more or less) directly from certain elements of the knowledge base? Or does any result essentially require the calculation of an appropriate model?

The explanation of a BLP inference result is given by the obtained local Bayes net which also encodes a (logical) derivation of the query. Therefore, it is obvious which clauses of the BLP knowledge base were involved in the calculation of the result. So the BLP approach offers some distinct level of explanation capability.

MLN inference is based on a log-linear model that has to be normalized in order to represent a probability distribution, cf. [7, Ch. 12]. The value of this normalization constant depends on the relationships among the formulas of an MLN knowledge base. Therefore, an inferred probability depends on all formulas of the knowledge base, because the weights of the formulas are relative values, where the higher the weight the greater the influence of the formula. Since MLN inference involves the construction of an appropriate ground Markov network, independencies among certain ground atoms are indicated by this network. So some independency aspects of inferred results can be explained by the net structure.

Inference in RPCL relies on solving the optimization problem (1). In some special cases (regarding the query and the conditionals in the knowledge base), the result of a query might be estimated directly considering how reasoning under the maximal entropy distribution "behaves". So in such rare cases, the inferred result can be explained by certain aspects of the knowledge base (having the principle of maximum entropy in mind). But in general, no intuitive explanation of inference results is evident for both the MLN and RPCL approaches.

4 Strict and Propositional Knowledge

In a probabilistic relational modeling language two essential dimensions are present that distinguish the respective approach from propositional logic: The *probabilistic* and the *relational* dimension. From a knowledge representation point of view, the following questions arise naturally. What happens if one cuts down any of these two dimensions? Which kind of logic does one obtain?

(SP-1) Strict Knowledge: Suppose one restricts the sentences occurring in a knowledge base such that only strict probabilistic knowledge can be expressed. What is the representation level of this degenerated case, and what are its semantics and inference properties? In particular, what is its relationship to classical non-probabilistic (first-order) logic?

Of the formalisms BLP, MLN, and RPCL, only MLNs allow for existential quantifiers (which in the Alchemy system are replaced by corresponding finite disjunctions over instantiations with the elements of the underlying universe). Looking at the language of logical MLN formulas we thus have first-order logic, restricted to a finite fixed universe. In order to express that a particular formula F represents strict knowledge, the weight of F must be set to infinity [5]. In this case, all possible worlds violating the strict formula are assigned zero probabilities by the MLN, and the probabilities of the satisfying worlds sum up to 1. Hence, the formulas that can be inferred with probability 1 from such an MLN $\mathcal{F}$ containing only strict formulas are the same as the formulas that can be derived from $\mathcal{F}$ in a classical way, provided $\mathcal{F}$ is satisfiable.

A Bayesian knowledge base containing only strict knowledge can be expressed by a BLP containing only conditional probabilities with values 0 and 1. In this case, also BLP semantics and BLP inference coincide with the semantics and inference in first-order logic. In RPCL, a strict knowledge base is also obtained by allowing just the two extreme probabilities 0 and 1. For a more detailed look at the relationship of the obtained logics to first-order logic, let $FOL_\forall$ be the set of quantifier-free first order formulas without function symbols, with all variables being implicitly universally quantified. For strict formulas of BLPs, we get only a subset of $FOL_\forall$ since in a BLP we can not express a disjunction like $A \vee B$.

Every set $\mathcal{F}$ of formulas of $FOL_\forall$ can be expressed by the RPCL knowledge base $\mathcal{F}_P = \{(A \mid \top)[1] \mid A \in \mathcal{F}\}$ containing only strict formulas. Then inference based on $\mathcal{F}$ and $\mathcal{F}_P$ is the same (independently of the actual used semantics for RPCL). Looking at the other direction, let KB be a strict RPCL knowledge base, and let $KB_{FOL_\forall} = \{\neg A \vee B \mid (B \mid A)[1] \in KB\} \cup \{A \wedge \neg B \mid (B \mid A)[0] \in KB\}$. If KB is consistent with respect to grounding, averaging, or aggregating semantics then inference in KB and $KB_{FOL_\forall}$ is the same. However, for the strict RPCL knowledge base $KB' = \{(B \mid A)[1],\ (A \mid \top)[0]\}$ we observe that KB' has no models since a probability distribution P can satisfy a conditional $\{(B \mid A)[x]$

only if $P(A) > 0$, independently of the actual semantics. On the other hand, $KB'_{FOL_\forall} = \{\neg A \vee B,\ \neg A\}$ does have a model. Thus, reducing a conditional to material implication is not adequate even in the case of only strict probabilistic conditionals (see also [1]).

Likewise, we can look at the degenerated knowledge representation formalism obtained by cutting out any relational representation aspects.

(SP-2) Propositional Knowledge: What kind of logic does one obtain if a knowledge base contains only ground knowledge? What are its semantics and inference properties, and in particular, what is its relationship to propositional probabilistic logic?

A BLP where all occurring atoms are ground obviously corresponds to a propositional Bayesian network. Restricting the formulas in an MLN to be variable-free yields the semantics of a propositional Markov net: If L is an MLN containing only ground atoms, then for any set C of constants the corresponding ground Markov net is independent of C. For a ground RPCL knowledge base grounding, averaging, and aggregating semantics coincide with classical probabilistic semantics in probabilistic conditional logic and inference based on the principle of maximum entropy is the same as in the propositional case, cf. [19].

5 Individuals and Universes

The core idea of relational knowledge representation is to talk about a set of elements (a *universe*) and the relations among them. Thus, methods are needed for specifying elements belonging to the universe, to refer to elements in the universe, and to reason about elements and their properties and relationships. In general, relational approaches may differ according to whether and how they support any of the following criteria.

(U-1) Listing of elements: Can universes be specified by explicitly listing all its elements?

The given facts in a BLP must all be ground; they determine the specific context of the BLP, thus allowing to list all elements of a universe by mentioning them in the atoms of the BLP. When defining an MLN, an explicit listing of all constants C must be given, and the semantics of an MLN requires that different constants denote different elements and that there are no elements other than the ones denoted by constants. Similarly, all constants in an RPCL knowledge base denote different elements, and there are no other elements.

(U-2) Open universe: Is it possible to have an *open* universe whose number of elements is not a-priori known?

In BLP, MLN, and RPCL it is not possible to specify such open universes directly. However, in all approaches the extensional part—i. e. the ground atoms resp. the given constants—can be exchanged while reusing the given generic knowledge. For instance, the constants occurring in a query Q together with the constants in a BLP P determine the Herbrand universe used to construct the ground Bayesian network for answering Q.

(U-3) Prototypical elements: Specification of prototypical elements of a universe.

A universally quantified variable X in a relational statement says that this statement applies to all elements of the considered universe. However, as Ex. 1 demonstrates, there is the need to also express knowledge about individuals, referred to by specific constants; in any of the five approaches, generic statements using variables may be combined with statements about individuals. In the elephant-keeper example, asking about a keeper *jim* will return the same probability as asking the same question about a keeper *james* since the respective knowledge bases do not contain any specific information neither about *jim* nor about *james*. More generally, let $C_{\mathcal{R}}$ be the set of constants occurring in a set of rules $\mathcal{R}$ and let C_U be the set of all constants under consideration. (Note that for MLN and ME, C_U is given explicitly, and that for a BLP, C_U is determined when a query is posed.) Then the elements in $C_{prot} = C_U \backslash C_{\mathcal{R}}$ are all prototypical as they can not be distinguished by any query asked w.r.t. $\mathcal{R}$: If d_1, $d_2 \in C_{prot}$ and Q is a query containing d_1, then the query Q' obtained from Q by replacing d_1 by d_2 (and possibly also d_2 by d_1) yields the same probability as Q. This observation holds for all of the five approaches.

(U-4) Inference for individuals: There should be a well-defined inference mechanism to infer probabilities for particular individuals (either prototypical individuals or specific, named individuals). Does such inference depend on the number of elements in a universe, and if so, what is the dependency?

Obviously, all approaches provide for querying about specific individuals. For example, given a BLP, a ground Bayes net can be constructed to infer probabilities for some ground query involving arbitrary constants. Similarly, this holds for MLNs and the approaches based on maximum entropy. Further, the number of elements in the universe might influence the probability of a query in all approaches. Consider the BLP B containing the clauses $(B(X) \mid A(X,Y))$ and $(A(X,Y))$. Given the query $B(c)$ for some constant c the probability of $B(c)$ depends on the number of instances of $A(c,Y)$, i. e., on the number of constants in the universe. If *noisy-or* is the combining rule for B then the probability of $B(c)$ tends towards one when the number of constants in the universe tends towards infinity, independently of the actual conditional probability distributions of $(B(X) \mid A(X,Y))$ and $(A(X,Y))$. A similar observation can be made for MLNs and RPCL. Another dependency of the number of elements in the universe and probabilities of queries arises for RPCL under averaging and aggregating semantics. Consider now the conditional $(B \mid A)[x]$ and averaging semantics. If $(B' \mid A')$ is an instance of $(B \mid A)$ that does not mention any constants in the knowledge base then it is easy to see that the probability of $(B' \mid A')$ tends towards x if the number of elements in the universe tends towards infinity, cf. [12].

(U-5) Grounding: Is there a mechanism for (consistent) grounding of a knowledge base?

The semantics of a BLP or an MLN knowledge base is defined via complete groundings yielding a (ground) Bayesian network or a (ground) Markov net, respectively. In a BLP, the logic part consists of Horn clauses which do not allow the specification of negated conclusions, so that inconsistencies on the logical level are avoided. Conflicting specifications on the quantitative level may arise when having syntactical variants of a clause, e. g. $(B(X) \mid A(X))$ and $(B(Y) \mid A(Y))$ with different cpd's. Such conflicts are resolved via the combining rules like *noisy-or* (cf. Ex. 2). An MLN might contain both (F, w) and $(\neg F, w)$, but the grounded semantics is still consistent and well defined. For RPCL under grounding semantics, complete grounding might generate an inconsistency; therefore, various more sophisticated instantiation strategies have been proposed [13].

Another important aspect connected to the notion of relational knowledge and universes is the question whether probabilities are interpreted statistically or as subjective degrees of belief, cf. the discussion in the context of (L-4).

6 Conclusion and Future Work

During the last years, many different approaches extending probabilistic propositional logic to a relational setting have been proposed. In this paper, we developed and illustrated various evaluation and comparison criteria and applied them to five different modeling and inference methods, thereby putting emphasis on the knowledge representation point of view.

There are several additional criteria that require further research and more investigation in detail. When considering the expressivity of a particular modeling method, it is easy to see that any of the approaches discussed in this paper can be used to define an arbitrary probability distribution over a finite domain, but the more interesting question is *how* this can be done. Jaeger [9] proposes a schema of comparing different formalisms by using two components: A generic component that is independent of a particular universe, and a component that takes into account a universe of constants. The sharp separation of generic and specific knowledge as required in the expressivity analysis proposed in [9] is problematic since it prohibits a modeling taking into account both types of knowledge in the form as it is done for instance in Ex. 1.

Another criterion is to ask what kind of queries can be answered, and which can be answered efficiently. In the context of (L-5), we already discussed the syntactic form of queries that can be answered in the considered approaches. With respect to the complexity of inference, further experimental and theoretical work is needed. For instance, inference in RPCL requires solving the numerical optimization problem (1) whose complexity grows in the number of possible groundings. Work on lifted first-order probabilistic inference is done in e.g. [16, 17], and in [6] for reasoning under maximimum entropy.

References

1. C. Beierle and G. Kern-Isberner. The relationship of the logic of big-stepped probabilities to standard probabilistic logics. In S. Link and H. Prade, editors, *Foundations of Information and Knowledge Systems (FoIKS 2010)*, volume 5956 of *Lecture Notes in Computer Science*, pages 191–210. Springer, 2010.
2. M. Bruynooghe *et al.* An Exercise with Statistical Relational Learning Systems. In P. Domingos and K. Kersting, editors, *International Workshop on Statistical Relational Learning (SRL-2009)*, Leuven, Belgium, 2009.
3. L. De Raedt and K. Kersting. Probabilistic inductive logic programming. In L. De Raedt, K. Kersting, N. Landwehr, S. Muggleton, and J. Chen, editors, *Probabilistic Inductive Logic Programming*, volume 4911 of *LNCS*, pages 1–27. Springer, 2008.
4. J. Delgrande. On first-order conditional logics. *Artificial Intelligence*, 105:105–137, 1998.
5. P. Domingos and D. Lowd. *Markov Logic: An Interface Layer for Artificial Intelligence.* Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan and Claypool, San Rafael, CA, 2009.
6. J. Fisseler. *Learning and Modeling with Probabilistic Conditional Logic*, volume 328 of *Dissertations in Artificial Intelligence.* IOS Press, Amsterdam, 2010.
7. L. Getoor and B. Taskar, editors. *Introduction to Statistical Relational Learning.* MIT Press, 2007.
8. M. Jaeger. Relational Bayesian Networks: A Survey. *Electronic Transactions in Artificial Intelligence*, 6, 2002.
9. M. Jaeger. Model-Theoretic Expressivity Analysis. In L. De Raedt, K. Kersting, N. Landwehr, S. Muggleton, and J. Chen, editors, *Probabilistic Inductive Logic Programming*, volume 4911 of *LNCS*, pages 325–339. Springer, 2008.
10. D. Jain, L. Mösenlechner, and M. Beetz. Equipping Robot Control Programs with First-Order Probabilistic Reasoning Capabilities. In *International Conference on Robotics and Automation (ICRA)*, pages 3130–3135, 2009.
11. G. Kern-Isberner. Characterizing the principle of minimum cross-entropy within a conditional-logical framework. *Artificial Intelligence*, 98:169–208, 1998.
12. G. Kern-Isberner and M. Thimm. Novel Semantical Approaches to Relational Probabilistic Conditionals. In *Proc. Twelfth International Conference on the Principles of Knowledge Representation and Reasoning (KR'10)*, pages 382–392, 2010.
13. S. Loh, M. Thimm, and G. Kern-Isberner. On the problem of grounding a relational probabilistic conditional knowledge base. In *Proceedings of the 14th International Workshop on Non-Monotonic Reasoning (NMR'10)*, Toronto, Canada, May 2010.
14. J.B. Paris. *The uncertain reasoner's companion – A mathematical perspective.* Cambridge University Press, 1994.
15. J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, 1998.
16. D. Poole. First-order probabilistic inference. In G. Gottlob and T. Walsh, editors, *Proc. IJCAI-03*, pages 985–991. Morgan Kaufmann, 2003.
17. P. Singla and P. Domingos. Lifted first-order belief propagation. In D. Fox and C. P. Gomes, editors, *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*, pages 1094–1099. AAAI Press/The MIT Press, 2008.
18. M. Thimm, M. Finthammer, G. Kern-Isberner, and C. Beierle. Comparing approaches to relational probabilistic reasoning: Theory and implementation, 2010. (submitted).
19. M. Thimm, G. Kern-Isberner, and J. Fisseler. Relational probabilistic conditional reasoning at maximum entroy. In *Proceedings ECSQARU-11*, 2011. (to appear).