Constructing Markov Logic Networks from First-Order Default Rules

Ondřej Kuželka[1], Jesse Davis[2], and Steven Schockaert[1]

[1] School of Computer Science & Informatics, Cardiff University, UK
{KuzelkaO,SchockaertS1}@cardiff.ac.uk
[2] Department of Computer Science, Katholieke Universiteit Leuven, Belgium
jesse.davis@cs.kuleuven.be

Abstract. Expert knowledge can often be represented using default rules of the form "if A then typically B". In a probabilistic framework, such default rules can be seen as constraints on what should be derivable by MAP-inference. We exploit this idea for constructing a Markov logic network $\mathcal{M}$ from a set of first-order default rules D, such that MAP inference from $\mathcal{M}$ exactly corresponds to default reasoning from D, where we view first-order default rules as templates for the construction of propositional default rules. In particular, to construct appropriate Markov logic networks, we lift three standard methods for default reasoning. The resulting Markov logic networks could then be refined based on available training data. Our method thus offers a convenient way of using expert knowledge for constraining or guiding the process of learning Markov logic networks.

1 Introduction

Markov logic is a popular framework for statistical relational learning [20]. Formulas in Markov logic essentially correspond to weighted first-order formulas, which act as soft constraints on possible worlds. In current applications, the weights are typically learned from data, while the first-order formulas are either hand crafted or obtained using standard rule learning methods.

The fact that a domain expert could manually specify (some of) the formulas, or could inspect learned formulas, is an important strength of Markov logic. Unfortunately, the weights associated with these formulas do not have an easily interpretable meaning. This limits the potential of Markov logic, as it means that domain experts cannot offer much guidance in terms of how the weights should be set (e.g. in applications with little or no training data) or cannot verify the quality of learned weights (e.g. in applications where the quality of the training data is in doubt). Often, however, Markov logic networks (MLN) are not used for evaluating probabilities but for finding the most likely truth assignment of unobserved variables, given the available evidence, i.e. for *maximum a posteriori* (MAP) reasoning. In such cases, the precise values of the weights are only relevant inasmuch as they influence the result of MAP queries. In this setting, we can instead ask for constraints on how MAP reasoning should behave as opposed to

asking a domain expert to specify weights. For example, the expert could specify constraints such as "if all we know is that x is a bird, then using MAP reasoning we should be able to conclude that x can fly", which is in agreement with the semantics of the default rule "birds can typically fly" in System P [13, 2].

Thus, a domain expert could be involved in the process of learning an MLN by providing a set of defaults, which are interpreted as constraints on the ranking of possible worlds induced by the MLN. Taking this idea one step further, in this paper, we show how a specific MLN can be constructed from the default rules provided by the expert. Constructing this MLN requires us to select a specific probability distribution that is compatible with the defaults. This selection problem is closely related to the problem of defining the closure of a set of defaults, which has been widely studied in the field of non-monotonic reasoning [14, 9, 10]. In particular, several proposals to define this closure are based on constructing a specific probability distribution [4, 10]. As we will show, it is possible to lift these approaches and thus obtain an efficient and principled way to construct MLNs that are compatible with a given set of defaults.

To date, the use of expert knowledge for guiding or even replacing weight learning has only received limited attention. One exception is [17], which constructs an MLN based on (potentially inconsistent) conditional probabilities provided by domain experts. While this can be useful in some applications, it relies on the ability of experts to provide meaningful probability estimates. However, humans are notoriously poor at judging likelihood. For example, properties that are common among the typical elements of a class of objects are often assumed to be likely in general [22]. Moreover, experts may be able to specify which properties are most likely to hold, in a given context, without being able to quantify their likelihood. In such situations, our default-rule-based approach would be more natural than approaches that force experts to estimate probabilities. On the other hand, our approach will only provide meaningful results for MAP queries: numerical input will be difficult to avoid if we want the constructed MLN to produce satisfactory conditional probability estimates.

This paper is structured as follows. The next section recalls some preliminaries from Markov logic and the non-monotonic reasoning literature. Then in Section 3 we show how three well-known approaches to non-monotonic reasoning can be implemented as MAP inference in a particular MLN. By lifting the constructions from Section 3, in Section 4 we show how MLNs can be constructed whose MAP-consequences are compatible with a given set of first-order default rules. Finally, Section 5 evaluates the performance of the resulting MLNs in a standard classification task.

2 Background

2.1 Markov logic networks

A Markov logic network (MLN) [20] is a set of weighted formulas (F, w_F), where F is a classical first-order formula and w_F is a real number, intuitively reflecting

a penalty that is applied to possible worlds that violate F. We will sometimes also use the notation $w_F : F$ to denote the formula (F, w_F). Given a set of constants C, an MLN $\mathcal{M}$ induces the following probability distribution on the set of possible worlds ω:

$$p_{\mathcal{M}}(\omega) = \frac{1}{Z} \exp \left(\sum_{(F, w_F) \in \mathcal{M}} w_F n_F(\omega) \right), \qquad (1)$$

where $n_F(x)$ is the number of true groundings of F in the possible world ω, and $Z = \sum_{\omega} p(\omega)$ is a normalization constant to ensure that p can be interpreted as a probability distribution. Sometimes, formulas (F, w_F) with $w_F = +\infty$ are considered to represent hard constraints. In such cases, we define $p_{\mathcal{M}}(\omega) = 0$ for all possible worlds that do not satisfy all of the hard constraints, and only formulas with a real-valued weight are considered in (1) for the possible worlds that do.

The main inference task for MLNs which we will consider is full MAP inference. Given a set of ground literals (the evidence), MAP inferences aims to compute the most probable configuration of all unobserved variables (the queries). Standard approaches for performing MAP inference include a strategy based on MaxWalkSAT [20] and a cutting plane based strategy [21, 16]. Given a set of ground formulas E, we write $\max(\mathcal{M}, E)$ for the set of most probable worlds of the MLN that satisfy E. We will also consider the following inference relation, initially proposed for penalty logic in [7]:

$$(\mathcal{M}, E) \vdash_{MAP} \alpha \quad \text{iff} \quad \forall \omega \in \max(\mathcal{M}, E) : \omega \models \alpha \qquad (2)$$

with $\mathcal{M}$ an MLN, α a ground formula and E a set of ground formulas. Note that $(\mathcal{M}, E) \vdash_{MAP} \alpha$ means that the formula α is satisfied in all the most probable worlds which are compatible with the available evidence.

2.2   Reasoning about default rules in System P

A variety of approaches have been proposed to reason about default rules of the form "if α then typically β holds", which we will denote as $\alpha \vdash\!\sim \beta$. Most approaches are based on the idea of defining a preference order over possible worlds and insisting that β is true in the most preferred (i.e. the most normal) of the worlds in which α is true [18, 13, 19, 9, 3]. The axioms of System P [13] capture a set of desirable properties for an inference relation for default rules:

Reflexivity $\alpha \vdash\!\sim \alpha$
Left logical equivalence If $\alpha \equiv \alpha'$ and $\alpha \vdash\!\sim \beta$ then $\alpha' \vdash\!\sim \beta$
Right weakening If $\beta \models \beta'$ and $\alpha \vdash\!\sim \beta$ then $\alpha \vdash\!\sim \beta'$
OR If $\alpha \vdash\!\sim \gamma$ and $\beta \vdash\!\sim \gamma$ then $\alpha \vee \beta \vdash\!\sim \gamma$
Cautious monotonicity If $\alpha \vdash\!\sim \beta$ and $\alpha \vdash\!\sim \gamma$ then $\alpha \wedge \beta \vdash\!\sim \gamma$
CUT If $\alpha \wedge \beta \vdash\!\sim \gamma$ and $\alpha \vdash\!\sim \beta$ then $\alpha \vdash\!\sim \gamma$

where $\alpha \equiv \alpha'$ and $\beta \models \beta'$ refer to equivalence and entailment from classical logic. Note that applying the axioms of System P to a set of defaults $\Delta = \{\alpha_1 \mathbin{|\!\sim} \beta_1, ..., \alpha_n \mathbin{|\!\sim} \beta_n\}$ corresponds to a form of monotonic reasoning about defaults. However, as the set of consequences that can be obtained in this way is limited, it is common to consider a non-monotonic inference relation whose set of consequences is closed under the axioms of System P as well as the following property:

Rational monotonicity If $\alpha \mathbin{|\!\sim} \beta$ and we cannot derive $\alpha \mathbin{|\!\sim} \neg\gamma$ then $\alpha \wedge \gamma \mathbin{|\!\sim} \beta$.

In this paper, we will consider three such inference relations: the rational closure [19], the lexicographic closure and the maximum entropy closure. A default $\alpha \mathbin{|\!\sim} \beta$ is tolerated by a set of defaults $\gamma_1 \mathbin{|\!\sim} \delta_1, ..., \gamma_m \mathbin{|\!\sim} \delta_m$ if the classical formula $\alpha \wedge \beta \wedge \bigwedge_i (\neg\gamma_i \vee \delta_i)$ is consistent. The rational closure is based on a stratification $\Delta_1, ..., \Delta_k$ of Δ, where each Δ_j contains all defaults $\alpha \mathbin{|\!\sim} \beta$ from Δ which are tolerated by $\Delta \setminus (\Delta_1 \cup ... \cup \Delta_{j-1})$. It can be shown that such a stratification always exists when Δ satisfies some natural consistency properties (see [19] for details). Intuitively, Δ_1 contains the most general default rules, Δ_2 contains exceptions to the rules in Δ_1, Δ_3 contains exceptions to the rules in Δ_2, etc. This stratification is known as the Z-ordering. Let j be the smallest index for which $\Delta_\alpha^{rat} = \Delta_j \cup ... \cup \Delta_k \cup \{\alpha\}$ is consistent. Then $\alpha \mathbin{|\!\sim} \beta$ is in the rational closure of Δ if $\Delta_\alpha^{rat} \models \beta$. When a set of hard rules Γ needs to be enforced, the Z-ordering can be generalized as follows [3]. Each set Δ_j then contains those defaults $\alpha \mathbin{|\!\sim} \beta$ for which $\Gamma \cup \{\alpha \wedge \beta\} \cup \{\neg\alpha_i \vee \gamma_i : (\alpha_i \mathbin{|\!\sim} \beta_i) \in \Delta \setminus (\Delta_1 \cup ... \cup \Delta_{j-1})\}$. Finally, we define $\Delta_k = \Gamma$, where $\Delta_1, ..., \Delta_{k-1}$ is the stratification of Δ that was obtained.

The rational closure encodes the intuition that in case of conflict, specific rules should have priority over more generic ones. However, it requires us to ignore all the defaults in $\Delta_1 \cup ... \Delta_{j-1}$, even defaults which are intuitively unrelated to this conflict. The lexicographic closure [1] addresses this issue as follows. For a propositional interpretation ω, we write $sat(\omega, \Delta_j)$ for the number of defaults satisfied by ω, i.e. $sat(\omega, \Delta_j) = |\{\alpha \mathbin{|\!\sim} \beta : (\alpha \mathbin{|\!\sim} \beta) \in \Delta_j, \omega \models \neg\alpha \vee \beta\}|$. We say that an interpretation ω_1 is lex-preferred over an interpretation ω_2, written $\omega_1 \prec \omega_2$, if there exists a j such that $sat(\omega_1, \Delta_j) > sat(\omega_2, \Delta_j)$ while $sat(\omega_1, \Delta_i) = sat(\omega_2, \Delta_i)$ for all $i > j$. The default $\alpha \mathbin{|\!\sim} \beta$ is in the lexicographic closure of Δ if β is satisfied in all the most lex-preferred models of α, i.e. $\forall \omega \in [\![\alpha]\!] : (\omega \not\models \beta) \Rightarrow \exists \omega' \in [\![\alpha]\!] : \omega' \prec \omega$.

Another approach to default reasoning is based on the principle of maximum entropy [10]. To describe how the maximum-entropy ranking of possible worlds can be computed, we need some additional terminology. A possible world ω is said to falsify a rule $\alpha \mathbin{|\!\sim} \beta$ if $\omega \models \alpha \wedge \neg\beta$ and said to verify it if $\omega \models \alpha \wedge \beta$. A set of default rules Δ is said to be a minimal core if for any rule $\alpha \mathbin{|\!\sim} \beta$, the set $\{\alpha \mathbin{|\!\sim} \neg\beta\} \cup (\Delta \setminus \{\alpha \mathbin{|\!\sim} \beta\})$ is consistent. Given a minimal core set of defaults Δ, the maximum-entropy ranking is obtained as follows [10]. Let Γ be the set of rules tolerated by Δ. For each rule $r \in \Gamma$, we set $\kappa_{ME}(r) = 1$. While $\Gamma \neq \Delta$ we repeat the following steps. Let Ω be the set of models ω which do not falsify any of the rules in $\Delta \setminus \Gamma$ and verify at least one of these rules. For each model

$\omega \in \Omega$, we compute $\kappa_{ME}(\omega) = \sum\{\kappa_{ME}(\alpha \mathrel{|\!\sim} \beta) \colon (\alpha \mathrel{|\!\sim} \beta) \in \Gamma, \omega \models \alpha \wedge \neg\beta\}$. Let ω^* be the model in Ω with minimum rank. Each rule $\alpha \mathrel{|\!\sim} \beta$ that is verified by ω^* is added to Γ, and its rank is computed as $\kappa_{ME}(\alpha \mathrel{|\!\sim} \beta) = 1 + \kappa_{ME}(\omega^*)$.

3   Encoding ground default theories in Markov logic

It is well-known that any set of defaults Δ which is closed under the axioms of System P and rational monotonicity corresponds to a linear ranking κ of possible worlds, such that $\alpha \models \beta$ iff $\kappa(\alpha \wedge \beta) > \kappa(\alpha \wedge \neg\beta)$, where we write $\kappa(\gamma)$ for a formula γ as an abbreviation for $\max\{\kappa(\omega) \colon \omega \models \gamma\}$. Since the ranking κ can be encoded as a probability distribution, and every probability distribution on possible worlds can be represented as an MLN, it is clear that there must exist an MLN $\mathcal{M}$ such that $(\alpha \mathrel{|\!\sim} \beta) \in \Delta$ iff $(\mathcal{M}, \alpha) \vdash_{MAP} \beta$. More generally, for any (i.e. not necessarily closed) set of defaults Δ, there exists an MLN $\mathcal{M}$ such that $(\mathcal{M}, \alpha) \vdash_{MAP} \beta$ iff $\alpha \mathrel{|\!\sim} \beta$ is in the rational closure of Δ, and similar for the lexicographic and maximum-entropy closures. We now show how the MLNs corresponding to these three closures can be constructed.

Transformation 1 (Rational closure) *Let Δ be a set of ground default rules and let Θ be a set of hard constraints (clauses). Let $\Delta_1, ..., \Delta_k$ be the Z-ordering of $\Delta \cup \Theta$. Let the MLN $\mathcal{M}$ be defined as follows: $\bigcup_{i=1}^{k}(\{(\neg a_i \vee \neg\alpha \vee \beta, \infty) \colon \alpha \mathrel{|\!\sim} \beta \in \Delta_i\} \cup \{(a_i, 1)\} \cup \{(\phi, \infty) \colon \phi \in \Theta\}) \cup \bigcup_{i=2}^{k}\{(a_i \vee \neg a_{i-1}, \infty)\}$ where a_i are auxiliary literals. Then $(\mathcal{M}, \alpha) \vdash_{MAP} \beta$ iff $\alpha \mathrel{|\!\sim} \beta$ is in the rational closure of (Δ, Θ).*

Transformation 2 (Lexicographic closure) *Let Δ be a set of ground default rules and let Θ be a set of hard constraints (clauses). Let $\Delta_1, ..., \Delta_k$ be the Z-ordering of $\Delta \cup \Theta$. Let the MLN $\mathcal{M}$ be defined as follows: $\bigcup_{i=1}^{k}\{(\neg\alpha \vee \beta, \lambda_i) \colon \alpha \mathrel{|\!\sim} \beta \in \Delta_i\} \cup \{(\phi, \infty) \colon \phi \in \Theta\}$ where $\lambda_i = 1 + \sum_{j=1}^{i-1}|\Delta_j| \cdot \lambda_j$ for $i > 1$ and $\lambda_1 = 1$. Then $(\mathcal{M}, \alpha) \vdash_{MAP} \beta$ iff $\alpha \mathrel{|\!\sim} \beta$ is in the lexicographic closure of (Δ, Θ).*

Transformation 3 (Maximum-entropy closure) *Let Δ be a set of ground default rules and let Θ be a set of hard constraints (clauses). Let κ be weights of rules corresponding to the maximum-entropy closure of $\Delta \cup \Theta$. Let the MLN $\mathcal{M}$ be defined as follows: $\{(\neg\alpha \vee \beta, \kappa(\alpha \mathrel{|\!\sim} \beta)) \colon \alpha \mathrel{|\!\sim} \beta \in \Delta\} \cup \{(\phi, \infty) \colon \phi \in \Theta\}$. Then $(\mathcal{M}, \alpha) \vdash_{MAP} \beta$ iff $\alpha \mathrel{|\!\sim} \beta$ is in the maximum-entropy closure of (Δ, Θ).*

Example 1. Consider the default rules $\Delta = \{bird \mathrel{|\!\sim} flies, antarctic \wedge bird \mathrel{|\!\sim} \neg flies\}$. Then $\mathcal{M}_1 = \{(\neg a_1 \vee \neg bird \vee flies, \infty), (\neg a_2 \vee \neg antarctic \vee \neg bird \vee \neg flies, \infty), (a_1, 1), (a_2, 1), (a_2 \vee \neg a_1, \infty)\}$ is the result of Transformation 1 , and $\mathcal{M}_2 = \{(\neg bird \vee flies, 1), (\neg antarctic \vee \neg bird \vee \neg flies, 2)\}$ is the result of Transformation 2, which in this example coincides with the result of Transformation 3.

4   Encoding non-ground default theories in Markov logic

While reasoning with default rules has mostly been studied at the propositional level, a few authors have considered first-order default rules [8, 12]. Similar as for probabilistic first-order rules [11], two rather different semantics for first-order defaults can be considered. On the one hand, a default such as $P(x) \sim Q(x)$ could mean that the most typical objects that have the property P also have the property Q. On the other hand, this default could also mean that whenever $P(x)$ holds for a given x, in the most normal worlds $Q(x)$ will also hold. In other words, first-order defaults can either model typicality of objects or normality of worlds [8]. In this paper, we will consider the latter interpretation. Given that we only consider finite universes (as is usual in the context of MLNs), we can then see a first order default as a template for propositional defaults. For example $P(x) \sim Q(x)$ can be seen as a compact notation for a set of defaults $\{P(c_1) \sim Q(c_1), ..., P(c_n) \sim Q(c_n)\}$. Note that this approach would not be possible for first-order defaults that model the typicality of objects.

In particular, the first-order default theories we will consider consist of first-order logic formulas (hard rules) and default rules of the form $\alpha \sim \beta$, where α is a conjunction of literals and β is a disjunction of literals. Our approach can be straightforwardly extended to quantified default rules, where the scopes of quantifiers may be the whole default rules, and not just either the antecedent or the consequent of a rule. While this could be of interest, we do not consider this for the ease of presentation.

Definition 1 (Markov logic model of a first-order default theory). *Let (Δ, Θ) be a first-order default theory with Δ the set of default rules and Θ the set of hard rules. A Markov logic network $\mathcal{M}$ is a model of the default logic theory $\Delta \cup \Theta$ if it holds that: (i) $P[X = \omega] = 0$ whenever $\omega \not\models \Theta$, and (ii) for any default rule $\alpha \sim \beta \in \Delta$ and any grounding substitution θ of the unquantified variables of $\alpha \sim \beta$, either $\{\alpha\theta\} \cup \Theta \vdash \bot$ or $(\mathcal{M}, \alpha\theta) \vdash_{MAP} \beta\theta$. We say that (Δ, Θ) is satisfiable if it has at least one model.*

Below we will describe three methods for constructing Markov logic models of first-order default theories, generalizing Transformations 1–3. For convenience, we will use *typed* formulas (nevertheless, we will assume that default rules given on input are not typed for simplicity). For instance, when we have the formula $\alpha = owns(person : X, thing : Y)$ and the set of constants of the type *person* is $\{alice, bob\}$ and the set of constants of the type *thing* is $\{car\}$ then α corresponds to the ground formulas $owns(alice, car)$ and $owns(bob, car)$. In cases where there is only one type, we will not write it explicitly. For a constant or variable x, we write $\tau(x)$ to denote its type. Two formulas F_1 and F_2 (either both conjunctions or both disjunctions of literals) are said to be isomorphic when there is a substitution θ of the variables of F_1 such that $F_1\theta \equiv F_2$ (where $\equiv$ denotes logical equivalence). Two default rules $D_1 = \alpha_1 \sim \beta_1$ and $D_2 = \alpha_2 \sim \beta_2$ are said to be isomorphic if there exists a substitution θ of the variables of D_1 such that $\alpha_1\theta \equiv \alpha_2$ and $\beta_1\theta \equiv \beta_2$. Two default theories $\Delta_1 \cup \Theta_1$ and $\Delta_2 \cup \Theta_2$ are said

to be isomorphic, denoted by $\Delta_1 \cup \Theta_1 \approx \Delta_2 \cup \Theta_2$, if there is a bijection i from elements of $\Delta_1 \cup \Theta_1$ to elements of $\Delta_2 \cup \Theta_2$ such that for any $F \in \Delta_1 \cup \Theta_1$, $i(F) \approx F$. When j is a permutation of a subset of constants from $\Delta \cup \Theta$ then $j(\Delta \cup \Theta)$ denotes the default theory obtained by replacing any constant c from the subset by its image $j(c)$.

Definition 2 (Interchangeable constants). *Let $\Delta \cup \Theta$ be a non-ground default theory. A set of constants $\mathcal{C}$ is said to be interchangeable in $\Delta \cup \Theta$ if $j(\Delta \cup \Theta) \approx \Delta \cup \Theta$ for any permutation j of the constants in $\mathcal{C}$.*

The set of maximal interchangeable subsets of a set of constants is the uniquely defined partition of this set and will be called the *interchangeable partition*. To check whether a set of constants $\mathcal{C}$ is interchangeable, it is sufficient to check that $j(\Delta \cup \Theta) \approx \Delta \cup \Theta$ for those permutations which swap just two constants from $\mathcal{C}$. Note that the constants do not actually need to appear in $\Delta \cup \Theta$. It trivially holds that constants which do not appear in $\Delta \cup \Theta$ are interchangeable. When $\mathcal{I} = \{\mathcal{C}_1, \ldots, \mathcal{C}_n\}$ is the interchangeable partition of a set of constants then we may introduce a new type $type_{lexmin(C_i)}$ for every $\mathcal{C}_i \in \mathcal{C}$ (where $lexmin(C)$ denotes the lexically[3] smallest element from C_i). When $D = \alpha \mathrel{\vert\!\sim} \beta$ is a ground default rule, we write $variabilize(D)$ to denote the following default rule: $\bigwedge \{V_c \neq V_d \colon c, d \in const(D), \tau(c) = \tau(d)\} \wedge \alpha' \mathrel{\vert\!\sim} \beta'$ where $const(D)$ is the set of constants appearing in D and α' and β' are obtained from α and β by respectively replacing all constants c by a new variable V_c of type $\tau(c)$. Here $\neq$ is treated as a binary predicate which is defined in the set of hard rules Θ.

Let $\mathcal{C}$ be a set of constants and let $\mathcal{I} = \{\mathcal{C}_1, \ldots, \mathcal{C}_n\}$ be the interchangeable partition of the constants from $\mathcal{C}$. Two ground default rules $\alpha_1 \mathrel{\vert\!\sim} \beta_1$ and $\alpha_2 \mathrel{\vert\!\sim} \beta_2$ are said to be *weakly isomorphic* if $variabilize(\alpha_1 \mathrel{\vert\!\sim} \beta_1)$ and $variabilize(\alpha_2 \mathrel{\vert\!\sim} \beta_2)$ are isomorphic.

Definition 3 (Ground representatives). *Let $D = \alpha \mathrel{\vert\!\sim} \beta$ be a default rule and let $\mathcal{I} = \{\mathcal{C}_1, \ldots, \mathcal{C}_n\}$ be the interchangeable partition of constants. A set of ground representatives of D w.r.t. $\mathcal{I}$ is a maximal set of groundings of D by constants from $\mathcal{I}$ such that no two of these groundings are weakly isomorphic. (If $\alpha \mathrel{\vert\!\sim} \beta$ is typed then we only consider groundings which respect the typing of variables.)*

A set of ground representatives of a default rule $D = \alpha \mathrel{\vert\!\sim} \beta$ can be constructed in time $O(|\mathcal{C}|^{|D|})$. While this is exponential in the size of the default rule (which is usually small), it is only polynomial in the number of classes in the interchangeable partition $\mathcal{I}$ and does not depend on the total number of constants.

Let $\Delta \cup \Theta$ be a first-order default theory and $\mathcal{C}$ a set of constants. Let $\mathcal{R} = \bigcup_{\alpha \mathrel{\vert\!\sim} \beta \in \Delta} \mathcal{R}_{\alpha \mathrel{\vert\!\sim} \beta}$ where $\mathcal{R}_{\alpha \mathrel{\vert\!\sim} \beta}$ denotes a set of ground representatives of $\alpha \mathrel{\vert\!\sim} \beta$. The rational closure for the first-order default theory is based on the partition[4]

---

[3] Here, we are just ordering the constants by the lexical ordering of their names.

[4] With a slight abuse of terminology, we will call $\Delta_1^* \cup \cdots \cup \Delta_k^*$ *the partition of $\Delta \cup \Theta$* even though it is strictly speaking only a partition of Δ^*.

$\Delta_1^* \cup ... \cup \Delta_k^*$ of the set

$$\Delta^* = \{variabilize(\alpha \mathrel{|\!\sim} \beta) \colon \alpha \mathrel{|\!\sim} \beta \in \mathcal{R} \text{ and } \{\alpha\} \cup \Theta \not\vdash \bot\}$$

where Δ_j^* is the set of default rules $variabilize(\alpha \mathrel{|\!\sim} \beta) \in \Delta^*$ such that

$$\{\alpha \wedge \beta\} \cup \{\neg\alpha_i \vee \beta_i \colon (\alpha_i \mathrel{|\!\sim} \beta_i) \in \Delta^* \setminus (\Delta_1^* \cup ... \cup \Delta_{j-1}^*)\} \cup \Theta \qquad (3)$$

has a model with Herbrand universe $\mathcal{C}$. When all rules $\alpha' \mathrel{|\!\sim} \beta'$ from the set

$$\Delta_{\alpha \mathrel{|\!\sim} \beta}^* = \{variabilize(\alpha' \mathrel{|\!\sim} \beta') \colon \alpha' \mathrel{|\!\sim} \beta' \text{ is a ground representative of } \alpha \mathrel{|\!\sim} \beta\},$$

are contained in the same partition class Δ_j^* then we can simplify Δ_j^* by setting $\Delta_j^* := \Delta_j^* \cup \{\alpha \mathrel{|\!\sim} \beta\} \setminus \Delta_{\alpha \mathrel{|\!\sim} \beta}^*$. Furthermore, note that checking the existence of Herbrand models can be carried out using cutting-plane inference which means that it is seldom needed to ground the set of default rules completely. We can now present the lifted counterparts to Transformations 1–3.

Transformation 4 (Lifted rational closure) *Let Δ be a set of default rules and let Θ be a set of hard constraints. Let $\Delta_1^* \cup \cdots \cup \Delta_k^*$ be the partition of $\Delta \cup \Theta$, defined by (3). Let the MLN $\mathcal{M}$ be defined as follows: $\bigcup_{i=1}^k \{(\neg a_i \vee \neg\alpha \vee \beta, \infty) \colon \alpha \mathrel{|\!\sim} \beta \in \Delta_i^*\} \cup \{(a_i, 1)\} \cup \{(\phi, \infty) \colon \phi \in \Theta\} \cup \{(a_i \vee \neg a_{i-1}, \infty)\}$ where a_i are auxiliary (ground) literals. If (Δ, Θ) is satisfiable then $\mathcal{M}$ is a Markov logic model of (Δ, Θ).*

Transformation 5 (Lifted lexicographic entailment) *Let Δ be a set of default rules, let Θ be a set of hard constraints, and let $\mathcal{U}$ be the considered set of constants. Let $\Delta_1^* \cup \cdots \cup \Delta_k^*$ be the partition of $\Delta \cup \Theta$, defined by (3). Let the MLN $\mathcal{M}$ be defined as follows: $\bigcup_{i=1}^k \{(\neg\alpha \vee \beta, \lambda_i) \colon \alpha \mathrel{|\!\sim} \beta \in \Delta_i\} \cup \{(\phi, \infty) \colon \phi \in \Theta\}$ where $\lambda_i = 1 + \sum_{j=1}^{i-1} \sum_{\alpha \mathrel{|\!\sim} \beta \in \Delta_j^*} |\mathcal{U}|^{vars(\alpha \mathrel{|\!\sim} \beta)} \cdot \lambda_j$ for $i > 1$ and $\lambda_1 = 1$. If (Δ, Θ) is satisfiable then $\mathcal{M}$ is a Markov logic model of (Δ, Θ).*

Note that lexicographic entailment may lead to MLNs with very large weights.[5]

Next, we describe a lifted variant of maximum-entropy entailment. Let $\Delta \cup \Theta$ be a first-order default theory and $\mathcal{I}$ the interchangeable partition of constants from a given set $\mathcal{C}$. Let $\Delta_1^* \cup \cdots \cup \Delta_k^*$ be the partition of $\Delta \cup \Theta$, defined as in (3) (without the simplification of merging default rules), and let $\Gamma := \Delta_1^*$. First, we construct an MLN $\mathcal{M}$ containing the rules from Γ and set their weights equal to 1. For every Δ_j^* with $j > 1$, while $\Delta_j^* \not\subseteq \Gamma$, we repeat the following steps. We construct a new MLN $\mathcal{M}'$ by adding to the MLN $\mathcal{M}$ all rules from the set $\{\neg\alpha \vee \beta \colon \alpha \mathrel{|\!\sim} \beta \in (\Delta_j^* \cup \ldots \Delta_k^*) \setminus \Gamma\}$ as hard constraints (i.e. with infinite weights). For every $\alpha \mathrel{|\!\sim} \beta \in \Delta_j^* \setminus \Gamma$, we construct its ground representative $\alpha' \mathrel{|\!\sim} \beta'$ (note that there is only one ground representative up to isomorphism for any rule in

---

[5] Although existing MLN systems are not able to work with weights as large as are sometimes produced, due to numerical issues, we have implemented an MLN system based on cutting-plane MAP inference which can work with arbitrarily large weights.

Δ_j^*, which follows from the construction of Δ_j^*) and we find a most probable world $\omega_{\alpha\,\vdash\,\beta}$ of $(\mathcal{M}', \alpha')$; let us write $p_{\alpha\,\vdash\,\beta}$ for its penalty, i.e. the sum of the weights of the violated rules. Note that $\omega_{\alpha\,\vdash\,\beta}$ verifies the default rule $\alpha\,\vdash\,\beta$ and only falsifies rules in Γ, exactly as in the propositional version of the maximum-entropy transformation. We then select the rules $\alpha\,\vdash\,\beta$ with minimum penalty $p_{\alpha\,\vdash\,\beta}$, add them to Γ and to the MLN $\mathcal{M}$ with the weight set to $1 + p_{\alpha\,\vdash\,\beta}$. If $\mathcal{M}'$ does not have any models, the initial set of defaults cannot be satisfiable, and we end the procedure.

Transformation 6 (Lifted maximum-entropy entailment) *Let Δ be a set of default rules, let Θ be a set of hard constraints, and let $\mathcal{U}$ be the considered set of constants. Let $\mathcal{M}$ be the MLN obtained in the last iteration of the procedure described above. If (Δ, Θ) is satisfiable then $\mathcal{M}$ is a Markov logic model of (Δ, Θ).*

Example 2. Let us consider the following defaults:

$$bird(X)\,\vdash\,flies(X) \qquad bird(X) \wedge antarctic(X)\,\vdash\,\neg flies(X)$$
$$bird(X) \wedge antarctic(X) \wedge (X \neq Y) \wedge sameSpecies(X,Y)\,\vdash\,antarctic(Y)$$

Let the set of constants be given by $\mathcal{C} = \{tweety, donald, beeper\}$. Then the lexicographic transformation yields the MLN $\{(\phi_1, 1), (\phi_2, 4), (\phi_3, 4)\}$ while the maximum entropy transformation yields $\{(\phi_1, 1), (\phi_2, 2), (\phi_3, 3)\}$, where $\phi_1 = \neg bird(X) \vee flies(Y)$, $\phi_2 = \neg bird(X) \vee \neg sameSpecies(X,Y) \vee \neg(X \neq Y) \vee \neg bird(Y) \vee \neg antarctic(X) \vee antarctic(Y)$ and $\phi_3 = \neg bird(X) \vee \neg antarctic(X) \vee \neg flies(X)$.

As the next example illustrates, it is sometimes necessary to split the initial default rules into several typed specializations.

Example 3. Consider the following defaults: $bird(X) \wedge (X \neq tweety)\,\vdash\,flies(X)$, $bird(X) \wedge antarctic(X)\,\vdash\,\neg flies(X)$ and $bird(X) \wedge antarctic(X) \wedge (X \neq Y) \wedge sameSpecies(X,Y)\,\vdash\,antarctic(Y)$. Then the lexicographic transformation yields the MLN $\{(\phi_1, 1), (\phi_2, 1), (\phi_3, 7), (\phi_4, 7)\}$, where:

$$\phi_1 = \neg bird(\tau_{tweety} : X) \vee \neg antarctic(\tau_{tweety} : X) \vee \neg flies(\tau_{tweety} : X),$$
$$\phi_2 = \neg bird(X) \vee \neg(X \neq \tau_{tweety} : tweety) \vee flies(Y),$$
$$\phi_3 = \neg bird(\tau_{beeper} : X) \vee \neg antarctic(\tau_{beeper} : X) \vee \neg flies(\tau_{beeper} : X),$$
$$\phi_4 = \neg bird(X) \vee \neg sameSpecies(X,Y) \vee \neg(X \neq Y), \neg bird(Y) \vee \neg antarctic(X)$$

Note that the transformation had to introduce new types corresponding to the interchangeable sets of constants $\{\{tweety\}, \{beeper, donald\}\}$. The maximum entropy transformation leads to six such rules.

5   Evaluation

We have evaluated the proposed methods using the well-known UW-CSE dataset, which describes the Department of Computer Science and Engineering at the

University of Washington [20]. The usual task is to predict the *advisedBy*(*person*, *person*) predicate from the other predicates. A set of rules for this domain has previously been collected for the experiments in [20]. These rules, however, cannot be used as default rules because they are not satisfiable in the sense of Definition 1. Therefore, in order to evaluate our method, we have used the following consistent set of default rules.

D_1 : $\mathbin{|\!\sim} \neg advisedBy(S, P)$

D_2 : $advisedBy(S, P_1) \mathbin{|\!\sim} \neg tempAdvisedBy(S, P_2)$

D_3 : $advisedBy(S, P) \wedge publication(Pub, S) \mathbin{|\!\sim} publication(Pub, P)$

D_4 : $(P_1 \neq P_2) \wedge advisedBy(S, P_1) \mathbin{|\!\sim} \neg advisedBy(S, P_2)$

D_5 : $advisedBy(S, P) \wedge ta(C, S, T) \mathbin{|\!\sim} taughtBy(C, P, T)$

D_6 : $professor(P) \wedge student(S) \wedge publication(Pub, P) \wedge publication(Pub, S)$
$\mathbin{|\!\sim} advisedBy(S, P)$

D_7 : $professor(P) \wedge student(S) \wedge publication(Pub, P) \wedge publication(Pub, S) \wedge$
$tempAdvisedBy(S, P2) \mathbin{|\!\sim} \neg advisedBy(S, P)$

D_8 : $(S_1 \neq S_2) \wedge advisedBy(S_2, P) \wedge ta(C, S_2, T) \wedge ta(C, S_1, T) \wedge$
$taughtBy(C, P, T) \wedge student(S_1) \wedge professor(P) \mathbin{|\!\sim} advisedBy(S_1, P)$

D_9 : $(S_1 \neq S_2) \wedge advisedBy(S_2, P) \wedge ta(C, S_2, T) \wedge ta(C, S_1, T) \wedge$
$taughtBy(C, P, T) \wedge student(S_1) \wedge professor(P) \wedge tempAdvisedBy(S_1, P_2)$
$\mathbin{|\!\sim} \neg advisedBy(S_1, P)$

Recall that default rules $\alpha \mathbin{|\!\sim} \beta$ in our setting correspond to statements of the form: *for any grounding substitution θ, $\beta\theta$ is true in all most probable worlds of* $(\mathcal{M}, \alpha\theta)$. Thus the default rules $\alpha \mathbin{|\!\sim} \beta$ we consider should be such that an expert believes that $\alpha\theta$ being the only evidence, it would make sense to conclude $\beta\theta$. Seen with this perspective in mind, rule D_1 states that in absence of any knowledge, we assume that persons S and P are not in the *advisedBy* relationship. Rule D_2 states that if we only know that S has an advisor then we conclude that S does not have a temporary advisor. Rule D_3 states that advisors are typically co-authors of their students' papers. Rule D_4 states that students typically only have one advisor. The rest of the rules can be interpreted similarly. Note that rules D_7 and D_9 encode exceptions to rules D_6 and D_8.

We computed the lexicographic and maximum-entropy transformations of these rules using our implementation of the described methods.[6] We evaluated the obtained MLNs on the five different subject areas of the UW-CSE dataset, which is the standard methodology. Specifically, we computed the average number of true positives and false positives for the *advisedBy* predicate over 10 runs of MAP inference, noting that the results can depend on the specific MAP state that is returned. For comparison, we have used an MLN with the same set of rules

---

[6] Our implementation is based on a cutting-plane inference method for MAP inference implemented using the SAT4J library [5] and the MLN system Tuffy [15].

	MaxEnt		LEX		ONES		LEARNED	
	TP	**FP**	**TP**	**FP**	**TP**	**FP**	**TP**	**FP**
AI	10 ± 0	7 ± 0	10 ± 0	7 ± 0	8.6 ± 0.7	4.9 ± 0.9	10 ± 0	2 ± 0
GRA.	4 ± 0	5 ± 0	4 ± 0	5 ± 0	3.5 ± 0.7	3.9 ± 0.7	2 ± 0	2 ± 0
LAN.	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	2 ± 0	1 ± 0
SYS.	10.5 ± 0.5	3.5 ± 0.5	11 ± 0	3 ± 0	7.2 ± 1.1	2.4 ± 0.5	4 ± 0	0 ± 0
THE.	3 ± 0	3 ± 0	3 ± 0	3 ± 0	3 ± 0	1.7 ± 0.7	2 ± 0	1 ± 0

Table 1. Experimental results for MLNs obtained by the described methods.

but using weights learned discriminatively using Tuffy [15] (LEARNED), and an MLN with the same set of rules but with all weights set to 1 (ONES). The results are shown in Table 1. The maximum entropy and lexicographic entailment have highest recall but at the cost of also having higher number of false positives (note that the number of pairs which can potentially be in the *advisedBy* relationship is in the order of hundreds or thousands). The baseline method ONES has largest variance. Overall, the best performance seems to be obtained by the MLNs with learned weights which was an outcome to be expected.

6   Conclusion

We have discussed the problem of constructing a Markov logic network (MLN) from a set of first-order default rules, where default rules are seen as constraints on what should be derivable using MAP inference. The proposed construction methods have been obtained by lifting three well-known methods for non-monotonic reasoning about propositional default rules: the rational closure, the lexicographic closure and the maximum-entropy closure. As our evaluation with the UW-CSE dataset illustrates, our method can be used to construct useful MLNs in scenarios where no training data is available. In the future, we would like to explore the connections between our proposed lifted transformations and the lifted inference literature. For example, identifying interchangeable constants is known as shattering in lifted inference [6].

References

1. S. Benferhat, J. F. Bonnefon, and R. d. S. Neves. An overview of possibilistic handling of default reasoning, with experimental studies. *Synthese*, 146(1-2):53–70, 2005.
2. S. Benferhat, D. Dubois, and H. Prade. Nonmonotonic reasoning, conditional objects and possibility theory. *Artificial Intelligence*, 92(1-2):259–276, 1997.
3. S. Benferhat, D. Dubois, and H. Prade. Practical handling of exception-tainted rules and independence information in possibilistic logic. *Applied intelligence*, 9(2):101–127, 1998.

4. S. Benferhat, D. Dubois, and H. Prade. Possibilistic and standard probabilistic semantics of conditional knowledge bases. *Journal of Logic and Computation*, 9(6):873–895, 1999.
5. D. L. Berre and A. Parrain. The Sat4j library, release 2.2. *Journal on Satisfiability, Boolean Modeling and Computation*, 7:50–64, 2010.
6. R. de Salvo Braz, E. Amir, and D. Roth. Lifted first-order probabilistic inference. In L. P. Kaelbling and A. Saffiotti, editors, *Proceeding of the 19th Joint Conference on Artificial Intelligence*, page 1319, 2005.
7. F. Dupin de Saint-Cyr, J. Lang, and T. Schiex. Penalty logic and its link with Dempster-Shafer theory. In *Proceedings of the 10th International Conference on Uncertainty in Artificial Intelligence*, pages 204–211, 1994.
8. N. Friedman, J. Y. Halpern, and D. Koller. First-order conditional logic for default reasoning revisited. *ACM Transactions on Computational Logic*, 1(2):175–207, 2000.
9. H. Geffner and J. Pearl. Conditional entailment: Bridging two approaches to default reasoning. *Artificial Intelligence*, 53(2):209–244, 1992.
10. M. Goldszmidt, P. Morris, and J. Pearl. A maximum entropy approach to nonmonotonic reasoning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3):220–232, 1993.
11. J. Y. Halpern. An analysis of first-order logics of probability. *Artificial intelligence*, 46(3):311–350, 1990.
12. G. Kern-Isberner and M. Thimm. A ranking semantics for first-order conditionals. In *Proceedings of the 20th European Conference on Artificial Intelligence*, pages 456–461, 2012.
13. S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44(1-2):167–207, 1990.
14. D. Lehmann and M. Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55(1):1–60, 1992.
15. F. Niu, C. Ré, A. Doan, and J. W. Shavlik. Tuffy: Scaling up statistical inference in markov logic networks using an RDBMS. *PVLDB*, 4(6):373–384, 2011.
16. J. Noessner, M. Niepert, and H. Stuckenschmidt. Rockit: Exploiting parallelism and symmetry for map inference in statistical relational models. In *Proceedings of the 27th Conference on Artificial Intelligence (AAAI)*, 2013.
17. T. Pápai, S. Ghosh, and H. Kautz. Combining subjective probabilities and data in training Markov logic networks. In P. Flach, T. De Bie, and N. Cristianini, editors, *Machine Learning and Knowledge Discovery in Databases*, volume 7523 of *Lecture Notes in Computer Science*, pages 90–105. 2012.
18. J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.
19. J. Pearl. System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In *Proceedings of the 3rd Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 121–135, 1990.
20. M. Richardson and P. Domingos. Markov logic networks. *Machine Learning*, 62(1-2):107–136, 2006.
21. S. Riedel. Improving the accuracy and efficiency of MAP inference for Markov logic. In *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence*, pages 468–475, 2008.
22. A. Tversky and D. Kahneman. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4):293, 1983.