
Encoding Markov Logic Networks in Possibilistic Logic

Ondřej Kuželka

School of CS & Informatics
Cardiff University
Cardiff, UK

Jesse Davis

Department of Computer Science
KU Leuven
Leuven, Belgium

Steven Schockaert

School of CS & Informatics
Cardiff University
Cardiff, UK

Abstract

Markov logic uses weighted formulas to compactly encode a probability distribution over possible worlds. Despite the use of logical formulas, Markov logic networks (MLNs) can be difficult to interpret, due to the often counter-intuitive meaning of their weights. To address this issue, we propose a method to construct a possibilistic logic theory that exactly captures what can be derived from a given MLN using maximum a posteriori (MAP) inference. Unfortunately, the size of this theory is exponential in general. We therefore also propose two methods which can derive compact theories that still capture MAP inference, but only for specific types of evidence. These theories can be used, among others, to make explicit the hidden assumptions underlying an MLN or to explain the predictions it makes.

1 INTRODUCTION

Markov logic [22] and possibilistic logic [9] are two popular logics for modelling uncertain beliefs. Both logics share a number of important characteristics. At the syntactic level, formulas correspond to pairs (α, λ) , consisting of a classical formula α and a certainty weight λ , while at the semantic level, sets of these formulas induce a mapping from possible worlds to $[0, 1]$, encoding the relative plausibility of each possible world.

Despite their close similarities, however, Markov logic and possibilistic logic have been developed in different communities and for different purposes: Markov logic has mainly been studied in a machine learning context whereas possibilistic logic has been studied as a knowledge representation language. This reflects the complementary strengths and weaknesses of these logics. On the one hand, the qualitative nature of possibilistic logic makes it challenging to use for learning; although a few interesting approaches for

learning possibilistic logic theories from data have been explored (e.g. [24]), their impact on applications to date has been limited. On the other hand, the intuitive meaning of Markov logic theories is often difficult to grasp, which limits the potential of Markov logic for knowledge representation. The main culprit is that the meaning of a theory can often not be understood by looking at the individual formulas in isolation. This issue, among others, has been highlighted in [26], where coherence measures are proposed that evaluate to what extent the formulation of a Markov logic theory is misleading.

Example 1. Consider the following Markov logic formulas:

$$\begin{aligned} +\infty : & \quad \text{antarctic-bird}(X) \rightarrow \text{bird}(X) \\ 10 : & \quad \text{bird}(X) \rightarrow \text{flies}(X) \\ 5 : & \quad \text{antarctic-bird}(X) \rightarrow \neg \text{flies}(X) \end{aligned}$$

While the last formula might appear to suggest that antarctic birds cannot fly, in combination with the other two formulas, it merely states that antarctic birds are less likely to fly than birds in general.

Possibilistic logic is based on a purely qualitative, comparative model of uncertainty: while a Markov logic theory compactly encodes a probability distribution over the set of possible worlds, a possibilistic logic theory merely encodes a ranking of these possible worlds. Even though a probability distribution offers a much richer uncertainty model, many applications of Markov logic are based on MAP inference, which only relies on the ranking induced by the probability distribution.

In this paper, we first show how to construct a possibilistic logic theory Θ , given a Markov logic theory \mathcal{M} , such that the conclusions that we can infer from Θ are exactly those conclusions that we can obtain from \mathcal{M} using MAP inference. Our construction can be seen as the syntactic counterpart of the probability-possibility transformation from [10]. In principle, it allows us to combine the best of both worlds, using \mathcal{M} for making predictions while using Θ for elucidating the knowledge that is captured by \mathcal{M} (e.g. to verify

that the theory \mathcal{M} is sensible). However, the size of Θ can be exponential in the size of \mathcal{M} , which is unsurprising given that the computational complexity of MAP inference is higher than the complexity of inference in possibilistic logic. To overcome this problem, we begin by studying ground (i.e. propositional) theories and propose two novel approaches for transforming a ground MLN into a compact ground possibilistic logic theory that still correctly captures MAP inference, but only for specific types of evidence (e.g. sets of at most k literals). Then we lift one of these approaches such that it can transform a first-order MLN into a first-order possibilistic logic theory. Finally, we present several examples that illustrate how the transformation process can be used to help identify unintended consequences of a given MLN, and more generally, to better understand its behaviour.

The remainder of the paper is structured as follows. In the next section, we provide some background on Markov logic and possibilistic logic. In Section 3, we analyse the relation between MAP inference in ground Markov logic networks and possibilistic logic inference, introducing in particular two methods for deriving compact theories. Section 4 then discusses how we can exploit the symmetries in the case of an ungrounded Markov logic network, while Section 5 provides some illustrative examples. Finally, we provide an overview of related work in Section 6.

Due to space limitations, some of the proofs have been omitted from this paper. These proofs can be found in an online appendix.¹

2 BACKGROUND

2.1 MARKOV LOGIC

A Markov logic network (MLN) [22] is a set of pairs (F, w_F) , where F is a formula in first-order logic and w_F is a real number, intuitively reflecting a penalty that is applied to possible worlds (i.e. logical interpretations) that violate F . In examples, we will also use the notation $w_F : F$ to denote the formula (F, w_F) . An MLN serves as a template for constructing a propositional Markov network. In particular, given a set of constants C , an MLN \mathcal{M} induces the following probability distribution on possible worlds ω :

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

where $n_F(x)$ is the number of true groundings of F in the possible world ω , and Z is a normalization constant to ensure that $p_{\mathcal{M}}$ can be interpreted as a probability distribution. Sometimes, formulas (F, w_F) are considered where $w_F = +\infty$, 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. Note that a Markov logic network can be seen as a weighted set of propositional formulas, which are obtained by grounding the formulas in \mathcal{M} w.r.t. the set of constants C in the usual way. In the particular case that all formulas in \mathcal{M} are already grounded, \mathcal{M} corresponds to a theory in penalty logic [13].

One common inference task in MLNs is full MAP inference. In this setting, given a set of ground literals (the evidence) the goal is to compute the most probable configuration of all unobserved variables (the queries). Two standard approaches for performing MAP inference in MLNs are to employ a strategy based on MaxWalkSAT [22] or to use a cutting plane based strategy [23, 18]. 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 . For each $\omega \in \max(\mathcal{M}, E)$, $\sum_{(F, w_F) \in \mathcal{M}} w_F n_F(\omega)$ evaluates to the same value, which we will refer to as $\text{sat}(\mathcal{M}, E)$. We define the penalty $\text{pen}(\mathcal{M}, E)$ of E as follows:

$$\text{pen}(\mathcal{M}, E) = \text{sat}(\mathcal{M}, \emptyset) - \text{sat}(\mathcal{M}, E)$$

We will sometimes identify possible worlds with the set of literals they make true, writing $\text{pen}(\mathcal{M}, \omega)$. We will also write $\text{pen}(\mathcal{M}, \alpha)$, with α a ground formula, as a shorthand for $\text{pen}(\mathcal{M}, \{\alpha\})$. We will consider the following inference relation, which has been considered among others in [13]:

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

with \mathcal{M} an MLN, α a ground formula and E a set of ground formulas. It can be shown that checking $(\mathcal{M}, E) \vdash_{\text{MAP}} \alpha$ for a ground network \mathcal{M} is Δ_2^P -complete² [4].

2.2 POSSIBILISTIC LOGIC

A possibility distribution in a universe Ω is a mapping π from Ω to $[0, 1]$, encoding our knowledge about the possible values that a given variable X can take; throughout this paper, we will assume that all universes are finite. For each $x \in \Omega$, $\pi(x)$ is called the possibility degree of x . By convention, in a state of complete ignorance, we have $\pi(x) = 1$ for all $x \in \Omega$; conversely, if $X = x_0$ is known, we have $\pi(x_0) = 1$ and $\pi(x) = 0$ for $x \neq x_0$. Possibility theory [27, 12] is based on the possibility measure Π and dual necessity measure N , induced by a possibility distribution π as follows ($A \subseteq \Omega$):

$$\Pi(A) = \max_{a \in A} \pi(a)$$

$$N(A) = 1 - \Pi(\Omega \setminus A)$$

²The complexity class Δ_2^P contains those decision problems that can be solved in polynomial time on a deterministic Turing machine with access to an NP oracle.

¹<http://arxiv.org/abs/1506.01432>

Intuitively, $\Pi(A)$ is the degree to which available evidence is compatible with the view that X belongs to A , whereas $N(A)$ is the degree to which available evidence implies that X belongs to A , i.e. the degree to which it is certain that X belongs to A .

A theory in possibilistic logic [9] is a set of formulas of the form (α, λ) , where α is a propositional formula and $\lambda \in [0, 1]$ is a certainty weight. A possibility distribution π satisfies (α, λ) iff $N(\llbracket \alpha \rrbracket) \geq \lambda$, with N the necessity measure induced by π and $\llbracket \alpha \rrbracket$ the set of propositional models of α . We say that a possibilistic logic theory Θ entails (α, λ) , written $\Theta \models (\alpha, \lambda)$, if every possibility distribution which satisfies all the formulas in Θ also satisfies (α, λ) . A possibility distribution π_1 is called less specific than a possibility distribution π_2 if $\pi_1(\omega) \geq \pi_2(\omega)$ for every ω . It can be shown that the set of models of Θ always has a least element w.r.t. the minimal specificity ordering, which is called the least specific model π^* of Θ . It is easy to see that $\Theta \models (\alpha, \lambda)$ iff π^* satisfies (α, λ) .

Even though the semantics of possibilistic logic is defined at the propositional level, we will also use first-order formulas such as $(p(X) \rightarrow q(X, Y), \lambda)$ throughout the paper. As in Markov logic, we will interpret these formulas as abbreviations for a set of propositional formulas, obtained using grounding in the usual way. In particular, we will always assume that first-order formulas are defined w.r.t. a finite set of constants.

The λ -cut Θ_λ of a possibilistic logic theory Θ is defined as follows:

$$\Theta_\lambda = \{\alpha \mid (\alpha, \mu) \in \Theta, \mu \geq \lambda\}$$

It can be shown that $\Theta \models (\alpha, \lambda)$ iff $\Theta_\lambda \models \alpha$, which means that inference in possibilistic logic can straightforwardly be implemented using a SAT solver.

An inconsistency-tolerant inference relation \vdash_{poss} for possibilistic logic can be defined as follows:

$$\Theta \vdash_{poss} \alpha \quad \text{iff} \quad \Theta_{con(\Theta)} \models \alpha$$

where the consistency level $con(\Theta)$ of Θ is the lowest certainty level λ for which Θ_λ is satisfiable (among the certainty levels that occur in Θ). Note that all formulas with a certainty level below $con(\Theta)$ are ignored, even if they are unrelated to any inconsistency in Θ . This observation is known as the drowning effect.

We will write $(\Theta, E) \vdash_{poss} \alpha$, with E a set of propositional formulas, as an abbreviation for $\Theta \cup \{(e, 1) \mid e \in E\} \vdash_{poss} \alpha$. Despite its conceptual simplicity, \vdash_{poss} has many desirable properties. Among others, it is closely related to AGM belief revision [8] and default reasoning [2]. It can be shown that checking $\Theta \vdash_{poss} (\alpha, \lambda)$ is a Θ_2^P complete problem³ [17]. In this paper, \vdash_{poss} will allow us to capture the non-monotonicity of MAP inference.

³The complexity class Θ_2^P contains those decision problems

Example 2. Let Θ consist of the following formulas:

$$\begin{aligned} &(\text{penguin}(X) \rightarrow \text{bird}(X), 1) \\ &(\text{penguin}(X) \rightarrow \neg \text{flies}(X), 1) \\ &(\text{bird}(X) \rightarrow \text{flies}(X), 0.5) \end{aligned}$$

Then we find:

$$\begin{aligned} &(\Theta, \{\text{bird}(\text{tweety})\}) \vdash_{poss} \text{flies}(\text{tweety}) \\ &(\Theta, \{\text{bird}(\text{tweety}), \text{penguin}(\text{tweety})\}) \vdash_{poss} \neg \text{flies}(\text{tweety}) \end{aligned}$$

In general, \vdash_{poss} allows us to model rules with exceptions, by ensuring that rules about specific contexts have a higher certainty weight than rules about general contexts.

3 ENCODING GROUND NETWORKS

Throughout this section, we will assume that \mathcal{M} is a ground MLN in which all the weights are strictly positive. This can always be guaranteed for ground MLNs by replacing formulas (α, λ) with $\lambda < 0$ by $(\neg\alpha, -\lambda)$, and by discarding any formula whose weight is 0. For a subset $X \subseteq \mathcal{M}$, we write X^* for the set of corresponding classical formulas, e.g. for $X = \{(F_1, w_1), \dots, (F_n, w_n)\}$ we have $X^* = \{F_1, \dots, F_n\}$. In particular, \mathcal{M}^* are the classical formulas appearing in the MLN \mathcal{M} .

The following transformation constructs a possibilistic logic theory that is in some sense equivalent to a given MLN. It is inspired by the probability-possibility transformation from [10].

Transformation 1. We define the possibilistic logic theory $\Theta_{\mathcal{M}}$ corresponding to an MLN \mathcal{M} as follows:

$$\{(\bigvee X^*, \phi(\neg \bigvee X^*)) \mid X \subseteq \mathcal{M}, \phi(\neg \bigvee X^*) > 0\} \quad (3)$$

where for a propositional formula α :

$$\phi(\alpha) = \begin{cases} \frac{K + \text{pen}(\mathcal{M}, \alpha)}{L} & \text{if } \alpha \text{ satisfies the hard constraints} \\ 1 & \text{otherwise} \end{cases}$$

and the constants K and L are chosen such that $0 = \phi(\top) \leq \phi(\alpha) < 1$ for every α that satisfies the hard constraints (i.e. the formulas with weight $+\infty$).

In the following we will use the notations $\phi(\omega)$ for a possible world ω and $\phi(E)$ for a set of formulas E , defined entirely analogously. Throughout the paper we will also write $(-K < x < L - K)$:

$$\lambda_x = \frac{K + x}{L}$$

The correctness of Transformation 1 follows from the next proposition, which is easy to show.

that can be solved in polynomial time on a deterministic Turing machine, by making at most a logarithmic number of calls to an NP oracle.

Proposition 1. Let \mathcal{M} be a ground MLN and $\Theta_{\mathcal{M}}$ the corresponding possibilistic logic theory. Let π be the least specific model of $\Theta_{\mathcal{M}}$. It holds that:

$$\pi(\omega) = 1 - \phi(\omega)$$

Corollary 1. Let \mathcal{M} be a ground MLN and $\Theta_{\mathcal{M}}$ the corresponding possibilistic logic theory. It holds that for $\lambda < 1$:

$$\Theta_{\mathcal{M}} \models (\alpha, \lambda) \quad \text{iff} \quad \text{pen}(\mathcal{M}, \neg\alpha) \geq \lambda L - K$$

and

$$\Theta_{\mathcal{M}} \models (\alpha, 1) \quad \text{iff} \quad \text{pen}(\mathcal{M}, \neg\alpha) = +\infty$$

Corollary 2. Let \mathcal{M} be a ground MLN and $\Theta_{\mathcal{M}}$ the corresponding possibilistic logic theory. For $p_{\mathcal{M}}$ the probability distribution induced by \mathcal{M} and π the least specific model of $\Theta_{\mathcal{M}}$, it holds that

$$p_{\mathcal{M}}(\omega_1) > p_{\mathcal{M}}(\omega_2) \quad \text{iff} \quad \pi(\omega_1) > \pi(\omega_2)$$

for all possible worlds ω_1 and ω_2 . In particular, it follows that for every propositional formula α and every set of propositional formulas E :

$$(\mathcal{M}, E) \vdash_{\text{MAP}} \alpha \quad \text{iff} \quad (\Theta, E) \vdash_{\text{poss}} \alpha \quad (4)$$

Example 3. Consider the MLN \mathcal{M} containing the following formulas:

$$5 : a \rightarrow x \quad 5 : a \rightarrow y \quad 10 : a \wedge b \rightarrow \neg y$$

Then $\Theta_{\mathcal{M}}$ contains the following formulas:

$$\begin{aligned} \lambda_5 : a \rightarrow x & & \lambda_5 : a \rightarrow y \\ \lambda_{10} : a \wedge b \rightarrow \neg y & & \lambda_{10} : a \rightarrow x \vee y \\ \lambda_{15} : a \wedge b \rightarrow x \vee \neg y & & \end{aligned}$$

It can be verified that:

$$(\Theta_{\mathcal{M}}, \{a\}) \vdash_{\text{poss}} x \wedge y \quad (\Theta_{\mathcal{M}}, \{a, b\}) \vdash_{\text{poss}} x \wedge \neg y$$

An important drawback of the transformation to possibilistic logic is that the number of formulas in $\Theta_{\mathcal{M}}$ is exponential in $|\mathcal{M}|$. This makes the transformation inefficient, and moreover limits the interpretability of the possibilistic logic theory. In general, the exponential size of $\Theta_{\mathcal{M}}$ cannot be avoided if we want (4) to hold for any E and α . However, more compact theories can be found if we focus on specific types of evidence. Sections 3.1 and 3.2 introduce two practical methods to accomplish this.

3.1 SELECTIVELY AVOIDING DROWNING

In many applications, we are only interested in particular types of evidence sets E . For example, we may only

be interested in evidence sets that contain at most k literals, or in evidence sets that only contain positive literals. In such cases, we can often derive a more compact possibilistic logic theory $\Theta^{\mathcal{E}}$ as follows. Let \mathcal{E} be the set of evidence sets that we wish to consider, where each $E \in \mathcal{E}$ is a set of ground formulas. Given $E \in \mathcal{E}$ we write \mathcal{S}_E for the set of all minimal subsets $\{F_1, \dots, F_l\}$ of $\mathcal{M}_E^* = \{F \mid F \in \mathcal{M}^*, \text{pen}(\mathcal{M}, \neg F) < \text{pen}(\mathcal{M}, E)\}$ s.t.

$$\text{pen}(\mathcal{M}, \bigwedge E \wedge \neg F_1 \wedge \dots \wedge \neg F_l) > \text{pen}(\mathcal{M}, E) \quad (5)$$

The following transformation constructs a possibilistic logic theory that correctly captures MAP inference for evidence sets in \mathcal{E} . The basic intuition is that we want to weaken the formulas in \mathcal{M}^* just enough to ensure that the resulting certainty level prevents them from drowning when the evidence E becomes available.

Transformation 2. Given a ground MLN \mathcal{M} and a set of evidence sets \mathcal{E} , we define the possibilistic logic theory $\Theta_{\mathcal{M}}^{\mathcal{E}}$ as follows:

$$\{(F_1, \phi(\neg F_1)) \mid F_1 \in \mathcal{M}^*\} \quad (6)$$

$$\cup \{(\neg \bigwedge E \vee \bigvee Z, \phi(\bigwedge E \wedge \neg \bigwedge Z)) \mid Z \in \mathcal{S}_E, \quad (7)$$

$$E \in \mathcal{E}\} \cup \{(\neg \bigwedge E, \phi(\bigwedge E)) \mid E \in \mathcal{E}\} \quad (8)$$

If \mathcal{M} is clear from the context, we will omit the subscript in $\Theta_{\mathcal{M}}^{\mathcal{E}}$. The formulas in (6) are the direct counterpart of the MLN. Intuitively, there are two reasons why these formulas are not sufficient. First, due to the drowning effect, formulas F such that $\text{pen}(\mathcal{M}, \neg F) < \text{pen}(\mathcal{M}, E)$ will be ignored under the evidence E . In such cases we should look at minimal ways to weaken these formulas such that the certainty level of the resulting formula is sufficient to avoid drowning under the evidence E . This is accomplished by adding the formulas in (7). Second, as $\Theta^{\mathcal{E}}$ contains less information than $\Theta_{\mathcal{M}}$, we need to ensure that the consistency level for $\Theta^{\mathcal{E}}$ is never lower than the consistency level for $\Theta_{\mathcal{M}}$, given an evidence set $E \in \mathcal{E}$. To this end, $\Theta^{\mathcal{E}}$ includes the formulas in (8). The following example illustrates why these formulas are needed.

Example 4. Consider the following MLN \mathcal{M} :

$$\begin{aligned} 3 : u & & 2 : a & & 10 : (a \vee b) \wedge (u \vee v) \rightarrow \neg x \\ 2 : b & & 1 : v & & \end{aligned}$$

and let $\mathcal{E} = \{\{x\}\}$, i.e. the only evidence set in which we are interested is $\{x\}$. It holds that

$$\mathcal{S}_E = \{\{a, u\}, \{b, u\}, \{a, v\}, \{b, v\}\} \quad (9)$$

and $\Theta^{\mathcal{E}} = \Theta \cup \Psi \cup \Gamma$, where:

$$\Theta = \{(u, \lambda_3), (a, \lambda_2), ((a \vee b) \wedge (u \vee v) \rightarrow \neg x, \lambda_{10}), (b, \lambda_2), (v, \lambda_1)\}$$

$$\Psi = \{(a \vee u \vee \neg x, \lambda_6), (b \vee u \vee \neg x, \lambda_6), (a \vee v \vee \neg x, \lambda_5), (b \vee v \vee \neg x, \lambda_5)\}$$

$$\Gamma = \{(\neg x, \lambda_4)\}$$

It is easy to verify that $(\Theta \cup \Psi, \{x\}) \vdash_{poss} u$ whereas $(\mathcal{M}, \{x\}) \not\vdash_{MAP} u$ and $(\Theta \cup \Psi \cup \Gamma, \{x\}) \not\vdash_{poss} u$.

We now prove the correctness of Transformation 2.

Proposition 2. *For any formula α and any evidence set $E \in \mathcal{E}$, it holds that $(\Theta_{\mathcal{M}}, E) \vdash_{poss} \alpha$ iff $(\Theta^{\mathcal{E}}, E) \vdash_{poss} \alpha$.*

Proof. Let us introduce the following notation:

$$\begin{aligned}\lambda_E &= \text{con}(\Theta_{\mathcal{M}} \cup \{(e, 1) \mid e \in E\}) \\ \lambda_E^{\mathcal{E}} &= \text{con}(\Theta^{\mathcal{E}} \cup \{(e, 1) \mid e \in E\}) \\ A &= (\Theta_{\mathcal{M}} \cup \{(e, 1) \mid e \in E\})_{\lambda_E} \\ A^E &= (\Theta^{\mathcal{E}} \cup \{(e, 1) \mid e \in E\})_{\lambda_E^{\mathcal{E}}}\end{aligned}$$

We need to show that A is equivalent to A^E , for any $E \in \mathcal{E}$.

By Corollary 1, we know that every formula (α, λ) in $\Theta_{\mathcal{M}}^{\mathcal{E}}$ is entailed by $\Theta_{\mathcal{M}}$, hence $\lambda_E^{\mathcal{E}} \leq \lambda_E$. Since λ_E is the smallest certainty level from $\Theta_{\mathcal{M}}$ which is strictly higher than $\phi(E)$, it follows that A^E contains every formula which appears in $\Theta^{\mathcal{E}}$ with a weight that is strictly higher than $\phi(E)$. Moreover, since $\Theta^{\mathcal{E}}$ by construction contains $(\neg \bigwedge E, \phi(\bigwedge E))$, we find that A^E can only contain such formulas:

$$A^E = E \cup \{\alpha \mid (\alpha, \lambda) \in \Theta^{\mathcal{E}}, \lambda > \phi(E)\} \quad (10)$$

It follows that $A \models A^E$.

Let $G_1 \vee \dots \vee G_s$ be a formula from A . From Corollary 1 we know that:

$$\text{pen}(\mathcal{M}, \neg G_1 \wedge \dots \wedge \neg G_s) > \text{pen}(\mathcal{M}, E)$$

and a fortiori

$$\text{pen}(\mathcal{M}, E \wedge \neg G_1 \wedge \dots \wedge \neg G_s) > \text{pen}(\mathcal{M}, E)$$

This means that for any formula $G_1 \vee \dots \vee G_s$ in A , either $\text{pen}(\mathcal{M}, \neg G_i) > \text{pen}(\mathcal{M}, E)$ for some i or \mathcal{S}_E contains a subset $\{H_1, \dots, H_r\}$ of $\{G_1, \dots, G_s\}$. Then $\Theta^{\mathcal{E}}$ contains either G_i or the formula $\neg E \vee H_1 \dots \vee H_r$ with a weight which is strictly higher than $\phi(E)$ and thus either G_i or $\neg E \vee H_1 \dots \vee H_r$ belongs to A^E . In both cases we find $A^E \models G_1 \vee \dots \vee G_s$. We conclude $A^E \models A$. \square

An alternative, which would make the approach in this section closer to the standard encoding in (1), is to define \mathcal{S}'_E as the set of minimal subsets $\{F_1, \dots, F_l\}$ of \mathcal{M}_E^* such that

$$\text{pen}(\mathcal{M}, \neg F_1 \wedge \dots \wedge \neg F_l) > \text{pen}(\mathcal{M}, E) \quad (11)$$

and then replace the formulas in (7) by

$$\{(\bigvee Z, \phi(\neg \bigwedge Z)) \mid Z \in \mathcal{S}'_E\} \quad (12)$$

The advantage of (7), however, is that we can expect many of the sets in \mathcal{S}_E to be singletons. To see why this is

the case, first note that for each world ω in $\max(\mathcal{M}, E)$, the set of formulas $\mathcal{Y} \subseteq \mathcal{M}^*$ satisfied by ω is such that $\text{pen}(\mathcal{M}, \neg \bigvee (\mathcal{M}^* \setminus \mathcal{Y}))$ is minimal among all sets $\mathcal{Y}' \subseteq \mathcal{M}^*$ for which $E \wedge \bigwedge \mathcal{Y}'$ is consistent. Let us write $\text{Cons}_E(\mathcal{M})$ for the set of all these maximally consistent subsets of \mathcal{M}^* . Note that $\max(\mathcal{M}, E) = \llbracket \bigvee \{ \bigwedge \mathcal{Y} \mid \mathcal{Y} \in \text{Cons}_E(\mathcal{M}) \} \rrbracket$.

Lemma 1. *For a set of formulas $\{F_1, \dots, F_l\} \subseteq \mathcal{M}^*$ it holds that $\text{pen}(\mathcal{M}, E \wedge \neg F_1 \wedge \dots \wedge \neg F_l) > \text{pen}(\mathcal{M}, E)$ iff $\{F_1, \dots, F_l\} \cap \mathcal{Y} \neq \emptyset$ for every \mathcal{Y} in $\text{Cons}_E(\mathcal{M})$.*

Corollary 3. *Let $\text{Cons}_E(\mathcal{M}) = \{\mathcal{Y}_1, \dots, \mathcal{Y}_s\}$. It holds that \mathcal{S}_E consists of the subset-minimal elements of $\{\{y_1, \dots, y_s\} \mid y_1 \in \mathcal{Y}_1 \cap \mathcal{M}_E^*, \dots, y_s \in \mathcal{Y}_s \cap \mathcal{M}_E^*\}$.*

Example 5. *Consider again the MLN \mathcal{M} from Example 4 and let $E = \{x\}$. It holds that $\text{Cons}_E(\mathcal{M}) = \{\mathcal{Y}_1, \mathcal{Y}_2\}$, where*

$$\begin{aligned}\mathcal{Y}_1 &= \{(a \vee b) \wedge (u \vee v) \rightarrow \neg x, a, b\} \\ \mathcal{Y}_2 &= \{(a \vee b) \wedge (u \vee v) \rightarrow \neg x, u, v\} \\ \mathcal{M}_E^* &= \{a, b, u, v\}\end{aligned}$$

From Corollary 3, it follows that \mathcal{S}_E is given by (9).

In practice, $\text{Cons}_E(\mathcal{M})$ will often contain a single element, in which case all the elements of \mathcal{S}_E will be singletons.

3.2 MAP INFERENCE AS DEFAULT REASONING

A large number of approaches has been proposed for reasoning with a set of default rules of the form “if α then typically β ” [15, 19, 14]. At the core, each of the proposed semantics corresponds to the intuition that a set of default rules imposes a preference order on possible worlds, where “if α then β ” means that β is true in the most preferred models of α . The approaches from [15] and [19] can be elegantly captured in possibilistic logic [2], by interpreting the default rule as the constraint $\Pi(\alpha \wedge \beta) > \Pi(\alpha \wedge \neg \beta)$. In Markov logic, the same constraint on the ordering of possible worlds can be expressed by imposing the constraint $(\mathcal{M}, \alpha) \vdash_{MAP} \beta$. In other words, we can view the MAP consequences of an MLN as a set of default rules, and encode these default rules in possibilistic logic. The following transformation is based on this idea.

Transformation 3. *Given a ground MLN \mathcal{M} and a positive integer k , we construct a possibilistic logic theory $\Theta_{\mathcal{M}}^k$ as follows:*

- For each hard rule F from \mathcal{M} , add $(F, 1)$ to $\Theta_{\mathcal{M}}^k$.
- For each set of literals E such that $0 \leq |E| \leq k$, let $X = \{x \mid (\mathcal{M}, E) \vdash_{MAP} x\}$ be the set of literals that are true in all the most plausible models of E . Unless there is a literal $y \in E$ such that $\bigwedge (E \setminus \{y\}) \vdash_{MAP} y$, add

$$\left(\bigwedge E \rightarrow \bigwedge X, \lambda_E \right)$$

to $\Theta_{\mathcal{M}}^k$, where $\lambda_E = \phi(\bigwedge E)$. If $\text{pen}(\mathcal{M}, E) > \text{pen}(\mathcal{M}, \emptyset)$, add also

$$(\neg(\bigwedge E \wedge \bigwedge X), \lambda'_E) \quad (13)$$

where λ'_E is the certainty level just below λ_E in $\Theta_{\mathcal{M}}^k$, i.e. $\lambda_{E'} = \max\{\lambda_F \mid \lambda_F < \lambda_E, |F| \leq k\}$.

If \mathcal{M} is clear from the context, we will omit the subscript in $\Theta_{\mathcal{M}}^k$. The possibilistic encoding of default rules used in Transformation 3 is similar in spirit to the method from [2], which is based on the Z-ranking from [19]. However, because $p_{\mathcal{M}}$ already provides us with a model of the default rules, we can directly encode default rules in possibilistic logic, without having to rely on the Z-ranking. Also note that although the method is described in terms of an MLN, it can be used for encoding any ranking on possible worlds (assuming a finite set of atoms).

As illustrated in the following example, (13) is needed to avoid deriving too much, serving a similar purpose to (8) in the approach from Section 3.1.

Example 6. Consider the following MLN \mathcal{M} :

$$2 : \neg a \vee b \quad 2 : a \vee b \quad 1 : a \vee \neg b$$

Then $\Theta^1 = \Theta \cup \Psi$, where

$$\begin{aligned} \Theta &= \{(\top \rightarrow a \wedge b, \lambda_0), (\neg a \rightarrow b, \lambda_1), (\neg b \rightarrow \top, \lambda_2)\} \\ \Psi &= \{(b, \lambda_1), (a \vee \neg b, \lambda_0)\} \end{aligned}$$

We find $(\Theta, \{\neg b\}) \vdash_{\text{poss}} a$ while $(\mathcal{M}, \{\neg b\}) \not\vdash_{\text{MAP}} a$. Accordingly, we have $(\Theta \cup \Psi, \{\neg b\}) \not\vdash_{\text{poss}} a$.

Transformations 2 and 3 have complementary strengths. For example, Transformation 2 may lead to more compact theories for relatively simple MLNs, e.g. if for most of the considered evidence sets, there is a unique set of formulas from the MLN that characterizes the most probable models of the evidence (cf. Lemma 1). On the other hand, Transformation 3 may lead to substantially more compact theories in cases where the number of formulas is large relative to the number of atoms.

We now show the correctness of Transformation 3.

Proposition 3. Let \mathcal{M} be an MLN, k a positive integer and Θ^k the proposed possibilistic logic encoding of \mathcal{M} . Furthermore, let E and C be sets of literals such that $|E| + |C| \leq k + 1$. It holds that $(\mathcal{M}, E) \vdash_{\text{MAP}} \bigvee C$ if and only if $(\Theta^k, E) \vdash_{\text{poss}} \bigvee C$.

Before we prove Proposition 3, we present a number of lemmas. In the lemmas and proofs below, \mathcal{M} will always be an MLN, Θ^k will be the corresponding possibilistic logic theory and k will be the maximum size of the evidence sets considered in the translation.

Lemma 2. If E is a set of literals, $|E| \leq k$, $\lambda = \phi(E)$ and $(\mathcal{M}, E) \vdash_{\text{MAP}} x$ then

$$\left\{ \left(\bigwedge E' \rightarrow \bigwedge X \right) \in \Theta_{\lambda}^k \text{ s.t. } |E'| \leq |E| \right\} \vdash \bigwedge E \rightarrow x$$

Lemma 3. If $\phi(\omega) \leq \lambda$ then ω is a model of Θ_{λ}^k .

Proof. If there were a formula $F = (\bigwedge E) \rightarrow (\bigwedge X)$ in Θ_{λ}^k that was not satisfied by ω , then its body would have to be true in ω but then necessarily

$$\lambda \leq \phi(E) \leq \phi(\omega) \leq \lambda.$$

The first inequality follows from the fact that, by the construction of Θ^k , if the certainty weight of F is at least λ then it must be the case that $\phi(E) \geq \lambda$. The second inequality follows from the fact that ω was assumed to be a model of $\bigwedge E$. It follows that:

$$\text{pen}(\mathcal{M}, \omega) = \text{pen}(\mathcal{M}, E).$$

However, this would mean that ω is also a most probable world of (\mathcal{M}, E) , but then $\omega \models F$ by construction of Θ^k .

If there were an unsatisfied formula $F = \neg(\bigwedge E \wedge \bigwedge X)$ in Θ_{λ}^k then by construction we would have $\phi(E \cup X) > \lambda$. However, from $\omega \models \bigwedge E \wedge \bigwedge X$ we find $\phi(E \cup X) \leq \phi(\omega) \leq \lambda$, a contradiction.

Since all formulas in Θ^k are of the two considered types, it follows that all formulas from Θ^k whose certainty weight is at least λ must be satisfied in ω . \square

Lemma 4. If $(\mathcal{M}, E) \vdash_{\text{MAP}} (y_1 \vee \dots \vee y_m)$ then

- (i) for any i , either $(\mathcal{M}, E \cup \{\neg y_i\}) \vdash_{\text{MAP}} (y_1 \vee \dots \vee y_{i-1} \vee y_{i+1} \vee \dots \vee y_m)$ or $(\mathcal{M}, E) \vdash_{\text{MAP}} y_i$,
- (ii) there exist a j and a set $\{y'_1, \dots, y'_m\} \subseteq \{y_1, \dots, y_m\} \setminus \{y_j\}$ such that $(\mathcal{M}, E \cup \{\neg y'_1, \dots, \neg y'_m\}) \vdash_{\text{MAP}} y_j$.

Lemma 5. If $|C| + |E| \leq k + 1$, and $\lambda = \phi(E)$ then $\Theta_{\lambda}^k \cup E \vdash \bigvee C$ if and only if $(\mathcal{M}, E) \vdash_{\text{MAP}} \bigvee C$.

We now turn to the proof of Proposition 3.

Proof of Proposition 3. Let E be an evidence set such that $|E| \leq k$ and let $\lambda = \phi(E)$. Given Lemma 5, it is sufficient to show that $\text{con}(\Theta^k, E) = \lambda$. It follows from Lemma 3 that $\text{con}(\Theta^k, E) \leq \lambda$. Let $X = \{x \mid (\mathcal{M}, E) \vdash_{\text{MAP}} x\}$ be the set of literals which can be derived from (\mathcal{M}, E) using MAP inference. By construction, Θ^k contains a formula $\neg(\bigwedge E \wedge \bigwedge X)$ with a certainty weight which is just below λ . Specifically, for $\lambda' < \lambda$ we either have $\Theta_{\lambda'}^k = \Theta_{\lambda}^k$, or $\Theta_{\lambda'}^k \models \neg \bigwedge E$, from which we find $\text{con}(\Theta^k, E) = \lambda$. \square

It is of interest to remove any formulas in Θ^k that are redundant, among others because this is likely to make the theory easier to interpret. Although we can use possibilistic logic inference to identify redundant formulas, in some cases we can avoid adding the redundant formulas altogether. For example, in the transformation procedure, we do not add any rules for E if it holds that $E \setminus \{y\} \vdash_{MAP} y$ for some $y \in E$. This pruning rule is the counterpart of the cautious monotonicity property, which is well-known in the context of default reasoning [15]. Any ranking on possible worlds also satisfies the stronger rational monotonicity property, which translated to our setting states that when $(\mathcal{M}, E \setminus \{y\}) \vdash_{MAP} x$ and $(\mathcal{M}, E \setminus \{y\}) \not\vdash_{MAP} \neg y$ it holds that $(\mathcal{M}, E) \vdash_{MAP} x$. Accordingly, when processing the evidence set E in the transformation procedure, instead of $(\bigwedge E \rightarrow \bigwedge X, \lambda_E)$ it is sufficient to add the following rule:

$$\left(\bigwedge E \rightarrow \bigwedge (X \setminus X_0), \lambda_E \right)$$

where

$$X_0 = \{x \mid E \setminus \{y\} \vdash_{MAP} x \text{ and } E \setminus \{y\} \not\vdash_{MAP} \neg y\}$$

The correctness of this pruning step follows from the following proposition.

Proposition 4. *Let x and y be literals. If $|E| < k$, $(\mathcal{M}, E) \vdash_{MAP} x$ and $(\mathcal{M}, E) \not\vdash_{MAP} \neg y$ then:*

$$\Theta^k \setminus \{F\} \models \left(\bigwedge E \wedge y \rightarrow x, \lambda_{E \cup \{y\}} \right)$$

where F is the formula in Θ^k corresponding to the evidence set $E \cup \{y\}$, i.e.:

$$F = \bigwedge (E \cup \{y\}) \rightarrow \bigwedge \{x \mid (\mathcal{M}, E \cup \{y\}) \vdash_{MAP} x\}$$

Proof. If $(\mathcal{M}, E) \vdash_{MAP} x$ and $(\mathcal{M}, E) \not\vdash_{MAP} \neg y$ then $(\mathcal{M}, E \cup \{y\}) \vdash_{MAP} x$ and $pen(\mathcal{M}, E) = pen(\mathcal{M}, E \cup \{y\}) = pen(\mathcal{M}, E \cup \{x, y\})$. Therefore using Lemma 2, we find that $\Theta^k_{\lambda_{E \cup \{y\}}} \vdash \bigwedge E \rightarrow x$. From Lemma 2, it furthermore follows that $\bigwedge E \rightarrow x$ can be derived from rules with antecedents of length at most $|E|$. In particular, we find that $\bigwedge E \rightarrow x$ can be derived without using the formula $\bigwedge E \wedge y \rightarrow \bigwedge X$. \square

Finally, note that formulas of the form (13) can be omitted when $\lambda'_E = \lambda_{(E \setminus \{y\})}$ for some $y \in E$. Indeed, in such a case we find from $pen(\mathcal{M}, E \setminus \{y\}) < pen(\mathcal{M}, E)$ that $(\mathcal{M}, E \setminus \{y\}) \vdash_{MAP} \neg y$, hence (13) will be entailed by a formula of the form $(\bigwedge (E \setminus \{y\}) \rightarrow \bigwedge X, \lambda_{E \setminus \{y\}})$ in Θ^k .

4 ENCODING NON-GROUND NETWORKS

We now provide the counterpart to the construction from Section 3.2 for non-ground MLNs. The first-order nature

of MLNs often leads to distributions with many symmetries which can be exploited by lifted inference methods [20]. We can similarly exploit these symmetries for constructing more compact possibilistic logic theories from MLNs.

For convenience, in the possibilistic logic theories, we will use *typed* formulas. 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.

Two typed formulas F_1 and F_2 are said to be isomorphic when there is a type-respecting substitution θ of the variables of F_1 such that $F_1\theta \equiv F_2$ (where \equiv denotes equivalence of logical formulas). Two MLNs \mathcal{M}_1 and \mathcal{M}_2 are said to be isomorphic, denoted by $\mathcal{M}_1 \approx \mathcal{M}_2$, if there is a bijection i from formulas of \mathcal{M}_1 to formulas of \mathcal{M}_2 such that for $i(F, w) = (F', w')$ it holds that $w = w'$ and the formulas F and F' are isomorphic. When j is a permutation of a subset of constants from \mathcal{M} then $j(\mathcal{M})$ denotes the MLN obtained by replacing any constant c from the subset by its image $j(c)$.

Given a non-ground MLN \mathcal{M} , we can first identify sets of constants which are *interchangeable*, where a set of constants \mathcal{C}_t is said to be interchangeable if $j(\mathcal{M}) \approx \mathcal{M}$ for any permutation j of the constants in \mathcal{C}_t . Note that to check whether a set of constants \mathcal{C}_t is interchangeable, it is sufficient to check that $j(\mathcal{M}) \approx \mathcal{M}$ for those permutations which swap just two constants from \mathcal{C}_t . For every maximal set \mathcal{C}_t of interchangeable constants, we introduce a new type t . For a constant c , we write $\tau(c)$ to denote its type. When F is a ground formula, *variabilize*(F) denotes the following formula:

$$\bigwedge \{V_c \neq V_d \mid c, d \in const(F), \tau(c) = \tau(d)\} \rightarrow F'$$

where $const(F)$ is the set of constants appearing in F and F' is obtained from F by replacing all constants c by a new variable V_c of type $\tau(c)$.

Transformation 4. *Given an MLN \mathcal{M} and a positive integer k , we construct a possibilistic logic theory $\Theta^k_{\mathcal{M}}$ as follows:*

- For each hard rule F from \mathcal{M} , add $(F, 1)$ to $\Theta^k_{\mathcal{M}}$.
- For each set of literals E such that $0 \leq |E| \leq k$, let $X = \{x \mid (\mathcal{M}, E) \vdash_{MAP} x\}$. For all $x \in X$, unless there is a literal $y \in E$ such that $(\mathcal{M}, E \setminus \{y\}) \vdash_{MAP} y$ and unless $\Theta^k_{\mathcal{M}}$ already contains a formula isomorphic to *variabilize*($\bigwedge E \rightarrow x$), add

$$\left(\text{variabilize} \left(\bigwedge E \rightarrow x \right), \lambda_E \right)$$

to $\Theta_{\mathcal{M}}^k$. If $\text{pen}(\mathcal{M}, E) > \text{pen}(\mathcal{M}, \emptyset)$ and $\Theta_{\mathcal{M}}^k$ does not already contain a formula isomorphic to $\text{variabilize}(\neg(\bigwedge E \wedge \bigwedge X))$, add also

$$\left(\text{variabilize}\left(\neg\left(\bigwedge E \wedge \bigwedge X\right)\right), \lambda'_E\right) \quad (14)$$

where λ'_E is the certainty level just below λ_E in $\Theta_{\mathcal{M}}^k$.

As before, we will usually omit the subscript in $\Theta_{\mathcal{M}}^k$. We can show that after grounding, Θ^k is equivalent to the theory that would be obtained by first grounding the MLN and then applying the method from Section 3.2. The correctness proof is provided in the online appendix.

Our implementation⁴ of Transformation 4 relies on an efficient implementation of inference in possibilistic logic and Markov logic, efficient generation of non-redundant candidate evidence sets and efficient filtering of isomorphic formulas. For MAP inference in MLNs, we used a cutting-plane inference algorithm based on a SAT-based optimization. For inference in possibilistic logic, we also used cutting-plane inference in order to avoid having to ground the whole theory. To find the ground rules that need to be added by the cutting-plane method, we used a modified querying system from [16]. For solving and optimizing the resulting ground programs, we used the SAT4J library [3].

Note that to check whether $\Theta_{\lambda}^k \vdash F$, where F is a (not necessarily ground) clause, it is sufficient to find one (type-respecting) grounding θ of F , and check whether $\Theta_{\lambda}^k \cup \{\neg(F\theta)\}$ is inconsistent. In this way, we can check whether a rule is implied by Θ^k without grounding the whole theory because, as for MLNs, inference in non-ground possibilistic logic theories can be carried out by cutting-plane inference methods.

We implemented the transformation as a modification of the standard best-first search (BFS) algorithm which constructs incrementally larger candidate evidence sets, checks their MAP consequences and adds the respective rules to the possibilistic logic theory being constructed. Like the standard BFS algorithm it uses a hash-table based data structure *closed*, in which already processed evidence sets are stored. In order to avoid having to check isomorphism with every evidence set in *closed*, each time a new evidence set is considered, the stored evidence sets are enriched by fingerprints which contain some invariants, guaranteeing that no two variabilized evidence sets with different fingerprints are isomorphic. In this way, we can efficiently check for a given evidence set E whether there is a previously generated evidence set E' such that $\text{variabilize}(E)$ and $\text{variabilize}(E')$ are isomorphic.

As a final remark, we note that for the non-ground transformation, it may be preferable to replace any

⁴The implementation can be downloaded from: <https://github.com/supertweety/mln2poss>.

rule ($\text{variabilize}(\neg(\bigwedge E \wedge \bigwedge X)), \lambda'_E$) by the rule ($\text{variabilize}(\neg \bigwedge E), \lambda'_E$). The reason is that the former rules may often become too long in the non-ground case. On the other hand, for the ground transformation, the advantage of the longer rules is that they will often be the same for different sets E , which, in effect, means a smaller number of rules in the possibilistic logic theory. The correctness of this alternative to Transformation 4 is also shown in the online appendix.

5 ILLUSTRATIVE EXAMPLES

The first example is a variation on a classical problem from non-monotonic reasoning. Here, we want to express that birds generally fly, but heavy antarctic birds do not fly, unless they have a jet pack. The MLN which we will convert into possibilistic logic contains the following rules: $10 : \text{bird}(X) \rightarrow \text{flies}(X)$, $1 : \text{antarctic}(X) \rightarrow \neg \text{flies}(X)$, $10 : \text{heavy}(X) \rightarrow \neg \text{flies}(X)$, $100 : \text{hasJetPack}(X) \rightarrow \text{flies}(X)$. When presented with this MLN, Transformation 4 produces the following possibilistic logic theory.

$$\begin{aligned} &(\neg \text{antarctic}(X) \vee \neg \text{flies}(X), \lambda_0) \\ &(\neg \text{bird}(X) \vee \text{flies}(X), \lambda_0) \\ &(\neg \text{heavy}(X) \vee \neg \text{flies}(X), \lambda_0) \\ &(\text{flies}(X) \vee \neg \text{hasJetPack}(X), \lambda_0) \\ &(\neg \text{bird}(X) \vee \text{flies}(X) \vee \text{hasJetPack}(X), \lambda_1) \\ &(\neg \text{heavy}(X) \vee \text{antarctic}(X) \vee \neg \text{flies}(X), \lambda_1) \\ &(\neg \text{bird}(X) \vee \neg \text{heavy}(X), \lambda_1) \\ &(\neg \text{antarctic}(X) \vee \neg \text{heavy}(X) \vee \neg \text{flies}(X), \lambda_{10}) \\ &(\text{flies}(X) \vee \neg \text{hasJetPack}(X) \vee \text{bird}(X), \lambda_{11}) \\ &(\neg \text{bird}(X) \vee \text{flies}(X) \vee \neg \text{hasJetPack}(X), \lambda_{100}) \end{aligned}$$

Let us consider the evidence set $E = \{\text{bird}(\text{tweety}), \text{heavy}(\text{tweety})\}$. Then the levels λ_0 and λ_1 drown because of the inconsistency with the rule $(\neg \text{bird}(X) \vee \neg \text{heavy}(X), \lambda_1)$ which was produced as one of the rules (14). We can see from the rest of the possibilistic logic theory that unless we add either $\text{antarctic}(\text{tweety})$ or $\text{hasJetPack}(\text{tweety})$, we cannot say anything about whether *tweety* flies or not. It can be verified that the same is true also for the respective MLN.

The second example consists of formulas from a classical MLN about smokers. There are three predicates in this MLN: a binary predicate $f(A, B)$ denoting that A and B are friends, and two unary predicates $s(A)$ and $c(A)$ denoting that A smokes and that A has cancer, respectively. The MLN contains the following hard rules: $\neg f(A, B) \vee f(B, A)$ and $\neg f(A, A)$. In addition, we have two soft rules. The first soft rule $10 : \neg s(A) \vee \neg f(A, B) \vee s(B)$ states that if A and B are friends and A smokes then B is more likely to smoke too. The second rule $10 : \neg s(A) \vee c(A)$ states that smoking increases the likelihood of cancer. The following possi-

bilistic logic theory was obtained using Transformation 4 with $k = 4$.

$$\begin{aligned}
& (s(B) \vee \neg f(A, B) \vee \neg s(A) \vee \neg \text{alldiff}(A, B), \lambda_0) \\
& \quad (\neg s(A) \vee c(A), \lambda_0) \\
& (\neg f(C, B) \vee \neg f(A, B) \vee s(A) \vee s(C) \\
& \quad \vee \neg \text{alldiff}(A, B, C) \vee \neg s(B), \lambda_{10}) \\
& (\neg f(C, B) \vee \neg s(A) \vee \neg f(A, C) \vee s(C) \\
& \quad \vee \neg \text{alldiff}(A, B, C) \vee \neg s(B), \lambda_{10}) \\
& (\neg s(A) \vee \neg f(C, A) \vee s(C) \vee c(B) \\
& \quad \vee \neg \text{alldiff}(A, B, C) \vee \neg s(B), \lambda_{10}) \\
& (\neg s(A) \vee c(A) \vee c(B) \vee \neg s(B) \vee \neg \text{alldiff}(A, B), \lambda_{10}) \\
& (s(B) \vee \neg f(A, B) \vee \neg s(A) \vee c(A) \vee \neg \text{alldiff}(A, B), \lambda_{10}) \\
& \quad (\neg f(A, B) \vee f(B, A), 1) \\
& \quad (\neg f(A, A), 1)
\end{aligned}$$

At the lowest level λ_0 we find the counterparts of the soft rules from the MLN, whereas at level 1 we find the hard rules. At the intermediate level we intuitively find weakened rules from the MLN. For instance, the rule $(\neg s(A) \vee c(A) \vee c(B) \vee \neg s(B) \vee \neg \text{alldiff}(A, B), \lambda_1)$ can be interpreted as: if A and B smoke then at least one of them has cancer. It is quite natural that this rule has higher certainty weight than the rule: if A smokes then A has cancer.

A final, more elaborate example is provided in the online appendix.

6 RELATED WORK

One line of related work focuses on extracting a comprehensible model from another learned model that is difficult or impossible to interpret. A seminal work in this area is the TREPAN [5] algorithm. Given a trained neural network and a data set, TREPAN learns a decision tree to mimic the predictions of the neural network. In addition to producing interpretable output, this algorithm was shown to learn accurate models that faithfully mimicked the neural network’s predictions. More recent research has focused on approximating complex ensemble classifiers with a single model. For example, Popovic et al. [21] proposed a method for learning a single decision tree that mimics the predictions of a random forest.

While, to the best of our knowledge, this is the first paper that studies the relation between Markov logic and possibilistic logic, the links between possibility theory and probability theory have been widely studied. For example, [10] has proposed a probability-possibility transformation based on the view that a possibility measure corresponds to a particular family of probability measures. Dempster-Shafer evidence theory [25] has also been used to provide a probabilistic interpretation to possibility degrees. In particular, a possibility distribution can be interpreted as the contour

function of a mass assignment; see [11] for details. In [13] it is shown how the probability distribution induced by a penalty logic theory corresponds to the contour function of a mass assignment, which suggests that it is indeed natural to interpret this probability distribution as a possibility distribution. Several other links between possibility theory and probability theory have been discussed in [6].

In this paper, we have mainly focused on MAP inference. An interesting question is whether it would be possible to construct a (possibilistic) logic base that captures the set of accepted beliefs encoded by a probability distribution, where A is accepted if $P(A) > P(\neg A)$. Unfortunately, the results in [7] show that this is only possible for the limited class of so-called big-stepped probability distributions. In practice, this means that we would have to define a partition of the set of possible worlds, such that the probability distribution over the partition classes is big-stepped, and only capture the beliefs that are encoded by the latter, less informative, probability distribution. A similar approach was taken in [1] to learn default rules from data.

7 CONCLUSIONS

This paper has focused on how a Markov logic network \mathcal{M} can be encoded in possibilistic logic. We started from the observation that it is always possible to construct a possibilistic logic theory $\Theta_{\mathcal{M}}$ that is equivalent to \mathcal{M} , in the sense that the probability distribution induced by \mathcal{M} is isomorphic to the possibility distribution induced by $\Theta_{\mathcal{M}}$. As a result, applying possibilistic logic inference to $\Theta_{\mathcal{M}}$ yields the same conclusions as applying MAP inference to \mathcal{M} . Although the size of $\Theta_{\mathcal{M}}$ is exponential in the number of formulas in \mathcal{M} , we have shown how more compact theories can be obtained in cases where we can put restrictions on the types of evidence that need to be considered (e.g. small sets of literals).

Our main motivation has been to use possibilistic logic as a way to make explicit the assumptions encoded in a given MLN. Among others, the possibilistic logic theory could be used to generate explanations for predictions made by the MLN, to gain insight into the data from which the MLN was learned, or to identify errors in the structure or weights of the MLN. Taking this last idea one step further, our aim for future work is to study methods for repairing a given MLN, based on the mistakes that have thus been identified.

Acknowledgements

We would like to thank the anonymous reviewers for their helpful comments. This work has been supported by a grant from the Leverhulme Trust (RPG-2014-164). JD is partially supported by the Research Fund KU Leuven (OT/11/051), EU FP7 Marie Curie Career Integration Grant (294068) and FWO-Vlaanderen(G.0356.12).

References

- [1] S. Benferhat, D. Dubois, S. Lagrue, and H. Prade. A big-stepped probability approach for discovering default rules. *Int. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 11:1–14, 2003.
- [2] S. Benferhat, D. Dubois, and H. Prade. Nonmonotonic reasoning, conditional objects and possibility theory. *Artificial Intelligence*, 92(1-2):259–276, 1997.
- [3] D. L. Berre and A. Parrain. The SAT4J library, release 2.2. *Journal on Satisfiability, Boolean Modeling and Computation*, 7:50–64, 2010.
- [4] C. Cayrol and M.-C. Lagasquie-Schiex. On the complexity of non-monotonic entailment in syntax-based approaches. In *Proceedings of the 11th ECAI Workshop on Algorithms, Complexity and Commonsense Reasoning*, 1994.
- [5] M. W. Craven and J. W. Shavlik. Extracting tree-structured representations of trained networks. In D. Touretzky, M. Mozer, and M. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 24–30. MIT Press, 1996.
- [6] D. Dubois. Possibility theory and statistical reasoning. *Computational statistics & data analysis*, 51(1):47–69, 2006.
- [7] D. Dubois, H. Fargier, and H. Prade. Ordinal and probabilistic representations of acceptance. *Journal of Artificial Intelligence Research*, 22:23–56, 2004.
- [8] D. Dubois, J. Lang, and H. Prade. Automated reasoning using possibilistic logic: semantics, belief revision, and variable certainty weights. *IEEE Trans. on Knowledge and Data Engineering*, 6(1):64–71, 1994.
- [9] D. Dubois, J. Lang, and H. Prade. Possibilistic logic. In D. N. D. Gabbay, C. Hogger J. Robinson, editor, *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 3, pages 439–513. Oxford University Press, 1994.
- [10] D. Dubois and H. Prade. On several representations of an uncertain body of evidence. In M. Gupta and E. Sanchez, editors, *Fuzzy Information and Decision Processes*, pages 167–181. North-Holland, 1982.
- [11] D. Dubois and H. Prade. Fuzzy sets, probability and measurement. *European Journal of Operational Research*, 40(2):135–154, 1989.
- [12] D. Dubois and H. Prade. Possibility theory: qualitative and quantitative aspects. In D. Gabbay and P. Smets, editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume 1, pages 169–226. Kluwer Academic, 1998.
- [13] F. Dupin de Saint-Cyr, J. Lang, and T. Schiex. Penalty logic and its link with Dempster-Shafer theory. In *Proc. of the 10th International Conference on Uncertainty in Artificial Intelligence*, pages 204–211, 1994.
- [14] H. Geffner and J. Pearl. Conditional entailment: Bridging two approaches to default reasoning. *Artificial Intelligence*, 53(2):209–244, 1992.
- [15] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44(1-2):167–207, 1990.
- [16] O. Kuželka and F. Železný. A restarted strategy for efficient subsumption testing. *Fundamenta Informaticae*, 89(1):95–109, 2008.
- [17] J. Lang. Possibilistic logic: complexity and algorithms. In J. Kohlas and S. Moral, editors, *Algorithms for Uncertainty and Defeasible Reasoning*, volume 5 of *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, pages 179–220. Kluwer Academic Publishers, 2001.
- [18] J. Noessner, M. Niepert, and H. Stuckenschmidt. RockIt: Exploiting parallelism and symmetry for map inference in statistical relational models. In *Proc. of the 27th Conf. on Artificial Intelligence (AAAI)*, 2013.
- [19] J. Pearl. System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In *Proc. of the 3rd Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 121–135, 1990.
- [20] D. Poole. First-order probabilistic inference. In *Proceedings of IJCAI*, volume 18, pages 985–991, 2003.
- [21] D. Popovic, A. Sifrim, Y. Moreau, and B. D. Moor. eXtasy simplified - towards opening the black box. In *Proc. of the IEEE International Conf. on Bioinformatics and Biomedicine (BIBM)*, pages 24–28, 2013.
- [22] M. Richardson and P. Domingos. Markov logic networks. *Machine Learning*, 62(1-2):107–136, 2006.
- [23] S. Riedel. Improving the accuracy and efficiency of MAP inference for Markov logic. In *Proc. of the 24th Conference on Uncertainty in Artificial Intelligence*.
- [24] M. Serrurier and H. Prade. Introducing possibilistic logic in ILP for dealing with exceptions. *Artificial Intelligence*, 171(16–17):939–950, 2007.
- [25] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, 1976.
- [26] M. Thimm. Coherence and compatibility of markov logic networks. In *Proc. of the 21st European Conference on Artificial Intelligence*, pages 891–896, 2014.
- [27] L. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.