Proceedings of the Twenty-Sixth International Florida Artificial Intelligence Research Society Conference

Some Notes on the Factorization of Probabilistic Logical Models under Maximum Entropy Semantics

Nico Potyka
Department of Computer Science
FernUniversität in Hagen,
Germany

Abstract

Probabilistic conditional logics offer a rich and well-founded framework for designing expert systems. The factorization of their maximum entropy models has several interesting applications. In this paper a general factorization is derived providing a more rigorous proof than in previous work. It yields an approach to extend Iterative Scaling variants to deterministic knowledge bases. Subsequently the connection to Markov Random Fields is revisited.

1 Introduction

Probabilistic logics (Nilsson 1986) combine logic and probability theory. For designing expert systems in particular probabilistic conditional logics (Kern-Isberner 2001) are interesting, as they allow the definition of intuitive conditional formulas like $(\mathit{Flies}\,|\,\mathit{Bird})[0.9]$ expressing that the probability that a bird can fly is 90%. To obtain a complete probability distribution over logical atoms, such conditionals can be regarded as constraints for an entropy maximization problem (Paris 1994). We call the distribution satisfying a knowledge base $\mathcal{R}$ and maximizing entropy the ME-model of $\mathcal{R}$.

In practice, ME optimization problems are often solved by introducing lagrange multipliers for the constraints. As noted in (Fisseler 2010), the solution is necessarily a Markov Random Field (MRF), that is, it has the form $\mathcal{P}(\omega) = \frac{1}{Z}\exp(\sum_i \lambda_i \cdot f_i(\omega)) = \frac{1}{Z}\prod_i \exp(\lambda_i \cdot f_i(\omega))$, where λ_i denotes a lagrange multiplier corresponding to the i-th constraint function f_i. This 'factorization' has several interesting applications, as it can be used for the design of learning (Kern-Isberner 2001) and inference algorithms (Finthammer and Beierle 2012). One can show that if a factorization exists that satisfies linear constraints it is necessarily the unique ME-model (Darroch and Ratcliff 1972) satisfying these constraints. However, existence of lagrange multipliers and hence existence of the factorization is not self-explanatory. In previous work the existence of a factorization has been investigated only for positive ME-models of propositional languages thoroughly.

We close this foundational gap. The most important basics are explained in Section 2. In Section 3 we show that for each 'regular' knowledge base under each 'linear semantics' the ME-model factorizes. Besides propositional languages and their usual semantics, the result in particular covers relational languages under 'grounding' (Fisseler 2010) and 'aggregating' semantics (Kern-Isberner and Thimm 2010). Furthermore, the proof yields an approach to extend Iterative Scaling (Darroch and Ratcliff 1972) variants like the highly optimized implementation presented in (Finthammer and Beierle 2012) to non-positive ME-models. As noted in (Fisseler 2010) the factorization of ME-models corresponds to the factorization of MRFs. We show how the corresponding minimal Markov network can be constructed for different languages and semantics. The Markov network is of great practical interest, as it can be used to decompose the ME-model into easier computable local distributions.

2 Languages and Semantics

We consider logical languages $\mathcal{L}_\mathcal{A}$ built up over *alphabets* $\mathcal{A} = (\mathit{Const}, \mathit{Var}, \mathit{Rel}, \mathit{Op})$ partitioned into a finite set *Const* of *constants*, a set *Var* of *variables*, a finite set *Rel* of *relation symbols* and a set of logical operators *Op*. A relation symbol of arity zero is called a proposition. For ease of notation we abbreviate conjunction by juxtaposition, $\mathfrak{f}\mathfrak{g} := \mathfrak{f} \wedge \mathfrak{g}$, and negation by an overbar, $\overline{\mathfrak{f}} := \neg\mathfrak{f}$ in the following. We can extend $\mathcal{L}_\mathcal{A}$ to a *probabilistic conditional language* $(\mathcal{L}_\mathcal{A}|\mathcal{L}_\mathcal{A}) := \{(\mathfrak{g}|\mathfrak{f})[x] \,|\, \mathfrak{f}, \mathfrak{g} \in \mathcal{L}_\mathcal{A}, x \in [0,1]\}$ (Nilsson 1986; Lukasiewicz 1999). A *(conditional) knowledge base* $\mathcal{R} \subseteq (\mathcal{L}_\mathcal{A}|\mathcal{L}_\mathcal{A})$ is a set of conditionals.

Example 2.1. *Consider the alphabet* $(\{a,b\}, \{X,Y\}, \{F, A\, K\}, \{\wedge, \neg\})$. *$F(X)$ has the intended meaning 'X is famous', $A(X)$' expresses X is an actor' and $K(X,Y)$ expresses 'X knows Y'. Then $(F(X)\,|\,A(X))[0.7]$ is a conditional. It can be regarded as implicitly universally quantified. Intuitively it expresses a degree of belief. If we learn about someone, she is an actor, we suppose she is famous with about 70%.*

A *possible world* assigns a truth value to each (ground) atom over $\mathcal{A}$, or just to a subset that is obtained by grounding a conditional knowledge base. This set of ground atoms is denoted by $\mathcal{B}$ and is called the *interpretation base* of $\mathcal{A}$. Let $\Omega_\mathcal{B}$ denote the set of all possible worlds over $\mathcal{B}$. An atom $\mathfrak{a}$ is satisfied by $\omega \in \Omega_\mathcal{B}$ iff it is evaluated to true. The definition is extended to complex formulas in $\mathcal{L}_\mathcal{A}$ in the usual way. We denote the classical satisfaction relation between possible worlds and formulas in $\mathcal{L}_\mathcal{A}$ w.r.t. $\mathcal{B}$ by $\models_\mathcal{B}$. For a formula $\mathfrak{f} \in \mathcal{L}_\mathcal{A}$ let $\mathrm{Mod}(\mathfrak{f}) := \{\omega \in \Omega_\mathcal{B} \,|\, \omega \models_\mathcal{B} \mathfrak{f}\}$ denote the set of its classical models. Probabilistic semantics can be defined by considering probability distributions over possible worlds. Let $\mathcal{P} : \Omega_\mathcal{B} \to [0,1]$ be a probability distribution assigning a degree of belief to each possible world. $\mathcal{P}$ is extended to the power set $2^{\Omega_\mathcal{B}}$ via $\mathcal{P}(W) := \sum_{\omega \in W} \mathcal{P}(\omega)$ for all $W \subseteq \Omega_\mathcal{B}$. Let $\mathfrak{P}_\mathcal{B}$ denote the set of all such probability distributions over $\Omega_\mathcal{B}$.

A conditional semantics $\mathcal{S}$ defines which $\mathcal{P} \in \mathfrak{P}_\mathcal{B}$ satisfy a certain conditional. For propositional languages usually the definition of conditional probability is used. That is, for two propositional formulas $\mathfrak{g}, \mathfrak{f}$ we define $\mathcal{P} \models_\mathcal{S} (\mathfrak{g}|\mathfrak{f})[x]$ iff $\mathcal{P}(\mathrm{Mod}(\mathfrak{g}\mathfrak{f})) = x \cdot \mathcal{P}(\mathrm{Mod}(\mathfrak{f}))$ (e.g. (Paris 1994)). We call this semantics the *standard semantics*. For relational languages several other semantics have been considered. As a detailed discussion would go beyond the scope of this paper, we only sketch the ideas and refer to the literature for a more detailed discussion. *Grounding semantics* (Fisseler 2010) interprete conditionals by interpreting all their ground instances. Therefore a grounding operator $\mathrm{gr} : (\mathcal{L}_\mathcal{A}|\mathcal{L}_\mathcal{A}) \to 2^{(\mathcal{L}_\mathcal{A}|\mathcal{L}_\mathcal{A})}$ is introduced mapping conditionals to the set of its ground instances. The number of instances can be restricted by variable restrictions like $X \neq Y$.

Example 2.2. *In Example 2.1,* gr *maps* $(F(X)\,|\,A(X))[0.7]$ *to* $(F(a)\,|\,A(a))[0.7]$ *and* $(F(b)\,|\,A(b))[0.7]$. *In general, a distribution* $\mathcal{P}$ *satisfies a relational conditional* $(\mathfrak{g}|\mathfrak{f})[x]$ *under grounding semantics iff* $\mathcal{P}$ *satisfies all ground instances under standard semantics, i.e., iff* $\mathcal{P}(\mathrm{Mod}(\mathfrak{g}_{\mathrm{gr}}\mathfrak{f}_{\mathrm{gr}})) = x \cdot \mathcal{P}(\mathrm{Mod}(\mathfrak{f}_{\mathrm{gr}}))$ *for all* $(\mathfrak{g}_{\mathrm{gr}}\,|\,\mathfrak{f}_{\mathrm{gr}})[x] \in \mathrm{gr}((\mathfrak{g}\,|\,\mathfrak{f})[x])$.

Another interesting semantics for relational languages is the aggregating semantics (Kern-Isberner and Thimm 2010). Instead of regarding a conditional containing variables as a hard template for the probability of each ground instance, their conditional probabilities just have to 'aggregate' to the stated probability (Kern-Isberner and Thimm 2010).

In general, a conditional is satisfied by a probability distribution under a given semantics if a certain equation over probabilities of possible worlds is satisfied. Usually these equations can be transformed into a normal form $f_c(\mathcal{P}) = 0$. Let $(\mathcal{L}_\mathcal{A}|\mathcal{L}_\mathcal{A})$ be a conditional language over an alphabet $\mathcal{A}$ along with an interpretation base $\mathcal{B}$. We say a satisfaction relation $\models_\mathcal{S} \subseteq \mathfrak{P}_\mathcal{B} \times (\mathcal{L}_\mathcal{A}|\mathcal{L}_\mathcal{A})$ defines a *conditional semantics* $\mathcal{S}$ (with respect to $\mathcal{B}$) iff for each conditional $c \in (\mathcal{L}_\mathcal{A}|\mathcal{L}_\mathcal{A})$ there is a k_c-dimensional *constraint function* $f_c : \mathfrak{P}_\mathcal{B} \to \mathbb{R}^{k_c}$ such that for all $\mathcal{P} \in \mathfrak{P}_\mathcal{B}$, $c \in (\mathcal{L}_\mathcal{A}|\mathcal{L}_\mathcal{A})$ it holds $\mathcal{P} \models_\mathcal{S} c$ iff $f_c(\mathcal{P}) = 0$. By $f^{[i]}$ we denote the i-th component of a multi-dimensional function f. The dimension k_c of the image of the constraint function is usually 1, only for grounding semantics it can be greater. Then it corresponds to the number of ground instances of the conditional c. All introduced semantics use constraint functions with a similar structure.

Definition 2.3. $\mathcal{S}$ *is called* linearly structured *iff for each conditional* $c \in (\mathcal{L}_\mathcal{A}|\mathcal{L}_\mathcal{A})$, *there are* k_c-*dimensional functions* $\mathrm{V}_c : \Omega_\mathcal{B} \to \mathbb{N}_0^{k_c}$ *and* $\mathrm{F}_c : \Omega_\mathcal{B} \to \mathbb{N}_0^{k_c}$ *such that* $f_c(\mathcal{P}) = \sum_{\omega \in \Omega_\mathcal{B}} \mathcal{P}(\omega) \cdot (\mathrm{V}_c(\omega) \cdot (1-x) - \mathrm{F}_c(\omega) \cdot x)$.

The functions V_c and F_c basically indicate whether a world verifies or falsifies a conditional, see (Potyka 2012) for details. The introduced standard semantics and aggregating semantics are linearly structured (Potyka 2012). Grounding semantics basically map a relational conditional language to a propositional language over ground atoms and use the standard semantics. Therefore they are also linearly structured.

For a conditional $c \in (\mathcal{L}_\mathcal{A}|\mathcal{L}_\mathcal{A})$ let $\mathrm{Mod}_\mathcal{S}(c) := \{\mathcal{P} \in \mathfrak{P}_\mathcal{B} \,|\, f_c(\vec{\mathcal{P}}) = 0\}$ denote the set of its probabilistic models under a given conditional semantics $\mathcal{S}$. For a knowledge base $\mathcal{R} \subseteq (\mathcal{L}_\mathcal{A}|\mathcal{L}_\mathcal{A})$ let $\mathrm{Mod}_\mathcal{S}(\mathcal{R}) := \bigcap_{c \in \mathcal{R}} \mathrm{Mod}_\mathcal{S}(c)$. We are interested in the best probability distribution in $\mathrm{Mod}_\mathcal{S}(\mathcal{R})$. An appropriate selection criterion is the principle of maximum entropy (Paris 1994; Kern-Isberner 2001). One important computational problem in this framework is the task of computing an *ME-Model* of $\mathcal{R}$, i.e., a model $\mathcal{P} \in \mathrm{Mod}_\mathcal{S}(\mathcal{R})$ maximizing the entropy $H(\mathcal{P}) := -\sum_{\omega \in \Omega_\mathcal{B}} \mathcal{P}(\omega) \cdot \log \mathcal{P}(\omega)$. For linearly structured semantics there is always a unique solution if $\mathcal{R}$ is consistent.

3 Factorization of ME-Models

Let $d = 2^{|\mathcal{B}|}$. We assume the possible worlds are ordered in a fixed sequence $\omega_1, \ldots, \omega_d \in \Omega_\mathcal{B}$ so that we can identify each $\mathcal{P} \in \mathfrak{P}_\mathcal{B}$ with a point $(\mathcal{P}(\omega_i))_{1 \leq i \leq d} \in \mathbb{R}^d$. We use the usual terminology and operators defined for $\mathbb{R}^d$ in the following. For a knowledge base $\mathcal{R}$ let $\mathcal{R}^= := \{(\mathfrak{g}|\mathfrak{f})[x] \in \mathcal{R} \,|\, x \in \{0,1\}\}$ be the subset of deterministic conditionals and let $\mathcal{R}^\approx := \mathcal{R} \setminus \mathcal{R}^=$. $\mathcal{R}^=$ enforces zero probabilities for a subset of worlds $\mathcal{N}_\mathcal{R} \subseteq \Omega_\mathcal{B}$ (Potyka 2012).

We call $\mathcal{R}$ *regular* if there is a positive distribution over $\Omega_\mathcal{B} \setminus \mathcal{N}_\mathcal{R}$ satisfying the constraints expressed by $\mathcal{R}^\approx$ with respect to $\mathcal{S}$. For consistent non-regular knowledge bases no finite factors might exist. We remark that regularity can be checked by means of linear algebra.

Proposition 3.1. *Let* $\mathcal{R} \subseteq (\mathcal{L}_\mathcal{A}|\mathcal{L}_\mathcal{A})$ *be a regular knowledge base interpreted by a linearly structured semantics* $\mathcal{S}$ *with respect to an interpretation base* $\mathcal{B}$ *over* $\mathcal{A}$. *Let* $\mathcal{P}_{ME} \in \mathfrak{P}_\mathcal{B}$ *be the ME-model of* $\mathcal{R}$ *subject to* $\mathcal{S}$. *Let* $0^0 := 1$ *and for* $c \in \mathcal{R}$ *let* $\mathrm{h}_c^{[i]}(\omega) := \mathrm{V}_c^{[i]}(\omega) \cdot (1-x) - \mathrm{F}_c^{[i]}(\omega) \cdot x$. *Then for each* $c \in \mathcal{R}^\approx$ *with* k_c-*dimensional constraint function there are* k_c *positive numbers* $a_{c,i} \in \mathbb{R}_0^+, 1 \leq i \leq k_c$ *such that*

$$\mathcal{P}_{ME}(\omega) = \frac{1}{Z} \prod_{c \in \mathcal{R}^\approx} \prod_{1 \leq i \leq k_c} a_{c,i}^{\mathrm{h}_c^{[i]}(\omega)} \prod_{c \in \mathcal{R}^=} \prod_{1 \leq i \leq k_c} 0^{|\,\mathrm{h}_c^{[i]}(\omega)|}.$$

Proof. It holds for all $c = (\mathfrak{g}|\mathfrak{f})[x] \in \mathcal{R}^=$ and for all $\omega \in \Omega_\mathcal{B}$ that if $x = 0$ and $\mathrm{V}_c(\omega) \neq 0$ then $\mathcal{P}_{ME}(\omega) = 0$, see (Potyka 2012), Lemma 4.2. Symmetrically, if $x = 1$ and $\mathrm{F}_c(\omega) \neq 0$ then $\mathcal{P}_{ME}(\omega) = 0$. Let $\mathcal{N}_\mathcal{R} = \{\omega \in \Omega_\mathcal{B} \,|\, \exists\, c = (\mathfrak{g}|\mathfrak{f})[x] \in \mathcal{R}^= : \exists\, 1 \leq i \leq k_c : \mathrm{h}_c^{[i]}(\omega) \neq 0\}$ denote the set of worlds verifying a 0-conditional or falsifying a 1-conditional. Let $\mathcal{P}'$ be the ME-optimal distribution over $\Omega_\mathcal{B}^+ := \Omega_\mathcal{B} \setminus \mathcal{N}_\mathcal{R}$ satisfying $\mathcal{R}^\approx$. Then for all $\omega \in \Omega_\mathcal{B}^+$ it holds $\mathcal{P}_{ME}(\omega) = \mathcal{P}'(\omega)$ and for all $\omega \in \mathcal{N}_\mathcal{R}$ it holds $\mathcal{P}_{ME}(\omega) = 0$, see (Potyka 2012), Proposition 4.3.

As $\mathcal{R}$ is regular, there is a positive distribution over $\Omega_{\mathcal{B}}^+$ satisfying $\mathcal{R}^{\approx}$. Since $\mathcal{S}$ is linearly structured, it induces linear constraints and therefore the Open-mindedness Principle (Paris 1994) implies that the ME-optimal distribution $\mathcal{P}'$ over $\Omega_{\mathcal{B}}^+$ is positive and therefore a local maximum in the interior of $\mathfrak{P}_{\mathcal{B}}$. If the (constant) gradients $\{\nabla f_c^{[i]} \mid c \in \mathcal{R}, 1 \leq i \leq k_c\}$ of the constraint functions are linearly independent, the lagrange multiplier rule (McShane 1973) guarantees the existence of lagrange multipliers. If they are not independent, there are multipliers for an independent subset and the multipliers for the remaining constraint functions can be set to zero. We introduce a multiplier λ_0 corresponding to the normalizing constraint $\sum_{\omega \in \Omega_{\mathcal{B}}^+} \mathcal{P}'(\omega) - 1 = 0$ and multipliers $\lambda_{c,i}$ for the constraint components f_c^i. Computing the partial derivative of $H(\mathcal{P}')$ w.r.t. $\mathcal{P}'(\omega)$ we obtain $-\log \mathcal{P}'(\omega) - 1 = \lambda_0 + \sum\limits_{c \in \mathcal{R}^{\approx}} \sum\limits_{1 \leq i \leq k_c} \lambda_{c,i} \cdot \mathrm{h}_c^{[i]}(\omega)$.

Reordering the terms and applying the exponential function to both sides we obtain with $Z := \exp(1 + \lambda_0)$

$$\mathcal{P}'(\omega) = \frac{1}{Z} \prod_{c \in \mathcal{R}^{\approx}} \prod_{1 \leq i \leq k_c} \exp(-\lambda_{c,i} \cdot \mathrm{h}_c^{[i]}(\omega)). \qquad (1)$$

As explained above, we can construct the ME-model from $\mathcal{P}'$ by defining $\mathcal{P}(\omega) := \mathcal{P}'(\omega)$ for $\omega \in \Omega_{\mathcal{B}}^+$ and $\mathcal{P}(\omega) := 0$ for $\omega \in \mathcal{N}_{\mathcal{R}}$. We extend the factorization in equation (1) appropriately. Consider $N(\omega) := \prod_{c \in \mathcal{R}^{=}} \prod_{1 \leq i \leq k_c} 0^{|\,\mathrm{h}_c^{[i]}(\omega)|}$. If $\omega \in \mathcal{N}_{\mathcal{R}}$ then, by definition of $\mathcal{N}_{\mathcal{R}}$, there is a $c \in \mathcal{R}^{=}$ such that $\mathrm{h}_c^{[i]}(\omega) \neq 0$ for some $1 \leq i \leq k_c$. Therefore the exponent of a 0-factor is greater than zero and $N(\omega) = 0$. Otherwise, if $\omega \in \Omega_{\mathcal{B}} \setminus \mathcal{N}_{\mathcal{R}}$, then always $\mathrm{h}_c^{[i]}(\omega) = 0$ holds. Hence $N(\omega) = \prod_{c \in \mathcal{R}^{=}} \prod_{1 \leq i \leq k_c} 0^0 = 1$ by definition. Hence we can extend the factorization of the reduced solution in equation (1) to the complete solution over $\Omega_{\mathcal{B}}$ by multiplying the factor $N(\omega)$. For $c \in \mathcal{R}^{\approx}$ with k_c-dimensional constraint function we define $a_{c,i} := \exp(-\lambda_{c,i})$ for $1 \leq i \leq k_c$. Then it holds $\mathcal{P}_{ME}(\omega) = \mathcal{P}'(\omega) \cdot N(\omega) = \frac{1}{Z} \prod_{c \in \mathcal{R}^{\approx}} \prod_{1 \leq i \leq k_c} a_{c,i}^{\mathrm{h}_c^{[i]}(\omega)} \prod_{c \in \mathcal{R}^{=}} \prod_{1 \leq i \leq k_c} 0^{|\,\mathrm{h}_c^{[i]}(\omega)|}$.

As $\exp(x) > 0$ for all $x \in \mathbb{R}$ all factors $a_{c,i}$ are positive. $\square$

For $\mathfrak{f} \in \mathcal{L}_{\mathcal{A}}$ let $1_{\{\mathfrak{f}\}} : \Omega_{\mathcal{B}} \to \{0, 1\}$ denote the indicator function that maps a world ω to 1 if $\omega \models_{\mathcal{B}} \mathfrak{f}$, and to 0 otherwise. By inserting the specific values for $\mathrm{V}_c^{[i]}$ and $\mathrm{F}_c^{[i]}$ from (Potyka 2012) for the introduced linearly structured semantics, we obtain the positive factorizations shown in Table 1. Thereby $c = (\mathfrak{g}|\mathfrak{f})[x]$ is supposed to be a conditional in the corresponding language and $\omega \in \Omega_{\mathcal{B}}$ a classical interpretation of this language. For aggregating semantics the factor a_c is indeed independent of the ground instance index i, because just a single constraint function is induced for all ground instances (Potyka 2012).

Standard semantics
$\frac{1}{Z} \prod\limits_{(\mathfrak{g}\|\mathfrak{f})[x] \in \mathcal{R}^{\approx}} a_c^{1_{\{\mathfrak{g}\mathfrak{f}\}}(\omega) \cdot (1-x) - 1_{\{\overline{\mathfrak{g}}\mathfrak{f}\}}(\omega) \cdot x}$
Grounding semantics
$\frac{1}{Z} \prod\limits_{c \in \mathcal{R}^{\approx}} \prod\limits_{(\mathfrak{g}_i \,\mid\, \mathfrak{f}_i)[x] \in \mathrm{gr}(c)} a_{c,i}^{1_{\{\mathfrak{g}_i\mathfrak{f}_i\}}(\omega) \cdot (1-x) - 1_{\{\overline{\mathfrak{g}_i}\mathfrak{f}_i\}}(\omega) \cdot x}$
Aggregating semantics
$\frac{1}{Z} \prod\limits_{c \in \mathcal{R}^{\approx}} \prod\limits_{(\mathfrak{g}_i \,\mid\, \mathfrak{f}_i)[x] \in \mathrm{gr}(c)} a_c^{1_{\{\mathfrak{g}_i\mathfrak{f}_i\}}(\omega) \cdot (1-x) - 1_{\{\overline{\mathfrak{g}_i}\mathfrak{f}_i\}}(\omega) \cdot x}$

Table 1: Factorization of linearly structured semantics.

Application to Iterative Scaling Algorithms

Given some linear constraints, Iterative Scaling algorithms (Darroch and Ratcliff 1972) compute the ME-optimal distribution over a set of elementary events Ω, given that the solution is positive. In our domain, we use the notation $\mathrm{IS}(\Omega, \mathcal{R}, \mathcal{S}) = \mathcal{P}^*$, to express that the Iterative Scaling algorithm IS computes the ME-optimal probability distribution $\mathcal{P}^*$ over a set of worlds Ω satisfying a knowledge base $\mathcal{R}$ interpreted by a linearly structured semantics $\mathcal{S}$. Usually IS introduces a factor for each (ground) conditional from $\mathcal{R}$ and initializes it with 1. Then the factors are successively 'scaled' until all constraints are satisfied. If the constraints are consistent and $\mathcal{P}^*$ is positive, IS will converge towards $\mathcal{P}^*$. We can extend such algorithms to include deterministic conditionals using the following scheme:

1. Compute $\Omega_{\mathcal{B}}^+$.
2. $P' := \mathrm{IS}(\Omega_{\mathcal{B}}^+, \mathcal{R}^{\approx}, \mathcal{S})$.
3. Define $\mathcal{P}(\omega) := P'(\omega)$ for $\omega \in \Omega_{\mathcal{B}}^+$ and $\mathcal{P}(\omega) := 0$ for $\omega \in \mathcal{N}_{\mathcal{R}}$.

Corollary 3.2. *Given that $\mathcal{R}$ is regular, the procedure above computes the ME-model $\mathcal{P}_{ME}$ of $\mathcal{R}$, i.e., $\mathcal{P} = \mathcal{P}_{ME}$.*

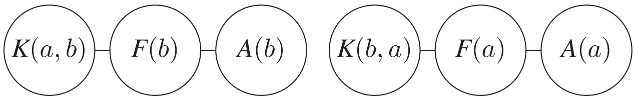
Proof. As argued before, since $\mathcal{R}$ is regular, the solution over $\Omega_{\mathcal{B}}^+$ with respect to $\mathcal{R}^{\approx}$ and $\mathcal{S}$ is positive and therefore can be computed by IS. $\mathcal{P} = \mathcal{P}_{ME}$ follows just like in Proposition 3.1. $\square$

Connection to Markov Random Fields

In this section we presuppose some familiarity with Markov Random Fields (MRFs) (Koller and Friedman 2009). Basically an MRF is a joint distribution P over a set of random variables $X = \{X_1, \cdots, X_n\}$ that factorizes as follows: $P_\Phi(X) = \frac{1}{Z} \prod_{\phi \in \Phi} \phi(X|_{D_\phi})$, where each factor $\phi : D_\phi \to \mathbb{R}, D_\phi \subseteq X$, depends only on a subset D_ϕ of X and $X|_{D_\phi}$ denotes the restriction of X to D_ϕ.

Note that the classical logical interpretation of a formula depends only on the interpretation of a subset of atoms in $\mathcal{B}$. For a ground formula $\mathfrak{f}$ let $\mathrm{scope}(\mathfrak{f})$ be the set of ground atoms contained in $\mathfrak{f}$. For a conditional c let $\mathrm{scope}(c) := \bigcup_{(\mathfrak{g}_i \,\mid\, \mathfrak{f}_i)[x] \in \mathrm{gr}(c)} \{\mathrm{scope}(\mathfrak{g}_i\mathfrak{f}_i)\}$, where $\mathrm{gr}(c)$ maps a propositional or ground conditionals c to $\{c\}$. Note that $\mathrm{scope}(c)$ is indeed a set of scopes, each set corresponding to a ground instance of c. Correspondingly, for a knowledge base $\mathcal{R}$ let $\mathrm{scope}(\mathcal{R}) := \bigcup_{c \in \mathcal{R}} \mathrm{scope}(c)$. Given an interpretation base $\mathcal{B}$ and a knowledge base $\mathcal{R}$ over an alphabet $\mathcal{A}$, the *induced graph* $\mathcal{G}_{\mathcal{B},\mathcal{R}}$ is the undirected graph, whose nodes correspond to the ground atoms in $\mathcal{B}$, and that contains an edge between $\mathfrak{a}_1, \mathfrak{a}_2 \in \mathcal{B}$ if and only if there is a set $S \in \mathrm{scope}(\mathcal{R})$ such that $\mathfrak{a}_1, \mathfrak{a}_2 \in S$, i.e., $\mathfrak{a}_1$ and $\mathfrak{a}_2$ both appear in the scope of a ground conditional.

Example 3.3. *We continue Example 2.1. Consider* $\mathcal{R} = \{(F(X) \,|\, A(X))[0.7], (K(Y, X) \,|\, F(X))[0.9]\}$ *and let the second conditional respect the instantiation restriction* $X \neq Y$. *We obtain* $\mathcal{B} = \{F(a), F(b), A(a), A(b), K(a, b), K(b, a)\}$ *by grounding* $\mathcal{R}$. *The following figure shows* $\mathcal{G}_{\mathcal{B},\mathcal{R}}$*:*



As Table 1 shows, for the introduced linearly structured semantics the factorization of $\mathcal{P}_{ME}$ contains one factor for each ground instance of a conditional in $\mathcal{R}$. In particular, the value of each of these factors relies only on the scope of the corresponding ground conditional. If we regard $\mathcal{P}_{ME}$ as a joint distribution over random variables $\mathcal{B}$, it is just an MRF. For the introduced semantics it has the form $\mathcal{P}_{ME}(\mathcal{B}) = \frac{1}{Z} \prod_{c \in \mathcal{R}^{\approx}} \prod_{(\mathfrak{g}_i \,|\, \mathfrak{f}_i)[x] \in \mathrm{gr}(c)} \phi_{c,i}(\mathrm{scope}(\mathfrak{g}_i \mathfrak{f}_i))$, where (letting $a_{c,i} := a_c$ for standard and aggregating semantics) $\phi_{c,i}(S) := a_{c,i}^{1_{\{\mathfrak{g}_i \mathfrak{f}_i\}}(S) \cdot (1-x) - 1_{\{\overline{\mathfrak{g}_i} \mathfrak{f}_i\}}(S) \cdot x}$. The induced graph is constructed by connecting each two ground atoms that appear together in the scope of a factor $\phi(\mathrm{scope}(\mathfrak{g}_i \mathfrak{f}_i))$. One can show that in this way the minimal Markov network for $\mathcal{P}_{ME}(\mathcal{B})$ is constructed, see (Koller and Friedman 2009), Prop. 9.1. Combining both findings, we obtain the following results.

Corollary 3.4. *Let* $\mathcal{A}$ *be an alphabet containing arbitrary logical connectives but no quantifiers. Let* $\mathcal{R} \subseteq (\mathcal{L}_{\mathcal{A}} | \mathcal{L}_{\mathcal{A}})$ *be a regular knowledge base interpreted by grounding or aggregating semantics with respect to an interpretation base* $\mathfrak{B}$ *over* $\mathcal{A}$. *Then the ME-model* $\mathcal{P}_{ME}$ *is an MRF and the induced graph* $\mathcal{G}_{\mathcal{B},\mathcal{R}}$ *is the minimal Markov network to* $\mathcal{P}_{ME}$*.*

Example 3.5. *We continue Example 3.3.* $\mathcal{R}$ *is indeed regular. Therefore, using MRF notation, the ME-model can be written as* $\mathcal{P}(A(a), A(b), F(a), F(b), K(a, b), K(b, a))$. *We assign a factor* α_1 *to* $(F(a) \,|\, A(a))[0.7]$*,* α_2 *to* $(F(b) \,|\, A(b))[0.7]$*,* β_1 *to* $(K(b, a) \,|\, F(a))[0.9]$ *and* β_2 *to* $(K(a, b) \,|\, F(b))[0.9]$. *Then, for instance,* $\mathcal{P}(1, 1, 1, 0, 0, 1) = \alpha_1^{0.3} \cdot \alpha_2^{-0.7} \cdot \beta_1^{0.1} \cdot \beta_2^{0}$.

Corollary 3.6. *Let* $\mathcal{A}$ *be a propositional alphabet. Let* $\mathcal{R} \subseteq (\mathcal{L}_{\mathcal{A}} | \mathcal{L}_{\mathcal{A}})$ *be a regular knowledge base interpreted by standard semantics with respect to an interpretation base* $\mathfrak{B}$ *over* $\mathcal{A}$. *Then the ME-model* $\mathcal{P}_{ME}$ *is an MRF and the induced graph* $\mathcal{G}_{\mathcal{B},\mathcal{R}}$ *is the minimal Markov network to* $\mathcal{P}_{ME}$*.*

Grounding and aggregating semantics are applied to quantifier-free languages only. However, in principle one can consider a quantified first-order language under standard semantics. Then Proposition 3.1 still guarantees that the ME-model of regular knowledge bases is an MRF. But the corresponding minimal Markov network will be much more complex, because for quantified formulas there is not one scope for the interpretation of each ground instance, but a single big scope containing the scopes of all ground instances.

4 Discussion

We revisited the factorization of ME-models and the connection to Markov Random Fields. Proposition 3.1 guarantees the existence of the factorization for regular knowledge bases under arbitrary linearly structured semantics. In particular, as we saw in Corollary 3.2, the ME-model of regular knowledge bases can be computed by Iterative Scaling variants even if deterministic conditionals are contained. The restriction to regular knowledge bases is not a heavy drawback, as Examples for consistent knowledge bases that are non-regular are usually pathological and do not appear in practice.

As noted in (Fisseler 2010), the factorization of ME-models establishes a connection to Markov Random Fields. We showed how the corresponding minimal Markov network can be constructed for the ME-model of regular knowledge bases for different languages and semantics. This network can be used to generate a junction tree or, more generally, a cluster graph that enables more efficient inference techniques (Koller and Friedman 2009). A junction tree representation, which has been derived in another way, has already been used profitably for propositional languages under standard semantics (Rödder, Reucher, and Kulmann 2006).

References

Darroch, J. N., and Ratcliff, D. 1972. Generalized Iterative Scaling for Log-Linear Models. *The Annals of Mathematical Statistics* 43(5):1470–1480.

Finthammer, M., and Beierle, C. 2012. Using equivalences of worlds for aggregation semantics of relational conditionals. In *German Conference on AI (KI)*, LNAI. Springer.

Fisseler, F. 2010. *Learning and Modeling with Probabilistic Conditional Logic*, volume 328 of *Dissertations in Artificial Intelligence*. Amsterdam: IOS Press.

Kern-Isberner, G., and Thimm, M. 2010. Novel semantical approaches to relational probabilistic conditionals. In *KR*, 382–392. AAAI Press.

Kern-Isberner, G. 2001. *Conditionals in nonmonotonic reasoning and belief revision*. LNAI. Springer.

Koller, D., and Friedman, N. 2009. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.

Lukasiewicz, T. 1999. Probabilistic deduction with conditional constraints over basic events. *J. Artif. Intell. Res* 10:380–391.

McShane, E. J. 1973. The lagrange multiplier rule. *The American Mathematical Monthly* 80(8):pp. 922–925.

Nilsson, N. J. 1986. Probabilistic logic. *Artif. Intell.* 28:71–88.

Paris, J. 1994. *The uncertain reasoner's companion – A mathematical perspective*. Cambridge University Press.

Potyka, N. 2012. Towards a general framework for maximum entropy reasoning. In *FLAIRS*. AAAI Press, Menlo Park, California.

Rödder, W.; Reucher, E.; and Kulmann, F. 2006. Features of the expert-system-shell spirit. *Logic Journal of the IGPL* 14(3):483–500.