A Two-Level Approach to Maximum Entropy Model Computation for Relational Probabilistic Logic Based on Weighted Conditional Impacts

Marc Finthammer and Christoph Beierle

Dept. of Computer Science, University of Hagen, Germany

Abstract. The principle of maximum entropy allows to define the semantics of a knowledge base consisting of a set of probabilistic relational conditionals by a unique model having maximum entropy. Using the concept of a conditional structure of a world, we define the notion of weighted conditional impacts and present a two-level approach for maximum entropy model computation based on them. Once the weighted conditional impact of a knowledge base has been determined, a generalized iterative scaling algorithm is used that fully abstracts from concrete worlds. The weighted conditional impact may be reused when only the quantitative aspects of the knowledge base are changed. As a further extension of previous work, also deterministic conditionals may be present in the knowledge base, and a special treatment of such conditionals reduces the problem size.

1 Introduction

When enriching propositional logic with probabilities for modeling uncertainty (e.g. [15,18,4,9]), can play a vital role. Relational probabilistic conditionals are useful for modeling uncertain knowledge in scenarios where relations among individual objects are important. For instance, given a set of connected personal computers, stating that the probability that a malware infected PC sends a message to another PC is 0.7 while for a non-infected PC it is only 0.1, could be formally denoted by the conditionals $(\mathit{sendsMail}(X,Y)|\mathit{infected}(X))[0.7]$ and $(\mathit{sendsMail}(X,Y)|\neg\mathit{infected}(X))[0.1]$. Having a knowledge base $\mathcal{R}$ consisting of a set of such conditionals, there may be many different probability distributions satisfying them. The idea of the *principle of maximum entropy (ME)* [20,17,10,11] is to select among all models the model adding as little information as possible and thus being the most unbiased one. Recently, different approaches to applying the ME principle not only to the propositional case, but also in a relational first-order setting have been proposed [13,2]. In these approaches, ME reasoning amounts to compute the probability of a formula F under the ME model of $\mathcal{R}$, and determining the ME model of a knowledge base is the most crucial step for reasoning under ME semantics.

In this paper, we present AggME, a system that implements the ME model computation for probabilistic relational conditionals under *aggregating ME semantics* [13] which requires solving a complex optimization problem. In [5], a

U. Straccia and A. Calì (Eds.): SUM 2014, LNAI 8720, pp. 162–175, 2014.

generalized iterative scaling (GIS) algorithm is proposed for this task. The approach implemented in AggME refines and extends the proposal of [5] in several directions. While in [5] only non-deterministic conditionals are allowed, AggME also treats deterministic conditionals having probability 0 or 1 which is required in many application scenarios. While [5] uses conditional structures introduced by Kern-Isberner [10] for defining equivalences of worlds, AggME extends the use of conditional structures and introduces a two-phase ME computation. For the first phase, an algorithm WCI is developed for computing what we call the *weighted conditional impact of* $\mathcal{R}$; this algorithm is based solely on the qualitative parts of the conditionals in $\mathcal{R}$. The second phase employs a GIS algorithm $\text{GIS}_{\odot}^{\gamma_{\mathcal{R}}}$ that fully abstracts from worlds by just using the weighted conditional impact and the probabilities given in the conditionals in $\mathcal{R}$. The modular design of AggME allows for an easy exchange of alternative computation methods for both phases. It also supports the reuse of the weighted conditional impact of $\mathcal{R}$ for a modified knowledge base $\mathcal{R}'$ obtained from $\mathcal{R}$ by just changing the probabilities of the conditionals, a situation that is quite common when developing a knowledge base. AggME is implemented in Java and is available as a plugin for KReator[1] [6], an integrated development environment for relational probabilistic logic.

After briefly recalling the basics of aggregating semantics (Sec. 2), Sec. 3 addresses the treatment of deterministic conditionals under ME semantics. In Sec. 4, weighted conditional impacts are defined and illustrated, leading to an alternative formulation of the ME optimization problem solved by the AggME algorithms presented in Sec. 5. Some examples and first evaluation results are given in Sec. 6, and in Section 7 we conclude and point out further work.

2 Background

We consider a quantifier-free first-order language $\mathcal{L}$ over a set of predicates *Pred* and a finite set of constants *Const*. For formulas $A, B \in \mathcal{L}$, AB abbreviates the conjunction $A \wedge B$, and $\text{gnd}(A)$ denotes the set of ground instances of A. By introducing the operator $|$, we obtain the language $(\mathcal{L}|\mathcal{L})^{prob}$ of *probabilistic conditionals* of the form $(B(\boldsymbol{X})|A(\boldsymbol{X}))[d]$ with $\boldsymbol{X}$ containing the variables of the formulas A and B, and where $d \in [0,1]$ is a probability; $(B(\boldsymbol{X})|\top))[d]$ is a *probabilistic fact*. The conditional is *deterministic* iff $d = 0$ or $d = 1$; otherwise, it is *non-deterministic*. A finite set $\mathcal{R} \subseteq (\mathcal{L}|\mathcal{L})^{prob}$ is called a *knowledge base*. We always implicitly consider $\mathcal{R}$ together with the respective sets *Pred* and *Const*.

$\mathcal{H}$ denotes the *Herbrand base*, i.e. the set containing all ground atoms over *Pred* and *Const*, and $\Omega = \mathfrak{P}(\mathcal{H})$ is the set of all possible worlds (i. e. *Herbrand interpretations*), where $\mathfrak{P}$ is the power set operator. The *probabilistic interpretations* for $(\mathcal{L}|\mathcal{L})^{prob}$ are given by the set $\mathcal{P}_\Omega$ of all probability distributions $P : \Omega \to [0,1]$ over possible worlds. P is extended to ground formulas $A(\boldsymbol{a})$, with $A(\boldsymbol{a}) \in \text{gnd}(A(X))$, by defining $P(A(\boldsymbol{a})) := \sum_{\omega \models A(\boldsymbol{a})} P(\omega)$. The *aggregation semantics* [13] extends P to conditionals and resembles the definition of a

[1] KReator and AggME can be found at `http://kreator-ide.sourceforge.net/`

conditional probability by summing up the probabilities of all respective ground formulas; it defines the satisfaction relation $\models_{\odot}$ for $r = (B(\boldsymbol{X})|A(\boldsymbol{X}))[d]$ by

$$P \models_{\odot} r \quad \text{iff} \quad \frac{\displaystyle\sum_{(B(\boldsymbol{a})|A(\boldsymbol{a}))\in \text{gnd}(B(\boldsymbol{X})|A(\boldsymbol{X}))} P\left(A(\boldsymbol{a})B(\boldsymbol{a})\right)}{\displaystyle\sum_{(B(\boldsymbol{a})|A(\boldsymbol{a}))\in \text{gnd}(B(\boldsymbol{X})|A(\boldsymbol{X}))} P\left(A(\boldsymbol{a})\right)} = d \tag{1}$$

Where $\sum_{(B(\boldsymbol{a})|A(\boldsymbol{a}))\in \text{gnd}(B(\boldsymbol{X})|A(\boldsymbol{X}))} P\left(A(\boldsymbol{a})\right) > 0$. If $P \models_{\odot} r$ holds, we say that *P satisfies r* or *P is a model of r*. *P satisfies* a set of conditionals $\mathcal{R}$ if it satisfies every element of $\mathcal{R}$, and $\mathit{Mod}(\mathcal{R}) := \{P \in \mathcal{P}_{\Omega} \mid P \models_{\odot} \mathcal{R}\}$. $\mathcal{R}$ is *consistent* iff $\mathit{Mod}(\mathcal{R}) \neq \emptyset$. The *entropy* $H(P) := -\sum_{\omega\in\Omega} P(\omega)\log P(\omega)$ of a probability distribution P measures the indifference within P. The principle of *maximum entropy* (*ME*) chooses the distribution P where $H(P)$ is maximal among all distributions satisfying $\mathcal{R}$ [17,10]. The ME model $P^*_{\mathcal{R}}$ for $\mathcal{R}$ based on aggregation semantics is uniquely defined [13] by the solution of the convex optimization problem

$$P^*_{\mathcal{R}} := \arg \max_{P\in\mathcal{P}_{\Omega}:P\models_{\odot}\mathcal{R}} H(P) \tag{2}$$

3 Null-Worlds and Maximum Entropy

For illustrating knowledge bases with relational probabilistic conditionals and as a running example, we consider the following scenario:

Example 1 (Antivirus, $\mathcal{R}_{\text{vir}}$). Suppose we want to model some knowledge about virus infected computers (cf. Sec. 1): If an infected computer sends mail to another computer without antivirus protection, the other computer is likely to get infected (with probability 0.9). Computers with antivirus on very rarely get infected (probability 0.01). Infected computers are likely to send email to any computer (0.7), while uninfected computers do this only with probability 0.1. Moreover, we know that in our scenario to be modeled, computers do not send email to themselves. The following knowledge base $\mathcal{R}_{\text{vir}}$ represents this:

$r_1 : (\mathit{infected}(Y)|\mathit{sendsMail}(X,Y) \wedge \mathit{infected}(X) \wedge \neg \mathit{antiVirOn}(Y))[0.9]$
$r_2 : (\mathit{infected}(X)|\mathit{antiVirOn}(X))[0.01]$
$r_3 : (\mathit{sendsMail}(X,Y)|\mathit{infected}(X))[0.7]$
$r_4 : (\mathit{sendsMail}(X,Y)|\neg\mathit{infected}(X))[0.1]$
$r_5 : (\mathit{sendsMail}(X,X)|\top)[0.0]$

Note that r_5 is a deterministic conditional, and the presence of the deterministic conditionals prohibits applying the GIS algorithm approach of [5] directly to $\mathcal{R}_{\text{vir}}$. In the following, we will show how the restriction to nondeterministic conditionals required in [5] can be removed. For the rest of this paper, we assume

$$\mathcal{R} := \mathcal{R}^{\approx} \cup \mathcal{R}^{=}, \quad \mathcal{R}^{\approx} := \underbrace{\{r_1,\ldots,r_m\}}_{m \text{ non-deterministic}}, \quad \mathcal{R}^{=} := \underbrace{\{r_{m+1},\ldots,r_{m+M}\}}_{M \text{ deterministic}} \tag{3}$$

Where $\mathcal{R}$ is a consistent set consisting of m non-deterministic and M deterministic conditionals. Furthermore, let $\mathcal{R}^{=0} := \{r_i \in \mathcal{R}^{=} \mid d_i = 0\}$ and $\mathcal{R}^{=1} := \{r_i \in \mathcal{R}^{=} \mid d_i = 1\}$ denote the set of deterministic conditionals with probability 0 and 1, respectively.

For a relational conditional $r_i = (B_i(\boldsymbol{X})|A_i(\boldsymbol{X}))[d_i]$, the *counting functions* (cf. [12] and also [5]) $ver_i, fal_i : \Omega \to \mathbb{N}_0$ are given by:

$$ver_i(\omega) := \left|\left\{(B_i(\boldsymbol{a})|A_i(\boldsymbol{a})) \in \mathrm{gnd}(B_i(\boldsymbol{X})|A_i(\boldsymbol{X})) \mid \omega \models A_i(\boldsymbol{a})B_i(\boldsymbol{a})\right\}\right| \quad (4)$$

$$fal_i(\omega) := \left|\left\{(B_i(\boldsymbol{a})|A_i(\boldsymbol{a})) \in \mathrm{gnd}(B_i(\boldsymbol{X})|A_i(\boldsymbol{X})) \mid \omega \models A_i(\boldsymbol{a})\neg B_i(\boldsymbol{a})\right\}\right| \quad (5)$$

For a world $\omega \in \Omega$, $ver_i(\omega)$ yields the number of ground instances of the qualitative part of r_i which are *verified* by ω; and analogously, $fal_i(\omega)$ yields the number of ground instances of the qualitative part of r_i which are *falsified* by ω. In the following, when talking about a conditional, we will not distinguish explicitly the qualitative part of a conditional and the conditional and we may just drop the probability if the context is clear.

Example 2. Consider the five conditionals of $\mathcal{R}_{\text{vir}}$ from Example 1 together with the set of constants $Const = \{a, b, c\}$. Then each of the conditionals r_1, r_3, and r_4 has nine ground instances and both r_2 and r_5 have three ground instances. When abbreviating *infected* by *in*, *antiVirOn* by *an*, and *sendsMail* by *se*, these ground instances are:

$r_{1,1}: (in(a)|se(a,a) \wedge in(a) \wedge \neg an(a))$ $r_{3,1}: (se(a,a)|in(a))$ $r_{4,1}: (se(a,a)|\neg in(a))$
$r_{1,2}: (in(a)|se(b,a) \wedge in(b) \wedge \neg an(a))$ $r_{3,2}: (se(a,b)|in(a))$ $r_{4,2}: (se(a,b)|\neg in(a))$
$r_{1,3}: (in(a)|se(c,a) \wedge in(c) \wedge \neg an(a))$ $r_{3,3}: (se(a,c)|in(a))$ $r_{4,3}: (se(a,c)|\neg in(a))$
$r_{1,4}: (in(b)|se(a,b) \wedge in(a) \wedge \neg an(b))$ $r_{3,4}: (se(b,a)|in(b))$ $r_{4,4}: (se(b,a)|\neg in(b))$
$r_{1,5}: (in(b)|se(b,b) \wedge in(b) \wedge \neg an(b))$ $r_{3,5}: (se(b,b)|in(b))$ $r_{4,5}: (se(b,b)|\neg in(b))$
$r_{1,6}: (in(b)|se(c,b) \wedge in(c) \wedge \neg an(b))$ $r_{3,6}: (se(b,c)|in(b))$ $r_{4,6}: (se(b,c)|\neg in(b))$
$r_{1,7}: (in(c)|se(a,c) \wedge in(a) \wedge \neg an(c))$ $r_{3,7}: (se(c,a)|in(c))$ $r_{4,7}: (se(c,a)|\neg in(c))$
$r_{1,8}: (in(c)|se(b,c) \wedge in(b) \wedge \neg an(c))$ $r_{3,8}: (se(c,b)|in(c))$ $r_{4,8}: (se(c,b)|\neg in(c))$
$r_{1,9}: (in(c)|se(c,c) \wedge in(c) \wedge \neg an(c))$ $r_{3,9}: (se(c,c)|in(c))$ $r_{4,9}: (se(c,c)|\neg in(c))$

$r_{2,1}: (in(a)|an(a))$ $r_{5,1}: (se(a,a)|\top)$
$r_{2,2}: (in(b)|an(b))$ $r_{5,2}: (se(b,b)|\top)$
$r_{2,3}: (in(c)|an(c))$ $r_{5,3}: (se(c,c)|\top)$

For the world $\omega' = \{in(a), in(b), an(c), se(a,c), se(b,c)\}$, we have

$$\omega' \models in(a) \wedge se(a,c) \quad \text{and} \quad \omega' \models in(a) \wedge \neg se(a,b)$$

Since $in(a) \in \omega'$, $se(a,c) \in \omega'$, and $se(a,b) \notin \omega'$. Thus, ω' verifies the ground instance $r_{3,3}: (se(a,c)|in(a))$, and ω' falsifies the ground instance $r_{3,2}: (se(a,b)|in(a))$. Overall, ω' verifies two and falsifies four ground instances of r_3, i. e. $ver_3(\omega') = 2$ and $fal_3(\omega') = 4$ holds.

For characterizing the behavior of $P^*_{\mathcal{R}}$ on $\mathcal{R}^{\approx}$ and $\mathcal{R}^{=}$, we employ the counting functions (4) and (5).

Definition 1 (Null-Worlds and Potentially Positive Worlds). *The set*

$$\Omega_{\text{null}(\mathcal{R})} := \left\{\omega \in \Omega \mid \left(\exists r_i \in \mathcal{R}^{=0} : ver_i(\omega) > 0\right) \ \vee \ \left(\exists r_i \in \mathcal{R}^{=1} : fal_i(\omega) > 0\right)\right\}$$

is called the set of null-worlds with respect to $\mathcal{R}$. *The* set of potentially positive worlds with respect to $\mathcal{R}$ *is given by* $\Omega_{\text{pos}(\mathcal{R})} := \Omega \setminus \Omega_{\text{null}(\mathcal{R})}$.

Extending Paris' *open-mindedness principle* [17], we can show:

Proposition 1 (Null-Worlds and ME). *If* $P \in Mod(\mathcal{R})$*, then* $P(\omega) = 0$ *for all* $\omega \in \Omega_{\text{null}(\mathcal{R})}$*, and* $P^*_{\mathcal{R}}(\omega) > 0$ *for all* $\omega \in \Omega_{\text{pos}(\mathcal{R})}$*.*

Thus, for the ME model $P^*_{\mathcal{R}}$ and for any null-world ω, $P^*_{\mathcal{R}}(\omega) = 0$ holds, but for every potentially positive world, $P^*_{\mathcal{R}}$ yields a non-zero probability, i. e. the worlds in $\Omega_{\text{pos}(\mathcal{R})}$ are indeed positive under $P^*_{\mathcal{R}}$.

Example 3. For $\mathcal{R}_{\text{vir}}$ from Example 2 the world

$$\omega'_0 = \{an(b), in(c), se(a,a), se(b,c)\}$$

Is a null-world, because ω'_0 verifies the ground instance $r_{5,1}$: $(se(a,a)|\top)$ of the deterministic conditional r_5, i. e. $ver_5(\omega'_0) > 0$ holds. So every world containing a ground atom $se(a,a)$, $se(b,b)$, or $se(c,c)$ is a null-world due to r_5. In fact, $28,672$ $(= 7 \cdot 2^{12})$ of the $32,768$ $(= 2^{15})$ worlds contained in Ω are null-worlds, i. e. there are merely $4,096$ $(= 2^{12})$ potentially positive worlds.

4 Weighted Conditional Impact

For propositional conditionals, the satisfaction relation can be expressed by using feature functions (e. g. [7]). For a relational conditional $r_i = (B_i(\boldsymbol{X})|A_i(\boldsymbol{X}))[d_i]$, the *feature function* $f_i : \Omega \to \mathbb{R}$ with

$$f_i(\omega) := ver_i(\omega)(1 - d_i) - fal_i(\omega)d_i \tag{6}$$

given in [5] uses the counting functions (4), (5). While the satisfaction relation (1) can be expressed using these feature functions by observing

$$P \models_{\odot} r_i \text{ iff } \sum_{\omega \in \Omega} P(\omega) f_i(\omega) = 0, \tag{7}$$

Here we will transform the f_i so that they do not have to consider worlds any more.

In [10], Kern-Isberner investigates the behavior of worlds with respect to propositional conditionals and introduces the concept of *conditional structure*, formalized as a product in a free Abelian group, of a world with respect to a set of propositional conditionals. Kern-Isberner's idea of a conditional structure carries over to the relational case by employing the functions ver_i, fal_i counting the number of verified and falsified ground instances [12]. In the following, we will employ the conditional structure of a world in a relational setting and extend it to the case where also deterministic conditionals may be present. Instead of a free Abelian group notation as in [12], we will use a concrete representation using ordered tuples of pairs of natural numbers as in [5] and call these tuples *conditional impact*.

Definition 2 (Conditional Impact). *Let* $\mathcal{R} = \mathcal{R}^{\approx} \cup \mathcal{R}^{=}$ *be as in (3). The* conditional impact *caused by a world* ω *is given by the function*

$$\gamma_{\mathcal{R}} : \Omega_{\mathrm{pos}(\mathcal{R})} \to (\mathbb{N}_0 \times \mathbb{N}_0)^m \quad \text{with}$$

$$\gamma_{\mathcal{R}}(\omega) := \Big(\big(ver_1(\omega), fal_1(\omega)\big), \ldots, \big(ver_m(\omega), fal_m(\omega)\big)\Big) \tag{8}$$

Note that the conditional impact caused by a world does neither take any deterministic conditionals nor any probabilities into account, i. e. it just considers the logical part of non-deterministic conditionals in $\mathcal{R}$.

Example 4. As in Example 2, consider again the world

$$\omega' = \{in(a), in(b), an(c), se(a, c), se(b, c)\}. \tag{9}$$

Then $\gamma_{\mathcal{R}}(\omega') = ((0, 0), (0, 1), (2, 4), (0, 3))$ holds, because ω'

- neither verifies nor falsifies any ground instances of r_1, i. e. $ver_1(\omega') = 0$, $fal_1(\omega') = 0$, and
- verifies none and falsifies one ground instance of r_2, i. e. $ver_2(\omega') = 0$, $fal_2(\omega') = 1$, and
- verifies two and falsifies four ground instances of r_3, i. e. $ver_3(\omega') = 2$, $fal_3(\omega') = 4$, and
- verifies none and falsifies three ground instances of r_4, i. e. $ver_4(\omega') = 0$, $fal_4(\omega') = 3$, .

So one can say that $\gamma_{\mathcal{R}}(\omega')$ indicates the conditional impact on ground instances caused by the world ω'. Now consider the worlds

$$\omega'' = \{in(b), in(c), an(a), se(b, a), se(c, a)\} \tag{10}$$

$$\omega''' = \{in(a), in(c), an(b), se(a, b), se(c, b)\} \tag{11}$$

Then determining the values of ver_i and fal_i for ω'' and ω''' reveals that all three worlds have the same conditional impact

$$\begin{aligned} \gamma' : &= ((0, 0), (0, 1), (2, 4), (0, 3)) \\ &= \gamma_{\mathcal{R}}(\omega') = \gamma_{\mathcal{R}}(\omega'') = \gamma_{\mathcal{R}}(\omega''') \end{aligned} \tag{12}$$

Since each of these worlds verifies and falsifies, respectively, the same number of ground instances with respect to each conditional. Note that it does not matter which concrete ground instances of a conditional are verified and falsified, but just the numbers of verified and of falsified ground instances of a conditional are relevant. For instance, each of the three worlds falsifies exactly one ground instance of r_2, i.e., ω' falsifies $(in(c)|an(c))$, ω'' falsifies $(in(a)|an(a))$, and ω''' falsifies $(in(b)|an(b))$.

We now fully abstract from worlds by introducing weighted conditional impacts obtained from the images of $\gamma_{\mathcal{R}}$ and their preimage cardinalities.

Definition 3 (Weighted Conditional Impact). *Let $\mathcal{R} = \mathcal{R}^{\approx} \cup \mathcal{R}^{=}$ as in (3).*

- *A tuple $\gamma \in (\mathbb{N}_0 \times \mathbb{N}_0)^m$ is a* conditional impact of $\mathcal{R}$ *iff there is a world ω with $\gamma_{\mathcal{R}}(\omega) = \gamma$.*
- *For such a γ,* $\mathrm{wgt}(\gamma) := |\gamma_{\mathcal{R}}^{-1}(\gamma)|$ *is the* weight *of γ, and* wgt *is called the* weighting function *of $\mathcal{R}$.*
- *$\Gamma_{\mathcal{R}}$ denotes the set of all conditional impacts of $\mathcal{R}$.*
- $(\Gamma_{\mathcal{R}}, \mathrm{wgt})$ *is called the* weighted conditional impact *of $\mathcal{R}$.*

Example 5. As in Example 4, consider again $\gamma' = ((0,0),(0,1),(2,4),(0,3))$ as given in (12). This tuple γ' is a conditional impact of $\mathcal{R}_{\mathrm{vir}}$, i. e. $\gamma' \in \Gamma_{\mathcal{R}}$ holds, because for instance $\gamma_{\mathcal{R}}(\omega') = \gamma'$ where ω' is as in (9).

Overall, there exist 328 different conditional impacts of $\mathcal{R}_{\mathrm{vir}}$, i. e. the set $\Gamma_{\mathcal{R}}$ has 328 elements. However, the tuple

$$((0,0),(\mathbf{3},\mathbf{2}),(2,4),(0,3)) \in (\mathbb{N}_0 \times \mathbb{N}_0)^m \tag{13}$$

Is not in $\Gamma_{\mathcal{R}}$. In fact, (13) cannot be a conditional impact of $\mathcal{R}_{\mathrm{vir}}$, because r_2 has only 3 ground instances and therefore it is not possible that both $ver_2(\omega) = 3$ and $fal_2(w) = 2$ holds for any world $\omega \in \Omega_{\mathrm{pos}(\mathcal{R})}$. Furthermore, also the tuple

$$((0,0),(0,1),(\mathbf{1},\mathbf{4}),(0,3)) \in (\mathbb{N}_0 \times \mathbb{N}_0)^m \tag{14}$$

Is not a conditional impact of $\mathcal{R}_{\mathrm{vir}}$ either. Here, the reason is not as obvious as it is for (13). A closer lock at the ground instances of r_3 reveals that $ver_3(\omega) + fal_3(w) \in \{0,3,6,9\}$ must hold for every world $\omega \in \Omega_{\mathrm{pos}(\mathcal{R})}$, since always three ground instances of r_3 share the same antecedence, implying that $ver_3(\omega) + fal_3(w)$ must be a multiple of 3. Thus, (14) cannot be a conditional impact of $\mathcal{R}_{\mathrm{vir}}$, since it cannot be caused by any world.

When determining the conditional impacts of $\mathcal{R}_{\mathrm{vir}}$ caused by each world, it becomes evident that apart from the three worlds ω', ω'', and ω''' given in (9), (10), and (11), there is no other world $\omega \in \Omega_{\mathrm{pos}(\mathcal{R})}$ with $\gamma_{\mathcal{R}}(\omega) = \gamma'$. Therefore, $\mathrm{wgt}(\gamma') = 3$ holds, i. e. the weight of γ is 3 since there are exactly three worlds which cause the conditional impact γ'.

Figure 1 shows some conditional impacts of $\mathcal{R}_{\mathrm{vir}}$ and their weights, together with the worlds causing these impacts.

For $\gamma = ((ver_1, fal_1), \ldots, (ver_m, fal_m)) \in \Gamma_{\mathcal{R}}$ let $\gamma_{|i,\mathrm{v}}$ denote the value ver_i and let $\gamma_{|i,\mathrm{f}}$ denote the value fal_i. Then for $r_i \in \mathcal{R}^{\approx}$, the *feature function* $f_i^{\Gamma} : \Gamma_{\mathcal{R}} \to \mathbb{R}$ *on conditional impacts* is given by:

$$f_i^{\Gamma}(\gamma) := \gamma_{|i,\mathrm{v}} \cdot (1 - d_i) - \gamma_{|i,\mathrm{f}} \cdot d_i \tag{15}$$

As in [5], we get *normalized feature functions* $\hat{f}_i^{\Gamma} : \Gamma_{\mathcal{R}} \to [0,1]$ and an additional *correctional feature function* $\hat{f}_{\hat{m}}^{\Gamma} : \Gamma_{\mathcal{R}} \to [0,1]$ with $\hat{m} = m + 1$ by

$$\hat{f}_i^{\Gamma}(\gamma) := \frac{f_i^{\Gamma}(\gamma) + d_i g_i}{G} \quad \text{and} \quad \hat{f}_{\hat{m}}^{\Gamma}(\gamma) := 1 - \sum_{i=1}^{m} \hat{f}_i^{\Gamma}(\gamma) \tag{16}$$

$\gamma \in \Gamma_{\mathcal{R}}$	wgt(γ)	$\omega \in \Omega_{\text{pos}(\mathcal{R})}$ with $\gamma_{\mathcal{R}}(\omega) = \gamma$
$((0,0),(0,1),(2,4),(0,3))$	3	{*in*(a), *in*(b), *an*(c), *se*(a, c), *se*(b, c)}, {*in*(b), *in*(c), *an*(a), *se*(b, a), *se*(c, a)}, {*in*(a), *in*(c), *an*(b), *se*(a, b), *se*(c, b)}
$((0,0),(0,0),(0,9),(0,0))$	1	{*in*(a), *in*(b), *in*(c)}
$((0,0),(0,2),(0,3),(0,6))$	3	{*in*(a), *an*(b), *an*(c)}, {*in*(b), *an*(a), *an*(c)}, {*in*(c), *an*(a), *an*(b)}
...	...	...

Fig. 1. Some conditional impacts of $\mathcal{R}_{\text{vir}}$ and their weights (Example 5)

where g_i denotes the number of ground instances of $r_i \in \mathcal{R}^{\approx}$ and $G := \sum_{r_i \in \mathcal{R}^{\approx}} g_i$. The expected values of these functions remain as in [5]:

$$\hat{\varepsilon}_i = \frac{d_i g_i}{G} \qquad \text{and} \qquad \hat{\varepsilon}_{\hat{m}} = 1 - \sum_{i=1}^{m} \hat{\varepsilon}_i \tag{17}$$

Example 6. As in Example 4, consider again the conditional impact $\gamma' = ((0,0),(0,1),(2,4),(0,3))$ of $\mathcal{R}_{\text{vir}}$ as given in (12). The four feature functions on conditional impacts f_1^{Γ} to f_4^{Γ} corresponding to the four probabilistic conditionals r_1 to r_4 have the following values on γ':

$$f_1^{\Gamma}(\gamma') = 0 \cdot (1-0.9) - 0 \cdot 0.9 = 0 \qquad f_3^{\Gamma}(\gamma') = 2 \cdot (1-0.7) - 4 \cdot 0.7 = -2.2$$
$$f_2^{\Gamma}(\gamma') = 0 \cdot (1-0.01) - 1 \cdot 0.01 = -0.01 \qquad f_4^{\Gamma}(\gamma') = 0 \cdot (1-0.1) - 3 \cdot 0.1 = -0.3$$

Since the conditionals have $g_1 = g_3 = g_4 = 9$ and $g_2 = 3$ ground instances, respectively, there is an overall number of $G = 30$ ground instances (cf. Example 2). Therefore, the corresponding normalized feature functions $\hat{f}_i^{\Gamma}$ have the following values on γ' and the following expected values $\hat{\varepsilon}_i$:

$$\hat{f}_1^{\Gamma}(\gamma') = (0 + 0.9 \cdot 9) \,/\, 30 = 0.27 \quad \text{and} \quad \hat{\varepsilon}_1 = \tfrac{0.9 \cdot 9}{30} = 0.27$$
$$\hat{f}_2^{\Gamma}(\gamma') = (-0.01 + 0.01 \cdot 3) \,/\, 30 = 0.000\overline{6} \quad \text{and} \quad \hat{\varepsilon}_2 = \tfrac{0.01 \cdot 3}{30} = 0.001$$
$$\hat{f}_3^{\Gamma}(\gamma') = (-2.2 + 0.7 \cdot 9) \,/\, 30 = 0.13\overline{6} \quad \text{and} \quad \hat{\varepsilon}_3 = \tfrac{0.7 \cdot 9}{30} = 0.21$$
$$\hat{f}_4^{\Gamma}(\gamma') = (-0.3 + 0.1 \cdot 9) \,/\, 30 = 0.02 \quad \text{and} \quad \hat{\varepsilon}_4 = \tfrac{0.1 \cdot 9}{30} = 0.03$$

So for the correctional feature function $\hat{f}_{\hat{m}}^{\Gamma}$ and the expected value $\hat{\varepsilon}_{\hat{m}}$ we get:

$$\hat{f}_{\hat{m}}^{\Gamma}(\gamma') = 1 - (0.27 + 0.000\overline{6} + 0.13\overline{6} + 0.02) = 0.572\overline{6}$$
$$\hat{\varepsilon}_{\hat{m}} = 1 - (0.27 + 0.001 + 0.21 + 0.03) = 0.489$$

For every $\omega', \omega'' \in \gamma_{\mathcal{R}}^{-1}(\gamma)$, we have $P_{\mathcal{R}}^*(\omega') = P_{\mathcal{R}}^*(\omega'')$ (cf. Corollary 1 in [5]). Thus, setting $P_{\mathcal{R}}^*(\gamma) := P_{\mathcal{R}}^*(\omega')$ for an arbitrary $\omega' \in \gamma_{\mathcal{R}}^{-1}(\gamma)$ yields a well-defined function $P_{\mathcal{R}}^* : \Gamma_{\mathcal{R}} \to [0,1]$. Using this function, we can express the satisfaction relation (7) with respect to the ME model $P_{\mathcal{R}}^*$ as follows:

Proposition 2 (Satisfaction for $P^*_{\mathcal{R}}$). *Let $\mathcal{R} = \mathcal{R}^{\approx} \cup \mathcal{R}^{=}$ be as in (3). Then for any probabilistic conditional $r_i \in \mathcal{R}^{\approx}$, we have*

$$P^*_{\mathcal{R}} \models_{\odot} r_i \quad \text{iff} \quad \sum_{\gamma \in \Gamma_{\mathcal{R}}} \hat{f}^{\Gamma}_i(\gamma) \cdot \mathrm{wgt}(\gamma) \cdot P^*_{\mathcal{R}}(\gamma) = \hat{\varepsilon}_i \qquad (18)$$

Compared to (1) and (7), the satisfaction relation (18) employs feature functions on conditional impacts. Thereby it allows us to solve the ME optimization problem induced by (2) by a two-level algorithmic approach: First, the weighted conditional impact $(\Gamma_{\mathcal{R}}, \mathrm{wgt})$ is determined, then a generalized iterating scaling algorithm working on $(\Gamma_{\mathcal{R}}, \mathrm{wgt})$ is used to determine the ME distribution $P^*_{\mathcal{R}}$.

In the following, we omit the index Γ of f^{Γ}_i and $\hat{f}^{\Gamma}_i$ in order to ease our notation as it will be clear from the context when we use feature functions operating on the set of conditional impacts $\Gamma_{\mathcal{R}}$ rather than on worlds.

5 Computing ME Models using Weighted Conditional Impacts

The algorithm WCI implemented in AggME for computing the weighted conditional impact $(\Gamma_{\mathcal{R}}, \mathrm{wgt})$ of any $\mathcal{R}$ is given in Fig. 2. The algorithm starts with an empty set $\Gamma_{\mathcal{R}}$ in step (1). Then the elements of the set $\Gamma_{\mathcal{R}}$ and the values for the weighting function wgt are successively determined by performing the following steps once for each world $\omega \in \Omega$: In step (2a), the deterministic conditionals are exploited to check if ω is a null-world. If ω is a null-world, no further steps are performed on ω. Otherwise ω is a positive world and in step (2b), the conditional impact $\gamma_{\mathcal{R}}(\omega)$ is determined. In step (2c), $\gamma_{\mathcal{R}}(\omega)$ is appended to $\Gamma_{\mathcal{R}}$ if its not already there, and its weight $\mathrm{wgt}(\gamma_{\mathcal{R}}(\omega))$ is adjusted.

Having determined $(\Gamma_{\mathcal{R}}, \mathrm{wgt})$ for a knowledge base $\mathcal{R}$, the algorithm $\mathrm{GIS}^{\gamma_{\mathcal{R}}}_{\odot}$ given in Fig. 3 is used for the second phase of computing the ME model $P^*_{\mathcal{R}}$. As in [5], a generalized iterative scaling approach is used, but $\mathrm{GIS}^{\gamma_{\mathcal{R}}}_{\odot}$ fully abstracts from worlds. Instead of referring to worlds as the algorithms $\mathrm{GIS}^{\alpha}_{\odot}$ and $\mathrm{GIS}^{\equiv_{\mathcal{R}}}_{\odot}$ in [5], $\mathrm{GIS}^{\gamma_{\mathcal{R}}}_{\odot}$ performs all steps on $(\Gamma_{\mathcal{R}}, \mathrm{wgt})$. That way, $\mathrm{GIS}^{\gamma_{\mathcal{R}}}_{\odot}$ can also cope with deterministic conditionals which are not allowed in [5].

For any consistent set $\mathcal{R}$ of probabilistic conditionals as in Fig. 3, $\mathrm{GIS}^{\gamma_{\mathcal{R}}}_{\odot}$ computes values $\alpha_0, \alpha_1, \ldots, \alpha_{m+M}$. Based on the method of Lagrange multipliers [3], these alpha-values determine the ME model as a Gibbs distribution [8] by

$$P^*_{\mathcal{R}}(\omega) = \alpha_0 \prod_{i=1}^{m+M} \alpha_i^{f_i(\omega)} \qquad (19)$$

With the feature functions f_i as given in (6).

6 Examples and First Evaluation Results

We apply the two-phase ME model computation implemented in AggME to different knowledge bases; the results are shown in Fig. 4.

Input: a set $\mathcal{R} = \{r_1, \ldots, r_m\} \cup \{r_{m+1}, \ldots, r_{m+M}\}$
of m non-deterministic and M deterministic probabilistic conditionals
Output: the weighted conditional impact $(\Gamma_{\mathcal{R}}, \mathrm{wgt})$ of $\mathcal{R}$

1. $\Gamma_{\mathcal{R}} := \emptyset$ // *initialize value*
2. for each $\omega \in \Omega$:
   (a) // *check if* ω *is a null-world by evaluating deterministic conditionals*
   for each $r_j = (B_j(\boldsymbol{X})|A_j(\boldsymbol{X})) \in \mathcal{R}^{=}$: // *for determ. cond.* r_j, $m{+}1 \leq j \leq M$
   // *consider all ground instances of* r_j
   for each $(B_j(\boldsymbol{a})|A_j(\boldsymbol{a})) \in \mathrm{gnd}(B_j(\boldsymbol{X})|A_j(\boldsymbol{X}))$:
   if $(r_j \in \mathcal{R}^{=0})$
   if $(\omega \models A_j(\boldsymbol{a})B_j(\boldsymbol{a}))$ // *if* ω *verifies this ground conditional*
   then break to step 2 // ω *is a null-world, so check is finished*
   else // $r_j \in \mathcal{R}^{=1}$ *holds*
   if $(\omega \models A_j(\boldsymbol{a})\overline{B_j(\boldsymbol{a})})$ // *if* ω *falsifies this ground conditional*
   then break to step 2 // ω *is a null-world, so check is finished*
   end loop
   end loop
   // ω *is a positive world, since no determ. cond. proved* ω *to be a null-world*
   (b) // *determine* $\gamma_{\mathcal{R}}(\omega)$
   $\gamma_{\mathcal{R}}(\omega) := ((0,0), \ldots, (0,0))$ // *initialize all vf-pairs with* $(0,0)$
   for each $r_i = (B_i(\boldsymbol{X})|A_i(\boldsymbol{X})) \in \mathcal{R}^{\approx}$: // *for non-determ. cond.* r_i, $1 \leq i \leq m$
   // *consider all ground instances of* r_i
   for each $(B_i(\boldsymbol{a})|A_i(\boldsymbol{a})) \in \mathrm{gnd}(B_i(\boldsymbol{X})|A_i(\boldsymbol{X}))$:
   if $(\omega \models A_i(\boldsymbol{a})B_i(\boldsymbol{a}))$ // *if* ω *verifies this ground conditional*
   then $\gamma_{\mathcal{R}}(\omega)_{|i,\mathrm{v}} := \gamma_{\mathcal{R}}(\omega)_{|i,\mathrm{v}} + 1$ // *then increment verify count*
   if $(\omega \models A_i(\boldsymbol{a})\overline{B_i(\boldsymbol{a})})$ // *if* ω *falsifies this ground conditional*
   then $\gamma_{\mathcal{R}}(\omega)_{|i,\mathrm{f}} := \gamma_{\mathcal{R}}(\omega)_{|i,\mathrm{f}} + 1$ // *then increment falsify count*
   end loop
   end loop
   (c) // *check if value* $\gamma_{\mathcal{R}}(\omega)$ *is already contained in* $\Gamma_{\mathcal{R}}$
   if $\gamma_{\mathcal{R}}(\omega) \in \Gamma_{\mathcal{R}}$
   then $\mathrm{wgt}(\gamma_{\mathcal{R}}(\omega)) := \mathrm{wgt}(\gamma_{\mathcal{R}}(\omega)) + 1$ // *increment cardinality of* $\gamma_{\mathcal{R}}(\omega)$
   else
   $\Gamma_{\mathcal{R}} := \Gamma_{\mathcal{R}} \cup \{\gamma_{\mathcal{R}}(\omega)\}$ // *add new value* $\gamma_{\mathcal{R}}(\omega)$ *to* $\Gamma_{\mathcal{R}}$
   $\mathrm{wgt}(\gamma_{\mathcal{R}}(\omega)) := 1$ // *initialize cardinality of* $\gamma_{\mathcal{R}}(\omega)$ *with 1*
   end loop

Fig. 2. Algorithm WCI computing the weighted conditional impact of $\mathcal{R}$

Input: - a consistent set $\mathcal{R} = \{r_1, \ldots, r_m\} \cup \{r_{m+1}, \ldots, r_{m+M}\}$ of m non-deterministic and M deterministic probabilistic conditionals

- the weighted conditional impact $(\Gamma_\mathcal{R}, \mathrm{wgt})$ of $\mathcal{R}$

Output: - alpha-values $\alpha_0, \alpha_1, \ldots, \alpha_{m+M}$ determining the ME-distribution $P^*_\mathcal{R}$

1. for each $1 \leq i \leq \hat{m}$: $\hat{\alpha}_{(0),i} := 1$ // *initialize normalized $\hat{\alpha}$-values*
2. for each $\gamma \in \Gamma_\mathcal{R}$: $P_{(0)}(\gamma) := \dfrac{1}{\sum_{\gamma' \in \Gamma_\mathcal{R}} \mathrm{wgt}(\gamma')}$ // *initial. to uniform probabilities*
3. $k := 0$ // *initialize iteration counter*
4. repeat until an abortion condition holds:
   (a) $k := k + 1$ // *increment iteration counter k*
   (b) for each $1 \leq i \leq \hat{m}$: // *determine current scaling factors $\beta_{(k),i}$*
   $$\beta_{(k),i} := \frac{\hat{\varepsilon}_i}{\sum_{\gamma \in \Gamma_\mathcal{R}} \mathrm{wgt}(\gamma) P_{(k-1)}(\gamma) \hat{f}_i(\gamma)}$$
   (c) for each $\gamma \in \Gamma_\mathcal{R}$: // *scale all probabilities $P'_{(k)}(\gamma)$*
   $$P'_{(k)}(\gamma) := P_{(k-1)}(\gamma) \prod_{i=1}^{\hat{m}} \left(\beta_{(k),i}\right)^{\hat{f}_i(\gamma)}$$
   (d) for each $\hat{\alpha}_{(k),i}, 1 \leq i \leq \hat{m}$: // *scale all $\hat{\alpha}$-values $\hat{\alpha}_{(k),i}$*
   $\hat{\alpha}_{(k),i} := \hat{\alpha}_{(k-1),i} \cdot \beta_{(k),i}$
   (e) for each $\gamma \in \Gamma_\mathcal{R}$: // *normalize all probabilities $P_{(k)}(\gamma)$*
   $$P_{(k)}(\gamma) := \frac{P'_{(k)}(\gamma)}{\sum_{\gamma' \in \Gamma_\mathcal{R}} \mathrm{wgt}(\gamma') P'_{(k)}(\gamma)}$$
   end loop
5. for each $1 \leq i \leq \hat{m}$: $\hat{\alpha}_i := \hat{\alpha}_{(k),i}$ // *define final $\hat{\alpha}$-values*
   $$\hat{\alpha}_0 := \left(\sum_{\gamma \in \Gamma_\mathcal{R}} \mathrm{wgt}(\gamma) \prod_{i=1}^{\hat{m}} \hat{\alpha}_i^{\hat{f}_i(\gamma)} \right)^{-1}$$ // *define $\hat{\alpha}_0$-value*
6. for each $1 \leq i \leq m$: $\alpha_i := \left(\frac{\hat{\alpha}_i}{\hat{\alpha}_{\hat{m}}}\right)^{\frac{1}{G}}$ // *define α-values for $\mathcal{R}^{\approx}$*
   for each $1 \leq j \leq M$: $\alpha_{m+j} := 0$ // *define α-values for $\mathcal{R}^{=}$*
   $\alpha_0 := \hat{\alpha}_0 \hat{\alpha}_{\hat{m}} \prod_{i=1}^{m} \alpha_i^{d_i g_i}$ // *define α_0-value*

Fig. 3. Algorithm $\mathrm{GIS}^{\gamma\mathcal{R}}_{\odot}$ for aggregation semantics operating on $(\Gamma_\mathcal{R}, \mathrm{wgt})$

Example 7 ($\mathcal{R}_{\mathrm{vir}}$ (cont.)). When considering $\mathcal{R}_{\mathrm{vir}}$ from Example 1 together with five constants, the size of Ω is 2^{35} and even $\Omega_{\mathrm{pos}(\mathcal{R})}$ still contains 2^{30} worlds. Since the method given in [5] for computing the ME model requires to keep all worlds in memory, it cannot be applied to that example due to memory limitations. However, the algorithms WCI and $\mathrm{GIS}^{\gamma\mathcal{R}}_{\odot}$ can easily cope with the example, since they just require to keep 18,720 conditional impacts in memory. Fig. 4 also illustrates that for increasing sizes of Ω, the computation of the weighted conditional impact becomes the dominating part in the overall computation time.

KB	$\|Const\|$	Size of Ω	Size of $\Omega_{\text{pos}(\mathcal{R})}$	Size of $\Gamma_{\mathcal{R}}$	Iteration Steps	Computation Time WCI	Computation Time $\text{GIS}_{\odot}^{\gamma_{\mathcal{R}}}$
$\mathcal{R}_{\text{mky}}$	4	2^{20}	$6{,}561 \approx 2^{12}$	156	6,721	< 1 sec	< 1 sec
$\mathcal{R}_{\text{mky}}$	5	2^{30}	$1{,}419{,}857 \approx 2^{20}$	530	7,912	6 min 41 sec	1 sec
$\mathcal{R}_{\text{cty}}$	3+4	2^{21}	$1{,}404{,}928 \approx 2^{20}$	992	4,228	5 sec	1 sec
$\mathcal{R}_{\text{cty}}$	4+4	2^{28}	$157{,}351{,}936 \approx 2^{27}$	3,601	4,947	9 min 41 sec	3 sec
$\mathcal{R}_{\text{vir}}$	4	2^{24}	$1{,}048{,}576 = 2^{20}$	2,742	5,730	13 sec	4 sec
$\mathcal{R}_{\text{vir}}$	5	2^{35}	$1{,}073{,}741{,}824 = 2^{30}$	18,720	4,088	4 h 36 min	15 sec

Fig. 4. Results for example knowledge bases (GIS accuracy threshold: $\delta_\beta = 0.001$)

Example 8 (Monkeys, $\mathcal{R}_{\text{mky}}$). Suppose we have a zoo with a population of monkeys exhibiting a peculiar feeding behavior. The predicate $\mathit{feeds}(X, Y)$ expresses that a monkey X feeds another monkey Y and $\mathit{hungry}(X)$ says that a monkey X is hungry. $\mathcal{R}_{\text{mky}}$ contains the following conditionals:

$$r_1 : (\mathit{feeds}(X, Y) \mid \neg \mathit{hungry}(X) \wedge \mathit{hungry}(Y)) \; [0.80]$$
$$r_2 : (\neg \mathit{feeds}(X, Y) \mid \mathit{hungry}(X)) \; [1.0]$$
$$r_3 : (\neg \mathit{feeds}(X, Y) \mid \neg \mathit{hungry}(X) \wedge \neg \mathit{hungry}(Y)) \; [0.90]$$
$$r_4 : (\mathit{feeds}(X, \mathit{charly}) \mid \neg \mathit{hungry}(X)) \; [0.95]$$
$$r_5 : (\mathit{feeds}(X, X) \mid \top) \; [0.0]$$

r_1 states that is very likely that a not-hungry monkey feeds a hungry monkey. r_2 expresses the certain knowledge that a hungry monkey never feeds another one. r_3 says that it is very probable that a not-hungry monkey is not fed by another one. r_4 makes a statement about an individual monkey: it is most probable that if a monkey is not hungry, he feeds the monkey Charly, i. e. albeit Charly is hungry or not (perhaps because Charly is an underfed baby monkey suffering from an eating disorder). Thus, r_4 describes a special case for Charly, because according to r_3, one would have suspected that the feeding of Charly (by a not-hungry monkey) depends on whether Charly is hungry or not. r_5 expresses that a monkey does not feed itself.

Example 9 (European Cities, $\mathcal{R}_{\text{cty}}$). This example makes use of *typed* constants and predicates; there are a certain number of constants of type *Person*, and the constants *london*, *paris*, *rome*, and *vienna* are of type *EuropeanCity*. The predicate $\mathit{visitsEUcity}(P, E)$ expresses that a person P visits a European city E. The predicates $\mathit{likesSightseeing}(P)$, $\mathit{livesInEurope}(P)$, and $\mathit{likesChurches}(P)$ express that a person P likes sightseeing, lives in Europe, and likes churches, respectively. The set $\mathcal{R}_{\text{cty}}$ contains four conditionals:

$$r_1 : (\mathit{visitsEUcity}(P, C) \mid \top) \; [0.1]$$
$$r_2 : (\mathit{visitsEUcity}(P, C) \mid \mathit{likesSightseeing}(P)) \; [0.3]$$
$$r_3 : (\mathit{visitsEUcity}(P, C) \mid \mathit{livesInEurope}(P)) \; [0.6]$$
$$r_4 : (\mathit{visitsEUcity}(P, \mathit{rome}) \mid \mathit{likesChurches}(P) \wedge \mathit{likesSightseeing}(P)) \; [1.0]$$

Looking at Examples 7–9 and the corresponding numbers in Fig. 4, we would like to point out the following aspects. By allowing deterministic conditionals, there is no more need to approximate probabilities 0 or 1 as a workaround. For instance, if the probabilities of r_2 and r_5 in $\mathcal{R}_{\text{mky}}$ were approximated by 0.999 and 0.001, respectively, then $\Omega_{\text{pos}(\mathcal{R})} = \Omega$ would hold and, in case of five constants, more than a billion worlds would have to be processed in the expensive step 2b of algorithm WCI. Furthermore, the size of $\Gamma_{\mathcal{R}}$ would increase significantly as well, increasing the runtime of $\text{GIS}_{\odot}^{\gamma_{\mathcal{R}}}$.

Using weighted conditional impacts $(\Gamma_{\mathcal{R}}, \text{wgt})$ and pre-computing them reduces the overall computation time. For instance, computing the ME model for $\mathcal{R}_{\text{cty}}$ with 3 constants of type *Person* and 4 constants of type *EuropeanCity* by a straightforward implementation of a GIS algorithm on $\Omega_{\text{pos}(\mathcal{R})}$, requires over 17 min., compared to just 6 sec. overall for WCI and $\text{GIS}_{\odot}^{\gamma_{\mathcal{R}}}$ as shown in Fig. 4.

Another important benefit of pre-computing $(\Gamma_{\mathcal{R}}, \text{wgt})$ is that it can be reused if the probabilities of some conditionals of $\mathcal{R}^{\approx}$ are modified since $(\Gamma_{\mathcal{R}}, \text{wgt})$ only depends on the logical part of $\mathcal{R}^{\approx}$. Since $|\Gamma_{\mathcal{R}}|$ is much smaller than $|\Omega_{\text{pos}(\mathcal{R})}|$, working with $(\Gamma_{\mathcal{R}}, \text{wgt})$ also reduces the memory requirements for the ME model computation significantly. In fact, while the algorithm from [5] has a memory requirement of $O(|\Omega_{\text{pos}(\mathcal{R})}|)$, preventing it to handle some of the examples given in Fig. 4, the memory requirements during all phases of the ME computation in AggME are limited by $O(|\Gamma_{\mathcal{R}}|)$.

As pointed out in Ex. 7, the numbers in Fig. 4 illustrate that increasing the size of *Const* and thus the size of Ω is the limiting factor for ME model computation in the current AggME version. An advantage of the two-level approach is that WCI can be replaced by another algorithm computing the weighted conditional impact of $\mathcal{R}$ more efficiently without having to change $\text{GIS}_{\odot}^{\gamma_{\mathcal{R}}}$.

7 Conclusions and Further Work

For knowledge bases $\mathcal{R}$ with probabilistic relational conditionals, we presented a two-level approach for computing the ME model $P_{\mathcal{R}}^*$ under aggregation semantics, thereby improving on previous work. While our approach can handle larger examples and also deterministic conditionals, it is desirable to develop alternative methods for computing the weighted conditional impact of $\mathcal{R}$ without having to enumerate all possible worlds as in step (2.) of the WCI algorithm. Therefore, we are currently working on employing a combinatorial approach to construct $(\Gamma_{\mathcal{R}}, \text{wgt})$ directly, without considering worlds explicitly. That way, the exponential blow-up in Ω could be circumvented when computing $(\Gamma_{\mathcal{R}}, \text{wgt})$, allowing to handle domains with significantly more constants. We are also investigating which alternative algorithms could be employed to solve the ME optimization problem on $(\Gamma_{\mathcal{R}}, \text{wgt})$. For instance, instead of using a generalized iterative scaling approach as in our $\text{GIS}_{\odot}^{\gamma_{\mathcal{R}}}$ algorithm, an alternative approach like L-BFGS [21] could be considered.

Furthermore, we will exploit the concept of weighted conditional impacts for actual ME inference, i. e. for determining the probability of an arbitrary conditional under the ME model $P_{\mathcal{R}}^*$. To accomplish that, a technique should be developed which operates on the impact of an arbitrary conditional, i. e. the actual query, and the already determined weighted conditional impact of $\mathcal{R}$; for this, methods of lifted inference [19,14] might be applicable.

References

1. Adams, E.: The Logic of Conditionals. D. Reidel, Dordrecht (1975)
2. Beierle, C., Finthammer, M., Kern-Isberner, G., Thimm, M.: Evaluation and comparison criteria for approaches to probabilistic relational knowledge representation. In: Bach, J., Edelkamp, S. (eds.) KI 2011. LNCS, vol. 7006, pp. 63–74. Springer, Heidelberg (2011)
3. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, New York (2004)
4. Fagin, R., Halpern, J.Y.: Reasoning about knowledge and probability. J. ACM 41(2), 340–367 (1994)
5. Finthammer, M., Beierle, C.: Using equivalences of worlds for aggregation semantics of relational conditionals. In: Glimm, B., Krüger, A. (eds.) KI 2012. LNCS, vol. 7526, pp. 49–60. Springer, Heidelberg (2012)
6. Finthammer, M., Thimm, M.: An integrated development environment for probabilistic relational reasoning. Logic Journal of the IGPL 20(5), 831–871 (2012)
7. Fisseler, J.: Learning and Modeling with Probabilistic Conditional Logic, Dissertations in Artificial Intelligence, vol. 328. IOS Press, Amsterdam (2010)
8. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721–741 (1984)
9. Halpern, J.: Reasoning About Uncertainty. MIT Press (2005)
10. Kern-Isberner, G.: Conditionals in Nonmonotonic Reasoning and Belief Revision. LNCS (LNAI), vol. 2087, p. 27. Springer, Heidelberg (2001)
11. Kern-Isberner, G., Lukasiewicz, T.: Combining probabilistic logic programming with the power of maximum entropy. Artif. Intell. 157(1-2), 139–202 (2004)
12. Kern-Isberner, G., Thimm, M.: A ranking semantics for first-order conditionals. In: ECAI 2012, pp. 456–461. IOS Press (2012)
13. Kern-Isberner, G., Thimm, M.: Novel semantical approaches to relational probabilistic conditionals. In: Proc. of KR 2010, pp. 382–392. AAAI Press (May 2010)
14. Milch, B., Zettlemoyer, L., Kersting, K., Haimes, M., Kaelbling, L.P.: Lifted probabilistic inference with counting formulas. In: AAAI 2008, pp. 1062–1068. AAAI Press (2008)
15. Nilsson, N.: Probabilistic logic. Artificial Intelligence 28, 71–87 (1986)
16. Nute, D., Cross, C.: Conditional logic. In: Gabbay, D., Guenther, F. (eds.) Handbook of Philosophical Logic, 2nd edn., vol. 4, pp. 1–98. Kluwer Academic Publishers (2002)
17. Paris, J.: The uncertain reasoner's companion – A mathematical perspective. Cambridge University Press (1994)
18. Pearl, J.: Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann, San Mateo (1988)
19. de Salvo Braz, R., Amir, E., Roth, D.: Lifted first-order probabilistic inference. In: IJCAI 2005, pp. 1319–1325. Professional Book Center (2005)
20. Shore, J., Johnson, R.: Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. IEEE Transactions on Information Theory IT-26, 26–37 (1980)
21. Zhu, C., Byrd, R.H., Lu, P., Nocedal, J.: Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. ACM Trans. Math. Softw. 23(4), 550–560 (1997), `http://doi.acm.org/10.1145/279232.279236`