Computing LP$^{\text{MLN}}$ *Using ASP and MLN Solvers (System Paper)*

Joohyung Lee, Samidh Talsania, and Yi Wang
School of Computing, Informatics, and Decision Systems Engineering
Arizona State University, Tempe, USA
(*e-mail:* {joolee, stalsani, ywang485}@asu.edu)

submitted ; revised ; accepted

Abstract

LP$^{\text{MLN}}$ is a recent addition to probabilistic logic programming languages. Its main idea is to overcome the rigid nature of the stable model semantics by assigning a weight to each rule in a way similar to Markov Logic is defined. We present two implementations of LP$^{\text{MLN}}$, LPMLN2ASP and LPMLN2MLN. System LPMLN2ASP translates LP$^{\text{MLN}}$ programs into the input language of answer set solver CLINGO, and using weak constraints and stable model enumeration, it can compute most probable stable models as well as exact conditional and marginal probabilities. System LPMLN2MLN translates LP$^{\text{MLN}}$ programs into the input language of Markov Logic solvers, such as ALCHEMY, TUFFY, and ROCKIT, and allows for performing approximate probabilistic inference on LP$^{\text{MLN}}$ programs. We also demonstrate the usefulness of the LP$^{\text{MLN}}$ systems for computing other languages, such as ProbLog and Pearl's Causal Models, that are shown to be translatable into LP$^{\text{MLN}}$.

1 Introduction

LP$^{\text{MLN}}$ is a simple extension of answer set programs with the concept of weighted rules, whose weight scheme is adopted from that of Markov Logic (Richardson and Domingos 2006). Like Markov Logic, not all LP$^{\text{MLN}}$ rules have to be true but, roughly speaking, the more rules are true, the larger weight is assigned to the corresponding stable model.

It is shown that both answer set programs and Markov Logic can be easily embedded in LP$^{\text{MLN}}$ (Theorems 1, 2 of (Lee and Wang 2016)). The other direction is more interesting from a computational point of view because it provides ways to compute LP$^{\text{MLN}}$ using existing implementations of ASP and MLN (Markov Logic Network) solvers. This paper further develops the translations in this direction, presents two implementations of LP$^{\text{MLN}}$ based on them along with a few new theorems justifying the implementations, and illustrates the usefulness of the systems.

It is shown in (Balai and Gelfond 2016) that LP$^{\text{MLN}}$ programs can be turned into P-log programs. We present a similar but simpler translation that turns LP$^{\text{MLN}}$ programs into answer set programs. However, for a non-ground LP$^{\text{MLN}}$ programs, the translation yields an answer set program that is unsafe, so a direct implementation of this idea has a drawback. Instead, we develop an implementation based on the translation from (Lee and Yang 2017) that turns (ground) LP$^{\text{MLN}}$ programs into answer set programs containing weak constraints so that the most probable stable models of an LP$^{\text{MLN}}$ program coincide with the optimal stable models of the translated ASP program. Going further, we show how to map the penalty of each stable model of the translated ASP program to the probability of each stable model of a (non-ground) LP$^{\text{MLN}}$ program, and use this result to compute probabilistic queries for an LP$^{\text{MLN}}$ program using an ASP solver CLINGO. The input language of LPMLN2ASP is familiar to the users of CLINGO because its syntax is a simple extension of CLINGO

rules with weights, thereby allowing many advanced constructs of the CLINGO language, such as aggregates and conditional literals, in the context of LP$^{\text{MLN}}$.

A different method to compute an LP$^{\text{MLN}}$ program is by converting it into a Markov Logic Network (Lee and Wang 2016, Theorem 3), similar to the reduction of answer set programs to propositional logic, and then invoking MLN solvers, such as ALCHEMY, TUFFY, and ROCKIT. While it is possible to turn any LP$^{\text{MLN}}$ program into an equivalent MLN by adding all loop formulas, in practice, this method does not yield an effective computation. Thus, we limit attention to the "tight" fragment of LP$^{\text{MLN}}$ programs that can be easily converted into MLNs using the process of completion. Even so, the straightforward encoding of completion formulas in Markov Logic may lead to a blow-up in CNF conversion performed by ALCHEMY because the conversion is naively implemented in ALCHEMY. Furthermore, the input languages of TUFFY and ROCKIT do not even allow nested formulas, which are required to encode completion formulas. So, LPMLN2MLN uses some equivalent transformation using auxiliary atoms to avoid the blow-up in CNF conversion and takes care of differences in the input language of different MLN solvers. The input language of LPMLN2MLN resembles that of ALCHEMY, and is converted into one of the input languages of ALCHEMY, TUFFY, and ROCKIT depending on the mode selected. The system utilizes approximate probabilistic inference methods or the exact optimization methods supported by each MLN solver.

The implementations are not only interesting for computing LP$^{\text{MLN}}$. System LPMLN2ASP can be used to derive the most probable stable models even when the standard answer set program is inconsistent. This feature could be useful in debugging an inconsistent answer set program or deriving some meaningful conclusions from an inconsistent knowledge base. Also, both implementations can be used to compute other formalisms, such as Markov Logic, ProbLog (De Raedt et al. 2007), Pearl's Probabilistic Causal Models (Pearl 2000), and P-log (Baral et al. 2009), which are shown to be translatable into LP$^{\text{MLN}}$ (Lee and Wang 2016; Lee et al. 2015; Lee and Yang 2017).

The systems are publicly available at

`http://reasoning.eas.asu.edu/lpmln/`,

along with the user manual and examples. We refer the reader to the system homepage for more details.

The paper is organized as follows. Section 2 reviews the language LP$^{\text{MLN}}$ based on the concept of reward and presents a reformulation of LP$^{\text{MLN}}$ based on the concept of penalty. Section 3 shows two translations of LP$^{\text{MLN}}$ programs into answer set programs and presents system LPMLN2ASP that implements one of them. Section 4 shows a translation of tight LP$^{\text{MLN}}$ programs into Markov Logic Networks and presents system LPMLN2MLN that implements the translation. Section 5 gives a comparison and running statistics of these implementations. Section 6 shows how to use these systems to compute other probabilistic logic languages that are shown to be translatable into LP$^{\text{MLN}}$.

2 Language LP$^{\text{MLN}}$

2.1 *Review:* LP$^{\text{MLN}}$

We review the definition of LP$^{\text{MLN}}$ from (Lee and Wang 2016). An LP$^{\text{MLN}}$ program is a finite set of weighted rules $w : R$ where R is a rule (as allowed in the input language of ASP solver CLINGO), and w is a real number (in which case, the weighted rule is called *soft*) or α for denoting the infinite weight (in which case, the weighted rule is called *hard*). An LP$^{\text{MLN}}$ program is called *ground* if its rules contain no variables. We assume a finite Herbrand Universe so that the ground program is finite. Each ground instance of a non-ground rule receives the same weight as the original non-ground formula.

For any ground LP$^{\text{MLN}}$ program Π and any interpretation I, $\overline{\Pi}$ denotes the usual (unweighted) ASP program obtained from Π by dropping the weights, and Π_I denotes the set of $w : R$ in Π such that $I \models R$, and $\text{SM}[\Pi]$ denotes the set $\{I \mid I \text{ is a stable model of } \overline{\Pi_I}\}$. The *unnormalized weight* of an interpretation I under Π is defined as

$$W_\Pi(I) = \begin{cases} exp\left(\sum\limits_{w:R \in \Pi_I} w\right) & \text{if } I \in \text{SM}[\Pi]; \\ 0 & \text{otherwise.} \end{cases}$$

The *normalized weight* (a.k.a. *probability*) of an interpretation I under Π is defined as

$$P_\Pi(I) = \lim_{\alpha \to \infty} \frac{W_\Pi(I)}{\sum\limits_{J \in \text{SM}[\Pi]} W_\Pi(J)}.$$

I is called a *(probabilistic) stable model* of Π if $P_\Pi(I) \neq 0$. The most probable stable models of Π are the stable models with the highest probability.

2.2 *Reformulating* LP$^{\text{MLN}}$ *Based on the Concept of Penalty*

In the definition of the LP$^{\text{MLN}}$ semantics from (Lee and Wang 2016), the weight assigned to each stable model can be regarded as "rewards": the more rules are true in deriving the stable model, the larger weight is assigned to it. In this section, we reformulate the LP$^{\text{MLN}}$ semantics in a "penalty" based way. More precisely, the penalty based weight of an interpretation I is defined as the exponentiated negative sum of the weights of the rules that are not satisfied by I (when I is a stable model of $\overline{\Pi_I}$). Let

$$W_\Pi^{\text{pnt}}(I) = \begin{cases} exp\left(-\sum\limits_{w:R \in \Pi \text{ and } I \not\models R} w\right) & \text{if } I \in \text{SM}[\Pi]; \\ 0 & \text{otherwise.} \end{cases}$$

and

$$P_\Pi^{\text{pnt}}(I) = \lim_{\alpha \to \infty} \frac{W_\Pi^{\text{pnt}}(I)}{\sum\limits_{J \in \text{SM}[\Pi]} W_\Pi^{\text{pnt}}(J)}.$$

The following theorem tells us that the LP$^{\text{MLN}}$ semantics can be reformulated using the concept of a penalty-based weight.

Theorem 1

For any LP$^{\text{MLN}}$ program Π and any interpretation I,

$$W_\Pi(I) \propto W_\Pi^{\text{pnt}}(I) \quad \text{and} \quad P_\Pi(I) = P_\Pi^{\text{pnt}}(I).$$

Although the penalty-based reformulation appears to be more complicated, it has a few desirable features. One of them is that adding a trivial rule does not affect the weight of an interpretation, which is not the case with the original definition. More importantly, this reformulation leads to a better translation of LP$^{\text{MLN}}$ programs into answer set programs as we discuss in Section 3.3.

3 Turning LP$^{\mathrm{MLN}}$ into ASP with Weak Constraints

3.1 Review: Weak Constraints

A *weak constraint* (Buccafurri et al. 2000; Calimeri et al. 2013) has the form

$$:\sim F \quad [\mathit{Weight}\ @\ \mathit{Level}].$$

where F is a conjunction of literals, *Weight* is a real number, and *Level* is a nonnegative integer.

Let Π be a program $\Pi_1 \cup \Pi_2$, where Π_1 is an answer set program that does not contain weak constraints, and Π_2 is a set of weak constraints. We call I a stable model of Π if it is a stable model of Π_1. For every stable model I of Π and any nonnegative integer l, the *penalty* of I at level l, denoted by $\mathit{Penalty}_\Pi(I,l)$, is defined as

$$\sum_{\substack{:\sim\ F[w@l]\in\Pi_2,\\ I\models F}} w.$$

For any two stable models I and I' of Π, we say I is *dominated* by I' if

- there is some nonnegative integer l such that $\mathit{Penalty}_\Pi(I',l) < \mathit{Penalty}_\Pi(I,l)$ and
- for all integers $k > l$, $\mathit{Penalty}_\Pi(I',k) = \mathit{Penalty}_\Pi(I,k)$.

A stable model of Π is called *optimal* if it is not dominated by another stable model of Π.

3.2 Turning LP$^{\mathrm{MLN}}$ into ASP: Reward Way

In (Balai and Gelfond 2016) it is shown that LP$^{\mathrm{MLN}}$ programs can be turned into P-log. In this section, we show that using a similar translation, it is even possible to turn LP$^{\mathrm{MLN}}$ programs into answer set programs.

We turn each (possibly non-ground) rule

$$w_i: \quad \mathit{Head}_i \leftarrow \mathit{Body}_i$$

(i is the index of the rule) in an LP$^{\mathrm{MLN}}$ program Π into ASP rules

$$\begin{array}{rcl} \mathtt{sat}(i,w_i,\mathbf{x}) & \leftarrow & \mathit{Head}_i \\ \mathtt{sat}(i,w_i,\mathbf{x}) & \leftarrow & \mathtt{not}\ \mathit{Body}_i \\ \mathit{Head}_i & \leftarrow & \mathit{Body}_i,\ \mathtt{not\ not}\ \mathtt{sat}(i,w_i,\mathbf{x}) \\ & :\sim & \mathtt{sat}(i,w_i,\mathbf{x}).\ \ [-w_i'@l,i,\mathbf{x}] \end{array} \tag{1}$$

where (i) $\mathbf{x}$ is the list of global variables in the rule; (ii) $w_i' = 1$ and $l = 1$ if w_i is α; and (iii) $w_i' = w_i$ and $l = 0$ otherwise.[1]

Intuitively, a ground `sat` atom is true if the corresponding ground rule in the original program is true, and for each true `sat` atom, the weak constraint imposes a reward (negative penalty) w_i.

By lpmln2asp$^{\mathrm{rwd}}(\Pi)$ we denote the resulting ASP program containing weak constraints.

Theorem 2
For any LP$^{\mathrm{MLN}}$ program Π, there is a 1-1 correspondence ϕ between SM$[\Pi]$ [2] and the set of stable

[1] CLINGO restricts the weights to be integers only. To implement the translation using CLINGO, we need to turn w_i' into an integer by multiplying some factor.
[2] Recall the definition in Section 2.1.

models of lpmln2asp$^{\text{rwd}}(\Pi)$, where $\phi(I) = I \cup \{sat(i) \mid w_i : Head_i \leftarrow Body_i \in \Pi, I \models Body_i \rightarrow Head_i\}$. Furthermore,

$$W_\Pi(I) = exp\left(\sum_{\texttt{sat}(i,w_i,\mathbf{c})\in\phi(I)} w_i\right). \tag{2}$$

While the translation is simple and modular, there are a few problems with using this translation to compute LP$^{\text{MLN}}$ using ASP solvers. First, the translation does not necessarily yield a program that is acceptable in CLINGO and requires a further translation. In particular, the first and the second rules of (1) may not be in the syntax of CLINGO. (The third rule contains double negations, which are allowed in CLINGO from version 4.) Second, more seriously, when we translate non-ground LP$^{\text{MLN}}$ rules into the input language of ASP solvers, the second rule of (1) may be unsafe, so CLINGO cannot ground the program. An alternative translation in the next section avoids these problems by basing on the penalty-based concept of weights.

3.3 Turning LP$^{\text{MLN}}$ *into ASP: Penalty Way*

Based on the reformulation of LP$^{\text{MLN}}$ in Section 2.2, we introduce another translation that turns LP$^{\text{MLN}}$ programs into ASP programs. The translation ensures that a safe LP$^{\text{MLN}}$ program is always turned into a safe ASP program, and the resulting program is readily acceptable in the input to CLINGO.[3]

We define the translation lpmln2asp$^{\text{pnt}}(\Pi)$ by translating each (possibly non-ground) rule

$$w_i \quad Head_i \leftarrow Body_i$$

in an LP$^{\text{MLN}}$ program Π into

$$\begin{array}{rcl} \texttt{unsat}(i,w_i,\mathbf{x}) & \leftarrow & Body_i,\ \texttt{not}\ Head_i \\ Head_i & \leftarrow & Body_i,\ \texttt{not}\ \texttt{unsat}(i,w_i,\mathbf{x}) \\ & :\sim & \texttt{unsat}(i,w_i,\mathbf{x}).\quad [w_i'@l,i,\mathbf{x}] \end{array} \tag{3}$$

where (i) $\mathbf{x}$ is the list of global variables in the rule; (ii) $w_i' = 1$ and $l = 1$ if w_i is α; and (iii) $w_i' = w_i$ and $l = 0$ otherwise.[4]

Intuitively, the first rule of (3) makes atom $\texttt{unsat}(i,w_i,\mathbf{x})$ true when the i-th rule in the original program is not satisfied. In that case, the second rule is not effective, and w_i is imposed on the penalty of the stable model. On the other hand, if the i-th rule is satisfied, atom $\texttt{unsat}(i,w_i,\mathbf{x})$ is false, the rule $Head_i \leftarrow Body_i$ is effective, and the penalty is not imposed.

The following theorem is an extension of Corollary 2 from (Lee and Yang 2017) to allow non-ground programs and to consider the correspondence between all stable models, not only the most probable ones.

Theorem 3
For any LP$^{\text{MLN}}$ program Π, there is a 1-1 correspondence ϕ between SM$[\Pi]$ and the set of stable models of lpmln2asp$^{\text{pnt}}(\Pi)$, where $\phi(I) = I \cup \{\texttt{unsat}(i) \mid w_i : Head_i \leftarrow Body_i \in \Pi, I \models Body_i, I \not\models Head_i\}$. Furthermore,

$$W_\Pi^{\text{pnt}}(I) = exp\left(-\sum_{\texttt{unsat}(i,w_i,\mathbf{c})\in\phi(I)} w_i\right). \tag{4}$$

[3] An LP$^{\text{MLN}}$ program Π is *safe* if its unweighted program $\overline{\Pi}$ is safe as defined in (Calimeri et al. 2013).
[4] In the case $Head_i$ is a disjunction $l_1;\ldots;l_n$, expression $\texttt{not}\ Head_i$ stands for $\texttt{not}\ l_1,\ldots,\texttt{not}\ l_n$.

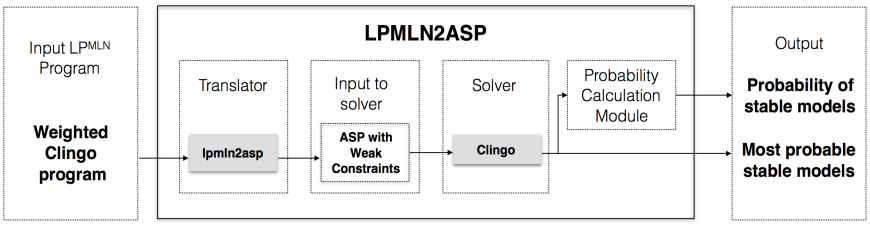

Fig. 1. Architecture of System LPMLN2ASP

Theorem 3, in conjunction with Theorem 1, provides a way to compute the probability of a stable model of an LP$^{\text{MLN}}$ program by examining the `unsat` atoms satisfied by the corresponding stable model of the translated ASP program.

3.4 System LPMLN2ASP

System LPMLN2ASP is an implementation of LP$^{\text{MLN}}$ based on the result in Section 3.3. It can be used for computing the probabilities of stable models, marginal/conditional probability of a query, as well as the most probable stable models.

In the input language of LPMLN2ASP, a soft rule is written in the form

$$w_i \;\; Head_i \leftarrow Body_i \tag{5}$$

where w_i is a real number in decimal notation, and $Head_i \leftarrow Body_i$ is a CLINGO rule. A hard rule is written without weights and is identical to a CLINGO rule.

The syntax of the command line is

```
lpmln2asp -i <input file> [-r <output file>] [-e <evidence file>]
                [-q <query atoms>] [-hr] [-all] [-clingo <clingo options>]
```

which follows the ALCHEMY command line syntax.

By default, the system finds a most probable stable model by leveraging CLINGO's built-in optimization method for weak constraints. The system computes the probabilities of query atoms specified by the option `-q`. In the presence of the `-e` option, the system computes the conditional probabilities of the query atoms given the evidence specified in the evidence file. The option `-all` instructs the system to enumerate all stable models and their probabilities.

Note that the probability computation involves enumerating all stable models so that it can be much more computationally expensive than the default MAP (Maximum A Posteriori) inference. On the other hand, the computation is exact, so compared to an approximate inference, the "gold standard" result is easy to understand. Also, conditional probability is more effectively computed than marginal probability because CLINGO effectively prunes many answer sets that do not satisfy the conditions specified in the evidence file.

The "Bird" example from (Lee and Wang 2016) can be represented in the input language of LPMLN2ASP as follows. The first three rules represent definite knowledge while the last two rules represent uncertain knowledge with different confidence.

```
% bird.lpmln
bird(X) :- residentbird(X).
bird(X) :- migratorybird(X).
:- residentbird(X), migratorybird(X).
```

```
2 residentbird(jo).
1 migratorybird(jo).
```

- For the command

  ```
  lpmln2asp -i bird.lpmln
  ```

  the system prints out a most probable stable model (i.e., by default it finds the MAP estimate). Appending `-all` prints out all stable models with their probabilities.
- For the command

  ```
  lpmln2asp -i bird.lpmln -e evid1.db -q residentbird
  ```

  with the "evid1.db" file containing the evidence

  ```
  :- not bird(jo).
  ```

  the system prints out the conditional probability P(`residentbird(X)` | `bird(jo)`).

By default, LPMLN2ASP does not translate hard rules. According to Proposition 2 from (Lee and Wang 2016), as long as the LP$^{\text{MLN}}$ program has a probabilistic stable model that satisfies all hard rules, the simplified translation gives the same result as the full translation and is more computationally efficient. Since in many cases hard rules represent definite knowledge that should not be violated, this is desirable.

On the other hand, translating hard rules could be relevant in some other cases, such as debugging an answer set program by finding which rules cause inconsistency. For example, consider a CLINGO input program `bird.lp`, that is similar to `bird.lpmln` but dropping the weights in the last two rules. CLINGO finds no stable models for this program. However, if we invoke LPMLN2ASP on the same program as

```
lpmln2asp -i bird.lp -hr -all
```

(`-hr` is the option to instruct the system to translate hard rules as well), the output of `lpmln2asp` shows three probabilistic stable models. Each of them shows a way to resolve the inconsistency by dropping one of the last three rules. For instance, one of the stable models is {`bird(jo)`,`residentbird(jo)`}, which disregards the last rule.

3.5 Computing MLN with LPMLN2ASP

A typical example in the MLN literature is a social networks domain that describes how smokers influence other people, which can be represented in LP$^{\text{MLN}}$ as follows. We assume three people *alice*, *bob*, and *carol*, and assume that *alice* is a smoker, *alice* influences *bob*, *bob* influences *carol*, and nothing else is known.

$$\begin{array}{l} w: \ \ smoke(x) \wedge influence(x,y) \rightarrow smoke(y) \\ \alpha: smoke(alice) \qquad \alpha: influence(alice,bob) \qquad \alpha: influence(bob,carol). \end{array} \tag{6}$$

(w is a positive number.) One may expect *bob* is less likely a smoker than *alice*, and *carol* is less likely a smoker than *bob*. Indeed, under the LP$^{\text{MLN}}$ semantics, $P(smoke(alice)) = 1$ and one can compute

$$P(smoke(bob)) = \frac{e^{8w}+e^{9w}}{2e^{8w}+e^{9w}} > P(smoke(carol)) = \frac{e^{9w}}{2e^{8w}+e^{9w}}.$$

This can be verified by LPMLN2ASP. For $w = 1$, the input program `smoke.lpmln` is

```
1 smoke(Y) :- smoke(X), influence(X, Y).

smoke(alice).
influence(alice, bob).
influence(bob, carol).
```

Executing `lpmln2asp -i smoke.lpmln -q smoke` outputs

```
smoke(alice) 1.00000000000000
smoke(bob) 0.788058442382915
smoke(carol) 0.576116884765829
```

as expected.

On the other hand, if (6) is understood under the MLN semantics (assuming `influence` relation is fixed as before), one can compute

$$P(smoke(bob)) = \frac{e^{8w} + e^{9w}}{3e^{8w} + e^{9w}} = P(smoke(carol)).$$

In other words, the degraded probability along the transitive relation does not hold under the MLN semantics. This is related to the fact that Markov logic cannot express the concept of transitive closure correctly as it inherits the FOL semantics.

According to Theorem 2 in (Lee and Wang 2016), MLN can be easily embedded in LP$^{\text{MLN}}$ by adding a choice rule for each atom with an arbitrary weight, similar to the way propositional logic can be embedded in ASP using choice rules. Consequently, it is possible to use system LPMLN2ASP to compute MLN, which is essentially using an ASP solver to compute MLN.

Let `smoke.mln` be the resulting program. Executing `lpmln2asp -i smoke.mln -q smoke` outputs

```
smoke(alice) 1.0
smoke(bob) 0.650244590946
smoke(carol) 0.650244590946
```

which agrees with the computation above.

4 Turning Tight LP$^{\text{MLN}}$ into MLN

4.1 Translation

Extending Theorem 3 from (Lee and Wang 2016) to non-ground programs, LP$^{\text{MLN}}$ programs (possibly containing variables) can be turned into MLN instances by first rewriting each rule into Clark normal form

$$w: \quad p(\mathbf{x}) \leftarrow Body$$

using equality in the body, where $\mathbf{x}$ is a list of distinct variables, and then adding the completion formulas

$$\alpha: \quad p(\mathbf{x}) \rightarrow \bigvee_{w:\ p(\mathbf{x}) \leftarrow Body\ \in\ \Pi} Body \tag{7}$$

for each atom $p(\mathbf{x})$. In fact, since the built-in algorithm in ALCHEMY for clausifying the completion formulas may yield an exponential blow-up, LPMLN2MLN implements an equivalent rewriting known as Tseytin transformation,[5] which introduces auxiliary atoms for each disjunctive term

[5] https://en.wikipedia.org/wiki/Tseytin_transformation

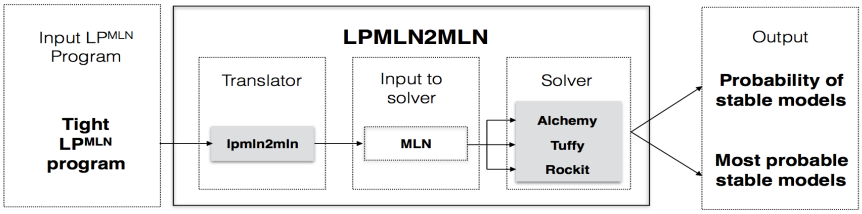

Fig. 2. Architecture of System LPMLN2MLN

in (7). The resulting MLN instance (possibly containing variables) is fed into ALCHEMY, which grounds the MLN and performs probabilistic inference on the ground network.[6]

The following proposition justifies the equivalent rewriting using auxiliary atoms.

Proposition 1
For any MLN $\mathbb{L}$ of signature σ, let $F(\mathbf{x})$ be a subformula of some formula in $\mathbb{L}$ where $\mathbf{x}$ is the list of all free variables of $F(\mathbf{x})$, and let $\mathbb{L}^F_{Aux}$ be the MLN program obtained from $\mathbb{L}$ by replacing $F(\mathbf{x})$ with a new predicate $Aux(\mathbf{x})$ and adding the formula

$$\alpha \ : \ \forall \mathbf{x}(Aux(\mathbf{x}) \leftrightarrow F(\mathbf{x})).$$

For any interpretation I of $\mathbb{L}$, let I_{Aux} be the extension of I of signature $\sigma \cup \{Aux\}$ defined by $I_{Aux}(Aux(\mathbf{c})) = (F(\mathbf{c}))^I$ for every list $\mathbf{c}$ of ground terms. We have

$$P_{\mathbb{L}}(I) = P_{\mathbb{L}^F_{Aux}}(I_{Aux}).$$

4.2 System LPMLN2MLN

The basic command of executing LPMLN2MLN is

```
lpmln2mln -i <input file> -r <output file> -q <query atoms> [-e <evidence file>]
     [-tuffy| -rockit| -alchemy] [-mln <options for mln solvers>]
```

which is similar to the command of executing LPMLN2ASP.

The syntax of the input language of LPMLN2MLN follows that of ALCHEMY, except that it uses a rule form. For example, consider again Example 1 in (Lee and Wang 2016). In the input language of LPMLN2MLN, it is encoded as

```
entity={Jo}

Bird(entity)
MigratoryBird(entity)
ResidentBird(entity)
```

```
Bird(x) <= ResidentBird(x).
Bird(x) <= MigratoryBird(x).
<= ResidentBird(x) ^ MigratoryBird(x).

2 ResidentBird(Jo)
1 MigratoryBird(Jo)
```

Executing

```
lpmln2mln -i bird.lpmln -r out -q Bird,ResidentBird,MigratoryBird
```

[6] Without introducing auxiliary atoms in the process of clausification, most examples cannot be run on ALCHEMY because the CNF conversion method implemented in ALCHEMY is naive.

gives[7]

```
Bird(Jo) 0.90296
ResidentBird(Jo) 0.667983
MigratoryBird(Jo) 0.235026
```

5 Comparison Between Two LPMLN Implementations

Both LPMLN2ASP and LPMLN2MLN can compute conditional/marginal probability, as well as finding the most probable stable models (MAP estimates). The implementations use ASP and MLN solvers as blackboxes, so their performance depends on the underlying solvers. Although ASP solvers do not have a built-in notion of probabilistic reasoning, it is interesting to note how the optimal answer set finding is related to MAP estimates in probabilistic reasoning. Grounding in ASP solvers is much more efficient than that in MLN solvers for the examples that we tested, but they have different characters. While grounding methods implemented in MLN solvers are not highly optimized, they do not ground the whole network; rather an essential part of a Markov network can be constructed from Markov blankets relevant to the query. Unlike LPMLN2ASP, LPMLN2MLN utilizes approximate sampling-based inference methods in underlying MLN solvers. Consequently, its solving is more scalable but gives less accurate results. Its input program is restricted to tight LP^{MLN} programs and does not support advanced ASP constructs, such as aggregates. When the domain is small, our experience is that it is much more convenient to work with LPMLN2ASP because it supports many useful ASP constructs and the exact computation yields outcomes that are easier to understand. Once we make sure the program is correct and we do not need advanced ASP constructs, we may use LPMLN2MLN for more scalable inference.

[7] When no MLN solver is specified in the command line, ALCHEMY is used by default.

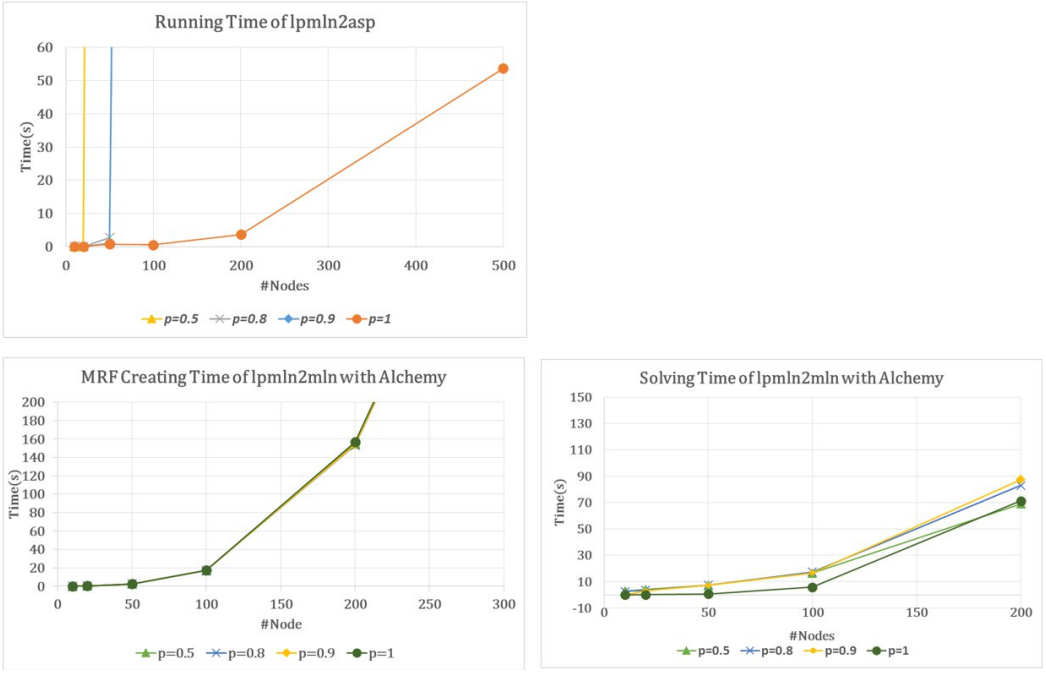

Fig. 3. Running Statistics on Finding Relaxed Clique

We report the running time statistics for both LPMLN2ASP and LPMLN2MLN on the example of finding a maximal "relaxed clique" in a graph, where the goal is to select as many nodes as possible while a penalty is assigned for each pair of disconnected nodes. The penalty assigned to disconnected nodes and the reward given to each node included in the subgraph define how much "relaxed" the clique is.

The LPMLN2ASP encoding of the relaxed clique example is

```
{in(X)} :- node(X).
disconnected(X, Y) :- in(X), in(Y), not edge(X, Y).
5  :- not in(X), node(X).
5  :- disconnected(X, Y).
```

The LPMLN2MLN encoding of the relaxed clique example is

```
{In(x)} <= Node(x).
Disconnected(x, y) <= In(x) ^ In(y) ^ !Edge(x, y).
5  <= !In(x) ^ Node(x)
5  <= Disconnected(x, y)
```

We use a Python script to generate a random graph with each edge generated with a fixed probability p. We experimented with $p = 0.5, 0.8, 0.9, 1$ and different numbers of nodes. For each problem instance, we perform MAP inference to find a maximal relaxed clique with both LPMLN2ASP and LPMLN2MLN. The timeout is 20 minutes. The experiments are performed on a machine powered by 4 Intel(R) Core(TM) i5-2400 CPU with OS Ubuntu 14.04.5 LTS and 8G memory.

Figure 3 shows running statistics of utilizing different underlying solvers. For LPMLN2ASP, grounding finishes almost instantly for all problem instances that we tested. We plot how solving times vary according to the number of nodes for different edge generation probabilities (top left graph). Roughly, solving time increases as the number of nodes increases. However, there is no clear correlation between solving time and the edge probability (i.e., the density of the graph). For $p = 0.5$, the LPMLN2ASP system first times out when $\#Nodes = 50$, while for both $p = 0.8$ and $p = 0.9$, it first times out when $\#Node = 100$. On the other hand, when $\#Node = 20$, solving time roughly increases as the edge probability increases except for $p = 0.5$. The running time is sensitive to particular problem instances, due to the exact optimization algorithm CDNL-OPT (Gebser et al. 2011) used by CLINGO, which only terminates when a true optimal solution is found. The non-deterministic nature of CDNL-OPT also brings randomness on the path through which an optimal solution is found, which makes the running time differ even among similar-sized problem instances, while in general, as the size of the graph increases, the search space gets larger, thus the solving time increases.

For LPMLN2MLN with ALCHEMY (bottom left and bottom right), grounding (MRF creating time) becomes the bottleneck. It increases much faster than solving time, and times out first when $\#Nodes = 500$. Again, the running time increases as the number of nodes increases. On the other hand, unlike LPMLN2ASP, ALCHEMY uses MaxWalkSAT for MAP inference, which allows a sub-optimal solution to be returned. The approximate nature of the method allows relatively consistent running times for different problem instances, as long as parameters such as the maximum number of iterations/tries are fixed among all experiments. The running times were not also much affected by the edge probability.

In general, LPMLN2MLN can be more scalable via parameter setting, while LPMLN2ASP grants better solution quality. LPMLN2MLN with TUFFY shows a similar behavior as LPMLN2MLN with ALCHEMY.

6 Using LP^{MLN} Implementations to Compute Other Languages

6.1 *Computing* PROBLOG

ProbLog (De Raedt et al. 2007) can be viewed as a special case of LP^{MLN} language (Lee and Wang 2016), in which soft rules are atomic facts only. The precise relation between the semantics of the two languages is stated in (Lee and Wang 2016). PROBLOG2 implements a native inference and learning algorithm which converts probabilistic inference problems into weighted model counting problems and then solves with knowledge compilation methods (Fierens et al. 2013). We compared the performance of LPMLN2ASP with that of PROBLOG2 on ProbLog input programs. We encoded the problem of reachability in a probabilistic graph in both languages, and performed MAP inference ("given that there is a path between two nodes, what is the most likely graph?") as well as marginal probability computation ("given two particular nodes, what is the probability that there exists a path between them?"). We used a Python script to generate edges with probabilities randomly assigned. For the probabilistic facts $p :: \texttt{edge}(n_1,n_2)$ $(0 < p < 1)$ in PROBLOG2, we write $ln(p/(1-p)) : \texttt{edge}(n_1,n_2)$ for LPMLN2ASP, which makes the probability of the edge being true to be p and being false to be $1-p$.

The path relation is defined in the input language of LPMLN2ASP as

```
path(X,Y) :- edge(X,Y).
path(X,Y) :- path(X,Z), path(Z, Y), Y != Z.
```

and in the input language of PROBLOG2 as

```
path(X,Y) :- edge(X,Y).
path(X,Y) :- path(X,Z), path(Z,Y), Y \== Z.
```

Parameter	MAP	
	LPMLN2CL (CLINGO 4.5)	PROBLOG2
$\#edges = 9$	0.013s	0.192s
$\#edges = 25$	0.021s	Timeout
$\#edges = 81$	0.308s	Timeout
$\#edges = 100$	0.756s	Timeout
$\#edges = 225$	6.121s	Timeout
$\#edges = 400$	29.706s	Timeout

Parameter	Marginal		
	LPMLN2ASP (CLINGO 4.5)	PROBLOG2 Default Setting	PROBLOG2 Sample Based $\#_{Sample} = 1000$
$\#edges = 9$	0.520s	0.137s	1.878s
$\#edges = 10$	0.676s	1.468s	1.656s
$\#edges = 11$	1.396s	1.480s	1.684s
$\#edges = 12$	2.524s	1.781s	1.672s
$\#edges = 13$	4.995s	Timeout	1.692s
$\#edges = 14$	9.744s	Timeout	1.796s
$\#edges = 15$	19.568s	Timeout	1.732s
$\#edges = 16$	38.476s	Timeout	1.748s
$\#edges = 18$	164.192s	Timeout	16.676s
$\#edges = 19$	306.000s	Timeout	4.564s
$\#edges = 20$	637.584s	Timeout	4.020s
$\#edges = 21$	Timeout	Timeout	4.344s

Fig. 4. Running Statistics on Reachability in a Probabilistic Graph

Figure 4 shows the running time of each experiment. LPMLN2ASP outperforms PROBLOG2 with the default setting (exact inference) in both MAP inference and marginal probability computation. However, both systems' marginal probability computations are not scalable because they enumerate all models. Using a sampling-based inference, PROBLOG2 was able to handle marginal probability computation effectively (the MAP inference in PROBLOG2 is exact inference only). In general, compared to running on tight programs, PROBLOG2 is slow for non-tight programs such as the program we use here. A possible reason is that it has to convert the input program, combined with the query, into weighted Boolean formulas, which is expensive for non-tight programs.

6.2 Reasoning about Probabilistic Causal Model

(Lee et al. 2015) shows how to represent Pearl's probabilistic causal model (Pearl 2000) by LP$^{\text{MLN}}$. Due to the acyclicity assumption on the causality, the LP$^{\text{MLN}}$ representation is tight, so we can use either implementation of LP$^{\text{MLN}}$ to compute probabilistic queries on a PCM, such as counterfactual queries. (Related to this, Section A.1 shows how Bayesian networks can be represented in LP$^{\text{MLN}}$.)

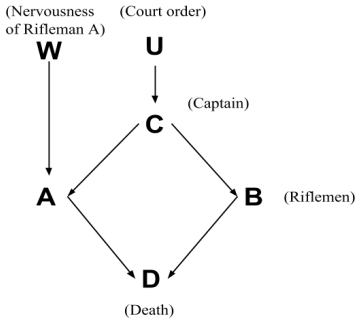

Fig. 5. Firing Squad Example

As an example, consider a probabilistic version of the firing squad example, shown in Figure 5. Court orders the execution (U) with probability p and Rifleman A is nervous (W) with probability q. The nervousness of Rifleman A causes him shooting at the prisoner (A). Court orders the execution causes the Captain to signal (C), which again causes Rifleman A and Rifleman B to shoot at the prisoner. Either of Rifleman A and Rifleman B shooting causes the prisoner's death (D). We illustrate how we use LP$^{\text{MLN}}$ systems to compute the counterfactual query "Given that the prisoner is dead, what is the probability that the prisoner would be alive if Rifleman A had not shot?" According to (Pearl 2000), the answer is $\frac{(1-p)q}{1-(1-p)(1-q)}$.

Theorem 4 from (Lee et al. 2015) states that the counterfactual reasoning in PCM can be reduced to LP$^{\text{MLN}}$ computation. The translation of PCM into LP$^{\text{MLN}}$ in Section 4.4 from (Lee et al. 2015) can be represented in the input language of LPMLN2ASP as follows, where `as`, `bs`, `cs`, `ds` are nodes in the twin network, `a1` means that `a` is true; `a0` means that `a` is false; other atoms are defined similarly.

```
@log(0.7/0.3)  u.                          a :- w.
@log(0.2/0.8)  w.                          b :- c.
                                           d :- a.
c :- u.                                    d :- b.
a :- c.

cs :- u, not do(c1), not do(c0).
as :- cs, not do(a1), not do(a0).          cs :- do(c1).
as :- w, not do(a1), not do(a0).           as :- do(a1).
bs :- cs, not do(b1), not do(b0).          bs :- do(b1).
ds :- as, not do(d1), not do(d0).          ds :- do(d1).
ds :- bs, not do(d1), not do(d0).
```

To represent the counterfactual query, the evidence file contains:

```
do(a0).
:- not d.
```

Note the different ways that intervention (`do(a0)`) and observation (`d`) is encoded in the query.

With the command `lpmln2asp -i pcm.lp -r out -e evid.db -q ds` we obtain `ds 0.921047297896`, which means there is a 8% chance that the prisoner would be alive.

7 Conclusion

We presented two implementations of LP$^{\text{MLN}}$ using ASP and MLN solvers. This is based on extending the translations that turn LP$^{\text{MLN}}$ into answer set programs and Markov logic to allow non-ground weighted rules. Although the input language of CLINGO does not have a built-in notion of

probabilistic reasoning, its optimal answer set finding algorithm is shown to be effective in finding MAP estimates (most probable stable models). It is also interesting that the efficient stable model enumeration leads to competitive exact probability computation.

The implementations also serve for other probabilistic logic languages that are shown to be embeddable in LP$^{\text{MLN}}$, such as ProbLog, Pearl's causal models, Bayesian networks, and P-log.

PrASP (Nickles 2016) is another system whose input language extends answer set programs with weights, although the semantics is different from that of LP$^{\text{MLN}}$. While LP$^{\text{MLN}}$ systems turn an input program into another input program that can be computed by existing systems, PrASP implements several native inference algorithms, including model counting, simulated annealing, flip-sampling, iterative refinement, etc. The formal relationships between the language of PrASP and other languages are not established.

The LP$^{\text{MLN}}$ implementations suggest how to combine the solving techniques from the two solvers. While CLINGO is efficient for grounding, MLN solvers consider subnetworks derived from the Markov blanket of query atoms and evidence. While CLINGO does exact inference only, MLN solvers can perform sampling based approximate inference. Future work includes building a native algorithm for LP$^{\text{MLN}}$ borrowing the techniques from the related systems.

References

BALAI, E. AND GELFOND, M. 2016. On the relationship between P-log and LP$^{\text{MLN}}$. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)*.

BARAL, C., GELFOND, M., AND RUSHTON, J. N. 2009. Probabilistic reasoning with answer sets. *TPLP 9,* 1, 57–144.

BUCCAFURRI, F., LEONE, N., AND RULLO, P. 2000. Enhancing disjunctive datalog by constraints. *Knowledge and Data Engineering, IEEE Transactions on 12,* 5, 845–860.

CALIMERI, F., FABER, W., GEBSER, M., IANNI, G., KAMINSKI, R., KRENNWALLNER, T., LEONE, N., RICCA, F., AND SCHAUB, T. 2013. Asp-core-2 input language format.

DE RAEDT, L., KIMMIG, A., AND TOIVONEN, H. 2007. ProbLog: A probabilistic prolog and its application in link discovery. In *IJCAI*. Vol. 7. 2462–2467.

FIERENS, D., VAN DEN BROECK, G., RENKENS, J., SHTERIONOV, D., GUTMANN, B., THON, I., JANSSENS, G., AND DE RAEDT, L. 2013. Inference and learning in probabilistic logic programs using weighted boolean formulas. *Theory and Practice of Logic Programming*, 1–44.

GEBSER, M., KAMINSKI, R., KAUFMANN, B., AND SCHAUB, T. 2011. Multi-criteria optimization in answer set programming. In *LIPIcs-Leibniz International Proceedings in Informatics*. Vol. 11. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

LEE, J., MENG, Y., AND WANG, Y. 2015. Markov logic style weighted rules under the stable model semantics. In Technical Communications of the 31st International Conference on Logic Programming.

LEE, J. AND WANG, Y. 2016. Weighted rules under the stable model semantics. In *Proceedings of International Conference on Principles of Knowledge Representation and Reasoning (KR)*.

LEE, J. AND YANG, Z. 2017. LPMLN, weak constraints, and p-log. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*.

NICKLES, M. 2016. A tool for probabilistic reasoning based on logic programming and first-order theories under stable model semantics. In *European Conference on Logics in Artificial Intelligence (JELIA)*. 369–384.

PEARL, J. 2000. *Causality: models, reasoning and inference*. Vol. 29. Cambridge Univ Press.

RICHARDSON, M. AND DOMINGOS, P. 2006. Markov logic networks. *Machine Learning 62,* 1-2, 107–136.

SANG, T., BEAME, P., AND KAUTZ, H. 2005. Solving bayesian networks by weighted model counting. In *Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI-05)*. Vol. 1. 475–482.

Appendix to "Computing LP$^{\text{MLN}}$ Using ASP and MLN Solvers"

Appendix A More about Using LP$^{\text{MLN}}$ to Compute Other Probabilistic Logic Languages

A.1 Bayesian Network in LP$^{\text{MLN}}$

It is easy to represent Bayesian network in LP$^{\text{MLN}}$ in a way similar to (Sang et al. 2005).

We assume all random variables are Boolean. Each conditional probability table associated with the nodes can be represented by a set of probabilistic facts. For each CPT entry $P(V = \mathbf{t} \mid V_1 = S_1, \ldots, V_n = S_n) = p$ where $S_1, \ldots, S_n \in \{\mathbf{t}, \mathbf{f}\}$, we include a set of weighted facts

- $ln(p/(1-p))$: $PF(V, S_1, \ldots, S_n)$ if $0 < p < 1$;
- α : $PF(V, S_1, \ldots, S_n)$ if $p = 1$;
- α : $\leftarrow$ *not* $PF(V, S_1, \ldots, S_n)$ if $p = 0$.

For each node V whose parents are $V_1, \ldots, V_n$, each directed edge can be represented by rules

$$\alpha : V \leftarrow V_1^{S_1}, \ldots, V_n^{S_n}, PF(V, S_1, \ldots, S_n) \qquad (S_1, \ldots, S_n \in \{\mathbf{t}, \mathbf{f}\})$$

where $V_i^{S_i}$ is V_i if S_i is $\mathbf{t}$, and *not* V_i otherwise.

For example, in the firing example (Figure A 1), the CPT for the node "alarm" can be represented by

```
@log(0.5/0.5)      pf(a,t1f1).          @log(0.99/0.01)      pf(a,t0f1).
@log(0.85/0.15)    pf(a,t1f0).          @log(0.0001/0.0009)  pf(a,t0f0).
```

The directed edges can be represented by hard rules as follows:

```
tampering :- pf(t).

fire :- pf(f).

alarm :- tampering, fire, pf(a,t1f1).
alarm :- tampering, not fire, pf(a,t1f0).
alarm :- not tampering, fire, pf(a,t0f1).
alarm :- not tampering, not fire, pf(a,t0f0).
```

```
smoke :- fire, pf(s,f1).
smoke :- not fire, pf(s,f0).

leaving :- alarm, pf(l,a1).
leaving :- not alarm, pf (l,a0).

report :- leaving, pf(r,l1).
report :- not leaving, pf(r,l0).
```

Theorem 4
For any Bayesian network where every random variable is Boolean, for any interpretation I, the probability of I according to the Bayesian network semantics coincides with the probability of I for the translated LP$^{\text{MLN}}$ program.

Since Bayesian networks are base on directed acyclic graphs, LP$^{\text{MLN}}$ programs that represent them are always tight. So both LPMLN2ASP and LPMLN2MLN can be used to compute Bayesian networks.

- To compute $P(fire \mid alarm)$, one can invoke

  ```
  lpmln2mln -i fire-bayes.lpmln -e evid1.db -r output -q fire
  ```

  where `evid1.db` contains the line `alarm`.

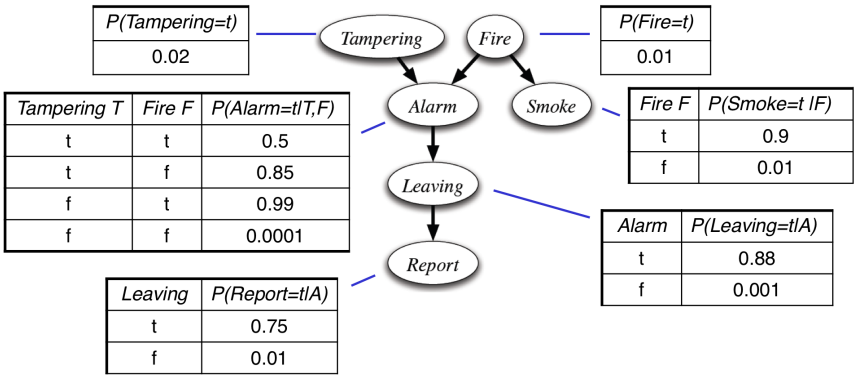

Fig. A 1. Bayes Net Example

- To compute $P(fire \mid alarm, \neg tampering)$, one can invoke

```
lpmln2mln -i fire-bayes.lpmln -e evid2.db -r output -q fire
```

where `evid1.db` contains two lines `alarm` and `!tampering`.

Appendix B Proof of Theorem 1

Theorem 1 *For any* LP$^{\text{MLN}}$ *program* Π *and any interpretation* I,

$$W_\Pi(I) \propto W_\Pi^{\text{pnt}}(I) \quad \text{and} \quad P_\Pi(I) = P_\Pi^{\text{pnt}}(I).$$

Proof

$$\begin{aligned}
W_{\mathbb{L}}(I) &= exp(\sum_{w\,:\,F \in \mathbb{L} \text{ and } I \models F} w) \\
&= exp(\sum_{w:F \in \mathbb{L}} w - \sum_{w\,:\,F \in \mathbb{L} \text{ and } I \not\models F} w) \\
&= exp(\sum_{w:F \in \mathbb{L}} w) \cdot exp(- \sum_{w\,:\,F \in \mathbb{L} \text{ and } I \not\models F} w) \\
&= TW_{\mathbb{L}} \cdot exp(- \sum_{w\,:\,F \in \mathbb{L} \text{ and } I \not\models F} w) \\
&= TW_{\mathbb{L}} \cdot W_{\mathbb{L}}^{\text{pnt}}(I)
\end{aligned}$$

Consequently,

$$\begin{aligned}
P_{\mathbb{L}}(I) &= \frac{W_{\mathbb{L}}(I)}{\sum_J W_{\mathbb{L}}(J)} \\
&= \frac{TW_{\mathbb{L}} \cdot W_{\mathbb{L}}^{\text{pnt}}(I)}{\sum_J TW_{\mathbb{L}} \cdot W_{\mathbb{L}}^{\text{pnt}}(J)} \\
&= \frac{W_{\mathbb{L}}^{\text{pnt}}(I)}{\sum_J W_{\mathbb{L}}^{\text{pnt}}(J)} \cdot \frac{\sum_J TW_{\mathbb{L}}}{\sum_J TW_{\mathbb{L}}} \\
&= \frac{W_{\mathbb{L}}^{\text{pnt}}(I)}{\sum_J W_{\mathbb{L}}^{\text{pnt}}(J)} \\
&= P_{\mathbb{L}}^{\text{pnt}}(I)
\end{aligned}$$

□

Appendix C Proof of Theorem 2

We divide lpmln2wc$^{\text{rwd}}(\Pi)$ into three parts:

$$\text{lpmln2wc}^{\text{rwd}}(\Pi) = SAT(\Pi) \cup ORIGIN(\Pi) \cup WC(\Pi)$$

where

$$\begin{aligned}
SAT(\Pi) = &\{\texttt{sat}(i) \leftarrow Head_i \mid w_i \quad Head_i \leftarrow Body_i \in \Pi\} \cup \\
&\{\texttt{sat}(i) \leftarrow \texttt{not}\ Body_i \mid w_i \quad Head_i \leftarrow Body_i \in \Pi\}
\end{aligned}$$

$$ORIGIN(\Pi) = \{Head_i \leftarrow Body_i, \texttt{not not sat}(i) \mid w_i \quad Head_i \leftarrow Body_i \in \Pi\}$$

and

$$WC(\Pi) = \{:\sim \texttt{sat}(i).\ \ [-w_i@l] \mid w_i \quad Head_i \leftarrow Body_i \in \Pi\}$$

Lemma 1
Let Π be an LP$^{\text{MLN}}$ program. $\phi(I) = I \cup \{sat(i) \mid w_i : Head_i \leftarrow Body_i \in \Pi, I \models Head_i \leftarrow Body_i\}$ is an $1-1$ correspondence between $SM[\Pi]$ and the stable models of $SAT(\Pi) \cup ORIGIN(\Pi)$.

Proof
Let σ be the signature of Π, and σ_{sat} be the set $\{sat(i) \mid w_i \quad Head_i \leftarrow Body_i \in \Pi\}$.

Let I be a member of $SM[\Pi]$. It can be seen that

- each strongly connected component of the dependency graph of $SAT(\Pi) \cup ORIGIN(\Pi)$ w.r.t. $\sigma \cup \sigma_{sat}$ is a subset of σ or a subset of σ_{sat};
- no atom in σ_{sat} has a strictly positive occurrence in $ORIGIN(\Pi)$;
- no atom in σ has a strictly positive occurrence in $SAT(\Pi)$.

so, according to splitting theorem, $\phi(I)$ is a stable model of $SAT(\Pi) \cup ORIGIN(\Pi)$ if and only if $\phi(I)$ is a stable model of $SAT(\Pi)$ w.r.t. σ_{sat} and is a stable model of $ORIGIN(\Pi)$ w.r.t. σ.

- $\phi(I)$ **is a stable model of** $SAT(\Pi)$ **w.r.t.** σ_{sat}**.** By the definition of ϕ, $\texttt{sat}(i) \in \phi(I)$ if and only if $I \models Head_i \leftarrow Body_i$, in which case either $I \models Head_i$ or $I \not\models Body_i$. This means, $I \models SAT(\Pi) \cup \{\texttt{sat}(i) \rightarrow Head_1 \vee Body_i \mid w_i : Head_i \leftarrow not\ \ Body_i \in \Pi\}$, which is the completion of $SAT(\Pi)$. It is obvious that $SAT(\Pi)$ is tight since every rule has its head not occurring in its body. So $\phi(I)$ must be a stable model of $SAT(\Pi)$ w.r.t. σ_{sat}.

- $\phi(I)$ **is a stable model of** $ORIGIN(\Pi)$ **w.r.t.** σ**.** Since I is a model of $\overline{\Pi_I}$, and $\mathtt{sat}(i) \in \phi(I)$ if and only if $I \vDash Head_i \leftarrow Body_i$, $\phi(I)$ is a model of $ORIGIN(\Pi)$. Next we show that $\phi(I)$ satisfies the loop formula of $ORIGIN(\Pi)$. Let L be any subset of σ that $\phi(I)$ satisfies. Since I is a stable model of $\overline{\Pi_I}$, we have

$$I \vDash LF_{\overline{\Pi_I}}(L)$$

i.e.,

$$I \vDash L^{\wedge} \rightarrow \bigvee_{\substack{Head_i \cap L \neq \emptyset, \\ Head_i \leftarrow Body_i \in \overline{\Pi_I}, \\ Body_i \cap L = \emptyset}} (Body_i \bigwedge_{b \in Head_i \setminus L} \neg b)$$

Since $I \vDash L$, $I \vDash Head_i$ for all $Head_i \cap L \neq \emptyset$, and thus $\phi(I) \vDash sat(i)$, so we have

$$\phi(I) \vDash L^{\wedge} \rightarrow \bigvee_{\substack{Head_i \cap L \neq \emptyset, \\ Head_i \leftarrow Body_i \in \overline{\Pi_I}, \\ Body_i \cap L = \emptyset}} (Body_i \wedge \neg\neg sat(i) \bigwedge_{b \in Head_i \setminus L} \neg b)$$

Since $\{Head_i \leftarrow Body_i, not\ not\ sat(i) \mid Head_i \leftarrow Body_i \in \overline{\Pi_i}\}$ is a subset of $ORIGIN(\Pi)$, we have

$$\phi(I) \vDash L^{\wedge} \rightarrow \bigvee_{\substack{Head_i \cap L \neq \emptyset, \\ Head_i \leftarrow Body_i, \\ not\ not\ sat(i) \in ORIGIN(\Pi), \\ Body_i \cap L = \emptyset}} (Body_i \wedge \neg\neg sat(i) \bigwedge_{b \in Head_i \setminus L} \neg b)$$

i.e.,

$$\phi(I) \vDash L^{\wedge} \rightarrow LF_{ORIGIN(\Pi)}(L)$$

So $\phi(I)$ is a stable model of $SAT(\Pi) \cup ORIGIN(\Pi)$.

Let $\phi(I)$ be a stable model of $SAT(\Pi) \cup ORIGIN(\Pi)$. By splitting theorem, $\phi(I)$ is a stable model of $ORIGIN(\Pi)$ w.r.t. σ. So $\phi(I)$ satisfies $ORIGIN(\Pi)$. It is easy to see that this imples $I \vDash \overline{\Pi_I}$. Consider any subset L of *sigma* that is satisified by I. Since $\phi(I)$ is a stable model of $ORIGIN(\Pi)$, we have

$$\phi(I) \vDash LF_{ORIGIN(\Pi)}(L)$$

, i.e.,

$$\phi(I) \vDash L^{\wedge} \rightarrow \bigvee_{\substack{Head_i \cap L \neq \emptyset, \\ Head_i \leftarrow Body_i, \\ not\ not\ sat(i) \in ORIGIN(\Pi), \\ Body_i \cap L = \emptyset}} (Body_i \wedge \neg\neg sat(i) \bigwedge_{b \in Head_i \setminus L} \neg b)$$

All $Head_i \leftarrow Body_i, not\ not\ sat(i)$ such that $Head_i \cap L \neq \emptyset$ are satisfied by $\phi(I)$ and thus $Head_i \leftarrow Body_i \in \overline{\Pi_I}$. So we have

$$\phi(I) \vDash L^{\wedge} \rightarrow \bigvee_{Head_i \cap L \neq \emptyset, Head_i \leftarrow Body_i, \in \overline{\Pi_I}, Body_i \cap L = \emptyset} (Body_i \wedge \neg\neg sat(i) \bigwedge_{b \in Head_i \setminus L} \neg b)$$

Since $\phi(I) \vDash sat(i)$ for all $Head_i \cap L \neq \emptyset$ (due to that $\phi(I) \vDash L$), they can be removed from the above formulas, resulting in

$$\phi(I) \vDash L^{\wedge} \rightarrow \bigvee_{Head_i \cap L \neq \emptyset, Head_i \leftarrow Body_i, \in \overline{\Pi_I}, Body_i \cap L = \emptyset} (Body_i \bigwedge_{b \in Head_i \setminus L} \neg b)$$

Since $\phi(I)$ and I agree on all atoms in σ, it can be further rewritten as

$$I \models L^{\wedge} \rightarrow \bigvee_{Head_i \cap L \neq \emptyset, Head_i \leftarrow Body_i, \in \overline{\Pi_I}, Body_i \cap L = \emptyset} (Body_i \bigwedge_{b \in Head_i \setminus L} \neg b)$$

which means

$$I \models LF_{\overline{\Pi_I}}(L)$$

So I is a stable model of $\overline{\Pi_I}$, and thus is a member of $SM[\Pi]$. $\square$

Theorem 2 *For any* LP$^{\text{MLN}}$ *program* Π*, there is a 1-1 correspondence* ϕ *between* SM$[\Pi]$ *and the set of stable models of* lpmln2asp$^{\text{rwd}}(\Pi)$*, where* $\phi(I) = I \cup \{sat(i) \mid w_i : Head_i \leftarrow Body_i \in \Pi, I \models Body_i \rightarrow Head_i\}$*. Furthermore,*

$$W_{\Pi}(I) = exp\left(\sum_{\texttt{sat}(i,w_i,\texttt{c}) \in \phi(I)} w_i\right). \tag{C1}$$

Proof

For any interpretation I of lpmln2wc$^{\text{rwd}}(\Pi)$, we use $Penalty_{\Pi}(I,l)$ to denote the total penalty it receives at level l defined by weak constraints:

$$Penalty_{\Pi}(I,l) = \sum_{:\sim sat(i).[-w_i@l] \in WC(\Pi), I \models sat(i)} -w_i$$

Let $\phi(I)$ be a stable model of lpmln2wc$^{\text{rwd}}(\Pi)$. By Lemma 1, $I \in SM[\Pi]$. So it is sufficient to prove

$$I \in \underset{J:J \in \underset{K:K \in SM[\Pi]}{argmax} W_{\Pi^{\text{hard}}}(K)}{argmax} W_{\Pi^{\text{soft}}}(J) \text{ iff}$$

$$\phi(I) \in \underset{J':J' \in \underset{K':K' \text{ is a stable model of } SAT(\Pi) \cup ORIGIN(\Pi)}{argmin} Penalty_{lpmln2wc^{rwd}}(K',1)}{argmin} Penalty_{\text{lpmln2wc}^{\text{rwd}}(\Pi)}(J',0)$$

which is equivalent to proving

$$I \in \underset{J:J \in \underset{K:K \in SM[\Pi]}{argmax} W_{\Pi^{\text{hard}}}(K)}{argmax} W_{\Pi^{\text{soft}}}(J) \text{ iff}$$

$$I \in \underset{J':J' \in \underset{K':K' \in SM[\Pi]}{argmin} Penalty_{\text{lpmln2wc}^{\text{rwd}}(\phi(K'),1)}}{argmin} Penalty_{\text{lpmln2wc}^{\text{rwd}}(\Pi)}(\phi(J'),0)$$

This is clear because

$$
\begin{aligned}
& \underset{J:\, J\in \underset{K:\, K\in \mathrm{SM}[\Pi]}{\operatorname{argmax}} W_{\Pi^{\mathrm{hard}}}(K)}{\operatorname{argmax}} W_{\Pi^{\mathrm{soft}}}(J) \\
= \quad & \text{(by Lemma 1 and definition)} \\
& \underset{J:\, J\in \underset{K:\, K \text{ is a stable model of } \mathrm{lpmln2wc}^{\mathrm{rwd}}(\Pi)}{\operatorname{argmax}} exp\Big(\sum_{\alpha: F\in (\Pi^{\mathrm{hard}})_K} \alpha\Big)}{\operatorname{argmax}} exp\Big(\sum_{w:F\in(\Pi^{\mathrm{soft}})_J} w\Big) \\
= & \\
& \underset{J:\, J\in \underset{K:\, K \text{ is a stable model of } \mathrm{lpmln2wc}^{\mathrm{rwd}}(\Pi)}{\operatorname{argmax}} exp\Big(\sum_{\alpha: F\in \Pi^{\mathrm{hard}}, K\models F} 1\Big)}{\operatorname{argmax}} exp\Big(\sum_{w:F\in \Pi^{\mathrm{soft}}, J\models F} w\Big) \\
= & \\
& \underset{J:\, J\in \underset{K:\, K \text{ is a stable model of } \mathrm{lpmln2wc}^{\mathrm{rwd}}(\Pi)}{\operatorname{argmin}} \Big(\sum_{\alpha: F\in \Pi^{\mathrm{hard}}, K\models F} -1\Big)}{\operatorname{argmin}} \Big(\sum_{w:F\in \Pi^{\mathrm{soft}}, J\models F} -w\Big) \\
= & \\
& \underset{J:\, J\in \underset{K:\, K \text{ is a stable model of } \mathrm{lpmln2wc}^{\mathrm{rwd}}(\Pi)}{\operatorname{argmin}} \Big(\sum_{:\sim F[-1@1]\in \mathrm{lpmln2wc}^{\mathrm{rwd}}(\Pi), K\models F} -1\Big)}{\operatorname{argmin}} \Big(\sum_{:\sim F[-w@0]\in \mathrm{lpmln2wc}^{\mathrm{rwd}}(\Pi), J\models F} -w\Big) \\
= & \\
& \underset{J:\, J\in \underset{K:\, K \text{ is a stable model of } \mathrm{lpmln2wc}^{\mathrm{rwd}}(\Pi)}{\operatorname{argmin}} Penalty_{\mathrm{lpmln2wc}(\Pi)}(K,1)}{\operatorname{argmin}} Penalty_{\mathrm{lpmln2wc}^{\mathrm{rwd}}(\Pi)}(J,0).
\end{aligned}
$$

The fact

$$W_\Pi(I) = exp\left(\sum_{\mathtt{sat}(i,w_i,\mathbf{c})\in\phi(I)} w_i\right). \tag{C2}$$

can be easily seen from how $\phi(I)$ is defined.

□

Appendix D Proof of Theorem 1

Theorem 1 *For any* LP$^{\mathrm{MLN}}$ *program* Π *and any interpretation* I,

$$W_\Pi(I) = W_\Pi^{\mathrm{pnt}}(I) \times TW_\Pi \quad \textit{and} \quad P_\Pi(I) = P_\Pi^{\mathrm{pnt}}(I).$$

where

$$TW_\Pi = exp\left(\sum_{w:F\in\Pi} w\right)$$

Proof

We first show that $W_\Pi(I) = TW_\Pi \cdot W_\Pi^{\mathrm{pnt}}(I)$. This is obviously true when $I \notin SM[\Pi]$. When $I \in$

$SM[\Pi]$, we have

$$\begin{aligned}
W_\Pi(I) &= exp(\sum_{w : F \in \Pi \text{ and } I \models F} w) \\
&= exp(\sum_{w:F\in\Pi} w - \sum_{w : F \in \Pi \text{ and } I \not\models F} w) \\
&= exp(\sum_{w:F\in\Pi} w) \cdot exp(-\sum_{w : F \in \Pi \text{ and } I \not\models F} w) \\
&= TW_\Pi \cdot exp(-\sum_{w : F \in \Pi \text{ and } I \not\models F} w) \\
&= TW_\Pi \cdot W_\Pi^{\text{pnt}}(I)
\end{aligned}$$

Consequently,

$$\begin{aligned}
P_\Pi(I) &= \frac{W_\Pi(I)}{\sum_J W_\Pi(J)} \\
&= \frac{TW_\Pi \cdot W_\Pi^{\text{pnt}}(I)}{\sum_J TW_\Pi \cdot W_\Pi^{\text{pnt}}(J)} \\
&= \frac{W_\Pi^{\text{pnt}}(I)}{\sum_J W_\Pi^{\text{pnt}}(J)} \cdot \frac{\sum_J TW_\Pi}{\sum_J TW_\mathbb{L}} \\
&= \frac{W_\Pi^{\text{pnt}}(I)}{\sum_J W_\Pi^{\text{pnt}}(J)} \\
&= P_\Pi^{\text{pnt}}(I)
\end{aligned}$$

□

Appendix E Proof of Proposition 1

Proposition 1 *For any MLN* $\mathbb{L}$ *of signature* σ*, let* $F(\mathbf{x})$ *be a subformula of some formula in* $\mathbb{L}$ *where* $\mathbf{x}$ *is the list of all free variables of* $F(\mathbf{x})$*, and let* $\mathbb{L}^F_{Aux}$ *be the MLN program obtained from* $\mathbb{L}$ *by replacing* $F(\mathbf{x})$ *with a new predicate* $Aux(\mathbf{x})$ *and adding the formula*

$$\alpha \quad : \quad \forall \mathbf{x}(Aux(\mathbf{x}) \leftrightarrow F(\mathbf{x})).$$

For any interpretation I *of* $\mathbb{L}$*, let* I_{Aux} *be the extension of* I *of signature* $\sigma \cup \{Aux\}$ *defined by* $I_{Aux}(Aux(\mathbf{c})) = (F(\mathbf{c}))^I$ *for every list* c *of element in the Herbrand universe. We have*

$$P_\mathbb{L}(I) = P_{\mathbb{L}^F_{Aux}}(I^F_{Aux})$$

Proof

For any formula G, let G^{Aux}_F be the formulas obtained from G by replacing subformulas $F(\mathbf{x})$ with $Aux_F(\mathbf{x})$. Since $I^{Aux}_F(Aux_F(\mathbf{c}))$ is defined as $I(F(\mathbf{c}))$ for every list c of element in the Herbrand universe, I^{Aux}_F satisfies $\forall(\mathbf{x}Aux_F(\mathbf{x}) \leftrightarrow F(\mathbf{x}))$. For a formula G_i, let n_i be the number of its ground

instances.

$$
\begin{aligned}
P_{\mathbb{L}_F^{Aux}}(I_F^{Aux}) &= \lim_{\alpha\to\infty} \frac{exp(\sum_{w_i:G_i\in\mathbb{L}_F^{Aux}} n_i w_i)}{\sum_J exp(\sum_{w_i:G_i\in\mathbb{L}_F^{Aux}} n_i w_i)} \\
&= \lim_{\alpha\to\infty} \frac{\alpha\cdot exp(\sum_{w_i:G_i\in\mathbb{L}} n_i w_i)}{\sum_{J\vDash\forall\mathbf{x}Aux_F(\mathbf{x})\leftrightarrow F(\mathbf{x})} exp(\sum_{G_i\in\mathbb{L}_F^{Aux}} n_i w_i) + \sum_{J\nvDash\forall\mathbf{x}Aux_F(\mathbf{x})\leftrightarrow F(\mathbf{x})} exp(\sum_{G_i\in\mathbb{L}_F^{Aux}} n_i w_i)} \\
&= \lim_{\alpha\to\infty} \frac{exp(\sum_{w_i:G_i\in\mathbb{L}} n_i w_i)}{\sum_{J\vDash\forall\mathbf{x}Aux_F(\mathbf{x})\leftrightarrow F(\mathbf{x})} exp(\sum_{w_i:G_i\in\mathbb{L}} n_i w_i) + \frac{1}{\alpha}\sum_{J\nvDash\forall\mathbf{x}Aux_F(\mathbf{x})\leftrightarrow F(\mathbf{x})} exp(\sum_{w_i:G_i\in\mathbb{L}} n_i w_i)} \\
&= \lim_{\alpha\to\infty} \frac{exp(\sum_{w_i:G_i\in\mathbb{L}} n_i w_i)}{\sum_{J\vDash\forall\mathbf{x}Aux_F(\mathbf{x})\leftrightarrow F(\mathbf{x})} exp(\sum_{w_i:G_i\in\mathbb{L}} n_i w_i)} \\
&= \lim_{\alpha\to\infty} \frac{exp(\sum_{G_i\in\mathbb{L}} n_i w_i)}{\sum_J exp(\sum_{G_i\in\mathbb{L}} n_i w_i)} \\
&= P_{\mathbb{L}}(I)
\end{aligned}
$$

□