Coherence and Compatibility of Markov Logic Networks

Matthias Thimm[1]

Abstract. Markov logic is a robust approach for probabilistic relational knowledge representation that uses a log-linear model of weighted first-order formulas for probabilistic reasoning. This log-linear model always exists but may not represent the knowledge engineer's intentions adequately. In this paper, we develop a general framework for measuring this *coherence* of Markov logic networks by comparing the resulting probabilities in the model with the weights given to the formulas. Our measure takes the interdependence of different formulas into account and analyzes the degree of impact they have on the probabilities of other formulas. This approach can be used by the knowledge engineer in constructing a well-formed Markov logic network if data for learning is not available. We also apply our approach to the problem of assessing the *compatibility* of multiple Markov Logic networks, i. e., to measure to what extent the merging of these networks results in a change of probabilities.

1 Introduction

Statistical relational learning [1] is a research area that deals with knowledge representation and learning in probabilistic first-order logics. Therein, a particularly popular approach is *Markov Logic* [15]. A *Markov logic network* (MLN) is a set of weighted first-order formulas where a larger weight means that the formula is more likely to be true. The semantics of an MLN is given via a log-linear model that takes the weights of formulas into account in order to determine probabilities for classical first-order interpretations. Markov logic networks have been used for e. g. diagnosis of bronchial carcinoma on ion mobility spectrometry data [3] or social network analysis [2].

In knowledge representation and reasoning *consistency* is a crucial issue and in order to cope with inconsistency different formalisms use different techniques. For example, most belief revision approaches [6] have to maintain consistency by altering the represented information, and default logics and the like [14, 5] use a non-monotonic inference procedure that bypasses classical inconsistency. Still, even a default theory can be inconsistent in a non-classical sense if there are two complementary defaults present in the theory. In Markov logic, inconsistency is not an issue as every MLN has a well-defined log-linear model (ignoring MLNs that contain infinite weights on two contradictory formulas). Therefore, every MLN is consistent by definition. However, whether the log-linear model is meaningful and adequately represents the information in the network may not be necessary true. For example, when representing weighted formulas such as $(\text{sunny}, 5)$ and $(\text{rain}, -20)$ one would probably expect that at least $P(\text{sunny}) > P(\text{rain})$ for the log-linear model P of the whole MLN. However, this is not guaranteed as other formulas may interfere in the computation of the final probabilities. Furthermore, consider the two weighted formulas $(\psi, 10)$ and $(\neg\psi, 10)$. The log-linear model P of only these two formulas is well-defined and has $P(\psi) = 0.5$. It is questionable whether these probabilities are appropriate and whether it would not be more appropriate to define this set of formulas as inconsistent. In particular, computing a log-linear model P' of an "empty" knowledge base yields $P'(\psi) = 0.5$ as well. Therefore, from the semantical point of view, the difference between inconsistency (contradicting formulas) and ignorance (no knowledge at all) cannot be recognized. This also makes it hard to detect modeling errors, particularly in large knowledge bases.

In this paper, we introduce the notion of *coherence* for MLNs. Informally, an MLN is coherent if it is "adequately" represented by its log-linear model. We develop a general framework for coherence measurement that bases on a notion of distance between the log-linear model and the weights of the formulas of the MLN. This measure is able to identify the amount of interference between different formulas of the MLN and thus gives an estimation of whether inference based on the log-linear model might result in counterintuitive results. We discuss one particular application of our framework for merging multiple MLNs into a single one. This is a typical scenario when multiple (domain) experts have to share their knowledge in order to solve a more general task. When merging multiple MLNs, the formulas of one MLN might influence the probabilities previously determined by another MLN which might give unintuitive results. By comparing the coherence of the merged MLN with the coherence of the individual MLNs we define a notion of compatibility for the merging scenario. In summary, the contributions of this paper are as follows:

1. We introduce the notion of *coherence* as a measure for assessing the adequateness of the log-linear model of an MLN (Section 3).
2. We show that our measure satisfies several desirable properties such as monotonicity and independence of irrelevant information. We also present a methodology for using the notion of coherence for knowledge engineering (Section 4).
3. We apply the notion of coherence to the problem of merging multiple MLNs and show that our measure is able to identify incompatibilities (Section 5).
4. We briefly describe our implementation of the coherence measurement framework (Section 6).

Proofs of technical results have been omitted due to space restrictions but can be found in an online appendix[2].

2 Markov Logic Networks

Markov logic [15] is a statistical relational framework which combines Markov networks [13] with aspects of first-order logic. The Markov logic syntax complies with first-order logic without functions where each formula is quantified by an additional weight. Let

[1] Institute for Web Science and Technologies, University of Koblenz-Landau, Germany, thimm@uni-koblenz.de

[2] http://www.mthimm.de/misc/mln_ecai2014_proofs.pdf

Pred be a finite set of predicates, C a finite set of constants, V a set of variables, and $\mathcal{L}_C$ be the functor-free first-order language on Pred, C, and V. For what remains we assume Pred and V to be fixed.

Definition 1. A *Markov logic network* (MLN) L on $\mathcal{L}_C$ is a finite ordered set of tuples $L = \langle(\phi_1, g_1), \ldots, (\phi_n, g_n)\rangle$ with $\phi_1, \ldots, \phi_n \in \mathcal{L}_C$ and $g_1, \ldots, g_n \in \mathbb{R}$.

In contrast to the original literature on MLNs [15] we define an MLN to be an *ordered* set of tuples (ϕ_i, g_i) $(i = 1, \ldots, n)$. This order can be arbitrary and has no special meaning other than to enumerate the elements of an MLN in an unambiguous manner. Any set operation on an MLN is defined in the same way as without an explicit order.

Note, that the weights of an MLN L have no obvious probabilistic interpretation [4] and are interpreted relative to each other when defining the joint probability function for L (see below).

Example 1. We adopt the standard example [2] to illustrate the intuition behind MLNs. Define $L_{sm} = \langle(\phi_1, 0.7), (\phi_2, 2.3), (\phi_3, 1.5), (\phi_4, 1.1), (\phi_5, \infty)\rangle$ via

$$\phi_1 = \mathsf{friends}(X, Y) \wedge \mathsf{friends}(Y, Z) \Rightarrow \mathsf{friends}(X, Z)$$
$$\phi_2 = \neg(\exists Y : \mathsf{friends}(X, Y)) \Rightarrow \mathsf{smokes}(X)$$
$$\phi_3 = \mathsf{smokes}(X) \Rightarrow \mathsf{cancer}(X)$$
$$\phi_4 = \mathsf{friends}(X, Y) \Rightarrow (\mathsf{smokes}(X) \Leftrightarrow \mathsf{smokes}(Y))$$
$$\phi_5 = \mathsf{friends}(X, Y) \Leftrightarrow \mathsf{friends}(Y, X)$$

The above MLN models uncertain relationships of smoking habits and people. Formula ϕ_1 means that being friends is a transitive relation, ϕ_2 means that people without friends usually smoke, ϕ_3 that smoking causes cancer, ϕ_4 that friends have similar smoking habits, and ϕ_5 that being friends is a symmetric relation. The formula ϕ_5 has an infinite weight which results in ϕ_5 being a *hard* constraint that must be satisfied.

Semantics are given to an MLN L by grounding L appropriately in order to build a Markov net and its corresponding log-linear model. Let $\Omega(C)$ be the set of (Herbrand) interpretations for Pred and C. For $\phi \in \mathcal{L}_C$ let $\mathsf{gnd}_C(\phi)$ denote the set of ground instances of ϕ wrt. C. Let $\omega \in \Omega(C)$ and define $n_\phi^C(\omega) = |\{\phi' \in \mathsf{gnd}_C(\phi) \mid \omega \models \phi'\}|$. The term $n_\phi^C(\omega)$ denotes the number of instances of ϕ that are satisfied in ω. Then a probability function $P_{L,C} : \Omega(C) \to [0, 1]$ can be defined as

$$P_{L,C}(\omega) = \frac{1}{Z_C} \exp\left(\sum_{(\phi,g)\in L} n_\phi^C(\omega) g\right) \tag{1}$$

with

$$Z_C = \sum_{\omega\in\Omega(C)} \exp\left(\sum_{(\phi,g)\in L} n_\phi^C(\omega) g\right)$$

being a normalization constant and $\exp(x) = e^x$ is the exponential function with base e. By defining $P_{L,C}$ in this way, worlds that violate fewer instances of formulas are more probable than worlds that violate more instances (depending on the weights of the different formulas). Hence, the fundamental idea for MLNs is that first-order formulas are not handled as hard constraints. Instead, each formula is more or less softened depending on its weight. Hence, a possible world *may* violate a formula without necessarily receiving a zero probability. A formula's weight specifies how strong the formula is, i. e., how much the formula influences the probability of a satisfying world versus a violating world. This way, the weights of all formulas influence the determination of a possible world's probability in a complex manner. One clear advantage of this approach is that MLNs can directly handle contradictions in a knowledge base, since the (contradictory) formulas are weighted against each other.

The probability function $P_{L,C}$ can be extended to sentences (ground formulas) of $\mathcal{L}_C$ via

$$P_{L,C}(\phi) = \sum_{\omega\models\phi} P_{L,C}(\omega) \tag{2}$$

for ground $\phi \in \mathcal{L}_C$.

Determining the probability of a sentence ϕ using Equations (1) and (2) is merely manageable for very small sets of constants, but intractable for domains of a more realistic size. While $P_{L,C}(\phi)$ can be approximated using Markov chain Monte-Carlo methods (MCMC methods) performance might still be too slow in practice [15]. There are more sophisticated and efficient methods to perform approximated inference if ϕ is a conjunction of ground literals, cf. [15]. Also, approaches for lifted inference exploit symmetries in the graph models which can speed up performance quite impressively, see e. g. [7] for an overview.

3 Measuring Coherence

Representing knowledge using Markov Logic requires defining the weights for the qualitative parts of the knowledge. In [15] it is suggested that weights of formulas have to be learned from data. Nonetheless, in [2] and [4] a heuristic is discussed that determines weights of formulas from probabilities. There, an interpretation of the weight g of a formula ϕ is provided as the log-odd between a world where ϕ is true and a world where ϕ is false, other things being equal, i. e., given some probability $p \in [0, 1]$ and a formula ϕ the corresponding *Markov weight* $g_{p,\phi}$ of p is defined by

$$g_{p,\phi} = \ln \frac{p}{1-p} r_\phi \tag{3}$$

where $\ln x$ is the natural logarithm of x and r_ϕ is the ratio of the number of worlds not satisfying and the number of worlds satisfying some ground instance of ϕ[3], see also [4] for a discussion. The justification for this heuristic comes from the general observation that for a ground formula ϕ and an MLN $L = \langle(\phi, g_{p,\phi})\rangle$, one exactly obtains $P_{L,C}(\phi) = p$. Arguably, it is easier for an expert to express uncertainty in the truth of a formula in form of a probability instead of a weight on a logarithmic scale. When defining an MLN L in this way one has to be aware of the fact that the probabilistic model $P_{L,C}$ of L and a set of constants C may not completely reproduce those intended probabilities.

Example 2. Consider the MLN $L = \langle(A(X), 2), (A(c_1), -5)\rangle$ and $C = \{c_1, c_2, c_3\}$. Assume that the weights of the formulas of L have been defined using the schema of Markov weights, i. e., the probability of $A(X)$ is intended to be approximately 0.881 ($g_{0.881,A(X)} \approx 2$) and of $A(c_1)$ it is approximately 0.0067 ($g_{0.0067,A(X)} \approx -5$). However, we obtain $P_{L,C}(A(c_1)) = 0.041$ which matches neither probability.

[3] For example, it is $r_\phi = 1$ for a ground atom ϕ and $r_\phi = 0.75$, $r_\phi = 0.25$ for a disjunction resp. conjunction of ground atoms.

In contrast to other probabilistic logics such as classical probabilistic logic [12] or Bayes nets [13], weights in Markov Logic are not handled as constraints but as factors that influence the determination of probabilities. By accepting this behavior the observation made in Example 2 is understandable. However, due to this behavior it is hard to verify whether some formalization is adequate for a representation problem and whether it is *robust* with respect to extensions:

Example 3. Assume we want to model an MLN L such that its model gives a probability 0.5 for each instance $A(c_1), A(c_2), A(c_3)$. This can be achieved by modeling $L = \langle (A(X), -10), (A(X), 10) \rangle$ and $C = \{c_1, c_2, c_3\}$. Assume now we want to incorporate a new piece of information such that $P_{L,C}(A(c_1)) = 0.9$ but still $P_{L,C}(A(c_2)) = P_{L,C}(A(c_3)) = 0.5$. In order to realize this one has to add a new weighted formula $(A(c_1), g)$ to L with some weight g. Due to the interference with the other formulas g cannot easily be determined. This results from the inadequate modeling of the initial knowledge via the MLN L. In this case, the empty MLN would have been a better fit to represent the intended uniform probability distribution. Also, the extended MLN $L' = \langle (A(c_1), 2.2) \rangle$ ($2.2 \approx \ln(0.9/1-0.9)$) yields $P_{L,C}(A(c_1) \approx 0.9$ and $P_{L,C}(A(c_2) = P_{L,C}(A(c_3) = 0.5$.

In the rest of this section, we investigate the issue of assessing how well the probabilistic model $P_{L,C}$ of an MLN L and a set of constants C reflects the probabilities used for defining L. For that we employ the Markov weights as a comparison criterion, i. e., we compare the probability of every formula of L in the probabilistic model $P_{L,C}$ with the probability this formula would have in the probabilistic model $P_{L',C}$ of the MLN L' that only consists of this formula. Note that our approach could also be formulated using any other (surjective) function g'_p that assigns weights to probabilities.

Similarly as consistency is defined for classical logics we also define a strict version of coherence. In particular, we say that L is *perfectly coherent* wrt. C if $P_{L,C}$ assigns to each formula the same probability as prescribed by the Markov weights. More formally:

Definition 2. Let $L = \langle (\phi_1, g_1), \ldots, (\phi_n, g_n) \rangle$ be an MLN. We say that L is *perfectly coherent* if and only if for all $i = 1, \ldots, n$ and $\phi' \in \mathsf{gnd}_C(\phi_i)$ it holds $P_{L,C}(\phi') = p$ and $g_i = g_{p,\phi}$.

If $g = g_{p,\phi}$ is a Markov weight observe that

$$p = p_{g,\phi} = \frac{\exp(g)}{r_\phi + \exp(g)}$$

with $p_{g,\phi} = 1$ if $g = \infty$ and $p_{g,\phi} = 0$ if $g = -\infty$. We also call $p_{g,\phi}$ a *Markov probability*. Following the spirit of inconsistency measures for probabilistic logics [16] we take a more graded approach to coherence analysis and, consequently, in the following we will consider the problem of defining coherence values.

Before formalizing our coherence measurement framework we need some further notation. Let C be a set of constants and $\phi \in \mathcal{L}_C$. The *ground vector* of ϕ with respect to C is defined via $\mathsf{gnd}^{\rightarrow}_C(\phi) = \langle \phi_1, \ldots, \phi_n \rangle$ where $\mathsf{gnd}_C(\phi) = \{\phi_1, \ldots, \phi_n\}$ and $\phi_1, \ldots, \phi_n$ is some arbitrary but fixed canonical ordering of $\mathsf{gnd}_C(\phi)$. If $\langle \phi_1, \ldots, \phi_n \rangle \in \mathcal{L}^n_C$ is a vector of formulas and P a probability function then we write

$$P(\langle \phi_1, \ldots, \phi_n \rangle) = \langle P(\phi_1), \ldots, P(\phi_n) \rangle$$

As a central tool for measuring coherence we use (weak) distance measures.

Definition 3. Let $n \in \mathbb{N}^+$. A function $d : [0,1]^n \times [0,1]^n \to [0, \infty)$ is called a (weak) *distance measure* if it satisfies 1.) $d(\vec{x}, \vec{y}) = 0$ if and only if $\vec{x} = \vec{y}$ (*reflexivity*) and 2.) $d(\vec{x}, \vec{y}) = d(\vec{y}, \vec{x})$ (*symmetry*), for all $\vec{x}, \vec{y} \in \mathbb{R}^n$.

Note that weak distance measures differ from standard distance measures by not requiring the *triangle equality* to hold. In this work we consider the following distance measures (let $\vec{x} = \langle x_1, \ldots, x_n \rangle, \vec{y} = \langle y_1, \ldots, y_n \rangle \in [0,1]^n$, $p \in \mathbb{N}^+$): 1.) $d_p(\vec{x}, \vec{y}) = \sqrt[p]{|x_1 - y_1|^p + \ldots + |x_n - y_n|^p}$ (p-norm distance), 2.) $d_{p,0}(\vec{x}, \vec{y}) = \sqrt[p]{|x_1 - y_1|^p + \ldots + |x_n - y_n|^p} / \sqrt[p]{n}$ (normalized p-norm distance), 3.) $d_{\max}(\vec{x}, \vec{y}) = \max\{|x_1 - y_1|, \ldots, |x_n - y_n|\}$ (max-distance), 4.) $d_{\min}(\vec{x}, \vec{y}) = \min\{|x_1 - y_1|, \ldots, |x_n - y_n|\}$ (min-distance), and 5.) $d_{\text{avg}}(\vec{x}, \vec{y}) = (|x_1 - y_1| + \ldots + |x_n - y_n|)/n$ (average distance).

In the following, we will use distance measures to measure the differences between vectors of probabilities that arise for each formula of an MLN upon grounding and the corresponding expected probabilities. In order to aggregate the distances of each formula we use aggregation functions.

Definition 4. A function $\theta : [0,1]^n \to [0,1]$ is called an *aggregation function.*

We consider the following aggregation functions (let $\vec{x} = \langle x_1, \ldots, x_n \rangle \in [0,1]^n$): 1.) $\theta_{\max}(\vec{x}) = \max\{x_1, \ldots, x_n\}$ (maximum), 2.) $\theta_{\min}(\vec{x}) = \min\{x_1, \ldots, x_n\}$ (minimum), and 3.) $\theta_{\text{avg}}(\vec{x}) = (x_1 + \ldots + x_n)/n$ (average).

Using distance measures and aggregation functions we define the *coherence* of an MLN L as how well L reflects the probabilities that are intended to be modeled by weights.

Definition 5. Let d be a distance measure, θ an aggregation function, $L = \langle (\phi_1, g_1), \ldots, (\phi_n, g_n) \rangle$ an MLN, and C a set of constants. Then the *coherence* $\mathsf{coh}^{d,\theta}_C(L)$ of L wrt. C and given d, θ is defined via

$$\mathsf{coh}^{d,\theta}_C(L) = 1 - \theta\left(\left\langle d\left(P_{L,C}(\mathsf{gnd}^{\rightarrow}_C(\phi_i)), \Pi^{|\mathsf{gnd}^{\rightarrow}_C(\phi_i)|}_{\phi_i, g_i}\right)\right\rangle_{i=1,\ldots,n}\right) \quad (4)$$

with

$$\Pi^n_{\phi,g} = \underbrace{\langle p_{g,\phi}, \ldots, p_{g,\phi} \rangle}_{\text{n times}}$$

The intuition behind the above definition is as follows. Assume that $(\phi(X), g) \in L$ and that $\{\phi(c_1), \ldots, \phi(c_n)\}$ are the groundings of $\phi(X)$. Then $P_{L,C}$ assigns to each of this ground formulas a (potentially different) probability $P_{L,C}(\phi(c_i))$ ($i = 1, \ldots, n$). First, we compute the distance of the vector $\langle P_{L,C}(\phi(c_1), \ldots, P_{L,C}(\phi(c_n)) \rangle$ to the vector $\langle p_{g,\phi(X)}, \ldots, p_{g,\phi(X)} \rangle$ (the uniform vector of the probability corresponding to the weight g). Finally, we aggregate the distances of all these vectors for all formulas in L. Therefore, $\mathsf{coh}^{d,\theta}_C(L)$ provides an aggregated assessment of how close the actual probabilities match the weights.

As we are in a probabilistic framework, one might wonder why we use ordinary distance measures and aggregation functions for defining a measure of coherence. A seemingly better alternative should be e. g. the Kullback-Leibler divergence [8] which has a well-defined meaning when measuring the difference between two probability functions. However, in our setting we compare a probability function $P_{L,C}$ with a set of probabilities derived from the weights of the

MLN L. In particular, the latter is usually contradictory (unless L is perfectly coherent), so the meaning of the Kullback-Leibler divergence in this context is not clear. We leave this issue for future work and consider now the distance measures defined so far.

4 Analysis

To further illustrate the meaning of the Definition 5 let us consider the coherence measure $\mathsf{coh}_C^{d_{\max},\theta_{\max}}$ and an MLN $L = \langle(\phi_1, g_1), \ldots (\phi_n, g_n)\rangle$. Then $\mathsf{coh}_C^{d_{\max},\theta_{\max}}(L)$ is one minus the maximum deviation of the probability of some ground instance ϕ_i of L in $P_{L,C}$ to the probability p_i estimated by its weight g_i, assumed that g_i has been determined by setting $g_i = \ln \frac{p_i}{1-p_i} r_{\phi_i}$.

Example 4. Consider the MLN $L = \langle(A(X), 2)\rangle$ and $C = \{c_1, c_2, c_3\}$. Note that the probability p intended to be modeled by the weight 2 is $p = p_{2,A(X)} = \frac{\exp(2)}{1+exp(2)} \approx 0.881$ (note that $r_{A(X)} = 1$). As there is only one formula in L it also follows directly that $P_{L,C}(A(c_1)) = P_{L,C}(A(c_2)) = P_{L,C}(A(c_3)) \approx 0.881$ as well. It follows that $\mathsf{coh}_C^{d_{\max},\theta_{\max}}(L) = 1 - 0 = 1$.

Example 5. We continue Example 2 and consider the MLN $L = \langle(A(X), 2), (A(c_1), -5)\rangle$ and $C = \{c_1, c_2, c_3\}$. Note that the probability p_1 intended to be modeled by the weight 2 is $p_1 = p_{2,A(X)} = \frac{\exp(2)}{1+\exp(2)} \approx 0.881$ and for the weight -5 it is $p_2 = p_{-1,A(c_1)} = \frac{\exp(-5)}{1+\exp(-5)} \approx 0.0067$. For $P_{L,C}$ we obtain $P_{L,C}(A(c_1)) \approx 0.041$ and $P_{L,C}(A(c_2)) = P_{L,C}(A(c_3)) \approx 0.881$. Then $\mathsf{coh}_C^{d_{\max},\theta_{\max}}(L)$ computes to

$$\begin{aligned}\mathsf{coh}_C^{d_{\max},\theta_{\max}}(L) = 1 - \max\{&|P_{L,C}(A(c_1)) - p_1|,\\ &|P_{L,C}(A(c_2)) - p_1|, |P_{L,C}(A(c_3)) - p_1|,\\ &|P_{L,C}(A(c_1)) - p_2|\}\\ \approx 0.16&\end{aligned}$$

In the introduction we gave an example illustrating that MLNs are not always capable of differentiating between (logical) inconsistency and ignorance. However, using our notion of coherence we are able to detect this difference.

Example 6. Consider the MLN $L = \langle(A, -10), (A, 10)\rangle$ with a proposition (a predicate without parameters) A and $C = \{c_1, c_2, c_3\}$. The probabilities p_1, p_2 intended to modeled by the weights -10, and 10 are (respectively) $p_1 = p_{-10,A} \approx 0$ and $p_2 = p_{10,A} \approx 1$ and for $P_{L,C}$ we obtain $P_{L,C}(A) = 0.5$. Then we have

$$\begin{aligned}\mathsf{coh}_C^{d_{\max},\theta_{\max}}(L) &= 1 - \max\{|P_{L,C}(A) - p_1|, |P_{L,C}(A) - p_2|\}\\ &= 0.5\end{aligned}$$

Furthermore, for the empty MLN $L' = \langle\rangle$ and an arbitrary C we always have $\mathsf{coh}_C^{d,\theta}(L) = 1$ for any $d \in \{d_p, d_{p,0}, d_{\max}, d_{\min}, d_{\text{avg}}\}$ and $\theta \in \{\theta_{\max}, \theta_{\min}, \theta_{\text{avg}}\}$.

We now turn to the formal properties of $\mathsf{coh}_C^{d,\theta}$.

Proposition 1. *For $d \in \{d_{p,0}, d_{\max}, d_{\min}, d_{avg}\}$ and $\theta \in \{\theta_{\max}, \theta_{\min}, \theta_{avg}\}$ we have $\mathsf{coh}_C^{d,\theta}(L) \in [0, 1]$ for every L and C.*

The above proposition shows that many coherence measures are normalized on $[0, 1]$ and, therefore, different MLNs can be compared and categorized by their coherence values. Note that the Proposition 1 does not hold in general for d_p.

Proposition 2. *If d satisfies reflexivity and θ satisfies $\theta(x_1, \ldots, x_n) = 0$ iff $x_1 = \ldots = x_n = 0$ then $\mathsf{coh}_C^{d,\theta}(L) = 1$ iff L is perfectly coherent wrt. C.*

The above proposition states that our framework satisfies the basic property of detecting whether an MLN is perfectly coherent, given some minimal requirements of both distance measure and aggregation function.

Corollary 1. *If $d \in \{d_p, d_{p,0}, d_{\max}, d_{avg}\}$ ($p \in \mathbb{N}^+$) and $\theta \in \{\theta_{\max}, \theta_{avg}\}$ then $\mathsf{coh}_C^{d,\theta}(L) = 1$ iff L is perfectly coherent wrt. C.*

Note that the above statement does not hold for $d_{\min}$ and $\theta_{\min}$.

Next we look into the behavior of $\mathsf{coh}_C^{d,\theta}$ under changes of L and C.

Proposition 3. *For any d it holds $\mathsf{coh}_C^{d,\theta_{\max}}(L)$ is monotonically decreasing in L, i. e. $\mathsf{coh}_C^{d,\theta_{\max}}(L) \geq \mathsf{coh}_C^{d,\theta_{\max}}(L \cup \{(\phi, g)\})$.*

This property states that $\mathsf{coh}_C^{d,\theta_{\max}}(L)$ cannot get more coherent under addition of formulas. This corresponds to the classical concept of inconsistency insofar that an inconsistent knowledge base of classical logical formulas cannot get consistent when adding new information. Note that the above property does not hold in general for $\theta_{\min}$ and θ_{avg}. For a special case of a new formula we make the following observation.

Proposition 4. *For any d, if a consistent ϕ shares no predicate with L then $\mathsf{coh}_C^{d,\theta_{\max}}(L) = \mathsf{coh}_C^{d,\theta_{\max}}(L \cup \{(\phi, g)\})$ for every $g \in \mathbb{R}$.*

In other words, if we add totally unrelated (but consistent) information to an MLN this does not change its coherence.

Proposition 5. *For $\theta \in \{\theta_{\max}, \theta_{\min}, \theta_{avg}\}$ it holds that $\mathsf{coh}_C^{d_{\min},\theta}(L)$ is monotonically increasing in C, i. e. $\mathsf{coh}_C^{d_{\min},\theta}(L) \leq \mathsf{coh}_{C\cup\{c\}}^{d_{\min},\theta}(L)$.*

This result states that considering more individuals increases the coherence of the MLN wrt. $d_{\min}$. The rationality of satisfying this property is evident as by taking more individuals into account exceptions to formulas become negligible. Consider the MLN L of Example 5 which specifies a general rule ($A(X)$ has to hold in general) and an exception (c_1 does not satisfy $A(X)$). However, the general rule dominates the coherence value the the more individuals actually satisfy it.

Example 7. We continue Example 5 but consider varying sizes of the domain. So let $L = \langle(A(X), 2), (A(c_1), -5)\rangle$ and $C_i = \{c_1, \ldots, c_i\}$ for $i \in \mathbb{N}$. Figure 1 shows the behavior of four different coherence measures when the domain increases in size.

The framework proposed so far can be utilized by a knowledge engineer when debugging MLNs. In particular, a coherence measure can be used to evaluate whether the semantics of an MLN adequately represents its intended meaning if no data for learning is available. Note that this tool can be applied even if the heuristic for defining weights from probabilities may not seem adequate as the tool uses these only for assessing the influence one formula has on another.

Example 7 showed that, in particular, distance measures based on the p-norm may give a more fine-grained view on the evolution of coherence values (however, note that these distance measures do not satisfy monotonicity wrt. the domain in general). Independently of the actually chosen combination of distance measure and aggregation function, by utilizing the framework of coherence measurement for analyzing a given MLN the knowledge engineer is already able to detect several design flaws:

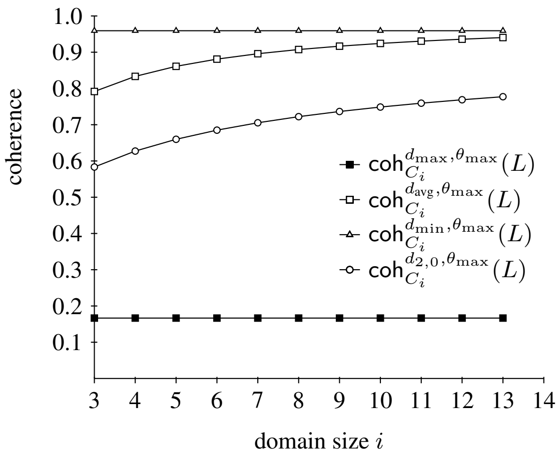

Figure 1. Coherence values for Example 7

1. If an MLN is coherent (i. e. has a comparatively large coherence value) but exhibits unintuitive inferences, then probably some weights have been chosen wrong (as there is only little interdependence between formulas).
2. If an MLN is coherent and exhibits no unintuitive inferences, then the MLN is a good representation of the given knowledge and it will probably be easier to extend it.
3. If an MLN is incoherent (i. e. has a comparatively low coherence value) and exhibits unintuitive inferences, then the knowledge engineer should have a look into the structure of the knowledge base as there may be unwanted interdependences amongst formulas.
4. If an MLN is incoherent but exhibits no unintuitive inferences, then the MLN may not be an adequate representation of the knowledge and further extensions might yield unintuitive results.

As a final remark, observe that our notion of coherence is also compatible with the usual notion of probabilistic consistency. In particular, starting from a consistent probabilistic view in form of a probability function, we can always find a perfectly coherent MLN representing this probability function.

Proposition 6. *Let $P : \Omega(C) \to [0,1]$ be any probability function. Then there is a perfectly coherent MLN L on $\mathcal{L}_C$ with $P_{L,C} = P$. In particular, it holds $\mathsf{coh}_C^{d,\theta}(L) = 1$ for any $d \in \{d_p, d_{p,0}, d_{\max}, d_{\min}, d_{avg}\}$ and $\theta \in \{\theta_{\max}, \theta_{\min}, \theta_{avg}\}$.*

As for every MLN L the probability function $P_{L,C}$ is always well-defined the above observation can also be used to transform an incoherent MLN L into a coherent MLN L' that weighs formulas more adequately. However, note that the formulas in L' need not necessarily to be the same as in L.

5 Application: Compatibility of MLNs

A particular use case for applying our framework arises when considering a knowledge merging scenario. Consider the case of multiple experts merging their knowledge in order to obtain a broader picture on some problem domain. Then, the individual pieces of information of each expert contribute to the overall probabilities obtained from the log-linear model of the merged MLN. Given that the experts have contradictory views on some parts of the modeled knowledge the merged MLN might not adequately reflect the joined knowledge. In order to analyze whether the merging of MLNs gives rise to a potentially meaningless joint MLN we employ our framework of coherence measurement as follows.

Definition 6. Let d be a distance measure, θ an aggregation function, $L_1, \ldots, L_m$ MLNs, and $C_1, \ldots, C_m$ sets of constants. The *compatibility* $\mathsf{comp}_{C_1,\ldots,C_m}^{d,\theta}(L_1, \ldots, L_m)$ of $L_1, \ldots, L_m$ wrt. $C_1, \ldots, C_m$ given d, θ is defined via

$$\mathsf{comp}_{C_1,\ldots,C_m}^{d,\theta}(L_1, \ldots, L_m) = \frac{1}{2}\left(1 + \mathsf{coh}_{C_1\cup\ldots\cup C_m}^{d,\theta}(L_1 \cup \ldots \cup L_m) - \frac{1}{m}\sum_{i=1}^{m} \mathsf{coh}_{C_i}^{d,\theta}(L_i)\right)$$

The value $\mathsf{comp}_{C_1,\ldots,C_m}^{d,\theta}(L_1, \ldots, L_m)$ describes how well the MLNs $L_1, \ldots, L_n$ can be merged. In essence, it measures how much the coherence of the joint MLN differs from the average coherence of all input MLNs. Intuitively, the larger the value of $\mathsf{comp}_{C_1,\ldots,C_m}^{d,\theta}(L_1, \ldots, L_m)$ the more *compatible* the MLNs should be. The exact form of the compatibility measure has been chosen like this to satisfy the *normalization property*, see Proposition 8 below.

Example 8. Consider the three MLNs $L_1 = \langle(\phi_1, 1.85), (\phi_2, 1.85)\rangle, L_2 = \langle(\phi_3, \infty)\rangle, L_3 = \langle(\phi_4, 1.1), (\phi_5, \infty)\rangle$ defined via

$$\phi_1 = \mathsf{quaker}(X) \Rightarrow \mathsf{pacifist}(X)$$
$$\phi_2 = \mathsf{republican}(X) \Rightarrow \neg\mathsf{pacifist}(X)$$
$$\phi_3 = \mathsf{republican}(nixon) \wedge \mathsf{quaker}(nixon) \wedge \mathsf{president}(nixon)$$
$$\phi_4 = \mathsf{president}(X) \Rightarrow \neg\mathsf{actor}(X)$$
$$\phi_5 = \mathsf{president}(reagan) \wedge \mathsf{actor}(reagan)$$

which model an extended version of the Nixon diamond. Using $\mathsf{coh}_C^{d_{\max},\theta_{\max}}$ we obtain

$$\mathsf{coh}_{\{d\}}^{d_{\max},\theta_{\max}}(L_1) \approx 0.982$$
$$\mathsf{coh}_{\{nixon\}}^{d_{\max},\theta_{\max}}(L_2) = 1$$
$$\mathsf{coh}_{\{reagan\}}^{d_{\max},\theta_{\max}}(L_3) = 0.9$$

and for the merged MLN $L = L_1 \cup L_2 \cup L_3$ we obtain

$$\mathsf{coh}_{\{d,nixon,reagan\}}^{d_{\max},\theta_{\max}}(L) \approx 0.55$$

This leads to

$$\mathsf{comp}_{\{d\},\{nixon\},\{reagan\}}^{d_{\max},\theta_{\max}}(L_1, L_2, L_3) \approx 0.295$$

Furthermore, note that $\mathsf{coh}_{\{d,nixon\}}^{d_{\max},\theta_{\max}}(L_1 \cup L_2) = 0.55$ and $\mathsf{coh}_{\{nixon,reagan\}}^{d_{\max},\theta_{\max}}(L_2 \cup L_3) = 0.85$ and, therefore, L_2 and L_3 are more compatible than L_1 and L_2:

$$\mathsf{comp}_{\{d\},\{nixon\}}^{d_{\max},\theta_{\max}}(L_1, L_2) \approx 0.2795$$
$$\mathsf{comp}_{\{nixon\},\{reagan\}}^{d_{\max},\theta_{\max}}(L_2, L_3) = 0.45$$

Our compatibility measure gives meaningful results in the above example. We now investigate how it behaves in the general case.

Proposition 7. *It holds $\mathsf{comp}_{C_1,\ldots,C_m}^{d,\theta_{\max}}(L_1, \ldots, L_m) \in [0,1]$ for every $d \in \{d_{p,0}, d_{\max}, d_{\min}, d_{avg}\}$,*

The statement above says that the compatibility measure is normalized and therefore comparable.

Proposition 8. *For every $d \in \{d_{p,0}, d_{\max}, d_{\min}, d_{avg}\}$ it is $\mathsf{comp}_{C_1,\ldots,C_m}^{d,\theta_{\max}}(L_1, \ldots, L_m) = 0$ if and only if $\mathsf{coh}_{C_1\cup\ldots\cup C_m}^{d,\theta_{\max}}(L_1 \cup \ldots \cup L_m) = 0$ and $\mathsf{coh}_{C_i}^{d,\theta_{\max}}(L_i) = 1$ for all $i = 1, \ldots, m$.*

The above proposition states that a set of MLNs is completely incompatible if and only if each individual MLN is perfectly coherent and the merged MLN is completely incoherent.

6 Implementation

The framework for measuring coherence of MLNs has been implemented in the *Tweety library for artificial intelligence*[4]. The framework contains implementations for all distance measures and aggregation functions discussed above and we provided both a naive and complete MLN reasoner and a wrapper for using the *Alchemy*[5] MLN reasoner. While the naive MLN reasoner implements Equations (1) and (2) in a straightforward way by simply computing the the probabilities $P_{L,C}(\omega)$ for all $\omega \in \Omega(C)$, the *Alchemy* MLN reasoner supports different approximate methods such as Markov chain Monte-Carlo. Computing the coherence value $\mathsf{coh}_C^{d,\theta}(L)$ is computationally quite expensive as it involves calls to the MLN reasoner for every ground instance of a formula in L. Therefore, using the naive MLN reasoner is only feasible for small examples. However, in its current version the *Alchemy* MLN reasoner does not support querying the probabilities of arbitrary ground formulas but only for ground atoms. In order to obtain the probability of an arbitrary ground formula ϕ using *Alchemy* it has to be incorporated into the MLN via adding a strict formula $\phi \Leftrightarrow a$ with some new ground atom a. Then the probability of a can be queried which is, in theory, the same as the probability of ϕ. However, during our experiments we discovered that internal optimization mechanisms of *Alchemy* might change the probabilities of other formulas when adding the strict formula $\phi \Leftrightarrow a$. This observation also raises the need for the development of an MLN reasoner that supports querying for arbitrary ground formulas. Recent developments such as [10] are gaining to close this gap.

7 Discussion and Conclusion

We introduced coherence as an indicator of how the weighted formulas of an MLN interact with each other. We used distance measures and aggregation functions to measure coherence by comparing the observed probabilities with the ones stemming from a naive probabilistic interpretation of the weights. By doing so, we came up with a meaningful assessment tool that satisfies several desirable properties. As an application for our framework we investigated the issue of merging and developed an indicator for quantifying the compatibility of different MLNs.

The approach presented in this paper can be used by a knowledge engineer to determine suitable weights for formulas, thus complementing the work of Pápai et al. [11] where MLNs are constructed by taking subjective probabilities of an expert into account. In particular, [11] already discusses the issue of consistent and inconsistent input probabilities and that in the latter case, parameters for the probability distribution have to be averaged, thus also resulting in an incoherent MLN in the sense of our work. By assessing the representation quality of MLNs using our approach experts can be guided to carefully choose correct weights/probabilities or re-structure the knowledge base.

To the best of our knowledge this work is the first that deals with quantifying the representation quality of an MLN by investigating the interrelationships of its formulas. The work presented is inspired by works on measuring the inconsistency in probabilistic conditional logic [16]. The work [16] defines an inconsistency measure by measuring the distance of an inconsistent knowledge base to the next consistent one. In this aspect, our framework uses similar methods. But as the concept of consistency is not applicable for MLNs we used a probabilistic interpretation of weights as a reference for assessing the coherence of an MLN. The term *coherence* has also been used before to describe "appropriateness" of a knowledge base or a model in other contexts. For example, in [9] a set of propositional formulas is said to be coherent with respect to a probability function if the probability of each single formulas increases when conditioning on the other formulas (there are also other similar notions considered).

Although MLNs are quite a mature framework for dealing with first-order probabilistic information, the lack of powerful and flexible MLN reasoner became evident in our experiments. Besides *Alchemy* we also looked at other available reasoning systems for MLNs such as thebeast[6] and Tuffy[7] but all lacked the crucial feature of computing the probabilities of arbitrary ground formulas. For future work, we consider to approach this problem and develop an MLN reasoner that can specifically be used for measuring coherence. Another direction for future work is the problem of deciding whether a coherent MLN can be learned from data and how to do this.

[4] http://tinyurl.com/MLNCoherence2

[5] http://alchemy.cs.washington.edu

[6] http://code.google.com/p/thebeast/

[7] http://hazy.cs.wisc.edu/hazy/tuffy/

REFERENCES

[1] L. De Raedt and K. Kersting, 'Probabilistic inductive logic programming', in *Probabilistic Inductive Logic Programming*, 1–27, Springer, (2008).

[2] P. Domingos and D. Lowd, *Markov Logic: An Interface Layer for Artificial Intelligence*, Morgan and Claypool, 2009.

[3] M. Finthammer, C. Beierle, J. Fisseler, G. Kern-Isberner, and J.I. Baumbach, 'Using probabilistic relational learning to support bronchial carcinoma diagnosis based on ion mobility spectrometry', *International Journal for Ion Mobility Spectrometry*, **13**, 83–93, (2010).

[4] J. Fisseler, 'Toward markov logic with conditional probabilities', in *Proceedings of the Twenty-First International FLAIRS Conference*, eds., D. C. Wilson and H. C. Lane, pp. 643–648. AAAI Press, (2008).

[5] M. Gelfond and N. Leone, 'Logic programming and knowledge representation - the a-prolog perspective', *Artificial Intelligence*, **138**(1–2), 3–38, (2002).

[6] S. O. Hansson, *A Textbook of Belief Dynamics*, Kluwer Academic Publishers, 2001.

[7] Kristian Kersting, 'Lifted probabilistic inference', in *Proceedings of 20th European Conference on Artificial Intelligence (ECAI'12)*. IOS Press, (August 2012).

[8] S. Kullback and R.A. Leibler, 'On information and sufficiency', *Annals of Mathematical Statistics*, **22**(1), 79–86, (1951).

[9] W. Meijs, *Probabilistic Measures of Coherence*, Haveka, 2005.

[10] M. Niepert, 'Symmetry-aware marginal density estimation', in *Proceedings of the 27th Conference on Artificial Intelligence (AAAI'13)*, (2013).

[11] T. Pápai, S. Ghosh, and H. Kautz, 'Combining subjective probabilities and data in training markov logic networks', in *Proceedings of the 2012 European conference on Machine Learning and Knowledge Discovery in Databases (ECML+PKDD'12)*, pp. 90–105, Berlin, Heidelberg, (2012). Springer-Verlag.

[12] Jeff B. Paris, *The Uncertain Reasoner's Companion – A Mathematical Perspective*, Cambridge University Press, 1994.

[13] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Springer, 1988.

[14] R. Reiter, 'A logic for default reasoning', *Artificial Intelligence*, **13**, 81–132, (1980).

[15] M. Richardson and P. Domingos, 'Markov logic networks', *Machine Learning*, **62**(1–2), 107–136, (2006).

[16] M. Thimm, 'Inconsistency measures for probabilistic logics', *Artificial Intelligence*, **197**, 1–24, (April 2013).