

Modal Markov Logic

Henry Kautz
University of Rochester
Rochester, NY 14627
kautz@cs.rochester.edu

7 July 2009

1 Introduction

Consider representing the following the statements about communication and cooperation:

If A tells B that P, then A believes P.

If A tells B that P, then B will believe that A wants B to believe P.

if A trusts B, and B believes P, then A will believe P.

if A is cooperative with B, and B wants P, then A will want P.

Two features stand out regarding the expressive requirements of a formal language that could express these bits of commonsense knowledge. First, the language must have a way of representing propositional *attitudes* such as believes and wants. This suggests the language should include modal operators [Che80]. Second, the language must be able to represent assertions that hold in general, but are not universally true. While all of the statements above are useful principles for reasoning about social behavior, it is easy to construct situations where each fails to hold. This suggests that the language should be able to represent *defeasible* assertions. Furthermore, it is clear that the degree to which each of these assertions holds differs. The first statement can be restated as “people are truthful”, and the second as “people want to be believed”; the second is surely more generally true than the first. The need for a defeasible representation whose statements can be assigned a degree of truth suggests that the representation in fact be *probabilistic*.

Markov Logic [RD06] is a probabilistic generalization of first-order logic. As we will see, it is particularly well suited for representing defeasible statements of varying strength. In this white paper, we will describe how Markov Logic’s semantics and inference methods can be extended to handle modal operators. We will concentrate on sentential modal operators, such as appear in the examples above. We will include a brief discussion of the significantly higher complexity of adding *functional* modal operators.

2 Background

The propositional subset of Modal Markov Logic is defined as follows; for the full first-order case see [RD06]. A Markov Logic knowledge base is a set of clauses with associated real-valued weights. We write w_i for the weight of clause c_i . A world X is a Boolean vector corresponding to a truth assignment to the propositions in the language. We define the Boolean feature function $f_i(X)$ as returning 1 if and only if c_i is satisfied by X . A knowledge base induces a probability distribution over the space of worlds by defining the following random field, where \exp is the exponential function and Z is a normalization constant:

$$\Pr(X) = (1/Z) \exp\left(\sum_i w_i f_i(X)\right)$$

Practical algorithms have been devised for inference, weight learning, and structure (clause) learning in Markov Logic.

Modal logic [Che80] extends propositional or first-order logic by adding one or more non-truth functional sentential operators; we will again only discuss the propositional case here. We use the symbol B_i to represent the i -th modal operator. Where α is a well formed sentence, then $B_i\alpha$ is a well formed sentence. For example, if we take B_i to mean “agent i knows that”, then the formula $B_iP \vee B_i\neg P$ means that agent i knows whether or not P holds. Note that this is quite different from the tautology $B_iP \vee \neg B_iP$.

Different modal operators for concepts such as belief, knowledge, desire, obligation, *etc.*, can be specified by the *axiom schemas* that they satisfy. In this paper, we will consider what are called *normal modal logics*, where the operator satisfy *deductive closure*. Deduction closure can be stated as follows: if $\alpha_1, \dots, \alpha_n \vdash \beta$ by ordinary propositional reasoning, then $M_i\alpha_1, \dots, M_i\alpha_n \vdash M_i\beta$. Following common practice in logic and AI, we assume that belief, the attitude on which we focus for the rest of this paper, satisfies deductive closure.¹

3 Modal Markov Logic

We use X to denote either a set of propositions or a Boolean vector over those propositions, as distinguished by context. The language of Modal Markov Logic (MML) theory over X and a set of modal operators B_i is defined as follows:

- A theory is a set of pairs (w_i, c_i) , where w_i is a real number and c_i is a clause.
- A clause is a (possibly unary) disjunction of literals.
- A literal is an atom or its negation.
- An atom is a member of X or a symbol of the form $B_i c$, where B_i is a modal operator and c is a clause. We call the former base atoms and the latter modal atoms.

For example, an MML theory that asserted the defeasible rule that “if John believes P then Mary believes P ” would include

$$(w, \neg B_{John}P \vee B_{Mary}P)$$

where w is a positive real number.

We next define the notion of *correspondence* between a set of modal atoms M and a set of sentences \mathcal{T} in propositional modal logic. Let \mathcal{T} contain the sentence $B_i c$ for each modal atom $B_i c$ assigned true by M , and contain the sentence $\neg B_i c$ for each modal atom assigned false by M .²

¹Although people are obviously not logically omniscience, it has proven difficult to develop more realistic models of human inference, because so many extraneous factors can influence a person’s reasoning abilities, such as familiarity, prejudices, cognitive load, *etc.* Deductive closure can at least be justified for a *normative* theory of inference.

²Note that in the context of MML, the string “ $B_i c$ ” is an atom, while in the context of modal logic, it is a sentence which begins with a modal operator.

We can now define the probability distribution induced by a MML theory. Let M denote either the modal atoms of a given MML theory or a Boolean vector over those propositions, as distinguished by context. A world is truth assignment to $X \cup M$. The probability of a world is given by

$$\Pr(X, M) = (1/Z)g(M) \exp\left(\sum_i w_i f_i(X, M)\right)$$

where $g(M)$ is a function that returns 1 if and only if the modal logic theory that corresponds to M is consistent, and 0 otherwise.³

We see that the semantics of MML generalize the semantics for ordinary Markov Logic by requiring worlds to interpret modal atoms in a consistent manner. For example, $g(M)$ would rule out (assign zero probability to) worlds M where $B_i(P \vee Q)$ were assigned false and $B_i P$ were assigned true.

The computation performed by $g(M)$ depends upon the choice of underlying modal logic for the MML theory. If we choose the weakest normal modal logic (called \mathcal{K}), then $g(M)$ should determine consistency under (modal) deductive closure.

One method for implementing inference in MML is to combine a modal logic decision procedure with an algorithm for sampling from the distribution defined by $(1/Z) \exp(\sum_i w_i f_i(X, M))$. The latter task can be performed, for example, by Gibbs sampling or slice sampling, as in the Alchemy implementation of Markov Logic (<http://www.alchemy.org>). Samples that fail the modal consistency test are rejected. Note that this does *not* mean that the modal decision procedure must be invoked on every sampling step. For example, when Gibbs sampling is performed, a step that samples a variable from X while holding the rest of $X \cup M$ constant will not invoke the modal decision procedure.

Rejection sampling can be made more efficient by automatically adding axioms to the MML theory that rule out many of the states that are modally inconsistent. For example, suppose the MML theory contains the atoms $B_a P$, $B_a Q$, and $B_a(P \vee Q)$. Then one could add hard (infinite weight) clauses that relate these atoms to the MML theory:

$$\begin{aligned} B_a P &\supset B_a(P \vee Q) \\ B_a Q &\supset B_a(P \vee Q) \\ \neg B_a P &\vee \neg B_a \neg P \\ \neg B_a Q &\vee \neg B_a \neg Q \\ B_a \neg P \wedge B_a(P \vee Q) &\supset B_a Q \\ B_a \neg Q \wedge B_a(P \vee Q) &\supset B_a P \end{aligned}$$

Adding clauses of this kind would be particularly appropriate when doing slice-sampling, because slice-sampling changes many variables simultaneously and is robust in the presence of hard clauses.⁴

A variety of modal decision procedures have been proposed in the literature. From a practical standpoint, the most promising approaches are based on extensions to the Davis-Putnam backtracking procedure for satisfiability testing [GGT02]. An interesting research question is whether there exist efficient decision procedures that leverage the particular syntactic form of the modal theories that are generated in the $g(M)$ step, and/or that take advantage of the top-level sampling strategy by reusing previous partial results.

4 Example

We now present a simple example of using MML to reason about perception and belief. Suppose we wish to represent the following assertions:

1. If John believes there is no fire, then Mary believes there is no fire.
2. If Sam believes there is smoke, then Mary believes there is smoke.

³We placed $g(M)$ outside of the exponential to make clear its role as a filter. Alternatively, $g(M)$ could be added to the summation in the exponent, by having it return 0 for a modally consistent theory and $-\infty$ for an inconsistent one.

⁴Note that if enough clauses were added, it would not be necessary to filter the samples using a modal decision procedure. In the worst case, however, it would be necessary to add an exponential number of new clauses.

3. Mary believes that if there is smoke, there is fire.

Suppose the statements are ordered in increasingly likelihood. Using the proposition F to mean “there is fire” and S to mean there is smoke, and glossing over the temporal aspects of the statements, we might represent these assertion in MML as:

1. $(0.7, (B_{John} \neg F) \supset (B_{Mary} \neg F))$
2. $(0.8, (B_{Sam} S) \supset B_{Mary} S)$
3. $(0.9, B_{Mary}(S \supset F))$

The weights on the clauses are not true probabilities, because they are not locally normalized. It is correct, however, to think of statements with higher weight as being more likely to be true. Specific weights can either be learned from examples or (in certain cases) computed from a set of conditional probability tables (see [Par02] for details).

Suppose we given the evidence $(B_{John} \neg F, B_{Sam} S)$, and wish to infer the probability that $B_{Mary} F$. The modal consistency check eliminates worlds where $B_{Mary} F$ and $B_{Mary} \neg F$ both hold. In this small example, one can enumerate all the remaining worlds, calculate the probability of each, and sum the probabilities of the worlds where $B_{Mary} F$, and so determine that $\Pr(B_{Mary} F) = 0.5$.

5 Functional Modal Operators

The language we described in this paper allows us to reason about uncertain information, defeasible rules, and modal attitudes such as belief. It allows us to state conditions under which beliefs come about, but says little about what beliefs *are*, beyond constraining one’s set of beliefs to be deductively consistent. In particular, it does not say that beliefs must be tautologies or universal truths. It can be useful to think of beliefs as attitudes held by an agent that are *actionable*, in the sense that an agent’s beliefs are made manifest by the agent’s actions.

Our account so far does not allow us to talk about *degrees of belief*. The statement $\Pr(B_{Mary} F) = 0.5$ does *not* mean that Mary’s degree of belief in F is 0.5. Mary either believes F or she does not believe F . The statement means that the probability that Mary believes F , given all the evidence we have at hand, is 0.5.

If we do wish to represent the degrees of belief that agents assign to statements, we must move to a more powerful language. One way this can be done is to treat modal operators as functions that map a sentence to a probability, and allow such functions to appear in mathematical expressions. For example, a statement such as

John is at least 70% sure that Mary is at least 80% sure that P .

can be represented as

$$B_{John}(B_{Mary} P \geq 0.8) \geq 0.7$$

[Hal05] developed a first-order logic axiomatization of this language. He showed that the axiomatization captured the semantic notion of a *multi-agent probabilistic structure*, which generalizes the Kripke structure semantics for modal logic by augmenting the accessibility relationship between possible worlds with a probability measure over accessible worlds. However, it is not yet known how to devise a *practical* reasoning system for a language this powerful. Although a complete first-order theorem prover would in principle provide a decision procedure for the language, in practice it could handle only the smallest toy examples. Recent work by [SA07] shows how one can do *model checking* in this language, that is, how one can *evaluate* a sentence in a multi-agent logic of probability against a mathematical structure that represents a higher-order probability distribution, but does not investigate how to answer queries against a knowledge base of sentences in the language.

6 Research Plan

The work described in this white paper is in its earliest stages. Our research plan includes the following milestones:

1. Prove that the semantics for Modal Markov Logic given above is equivalent to defining a probability distribution over Kripke structures (the standard semantics for modal logic). Prove that the rejection-sampling inference algorithm outlined above is sound.
2. Extend propositional MML to first-order quantification over finite domains, as in general Markov logic. In particular, ensure that MML correctly handles “quantifying in” to distinguish an agent’s attitude toward a quantified variable. For example, while the statement $B_{John}\exists x.P(x)$ means that John believes there is something that is a P , the statement $\exists x.B_{John}P(x)$ makes the stronger claim that John believes P holds of an entity that he can identify.
3. Extend the incremental-instantiation method used for inference in Markov Logic to handle *axiom schemas* in the context of MML. Many important defeasible rules for inferring the mental states of agents are better captured by schemas than by first-order statements. For example, consider the rule that states that John believes Mary believes whatever he believes. This can be captured by the schema

$$(w, (B_{John}\alpha) \supset (B_{John}B_{Mary}\alpha))$$

where α ranges over *sentences*.

4. Devise, implement, and evaluate the Gibbs and slice sampling algorithms for approximate inference described above. In particular, test the methods of using an incremental modal decision procedure and of adding new clauses to the MML theory in order to improve the efficiency of rejection sampling.
5. Determine how to extend the belief propagation inference algorithm for Markov Logic to MML.
6. Apply MML to a series of real-world problems. One potential application is disambiguating speech acts and tracking the beliefs, wants, and obligations of a group of people in conversation. Data for such an effort could come from annotated transcripts of business meetings [SDB⁺04]. Another potential application is in modeling developmental theories of mind (*e.g.*, see [BBC07]).

References

- [BBC07] Paul Bello, Perrin Bignoli, and Nicholas Cassimatis. Attention and association explain the emergence of reasoning about false beliefs in young children. In *Proceedings of ICCM-2007, the Eighth International Conference on Cognitive Modeling*, 2007.
- [Che80] Brian F. Chellas. *Modal logic: an introduction*. Cambridge University Press, 1980.
- [GGT02] Enrico Giunchiglia, Fausto Giunchiglia, and Armando Tacchella. SAT-based decision procedures for classical modal logics. *Journal of Automated Reasoning*, 28(2), 2002.
- [Hal05] Joseph Y. Halpern. *Reasoning About Uncertainty*. The MIT Press, 2005.
- [Par02] James D. Park. Using weighted MAX-SAT engines to solve MPE. In *Proceedings of the 18th National Conference on Artificial Intelligence (AAAI-2002)*, pages 682–687, 2002.
- [RD06] Matt Richardson and Pedro Domingos. Markov logic networks. *Markov Logic Networks*, 62, 2006.
- [SA07] Afsaneh Shirazi and Eyal Amir. Probabilistic modal logic. In *Proceedings of the 22nd National Conference on Artificial Intelligence (AAAI-2007)*, 2007.
- [SDB⁺04] E. Shriberg, R. Dhillon, S. Bhagat, J. Ang, and H. Carvey. The ICSI meeting recorder dialog act (MRDA) corpus. In *Proceedings of the SigDial Workshop*, 2004.