A Summary of the Doctoral Thesis

"Results on Modal Reasoning with Applications to Modal Deductive Databases"

by Nguyen Anh Linh

Deductive databases and logic programming are fields that have gained a great deal of attention due to their usefulness and a wide range of applications. These fields are closely related to each other, they both use classical logic rules to represent mutual dependencies between relations. While logic programming is mostly used in artificial intelligence systems, deductive databases are systems in which the amount of data is much greater than the number of rules. Primarily, logic rules in such systems are restricted to the positive form, i.e. to universally quantified Horn clauses of the classical first-order logic. It is known that with such restriction the least model for a program (or a database) is computable in polynomial time and its size is bounded by a polynomial.

When trying to extend the fields of deductive databases and logic programming to deal with modal logics, a natural attempt is to consider the modal Horn fragment. Unfortunately, it has been shown by Chen and Lin [3] that the problem of checking satisfiability of a set of modal Horn formulae in the propositional modal logics K, KD, T, KB, KDB, B, $K4$, $KD4$ and $S4$ is PSPACE-complete. Due to this fact, modal logic programming has not gained much interest from researchers. Moreover, the modal Horn fragment has been only, as known to us, studied in the context of top-down methods, which are suitable for logic programming but not deductive databases. With bottom-up methods, we often deal with constructing the least models for positive programs. Having the least model of a positive program, all queries for that program can be effectively answered.

Complexity of modal logics has been studied by several authors. In [10] Ladner showed that the complexity of provability in the propositional modal logics K, T, and $S4$ is PSPACE-complete, and in $S5$ is co-NP-complete. In [4] Fariñas del Cerro and Penttonen showed that the problem of checking whether a set of modal Horn formulae is satisfiable in $S5$ is solvable in PTIME. In [3] Chen and Lin showed that the similar problem for a propositional modal logic L which is a normal extension of $K5$, write $K5 \leq L$, is also solvable in PTIME. Chen and Lin also proved that for a propositional normal modal logic L such that $K \leq L \leq S4$ or $K \leq L \leq B$, the problem is PSPACE-complete. Recently, some authors have analyzed space requirement for propositional modal logics. In [7] Hudelmaier showed that provability for the logic $S4$ is decidable in $O(n^2.\log n)$-space, for the logic K in $O(n.\log n)$-space. In [1, 12, 2] Basin, Matthews and Viganò showed that provability for the logic KD is decidable in $O(n.\log n)$-space, and for the logics T, $K4$ and $KD4$ in $O(n^2.\log n)$-space.

Motivation

The primary aim of our thesis is to develop a modal query language. We want to select fragments of modal logics that are solvable in polynomial time, and then formulate them into a query language form. We also want to develop a bottom-up computation method for the selected fragments. Another problem we want to study is whether decision procedures and upper space bounds for propositional modal logics can be improved.

Definitions for Propositional Modal Logics

Formulae in propositional modal logics are built from primitive propositions, classical connectives, the unary modal connectives $\Box$ and $\Diamond$, and parentheses, in a usual way.

Definition A *Kripke frame* is a triple $\langle W, \tau, R\rangle$, where W is a nonempty set (of possible worlds), $\tau \in W$ is the actual world, and R is a binary relation on W, called the accessibility relation. A (propositional) *Kripke model* is a tuple $\langle W, \tau, R, h\rangle$, where $\langle W, \tau, R\rangle$ is a Kripke frame, and h is a mapping from worlds to sets of primitive propositions; that is, $h(w)$ is the set of primitive propositions which are "true" at the world w.

Axiom	Schemata	First-Order Formula
D	$\Box\Phi \to \Diamond\Phi$	$\forall w\ \exists w'\ R(w,w')$
T	$\Box\Phi \to \Phi$	$\forall w\ R(w,w)$
B	$\Phi \to \Box\Diamond\Phi$	$\forall w,w'\ R(w,w') \to R(w',w)$
4	$\Box\Phi \to \Box\Box\Phi$	$\forall w,u,v\ R(w,u) \wedge R(u,v) \to R(w,v)$
5	$\Diamond\Phi \to \Box\Diamond\Phi$	$\forall w,u,v\ R(w,u) \wedge R(w,v) \to R(u,v)$

Table 1: Axioms and corresponding first-order conditions on R

The *satisfiability relation* is defined in a usual way (see Definition 2.1.4).

The simplest normal (propositional) modal logic (called K) is axiomatized by the standard axioms for the classical propositional logic and the *modus ponens* inference rule together with the axiom schema $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$, which is usually called the *K-axiom*, plus the additional *necessitation rule*

$$\frac{\vdash \phi}{\vdash \Box\phi}$$

It can be shown that a modal logic formula is provable in this axiomatization iff it is satisfied in every Kripke model (i.e. without any special R-properties) [9]. It is known that certain axiom schemata added to this axiomatization correspond to certain properties of the accessibility relation.

Many of such correspondences are definable as formulae of first-order logic where the binary predicate $R(x, y)$ represents the accessibility relation, as shown in Table 1. Different modal logics are distinguished by their respective additional axiom schemata. Some of the most popular modal logics together with their axiom schemata are listed in Table 2. We refer to properties of the accessibility relation of a modal logic L as *L-frame axioms* or *L-frame restrictions*. A Kripke frame is called *L-frame* if its accessibility relation satisfies all L-frame restrictions.

We call a Kripke model M a *L-model* if the accessibility relation of M satisfies all L-frame restrictions. We say that ϕ is L-satisfiable if there exists a L-model of ϕ.

We call modal logics which are characterized by classes of Kripke models *normal modal logics*. Given two normal modal logics L and L', we say that L' is a *normal extension* of L if all L-frame restrictions are also L'-frame restrictions.

We say that two sets X and Y of formulae are equisatisfiable in a logic L if (X is L-satisfiable iff Y is L-satisfiable).

Modal Horn Formulae and Positive Propositional Modal Logic Programs

A formula is in the *negative normal form* if each of its negations occurs immediately before a primitive proposition. A formula is called *negative* if in its negative normal form every primitive proposition is prefixed by negation. A formula is called *non-negative* if it is not negative, and *positive* if its negation is a negative formula.

Definition A formula ϕ is called a *Horn formula* iff one of the following conditions holds:

- ϕ is a primitive proposition
- ϕ is a negative formula
- $\phi = \Box\psi$, or $\phi = \Diamond\psi$, or $\phi = \psi \wedge \zeta$, where ψ and ζ are Horn formulae
- ϕ is a disjunction of a negative formula and a Horn formula.

Logic	Axioms	Frame Restriction
K	K	no restriction
KD	KD	serial
T	KT	reflexive
KB	KB	symmetric
KDB	KDB	serial and symmetric
B	KTB	reflexive and symmetric
$K4$	K4	transitive
$KD4$	KD4	serial and transitive
$S4$	KT4	reflexive and transitive
$K5$	K5	euclidean
$KD5$	KD5	serial and euclidean
$K45$	K45	transitive and euclidean
$KD45$	KD45	serial, transitive and euclidean
$KB5$	KB5	symmetric and euclidean
$S5$	KT5	reflexive and euclidean

Table 2: Modal logics and frame restriction

Definition A *positive propositional modal logic program*, sometimes just called a positive program, is a finite set of rules of the following form:

$$\Box^s(B_1 \wedge \ldots \wedge B_k \rightarrow A)$$

where $s \geq 0$, $k \geq 0$ and $A, B_1, \ldots, B_k$ are atoms of one of the forms p, $\Box p$, $\Diamond p$, where p is a primitive proposition. We often write program rules in the form:

$$\Box^s(A \leftarrow B_1, \ldots, B_k).$$

It is shown in the thesis that we can translate any set X of Horn formulae to a positive program P and a positive formula Q such that for any normal modal logic L, X is L-satisfiable iff $P \not\models_L Q$.

Definition Let $M = \langle W, \tau, R, h\rangle$ and $N = \langle W', \tau', R', h'\rangle$ be Kripke models. We say that M is less than or equal to N wrt. a binary relation $r \subseteq W \times W'$, and write $M \leq N$ wrt. r, if:

1. $r(\tau, \tau')$
2. $\forall x, x' \;\; r(x, x') \rightarrow (h(x) \subseteq h'(x'))$
3. $\forall x, x', y \;\; R(x, y) \wedge r(x, x') \rightarrow \exists y' \; R'(x', y') \wedge r(y, y')$
4. $\forall x, x', y' \;\; R'(x', y') \wedge r(x, x') \rightarrow \exists y \; R(x, y) \wedge r(y, y')$

We say that a model M is less than or equal to N, and write $M \leq N$, if $M \leq N$ wrt. some r.

It is shown in the thesis that the relation $\leq$ between Kripke models is a pre-order, and that if $M \leq N$, then for every positive formula ϕ, if $M \models \phi$ then $N \models \phi$.

Given a positive program P in a normal modal logic L, we say that M is the least L-model of P if M is a L-model of P and M is less than or equal to every L-model of P.

Clausal Tableau Systems for Propositional Modal Logics

In the thesis, we propose so called clausal tableau systems for the propositional modal logics K, KD, T, KB, KDB, B, $K4$, $KD4$, and $S4$ (see Sections 2.5 and 3.1, and Tables 3.1, 3.2, and 3.3). Basing on these systems we give decision procedures for the logics with improved upper space bounds.

Our tableau systems are defined only for sets of *basic clauses*. If we use letters a and b to denote primitive propositions and their negations, then by a basic clause we mean a formula of one of the forms $\Box^s \neg\Box a$, $\Box^s(a_1 \vee \ldots \vee a_k)$, and $\Box^s(a \vee B)$, where $s \geq 0$, $k \geq 1$, and B is a formula of the form $\Box b$ or $\neg\Box b$. Given a formula, we can translate it to an equisatisfiable set of basic clauses to be usable by our systems (see Section 2.3).

Our formulation of tableau systems is based on the work of Hintikka [6], Rautenberg [11], Goré [5], and Hudelmaier [8]. For the logics K, KD, T, KB, KDB, and B, we introduce the connective $\blacksquare$, which has the same semantics as $\Box$, but it plays a specific syntactical role, namely, formulae of the form $\blacksquare^s\phi$ are blocked from being principal clauses. For the logics $K4$ and $KD4$, we introduce the connective $\boxdot$ to simulate the connective $\Box$ of $S4$, in the way such that $M, w \vDash \boxdot\phi$ iff $M, w \vDash (\Box\phi \wedge \phi)$. For the logics $K4$ and $KD4$, the connective $\blacksquare$ has the same semantics as $\boxdot$, but formulae of the form $\blacksquare^s\phi$ are blocked from being principal clauses.

Our tableau systems are sound and complete (see Theorems 3.4.9, 3.5.9, and 3.6.9), and have the finite tableau property (see Section 3.3).

Definitions for Fixed-domain First-order Modal Logics over Signatures without Functions

Formulae in such logics are built from constant symbols, variables, relational symbols, classical connectives, the modal connectives $\Box$ and $\Diamond$, the quantifiers $\forall$ and $\exists$, and parentheses, in a usual way. A *term* is either a constant symbol or a variable. We call $R_i(t_1, \ldots, t_n)$, where R_i is a relation symbol of arity n and $t_1, \ldots, t_n$ are terms, a *classical atom*. A formula is called *ground* if it contains no quantifiers and no variables.

We define the modal depth of a formula ϕ to be the maximal nesting depth of modalities (i.e. the connectives $\Box$ and $\Diamond$) in ϕ.

Definition A *fixed-domain first-order Kripke model over a signature without functions* Σ, simply called first-order Kripke model, is a tuple $M = \langle D, W, \tau, R, m\rangle$, where D is a set called *domain*, $\langle W, \tau, R\rangle$ is a Kripke frame, and m is an interpretation of constant symbols and relation symbols. For a constant symbol c_i, $m(c_i)$ is an element of D, denoted also by ${c_i}^M$. For a relation symbol R_j of arity n and a world $w \in W$, $m(w)(R_j)$ is a n-ary relation on D, denoted also by ${R_j}^{M,w}$.

The *satisfiability relation* is defined in a usual way (see Definition 5.1.3).

Let L be the name of some propositional normal modal logic, e.g. KD. To define what is the *first-order modal logic* L we define the class of admissible interpretations to be the class of first-order Kripke models $M = \langle D, W, \tau, R, m\rangle$ with $\langle W, \tau, R\rangle$ being a L-frame.

The definitions of L-satisfiability and L-models, for a first-order normal modal logic L, are straightforward.

If X is a set of formulae and ϕ is a formula then we write $X \vDash_L \phi$ to denote the fact that for any first-order L-model M, if $M \vDash X$ then $M \vDash \phi$.

The Modal Query Language MDatalog

The modal query language MDatalog is defined and interpreted in fixed-domain first-order modal logics over signatures without functions.

Definition A *MDatalog program* is a finite set of rules of the form

$$\forall\overline{x}\ \Box^s(A \leftarrow B_1, \ldots, B_k)$$

where $s \geq 0$, $k \geq 0$, and

- A, B_1, ..., B_k are formulae of the forms $\Box\phi$, $\Diamond\phi$, ϕ, with ϕ being a classical atom.
- $\overline{x}$ are all variables occurring in $(A \leftarrow B_1, \ldots, B_k)$.

- If $k = 0$ then the rule is a ground fact, i.e. there is no quantifiers prefixing the rule.
- All variables occurring in A also occur in $(B_1, \ldots, B_k)$.

Definition A *query* is a formula of the form $\exists \overline{x}\ \phi(\overline{x})$, where ϕ is a positive formula without quantifiers, and $\overline{x}$ is a vector of all variables occurring in ϕ, and can be empty.

Fix a first-order normal modal logic L. Given a MDatalog program P and a query $\exists \overline{x}\ \phi(\overline{x})$, the goal is to check whether $P \models_L \exists \overline{x}\ \phi(\overline{x})$, and to find tuples of constant symbols $\overline{c}$ such that $P \models_L \phi(\overline{c})$.

First-order Kripke models can be ordered in a similar way as propositional Kripke models.

Definition Let $M = \langle D, W, \tau, R, m \rangle$ and $N = \langle D', W', \tau', R', m' \rangle$ be first-order Kripke models. We say that M is less than or equal to N with respect to a function $f : D \to D'$ and a binary relation $r \subseteq W \times W'$ if

1. $r(\tau, \tau')$
2. $\forall x, x', y\ \ R(x, y) \land r(x, x') \to \exists y'\ R'(x', y') \land r(y, y')$
3. $\forall x, x', y'\ \ R'(x', y') \land r(x, x') \to \exists y\ R(x, y) \land r(y, y')$
4. For any constant symbol c_i, $f({c_i}^M) = {c_i}^N$.
5. For any $x \in W$ and $x' \in W'$ such that $r(x, x')$, for any n-ary relation symbol R_j and elements $a_1, \ldots, a_n$ of D, if ${R_j}^{M,x}(a_1, \ldots, a_n)$ then ${R_j}^{N,x'}(f(a_1), \ldots, f(a_n))$.

We say that a model M is less then or equal to N, and write $M \leq N$, if M is less then or equal to N with respect to some f and some r.

It is shown in the thesis that the relation $\leq$ between first-order Kripke models is a pre-order, and that if M and N are first-order Kripke models and $M \leq N$, then for every positive ground formula ϕ, if $M \models \phi$ then $N \models \phi$.

Given a MDatalog program P in a first-order normal modal logic L, we say that M is the least L-model of P if M is a L-model of P and M is less than or equal to every L-model of P.

Methods

- To estimate space bounds for propositional modal logics, we define decision procedures based on clausal tableau systems for them, in which saturating operations play an important role.
- For L being a propositional normal modal logic, to construct the least L-model for a positive propositional modal logic program P we build a L-model graph realizing P.
- Given a MDatalog program P in a first-order normal modal logic L, to construct the least L-model of P we can make the ground instantiation of P, treat it as a positive propositional modal logic program, and construct the least propositional L-model for it. (This method is for estimating computational complexity. In the thesis we also discuss a more efficient method.)

Results

Our main contributions are the following:

- clausal tableau systems and improved decision procedures for the propositional modal logics K, KD, T, KB, KDB, B, $K4$, $KD4$, and $S4$ (see Chapter 3);
- establishment of the best currently known upper space bound $O(n.\log n)$ for the propositional modal logics $K4$, $KD4$, and $S4$, and $O(n^2)$ for KB, KDB, and B (see Section 3.7 and Theorems 3.7.2, 3.7.6, and 3.7.9);

- a new proof for the results of other authors [8, 1] such that the propositional modal logics K and KD are decidable in $O(n.\log n)$-space (see Section 3.7.1 and Theorem 3.7.2);

- the formulation of positive propositional modal logic programs and MDatalog programs (see Definitions 2.4.2 and 5.2.1);

- algorithms of constructing the least models for positive propositional modal logic programs and MDatalog programs in the modal logics KD, T, KDB, B, $KD5$, $KD45$, $S5$, $KD4$, and $S4$ (see Algorithms 4.2.1, 4.4.1, and 5.5.1), and conditions that guarantee the algorithms terminating in polynomial time (see Theorems 4.2.6 and 4.4.9, and Corollary 5.4.5);

- a characterization of positive modal logic programs and MDatalog programs in KB, $K5$, $K45$, and $KB5$ by one or two models, which can be effectively constructed (see Section 4.3 and Corollaries 4.3.5 and 5.4.7);

- a theorem showing PTIME complexity of computing queries for a given MDatalog program in the logics KD, T, KB, KDB, B, $K5$, $KD5$, $K45$, $KD45$, $KB5$, and $S5$, provided that the quantifier depths of queries and the program are finitely bounded, and that the modal depth of the program is finitely bounded in the case when the considered logic is not a normal extension of $K5$ (see Theorem 5.4.8);

- a theorem stating that the problem of checking satisfiability of a set of Horn formulae with finitely bounded modal depths in the propositional modal logics KD, T, KB, KDB, and B is decidable in PTIME (see Corollaries 4.2.7 and 4.3.6);

- a new proof for the results of other authors [4, 3] such that the problem of checking satisfiability of a set of Horn formulae in the propositional modal logics $K5$, $KD5$, $K45$, $KD45$, $KB5$, and $S5$ is decidable in PTIME (see Corollaries 4.2.8 and 4.3.7).

The thesis also contains examples of application of the techniques to modal deductive databases. Our methods for both of the problems of estimating space requirement for propositional modal logics and of constructing the least models for positive modal logic programs are applicable not only for the logics considered in the thesis.

References

[1] D. Basin, S. Matthews, and L. Viganò. A new method for bounding the complexity of modal logics. In *G. Gottlob, A. Leitsch, and D. Mundici, editors, Proceedings of the 5th Kurt Gödel Colloquium on Computational Logic and Proof Theory (KGC'97), LNCS 1289*, pages 89–102. Springer, 1997.

[2] D. Basin, S. Matthews, and L. Viganò. *Complexity Bounds for Propositional Modal Logics.* Technical Report, Institut für Informatik, Albert-Ludwigs-Universität Freiburg, Germany. To appear. Available at http://www.informatik.uni-freiburg.de/˜luca, 1999.

[3] C.C. Chen and I.P. Lin. The computational complexity of the satisfiability of modal Horn clauses for modal propositional logics. *Theoretical Computer Science*, 129:95–121, 1994.

[4] L. Fariñas del Cerro and M. Penttonen. A note on the complexity of the satisfiability of modal Horn clauses. *Logic Programming*, 4:1–10, 1987.

[5] R. Goré. *Tableau Methods for Modal and Temporal Logics.* Technical Report TR-ARP-15-95, Australian National University, Available via http://arp.anu.edu.au on WWW, 1995.

[6] K.J.J. Hintikka. Form and content in quantification theory. *Acta Philosophica Fennica*, 8:3–55, 1955.

[7] J. Hudelmaier. A contraction-free sequent calculus for S4. In *H. Wansing, editor, Proof Theory of Modal Logic*, pages 3–15. Kluwer, Dordrecht, 1996.

[8] J. Hudelmaier. Improved decision procedures for the modal logics K, T, S4. In *H. Kleine Büning, editor, Proceedings of CLS'95, LNCS 1092*, pages 320–334. Springer, 1996.

[9] S.A. Kripke. Semantical analysis of modal logic, I. normal modal propositional calculus. *Z. Math. Logic Grundlag*, 9:67–96, 1963.

[10] R. Ladner. The computational complexity of provability in systems of modal propositional logic. *SIAM Journal of Computing*, 6:467–480, 1977.

[11] W. Rautenberg. Modal tableau calculi and interpolation. *Journal of Philosophical Logic*, 12:403–423, 1983.

[12] L. Viganò. *A framework for non-classical logics.* PhD thesis, Universität des Saarlandes, Saarbrücken, Germany, 1997.