Graded Modal Logic GS5 and Itemset Support Satisfiability

Yakoub SALHI and Saïd JABBOUR and Lakhdar SAIS

Université Lille-Nord de France
CRIL - CNRS UMR 8188
Artois, F-62307 Lens
{salhi,jabbour,sais}@cril.fr

Abstract. Graded modal logic GS5 is an extension of S5 by the modal connective $\Diamond_\lambda$: the formula $\Diamond_\lambda A$ means that there are at least λ worlds satisfying A. In this paper, we show how to reduce GS5 satisfiability to propositional satisfiability (SAT). Furthermore, we consider a satisfiability problem related to the frequent itemset mining problem: $\mathsf{SUPPSAT}^n$ (where n is a strictly positive integer). We show how $\mathsf{SUPPSAT}^n$ can be encoded in GS5 satisfiability and consequently in SAT.

1 Introduction

The modal logic S5 is among the most studied normal modal logics. In a possible world semantics, a model frame is a non empty set of *worlds* with an accessibility relation which is an equivalence relation [4]. However, there is another equivalent possible world semantics where a model frame is just a non empty set of worlds without any accessibility relation [9]. The formula $\Diamond A$ means simply that there exists a world where A is true and $\Box A$ means that A is true in every world. Let us note that there exist various formal systems dedicated to proof-search in this logic [9, 22, 15].

Graded modal logic GS5 includes modal connectives allowing to reason on the number of worlds satisfying a formula [7, 21]. For example, the formula $\Diamond_5 A$ means that there are at least 5 worlds satisfying A. The first sound and complete axiomatization for this logic was provided by Fine in [7] and other results on axiomatic aspects were provided in [21]. Let us note that the modal connectives in GS5 coincide with concept cardinality restrictions in description logics defined in [3] and count operators in modal logics defined in [2].

In the literature, the majority of decision algorithms for modal logics are based on either the use of the formalism style in structural proof theory called sequent calculus and its variant called tableau method [8], or encodings in first order logic [16]. In the last decade, efficient SAT-based algorithms have been proposed to decide satisfiability of a formula in modal logics [10, 11]. It has been formally proved that these algorithms are better than those based on tableau method [17]. One of the main ideas in the SAT-based approach consists simply in considering the modal subformulas as propositional variables and trying to extend propositional models to modal models. In this work, we are interested in finding reductions of satisfiability problem in S5 and GS5 to SAT.

After having introduced syntax and semantics of the modal logic S5, we present a reduction of S5 satisfiability problem to SAT. This reduction is obtained by using the property that says that a formula is satisfiable in S5 if and only if it is satisfiable in a model with at most the number of its modal subformulas as the number of worlds. Using a similar approach, we also show how to reduce GS5 satisfiability to SAT. Thus, efficient propositional SAT solvers can be used to perform satisfiability in S5 and its graded version GS5.

In order to show how GS5 satisfiability, and consequently SAT, can be used to reason over some problems in data mining, we consider a satisfiability problem in data mining: $\mathsf{SUPPSAT}^n$ (where n is a strictly positive integer) [5, 6]. It is related to the frequent itemset mining problem which is one of the most studied problems in data mining [1]. It consists in computing frequent itemsets in a transaction database. We show how $\mathsf{SUPPSAT}^n$ can be encoded in GS5 satisfiability.The $\mathsf{SUPPSAT}^n$ problem has important applications. For instance, in the privacy preserving data mining where the $\mathsf{SUPPSAT}^n$ problem can be used for trying to reconstruct parts of the original database from data mining outcomes: *inverse data mining* [14] (see [5, 6] for some details). Moreover, the $\mathsf{SUPPSAT}^n$ problem can be used to improve the pruning of infrequent candidates in frequent itemset mining algorithms.

2 Modal Logic S5

2.1 Syntax and Semantics

The set of propositional formulae of S5, denoted $\mathsf{Form}_{\mathsf{S5}}$, is inductively defined from a set of propositional variables, denoted Prop (we use $p, q, r, \dots$ to range over Prop), by using the propositional connectives $\wedge$ and $\neg$ and the modal connective $\Diamond$. The other propositional connectives and the modal connective $\Box$ can be expressed using $\wedge$, $\neg$ and $\Diamond$ as follows: $A \vee B =_{def} \neg(\neg A \wedge \neg B)$, $A \rightarrow B =_{def} \neg A \vee B$ and $\Box A =_{def} \neg\Diamond\neg A$. In other words, the language of S5 extends that of classical propositional logic by the modal connectives $\Box$ and $\Diamond$.

A Hilbert axiomatic system for S5 is given by the following axioms and rules:
0. Any substitution instance of a propositional tautology.
$\mathcal{K}$. $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$ $\quad$ $\mathcal{T}$. $\Box A \rightarrow A$ $\quad$ $\mathcal{B}$. $A \rightarrow \Box\Diamond A$ $\quad$ 4. $\Box A \rightarrow \Box\Box A$

$$\frac{A \rightarrow B \qquad A}{B}\ [mp] \quad \text{and} \quad \frac{A}{\Box A}\ [nec]$$

Le us now define the simplest possible world semantics for S5. We first define the structure of S5 modal model, and then we define a relation between the worlds and $\mathsf{Form}_{\mathsf{S5}}$, called forcing relation, that allows us to define S5 satisfiability problem.

Definition 1 (Modal Model). *A modal model is a couple $\mathcal{M} = (W, V)$ where W is a non-empty set (of worlds) and V is a function from W to $\mathcal{P}(\mathsf{Prop})$, where $\mathcal{P}(\mathsf{Prop})$ is the powerset of* Prop*, i.e. its set of subsets.*

Definition 2 (Forcing Relation). *Let $\mathcal{M} = (W, V)$ be a modal model. The forcing relation, denoted $\vDash_{\mathcal{M}}$, between W and $\mathsf{Form}_{\mathsf{S5}}$ is inductively defined on formula structure as follows:*

$w \vDash_{\mathcal{M}} p$ *iff* $p \in V(w)$*;*
$w \vDash_{\mathcal{M}} A \wedge B$ *iff* $w \vDash_{\mathcal{M}} A$ *and* $w \vDash_{\mathcal{M}} B$*;*
$w \vDash_{\mathcal{M}} \neg A$ *iff* $w \nvDash_{\mathcal{M}} A$*;*
$w \vDash_{\mathcal{M}} \Diamond A$ *iff* $\exists w' \in W$, $w' \vDash_{\mathcal{M}} A$*.*

Definition 3 (S5 Satisfiability Problem). *A formula A is satisfiable in* S5 *iff there exists a modal model $\mathcal{M} = (W, V)$ sand a world $w \in W$ such that $w \vDash_{\mathcal{M}} A$.*

A formula A is valid in $\mathcal{M}$, denoted $\mathcal{M} \vDash A$, if and only if, for all $w \in W$, $w \vDash_{\mathcal{M}} A$. A is a theorem of S5 if and only if A is valid in every modal model. Satisfiability and validity are complementary in S5. In fact, S5 satisfiability (resp. validity) is NP-complete (resp. co-NP-Complete).

We define the size of a formula, denoted $|\cdot|$, as follows:

$$|p| = 1 \qquad |\neg A| = |A| + 1 \qquad |A \wedge B| = |A| + |B| + 1 \qquad |\Diamond A| = |A| + 1$$

Theorem 1 ([13]). *A is satisfiable in* S5 *iff it is satisfiable in a model with at most $|A|$ worlds.*

2.2 Reducing S5 satisfiability to SAT

We first introduce the satisfiability problem and some necessary notations that will be used in the different reductions of our problems to SAT. A *literal* is a positive (p) or negated ($\neg p$) propositional variable. The two literals p and $\neg p$ are called *complementary*. We denote by $\bar{l}$ the complementary literal of l. Let us recall that any propositional formula can be translated to the conjunctive normal form (CNF) using linear Tseitin encoding [20]. We denote by $Var(F)$ the set of propositional variables appearing in F. An *interpretation* $\mathcal{B}$ of a propositional formula F is a function which associates a value $\mathcal{B}(p) \in \{0, 1\}$ (0 corresponds to $false$ and 1 to $true$) to the variables $p \in Var(F)$. A *model* of a formula F is an interpretation $\mathcal{B}$ that satisfies the formula. The SAT problem consists in deciding if a given formula admits a model or not.

Let us now show how to reduce S5 satisfiability to SAT. We first associate to each S5 formula A and strictly positive integer n a propositional formula so that A is satisfiable in a model with n worlds if and only if its associated propositional formula is satisfiable. Thus, using the property that a formula is satisfiable in S5 if and only if it is satisfiable in a model with at most the number of its modal subformulas as the number of worlds, we provide a polynomial reduction of S5 satisfiability to SAT.

We define a *deep modal subformula* as a subformula of the form $\Diamond A$ that is in the scope of only one modal connective. For instance, the deep modal subformulas of $\Diamond(\Diamond A \wedge \Diamond(C \vee \Diamond D))$ are $\Diamond A$ and $\Diamond(C \vee \Diamond D)$.

Definition 4. *Let A be an* $\mathsf{S5}$ *formula,* $\{\Diamond B_1, \ldots, \Diamond B_k\}$ *the set of the deep modal subformulas of A and* $\{x_{\Diamond B_1}, \ldots, x_{\Diamond B_k}\}$ *a set of k fresh variables. We define inductively the* $\mathsf{S5}$ *formula $\mathcal{D}(A)$ as follows:*

- $\mathcal{D}(A) = A$ *if A does not contain any deep modal subformula;*
- $\mathcal{D}(A) = A[\Diamond B_1/x_{\Diamond B_1}, \ldots, \Diamond B_k/x_{\Diamond B_k}] \bigwedge_{i=1}^{k} \mathcal{D}(\Diamond x_{\Diamond B_i} \to \Diamond B_i)$ $\bigwedge_{i=1}^{k} \mathcal{D}(\Diamond B_i \to \neg\Diamond\neg x_{\Diamond B_i})$.

where $A[\Diamond B_1/x_{\Diamond B_1}, \ldots, \Diamond B_k/x_{\Diamond B_k}]$ denotes the result of substituting each deep modal subformula by its associated fresh variable.

It is easy to see that $\mathcal{D}(A)$ is of polynomial size in $|A|$. Intuitively, the two formulae $\Diamond x_{\Diamond B_i} \to \Diamond B_i$ and $\Diamond B_i \to \neg\Diamond\neg x_{\Diamond B_i}$ correspond to $\Box(x_{\Diamond B_i} \leftrightarrow \Diamond B_i)$. Indeed, we have $\Box(x_{\Diamond B_i} \to \Diamond B_i) \equiv \Box(\neg x_{\Diamond B_i} \vee \Diamond B_i) \equiv \neg\Diamond x_{\Diamond B_i} \vee \Box\Diamond B_i \equiv \Diamond x_{\Diamond B_i} \to \Diamond B_i$ (because of $\Box\Diamond B_i \equiv \Diamond B_i$ and $\Box\neg x_{\Diamond B_i} \equiv \neg\Diamond x_{\Diamond B_i}$). Furthermore, $\Box(\Diamond B_i \to x_{\Diamond B_i}) \equiv \Box\neg\Diamond B_i \vee \Box x_{\Diamond B_i} \equiv \neg\Diamond B_i \vee \neg\Diamond\neg x_{\Diamond B_i}$. Thus, one can easily prove that A is satisfiable if and only if $\mathcal{D}(A)$ is satisfiable. Moreover, we have for all model $\mathcal{M}$, if $\mathcal{M} \vDash \mathcal{D}(A)$ then $\mathcal{M} \vDash A$. In fact, it is well-known that one can disallow nested modalities in $\mathsf{S5}$ without any loss of generality. Here, we use an approach similar to that using in Tseitin encoding in propositional logic [20].

Definition 5. *Let A be an* $\mathsf{S5}$ *formula and n a strictly positive integer. We associate to each propositional variable p of A and $i \in \{1, \ldots, n\}$ a new propositional variable p_i. Then, we define the propositional formula $(A)_n^i$ by induction on the structure of A as follows:*

$$(p)_n^i = p_i \qquad (B \wedge C)_n^i = (B)_n^i \wedge (C)_n^i$$
$$(\neg B)_n^i = \neg(B)_n^i \qquad (\Diamond B)_n^i = \bigvee_{1 \leqslant j \leqslant n} (B)_n^j$$

Let us note that $(A)_n^i$ is not always of polynomial size in $|A|$ (we consider that n is polynomial in $|A|$). For instance, we have with a naive enconding:

$$(\overbrace{\Diamond \cdots \Diamond}^{k\ times} p)_n^i = \overbrace{(p_1 \vee \cdots \vee p_n) \vee \cdots \vee (p_1 \vee \cdots \vee p_n)}^{k^{n-1}\ times}$$

However, one can see that the propositional formula $(\mathcal{D}(A))_n^i$ is in the polynomial size of $|A|$. Moreover, $(\mathcal{D}(A))_n^i$ is satisfiable if and only if $(A)_n^i$ is satisfiable, and if an interpretation $\mathcal{B}$ is a model of $(\mathcal{D}(A))_n^i$ then it is also a model of $(A)_n^i$.

Proposition 1. *A is* $\mathsf{S5}$ *satisfiable in a modal model of n worlds iff $(\mathcal{D}(A))_n^1$ is satisfiable.*

Proof. We only have to prove that A is $\mathsf{S5}$ satisfiable in a modal model of n worlds iff $(A)_n^1$ is satisfiable.
Part $\Rightarrow$. Let $\mathcal{M} = (W, V)$ be a modal model with n worlds such that there exists w in W where $w \vDash_{\mathcal{M}} A$. Let f be a function associating to each integer $i \in \{1, \ldots, n\}$ a world in W such that $f(1) = w$ and for all $i, j \in \{1, \ldots, n\}$, if $i \neq j$ then $f(i) \neq f(j)$. We define the boolean interpretation $\mathcal{B}$ as follows: $\mathcal{B}(p_i)$ is equal to 1 if $p \in V(f(i))$, 0 otherwise. One can easily prove by simultaneous induction on formula structure that for all $F \in \mathsf{Form}_{\mathsf{S5}}$ and for all $i \in \{1, \ldots, n\}$:

- if $f(i) \models_{\mathcal{M}} F$ then $\mathcal{B}((F)^i_n) = 1$; and
- if $f(i) \not\models_{\mathcal{M}} F$ then $\mathcal{B}((F)^i_n) = 0$.

Therefore, we deduce that $\mathcal{B}$ is a model of $(A)^1_n$.
Part $\Leftarrow$. Let $\mathcal{B}$ be a model of $(A)^1_n$ and $\mathcal{M} = (\{1, \dots, n\}, V)$ a modal model such that for all $i \in \{1, \dots, n\}$, $V(i) = \{p \mid B(p_i) = 1\}$. Similarly to Part $\Rightarrow$, one can easily prove by simultaneous induction on formula structure that for all $F \in \mathsf{Form}_{\mathsf{S5}}$ and for all $i \in \{1, \dots, n\}$:

- if $\mathcal{B}((F)^i_n) = 1$ then $i \models_{\mathcal{M}} F$; and
- if $\mathcal{B}((F)^i_n) = 0$ then $i \not\models_{\mathcal{M}} F$.

Since $\mathcal{B}$ satisfies $(A)^1_n$, we deduce that $1 \models_{\mathcal{M}} A$, and consequently $\mathcal{M}$ satisfies A.

Theorem 2. *A is* S5 *satisfiable iff $\bigvee_{1 \leqslant i \leqslant |A|} (\mathcal{D}(A))^1_i$ is satisfiable.*

Proof. It is a direct consequence of Theorem 1 and Proposition 1.

3 Graded Modal Logic GS5

3.1 Syntax and Semantics

The graded modal logic GS5 corresponds to an extension of S5. Indeed, in this logic the modal connective $\Diamond$ is generalized to a new one $\Diamond_\lambda$ where λ is a strictly positive integer. The set of formulae of GS5 ($\mathsf{Form}_{\mathsf{GS5}}$) is given by the following grammar: $A, B ::= p \mid \neg A \mid A \wedge B \mid \Diamond_\lambda A$.
Semantically, the formula $\Diamond_\lambda A$ is satisfied in a model if and only if there exist at least λ different worlds satisfying A. Thus, such a modal logic allows to reason on propertiy supports. Concerning the dual modal connective, $\Box_\lambda A$ is satisfied if and only if there exist less than λ distinct worlds satisfying $\neg A$. Like in the case of S5, GS5 satisfiability is NP-complete.

The definition of forcing relation in the case of $\Diamond_\lambda$ is as follows:

$$w \models_{\mathcal{M}} \Diamond_\lambda A \quad \text{iff there exist } \lambda \text{ distinct worlds } w_1, \dots, w_\lambda \text{ in } W \text{ such that } w_1 \models_{\mathcal{M}} A, \dots, w_\lambda \models_{\mathcal{M}} A.$$

The formulae size $|\cdot|$ is extended to $\mathsf{Form}_{\mathsf{GS5}}$ by $|\Diamond_\lambda A| = |A| + \lambda$.

Theorem 3 ([21]). *A is* GS5 *satisfiable iff it is satisfiable in a model with at most $|A|$ worlds.*

Definition 6. *Let A be a* GS5 *formula and n a strictly positive integer. The* GS5 *formula $s(A, n)$ is inductively defined on the structure of A as follows:*

$$\begin{array}{ll} s(p, n) = p & s(B \wedge C, n) = s(B, n) \wedge s(C, n) \\ s(\Diamond_\lambda B, n) = \Diamond_\lambda s(B, n) & s(\neg p, n) = \neg p \\ s(\neg\neg B, n) = s(B, n) & s(\neg(B \wedge C), n) = s(\neg B, n) \vee s(\neg C, n) \\ s(\neg \Diamond_\lambda B, n) = \Diamond_{n-\lambda+1} s(\neg B, n) & \end{array}$$

Proposition 2. *A is* GS5 *satisfiable in a model $\mathcal{M}$ with n worlds iff $s(A, n)$ is* GS5 *satisfiable in $\mathcal{M}$*

Proof. By induction on the structure of A.

3.2 Reducing GS5 satisfiability to SAT

In the case of $\Diamond_\lambda$ in GS5 satisfiablity, we have to determine if a formula is satisfied in at least λ worlds. In our reduction (Definition 7), such counting argument is encoded using the well known Horn cardinality constraint $\sum_{j=1}^{n} x_j \geqslant \lambda$. It is important to note that this kind of constraints and its generalized form $\sum_{j=1}^{n} a_j x_j \geqslant \lambda$ (where a_j are positive integers) can be polynomialy encoded as a propositional formula in CNF [12]. As mentioned by J. P. Warners in [12], the first polynomial CNF expansion of Horn cardinality constraint is first proposed by Hooker in an unpublished note. The Hooker encoding of $\sum_{j=1}^{n} x_j \geqslant \lambda$ to CNF is obtained as follows:

$$\neg z_{ik} \vee x_i, \quad i = 1, \ldots, n \quad k = 1, \ldots, \lambda \tag{1}$$

$$\bigvee_{i=1}^{n} z_{ik}, \quad k = 1, \ldots, \lambda \tag{2}$$

$$\neg z_{ik} \vee \neg z_{ik'}, \quad i = 1, \ldots, n \quad k, k' = 1, \ldots, \lambda, k \neq k' \tag{3}$$

The two equations 2 and 3 encode the pigeon hole problem $PH_{\lambda,n}$, where λ is the number of pigeons and n is the number of holes (z_{ik} expresses that pigeon k is in hole i). The mapping between the models of $PH_{\lambda,n}$ and those of $\sum_{j=1}^{n} x_j \geqslant \lambda$ are obtained thanks to the equation 1. In this elegant transformation, the number of additional variables is $\lambda \times n$ and the number of clauses required is $\frac{1}{2}\lambda(n^2+n+2)$. Let us mention that in [12] (see Section 5), the equation 3 is written as follows:

$$\neg z_{ik} \vee \neg z_{jk}, \quad i, j = 1, \ldots, n, i \neq j \quad k = 1, \ldots, \lambda. \tag{4}$$

One can obviously check that the formulation given in [12], where the equation 3 is substituted by the equation 4 is not correct. Even the description associated to the formulation in Warners' paper is not correct. This brief recall allows us to give a correct formulation described by the three equations 1, 2 and 3. Several improvements of the CNF encoding of both Horn cardinality constraints (e.g. [19, 18]) have been proposed since 1996. In these recent encodings the propagation capabilities of the obtained CNF has been significantly enhanced.

Let us now introduce our reduction of GS5 satisfiability to SAT. To this end, we use an approach similar to that used in the case of S5 satisfiability.

Definition 7. *Let A be a GS5 formula and n a strictly positive integer. We associate to each propositional variable p and $i \in \{1, \ldots, n\}$ a new propositional variable p_i. We define the propositional formula $(s(A, n))^i$ by induction on the structure of $s(A, n)$ as follows:*

$(p)^i = p_i \qquad (\neg p)^i = \neg p_i$

$(B \wedge C)^i = (B)^i \wedge (C)^i \qquad (B \vee C)^i = (B)^i \vee (C)^i$

$$(\Diamond_\lambda B)^i = \begin{cases} \bot & if \quad \lambda > n \\ (\bigwedge_{j=1}^{n} x_j \leftrightarrow (B)^j) \wedge \sum_{j=1}^{n} x_j \geqslant \lambda & otherwise \end{cases}$$

where $x_1, \ldots, x_n$ are n fresh propositional variables.

Similarly to $(A)_n^i$, $(s(A,n))^i$ is not always polynomial in $|A|$. In order to obtain an equivalent polynomial size formula, we only have to proceed, like in the case of $\mathsf{S5}$ (see Definition 4), by introducing the formula $\mathcal{D}(A)$ obtained by substituting inductively the deep modal subformulas by fresh variables:

- $\mathcal{D}(A) = A$ if A does not contain any deep modal subformula;
- $\mathcal{D}(A) = A[\Diamond_{\lambda_1} B_1/x_{\Diamond_{\lambda_1} B_1}, \ldots, \Diamond_{\lambda_k} B_k/x_{\Diamond_{\lambda_k} B_k}] \bigwedge_{i=1}^{k} \mathcal{D}(\Diamond_1 x_{\Diamond_{\lambda_i} B_i} \rightarrow \Diamond_{\lambda_i} B_i)$ $\bigwedge_{i=1}^{k} \mathcal{D}(\Diamond_{\lambda_i} B_i \rightarrow \neg\Diamond_1 \neg x_{\Diamond_{\lambda_i} B_i})$.

Proposition 3. *A is $\mathsf{GS5}$ satisfiable in a modal model of n worlds iff $(\mathcal{D}(s(A,n)))^1$ is satisfiable.*

Proof. Similar to the proof of Proposition 1.

As a consequence of Proposition 3 and Theorem 3, we obtain the following theorem:

Theorem 4. *A is $\mathsf{GS5}$ satisfiable iff $\bigvee_{1 \leqslant i \leqslant |A|} (s(A,i))^1$ is satisfiable.*

4 Itemset Support Satisfiability

4.1 Preliminary

Let $\mathcal{I}$ be a finite set of *items*, a set $I \subseteq \mathcal{I}$ is called an *itemset*. A *transaction* is a couple (tid, I) where *tid* is a *transaction identifier* and I is an itemset. A *transaction database* $\mathcal{D}$ is a finite set of transactions over $\mathcal{I}$ where for all two different transactions, they do not have the same identifier. We say that a transaction (tid, I) *supports* an itemset J if $J \subseteq I$.

The *support* of an itemset I in $\mathcal{D}$ is defined by: $\mathcal{S}(I, \mathcal{D}) = |\{tid \mid (tid, J) \in \mathcal{D} \ \texttt{and} \ I \subseteq J\}|$. Moreover, the frequency of I in $\mathcal{D}$ is defined by:
$\mathcal{F}(I, \mathcal{D}) = \frac{\mathcal{S}(I,\mathcal{D})}{|\mathcal{D}|}$.

Let us consider the following example of transaction database over the set of items $\mathcal{I} = \{Spaghetti, Tomato, Parmesan, Beef, Olive\ oil\}$:

tid	itemset
1	$Spaghetti, Tomato, Olive\ oil$
2	$Spaghetti, Parmesan, Olive\ oil$
3	$Spaghetti, Olive\ oil$
4	$Salad, Olive\ oil$
5	$Spaghetti, Beef, Olive\ oil$

Transaction database $\mathcal{D}$

For instance, we have $\mathcal{S}(\{Spaghetti, Olive\ oil\}, \mathcal{D}) = |\{1,2,3,5\}| = 4$ and $\mathcal{F}(\{Spaghetti, Olive\ oil\}, \mathcal{D}) = \frac{4}{5}$.

It is easy to see that a transaction databases can be considered as a modal model where the items corresponds to propositional variables and the transactions to worlds. For example, the transaction database in the previous example corresponds to the model $\mathcal{M} = (\{1, 2, 3, 4, 5\}, V)$ where $(w, V(w)) \in \mathcal{D}$ (e.g. $V(1) = \{Spaghetti, Tomato, Olive\ oil\}$). We denote by $\mathcal{M}_{\mathcal{D}}$ (resp. $\mathcal{D}_{\mathcal{M}}$) the modal model (resp. the transaction database) associated to the transaction database (resp. the modal model) $\mathcal{D}$ (resp. $\mathcal{M}$).

Using the fact that each transaction database has a corresponding modal model, one can encode some transaction database problems in GS5. Thus, the verification techniques in GS5, in particular using SAT, can be used to prove properties on such problems. For instance, let us consider the problem of computing frequent itemsets which is one of the most studied problems in data mining. Let $\mathcal{D}$ be a transaction database over $\mathcal{I}$ and λ a minimal support threshold. The frequent itemset mining problem consists in computing the following set:

$$\mathcal{FIM}(\mathcal{D}, \lambda) = \{I \subseteq \mathcal{I} \mid \mathcal{S}(I, \mathcal{D}) \geqslant \lambda\}$$

It can be encoded in GS5 as follows: $I \in \mathcal{FIM}(\mathcal{D}, \lambda) \Leftrightarrow \mathcal{M}_{\mathcal{D}} \models \Diamond_{\lambda} \bigwedge_{i \in I} i$. For example, in order to prove that for all transaction database, for all minimal support threshold λ and for all itemsets I and J, if $I \subseteq J$ and $J \in \mathcal{FIM}(\mathcal{D}, \lambda)$ then $I \in \mathcal{FIM}(\mathcal{D}, \lambda)$ (anti-monotonicity), it suffices to prove that $F = \Diamond_{\lambda}(p \wedge q) \rightarrow \Diamond_{\lambda} p$ is a theorem of GS5. Indeed, if F is not a theorem then there exists a transaction database (a modal model) where the anti-monotonicity property is not true. Otherwise, it suffices to notice that if $I \subseteq J$ then there exists K such that $\bigwedge_{j \in J} j \equiv (\bigwedge_{i \in I} i) \wedge (\bigwedge_{k \in K} k)$. Since $\mathcal{M}_{\mathcal{D}} \models \Diamond_{\lambda} \bigwedge_{j \in J} j$ and F is a theorem of GS5, we deduce that $\mathcal{M}_{\mathcal{D}} \models \Diamond_{\lambda} \bigwedge_{i \in I} i$ and then $I \in \mathcal{FIM}(\mathcal{D}, \lambda)$.

4.2 SUPPSAT^n Problem

A *support constraint* is an expression of the form $I \lhd k$ where I is a an itemset, $\lhd \in \{\leqslant, \geqslant\}$ and k is a natural number. A transaction database $\mathcal{D}$ satisfies a support constraint $I \lhd k$ if and only if $\mathcal{S}(I, \mathcal{D}) \lhd k$.

*Problem 1 (*SUPPSAT^n*).* Given a set of support constraints $\mathcal{E}$, determine if there exists a database $\mathcal{D}$ with n transaction satisfying all the constraints in $\mathcal{E}$.

Now, let us encode SUPPSAT^n in GS5 satsifiability (and consequently in SAT).

Definition 8. *Let $\mathcal{E} = \{I_1 \lhd_1 k_1, \ldots, I_m \lhd_m k_m\}$ be a set of support constraints. The* GS5 *formula associated to $\mathcal{E}$, denoted $F_{\mathcal{E}}$, is $\bigwedge_{i=1}^{m} F(I_i \lhd_i k_i)$ where $F(I \geqslant k) = \Diamond_k \bigwedge_{j \in I} j$ and $F(I \leqslant k) = \neg\Diamond_{k+1} \bigwedge_{j \in I} j$.*

Proposition 4. *A set of support constraints $\mathcal{E}$ is* SUPPSAT^n *iff $F_{\mathcal{E}} \wedge \neg\Diamond_{n+1}\top$ is* GS5 *satsifiable.*

Proof.
- *Part* $\Rightarrow$. Let $\mathcal{D}$ be a database with n transaction satisfying $\mathcal{E}$, then it is easy to see that $\mathcal{M}_{\mathcal{D}}$ satisfies $F_{\mathcal{E}}$. Moreover, since $\mathcal{M}_{\mathcal{D}}$ contains n worlds, it satisfies

also $\neg\Diamond_{n+1}\top$.
- *Part* $\Leftarrow$. Let $\mathcal{M}$ be a model satisfying $F_{\mathcal{E}} \wedge \neg\Diamond_{n+1}\top$. Then $\mathcal{D}_{\mathcal{M}}$ is a database with m transactions satisfying $\mathcal{E}$ where $m \leqslant n$ because of $\neg\Diamond_{n+1}\top$. Let $\mathcal{D}'$ be the transaction database obtained from $\mathcal{D}$ by adding $n-m$ empty transactions (transaction without items). It is trivial that $\mathcal{D}'$ is a database with n transactions satisfying $\mathcal{E}$.

5 Acknowledgment

We thank the CNRS (Centre National de la Recherche Scientifique) and the ANR (Agence Nationale de la Recherche) for supporting this work.

6 Conclusion and Perspectives

In this work, we study the satisfiability problem in the modal logic S5 and its graded version GS5. We provide reductions of S5 satsifiability and GS5 satisfiability to propositional satisfiability (SAT). The key point is that a modal formula is satisfiable in S5 or GS5 if and only if it is satisfiable in a model with at most the number of the modal subformulas as the number of worlds. Moreover, in order to show how GS5 satisfiablity can be used in data mining, we consider a satisfiability problem related to the known frequent itemset mining problem, called itemset support satisfiability and denoted by $\mathsf{SUPPSAT}^n$. We show how $\mathsf{SUPPSAT}^n$ can be encoded simply in GS5 satisfiability.

In further works we will study the possibility to develop a suitable resolution method for GS5 satisfiability problem. In this context, we think that we have to find a useful form for the modal formulas similar to conjunctive normal form in the propositional case. We also have to propose additional rules to deal with the modal connectives. Such a method can be more appropriate for GS5 satisfiability and also for itemset support satisfiability than a reduction to SAT. Moreover, we will study the satisfiability problem in some extensions of GS5 with new connectives, for example, with the connective $>>$ where $A >> B$ means that the number of worlds satisfying A are greater than the number of those satisfying B. Such a connective will allow us to reason on relations between property supports.

References

1. R. Agrawal, T. Imielinski, and A. N. Swami. Mining association rules between sets of items in large databases. In *SIGMOD*, pages 207–216. ACM Press, 1993.
2. C. Areces, G. Hoffmann, and Al. Denis. Modal logics with counting. In *WoLLIC*, volume 6188 of *LNCS*, pages 98–109. Springer, 2010.
3. F. Baader, M. Buchheit, and B. Hollunder. Cardinality restrictions on concepts. *Artificial Intelligence*, 88(1-2):195–213, 1996.
4. P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, 2001.

5. T. Calders. *Axiomatization and Deduction Rules for the Frequency of Itemsets*. PhD Thesis. Universiteit Antwerpen, 2003.
6. T. Calders. Itemset frequency satisfiability: Complexity and axiomatization. *Theoretical Computer Science*, 394(1-2):84–111, 2008.
7. K. Fine. Cut-free modal sequents for normal modal logics. *Notre-Dame Journal of Formal Logic*, 13(4):516–520, 1972.
8. M. Fitting. *Proof methods for modal and intuitionistic logics*, volume 169 of *Synthese Library*. Kluwer, 1983.
9. M. Fitting. A simple propositional S5 tableau system. *the Annals of Pure and Applied Logic*, 96:107–115, 1999.
10. E. Giunchiglia, F. Giunchiglia, R. Sebastiani, and A. Tacchella. Sat vs. Translation Based decision procedures for modal logics: a comparative evaluation. *Journal of Applied Non-Classical Logics*, 10(2), 2000.
11. E. Giunchiglia, A. Tacchella, and F. Giunchiglia. Sat-based decision procedures for classical modal logics. *J. Autom. Reasoning*, 28(2):143–171, 2002.
12. j. P. Warners. A linear-time transformation of linear inequalities into conjunctive normal form. *Information Processing Letters*, 1996.
13. R. E. Ladner. The computational complexity of provability in systems of modal propositional logic. *SIAM Journal Computation*, 6(3):467–480, 1977.
14. T. Mielikäinen. On Inverse Frequent Set Mining. In *2nd IEEE ICDM Workshop on Privacy Preserving Data Mining (PPDM)*, pages 18–23. IEEE, 2003.
15. S. Negri. Proof analysis in modal logic. *Journal of Philosophical Logic*, 34:507–534, 2005.
16. H. J. Ohlbach. Semantics-based translation methods for modal logics. *J. Log. Comput.*, 1(5):691–746, 1991.
17. R. Sebastiani and D. McAllester. New upper bounds for satisfiability in modal logics – the case-study of modal K, 1997.
18. João P. Marques Silva and Inês Lynce. Towards robust cnf encodings of cardinality constraints. In *CP*, pages 483–497, 2007.
19. Carsten Sinz. Towards an optimal cnf encoding of boolean cardinality constraints. In *11th International Conference on Principles and Practice of Constraint Programming - CP 2005*, pages 827–831, 2005.
20. G.S. Tseitin. On the complexity of derivations in the propositional calculus. In H.A.O. Slesenko, editor, *Structures in Constructives Mathematics and Mathematical Logic, Part II*, pages 115–125, 1968.
21. W. van der Hoek and M. de Rijke. Counting Objects. *Journal of Logic and Computation*, 5(3):325–345, 1995.
22. H. Wansing. Sequent systems for modal logics. In D. Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic, 2nd edition*, volume 8, pages 61–145. Kluwer, Dordrecht, 2002.