De la déduction dans le fragment $\{\exists, \wedge, \neg_a\}$ de la logique du premier ordre à SAT

Khalil Ben Mohamed, Michel Leclère, Marie-Laure Mugnier

LIRMM, Université de Montpellier,
161 rue Ada, 34392 Montpellier Cedex 5
{benmohamed,leclere,mugnier}@lirmm.fr

Résumé : nous considérons le problème de déduction dans le fragment existentiel conjonctif muni de la négation atomique en logique du 1^{er} ordre. Nous proposons une réécriture de ce problème en un problème de test de validité d'une formule propositionnelle, tout en exploitant les résultats antérieurs obtenus sur un problème équivalent, celui de l'inclusion de requêtes conjonctives avec négation (Leclère & Mugnier, 2007). Ces résultats sont basés sur la notion d'homomorphisme de graphes. Premièrement, nous faisons une synthèse de ces résultats et les reformulons dans un cadre logique. Deuxièmement, nous présentons notre nouvelle approche qui conduit à tester la validité d'une forme disjonctive propositionnelle, autrement dit l'insatisfiabilité d'une forme clausale propositionnelle, ce qui permet d'utiliser un solveur SAT.

Mots-clés : logique du 1^{er} ordre, déduction, négation atomique, SAT.

1 Introduction

Dans ce papier, nous nous intéressons à la résolution du problème de *déduction dans le fragment existentiel conjonctif* de la logique du 1^{er} ordre (sans fonctions) : « Étant données deux conjonctions de littéraux fermées existentiellement F et G, F se déduit-elle de G (noté $G \models F$) ? » Ce problème est Π_2^P-$complet$[1]. Il peut être reformulé sous la forme de deux problèmes fondamentaux en informatique : le problème d'inclusion de requêtes en base de données (Abiteboul *et al.*, 1995), fondamental pour l'optimisation des mécanismes d'évaluation de requêtes (Chandra & Merlin, 1977) ou la réécriture de requêtes avec vues (Halevy, 2001), et le problème d'implication de clauses qui est à la base des techniques d'ILP (Inductive Logic Programming), domaine à la croisée de l'apprentissage automatique et de la programmation logique (Muggleton, 1992)(Lavrac & Dzeroski, 1994). Dans le cadre du problème d'inclusion de requêtes, nous considérons des requêtes conjonctives avec négation, i.e. de la forme $ans(u) \leftarrow r_1(u_1), \ldots, r_n(u_n), \neg s_1(y_1), \ldots, \neg s_m(y_m)$, où $n \geq 1$ et $m \geq 0$. Étant données deux requêtes conjonctives avec négation q_1 et q_2, il s'agit de déterminer si q_1 est incluse dans q_2, c'est-à-dire si l'ensemble des réponses à q_1 est inclus dans l'ensemble

[1] $\Pi_2^P = (co\text{-}NP)^{NP}$

des réponses à q_2 pour toute base de données (voir (Ullman, 1997)(Wei & Lausen, 2003)(Leclère & Mugnier, 2007)). Le problème d'implication de clauses est le suivant : « Étant données deux clauses C_1 et C_2, C_1 implique-t-elle C_2, i.e. C_2 se déduit-elle de C_1 ? » Si nous considérons des clauses du premier ordre sans fonctions (comme dans (Gottlob, 1987)), nous obtenons par contraposition une instance du problème de déduction dans le fragment existentiel conjonctif de la logique du 1^{er} ordre sans fonctions. Enfin, si l'on ajoute un ordre partiel sur les prédicats, on obtient exactement le problème de déduction dans les graphes conceptuels polarisés (Leclère & Mugnier, 2006).

Dans (Leclère & Mugnier, 2007), le problème d'inclusion de requêtes conjonctives avec négation est étudié par le biais de la notion d'homomorphisme de graphes. Des cas particuliers de requêtes où la complexité du problème décroît sont mis en exergue, et un algorithme de résolution pour le cas général, améliorant ceux existant en base de données, est proposé. Ces résultats seront rapidement présentés dans la section 2, reformulés en termes logiques.

Dans cet article, nous proposons une nouvelle façon de résoudre le problème : nous le reformulons en un problème de test de validité d'une formule propositionnelle, tout en exploitant la notion d'homomorphisme et les résultats antérieurs. La formule obtenue est une forme disjonctive propositionnelle. Sa négation est donc une forme clausale propositionnelle, dont nous pouvons tester l'insatisfiabilité en utilisant un solveur SAT. Cette nouvelle approche sera présentée en section 3.

2 Contexte de l'étude

Nous allons dans un premier temps rappeler les définitions de base et faire une rapide synthèse des résultats de (Leclère & Mugnier, 2007).

2.1 Notions de base

Nous nous plaçons dans le fragment conjonctif existentiel de la logique du 1^{er} ordre muni de la négation atomique, que nous noterons $FOL\{\exists, \wedge, \neg_a\}$, et nous ne considérons pas de symbole de fonctions hormis des constantes.

Définition 1 (Formule et sous-formule)
Une formule *de* $FOL\{\exists, \wedge, \neg_a\}$ *est une conjonction de littéraux positifs ou négatifs fermée existentiellement. On la verra aussi comme un ensemble* $F = \{l_1, \ldots, l_p\}$*, où* l_i *est un littéral pour* $i = 1 \ldots p$*, composé d'un symbole de prédicat d'arité* k *(ou de sa négation) et d'un* k*-tuple de termes. Une* sous-formule F' *de* F *est alors un ensemble de littéraux tel que* $F' \subseteq F$*.*

Définition 2 (Formule consistante (ou satisfiable))
Une formule de $FOL\{\exists, \wedge, \neg_a\}$ *est* consistante *si elle ne contient pas deux littéraux opposés.*

Dans la suite, F et G représenteront deux formules consistantes de $FOL\{\exists, \wedge, \neg_a\}$.

De façon classique, une substitution de F dans G est une application de l'ensemble des variables de F dans l'ensemble des termes (variables ou constantes) de G. Nous l'étendons à une application des termes de F dans les termes de G, une constante ayant pour image elle-même. On note $s(F)$ l'application d'une substitution s à une formule F. Nous définissons un homomorphisme comme une substitution particulière :

Définition 3 (Homomorphisme)
Un homomorphisme h *de* F *dans* G *est une substitution de* F *dans* G *telle que* $h(F) \subseteq G$.

2.2 Résultats antérieurs

Nous transposons les résultats précédents établis pour le problème d'inclusion de requêtes conjonctives (Leclère & Mugnier, 2007) sous la forme déduction de formules conjonctives existentielles afin de faciliter la comparaison avec la nouvelle approche proposée.

Pour des formules uniquement positives F et G, $G \models F$ si et seulement si il existe un homomorphisme de F dans G. Par contre, dès que l'on considère la négation atomique, un seul sens de la propriété reste vrai :

Propriété 1
Soient deux formules F *et* G *de* $FOL\{\exists, \wedge, \neg_a\}$*, s'il existe un homomorphisme de* F *dans* G *alors* $G \models F$*.*

En effet, l'homomorphisme ne suffit plus pour répondre au problème de déduction :

Exemple 1
Considérons les formules $F = \exists u \exists v\ (p(u) \wedge r(u,v) \wedge \neg p(v))$ *et* $G = \exists w \exists x \exists y \exists z\ (p(w) \wedge r(w,x) \wedge r(x,y) \wedge r(y,z) \wedge \neg p(z))$*. Il n'y a pas d'homomorphisme de* F *dans* G *alors que* $G \models F$ *: il suffit d'introduire* $(p(x) \vee \neg p(x)) \wedge (p(y) \vee \neg p(y))$ *dans* G *pour s'en persuader. On obtient alors la formule* G' *(équivalente à* G*) suivante :* $G' = \exists w \exists x \exists y \exists z\ ((p(w) \wedge r(w,x) \wedge r(x,y) \wedge r(y,z) \wedge \neg p(z) \wedge p(x) \wedge p(y)) \vee (p(w) \wedge r(w,x) \wedge r(x,y) \wedge r(y,z) \wedge \neg p(z) \wedge \neg p(x) \wedge p(y)) \vee (p(w) \wedge r(w,x) \wedge r(x,y) \wedge r(y,z) \wedge \neg p(z) \wedge p(x) \wedge \neg p(y)) \vee (p(w) \wedge r(w,x) \wedge r(x,y) \wedge r(y,z) \wedge \neg p(z) \wedge \neg p(x) \wedge \neg p(y)))$*. Chacune des quatre conjonctions de* G' *correspond à une façon de compléter* G *par rapport au prédicat* p*. On peut vérifier qu'il existe un homomorphisme de* F *dans chacune de ces quatre conjonctions.* F *se déduit donc de* G'*.*

Cet exemple servira de fil rouge à notre présentation et sera complété au fur et à mesure. Une façon de résoudre le problème consiste donc à générer toutes les formules conjonctives « complètes » que l'on peut obtenir à partir de G en utilisant les prédicats apparaissant dans G, puis à tester s'il existe un homomorphisme de F dans chacune de ces formules.

Définition 4 (Formule complète)
Une formule consistante G *est dite* complète *par rapport à un ensemble de prédicats* $\mathcal{P}$ *si pour chaque prédicat* p *de* $\mathcal{P}$ *d'arité* k *et chaque k-tuple de termes (non nécessairement distincts)* $t_1, \ldots, t_k$ *de* G*,* G *contient soit* $p(t_1, \ldots, t_k)$ *soit* $\neg p(t_1, \ldots, t_k)$*.*

Définition 5 (Complétion)
Une complétion G' *de* G *est une formule obtenue de* G *par ajouts successifs de nouveaux littéraux (composés de prédicats et de termes apparaissant déjà dans* G*) aussi longtemps qu'aucune inconsistance n'apparaît (*$G \subseteq G'$*). Chaque ajout représente une* étape de complétion. *Une complétion de G est dite* totale *si c'est une formule complète par rapport à l'ensemble de prédicats considéré.*

Théorème 1
Soient deux formules F *et* G *(avec* G *consistante),* $G \models F$ *si et seulement si quelque soit* G^c*, une formule complète obtenue de* G *par rapport à l'ensemble de prédicats apparaissant dans* G*, il existe un homomorphisme de* F *dans* G^c*.*

Une première propriété de (Leclère & Mugnier, 2007) est que les prédicats n'apparaissant pas à la fois en positif et en négatif dans F et dans G ne sont pas utiles. On peut donc restreindre l'ensemble de prédicats considéré à un ensemble appelé le vocabulaire de complétion :

Définition 6 (Vocabulaire de complétion)
Soient deux formules F *et* G *de* $FOL\{\exists, \wedge, \neg_a\}$*, un* prédicat de complétion *de* F *et* G *est un prédicat apparaissant dans des littéraux de signe opposé, à la fois dans* F *et dans* G*. Le* vocabulaire de complétion *est composé de l'ensemble des prédicats de complétion.*

Une approche brutale (introduite dans (Ullman, 1997)) consiste à calculer l'ensemble des formules complètes de G et à effectuer un test d'homomorphisme de F dans chaque formule complète obtenue. Néanmoins, la complexité d'une telle méthode est exorbitante : $\mathcal{O}(2^{(n_G)^k \times |\mathcal{V}|} \times hom(F, G^c))$, où n_G est le nombre de termes dans G, k est l'arité maximum d'un prédicat de G, $\mathcal{V}$ est le vocabulaire de complétion considéré et $hom(F, G^c)$ est la complexité du test d'existence d'un homomorphisme[2] de F dans G^c.

Deux types d'amélioration de cette méthode sont proposés dans (Leclère & Mugnier, 2007). Premièrement, une exploration incrémentale de l'espace de recherche menant de G à ses complétions totales est effectuée. Cet espace de recherche est partiellement ordonné par la relation d'inclusion. Il est exploré sous la forme d'une arborescence binaire de racine G. Les fils d'un noeud sont obtenus en ajoutant à la formule associée à ce noeud (soit G') un littéral sous sa forme positive et sous sa forme négative (chacune des deux nouvelles formules est donc obtenue par une étape de complétion à partir de G'). Au lieu de construire et tester toutes les complétions totales de G, on recherche un ensemble de complétions partielles couvrant ces complétions totales, i.e. la question de savoir s'il existe un homomorphisme de F dans chaque complétion totale de G devient : « Existe-t-il un ensemble de complétions partielles $\{G_1, \ldots, G_n\}$ tel que (1) il existe un homomorphisme de F dans chaque G_i pour $i = 1 \ldots n$; (2) chaque complétion totale G^c de G est couverte par un G_i, i.e. $G_i \subseteq G^c$? » Après chaque étape de complétion, un test d'homomorphisme de F dans la complétion courante est effectué ; si le résultat est positif, cette complétion partielle est l'un des G_i recherchés, sinon l'exploration continue. La figure 1 donne un aperçu de cette méthode sur le cas très

[2] Ce problème est $NP-Complet$. Il est borné par $n_G^{n_F}$, où n_F est le nombre de termes de F.

simple de l'exemple 1. On crée deux nouvelles formules G_1 et G_2, en ajoutant respec-

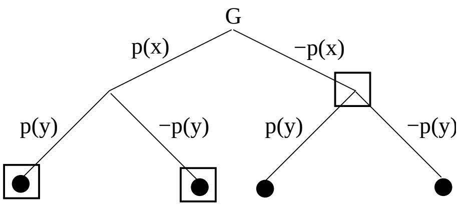

FIG. 1 – Exemple d'arbre de recherche de l'exemple 1. Chaque point noir représente un G^c et chaque carré un G_i.

tivement $p(x)$ et $\neg p(x)$ à G. Il y a un homomorphisme de F dans G_2, on arrête donc de compléter G_2. Il n'y a pas d'homomorphisme de F dans G_1 : on crée deux nouvelles formules G_3 et G_4, en ajoutant respectivement $p(y)$ et $\neg p(y)$ à G_1. Il y a un homomorphisme de F dans G_3 et de F dans G_4. Finalement, l'ensemble prouvant que F se déduit de G est $\{G_2, G_3, G_4\}$ (alors qu'il existe 4 complétions totales de G par rapport à p). Un deuxième niveau d'amélioration consiste à identifier des sous-formules de F pour lesquelles il existe nécessairement un homomorphisme dans G quand $G \models F$. Les deux objectifs principaux liés à de telles sous-formules sont (1) de mettre en place un filtrage (s'il n'y a pas d'homomorphisme d'une de ces sous-formules dans G alors $G \not\models F$) ; (2) de les utiliser comme heuristique de choix du littéral à ajouter lors de la prochaine étape de complétion. Pour améliorer le filtrage et pouvoir les utiliser comme heuristique, on impose que l'homomorphisme de ces sous-formules soit « compatible » avec la formule entière (i.e. que l'homomorphisme de ces sous-formules constitue un homomorphisme de F dans une complétion de G). Cette condition est assurée par la notion d'homomorphisme partiel :

Définition 7 (Homomorphisme partiel)
Soient deux formules F et G de $FOL\{\exists, \wedge, \neg_a\}$ et F' une sous-formule de F, une substitution h de F dans G est un homomorphisme partiel *de F dans G par rapport à F' si :*
- *$h(F') \subseteq G$, i.e. h est un homomorphisme de F' dans G.*
- *pour chaque littéral de prédicat p sur $(c_1, \ldots, c_k)$ de $F - F'$, il n'existe pas un littéral de signe opposé de prédicat p sur $(h(c_1), \ldots, h(c_k))$ dans G.*
- *pour chaque paire de littéraux de signe opposé de prédicat p respectivement sur $(c_1, \ldots, c_k)$ et $(d_1, \ldots, d_k)$ de $F - F'$, $(h(c_1), \ldots, h(c_k)) \neq (h(d_1), \ldots, h(d_k))$.*[3]

Un exemple de sous-formule est celui des *sous-formules pures* : une sous-formule pure de F est une sous-formule ne contenant pas d'occurrences opposées d'un même prédicat (i.e. tout prédicat n'apparaît que sous une forme positive ou négative dans la sous-formule). Une notion plus forte est celle de sous-formule ne possédant pas de

[3]Cette condition n'était pas présente dans (Leclère & Mugnier, 2007) (ce qui ne remet pas en cause les résultats de ce papier), mais elle est nécessaire pour assurer que h constitue un homomorphisme de F dans une complétion de G.

littéraux échangeables, c'est-à-dire de littéraux $p(c_1, \ldots, c_k)$ et $\neg p(d_1, \ldots, d_k)$ tels qu'il existe deux complétions totales de G, G_1 et G_2, et deux homomorphismes h_1 et h_2, respectivement de F dans G_1 et de F dans G_2 avec $(h_1(c_1), \ldots, h_1(c_k)) = (h_2(d_1), \ldots, h_2(d_k))$. On a ainsi :

Théorème 2
(Leclère & Mugnier, 2007) Si $G \models F$ alors pour toute sous-formule F' de F sans littéraux échangeables, il existe un homomorphisme partiel de F dans G par rapport à F'.

L'algorithme finalement proposé dans (Leclère & Mugnier, 2007) prend en compte le filtrage associé à différentes sous-formules, choisit le littéral à ajouter selon un homomorphisme partiel par rapport à une sous-formule spéciale (par exemple, une sous-formule pure maximale pour l'inclusion) et effectue un test d'homomorphisme après chaque étape de complétion.

3 Nouvelle approche

Dans cette partie nous présentons notre nouvelle approche, qui consiste en une réécriture du problème de déduction dans $FOL\{\exists, \wedge, \neg_a\}$ en un problème de validité en logique propositionnelle. Nous allons procéder en plusieurs étapes :

1. Calcul de tous les homomorphismes partiels de F dans G par rapport à une sous-formule spéciale de F.
2. Création d'une formule propositionnelle F_{prop} qui fait la disjonction de toutes les conjonctions de littéraux manquants dans G pour que chaque homomorphisme partiel de l'étape 1 soit total.
3. Test de la validité de F_{prop}, c'est-à-dire de l'insatisfiabilité de sa négation.

Remarquons tout d'abord que l'on peut, et ce sans incidence sur notre problème de déduction, remplacer chaque variable de G par une **nouvelle** constante. Cette modification préserve tous les homomorphismes dans G. Dans la suite, G sera donc supposé contenir uniquement des constantes.

Exemple 2
La formule G de l'exemple 1 devient $G = p(a) \wedge r(a, b) \wedge r(b, c) \wedge r(c, d) \wedge \neg p(d)$.

3.1 Calcul des homomorphismes partiels

Nous pouvons constater qu'une complétion totale G^c code généralement beaucoup plus d'informations que nécessaire pour la recherche d'homomorphismes de F dans G^c. L'idée est d'utiliser la notion d'homomorphisme partiel pour éviter des ajouts superflus en ciblant dès le départ des ensembles de littéraux nécessaires.

Nous observons également qu'un homomorphisme partiel de F dans G par rapport à F' (une sous-formule de F) peut toujours s'étendre à un homomorphisme de F dans une complétion partielle de G. De plus, d'après le théorème 1, si $G \models F$ alors il existe

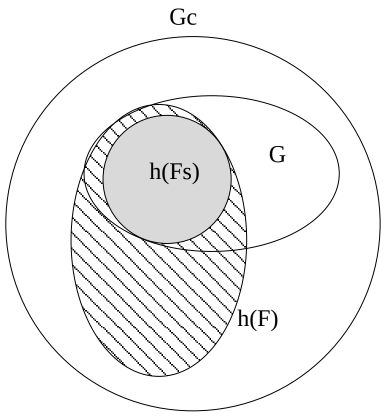

FIG. 2 – Illustration de la propriété 2

un homomorphisme de F dans chaque G^c. Avec le calcul de tous les homomorphismes partiels, on va pouvoir construire un ensemble de complétions partielles $\{G_1, \ldots, G_n\}$ tel qu'il y aura un homomorphisme de F dans G_i pour $i = 1 \ldots n$. Dans le cas où $G \models F$, cet ensemble couvrira l'ensemble des G^c (i.e. pour chaque G^c il existera un G_i tel que $G_i \subseteq G^c$).

Pour obtenir cet ensemble, nous avons besoin d'une sous-formule de F spéciale, que nous appellerons *stable* :

Définition 8 (Sous-formule stable)
La sous-formule stable de F, notée F^s, est la sous-formule composée de tous les littéraux de F dont le prédicat n'est pas un prédicat de complétion de F et G (le vocabulaire de complétion de F^s et G est donc vide). Elle est éventuellement vide.

La propriété ci-dessous est illustrée par la Figure 2.

Propriété 2
Soient deux formules F et G de $FOL\{\exists, \wedge, \neg_a\}$. Soit h un homomorphisme de F dans G^c, une complétion totale de G. h est un homomorphisme partiel de F dans G par rapport à F^s.

Preuve. On a h un homomorphisme de F dans G^c. En restreignant h à F^s, h définit un homomorphisme de F^s dans G. Dès lors que G est consistante, pour chaque littéral l de F qui n'apparaît pas dans F^s, il n'existe pas $h(\bar{l})$ dans G, et G^c est consistante par construction, donc il n'existe pas de paire de littéraux de signe opposé $p(c_1, \ldots, c_k)$ et $\neg p(d_1, \ldots, d_k)$ de $F - F^s$ tels que $(h(c_1), \ldots, h(c_k)) = (h(d_1), \ldots, h(d_k))$. $\square$

Il est à noter que la propriété n'est plus vraie si l'on utilise une sous-formule plus grande que la stable, comme le montre l'exemple suivant :

Exemple 3
Soient F et G les formules de l'exemple 1. $F^s = r(u, v)$. Soient $F^{p_1} = p(u) \wedge r(u, v)$

et $F^{p_2} = r(u,v) \wedge \neg p(v)$ les deux sous-formules pures de F plus grandes que F^s. Soit $G_1^c = p(a) \wedge r(a,b) \wedge p(b) \wedge r(b,c) \wedge \neg p(c) \wedge r(c,d) \wedge \neg p(d)$ une complétion totale de G. Il y a un homomorphisme de F dans G_1^c, $h = \{u \mapsto b, v \mapsto c\}$, mais ce n'est pas un homomorphisme partiel de F dans G que ce soit par rapport à F^{p_1} ou à F^{p_2}.

Ainsi, la sous-formule stable est de taille maximale pour la propriété 2. Cela est dû au fait que les seuls littéraux pouvant servir à la complétion (et ainsi être ajoutés à G) sont ceux possédant un prédicat du vocabulaire de complétion de F et G et la sous-formule stable en est dépourvue.

Exemple 4
Il y a 3 homomorphismes de F dans G par rapport à F^s : $h_1 = \{u \mapsto a, v \mapsto b\}$, $h_2 = \{u \mapsto b, v \mapsto c\}$ et $h_3 = \{u \mapsto c, v \mapsto d\}$

3.2 Création de la formule propositionnelle

Cette étape consiste à se servir des homomorphismes partiels précédemment calculés afin de construire une formule propositionnelle. Comme vu précédemment, un homomorphisme partiel de F dans G peut s'étendre à un homomorphisme complet par l'ajout d'un certain nombre de littéraux à G. Ce sont ces « informations manquantes » que nous codons dans la formule propositionnelle et qui vont nous permettre de répondre au problème de déduction.

Il nous faut donc définir, pour un homomorphisme partiel h de F dans G par rapport à F^s, ce qu'est l'ensemble des littéraux manquants à G afin que h soit égal à un homomorphisme total de F dans une complétion partielle de G. Dans la suite, nous considérons les formules F, G et F^s, et h désigne un homomorphisme partiel de F dans G par rapport à F^s.

Définition 9 (Conjonction minimale)
On appelle conjonction minimale *de G par rapport à h, notée C^m, la conjonction composée de l'atome $\blacksquare$[4] (représentant la tautologie) et des littéraux l tels que $l \in (F - F^s)$ et $h(l) \notin G$.*

Exemple 5
Pour h_1 : $C_1^m = \neg p(b)$; pour h_2 : $C_2^m = p(b) \wedge \neg p(c)$; pour h_3 : $C_3^m = p(c)$.

Définition 10 (Complétion minimale)
On appelle complétion minimale *de G, notée $G^m = G \wedge C^m$, la conjonction composée des littéraux de G et de la conjonction minimale de G par rapport à h.*

Propriété 3
Soit h un homomorphisme partiel de F dans G par rapport à F^s. h est un homomorphisme de F dans la complétion minimale de G issue de h.

[4] nécessaire lorsque h est un homomorphisme total de F dans G car sinon la conjonction est vide.

Propriété 4
Soit C^m une conjonction minimale de G issue d'un homomorphisme partiel de F dans G par rapport à F^s, alors $G \wedge C^m \models F$.

La formule propositionnelle est construite à partir de tous les homomorphismes partiels de F dans G par rapport à F^s :

Définition 11 (Disjonction des Conjonctions Minimales du Stable ($\mathcal{DCMS}$))
On appelle $\mathcal{DCMS}(F, G)$ la disjonction de l'atome $\square$[5] (représentant l'absurde) et des conjonctions minimales de tous les homomorphismes partiels de F dans G par rapport à F^s (i.e. $\square \vee C_1^m \vee \cdots \vee C_n^m$, noté également $\bigvee C^m$).

Exemple 6
Création de la formule propositionnelle : $\mathcal{DCMS}(F, G) = \neg p(b) \vee (p(b) \wedge \neg p(c)) \vee p(c)$.

Nous pouvons maintenant énoncer le théorème validant notre approche (rappelons que G est consistante) :

Théorème 3
$G \models F$ si et seulement si $\mathcal{DCMS}(F, G)$ est valide.

Pour la démonstration, nous avons besoin des définitions et lemmes suivants :

Définition 12 (Conjonction totale)
Soit G^c une complétion totale de G. On appelle conjonction totale, *notée C, la conjonction composée des littéraux $l \in (G^c - G)$.*

Définition 13 (Disjonction des Conjonctions Totales ($\mathcal{DCT}$))
On appelle $\mathcal{DCT}(G)$ la disjonction des conjonctions totales de G (i.e. $C_1 \vee \cdots \vee C_n$).

Lemme 1
$\mathcal{DCT}(G)$ est valide.

Preuve. Par récurrence sur la profondeur de l'arborescence de complétion. $\square$

Lemme 2
Si $G \models F$ alors quelque soit $C \in \mathcal{DCT}(G)$, il existe $C^m \in \mathcal{DCMS}(F, G)$ tel que $C^m \subseteq C$.

Preuve. Soit C une conjonction totale apparaissant dans $\mathcal{DCT}(G)$. Comme $G \models F$, il existe un homomorphisme h de F dans $G \wedge C$ (cf. théorème 1). D'après la propriété 2, h est un homomorphisme partiel de F dans G par rapport à F^s. D'après la propriété 3, h est un homomorphisme de F dans la complétion minimale $G \wedge C^m$ (où C^m est la conjonction minimale de G issue de h). Enfin par définition, C^m ne contient que des littéraux l' tels que $l' = h(l)$ où l est un littéral de F, donc l' est aussi un littéral de C. $\square$

[5] nécessaire lorsqu'il n'y a pas d'homomorphisme partiel de F dans G par rapport à F^s car sinon la disjonction est vide.

Preuve du théorème 3 :

$\Longleftarrow$ Puisque $\mathcal{DCMS}(F,G)$ est valide, $G \equiv G \wedge \bigvee C^m \equiv \bigvee (G \wedge C^m)$. D'après la propriété 4, on a $\bigvee (G \wedge C^m) \models F$. Donc $G \models F$.

$\Longrightarrow$ Soit $G \models F$. Alors d'après le lemme 2 quelque soit $C \in \mathcal{DCT}(G)$, il existe $C^m \in \mathcal{DCMS}(F,G)$ tel que $C = C^m \wedge C'$ où C' est une conjonction de littéraux. *Par l'absurde* : supposons $\mathcal{DCMS}(F,G)$ non valide. Alors il existe une interprétation $\mathcal{I}$ telle que pour tout $C^m \in \mathcal{DCMS}(F,G)$, $v(C^m,\mathcal{I}) = 0$. Donc quelque soit $C \in \mathcal{DCT}(G)$, $v(C,\mathcal{I}) = 0$. Donc $\mathcal{DCT}(G)$ n'est pas valide, ce qui contredit le lemme 1. Donc $\mathcal{DCMS}(F,G)$ est valide. □

3.3 Passage à SAT/UNSAT

Nous avons donc transformé le problème de déduction initial en un problème de test de validité d'une formule propositionnelle, $\mathcal{DCMS}(F,G)$. La négation de cette formule est directement une forme clausale, dont il nous faut tester l'insatisfiabilité. Nous pouvons pour cela utiliser un solveur SAT.

Exemple 7

FC = $\overline{\mathcal{DCMS}(F,G)} = p(b) \wedge (\neg p(b) \vee p(c)) \wedge \neg p(c)$. FC est insatisfiable, donc $\mathcal{DCMS}(F,G)$ est valide, d'où $G \models F$.

3.4 Taille de la formule obtenue

On passe ainsi d'un problème Π_2^P-$complet$ à un problème co-NP-$complet$; toutefois ce deuxième problème prend en entrée une formule dont la taille peut être exponentielle en la taille des formules de départ. Bornons plus précisément la taille de la formule propositionnelle obtenue.

Appelons nb_{hom} le nombre d'homomorphismes partiels de F dans G par rapport à F^s. Ce nombre représente le nombre de conjonctions minimales obtenues dans la formule propositionnelle finale. Il est borné par $n_G^{n_F}$, où n_G et n_F sont respectivement le nombre de termes dans G et F. Néanmoins en pratique, nous nous attendons à ce que plus la taille de F^s soit grande, plus le nombre d'homomorphismes partiels de F dans G par rapport à F^s soit faible, puisque chaque littéral de F^s représente potentiellement une contrainte en plus lors de la recherche d'un homomorphisme partiel de F dans G par rapport à F^s.

La taille d'une conjonction minimale est quant à elle bornée par : $L = |F - F^s|$.

La taille de la formule propositionnelle obtenue est donc bornée par : $nb_{hom} * |F - F^s|$.

D'une part, plus la taille de F^s est grande, plus la taille d'une conjonction minimale est petite, d'autre part on s'attend à ce que nb_{hom} diminue avec la croissance de F^s : ces deux points laissent penser que l'efficacité de cette méthode sera directement corrélée à l'importance de la sous-formule stable dans F, ce qui reste toutefois à vérifier expérimentalement.

Il est également à noter que nb_{hom} peut être borné plus précisément : ce n'est pas le nombre d'homomorphismes partiels de F dans G par rapport à F^s qui importe véritablement, mais le nombre de substitutions différentes sur les termes de L que l'on obtient par ces homomorphismes partiels. En effet, deux homomorphismes partiels de F dans G par rapport à F^s différents h_1 et h_2 mais tels que $h_1(L) = h_2(L)$ amèneront la création de la même conjonction minimale.

4 Conclusion et perspectives

La démarche présentée dans (Leclère & Mugnier, 2007) consistait à construire incrémentalement un ensemble couvrant toutes les formules complètes générables à partir de G, tel que F se déduise de chaque élément de cet ensemble. Le point crucial est le choix, à chaque étape de complétion, du littéral à ajouter pour construire un ensemble couvrant de taille minimale.

L'approche présentée dans ce papier consiste à construire une formule propositionnelle en tirant parti d'un ensemble d'homomorphismes partiels préalablement calculé, puis d'en tester la validité, ramenant ainsi le problème étudié à un problème de validité en logique propositionnelle. Ceci nous permet de nous servir des avancées réalisées ces dernières années sur le problème SAT.

Nous expérimentons actuellement cette approche afin de la comparer aux précédentes (notamment (Leclère & Mugnier, 2007)). Nous pensons pouvoir identifier des critères permettant de choisir une méthode plutôt qu'une autre selon la structure de F (et peut-être de G) comme par exemple la taille de la sous-formule stable. Enfin, nous espérons trouver un certain nombre de propriétés sur la formule obtenue, allant de bornes sur la taille de cette dernière à des propriétés de filtrage et d'amélioration du test de validité.

Références

ABITEBOUL S., HULL R. & VIANU V. (1995). *Foundations of Databases : The Logical Level*. Addison-Wesley.

CHANDRA A. & MERLIN P. (1977). Optimal implementation of conjunctive queries in relational databases. In *9th ACM Symposium on Theory of Computing*, p. 77–90.

GOTTLOB G. (1987). Subsumption and implication. *Inf. Process. Lett.*, **24**(2), 109–111.

HALEVY A. Y. (2001). Answering queries using views : A survey. *VLDB Journal : Very Large Data Bases*, **10**(4), 270–294.

LAVRAC N. & DZEROSKI S. (1994). *Inductive Logic Programming : Techniques and Applications*. Ellis Horwood.

LECLÈRE M. & MUGNIER M.-L. (2006). Simple Conceptual Graphs with Atomic Negation and Difference. In *Proc. of ICCS'06*, volume 4068 of *LNAI*, p. 331–345 : Springer.

LECLÈRE M. & MUGNIER M.-L. (2007). Some Algorithmic Improvments for the Containment Problem of Conjunctive Queries with Negation. In *Proc. of ICDT'07*, volume 4353 of *LNCS*, p. 401–418 : Springer.

MUGGLETON S. (1992). *Inductive Logic Programming*. Academic Press.

ULLMAN J. D. (1997). Information Integration Using Logical Views. In *Proc. of ICDT'97*, volume 1186 of *LNCS*, p. 19–40 : Springer.

WEI F. & LAUSEN G. (2003). Containment of Conjunctive Queries with Safe Negation. In *International Conference on Database Theory (ICDT).*