THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Lossless Selection Views under Constraints

Link:
[Link to publication record in Edinburgh Research Explorer](#)

Document Version:
Peer reviewed version

OPEN ACCESS

Lossless Selection Views under Constraints

Ingo Feinerer[1], Enrico Franconi[2], and Paolo Guagliardo[2]

[1] Vienna University of Technology
`feinerer@logic.at`
[2] Free University of Bozen-Bolzano
`{franconi,guagliardo}@inf.unibz.it`

1 Introduction

The problem of updating a database through a set of views consists in propagating updates of the views to the base relations over which the view relations are defined, so that the changes to the database reflect *exactly* those to the views. This is a classical problem in database research, known as the *view update problem* [1], which in recent years has received renewed attention.

View updates can be consistently propagated in an unambiguous way under the condition that the mapping between database and view relations is *lossless*, and that each database relation can be expressed in terms of the view relations by means of a query, in much the same way the latter are defined from the former [3]. In such a context, database decompositions play an important role, in that their losslessness is associated with the existence of an explicit *reconstruction operator* that prescribes how a database relation can be rebuilt from the pieces into which it has been decomposed. In the case of *horizontal decomposition*, in which a relation is split into sub-relations containing each a subset of its rows, the reconstruction operator is the union.

The study of horizontal decomposition in the literature has mostly focused on settings where data values can only be compared for equality. However, most real-world applications make use of data values coming from domains with a richer structure (e.g., ordering) on which a variety of other restrictions besides equality can be expressed (e.g., that of being within a range or above a threshold). It is therefore of practical interest to consider a scenario where some of the attributes in the database schema are *interpreted* over a specific domain, such as the reals or the integers, on which a set of special predicates (e.g., smaller/greater than) and functions (e.g., addition/subtraction) are defined, according to a first-order language $\mathfrak{C}$. We consider horizontal decomposition in such a setting and investigate its *losslessness* in the presence of integrity constraints. The only restriction we impose on the language $\mathfrak{C}$ is that of being *closed under negation*.

2 Fundamentals

We consider a source relation symbol R under so-called *conditional domain constraints* (CDCs) of the form

$$\forall \overline{x}, \overline{y} . R(\overline{x}, \overline{y}) \wedge \lambda(\overline{x}) \rightarrow \delta(\overline{y}) \;, \tag{1}$$

where $\lambda(\overline{x})$ is a Boolean combination of equalities $x = a$ between a variable from $\overline{x}$ and a constant, and $\delta(\overline{y}) \in \mathfrak{C}$. The variables in $\overline{x}$ and $\overline{y}$ are associated with non-interpreted and interpreted positions (that is, attributes) of R, respectively. By means of formulae in $\mathfrak{C}$, CDCs restrict the admissible values at interpreted positions, whenever a certain condition holds on the non-interpreted ones.

We consider a decomposed schema $\mathbf{V}$ consisting of view symbols having the same arity as R, each of which is defined by a formula of the form

$$\forall \overline{x}, \overline{y} \,.\, V(\overline{x}, \overline{y}) \leftrightarrow R(\overline{x}, \overline{y}) \wedge \big(\lambda(\overline{x}) \,\mathsf{op}\, \sigma(\overline{y})\big) \ , \tag{2}$$

where $\lambda(\overline{x})$ is as in (1), $\sigma(\overline{y}) \in \mathfrak{C}$, and op is $\wedge$. A set Σ of formulae of the form (2), one for each $V \in \mathbf{V}$, specifies a horizontal decomposition of R.

3 Losslessness

Lossless horizontal decomposition under CDCs can be characterised in terms of unsatisfiability in $\mathfrak{C}$, which gives a decision procedure for losslessness whenever the satisfiability of sets of formulae in $\mathfrak{C}$ is decidable. The main idea is as follows:

1) With each equality between a variable and a constant a we associate a propositional variable p_i^a, whose truth-value indicates whether the value at (non-interpreted) position i is a.
2) For a given valuation α of such propositional variables, we consider the set consisting of the $\mathfrak{C}$-formulae in the r.h.s. of all the CDCs that are "applicable" (i.e., the propositional representation of their l.h.s. is true) under α and the *negation* of any $\mathfrak{C}$-formula $\delta(\overline{y})$ appearing in some selection in Σ where the propositional representation of $\lambda(\overline{x})$ is satisfied by α.
3) Checking losslessness is equivalent to checking that there exists no valuation α for which the above set of $\mathfrak{C}$-formulae is satisfiable.

The class of *Unit Two-Variable Per Inequality* formulae (UTVPIs) is a fragment of linear arithmetic over the integers for which the satisfiability problem is decidable in polynomial time [4]. UTVPI have the form $ay + by' \leq d$, where y and y' are integer variables, $a, b \in \{-1, 0, 1\}$ and $d \in \mathbb{Z}$. We refer to Boolean combinations of UTVPIs as BUTVPIs; the satisfiability problem for sets of BUTVPIs is NP-complete [5]. It can be shown that, when $\mathfrak{C}$ is the language of either UTVPIs or BUTVPIs, the problem of deciding losslessness is co-NP-complete.

4 Separability

We also investigated lossless horizontal decomposition under CDCs in combination with traditional integrity constraints, showing that functional dependencies (FDs) do not interact with CDCs and can therefore be allowed without any restriction, whereas this is not the case for unary inclusion dependencies (UINDs).

The main technical tool we employ in our study is the notion of *separability*: a class of constraints is separable from CDCs if, after making explicit the result of their interaction, captured by a suitable finite set $\mathcal{S}$ of sound inference rules, we can focus solely on CDCs and disregard the other constraints, as far as lossless

horizontal decomposition is concerned. Given a horizontal decomposition Σ and a combination Δ of CDCs with other constraints that are separable from them w.r.t. $\mathcal{S}$, we can check whether Σ is lossless under Δ as follows:

1) compute the deductive closure Δ^* of Δ w.r.t. $\mathcal{S}$, which makes explicit the interaction between CDCs and the other constraints in Δ by adding entailed constraints;

2) by using the technique of Section 3, check whether Σ is lossless under the set $\mathsf{cdc}(\Delta^*)$ obtained from Δ^* by discarding all of the constraints that are not CDCs.

A UIND $R[i] \subseteq R[j]$ is satisfied if, for every source instance I, every value in the i-th column of the extension R^I of R under I appears also in the j-th column of R^I. Let y_i and y_j be the variables associated with the interpreted positions i and j of R, respectively; by means of the following *domain propagation rule*

$$\frac{\top \to \delta(y_j) \qquad R[i] \subseteq R[j]}{\top \to \delta(y_i)} \quad , \tag{dp}$$

it is possible to derive a set of CDCs that fully captures the interaction between a given set of UINDs and opportunely restricted CDCs w.r.t. lossless horizontal decomposition, which makes possible to employ the general technique for deciding losslessness also in the presence of UINDs.

5 Discussion

The results on losslessness and separability briefly reported in this short paper are under revision for journal publication; a preliminary version appeared in [2]. We remark that our technique for checking losslessness can also be applied when op in (2) is $\vee$.

We are currently investigating the possibility of generalising the separability results for UINDs to arbitrary inclusion dependencies (INDs). We also strongly believe that equalities between non-interpreted variables (i.e., from $\overline{x}$ in (1) and (2) can be allowed and our technique extended by representing such equalities by means of additional propositional variables and by adding suitable propositional axioms to handle transitivity and symmetry.

References

1. Bancilhon, F., Spyratos, N.: Update semantics of relational views. ACM Trans. Database Syst. 6(4), 557–575 (Dec 1981)
2. Feinerer, I., Franconi, E., Guagliardo, P.: Lossless horizontal decomposition with domain constraints on interpreted attributes. In: Big Data – Proc. of BNCOD 2013, LNCS, vol. 7968, pp. 77–91. Springer (2013)
3. Franconi, E., Guagliardo, P.: Effectively updatable conjunctive views. In: Proc. of AMW 2013. CEUR Workshop Proceedings, vol. 1087. CEUR-WS.org (2013)
4. Schutt, A., Stuckey, P.J.: Incremental satisfiability and implication for UTVPI constraints. INFORMS Journal on Computing 22(4), 514–527 (2010)
5. Seshia, S.A., Subramani, K., Bryant, R.E.: On solving boolean combinations of UTVPI constraints. JSAT 3(1-2), 67–90 (2007)