Loop Restricted Existential Rules and First-Order Rewritability for Query Answering

Vernon Asuncion, Yan Zhang, Heng Zhang, Yun Bai, Weisheng Si

School of Computing Engineering and Mathematics, Western Sydney University, Australia
School of Computer Software, Tianjin University, China
Huazhong University of Science & Technology, China

Introduction
Under the language of TGDs, queries are answered against an ontology represented by a set of TGDs and an input database. In particular, given a database instance $D$, a finite set $\Sigma$ of TGDs, and a query $q$, we want to decide whether $D \cup \Sigma \models q$. However, this problem is undecidable generally, due to the potential cyclic applications of TGDs in $\Sigma$.

In recent years, considerable research has been carried out to identify various expressive decidable classes of TGDs. Among all these decidable classes, some are of special interests for OBDA, i.e., the classes of first-order rewritable TGDs, where conjunctive query answering can be reduced to the evaluation of a first-order query over the database. So far, several useful first-order rewritable classes of TGDs have been discovered: acyclic (AC), domain restricted (DR) (Baget et al. 2011), aGRD (Baget et al. 2011), linear and multi-linear (ML), sticky and sticky-join (SJ), while multi-linear and sticky-join generalise linear TGDs and sticky TGDs, respectively (Cali, Gottlob, and Pieris 2012). Civili and Rosati (Civili and Rosati 2012) further identified another first-order rewritable class called weakly recursive TGDs, and showed that by restricting to simple TGDs, weakly recursive class contains all other first-order rewritable classes.

Unfortunately, there are still real life scenarios that are simple and intuitive but not syntactically recognisable by any of the existing first-order rewritable TGDs classes, which this work on loop restricted TGDs tries to address.

Then main contributions of this paper are summarised here:

1. We define notations of derivation paths and derivation trees for query answering over TGDs (existential rules), and provide a precise characterisation for the traditional TGDs chase procedure through the corresponding derivation tree.

2. Based on the concept of derivation paths, we introduce a new class called loop restricted (LR) TGDs, which are TGDs with certain restrictions on the loops embedded in the underlying rule set.

3. Under our derivation tree framework, we show that the conjunctive query answering (CQA) under LR TGDs satisfies a property called bounded derivation tree depth property (BDTDP). We further prove that BDTDP implies the well-known bounded derivation-depth property (BDDP). This result implies that conjunctive query answering under LR TGDs is not only decidable but also first-order rewritable.

4. We further extend LR TGDs to generalised loop restricted (GLR) TGDs, and prove that the class of GLR TGDs is also first-order rewritable and contains most of other first-order rewritable TGD classes discovered in the literature so far.

Preliminaries
A tuple-generating dependency (TGD) $\sigma$, also called existential rule, over a schema $R$ is a first-order formula of the form

$$\sigma : \forall XY \varphi(X, Y) \rightarrow \exists Z \psi(X, Z),$$

where $X \cup Y \cup Z \subseteq \Gamma \cup \Gamma_V$, $\varphi$ and $\psi$ are conjunctions of atoms over $\mathcal{R}$. When there is no confusion, we usually omit the universal quantifiers from (1). In this case, we also use $\text{head}(\sigma)$ and $\text{body}(\sigma)$ to denote formulas $\exists Z \psi(X, Z)$ and $\varphi(X, Y)$ respectively. In this case, we also use $\text{head}(\sigma)$ and $\text{body}(\sigma)$ to denote formulas $\exists Z \psi(X, Z)$ and $\varphi(X, Y)$ respectively.

Definition 1 (Derivation path). Let $\Sigma$ be a set of TGDs. A derivation path $P$ of $\Sigma$ is a finite sequence of pairs of an atom and a rule: $(\alpha_0, \rho_1), \ldots, (\alpha_n, \rho_n)$, such that

- for each $1 \leq i \leq n$, $\alpha_i = \text{head}(\rho_i)$;
- for each $1 \leq i \leq n$, $\rho_i = \sigma_i \theta_i$ for some $\sigma_i \in \Sigma$ and substitution $\theta_i$;
- for each $1 \leq i < n$, $\alpha_{i+1} \in \text{body}(\rho_i)$;
- for each $1 \leq i \leq n$, if a null $n \in \text{head}(\alpha_i)$ is introduced due to the elimination of existentially quantified variable, then this $n$ must not occur in $\rho_j$, for all $j \in \{i+1, \ldots, n\}$.

A conjunctive query (CQ) $q$ of arity $n$ over a schema $\mathcal{R}$ has the form $p(X) \leftarrow \exists Y \varphi(X, Y)$, where $\varphi(X, Y)$ is a conjunction of atoms with the variables $X$ and $Y$ from $\Gamma_V$ and constants from $\Gamma$, but without nulls, and $p$ is an $n$-ary predicate not occurring in $\mathcal{R}$. We allow $\varphi(X, Y)$ to contain equalities but no inequalities. When $\varphi(X, Y)$ is just a single atom, then we say that the CQ $q$ is atomic. A Boolean Conjunctive Query (BCQ) over $\mathcal{R}$ is a CQ of zero arity. In
this case, we can simply write a BCQ $q$ as $\exists Y \varphi(Y)$. A CQ answering problem, or called CQA problem, defined to be the answer to a CQ $q$ with $n$ arity over an instance $I$, denoted as $q(I)$, is the set of all $n$-tuples $t \in \Gamma^n$ for which there exists a homomorphism $h: X \cup Y \to \Gamma \cup \Gamma_Y$ such that $h(\varphi(X, Y)) \subseteq I$ and $h(X) = t$. The answer to a BCQ is positive over $I$, denoted as $I \models q$, if $\emptyset \in q(I)$.

**Generalised Loop Restricted TGDs**

First we introduce the notion of the bounded derivation depth property.

**Definition 2 (BDDP).** A class $C$ of TGDs satisfies the bounded derivation-depth property ($BDDP$) if for each BCQ $q$ over a schema $R$, for every input database $D$ for $R$ and for every set $\Sigma \subseteq C$ over $R$, $D \cup \Sigma \models q$ implies that there exists some $k \geq 0$ which only depends on $q$ and $\Sigma$ such that $\chi^k (D, \Sigma) \models q$.

It has been shown that the BDDP implies the first-order rewritability (Cali, Gottlob, and Lukasiewicz 2012; Cali, Gottlob, and Pieris 2012). Formally, the BCQA problem is first-order rewritable for a class $C$ of sets of TGDs if for each $\Sigma \subseteq C$, and each BCQ $q$, there exists a first-order query $q_S$ such that $D \cup \Sigma \models q_S$ iff $D \models q_S$; for every input database $D$. In this case, we also simply say that the class $C$ of TGDs is first-order rewritable.

**Definition 3 (BDTP).** A class $C$ of TGDs satisfies the bounded derivation tree depth property ($BDTP$) if for each $\Sigma \subseteq C$, there exists some $k \geq 0$ such that for every BCQ query $\exists Z p(Z)$ and every database $D$, $D \cup \Sigma \models \exists Z p(Z)$ iff $T(D, \Sigma) \models p(n)$ for some instantiated derivation tree $T(D, \Sigma)$ and atom $p(n)$, where depth$(T(D, \Sigma)) \leq k$ and $h(Z) = n$ for some homomorphism $h$.

Basically, Definition 3 says that if a class of TGDs satisfies BDTP, then every BCQ query answering problem can be always decided within a fixed number $k$ of derivation steps with respect to the corresponding instantiated derivation trees.

Now we are ready to formally define the notion of generalised loop restricted patterns. Let $A$ be a set of atoms, we use $\var(n, A)$ (resp., nulls$(A)$) to denote the set of all variables (resp., nulls) occurring in $A$.

**Definition 4 (Generalised loop restricted (GLR) patterns).** Let $\Sigma$ be a set of TGDs. $\Sigma$ is generalised loop restricted (GLR), if each loop pattern $L = (\alpha_1, \rho_1) \cdots (\alpha_n, \rho_n)$ of $\Sigma$ falls into one of the following four types:

**Type I** For each pair $(\alpha_i, \rho_i)$ in $L$ $(1 \leq i < n)$, body$(\rho_i)$ can be separated into two disjoint parts body$(\rho_i) = body_b(\rho_i) \cup body_y(\rho_i)$ such that the following three conditions hold:

1. body$_b(\rho_i) \cap body_y(\rho_i) = \emptyset$,
2. $\alpha_{i+1} \in body_y(\rho_i)$,
3. $\var(\alpha_i) \cup body_y(\rho_i) \cap \var(body_b(\rho_i)) = \cap_{j=1}^n \var(\alpha_j)$.

**Type II** There exists a pair $(\alpha_i, \rho_i)$ in $L$ $(1 \leq i < n)$ such that body$(\rho_i)$ can be separated into two disjoint parts body$(\rho_i) = body_y(\rho_i) \cup body_b(\rho_i)$, where the following three conditions hold:

1. body$_y(\rho_i) \cap body_b(\rho_i) = \emptyset$,
2. $\alpha_{i+1} \in body_b(\rho_i)$,
3. $\var(\alpha_i) \cup body_b(\rho_i) \cap \var(body_y(\rho_i)) = \emptyset$.

**Type III** For each pair $(\alpha_i, \rho_i)$ in $L$ $(1 \leq i < n)$ and each $\beta \in body(\rho_i)$, $\var(\rho_i) \subseteq \var(\beta)$.

**Type IV** For each pair $(\alpha_i, \rho_i)$ in $L$ $(1 \leq i < n)$ and each $\beta \in body(\rho_i) \setminus \{\alpha_{i+1}\}$, $(\var(\alpha_{i+1}) \cap \var(\beta)) \neq \emptyset$ implies

$(\var(\alpha_{i+1}) \cap \var(\beta)) \subseteq \cap_{j=1}^n \var(\alpha_j)$.

**Type V** There exists a pair $(\alpha_i, \rho_i)$ in $L$ $(1 \leq i < n)$, such that body$(\rho_i)$ can be separated into two disjoint parts body$(\rho_i) = body_y(\rho_i) \cup body_b(\rho_i)$, where the following three conditions hold:

1. body$_y(\rho_i) \cap body_y(\rho_i) = \emptyset$,
2. $(\cup_{j=1}^n \alpha_{i,j}) \cap body_y(\rho_i) = \emptyset$,
3. $\null(\rho_i) \neq \emptyset$.

**Main Results**

**Theorem 1.** If a class $C$ of TGDs satisfies BDTP then $C$ also satisfies BDDP. Therefore, since we can also prove that the class GLR of TGDs satisfies BDTP, then it follows that the class GLR is also first-order rewritable.

**Theorem 2.** Consider the BCQA problem for a given set of GLR TGDs. Its data complexity is in $AC^0$, and its combined complexity is $EXPTIME$ complete.

**Theorem 3.** Deciding whether a set of TGDs is generalised loop restricted is $PSPACE$ complete.

GLR actually captures a large class of first-order rewritable TGDs. In fact, we have the following result.

**Proposition 1.** Let GLR be the class of generalised loop restricted TGDs defined in Definition 4. Then we have that:

1. $AC \subseteq GLR$;
2. $ML \subseteq GLR$;
3. $SJ \subseteq GLR$;
4. $aGRD \subseteq GLR$;
5. $DR \subseteq GLR$.

**References**


