Reachability Problems for Controlled Rewriting Systems

Yoshiharu Kojima[1,2]
joint work with Masahiko Sakai[1] and Florent Jacquemard[3]

[1]Nagoya University

[2]Research Fellow of JSPS

[3]INRIA Sacray, LSV

AJSW2010, Obergurgl, 3 August, 2010

Motivation

I have ever studied about reachability and regularity preservation under some strategies (e.g. innermost, context-sensitive, and etc.), but in some problems, these strategies are not sufficient to represent (e.g. XML update).
Hence we should formalize another strategy.

Related work

- F. Jacquremard and M. Rusinowitch, *Rewrite based verification for XML update*, PPDP 2010.

This paper formalize controlled TRS by XPath expression and global constrained TRS to express XML update.

Outline

- Introduction of controlled string rewriting systems (CSRS).
- Proofs of undecidability and decidability of reachability and regular model checking ($R^*(L_{\tt in}) \cap L_{\tt err} \neq \emptyset$ where R is CSRS, $L_{\tt in}$ and $L_{\tt err}$ are regular language) for controlled string rewriting systems (CSRS).
- Extension of CSRS to the following three kinds of controlled term rewriting systems (CTRS) by means of tree automata:
  1. full-control,
  2. monadic-control, and
  3. prefix-control.
- Proofs of undecidability and decidability of reachability and regular model checking for above CTRS.

Controlled string rewriting systems (CSRS)

[P. Butzbach 73] and [L. Chottin 79] formalized. CSRS can control prefix of redexes by regular languages.

Definition (rewrite rule)

Each rewrite rule is of the form:

$$v_1 \to v_2 \text{ if } L\underline{v_1}$$

where $v_1, v_2 \in \Sigma^*$ and L is regular language over Σ for the set of alphabet Σ.

Definition (rewrite relation)

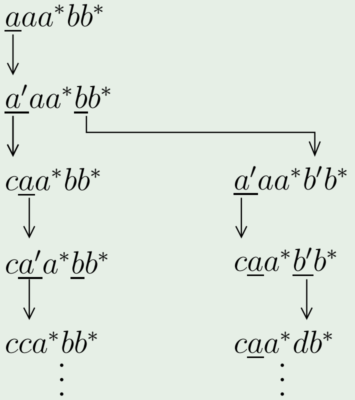
s can be rewritten to t by an above rewrite rule if

$$s = uv_1w, t = uv_2w \text{ and } u \in L.$$

Controlled string rewriting systems

Example

$$R = \left\{ \begin{array}{llll} a \to a' & \text{if} & c^* \underline{a} \\ a' \to c & \text{if} & c^* \underline{a'} \\ b \to b' & \text{if} & c^* a' a^* d^* \underline{b} \\ b' \to d & \text{if} & c^* a^* d^* \underline{b'} \end{array} \right\}$$



$R^*(a^*b^*) \cap \{c,d\}^* = \{c^i d^j \mid j \leq i\}$ (context-free language).

Undecidable properties for CSRS

Theorem

- *Regular model checking is undecidable for length-one CSRS.*
- *Reachability is undecidable for flat CSRS.*

  Flat CSRS

  - *CSRS where each rule is of the form $a \to b, a \to bc$, or $ab \to c$ and latter two rules can be only applied at right-end of strings.*

Proof (Undecidability of reachability).

Simulating emptiness problem of context-sensitive grammar.

Context-sensitive grammar

- Grammar composed by quadruple $\langle N, T, P, S\rangle$ where each rule in P is of the form $\alpha A\beta \to \alpha\gamma\beta$ for $\alpha, \beta \in (N \cup T)^*, A \in N$, and $\gamma \in (N \cup T)^+$.
- Simulated by the rules $S \to SI, AB \to AC$ and $A \to a$ [M. Penttonen 74] where $I, A, B, C \in N$ and $a \in T$.
- Membership problems are PSPACE-complete, and emptiness problems are undecidable.

□

Proof (Undecidable properties of reachability).

Simulate a CSG $G = \langle N, T, P, S \rangle$ where each production rule in P is of the form $S \to SI$, $AB \to AC$, or $A \to a$.
Let CSRS R as the following:

$R_1 = \{S \to SI \text{ if } \underline{S} \mid S \to SI \in P\}$,

$R_2 = \{B \to C \text{ if } (N \cup T)^* A\underline{B} \mid AB \to AC \in P\}$,

$R_3 = \{A \to a \text{ if } (N \cup T)^* \underline{A} \mid A \to a \in P\}$, and

$R_4 = \{ab \to a \text{ if } T^* \underline{ab}, a \to \sharp \text{ if } \underline{a} \mid a, b \in T\}$,

$R = R_1 \cup R_2 \cup R_3 \cup R_4$.

Then, we have

$$S \xrightarrow[R_1]{*} SI \cdots I \xrightarrow[R_2]{*} A_1 \cdots A_n \xrightarrow[R_3]{*} a_1 \cdots a_n \xrightarrow[R_4]{*} \sharp$$

iff $a_1 \cdots a_n$ is produced by CSG G and hence we have $S \xrightarrow[R]{*} \sharp$ iff $\mathcal{L}(G) \neq \emptyset$. □

Decidable property for CSRS

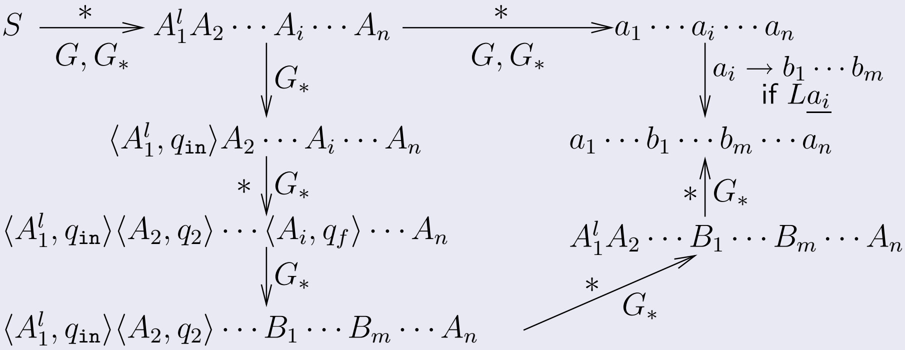
Theorem

Reachability for non-length-decreasing CSRS is PSPACE-complete.

Lemma

For CSRS R over Σ and CSG G s.t. $L(G) = \{a_1 \cdots a_n\}$, we can construct a CSG G_* s.t. $L(G_*) = \{t \mid a_1 \cdots a_n \xrightarrow[R]{} t\}$.

Proof (sketch).



Input: CSRS $R = \{a_i \to b_1 \cdots b_m \text{ if } L\underline{a_i}\}$ and CSG G s.t. $\mathcal{L}(G) = \{a_1 \cdots a_i \cdots a_n\}$.

Output: CSG G_* s.t. $\mathcal{L}(G_*) = \{a_1 \cdots a_i \cdots a_n, a_1 \cdots b_1 \cdots b_m \cdots a_n\}$

□

Proof (sketch).

$$S \xrightarrow[G]{*} A_1^l A_2 \cdots A_i \cdots A_n \xrightarrow[G]{*} a_1 \cdots a_i \cdots a_n$$

$$a_1 \cdots a_i \cdots a_n \xrightarrow{a_i \to b_1 \cdots b_m \text{ if } L\underline{a_i}} a_1 \cdots b_1 \cdots b_m \cdots a_n$$

- $a_1 \cdots a_i \cdots a_n$ is rewritten to $a_1 \cdots b_1 \cdots b_m \cdots a_n$ by the rule $a_i \to b_1 \cdots b_m$ if $L\underline{a_i}$.
- CSG G produce $a_1 \cdots a_i \cdots a_n$ as follows:
  1. G has non-terminal A that corresponds to $a \in \Sigma$ and its terminal is Σ
  2. Firstly, G produces $A_1^l \cdots A_i \cdots A_n$.
  3. Next, G produces $a_1 \cdots a_i \cdots a_n$ by the rules of the form $A_k \to a_k \in P$. Terminals are only produced by such rules.

□

Proof (sketch).

$$S \xrightarrow[G,G_*]{*} A_1^l A_2 \cdots A_i \cdots A_n \xrightarrow[G,G_*]{*} a_1 \cdots a_i \cdots a_n$$

$$a_1 \cdots a_i \cdots a_n \xrightarrow{a_i \to b_1 \cdots b_m \text{ if } L\underline{a_i}} a_1 \cdots b_1 \cdots b_m \cdots a_n$$

$$A_1^l A_2 \cdots A_i \cdots A_n \xrightarrow{G_*} \langle A_1^l, q_{\mathtt{in}} \rangle A_2 \cdots A_i \cdots A_n$$

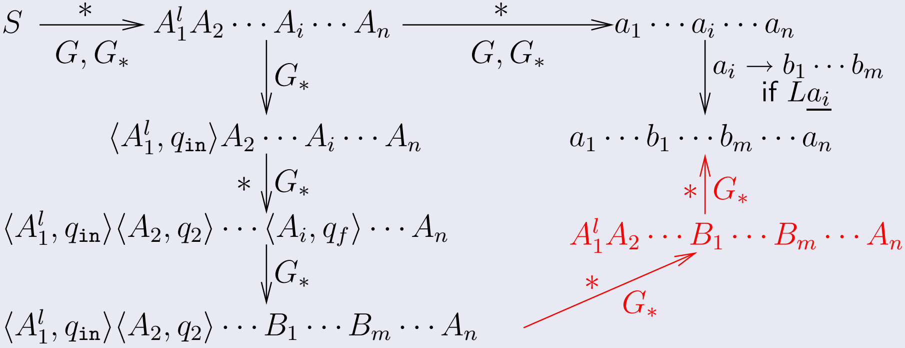
$$\xrightarrow[G_*]{*} \langle A_1^l, q_{\mathtt{in}} \rangle \langle A_2, q_2 \rangle \cdots \langle A_i, q_f \rangle \cdots A_n$$

$$\xrightarrow{G_*} \langle A_1^l, q_{\mathtt{in}} \rangle \langle A_2, q_2 \rangle \cdots B_1 \cdots B_m \cdots A_n$$

G_* simulate DFA $\mathcal{A}$ that corresponds to L in the rewrite rule.

1. Produce non-terminal $\langle A_1^l, q_{\mathtt{in}} \rangle$ from A_1^l (l is marker for left-end).
2. Simulate transition $\delta(a_k, q_k) \to q'_k$ of $\mathcal{A}$ by the production rule $\langle A_k, q_k \rangle A_{k+1} \to \langle A_k, q_k \rangle \langle A_{k+1}, q'_k \rangle$.
3. If q_f is final state of $\mathcal{A}$, then it implies $a_1 \cdots a_{i-1} \in L(\mathcal{A})$ and $B_1 \cdots B_m$ is produced by the rule $\langle A_i, q_f \rangle \to B_1 \cdots B_m$.

Proof (sketch).



Finally, $a_1 \cdots b_1 \cdots b_m \cdots a_n$ is produced by the rules $\langle A_k, q_k \rangle \to A_k$ and $A_k \to a_k$. □

Results for CSRS

Undecidable properties

- Regular model checking for length-one CSRS.
- Reachability for flat CSRS.

Decidable property

- Reachability for non-length decreasing CSRS is PSPACE-complete.

Controlled term rewriting systems (CTRS)

CTRS $(R, \langle \mathcal{A}, S \rangle)$ is a pair of rewrite rules R and a tree automaton $\langle \mathcal{A}, S \rangle$ named selection automaton.

Definition (rewrite rules)

Rewrite rules are of the form:

$$(s_1 \to t_1, \langle \mathcal{A}, S \rangle)$$

where $s, t \in \mathcal{T}(F, X)$, $\mathcal{A}(= \langle Q, Q^f, \Delta \rangle)$ is a usual tree automaton and $S \subseteq Q$. $\langle A, S \rangle$ is named selection automaton.

Definition (rewrite relation)

s can be rewritten to t by the above rule if:

$$s = C[s_1\sigma], t = C[t_1\sigma], \text{ and } \exists q \in S. C[s_1\sigma] \xrightarrow[\mathcal{A}]{*} C[q] \xrightarrow[\mathcal{A}]{*} q^f \in Q^f.$$

Three kinds of CTRS

Definition

For every selection automaton $\langle \mathcal{A}(= \langle Q, Q^f, \Delta\rangle), S\rangle$ in CTRS R, if Δ does not contain ε-transition and there exists a state $q_0 \in Q \backslash S$ s.t. every transition rule $f(q_1, \ldots, q_n) \rightarrow q$ meets the following condition:

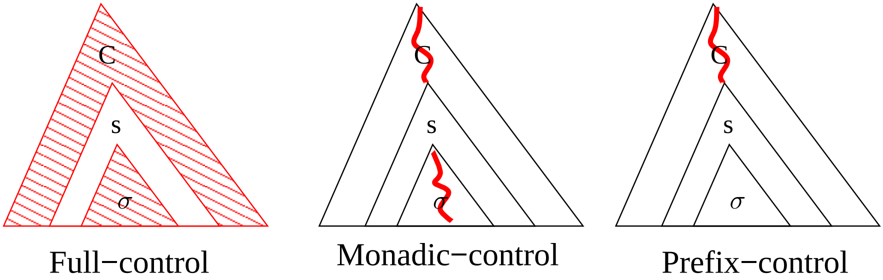
- if $q \neq q_0$, then there exists at most one i s.t. $q_i \neq q_0$,
- otherwise, $q = q_1 = \cdots = q_n = q_0$,

then the CTRS is monadic control.
In addition to condition of monadic control, if every transition rule $f(q_1, \ldots, q_n) \rightarrow q$ meets following condition, then the CTRS is prefix control

- If $q \in S$, then $q_1 = \cdots = q_n = q_0$.

Otherwise, the CTRS is full-control.

Three kinds of CTRS



Red part in the above figures represent controlled part.

- Full-control can control whole context and substitution.
- Monadic-control can control path.
- Prefix-control can control prefix.

Generality

(Context-sensitive $\leq$) Prefix-control $\leq$ Monadic-control $\leq$ Full-control.

Example

Full-CTRS can represent a rule like:

$$a \to c \text{ if "} a \text{ is brother of } b \text{ and descendant of } f\text{"}$$

by the TA $\mathcal{A} = \langle \{q, \tilde{q_a}, q_b, q_{qb}, q_f\}, \{q^f\}, \Delta \rangle$ where Δ is the union of the followings and $S = \{\tilde{q_a}\}$:

$\{a \to \tilde{q_a}, b \to q_b\}$,
$\{g(q_1, \ldots, q_n) \to q \mid g \neq f, \neg\exists i, j.(q_i = \tilde{q}_a \wedge q_j = q_b)\}$,
$\{g(q_1, \ldots, q_n) \to q_{ab} \mid g \neq f, \exists i, j.(q_i = \tilde{q}_a \wedge q_j = q_b)\}$,
$\{g(q_1, \ldots, q_n) \to q_{ab} \mid g \neq f, \forall i.q_i \neq q_f \wedge \exists i.q_i = q_{ab}\}$,
$\{g(q_1, \ldots, q_n) \to q_f \mid g \neq f, \exists i.q_i = q_f\}$,
$\{f(q_1, \ldots, q_n) \to q \mid \neg\exists i.q_i = q_{ab}\}$,
$\{f(q_1, \ldots, q_n) \to q_f \mid \exists i.(q_i = q_{ab} \vee q_i = q_f)\}$.

We have the transition

$C_1[f(C_2[g(a, b, t_1, \ldots, t_n)])] \xrightarrow[\mathcal{A}]{*} C_1[f(C_2[g(\tilde{q_a}, q_b, q, \ldots, q)])] \xrightarrow[\mathcal{A}]{*}$
$C_1[f(q_{ab})] \xrightarrow[\mathcal{A}]{} C_1[q_f] \xrightarrow[\mathcal{A}]{*} q_f \in Q^f$

Undecidability and decidability for CTRS

Undecidability

- Reachability is undecidable for flat prefix-CTRS.
- Regular model checking is undecidable for depth-one prefix-CTRS.
- Reachability is undecidable for ground full-CTRS.

Decidability

- Reachability is decidable for non-size-decreasing full-CTRS.

Theorem (Results of undecidability)

The following problems are undecidable:

1. *Reachability is undecidable for flat prefix-CTRS.*
2. *Regular model checking is undecidable for depth-one prefix-CTRS.*
3. *Reachability is undecidable for ground full-CTRS.*

Proof.

By Simulating the problems for string case.

1,2 By representing every string $a_1 \cdots a_n$ as the tree $a_1(\cdots(a_n))$, depth-one CTRS can simulate length-one CSRS and flat prefix-CTRS can simulate flat CSRS.

3 By introducing binary symbol f and representing every string $a_1 \cdots a_n$ as the tree $f(a_1, f(\cdots f(a_{n-1}, a_n)))$, ground full-CTRS can simulate flat-CSRS.

□

Proof. (ground full-CTRS).

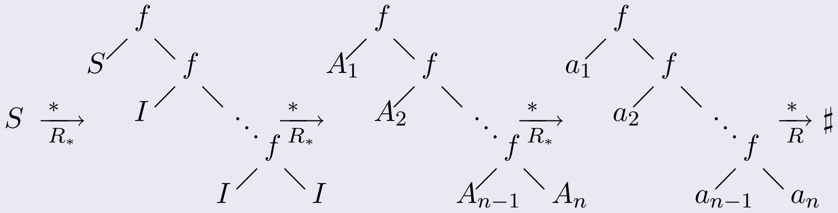
Ground full-CTRS R_* can simulate CSRS R that simulates CSG G where each production rule is of the form: $S \to SI$, $AB \to AC$, $A \to a$

$S \underset{R}{\to} SI$ is simulated by $S \to f(S, I) \in R_*$ and $I \to f(I, I) \in R_*$.

$A \underset{R}{\to} a$ is simulated by $A \to a \in R_*$.

$ab \underset{R}{\to} a$ and $a \underset{R}{\to} \sharp$ is simulated by $f(a, b) \to a \in R_*$ and $a \to \sharp \in R_*$.

$AB \underset{R}{\to} AC$ is simulated by $B \to C$ if "parents' brother is A" $\in R_*$. This reduction is like $\cdots f(A, f(B, \cdots)) \to \cdots f(A, f(C, \cdots))$.



□

Theorem (Results of decidability)

Reachability is decidable for non-size-decreasing full-CTRS.

Proof.

For every term t, the set of terms reachable to t is finite. □

However, complexity of this problem is still open.
In string cases, we have a corresponding grammar (context-sensitive grammar), but we don't have a corresponding grammar for tree (context-sensitive tree grammar?).

Conclusion

- Proofs of decidability and undecidability for some problems for CSRS.
  - Reachability is undecidable for flat CSRS.
  - Regular model checking is undecidable for length-one CSRS .
  - Reachability for non-length-decreasing CSRS is PSPACE-complete.
- Formalizing CTRS by selection automata.
- Proofs of decidability and undecidability for some problems for CTRS.
  - Reachability is undecidable for flat prefix-CTRS.
  - Regular model checking is undecidable for depth-one CTRS .
  - Reachability is decidable for non-size-decreasing CTRS.

Future works

The following problems are still open:

- Reachability for ground monadic- or ground prefix-CTRS.
- Complexity of reachability for non-size-decreasing CTRS.
  - We can prove this by formalizing context-sensitive tree grammar?
- Subclasses of CTRS s.t. regular model-checking is decidable.
  - There exists decidable results of confluence, equivalence problem, and or so for basic strict CSRS [G. Senizergues 90].