Rewrite Closure and CF Hedge Automata

Florent Jacquemard¹ Michael Rusinowitch²

¹Ircam, INRIA Paris-Rocquencourt

²LORIA, INRIA Nancy

April 4, 2013

LATA 2013 – Bilbao

Static Typechecking [Milo Suciu Vianu 03 JCSS]

Consider a tree transformation modeled by a rewrite system \mathcal{R} .

Typechecking:

<u>Iteration</u> of \mathscr{R} always converts valid input data from a tree set L_{in} into valid output data from a tree set L_{out}



Rewrite Closure and CF Hedge Automata 2/ 33

Static Typechecking

[Milo Suciu Vianu 03 JCSS]



- hence TA are also useful for XML reasoning tasks
- like for instance the problem of typechecking.
- It is the verification that a tree transformations *T* defined by a program in a language like XSLT always converts valid input data into valid output data,
- · input and output types are defined by tree automata
- and if the forward closure by T of input data L_{in} is regular, then type checking is reduced to the problem of inclusion for TA (decidable in EXPTIME)
- if this is not the case, like in the cited paper, a second option is to consider the backward closure of the complement of the output data *L*_{out}, and, if it is regular, typechecking is reduced to testing the emptiness of the intersection with the input type *L*_{in}.

Consistency of R/W Access Control Policies

atomic r/w access (updates) modeled by rewrite rules

An ACP is defined by two rewrite systems:

- \mathscr{R}_+ : authorized operations,
- \mathcal{R}_{-} : forbidden operations.

lt is

- ► inconsistent if one rule of *R*₋ can be simulated through a sequence of rules of *R*₊.
- ▶ locally inconsistent for a document *t* if there exists *u* such that $t \xrightarrow{\mathscr{R}_{-}} u$ and $t \xrightarrow{*}{\mathscr{R}_{+}} u$, i.e. $\mathscr{R}_{-}^{1}(t) \cap \mathscr{R}_{+}^{*}(t) \neq \emptyset$.

Rewrite Closure & Tree Automata

if the closure $\mathscr{R}^*(L)$ is effectively regular (for L regular)

- reduce typechecking and local inconsistency to tree automata decision problems
- if not
 - approximate
 - extend the tree automata model

Rewrite Closure and CF Hedge Automata 4/ 33

Rewrite Closure & Tree Automata



- plan: introduce new model automata for unranked trees, given 2 results of automata construction for rewrite closure

- here: we consider unranked - but start with case of ranked trees (better known and simpler)

(parenthesis)

The Case of Ranked Trees

The Case of Ranked Trees

 Σ is a ranked alphabet, every symbol has a fixed arity

A Term Rewriting System \mathscr{R} is a finite set of rewrite rules $\ell \to r$, where ℓ and r are ranked trees (terms) with variables.

Rewriting:

- pattern matching of ℓ and replacement by r
- substitution of variables by ranked trees

Rewrite relation denoted $s \xrightarrow{*} t$, reflexive transitive closure $s \xrightarrow{*} t$.

Rewrite closure of a term set *L*: $\mathscr{R}^*(L) = \{t \mid \exists s \in L, s \xrightarrow{*}{\mathscr{R}} t\}$

Ranked Tree Automata

Ranked Tree Automata

 $\mathscr{A} = \langle \Sigma, Q, F, \Delta \rangle$ with

- ranked alphabet Σ, every symbol has a fixed arity
- finite state set Q,
- final states $F \subseteq Q$,
- set Δ of transitions $a(q_1, \ldots, q_n) \rightarrow q$ with
 - $a \in \Sigma$, *a* of arity *n*
 - $q_1,\ldots,q_n,q\in Q$

Language
$$L(\mathscr{A}, \mathbf{q}) = \{t \mid t \xrightarrow{*}{\Delta} \mathbf{q}\}$$

 $L(\mathscr{A}) = \bigcup_{\mathbf{q} \in F} L(\mathscr{A}, \mathbf{q})$

Rewrite closure of Ranked Tree Automata Languages

Ranked tree automata completion:

- given $\mathscr{A} = \langle \Sigma, Q, F, \Delta_0 \rangle$ and \mathscr{R} over Σ ,
- compute \mathscr{A}^* such that $L(\mathscr{A}^*) = \mathscr{R}^*(L(\mathscr{A}))$.

For linear and right-flat TRS \mathscr{R} : superposition of \mathscr{R} 's rules into \mathscr{A} 's transitions

for
$$a(b(\mathbf{x}_1), c(\mathbf{x}_2)) \to c(\mathbf{x}_2, \mathbf{x}_1) \in \mathscr{R}$$
 $a(b(\mathbf{q}_1), c(\mathbf{q}_2)) \xrightarrow{*}_{\Delta_i} q$
 $* \downarrow \mathscr{R} \xrightarrow{*}_{\Delta_{i+1}} c(\mathbf{q}_2, \mathbf{q}_1)$

Application to static analysis of functional programs: Tree Regular Model Checking [Bouajjani et al 2002]

(end of parenthesis)

Unranked Trees

Unranked Ordered Trees and Rewriting

 $\boldsymbol{\Sigma}$ unranked alphabet

Rewrite System: finite set of rewrite rules $\ell \rightarrow r$, where ℓ and r are hedges with variables.

Rewriting:

- pattern matching and replacement
- replacement of variables by hedges

Hedge Automata and CF Hedge Automata

- $\mathscr{A} = (\Sigma, Q, F, \Delta)$ with
 - alphabet Σ,
 - finite state set Q,
 - final states $F \subseteq Q$,
 - set Δ of transitions $a(L) \rightarrow p$ with

Language
$$L(\mathscr{A}, \mathbf{q}) = \{t \mid t \xrightarrow{*}{\Delta} q\}$$

 $L(\mathscr{A}) = \bigcup_{\mathbf{q} \in F} L(\mathscr{A}, \mathbf{q})$

Regular set of unranked tree = HA language.

Rewriting Example 1: node renaming

 $c(x) \rightarrow d(x)$



Closure with rules of this form effectively preserve regularity.

for $c(L) \rightarrow q \in \Delta$, add $d(L) \rightarrow q$ to Δ .

Rewrite Closure and CF Hedge Automata 12/33

Rewriting Example 1: node renaming



- the rule can be applied to any node labeled by c
- the variable x represents a finite sequence of trees (hedge)

Rewriting Example 2: insert

 $n(x) \rightarrow n(x).ph(1)$



Closure with rules of this form effectively preserves regularity.

by transformation of languages in transition rules (add a loop with q_{ph} recognizing {ph(1)}) [RJ PPDP 2010]



Rewrite Closure and CF Hedge Automata 13/33

Rewriting Example 2: insert



 the right hand side of this rule is an hedge of length 2 (not a tree) **Rewriting Example 3: collapse**

$$\mathscr{R}^*(L_{in}) \cap c\bigl(\{a,b\}^*\bigr) = \{c(a^n.b^n) \mid n \ge 0\}$$

The rewrite closure is a CF-HA language.

also for inserts $\mathscr{R} = \{c(x) \rightarrow c(a.x.b)\}$ and $\mathscr{R} = \{c_0(x) \rightarrow c_1(a.x), c_1(x) \rightarrow c_0(x.b)\}$ **Rewriting Example 4: T-patterns**

$$\mathscr{R} = \left\{ \begin{array}{ll} p_0(\mathbf{x}) \to a.p_1(\mathbf{x}), & p_1(\mathbf{x}) \to p_2(\mathbf{x}).c, \\ p_2(\mathbf{x}) \to p_0(b(\mathbf{x})), & p_2(\mathbf{x}) \to b(\mathbf{x}) \end{array} \right\}$$
$$p_0 \quad \stackrel{*}{\xrightarrow{\mathscr{R}}} a.p_1 \stackrel{*}{\xrightarrow{\mathscr{R}}} a.p_2.c \stackrel{*}{\xrightarrow{\mathscr{R}}} a.p_0(b).c \stackrel{*}{\xrightarrow{\mathscr{R}}} a.a.p_1(b).c \\ \stackrel{*}{\xrightarrow{\mathscr{R}}} a.a.p_2(b).c.c \stackrel{*}{\xrightarrow{\mathscr{R}}} a.a.p_0(b(b)).c.c \stackrel{*}{\xrightarrow{\mathscr{R}}} \dots$$

 $\mathscr{R}^*(\{p_0\})$ without p_0, p_1, p_2 is the set of trees



with the same number of a, b and c.

The rewrite closure is not a CF-HA language.

$\mathsf{CF}^2\mathsf{HA}$

- $\mathscr{A} = (\Sigma, Q, F, \Delta)$ with
 - alphabet Σ,
 - finite state set Q,
 - final states $F \subseteq Q$,
 - set Δ of transitions of the form,

$$p_1(x_1)\dots p_n(x_n) \rightarrow q(x_1\dots x_n)$$

 $p_1(p_2(x)) \rightarrow q(x)$

horizontal transitions, vertical transitions.

where $p_1, \ldots, p_n \in Q \cup \Sigma$, $q \in Q$ and $n \ge 0$

equivalent form:

1

$$\begin{array}{rccc} p_1(\delta_1) \dots p_n(\delta_n) & \to & \boldsymbol{q}(\delta_1 \dots \delta_n) \\ p_1(p_2(\delta_1)) & \to & \boldsymbol{q}(\delta_1) \end{array} & n > 0 \end{array}$$

where every δ_i is either a variable x_i or ε .

CF²HA: T-patterns



 $\langle \{a,b,c\}, \{ q_0,q_1,q_2\}, \{ q_0\}, \Delta \rangle$ with

$$\Delta = \left\{ \begin{array}{cc} b(x_1) \to q_2(x_1), & a.q_1(x_2) \to q_0(x_2), \\ q_2(x_1).c \to q_1(x_1), & q_0(b(x)) \to q_2(x) \end{array} \right\}$$

$CF^{2}HA \supset HA \text{ and } CF-HA$

CF-HA with variable-free transitions

$$\begin{array}{cccc} q_1 \dots q_n & o & q \\ a(q_1) & o & q_2 \end{array}$$

where $a \in \Sigma$ and q_1, \ldots, q_n, q are states.

HA with $Q = Q_h \uplus Q_v$, transitions:

$$egin{array}{ccc} m{arepsilon} & m{arepsilon} & q_{\mathsf{h}} \ q_{\mathsf{h}}.q_{\mathsf{v}} &
ightarrow & q'_{\mathsf{h}} \ a(q_{\mathsf{h}}) &
ightarrow & q_{\mathsf{v}} \end{array}$$

where $a \in \Sigma$, $q_h, q'_h \in Q_h$, $q_v \in Q_v$.

Membership is decidable for CF^2HA .

PSPACE procedure

Emptiness is decidable in PTIME for CF^2HA .

State marking with 2 marks.

Rewrite Closure under Inverse-Monadic and 1-childvar Rewrite Systems

Inverse-Monadic and 1-childvar Rewrite Systems

Rewrite closure of CF²HA is CF²HA for rules of the form $a(x) \rightarrow r$ where x

- is the only variable in r,
- ▶ has at most 1 occurrence in *r*,
- has no siblings in r.

example: T-patterns

$$\mathscr{R} = \left\{ \begin{array}{rrr} p_0(\mathbf{x}) & \to & a.p_1(\mathbf{x}), & & p_1(\mathbf{x}) & \to & p_2(\mathbf{x}).c, \\ p_2(\mathbf{x}) & \to & p_0(\mathbf{b}(\mathbf{x})), & & p_2(\mathbf{x}) & \to & b(\mathbf{x}) \end{array} \right\}$$

Inverse-Monadic and 1-childvar Rewrite Systems

Rewrite closure of CF²HA is CF²HA for rules of the form $a(x) \rightarrow r$ where *x* only var., has 1 occ. and no siblings in *r*.

Given a CF²HA $\mathscr{A}_{in} = \langle \Sigma, Q_{in}, F_{in}, \Delta_{in} \rangle$, we construct a CF²HA \mathscr{A} with state set

 $Q = Q_{\text{in}} \uplus \{ \frac{h}{a} \mid h \text{ non-var subhedge of a rhs of } \mathscr{R} \} \uplus \{ \frac{a}{a} \mid a \in \Sigma \} \uplus \{ \frac{q}{a} \}$

and transitions

1-childvar Condition

$$\mathscr{R} = \{a(x) \to c.a(e.x.g).d\}$$

$$\mathscr{R}^*ig(\{a\}ig)=\{c^n.a(e^n.g^n).d^n\mid n\geq 1\}$$
 seemingly not $\mathsf{CF}^2\mathsf{HA}.$

Rewrite Closure under Update Rewrite Systems

Update Rewrite Rules

We assume a fixed HA $\mathscr{A} = \langle \Sigma, Q, F, \Delta \rangle$.

a (x)	\rightarrow	b(x)		node renaming	(ren)
a (x)	\rightarrow	$a(u_1.x.u_2)$	$u_1, u_2 \in Q^*$	insertion of child nodes	(ins.c)
a (x)	\rightarrow	$v_1.a(x).v_2$	$v_1, v_2 \in Q^*$	insertion of sibling nodes	(ins.s)
a (x)	\rightarrow	b(a(x))		insertion of a parent node	(ins.p)
a (x)	\rightarrow	и	$u \in Q^*$	node replacement recursive deletion	(\mathbf{rpl})
a (x)	\rightarrow	x		node deletion	(del)

Rewrite closure of CF-HA is CF-HA for every update rewrite system.

Loop-Free Update Rewrite Systems

 \mathscr{R} loop-free if there exists no sequence a_1, \ldots, a_n (n > 1) such that for all $1 \le i < n$, $a_i(x) \to a_{i+1}(x) \in \mathscr{R}$ and $a_1 = a_n$.

Transformation of an update rewrite system \Re into a loop-free update rewrite system $\hat{\Re}$ by selecting a representative \hat{a} of a in every a's loop and suppressing loops.

In the construction of an automaton for \mathscr{R}^* , it is sufficient to consider the rewrite closure by $\hat{\mathscr{R}}$.

Closure under Update Rewrite Systems

Rewrite closure of CF-HA is CF-HA for every loop-free update rewrite system.

Initialize, given CF-HA $\mathscr{A}_{in} = \langle \Sigma, Q_{in}, F_{in}, \Delta_{in} \rangle$

$$\begin{array}{rcl} \Delta_0 = \Delta_{\mathsf{in}} & \cup & \{ q_{a_1} \to q \} \\ & \cup & \{ a_n \left(q^{a_1 \dots a_n} \right) \to q_{a_1 \dots a_n} \mid a_1, \dots, a_n \text{ is a renaming chain} \} \end{array}$$

Completion

	\mathscr{R} contains	$\Delta_{i+1} {=} \Delta_i \cup$
(ren)	$a_n(\mathbf{x}) \to b(\mathbf{x})$	$\{q^{a_1\dots a_n} \to q^{a_1\dots a_n b} \mid q^{a_1\dots a_n b} \in Q\}$ $\cup \{q_{a_1\dots a_n b} \to q_{a_1\dots a_n} \mid q_{a_1\dots a_n b} \in Q\}$
(ins.c)	$a_n(\mathbf{x}) \rightarrow a_n(\mathbf{u}.\mathbf{x}.\mathbf{v})$	$\{u.q^{a_1a_n}, v \to q^{a_1a_n} \mid q^{a_1a_n} \in Q\}$
(ins.s)	$a_n(x) \rightarrow \mathbf{u}.a_n(x).\mathbf{v}$	$\{u.q_{a_1a_n}.v \to q_{a_1a_n} \mid q_{a_1a_n} \in Q\}$
(ins.p)	$a_n(x) \to b(a_n(x))$	$\{b(q_{a_1\dots a_n}) ightarrow q_{a_1\dots a_n} \mid q_{a_1\dots a_n} \in Q\}$
(rpl)	$a_n(x) \rightarrow u$	$\{u ightarrow q_{a_1a_n} \mid q_{a_1a_n} \in Q\}$
(del)	$a_n(x) \rightarrow x$	$\{q^{a_1a_n} ightarrow q_{a_1a_n}\mid q^{a_1a_n}\in Q\}$

Rewrite Closure and CF Hedge Automata 27/33

Closure under Update Rewrite Systems

Clo	sure u	nder Update Re	ewrite Systems		
F	$\label{eq:action} \begin{array}{l} \mbox{Rewith} covery loop fields rewith system. \\ \mbox{Institutes, given CF-HA act } \\ \mbox{Act} = \Delta_{0c} \cup \{ q_{cd} \in \sigma^{-1} \\ \cup \{ a_{cd} \in \sigma^{-1} \\ o \in a_{cd} = \sigma^{-1} \\ o \in a_{cd} = \sigma^{-1} \\ \mbox{Completes} \end{array} \right) \rightarrow q_{cs}, \\ \mbox{Act} = \Delta_{0c} \cup \{ a_{cd} \in \sigma^{-1} \\ o \in a_{cd} \\ \mbox{Act} \\ \m$				
1					
c					
ſ		<i>ℜ</i> contains	$\Delta_{i+1} = \Delta_i \cup$		
ĺ	(ren)	$a_n(x) \to b(x)$	$\{q^{a_1a_n} \rightarrow q^{a_1a_nb} q^{a_1a_nb} \in Q\}$ $\cup \{q_{a_1a_nb} \rightarrow q_{a_1a_n} q_{a_1a_nb} \in Q\}$		
	(ins.c)	$a_n(\mathbf{x}) \rightarrow a_n(\mathbf{u}.\mathbf{x}.\mathbf{v})$	$\{u.q^{a_1a_n}, v \to q^{a_1a_n} \mid q^{a_1a_n} \in Q\}$		
	(ins.s)	$a_n(\mathbf{x}) \rightarrow \mathbf{u}.a_n(\mathbf{x}).\mathbf{v}$	$\{u.q_{c_1\ldots d_n}.v\rightarrow q_{c_1\ldots d_n}\mid q_{c_1\ldots d_n}\in Q\}$		
	(ins.p)	$a_n(\mathbf{x}) \rightarrow b(a_n(\mathbf{x}))$	$\{b(q_{a_1\dots a_n}) \rightarrow q_{a_1\dots a_n} \mid q_{a_1\dots a_n} \in Q\}$		
	(rpl)	$a_n(x) \rightarrow u$	$\{u \rightarrow q_{a_1\dots a_n} \mid q_{a_1\dots a_n} \in Q\}$		
	(del)	$a_n(\mathbf{x}) \rightarrow \mathbf{x}$	$\{q^{a_1\cdots a_n}\rightarrow q_{a_1\cdots a_n}\mid q^{a_1\cdots a_n}\in Q\}$		

principle of proof:

- construction of a new transition set for the rewrite closure, by iterative completion of $\Delta_{\rm in}$ according to \mathscr{R}
- insertion rules depend on the label of the current node,
- \Rightarrow the automaton must check the label in order to be sure that an insert was possible
- there can be interleavings of (ren) and (ins) of \neq kinds (e.g. child and sibling), \neq (ins) checked at \neq levels of bottom-up computation \Rightarrow store the renaming chains in states.
- ${\mathscr R}$ loop-free \Rightarrow length of every renaming chain is bounded by $|{\mathscr R}|.$
- 2 modes for states:
- *push* mode for $q^{a_1...a_n}$: for ins sibling (performed first)
- *pop* mode for $q_{a_1...a_n}$: for ins child (performed second)

Synchronized rename and insert

$$\mathscr{R} = \{a(x) \to c.a(e.x.g).d\}$$

$$\mathscr{R}^*(\lbrace a\rbrace) = \lbrace c^n.a(e^n.g^n).d^n \mid n \ge 1\rbrace$$

$$\mathscr{R}' = \left\{ \begin{array}{rrr} a(x) & \to & c.a'(x).d, & \text{inv-monadic, 1-childvar} \\ & & \in (\text{ins.s}) + (\text{ren}) \\ a'(x) & \to & a(e.x.g) & \notin \text{1-childvar} \\ & & \in (\text{ins.c}) + (\text{ren}) \end{array} \right\}$$

Conclusion

- decidable model CF²HA of unranked tree recognizer extending hedge automata
- captures the rewrite closure under
 - linear, inverse-monadic, 1-childvar rewrite systems
 - update (parametric) rewrite systems

Perspectives

- intersection with regular tree languages
- counting constraints on horizontal and vertical paths
- closure under (maximal) parallel rewriting
 [Solimando, Delzanno, Guerrini Games 2012]

Т

Rule based ACPs

[Fundulaki Maneth][Bravo et al, ACCOn]

An access control policy (ACP) given by two finite sets of rules

- ▶ 𝔐₊: authorized operations
- \mathcal{R}_{-} : forbidden operations.

example

$$\blacktriangleright \ \mathscr{R}_{+} = \left\{ \begin{array}{cc} \operatorname{addressbook}(x) & \to & \operatorname{addressbook}(p_{\operatorname{ec}} x), \\ & & \operatorname{card}(x) & \to & () \end{array} \right\}$$

user can insert card with name, delete card.

•
$$\mathscr{R}_{-} = \{\operatorname{name}(x) \to p_n\}$$

user cannot change a name.

Closure under Update Rewrite Systems

Rewrite closure of CF-HA is CF-HA for every loop-free update rewrite system.

normalized CF-HA $\langle \Sigma, Q, Q^{f}, \Delta \rangle$ for all $a \in \Sigma$ and $q \in Q$, there exists one unique $q^{a} \in Q$ such that $a(q^{a}) \rightarrow q \in \Delta$, and moreover, q^{a} does neither occur in a left hand side of an horizontal transition of Δ nor in a right hand side of a vertical transition of Δ .

$$Q = P \cup \{ q_a \mid q^a \in P_{\text{in}} \} \cup \left\{ \begin{array}{c} q^{a_1 \dots a_n} \\ q_{a_1 \dots a_n} \end{array} \middle| \begin{array}{c} q \in P \setminus P_{\text{in}}, n \ge 2, \\ a_1, \dots, a_n \end{array} \right\}$$

Closure under Update Rewrite Systems

$$egin{array}{lll} \Delta_0 = \Gamma_h & \cup & \{ q_{a_1} o q \mid q_{a_1} \in Q \} \ & \cup & \{ a_n ig(q^{a_1 \dots a_n} ig) o q_{a_1 \dots a_n} \mid q^{a_1 \dots a_n}, q_{a_1 \dots a_n} \in Q, n \geq 1 \} \end{array}$$

where Γ_h is the subset of horizontal transitions of Γ .

	\mathscr{R} contains	$\Delta_{i+1} {=} \Delta_i \cup$
(ren)	$a_n(x) \rightarrow b(x)$	$\{q^{a_1a_n} o q^{a_1a_n b} \mid q^{a_1a_n b} \in Q\} \ \cup \ \{q_{a_1a_n b} o q_{a_1a_n} \mid q_{a_1a_n b} \in Q\}$
(ins.c)	$a_n(x) \rightarrow a_n(uxv)$	$\{uq^{a_1\dots a_n}v \to q^{a_1\dots a_n} \mid q^{a_1\dots a_n} \in Q\}$
(ins.s)	$a_n(x) \to u a_n(x) v$	$\{uq_{a_1\dots a_n}\nu \to q_{a_1\dots a_n} \mid q_{a_1\dots a_n} \in Q\}$
(ins.p)	$a_n(x) \to b(a_n(x))$	$\{b(q_{a_1\dots a_n}) ightarrow q_{a_1\dots a_n}\mid q_{a_1\dots a_n}\in Q\}$
(rpl)	$a_n(x) \rightarrow u$	$\{u ightarrow q_{a_1a_n} \mid q_{a_1a_n} \in Q\}$
(del)	$a_n(x) \rightarrow x$	$\{q^{a_1a_n} ightarrow q_{a_1a_n} \mid q^{a_1a_n} \in Q\}$