Decidability for Clark-congruential CFGs

Tobias Kappé Makoto Kanazawa

NII Logic Seminar, January 10, 2018

Introduction

Context Free Grammars are surrounded by undecidable questions:

- Universality
- Equivalence
- Congruence

These are all decidable for regular languages.

Introduction

Idea: restrict CFGs, such that:

- Regular languages are contained (and then some)
- Some questions become decidable

Preliminaries

Let us fix a (finite) *alphabet* Σ.

Σ^* denotes the *set of words* over Σ.

The *empty word* is denoted by ϵ.

Σ^+ denotes the *non-empty words* over Σ.

For $w, x \in \Sigma^*$, wx denotes the *concatenation* of w and x.

Preliminaries

A *congruence* on Σ^* is an equivalence $\equiv$ on Σ^* such that

$$\frac{w \equiv w' \qquad x \equiv x'}{wx \equiv w'x'}$$

$\equiv$ is *finitely generated* if it is the smallest congruence contained in a finite relation.

We write $[w]_\equiv$ for the *congruence class* of $w \in \Sigma^*$ modulo $\equiv$, i.e.,

$$[w]_\equiv = \{x \in \Sigma^* : w \equiv x\}$$

Preliminaries

Every language L induces a *syntactic congruence* $\equiv_L$:

$$\frac{\forall u, v \in \Sigma^*.\ uwv \in L \iff uxv \in L}{w \equiv_L x}$$

Preliminaries

A *reduction* on Σ^* is a reflexive, transitive and Noetherian relation $\rightsquigarrow$ on Σ^* such that

$$\frac{w \rightsquigarrow w' \qquad x \rightsquigarrow x'}{wx \rightsquigarrow w'x'}$$

Preliminaries

A *Context-Free Grammar* (*CFG*) is a tuple $G = \langle V, \rightarrow, I \rangle$, s.t.

- V is a finite set of *non-terminals*
- $\rightarrow \subseteq V \times (V \cup \Sigma)^*$ is a finite *production relation*
- $I \subseteq V$ is a finite set of *initial non-terminals*

Elements of $\rightarrow$ are known as *productions*. We write $\hat{\Sigma}$ for $V \cup \Sigma$.

We fix $G = \langle V, \rightarrow, I \rangle$ throughout this talk.

Preliminaries

$\Rightarrow_G$ is the smallest relation on $\hat{\Sigma}^*$ such that

$$\frac{\alpha B \gamma \in \hat{\Sigma}^* \qquad B \to \beta}{\alpha B \gamma \Rightarrow_G \alpha \beta \gamma}$$

We write $\Leftrightarrow_G$ for the symmetric closure of $\Rightarrow_G$.

Preliminaries

For $A \in V$, we define:

$$\ell(G, A) = \{\alpha \in \hat{\Sigma}^* : A \Rightarrow_G^* \alpha\} \qquad \ell(G) = \bigcup_{A \in I} \ell(G, A)$$

$$L(G, A) = \{w \in \Sigma^* : A \Rightarrow_G^* w\} \qquad L(G) = \bigcup_{A \in I} L(G, A)$$

Convention

If $A \in V$, then $L(G, A) \neq \emptyset$.

Preliminaries

Congruence problem

Given a grammar G, and $w, x \in \Sigma^*$, does $w \equiv_{L(G)} x$ hold?

Equivalence problem

Given grammars G_1 and G_2, does $L(G_1) = L(G_2)$ hold?

Equivalence[1] and congruence are undecidable for general CFGs.

Recognition problem

Given a class of grammars $\mathcal{G}$ and a grammar G, does $G \in \mathcal{G}$ hold?

[1] Bar-Hillel, Perles, and Shamir 1961.

Classes of grammars

G is NTS[2] when for $A \in V$ and $\alpha \in \hat{\Sigma}^*$, we have $A \Rightarrow_G^* \alpha$ iff $A \Leftrightarrow_G^* \alpha$.

Example

Consider the grammars

$$G_1 = \langle \{S\}, \{S \rightarrow SS + a + b\}, \{S\} \rangle$$
$$G_2 = \langle \{S\}, \{S \rightarrow aS + bS + a + b\}, \{S\} \rangle$$

Here $\ell(G_1, S) = \{a, b, S\}^+ = \bar{\ell}(G_1, S)$, and thus G_1 is NTS.

Contrarily, $S \Leftrightarrow_{G_2}^* SS$ while $S \not\Rightarrow_{G_2}^* SS$, and thus G_2 is not NTS.

[2] Boasson 1980.

Classes of grammars

G is pre-NTS[3] when for $A \in V$ and $w \in \Sigma^*$, we have $A \Rightarrow_G^* w$ iff $A \Leftrightarrow_G^* w$.

Example

Consider the grammars

$$G_2 = \langle \{S\}, \{S \to aS + bS + a + b\}, \{S\} \rangle$$
$$G_3 = \langle \{S, T\}, \{S \to SS + a + b,\ T \to b\}, \{S, T\} \rangle$$

Here $L(G_2, S) = \{a, b\}^+ = \bar{L}(G_2, S)$, and thus G_2 is pre-NTS.

Contrarily, $T \Leftrightarrow_{G_3}^* a$ while $T \not\Rightarrow_{G_3}^* a$, and thus G_3 is not pre-NTS.

[3] Autebert and Boasson 1992.

Classes of grammars

G is *Clark-congruential*[4] when for $A \in V$ and $w, x \in L(G, A)$ it holds that $w \equiv_{L(G)} x$.

Example

Consider the grammars

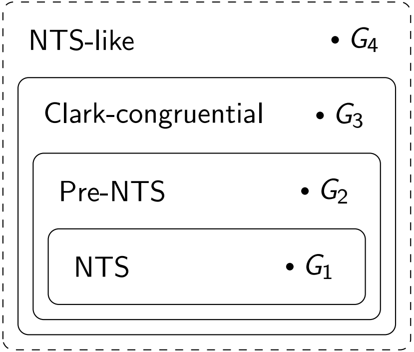
$$G_3 = \langle \{S, T\}, \{S \to SS + a + b,\ T \to b\}, \{S, T\} \rangle$$
$$G_4 = \langle \{S, T\}, \{S \to SS + a + b + aT,\ T \to c + cc\}, \{S\} \rangle$$

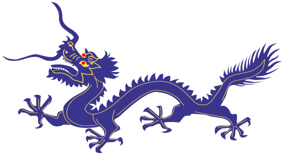
Here $L(G_3, S), L(G_4, T) \subseteq [a]_{\equiv_{L(G_3)}}$, and thus G_3 is Clark-congruential.

Contrarily, $a, \epsilon \in L(G_4, T)$ while $c \not\equiv_{L(G_4)} cc$, and thus G_4 is not Clark-congruential.

[4] Clark 2010.

Classes of grammars





Classes of grammars

	Congruence	Equivalence	Recognition
NTS	✓[5]	✓[5]	✓[5,6]
Pre-NTS	✓[7]	✓[7]	✗[8]
Clark-congruential	✓	✓	†

[5] Sénizergues 1985.
[6] Engelfriet 1994.
[7] Autebert and Boasson 1992.
[8] Zhang 1992.

Deciding congruence

We assume a total order $\preceq$ on Σ.

This order extends to a total order on Σ^*:

- If w is shorter than x, then $w \preceq x$.
- If w and x are of equal length, compare lexicographically.

For $\alpha \in \hat{\Sigma}^*$ with $L(G, \alpha) \neq \emptyset$, write $\vartheta_G(\alpha)$ for the $\preceq$-minimal element of $L(G, \alpha)$.

Deciding congruence

We mimic an earlier method to decide congruence.[9]

Let $\rightsquigarrow_G$ be the smallest reduction such that

$$\frac{A \to \alpha \qquad L(G, \alpha) \neq \emptyset}{\vartheta_G(\alpha) \rightsquigarrow_G \vartheta_G(A)}$$

Lemma

If $w \rightsquigarrow_G x$, then $w \equiv_{L(G)} x$.

[9] Autebert and Boasson 1992.

Deciding congruence

Lemma

$w \in L(G)$ if and only if $w \rightsquigarrow_G \vartheta_G(A)$ for some $A \in I$.

Proof.

($\Rightarrow$) If $w \in L(G)$, then $w \in L(G, A)$ for some $A \in I$. Work "backwards" through the derivation $A \Rightarrow_G^* w$ to go from w to $\vartheta_G(A)$.

($\Leftarrow$) If $w \rightsquigarrow_G \vartheta_G(A)$, then $w \equiv_{L(G)} \vartheta_G(A)$, and thus $w \in L(G)$. □

Deciding congruence

Example

Let $G = \langle \{S\}, \{S \to SS + qSp + \epsilon\}, \{S\} \rangle$.

Then $\rightsquigarrow_G$ is generated by $qp \rightsquigarrow_G \epsilon$, and thus

$$qqp\underline{qp}pqp \rightsquigarrow_G q\underline{qp}pqp \rightsquigarrow_G qp\underline{qp} \rightsquigarrow_G \underline{qp} \rightsquigarrow_G \epsilon = \vartheta_G(S)$$

and therefore $qqpqppqp \in L(G)$.

From $pqpq$, we can only "reach" pq, which is irreducible; thus, $pqpq \notin L(G)$.

Deciding congruence

Given G, we write $\mathcal{I}_G$ for the set of words *irreducible* by $\rightsquigarrow_G$.

Let us fix $w, x \in \Sigma^*$.

Lemma

We can create a DPDA M_w such that $L(M_w) = \{u\sharp v : uwv \in L(G),\ u, v \in \mathcal{I}_G\}$.

"I have a truly marvelous proof which this margin is too narrow to contain..."

Deciding congruence

Recall: $L(M_w) = \{u\sharp v : uwv \in L(G),\ u, v \in \mathcal{I}_G\}$.

Lemma

$L(M_w) = L(M_x)$ if and only if $w \equiv_L x$.

Proof.

($\Rightarrow$) If $uwv \in L(G)$, let $u', v' \in \mathcal{I}_G$ be such that $u \rightsquigarrow_G u'$ and $v \rightsquigarrow_G v'$. Then $u'\sharp v' \in L(M_w) = L(M_x)$. But then $u'xv' \in L(G)$; since $u'xv' \equiv_{L(G)} uxv$, also $uxv \in L(G)$. Analogously, $uxv \in L(G)$ implies $uwv \in L(G)$.

($\Leftarrow$) If $y \in L(M_w)$, then $y = u\sharp v$ such that $uwv \in L(G)$ and $u, v \in \mathcal{I}_G$. But then $uxv \in L(G)$, and so $u\sharp v \in L(M_x)$. Analogously, $L(M_x) \subseteq L(M_w)$. □

Deciding congruence

Since equivalence of DPDAs is decidable,[10] we have

Theorem

Let $w, x \in \Sigma^*$. We can decide whether $w \equiv_{L(G)} x$.

[10] Sénizergues 1997.

Deciding equivalence

Lemma

Let $\approx_G$ be the smallest congruence containing $\rightsquigarrow_G$. Then

$$L(G) = \bigcup_{A \in I} [\vartheta_G(A)]_{\approx_G}$$

Proof.

($\subseteq$) If $w \in L(G)$, then $w \rightsquigarrow_G \vartheta_G(A)$ for some $A \in I$, and so $w \approx_G \vartheta_G(A)$.

($\supseteq$) If $w \approx_G \vartheta_G(A)$, then $w \equiv_{L(G)} \vartheta_G(A)$; but then $w \in L(G)$. □

Note: $\approx_G$ is finitely generated.

Deciding equivalence

Analogous to a result about NTS grammars,[11] we find

Lemma

Let $G_1 = \langle V_1, \rightarrow_1, I_1 \rangle$ *and* $G_2 = \langle V_2, \rightarrow_2, I_2 \rangle$ *be Clark-congruential.*

Then $L(G_1) = L(G_2)$ *if and only if*

(i) *for all* $A \in I_1$*, it holds that* $\vartheta_{G_1}(A) \in L(G_2)$

(ii) *for all* $A \in I_2$*, it holds that* $\vartheta_{G_2}(A) \in L(G_1)$

(iii) *for all pairs* $u \approx_{G_1} v$ *generating* $\approx_{G_1}$*, also* $u \equiv_{L(G_2)} v$

(iv) *for all pairs* $u \approx_{G_2} v$ *generating* $\approx_{G_2}$*, also* $u \equiv_{L(G_1)} v$

[11] Sénizergues 1985.

Deciding equivalence

Theorem

Let G_1 and G_2 be Clark-congruential. We can decide whether $L(G_1) = L(G_2)$.

Deciding Clark-congruentiality

Given a congruence $\equiv$, we can extend it a congruence $\hat{\equiv}$ on $\hat{\Sigma}^*$, by stipulating

$$\frac{\vartheta_G(\alpha) \equiv \vartheta_G(\beta)}{\alpha \mathrel{\hat{\equiv}} \beta}$$

Deciding Clark-congruentiality

Lemma

Let $\equiv$ be a congruence on Σ^*.

The following are equivalent:

(i) *For all productions $A \to \alpha$, it holds that $A \triangleq \alpha$*

(ii) *For all $A \in V$ and $w, x \in L(G, A)$, it holds that $w \equiv x$.*

Proof.

(i) $\Rightarrow$ (ii): If $\beta \Rightarrow_G^* \gamma$, then $\beta \triangleq \gamma$. Thus, if $w, x \in L(G, A)$, then $A \Rightarrow_G^* w, x$, and so $w \triangleq A \triangleq x$. We conclude that $w = \vartheta_G(w) \equiv \vartheta_G(x) = x$.

(ii) $\Rightarrow$ (i): If $A \to \alpha$, then $\vartheta_G(A), \vartheta_G(\alpha) \in L(G, A)$, and so $\vartheta_G(A) \equiv \vartheta_G(\alpha)$. From this, we conclude that $A \triangleq \alpha$. □

Deciding Clark-congruentiality

Theorem

If $\equiv_{L(G)}$ is decidable, then we can decide whether G is Clark-congruential.

Proof.

For $A \to \alpha$, check whether $A \mathrel{\hat{\equiv}}_{L(G)} \alpha$, i.e., whether $\vartheta_G(A) \equiv_{L(G)} \vartheta_G(\alpha)$. □

Corollary

If $L(G)$ is a deterministic CFL, then it is decidable whether G is Clark-congruential.

Further work

Many open questions:

- Are pre-NTS grammars more expressive than NTS grammars?
- Are Clark-congruential grammars more expressive than pre-NTS grammars?

- Is the language of every pre-NTS grammar a DCFL?
- Is the language of every Clark-congruential grammar a DCFL?

- Is it decidable in general whether a given grammar is Clark-congruential?
- Is it decidable whether the grammar of a DCFL is pre-NTS?

Bonus: NTS-like grammars

G is *NTS-like* when $L(G, A) \cap L(G, B) \neq \emptyset$ implies that adding $A \to B$ and $B \to A$ does not change $L(G)$.

Example

Consider the grammars

$$G_5 = \langle \{S, T\}, \{S \to aS + bT + \epsilon,\ T \to bS + aT + \epsilon\}, \{S\} \rangle$$
$$G_6 = \langle \{S, T\}, \{S \to aS + bT + \epsilon,\ T \to aS + a\}, \{S\} \rangle$$

Here $L(G_5) = L(G_5, A) = L(G_5, T) = \{a, b\}^*$; thus, G_5 is NTS-like.

Contrarily, $a \in L(G_6, S) \cap L(G_6, T)$, but adding $T \to S$ changes $L(G_6)$.

Bonus: grammar to DPDA

Lemma

Let G be Clark-congruential, and let R be regular.

We can create a Clark-congruential grammar G_R such that $L(G_R) = L(G) \cap R$.

Lemma

Let $h : \Sigma^* \to \Sigma^*$ be a strictly alphabetic morphism, that is, $h(a) \in \Sigma$ for all $a \in \Sigma$.

We can create a Clark-congruential grammar G^h such that $L(G^h) = h^{-1}(L(G))$.

Bonus: grammar to DPDA

For $a \in \Sigma$, add $\bar{a}$ to Σ.

Let $h : \Sigma \to \Sigma$ be such that $h(a) = h(\bar{a}) = a$.

Create G' such that $L(G') = h^{-1}(L(G))$.

Intuition

G' is the same as G, but positions in every word can be "marked" by $\bar{}$.

Bonus: grammar to DPDA

Note that $\mathcal{I}_G$ is a regular language.

Create G'_w such that $L(G'_w) = L(G') \cap \mathcal{I}_G \bar{w} \mathcal{I}_G$.

Now $G'_w = \{u\bar{w}v : uwv \in L(G),\ u, v \in \mathcal{I}_G\}$.

Intuition

$L(G'_w)$ has words in $L(G)$ with w as a marked substring, with context reduced by $\rightsquigarrow_G$.

Bonus: grammar to DPDA

Lemma

If G is Clark-congruential, we can create a grammar G_ω such that:

(i) *G_ω is Clark-congruential.*

(ii) *G_ω is equivalent to G, i.e., $L(G) = L(G_\omega)$.*

(iii) *If $A \in V$, then $A \in I$ or $L(G, A)$ is infinite.*

(iv) *If $A \to \alpha$ and $L(G, A)$ is finite, then $\alpha \in \Sigma^*$.*

Let $G_w = \langle V_w, \to_w, I_w \rangle$ be such a grammar obtained from G'_w.

Bonus: grammar to DPDA

Lemma

If $A \to \alpha$ exists in G_w, then one of the following holds:

(i) $\vartheta_{G_w}(A) = x_A \bar{w}_\ell$ *and* $\vartheta_{G_w}(\alpha) = x_\alpha \bar{w}_\ell$, *for* $x_A, x_\alpha \in \Sigma_0^*$ *and* $\bar{w}_\ell$ *a prefix of* $\bar{w}$.

(ii) $\vartheta_{G_w}(A) = \bar{w}_r y_A$ *and* $\vartheta_{G_w}(\alpha) = \bar{w}_r y_\alpha$, *for* $y_A, y_\alpha \in \Sigma_0^*$ *and* $\bar{w}_r$ *a suffix of* $\bar{w}$.

(iii) $\vartheta_{G_w}(A) = x_A \bar{w} y_A$ *and* $\vartheta_{G_w}(\alpha) = x_\alpha \bar{w} y_\alpha$, *for* $x_A, y_A, x_\alpha, y_\alpha \in \Sigma_0^*$.

Intuition

Every rule generating $\rightsquigarrow_{G_w}$ overlaps and preserves $\bar{w}$.

Bonus: grammar to DPDA

We can now create a reduction $\rightsquigarrow_{G[w]}$ and a finite set S_w such that

- Every rule generating $\rightsquigarrow_{G[w]}$ contains and preserves $\sharp$.
- $\{x \in \Sigma^* : x \rightsquigarrow_{G[w]} y \in S_w\} = \{u\sharp v : uwv \in L(G),\ u, v \in \mathcal{I}_G\}$

The DPDA M_w acts by reading $u\sharp v$ up to $\sharp$, putting the input on the stack. Then:

- Pop from the stack or read from input into two buffers (encoded in state).
- Whenever possible, reduce according to the rules from $\rightsquigarrow_{G[w]}$.
- When the buffer resembles S_w and the input and stack are empty, accept.

With some analysis, we find that $L(M_w) = \{u\sharp v : uwv \in L(G),\ u, v \in \mathcal{I}_G\}$.