CS:4330 Theory of Computation
Spring 2018

Automata and Languages

Summary

Haniel Barbosa

The University of Iowa

Readings for this lecture

Chapters 1-3 of [Sipser 1996], 3rd edition.

A hierachy of languages

- Regular: $a^n b^m$
- Deterministic Context-free: $a^n b^n$
- Context-free: $a^n b^n \cup a^n b^{2n}$
- Turing decidable: $a^n b^n c^n$
- Turing recognizable: A_{TM}

How to show a language is in a given class?

- Regular language: a DFA, NFA or regexp
  - DFA: $(Q, \Sigma, \delta, q_0, F)$, $\delta : Q \times \Sigma \to Q$
  - NFA: $(Q, \Sigma, \delta, q_0, F)$, $\delta : Q \times (\Sigma \cup \{\epsilon\}) \to \mathscr{P}(Q)$
  - Regular Expression: base cases and recursive cases
- Context-free language: PDA, context-free grammar
  - PDA: $(Q, \Sigma, \Gamma, \delta, q_0, F)$, $\delta : Q \times \Sigma_\epsilon \times \Gamma_\epsilon \to \mathscr{P}(Q \times \Gamma_\epsilon)$
  - CFG: (V, Σ, R, S), R is a finite set of *rules* of the form $lhs \to rhs$, in which $lhs \in V$ and $rhs \in (V \cup \Sigma)^*$
- Turing recognizable: TM and variants, i.e. multi-tape TM, nondeterministic TM, 2-PDA etc.
  - TM: $(Q, \Sigma, \Gamma, \delta, q_0, q_{\text{accept}}, q_{\text{reject}})$, $\delta : Q \times \Gamma \to Q \times \Gamma \times \{L, R\}$
- Turing decidable: Turing machine decider, i.e. a TM that always halts

Closure properties

Operation	Regular	Ctx-free	Deterministic Ctx-free	Decidable	Turing recognizable
union	yes	yes	no	yes	yes
concatenation	yes	yes	no	yes	yes
star	yes	yes	no	yes	yes
intersection	yes	no	no	yes	yes
intersect w/ regular	yes	yes	yes	yes	yes
complement	yes	no	yes	yes	no
reversal	yes	yes	no	yes	yes
shuffle	yes	no	no	yes	yes

How to show a language is *not* in a given class?

- ▷ Regular language: Pumping lemma
- ▷ Context-free language: Pumping lemma
- ▷ Turing decidable: TM decider as a witness or through reducibility techniques
- ▷ Turing recognizable: diagonalization method and reducibility techniques

Pumping lemma for regular languages

Theorem

If A is a regular language, then there is a number p (the pumping length) such that, if s is any string in A of length at least p, then s may be divided intro three pieces, $s = xyz$, satisfying the conditions:

(1) for each $i \geq 0$, $xy^iz \in A$

(2) $|y| > 0$

(3) $|xy| \leq p$

Example

Prove that $B = \{0^n 1^n \mid n \geq 0\}$ is not regular

Proof

Assume that B is regular and let p be the pumping length of B. Choose $s = 0^p 1^p \in B$; therefore $|0^p 1^p| > p$. By pumping lemma, $s = xyz$ such that for any $i \geq 0$, $xy^i z \in B$

Consider the cases:

1. y consists of $0s$ only. Then for some i, $xy^i z$ has more $0s$ than $1s$ and is not in B, violating condition (1).
2. y consists of $1s$ only. This leads to the same contraction.
3. y consists of $0s$ and $1s$. Now $xy^i z$ may have the same number of $0s$ and $1s$ but they are out of order with some $1s$ before some $0s$. Therefore it cannot be in B either.

The contraction is unavoidable if we make the assumption that B is regular, therefore B is not regular.

Pumping lemma for context-free languages

Theorem

If A is a context-free language, then there is a number p (the pumping length) such that if $s \in A$ and $|s| \geq p$, then s may be divided into five pieces, $s = uvxyz$, satisfying the conditions:

1. *For each $i \geq 0$, $uv^ixy^iz \in A$*
2. $|vy| \geq 0$
3. $|vxy| \leq p$

Example

We will use the pumping lemma to show that the language $B = \{\mathrm{a}^n\mathrm{b}^n\mathrm{c}^n \mid n \geq 0\}$ is not context-free by assuming it is and deriving a contradiction.

Proof

- Let p be the pumping length for B. Consider the string $s = \mathrm{a}^p\mathrm{b}^p\mathrm{c}^p$, for which it holds that $s \in B$ and $|s| \geq p$.
- By condition 2 of pumping lemma, in any division $s = uvxyz$ either v or y is not empty
  - When both v and y contain only one type of symbol ($\mathrm{a}, \mathrm{b}, \mathrm{c}$), v does not contain both a's and b's or both b's and c's; the same holds for y.
    In this case uv^2xy^2z cannot contain an equal number of $\mathrm{a}, \mathrm{b}, \mathrm{c}$.
  - When either v or y contain more than one type of symbols $\mathrm{a}, \mathrm{b}, \mathrm{c}$, uv^2xy^2z may contain equal number of $\mathrm{a}, \mathrm{b}, \mathrm{c}$, but they do not occur in the right order.
- Since one of these cases must occur, a contradiction is unavoidable, and therefore B is not a CFL.