Schutzenberger, McNaughton-Papert Theorems: Linking logic, automata, regular expressions and monoids

Harsha Gurnani Soumyo Biswas

ATC Seminar

Nov 13, 2015

Overview

Theorem

Schutzenberger (1965), McNaughton-Papert (1971)

Let $L \subseteq \mathcal{A}^*$. Then the following are equivalent:

1. $\mathcal{A}_{\equiv L}$ is counter-free DFA.
2. L is accepted by a counter-free DFA.
3. L is definable in FO $(<)$
4. L is definable by a star-free extended regular expression.
5. L is recognized by an aperiodic finite monoid.
6. $M(L)$, the syntactic monoid is aperiodic.

Plan of the proof

- $\mathcal{A}_{\equiv L}$ is counter-free $\Rightarrow$ L is accepted by a counter-free DFA
- L is accepted by a counter-free DFA $\Rightarrow$ $\mathcal{A}_{\equiv L}$ is counter-free

Equivalence of (1) and (2)

- Star-free expression $\Rightarrow$ FO $(<)$ - definable
- FO $(<)$ - definable $\Rightarrow$ $M(L)$ is aperiodic
- $M(L)$ is aperiodic $\Rightarrow$ Recognized by an aperiodic finite monoid
- Recognized by an aperiodic finite monoid $\Rightarrow$ Star-free expression

Equivalence of (3), (4), (5) and (6)

- Counter-free $\mathcal{A}_{\equiv L}$ $\Rightarrow$ $M(L)$ is aperiodic
- $M(L)$ is aperiodic $\Rightarrow$ L is non-counting $\Rightarrow$ $\mathcal{A}_{\equiv L}$ is counter-free

Equivalence of (1) and (6)

Counter-free DFA-I

- $\mathcal{A}_{\equiv L}$ is counter-free $\Rightarrow$ L is accepted by a counter-free DFA

  Trivial

Counter-free DFA-II

- L is accepted by a counter-free DFA $\Rightarrow$ $\mathcal{A}_{\equiv L}$ is counter-free

Proof. Suppose $\mathcal{A}_{\equiv L}$ has a counter: w on $Q_0 = \{q_0, q_1, \ldots q_{k-1}\}$. Every DFA $\mathcal{A}'$ accepting L is a refinement of $\mathcal{A}_{\equiv L}$. Choose a state p_0 in $\mathcal{A}'$ that is in $[q_0]$, and create:

$Q' = \{\ p_0, \widehat{\delta'}(p_0, w), \widehat{\delta'}(p_0, w^2) \ldots\ \}$. Then $\widehat{\delta'}(p_0, w^{nk+j}) = [q_j]$.

Counter-free DFA-II

- L is accepted by a counter-free DFA $\Rightarrow$ $\mathcal{A}_{\equiv L}$ is counter-free

Proof. Suppose $\mathcal{A}_{\equiv L}$ has a counter: w on $Q_0 = \{q_0, q_1, \ldots q_{k-1}\}$. Every DFA $\mathcal{A}'$ accepting L is a refinement of $\mathcal{A}_{\equiv L}$. Choose a state p_0 in $\mathcal{A}'$ that is in $[q_0]$, and create:

$Q' = \{\ p_0, \widehat{\delta'}(p_0, w), \widehat{\delta'}(p_0, w^2) \ldots\ \}$. Then $\widehat{\delta'}(p_0, w^{nk+j}) = [q_j]$.

As it is a finite automaton, each q_j only has finite copies in $\mathcal{A}'$. Thus for some n, n', j

$$\widehat{\delta'}(p_0, w^{nk+j}) = \widehat{\delta'}(p_0, w^{n'k+j}) = p'$$

Then w is a counter in $\mathcal{A}'$ over $Q'_0 = \{p', \widehat{\delta'}(p', w), \widehat{\delta'}(p', w^2) \ldots p''\}$

Star-free to first order logic

- Star-free expression $\Rightarrow$ FO($<$) definable

Lemma 2.1 For every star-free expression r there is an FO-formula $\phi_r(x, y)$ that expresses that "the segment from x up to y is in $L(r)$", that is $w \in L(r)$ iff $w \models \phi_r$[min,max] for all $w \in \Sigma^+$. [1]

[1] *Applied Automata Theory*, Thomas, Wolfgang

Star-free to first order logic

Proof by induction. For atomic formulae:

- $r = a \ : \varphi_r(x, y) := \quad (x = y) \wedge Q_a(x)$
- $r = \overline{\phi} \ : \varphi_r(x, y) := \quad \exists z(x \leq z \wedge z \leq y \wedge \neg(z = z))$

Star-free to first order logic

Proof by induction. For atomic formulae:

- $r = a \ : \varphi_r(x,y) := \quad (x = y) \wedge Q_a(x)$
- $r = \overline{\phi} \ : \varphi_r(x,y) := \quad \exists z(x \leq z \wedge z \leq y \wedge \neg(z = z))$

Let r and s be two star-free expressions with FO sentences $\varphi_r(x,y)$ and $\varphi_s(x,y)$. We show for

- $r + s : \varphi_r(x,y) \vee \varphi_s(x,y)$
- $r \cap s : \varphi_r(x,y) \wedge \varphi_s(x,y)$
- $\overline{r} \quad : \neg\varphi_r(x,y)$
- $r \cdot s :$
  $\exists z, z'(x \leq z \wedge Succ(z,z') \wedge z' \leq y \wedge \varphi_r(x,z) \wedge \varphi_s(z',y))$

First order logic to aperiodic $M(L)$

- FO($<$) definable $\Rightarrow$ $M(L)$ is aperiodic

Proof. [2] Atomic first-order formulae and using quantifiers:

- For free variables $\{x_1, x_2 \dots x_j\}$, j-ary relations R_i^j
- For $a \in A$, $Q_a(x)$
- For formulae φ and ψ, and free variable x: $\varphi \wedge \psi$, $\varphi \vee \psi$, $\neg\varphi$, $\forall x\varphi$ and $\exists x\varphi$

[2] *Algebraic and logical characterization of star-free languages*, Thomas Zeume

First order logic to aperiodic $M(L)$

Defn. *(V-structure)* Let V be a finite set of first-order variables. A V -structure over A is a word

$$(a_1, U_1)(a_2, U_2)\ldots(a_n, U_n)$$

from the alphabet $\mathcal{A} = Ax\mathcal{P}(V)$ such that

- $U_i \cap U_j = \emptyset \quad \text{for } i \neq j$
- $\cup_i U_i = V$

First order logic to aperiodic $M(L)$

Defn. *(Satisfaction)* A V -structure $w = (a_1, U_1) \dots (a_n, U_n)$ satisfies a formula φ with respect to an interpretation I, $(w \models_I \varphi)$, if inductively

- $w \models_I Q_a(x)$ if and only if w contains a letter (a, S) and $x \in S$
- $w \models_I R_i^j(x_1, \dots, x_j)$ if and only if $(p_1, \dots, p_j) \in P_i^j$, where the p_i are defined by $x_i \in U_{p_i}$.

First order logic to aperiodic $M(L)$

Defn. *(Satisfaction)* A V -structure $w = (a_1, U_1) \dots (a_n, U_n)$ satisfies a formula φ with respect to an interpretation I, $(w \models_I \varphi)$, if inductively

- $w \models_I Q_a(x)$ if and only if w contains a letter (a, S) and $x \in S$
- $w \models_I R_i^j(x_1, \dots, x_j)$ if and only if $(p_1, \dots, p_j) \in P_i^j$, where the p_i are defined by $x_i \in U_{p_i}$.
- $w \models_I \varphi \wedge \psi$ if and only if $w \models_I \varphi$ and $w \models_I \psi$.
- $w \models_I \neg\varphi$ if and only if $w \not\models_I \varphi$.
- $w \models_I \exists x\varphi$ if and only if there is an i, $1 \leq i \leq n$ such that

$$(a_1, U_1) \dots (a_{i1}, U_{i1})(a_i, U_i \cup \{x\})(a_{i+1}, U_{i+1}) \dots (a_n, U_n) \models \varphi$$

$$\text{Thus} \qquad L(\varphi) = \{w \in (A \times \mathcal{P}(V))^* \mid w \models_I \varphi\}$$

In FO($<$), there is a single binary relation - $<$.

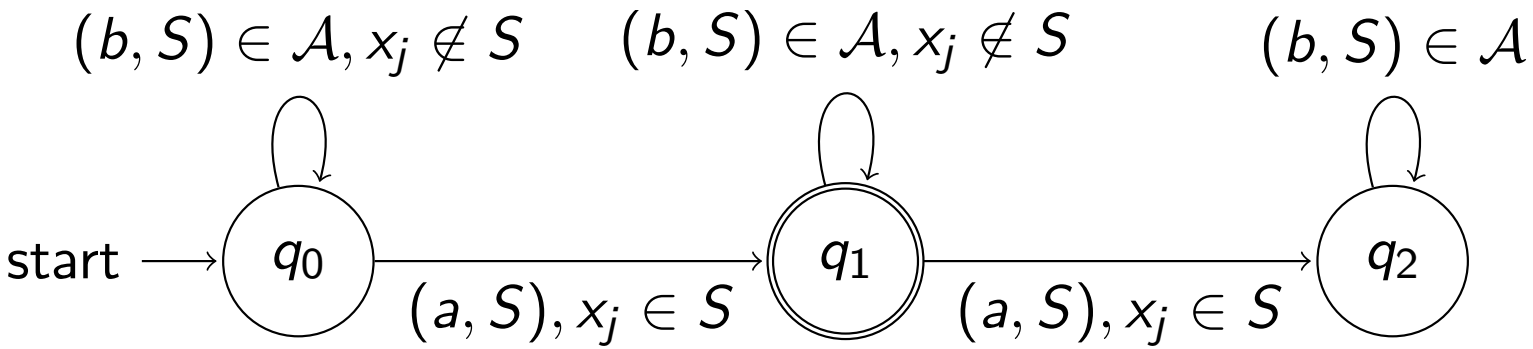
First order logic to aperiodic $M(L)$

Finiteness of $M(L(\varphi))$

- DFA to check each x_j belongs to only one U_i
- DFA for $Q_a(x)$
- DFA for $x < y$
- Closure under intersection and complementation of DFA
- $\exists x\varphi$ - Non-deterministic - $x \in U_i$ for any i

First order logic to aperiodic $M(L)$

Fig1: Check x_j in only one set U_i



First order logic to aperiodic $M(L)$

Fig2: DFA for $Q_a(x)$

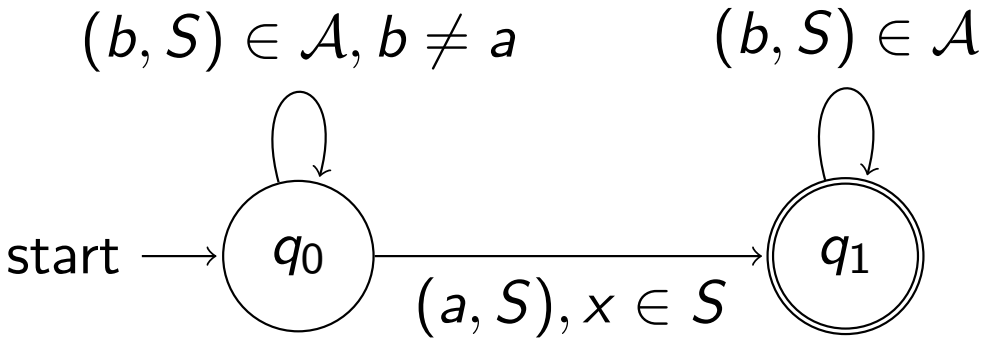
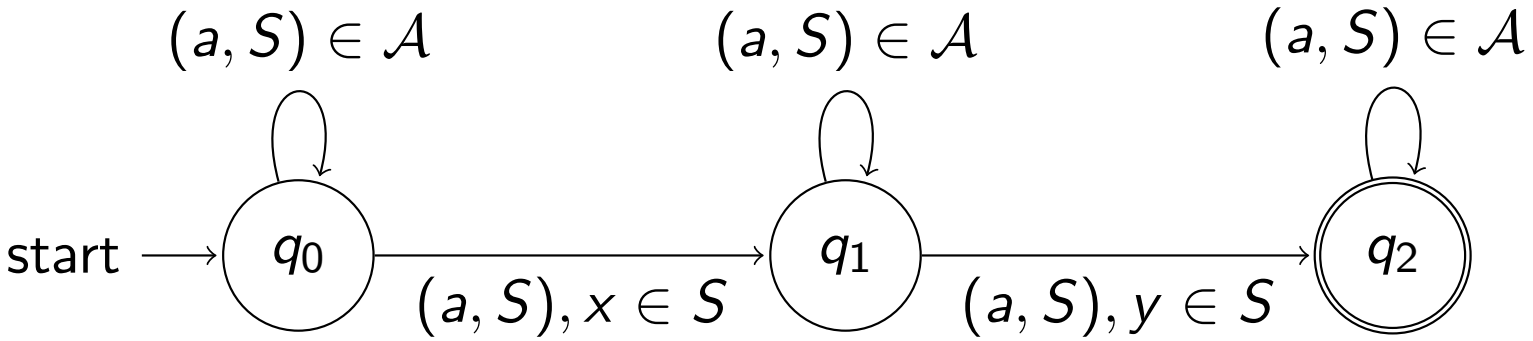


Fig3: DFA for $x < y$



First order logic to aperiodic $M(L)$

Aperiodicity of $M(L(\varphi))$

$\mathcal{A} = A \times \mathcal{P}(V)$. For all $w \in \mathcal{A}^*$, we need to find $k > 0$ such that $[w]^k_{\equiv_\varphi} = [w]^{k+1}_{\equiv_\varphi}$

- For $\varphi := Q_a(x)$ and $\varphi := x < y$, $k = 2$ works. As $uw^2v \in L(\varphi)$ does not allow repetition of free variables in the V-structure, the required alphabets are present in u and v.

First order logic to aperiodic $M(L)$

Aperiodicity of $M(L(\varphi))$

$\mathcal{A} = A \times \mathcal{P}(V)$. For all $w \in \mathcal{A}^*$, we need to find $k > 0$ such that $[w]^k_{\equiv_\varphi} = [w]^{k+1}_{\equiv_\varphi}$

- For $\varphi := Q_a(x)$ and $\varphi := x < y$, $k = 2$ works. As $uw^2v \in L(\varphi)$ does not allow repetition of free variables in the V-structure, the required alphabets are present in u and v.
- For $\varphi \wedge \psi$, take maximum of k_φ and k_ψ
- For $\neg\varphi$, the same k as φ works
- For $\exists x\varphi$, use $2k + 1$.

M(L) to recognition by aperiodic monoid

- $M(L)$ is finite and aperiodic $\Rightarrow$ L is recognized by a finite, aperiodic monoid

  Trivial, as $M(L)$ recognizes L

Aperiodic Monoids to star-free expressions

- L is recognized by a finite, aperiodic monoid $\Rightarrow$ L has a star-free extended regular expression

Lemma: Let $h : A^* \to M$ be a morphism into a finite aperiodic monoid. Then for every $m \in M$, the inverse image $h^{-1}(m)$ is definable by a star-free expression.[3]

[3] Course notes: *Advance Topics in Automata*, Mikolaj Bojanczyk

Aperiodic Monoids to star-free expressions

Some primers to monoids

$(M, \cdot, e)$ is finite, (aperiodic). For any $m, n \in M$, we define:

- m is a **prefix** of n if $n = m \cdot x$
- m is a **suffix** of n if $n = x \cdot m$
- m is an **infix** of n if $n = y \cdot m \cdot x$

Consequently, we get 3 "quasi-orders": $\mathcal{R}$-ordering, $\mathcal{L}$-ordering and $\mathcal{J}$-ordering.

If m is a prefix of n, $nM \subseteq mM$, where $mM = \{m \cdot x \mid x \in M\}$.
We write m is $\mathcal{R}$-simpler than n or:

$$m \geq_{\mathcal{R}} n$$

We also have the corresponding 'equivalence' classes.

Aperiodic Monoids to star-free expressions

Lemma (Eggbox Lemma) : For $m, n \in M$, if m is a prefix of n and they are both in the same $\mathcal{J}$-class, then also n is a prefix of m.

Proof. $n = mx$ and $m = ynz$. By iterative substitution, we get:

$$m = y^i m (xz)^i$$

Using the idempotent power $x^{i+1} = x^i$

$$m = y^i m (xz)^i (xz) = mxz = nz$$

Aperiodic Monoids to star-free expressions

m and n are $\mathcal{H}$-equivalent if they have the same $\mathcal{R}$-class and $\mathcal{L}$-class.

Lemma ($\mathcal{H}$-Dichotomy lemma) : Let H be an $\mathcal{H}$-class included in a $\mathcal{J}$-class J. Then either

- $mn \notin J$ for every $m, n \in H$; or
- H is a group.

Proof. If $mn \in J$ for some $m, n \in H$, show that H is a group. (Use the Eggbox lemma and existence of idempotent power.)

Corollary 1: H is a group iff it contains an idempotent.
Corollary 2: The $\mathcal{J}$-class of the identity is a group.

Aperiodic Monoids to star-free expressions

Lemma: In an aperiodic monoid, every $\mathcal{H}$-class has a single element.

Proof. If m, n are both $\mathcal{R}$ and $\mathcal{L}$ equivalent:

$$n = mx \qquad m = yn$$

Then by iteratively substituting, we get

- $m = x^i \cdot m \cdot y^i$
- $n = x^i \cdot m \cdot y^{i+1}$

Using the idempotent power i, we get $m = n$

Aperiodic Monoids to star-free expressions

Back to proof for main lemma: By induction on the position of m in the $\mathcal{J}$-ordering

Base: Simplest $\mathcal{J}$-class which contains the identity.
Contains *only* the identity as no non-trivial groups in M. Then, a star-free expression for $h^{-1}(J_e)$ is

$$\neg(\neg\emptyset \cdot B \cdot \neg\emptyset)$$

where $B = \{a_{i_1}, a_{i_2}, \ldots a_{i_k}\} \subseteq A$ such that $h(a_{i_j}) \neq e$

Aperiodic Monoids to star-free expressions

Induction step: Fix some $\mathcal{J}$-class J in the monoid, and assume that the inverse image $h^{-1}(m)$ is star-free for every m which has a $\mathcal{J}$-class strictly simpler than J. We will prove that the same holds for every $m \in J$.

Define $R_m \subseteq A^*$ to be the set of words which have a *prefix* with image in M that is $\mathcal{R}$-equivalent to m.
Likewise define L_m to be the set of words which have a *suffix* whose image is $\mathcal{L}$-equivalent to m.

Every word with image m belongs to both R_m and L_m.

Aperiodic Monoids to star-free expressions

Lemma: $\exists$ a star-free expression for R_m and L_m for all $m \in J$.

Proof. Let $w \in R_m$. Let $w = uv$, such that u is the shortest prefix of w with $h(u)$ in same $\mathcal{J}$-class as m. Then $u \in R_m$.
u will be of the form $u'a$, where $h(u')$ is $\mathcal{J}$-simpler than m. Then

$$R_m = \cup_{n,a}\ h^{-1}(n) \cdot a \cdot \neg\emptyset$$

where n is $\mathcal{J}$-simpler than m, $a \in A$ and $n \cdot h(a)$ is $\mathcal{R}$-equivalent to m.

Aperiodic Monoids to star-free expressions

Lemma: For every $m \in J$, $h^{-1}(m)$ is given by

$$L_m \cap R_m \; - \; \cup_{n,a} \; L_n \cdot a \cdot \neg\emptyset$$

where union is over n which are $\mathcal{J}$-equivalent to m, but $n \cdot h(a)$ is not.

Due to finiteness of monoid, these are finite unions and hence we have a star-free expression for every $m \in J$.

Counter-freeness to Aperiodicity

- $\mathcal{A}_{\equiv L}$ is counter-free $\Rightarrow$ $M(L)$ is aperiodic

Proof. [4] Assume $M(L)$ is not aperiodic. Thus, there exists non-trivial group G. Pick $g \in G$, not the identity element, such that $< g >= \{g, g^2, \ldots, g^k\}$ is a group with $g^k = e$ and $g^j \neq e$ for $j < k$.

As $M(L) \cong A^* / \equiv_L$, let $g = [u]_{\equiv_L}$. Then $g^i = [u^i]_{\equiv_L}$. But $g^k \neq g^{k+1}$, so $[u^k] \neq [u^{k+1}]$.
Thus, $\exists$ state q such that $\widehat{\delta}(q, u^k) \neq \widehat{\delta}(q, u^{k+1})$

Then u is a counter over

$$Q_0 = \{\widehat{\delta}(q, u^k), \widehat{\delta}(q, u^{k+1}), \ldots, \widehat{\delta}(q, u^{2k-1})\}$$

[4] *Applied Automata Theory*, Wolfgang Thomas

Aperiodicity to Counter-freeness

Defn. L is non-counting if

$$\exists n_0 \; \forall n \geq n_0 \; \forall u, v, w \in A^* \quad uv^n w \in L \Leftrightarrow uv^{n+1} w$$

This means for $n \geq n_0$ either all $uv^n w$ are in L, or none is.

Aperiodicity to Counter-freeness

- Recognition by finite, aperiodic $\Rightarrow$ L is non-counting monoid

Proof. Assume L is not non-counting.

In a group, every element m has an idempotent power i such that $(m^i)^2 = m^i$. Take LCM of all i to get K.
As L is not non-counting, $\exists u, v, w$ and arbitrarily large $n > K$ such that $uv^n w \in L$ but $uv^{n+1} w \notin L$.
If $[x]$ is image of word x in M, then $[v^n] \neq [v^{n+1}]$.

As n can be arbitrarily large $[v^K] \neq [v^{K+1}]$ but $[v^K] = [v^{2K}]$.
Then $\{[v^K], [v^{K+1}], \ldots [v^{2K-1}]\}$ form a cyclic group in M.

Aperiodicity to Counter-freeness

- L is non-counting $\Rightarrow \mathcal{A}_{\equiv_L}$ is counter-free

Proof. Assume $\mathcal{A}_{\equiv_L}$ has a counter v on $Q_0 = \{q_0, q_1 \dots q_m\}$. Then $\delta(q_0, v^{km+1}) = q_1 \neq q_0$. Thus $\exists u, w$ such that

$$\delta(s, u) = q_0,\ \delta(q_0, w) \in F \wedge \delta(q_1, w) \notin F$$

Then $uv^{km}w \in L$ but $uv^{km+1}w \notin L$, for arbitrarily large k.

So L cannot be non-counting.

References


Dr. Wolfgang Thomas

Applied automata theory


Dr. Thomas Zeune

Algebraic and logical characterization of star-free languages


Mikolaj Bojanczyk

Advanced Topics in Automata