Schützenberger's Aperiodic Monoid Characterization of Star-Free Languages

November 25, 2021

Preliminaries: Star-Free Languages

- A language that can be described by a regular expression of the alphabet letters, the empty set symbols and Boolean operators is known as a star-free language
- Concatenation is allowed but Kleene Star is not (hence, Star-Free)
- As is the point of this presentation we'll see later that Star-Free languages can be characterized as those having aperiodic syntactic monoids.

Monoids

Definition 1

A monoid is a set M with a binary operation $\cdot : M \times M \to M$ with the following properties

- **Associativity**: $(a \cdot b) \cdot b = a \cdot (b \cdot c) \quad \forall a, b, c \in M$
- **Identity element**: $\exists e \in M$ (called the identity) such that $e \cdot m = m \cdot e = m \quad \forall m \in M$

The monoid as a whole is represented as the triplet $(M, \cdot, e)$

Monoid Examples

- All groups are also monoids (with the extra group structure). Eg. $(\mathbb{Z}, +, 0), (\mathbb{R}, \cdot, 1)$ (Since monoids don't require inverses, 0 can be included with $\mathbb{R}$)
- $(A^*, \cdot, \epsilon)$, where A is an alphabet, $\cdot$ is the concatentation operation and ϵ is the empty string.
- $(2^A, \cap, A)$, where A is a set, 2^A is its power set and $\cap$ is the intersection operation

Idempotents

An element $m \in M$ is called idempotent if $m^2 := m \cdot m = m$

Proposition 2

If M is a finite monoid, then $\forall m \in M, \exists k \in \mathbb{N}$ such that m^k is idempotent

Proof: Let $m \in M$. Then $\forall n \in \mathbb{N}, m^n \in M$. But theres only finitely many elements in M. So, there must exist $p, q \in \mathbb{N}$ such that $m^{p+q} = m^p$.
Then, $m^{p+q} \cdot m^q = m^{p+q}$
$m^{p+2q} = m^{p+q} = m^p$. Clearly, this holds for all multiples of q.
(Seen by induction)

Idempotents

Take b such that $bq > p$,
$(m^{bq})^2 = m^{2bq} = m^{bq-p} \cdot mbq + p = m^{bq-p} \cdot m^p = m^{bq}$
Thus, m^{bq} is idempotent.

Corollary 3

$\exists \omega \in \mathbb{N}$ such that m^{ω} is idempotent $\forall m \in M$,

Proof: Since there are only finitely many $m \in M$, take all the elements, take the powers for which each element in idempotent, and take the LCM of all the powers.
This ω is called the exponent of M

Green's Relations

Let M be a monoid. We define the following four relations on M

1. $s \leq_R t$ iff $s = tu$ for some $u \in M$
2. $s \leq_L t$ iff $s = ut$ for some $u \in M$
3. $s \leq_J t$ iff $s = vut$ for some $u, v \in M$
4. $s \leq_H t$ iff $s \leq_R tands \leq_L t$

Equivalently

1. $s \leq_R t$ iff $sM \subseteq tM$
2. $s \leq_L t$ iff $Ms \subseteq Mt$
3. $s \leq_J t$ iff $MsM \subseteq MtM$
4. $s \leq_H t$ iff $s \leq_R tands \leq_L t$

Green's Relations

We'll also proceed to define the following equivalence relations

1. $s\mathcal{R}t$ iff $sM = tM$
2. $s\mathcal{L}t$ iff $Ms = Mt$
3. $s\mathcal{J}t$ iff $MsM = MtM$
4. $s\mathcal{H}t$ iff $s\mathcal{R}t$ and $s\mathcal{L}t$
5. $s\mathcal{D}t$ iff $\exists u \in M$ such that $s\mathcal{R}u$ and $u\mathcal{L}t$
   or $\exists v \in M$ such that $s\mathcal{L}v$ and $v\mathcal{R}t$

Green's Relations

Theorem 4

We claim that in a finite monoid, the relations $\mathcal{J}$ and $\mathcal{D}$ are equal. Additionally,

1. $s \leq_J sm \implies s\mathcal{R}(sm)$
2. $s \leq_J ms \implies s\mathcal{L}(ms)$
3. $s\mathcal{J}t \wedge s \leq_R t \implies s\mathcal{R}t$
4. $s\mathcal{J}t \wedge s \leq_L t \implies s\mathcal{L}t$
5. $\exists u, v \in M$ such that $(s = usv) \implies (us)\mathcal{H}s\mathcal{H}(sv)$

Green's Lemma

Theorem 5

Let the elements $s, t \in M$ such that $s\mathcal{R}t$. Let $s = tp$ and $t = sq$. Then, the following holds:

1. The map $x \to xp$ is a bijection from $\mathcal{L}(t)$ onto $\mathcal{L}(s)$
2. The map $x \to xq$ is a bijection from $\mathcal{L}(s)$ onto $\mathcal{L}(t)$
3. The above two maps are inverses of each other
4. $\mathcal{H}$ classes are preserved by these maps

Ordered Monoids

Definition 6

An ordered monoid $(M, \leq)$ is a monoid along with an order relation *leq*, such that $x \leq y \implies uxv \leq uyv, \ \forall u, v \in M$

Definition 7

A subset P of an ordered monoid M is called an upper set if $u \in P$ and $u \leq v \implies v \in P$

Definition 8

Given an upper set P define the **syntantic order** relation on $(M, \leq)$ as, $u \leq_P v$ iff $xuy \in P \implies xvy \in P \ \forall x, y \in M$

Examples of Ordered Monoids

- The monoid $(2^A, \cap, A, \subseteq)$ is an ordered monoid with the subset relation
- We can equip all monoids with the order relation '$=$' to obtain an ordered monoid $(M, =)$
- $(\mathbb{N}, \times, 1, \leq)$ is an ordered relation with the usual definitions of $\times$ and $\leq$

Homomorphisms

Definition 9

A homomorphism (sometimes called a 'morphism') is a map between two monoids $(M, \cdot, id_M)$ and $(N, \times, id_N)$, $\varphi : M \to N$ such that
$\varphi(m_1 \cdot m_2) = \varphi(m_1) \times \varphi(m_2)\ \forall m_1, m_2 \in M$
It is easy to make the two following observations :

- $\varphi(id_M) = id_N$
- $(\varphi(M), \times, \varphi(id_M))$

Recognisable Subsets and Languages

Definition 10

A subset L of a monoid M is said to be recognisable if there exists a finite monoid N and a morphism $\varphi : M \to N$ and a subset $X \subseteq N$ such that $L = \varphi^{-1}(X)$
We say that (φ, X) recognises L

Definition 11

A language L under an alphabet A is said to be recognisable if it is a recognisable subset of the monoid $(A^*, \cdot, \epsilon)$ (where $\cdot$ is the concatenation operator)

Transition Monoids

Definition 12

Given a DFA $\mathcal{A} = (Q, A, \delta, s, F)$ (the same terminology used in class), we define its transition monoid as follows:
Given a word $w \in A^*$, we define the map $f_w : Q \to Q$ as
$f_w(q) = \hat{\delta}(q, w)$

Then, the set $F_\mathcal{A} = \{f_w | w \in A^*\}$ is a monoid with the composition operation and the identity map as the identity, i.e
$(F_\mathcal{A}, \circ, id)$ is a monoid called the transition monoid.

Transition Monoids

Given a DFA $\mathcal{A}$, the language $L(\mathcal{A})$ is recognised by the morphism-set pair $(\varphi, \varphi(L(\mathcal{A})))$, where φ is the map that sends w to f_w as defined in the last slide.

Given a recognisable language $L \subseteq A^*$, recognised by the finite monoid M via the morphism-set pair (φ, X), we can define the following DFA, $\mathcal{A} := (M, A, \delta, id, X)$, where
$\delta(m, a) := m \cdot \varphi(a), \ \forall m \in Ma \in A$
It can be easily seen that $L = L(\mathcal{A})$
Thus, the set of regular languages over an alphabet A is same as the set of recognisable languages over A

Syntactic Monoid

Definition 13

Given a subset X of a monoid M, we say that u is syntactically congruent to v over X, denoted as $u \cong_X v$ iff
$xuy \in X \iff xvy \in X \; \forall x, y \in M$
$\cong_X$ is an equivalence relation on M

Definition 14

The syntactic monoid of $X \subseteq M$ is the monoid $M/\cong_X$, i.e its the monoid with the same operation as M, but the elements and identity of $M/\cong_X$ are the equivalence classes of M under the equivalence relation $\cong_X$.
The ordered syntactic monoid is the ordered monoid $(M/\cong_X, \leq_X)$ (with $\leq_X$ being the syntactic order)

Syntactic Monoid of a Language

Definition 15

A morphism between two monoids M and N is said to be an isomorphism if it is bijective.
Two monoids M and N are said to be isomorphic if there exists an isomorphism between the two.

Proposition 16

The syntactic morphism of a recognisable language under an alphabet is isomorphic to the transition monoid of its minimal DFA

Aperiodic Monoid

Definition 17

We say that a finite monoid is aperiodic if $\exists n \in \mathbb{N}$ such that $m^n = m^{n+1}\ \forall m \in M$

Proposition 18

A finite ordered monoid is aperiodic iff $\forall \mathrm{m} \in M,\ \exists n \in \mathbb{N}$ such that $m^{n+1} \leq m^n$

Proof:

If M is aperiodic, then $\exists n \in \mathbb{N}$ such that
$m^{n+1} = m\ \forall m \in M \implies m^{n+1} \leq m^n$
Now, if $\forall m \in M,\ \exists m_n \in \mathbb{N}$ such that $m^{n+1} \leq m^n$, take a multiple ω of the exponent ω' that is greater than all the m_n's. So, we get, $\forall m \in M,\ m^\omega = m^{2\omega} \leq m^{2\omega-1} \leq m^{2\omega-2} \ldots m^{\omega+1} \leq m^\omega$
So, we have $m^\omega \leq m^{\omega+1} \leq m^{omega}$ and so, $m^\omega = m^{\omega+1}$
Thus, the monoid is aperiodic.

Some Lemmas on Aperiodic Monoids

We'll state some lemmas (without proof) that we will be using in the proof later.

Lemma 19

If $L_1, L_2 \subseteq A^*$ are recognisable languages recognised by the aperiodic monoids M_1 and M_2, then let $L = L_1L_2$ be recognised by M.
Then, M is aperiodic.

Lemma 20

If a finite monoid M is aperiodic then it is $\mathcal{H}$-trivial

Lemma 21

Let m be an element of an aperiodic monoid M. Then,
$\{m\} = (mM \cap Mm) \backslash J_m$, with $J_m := \{s \in M : m \notin MsM\}$

Simplification Lemma

Lemma 22

If M is an aperiodic monoid and $pqr = q$ for some $p, q, r \in M$, then $pq = q = qr$

Proof:

$pqr = q \implies p(pqr)r = q \implies p^2qr^2 = q$

Extending this via induction, we get $p^nqr^n = q \; \forall n \in \mathbb{N}$

Since M is aperiodic, this holds for the $n_0 \in \mathbb{N}$ such that $p^{n_0} = p^{n_0+1}$

So, $p^{n_0+1}qr^{n_0} = p(p^{n_0}qr^{n_0}) = pq$

Similarly, we get $qr = q$ (Using $r^{n_0} = r^{n_0+1}$)

Star Free Languages

Definition 23

Given an alphabet A, the set of star-free languages $\mathfrak{R}$ is the smallest subset of 2^{A^*} such that:

- $\phi \in \mathfrak{R}$, $\{\epsilon\} \in \mathfrak{R}$, $\{a\} \in \mathfrak{R}$, $\forall a \in A$
- $S, T \in \mathfrak{R} \implies A^* \backslash S \in \mathfrak{R},\ S \cup T \in \mathfrak{R},\ S \cdot T \in \mathfrak{R}$

Notation

We will, later on, for brevity use $+$ to denote $\cup$, 0 to denote ϕ, 1 to denote $\{\epsilon\}$, a to denote $\{a\}$ $\forall a \in A$, L^c to denote complementation and $L_1 L_2$ to denote $L_1 \cdot L_2$

Examples of Star Free Languages

- All finite languages are Star-Free
- A^* is star-free, since $A^* = \phi^c = 0^c$
- More examples in the proof

Schützenberger's Theorem

Theorem (Schützenberger)
A language is star-free iff its syntactic monoid is aperiodic.

Proof of Schützenberger's Theorem

Proof: [Sch65; Kum;]
Let A be the alphabet. For one direction, Define $\mathcal{A}(A)$ as the set of recognisable languages over A, whose syntactic monoids are aperiodic. Then,

- $\phi, \{\epsilon\}, \{a\} \in \mathcal{A}(A), \forall a \in A$ (Using the trivial monoid and the unique aperiodic monoid of cardinality 2 in different ways, with the natural homomorphisms)
- $\mathcal{A}(A)$ is closed under complementation.
- $\mathcal{A}(A)$ is closed under finite intersection, which along with the previous property gives that it is closed under finite union.
- $\mathcal{A}(A)$ is closed under finite product.

Thus $\mathcal{A}(A)$ contains all the star-free languages over A.

Proof of Schützenberger's Theorem

In the other direction, we must prove that the set of all star-free languages over A contains $\mathcal{A}(A)$.
Let M be an aperiodic monoid, and let $\varphi : A^* \to M$ be any monoid morphism over it. It is enough to show that $\varphi^{-1}(P)$ is star-free $\forall P \subseteq M$.
Since M is finite, P is finite and $\varphi^{-1}(P) = \sum_{m \in P} \varphi^{-1}(m)$, so WLOG we may assume that P is a singleton set.

Proof of Schützenberger's Theorem

Claim 1 $\varphi^{-1}(m)$ is a star-free language, $\forall m \in M$.
Proof: The proof will proceed by induction over
$r(m) := |M \setminus MmM|$
For the base case, $r(m) = 0 \implies M = MmM$. Thus $\exists u, v \in M$ s.t. $umv = 1$. Multiplying by 1, $(um)1(v) = 1$ and $(u)1(mv) = 1$. Using the simplification lemma this gives $u = v = 1$ and thus $m = 1$.

Proof of Schützenberger's Theorem

Let $B := \{a \in A : \varphi(a) = 1\}$. Then $\forall u \in B^*$,
$u \in \varphi^{-1}(1) \implies B^* \subseteq \varphi^{-1}(1)$.
$u \in \varphi^{-1}(1)$ implies that $\varphi(b) = 1$ for each letter b in u, by using the simplification lemma.
Thus $\varphi^{-1}(1) \subseteq B^* \implies \varphi^{-1}(1) = B^*$, which is star-free since $B^* = \left(\sum_{a \in A \setminus B} A^* a A^*\right)^c$.

Proof of Schützenberger's Theorem

Now for the induction step, let $r(m) > 0$ and let $\phi^{-1}(s)$ be star-free for all s with $r(s) < r(m)$.

Claim 2

$$\varphi^{-1}(m) = L, \text{ where}$$

$$L = (UA^* \cap A^*V) \setminus (A^*CA^* \cup A^*WA^*), \text{ with}$$

$$U := \sum_{(n,a)\in E} \varphi^{-1}(n)a,$$

$$E := \{(n,a) \in M \times A : n\varphi(a)\mathcal{R}m \wedge n \notin mM\}$$

Proof of Schützenberger's Theorem

$$V := \sum_{(a,n)\in F} a\varphi^{-1}(n),$$

$$F := \{(a,n) \in A \times M : \varphi(a)n\mathcal{L}m \wedge n \notin Mm\}$$

$$C := \{a \in A : m \notin M\varphi(a)M\}$$

$$W := \sum_{(a,n,b)\in G} a\varphi^{-1}(n)b,$$

$$G := \{(a,n,b) \in A \times M \times A :$$

$$m \in (M\varphi(a)nM \cap Mn\varphi(b)M) \setminus M\varphi(a)n\varphi(b)M\}$$

Proof of Schützenberger's Theorem

Proof of Claim 2: Given any $u \in \varphi^{-1}(m)$, let p be the shortest prefix of u s.t. $\varphi(p)\mathcal{R}m$. We can assume $p \neq \epsilon$, as this would imply that $m\mathcal{R}1 \implies m = 1$ (By the simplification lemma), and we would be done.
So we can write $p = ra$, with $r \in A^*$ and $a \in A$. Let $n = \varphi(r)$. By construction, $(n, a) \in E$, explained next.
$n\varphi(a) = \varphi(r)\varphi(a) = \varphi(p)$, and $\varphi(p)\mathcal{R}m$.
Also since $m \leq_R \varphi(p)$ and $\varphi(p) = n\varphi(a) \leq_R n$, $n \notin mM$ as otherwise this would imply that $n\mathcal{R}m$, a contradiction.

$$(\text{Recall, } E := \{(n, a) \in M \times A : n\varphi(a)\mathcal{R}m \wedge n \notin mM\})$$

Proof of Schützenberger's Theorem

Thus $p \in \varphi^{-1}(n)a$ and $u \in UA^*$
Similarly it can be proven that $u \in A^*V$.
If $u \in A^*CA^*, \exists a \in C$ s.t. $m = \varphi(u) \in M\varphi(a)M$, a contradiction since $a \in C$.

$$\text{(Recall, } C := \{a \in A : m \notin M\varphi(a)M\})$$

Proof of Schützenberger's Theorem

If $u \in A^* W A^*, \exists (a, n, b) \in G$ s.t. $m \in M\varphi(a)n\varphi(b)M$, a contradiction since $(a, n, b) \in G$.

$$\text{(Recall, } G := \{(a, n, b) \in A \times M \times A :$$

$$m \in (M\varphi(a)nM \cap Mn\varphi(b)M) \setminus M\varphi(a)n\varphi(b)M\})$$

Thus by definition of L, $u \in L \implies \varphi^{-1}(m) \subseteq L$

Proof of Schützenberger's Theorem

To prove the other direction, given any $u \in L$, we must show that $u \in \varphi^{-1}(m)$. Let $s = \varphi(u)$
Since $u \in UA^*$, we have $u \in \varphi^{-1}(n)aA^*$, for some $(n, a) \in E$.
Thus $s = \varphi(u) \in n\varphi(a)M$.
Also $(n, a) \in E$, $n\varphi(a)M = mM$ implying $s \in mM$.

$$(\text{Recall, } E := \{(n, a) \in M \times A : n\varphi(a)\mathcal{R}m \wedge n \notin mM\})$$

Proof of Schützenberger's Theorem

A similar argument using $u \in A^*V$ can be used to see that $s \in Mm$.
Thus, via the Lemma 21 about Aperiodic Monoids, $s \notin J_M$ would give that $s = m$. So it suffices to prove that $m \in MsM$.
Assume otherwise, that $m \notin MsM$. Then there exists a factor of minimal length of u, f s.t. $f \neq \epsilon$ and $m \notin M\varphi(f)M$.

Proof of Schützenberger's Theorem

If $f \in A$ then $f \in C$ and thus $u \in A^* C A^*$, which is is a contradiction.

$$(\text{Recall, } C := \{a \in A : m \notin M\varphi(a)M\})$$

Proof of Schützenberger's Theorem

So f contains more than one alphabet. Let $f = agb$, with $a, b \in A$ and $g \in A^*$. Now let $n = \varphi(g)$.
By the minimal length definition of f, this implies that
$m \in M\varphi(a)nM$ and $m \in Mn\varphi(b)M$.
But this would imply that $(a, n, b) \in G$ and thus $f \in W$, which is a contradiction.

$$\text{(Recall, } G := \{(a, n, b) \in A \times M \times A :$$

$$m \in (M\varphi(a)nM \cap Mn\varphi(b)M) \setminus M\varphi(a)n\varphi(b)M\}$$

$$\text{and } W := \sum_{(a,n,b)\in G} a\varphi^{-1}(n)b)$$

Proof of Schützenberger's Theorem

This proves Claim 2.

Now it is enough to show that all the languages involved in the definition of L are star-free.

A^*CA^* is star-free by definition.

Let $(n, a) \in E$. Then $n\varphi(a)M = mM$, $MmM \subseteq MnM$, and hence $r(n) \leq r(m)$.

If $r(n) = r(m)$, then $MmM = MnM$. Since $m \leq_R n$, this along with Theorem 1 would imply that $n\mathcal{R}m$, which is a contradiction since by definition $n \not mM$.

Thus $r(n) < r(m)$ and U is star free, by Claim 2.

Proof of Schützenberger's Theorem

This proves Claim 2.

Now it is enough to show that all the languages involved in the definition of L are star-free.

A^*CA^* is star-free by definition.

Let $(n, a) \in E$. Then $n\varphi(a)M = mM$, $MmM \subseteq MnM$, and hence $r(n) \leq r(m)$.

If $r(n) = r(m)$, then $MmM = MnM$. Since $m \leq_R n$, this along with Theorem 4 would imply that $n\mathcal{R}m$, which is a contradiction since by definition $n \notin mM$.

Thus $r(n) < r(m)$ and U is star free, by the induction step.

Proof of Schützenberger's Theorem

We can similarly argue that V is star-free.
Now let $(a, n, b) \in G$.

$$(\text{Recall, } G := \{(a, n, b) \in A \times M \times A :$$
$$m \in (M\varphi(a)nM \cap Mn\varphi(b)M) \setminus M\varphi(a)n\varphi(b)M\})$$

Proof of Schützenberger's Theorem

Then $r(n) \leq r(m)$ since $m \in MnM$.So $MmM \subseteq MnM$. Suppose $MmM = MnM$. Then $n \in MmM$. We also have $m \in M\varphi(a)nM$ and $m \in Mn\varphi(b)M$.
Thus $n \in M\varphi(a)nM$ and $n \in Mn\varphi(b)M$ as well. It follows that $n\mathcal{L}\varphi(a)n$ and $n\mathcal{R}n\varphi(b)$.

Proof of Schützenberger's Theorem

So we have $n\varphi(b)\mathcal{L}\varphi(a)n\varphi(b)$, $\varphi(a)n\mathcal{R}\varphi(a)n\varphi(b)$ and $m \in (M\varphi(a)nM \cap Mn\varphi(b)M)$ which by Green's Lemma gives $m \in M\varphi(a)n\varphi(b)$M, which is a contradiction as $(a, n, b) \in G$.

Proof of Schützenberger's Theorem

Thus $r(n) < r(m)$ and it follows that the W is a star-free language by the induction step.
This completes the proof of the theorem.

[Sch65] M.P. Schützenberger. "On finite monoids having only trivial subgroups". In: *Information and Control* 8.2 (1965), pp. 190–194. ISSN: 0019-9958. DOI: https://doi.org/10.1016/S0019-9958(65)90108-7. URL: https://www.sciencedirect.com/science/article/pii/S0019995865901087.

[Kum] K Narayan Kumar. *Lecture 5: Schutzenberger's Theorem*. https://www.cmi.ac.in/~kumar/words/lecture05.pdf.

[] *Schützenberger's Aperiodic Monoid Characterization of Star-Free Languages*. https://www.csa.iisc.ac.in/~deepakd/atc-2019/Ankur-Naskar_ATC_seminar_aperiodic_monoids_star_free_languages.pdf.