On the Generating Series of Languages

Null

April 15, 2025

Σ will always denote a finite set called the *alphabet*, and Σ^* will denote the free monoid on Σ, as well as its underlying set.

Definition 0.1 Given $\mathcal{L} \subseteq \Sigma^*$ a language, the generating series $\mathbb{S}(\mathcal{L})$ of $\mathcal{L}$ is defined by:

$$\mathbb{S}(\mathcal{L})(x) = \sum_{n \geq 0} |\mathcal{L} \cap \Sigma^n| \, x^n \tag{1}$$

It always converges on the open disk $\mathcal{D}\left(0, \frac{1}{|Sigma|}\right)$.

Proof. To show the convergence it suffices to see that the sequence $\frac{1}{\Sigma} \times |\mathcal{L} \cap \Sigma^n|$ is bounded in n and $\mathcal{L}$ by $\frac{|\Sigma^n|}{|\Sigma^n|} = 1$ and conclude by Abel's Lemma. ∎

We will now prove the following theorem:

Theorem 0.1 If $\mathcal{L}$ is a rational language, its generating series $\mathbb{S}(\mathcal{L})$ is a real rational function.

Proof by Induction. The generating series of a word of length $n \geq 0$ in Σ^* is $x \mapsto x^n$. Assume that $\mathcal{L}$ is a rational language:

- If we can write $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$ where $\mathcal{L}_1$ and $\mathcal{L}_2$ are rational and disjoint[1]. Then, for all $n \geq 0$:

  $$|\mathcal{L} \cap \Sigma^n| = |\mathcal{L}_1 \cap \Sigma^n| + |\mathcal{L}_2 \cap \Sigma^n|$$

  which means that:

  $$\mathbb{S}(\mathcal{L}) = \mathbb{S}(\mathcal{L}_1) + \mathbb{S}(\mathcal{L}_2)$$

  and that $\mathbb{S}(\mathcal{L})$ is a real rational function.

- If we can write $\mathcal{L} = \mathcal{L}_1 \cdot \mathcal{L}_2$ where $\cdot$ is the monoidal operation on Σ^* then, for $n \geq 0$, words of length n in $\mathcal{L}$ are made of a word of length k in $\mathcal{L}_1$ and a word of length $n-k$ in $\mathcal{L}_2$ for $k \leq n$ and thus:

  $$|\mathcal{L} \cap \Sigma^n| = \sum_{k=0}^{n} \left|\mathcal{L}_1 \cap \Sigma^k\right| \times \left|\mathcal{L}_2 \cap \Sigma^{n-k}\right|$$

  Replacing this in the definition of $\mathbb{S}(\mathcal{L})$:

  $$\mathbb{S}(\mathcal{L})(x) = \sum_{n \geq 0} \sum_{k=0}^{n} \left|\mathcal{L}_1 \cap \Sigma^k\right| \times \left|\mathcal{L}_2 \cap \Sigma^{n-k}\right| x^n = \sum_{k=0}^{n} \left|\mathcal{L}_1 \cap \Sigma^k\right| \times \left|\mathcal{L}_2 \cap \Sigma^{n-k}\right| x^k x^{n-k} = \mathbb{S}(\mathcal{L}_1)(x) \, \mathbb{S}(\mathcal{L}_2)(x)$$

  As a product of real rational functions, $\mathbb{S}(\mathcal{L})$ is a real rational function.

[1] The case where their intersection is not null is solved later with the complement case

- If $\mathcal{L} = \mathcal{L}'^{\complement}$ taken in Σ^*, clearly:

$$\mathbb{S}\left(\mathcal{L}\right)(x) = \sum_{n \geq 0} \left(\left|Sigma\right|^n - \left|\mathcal{L}' \cap \Sigma^n\right|\right) x^n = \frac{1}{1 - \left|\Sigma\right| x} - \mathbb{S}\left(\mathcal{L}'\right)(x)$$

  As a difference of real rational functions, $\mathbb{S}\left(\mathcal{L}\right)$ is a real rational function.

- If $\mathcal{L} = \mathcal{L}'^*$ is the Kleene star of a rational language, since a word in $\mathcal{L}$ is a sequence of words on $\mathcal{L}$, $\mathcal{L} = \bigcup_{n \geq 0} \mathcal{L}'^n$, from the previous points on union and concatenation, we get that:

$$\mathbb{S}\left(\mathcal{L}\right) = \sum_{n \geq 0} \mathbb{S}\left(\mathcal{L}'^n\right) = \sum_{n \geq 0} \mathbb{S}\left(\mathcal{L}'\right)^n$$

  and thus:

$$\mathbb{S}\left(\mathcal{L}\right) = \frac{1}{1 - \mathbb{S}\left(\mathcal{L}'\right)}$$

  which is a real rational function.

Thus, by induction, all rational languages have real rational generating series. ■

Proof by Automatons. Given a rational language $\mathcal{L}$, we take an automaton $\mathcal{M}$ that recognizes it. We know that $\left|\mathcal{L} \cap \Sigma^n\right|$ is the number of walks of length n on $\mathcal{M}$ from an initial state to a terminal state. If we set e to be the standard unit row vector, c to be the characteristic vector of the terminal states and A to be the adjacency matrix of the automaton, then:

$$\mathbb{S}\left(\mathcal{L}\right)(x) = \sum_{n \geq 0} eA^n c x^n = e\left[\sum_{n \geq 0} A^n x^n\right] c = e\left(\mathrm{Id} - xA\right)^{-1} c$$

This proves that $\mathbb{S}\left(\mathcal{L}\right)$ is a real rational function (since $\det\left(I - xA\right)^{-1}$ is a polynomial in X). ■