Metric Spaces for Regular Languages

Anton Xue

1 Introduction

Filler introduction text

2 Background

Filler text for background

2.1 Regular Languages and Finite Automata

A regular language is a language that can be recongnized by a regular expression. Such languages play a central role in theoretical comptuer science and formal language theory due to their ability to simply describe a large class of strings. We now formalize these terms.

Given a finite set of unique symbols Σ called the alphabet, write the set of all finite strings constructed over this alphabet as $\Sigma^\star$. A string is nothing more than a finite sequences of symbols from an alphabet. Let ε denote the empty string which contains no symbols.

Example 1. *Consider an alphabet $\Sigma = \{a, \alpha, b, \beta, \div\}$, examples of strings (finite sequences) that can be constructed with this alphabet include*

$$ab\beta\beta\alpha \qquad \div\div a\beta \qquad aaaaaaa \qquad \varepsilon$$

A language L is nothing more than a set of strings. In other words, $L \subseteq \Sigma^\star$. We are now ready to introduce the concept of a regular language. Formally, regular languages are a family of languages inductively defined as follows:

(1) The empty set $\emptyset$ and the empty string language $\{\varepsilon\}$ are regular languages.

(2) For each symbol $a \in \Sigma$, the singleton language $\{a\}$ is a regular language.

(3) If A and B are regular languages, then so is their union $A \cup B$, their concatenation $A \cdot B$, and their Kleene star $A^\star$, defined as:

$$\begin{aligned} A \cup B &= \{w \ : \ w \in A \cup B\} \\ A \cdot B &= \{w_a \cdot w_b \ : \ w_a \in A, w_b \in B\} \\ A^\star &= \left\{w^k \ : \ k \in \mathbb{N}, w \in A\right\} \end{aligned}$$

Where w^k is the k-fold concatenation of a string to itself, and $w_a \cdot w_b$ is the concatenaiton of strings w_a and w_b. Sometimes we write $w_a w_b$ for concatenation when context is clear.

(4) No other languages are regular.

2.2 Metric Spaces

Filler text on metric spaces

2.3 Measure Theory

Filler text on measure theory

3 Measure Theoretic Approaches

Consider a language $L \subseteq \Sigma^\star$. The n-splice of a language written as L^n is defined as:

$$L^n = L \cap \Sigma^n$$

We then have the following relations:

$$L = \bigcup_{n=0}^{\infty} L^n = \bigcup_{n=0}^{\infty} (L \cap \Sigma^n) = L \cap \bigcup_{n=0}^{\infty} \Sigma^n = L \cap \Sigma^\star = L$$

Suppose that $\left(\mathbb{N}, 2^{\mathbb{N}}, \eta\right)$ is a probabiliy measure space on $\mathbb{N}$ with the probability measure η, one way to define a measure λ_η is as follows:

$$\lambda_\eta (L) = \sum_{n=0}^{\infty} \frac{|L^n|}{|\Sigma^n|} \eta (n)$$

Theorem 1. *$\left(\Sigma^\star, 2^{\Sigma^\star}, \lambda_\eta\right)$ is a measure space.*

Proof. Since the $2^{\Sigma^\star}$ is the largest σ-algebra on $\Sigma^\star$, it suffices to show that λ_η is a measure.

To see that $\emptyset$ is mapped to 0:

$$\lambda_\eta(\emptyset) = \sum_{n=0}^{\infty} \frac{|\emptyset|}{|\Sigma^n|} \eta(n) = \sum_{n=0}^{\infty} 0 = 0$$

Now take $(A_n) \subseteq \Sigma^\star$ to be a countable collection of disjoint sets. Write A_n^k to denote the k splice of the nth set. In other words:

$$A_n = \bigcup_{k=0}^{\infty} A_n^k$$

Observe that all such A_n^k are pairwise disjoint by construction, and so:

$$\lambda_\eta\left(\bigcup_{n=0}^{\infty} A_n\right) = \lambda_\eta\left(\bigcup_{n=0}^{\infty}\bigcup_{k=0}^{\infty} A_n^k\right) = \sum_{n=0}^{\infty} \lambda_\eta\left(\bigcup_{k=0}^{\infty} A_n^k\right) = \sum_{n=0}^{\infty}\sum_{k=0}^{\infty} \frac{|A_n^k|}{|\Sigma^n|}\eta(k) = \sum_{n=0}^{\infty} \lambda_\eta(n)$$

We conclude that $\left(\Sigma^\star, 2^{\Sigma^\star}, \lambda_\eta\right)$ forms a measure space.

$\square$

We can generalize this more. Suppose that $\nu = (\nu_n)$ is a countable collection of measures where each ν_n is defined on the splice Σ^n. Then we can extend a definition of $\lambda_{\eta,\nu}$ as:

$$\lambda_{\eta,\nu}(A) = \sum_{n=0}^{\infty} \nu(A^n)\,\eta(n)$$

Theorem 2. *$\left(\Sigma^\star, 2^{\Sigma^\star}, \lambda_{\eta,\nu}\right)$ is a measure space.*

Proof. As with before, we only show that $\lambda_{\eta,\nu}$ is a measure.

For $\emptyset$ we have again:

$$\lambda_{\eta,\nu}(\emptyset) = \sum_{n=0}^{\infty} 0 = 0$$

Again take $(A_n) \subseteq \Sigma^\star$ to be a countable disjoint collection of sets, and A_n^k to be the k splice of A_n. Then:

$$\lambda_{\eta,\nu}\left(\bigcup_{n=0}^{\infty} A_n\right) = \lambda_{\eta,\nu}\left(\bigcup_{n=0}^{\infty}\bigcup_{k=0}^{\infty} A_n^k\right) = \sum_{n=0}^{\infty} \lambda_{\eta,\nu}\left(\bigcup_{k=0}^{\infty} A_n^k\right) = \sum_{n=0}^{\infty}\sum_{k=0}^{\infty} \nu_k\left(A_n^k\right)\eta(k) = \sum_{n=0}^{\infty} \lambda_{\eta,\nu}(A_n)$$

This shows that $\left(\Sigma^\star, 2^{\Sigma^\star}, \lambda_{\eta,\nu}\right)$ is a measure space. $\square$

4 Linear Operators

5 Separating Automata