Natural Language Processing - Assignment 3

Detailed Conceptual Explanations

Mohammad Taha Majlesi - 810101504

ETH Zürich, Fall 2022

Abstract

This document provides a detailed, step-by-step explanation of the concepts and solutions presented in Assignment 3. It covers the abstract algebraic structure of semirings, the definition and properties of the Kleene star operator, and its application to different domains, including formal language theory and graph algorithms for pathfinding problems. Each section is designed to build a strong conceptual understanding of the underlying theory.

Contents

1 The Kleene Star in General Semirings

1.1 Introduction to Semirings

Before diving into the questions, it's crucial to understand the algebraic structure known as a **semiring**. A semiring is a set of elements, S, equipped with two binary operations, which we can think of as a generalized "addition" ($\oplus$) and a generalized "multiplication" ($\otimes$).

A structure $\langle S, \oplus, \otimes, \mathbf{0}, \mathbf{1} \rangle$ is a semiring if it satisfies the following properties:

1. $\langle S, \oplus, \mathbf{0} \rangle$ is a **commutative monoid**. This means:
   - **Associativity of** $\oplus$: $a \oplus (b \oplus c) = (a \oplus b) \oplus c$
   - **Commutativity of** $\oplus$: $a \oplus b = b \oplus a$
   - **Identity for** $\oplus$: There exists an element $\mathbf{0}$ such that $a \oplus \mathbf{0} = a$.
2. $\langle S, \otimes, \mathbf{1} \rangle$ is a **monoid**. This means:
   - **Associativity of** $\otimes$: $a \otimes (b \otimes c) = (a \otimes b) \otimes c$
   - **Identity for** $\otimes$: There exists an element $\mathbf{1}$ such that $a \otimes \mathbf{1} = \mathbf{1} \otimes a = a$.
3. **Distributivity**: $\otimes$ distributes over $\oplus$:
   - $a \otimes (b \oplus c) = (a \otimes b) \oplus (a \otimes c)$
   - $(a \oplus b) \otimes c = (a \otimes c) \oplus (b \otimes c)$
4. **Annihilation**: The additive identity $\mathbf{0}$ annihilates under $\otimes$:
   - $a \otimes \mathbf{0} = \mathbf{0} \otimes a = \mathbf{0}$

Semirings are fundamental in computer science because they provide a general framework for problems involving paths, probabilities, and languages.

1.2 Part (a): Recursive Definition of the Kleene Star

1.2.1 Concept: The Kleene Star

The Kleene star of an element a, denoted a^*, represents the infinite sum of all "powers" of a. The powers are formed using the semiring's multiplication ($\otimes$), and the sum is performed using the semiring's addition ($\oplus$).

1.2.2 Formal Derivation

The Kleene star is defined as the infinite sum:

$$a^* = \bigoplus_{n=0}^{\infty} a^{\otimes n} \tag{1}$$

where $a^{\otimes n}$ means $a \otimes a \otimes \cdots \otimes a$ (n times). By convention, $a^{\otimes 0} = \mathbf{1}$ (the multiplicative identity) and $a^{\otimes 1} = a$.

To derive the recursive form, we can separate the $n = 0$ term from the rest of the sum:

$$\begin{aligned} a^* &= a^{\otimes 0} \oplus \left(\bigoplus_{n=1}^{\infty} a^{\otimes n} \right) \\ &= \mathbf{1} \oplus \left(a^{\otimes 1} \oplus a^{\otimes 2} \oplus a^{\otimes 3} \oplus \dots \right) \end{aligned}$$

Now, we can use the distributive property of the semiring to factor out a single a from the terms in the parenthesis:

$$\begin{aligned} a^* &= \mathbf{1} \oplus \left(a \otimes a^{\otimes 0} \oplus a \otimes a^{\otimes 1} \oplus a \otimes a^{\otimes 2} \oplus \dots \right) \\ &= \mathbf{1} \oplus a \otimes \left(a^{\otimes 0} \oplus a^{\otimes 1} \oplus a^{\otimes 2} \oplus \dots \right) \end{aligned}$$

We recognize that the expression in the parenthesis is the original definition of a^*. Substituting this back gives us the fundamental recursive identity:

$$a^* = \mathbf{1} \oplus (a \otimes a^*) \tag{2}$$

This recursive definition is extremely powerful and forms the basis for many dynamic programming algorithms, as it relates the solution of a problem to a smaller version of the same problem.

1.3 Part (b): The Kleene Star in the Log-Semiring

1.3.1 Concept: The Log-Semiring

The Log-semiring is used to compute sums of products of small probabilities without running into numerical underflow. By working in the log domain, multiplication becomes addition, and addition becomes a more complex 'log-sum-exp' operation.

The Log-semiring is defined as:

- **Set**: Real numbers $\mathbb{R}$.
- **Addition ($\oplus$)**: $x \oplus y = \log(e^x + e^y)$.
- **Multiplication ($\otimes$)**: $x \otimes y = x + y$.
- **Additive Identity ($\mathbf{0}$)**: $-\infty$.
- **Multiplicative Identity ($\mathbf{1}$)**: 0.

1.3.2 Derivation

We apply the definition $a^* = \bigoplus_{n=0}^{\infty} a^{\otimes n}$ to this specific semiring. First, let's compute the "powers" of a using the multiplication operator, which is standard addition:

- $a^{\otimes 0} = \mathbf{1} = 0$
- $a^{\otimes 1} = a$
- $a^{\otimes 2} = a \otimes a = a + a = 2a$
- $a^{\otimes n} = n \cdot a$

Next, we sum these terms using the semiring's addition operator ($\log(e^x + e^y)$):

$$\begin{aligned} a^* &= \bigoplus_{n=0}^{\infty}(n \cdot a) \\ &= (0 \cdot a) \oplus (1 \cdot a) \oplus (2 \cdot a) \oplus \dots \\ &= \log(e^0 + e^a + e^{2a} + e^{3a} + \dots) \\ &= \log\left(\sum_{n=0}^{\infty}(e^a)^n\right) \end{aligned}$$

The sum inside the logarithm is an infinite **geometric series** with first term 1 and common ratio $r = e^a$. This series converges only if the absolute value of the common ratio is less than 1, i.e., $|e^a| < 1$. Since e^a is always positive, this condition simplifies to $e^a < 1$, which is true if and only if $a < 0$.

When it converges, the sum of a geometric series is $\frac{1}{1-r}$. Therefore:

$$a^* = \log\left(\frac{1}{1 - e^a}\right), \quad \text{for } a < 0 \tag{3}$$

1.4 Part (d): The Language Semiring

1.4.1 Concept: Formal Languages as a Semiring

This part demonstrates that the set of all formal languages over an alphabet Σ forms a semiring. This is a profound result because it means that operations on languages (like union and concatenation) follow the same abstract rules as operations in other domains (like addition and multiplication), allowing us to apply general algorithms to them.

The language semiring is defined as:

- **Set**: 2^{Σ^*}, the power set of Σ^*. An element of this set is a language.
- **Addition ($\oplus$)**: Set Union ($\cup$).
- **Multiplication ($\otimes$)**: Language Concatenation, $A \otimes B = \{ab \mid a \in A, b \in B\}$.
- **Additive Identity (0)**: The empty set, $\emptyset$ or $\{\}$.
- **Multiplicative Identity (1)**: The set containing only the empty string, $\{\varepsilon\}$.

1.4.2 Proof of Semiring Properties

We must verify the four main axioms.

1. $\langle 2^{\Sigma^*}, \cup, \{\}\rangle$ **is a commutative monoid**:
   - **Identity**: $A \cup \{\} = A$. This holds.
   - **Associativity**: $(A \cup B) \cup C = A \cup (B \cup C)$. This is a standard property of set union.
   - **Commutativity**: $A \cup B = B \cup A$. This is also a standard property of set union.
2. $\langle 2^{\Sigma^*}, \otimes, \{\varepsilon\}\rangle$ **is a monoid**:

- **Identity**: $A \otimes \{\varepsilon\} = \{a\varepsilon \mid a \in A\} = A$. This holds.
- **Associativity**: String concatenation is associative, so $(ab)c = a(bc)$. This property extends to languages: $(A \otimes B) \otimes C = A \otimes (B \otimes C)$.

3. $\otimes$ **distributes over** $\cup$: We must show $A \otimes (B \cup C) = (A \otimes B) \cup (A \otimes C)$.

$$\begin{aligned} A \otimes (B \cup C) &= \{ad \mid a \in A, d \in (B \cup C)\} \\ &= \{ad \mid a \in A \text{ and } (d \in B \text{ or } d \in C)\} \\ &= \{ad \mid a \in A, d \in B\} \cup \{ad \mid a \in A, d \in C\} \\ &= (A \otimes B) \cup (A \otimes C) \end{aligned}$$

The distributive property holds.

4. $\{\}$ **is the annihilator for** $\otimes$:

$$A \otimes \{\} = \{ab \mid a \in A, b \in \{\}\}$$

Since there are no elements b in the empty set, the resulting set of concatenations must also be empty. Thus, $A \otimes \{\} = \{\}$.

Since all conditions are met, the structure is a valid semiring.

1.4.3 Kleene Star in the Language Semiring

Applying the general definition of the Kleene star to this semiring yields:

$$\begin{aligned} A^* &= \bigoplus_{n=0}^{\infty} A^{\otimes n} \\ &= A^{\otimes 0} \cup A^{\otimes 1} \cup A^{\otimes 2} \cup \dots \\ &= \{\varepsilon\} \cup A \cup (A \otimes A) \cup (A \otimes A \otimes A) \cup \dots \end{aligned}$$

This is exactly the definition of the **Kleene closure** in formal language theory: the set of all strings formed by concatenating zero or more strings from language A.

2 Semirings in Graph Algorithms

2.1 Introduction: A General Framework for Path Problems

Many classic graph algorithms, like Dijkstra's for shortest paths or Floyd-Warshall for all-pairs shortest paths, can be generalized using the framework of semirings. The choice of semiring determines the problem being solved.

- A path's "cost" is computed by $\otimes$-multiplying the edge weights along it.
- The "best" path among several alternatives is found by $\oplus$-adding their costs.

2.2 Part (a): Tropical and Arctic Semirings

2.2.1 Concept: Two Key Semirings for Pathfinding

- **Tropical Semiring (Min-Plus):** Used for standard shortest-path problems.
  - **Set**: $\mathbb{R}_{\geq 0} \cup \{\infty\}$
  - $\oplus = \min$
  - $\otimes = +$
  - $\mathbf{0} = \infty$ (identity for min)
  - $\mathbf{1} = 0$ (identity for $+$)
- **Arctic Semiring (Max-Plus):** Used for finding paths with maximum capacity or reliability (often called "widest path" problems).
  - **Set**: $\mathbb{R} \cup \{-\infty, \infty\}$
  - $\oplus = \max$
  - $\otimes = +$
  - $\mathbf{0} = -\infty$ (identity for max)
  - $\mathbf{1} = 0$ (identity for $+$)

A semiring is **0-closed** if a^* is well-defined for all a. For these semirings, this is true. For any $a \geq 0$ in the Tropical semiring, $a^* = \min(0, a, 2a, \dots) = 0$.

2.3 Part (b): Inductive Proof for Path Sums

2.3.1 Concept: Matrix Powers and Path Lengths

If M is the adjacency matrix of a graph, where M_{ij} is the weight of the edge $i \to j$, then the entry $(M^n)_{ik}$ in the n-th matrix power of M represents the total "cost" of all paths of length exactly n from node i to node k.

2.3.2 Proof by Induction

Statement $P(n)$**:** The matrix entry $(M^n)_{ik}$ contains the semiring-sum of weights of all paths of length exactly n from node i to node k.

- **Base Case (**$n = 1$**):** $P(1)$ is true by definition. $(M^1)_{ik} = M_{ik}$, which is the weight of the direct edge (path of length 1) from i to k.
- **Inductive Hypothesis:** Assume $P(n)$ is true for some $n \geq 1$.
- **Inductive Step (Show** $P(n+1)$ **is true):** By the definition of semiring matrix multiplication:

$$(M^{n+1})_{ik} = (M \otimes M^n)_{ik} = \bigoplus_{j=1}^{N} (M_{ij} \otimes (M^n)_{jk}) \tag{4}$$

  Let's analyze this expression:

  - A path of length $n+1$ from i to k must consist of a first step from i to some intermediate node j, followed by a path of length n from j to k.
  - The term M_{ij} represents the cost of the first step.
  - By the inductive hypothesis, $(M^n)_{jk}$ represents the sum of costs for all paths of length n from j to k.
  - The product $M_{ij} \otimes (M^n)_{jk}$ calculates the total cost of all paths of length $n+1$ that go through j as their first intermediate stop.
  - The sum $\bigoplus_{j=1}^{N}$ aggregates these costs over all possible first intermediate stops j, thus correctly computing the total cost of all paths of length $n+1$ from i to k.

Thus, $P(n+1)$ is true, completing the proof.

2.4 Parts (c-d): All-Paths Sum and the Kleene Star

2.4.1 Concept: The All-Paths Problem

The goal is now to find the sum of weights of **all paths of any length** from i to j. This can be expressed as the infinite sum of the costs of paths of length $0, 1, 2, \ldots$.

2.4.2 Derivation

Let Z be the matrix of all-paths sums. Then:

$$Z_{ij} = \bigoplus_{n=0}^{\infty} (M^n)_{ij} \tag{5}$$

This is precisely the definition of the Kleene star of the matrix M. Therefore, $Z = M^*$.

In many practical semirings (like Tropical with non-negative weights), paths with cycles are not optimal. A simple path in a graph with N nodes can have at most length $N-1$. Any longer path must contain a cycle, which does not improve the result. Therefore, the infinite sum can be truncated:

$$M^* = \bigoplus_{n=0}^{N-1} M^n \tag{6}$$

2.5 Part (e): A Simple Algorithm and its Complexity

A direct algorithm to compute $M^* = \bigoplus_{n=0}^{N-1} M^n$:

1. Initialize result matrix $M_{star} = I$ (where $I = M^0$).
2. Initialize power matrix $M_{power} = I$.
3. For $n = 1, \ldots, N-1$:
   - Update power: $M_{power} \leftarrow M_{power} \otimes M$. (Computes M^n)
   - Update sum: $M_{star} \leftarrow M_{star} \oplus M_{power}$.
4. Return M_{star}.

Complexity Analysis:

- The loop runs $O(N)$ times.
- Inside the loop, the dominant operation is matrix multiplication ($M_{power} \otimes M$), which takes $O(N^3)$ for $N \times N$ matrices. Matrix addition takes $O(N^2)$.
- Total complexity $= O(N) \times O(N^3) = O(N^4)$.

2.6 Parts (g-h): A Faster Algorithm via Repeated Squaring

2.6.1 Concept: Exploiting Idempotency

A semiring is **idempotent** if $a \oplus a = a$ for all a. The Tropical and Arctic semirings are both idempotent (since $\min(a,a) = a$ and $\max(a,a) = a$). For such semirings, a powerful identity holds:

$$\bigoplus_{n=0}^{K} M^n = (I \oplus M)^K \tag{7}$$

This transforms the problem from a sum of matrix powers into a single matrix power. So, we only need to compute $M^* = (I \oplus M)^{N-1}$.

2.6.2 Algorithm: Binary Exponentiation (Repeated Squaring)

We can compute A^k in $O(\log k)$ multiplications instead of $O(k)$. The idea is to build up powers of 2. For example, $A^{13} = A^{8+4+1} = A^8 \otimes A^4 \otimes A^1$. We can get A^2, A^4, A^8 by repeatedly squaring: $A^2 = A \otimes A$, $A^4 = A^2 \otimes A^2$, etc.

Faster Algorithm:

1. Compute $B = I \oplus M$. This takes $O(N^2)$ time.
2. Compute B^{N-1} using binary exponentiation.

Complexity Analysis:

- Binary exponentiation requires $O(\log N)$ matrix multiplications.
- Each matrix multiplication costs $O(N^3)$.
- Total complexity $= O(\log N) \times O(N^3) = O(N^3 \log N)$.

This is a significant improvement over the $O(N^4)$ naive algorithm.

2.7 Parts (i-k): Error Bound for Kleene Star Approximation

2.7.1 Concept: Error Analysis in Standard Algebra

Here we switch from abstract semirings back to standard matrix algebra over real numbers. The Kleene star is the matrix inverse $A^* = (I - A)^{-1} = \sum_{n=0}^{\infty} A^n$. This series converges if the spectral radius of A is less than 1. We want to bound the error when we approximate this infinite sum by a finite one.

2.7.2 Derivation of the Error Bound

The error is the norm of the truncated tail of the series:

$$\text{Error} = \left\| A^* - \sum_{n=0}^{K} A^n \right\|_2 = \left\| \sum_{n=K+1}^{\infty} A^n \right\|_2 \tag{8}$$

Using properties of matrix norms (triangle inequality and sub-multiplicativity):

$$\begin{aligned} \left\| \sum_{n=K+1}^{\infty} A^n \right\|_2 &\leq \sum_{n=K+1}^{\infty} \|A^n\|_2 \\ &\leq \sum_{n=K+1}^{\infty} \|A\|_2^n \end{aligned}$$

The operator 2-norm, $\|A\|_2$, is equal to the largest singular value of A, denoted $\sigma_{\max}(A)$. Let $\sigma = \sigma_{\max}(A)$. The condition for convergence is $\sigma < 1$. The error is bounded by a geometric series:

$$\text{Error} \leq \sum_{n=K+1}^{\infty} \sigma^n = \frac{\sigma^{K+1}}{1 - \sigma} \tag{9}$$

This result is highly significant. It shows that the error is of the order $O(\sigma^K)$. Since $\sigma < 1$, the error decreases **exponentially** as the number of terms K increases. This means that truncating the series is a very effective approximation method.

3 Question 3: Colab Notebook

The work for this question involves practical implementation and experimentation, which is presented in codes folder.