Universal Rewriting Rules for the Parikh Matrix Injectivity Problem

Week 2 Summer

Yonsei CS Theory Student Group

Basic Notation

- Number of occurrences of u in w as subsequence is denoted as $|w|_u$.
- E.g. $|abbabc|_{ab} = 4$
- E.g. $|aab|_{ab} = 2$
- E.g. $|abbbc|_{abc} = 3$

- Infix, substring: v is infix of x if $x = uvw$ for some $u, w \in \Sigma^*$.

Parikh Vector

- $\Psi: \Sigma^* \to \mathbb{N}^{|\Sigma|}$
- For an ordered alphabet $\Sigma = \{a_1, a_2, \cdots, a_n\}$ and word w, Parikh vector of w is defined as follows:

  $\Psi(w) = (|w|_{a_1}, |w|_{a_2}, \cdots, |w|_{a_n})$.
- E.g. $\Sigma \coloneqq \{a < b < c\}$, $\Psi(aaabbbac) = (4,3,1)$

Parikh Matrix

- Given ordered alphabet Σ, Parikh matrix mapping Ψ_M is a homomorphism from Σ^*(with concatenation) to $\mathbb{N}^{(|\Sigma|+1)\times(|\Sigma|+1)}$ (with matmul).
- E.g. For $\Sigma = \{a < b < c\}$,

$$\Psi_M(a) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \Psi_M(b) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \Psi_M(c) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Parikh Matrix

- E.g. $\Psi_M(abca) = \Psi_M(a)\Psi_M(b)\Psi_M(c)\Psi_M(a)$

$$= \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Parikh Matrix

- **Proposition 1. [Mateescu et al. 2001]**

For an ordered alphabet $\Sigma = \{a_1 < a_2 < \cdots < a_k\}$, and $w \in \Sigma^*$,

$[\Psi_M(w)]_{i,j} = 0$ for $1 \leq j < i \leq k + 1$

$[\Psi_M(w)]_{i,i} = 1$ for $1 \leq i \leq k + 1$

$[\Psi_M(w)]_{i,j} = |w|_{a_i a_{i+1} \cdots a_j}$ for $1 \leq i \leq j \leq k$.

Parikh Matrix

E.g.

$\Sigma = \{a < b\}$, and $w \in \Sigma^*$,

$$\Psi_M(w) = \begin{pmatrix} 1 & |w|_a & |w|_{ab} \\ 0 & 1 & |w|_b \\ 0 & 0 & 1 \end{pmatrix}.$$

Parikh Matrix

E.g.

$\Sigma = \{a < b < c\}$, and $w \in \Sigma^*$,

$$\Psi_M(w) = \begin{pmatrix} 1 & |w|_a & |w|_{ab} & |w|_{abc} \\ 0 & 1 & |w|_b & |w|_{bc} \\ 0 & 0 & 1 & |w|_c \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Parikh Matrix

E.g.

$\Sigma = \{a < b < c < d\}$, and $w \in \Sigma^*$,

$$\Psi_M(w) = \begin{pmatrix} 1 & |w|_a & |w|_{ab} & |w|_{abc} & |w|_{abcd} \\ 0 & 1 & |w|_b & |w|_{bc} & |w|_{bcd} \\ 0 & 0 & 1 & |w|_c & |w|_{cd} \\ 0 & 0 & 0 & 1 & |w|_d \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Parikh Matrix

Proof(sketch).

Proof can be done by induction on length of the word. Compare two:

$$|aw|_{av} = (pick\ av\ in\ w) + (use\ first\ a, pick\ v\ in\ w) = |w|_{av} + |w|_v \cdots (1)$$

$$\Psi_M(aw) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & |w|_a & |w|_{ab} & |w|_{abc} \\ 0 & 1 & |w|_b & |w|_{bc} \\ 0 & 0 & 1 & |w|_c \\ 0 & 0 & 0 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} 1 & |w|_a + 1 & |w|_{ab} + |w|_b & |w|_{abc} + |w|_{bc} \\ 0 & 1 & |w|_b & |w|_{bc} \\ 0 & 0 & 1 & |w|_c \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad \cdots (2)$$

Parikh Matrix

Definition(M-equivalence). Two words w_1 and w_2 are M-equivalent($w_1 \equiv_M w_2$) if and only if $\Psi_M(w_1) = \Psi_M(w_2)$.

Q. Characterize M-equivalence. In other words, come up with a nice condition p, s.t. $w_1 \equiv_M w_2$ if and only if $p(w_1, w_2)$.

M-equivalence

Fact 1. For $\forall x, y \in \Sigma^*$, $w_1 \equiv_M w_2$ if and only if $xw_1y \equiv_M xw_2y$.

Proof(⇒). If $\Psi_M(w_1) = \Psi_M(w_2)$, then

$$\Psi_M(xw_1y) = \Psi_M(x)\Psi_M(w_1)\Psi_M(y) = \Psi_M(x)\Psi_M(w_2)\Psi_M(y) = \Psi_M(xw_2y).$$

*(⇐)*For reverse direction, we should use the fact that Parikh matrix is invertible.

If $\Psi_M(xw_1y) = \Psi_M(xw_2y)$,

then

$[\Psi_M(x)]^{-1}\Psi_M(xw_1y)[\Psi_M(y)]^{-1} = \Psi_M(w_1)$ and

$[\Psi_M(x)]^{-1}\Psi_M(xw_2y)[\Psi_M(y)]^{-1} = \Psi_M(w_2)$. □

M-equivalence

Fact 2. If $w_1 \equiv_M w_2$ and $v_1 \equiv_M v_2$, then $w_1 v_1 \equiv_M w_2 v_2$

(i.e. $\equiv_M$ is a congruence relation w.r.t. concatenation.).

M-equivalence

Q. Characterize M-equivalence. In other words, come up with a nice condition p, s.t. $w_1 \equiv_M w_2$ if and only if $p(w_1, w_2)$.

Definition(Atanasiu07). For binary alphabet $\Sigma = \{a < b\}$, rewriting rule A is defined as follows:

$A.\ \ ab\gamma ba \leftrightarrow ba\gamma ab$ for $\forall\gamma \in \Sigma^*$.

E.g. $abbaaab \leftrightarrow baabaab \leftrightarrow abababa \leftrightarrow \dots$

Rewriting rule is applied to arbitrary infix of the string(Context freely).

Note that $\equiv_M$ is already congruence relation.

M-equivalence

Theorem 1(Atanasiu07). For binary alphabet $\Sigma = \{a < b\}$, $w_1 \equiv_M w_2$ if and only if $w_1 \leftrightarrow_A^* w_2$.

Remark. $\square^*$ means transitive and reflexive closure.

M-equivalence

Theorem 1(Atanasiu07). For binary alphabet $\Sigma = \{a < b\}$, $w_1 \equiv_M w_2$ if and only if $\mathrm{w}_1 \leftrightarrow_A^* w_2$.

A. $ab\gamma ba \leftrightarrow ba\gamma ab$ for $\forall\gamma \in \Sigma^*$.

Proof(⇐). 1. $|ab\gamma ba|_a = |ba\gamma ab|_a$ (Trivial.)

2. $|ab\gamma ba|_b = |ba\gamma ab|_b$ (Trivial.)

3. $|ab\gamma ba|_{ab} = |abba|_{ab} + |ab|_a|\gamma|_b + |\gamma|_{ab} + |\gamma|_a|ba|_b$

$= 2 + |\gamma|_b + |\gamma|_{ab} + |\gamma|_a$.

$|ba\gamma ab|_{ab} = |baab|_{ab} + |ba|_a|\gamma|_b + |\gamma|_{ab} + |\gamma|_a|ab|_b$

$= 2 + |\gamma|_b + |\gamma|_{ab} + |\gamma|_a$. □

M-equivalence

Theorem 2(Atanasiu02). For binary alphabet $\Sigma = \{a < b\}$, $w_1 \equiv_M w_2$ if and only if $\mathrm{w}_1 \leftrightarrow^*_{A'} w_2$.

$A'.\, x \leftrightarrow y$ for $\forall x, y \in \Sigma^*$ where *1.* x and y are palindrome, *2.* x and y have same Parikh vector.

Proof(⇐). 1. $|x|_a = |y|_a$ 2. $|x|_b = |y|_b$ (Same Parikh vector).

3. First, note that $|w|_{ab} + |w|_{ba} = |w|_a |w|_b$ for any $w \in \Sigma^*$.

As x and y are palindrome, $|x|_{ab} = |y|_{ab} = \frac{|x|_a |x|_b}{2}$. □

M-equivalence

Proposition 1(Atanasiu02). For an alphabet $\Sigma = \{a_1 < a_2 < \cdots < a_n\}$, $w_1 \equiv_M w_2$ if $\mathrm{w}_1 \leftrightarrow^* w_2$ with respect to following rules.

A1. $a_i a_{i+1} a_{i+1} a_i \leftrightarrow a_{i+1} a_i a_i a_{i+1}$ for $1 \leq i \leq n-1$.

A2. $a_i a_j \leftrightarrow a_j a_i$ for $|i-j| \geq 2$.

Proof(A2). WLOG let $i > j$. $\left|a_i a_j\right|_{a_i a_{i+1} \cdots a_j} = 0 = \left|a_j a_i\right|_{a_i a_{i+1} \cdots a_j}$.

Similar can be seen for every other subsequence occurrences that appears in Parikh matrix.

M-equivalence

Problem with Atanasiu's rewriting system:

For $\Sigma = \{a < b < c\}$,

$abcba\ bacab \equiv_M bacab\ abcba$. However,

$abcba\ bacab \leftrightarrow abcba\ bcaab$

$\leftrightarrow abcba\ baacb$ and

$bacab\ abcba \leftrightarrow bcaab\ abcba$

$\leftrightarrow baacb\ abcba,$

but $abcba\ bacab \not\leftrightarrow bacab\ abcba$. *Need more powerful representation!*

Exponent-string

String(Finite sequence of symbols)

Notation using exponents

$aabbbcc$

$a^2b^3c^2$

Fact. When denoting words with power notation, only natural numbers can be exponent.

Exponent-string

Exponent-string

(Finite sequence of pairs)

$(a, 7.1), (\mathrm{b}, \sqrt{2}), (c, 0.2)$

Notation of exponent-string

$a^{7.1} b^{\sqrt{2}} c^{0.2}$

Take contrapositive:

If exponents are not integers, we are not notating words.

⇒ We call this discovery an exponent-string!

Exponent-string

Quick remarks about exponent-strings:

1. For semigroup S, S-exponent-strings are allowed to have elements of S as exponents.

1-2. Let $S := (Q^+, +)$ and $\Sigma = \{a, b, c\}$.

Then, we let $a^2 b^{3.5} c^{\frac{7}{3}} = a^{1.3} a^{0.7} b^{3.5} c c c^{\frac{1}{3}} = a^2 b^{2.5} b c^{\frac{7}{3}} = \cdots$ (Same S-exponent-string with different notation.).

1-3. For semigroup $\mathbb{N}' := (N, \times)$ and $\Sigma = \{a, b, c\}$,

$a^8 b^6 c^2 = a^{2\times 4} b^6 c^2 = a^2 a^4 b^6 c^2 = \cdots$.

Exponent-string

Quick remarks about exponent-strings:

2. For semigroup of natural numbers $\mathbb{N} \coloneqq (N, +)$,

monoid of $\mathbb{N}$-exponent-strings is isomorphic with monoid of strings.

2-1. $\mathbb{N}$-exponent-string	Strings
$a^1 b^2 \cdot b^1 c^1 = a^1 b^3 c^1$	$abb \cdot bc = abbbc$

Exponent-string

Quick remarks about exponent-strings:

3. For semigroups S_1 and S_2, if S_1 is a subsemigroup of S_2, then monoid of S_1-exponent-strings is a submonoid of the monoid of S_2-exponent-strings.

Let $\mathbb{Q}^+ \coloneqq (Q^+, +)$ and $\mathbb{R}^+ \coloneqq (R^+, +)$.

4. From 2 and 3, $\mathbb{Q}^+$-exponent string and $\mathbb{R}^+$-exponent string are extensions of string($\mathbb{N}$-exponent-string)!

Parikh Matrix defined over $\mathbb{Q}^+$-exponent-string

- *Recall*

- For $\Sigma = \{a < b < c\}$,

$$\Psi_M(a) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \ \Psi_M(b) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \Psi_M(c) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\Psi_M(abca) = \Psi_M(a)\Psi_M(b)\Psi_M(c)\Psi_M(a) \ .$$

Parikh Matrix defined over $\mathbb{Q}^+$-exponent-string

- For $\Sigma = \{a < b < c\}$,

$$\Psi_M(a) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \ \Psi_M(b) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \Psi_M(c) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\Psi_M^{\mathbb{Q}^+}\left(a^{\frac{3}{2}} b c^{\frac{1}{5}} a^{\frac{4}{3}}\right) = [\Psi_M(a)]^{\frac{3}{2}} [\Psi_M(b)]^{1} [\Psi_M(c)]^{\frac{1}{5}} [\Psi_M(a)]^{\frac{4}{3}} .$$

Parikh Matrix defined over $\mathbb{Q}^+$-exponent-string

- How to calculate $\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^{\frac{7}{11}}$?

Parikh Matrix defined over $\mathbb{Q}^+$-exponent-string

- $\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^{\frac{7}{11}} = \left[I + \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \right]^{\frac{7}{11}} = [I + A]^{\frac{7}{11}}$

$$= I + \frac{\frac{7}{11}}{1!} A + \frac{\frac{7}{11}\left(\frac{7}{11} - 1\right)}{2!} A^2 + \frac{\frac{7}{11}\left(\frac{7}{11} - 1\right)\left(\frac{7}{11} - 2\right)}{3!} A^3 + \cdots$$

Parikh Matrix defined over $\mathbb{Q}^+$-exponent-string

- $= I + \frac{7}{11}A + O + O + \cdots = \begin{pmatrix} 1 & \frac{7}{11} & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$

Parikh Matrix defined over $\mathbb{Q}^+$-exponent-string

- *Recall*

$\Sigma = \{a < b\}$, and $w \in \Sigma^*$,

$$\Psi_M(w) = \begin{pmatrix} 1 & |w|_a & |w|_{ab} \\ 0 & 1 & |w|_b \\ 0 & 0 & 1 \end{pmatrix}.$$

$$\Psi_M(a^2b^3a) = \begin{pmatrix} 1 & 3 & 6 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}$$

Parikh Matrix defined over $\mathbb{Q}^+$-exponent-string

- For $\mathbb{Q}$-exponent-strings:

- $\Psi_M^{\mathbb{Q}^+}\left(a^3 b^{\frac{5}{3}} a^{\frac{1}{2}}\right) = \begin{pmatrix} 1 & \frac{7}{2} & 5 \\ 0 & 1 & \frac{5}{3} \\ 0 & 0 & 1 \end{pmatrix}$ for $\Sigma = \{a < b\}$.

- $\Psi_M^{\mathbb{Q}^+}\left(a^2 b^{\frac{1}{2}} c^{\frac{1}{2}}\right) = \begin{pmatrix} 1 & 2 & 1 & 0.5 \\ 0 & 1 & 0.5 & 0.25 \\ 0 & 0 & 1 & 0.5 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ for $\Sigma = \{a < b < c\}$.

EM-equivalence

- **Definition(EM-equivalence).** For $p, q \in \Sigma^*_{\mathbb{Q}^+}$, $p \equiv_{EM} q$ if and only if $\Psi_M^{\mathbb{Q}^+}(p) = \Psi_M^{\mathbb{Q}^+}(q)$.

- *Remark.* $\equiv_M$ is restriction of $\equiv_{EM}$ to Σ^*.

- Characterization of M-equivalence is followed if we characterize EM-equivalence!

EM-equivalence

- **Theorem 3(Universal rewriting rule).** Let $\Sigma = \{a_1 < a_2 < \cdots < a_n\}$. Then for $\forall p, q \in \Sigma^*_{\mathbb{Q}^+}$, $p \equiv_{EM} q$ if and only if $p \leftrightarrow^* q$.

R1. ${a_i}^x {a_j}^y \leftrightarrow {a_j}^y {a_i}^x$ for $|i - j| \geq 2$, and $x, y \in \mathbb{Q}^+$.

R2. ${a_i}^x {a_{i+1}}^{2y} a_i^x \leftrightarrow {a_{i+1}}^y {a_i}^{2x} a_{i+1}^y$ for $1 \leq i \leq n - 1$, and $x, y \in \mathbb{Q}^+$.

- **Proposition 1(Atanasiu02).** For an alphabet $\Sigma = \{a_1 < a_2 < \cdots < a_n\}$, $w_1 \equiv_M w_2$ if $\mathrm{w}_1 \leftrightarrow^* w_2$ with respect to following rules.

A1. $a_i a_{i+1} a_{i+1} a_i \leftrightarrow a_{i+1} a_i a_i a_{i+1}$ for $1 \leq i \leq n - 1$.

A2. $a_i a_j \leftrightarrow a_j a_i$ for $|i - j| \geq 2$.