Equivalence Relations Defined by Numbers of Occurrences of Factors

Aleksi Saarela

Department of Mathematics and Statistics
and FUNDIM Centre,
University of Turku, Finland

RuFiDiM 2014
Petrozavodsk
17.9.2014

Outline

Notation

- $|u|_x$ is the number of occurrences of x as a factor in u
  - $|u|_\varepsilon = |u| + 1$
- u being a proper prefix of v is denoted by $u < v$
- u being a proper suffix of v is denoted by $v > u$
- $\mathrm{pref}_n(u)$ is the prefix of u of length n (or $\mathrm{pref}_n(u) = u$ if $|u| < n$)
- $\mathrm{suff}_n(u)$ is the suffix of u of length n (or $\mathrm{suff}_n(u) = u$ if $|u| < n$)
- $[P] = \begin{cases} 1 & \text{if } P \text{ is true} \\ 0 & \text{if } P \text{ is false} \end{cases}$
- $\mathcal{L}(V)$ is the vector space generated by a set of vectors V
- $\mathrm{rank}(V) = \dim(\mathcal{L}(V))$

Motivating question

Let $k \geq 1$ and $S \subseteq \Sigma^{\leq k}$.
If we know $|u|_s$ for all $s \in S$, $\mathrm{pref}_{k-1}(u)$, and $\mathrm{suff}_{k-1}(u)$, what can we say about u?

Example

Suppose that we do not know the word $u \in \{0,1\}^*$, but we know $|u|_\varepsilon$ and $|u|_0$. Then we can of course deduce the number of 1's:
$|u|_1 = |u|_\varepsilon - 1 - |u|_0$.

Example

Suppose that we do not know the word $u \in \{0,1\}^+$, but we know $\mathrm{pref}_1(u)$, $\mathrm{suff}_1(u)$, and $|u|_{01}$. Then we can deduce the number of 10's:
$|u|_{10} = |u|_{01} + \mathrm{pref}_1(u) - \mathrm{suff}_1(u)$.

k-abelian equivalence

Words $u, v \in \Sigma^*$ are *k-abelian equivalent* if

- $|u|_s = |v|_s$ for all $s \in S$,
- $\mathrm{pref}_{k-1}(u) = \mathrm{pref}_{k-1}(v)$, and
- $\mathrm{suff}_{k-1}(u) = \mathrm{suff}_{k-1}(v)$,

where the set S can be any of the following:

- $\Sigma^{\leq k}$,
- Σ^k,
- $\Sigma^{\leq k} \setminus 0\Sigma^* \setminus \Sigma^* 0$ $(0 \in \Sigma)$.

Can we characterize all possible sets S?

Outline

(k, S)-equivalence

For an alphabet Σ, positive integer k, and set $S \subseteq \Sigma^{\leq k}$, words $u, v \in \Sigma^*$ are called (k, S)-*equivalent* if

- $|u|_s = |v|_s$ for all $s \in S$,
- $\mathrm{pref}_{k-1}(u) = \mathrm{pref}_{k-1}(v)$, and
- $\mathrm{suff}_{k-1}(u) = \mathrm{suff}_{k-1}(v)$

Example

- $S = \Sigma$: abelian equivalence
- $S = \Sigma^k$: k-abelian equivalence
- $\Sigma = \{0, 1\}$, $S = \{0\}$: equivalence classes are $1^*, 1^*01^*, 1^*01^*01^*, 1^*01^*01^*01^*, \dots$

Maximal and minimal sets

The set $\overline{S}$ is defined to consist of all words $t \in \Sigma^{\leq k}$ such that $|u|_t$ depends only on the (k, S)-equivalence class of u.

- The set $\overline{S}$ does not depend on k (but it depends on Σ).
- For $S_1, S_2 \subseteq \Sigma^{\leq k}$, (k, S_1)-equivalence and (k, S_2)-equivalence are the same if and only if $\overline{S_1} = \overline{S_2}$.

A set R is S-*minimal* if $\overline{R} = \overline{S}$ and $\overline{Q} \neq \overline{S}$ for all $Q \subsetneq R$.

- Every set S has an S-minimal subset, but an S-minimal set need not be a subset of S.
- All S-minimal sets are of the same size (this is not trivial).

Example

Let $\Sigma = \{0, 1\}$.

- $\overline{\Sigma} = \{\varepsilon, 0, 1\}$ and $\overline{\{01\}} = \{01, 10\}$.
- The Σ-minimal sets are $\{\varepsilon, 0\}, \{\varepsilon, 1\}, \Sigma$.

Questions

- Given S, how big are S-minimal sets?
- Which sets S are S-minimal?
- Given k and S, for which $t \in \Sigma^{\leq k}$ can we deduce $|u|_t$ based on the (k, S)-equivalence class of u? In other words, what is the set $\overline{S}$?
- For which S_1, S_2 is (k, S_1)-equivalence the same as (k, S_2)-equivalence, or in other words, $\overline{S}_1 = \overline{S}_2$.
- For which S is (k, S)-equivalence the same as k-abelian equivalence, or in other words, $\overline{S} = \Sigma^{\leq k}$.

Outline

Vectors

Let $t_1, \ldots, t_M$ be the words in $\Sigma^{\leq k}$ in radix order.
The *extended Parikh vector* of a word $u \in \Sigma^*$ is $P_u = (|u|_{t_1}, \ldots, |u|_{t_M})$.
We define families of vectors:

$$\begin{aligned}
U_t &= (a_1, \ldots, a_M), \quad \text{where } a_i = [t_i > t] - [t < t_i] & \text{and } t \in \Sigma^{k-1}, \\
U'_t &= (a_1, \ldots, a_M), \quad \text{where } a_i = [t_i = t] - [t < t_i \in \Sigma^k] & \text{and } t \in \Sigma^{\leq k-1}, \\
V_t &= (a_1, \ldots, a_M), \quad \text{where } a_i = [t_i = t] & \text{and } t \in \Sigma^{\leq k}.
\end{aligned}$$

Lemma

Let $u \in \Sigma^*$. Then

$$\begin{aligned}
U_t \cdot P_u &= [u > t] - [t < u] & \textit{for } t \in \Sigma^{k-1}, \\
U'_t \cdot P_u &= |\mathrm{suff}_{k-1}(u)|_t & \textit{for } t \in \Sigma^{\leq k-1}, \\
V_t \cdot P_u &= |u|_t & \textit{for } t \in \Sigma^{\leq k}.
\end{aligned}$$

Vectors

Let $\mathcal{U} = \{U_t \mid t \in \Sigma^{k-1}, t \neq 0^{k-1}\} \cup \{U'_t \mid t \in \Sigma^{\leq k-1}\}$ and $\mathcal{V}_S = \{V_s \mid s \in S\}$ for $S \subseteq \Sigma^{\leq k}$.

Lemma

The set $\mathcal{U}$ is linearly independent.

Lemma

Let $S \subseteq \Sigma^{\leq k}$ and $t \in \Sigma^{\leq k}$. If $V_t \in \mathcal{L}(\mathcal{U} \cup \mathcal{V}_S)$, then $t \in \overline{S}$.

Example

Let $\Sigma = \{0, 1\}$ and $S = \{01\}$. Then

$$\begin{aligned}
\mathcal{U} = \{U_1, U'_\varepsilon, U'_0, U'_1\} &= \{(0,0,0,0,1,-1,0), (1,0,0,-1,-1,-1,-1), \\
&\qquad (0,1,0,-1,-1,0,0), (0,0,1,0,0,-1,-1)\}, \\
\mathcal{V}_S = \{V_{01}\} &= \{(0,0,0,0,1,0,0)\}, \\
V_{10} = (0,0,0,0,0,1,0) &\in \mathcal{L}(\mathcal{U} \cup \mathcal{V}_S).
\end{aligned}$$

Vectors

Lemma

If $S \subseteq \Sigma^{\leq k}$ is S-minimal, then $\mathcal{U} \cup \mathcal{V}_S$ is linearly independent.

Proof.

Let $\mathcal{U} \cup \mathcal{V}_S$ be linearly dependent. The set $\mathcal{U}$ is linearly independent, so there exists $s \in S$ such that $V_s \in \mathcal{L}(\mathcal{U} \cup \mathcal{V}_R)$, where $R = S \setminus \{s\}$. Then $s \in \overline{R}$, so $\overline{R} = \overline{S}$ and S is not S-minimal. $\square$

Outline

Number of equivalence classes

For a finite set $A \subset \Sigma^*$, the number of (k, S)-equivalence classes of words in A is denoted by $\mathrm{nec}_{k,S}(A)$.

Lemma

Let $S \subseteq \Sigma^{\leq k}$ *and* $\overline{S} = \Sigma^{\leq k}$.
Then $\mathrm{nec}_{k,S}(\Sigma^{\leq n}) = \Theta(n^m)$, *where* $m = \#\Sigma^k - \#\Sigma^{k-1} + 1$.

Lemma

Let $S \subseteq \Sigma^{\leq k}$ *and let* R *be* S*-minimal.*
Then $\mathrm{nec}_{k,S}(\Sigma^{\leq n}) = O(n^{\#R})$.

Lemma

Let $S \subseteq \Sigma^{\leq k}$ *and* $m = \mathrm{rank}(\mathcal{U} \cup \mathcal{V}_{\overline{S}}) - \mathrm{rank}(\mathcal{U})$.
Then $\mathrm{nec}_{k,S}(\Sigma^{\leq n}) = \Omega(n^m)$.

Number of equivalence classes

Theorem

Let $S \subseteq \Sigma^{\leq k}$ and let R be S-minimal. Then

$$\#R = \mathrm{rank}(\mathcal{U} \cup \mathcal{V}_S) - \mathrm{rank}(\mathcal{U}) \qquad \textit{and} \qquad \mathrm{nec}_{k,S}(\Sigma^{\leq n}) = \Theta(n^{\#R}).$$

Corollary

Let $S_1, S_2 \subseteq \Sigma^{\leq k}$. If $\overline{S}_1 = \overline{S}_2$, then $\mathrm{rank}(\mathcal{U} \cup \mathcal{V}_{S_1}) = \mathrm{rank}(\mathcal{U} \cup \mathcal{V}_{S_2})$.

Outline

Which sets S are S-minimal?

Theorem

A set $S \subseteq \Sigma^{\leq k}$ is S-minimal if and only if $\mathcal{U} \cup \mathcal{V}_S$ is linearly independent.

Proof.

If S is S-minimal, then the claim follows from an earlier lemma.
If S is not S-minimal, then it has a proper subset R such that $\overline{R} = \overline{S}$.
Then $\mathrm{rank}(\mathcal{U} \cup \mathcal{V}_S) = \mathrm{rank}(\mathcal{U} \cup \mathcal{V}_R) \leq \#(\mathcal{U} \cup \mathcal{V}_R) < \#(\mathcal{U} \cup \mathcal{V}_S)$,
so $\mathcal{U} \cup \mathcal{V}_S$ cannot be linearly independent. □

What is the set $\overline{S}$?

Theorem

Let $S \subseteq \Sigma^{\leq k}$ and $t \in \Sigma^{\leq k}$. Then $t \in \overline{S}$ if and only if $V_t \in \mathcal{L}(\mathcal{U} \cup \mathcal{V}_S)$.

Proof.

If $V_t \in \mathcal{L}(\mathcal{U} \cup \mathcal{V}_S)$, then the claim follows from an earlier lemma.
If $V_t \notin \mathcal{L}(\mathcal{U} \cup \mathcal{V}_S)$, then $\mathrm{rank}(\mathcal{U} \cup \mathcal{V}_{S \cup \{t\}}) > \mathrm{rank}(\mathcal{U} \cup \mathcal{V}_S)$.
Then $\overline{S \cup \{t\}} \neq \overline{S}$, so $t \notin \overline{S}$. $\square$

When is $\overline{S}_1 = \overline{S}_2$?

Corollary

Let $S_1, S_2 \subseteq \Sigma^{\leq k}$. *Then* $\overline{S}_1 = \overline{S}_2$ *if and only if* $\mathcal{L}(\mathcal{U} \cup \mathcal{V}_{S_1}) = \mathcal{L}(\mathcal{U} \cup \mathcal{V}_{S_2})$.

Corollary

Let $S \subseteq \Sigma^{\leq k}$. *Then* $\overline{S} = \Sigma^{\leq k}$ *if and only if* $\mathrm{rank}(\mathcal{U} \cup \mathcal{V}_S) = \#\Sigma^{\leq k}$.

Thank You!