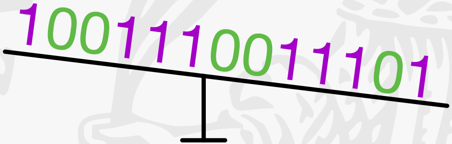
Weighted Prefix Normal Words: Mind the Gap

Yannik Eikmeier, **Pamela Fleischmann**, Mitja Kulcynski, Dirk Nowotka

Kiel University

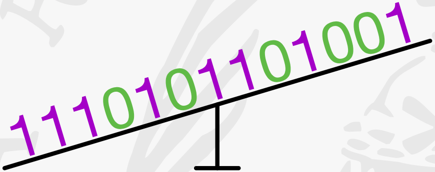
The Beginning

Compare number of 1 in any factor with the number of 1 in prefix of same length.



The Beginning

Compare number of 1 in any factor with the number of 1 in prefix of same length.



Prefixes have at least the same number of 1 as any factor of the same length.

Formally

Definition (Fici and Liptak 2011).
Given $w \in \{0, 1\}^*$

Formally

Definition (Fici and Liptak 2011).
Given $w \in \{0,1\}^*$

- maximum-ones-function
  $f_w : [|w|] \to \mathbb{N}_0;\ n \mapsto \max\{|u|_1 \mid u \in \mathrm{Fact}_n(w)\}$

Formally

Definition (Fici and Liptak 2011).
Given $w \in \{0,1\}^*$

- maximum-ones-function
  $f_w : [|w|] \to \mathbb{N}_0;\ n \mapsto \max\{|u|_1 \mid u \in \mathrm{Fact}_n(w)\}$
- prefix-ones-function
  $p_w : [|w|] \to \mathbb{N}_0;\ n \mapsto |\mathrm{Pref}_n(w)|_1$

Formally

Definition (Fici and Liptak 2011).
Given $w \in \{0,1\}^*$

- maximum-ones-function
  $f_w : [|w|] \to \mathbb{N}_0;\ n \mapsto \max\{|u|_1 \mid u \in \mathrm{Fact}_n(w)\}$
- prefix-ones-function
  $p_w : [|w|] \to \mathbb{N}_0;\ n \mapsto |\mathrm{Pref}_n(w)|_1$
- $w \sim v$ iff $f_w = f_v$

Formally

Definition (Fici and Liptak 2011).
Given $w \in \{0,1\}^*$

- maximum-ones-function
  $f_w : [|w|] \to \mathbb{N}_0;\ n \mapsto \max\{|u|_1 \mid u \in \mathrm{Fact}_n(w)\}$
- prefix-ones-function
  $p_w : [|w|] \to \mathbb{N}_0;\ n \mapsto |\mathrm{Pref}_n(w)|_1$
- $w \sim v$ iff $f_w = f_v$
- w prefix normal iff $p_w = f_w$

Main Results

Theorem (Fici and Liptak 2011).
Each class of prefix equivalent words contains exactly one prefix normal word.

Main Results

Theorem (Fici and Liptak 2011).
Each class of prefix equivalent words contains exactly one prefix normal word.

Theorem (Fici and Liptak 2016).
For given $x, y \in \mathbb{N}$, the answer for an indexed binary jumbled pattern matching query $q = (x, y)$ is yes if and only if

$$p_{pnf_1(w)}(x+y) \geq x \geq p_{pnf_0(w)}(x+y).$$

Main Results

Theorem (Fici and Liptak 2011).
Each class of prefix equivalent words contains exactly one prefix normal word.

Theorem (Fici and Liptak 2016).
For given $x, y \in \mathbb{N}$, the answer for an indexed binary jumbled pattern matching query $q = (x, y)$ is yes if and only if

$$p_{pnf_1(w)}(x+y) \geq x \geq p_{pnf_0(w)}(x+y).$$

Generalsing the Setting

Arbitrary Finite Alphabets

How to add a weight to each letter? Just using integer alphabets?

Generalsing the Setting

How to add a weight to each letter? Just using integer alphabets?

Definition.
totally ordered and cancellative monoid A

- morphism $\mu : \Sigma^* \to A$ weight measure iff for all $w \in \Sigma^*, v \in \Sigma^+$
  - $\mu(wv) = \mu(vw)$

Generalising the Setting

How to add a weight to each letter? Just using integer alphabets?

Definition.
totally ordered and cancellative monoid A

- morphism $\mu : \Sigma^* \to A$ weight measure iff for all $w \in \Sigma^*$, $v \in \Sigma^+$
  - $\mu(wv) = \mu(vw)$
  - $\mu(w) <_A \mu(wv)$ (increasing property)

Remark. $\mu(\varepsilon) = \mathbb{1}_A$.

Generalising the Setting

Definition.
Given $w \in \Sigma^*$, μ weight measure

Generalising the Setting

Definition.
Given $w \in \Sigma^*$, μ weight measure

- factor weight function
  $f_{w,\mu} : [|w|] \to \mathbb{N}_0;\ n \mapsto \max\{\mu(u) \mid u \in \mathrm{Fact}_n(w)\}$

Generalising the Setting

Definition.
Given $w \in \Sigma^*$, μ weight measure

- factor weight function
  $f_{w,\mu} : [|w|] \to \mathbb{N}_0;\ n \mapsto \max\{\mu(u) \mid u \in \mathrm{Fact}_n(w)\}$
- prefix weight function
  $p_{w,\mu} : [|w|] \to \mathbb{N}_0;\ n \mapsto \mu(\mathrm{Pref}_n(w))$

Generalising the Setting

Definition.
Given $w \in \Sigma^*$, μ weight measure

- factor weight function
  $f_{w,\mu} : [|w|] \to \mathbb{N}_0;\ n \mapsto \max\{\mu(u) |\ u \in \mathrm{Fact}_n(w)\}$
- prefix weight function
  $p_{w,\mu} : [|w|] \to \mathbb{N}_0;\ n \mapsto \mu(\mathrm{Pref}_n(w))$
- $w \sim_\mu v$ iff $f_{w,\mu} = f_{v,\mu}$

Generalising the Setting

Definition.
Given $w \in \Sigma^*$, μ weight measure

- factor weight function
  $f_{w,\mu} : [|w|] \to \mathbb{N}_0;\ n \mapsto \max\{\mu(u) \mid u \in \mathrm{Fact}_n(w)\}$
- prefix weight function
  $p_{w,\mu} : [|w|] \to \mathbb{N}_0;\ n \mapsto \mu(\mathrm{Pref}_n(w))$
- $w \sim_\mu v$ iff $f_{w,\mu} = f_{v,\mu}$
- w prefix normal iff $p_{w,\mu} = f_{w,\mu}$

Example

Given $\mu(\text{a}) = 1, \mu(\text{p}) = 2, \mu(\text{y}) = 3$

papaya

	1	2	3	4	5	6
$f_{w,\mu}$						
$p_{w,\mu}$						

Example

Given $\mu(\text{a}) = 1, \mu(\text{p}) = 2, \mu(\text{y}) = 3$

papaya

	1	2	3	4	5	6
$f_{w,\mu}$	3					
$p_{w,\mu}$	2					

Example

Given $\mu(\text{a}) = 1, \mu(\text{p}) = 2, \mu(\text{y}) = 3$

papaya

	1	2	3	4	5	6
$f_{w,\mu}$	3	4				
$p_{w,\mu}$	2	3				

Example

Given $\mu(\text{a}) = 1, \mu(\text{p}) = 2, \mu(\text{y}) = 3$

papaya

	1	2	3	4	5	6
$f_{w,\mu}$	3	4	6			
$p_{w,\mu}$	2	3	5			

Example

Given $\mu(\mathrm{a}) = 1, \mu(\mathrm{p}) = 2, \mu(\mathrm{y}) = 3$

papaya

	1	2	3	4	5	6
$f_{w,\mu}$	3	4	6	7		
$p_{w,\mu}$	2	3	5	6		

Example

Given $\mu(\text{a}) = 1, \mu(\text{p}) = 2, \mu(\text{y}) = 3$

papaya

	1	2	3	4	5	6
$f_{w,\mu}$	3	4	6	7	9	
$p_{w,\mu}$	2	3	5	6	9	

Example

Given $\mu(\mathrm{a}) = 1, \mu(\mathrm{p}) = 2, \mu(\mathrm{y}) = 3$

papaya

	1	2	3	4	5	6
$f_{w,\mu}$	3	4	6	7	9	10
$p_{w,\mu}$	2	3	5	6	9	10

Example

Given $\mu(\text{a}) = 1, \mu(\text{p}) = 2, \mu(\text{y}) = 3$

papaya

	1	2	3	4	5	6
$f_{w,\mu}$	3	4	6	7	9	10
$p_{w,\mu}$	2	3	5	6	9	10

papaya is not μ-prefix normal!

Example

Given $\mu(a) = 1, \mu(n) = 2, \mu(b) = 3$ and

banana

we obtain

	1	2	3	4	5	6
$f_{w,\mu}$	3	4	6	7	9	10
$p_{w,\mu}$	3	4	6	7	9	10

Prefix Normal Form

Consider $\Sigma = \{a, b, n, p, y\}$ and $\mu(a) = 1$, $\mu(n) = \mu(p) = 2$, $\mu(b) = \mu(y) = 3$:

Prefix Normal Form

Consider $\Sigma = \{a, b, n, p, y\}$ and $\mu(a) = 1$, $\mu(n) = \mu(p) = 2$, $\mu(b) = \mu(y) = 3$:

$$[\mathtt{banana}]_{\sim_\mu} = \{\, x_1 a x_2 a x_2 a,\ a x_2 a x_2 a x_1,\ a x_2 a x_1 a x_2,\ a x_1 a x_2 a x_2,\ x_2 a x_2 a x_1 a,\ x_2 a x_1 a x_2 a \mid x_1 \in \{b, y\}, x_2 \in \{n, p\} \,\}$$

Prefix Normal Form

Consider $\Sigma = \{a, b, n, p, y\}$ and $\mu(a) = 1$, $\mu(n) = \mu(p) = 2$, $\mu(b) = \mu(y) = 3$:

$$[\text{banana}]_{\sim_\mu} = \{\, x_1 a x_2 a x_2 a,\ a x_2 a x_2 a x_1,\ a x_2 a x_1 a x_2,\ a x_1 a x_2 a x_2,\ x_2 a x_2 a x_1 a,\ x_2 a x_1 a x_2 a \mid x_1 \in \{b, y\}, x_2 \in \{n, p\} \,\}$$

$\rightsquigarrow$ $x_1 a x_2 a x_2 a$ for all $x_1 \in \{b, y\}, x_2 \in \{n, p\}$ are μ-prefix normal!

Prefix Normal Form

Consider $\Sigma = \{a, b, n, p, y\}$ and $\mu(a) = 1$, $\mu(n) = \mu(p) = 2$, $\mu(b) = \mu(y) = 3$:

$$[\text{banana}]_{\sim_\mu} = \{\, x_1 a x_2 a x_2 a,\ a x_2 a x_2 a x_1,\ a x_2 a x_1 a x_2,\ a x_1 a x_2 a x_2,\ x_2 a x_2 a x_1 a,\ x_2 a x_1 a x_2 a \mid x_1 \in \{b, y\}, x_2 \in \{n, p\} \,\}$$

$\rightsquigarrow$ $x_1 a x_2 a x_2 a$ for all $x_1 \in \{b, y\}, x_2 \in \{n, p\}$ are μ-prefix normal!

Notice.
Only if the weight measure is not injective, we have more than one prefix normal form.

Prefix Normal Forms

Consider $\Sigma = \{\mathrm{d}, \mathrm{e}, \mathrm{u}\}$ and $\mu(\mathrm{u}) = 1$, $\mu(\mathrm{e}) = 2$, $\mu(\mathrm{d}) = 4$:

Prefix Normal Forms

Consider $\Sigma = \{\mathrm{d}, \mathrm{e}, \mathrm{u}\}$ and $\mu(\mathrm{u}) = 1$, $\mu(\mathrm{e}) = 2$, $\mu(\mathrm{d}) = 4$:

$$[\mathrm{dude}]_{\sim_\mu} = \{\mathrm{dude}, \mathrm{edud}\}.$$

Prefix Normal Forms

Consider $\Sigma = \{\mathrm{d}, \mathrm{e}, \mathrm{u}\}$ and $\mu(\mathrm{u}) = 1$, $\mu(\mathrm{e}) = 2$, $\mu(\mathrm{d}) = 4$:

$$[\mathrm{dude}]_{\sim_\mu} = \{\mathrm{dude}, \mathrm{edud}\}.$$

$\rightsquigarrow$ no normal form witnessed by

	1	2	3	4
$f_{\mathrm{dude},\mu}, f_{\mathrm{edud},\mu}$	4	6	9	11
$p_{\mathrm{dude},\mu}$	4	5	9	11
$p_{\mathrm{edud},\mu}$	2	6	7	11

Prefix Normal Forms

Consider $\Sigma = \{\mathrm{d}, \mathrm{e}, \mathrm{u}\}$ and $\mu(\mathrm{u}) = 1$, $\mu(\mathrm{e}) = 2$, $\mu(\mathrm{d}) = 4$:

$$[\mathrm{dude}]_{\sim_\mu} = \{\mathrm{dude}, \mathrm{edud}\}.$$

$\rightsquigarrow$ no normal form witnessed by

	1	2	3	4
$f_{\mathrm{dude},\mu}, f_{\mathrm{edud},\mu}$	4	6	9	11
$p_{\mathrm{dude},\mu}$	4	5	9	11
$p_{\mathrm{edud},\mu}$	2	6	7	11

Notice.
If the weight measure has gaps, we do not have a prefix normal form.

Gapfree Weight Measure

Definition.
Weight measure μ gapfree iff for all $i \in [|w|]$ there exists $\mathtt{a} \in \Sigma$ with

$$f_{w,\mu}(i) = f_{w,\mu}(i-1) \circ_A \mu(\mathtt{a}).$$

Prefix Normal Form

$\mathcal{P}_\mu(w)$ set of all prefix normal words in $[w]_{\sim_\mu}$

Theorem.

- there exists $w \in \Sigma^*$ with $|\mathcal{P}_\mu(w)| = 0$ iff μ is gapful
- there exists $w \in \Sigma^*$ with $|\mathcal{P}_\mu(w)| > 1$ iff μ is not injective
- for all $w \in \Sigma^*$, $|\mathcal{P}_\mu(w)| = 1$ iff μ is gapfree, injective

Gapfree Weight Measures

Definition.
μ_A, μ_B weight measures over Σ w.r.t. monoids A, resp. B

- μ_A equivalent to μ_B iff for all $n \in \mathbb{N}$, for all $v, w \in \Sigma^n$

$$\mu_A(v) <_A \mu_A(w) \text{ iff } \mu_B(v) <_B \mu_B(w).$$

Equivalent weight measures behave in the same way, e.g. $\mathcal{P}_{\mu_A}(w) = \mathcal{P}_{\mu_B}(w)$ for all $w \in \Sigma^*$.

Standard Weight Measure

Definition.
$\Sigma = \{a_1, \ldots, a_n\}$ strictly totally ordered alphabet

- μ_Σ standard weight measure iff $\mu_\Sigma(a_i) = i$

The standard weight measure is gapfree, injective, an alphabetically ordered ($a_i <_\Sigma a_j$ implies $\mu(a_i) <_A \mu(a_j)$).

Characterisation of Gapfree

Theorem.
For any non-binary, injective, alphabetically ordered weight measure μ, we have the following equivalences

1. μ is gapfree
2. μ has no gap over any word of the form $\mathtt{cacb}$ with $\mathtt{a} <_\Sigma \mathtt{b} <_\Sigma \mathtt{c}$
3. μ is equivalent to the standard weight measure μ_Σ

Characterisation of Gapfree

Theorem.
For any non-binary, injective, alphabetically ordered weight measure μ, we have the following equivalences

1. μ is gapfree
2. μ has no gap over any word of the form $\mathtt{cacb}$ with $\mathtt{a} <_\Sigma \mathtt{b} <_\Sigma \mathtt{c}$
3. μ is equivalent to the standard weight measure μ_Σ

The gapfree property is decidable in time $\mathcal{O}(|\Sigma|^4)$: test for all $\binom{|\Sigma|}{3}$ possible enumeration $\mathtt{a} <_\Sigma \mathtt{b} <_\Sigma \mathtt{c}$ whether there exists $\mathtt{x} \in \Sigma$ with $\mu(\mathtt{bx}) = \mu(\mathtt{ac})$.

The Prefix Normal Form

Definition.
Σ strictly totally ordered alphabet, $w \in \Sigma^*$

- the μ_Σ prefix normal form is the unique prefix normal form of w

The Prefix Normal Form

Definition.
Σ strictly totally ordered alphabet, $w \in \Sigma^*$

- the μ_Σ prefix normal form is the unique prefix normal form of w

$\Sigma = \{a_1, a_2, a_3\} \rightsquigarrow a_3 a_2 a_2 a_2$ is the prefix normal form of $a_2 a_3 a_1 a_3$

Thank you for your attention!