Subword Metrics for Infinite Words

Stefan Hoffmann and Ludwig Staiger

University of Trier
and
Martin-Luther-University Halle-Wittenberg

CIAA, Umeå, 20. August 2015

Notation: Strings and Languages

Finite Alphabet $X = \{0, \ldots, r-1\}$, cardinality $|X| = r$

Finite strings (words) $w = x_1 \cdots x_n \in X^*, x_i \in X$, length $|w| = n$

Prefixes of infinite words $\xi[1 \ldots n] \in X^*, |\xi[1 \ldots n]| = n$,

$$\mathbf{pref}(\xi) := \{\xi[1 \ldots n] : n \in \mathbb{N}\}$$

Infixes (factors) of infinite strings

$$\mathbf{infix}(\xi) := \{v \in X^* : \xi = uv\eta \text{ for some } u \in X^*, \eta \in X^\omega\}$$

Infixes occuring infinitely often

$$\mathrm{infix}^\infty(\xi) := \{v \in X^* : \xi = uv\eta \text{ for inf. many } u \in X^*, \eta \in X^\omega\}$$

Languages and ω-Languages $W \subseteq X^*$ and $F \subseteq X^\omega$

X^ω as Cantor Space

Metric: $\rho(\xi, \eta) := \inf\{r^{-n} : \xi[1 \ldots n] = \eta[1 \ldots n]\}$

Balls: $w \cdot X^\omega = \{\xi : \xi[1 \ldots |w|] = w\}$.

Open sets: $W \cdot X^\omega = \bigcup_{w \in W} w \cdot X^\omega$

Closure: (Smallest closed set containing F)

$$\mathcal{C}(F) := \{\xi : \mathbf{pref}(\xi) \subseteq \mathbf{pref}(F)\}$$

Remark

A set $F \subseteq X^\omega$ is *closed* if and only if the property that each prefix of ξ is the prefix of some word in F implies that $\xi \in F$.

Shift-Invariance

Definition (Left and right shift)

- $F/w := \{\xi : w \cdot \xi \in F\}$ is the *left shift* (or *left derivative*) of F induced by $w \in X^*$
- $w \cdot F := \{w \cdot \xi : \xi \in F\}$ is the *right shift* of F induced by $w \in X^*$

Fact

The Cantor space is shift-invariant, that is, for each open (closed) set $E \subseteq X^\omega$ also $w \cdot E$ and E/w are open (closed).

Automata on ω-words: Büchi-automata

Automaton: $\mathcal{A} = (X, Q, \Delta, q_0, Q_{\text{fin}})$ with
$\Delta \subseteq Q \times X \times Q,\ q_0 \in Q,\ Q_{\text{fin}} \subseteq Q$

Run on ξ: $(q_i)_{i \in \mathbb{N}}$ with $\forall i \geq 0 : (q_i, \xi(i+1), q_{i+1}) \in \Delta$

$$\begin{array}{ccccccccccc} q_0 & & q_1 & & q_2 & & q_{i-1} & & q_i & & \\ & \searrow & \uparrow & \searrow & \uparrow & \cdots & \uparrow & \searrow & \uparrow & \searrow & \cdots \\ & & \xi(1) & & \xi(2) & & \xi(i-1) & & \xi(i) & & \end{array}$$

$\mathcal{A}$ accepts ξ: $\exists (q_i)_{i \in \mathbb{N}} \quad \forall i \geq 0 : (q_i, \xi(i+1), q_{i+1}) \in \Delta \quad \wedge$
$\exists^\infty k : q_k \in Q_{\text{fin}}$

$\mathcal{A}$ accepts F: $F = \{\xi : \mathcal{A} \text{ accepts } \xi\}$

Regular ω-languages

Definition (Regular ω-language)

An ω-language $F \subseteq X^\omega$ is called *regular* if and only if F is accepted by a finite automaton.

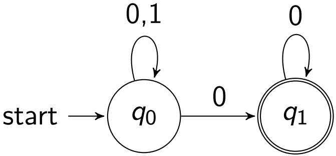
Theorem (Büchi 1962)

1. *The family of regular ω-languages is a Boolean algebra.*
2. *If $F \subseteq X^\omega$ is regular, then the shifts $u \cdot F$ and F/w are also regular.*

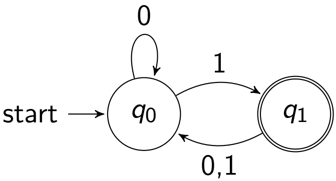
Automata on ω-words: Büchi-automata

Example

- $F = \{0, 1\}^* \cdot 0^\omega$



- $F = (0^*1)^\omega$



Regular ω-languages: Deterministic Case

Fact (Landweber'69)

$$F \text{ deterministic Büchi} \iff \text{regular and } G_\delta$$

$$F \text{ co-deterministic Büchi} \iff \text{regular and } F_\sigma$$

Example

For $u \in X^* \setminus \{e\}$ the ω-language $X^* \cdot u^\omega$ is regular, but not accepted by any deterministic Büchi-automaton.

Subword Metrics

Definition (Subword metrics)

For $\xi, \eta \in X^\omega$ define

$$\rho_I(\xi,\eta) := \inf\{r^{-n} : \xi[1\ldots n] = \eta[1\ldots n] \text{ and } \mathbf{infix}(\xi)\cap X^n = \mathbf{infix}(\eta)\cap X^n\}$$

and

$$\rho_\infty(\xi,\eta) := \inf\{r^{-n} : \xi[1\ldots n] = \eta[1\ldots n] \text{ and } \mathrm{infix}^\infty(\xi)\cap X^n = \mathrm{infix}^\infty(\eta)\cap X^n\}.$$

Remark

As $\rho_I(\xi,\eta) \geq \rho(\xi,\eta)$ and $\rho_\infty(\xi,\eta) \geq \rho(\xi,\eta)$ both spaces (X^ω, ρ_I) and (X^ω, ρ_∞) refine the Cantor space.

Shift-Invariance - Proof

Lemma

Let $u \in X^*$ *and* $|v| = |w| = m$*. Then*

$$\begin{aligned}
\rho_\infty(u \cdot \xi, u \cdot \eta) &\leq \rho_\infty(\xi, \eta), &(1)\\
\rho_\infty(\xi, \eta) &\leq r^m \cdot \rho_\infty(w \cdot \xi, v \cdot \eta), &(2)\\
\rho_I(u \cdot \xi, u \cdot \eta) &\leq \rho_I(\xi, \eta), \textit{ and} &(3)\\
\rho_I(\xi, \eta) &\leq r^m \cdot \rho_I(w \cdot \xi, v \cdot \eta). &(4)
\end{aligned}$$

Hence the maps $\Phi_u(\xi) := u \cdot \xi$ and $\Phi_m(u \cdot \xi) := \xi$ for $|u| = m$ are continuous, and we get:

Corollary

The spaces (X^ω, ρ_I) *and* (X^ω, ρ_∞) *are shift invariant.*

Balls in (X^ω, ρ_I) and (X^ω, ρ_∞)

Let $K_I(\xi, r^{-n})$ and $K_\infty(\xi, r^{-n})$ be the open balls of radius r^{-n} in the spaces (X^ω, ρ_I) and (X^ω, ρ_∞), respectively. Then

$$\begin{aligned} K_I(\xi, r^{-n}) &= w \cdot X^\omega \cap \bigcap_{u \in W} X^* \cdot u \cdot X^\omega \setminus \bigcup_{u \in \overline{W}} X^* \cdot u \cdot X^\omega, \\ K_\infty(\xi, r^{-n}) &= w \cdot X^\omega \cap X^* \cdot \big((\prod_{u \in V} X^* \cdot u)^\omega \setminus \bigcup_{u \in \overline{V}} X^* \cdot u \cdot X^\omega \big). \end{aligned}$$

with

- $w := \xi[1 \ldots n+1]$,
- $W := X^{n+1} \cap \mathbf{infix}(\xi)$, $\quad \overline{W} := X^{n+1} \setminus \mathbf{infix}(\xi)$,
- $V := X^{n+1} \cap \mathrm{infix}^\infty(\xi)$, $\quad \overline{V} := X^{n+1} \setminus \mathrm{infix}^\infty(\xi)$.

Remark

The open balls are regular.

Isolated Points in (X^ω, ρ_I)

Lemma

Let $w, u \in X^*, u \neq e$ and $\xi \in X^\omega$. Then $w \cdot u \sqsubset \xi$ and $\mathbf{infix}(\xi) \cap X^{|w \cdot u|} = \mathbf{infix}(w \cdot u^\omega) \cap X^{|w \cdot u|}$ imply $\xi = w \cdot u^\omega$.

Lemma

1. *Let $w \cdot u^\omega \in X^\omega$ where $|w| \leq |u|, |u| > 0$, and let $m > |w| + |u|$ and $n > |u|$.*
   Then $\rho_I(\xi, w \cdot u^\omega) \geq r^{-m}$ for all $\xi \neq w \cdot u^\omega$.
2. *Every ultimately periodic ω-word $w \cdot u^\omega$ is an isolated point in (X^ω, ρ_I).*

Lemma

1. *$\{uw^\omega : u, w \in X^*\}$ is the set of isolated points in (X^ω, ρ_I).*
2. *(X^ω, ρ_∞) has no isolated point.*

Non-preservation of regular ω-languages I

Fact

In Cantor space the closure $\mathcal{C}(F)$ is regular if F is regular.

Theorem

1. *In (X^ω, ρ_I) the closure of the regular ω-language $\{0,1\}^* \cdot 0^\omega$ is not regular.*
2. *In (X^ω, ρ_∞) the closure of the regular ω-language $\{0,1\}^* \cdot ((00)^*1)^\omega$ is not regular.*

Language of ultimately periodic words $\mathrm{Ult} := \{uw^\omega : u, w \in X^*\}$

Theorem (Büchi)

Let $E, F \subseteq X^\omega$ be regular ω-languages. Then

$$E = F \quad \textit{if and only if} \quad E \cap \mathrm{Ult} = F \cap \mathrm{Ult}\,.$$

Non-preservation of regular ω-languages II

Example

In (X^ω, ρ_I) the closure $C_I(X^* \cdot 0^\omega)$ of $X \cdot 0^\omega$

- does not contain any ultimately periodic ω-word $w \cdot u^\omega$ other than $w \cdot 0^\omega$,

$\Longrightarrow$ $\mathcal{C}_I(X^* \cdot 0^\omega) \cap \mathrm{Ult} = X^* \cdot 0^\omega$, and

- contains every ω-word ξ having **infix**$(\xi) = X^*$.

$\Longrightarrow$ $\mathcal{C}_I(X^* \cdot 0^\omega) \neq X^* \cdot 0^\omega$.

$\Longrightarrow$ $C_I(X^* \cdot 0^\omega)$ is not a regular ω-language.

Construction

Let **infix**$(\xi) = X^*$ and $v_n \in$ **pref**(ξ) such that **infix**$(v_n) = X^n$ and $|v_n| \geq n$.
Then $\rho_I(\xi, v_n \cdot 0^\omega) \leq r^{-n}$ and $\lim_{n\to\infty} v_n \cdot 0^\omega = \xi$.

Subword Complexity: Definition

Definition (Asymptotic subword complexity)

$$\tau(\xi) := \lim_{n\to\infty} \frac{\log_r |\mathbf{infix}(\xi)\cap X^n|}{n} = \inf\left\{ \frac{\log_r |\mathbf{infix}(\xi)\cap X^n|}{n} : n \in \mathbb{N} \right\}$$

Lemma (*St'93*)

$$\tau(\xi) := \lim_{n\to\infty} \frac{\log_r |\mathrm{infix}^\infty(\xi)\cap X^n|}{n} = \inf\left\{ \frac{\log_r |\mathrm{infix}^\infty(\xi)\cap X^n|}{n} : n \in \mathbb{N} \right\}$$

Definition (Level Set)

Let $0 < \gamma \leq 1$.

$$F_\gamma^{(\tau)} := \{\xi \in X^\omega : \tau(\xi) < \gamma\} \text{ and}$$

$$F_0^{(\tau)} := \mathrm{Ult}\,.$$

Subword Complexity

For $\xi \in X^\omega$ set

$$E_n(\xi) := \{\eta \in X^\omega : \mathbf{infix}(\eta) \cap X^n \subseteq \mathbf{infix}(\xi)\} \quad \text{and}$$

$$E'_n(\xi) := \{\eta \in X^\omega : \mathrm{infix}^\infty(\eta) \cap X^n \subseteq \mathrm{infix}^\infty(\xi)\}.$$

Fact

$$E_n(\xi) = X^\omega \setminus X^* \cdot (X^n \setminus \mathbf{infix}(\xi))) \cdot X^\omega$$

$$E'_n(\xi) = X^* \cdot (X^\omega \setminus X^* \cdot (X^n \setminus \mathrm{infix}^\infty(\xi)) \cdot X^\omega)$$

Subword Complexity

$$E_n(\xi) := \{\eta \in X^\omega : \mathbf{infix}(\eta) \cap X^n \subseteq \mathbf{infix}(\xi)\} \quad \text{and}$$

$$E'_n(\xi) := \{\eta \in X^\omega : \text{infix}^\infty(\eta) \cap X^n \subseteq \text{infix}^\infty(\xi)\}.$$

Lemma

Let $\xi \in X^\omega$. Then

- *the set $E_n(\xi)$ is open in (X^ω, ρ_I) and*
- *the set $E'_n(\xi)$ is open in (X^ω, ρ_∞).*

Proof.

- We show that $\eta \in E_n(\xi)$ implies $K_I(\eta, r^{-n}) \subseteq E_n(\xi)$.
- If $\eta \in E_n(\xi)$ and $\zeta \in K_I(\eta, r^{-n})$ then $\rho_I(\zeta, \eta) < r^{-n}$, in particular $\mathbf{infix}(\zeta) \cap X^n = \mathbf{infix}(\eta) \cap X^n$, whence $\zeta \in E_n(\xi)$.
- The proof for $E'_n(\xi)$ is similar.

□

Topological Properties of Level Sets

$$\begin{aligned} F_\gamma^{(\tau)} &:= \{\xi \in X^\omega : \tau(\xi) < \gamma\}, \gamma > 0, \text{ and} \\ F_0^{(\tau)} &:= \mathsf{Ult}\,. \end{aligned}$$

Theorem

Let $0 \leq \gamma \leq 1$. Then the sets $F_\gamma^{(\tau)}$ are open in (X^ω, ρ_I) and (X^ω, ρ_∞).

Corollary

The set $X^\omega \setminus F_1^{(\tau)} = \{\xi : \mathbf{infix}(\xi) = X^*\}$ is closed in (X^ω, ρ_I) and (X^ω, ρ_∞).

References

- Hoffmann, S.,
  Metriken zur Verfeinerung des Cantor-Raumes auf X^ω.
  Diploma thesis, Martin-Luther-Universität Halle-Wittenberg, 2014.
- Prodinger, H..
  Topologies on free monoids induced by closure operators of a special type.
  RAIRO Inform. Théor. Appl., 14(2):225–237, (1980).
- Schwarz, S. and Staiger, L.
  Topologies refining the Cantor topology on X^ω,
  in: IFIP Advances in Information and Communication Technology Vol. 323, Springer-Verlag, Berlin 2010, 271 - 285,
- Staiger, L.
  Kolmogorov complexity and Hausdorff dimension.
  Inform. and Comput., 103(2):159–194, (1993).

Thank you