CAI 2022,
ARISTOTLE UNIVERSITY OF
THESSALONIKI

When Variable-Length Codes Meet the Field of Error Detection

Jean Néraud

University of Rouen, France

A classical scheme to model information transmission

A classical scheme to model information transmission

$\phi : B^* \to A^*$ is a fixed one-to-one (injective) monoid homorphism

$\phi(uv) = \phi(u)\phi(v)$; $X = \phi(B)$ is a *code*; $w = \phi(u) \in \phi(B^*) = X^*$.

A classical scheme to model information transmission

$\phi : B^* \to A^*$ is a fixed one-to-one (injective) monoid homorphism
$\phi(uv) = \phi(u)\phi(v)$; $X = \phi(B)$ is a *code*; $w = \phi(u) \in \phi(B^*) = X^*$.

- Traditionally, the code-words have a common length.
  Example: $B = A = \{0, 1\}$, $\phi(0) = 000$, $\phi(1) = 111$
  $\to$ Theory of error-detecting codes

A classical scheme to model information transmission

$\phi : B^* \to A^*$ is a fixed one-to-one (injective) monoid homorphism
$\phi(uv) = \phi(u)\phi(v)$; $X = \phi(B)$ is a *code*; $w = \phi(u) \in \phi(B^*) = X^*$.

- Traditionally, the code-words have a common length.
  Example: $B = A = \{0, 1\}$, $\phi(0) = 000$, $\phi(1) = 111$
  $\to$ Theory of error-detecting codes

- In our paper, the code-words may have different lengths,
  $x_1 \cdots x_n = y_1 \cdots y_p \quad \Rightarrow n = p, \quad x_i = y_i \ \ (i \in [1, n])$.
  Counterexample: $X = \{a, ab, ba\}$: $(ab)a = a(ba)$
  $\to$ Theory of variable-length codes (codes, for short)

A classical scheme to model information transmission

Only highly likely errors need to be taken into account
$\rightarrow$ We overcome probabilistic aspect.

Several conditions have to be satisfied by the system.
[Jürgensen, Konstantinidis 1995, Néraud 2020]

- Synchronization condition $\rightarrow$ mandatory.
  (c0) $w \in X^* = \phi(B^*) \rightarrow w = x_1 \cdots x_n, \Rightarrow w' = x'_1 \cdots x'_n$.
  $x_i \in X$ transmitted into x'_i ($i \in [i, n]$).

A classical scheme to model information transmission

Only highly likely errors need to be taken into account
$\rightarrow$ We overcome probabilistic aspect.

Several conditions have to be satisfied by the system.
[Jürgensen, Konstantinidis 1995, Néraud 2020]

- Synchronization condition $\rightarrow$ mandatory.
  (c0) $w \in X^* = \phi(B^*) \rightarrow w = x_1 \cdots x_n, \Rightarrow w' = x'_1 \cdots x'_n$.
  $x_i \in X$ transmitted into x'_i $(i \in [i, n])$.

  $\rightarrow x_1, \cdots, x_n$ can be transmitted one by one.

Transmission of code-words

- Error detection condition

Transmission of code-words

- Error detection condition $\rightarrow$ Fix $k \geq 1$

Transmission of code-words

- Error detection condition $\rightarrow$ Fix $k \geq 1$

The system is capable to detect whether no more than k errors have occurred in the transmission of $x \in X$ into $x' \in A^*$.

Transmission of code-words

- Error detection condition $\rightarrow$ Fix $k \geq 1$

The system is capable to detect whether no more than k errors have occurred in the transmission of $x \in X$ into $x' \in A^*$.

$\rightarrow$ Fix a quasi-metric d over A^* (symmetry not required).

(c1) either $x' = x$ or $1 \leq d(x, x') \leq k$.

Transmission of code-words

- Error detection condition $\rightarrow$ Fix $k \geq 1$

The system is capable to detect whether no more than k errors have occurred in the transmission of $x \in X$ into $x' \in A^*$.

$\rightarrow$ Fix a quasi-metric d over A^* (symmetry not required).

(c1) either $x' = x$ or $1 \leq d(x, x') \leq k$.

$\rightarrow \tau_{d,k} \subseteq A^* \times A^*$ s.t. $\rightarrow x' \in \tau_{d,k}(x)$ iff. $d(x, x') \leq k$.
$\rightarrow \underline{\tau_{d,k}} = \tau_{d,k} \setminus \{(w, w') | w \in A^*\} = \tau_{d,k} \cap \overline{id_{A^*}}$.

Transmission of code-words

- Error detection condition $\rightarrow$ Fix $k \geq 1$

The system is capable to detect whether no more than k errors have occurred in the transmission of $x \in X$ into $x' \in A^*$.

$\rightarrow$ Fix a quasi-metric d over A^* (symmetry not required).

(c1) either $x' = x$ or $1 \leq d(x, x') \leq k$.

$\rightarrow \tau_{d,k} \subseteq A^* \times A^*$ s.t. $\rightarrow x' \in \tau_{d,k}(x)$ iff. $d(x, x') \leq k$.
$\rightarrow \underline{\tau_{d,k}} = \tau_{d,k} \setminus \{(w, w') | w \in A^*\} = \tau_{d,k} \cap \overline{id_{A^*}}$.

(c1) either $x' = x$ or $x' \in \underline{\tau_{d,k}}(x)$.

Transmission of code-words

- Error detection condition $\rightarrow$ Fix $k \geq 1$

The system is capable to detect whether no more than k errors have occurred in the transmission of $x \in X$ into $x' \in A^*$.

$\rightarrow$ Fix a quasi-metric d over A^* (symmetry not required).

(c1) either $x' = x$ or $1 \leq d(x, x') \leq k$.

$\rightarrow \tau_{d,k} \subseteq A^* \times A^*$ s.t. $\rightarrow x' \in \tau_{d,k}(x)$ iff. $d(x, x') \leq k$.
$\rightarrow \underline{\tau_{d,k}} = \tau_{d,k} \setminus \{(w, w') | w \in A^*\} = \tau_{d,k} \cap \overline{id_{A^*}}$.

(c1) either $x' = x$ or $x' \in \underline{\tau_{d,k}}(x)$.

(c1) $\underline{\tau_{d,k}}(X) \cap X = \emptyset$ (X is $\underline{\tau_{d,k}}$-independent).

Transmission of code-words

- Error correction condition
  (c2) $(x, y \in X$ and $\tau_{d,k}(x) \cap \tau_{d,k}(y) = \emptyset) \Rightarrow x = y$.

Transmission of code-words

- Maximality condition
  (c3) X maximal in the family of $\tau_{d,k}$-independent codes.
  ($\not\Rightarrow$ X maximal in the whole family of codes)

Transmission of code-words

- Maximality condition
  (c3) X maximal in the family of $\tau_{d,k}$-independent codes.
  ($\not\Rightarrow$ X maximal in the whole family of codes)

  $\rightarrow$ Uniform Bernoulli measure:
  $\pi : 2^{A^*} \rightarrow \mathbb{R}^+$, $a \in A \Rightarrow \pi(a) = 1/|A|$, $\pi(ww') = \pi(w)\pi(w')$,
  $\pi(X) = \sum_{x \in X} \pi(x)$.

Transmission of code-words

- Maximality condition
  (c3) X maximal in the family of $\tau_{d,k}$-independent codes.
  ($\not\Rightarrow$ X maximal in the whole family of codes)

  $\rightarrow$ Uniform Bernoulli measure:
  $\pi : 2^{A^*} \rightarrow \mathbb{R}^+$, $a \in A \Rightarrow \pi(a) = 1/|A|$, $\pi(ww') = \pi(w)\pi(w')$,
  $\pi(X) = \sum_{x \in X} \pi(x)$.

  - $X \subseteq A^*$ code $\Rightarrow \pi(X) \leq 1$ (Kraft-McMillan).

Transmission of code-words

- Maximality condition
  (c3) X maximal in the family of $\tau_{d,k}$-independent codes.
  ($\not\Rightarrow$ X maximal in the whole family of codes)

  $\rightarrow$ Uniform Bernoulli measure:
  $\pi : 2^{A^*} \rightarrow \mathbb{R}^+$, $a \in A \Rightarrow \pi(a) = 1/|A|$, $\pi(ww') = \pi(w)\pi(w')$,
  $\pi(X) = \sum_{x \in X} \pi(x)$.

  - $X \subseteq A^*$ code $\Rightarrow \pi(X) \leq 1$ (Kraft-McMillan).
  - $X \subseteq A^*$ regular code:
    X maximal $\Leftrightarrow A^* = F(X^*)$ (X complete) $\Leftrightarrow \pi(X) = 1$.
    [Schützenberger, 1965]

Conditions over X, regular code

(c1) Error-detection: X is $\underline{\tau}$-independent

$X \cap \tau_{d,k}(X) = \emptyset$.

(c2) Error-correction:

$x, y \in X$ and $\tau_{d,k}(x) \cap \tau_{d,k}(y) = \emptyset \Rightarrow x = y$.

(c3) X maximal in the family of $\underline{\tau_{d,k}}$-independent codes.

(c4) $X \cup \tau_{d,k}(X)$ is a variable-length code.

Conditions over X, regular code

(c1) Error-detection: X is $\underline{\tau}$-independent

$X \cap \tau_{d,k}(X) = \emptyset$.

(c2) Error-correction:

$x, y \in X$ and $\tau_{d,k}(x) \cap \tau_{d,k}(y) = \emptyset \Rightarrow x = y$.

(c3) X maximal in the family of $\tau_{d,k}$-independent codes.

(c4) $X \cup \tau_{d,k}(X)$ is a variable-length code.

Conditions over X, regular code

(c1) Error-detection: X is $\underline{\tau}$-independent
$X \cap \tau_{d,k}(X) = \emptyset$.

(c2) Error-correction:
$x, y \in X$ and $\tau_{d,k}(x) \cap \tau_{d,k}(y) = \emptyset \Rightarrow x = y$.

(c3) X maximal in the family of $\tau_{d,k}$-independent codes.

(c4) $X \cup \tau_{d,k}(X)$ is a variable-length code.

Can we decide whether or not X satisfies these conditions?

What (quasi)metric to fix ?

- Hamming, Levenshtein metrics *[Néraud, 2020]*
- d_P = prefix metric $d_P(x, x') = |x| + |x'| - 2|p|$, $p \in P(x) \cap P(x')$ and $|p|$ maximum
- d_F = factor metric $d_F(x, x') = |x| + |x'| - 2|f|$, $f \in F(x) \cap F(x')$ and $|f|$ maximum
- θ automorphism (resp., anti-automorphism) onto A^*: $\theta(A) = A$, $\theta(ww') = \theta(w)\theta(w')$ (resp., $\theta(ww') = \theta(w')\theta(w)$):
  Example: Watson-Crick transformation.

  $d_\theta(x, x') = 0$ iff. $x' = x$,
  $d_\theta(x, x') = 1$ iff. $x' = \theta(x)$ (in general $x \neq \theta(x')$),
  $d_\theta(x, x') = 2$ otherwise.

What (quasi)metric to fix ?

- Hamming, Levenshtein metrics *[Néraud, 2020]*
- d_P = prefix metric $d_P(x, x') = |x| + |x'| - 2|p|$,
  $p \in P(x) \cap P(x')$ and $|p|$ maximum
- d_F = factor metric $d_F(x, x') = |x| + |x'| - 2|f|$,
  $f \in F(x) \cap F(x')$ and $|f|$ maximum
- θ automorphism (resp., anti-automorphism) onto A^*:
  $\theta(A) = A$, $\theta(ww') = \theta(w)\theta(w')$ (resp., $\theta(ww') = \theta(w')\theta(w)$:
  Example: Watson-Crick transformation.

  $d_\theta(x, x') = 0$ iff. $x' = x$,
  $d_\theta(x, x') = 1$ iff. $x' = \theta(x)$ (in general $x \neq \theta(x')$),
  $d_\theta(x, x') = 2$ otherwise.

Wrt. each of the preceding metrics, $\tau_{d,k} \subseteq A^* \times A^*$ is the behavior of some finite automaton (finite transducer) with arrows labeled in $A^* \times A^*$: $\tau_{d,k}$ is a regular relation.

However this does not guarantee that $\underline{\tau_{d,k}} = \tau_{d,k} \cap \overline{id_{A^*}}$ is regular.

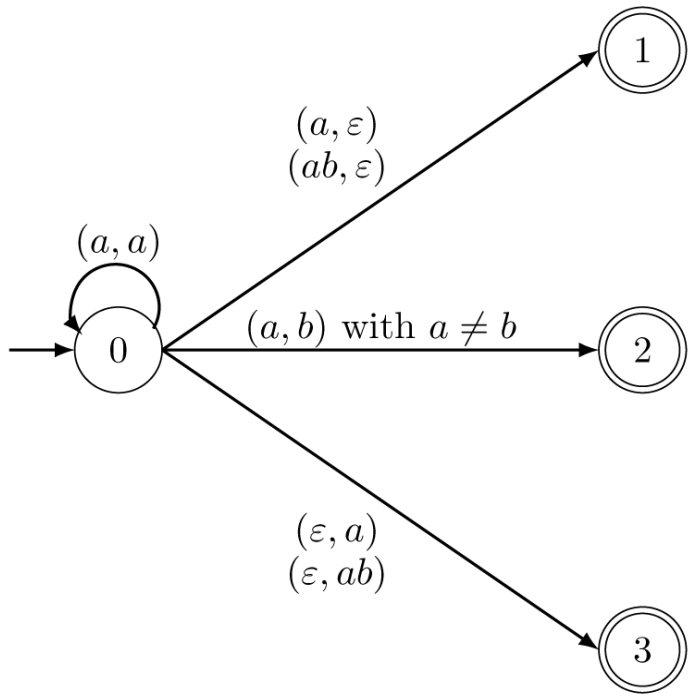
Framework of the prefix metric

$d_P(x, x') = |x| + |x'| - 2|p'|$ with $p \in P(x) \cap P(x')$, $|p|$ maximum;
$(x, y) \in \underline{\tau_{d_P,k}}$ iff. $1 \leq d_P(x, y) \leq k$.

Framework of the prefix metric

$d_P(x, x') = |x| + |x'| - 2|p'|$ with $p \in P(x) \cap P(x')$, $|p|$ maximum;
$(x, y) \in \underline{\tau_{d_P,k}}$ iff. $1 \leq d_P(x, y) \leq k$.

$A = \{a, b\}$, $k = 2$:



$\Rightarrow \underline{\tau_{d_P,k}}$ *is a regular relation.*

- (c1) Error detection condition can be decided
  $\underline{\tau_{d_P,k}}(X) \cap X = \emptyset$.
- (c2) Error correction condition can be decided
  Equivalently: $(X \times X) \cap \underline{\tau_{d_P,2k}} = \emptyset$.
- (c4) It can be decided whether $Y = X \cup \tau_{d_P,k}(X)$ is a variable-length code.

  $\rightarrow$ By applying Sardinas and Patterson algorithm to Y.
  $U_0 = Y^{-1}Y$, $U_{n+1} = Y^{-1}U_n \cup U_n^{-1}Y$: $(\exists i, j)\ i < j, U_i = U_j$
  Y is a code iff. $(\forall n \geq 0)\ \varepsilon \notin U_n$.

- (c3) It can be decided whether X is maximal in the family of $\tau_{d_P,k}$-independent codes.

Proposition *X regular $\tau_{d_P,k}$-independent code.*
(i) (c3) X maximal as a $\tau_{d_P,k}$-independent codes,
(ii) iff. X maximal code in A^*,
(iii) iff. X complete : $A^* = F(X^*)$,
(iv) iff. $\pi(X) = 1$.

(one can decide whether Cond. (iv) is satisfied)

- (c3) It can be decided whether X is maximal in the family of $\tau_{d_P,k}$-independent codes.

Proposition *X regular $\tau_{d_P,k}$-independent code.*
(i) (c3) X maximal as a $\tau_{d_P,k}$-independent codes,
(ii) iff. X maximal code in A^*,
(iii) iff. X complete : $A^* = F(X^*)$,
(iv) iff. $\pi(X) = 1$.

(one can decide whether Cond. (iv) is satisfied)

Lemma *Every non-complete regular $\tau_{d_P,k}$-independent code can be embedded into some complete one.*

[Ehrenfeucht and Rozenberg]
$\widehat{X} = X \cup Y$, with $Y = (zU)^* z$, $U = A^* \setminus (X^* \cup A^* z A^*)$, and $z = z_0 ab^{|z_0|}$ ($z_0 \in A^* \setminus F(X^*)$, $|z_0| \geq k$).

Framework of the factor metric

$d_F(x, x') = |x| + |x'| - 2|f|$, $f \in F(x) \cap F(x')$, $|f|$ maximum;
$(x, y) \in \underline{\tau_{d_P,k}}$ iff. $1 \leq d_P(x, y) \leq k$.

$\tau_{d_F,k}$ is regular,
but we do not know whether $\underline{\tau_{d_F,k}}$ is regular or not.

- We do not know whether Conds. (c1), (c2) can be decided for a regular code X (X finite code: OK).

Framework of the factor metric

$d_F(x, x') = |x| + |x'| - 2|f|$, $f \in F(x) \cap F(x')$, $|f|$ maximum;
$(x, y) \in \underline{\tau_{d_P,k}}$ iff. $1 \leq d_P(x, y) \leq k$.

$\tau_{d_F,k}$ is regular,
but we do not know whether $\underline{\tau_{d_F,k}}$ is regular or not.

- We do not know whether Conds. (c1), (c2) can be decided for a regular code X (X finite code: OK).
- One can decide whether X satisfies (c4): $X \cup \tau_{d_F,k}(X)$ is a code.

Framework of the factor metric

$d_F(x, x') = |x| + |x'| - 2|f|$, $f \in F(x) \cap F(x')$, $|f|$ maximum;
$(x, y) \in \underline{\tau_{d_P,k}}$ iff. $1 \leq d_P(x, y) \leq k$.

$\tau_{d_F,k}$ is regular,
but we do not know whether $\underline{\tau_{d_F,k}}$ is regular or not.

- We do not know whether Conds. (c1), (c2) can be decided for a regular code X (X finite code: OK).
- One can decide whether X satisfies (c4): $X \cup \tau_{d_F,k}(X)$ is a code.
- One can decide whether X satisfies (c3):

  Proposition *X regular $\underline{\tau_{d_F,k}}$-independent code.*
  (i) X satisfies (c3)
  (ii) iff. $\pi(X) = 1$.

  Lemma *Every non-complete regular $\underline{\tau_{d_F,k}}$-independent code can be embedded into some complete one.*

  $X_1 = X \cup Y_1$, $Y_1 = (z_1 U_1)^* z_1$, $U_1 = A^* \setminus (X^* \cup A^* z_1 A^*)$,
  $z_1 = a^{|z|} bz = a^{2|z_0|+1} bz_0 ab^{|z_0|} \notin F(X^*)$.

Quasi-metrics associated to (anti-)automorphisms

$d_\theta(x, x') = 0$ iff. $x' = x$,
$d_\theta(x, x') = 1$ iff. $x' = \theta(x)$,
$d_\theta(x, x') = 2$ otherwise.
$\rightarrow k = 1$,
$\rightarrow \tau_{d_\theta,1}(x) = \{x\} \cup \{\theta(x)\}, \underline{\tau_{d_\theta,1}}(x) = \{\underline{\theta}(x)\}$.

– In any case X satisfies Conds. (c1), (c2),
– One can decide whether X satisfies Conds. (c3), (c4).

Lemma *Every non-complete regular $\underline{\tau_{d_\theta,1}}$-independent code X can be embedded into some complete one.*

$X \cup Y_2$, $Y_2 = (z_2U_2)^*z_2$, $U_2 = A^* \setminus (A^*z_2A^* \cup X^*)$,
$z_2 = z_0\theta(z_0)\cdots\theta^{n-1}(z_0)ab^{n|z_0|} \notin F(X^*)$, with $\theta^n = id_{A^*}$.

Thank you for your attention.