TWO-SIDED LOCALLY TESTABLE LANGUAGES

Martin Kutrib[1], Friedrich Otto[2]

[1]Fachbereich Informatik, Universität Giessen
Arndtstr. 2, D-35392 Giessen
kutrib@informatik.uni-giessen.de

[2]Charles University, Faculty of Mathematics and Physics
Malostranké nám. 25, CZ-11800 Prague 1
otto@ktiml.mff.cuni.cz

NCMA 2018

Košice

August 21–22, 2018

Overview

1. Introduction
2. Definitions and Preliminaries
3. Closure Properties
4. Expressive Capacity
5. Conclusion

1. Introduction

In 1971 McNaughton and Papert introduced the

strictly locally testable and the locally testable languages.

These languages have received much attention in the literature because of their elegance and simplicity.

A language $L \subseteq \Sigma^*$ is strictly k-testable iff

$$\exists A, B, C \subseteq \Sigma^k : L \cap \Sigma^{\geq k} = (A \cdot \Sigma^* \cap \Sigma^* \cdot B) \setminus (\Sigma^+ \cdot (\Sigma^* \setminus C) \cdot \Sigma^+).$$

L is strictly locally testable if it is strictly k-testable for some $k \geq 1$.

For $w = a_1 a_2 \ldots a_n \in \Sigma^n$, $1 \leq i \leq n$, and $1 \leq k \leq n - i + 1$,

$$w[i,k] = a_i a_{i+1} \ldots a_{i+k-1},$$

that is, $w[i,k]$ is the factor of w of length k that starts with a_i.

For $w = a_1 a_2 \dots a_n \in \Sigma^n$, $1 \leq i \leq n$, and $1 \leq k \leq n - i + 1$,

$$w[i, k] = a_i a_{i+1} \dots a_{i+k-1},$$

that is, $w[i, k]$ is the factor of w of length k that starts with a_i.

For $k \geq 1$ and $w \in \Sigma^{\geq k}$,

$$\begin{array}{lcll} P_k(w) & = & w[1, k] & \text{is the prefix of } w \text{ of length } k, \\ S_k(w) & = & w[n-k+1, k] & \text{is the suffix of } w \text{ of length } k, \text{ and} \\ I_k(w) & = & \{\, w[i, k] \mid 2 \leq i \leq |w| - k \,\} & \text{are the infixes of } w \text{ of length } k. \end{array}$$

For $w = a_1 a_2 \ldots a_n \in \Sigma^n$, $1 \le i \le n$, and $1 \le k \le n - i + 1$,

$$w[i,k] = a_i a_{i+1} \ldots a_{i+k-1},$$

that is, $w[i,k]$ is the factor of w of length k that starts with a_i.

For $k \ge 1$ and $w \in \Sigma^{\ge k}$,

$$\begin{array}{lcll} P_k(w) & = & w[1,k] & \text{is the prefix of } w \text{ of length } k, \\ S_k(w) & = & w[n-k+1,k] & \text{is the suffix of } w \text{ of length } k, \text{ and} \\ I_k(w) & = & \{\, w[i,k] \mid 2 \le i \le |w| - k \,\} & \text{are the infixes of } w \text{ of length } k. \end{array}$$

A language $L \subseteq \Sigma^*$ is k-testable iff, for all $u, v \in \Sigma^{\ge k}$,

if $P_k(u) = P_k(v) \wedge S_k(u) = S_k(v) \wedge I_k(u) = I_k(v)$, then $(u \in L \Leftrightarrow v \in L)$.

L is locally testable if it is locally k-testable for some $k \ge 1$.

For $w = a_1 a_2 \ldots a_n \in \Sigma^n$, $1 \leq i \leq n$, and $1 \leq k \leq n - i + 1$,

$$w[i,k] = a_i a_{i+1} \ldots a_{i+k-1},$$

that is, $w[i,k]$ is the factor of w of length k that starts with a_i.

For $k \geq 1$ and $w \in \Sigma^{\geq k}$,

$$\begin{array}{lcll} P_k(w) & = & w[1,k] & \text{is the prefix of } w \text{ of length } k, \\ S_k(w) & = & w[n-k+1,k] & \text{is the suffix of } w \text{ of length } k, \text{ and} \\ I_k(w) & = & \{\, w[i,k] \mid 2 \leq i \leq |w| - k \,\} & \text{are the infixes of } w \text{ of length } k. \end{array}$$

A language $L \subseteq \Sigma^*$ is k-testable iff, for all $u, v \in \Sigma^{\geq k}$,

if $P_k(u) = P_k(v) \wedge S_k(u) = S_k(v) \wedge I_k(u) = I_k(v)$, then $(u \in L \Leftrightarrow v \in L)$.

L is locally testable if it is locally k-testable for some $k \geq 1$.

By LT(k) (SLT(k)) we denote the class of (strictly) k-testable languages, and LT (SLT) is the class of (strictly) locally testable languages.

A DFA can check membership of a word w, $|w| \geq k$, in a (strictly) k-testable language by scanning w from left to right, just remembering the prefix, the inner factors, and the suffix of length k that it encounters.

A DFA can check membership of a word w, $|w| \geq k$, in a (strictly) k-testable language by scanning w from left to right, just remembering the prefix, the inner factors, and the suffix of length k that it encounters.

In the literature many devices have been studied that scan their inputs in a more flexible way, e.g., two-head automata, where the heads scan the input starting at the two ends of the word (Rosenberg 1967).

A DFA can check membership of a word w, $|w| \geq k$, in a (strictly) k-testable language by scanning w from left to right, just remembering the prefix, the inner factors, and the suffix of length k that it encounters.

In the literature many devices have been studied that scan their inputs in a more flexible way, e.g., two-head automata, where the heads scan the input starting at the two ends of the word (Rosenberg 1967).

At NCMA 2017 we presented a generalization of the strictly locally testable languages that corresponds to such a two-way scanning, the two-sided strictly locally testable languages.

Here we further extend these considerations to the two-sided locally testable languages.

2. Definitions and Preliminaries

As the class SLT(k) is closed under intersection, using two windows of size k to scan a given word simultaneously from left to right and from right to left would just yield another strictly k-testable language. Hence, the two factors scanned concurrently must be put into some relation.

Definition 1

Let $k \geq 1$, let Σ be an alphabet, and let $R \subseteq \Sigma^k \times \Sigma^k$ be a symmetric relation. By $\Sigma_R^{\geq k}$ we denote the set of all words $w \in \Sigma^{\geq k}$ that satisfy the following condition:

$$\forall i \in \{1, 2, \ldots, |w| - k + 1\}, (w[i, k], w([|w| + 2 - k - i, k]) \in R.$$

We call these words the R-symmetric words.

Definition 2

A language $L \subseteq \Sigma^*$ *is two-sided strictly* k*-testable if*

$\exists A, B, C \subseteq \Sigma^k \,\exists$ *a symmetric relation* $R \subseteq \Sigma^k \times \Sigma^k$:

$\forall w \in L \cap \Sigma^{\geq k} : P_k(w) \in A \wedge S_k(w) \in B \wedge I_k(w) \subseteq C \wedge w \in \Sigma_R^{\geq k}$.

Definition 2

A language $L \subseteq \Sigma^*$ *is two-sided strictly* k*-testable if*

$\exists A, B, C \subseteq \Sigma^k \; \exists$ *a symmetric relation* $R \subseteq \Sigma^k \times \Sigma^k$:

$\forall w \in L \cap \Sigma^{\geq k} : P_k(w) \in A \wedge S_k(w) \in B \wedge I_k(w) \subseteq C \wedge w \in \Sigma_R^{\geq k}$.

By $L(A, B, C, R)$ *we denote the set of all words of length at least* k *that satisfy the conditions above. Hence,* $L = L(A, B, C, R) \cup F_L$*, where* $F_L \subseteq \Sigma^{\leq k-1}$.

Definition 2

A language $L \subseteq \Sigma^*$ is two-sided strictly k-testable if

$\exists A, B, C \subseteq \Sigma^k \; \exists$ *a symmetric relation* $R \subseteq \Sigma^k \times \Sigma^k$:

$\forall w \in L \cap \Sigma^{\geq k} : P_k(w) \in A \wedge S_k(w) \in B \wedge I_k(w) \subseteq C \wedge w \in \Sigma_R^{\geq k}$.

By $L(A, B, C, R)$ we denote the set of all words of length at least k that satisfy the conditions above. Hence, $L = L(A, B, C, R) \cup F_L$, where $F_L \subseteq \Sigma^{\leq k-1}$.

A language $L \subseteq \Sigma^*$ is two-sided strictly locally testable if it is two-sided strictly k-testable for some $k \geq 1$.

By $\mathrm{2SLT}(k)$ ($\mathrm{2SLT}$) we denote the class of two-sided strictly (k-) testable languages. If the binary relation R is to be emphasized, then we write the corresponding class as $\mathrm{2SLT}_R(k)$.

Example 1

Let $\Sigma = \{a, b\}$, and let $A = \{a, b\} = B = C$.
The triple (A, B, C) defines the strictly 1-testable language $L = L(A, B, C) = \Sigma^+$.

Let $R = \{(a, b), (b, a)\} \subset \Sigma \times \Sigma$ and let $L' = L(A, B, C, R)$ be the resulting two-sided strictly 1-testable language.

Then $w \in \Sigma^+$ belongs to L' iff, for all $i = 1, 2, \ldots, |w|$, the i-th letter from the left differs from the i-th letter from the right.

Thus, L' is not even regular, as $L' \cap (a^* \cdot b^*) = \{\, a^n b^n \mid n \geq 1 \,\}$ holds.
Hence, L' is in particular not locally testable.

Definition 3

Let $k \geq 1$, Σ an alphabet, and $R \subseteq \Sigma^k \times \Sigma^k$ a symmetric relation.

1. *A language $L \subseteq \Sigma^*$ is k-R-testable, if $L \cap \Sigma^{\geq k} \subseteq \Sigma_R^{\geq k}$, and the following conditions are met for all $x, y \in \Sigma_R^{\geq k}$:*

$$\text{if } P_k(x) = P_k(y) \wedge S_k(x) = S_k(y) \wedge I_k(x) = I_k(y), \\ \text{then } (x \in L \iff y \in L).$$

Definition 3

Let $k \geq 1$, Σ an alphabet, and $R \subseteq \Sigma^k \times \Sigma^k$ a symmetric relation.

1. *A language $L \subseteq \Sigma^*$ is k-R-testable, if $L \cap \Sigma^{\geq k} \subseteq \Sigma_R^{\geq k}$, and the following conditions are met for all $x, y \in \Sigma_R^{\geq k}$:*

$$\begin{array}{l} \text{if } P_k(x) = P_k(y) \wedge S_k(x) = S_k(y) \wedge I_k(x) = I_k(y), \\ \text{then } (x \in L \iff y \in L). \end{array}$$

2. *A language $L \subseteq \Sigma^*$ is called two-sided k-testable if there exists a symmetric relation $R \subseteq \Sigma^k \times \Sigma^k$ such that L is k-R-testable.*

Definition 3

Let $k \geq 1$, Σ an alphabet, and $R \subseteq \Sigma^k \times \Sigma^k$ a symmetric relation.

1. *A language $L \subseteq \Sigma^*$ is k-R-testable, if $L \cap \Sigma^{\geq k} \subseteq \Sigma_R^{\geq k}$, and the following conditions are met for all $x, y \in \Sigma_R^{\geq k}$:*

   $$\begin{array}{l} \text{if } P_k(x) = P_k(y) \wedge S_k(x) = S_k(y) \wedge I_k(x) = I_k(y), \\ \text{then } (x \in L \iff y \in L). \end{array}$$

2. *A language $L \subseteq \Sigma^*$ is called two-sided k-testable if there exists a symmetric relation $R \subseteq \Sigma^k \times \Sigma^k$ such that L is k-R-testable.*
3. *A language $L \subseteq \Sigma^*$ is called two-sided locally testable if it is two-sided k-testable for some $k \geq 1$.*

Example 2

The language $L = \{\, ad^nb, ae^nc \mid n \geq 1 \,\} \in 2\mathrm{SLT}(2) \smallsetminus 2\mathrm{SLT}(1)$.
However, this language is two-sided 1-testable.

For $R = \{(a,b),(b,a),(a,c),(c,a),(d,d),(e,e)\}$, we have

$$\Sigma_R^{\geq 1} = \{\, w \in \Sigma^{\geq 1} \mid \forall i = 1,\ldots,|w|: (w[i] = a \text{ iff } w[|w|+1-i] \in \{b,c\}) \text{ and if } w[i] \in \{d,e\}, \text{ then } w[i] = w[|w|+1-i] \,\}.$$

Example 2

The language $L = \{\, ad^n b, ae^n c \mid n \geq 1 \,\} \in 2\mathrm{SLT}(2) \smallsetminus 2\mathrm{SLT}(1)$.
However, this language is two-sided 1-testable.

For $R = \{(a,b),(b,a),(a,c),(c,a),(d,d),(e,e)\}$, we have

$$\Sigma_R^{\geq 1} = \{\, w \in \Sigma^{\geq 1} \mid \forall i = 1,\ldots,|w| \colon (w[i] = a \text{ iff } w[|w|+1-i] \in \{b,c\}) \text{ and if } w[i] \in \{d,e\}, \text{ then } w[i] = w[|w|+1-i] \,\}.$$

Now $L = \{\, w \in \Sigma_R^{\geq 1} \mid \quad P_1(w) = a \wedge ((I_1(w) = \{d\} \wedge S_1(w) = b) \vee (I_1(w) = \{e\} \wedge S_1(w) = c)) \,\}.$

Example 2

The language $L = \{\, ad^nb, ae^nc \mid n \geq 1 \,\} \in 2\mathrm{SLT}(2) \setminus 2\mathrm{SLT}(1)$.
However, this language is two-sided 1-testable.

For $R = \{(a,b),(b,a),(a,c),(c,a),(d,d),(e,e)\}$, we have

$$\Sigma_R^{\geq 1} = \{\, w \in \Sigma^{\geq 1} \mid \forall i = 1,\ldots,|w| : (w[i] = a \text{ iff } w[|w|+1-i] \in \{b,c\}) \text{ and if } w[i] \in \{d,e\}, \text{ then } w[i] = w[|w|+1-i] \,\}.$$

Now $L = \{\, w \in \Sigma_R^{\geq 1} \mid P_1(w) = a \wedge ((I_1(w) = \{d\} \wedge S_1(w) = b) \vee (I_1(w) = \{e\} \wedge S_1(w) = c)) \,\}.$

Actually, L is the union of the two-sided strictly 1-testable languages

$$L_1 = \{\, ad^nb \mid n \geq 1 \,\} \text{ and } L_2 = \{\, ae^nc \mid n \geq 1 \,\}.$$

In [NCMA 2017] we have shown that $\mathsf{SLT}(k) \subsetneq \mathsf{2SLT}(k)$ and that $\mathsf{2SLT}(k) \subsetneq \mathsf{2SLT}(k+1)$ for all $k \geq 1$.

Also it is known that $\mathsf{LT}(k) \subsetneq \mathsf{LT}(k+1)$.

In [NCMA 2017] we have shown that $\mathsf{SLT}(k) \subsetneq \mathsf{2SLT}(k)$ and that $\mathsf{2SLT}(k) \subsetneq \mathsf{2SLT}(k+1)$ for all $k \geq 1$.

Also it is known that $\mathsf{LT}(k) \subsetneq \mathsf{LT}(k+1)$.

For all $k \geq 1$, let L_k be the finite language $L_k = \{a^k, a^{k+1}\}$.

Lemma 4

For all $k \geq 1$, $L_{k+1} \in \mathsf{LT}(k+1) \smallsetminus \mathsf{2LT}(k)$.

In [NCMA 2017] we have shown that $\mathsf{SLT}(k) \subsetneq \mathsf{2SLT}(k)$ and that $\mathsf{2SLT}(k) \subsetneq \mathsf{2SLT}(k+1)$ for all $k \geq 1$.

Also it is known that $\mathsf{LT}(k) \subsetneq \mathsf{LT}(k+1)$.

For all $k \geq 1$, let L_k be the finite language $L_k = \{a^k, a^{k+1}\}$.

Lemma 4

For all $k \geq 1$, $L_{k+1} \in \mathsf{LT}(k+1) \smallsetminus \mathsf{2LT}(k)$.

In [NCMA 2017] we have shown that $\mathsf{SLT}(k) \subsetneq \mathsf{2SLT}(k)$ and that $\mathsf{2SLT}(k) \subsetneq \mathsf{2SLT}(k+1)$ for all $k \geq 1$.

Also it is known that $\mathsf{LT}(k) \subsetneq \mathsf{LT}(k+1)$.

For all $k \geq 1$, let L_k be the finite language $L_k = \{a^k, a^{k+1}\}$.

Lemma 4

For all $k \geq 1$, $L_{k+1} \in \mathsf{LT}(k+1) \smallsetminus \mathsf{2LT}(k)$.

Corollary 5

1. *The (two-sided) k-testable languages form an infinite ascending hierarchy with respect to the parameter k.*
2. *For all $k \geq 1$, $\mathsf{LT}(k) \subsetneq \mathsf{2LT}(k)$, but $\mathsf{2LT}(k) = \mathsf{SLT}(k)$ for unary languages.*

3. Closure Properties

Let $k \geq 1$, Σ an alphabet, and $R \subseteq \Sigma^k \times \Sigma^k$ a symmetric relation.
For all words $u, v \in \Sigma^k$ and all subsets $C \subseteq \Sigma^k$, we define

$$L_R(u, v, C) = \{\, w \in \Sigma_R^{\geq k} \mid P_k(w) = u, S_k(w) = v, \text{ and } I_k(w) = C \,\},$$

and for a k-R-testable language $L \subseteq \Sigma^*$, we define

$$\text{triple}(L) = \{\, (u, v, C) \mid u, v \in \Sigma^k \text{ and } C \subseteq \Sigma^k \text{ s.t. } L_R(u, v, C) \cap L \neq \emptyset \,\}.$$

3. Closure Properties

Let $k \geq 1$, Σ an alphabet, and $R \subseteq \Sigma^k \times \Sigma^k$ a symmetric relation.
For all words $u, v \in \Sigma^k$ and all subsets $C \subseteq \Sigma^k$, we define

$$L_R(u, v, C) = \{\, w \in \Sigma_R^{\geq k} \mid P_k(w) = u, S_k(w) = v, \text{ and } I_k(w) = C \,\},$$

and for a k-R-testable language $L \subseteq \Sigma^*$, we define

$$\text{triple}(L) = \{\, (u, v, C) \mid u, v \in \Sigma^k \text{ and } C \subseteq \Sigma^k \text{ s.t. } L_R(u, v, C) \cap L \neq \emptyset \,\}.$$

If $(u, v, C) \in \text{triple}(L)$, then there is a word $w \in \Sigma_R^{\geq k}$ with $P_k(w) = u$, $S_k(w) = v$, $I_k(w) = C$ that belongs to L and to $L_R(u, v, C)$.

3. Closure Properties

Let $k \geq 1$, Σ an alphabet, and $R \subseteq \Sigma^k \times \Sigma^k$ a symmetric relation.
For all words $u, v \in \Sigma^k$ and all subsets $C \subseteq \Sigma^k$, we define

$$L_R(u, v, C) = \{ w \in \Sigma_R^{\geq k} \mid P_k(w) = u, S_k(w) = v, \text{ and } I_k(w) = C \},$$

and for a k-R-testable language $L \subseteq \Sigma^*$, we define

$$\text{triple}(L) = \{ (u, v, C) \mid u, v \in \Sigma^k \text{ and } C \subseteq \Sigma^k \text{ s.t. } L_R(u, v, C) \cap L \neq \emptyset \}.$$

If $(u, v, C) \in \text{triple}(L)$, then there is a word $w \in \Sigma_R^{\geq k}$ with $P_k(w) = u$, $S_k(w) = v$, $I_k(w) = C$ that belongs to L and to $L_R(u, v, C)$.

Since L is k-R-testable, all words $w' \in \Sigma_R^{\geq k}$ with $P_k(w') = u$, $S_k(w') = v$, $I_k(w') = C$ belong to L as well.

We conclude that $L_R(u, v, C) \subseteq L$ if $(u, v, C) \in \text{triple}(L)$.

3. Closure Properties

Let $k \geq 1$, Σ an alphabet, and $R \subseteq \Sigma^k \times \Sigma^k$ a symmetric relation. For all words $u, v \in \Sigma^k$ and all subsets $C \subseteq \Sigma^k$, we define

$$L_R(u, v, C) = \{ w \in \Sigma_R^{\geq k} \mid P_k(w) = u, S_k(w) = v, \text{ and } I_k(w) = C \},$$

and for a k-R-testable language $L \subseteq \Sigma^*$, we define

$$\text{triple}(L) = \{ (u, v, C) \mid u, v \in \Sigma^k \text{ and } C \subseteq \Sigma^k \text{ s.t. } L_R(u, v, C) \cap L \neq \emptyset \}.$$

If $(u, v, C) \in \text{triple}(L)$, then there is a word $w \in \Sigma_R^{\geq k}$ with $P_k(w) = u$, $S_k(w) = v$, $I_k(w) = C$ that belongs to L and to $L_R(u, v, C)$.

Since L is k-R-testable, all words $w' \in \Sigma_R^{\geq k}$ with $P_k(w') = u$, $S_k(w') = v$, $I_k(w') = C$ belong to L as well.

We conclude that $L_R(u, v, C) \subseteq L$ if $(u, v, C) \in \text{triple}(L)$.

Thus,

$$L = F_L \cup \bigcup_{(u,v,C) \in \text{triple}(L)} L_R(u, v, C),$$

where $F_L = \{ w \in L \mid |w| \leq k - 1 \}$.

Lemma 6

Let $k \geq 1$, let Σ be an alphabet, and let $R \subseteq \Sigma^k \times \Sigma^k$ be a symmetric relation. If $L = \bigcup_{i=1}^{m} L_R(u_i, v_i, C_i)$, then L belongs to $2\mathrm{LT}_R(k)$.

Lemma 6

Let $k \geq 1$, let Σ be an alphabet, and let $R \subseteq \Sigma^k \times \Sigma^k$ be a symmetric relation. If $L = \bigcup_{i=1}^{m} L_R(u_i, v_i, C_i)$, then L belongs to $2\mathrm{LT}_R(k)$.

If $L \subseteq \Sigma^*$ is a two-sided strictly k-testable language, then there exist a symmetric binary relation R on Σ^k, sets $A, B, C \subseteq \Sigma^k$, and a finite subset $F \subseteq \Sigma^{\leq k-1}$ such that

$$L = F \cup \{\, w \in \Sigma_R^{\geq k} \mid P_k(w) \in A, S_k(w) \in B, \text{ and } I_k(w) \subseteq C \,\}.$$

Hence, L can be written as

$$L = F \cup \bigcup_{u \in A, v \in B, C' \subseteq C} L_R(u, v, C'),$$

which implies that L is k-R-testable.

Lemma 6

Let $k \geq 1$, let Σ be an alphabet, and let $R \subseteq \Sigma^k \times \Sigma^k$ be a symmetric relation. If $L = \bigcup_{i=1}^{m} L_R(u_i, v_i, C_i)$, then L belongs to $2\mathrm{LT}_R(k)$.

If $L \subseteq \Sigma^*$ is a two-sided strictly k-testable language, then there exist a symmetric binary relation R on Σ^k, sets $A, B, C \subseteq \Sigma^k$, and a finite subset $F \subseteq \Sigma^{\leq k-1}$ such that

$$L = F \cup \{\, w \in \Sigma_R^{\geq k} \mid P_k(w) \in A, S_k(w) \in B, \text{ and } I_k(w) \subseteq C \,\}.$$

Hence, L can be written as

$$L = F \cup \bigcup_{u \in A, v \in B, C' \subseteq C} L_R(u, v, C'),$$

which implies that L is k-R-testable.

Corollary 1

For all $k \geq 1$, $2\mathrm{SLT}(k) \subseteq 2\mathrm{LT}(k)$ and $2\mathrm{SLT} \subseteq 2\mathrm{LT}$.

Now we turn to the Boolean operations union and intersection and a variant of the operation of complementation.

Let $k \geq 1$, Σ an alphabet, and $R \subseteq \Sigma^k \times \Sigma^k$ a symmetric relation. For a k-R-testable language $L \subseteq \Sigma^*$, the R-complement is the set

$$L_R^c = (\Sigma^{\leq k-1} \cap L^c) \cup (\Sigma_R^{\geq k} \cap L^c).$$

It is easily seen that $L^c = L_R^c \cup (\Sigma^{\geq k} \setminus \Sigma_R^{\geq k})$, that is, the complement L^c of L differs from the R-complement L_R^c by the words of length at least k that are not R-symmetric.

Now we turn to the Boolean operations union and intersection and a variant of the operation of complementation.

Let $k \geq 1$, Σ an alphabet, and $R \subseteq \Sigma^k \times \Sigma^k$ a symmetric relation. For a k-R-testable language $L \subseteq \Sigma^*$, the R-complement is the set

$$L_R^c = (\Sigma^{\leq k-1} \cap L^c) \cup (\Sigma_R^{\geq k} \cap L^c).$$

It is easily seen that $L^c = L_R^c \cup (\Sigma^{\geq k} \setminus \Sigma_R^{\geq k})$, that is, the complement L^c of L differs from the R-complement L_R^c by the words of length at least k that are not R-symmetric.

Proposition 7

Let $k \geq 1$, let Σ be an alphabet, and let $R \subseteq \Sigma^k \times \Sigma^k$ be a symmetric relation. Then the family $2\mathrm{LT}_R(k)$ is closed under the Boolean operations intersection and union and under the operation of R-complementation.

Proof of Prop. 7.

Let $L, L_1, L_2 \subseteq \Sigma^*$ be k-R-testable languages.

The union $L_1 \cup L_2$ is represented by

$$F_{L_1} \cup F_{L_2} \cup \bigcup_{(u,v,C)\in(\text{triple}(L_1)\,\cup\,\text{triple}(L_2))} L_R(u, v, C).$$

By Lemma 6 it follows that $L_1 \cup L_2 \in 2\text{LT}_R(k)$.

Proof of Prop. 7.

Let $L, L_1, L_2 \subseteq \Sigma^*$ be k-R-testable languages.

The union $L_1 \cup L_2$ is represented by

$$F_{L_1} \cup F_{L_2} \cup \bigcup_{(u,v,C)\in(\text{triple}(L_1)\cup\text{triple}(L_2))} L_R(u, v, C).$$

By Lemma 6 it follows that $L_1 \cup L_2 \in 2\text{LT}_R(k)$.

Similarly, the intersection $L_1 \cap L_2$ is represented by

$$(F_{L_1} \cap F_{L_2}) \cup \bigcup_{(u,v,C)\in(\text{triple}(L_1)\cap\text{triple}(L_2))} L_R(u, v, C),$$

which shows that $L_1 \cap L_2 \in 2\text{LT}_R(k)$.

Proof of Prop. 7.

Let $L, L_1, L_2 \subseteq \Sigma^*$ be k-R-testable languages.

The union $L_1 \cup L_2$ is represented by

$$F_{L_1} \cup F_{L_2} \cup \bigcup_{(u,v,C)\in(\mathrm{triple}(L_1)\cup\mathrm{triple}(L_2))} L_R(u, v, C).$$

By Lemma 6 it follows that $L_1 \cup L_2 \in 2\mathrm{LT}_R(k)$.

Similarly, the intersection $L_1 \cap L_2$ is represented by

$$(F_{L_1} \cap F_{L_2}) \cup \bigcup_{(u,v,C)\in(\mathrm{triple}(L_1)\cap\mathrm{triple}(L_2))} L_R(u, v, C),$$

which shows that $L_1 \cap L_2 \in 2\mathrm{LT}_R(k)$.

Finally, for the R-complement L_R^c we have

$$L_R^c = \left(\Sigma^{\leq k-1} \cup \Sigma_R^{\geq k}\right) \setminus L = (\Sigma^{\leq k-1} \setminus F_L) \cup \left(\bigcup_{(u,v,C)\notin\mathrm{triple}(L)} L_R(u, v, C) \right).$$

Since the number of triples (u, v, C) that do not belong to $\mathrm{triple}(L)$ is finite, Lemma 6 shows that $L_R^c \in 2\mathrm{LT}_R(k)$. $\square$

Proposition 8

For all $k \geq 1$, the families $2\mathrm{LT}(k)$ and $2\mathrm{LT}$ are closed under intersection and R-complementation.

Proposition 8

For all $k \geq 1$, the families $2\mathrm{LT}(k)$ and $2\mathrm{LT}$ are closed under intersection and R-complementation.

To establish closure under intersection, we observed that $\Sigma^{\geq k}_{R_1 \cap R_2} = \Sigma^{\geq k}_{R_1} \cap \Sigma^{\geq k}_{R_2}$ for any $k \geq 1$ and any two binary relations R_1, R_2 on Σ^k. A similar argument does not hold for union, as $\Sigma^{\geq k}_{R_1} \cup \Sigma^{\geq k}_{R_2}$ is in general a proper subset of $\Sigma^{\geq k}_{R_1 \cup R_2}$.

Example 3.

Let $\Sigma = \{a, b\}$, $k = 1$, $R_1 = \{(a, a)\}$ and $R_2 = \{(b, b)\}$.
Then $\Sigma^{\geq 1}_{R_1} = a^+$ and $\Sigma^{\geq 1}_{R_2} = b^+$.
However, $R_1 \cup R_2 = \{(a, a), (b, b)\}$ and

$$\Sigma^{\geq 1}_{R_1 \cup R_2} = \{\, w \in \Sigma^+ \mid \forall i = 1, 2, \ldots, |w| \colon w[i] = w[|w| + 1 - i] \,\},$$

which also contains the word $w = abba \notin \Sigma^{\geq 1}_{R_1} \cup \Sigma^{\geq 1}_{R_2}$.

Proposition 9

For all $k \geq 1$*, the families* $2\mathrm{LT}(k)$ *and* $2\mathrm{LT}$ *are not closed under union.*

Proposition 9

For all $k \geq 1$, the families $2\mathrm{LT}(k)$ and $2\mathrm{LT}$ are not closed under union.

Proof.

Let $k = 1$, $\Sigma = \{a, b, c\}$, $R_1 = \{(a, a), (b, c), (c, b)\}$ and

$$L_1 = \{\, w \in \Sigma_{R_1}^{\geq 1} \mid P_1(w), S_1(w) \in \{a, b, c\}, I_1(w) \subseteq \{a, b, c\} \,\}$$
$$= \{\, w \in \{a, b, c\}^* \mid \forall i = 1, \ldots, |w| \colon (w[i] = a \Rightarrow w[|w| + 1 - i] = a)$$
$$\wedge (w[i] = b \Rightarrow w[|w| + 1 - i] = c)$$
$$\wedge (w[i] = c \Rightarrow w[|w| + 1 - i] = b) \,\}.$$

Further, let $R_2 = \{(a, a), (b, b), (c, c)\}$ and

$$L_2 = \{\, w \in \Sigma_{R_2}^{\geq 1} \mid P_1(w), S_1(w) \in \{a, b, c\}, I_1(w) \subseteq \{a, b, c\} \,\}$$
$$= \{\, w \in \{a, b, c\}^* \mid \forall i = 1, \ldots, |w| \colon w[i] = w[|w| + 1 - i] \,\}.$$

Clearly, $L_1 \in 2\mathrm{LT}_{R_1}(1)$ and $L_2 \in 2\mathrm{LT}_{R_2}(1)$ and, hence, $L_1, L_2 \in 2\mathrm{LT}(1) \subseteq 2\mathrm{LT}$.

Proof of Prop. 9 (cont.)

Assume $L_1 \cup L_2 \in 2\mathrm{LT}$. Then there are $k \geq 1$ and a symmetric binary relation R on $\{a, b, c\}^k$ such that $L_1 \cup L_2 \in 2\mathrm{LT}_R(k)$.

Proof of Prop. 9 (cont.)

Assume $L_1 \cup L_2 \in 2\mathrm{LT}$. Then there are $k \geq 1$ and a symmetric binary relation R on $\{a, b, c\}^k$ such that $L_1 \cup L_2 \in 2\mathrm{LT}_R(k)$.

We consider the word $v_1 = a^k b^k a^k c^k a^k$. Since $v_1 \in L_1$, the relation R contains the pairs $(a^{k-i}b^i, c^i a^{k-i})$, $(c^i a^{k-i}, a^{k-i}b^i)$, $(b^{k-i}a^i, a^i c^{k-i})$, and $(a^i c^{k-i}, b^{k-i}a^i)$ for all $0 \leq i \leq k$.

Since $v_2 = a^k c^k a^k c^k a^k \in L_2$, the relation R also contains the pairs $(a^{k-i}c^i, c^i a^{k-i})$ and $(c^i a^{k-i}, a^{k-i}c^i)$ for all $0 \leq i \leq k$.

Proof of Prop. 9 (cont.)

Assume $L_1 \cup L_2 \in 2\mathrm{LT}$. Then there are $k \geq 1$ and a symmetric binary relation R on $\{a, b, c\}^k$ such that $L_1 \cup L_2 \in 2\mathrm{LT}_R(k)$.

We consider the word $v_1 = a^k b^k a^k c^k a^k$. Since $v_1 \in L_1$, the relation R contains the pairs $(a^{k-i} b^i, c^i a^{k-i})$, $(c^i a^{k-i}, a^{k-i} b^i)$, $(b^{k-i} a^i, a^i c^{k-i})$, and $(a^i c^{k-i}, b^{k-i} a^i)$ for all $0 \leq i \leq k$.

Since $v_2 = a^k c^k a^k c^k a^k \in L_2$, the relation R also contains the pairs $(a^{k-i} c^i, c^i a^{k-i})$ and $(c^i a^{k-i}, a^{k-i} c^i)$ for all $0 \leq i \leq k$.

Thus, also the word $w = a^k b^k a^k c^k a^k c^k a^k c^k a^k$ belongs to $\{a, b, c\}^{\geq k}_R$.

Since $P_k(w) = P_k(v_1)$, $S_k(w) = S_k(v_1)$, and $I_k(w) = I_k(v_1)$, the word w belongs to the k-R-testable language $L_1 \cup L_2$.

However, as $w \notin L_1$ and $w \notin L_2$, this is a contradiction.

Hence, neither $2\mathrm{LT}(k)$ nor $2\mathrm{LT}$ is closed under union. $\square$

4. Expressive Capacity

A language $L \subseteq \Sigma^*$ is even linear (ELIN) if it is generated by a grammar $G = (V, \Sigma, S, P)$ s.t. all productions in P are of the form $(A \to uBv)$ or $(A \to w)$, where $A, B \in V$ and $u, v, w \in \Sigma^*$ satisfying $|u| = |v|$ [AP64].

4. Expressive Capacity

A language $L \subseteq \Sigma^*$ is even linear (ELIN) if it is generated by a grammar $G = (V, \Sigma, S, P)$ s.t. all productions in P are of the form $(A \to uBv)$ or $(A \to w)$, where $A, B \in V$ and $u, v, w \in \Sigma^*$ satisfying $|u| = |v|$ [AP64].

Theorem 10

The family 2LT *is properly included in* ELIN*.*

4. Expressive Capacity

A language $L \subseteq \Sigma^*$ is even linear (ELIN) if it is generated by a grammar $G = (V, \Sigma, S, P)$ s.t. all productions in P are of the form $(A \to uBv)$ or $(A \to w)$, where $A, B \in V$ and $u, v, w \in \Sigma^*$ satisfying $|u| = |v|$ [AP64].

Theorem 10

The family 2LT is properly included in ELIN.

Proof.

For each $L \in \mathsf{2LT}$, there are $k \geq 1$ and a symmetric binary relation R on Σ^k such that $L \in \mathsf{2LT}_R(k)$. Hence, L can be represented as $L = F \cup \bigcup_{(u,v,C) \in \mathsf{triple}(L)} L_R(u, v, C)$, where $F \subseteq \Sigma^{\leq k-1}$ and $L_R(u, v, C) = \{ w \in \Sigma_R^{\geq k} \mid P_k(w) = u, S_k(w) = v, \text{ and } I_k(w) = C \}$.

Since the set triple(L) is finite and ELIN is closed under union, it is sufficient to construct even linear grammars for the sets $L_R(u, v, C)$.

Proof of Theorem 10 (cont.)

To this end, let $G = (V, \Sigma, S, P)$, where $V = \{S\} \cup \{\, T^{Q}_{[x,y]} \mid Q \subseteq \Sigma^k, x, y \in \Sigma^k \,\}$ is the set of nonterminals, Σ is the set of terminals, S is the start symbol, and P contains the following productions, where $x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_k, p, q \in \Sigma$:

- $S \rightarrow T^{\emptyset}_{[u,v]}$ if $(u, v) \in R$,
- $T^{Q}_{[x_1 x_2 \cdots x_k, y_1 y_2 \cdots y_k]} \rightarrow x_1 T^{Q \cup \{x_2 \cdots x_k p, q y_1 y_2 \cdots y_{k-1}\}}_{[x_2 \cdots x_k p, q y_1 y_2 \cdots y_{k-1}]} y_k$
  if $(x_2 \cdots x_k p, q y_1 y_2 \cdots y_{k-1}) \in R$ and $x_2 \cdots x_k p, q y_1 y_2 \cdots y_{k-1} \in C$,
- $T^{Q}_{[x,y]} \rightarrow x$ if $x = y$ and $Q = C$,
- $T^{Q}_{[x_1 x_2 \cdots x_k, y_1 y_2 \cdots y_k]} \rightarrow x_1 x_2 \cdots x_k y_k$
  if $x_1 x_2 \cdots x_k y_k = x_1 y_1 y_2 \cdots y_k$ and $Q = C$.

Proof of Theorem 10 (cont.)

To this end, let $G = (V, \Sigma, S, P)$, where $V = \{S\} \cup \{\, T^{Q}_{[x,y]} \mid Q \subseteq \Sigma^k, x, y \in \Sigma^k \,\}$ is the set of nonterminals, Σ is the set of terminals, S is the start symbol, and P contains the following productions, where $x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_k, p, q \in \Sigma$:

- $S \to T^{\emptyset}_{[u,v]}$ if $(u, v) \in R$,
- $T^{Q}_{[x_1 x_2 \cdots x_k, y_1 y_2 \cdots y_k]} \to x_1 \, T^{Q \cup \{x_2 \cdots x_k p, q y_1 y_2 \cdots y_{k-1}\}}_{[x_2 \cdots x_k p, q y_1 y_2 \cdots y_{k-1}]} y_k$
  if $(x_2 \cdots x_k p, q y_1 y_2 \cdots y_{k-1}) \in R$ and $x_2 \cdots x_k p, q y_1 y_2 \cdots y_{k-1} \in C$,
- $T^{Q}_{[x,y]} \to x$ if $x = y$ and $Q = C$,
- $T^{Q}_{[x_1 x_2 \cdots x_k, y_1 y_2 \cdots y_k]} \to x_1 x_2 \cdots x_k y_k$
  if $x_1 x_2 \cdots x_k y_k = x_1 y_1 y_2 \cdots y_k$ and $Q = C$.

Then $L(G) = L_R(u, v, C)$. $\square$

Example 4.

Let $k = 1$, $\Sigma = \{a, b\}$, and $R = \{(a, a), (b, b)\} \subset \Sigma \times \Sigma$.
Then $\Sigma_R^{\geq 1} = L_{\text{pal}} \cap \Sigma^+$. So, L_{pal} is a (strictly) 1-testable language.

Example 4.

Let $k = 1$, $\Sigma = \{a, b\}$, and $R = \{(a, a), (b, b)\} \subset \Sigma \times \Sigma$.
Then $\Sigma_R^{\geq 1} = L_{\text{pal}} \cap \Sigma^+$. So, L_{pal} is a (strictly) 1-testable language.

As L_{pal} is neither deterministic context-free nor a Church-Rosser language [JL2007], while the regular language $(aa)^*$ is not two-sided locally testable, we obtain the following incomparability result.

Example 4.

Let $k = 1$, $\Sigma = \{a, b\}$, and $R = \{(a, a), (b, b)\} \subset \Sigma \times \Sigma$.
Then $\Sigma_R^{\geq 1} = L_{\mathrm{pal}} \cap \Sigma^+$. So, L_{pal} is a (strictly) 1-testable language.

As L_{pal} is neither deterministic context-free nor a Church-Rosser language [JL2007], while the regular language $(aa)^*$ is not two-sided locally testable, we obtain the following incomparability result.

Corollary 11

For all $k \geq 1$, 2LT and 2LT(k) are incomparable to the families of regular, deterministic linear, deterministic context-free, and Church-Rosser languages. There exist a binary alphabet Σ and a symmetric relation $R \subseteq \Sigma^k \times \Sigma^k$ such that $2\mathrm{LT}_R(k)$ is incomparable to the families of regular, deterministic linear, deterministic context-free, and Church-Rosser languages.

Let $2\mathrm{SLT}_R(k)$ denote the family of two-sided strictly k-testable languages that are based on the symmetric relation R.

Theorem 12

For each $k \geq 1$ and each symmetric relation $R \subseteq \Sigma^k \times \Sigma^k$, the family $2\mathrm{LT}_R(k)$ is the closure of the family $2\mathrm{SLT}_R(k)$ under union, intersection, and R-complementation.

Let $2SLT_R(k)$ denote the family of two-sided strictly k-testable languages that are based on the symmetric relation R.

Theorem 12

For each $k \geq 1$ and each symmetric relation $R \subseteq \Sigma^k \times \Sigma^k$, the family $2LT_R(k)$ is the closure of the family $2SLT_R(k)$ under union, intersection, and R-complementation.

Proof.

By Proposition 7, $2LT_R(k)$ is closed under the operations considered. Since $2SLT_R(k) \subseteq 2LT_R(k)$, it is sufficient to show that every language from $2LT_R(k)$ can be represented as a combination of languages from $2SLT_R(k)$ using the operations mentioned.

Let $2\mathrm{SLT}_R(k)$ denote the family of two-sided strictly k-testable languages that are based on the symmetric relation R.

Theorem 12

For each $k \geq 1$ and each symmetric relation $R \subseteq \Sigma^k \times \Sigma^k$, the family $2\mathrm{LT}_R(k)$ is the closure of the family $2\mathrm{SLT}_R(k)$ under union, intersection, and R-complementation.

Proof.

By Proposition 7, $2\mathrm{LT}_R(k)$ is closed under the operations considered. Since $2\mathrm{SLT}_R(k) \subseteq 2\mathrm{LT}_R(k)$, it is sufficient to show that every language from $2\mathrm{LT}_R(k)$ can be represented as a combination of languages from $2\mathrm{SLT}_R(k)$ using the operations mentioned.

Recall that each language $L \in 2\mathrm{LT}_R(k)$ has a representation of the form $L = F_L \cup \bigcup_{(u,v,C)\in\mathrm{triple}(L)} L_R(u,v,C)$, where

$$L_R(u,v,C) = \{\, w \in \Sigma_R^{\geq k} \mid P_k(w) = u, S_k(w) = v, \text{ and } I_k(w) = C \,\}.$$

Proof of Theorem 12 (cont.)

Since triple(L) is a finite set, it remains to be shown that any language $L_R(u, v, C)$ with $(u, v, C) \in$ triple(L) can be represented as a combination of languages from $2SLT_R(k)$ using the above operations.

Proof of Theorem 12 (cont.)

Since $\text{triple}(L)$ is a finite set, it remains to be shown that any language $L_R(u, v, C)$ with $(u, v, C) \in \text{triple}(L)$ can be represented as a combination of languages from $2\text{SLT}_R(k)$ using the above operations.

Starting with $L(\{u\}, \{v\}, C, R) \in 2\text{STL}_R(k)$, we obtain the inclusion $L_R(u, v, C) \subseteq L(\{u\}, \{v\}, C, R)$. The problem to cope with is that there may be words w in the latter language such that $I_k(w) \subsetneq C$.

These words can be filtered out by building the intersection of the R-complements of all languages $L(\{u\}, \{v\}, C', R)$, where the intersection is taken over all proper subsets C' of C.

So, we have the representation

$$L_R(u, v, C) = L(\{u\}, \{v\}, C, R) \cap \bigcap_{C' \subset C} L_R^c(\{u\}, \{v\}, C', R).$$

Thus, each language $L \in 2\text{LT}_R(k)$ is a finite combination of two-sided strictly k-testable languages with respect to the relation R. □

Since $2\mathrm{LT}(k)$ and 2LT are not closed under union, a similar characterization does not exist.

However, for every language L from these families, there are an integer $k \geq 1$ and a symmetric relation $R \subseteq \Sigma^k \times \Sigma^k$ such that $L \in 2\mathrm{LT}_R(k)$. Hence, L has a representation as a combination of two-sided strictly k-testable languages and, thus, belongs to the closure of $2\mathrm{SLT}_R(k)$ with respect to the operations above.

Since $2\mathrm{LT}(k)$ and 2LT are not closed under union, a similar characterization does not exist.

However, for every language L from these families, there are an integer $k \geq 1$ and a symmetric relation $R \subseteq \Sigma^k \times \Sigma^k$ such that $L \in 2\mathrm{LT}_R(k)$. Hence, L has a representation as a combination of two-sided strictly k-testable languages and, thus, belongs to the closure of $2\mathrm{SLT}_R(k)$ with respect to the operations above.

By R-Boolean closure we mean the closure under the operations of union, intersection, and R-complementation.

Corollary 13

1. *The family 2LT ($2\mathrm{LT}(k)$) is properly included in the R-Boolean closure of 2SLT ($2\mathrm{SLT}(k)$).*
2. *The R-Boolean closure of 2LT ($2\mathrm{LT}(k)$) coincides with the R-Boolean closure of 2SLT ($2\mathrm{SLT}(k)$).*

It has been observed in [NCMA2017] that there are regular languages that are two-sided strictly testable, but not (one-sided) strictly testable. A corresponding result holds for k-testability.

It has been observed in [NCMA2017] that there are regular languages that are two-sided strictly testable, but not (one-sided) strictly testable. A corresponding result holds for k-testability.

Theorem 14

For all $k \geq 2$, $\mathrm{LT}(k) \subsetneq 2\mathrm{LT}(k) \cap \mathrm{REG}$.

It has been observed in [NCMA2017] that there are regular languages that are two-sided strictly testable, but not (one-sided) strictly testable. A corresponding result holds for k-testability.

Theorem 14

For all $k \geq 2$, $\mathrm{LT}(k) \subsetneq 2\mathrm{LT}(k) \cap \mathrm{REG}$.

A corresponding result does not hold for $k = 1$.

Theorem 15

$2\mathrm{LT}(1) \cap \mathrm{REG} = \mathrm{LT}(1)$.

5. Conclusion

We have extended 2SLT to 2LT and shown that the latter are obtained as the R-Boolean closure of the former.

Further, we have established some closure and non-closure properties for 2LT and it can be shown that two-sided k-testable languages are learnable in the limit from positive data.

5. Conclusion

We have extended 2SLT to 2LT and shown that the latter are obtained as the R-Boolean closure of the former.

Further, we have established some closure and non-closure properties for 2LT and it can be shown that two-sided k-testable languages are learnable in the limit from positive data.

Some problems that remain open for future work include the following.

Inside the Boolean closure: How about the union closure of 2SLT(k)?

5. Conclusion

We have extended 2SLT to 2LT and shown that the latter are obtained as the R-Boolean closure of the former.

Further, we have established some closure and non-closure properties for 2LT and it can be shown that two-sided k-testable languages are learnable in the limit from positive data.

Some problems that remain open for future work include the following.

Inside the Boolean closure: How about the union closure of 2SLT(k)?

Separation by regular languages: For all $k \geq 2$, $\mathrm{LT}(k) \subsetneq 2\mathrm{LT}(k) \cap \mathrm{REG}$. Can we can separate LT from 2LT by a regular language?

5. Conclusion

We have extended 2SLT to 2LT and shown that the latter are obtained as the R-Boolean closure of the former.

Further, we have established some closure and non-closure properties for 2LT and it can be shown that two-sided k-testable languages are learnable in the limit from positive data.

Some problems that remain open for future work include the following.

Inside the Boolean closure: How about the union closure of 2SLT(k)?

Separation by regular languages: For all $k \geq 2$, $\mathrm{LT}(k) \subsetneq \mathrm{2LT}(k) \cap \mathrm{REG}$. Can we can separate LT from 2LT by a regular language?

Decidability: It can be shown that emptiness, finiteness, containment of a given regular set, and equality to a given regular set are decidable for two-sided locally testable languages. Also inclusion and equivalence are decidable for k-R-testable languages, but it remains open whether these problems are decidable for 2LT(k) and for 2LT.

Thank you for your attention!