Duplication Closure of Regular Languages

Masami ITO

(Received September 14, 2009, Revised December 2, 2009)

Abstract

In the present paper, we prove that the n-bounded duplication closure of a regular language is regular for n = 1, 2.

Keywords: duplication (closure), n-bounded duplication (closure), regular language, automaton

1. Introduction

Let X be a nonempty finite set, called an alphabet. An element of X is called a letter. By X*, we denote the free monoid generated by X. Let X* = X* \ {ε} where ε denotes the empty word of X*, i.e. the identity of X*. An element of X* is called a word over X. If u ∈ X*, then |u| denotes the length of u, i.e. the number of letters appearing in u. Notice that we also denote the cardinality of a finite set A by |A|. Regarding more details on languages, see [2] and [3].

Definition 1. Let A = (S, X, δ, s₀, F) where (1) S and X are nonempty finite sets called a state set and an alphabet, respectively, (2) δ is a function called a state transition function such that δ(s, a) ∈ S for any s ∈ S and any a ∈ X, (3) s₀ ∈ S called an initial state and (4) F ⊆ S called the set of final states.

Then A is called a finite automaton.

Notice that the above δ can be extended to the following function in a natural way, i.e. δ(s, ε) = s and δ(s, au) = δ(δ(s, a), u) for any s ∈ S, any u ∈ X* and any a ∈ X.

Definition 2. Let A = (S, X, δ, s₀, F) be a finite automaton. Then the language \{u ∈ X* | δ(s₀, u) ∈ F\} is said to be accepted by A and denoted by L(A). A language accepted by a finite automaton is called a regular language.

Finite automata can be generalized in the following way.

Definition 3. Let A = (S, X, δ, s₀, F) where (1) S and X are nonempty finite sets called a state set and an alphabet, respectively, (2) δ is a relation called a state transition relation such that δ(s, a) ⊆ S for any s ∈ S and any a ∈ X, (3) s₀ ⊆ S called the set of initial state and (4) F ⊆ S called the set of final states.

Then A is called a nondeterministic automaton.

The above δ can be extended to the following relation in a natural way, i.e. δ(s, ε) = \{s\} and δ(s, au) = \bigcup_{t \in \delta(s, a)} δ(t, u) for any s ∈ S, any u ∈ X* and any a ∈ X.
Definition 4. Let $\mathcal{A} = (S, X, \delta, S_0, F)$ be a nondeterministic automaton. Then the language $\{u \in X^* \mid \exists s_0 \in S_0, \delta(s_0, u) \cap F \neq \emptyset\}$ is said to be accepted by $\mathcal{A}$ and denoted by $L(\mathcal{A})$.

It seems that a language accepted by a nondeterministic automaton is not necessarily regular. However, we have the following result.

Fact. A language accepted by a nondeterministic automaton is regular.

Regarding more details on regular languages and automata, see [2] and [3].

Let $u \in X^*$ and let $n$ be a positive integer. Then we introduce operations, called duplication operations. Let $u = vwx$ where $v, w, x \in X^*$. Then we denote $u \rightarrow vwvx$ and $\rightarrow$ is called a duplication. Moreover, if $|w| \leq n$ in the above, then we denote $u \rightarrow_{\leq n} vwvx$ and $\rightarrow_{\leq n}$ is called an $n$-bounded duplication.

By $\rightarrow^*$ and $\rightarrow_{\leq n}^*$, we denote the reflexive and transitive closures of $\rightarrow$ and $\rightarrow_{\leq n}$, respectively.

Definition 5. Let $u \in X^*$ and let $n$ be a positive integer. Then $u^\circ = \{v \in X^* \mid u \rightarrow^* v\}$ and $u^{\circ_{\leq n}} = \{v \in X^* \mid u \rightarrow_{\leq n}^* v\}$, called the duplication closure of $u$ and $n$-bounded duplication closure of $u$, respectively. Moreover, let $n$ be a positive integer. Then $L^\circ = \{u^\circ \mid u \in L\}$ and $L^{\circ_{\leq n}} = \{u^{\circ_{\leq n}} \mid u \in L\}$, called the duplication closure of $L$ and $n$-bounded duplication closure of $L$, respectively.

2. 1-Bounded Duplication Closures

In this section, we prove that the 1-bounded duplication closure of a regular language is regular.

Theorem 1. Let $L \subseteq X^*$ be a regular language. Then $L^{\circ_{\leq 1}}$ is regular.

Proof. Let $\mathcal{A} = (S, X, \delta, s_0, F)$ be a finite automaton accepting $L$. We construct the following nondeterministic automaton $\overline{\mathcal{A}} = (\overline{S}, X, \overline{\delta}, s_0, F)$: (1) $\overline{S} = \{s_a \mid s \in S, a \in X \cup \{\varepsilon\}\}$ where $s_\varepsilon$ can be regarded as $s$. (2) $\overline{F} = \{s_a \mid a \in X \cup \{\varepsilon\}, s \in F\}$. (3) $\overline{\delta}(s_\alpha, a) = \{\delta(s, a) \cup \{s_a \mid \alpha = a\}$ for $\alpha \in X \cup \{\varepsilon\}$ and $a \in X$.

Now we prove that $L(\overline{\mathcal{A}}) = L^{\circ_{\leq 1}}$. Let $u \in L(\overline{\mathcal{A}})$. Then $u$ can be represented as follows: $u = u_1v_1u_2v_2 \cdots u_nv_n$, where $u_i = u_i' a_i, u_i' \in X^*$, $a_i \in X$ and $v_i \in a_i^+$ for any $i = 1, 2, \ldots, r$, and $\overline{\delta}(s_0, u_1u_2 \cdots u_r) \in F$, i.e. $u_1u_2 \cdots u_r \in L$. Hence $u \in L^{\circ_{\leq 1}}$, i.e. $L(\overline{\mathcal{A}}) \subseteq L^{\circ_{\leq 1}}$.

Now let $u \in L^{\circ_{\leq 1}}$. If $u \in L$, then obviously $u \in L(\overline{\mathcal{A}})$. Assume that $v \rightarrow_{\leq 1} u$ for some $v \in L^{\circ_{\leq 1}} \cap L(\overline{\mathcal{A}})$. Then $v = v_1a_1, v_1, v_2 \in X^*, a \in X$ and $u = v_1a_1^2v_2$. Let $s_\alpha \in \overline{\delta}(s_0, v_1a)$ where $s \in S$. Then $s_\alpha \in \overline{\delta}(s_0, v_1a)$. Hence $\overline{\delta}(s_0, v_1a^2) = \overline{\delta}(s_0, v_1a^2v_2)$, i.e. $u \in L(\overline{\mathcal{A}})$ and $L^{\circ_{\leq 1}} \subseteq L(\overline{\mathcal{A}})$.

This completes the proof of the theorem.

3. 2-Bounded Duplication Closures

In this section, we prove that the 2-bounded duplication closure of a regular language is regular.

Lemma 1. Let $a, b \in X$. Then $(ab)^{\circ_{\leq 2}} = a^*(a, b)b$. 

Masami ITO
Proposition 1. Let $L$ be a regular language. Then $L^{\circ \leq 2}$ is regular.

Theorem 2. Let $L \subseteq X^*$ be a regular language. Then $L^{\circ \leq 2}$ is regular.

Proof. Let $\mathcal{A} = (S, X, \delta, s_0, F)$ be a finite automaton accepting $L$. We construct the following nondeterministic automaton $\mathcal{B} = (T, X, \gamma, T_0, G)$: (1) $T = \{[s_0]_a | a \in X\} \cup \{[s]_{ab} | s \in S, a, b \in X\}$ where # is a new symbol. (2) $G = \{[s]_{ab} | a, b \in X, s \in F\}$ if $s_0 \notin F$ and $G = \{[s]_{ab} | a, b \in X, s \in F\} \cup \{[s_0]_a | a \in X\}$ if $s_0 \in F$. (3) $T_0 = \{[s_0]_a | a \in X\}$. (4) $\gamma([s_0]_a, a) = ([\delta(s_0, a)]_a | b \in X), \gamma([s]_{ab}, a) = ([s]_{ab})$ and $\gamma([s]_{ab}, b) = ([\delta(s, b)]_{cb} | c \in X \cup \{[s]_{ab}\}$.

Let $a, b \in X$ and let $u, v \in X^*$. By Lemma 1 and the structure of the automaton $\mathcal{B}$, the configuration $uabv \rightarrow^{*\leq 2} uaxbv$ with $x \in \{a, b\}^*$ corresponds to $\gamma([s]_{ab}, v) \subseteq \gamma([s_0]_{ab}, uaxbv)$ where $u \in cX^*$ and $s = \delta(s_0, ua)$. It can be proved that $L(\mathcal{B}) = L^{\circ \leq 2}$.

Actually, Theorem 2 has been already proved in [5] based on the proposition below. However, the proof was complicated. On the contrary, the above proof is simpler and more constructive.

Definition 6. Let $L \subseteq X^*$. The following equivalence relation $P_L$ on $X^* \times X^*$ is called the principal congruence on $L$: $\forall u, v \in X^*, uP_Lv \iff \forall x, y \in X^* (xuv \in L \iff xvy \in L)$.

Proposition 1. Let $L \subseteq X^*$. Then $L$ is regular if and only if the number of equivalence classes is finite.

4. Duplication Closure of a Regular Language over a Binary Alphabet

In this section, we prove that the duplication closure of a regular language over a binary alphabet is regular. By Lemma 1, we can obtain the following lemma.

Lemma 2. Let $X = \{a, b\}$ and let $u \in X^*$. Then $u^{\circ \leq 2} = u^{\circ \leq 3}$.

Theorem 3. Let $X = \{a, b\}$ and let $L \subseteq X^* \rho$ be regular. Then $L^{\circ}$ is regular.

Proof. By Lemma 2, $L^{\circ} = L^{\circ \leq 2}$. It follows from Theorem 2 that $L^{\circ}$ is regular.

Actually, it was proved in [1] that the duplication closure of any language over a binary alphabet is regular. However, Theorem 3 provides concrete relations between languages and their duplication closures for regular languages.

5. Conclusion

In a similar way as before, we can construct a nondeterministic automaton accepting the 3-bounded duplication closure of a regular language (see [4]). Thus we have:

Theorem 4. Let $L \subseteq X^*$ be a regular language, then $L^{\circ \leq 3}$ is regular.
Acknowledgements

This work was supported by the JSPS-HAS Joint Research Grant and the Grant-in-Aid for Organizing International Conferences from Kyoto Sangyo University. The author also thanks the referees for their useful comments.

References

正規言語の重複閉包

伊藤 正美

要 旨
本論文では、正規言語の1-有界重複閉包および2-有界重複閉包が正規言語になることを示す。

キーワード：重複（閉包）、n-有界重複（閉包）、正規言語、オートマトン