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 $X^+ = X^* \setminus \{\epsilon\}$ 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 $u \in X^*$, 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 $\mathcal{A} = (S, X, \delta, s_0, F)$ 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 $\delta(s, a) \in S$ for any $s \in S$ and any $a \in X$, (3) $s_0 \in S$ called an *initial state* and (4) $F \subseteq S$ called the *set of final states*.

Then $\mathcal{A}$ is called a *finite automaton*.

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

Definition 2. Let $\mathcal{A} = (S, X, \delta, s_0, F)$ be a finite automaton. Then the language $\{u \in X^* \mid \delta(s_0, u) \in F\}$ is said to be *accepted* by $\mathcal{A}$ and denoted by $\mathcal{L}(\mathcal{A})$. A language accepted by a finite automaton is called a *regular* language.

Finite automata can be generalized in the following way.

Definition 3. Let $\mathcal{A} = (S, X, \delta, S_0, F)$ 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 $\delta(s, a) \subseteq S$ for any $s \in S$ and any $a \in X$, (3) $S_0 \subseteq S$ called the *set of initial state* and (4) $F \subseteq S$ called the *set of final states*.

Then $\mathcal{A}$ is called a *nondeterministic automaton*.

The above δ can be extended to the following relation in a natural way, i.e. $\delta(s, \epsilon) = \{s\}$ and $\delta(s, au) = \bigcup_{t \in \delta(s,a)} \delta(t, u)$ for any $s \in S$, any $u \in X^*$ and any $a \in X$.

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 $\mathcal{L}(\mathcal{A})$.

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 $u \to vwwx$ and $\to$ is called a *duplication*. Moreover, if $|w| \leq n$ in the above, then we denote $u \to_{\leq n} uwwx$ and $\to_{\leq n}$ is called an *n-bounded duplication*.

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

Definition 5. Let $u \in X^*$ and let n be a positive integer. Then $u^\heartsuit = \{v \in X^* \mid u \to^* v\}$ and $u^{\heartsuit \leq n} = \{v \in X^* \mid u \to^*_{\leq n} v\}$, called the *duplication closure* of u and *n-bounded duplication closue* of u, respectively. Moreover, let n be a positive integer. Then $L^\heartsuit = \{u^\heartsuit \mid u \in L\}$ and $L^{\heartsuit \leq n} = \{u^{\heartsuit \leq n} \mid u \in L\}$, called the *duplication closure* of L and *n-bounded duplication closue* 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 $L^{\heartsuit \leq 1}$ is regular.*

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

Now we prove that $\mathcal{L}(\overline{\mathcal{A}}) = L^{\heartsuit \leq 1}$. Let $u \in \mathcal{L}(\overline{\mathcal{A}})$. Then u can be represented as follows: $u = u_1 v_1 u_2 v_2 \cdots u_r v_r$ 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 $\delta(s_0, u_1 u_2 \cdots u_r) \in F$, i.e. $u_1 u_2 \cdots u_r \in L$. Hence $u \in L^{\heartsuit \leq 1}$, i.e. $\mathcal{L}(\overline{\mathcal{A}}) \subseteq L^{\heartsuit \leq 1}$.

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

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 $(ab)^{\heartsuit \leq 2} = a\{a, b\}^* b$.*

Proof. It can be easily shown that the theorem holds true if $a = b$. Hence we assume $a \neq b$. Let $u \in X^*$. If $u = a^i, i \geq 0$, then $ab \to^*_{\leq 2} a^{i+1}b = aub$. If $u = b^i, i \geq 0$, then $ab \to^*_{\leq 2} ab^{i+1} = aub$. If $u = a^{i_1}b^{j_1}a^{i_2}b^{j_2}\cdots a^{i_p}$ such that $i_1, i_2, \ldots i_p, j_1, j_2, \ldots j_{p-1} \geq 1$, then $ab \to^*_{\leq 2} (ab)(ab)\cdots(ab) \to^*_{\leq 2} aa^{i_1}b^{j_1}a^{i_2}b^{j_2}\cdots a^{i_p}b = aub$. If $u = a^{i_1}b^{j_1}a^{i_2}b^{j_2}\cdots a^{i_p}b^{j_p}$ such that $i_1, i_2, \ldots, i_p, j_1, j_2, \ldots, j_p \geq 1$, then $ab \to^*_{\leq 2} (ab)(ab)\cdots(ab) \to^*_{\leq 2} a^{i_1+1}b^{j_1}a^{i_2}b^{j_2}\cdots a^{i_p}b^{j_p+1} = aub$. In the same way, we can prove that $ab \to^*_{\leq 2} aub$ for any $u \in b\{a, b\}^*$.

Theorem 2. *Let $L \subseteq X^*$ be a regular language. Then $L^{\heartsuit\leq 2}$ 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) $T = \{[s_0]_{\#a} \mid a \in X\} \cup \{[s]_{ab} \mid s \in S, a, b \in X\}$ where # is a new symbol. (2) $G = \{[s]_{ab} \mid a, b \in X, s \in F\}$ if $s_0 \notin F$ and $G = \{[s]_{ab} \mid a, b \in X, s \in F\} \cup \{[s_0]_{\#a} \mid a \in X\}$ if $s_0 \in F$. (3) $T_0 = \{[s_0]_{\#a} \mid a \in X\}$. (4) $\gamma([s_0]_{\#a}, a) = \{[\delta(s_0, a)]_{ab} \mid b \in X\}, \gamma([s]_{ab}, a) = \{[s]_{ab}\}$ and $\gamma([s]_{ab}, b) = \{[\delta(s, b)]_{bc} \mid c \in X\} \cup \{[s]_{ab}\}$.

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

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: $\forall u, v \in X^*, uP_Lv \Leftrightarrow \forall x, y \in X^*(xuy \in L \leftrightarrow xvy \in L)$.

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 $u^\heartsuit = u^{\heartsuit\leq 2}$.*

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

Proof. By Lemma 2, $L^\heartsuit = L^{\heartsuit\leq 2}$. It follows from Theorem 2 that $L^\heartsuit$ 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 $L^{\heartsuit\leq 3}$ 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

[1] D.P. Bovet and S. Varricchio, On the regularity of languages on a binary alphabet generated by copying systems, Inf. Process. Lett. 44, 119–123, 1992.

[2] J.E. Hopcroft and J.D. Ullman, *Introduction to Automata Theory, Languages and and Computation*, Addison-Wesley, Reading MA, 1979.

[3] M. Ito, *Algebraic Theory of Automata and Languages*, World Scientific, Singapore, 2004.

[4] M. Ito, On n-bounded duplication closures of languages, submitted.

[5] M. Ito, P. Leupold and K. Shikishima-Tsuji, Closure of language classes under bounded duplication, Lecture Notes in Computer Science 4036 (Springer, Heidelberg), 238–247, 2006.

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