Learning Reversible Languages with Terminal Distinguishability*

José M. Sempere

Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia (Spain)
jsempere@dsic.upv.es
http://www.dsic.upv.es/users/tlcc/tlcc.html

Abstract. k-reversible languages are regular ones that offer interesting properties under the point of view of identification of formal languages in the limit. Different methods have been proposed to identify k-reversible languages in the limit from positive samples. Non-regular language classes have been reduced to regular reversible languages in order to solve their associated learning problems. In this work, we present a hierarchy of reversible languages which can be characterized by some properties related to the set of terminal segments of the automata (*terminal distinguishability*). Terminal distinguishability is a property that has been previously used to characterize other language families which can be identified in the limit from positive data. In the present work we combine reversibility and terminal distinguishability in order to define a new hierarchy of regular languages which is highly related to the k-reversible hierarchy. We will provide an efficient method to identify any given language in the hierarchy from only positive examples.

Definitions

Σ denotes an alphabet, Σ^* the infinite set of strings defined by Σ, Σ^l denotes the set of strings with length l. The reverse of any string u will be denoted by u^{inv}, and the set of segments of u with length k will be denoted by $seg(u,k)$. Given a finite automaton A, we denote the reverse automaton of A by A^{inv}. Given a finite automaton $A = (Q, \Sigma, \delta, I, F)$ and $q \in Q$, and given an integer value $k \geq 0$, we will say that $u \in \Sigma^k$, is a k-follower (k-leader) of q if $\delta(q,u) \neq \oslash$ (resp. $\delta^{inv}(q, u^{inv}) \neq \oslash$). A finite automaton A is *deterministic with lookahead* k if and only if for every three states q_1, q_2 and q_3, if $q_1, q_2 \in \delta(q_3, a)$, or $q_1, q_2 \in I$, then there is no common k-follower of q_1 and q_2.

k-reversible Languages with r-terminal Distinguishability: The $\mathcal{REV}(k,r)$ Class

Definition 1. *Let* $A = (Q, \Sigma, \delta, I, F)$ *be a finite automaton, and let* k, r *be integer values with* $0 < r \leq k$. *We will say that A is deterministic with lookahead*

* Work supported by the Spanish CICYT under contract TIC2003-09319-C03-02 and the Generalitat Valenciana GV06/068.

Y. Sakakibara et al. (Eds.): ICGI 2006, LNAI 4201, pp. 354–355, 2006.

k and r-terminal distinguishability iff for every three states q_1, q_2 and q_3, if $q_1, q_2 \in \delta(q_3, a)$, or $q_1, q_2 \in I$, then at least one of the following conditions holds:

1. *$\exists v_1 \in \Sigma^k$ being a k-follower of q_1 such that for all $v_2 \in \Sigma^k$ being a k-follower of q_2 $seg(v_1, r) \neq seg(v_2, r)$.*
2. *$\exists v_2 \in \Sigma^k$ being a k-follower of q_2 such that for all $v_1 \in \Sigma^k$ being a k-follower of q_1 $seg(v_2, r) \neq seg(v_1, r)$.*

Definition 2. *We will say that a finite automaton A is k-reversible with r-terminal distinguishability iff A is deterministic and A^{inv} is deterministic with lookahead k and r-terminal distinguishability.*

A language L is k-reversible with r-terminal distinguishability, with $0 < r \leq k$, if there exists a k-reversible finite automaton with r-terminal distinguishability A such that $L(A) = L$. We will denote the family of k-reversible languages with r-terminal distinguishability by $\mathcal{REV}(k, r)$. Observe that there exists narrow relation between the class $\mathcal{REV}(k)$ in Angluin's work [An82] and the class $\mathcal{REV}(k, r)$. The relationship between different language classes is showed through the following results.

Lemma 1. *For any $k > 0$, $\mathcal{REV}(k) \subsetneq \mathcal{REV}(k, k)$.*

Lemma 2. *For every integer values k and r such that $1 \leq r < k$, $\mathcal{REV}(k, r) \subset \mathcal{REV}(k, r + 1)$.*

Identifying $\mathcal{REV}(k, r)$ Languages in the Limit

In order to identify any $\mathcal{REV}(k, r)$ language in the limit we propose a modification of Angluin's algorithm [An82]. Observe that Angluin proposes the merging of two states (*blocks*) due to the following two criteria:

1. The two states produce non determinism
2. The two states have at least one transition with the same symbol to the same state and they have a common k-leader.

Here we propose an additional criterion which can be enunciated as follows: Two states (*blocks*) are merged if they have at least one transition with the same symbol to the same state and the conditions of definition 1 do not hold. The modified algorithm obtains the smallest language L such that L is in $\mathcal{REV}(k, r)$ and L contains the sample given as input. The proof follows from Angluin's work [An82]. The complexity of the modified algorithm is still polynomial, in fact $\mathcal{O}(C(k, r) \cdot n^3)$, where $C(k, r)$ is a constant that depends on the number of k-leaders of every state and the number of r segments.

Reference

[An82] D. Angluin. *Inference of Reversible Languages.* Journal of the Association for Computing Machinery. Vol 29 No 3, pp 741-765. July 1982.