Learning of Regular $\omega$-Tree Languages

M. Jayasrirani$^1$, M.H. Begam$^1$, and D.G. Thomas$^2$

$^1$ Arignar Anna Government Arts College, Walajapet
$^2$ Madras Christian College, Chennai - 600 059
dgthomasmcc@yahoo.com

Abstract. We introduce two subclasses of regular $\omega$-tree languages called local $\omega$-tree languages and Buchi local $\omega$-tree languages. Automata characterization for these $\omega$-tree languages is given. For these subclasses and $\omega$-regular tree languages learning algorithms are given.

1 Introduction

The theory of tree automata and tree languages emerged in the middle of 1960s. Saoudi et al. [2] have considered infinite trees ($\omega$-trees), recognizable $\omega$-tree languages and regular $\omega$-tree languages. Infinite trees are useful to decide second order theories. In this paper local $\omega$-tree languages and Buchi local $\omega$-tree languages are defined and automata characterization for $\omega$-regular tree languages in terms of local $\omega$-tree languages and Buchi local $\omega$-tree languages is given. There is no learning algorithm so far in the literature for the local $\omega$-tree languages, Buchi local $\omega$-tree languages and regular $\omega$-tree languages. We give learning algorithms for these classes of $\omega$-tree languages. Our approach is similar to the one given in [3].

2 Definitions and Results

Definitions concerning trees, root of a tree, frontier of a tree, forks of a tree, infinite trees, automata on infinite trees and ultimately periodic infinite trees can be found in [1,2].

$T_\Sigma$ stands for the set of all finite trees over $\Sigma$.

$T_\omega^\Sigma$ stands for the set of all infinite trees over $\Sigma$.

root$(t)$ stands for root of a tree $t$.

fork$(t)$ stands for fork of a tree $t$.

fork$(\Sigma)$ stands for the set of all forks of $\Sigma$-trees.

Frfork$(t)$ stands for the set of all forks of a tree $t$ that end with frontiers of $t$.

Definition 1. A $\omega$-tree language $L \subseteq T_\omega^\Sigma$ is called a local $\omega$-tree language if there exists a pair $S = \{R, E\}$ (called a local system) where $R \subseteq \Sigma$ and $E \subseteq$ fork$(\Sigma)$ such that

$L = \{t \in T_\omega^\Sigma : \text{root}(t) \in R, \text{fork}(t) \subseteq E\}$

The elements in fork$(\Sigma)$ occur infinitely many times. In this case we write $L = L^\omega(R, E)$. The set of all local $\omega$-tree languages is denoted by $L^\omega$. $L = \{a(b^\omega, c^\omega), a(c^\omega, b^\omega)\}$ is a local $\omega$-tree language.
Definition 2. A Buchi local system over \( \Sigma \) is an ordered triple \( S = \{ R, E, E' \} \) where \( R \subseteq \Sigma \), \( E \subseteq \text{fork}(\Sigma) \) and \( E' \subseteq E \). We denote \( L^\omega(R, E, E') \) a Buchi local \( \omega \)-tree language defined as

\[
L'(R, E, E') = \{ t \in T_\Sigma : \text{root}(t) \in R, \text{fork}(t) \subseteq E, \text{inf\,fork}(t) \cap E' \neq \phi \}
\]

where \( \text{inf\,fork}(t) \) is the set of elements in \( \text{fork}(t) \) which occur infinitely many times in \( t \). An \( \omega \)-tree language \( L \subseteq T_\omega^\Sigma \) is called a Buchi local \( \omega \)-tree language if there exists a Buchi local system such that \( L = L^\omega(R, E, E') \). The set of all Buchi local \( \omega \)-tree languages is denoted by \( L_B^\omega \). \( L = \{ a(b^\omega, c^\omega), a(c^\omega, b^\omega) \} \) is a Buchi local \( \omega \)-tree language.

Theorem 1. Every regular \( \omega \)-tree language (recognizable \( \omega \)-tree language) is an alphabetic homomorphic image of a Buchi local (local) \( \omega \)-tree language.

We can give construction procedures for deterministic Buchi \( k \)-ary \( \omega \)-tree automaton \( M \) such that \( L = L^\omega(M) \) where \( L \) is a local (Buchi local) \( \omega \)-tree language.

3 Learning Buchi Local \( \omega \)-Tree Languages

Definition 3. Let \( L \in L_B^\omega \) be such that \( L = L^\omega(S) \) for some Buchi local system \( S = \{ R, E, E' \} \) over an alphabet \( \Sigma \). \( S \) is said to be minimal for \( L \), if for any other Buchi local system \( S_1 = \{ R_1, E_1, E'_1 \} \) over \( \Sigma \), with \( L = L^\omega(S_1) \), we have \( R \subseteq R_1, E \subseteq E_1 \) and \( E' \subseteq E'_1 \).

Definition 4. Let \( K \) be a finite sample of ultimately periodic infinite trees. Let \( R_K = \text{root}(K) = \{ \text{root}(t) : t \in K \} \), \( E_K = \text{fork}(K) = \cup_{t \in K} \text{fork}(t) \)

\[
E'_K = \cup_{a(b^\omega, c^\omega)} F_r \text{fork}(t)
\]

\( S_K = \{ R_K, E_K, E'_K \} \) is called a Buchi local system associated with \( K \) and \( L = L^\omega(S_K) \) is called Buchi local \( \omega \)-tree language associated with \( K \).

Theorem 2. If \( K, K' \) are finite samples of ultimately periodic \( \omega \)-trees of \( T_\omega^\Sigma \) then

1. \( K \subseteq L^\omega(S_K) \)
2. \( K \subseteq K' \) implies \( L^\omega(S_K) \subseteq L^\omega(S_K') \)
3. \( L \in L_B^\omega \) with \( K \subseteq L \) implies \( L^\omega(S_K) \subseteq L \)

Definition 5. Let \( L \) be a local (Buchi local) \( \omega \)-tree language. A finite subset \( F \) of \( T_\omega^\Sigma \) is called a characteristic sample for \( L \) if \( L \) is the smallest local (Buchi local) \( \omega \)-tree language containing \( F \).

Theorem 3. If \( F \) is the characteristic sample for a local (Buchi local) \( \omega \)-tree language and \( F \subseteq K \subseteq L \) then \( L = L^\omega(S_K) \).

Theorem 4. There effectively exists a characteristic sample for any local (Buchi local) \( \omega \)-tree language.

Theorem 5. Given an unknown local (Buchi local) \( \omega \)-tree language we give an algorithm that learns in the limit from positive data, a local system (Buchi local system) \( S_F \) such that \( L^\omega(S_F) = L \).