Journal of Automata, Languages and Combinatorics **12** (2007) 4, 501–524
© Otto-von-Guericke-Universität Magdeburg

LEARNING REGULAR TREE LANGUAGES FROM CORRECTION AND EQUIVALENCE QUERIES

CĂTĂLIN IONUŢ TÎRNĂUCĂ and CRISTINA TÎRNĂUCĂ [1]
Research Group on Mathematical Linguistics, Rovira i Virgili University
Plaça Imperial Tàrraco 1, 43005 Tarragona, Spain
e-mail: catalinionut.tirnauca@estudiants.urv.cat
cristina.bibire@urv.cat

ABSTRACT

Inspired by the results obtained in the string case, we present in this paper the extension of the correction queries to regular tree languages. Relying on Angluin's and Sakakibara's work, we introduce the algorithm $L_{RTL}C$ and we show that regular tree languages are learnable from equivalence and correction queries when the set of contexts is ordered by a Knuth-Bendix order. Moreover, a subclass of regular tree languages, called injective languages, is learned without equivalence queries. This can be extended for other subclasses and may have some practical relevance in fields like machine translation, pattern and speech recognition, building XML documents.

Keywords: Regular tree languages, learning from queries, weight function, correcting context

1. Introduction

The field of grammatical inference was practically introduced by Gold in 1967 [8], when he suggested that learning is an infinite process about making guesses of grammars, and which does not terminate in a finite number of steps but only converges in the limit. Twenty years later, Angluin [1] proposes another popular learning criterion: exact identification using queries. The algorithm, called L^*, allows the Learner to ask questions about a regular language from an oracle (also called MAT - minimally adequate teacher) and halts in polynomial time with a correct description of the language.

The first attempt to extend this inference method to context-free grammars (CFGs) belongs to Angluin as well. In the same paper [1] she designs the algorithm L^{cf} and shows that in this framework context-free languages (CFLs) are learnable. But the learnability is proved under powerful restrictions. Among them, we mention two important ones: the grammar should be in Chomsky normal form and the set of

[1] This work was possible thanks to a FPU Fellowship (AP2004-6968) from the Spanish Ministry of Education and Science