Mach Learn
DOI 10.1007/s10994-013-5403-2

Distributional learning of parallel multiple context-free grammars

Alexander Clark · Ryo Yoshinaka

Received: 8 December 2012 / Accepted: 26 July 2013

Abstract Natural languages require grammars beyond context-free for their description. Here we extend a family of distributional learning algorithms for context-free grammars to the class of Parallel Multiple Context-Free Grammars (PMCFGs). These grammars have two additional operations beyond the simple context-free operation of concatenation: the ability to interleave strings of symbols, and the ability to copy or duplicate strings. This allows the grammars to generate some non-semilinear languages, which are outside the class of mildly context-sensitive grammars. These grammars, if augmented with a suitable feature mechanism, are capable of representing all of the syntactic phenomena that have been claimed to exist in natural language.

We present a learning algorithm for a large subclass of these grammars, that includes all regular languages but not all context-free languages. This algorithm relies on a generalisation of the notion of distribution as a function from tuples of strings to entire sentences; we define nonterminals using finite sets of these functions. Our learning algorithm uses a nonprobabilistic learning paradigm which allows for membership queries as well as positive samples; it runs in polynomial time.

Keywords Mildly context-sensitive · Grammatical inference · Semilinearity

1 Introduction and motivation

Natural languages present some particular challenges for machine learning—primarily the fact that the classes of representations that will ultimately be required for a satisfactory description of natural language syntax are clearly much richer than the simple Markov models

Editors: Jeffrey Heinz, Colin de la Higuera, and Tim Oates.

A. Clark (✉)
Department of Philosophy, King's College London, London, UK
e-mail: alexander.clark@kcl.ac.uk

R. Yoshinaka
Graduate School of Informatics, Kyoto University, Kyoto, Japan
e-mail: ry@i.kyoto-u.ac.jp

Published online: 03 October 2013