In proceedings, Twelfth Cognitive Science Conference, Cambridge, Mass., July 25-28, 1990, pp. 455-562.

Efficient Learning of Language Categories: The Closed-Category Relevance Property and Auxiliary Verbs

Sheldon Nicholl and David C. Wilkins

nicholl@cs.uiuc.edu
wilkins@cs.uiuc.edu
Department of Computer Science
University of Illinois
405 North Mathews Avenue
Urbana, IL 61801

Abstract

This paper describes the mechanism used by the ALACK language acquisition program for identification of auxiliary verbs. Pinker's approach to this problem (Pinker, 1984) is a general learning algorithm that can learn any Boolean function but takes time exponential in the number of feature dimensions. In this paper, we describe an approach that improves upon Pinker's method by introducting the Closed-Category Relevance Property, and showing how it provides the basis of an algorithm that learns the class of Boolean functions that is believed sufficient for natural language, and does not require more than linear time as feature dimensions are added.

1 Introduction

Within the study of language acquisition, the problem of category identification is still a challenge to formal theories of language acquisition. Even the identification of the members of a closed category such as the Auxiliary Verbs (hereinafter referred to as AUX) stands unresolved. The principal approaches to this problem include (Anderson, 1983), (Berwick, 1985), and (Pinker, 1984); Our approach is closest to that of Pinker. We share with Pinker the following five assumptions: (1) Language is learned not from a string of words alone but from the corresponding meaning (and possibly other attributes) as well. (2) Some components of the meaning can be represented with features. (3) The features are drawn from sets we will call *feature dimensions*. Examples of several feature dimensions are shown below in Table 1. (4) Candidate AUXes are not annotated with syntactic features in the input, e.g. (Berwick, 1985), nor is the input prechunked into phrase-like groupings, e.g., (Anderson, 1983). (5) Steele's cross-linguistic generalization (Steele et al., 1981), holds for AUXes: AUXes encode tense or modality or both.

Pinker's approach is based on a general learning algorithm that (i) can learn any Boolean function but (ii) takes time exponential in the number of feature dimensions. In contrast, our approach, which depends on the Closed-Category Relevance Property, defined Section 2, (i) cannot learn an arbitrary Boolean function but is conjectured to be sufficient for natural language, and (ii) does not require more than linear time as feature dimensions are added.