

Grammatical Inference and Learning

Michal Bančák and Roman Vanek

Faculty of Information Technology, Brno University of Technology

{xbanca00,xvanek36}@stud.fit.vutbr.cz

Abstract

Grammatical inference, also called grammar induction is a process of learning a language from the set of example strings. This approach is based on the observation of children learning their first language. The involved set of strings is a concept. The form of the concept in largest research interest is in the form of positive presentation which denotes a set of strings from the specific language. The negative presentation means the opposite. One of the most important models of learning are the deterministic finite automata and their modifications.

The positive presentation learning is the main goal of the research. It focuses on the algorithms of effective learning including the problems of time complexity. One of the learning models includes a teacher, which acts as an oracle, answering queries such as membership or equivalence queries. The grammatical inference can also use the statistical probability methods for decision.

The research on the grammatical inference and learnability has yet been extended to the context-free languages or to the class of boolean formulas. The „query learning“ is a broader framework of learning paradigms including various query types, which are efficiently usable on specific problems. The principles of grammar induction have been applied to the lossless data compression and also other aspects of natural language processing such as example-based translation or morpheme analysis.

This presentation demonstrates a brief introduction to the principles of the grammatical learning methods.