

## Grammatical Inference

Pavol Vargovčík, e-mail: [xvargo01@stud.fit.vutbr.cz](mailto:xvargo01@stud.fit.vutbr.cz),  
Diogo Domingues F. L. C. S., e-mail: [xdomin05@stud.fit.vutbr.cz](mailto:xdomin05@stud.fit.vutbr.cz)

Grammatical inference is a process of learning of grammars from data, similarly to a child learning its language only from analysis of sentences belonging to the language. It is useful in situations when we want to describe a structure of some concept. The concept may be in form of positive presentation, which is a set of strings from the target language, or negative presentation, which also contains strings that do not belong to the target language and are marked as such. The structure is mostly expressed as a minimal automaton that accepts the positive set and rejects the negative set of input data.

Research concerning various situations has been done, it has been mostly focused on subclasses of regular and context-free languages. In some situations there is an Oracle, that is able to answer some types of queries to the learning algorithm. Sometimes it is sufficient to get a probabilistic automaton that generates probabilistic distributions over strings.

Grammatical inference has found its place in syntactic pattern recognition, adaptive intelligent agents, computational biology, log diagnosis, prediction, data mining, natural language acquisition, malware detection, and many more. This work briefly introduces the theory of grammatical inference, describes the actual state of the research, and shows the ideas of some of the basic algorithms from this field in more detail.