PAUTOMAC: a probabilistic automata and hidden Markov models learning competition

Sicco Verwer · Rémi Eyraud · Colin de la Higuera

Abstract Approximating distributions over strings is a hard learning problem. Typical techniques involve using finite state machines as models and attempting to learn these; these machines can either be hand built and then have their weights estimated, or built by grammatical inference techniques: the structure and the weights are then learned simultaneously. The Probabilistic Automata learning Competition (PAUTOMAC), run in 2012, was the first grammatical inference challenge that allowed the comparison between these methods and algorithms. Its main goal was to provide an overview of the state-of-the-art techniques for this hard learning problem. Both artificial data and real data were presented and contestants were to try to estimate the probabilities of strings. The purpose of this paper is to describe some of the technical and intrinsic challenges such a competition has to face, to give a broad state of the art concerning both the problems dealing with learning grammars and finite state machines and the relevant literature. This paper also provides the results of the competition and a brief description and analysis of the different approaches the main participants used.

S. Verwer (✉)
Institute for Computing and Information Sciences, Radboud University Nijmegen, Nijmegen, The Netherlands
e-mail: siccoverwer@gmail.com

R. Eyraud
QARMA team, Laboratoire d'Informatique Fondamentale de Marseille, Marseille, France
e-mail: remi.eyraud@lif.univ-mrs.fr

C. de la Higuera
TALN team, Laboratoire d'Informatique de Nantes Atlantique, Nantes University, Nantes Cedex 1, France
e-mail: cdlh@univ-nantes.fr

       ⓓ Springer