

Learning a Deterministic Finite Automaton With a Recurrent Neural Network

Laura Firoiu, Tim Oates, Paul R. Cohen

Abstract

We consider the problem of learning a finite automaton with recurrent neural networks from positive evidence. We train Elman recurrent neural networks with a set of sentences in a language and extract a finite automaton by clustering the states of the trained network. We observe that the generalizations beyond the training set, in the language recognized by the extracted automaton, are due to the training regime: the network performs a “loose” minimization of the of the training set (the automaton that has a state for each prefix of the sentences in the set).