Deep neural networks learn cellular automaton rules in many-valued logic

Yani Zhang
Chair for Mathematical Information Science, ETH Zurich

Abstract

We develop a theory characterizing the fundamental capability of deep neural networks to learn, from evolution traces, the logical rules governing the behavior of cellular automata (CA). This is accomplished by first establishing a novel connection between CA and Łukasiewicz propositional logic. While binary CA have been known for decades to essentially perform operations in Boolean logic, no such relationship exists for general CA. We demonstrate that many-valued (MV) logic, specifically Łukasiewicz propositional logic, constitutes a suitable language for characterizing general CA as logical machines. This is done by interpolating CA transition functions to continuous piecewise linear functions, which, by virtue of the McNaughton theorem, yield formulae in MV logic characterizing the CA. Recognizing that deep rectified linear unit (ReLU) networks realize continuous piecewise linear functions, it follows that these formulae are naturally extracted from CA evolution traces by deep ReLU networks. A corresponding algorithm together with a software implementation is provided. As the algorithm applies to networks with general, in particular also real-valued, weights, it can also be used to extract logical formulae from deep ReLU networks trained on data.