Capturing Computation with Algorithmic Alignment

Dr Petar Veličković

Google DeepMind

Monday 13 May 2024
1300hrs
LR7, IEB

Abstract:

What makes a neural network better, or worse, at fitting certain tasks? This question is arguably at the heart of neural network architecture design, and it is remarkably hard to answer rigorously. Over the past few years, there have been a plethora of attempts, using various facets of advanced mathematics, to answer this question under various assumptions. One of the most successful directions -- **algorithmic alignment** -- assumes that the target function, and a mechanism for computing it, are completely well-defined and known (i.e. the target is to learn to execute an algorithm). In this setting, fitting a task is equated to capturing the computations of an algorithm, inviting analyses from diverse branches of mathematics and computer science. I will present some of my personal favourite works in algorithmic alignment, along with their implications for building intelligent systems of the future.