COMPUTER SCIENCE & ENGINEERING
UNIVERSITY OF MICHIGAN

DISSERTATION DEFENSE

CHRISTOPHER GRIMM

The Value Equivalence Principle for Model-Based Reinforcement Learning

Tuesday, April 26, 2022
1:00 – 3:00pm
3725 Beyster
Virtual

ABSTRACT: This thesis focuses on model-based reinforcement learning which studies the problem of an agent maximizing reward in an environment using a learned environment simulator (i.e., a model) that enables counterfactual reasoning. Though it is accepted that intelligent agents should be able to reason counterfactually, model-based reinforcement learning methods have only recently begun to outperform model-free methods at scale. Interestingly, these recently successful models are learned to predict the values of future states rather than the future states themselves. This thesis proposes the value equivalence principle as a theoretical account of such models and as a tool to provide insights into how model-based reinforcement learning can be further improved.

This thesis consists of four parts. **First**, we define the value equivalence principle, which divides models into value equivalent model classes based upon properties of their Bellman operators. We study relationships between these classes and show that, when model capacity is limited, models learned to be in certain value equivalent classes outperform conventionally learned models. **Second**, we extend the value equivalence principle to include model classes based upon n-step Bellman operators. We study topological relationships between the resulting hierarchy of higher order value equivalences and highlight proper value equivalent model classes--- which divide models according to which of their value functions match the environment. We show that MuZero learns a proper value equivalent model and leverage our theory to propose a modification to its loss function which results in increased performance. **Third**, we generalize our prior results to the approximate setting, allowing for the analysis of models which are approximately value equivalent or that depend on quantities that can only be approximately computed. These generalizations allow us to provide performance guarantees on significantly broader classes of value equivalent models than before. **Fourth**, we propose a method of refining a given value equivalent model by predicting the future value functions that will arise in its own planning process. We term this approach model foresight, and provide theoretical support that models refined in this way can plan accurately over longer horizons. We conclude with a set of experiments corroborating these claims.

CHAIR: Prof. Satinder Singh Baveja