Multi-Stage Learning: An Effective Learning Paradigm for Multistage Systems

Jiaxing Shen

- Abstract

Data analytics is gaining popularity with the increasing demand for understanding real-world systems. It enables a wide range of system services ranging from diagnosis to prediction. However, there is a huge gap between the large ratio of multistage systems in the real world and little effort of specialized data analytics methods for multistage systems. Multistage systems are ubiquitous including manufacturing systems, medical systems, and education systems. Unlike single-stage systems where all the data are assumed available for analysis, multistage systems have separable yet interdependent subsystems (or stages) whose interactions remain complex and unknown.

Existing data analytics approaches largely ignore the characteristics of multistage systems which leads to limited interpretability and large variances. They focus on only observable factors while the latent information of multistage systems including the interstage structure and relationships is ignored. The information of structure and relationships correspond to how and why subsystems are connected respectively and thus is important in multistage systems.

In this project, we propose Multi-Stage Learning, an effective learning paradigm for data analytics of multistage systems. We aim to optimize the various objectives of multistage systems by modeling interstage structure and relationships, capturing their changing dynamics, and integrating them into machine learning models. The vision also entails several grand challenges including relation complexity, data unavailability, and error propagation. First, different interstage structures and various types of interstage relationships result in the dynamics of modeling inter-stage relationships which is the key factor of MSL. Second, as data flow from one stage to another, upstream stages mostly get little data from downstream stages and thus suffer from data insufficiency. Lastly, both errors and uncertainties propagate and accumulate among various stages which leads to performance gaps.

To address the challenges, we proposed a general research framework that mainly consists of the following three tasks.

- Task 1: Scalable stage relation mining which is to represent interstage structures, interstage relationships as well as their changing dynamics. The uncovered relation would reduce the uncertainties of relation complexity and relief the dilemma of data unavailability.

- Task 2: Hybrid multimodal data fusion including data-driven and model-driven methods so that the interstage knowledge and various data from different stages could be fused. Through an appropriate fusion of knowledge, the errors at each stage will be reduced.

- Task 3: Predictive applications in learning analytics to demonstrate the effectiveness of the proposed MSL methods. This task serves as a means of evaluation.

The proposed research will bridge the gap between multistage systems and data analytics. It will also benefit extensive real-life applications including manufacturing and learning analytics. The progress of this project will enable HK to break new ground as a leading region for big data analytics in multistage systems.