STEVENS
INSTITUTE OF TECHNOLOGY
1870

Ph.D. DISSERTATION DEFENSE

Candidate: Yu Yu
Degree: Doctor of Philosophy
School/Department: School of Engineering and Science
Date: Monday, June 5th, 2023
Time/Location: 11AM EST/ https://stevens.zoom.us/my/xujia
Title: Robust Machine Learning

Chairperson:

Dr. Jia Xu, Department of Computer Science, Stevens Institute of Technology

Committee Members:

Dr. Philippos Mordohai, Department of Computer Science, Stevens Institute of Technology
Dr. Ionut Florescu, School of Business, Stevens Institute of Technology
Dr. Hassan Sajjad, Department of Computer Science, Dalhousie University
Dr. Shusen Wang, Xiaohongshu, Xingin Information Technology Co Ltd

ABSTRACT

Machine learning is a data-driven process that heavily relies on the quality of the data being used. Trivial approaches trade data for performance resulting in heavy, noisy, and erroneous results. In particular, when the real-world test domain is different from the training, the laboratory experiments will not find practical applications.

To bridge this gap, previous work has explored selecting data from a specific target domain to address this discrepancy. However, there are two problems with domain-specific data selection: First, shifting data toward one target domain may fail in the source and other domains. Second, when target domains are unknown, as in the case of most real-world applications, we do not know what future data to receive before model launches.

We select training data without using target-domain information to address these challenges to achieve learning generalization. We propose a set of robustness metrics from a probabilistic perspective to approximate the generalization bound. These metrics are then incorporated into a reinforcement learning framework to enhance the generalization ability of NLP models for classification or generation tasks.

Additionally, we investigate the inter-dependencies among samples to improve generalization performance. We first explore the geometric data diversity with the concept of convex hull volume, dispersion, and graph entropy. We then propose a maximum-entropy rewarded reinforcement learning framework based on information bottleneck theory.

Our design of novel robustness evaluation metrics and our experiments on multiple domain data in NLP tasks such as sentiment analysis, named entity recognition, and language modeling demonstrates the strong generalization power of the reinforcement learning-based data selection framework with almost half of the data.