Statistica Sinica **35** (2025), 1-20
doi:https://doi.org/10.5705/ss.202022.0396

TRANSFER LEARNING FOR HIGH-DIMENSIONAL QUANTILE REGRESSION VIA CONVOLUTION SMOOTHING

Yijiao Zhang and Zhongyi Zhu*

Fudan University

Abstract: This paper studies the high-dimensional quantile regression problem under the transfer learning framework, where possibly related source datasets are available to make improvements on the estimation or prediction based solely on the target data. In the oracle case with known transferable sources, a smoothed two-step transfer learning algorithm based on convolution smoothing is proposed and the ℓ_1/ℓ_2 estimation error bounds of the corresponding estimator are also established. To avoid including non-informative sources, we propose to select the transferable sources adaptively and establish its selection consistency under regular conditions. Monte Carlo simulations as well as an empirical analysis of gene expression data demonstrate the effectiveness of the proposed procedure.

Key words and phrases: High-dimensional data, quantile regression, regularization, smoothing, transfer learning.

1. Introduction

The increasing availability of datasets from multiple sources has provided us with unprecedented opportunities to get a better understanding of the data-limited target problem. For example, for the task of drug sensitivity prediction, the drug response data for the target type of cancer may be limited, but source data for another cancer type may be sufficient (Turki, Wei and Wang, 2017). However, there is no free lunch. Along with the satisfactory sample size of source studies comes the heterogeneity between the source and the target. Intuitively, the more related the source to the target, the more improvement may be made in learning about the target. This motivates transfer learning (Torrey and Shavlik, 2010; Pan and Yang, 2009; Weiss, Khoshgoftaar and Wang, 2016; Niu et al., 2020), which attempts to improve a learner from one domain by transferring information from a related but different domain. A considerable amount of research has clearly shown the success of transfer learning in many real-world applications, including ride dispatching (Wang et al., 2018), medical images analyses (Yu et al., 2022), and human activity recognition (Hirooka et al., 2022), etc.

This paper aims to investigate the effect of transfer learning on quantile regression (QR) in a high-dimensional setting. Ever since the influential work of

*Corresponding author. E-mail: zhuzy@fudan.edu.cn