Commun. Math. Res.
doi: 10.4208/cmr.2022-0046

Vol. **39**, No. 3, pp. 386-413
August 2023

Robust Model Averaging Method Based on LOF Algorithm

Fan Wang, Kang You and Guohua Zou*

School of Mathematical Sciences, Capital Normal University, Beijing 100048, P.R. China.

Received 11 August 2022; Accepted 28 December 2022

Abstract. Model averaging is a good alternative to model selection, which can deal with the uncertainty from model selection process and make full use of the information from various candidate models. However, most of the existing model averaging criteria do not consider the influence of outliers on the estimation procedures. The purpose of this paper is to develop a robust model averaging approach based on the local outlier factor (LOF) algorithm which can downweight the outliers in the covariates. Asymptotic optimality of the proposed robust model averaging estimator is derived under some regularity conditions. Further, we prove the consistency of the LOF-based weight estimator tending to the theoretically optimal weight vector. Numerical studies including Monte Carlo simulations and a real data example are provided to illustrate our proposed methodology.

AMS subject classifications: C51, C53

Key words: Outliers, LOF algorithm, robust model averaging, asymptotic optimality, consistency.

---

1 Introduction

It is no doubt that model selection plays a major role in statistical analysis of real data. In the past decades, there have been many research activities on this

*Corresponding author. *Email addresses:* `wf18801498340@163.com` (F. Wang), `2020160637@jsnu.edu.cn` (K. You), `ghzou@amss.ac.cn` (G. Zou)

topic, and various criteria have been suggested, such as Akaike information criterion (AIC) [1], Bayesian information criterion (BIC) [27], and Mallows's C_p [22]. However, the inference based on model selection ignores the uncertainty from model selection process. Also, the final selected model could still miss important variables, and consequently lead to poor performance [8, 23]. As an alternative to model selection, model averaging combines estimators from various models which can avoid the shortcomings with model selection. Inspired by [12, 14], a number of literature about frequentist model averaging (FMA) have emerged. A main focus of these literature is on finding optimal weights for computing the model averaging estimators. Different from Bayesian model averaging [15], FMA weights do not need to set a prior in advance, and are selected by data only. In recent years, many model averaging criteria have been developed based on different ideas. For instance, [3] presented two weighting schemes by smoothing AIC (S-AIC) and BIC (S-BIC). [12] used the Mallows criterion to select the optimal weights which can be denoted as Mallows model averaging (MMA), and proved that MMA estimator is asymptotically optimal in the sense of minimizing the lowest possible squared error. [20] suggested an optimal weighting scheme that minimizes the trace of an approximately unbiased estimator of the mean squared error (MSE) matrix of the model averaging estimator. [13] proposed a jackknife model averaging (JMA) estimator, where the weights are selected by minimizing the well-known cross-validation criterion. [10] developed a leave-subject-out cross-validation criterion to average time series models. [34] applied a cross-validation model averaging approach to the prediction of a functional response variable.

Note that the aforementioned model selection and averaging methods are built on ordinary least squares or maximum likelihood, which can be badly influenced by the outliers. To reduce the impact of outliers on the traditional model selection and averaging approaches, many robust methods have been proposed. For instance, [24] developed a robust version of the well-known Mallows's C_p. Based on a wide variety of loss functions including absolute and Huber's loss functions, [4] proposed a general robust Akaike-type criterion. [25] suggested a robust model selection approach by using the fast and robust bootstrap. Under high-dimensional regression setting, [18] developed a nonconcave penalized M-estimation method which can perform parameter estimation and variable selection simultaneously. For now, there have been a few literature on robust model averaging. For mean regression, [32] developed an outlier-robust model averaging approach by Mallows-type criterion, and [11] suggested an outlier-robust S_p-type model averaging (SMA) procedure. In the case of quantile regression, [21] developed a jackknife model averaging criterion when the number of parame-