Information Sciences 432 (2018) 199–209

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Locally linear ensemble for regression

Seokho Kang[a], Pilsung Kang[b,*]

[a] Department of Systems Management Engineering, Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu, Suwon 16419, Republic of Korea
[b] School of Industrial Management Engineering, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, Republic of Korea

ARTICLE INFO

Article history:
Received 26 January 2017
Revised 30 October 2017
Accepted 11 December 2017
Available online 11 December 2017

Keywords:
Regression
Locally linear ensemble
Linear model
Ensemble learning
Local expert

ABSTRACT

Considerable research effort has been dedicated to the development of prediction models for yielding greater prediction accuracy in regression problems. Although non-linear models have achieved superior prediction accuracy by addressing the non-linearity of complex data, linear models are still favored because of their high prediction speed. In this study, a locally linear ensemble regression (LLER) is proposed in order to effectively address non-linearity while maintaining the advantage of linear models. The LLER predicts new instances based on multiple linear models that are trained on the regions that identify the local linearity of data. To achieve this, data are decomposed into several locally linear regions based on an expectation-maximization procedure, and linear models are built as local experts for each region to constitute an ensemble. We demonstrate the effectiveness of the LLER through experimental validation with benchmark datasets.

1. Introduction

Regression is a supervised learning task for inferring an underlying functional relationship $y = f(\boldsymbol{x}) + e$ from a set of training instances $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$, where $\boldsymbol{x}_i$ is a vector of input variables and y_i is the corresponding value of a continuous output variable. As earlier work on regression was based on linear models [1–4], they were limited to solving more complex regression problems, particularly when the relationship between input and output variables was not linear. To effectively address non-linear data structures, most recent advances in regression have focused on developing non-linear models, such as tree-based ensembles [5–7], kernel machines [8–11], and artificial neural networks [12–14].

Non-linear models have successfully achieved superior predictive performance for regression problems with non-linear structures; however, they have also introduced several drawbacks. They generally incur a greater computational cost in both the training and test phases, which is undesirable in practical deployment [15]. Moreover, complex non-linear models sometimes fail to provide accurate predictions owing to the possibility of overfitting during model selection as well as model training [16,17].

The aim of this study is to address non-linearity while retaining the advantages of linear models. Although an individual linear model cannot capture a global non-linear structure of the data, such a model is capable of fitting a locally linear region, where the relationship between input and output variables is linear [18,19]. Thus, a non-linear structure comprising multiple locally linear regions can be approximated using multiple linear models, each working as a local expert, *i.e.* a prediction model that is specialized in local prediction, to make a prediction in the corresponding region [20,21]. An

* Corresponding author.
E-mail addresses: s.kang@skku.edu (S. Kang), pilsung_kang@korea.ac.kr (P. Kang).

https://doi.org/10.1016/j.ins.2017.12.022
0020-0255/© 2017 Published by Elsevier Inc.

important consideration is to identify locally linear regions in order to determine the locality in the data while preserving the advantages of linear models.

Here we propose an ensemble learning method, called a locally linear ensemble regression (LLER), to effectively deal with non-linearity using multiple linear models. The LLER decomposes a global non-linear structure of the data into several locally linear regions based on an expectation-maximization (EM) procedure, and builds a linear model for each region to constitute the ensemble. Prediction for a test instance is performed by dynamically combining the linear models that correspond to the local regions near the instance. While the accuracy of the LLER is comparable to that of non-linear models, the LLER also provides fast prediction speed based on the characteristics of the linear models. Furthermore, the reliability of the prediction can be estimated as the disagreement between the outputs from the corresponding linear models, which can be extended to implementing a reject option to avoid low confidence predictions.

The rest of this paper is organized as follows. Section 2 provides a brief overview of related work. Section 3 describes the proposed method. Section 4 reports the experimental results for benchmark datasets. Finally, conclusions and future work are discussed in Section 5.

2. Related work

2.1. Ensemble learning

Ensemble learning refers to methods that integrate the outputs from multiple models to make a prediction. Various ensemble learning methods have been developed for different purposes with the aim of providing improved accuracy and robustness over individual models [22–24]. Most research efforts on ensemble learning have focused on classification problems, while relatively few have paid attention to regression problems [24,25].

An ensemble learning process can be generalized into generation and integration steps. The generation step involves building a set of base models that are diversified by strategies such as data manipulation or using different learning algorithms. The integration step involves combining these base models to make the final prediction. The combination can be performed statically or dynamically, where a static combination uses the same combination rule for all test instances and a dynamic combination uses different rules for different test instances [26,27].

In this study, the learning process of an LLER follows the generation and integration steps; the LLER itself has a focus on utilizing local experts in order to enhance prediction accuracy. Related work on the methods using local experts for prediction is provided in Section 2.2.

2.2. Local experts for prediction

A globally best prediction model for a dataset does not guarantee that the model is more accurate than others for every instance in the dataset [26,28–30]. As local regions can have different characteristics, the most accurate model can also be different for each local region in the dataset [31]. In addition, the complexity of local regions is typically lower than the global complexity of the dataset [32]. Therefore, we can use local experts, the prediction models that are specialized in local prediction, to achieve better prediction accuracy than globally best models. Considerable research effort has been dedicated to the development of various methods to obtain local experts in different ways. We categorize these methods into the three approaches.

The first approach is to choose the locally best prediction models from a pool of global models that are trained on the whole dataset. In this approach, the prediction is performed in a dynamic fashion, where different subsets of the models are combined to predict different instances [23,24]. At the prediction stage, the locally best models for each test instance are identified based on estimating the local accuracy in the neighborhood region of the instance [26,28–30]. Although this approach has been shown to improve prediction accuracy, linear models cannot be used as global models when the global structure of the dataset is not linear.

The second approach is to build prediction models independently for different test instances, which is referred to as instance-based learning [33]. A specialized local model is built for each test instance by focusing on the training instances in the vicinity of the test instance. Subsequently, prediction for an instance is only based on its nearest neighbors. This approach has been implemented with various learning algorithms [19,34–38]. Although local models often defeat global ones in terms of prediction accuracy, they require model training for each test instance during the prediction stage. Linear models are applicable for this approach [19,39]. However, the prediction accuracy is degraded for some test instances whose local neighborhood region is not linear.

The third approach is to use prediction models that are built on different local regions of the dataset. Similar to the second approach, a prediction model is trained by selecting those training instances that are most appropriate for explaining each local region. However, this approach derives the local regions based on a divide-and-conquer strategy, where the original dataset is decomposed into several smaller and less complex subsets [40–42]. The prediction stage is performed in a similar way to its treatment in the first approach, where corresponding prediction models are selected for each test instance to make the prediction by combining their outputs. Linear models are applicable when the local regions are shown to have primarily linear structures.