Neurocomputing 194 (2016) 34–44

Contents lists available at ScienceDirect

Neurocomputing

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

A Gaussian mixture framework for incremental nonparametric regression with topology learning neural networks

Zhiyang Xiang, Zhu Xiao, Dong Wang *, Xiaohong Li

College of Computer Science and Electronics Engineering, Hunan University, Changsha 410082, China

ARTICLE INFO

Article history:
Received 10 May 2015
Received in revised form
29 January 2016
Accepted 12 February 2016
Communicated by B. Hammer
Available online 23 February 2016

Keywords:
Nonparametric regression
Self-organizing incremental neural network
Growing neural gas
Incremental regression
Kernel density estimation

ABSTRACT

Incremental learning is important for memory critical systems, especially when the growth of information technology has pushed the memory and storage costs to limits. Despite the great amount of effort researching into incremental classification paradigms and algorithms, the regression is given far less attention. In this paper, an incremental regression framework that is able to model the linear and nonlinear relationships between response variables and explanatory variables is proposed. A three layer feed-forward neural network structure is devised where the weights of the hidden layer are trained by topology learning neural networks. A Gaussian mixture weighted integrator is used to synthesize from the output of the hidden layer to give smoothed predictions. Two hidden layer parameters learning strategies whether by Growing Neural Gas (GNG) or the single layered Self-Organizing Incremental Neural Network (SOINN) are explored. The GNG strategy is more robust and flexible, and single layered SOINN strategy is less sensitive to parameter settings. Experiments are carried out on an artificial dataset and 6 UCI datasets. The artificial dataset experiments show that the proposed method is able to give predictions more smoothed than K-nearest-neighbor (KNN) and the regression tree. Comparing to the parametric method Support Vector Regression (SVR), the proposed method has significant advantage when learning on data with multi-models. Incremental methods including Passive and Aggressive regression, Online Sequential Extreme Learning Machine, Self-Organizing Maps and Incremental K-means are compared with the proposed method on the UCI datasets, and the results show that the proposed method outperforms them on most datasets.

1. Introduction

With the amount of data growing rapidly in today's social and industrial life, it is beneficial for the data mining applications in these areas to be more space efficient. Not only can space efficient algorithms reduce the cost of data storage, they are important for memory critical systems such as embedded systems and autonomous robots. There has been a great amount of research into incremental and online data mining techniques which can reduce the memory required in learning. However, those researches mostly focus on classification and clustering. For incremental regression there has not been enough attention.

Conventional non-incremental regression methods are divided into two categories by their different approaches to nonlinear predictions, namely the parametric and nonparametric methods. The parametric methods assume that the model generated the data is an analytical model. For example, in Support Vector Regression (SVR) [1], the nonlinear model is assumed to a polynomial or a radiant basis function. The SVR learning is to tune the parameters of the models to reach the minimum error on the training data. In nonparametric regression methods like K-nearest-neighbor (KNN), data are not generalized by an analytical model but instead represented by a subset of the data.

A regression method is to learn a model $Y=f(X)$, where X is the explanatory variables and Y is the response variables. The incremental learning methods process the data in a sequential manner $(X(1),Y(1))$, $(X(2),Y(2))$, …, $(X(t-1),Y(t-1))$, $(X(t),Y(t))$, … In each step, the model is updated by the input $(X(t),Y(t))$ as $f_t \leftarrow f_{t-1}$. In some incremental learning strategies, the input is stored in a buffer with size k and the model is updated every k steps as $f_t \leftarrow f_{t-k}$. We refer a regression method as strictly incremental if and only if $k=1$. There are mainly two kinds of incremental methods like the non-incremental regression methods, namely the parametric and nonparametric methods.

Incremental parametric regression methods are often implemented as stochastic approximations to their non-incremental counterparts. The nonlinear prediction problem is a major challenge to adapt

* Corresponding author. Tel.: +86 13808451678.
E-mail addresses: z_xiang@hnu.edu.cn (Z. Xiang), zhxiao@hnu.edu.cn (Z. Xiao), wangd@hnu.edu.cn (D. Wang), lxh@hnu.edu.cn (X. Li).

http://dx.doi.org/10.1016/j.neucom.2016.02.008
0925-2312/

parametric regression methods for incremental learning. Researches such as the passive aggressive regression [2] are focusing on the linear regression. In parametric methods such as online SVR [3], the kernel parameter tuning which is important for the accuracy would be nearly impossible, since retraining is not allowed when the training is strictly incremental. In [4,5], the nonlinear problem is transformed linear by random features [6]. However, through random feature techniques, those methods introduce low efficiency when dealing with high dimensional prediction problems. There is another approach utilizing the random feature mapping technique, namely the Online Sequential Extreme Learning Machine (OS-ELM) [7]. The main problem of OS-ELM is stability, which can be remedied by ensembles [8]. However, by ensemble the benefit of incremental learning is reduced.

Incremental nonparametric regression methods do not assume linearity or a pre-defined nonlinear model. There are mainly two types of incremental nonparametric regression methods. One is the decision tree methods such as [9]. The other is KNN used in combine of incremental clustering methods such as self-organizing map (SOM) [10]. In decision tree methods, the data space is divided into subspaces such that the nonlinear prediction problem is rendered linear in each of the subspaces. This subspace division approach is less efficient if the data increase in sample size and dimension. Besides, regression trees are unable to give smooth predictions as the parametric methods. KNN has to balance accuracy and efficiency, since the larger the subset selected for knowledge representation, the less efficient the algorithm becomes. Moreover, KNN is flawed in its generalization abilities when the distribution of the data for prediction is different from that of the training data. Another way to make use of incremental clustering abilities for regression is the clusterwise regression [11]. The drawbacks of this framework are the same as regression trees because that it is implemented in a subspace division manner as the regression tree. In summery, comparing to the parametric regression, there are generalization issues in nonparametric methods such as decision tree and incremental clustering based regression.

Incremental clustering methods are adapted for regression as mentioned above. There have been new advances in the area of incremental clustering since the self-organizing maps (SOM) used in [10,11]. One is the data stream clustering such as data stream K-means [12], K-medians [13]. In these methods the clustering objective is to find the low density areas that separate the dense areas where data items are more crowded. Another category is the topology preserving and topology learning methods. SOM is such topology preserving learning methods that learns not only the vector quantization which is similar to the objective of conventional clustering methods, but also keeps the topology relationships of the clustering centers. One advantage of such topology preserving learning is that the data distribution can be represented more accurately [14,15]. SOM is limited in its clustering ability, because it needs a pre-defined topology structure which might contradict the true distribution. Growing neural gas (GNG) [16], on the contrary, can perform topology learning incrementally. The drawback of GNG is that there is no limit for the growing of neurons even when there is no new information in the incoming data. This drawback is remedied in Self-Organizing Incremental Neural Networks (SOINN) [17]. The original SOINN contains two layered structure, but is reduced to single layered structures in later works such as [18,19]. SOINN's incremental topology learning depends on the applicability of Delaunay triangulation construction from the data, thus a limitation is present when the training data are highly concentrated on some dimensions. Theoretically there is no such limitation for GNG. As a result, GNG and SOINN both have their own advantages.

In this paper, we propose an incremental nonparametric regression framework, with the topology learning neural networks as the solution for nonparametric distribution learning, and a Gaussian mixture regression framework for giving smoothed predictions. Two different approaches, GNG regression (GNGR) and the single layered SOINN regression (SOINNR), are explored. Our main contributions are listed as follows:

1. An incremental nonparametric regression framework based on topology learning neural networks is proposed.
2. A two steps regression mechanism is proposed. First, the joint density of explanatory variables and response variables are represented by the clustering results of topology learning neural networks. Second, the joint density function is used in a Gaussian mixture regression model to accomplish the regression task. Moreover, deductions are made to construct the regression function directly from the clustering results of topology learning neural networks.
3. Two different approaches, GNGR and SOINNR, are proposed. Experimental results confirm that GNGR is more scalable, while SOINNR is less sensitive to training parameter settings.
4. Comparing to parametric regression, the parameters in the proposed method can be reset without retraining the model. Comparing to nonparametric regression, the proposed framework is capable of smoothed prediction thus gives better generalization abilities.

Comparison experiments are carried out on an artificial dataset and six UCI datasets. The experiment on the artificial dataset shows that the proposed framework is more smoothed in prediction than KNN and the regression tree. Experiments on the UCI datasets show that the proposed framework outperforms KNN in accuracy and has better performance on most of the datasets than the existing incremental methods.

The rest of the paper is organized as follows. In Section 2, the algorithms of GNG and single layered SOINN which are preliminaries for later sections are introduced. Section 3 details our proposed framework and the algorithms. Section 4 is the experimental results and in Section 5 there are the conclusions.

2. Topology learning neural networks

Assume a data set $\{X\}$ with data points $X(1), X(2), X(3), \ldots, X(i) \in \mathcal{R}^d$ the learning task of GNG [16] and single layered SOINN [19] is that after a single pass scan of the dataset to represent the data by neurons i with weights $W_i \in \mathcal{R}^d$. Their learning objective is formally defined in most literature as a minimization of the reconstruction error [16]

$$\sum_{t=1}^{|\{X\}|} \sum_{i \in N} \omega_i \| X(t) - W_i \|^2 \tag{1}$$

where N is the set of neurons and

$$\omega_i = \begin{cases} 1, & \text{nearest neuron to input } X(t) \text{ is } i \\ 0, & \text{else} \end{cases} \tag{2}$$

The minimization goal stated in Eq. (1) is not so different from a clustering method such as K-means. The unique characteristics of topology learning is that the topology relationships of neurons are represented by a neighborhood function $\eta_{i,j}$. $\eta_{i,j}$ is defined by a graph G, each vertex of which stands for a neuron. If there is an edge between i and j in the graph G, then $\eta_{i,j} = 1$, else $\eta_{i,j} = 0$. In G, if two vertices i, j representing two neurons share one edge, they are referred as topological neighbors in this paper.