Engineering Applications of Artificial Intelligence 45 (2015) 204–219

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

A novel Boosted-neural network ensemble for modeling multi-target regression problems

Esmaeil Hadavandi [a], Jamal Shahrabi [a,*], Shahaboddin Shamshirband [b]

[a] Department of Industrial Engineering and management systems, Amirkabir University of Technology, P.O. Box 15875-4413, Tehran, Iran
[b] Department of Computer System and Technology, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 21 September 2014
Received in revised form
20 May 2015
Accepted 26 June 2015
Available online 25 July 2015

Keywords:
Neural network ensemble
Constrained least mean square
Negative correlation learning
Subspace projection method
Boosting
Multi-target regression

ABSTRACT

In this paper, the concept of ensemble learning is adopted and applied to modeling multi-target regression problems with high-dimensional feature spaces and a small number of instances. A novel neural network ensemble (NNE) model is introduced, called Boosted-NNE based on notions from boosting, subspace projection methods and the negative correlation learning algorithm (NCL). Rather than using an entire feature space for training each component in the Boosted-NNE, a new cluster-based subspace projection method (CLSP) is proposed to automatically construct a low-dimensional input space with focus on the difficult instances in each step of the boosting approach. To enhance diversity in the Boosted-NNE, a new, sequential negative correlation learning algorithm (SNCL) is proposed to train negatively correlated components. Furthermore, the constrained least mean square error (CLMS) algorithm is employed to obtain the optimal weights of components in the combination module. The proposed Boosted-NNE model is compared with other ensemble and single models using four real cases of multi-target regression problems. The experimental results indicate that using the SNCL in combination with the CLSP method offers the capability to improve the diversity and accuracy of the Boosted-NNE. Thus, this model seems a promising alternative for modeling high-dimensional multi-target regression problems.

1. Introduction and literature review

In commercial decision making, planning and controlling are critical. Prediction, a basis for planning and controlling, attempts to predict future trends and is vital for industry planning and operation (Shahrabi et al., 2013). In most common prediction problem settings, the value of a single target numeric attribute is predicted. A natural generalization of this setting is to predict multi-target numeric attributes (Aho et al., 2012) simultaneously, because in many real-life predictive modeling problems, the output is structured, meaning there can be dependencies or internal relations between targets. These problems generally exhibit complex behaviors, high-dimensional feature spaces and a small number of instances (Kocev et al., 2013).

Artificial intelligence (AI) methods, such as artificial neural networks (ANNs) (Rumelhart and McClelland, 1986) are popular tools for solving complex engineering and optimization problems. One approach to deal with complex, real-world problems is to combine AI prediction models and form an ensemble of predictors that exploit the different local behaviors of the base models to improve the overall prediction system's performance (Masoudnia et al., 2012). The main objective of ensemble learning methods is to simplify a difficult prediction task by dividing it into some relatively easy prediction subtasks and formulating a consensus prediction result for the original data (García-Pedrajas et al., 2012). From another perspective, ensemble learning is an approach to enhance the prediction accuracy for complex problems, such as those involving a limited number of instances, high-dimensional feature sets, and highly complex trends and behaviors (Kotsiantis, 2011).

Among the prevalent ensemble approaches is the neural network ensemble (NNE) (Hansen and Salamon, 1990) that consists of a finite number of NNs and has been intensively studied in recent years (Tian et al., 2012; Zhai et al., 2012). Techniques using NNE models usually comprise two independent phases: using a method to create individual NN components and using a method to combine NN components. Both theoretical and experimental studies have shown that an NNE model is more effective when its components' estimates are negatively correlated, it is moderately effective when its components are uncorrelated and it is only mildly effective when its components are

* Corresponding author. Tel.: +98 21 66413034; fax: +98 21 66413025.
E-mail addresses: es.hadavandi@aut.ac.ir (E. Hadavandi), jamalshahrabi@aut.ac.ir (J. Shahrabi).

http://dx.doi.org/10.1016/j.engappai.2015.06.022
0952-1976/

positively correlated (Brown et al., 2005). There are some common approaches to produce accurate and negatively correlated components for NNE models, such as manipulating the training data set or using the penalty method, whereby a penalty term is added to the error function of the NNE model (Brown et al., 2005).

Two popular methods of constructing ensembles by manipulating the instances to independently and sequentially train individual NN components are bagging (bootstrap aggregating) (Breiman, 1996) and boosting (Freund and Schapire, 1996), respectively. Bagging only generates different bootstrap samples from the original training set for training various components. Boosting adaptively changes the distribution of the training set based on the performance of the previously added components to the ensemble. The most widely employed boosting method is AdaBoost. It is based on adaptively increasing the probability of sampling instances with greater prediction errors using previous components. Different versions of AdaBoost available for regression problems include AdaBoost. RT (Solomatine and Shrestha, 2004). The idea behind it is to filter out examples with higher relative estimation error than a pre-set threshold value, and then to follow the AdaBoost procedure. Asymmetric-AdaBoost is another extension of AdaBoost, which only alters the procedure for weight updating. It is a cost-sensitive boosting algorithm based on the statistical interpretation of boosting to enable deriving a principled asymmetric boosting loss, which, much like the original AdaBoost, is then minimized by gradient descent in the functional space of convex combinations of weak learners (Masnadi-Shirazi and Vasconcelos, 2007). One of the causes of boosting failure is putting too much emphasis on correctly classifying all instances. Outliers, or noisy instances become too relevant in the training set, undermining the ensemble's performance (García-Pedrajas et al., 2012).

The feature space can be manipulated by randomly selecting feature subspaces from the original feature space (Garc´ıa-Pedrajas et al., 2007). The most widely applied ensemble learning method that manipulates the feature space is the Random Subspaces Method (RSM) (Ho, 1998). Some ensemble methods are based on constructing each component using features obtained by rotating subspaces of the original dataset (Rodriguez et al., 2006; Xiong et al., 2015). The notion of using subspace projection methods to construct NNE models has been applied in different works (Garc´ıa-Pedrajas et al., 2007; Garc´ıa-Pedrajas and Ortiz-Boyer, 2008). More recently, García-Pedrajas et al. (2012) proposed a new boosting-based method for designing classifier ensembles according to unsupervised and supervised projection of random subspaces using different projection methods. The projections were constructed using only misclassified instances to focus the next classifier on the most difficult dataset instances.

A popular penalty method for creating explicitly diverse components is negative correlation learning (NCL) (Liu and Yao, 1999). The key idea behind NCL is to introduce a correlation penalty term to the cost function of individual NN components so that each component minimizes its mean square error (MSE) together with the ensemble's error correlation (Masoudnia et al., 2012). Alhamdoosh and Wang (2014) incorporated random vector functional link (RVFL) networks as base components with the NCL strategy to build neural network ensembles. Lee et al. (2012) proposed a new selective neural network ensemble with negative correlation. A set of component networks were chosen to build an ensemble so the generalization error would be minimized and the negative correlation maximized. The mentioned advantages and limitations of NCL and subspace projection methods are summarized in Table 1.

In this paper, as a first study in the literature, a novel Boosted-NNE model is developed for high-dimensional multi-target regression problems with a small number of instances based on several ideas from boosting, subspace projection and a new NCL algorithm. The proposed model is constructed according to the following contributions:

(1) The first contribution is to develop a new cluster-based subspace projection method (CLSP) to construct a low-dimensional feature space for training each component network rather than using the entire input feature space to train the components. CLSP uses a clustering algorithm on the original input features, which is combined with the boosting principle of directing learning on the difficult instances. As for subspaces, a smaller number of input features allows obtaining diverse and negatively correlated components for the Boosted-NNE.
(2) As the second contribution, the sequential negative correlation learning algorithm (SNCL) is proposed to train the Boosted-NNE model components in order to improve the diversity among individual component networks. In SNCL, the penalty term in the error function of component t measures the error correlation between itself and previous components $(1, 2 \ldots, t-1)$. This way, it is hoped the NNE model components trained by SNCL are pushed away from each other so that the total distance between them expands.
(3) Moreover, the constrained least mean square error (CLMS) algorithm is utilized to obtain the optimal weights of the components in the Boosted-NNE model's combination module.

The Boosted-NNE model features are compared with other ensemble models to demonstrate its advantages (see Table 2). The listed models are constructed based on combinations of a number of ensemble creation methods.

The rest of the paper is organized as follows. In Section 2, the Boosted-NNE model is presented in detail. In Sections 3 and 4, the proposed model is compared with the existing ensemble models using four real datasets of multi-target regression problems. Finally, conclusions and future works are discussed in Section 5.

2. Methodology formulation

A multi-target regression problem is defined as follows:

An input space X that consists of multiple features (in our case these are continuous), i.e., $\forall X_n \in X$, $X_n = (x_{n1}, x_{n2}, \ldots, x_{nD})$, where D is the number of input features.

A target space Y that consists of multiple features (in our case these are continuous), i.e., $\forall Y_n \in Y, Y_n = (y_{n1}, y_{n2}, \ldots, y_{nP})$, where P is the number of target features.

A set of instances S, where each instance is a pair of input-target features, i.e., $S = \{(X_n, Y_n) | X_n \epsilon X,\ Y_n \epsilon Y\ ,\ 1 \le n \le N\}$ and N is the number of instances of S ($N = |S|$).

A quality criterion such as error function E rewards models with low predictive error.