Computers in Human Behavior 64 (2016) 641–655

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Full length article

A novel evolutionary-negative correlated mixture of experts model in tourism demand estimation

S.M.R. Kazemi [a], Esmaeil Hadavandi [a, *], Shahaboddin Shamshirband [b, *], Shahrokh Asadi [c]

[a] Department of Industrial Engineering, Birjand University of Technology, Birjand, Iran
[b] Department of Computer System and Information Technology, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia
[c] Faculty of Engineering, Farabi Campus, University of Tehran, Iran

ARTICLE INFO

Article history:
Received 1 May 2016
Received in revised form
29 June 2016
Accepted 6 July 2016

Keywords:
Mixture of experts
Negative correlation learning
Subspace projection
Time series estimation
Tourism demand

ABSTRACT

Mixtures of experts (ME) model are widely used in many different areas as a recognized ensemble learning approach to account for nonlinearities and other complexities in the data, such as time series estimation. With the aim of developing an accurate tourism demand time series estimation model, a mixture of experts model called LSPME (Lag Space Projected ME) is presented by combining ideas from subspace projection methods and negative correlation learning (NCL). The LSPME uses a new cluster-based lag space projection (CLSP) method to automatically obtain input space to train each expert focused on the difficult instances at each step of the boosting approach. For training experts of the LSPME, a new NCL algorithm called Sequential Evolutionary NCL algorithm (SENCL) is proposed that uses a moving average for the correlation penalty term in the error function of each expert to measure the error correlation between it and its previous experts. The LSPME model was compared with other ensemble models using monthly tourist arrivals to Japan from four markets: The United States, United Kingdom, Hong Kong and Taiwan. The experimental results show that the estimation accuracy of the proposed LSPME model is significantly better than the other ensemble models and can be considered to be a promising alternative for time series estimation problems.

1. Introduction and literature review

Tourism demand estimation is an important issue in the tourism industry and is generally seen to be one of the most complex functions of tourism management(Hadavandi, Ghanbari, Shahanaghi, & Abbasian-Naghneh, 2011). The studies on the pattern of the tourism demand from different origins are essential for the tourism-related industries to formulate efficient strategies on maintaining and boosting the tourism industry in a country. Planning strategic and operational decisions relies on accurate and robust estimation models. Government would be able of developing well-organized tourism strategies and proper infrastructures for serving visitors via these estimated trends on the coming tourism demand. Moreover, private sector can enjoy the merits by developing appropriate marketing strategies (Gretzel, Werthner, Koo, & Lamsfus, 2015; Hadavandi, Ghanbari, Shahanaghi, & Abbasian, 2011; Shahrabi, Hadavandi, & Asadi, 2013; Sigala, 2011). The most popular tourism demand time series estimation models are Autoregressive Integrated Moving Average (ARIMA), error correction model (ECM) and the vector autoregressive (VAR) models (Song & Witt, 2006) that use piecewise linear functions as basic elements of the estimation model. However, relationships between the demand trends and influencing factors in many time series are nonlinear and complex and it is hard or even impossible to obtain an accurate linear model describing the relationships (Shahrabi et al., 2013).

Nowadays, Artificial Intelligence (AI) (Asadi & Shahrabi, 2016a,b) models such as artificial neural networks (ANNs) (Rumelhart & McClelland, 1986) are used in time series estimation problems (Hadavandi, Ghanbari, Shahanaghi, & Abbasian, 2011; Hadavandi, Ghanbari, Shahanaghi, & Abbasian-Naghneh, 2011). These models have more flexibility than traditional linear ones and can be used to estimate the nonlinear relationships (Hadavandi, Ghanbari, Shahanaghi, & Abbasian, 2011; Hadavandi, Ghanbari,

* Corresponding author.
E-mail addresses: kazemi@birjandut.ac.ir (S.M.R. Kazemi), es.hadavandi@aut.ac.ir (E. Hadavandi), shamshirband@um.edu.my (S. Shamshirband), s.asadi520@gmail.com (S. Asadi).

http://dx.doi.org/10.1016/j.chb.2016.07.004
0747-5632/

Shahanaghi, & Abbasian-Naghneh, 2011; Shahrabi et al., 2013). One approach to deal with time series problems is combining the estimation models and forming an ensemble of models to exploit local different behavior of the base models (experts) to improve accuracy of the overall estimation system (Hadavandi, Shahrabi, & Shamshirband, 2015; Masoudnia & Ebrahimpour, 2012). Ensemble learning as one of the fields of AI consisting of a combination of different experts, homogeneous and heterogenous, is capable of modeling a complex problem jointly (Dietterich, 2000). From another view-point, ensemble learning is an approach for improving accuracy of a single model in complex problems such as those involving high-dimensional feature spaces, and highly complex behaviors (Hadavandi, Shahrabi, & Hayashi, 2015).

There are two stages for constructing an ensemble model: (i) creation of individual experts and (ii) combination of the experts. Both theoretical and experimental studies illustrate that an ensemble model is more effective when the experts' estimates are negatively correlated, but this procedure is moderately effective when the experts are uncorrelated and only mildly effective when the experts are positively correlated (Brown, W, H, & Y, 2005; Hadavandi, Shahrabi, & Hayashi, 2015; Hadavandi, Shahrabi, & Shamshirband, 2015; Masoudnia, Ebrahimpour, & Arani, 2012). Some methods are common in constructing accurate and diverse (negatively correlated) experts for the ensemble such as manipulating the training set or using penalty methods, where a penalty term is added to the error function of a neural network ensemble (Brown et al., 2005).

Bagging (Bootstrap aggregating) (Breiman, 1996) and Boosting (Freund & Schapire, 1996), two popular methods for training individual experts of an ensemble model, can be applied independently and sequentially, respectively. Bagging methods use bootstrap sampling for training different experts of an ensemble. Boosting methods adaptively change the distribution of the training instances based on the performance of the previous experts. The most widely used boosting method is AdaBoost. It is based on adaptively increasing the probability of sampling the instances that have higher estimation errors by the previous experts.

The manipulation of feature space can be done by random selection of feature subspaces from the original feature space (Garćıa-Pedrajas, Garćıa-Osorio, & Fyfe, 2007). The most widely used ensemble learning method that manipulates the feature space is the Random Subspaces Method (RSM) (Ho, 1998). Recently, García-Pedrajas et al. (2012), (García-Pedrajas, Maudes-Raedo, García-Osorio, & Rodríguez-Díez, 2012) proposed a method based on the combination of subspace projection methods (SPM) and boosting approach for ensemble construction. The method uses feature extraction models for rotating original training space to a lower dimension space.

Negative correlation learning (NCL) (Liu & Yao, 1999) and mixture of experts (ME) (Jacobs, Jordan, Nowlan, & Hinton, 1991) are popular penalty approaches for constructing diverse ensemble models with neural networks (NNs) experts. In the NCL a correlation penalty term is introduced to the cost function of individual NNs so that each NN can minimize its mean square error (MSE) together with the error correlation of the ensemble. There is a regularization term that provides a convenient way to balance the bias–var–cov trade-off, thus improving the generalization ability (Alhamdoosh & Wang, 2014; Hadavandi, Shahrabi, & Hayashi, 2015). In NCL, previously used static combiner methods do not have the capability to model the local competence of the experts (Chen & Yao, 2009). The ME is based on the Divide-and-Conquer (D&C) principle in which the complex problems are partitioned into a set of simpler sub-problems and are distributed among the experts. It uses a gating network to compute the combination weights dynamically from the inputs, according to the local efficiency of each expert (Ebrahimpour, Sadeghnejad, SAAA, & Mohammadi, 2012; Kheradpisheh, Sharifizadeh, Nowzari-Dalini, Ganjtabesh, & Ebrahimpour, 2014; Yuksel, Wilson, & Gader, 2012). The classical ME models are Inefficient in high-dimensional feature space problems and do not control over the bias–var–cov trade-off (Hadavandi, Shahrabi, & Hayashi, 2015).

Recently some mixture models have been proposed for time series estimation in the literature. The self-organizing mixture autoregressive (SOMAR) (Ni & Yin, 2009) and generalized SOMAR (GSOMAR) models (Yin & Ni, 2009) were presented to tackle nonlinear and nonstationary time series. These models contain a number of autoregressive models that are learnt and organized in a self-organized manner by the adaptive least mean square algorithm. In the SOMAR model, nonlinear time series is considered as a mixture of linear AR models. As an extension of GSOMAR, Ouyang and Yin (2014), (Ouyang & Yin, 2014) proposed the Neural gas mixture of autoregressive models (NGMAR). It organized the local mixture autoregressive models as neural gas and used them to describe the behavior of time series. The characteristics of the LSPME in comparison with mixture models for time series estimation are presented in Table 1:

In this paper, we develop a new ME model called LSPME (Lag Space Projected ME) for tourism demand estimation problems based on a combination of ideas from subspace projection methods, Boosting and a new NCL algorithm. The proposed model uses a new cluster-based lag space projection method (CLSP) to automatically construct subspaces for training experts of the ME model using the set of candidate lags of a time series.

To improve the diversity among the experts, training the LSPME is performed using a new NCL method called the sequential evolutionary negative correlation learning algorithm (SENCL). In SENCL, the penalty term in the error function of expert t is a moving average and measures the error correlation between it and its previous experts ($1,2\ldots,t-1$). It uses a genetic algorithm to evolve the weights of NN experts, so we hope the experts of the LSPME trained by SENCL are pushed away from each other, and thus the total distance between them grows.

Table 1
Comparison of the LSPME model with the other mixture models in time series forecasting.

Model	Wight updating	Input space construction
SOMAR, GSOMAR (Ni & Yin, 2009) and NGMAR (Ouyang & Yin, 2014)	**Static** ✓ Uses a static set of weights for AR models	**Information criteria tests** ✓ Uses Information criteria test such as BIC to select the lag order of AR models
LSPME	**Dynamic** ✓ Uses a gating network to compute the combination weights dynamically from the inputs, according to the local efficiency of each expert	**CLSP** ✓ Uses the idea of clustering lags of time series for constructing input space in the ME model focused on difficult instances at each step of boosting approach.