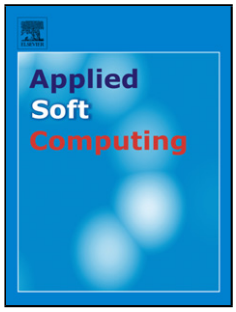

Title: Model forecasting based on two-stage feature selection procedure using orthogonal greedy algorithm

Author: He Jiang

PII: S1568-4946(17)30710-X
DOI: https://doi.org/10.1016/j.asoc.2017.11.047
Reference: ASOC 4591

To appear in: *Applied Soft Computing*

Received date: 23-6-2017
Revised date: 27-11-2017
Accepted date: 28-11-2017

Please cite this article as: He Jiang, Model forecasting based on two-stage feature selection procedure using orthogonal greedy algorithm, Applied Soft Computing Journal https://doi.org/10.1016/j.asoc.2017.11.047

Model forecasting based on two-stage feature selection procedure using orthogonal greedy algorithm

He Jiang[ab]

[a] *School of Statistics, Jiangxi University of Finance and Economics, Nanchang 330013, China*

[b] *Applied Statistics Research Center, Jiangxi University of Finance and Economics, Nanchang 330013, China*

*Corresponding author: He Jiang. Email: jiangsky2005@aliyun.com

---

Highlights

- BPNN is applied to depict the nonlinear relationships between the features
- OGA is used to select the important features and interaction terms
- A simple-to-implement and efficient algorithm is designed
- The forecast error bounds of the proposed model TSOGA are derived
- The real data experiments are given to support the proposed model

Abstract

Currently, forecasting and feature selection tasks are attracting considerable attention from various scientific fields including global solar radiation forecasting, signal processing, microarray data analysis, finance, medicine and others. However, both selection inconsistency and the intractable computational cost pose critical difficulties when implementing forecasting tasks. Although artificial neural networks (ANNs) are useful for forecasting, a large number of nuisance features are employed. To establish an interpretable forecasting model, feature selection techniques are combined with ANNs to reduce the number of inputs and the complexity of network structures. However, these approaches shrink the estimate, which results in inaccurate forecasting results. To overcome these drawbacks, this paper successfully investigates a novel soft computing approach referred to as a two-stage feature selection procedure using the orthogonal greedy algorithm (TSOGA) to select the important features as inputs of ANNs. A simple-to-implement and efficient computational algorithm is designed, and the theoretical analysis is also provided. Furthermore, the high dimensional Bayesian information criterion (HDBIC) is utilized to select the optimal forecasting model. Real data experiments directly demonstrate the outstanding