Applied Soft Computing 9 (2009) 563–574

Contents lists available at ScienceDirect

Applied Soft Computing

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

A neural-network-based nonlinear metamodeling approach to financial time series forecasting

Lean Yu [a,*], Shouyang Wang [a], Kin Keung Lai [b]

[a] Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
[b] Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 17 December 2007
Received in revised form 29 July 2008
Accepted 10 August 2008
Available online 19 August 2008

Keywords:
Artificial neural networks
Metamodeling
Data sampling
Meta-learning
PCA
Financial time series forecasting

ABSTRACT

In financial time series forecasting, the problem that we often encounter is how to increase the prediction accuracy as possible using the financial data with noise. In this study, we discuss the use of supervised neural networks as a meta-learning technique to design a financial time series forecasting system to solve this problem. In this system, some data sampling techniques are first used to generate different training subsets from the original datasets. In terms of these different training subsets, different neural networks with different initial conditions or training algorithms are then trained to formulate different prediction models, i.e., base models. Subsequently, to improve the efficiency of predictions of metamodeling, the principal component analysis (PCA) technique is used as a pruning tool to generate an optimal set of base models. Finally, a neural-network-based nonlinear metamodel can be produced by learning from the selected base models, so as to improve the prediction accuracy. For illustration and verification purposes, the proposed metamodel is conducted on four typical financial time series. Empirical results obtained reveal that the proposed neural-network-based nonlinear metamodeling technique is a very promising approach to financial time series forecasting.

1. Introduction

World financial markets function in a very complex and dynamic manner where high volatility and noisy data are routine. Many factors impact financial markets, including political events, general economic conditions, and even traders' expectations. Due to the high degrees of irregularity and nonlinearity, financial time series forecasting is regarded as a rather challenging task [1,2]. Empirical investigations have increasingly revealed that movements in financial markets are not random and the markets behave in a highly nonlinear, dynamic manner. Future prices movements are often assumed to be "the standard random walk" but in reality the randomness may only be a veil that shrouds a noisy nonlinear process [3–5]. Therefore, for traditional linear models such as autoregressive integrated moving average (ARIMA), it is extremely difficult to capture the irregularity and nonlinearity hidden in financial time series. In the past decades, many emerging artificial intelligent (AI) techniques such as artificial neural networks (ANNs) were widely used in financial time series forecasting and good prediction performance was obtained [2,6].

However, ANNs are a kind of unstable learning methods, i.e., small changes in the training set and/or parameter selection can produce large changes in the prediction. Results of many experiments have shown that the generalization of a single neural network is not unique. In other words, a neural network's results are not stable. Even for some simple problems, different structures of neural networks (e.g., different numbers of hidden layers, nodes and initial conditions) often result in different patterns of generalization. In addition, even the most powerful neural-network model cannot cope well with complex data sets containing some random errors or insufficient training data. Thus, the performance for these data sets may not be as good as expected [1,6,8].

Furthermore, many studies have found that ANNs are far from being optimal learners. For example, some existing studies (e.g., [7–9]) have found that the ways that the ways in which neural networks get to the global minima vary; some networks just settle into local minima instead of global minima through the analysis of error distributions. In this case, it is hard to identify which neural network's error reaches the global minima, if the error rate is not zero. Thus, it is not wise to select a single neural-network model, even with the best generalization, from a limited number of neural networks, if the error is larger than zero. In financial time series forecasting, it is difficult to obtain a consistently good result by using a single neural-network model due to high volatility and

* Corresponding author. Tel.: +86 10 62565817; fax: +86 10 62568364.
E-mail address: yulean@amss.ac.cn (L. Yu).

1568-4946/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2008.08.001

Nomenclature

a_j	the bias of the jth node in FNN model
A	parameter vector of model A
ANN	artificial neural networks
ARIMA	autoregressive integrated moving average
b_{ij}	the connection weights between layers in FNN models
CV	cross-validation
$d(x)$	target value or expected value
d_t	an indicator function, $d_t = \begin{cases} 1, & (x_t - x_{t-1})(\hat{x}_t - x_{t-1}) \geq 0 \\ 0, & \text{otherwise} \end{cases}$
DS	original data set
D_{cv}	a cross-validation dataset
D_{stat}	directional statistics
EUR	Euros
$f(x)$	an output function depending on x
$\hat{f}^{(j)}$	the j-step ahead prediction results
FNN	feed-forward neural network
G	data dimensions
H	the number of the components
JPY	Japanese yen
L	learner
M	the number of the selected based models
ML	meta-learner
MSE	mean squared error
MT	meta-training set
N	the number of the base models
NRMSE	normalized root mean squared error
NYSE	New York Stock Exchange index
P	the number of input nodes in FNN models
PCA	principal component analysis
Q	the number of hidden nodes in FNN models
R	eigenvector
SVM	support vector machines
S&P500	Standard & Poor 500 stock index
T	the number of testing samples
TR	training set
TR_i	the ith training subset
TS	testing set
VS	validation set
w_i	the assigned weight of each base model
(x, y)	a training pair, where x is independent variable and y is the dependent variable
$(x_{t-1}, \ldots, x_{t-s})$	a time series with s observations
$\hat{x}_t$	the prediction value at time t
Y	forecasting result matrix

Greek letters

Θ	vectors of all parameters in FNN models
θ	a specified threshold to determine principal components; all parameters in FNN models
λ	eigenvalue
σ^2	error variance
$\psi(\cdot)$	a nonlinear function determined by FNN model
$\varphi(\cdot)$	the transfer function of hidden layer

irregularity in financial markets. More and more researchers have realized that just selecting a single neural-network model with the best performance may lead to loss of potentially valuable information contained by other neural-network models that may have slightly weaker performances, relative to the best neural-network model [6–10]. Therefore, some different learning strategies such as combined or ensemble learning [2,6–9] and meta-learning [1,6,10,11] that consider the discarded neural networks also, whose performances are less accurate than the best neural-network model, have been presented. For this purpose, neural-network-based metamodeling using meta-learning strategy [10,11] is introduced in this study.

Meta-learning [1,6,10,11], which is defined as an attempt to learn from the learning process itself, provides a promising solution and a novel approach to the above challenges. In fact, meta-learning is similar to ensemble learning. But they belong to two different types or two different subfields of machine learning, which study algorithms and architectures that build collections or combinations of some algorithms that are more accurate than a single algorithm. The generic idea of meta-learning is to use some individual learning algorithms to extract knowledge from several different data subsets and then use the knowledge from these individual learning algorithms to create a unified body of knowledge that adequately represents the entire dataset. Therefore, meta-learning seeks to construct a metamodel that integrates, following certain principles, the separately learned models to boost overall predictive accuracy [11]. Different from combined or ensemble learning, which is defined as combining multiple learning models, under the assumption that "two (or more) heads are better than one" to produce more generalized results, the goal of meta-learning is to use metadata to understand how learning itself can become flexible and/or adaptable. It takes into account the domain or task under study, while solving different kinds of learning problems, and hence produces a metamodel to improve the performance of existing learning algorithms.

Although there are many studies on meta-learning [6,10–12], we find that there are two main problems in the existing meta-learning models. In the existing studies, metamodels are often produced in the linear weighted form. That is, some metadata from different base learning algorithms are combined in a linear way. However, a linear weighted approach is not necessarily appropriate for all circumstances (Problem I). Furthermore, it is not easy to determine the number of individual learning models. It is well known to us that the rule of "the more, the better" is not valid under all circumstances. Thus, it is necessary to choose an appropriate method to determine the number of individual learning algorithms required for producing a metamodel with better generalization ability (Problem II).

In view of the first problem (i.e., linear weighted drawback), a nonlinear weighted metamodel utilizing the neural-network technique is introduced in this paper. For the second problem (i.e., determining the number of individual learning models for formulating a metamodel), the principal component analysis (PCA) technique is introduced. That is, we utilize the PCA technique to choose the number of individual learning algorithms.

Considering the previous two main problems, this study proposes a four-stage neural-network-based nonlinear weighted metamodeling technique to predict financial time series. In the first stage, some data sampling techniques are used to generate different training sets, a validation set and a testing set. Based on the different training sets, different neural-network models with different initial conditions or different training algorithms are then trained to formulate different neural-network base models in the second stage. These base models' training processes do not affect the overall efficiency of time series forecasting system due to the