Neurocomputing 147 (2015) 239–250

Contents lists available at ScienceDirect

Neurocomputing

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

Port throughput forecasting by MARS-*R*SVR with chaotic simulated annealing particle swarm optimization algorithm

Jing Geng [a], Ming-Wei Li [a,*], Zhi-Hui Dong [a], Yu-Sheng Liao [b]

[a] College of Shipbuilding Engineering, Harbin Engineering, University, Harbin 150001, Heilongjiang, China
[b] Department of Healthcare Administration, Oriental Institute of Technology, No. 58, Sec. 2, Sichuan Rd., Panchiao, New Taipei, Taiwan, ROC

ARTICLE INFO

Article history:
Received 30 April 2014
Received in revised form
22 June 2014
Accepted 22 June 2014
Communicated by Wei Chiang Hong
Available online 8 July 2014

Keywords:
Port throughput
Forecasting
Chaotic mapping
Particle swarm optimization (PSO)
Simulated annealing (SA)
Robust ν-support vector regression (*R*SVR)

ABSTRACT

Port throughput forecasting is a very complex nonlinear dynamic process, prediction accuracy is influenced by uncertainty of socio-economic factors, especially by the mixed noise (singular point) produced in the collection, transfer and calculation of statistical data; consequently, it is difficult to obtain a satisfactory port throughput forecasting result. Thus, establishing an effective port throughput forecasting scheme is still a significant research issue. Since the robust ν-support vector regression model (*R*SVR) has the ability to solve the nonlinear and mixed noise in the port throughput history data and its related socio-economic factors, this paper introduces the *R*SVR model to forecast port throughput. In order to search the more appropriate parameters combination for the *R*SVR model, considering the proposed simulated annealing particle swarm optimization (SAPSO) algorithm and the original PSO algorithm still have the drawbacks of immature convergence and is time consuming, this study presents chaotic simulated annealing particle swarm optimization(CSAPSO) algorithm to determine the parameter combination. Aiming to identify the final input vectors for *R*SVR model, the multivariable adaptive regression splines (MARS) is adopted to select the final input vectors from the candidate input variables. This study eventually proposes a port throughput forecasting scheme that hybridizes the *R*SVR, CSAPSO and MARS to obtain a more accurate forecasting result. Subsequently, this study compiles the port throughput data and the corresponding socio-economic indicators data of Shanghai as the illustrative example to evaluate the feasibility and performance of the proposed scheme. The experimental results indicate that the proposed port throughput forecasting scheme obtains better forecasting result than the six competing models in terms of forecasting error.

1. Introduction

The prediction of port throughput has the basic function of making port strategic decision, port developing scale, port general layout and port district division. If this prediction is not accurate enough, bias in policy will occur, which may cause huge financial losses. Therefore, developing an effective port throughput forecasting model has become a crucial and challenging task.

Numerous forecasting approaches have been developed for port throughput prediction. In the conventional quantitative forecasting approaches, the autoregressive moving integrated moving average (ARIMA) models [1] are the most popular and practical time series forecasting. It is often applied to forecast the series when data are inadequate to construct econometric, or when knowledge of the structure of forecasting models is limited. Time series models are simple in calculation and fast in speed, and are likely to outperform other models in some cases, especially in short-term forecasting [2,3]. Therefore it is widely used in port throughput prediction [4–7]. However, time-series forecasting models fail to reflect other related factors of the predicting series.

Artificial neural network (ANN) is primarily based on a model of emulating the processing of human neurological system identify related spatial and temporal characteristics from the historical data patterns (particularly for nonlinear and dynamic evolutions); therefore, they can approximate any level of complexity and do not need prior knowledge of problem solving. Since port throughput prediction is too complex to be solved by a single linear statistical algorithm, ANN should be considered as an alternative for solving port throughput forecasting. Owing to the superior performance to approximate any degree of complexity and requiring no prior knowledge of problem solving, ANN models [8–10] have been widely applied in port throughput forecasting [11]. Although ANN-based forecasting models can approximate to any function, particularly nonlinear functions, they have difficulties in the non-convex problem of network training errors, explaining

* Corresponding author. Tel.: +886 18845615108; fax: +886 0451 82568317.
E-mail addresses: gengjing@hrbeu.edu.com (J. Geng), limingwei@hrbeu.edu.cn, mingweili123@gmail.com (M.-W. Li), fg002@mail.oit.edu.tw (Y.-S. Liao).

http://dx.doi.org/10.1016/j.neucom.2014.06.070
0925-2312/

black-box operations, can easily be trapped in local minima [12,13], have time-consuming training procedures, and have subjectivity in selecting an ANN model architecture [14]. Additionally, the training of an ANN model requires a large amount of training samples, while port throughput and related impact indicators have limited datum. Thus, ANN models cannot achieve satisfactory performance in port throughput forecasting.

Support vector regression (SVR) has overcome the inherent defects of the ANN model [15], with the structure risk minimization criterion. It possesses not only greater nonlinear modeling ability, but also several superior advantages, such as theoretically ensuring global optimum, simple modeling structure and processes, and small sample popularization requirement; therefore, SVR-based models [16] have been successfully applied to solve forecasting problems in many fields, such as financial time series forecasting [17–21], tourist arrival prediction [22,23], atmospheric science forecasting [24–27], traffic flow prediction [28–30], and electric load forecasting [31–36]. In addition, it has also successfully been used for prediction of port throughput [37,38].

The prediction of port throughput is a complex nonlinear dynamic procedure, and is affected by numerous factors, such as Gross Domestic Product, Gross National Product, Total Imports and Exports, Industrial Output, etc. These factors mostly have random and nonlinear characteristics, and may have complex nonlinear connections among them; thus, it is difficult to express them by a definite method in low dimensional space. Noise may emerge when collecting, transmitting and analyzing datum cause of stochastic error, and can exist both in port throughput sequence and in its related factors. The distribution of noise is subject to normal distribution, and has high amplitude value in some point. Theoretically, these models mentioned above do not consider about the drawback of noise. The mixed noise in throughput sequence data and related impact factors datum will largely affect the final prediction results, especially on sensitive SVR model (the matrix is not spare cause of outlier approaching decision boundary, while numbers of support vector grows very fast cause of outlier influence). Wu [39] designed a robust loss function (according to different situations, uses different loss functions), considering mixed noise of normal distribution, high amplitude values, singular point features in datum of prediction sequence and related impact factors, and gained a new support vector regression (namely *R*SVR), and applied it to products sale timing sequence prediction. The numerical calculation results indicate that the estimated model can effectively suppress noise, and lead to better prediction results. To deal well with the mixed noise in port throughput sequence and its related impact factors, this paper adopts the *R*SVR model to improve robustness and accuracy of port throughput prediction.

The practical results indicate that the forecasting accuracy of SVR-based models is influenced by the determination of the parameters significantly [40]. Although there are some recommendations on the appropriate setting of SVR parameters in the literature [41], those approaches do not simultaneously consider the interaction effects among the parameters. The common cross-validation method used for selecting SVR parameters has certain cross error [42], especially in complex forecasting problems, it cannot guarantee high forecasting accurate level.

To identify which approach is suitable for specified data patterns, researchers have employed different hybrid evolutionary algorithms [43–45] (such as particle swarm optimization, simulated annealing algorithms, genetic algorithms and immune algorithms) to determine the parameters. Here, all SVR models with parameters determined by different evolutionary algorithms are superior to other competitive forecasting models (ARIMA, ANNs, etc.); however, these evolutionary algorithms still suffer from the shortcoming of being time consuming or inefficient in optimizing the parameters for SVR models.

Therefore, a more alternative evolutionary algorithms needs to be developed to improve the optimizing approach of parameters. PSO is a stochastic optimization algorithm based on the groups theory, first introduced by Kennedy and Eberhart, and currently is widely used in function optimization, neural network training, pattern classification, fuzzy control system and other engineering fields. PSO is influenced by random oscillation effect in late evolution, making it time-consuming when near globally optimal values, which means convergence rate is very low, and is easily trapped in local minima, and becomes a bottleneck of the PSO algorithm for further development. The SA algorithm is a kind of heuristic random search algorithm based on the Monte-Carlo iteration solving method, originally introduced by N. Metropolis. The algorithm has a strong ability of global optimization search, and accepts both good solutions and inferior solutions by probability. Thus when SA falls into the trap of local optimization, theoretically it can also jump out of the trap after sufficient time, and finally obtain the global optimum. Recently, SA is used widely in engineering, such as production management, control engineering, machine learning, neural networks, image processing and other areas. PSO-SA, which combines SA and PSO, is proposed to enhance the algorithm's ability to jump out of the local maxima, and reduce convergence time [46]. Employing the PSO-SA algorithm to select parameters of the SVR model can improve the optimization effect of parameters for SVR [47,48].

However, due to the lack of diversity in the late evolutionary, the PSO-SA algorithm still easily falls into local optimum (immature convergence) because there is a lack of diversity in the late evolutionary. To enhance the diversity of evolution population and improve the ability of global exploration of PSO-SA, this paper decides to employ the chaotic sequence to transform the three hyper-parameters of an SVR model from the solution space to the chaotic space; any variable in this kind of chaotic space can travel ergodically over the whole space of interest to determine the improved solution eventually.

Recently, numerous chaotic sequences adopt the Logistic mapping function, which is distributed at both ends in the interval [0,1]; it could not excellently strengthen the chaotic distribution characteristics [49]. By comparing with the analysis on chaotic distribution characteristics after mapping the hyper-parameters into chaotic space, the author concludes that the Cat mapping function has good ergodic uniformity in the interval [0,1] and does not easily to fall into a minor cycle [50].

Therefore, this study employs the Cat mapping function to implement the chaos disturbance for the evolutionary population of PSO-SA, designs the coupling evolution mechanism of Cat mapping, SA and PSO, and proposes the CSAPSO algorithm to improve the optimizing performance of *R*SVR's parameters.

In the meanwhile, port throughput is influenced by various socio-economic factors (the candidate input variables, such as gross domestic product, total investment in fixed assets, total imports and exports, industrial output, etc.), but the major disadvantage of SVR-based forecasting models is that it cannot select the final input vectors from the candidate input variables. Thus, selecting the final input vectors is crucial in constructing an SVR-based forecasting model.

The MARS is a multivariate, nonlinear, nonparametric regression approach [51]; it not only has excellent variable selection capabilities, but also can analyze the difference between the degrees of significance for different variables effectively. Thus MARS has been widely used in various fields such as sales prediction [52], credit evaluation [53–55], stock price forecasting [56–58], software reliability analysis [59–61] and predicting species distribution [62,63]. However, MARS has been rarely used in port throughput forecasting in the existing literature; therefore, this paper employs MARS to determine the final input vectors for *R*SVR and analyze significance degrees between different factors for further port throughput generation mechanism research.