Expert Systems with Applications 41 (2014) 8049–8061

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Time series forecasting by neural networks: A knee point-based multiobjective evolutionary algorithm approach

Wei Du [a], Sunney Yung Sun Leung [a,*], Chun Kit Kwong [b]

[a] Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hong Kong, China
[b] Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Article history:
Available online 15 July 2014

Keywords:
Artificial neural network (ANN)
Multiobjective evolutionary algorithm (MOEA)
Time series forecasting (TSF)
Knee point

ABSTRACT

In this paper, we investigate the problem of time series forecasting using single hidden layer feedforward neural networks (SLFNs), which is optimized via multiobjective evolutionary algorithms. By utilizing the adaptive differential evolution (JADE) and the knee point strategy, a nondominated sorting adaptive differential evolution (NSJADE) and its improved version knee point-based NSJADE (KP-NSJADE) are developed for optimizing SLFNs. JADE aiming at refining the search area is introduced in nondominated sorting genetic algorithm II (NSGA-II). The presented NSJADE shows superiority on multimodal problems when compared with NSGA-II. Then NSJADE is applied to train SLFNs for time series forecasting. It is revealed that individuals with better forecasting performance in the whole population gather around the knee point. Therefore, KP-NSJADE is proposed to explore the neighborhood of the knee point in the objective space. And the simulation results of eight popular time series databases illustrate the effectiveness of our proposed algorithm in comparison with several popular algorithms.

1. Introduction

Artificial neural network (ANN) is a mathematical model consisting of a group of artificial neurons connecting with each other, the original idea of which comes from biological neural networks. An ANN is like a human's brain, capable of computing and storing specific information. Because of its satisfactory ability of detecting and extracting nonlinear relationships of the given information, ANNs have been widely utilized in pattern recognition, image processing, data mining, time series forecasting and so on (Bhaduri, Stefanski, & Srivastava, 2011; Leung, Tang, & Wong, 2012; Tang, Gao, & Kurths, 2014;Tang & Wong, 2013; Wong, Leung, & Guo, 2012; Wong, Seng, & Ang, 2011; Xu, Cao, & Qiao, 2011; Zhang, Tang, Miao, & Du, 2013). Among all these applications, time series forecasting (TSF) is quite an intriguing one. Traditional TSF has been performed via statistical-based methods, such as exponential smoothing models (Gardner, 2006; Taylor, 2006) and autoregressive integrated moving average (ARIMA) models (Contreras, Espínola, Nogales, & Conejo, 2003), which are categorized as linear models. While in the past few decades, ANNs play a dominated role in TSF problems owing to their superior performance on regression and classification problems. The main difference that distinguishes ANNs from traditional methods is their ability of generating nonlinear relations hidden in time series data. Recently, a large number of empirical studies have indicated that ANNs have better performance than traditional methods on TSF problems (Hill, O'Connor, & Remus, 1996; Wong & Guo, 2010; Zhang, Cao, & Schniederjans, 2004).

A variety of ANNs have been used in forecasting time series data. For instance, the single multiplicative neuron (SMN) model, a novel neural network introduced recently, was used for TSF (Yeh, 2013); the generalized regression neural network (GRNN) was proven effective in the prediction of time series data (Yan, 2012); radial basis function neural network (RBFNN) also performed well on TSF problems (Du & Zhang, 2008). Generally speaking, during the course of training, training error is the first objective to minimize through different learning algorithms (Yeh, 2013). While actually, there are other objectives that need to be optimized besides training error, like the number of hidden layer nodes and L_2-norm of the hidden layer weights (Goh, Teoh, & Tan, 2008). However, these objectives naturally conflict with each other, which means that the improvement of one objective may lead to deterioration of another (Zhou et al., 2011).

In the past two decades, multiobjective evolutionary algorithms (MOEAs) have attracted much attention in the field of evolutionary computation, which is largely due to their capability of dealing with a multiobjective optimization problem as well as finding nondominated sets in a single run (Deb, 2002; Tang, Wang, Gao, Swift, & Kurths, 2012). Among them, nondominated sorting genetic

* Corresponding author.
E-mail address: sunney.leung@polyu.edu.hk (S.Y.S. Leung).

http://dx.doi.org/10.1016/j.eswa.2014.06.041
0957-4174/

algorithm-II (NSGA-II) (Deb, Pratap, Agarwal, & Meyarivan, 2002) is one of the most popular one. Although NSGA-II finds solutions with satisfactory convergence and diversity in most of the test problems, it gets into trouble when facing multimodal problems (Deb et al., 2002). We impute it to the genetic algorithm in NSGA-II, since the genetic algorithm does not utilize the useful information to explore the preferred region effectively. Recently, a new evolutionary algorithm called adaptive differential evolution (JADE) was proposed, which has been proven efficient and versatile in exploration and exploitation (Das & Suganthan, 2011). Therefore, considering that there are many local optimal solutions in TSF problems solved by ANNs (Zhang, Patuwo, & Hu, 1998), replacing the non-adaptive genetic algorithm part by JADE comes to the first incentive of this paper.

The second incentive of this paper comes from the analysis of the prediction results of several time series databases. According to the results published so far, there are only a few results of applying MOEAs to tackle TSF problems (Chiam, Tan, & Mamun, 2007; Katagiri, Nishizaki, Hayashida, & Kadoma, 2012). And normally, their works have not deeply explored the distribution of the ANN which has the best forecast results in the whole population of ANNs. While in our paper, after employing the proposed nondominated sorting adaptive differential evolution (NSJADE) in several TSF problems, we find the individuals with the best results of forecasting problems mostly gather around the "corner" of corresponding Pareto front (PF). This phenomenon reminds us of the concept of *Knee Point* proposed in the past few years, which has been proven useful in the optimization of many engineering problems, like signalized intersection design, metal cutting process design, and so on (Branke, Deb, Dierolf, & Osswald, 2004; Das, 1999; Deb & Gupta, 2011). Therefore, to the best of the authors' knowledge, this paper makes the first attempt to employ the concept of knee point in TSF problems, which will be manifested as a promising way to ensure both accuracy and reliability at the same time.

When TSF problems were solved by ANNs, most previous research utilized single objective evolutionary algorithms to optimize ANNs (Gu, Zhu, & Jiang, 2011; Ren & Zhou, 2011; Wong et al., 2012). In their researches, training error was the only objective optimized by evolutionary algorithms; the ANN with the minimum training error was selected as the final network for the TSF problem. However, the features of the training samples do not represent the inherent underlying distribution of the new observations due to the existence of noise. So it is not reasonable to merely minimize the training error of ANNs when executing the forecasting. Fortunately, MOEA serves as a promising candidate to optimize ANNs in the TSF problems. As introduced before, according to the results published so far, there are only a few results of applying MOEAs to tackle TSF problems (Chiam et al., 2007; Katagiri et al., 2012). While in their works, they did not deeply explored the distribution of the ANN which has the best forecast results in the whole population of ANNs. However, this distribution information is quite important and helpful when solving the TSF problems, which will be utilized in our research to make the forecasting more accurate and reliable. Briefly, the contributions of this paper can be summarized as follows: (1) JADE is introduced in MOEAs, which delivers a promising performance in exploring the search space; (2) the knee point mechanism helps investigate the intrinsic properties of TSF problems, where our proposed KP-NSJADE guarantees both accuracy and reliability of the prediction of time series data.

The organization of this paper is as follows. Some preliminaries of multiobjective optimization and a new NSJADE, are given in Section 2. In Section 3, NSJADE is used to optimize SLFNs, and experiments are performed to validate the effectiveness of NSJADE. In Section 4, the concept of knee point is introduced. And based on it, KP-NSJADE, an improved version of NSJADE is proposed. In Section 5, eight benchmark time series data sets are predicted by KP-NSJADE-trained SLFNs; in addition, several experiments show the superiority of KP-NSJADE. Finally, conclusions are given in Section 6.

2. Background and methods

In this section, we first give some preliminaries of multiobjective optimization, as well as our proposed NSJADE. In addition, for easy reading, some abbreviations used in this paper are introduced below.

ANN: artificial neural network
SLFN: single hidden layer feedforward neural network
MOEA: multiobjective evolutionary algorithm
DE: differential evolution
JADE: adaptive differential evolution
CoDE: composite differential evolution
CLPSO: comprehensive learning particle swarm optimizer
SPEA2: strength Pareto evolutionary algorithm 2
NSGA-II: nondominated sorting genetic algorithm II
NSJADE: nondominated sorting adaptive differential evolution
KP-NSJADE: knee point based nondominated sorting adaptive differential evolution
TSF: time series forecasting
PF: Pareto front
ELM: extreme learning machine

2.1. Multiobjective optimization

Nowadays, many real-world optimization problems involve various objectives that often conflict with each other. A multiobjective optimization problem can be formulated as follows (without any loss of generality, a minimization problem is considered with a decision space Ω):

$$\begin{array}{ll} \text{minimize} & F(x) = (f_1(x), \ldots, f_n(x)) \\ \text{s.t.} & x \in \Omega, \end{array} \tag{1}$$

where Ω is a decision space and $x \in \Omega$ is a decision vector. $F(x) = (f_1(x), \ldots, f_n(x))$ is the objective vector with n objectives to be minimized.

The objectives in (1) are conflicting pairs, which means that there is not a single solution optimizing all the objectives simultaneously. So it is necessary to seek a group of solutions that can balance all the objectives. Here we introduce the definitions of Pareto dominance, Pareto optimal solution and Pareto front.

Definition 1. (*Pareto Dominance*): Given two objective vectors $X_1, X_2 \in \mathbb{R}^n$, then X_1 dominates X_2, denoted as $X_1 \prec X_2$, iff $x_{1i} \leqslant x_{2i}, \forall i \in \{1, 2, \ldots, n\}$ and $x_{1j} < x_{2j}, \exists j \in \{1, 2, \ldots, n\}$.

Definition 2. (*Pareto Optimal Solution*): A feasible solution $x^* \in \Omega$ of (1) is called a Pareto optimal solution, iff $\nexists x \in \Omega$ such that $F(x) < F(x^*)$.

Definition 3. (*Pareto Front*): The image of all the Pareto optimal solutions in the objective space is called the Pareto front (PF).

Fig. 1 illustrates the dominance relationships between different solutions, where the solutions represented by closed blue circles are dominated by the solutions denoted by closed red squares.