Applied Soft Computing 35 (2015) 52–65

Contents lists available at ScienceDirect

Applied Soft Computing

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

A novel hybrid learning algorithm for full Bayesian approach of artificial neural networks

Ozan Kocadağlı*

Department of Statistics, Mimar Sinan Fine Arts University, Silahsör Cad., No: 71, Bomonti/Sisli, Istanbul, Turkey

ARTICLE INFO

Article history:
Received 16 November 2013
Received in revised form 24 April 2015
Accepted 2 June 2015
Available online 25 June 2015

Keywords:
Bayesian neural networks
Bayesian learning
Hierarchical Bayesian models
Genetic algorithms
Markov chain Monte Carlo
Hybrid Monte Carlo

ABSTRACT

The Bayesian neural networks are useful tools to estimate the functional structure in the nonlinear systems. However, they suffer from some complicated problems such as controlling the model complexity, the training time, the efficient parameter estimation, the random walk, and the stuck in the local optima in the high-dimensional parameter cases. In this paper, to alleviate these mentioned problems, a novel hybrid Bayesian learning procedure is proposed. This approach is based on the full Bayesian learning, and integrates Markov chain Monte Carlo procedures with genetic algorithms and the fuzzy membership functions. In the application sections, to examine the performance of proposed approach, nonlinear time series and regression analysis are handled separately, and it is compared with the traditional training techniques in terms of their estimation and prediction abilities.

1. Introduction

Within Bayesian perspective, it is assumed that uncertainty of any quantity of interest can be expressed and measured by probabilistic distributions. This framework provides a natural way to estimate the functional structure in the nonlinear systems. For this reason, the Bayesian learning is mostly treated in the regression, the time series, the classification and the density estimation applications of the artificial neural networks (ANNs). Bayesian learning in the ANNs are typically based on Gaussian approximation, ensemble learning and Markov chain Monte Carlo (MCMC) simulations known as full Bayesian approach. For ANNs, Gaussian approximation was introduced by Buntine and Weigend [1] and MacKay [2] in which one well-established procedure for approximating the integrals over parameter space, known as Laplace's method. This approach is to model the posterior distribution by a Gaussian distribution, centered locally at a mode of posterior distribution of parameters [2]. The ensemble learning was introduced by Hinton and Camp [3] in which the approximating distribution is fitted globally by minimizing a Kullback–Leibler divergence rather than locally. In the context of Full Bayes approach, Neal [4] introduced advanced Bayesian simulation methods in which MCMC simulations are used to generate samples from the posterior distribution. However, MCMC techniques can be computationally expensive, and also suffer from assessing the convergence. For this reason, Neal [4] integrated Bayesian learning with Hybrid Monte Carlo (HMC) method introduced by Duane et al. [5] to overcome the mentioned shortcomings. Afterwards, the Bayesian applications to ANNs was reviewed thoroughly in [6–8].

In the literature, there are the remarkable studies that deal with the specific problems related to ANNs from Bayesian perspective. For instances, Insua and Müller [9], Marrs [10], Holmes and Mallick [11] worked on the issue of selecting the number of hidden neurons with the growing and the pruning algorithms for the dimensionality problem in the ANNs. In these studies, they applied the reversible jump MCMC algorithm introduced by Green [12], Richardson and Green [13]. Freitas [14] incorporated the particle filters and the sequential Monte Carlo (MC) methods in the BNNs. Liang and Wong [15] proposed to the evolutionary MC algorithm which samples the parameters in the ANNs from the Boltzmann distribution using the mutation, the crossover and the exchange operations defined in the Genetic Algorithms (GAs). Chua and Goh [16] used the Gaussian approach for the multivariate modeling in the ANNs. They used the stochastic gradient descent algorithm integrated with the evolutionary operators to produce the parameters in the ANNs. Liang [17] and Xie et al. [18] used the truncated Poison priors to determine the neuron numbers in the hidden layers, and estimated the parameters in the ANNs via the evolutionary MC algorithms proposed by Liang and Wong [15]. Lampinen and Vehtari [19], Vanhatalo and Vehtari [20] improved

* Tel.: +90 212 246 0011x5511; fax: +90 2122611121.
E-mail address: ozankocadagli@gmail.com

http://dx.doi.org/10.1016/j.asoc.2015.06.003
1568-4946/

a hybrid and reversible MCMC algorithm based on Neal [4]. Marwala [21] adapted to mutation and crossover operators defined in GAs into Bayesian learning, and estimated the parameters using a hybrid learning algorithm. Mirikitani [22] proposed a probabilistic approach to recursive the second-order training of recurrent neural networks for improved time-series modeling in which the regularization hyperparameters leads to better generalization, and stable numerical performance. Goodrich [23] developed a powerful methodology for estimating the full residual uncertainty in the network weights and making predictions by using a modified Jeffery's prior combined with a Metropolis MCMC method. Martens and Sutskever [24] resolved the long-outstanding problem of how to effectively train recurrent neural networks on complex and difficult sequence modeling problems which may contain the long-term data dependencies. Niu et al. [25] adapted Hybrid MC, proposed by Neal [4] and Duane et al. [5], to Gaussian approximation of BNNs. Beam et al. [26] examined Hybrid Monte Carlo proposed by Neal [4] in the context of full Bayesian approach, and then they compared this procedure with the existing methods using artificial and real data sets. Kocadağlı [27] integrated the hierarchical Bayesian learning with GAs and the fuzzy numbers, and then he proposed Genetic MC algorithm in the context of full Bayesian approach. Kocadağlı and Aşıkgil [28] adapted Genetic MC, proposed by Kocadağlı [27], to Gaussian approximation of BNNs.

2. Motivation and overview

In the model building, main difficulty is generally controlling the model complexity. During training ANNs, measuring the model complexity are mostly overlooked, since it is computationally very expensive. In the Bayesian learning, this issue can be handled in a natural and consistent way. However, Bayesian learning requires evaluating the high-dimensional integrals. Although these high-dimensional integrals can be solved by using the analytical integration, the approximation methods and the numerical integration, the analysts suffer from non-Gaussianity, nonlinearity and non-stationarity involved in the real-world problems that prevent using the analytical integration [29]. In this case, MC methods might provide better estimates than the approximate methods in context of BNNs. In spite of their superior advantage, MC methods suffer from the time consuming in addition to the efficient parameter estimation, the convergence, the random walk and the stuck in local optima in the high dimensional parameter cases. For this reason, the hybrid approaches are mostly preferred to overcome these problems. In these approaches, MC techniques are integrated with the gradient-based optimization algorithms, the simulated annealing method or the evolutionary algorithms. The hybrid methods based on gradient algorithms are capable of searching the regions with high density using the gradient information. However, gradient-based algorithms might not explore freely the high dimensional parameter spaces including many local optima. Besides, adjusting the control parameters or Hessian matrix evaluations in these algorithms causes a computational burden. Besides, this kind of optimization algorithms is not suitable for the optimization problems included non-derivatives functions as well. Some shortcomings of these algorithms are discussed in [7,30–32].

Instead of the gradient based algorithms, integrating GAs with MCMC methods provides much more efficient approaches that overcome the mentioned shortcomings in context of training ANNs because they do not require the gradient information and make the parallel searching through the high dimensional parameter spaces to achieve the global solutions rather than local ones. GAs developed by John Holland in 1970s are heuristic optimization methods based on concepts of natural evolution, and belongs to the larger class of evolutionary algorithms [33]. They consist of the artificial operators such as selection, mutation, crossover and migration involved in the natural evaluation. In the treatments of GAs, different kinds of uncertainties arise inherently. For instance, the possibilistic (or linguistic) and the probabilistic uncertainties appear simultaneously [27]. The possibilistic uncertainty concept is first introduced by Zadeh [34]. According to Zadeh [34], this kind of uncertainty deals with problems in which the source of imprecision is the absence of sharply defined criteria of class membership rather than the presence of random variables. In this case, the fuzzy membership functions provide a natural way to intuitively make the plausible semantic description of imprecise properties of data used in the natural system [35].

In this paper, a novel hybrid learning algorithm is introduced for the hierarchal full Bayesian approach. This algorithm called as Genetic MC integrates GAs with Gibbs sampling, Metropolis–Hastings and Simulated Annealing methods. By means of this hybrid learning framework, Genetic MC is capable of overcoming some complicated problems such as efficient parameter estimation, model complexity, random walk, convergence and stuck in the local optima in the high-dimensional parameter spaces. In spite of incorporating many complicated concepts in its structure, Genetic MC can be implemented to the nonlinear regression and time series analyses easily, and then it allows estimating the robust models in context of training BNNs. Essentially, this novel approach is based on the hybrid learning procedures proposed by Kocadağlı [27], Kocadağlı and Asikgil [28] in which GAs is integrated with MCMC methods for Gaussian approach of BNNs. However, in this study, Genetic MC is adapted to the full Bayesian approach instead of Gaussian approach. Besides, although this procedure has some similarities with the approaches developed by Liang and Wong [15] and Marwala [21], it consists of different procedure in terms of using evolutionary operators and MCMC methods. Specifically, Marwala's approach [21] creates only one parameter vector (child) in the any generation of GAs, and then this candidate parameter vector is compared with that held in the previous iteration of Metropolis–Hastings algorithm in terms of their performance over the fitting functions. However, this approach conflicts with diversity principle of GAs because the diversity in the population should be accomplished by keeping the members having different features in the selection process [27]. In Liang and Wong [15], GAs is integrated with Simulated Annealing method in context of MCMC procedure, and the system parameters are produced by means of different temperature values. Basically, this approach is different from Genetic MC approach in context of the estimation procedure and designing of the hierarchal full Bayesian learning.

In order to introduce Genetic MC thoroughly, this paper is organized as following. Section 3 includes an overview of a simple feed-forward ANN, the different types of error functions and the model complexity concepts. In Section 4, the framework of full Bayesian approach is given in context of BNNs. In Section 5, a novel Bayesian learning algorithm for full Bayesian approach is introduced in detail. In Section 6, to examine the implementations of the proposed algorithm, the nonlinear time series and regression analysis are handled, separately and then it is compared with the traditional training procedures for ANNs and the Gaussian approximation of BNNs in terms of their model estimation and prediction performances over a benchmark and an artificial data set. Sections 7 and 8 cover the results and conclusions of study, respectively.

3. Neural network structures

In this paper, as seen in Fig. 1, the feed-forward ANNs with one hidden layer are used due to their simplicity and sufficient performance over the related problems in the application section. According to network structure in Fig. 1, the mathematical