Neurocomputing 173 (2016) 1276–1287

Contents lists available at ScienceDirect

Neurocomputing

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

Set-valued functional neural mapping and inverse system approximation

Jiann-Ming Wu*, Chun-Chang Wu, Jia-Ci Chen, Yi-Ling Lin

Department of Applied Mathematics, National Dong Hwa University, Shoufeng, Hualien 974, Taiwan

ARTICLE INFO

Article history:
Received 4 January 2015
Received in revised form
9 May 2015
Accepted 2 September 2015
Communicated by M. Bianchini
Available online 12 September 2015

Keywords:
Supervised learning
State-regulated neural networks
Set-valued mapping
Inverse neural systems
Mean field annealing
Levenberg–Marquardt learning

ABSTRACT

This work explores set-valued functional neural mapping and inverse system approximation by learning state-regulated multilayer neural networks. Multilayer neural organization is extended to recruit a discrete regulating state in addition to predictive attributes in the input layer. The network mapping regulated over a set of finite discrete states translates a predictor to many targets. Stimuli and responses clamped at visible units are assumed as mixtures of paired predictors and targets sampled from many joined elementary mappings. Unknown regulating states are related to missing exclusive memberships of paired training data to distinct sources. Learning a state-regulated neural network for set-valued mapping approximation involves retrieving unknown exclusive memberships and refining network interconnections. The learning process is realized by a hybrid of mean field annealing and Levenberg–Marquardt methods that simultaneously track expectations of unknown regulating states and optimal interconnections among consecutive layers along a physical-like annealing process. Numerical simulations show the presented learning process well reconstructing many joined elementary functions for set-valued functional mapping and inverse system approximation.

1. Introduction

A multilayer neural network [11,6,20,24], typically consisting of layer-structured nonlinear processing elements, performs parallel and distributed computations for algebraic elementary single-valued mapping. The layer-structured neural organization synchronously translates high-dimensional predictors in the input layer through hidden layers to desired targets in the output layer. Since nonlinear processing elements in the hidden layer ideally carry out radial basis functions [11,6] or projective basis functions [20,24] and adaptable interconnections among consecutive layers are multiplicative, the network mapping can be mathematically expressed as an adaptive algebraic function, $F(x|\theta)$, where θ collects adaptive interconnections, including receptive fields and posterior weights. The expression $y=F(x|\theta)$ therefore characterizes an adaptive single-valued mapping that delicately translates a high-dimensional predictor x to one and only one target y in the function range.

Supervised learning of a multilayer neural network subject to paired training data mainly addresses the minimization of the mean square error of approximating targets by network responses to predictors with respect to θ. Generalization by supervised learning is typically verified by paired testing data during testing phase. Data driven supervised learning stands for function approximation as paired predictors and targets for training and testing are oriented from an elementary single-valued mapping. Supervised learning based on powerful computational methodologies, including mean field annealing [17,19,26] of constrained optimization, and gradient-based iterative approaches, such as the backpropagation method, the nonlinear conjugate gradient method, the Newton–Gauss method, and the Levenberg–Marquardt method [5,15,4] of unconstrained optimization, has been proposed in the field of neural networks, and extensively applied to signal processes, pattern recognition, control and system identification in the past decade.

However learning an adaptive single-valued mapping is not feasible for discrete set-valued mapping approximation in case that paired predictors and targets are mixtures of samples oriented from many joined elementary mappings. Similar predictors are probably mapped to very different desired targets according to constraints proposed by paired training data following the mixture assumption. A discrete set-valued mapping validly translates an identical predictor to many distinct targets. The problem of set-valued mapping approximation has been revealed in applications to inverse control [10,18], complex economic data prediction [12,13] and system inverting, where the mapping underlying paired training data is no more single-valued.

* Corresponding author. Tel.: +886 3 8633531; fax: +886 3 8633510.
E-mail addresses: jmwu@livemail.tw, jmwu@mail.ndhu.edu.tw (J.-M. Wu).

http://dx.doi.org/10.1016/j.neucom.2015.09.002
0925-2312/

For discrete set-valued mapping approximation, given training data are considered as mixtures of paired data sampled from many joined elementary mappings. Under the mixture assumption [8,16,22], the constraints proposed by paired training data allow translation of an identical predictor to many targets. A multilayer neural network only equipped with input units of receiving predictive attributes performs an adaptive elementary single-valued mapping. Subject to mixed training data, refining interconnections is ineffective for reducing the mean square approximating error toward resolving set-valued mapping approximation. The difficulty cannot be resolved by improving learning methodologies or generalizing transfer functions of hidden units in architecture.

For set-valued mapping approximation, the previous work [23] has related distinct targets of a predictor to stable outputs of a recurrent multilayer neural network,

$$y_n = F(x, y_{n\ 1} | \theta) \tag{1}$$

where the network output $y_{n\ 1}$ is transmitted through delayed circular connections to the input layer. For fixed x and y_0, the recurrent relation (1) searches for stable outputs by different initializations [23] of retrieving multiple targets in response to fixed x. However difficulties, including effective optimization of interconnections, reliable enumeration of stable outputs and accurate extraction of many elementary mappings embedded within the recurrent multilayer neural network, still challenge this line of researches. Alternatively, set-valued mapping approximation has been approached by organizing regularization networks [21] and mixture density networks [2,3]. A regularization network [21] consists of two cascaded multilayer neural networks, the first translating a predictor to coefficients of a polynomial with zeros well storing desired targets and the second mapping retrieved coefficients to desired targets. Learning a regularization network requires all targets corresponding to every predictor for determining polynomial coefficients. But the requirement is not satisfied following the mixture assumption. The mixture density network [2,3], organized for conditional probability density function (pdf) reconstruction, mainly translates a predictor to semi-parameters of Gaussian mixtures, including variances, mean vectors and weights. The domain or support of the reconstructed conditional probability density function is expected to contain continuous targets in response to a given predictor. Learning mixture density networks is translated to a task of estimating Gaussian mixtures for conditional density function approximation.

This work approaches discrete set-valued mapping approximation by learning a state-regulated multilayer neural network, which recruits a finite discrete regulating state in the input layer. The proposed state-regulated neural network inherits feedforward synchronous transmission through multilayer neural networks. Conditional to a fixed regulating state, a state-regulated neural network realizes an elementary mapping, insisting on translating a predictor to one and only one target in the range. The network mapping regulated over all finite discrete states essentially translates a predictor to many targets. It is notable that a state-regulated multilayer neural network maintains only one copy of adaptive interconnections among consecutive layers, instead of many multilayer neural networks [7], for discrete set-valued mapping approximation.

This work proposes a hybrid of mean field annealing and Levenberg–Marquardt methods for learning a state-regulated multilayer neural network. Under the mixture assumption, each paired predictor and target has its own exclusive membership, which is encoded by an unknown regulating state and represented by a Potts variable [19,25], to joined elementary functions. Supervised learning of a state-regulated multilayer neural network thus involves retrieving missed regulating states and optimizing adaptive interconnections for set-valued mapping approximation. Since the mean square approximating error is not differentiable with respect to discrete Potts variables, it is minimized by a hybrid of mean field annealing and Levenberg–Marquardt (LM) methods. Under a physical-like annealing process, the proposed hybrid approach iteratively tracks the mean configuration of multi-state Potts variables and applies the LM method to minimize the mean square error of approximating desired targets by network responses to predictors and expectations of regulating states. The mean configuration of Potts variables is eventually forced to discrete regulating states for representing exclusive memberships at the end of the annealing process.

This paper is organized as follows. Section 2 presents learning a state-regulated multilayer neural network for discrete set-valued mapping approximation. The learning task is translated to a mixed integer programming and resolved by a hybrid of mean field annealing and LM methods. Section 3 explores quantitative performance of the proposed learning approach for discrete set-valued mapping approximation by numerical simulations. Section 4 further extends the proposed learning approach for inverse system approximation. Conclusions are given in the final section.

2. Discrete set-valued mapping approximation

2.1. A mixed integer programming

Supervised learning of a state-regulated multilayer neural network is formulated as a mixed integer programming. Given training data are assumed as mixtures of paired predictors and targets oriented from many elementary functions. A discrete set-valued mapping, $\xi = \{f_i\}_i$ in Fig. 1, contains many elementary functions, where f_i denotes an algebraic single-valued mapping, always translating a predictor to one and only one target. Let $S_i = \{(x[t], y[t])\}_t$ collect paired predictors and targets sampled from the ith elementary function, where $y[t] = f_i(x[t]) + n[t]$ and $n[t]$ denotes a noise. Then $S = \bigcup_i S_i$ denotes mixtures of paired predictors and targets oriented from many elementary functions.

A multilayer neural network only equipped with input units of receiving attributes of predictors is unable to faithfully approximate the set-valued mapping ξ underlying S, since the network mapping translates a predictor to one and only one target. Discrete set-valued mapping approximation is essential for inverse system reconstruction. Figs. 4a and 5a show paired training data obtained by inverting the forward MIMO (multiple inputs and multiple outputs) system in Section 4. The forward MIMO system can be approximated by learning multilayer neural networks. But coordinate mappings of the inverse system are set-valued and cannot be faithfully approximated by single-valued mappings.

A high-dimensional predictor x is concatenated with a regulating state δ in the input layer of a state-regulated multilayer neural network, where $\delta = [\delta_1, \dots, \delta_K]$ is a Potts variable with $\delta_i \in \{0, 1\}$ and

$$\sum_{k=1}^{K} \delta_i = 1.$$

The mapping of a state-regulated multilayer neural network translates concatenated x and δ to a target,

$$\hat{y} = F(x, \delta | \theta). \tag{2}$$

The circular connection from the network output in Eq. (1) has been replaced with the discrete regulating state.

Let $\delta \in \Xi_K = \{e_1, \dots, e_K\}$, where e_k denotes a unitary vector with the kth bit one and others zero. Regulated by $\delta = e_k$, the network response to x, is

$$\hat{y}_k \equiv F_k(x | \theta) = F(x, \delta = e_k | \theta), \tag{3}$$

where F_k denotes an elementary mapping, always translating x to one target. Regulated over all finite discrete states in Ξ_K, F translates x to many targets, denoted by $\{\hat{y}_k\}_k$, for set-valued mapping