Signal Processing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Signal Processing

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

Deep transfer learning for classification of time-delayed Gaussian networks

Iti Chaturvedi [a], Yew-Soon Ong [a,*], Rajesh Vellore Arumugam [b]

[a] School of Computer Engineering, Nanyang Technological University, Singapore
[b] Data Storage Institute, Singapore

ARTICLE INFO

Article history:
Received 18 May 2014
Received in revised form
6 September 2014
Accepted 8 September 2014

Keywords:
Transfer learning
Manifold
Time-delays
Variable-order
Gaussian networks
Deep neural networks

ABSTRACT

In this paper, we propose deep transfer learning for classification of Gaussian networks with time-delayed regulations. To ensure robust signaling, most real world problems from related domains have inherent alternate pathways that can be learned incrementally from a stable form of the baseline. In this paper, we leverage on this characteristic to address the challenges of complexity and scalability. The key idea is to learn high dimensional network motifs from low dimensional forms through a process of transfer learning. In contrast to previous work, we facilitate positive transfer by introducing a triangular inequality constraint, which provides a measure for the feasibility of mapping between different motif manifolds. Network motifs from different classes of Gaussian networks are used collectively to pre-train a deep neural network governed by a Lyapunov stability condition. The proposed framework is validated on time series data sampled from synthetic Gaussian networks and applied to a real world dataset for the classification of basketball games based on skill level. We observe an improvement in the range of [15–25]% in accuracy and a saving in the range of [25–600]% in computational cost on synthetic as well as realistic networks with time-delays when compared to existing state-of-the-art approaches. In addition, new insights into meaningful offensive formations in the Basketball games can be derived from the deep network.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Time-delayed Gaussian networks (GNs) can be derived by estimating the output measurements of each variable using a multivariate Gaussian function over the available measurements of input parent variables, such that for all admissible time-delays, the error of estimation is minimal [1,2]. The edges in such a network represent causal interactions in dynamic systems and provide a basis for signal transduction in pathways. Signal transduction is transient; hence, the study on dynamics of the transduction is essential.

The classical time series models used ordinary differential equations to capture complex regulatory dynamics [3–5]. However, they do not work well on multivariate data with variable-order delays. In [6], the authors introduced stochastic models; these assume an underlying hidden state of a dynamic system evolving over time. The earliest stochastic networks were Boolean, which were built using mutual information among discrete nodes [7,8]. In order to predict the causality in networks with Gaussian or mixed nodes, Bayesian networks have been proposed [9–12].

State-of-the-art Bayesian network is a directed acyclic graph where variables are present at the nodes and edges represent causal interactions among them. For each node and parent set in the graph, conditional probabilities are computed from the time series data. The variable-order Bayesian network is attained by learning transition

* Corresponding author.
E-mail addresses: iti@ntu.edu.sg (I. Chaturvedi), asysong@ntu.edu.sg (Y.-S. Ong), Rajesh_VA@dsi.a-star.edu.sg (R.V. Arumugam).

http://dx.doi.org/10.1016/j.sigpro.2014.09.009
0165-1684/© 2014 Elsevier B.V. All rights reserved.

Please cite this article as: I. Chaturvedi, et al., Deep transfer learning for classification of time-delayed Gaussian networks, Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.09.009

ARTICLE IN PRESS

Nomenclature

n_d	number of domains or classes
$x_i(\tau)$	expression level of node i at time instant τ
$y(\tau)$	class label for sample τ
β	regression co-efficient matrix
$\boldsymbol{a}_i$	parent set of a node i
$\boldsymbol{\theta}_{i,\boldsymbol{a}_i}$	parameters for node i in the Bayesian network given parent set $\boldsymbol{a}_i$
N	number of variables in the system
Σ	covariance matrix of Gaussian node
μ	mean vector of Gaussian nodes
v_i	node i in the visible layer
h_j	neuron j in the hidden layer
T	number of data samples
r	the upper-bound of delay
X^s	time series data for class s
l	index for a hidden layer
f_l	activation function of the hidden layer l
E	global energy function for a DBN
W_l	weights of the hidden layer l
α	learning rate of a DBN
λ	transformation factor for a motif
$\Delta\epsilon$	change in classification precision error
S	Gaussian network

networks between three or more structures. Now, the expression of a node depends on the expression of parents from $(r>1)$ previous time points [13]. In [14], general tensor discriminant analysis was used to collectively model such higher-order networks during classification.

Accurately predicting joint multivariate probabilities with Gaussian nodes requires a lot of computational effort. For very large networks, it is customary to limit the connectivity of each node using some geometric prior [15]. This is unrealistic in practice since most real world networks require hub nodes with very high connectivity, so as to ensure robust signaling [4]. Lastly, when modeling networks with variable delays, robustness of the prediction has been established to be heavily reliant on the quality of time samples and prior knowledge. However, collecting data from real world networks is often plagued with practical difficulties and high experimental costs.

Further, the distributed nature of most real world networks manifests itself as intense crosstalk between pathways. In particular, states of Gaussian networks are often presumed to be stable, meaning that slight changes in the state of a few parents does not change the expression state of the child node. This relates to the natural phenomenon of real world complex networks, where a system retains functionality in spite of turbulences by maintaining redundancy. Hence, most real world problems have inherent alternate pathways that can be learned incrementally from a stable form of the baseline. In this paper, we leverage on this characteristic and look at transfer learning to address the challenges of complexity and scalability.

1.1. Related work

Domain adaptation aims to generalize one or more source class network(s) that are easier to learn, to augment the target network, for which data is scarce. Since data in networks of different classes are distributed differently, previous authors have collectively used data from the source(s) and the target class networks to learn shared feature representations [16]. In the case of networks, we can consider motifs, which are recurrent, or statistically significant sub-graphs shared across different classes. Deep neural networks are ideal for learning a set of shared motifs from different classes of Gaussian networks [17–19]. However, simply learning different class networks together can be detrimental when the networks are unrelated.

Transfer learning has been previously used in the Bayesian frameworks to estimate the prior covariance matrix of the target network by simply averaging the covariance matrices of previously determined structures in similar classes of networks [20,21]. In their approach, the history of past maximum likelihood (ML) estimates for motifs can be reused by transferring to suitable new structures of higher dimensions. However, the discriminative information in the covariance matrices is often lost due to undersampling of data. Previously, the use of normalized divergences was shown to preserve small differences between classes [22]. Transfer may be transductive which means across different dynamic systems or inductive which means transfer among related structures in the same network. The objective function tries to minimize the loss in prediction accuracy due to the transfer [23]. For example, in [24], labeled and unlabeled samples are combined with a trade-off parameter.

In document classification, instead of assuming that all word probabilities are independent in the target class, transductive transfer learning was used to estimate the dependencies between words using other source classes with known labels. Since, the number of possible covariance estimates would increase exponentially with size of the vocabulary, in [20] the authors learn the covariance for only a subset of words from the source classes and combine it with the rest of the vocabulary under a semi-definite constraint. However, their approach will not be able to capture the underlying structure due to higher-order dependencies among groups of words that can occur at variable positions in a document. It will also not be able to transfer effectively among words that are synonyms and used alternatively across documents. In [25] the authors tried to address this issue through adaptive learning of higher-order dependencies in the neighbourhood of a word. Similarly, to account for lack of data in click-based methods for image ranking in web search, we can predict clicks for new images using click data from associated images. In [26], the authors achieved this through learning of manifolds for each image feature separately using a group of weights and hyper-graphs to model higher-order

Please cite this article as: I. Chaturvedi, et al., Deep transfer learning for classification of time-delayed Gaussian networks, Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.09.009

Download English Version:

https://daneshyari.com/en/article/6959480

Download Persian Version:

https://daneshyari.com/article/6959480

Daneshyari.com