ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 37 (2007) 749–760

Learning effective brain connectivity with dynamic Bayesian networks

Jagath C. Rajapakse[a,b,*] and Juan Zhou[a]

[a]BioInformatics Research Center, Nanyang Technological University, Singapore
[b]Singapore-MIT Alliance, Singapore

Received 14 November 2006; revised 22 May 2007; accepted 5 June 2007
Available online 14 June 2007

We propose to use dynamic Bayesian networks (DBN) to learn the structure of effective brain connectivity from functional MRI data in an exploratory manner. In our previous work, we used Bayesian networks (BN) to learn the functional structure of the brain (Zheng, X., Rajapakse, J.C., 2006. Learning functional structure from fMR images. NeuroImage 31 (4), 1601–1613). However, BN provides a single snapshot of effective connectivity of the entire experiment and therefore is unable to accurately capture the temporal characteristics of connectivity. Dynamic Bayesian networks (DBN) use a Markov chain to model fMRI time-series and thereby determine temporal relationships of interactions among brain regions. Experiments on synthetic fMRI data demonstrate that the performance of DBN is comparable to Granger causality mapping (GCM) in determining the structure of linearly connected networks. Dynamic Bayesian networks render more accurate and informative brain connectivity than earlier methods as connectivity is described in complete statistical sense and temporal characteristics of time-series are explicitly taken into account. The functional structures inferred on two real fMRI datasets are consistent with the previous literature and more accurate than those discovered by BN. Furthermore, we study the effects of hemodynamic noise, scanner noise, inter-scan interval, and the variability of hemodynamic parameters on the derived connectivity.

Keywords: Bayesian networks; Causal effects; Dynamic Bayesian networks; Effective connectivity; Functional MRI; Hemodynamic response function; Markov chains; Markov chain Monte Carlo methods

Introduction

The brain areas involved in various cognitive tasks can now be identified quite accurately and reliably through functional Magnetic Resonance Imaging (fMRI) experiments (Friston et al., 1995; Rajapakse and Piyaratna, 2001; Wang and Rajapakse, 2006). However, functional specialization of the brain does not provide a holistic view of brain function and does not describe how different brain regions communicate and interact with one another. Considering the multiple processes taking place at different brain regions and interacting with one another in executing a specific task, extracting brain connectivity from fMRI data facilitates our understanding of brain function (Buchel and Friston, 1997). Recently, there has been an increasing interest in functional integration studies to infer brain connectivity, especially for high-order brain functions. In fMRI, the activity of brain is measured by time-series of signals depending on blood-oxygenation-level-dependent (BOLD) contrast. Given multivariate voxel-based time-series, several techniques have been proposed to use fMRI to characterize effective connectivity of the brain (Friston, 2003; Goebel et al., 2003; McIntosh and Gonzalez-Lima, 1994; Rajapakse et al., 2006; Zheng and Rajapakse, 2006).

Structural equation modelling (SEM) decomposes interregional covariances of fMRI time-series to find functional interactions among brain regions (Bullmore et al., 2000; McIntosh and Gonzalez-Lima, 1994; Mechelli et al., 2002). The covariance structure models the interactions of underlying neural systems only in second-order statistical sense and therefore does not render effective connectivity or the "cause and effect" relationships among brain regions. Dynamic causal modelling (DCM) characterizes the dynamics of interactions among states (of brain regions) with bilinear approximations of intrinsic coupling (among neuronal states) and the influence of external inputs. An extended balloon model is used in DCM to model hemodynamic response, which enables inference of interactions at the neuronal level (Friston, 2003). Both SEM and DCM are confirmatory in the sense that the analysis of brain connectivity requires a priori model to begin with and is inapplicable for higher-order functions unique to human such as language or cognition (Bullmore et al., 2000).

Granger causality mapping (GCM) extends the vector auto-regressive (VAR) technique to capture interactions among brain regions, assuming a causal and dynamic system of linear interactions, driven by stochastic innovations (Goebel et al., 2003; Harrison et al., 2003). A graphical approach linking the notions

* Corresponding author. BioInformatics Research Center, Nanyang Technological University, Singapore.
E-mail address: asjagath@ntu.edu.sg (J.C. Rajapakse).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2007.06.003

of graphical models and Granger causality has been applied to describe dynamic dependencies in neural systems (Eichler, 2005). Nevertheless, a multi-step procedure fitting autoregressive models at each step is required to identify networks and therefore limits its applicability for large networks.

Recently, two techniques based on Bayesian networks (BN) (Zheng and Rajapakse, 2006) and independent component analysis (ICA) (Rajapakse et al., 2006) were proposed to derive effective connectivity of the brain from functional MRI data in an exploratory manner. Bayesian networks do not provide an explicit mechanism to represent temporal dependencies among multiple processes at brain regions and instead give one snapshot of brain connectivity, taking into consideration the whole experiment. Therefore, neural systems derived with BN do not fully describe causal relationships among brain regions. Moreover, because of equivalent properties of BN, directions of some edges are indeterminate and could be bi-directional (Chickering, 1995).

In this paper, we propose dynamic Bayesian networks (DBN) to derive the effective connectivity of the brain by modelling fMRI time-series in a Markov chain. DBN, an extension of BN, admits a class of nonlinear continuous time interactions and provides a direct mechanism to model temporal relationships among brain regions. Functional MRI time-series of activated voxels are modelled with first-order stationary Markov chains. The inter-scan interval (ISI) of fMRI is used as the interval between two consecutive instances of the Markov chain. The connectivity between two time instances (or scans) is modelled in a transition network of two layers of brain regions (or nodes). In a stationarity setting, the connectivity of the transition network renders the effective connectivity of the brain.

Dynamic Bayesian networks may assume a known or unknown structure, and full or partial observability of states at the nodes. The states of activated brain regions are fully observed as intensity variations of fMRI time-series. Beginning with an unknown connectivity structure, we find the best structure fitting fMRI data in an exploratory manner. A greedy search or an expectation maximization (EM) provide only a local search of the structure of DBN. Starting with a partly connected structure, we use a Markov chain Monte Carlo (MCMC) method to derive the structure of the connectivity among brain regions from fMRI data. The MCMC method attempts to find a globally optimal solution by sampling a collection of highly probable structures from the equilibrium distribution of the Markov chain (Husmeier, 2003b).

We describe DBN and structure learning algorithm in the Method section. In experiments, synthetic fMRI data is used to illustrate the robustness of our approach and compare with GCM. The method is further demonstrated by exploring functional structures from real fMRI data obtained in two experiments: a silent word reading task and a counting Stroop task. A comparison between structures derived from BN and DBN is also provided.

Method

This section introduces DBN for modelling effective brain connectivity from functional MRI data. Then, a MCMC algorithm for structure learning is described.

Neural system modelling with DBN

When modelling the brain connectivity, the nodes in the Bayesian network are associated with activated brain regions while the edges characterize the interactions among regions. Consider a neural system of n brain regions activated by a sensory or cognitive task and let the regions be indexed in a set $I=\{i : i=1, 2, \ldots n\}$. The activation of a brain region is measured by the average fMRI time-series over the region. Let x_i be the activation measuring the hemodynamic response of region i.

Bayesian networks (BN) describe the probability distribution over the activation of brain regions, where the graphical structure provides an easy way to specify conditional interdependencies for a compact parameterization of the distribution. The BN is defined by a structure s and a joint distribution over the set of time-series $x=\{x_i : i\in I\}$. The BN structure is a directed acyclic graph (DAG) characterized by the absence of directed cycles. If a_i denote the set of activations of the parents of the region i, a DAG offers a simple and unique way to decompose the likelihood of activation in terms of conditional probabilities:

$$P(x|\theta) = \prod_{i\in I} P(x_i|a_i, \theta_i) \tag{1}$$

where $\theta=\{\theta_i : i\in I\}$ represents the parameters of the conditional probabilities.

Dynamic Bayesian network extends BN model to incorporate temporal characteristics of the time-series x. Let us explicitly represent temporal processes of brain regions and $x(t)=\{x_i(t): i\in I\}$ representing the activations of n brain regions at time t. The instances $t=1, 2, \ldots T$ correspond to the times when brain scans are taken and T denotes the total number of scans. In order to model the temporal dynamics of brain processes, we need to model a probability distribution over the set of random variables $\bigcup_{t=1}^{T} x(t)$ which is rather complex and practically prohibitive.

To avoid an explosion of the model complexity, we assume the temporal changes of activations of brain regions are stationary and first-order Markovian:

$$P(x(t+1)|x(t), \ldots x(1)) = P(x(t+1)|x(t)) \tag{2}$$

where the transition probabilities $P(x(t+1)|x(t))$ are independent of t. The transition network represents the connectivity structure between two consecutive brain scans, which renders the joint distribution of all possible trajectories of temporal processes. The structure of the DBN is obtained by unrolling the transition network over consecutive scans for all $t=1, 2, \ldots T$. The first-order stationary assumption provides a tractable causal model that explicitly takes into account the temporal dependencies of brain processes. Higher-order and non-stationary Markov models allow more complex temporal processes and connectivity patterns. However, such complex models pose obvious challenges in estimating structures and parameters.

Unlike BN, DBN is capable of modelling recurrent networks while still satisfying the acyclic constraint of the transition network. This is an important advantage of modelling neural system with DBN as there exist cyclic functional networks in the brain, such as cortico-subcortical loops. Inter-scan connections to same brain region itself are considered as default prior connections and their parameters are allowed to adapt. We do not allow intra-scan connections because the effect on a brain region takes place with a time delay after its cause. Although instantaneous interactions may exist due to low temporal sampling and hemodynamic modulation of fMRI, the determination of such interactions remains as a limitation of neural systems modelling with functional MRI.

Structure learning of DBN

With stationary and first-order Markovian assumptions, the transition network of DBN has two layers of n random variables, each representing n brain regions, connected in a structure s. As the inter-scan connections are always forward, conditional distributions are only defined for the nodes of the second layer given the first layer. As we consider the connectivity in the transition network, in what follows, the time variable is dropped. The optimal structure s^* of brain connectivity is obtained by the maximum a posteriori (MAP) estimation:

$$s^* = \arg\max_s P(s|x) \tag{3}$$

From Bayes rule, $P(s|x) \propto P(x|s)P(s)$ where $P(s)$ denotes the prior probability of the structure. The likelihood of the data given a particular structure s, requires marginalization over the parameters θ:

$$P(x|s) = \prod_{i\in I} \int P(x_i|a_i,\theta_i)P(\theta_i)d\theta_i \tag{4}$$

We use multinomial conditional distribution in order to capture non-linear dependence relations among brain regions, which is more appropriate for modelling effective connectivity of human brain than normal distribution. With multinomial distribution assumption, the integral in Eq. (4) is analytically tractable (Husmeier, 2003b) and the Bayesian score can be used to measure the fitness of structure. However, this choice requires discretization of the data.

Suppose there are K discretized levels and $\Lambda = \{k: k=1, 2, \ldots K\}$, each random variable $x_i \in \Lambda$ and the conditional probabilities among nodes are represented as a stochastic matrix of co-occurrences. If the random node x_i has J_i number of parent nodes, the parameters for conditional probabilities of x_i include a set of parameters $\theta_i = \{\theta_{ij} : j=1, 2,\ldots K^{J_i}\}$ where $\theta_{ij} = \{\theta_{ijk} : k \in \Lambda\}$. Dirichlet priors are assumed for this set of probabilities with the corresponding prior model parameters $\alpha_{ij} = \{\alpha_{ijk} : k \in \Lambda\}$:

$$P(\theta_{ij}|\alpha_{ij}) \frac{\Gamma(\sum_{k\in\Lambda}\alpha_{ijk})}{\prod_{k\in\Lambda}\Gamma(\alpha_{ijk})} \prod_{k\in\Lambda} \theta_{ijk}^{\alpha_{ijk}-1} \tag{5}$$

where $\alpha_{ijk} > 0$ and $\Sigma_k\, \theta_{ijk} = 1$. The values of θ_{ijk} are estimated by the number of corresponding co-occurrences N_{ijk} throughout the time-series based on the maximum likelihood criterion. Bayesian Dirichlet metric is used for the assignment of α_{ijk} and the total likelihood can be calculated by (Heckerman et al., 1995):

$$P(x|s) = \prod_{i\in I} \prod_{j=1}^{K^{J_i}} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}+N_{ij})} \prod_{k\in\Lambda} \frac{\Gamma(\alpha_{ijk}+N_{ijk})}{\Gamma(\alpha_{ijk})} \tag{6}$$

The number of network structures increases super-exponentially with the number of regions and hence finding the optimal structure is NP-hard. Heuristic local search algorithms like hill-climbing have been attempted. We resort to Markov Chain Monte Carlo (MCMC) algorithm for structure learning with DBN in a global search sense since direct sampling from the posterior probability is impossible. MCMC simulation converges to an equilibrium distribution, sampling of which provides the solution of the optimal structure. In what follows, we describe the MCMC algorithm adopted for structure learning (Husmeier, 2003b).

Given a network structure s^{old}, a new structure s^{new} is proposed with a probability $Q(s^{\text{new}}|s^{\text{old}})$ and accepted with Metropolis–Hastings (MH) acceptance criterion (Hastings, 1970):

$$\min\left\{1, \frac{P(s^{\text{new}}|D)}{P(s^{\text{old}}D)} \frac{Q(s^{\text{old}}|s^{\text{new}})}{Q(s^{\text{new}}|s^{\text{old}})}\right\} \tag{7}$$

The iteration of the above procedure generates a Markov chain converging in distribution to the true posterior distribution. In practice, a new network structure is proposed by applying one of the elementary operations such as deleting, reversing, or adding an edge, and then discarding those structures that violate the acyclic condition. The Hastings ratio is found by

$$\frac{Q(s^{\text{old}}|s^{\text{new}})}{Q(s^{\text{new}}|s^{\text{old}})} = \frac{N(s^{\text{old}})}{N(s^{\text{new}})} \tag{8}$$

where $N(s)$ is the size of the neighborhood of structure s. That is, the number of acyclic structures that can be obtained from s by application of one of the elementary operations such as deletion, addition, or reversion of an edge.

The parameters of MCMC simulation are the lengths of the burn-in phase, the sampling phase, and the sampling interval. We make no constraint on the maximum fan-in and the flat prior expectation of nodes with non-zero fan-out of nodes, leading to a uniform prior distribution for network structures. Bayesian score is used to measure the fitness of the structure, which integrates over all parameters.

The steps in the structure learning algorithm are as follows:

(1) *Initialization*: initialize the network structure s with all connections set to zero, except that inter-scan self-connections set to 1. Time-series $x_i(t)$ of every region $i \in I$ is discretized into ternary form $[-1, 0, 1]$ with

$$d_i(t) \begin{cases} 1 & \text{if } x_i(t) \geq \bar{x}_i + (x_{i,\max} - \bar{x}_i)/3; \\ -1 & \text{if } x_i(t) \leq \bar{x}_i - (\bar{x}_i - x_{i,\min})/3; \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

where $\bar{x}_i$, $x_{i,\min}$, $x_{i,\max}$ are the mean, minimum, and maximum values of time-series $x_i = (x_i(t): t=1, 2,\ldots T)$, respectively, and $d_i(t)$ represents the corresponding discretized time-series.

(2) *Burn-in phase*: propose a new network structure by applying elementary operations such as addition, deletion, or reversion of edges to the old network. Structures violating the acyclic constraint are discarded. Accept the structure based on Metropolis–Hastings (MH) acceptance criterion. Repeat the same procedure for a sufficient number of times until convergence to the true posterior probability is achieved.

(3) *Sampling phase*: follow the same procedure as in step 2 and collect the samples for every fixed interval.

(4) *Inference*: calculate the final structure and parameters by averaging over all sampled structures in step 3.

Experiments and results

We illustrate our method with experiments on synthetic data as well as on two real fMRI datasets from fMRI Data Center, Dartmouth College (fMRIDC, 2004): silent reading task (access

number 2-2000-11189) and counting Stroop task (access number: 2-2000-1123B). Bayesian network simulations were done in MATLAB, using two toolboxes: Bayes Net Toolbox (BNT) (Murphy, 2001) and Inferring Dynamic Bayesian Networks with MCMC toolbox (Husmeier, 2003a).

Synthetic data

Synthetic fMRI datasets were generated to test the robustness of the proposed method in detecting the underlying neural system and to compare its performance with Granger causality modelling (GCM). The effects of temporal sampling, scanner noise, hemodynamic noise, and HRF variability on deriving brain connectivity were also evaluated.

Data generation

In order to investigate whether and to what extent DBN is capable of detecting neuronal interactions, we constructed synthetic datasets emulating hemodynamic modulation and temporal sampling of BOLD responses in fMRI (Roebroeck et al., 2005). The steps (see Fig. 1(a)) involved in the generation of fMRI data are:

(1) *VAR modelling*: generate time-series x of brain regions with a predefined VAR model and additive Gaussian noise. The time interval between consecutive time instances is equal to inter-scan interval (ISI) of 100 ms and the length of time-series $T=3000$ (i.e., 300 s). A first-order vector autoregressive (VAR) model was used to generate the time series:

$$x(t+1) = Cx(t) + u(t) \tag{10}$$

where C denotes the linear connectivity matrix and $u(t)$ is the uncorrelated Gaussian innovation with zero mean and covariance matrix Σ with diagonal elements equal to 1 and off-diagonal elements equal to 0.

$$C = \begin{bmatrix} -0.9 & 0 \\ 0.4 & -0.9 \end{bmatrix} \tag{11}$$

For an example with two regions $n=2$, the connectivity matrix $C = \{c_{ij}\}_{2\times 2}$ of the neural system in Fig. 1(b) was set

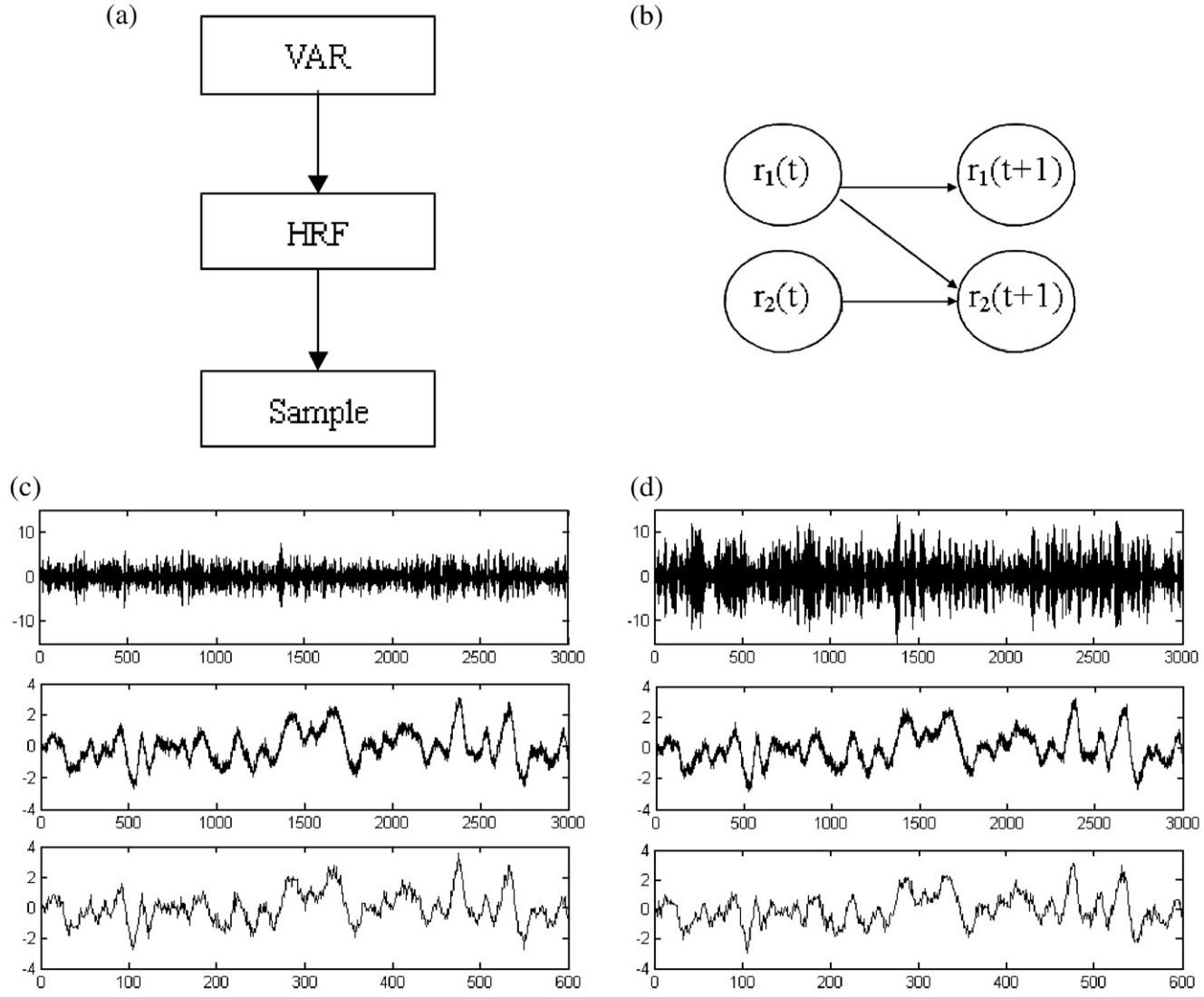

Fig. 1. Generation of synthetic data: (a) three steps involved in generation of time-series; (b) an example structure of two regions; (c) and (d) correspond to time-series generated at regions r_1 and r_2, respectively.

as in Eq. (11). The synthetic time-series generated are shown in Figs. 1(c) and (d). The connectivity matrix C defines directed inter-regional connection from region r_1 to r_2 but not in the reverse direction and the time-series of all the regions are generated in the lower frequency range because of the absence of high-frequency influences due to hemodynamic modulation in fMRI. As BN has been shown to be superior than SEM in obtaining the connectivity structure (Zheng and Rajapakse, 2006), we chose the above model of synthetic data to enable comparison with GCM.

(2) *HRF modulation*: modulate time-series by convolving with the canonical hemodynamic response function (HRF) consisting of a mixture of two gamma functions:

$$f(t) = \frac{1}{\Gamma\left(\frac{\tau_1}{\tau_3}\right)}\left(\frac{\delta}{\tau_3}\right)^{\frac{\tau_1}{\tau_3}} w(t)^{\frac{\tau_1}{\tau_3}-1} e^{-\frac{\delta}{\tau_3}w(t)} - \frac{1}{\tau_5\Gamma\left(\frac{\tau_2}{\tau_4}\right)}\left(\frac{\delta}{\tau_4}\right)^{\frac{\tau_2}{\tau_4}} w(t)^{\frac{\tau_2}{\tau_4}-1} e^{-\frac{\delta}{\tau_4}w(t)} \quad (12)$$

where $\delta = \frac{RT}{16}$, repetition time RT = 100 ms, delay of response $\tau_1 = 6$ s, delay of undershoot $\tau_2 = 16$ s, dispersion of response $\tau_3 = 1$ s, dispersion of undershoot $\tau_4 = 1$ s, ratio of response to undershoot $\tau_5 = 6$, and the length of kernel $\tau_6 = 32$ s. $\Gamma(\cdot)$ represents the gamma function and $w(t) \in \left\{0, \frac{1}{\delta}, \ldots \frac{\tau_6}{\delta}\right\}$. After normalized to zero mean and unit variance, a certain amount σ_1 of white Gaussian noise was added to emulate hemodynamic noise.

(3) *MR sampling*: acquire final time-series by down-sampling the signals at every Δ second, corresponding to the ISI of MR scanner. After normalization of signals individually, amount σ_2 of Gaussian noise was added to represent errors in measurement and acquisition (or scanner noise).

Using the above procedure, we generated neural systems consisting of different number of brain regions. The default parameters were $u(t) \sim N(0, 1)$, hemodynamic noise $\sigma_1 = 0$, scanner noise $\sigma_2 = 0$, and inter-scan interval $\Delta = 0.5$ s.

Parameter settings

For MCMC simulation, two sets of parameters were tested: [3000, 3000, 5] and [50000, 50000, 100] for lengths of burn-in phase, sampling phase, and sampling interval, respectively. The networks derived for synthetic data with two sets were approximately the same and only the results of [3000, 3000, 5] are reported here. Longer burn-in phase or sampling phase may be more likely to find the globally optimal solution at the expense of computation time.

The robustness of our method of deriving DBN for inferring connectivity structure was evaluated with respect to four aspects: (1) the number of brain regions n, (2) the ISI Δ in sampling, (3) the hemodynamic noise σ_1, and (4) the scanner noise σ_2. For different values of above parameters, synthetic fMRI datasets were generated: n from 3 to 8; Δ from 0.5 s to 3.0 s; the length of time-series T from 600 to 100 points covering 300 s and representing fMRI experiments with RT from 0.5 s to 3.0 s; the mean of both signal and noise were set to zero; the standard deviation of signal was set to 1.00; the standard deviations of noises $\sigma_1, \sigma_2 \in \{0.00, 0.05, \ldots 0.50\}$.

An example network of five regions $n = 5$ is shown in Fig. 2. Time-series data of $T = 600$ time points were generated. The

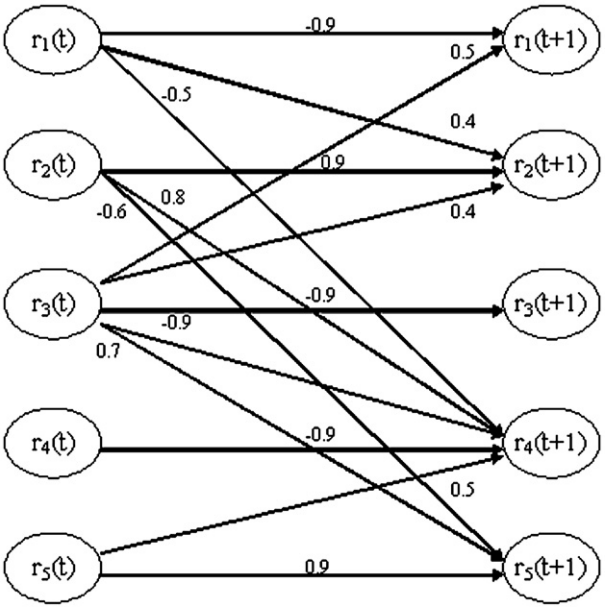

Fig. 2. A transition network of a DBN representing a neural system consisting of five brain regions. The values at the edges represent the strengths of connections.

execution time in deriving the DBN with MCMC parameters [3000, 3000, 5] was approximately 4 min on a Pentium IV 2.4 GHz machine having 512 MB memory and running MATLAB (version 7.0) in Windows XP environment.

Performance evaluation

The square error e^2 between the true connectivity structure $C = \{c_{ij}\}$ and the estimated structure $\hat{C} = \{\hat{c}_{ij}\}$ is measured by the square error between the elements of the matrices defining the structures:

$$e^2 = \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(c_{ij} - \hat{c}_{ij}\right)^2; \quad (13)$$

the elements C and $\hat{C}$ were scaled to the range [0, 1] for comparison with GCM results. For a given structure, five different datasets were generated following the above procedure with random initializations. MCMC simulation used the same set of parameters [3000, 3000, 5]. The mean and standard deviation of square errors e^2 are reported.

Fig. 3 shows the effect of error in deriving the connectivity structure with our approach and GCM. Each curve in the plots corresponds to a structure with a specified number of regions. Small standard deviation of errors (<0.03) in estimating the structure of effective connectivity in all simulations indicate the effectiveness of the algorithm. The errors at various sampling step sizes Δ are shown in Fig. 3(a). As seen, derived DBN structures are capable of representing neural systems closer to the ground truth, with negligible error up to 3 s of sampling steps. As the number of regions of the structure increases, the error increases, indicating a higher probability of the algorithm to trap in a local minimum. Figs. 3(c) and (e) show an increase in error with the amounts of hemodynamic noise σ_1 and scanner noise σ_2. Effects of both hemodynamic and scanner noise are more severe than those contributed by sampling and largely responsible for error irrespective of the number of regions.

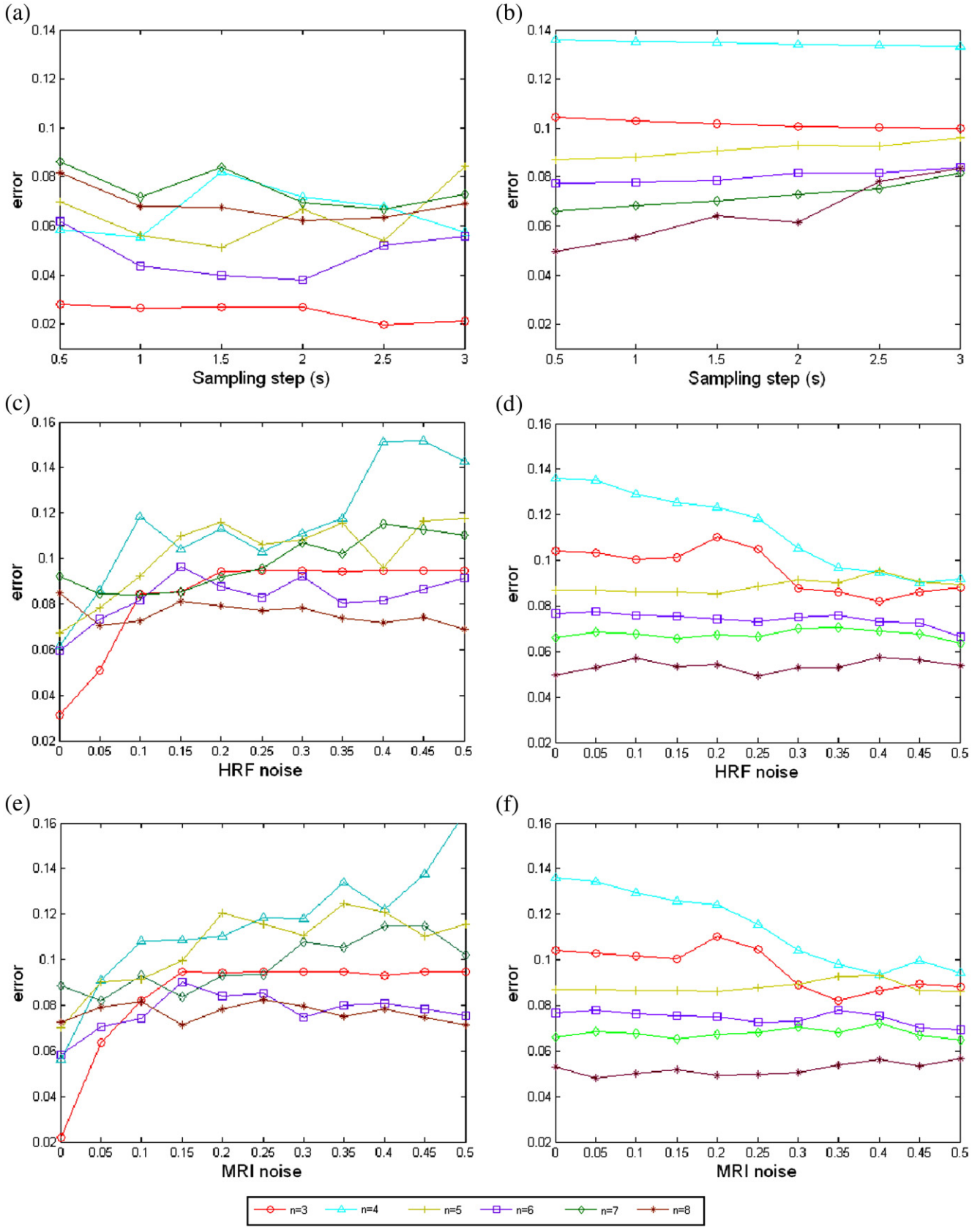

Fig. 3. Comparing the performance of DBN (left column) and GCM (right column) for learning connectivity structure of synthetic fMRI data: the effects of (a) and (b) sampling step size, (c) and (d) the amount of hemodynamic noise, and (e) and (f) the amount of scanner noise, in estimating the connectivity structure.

Comparison with GCM

Granger causality mapping (GCM) (Goebel et al., 2003; Roebroeck et al., 2005) was used to derive neural systems of synthetic datasets, using the algorithm proposed in Bagarinao and Sato (2002). Schwarz–Bayesian Information Criterion is used to estimate the optimal order of VAR model.

Three figures in the right column, (b), (d), and (f), of Fig. 3 demonstrate the effects of sampling step size, hemodynamic noise, and scanner noise on the accuracy of the networks derived by GCM, respectively. The order of 1 was always selected as the optimal order based on Bayesian Information Criteria. Despite the information loss in the discretization of time-series, DBN show comparable performance with GCM. Both methods had similar effects but DBN had less error than GCM at all sampling step sizes. Both techniques suffered from increased scanner and measurement noise as the number of regions is increased from 3 to 8. GCM is less vulnerable to hemodynamic and scanner noise than DBN, especially for network structures consisting of 3 or 4 regions. As the level of noise increases, the loss of information resulting from discretization influences the accuracy of DBN derived networks. Interestingly, sometimes larger number of regions showed lower average square error for both DBN and GCM.

Robustness to HRF variability

The characteristics of HRF vary across brain regions and subjects (Rajapakse et al., 1998). In order to evaluate the robustness of DBN to the variation of HRF, different HRF were generated by randomly selecting the values of six parameters within their practical ranges: delay of response $\tau_1 \in [4.0, 8.0]$s, delay of undershoot $\tau_2 \in [12.0, 20.0]$s, dispersion of response $\tau_3 \in [0.5, 1.5]$s, dispersion of undershoot $\tau_4 \in [0.5, 1.5]$s, ratio of response to undershoot $\tau_5 \in [4.0, 8.0]$, and the length of the kernel $\tau_6 \in [28.0, 36.0]$s. The neural activity of every region of different subjects was convolved with a randomly selected HRF.

For the same structure *s*, at a given sampling step size, hemodynamic noise level, and scanner noise level, five different datasets were generated with random initialization, hemodynamic modulation, and temporal sampling. The synthetic datasets were generated and the effects of sampling, hemodynamic noise, and scanner noise were tested, in the manner as described in Parameter settings section. In order to evaluate the capability of our approach handling the variability of HRF among different brain regions and subjects, we compared the averages of square error in learning structures on data generated with the canonical HRF and with varying HRF parameters. Paired *T*-test was used to examine whether there is a significant increase in the error for structures learned. Table 1 gives *p*-values of comparisons. As seen, the variability of HRF significantly affects the learning of structures with varying sampling rate but less with hemodynamic or scanner noise.

Table 1
Comparison of error in structure learning on synthetic data, considering HRF variability

Regions	Significance of error difference (*p*-value)		
n	Sampling interval	Hemodynamic noise	Scanner noise
3	$p<0.01$	$p<0.01$	–
4	$p=0.03$	$p=0.02$	$p=0.03$
5	$p<0.01$	$p=0.04$	–
6	$p<0.01$	$p=0.02$	$p<0.01$
7	$p<0.01$	–	–
8	$p=0.01$	$p<0.01$	$p<0.01$

A '–' indicates no significant increase in the error of detecting connectivity due to the variability of HRF parameters at the level of $p<0.05$.

Silent reading task

Functional MRI data on a silent reading task were collected from six normal subjects. The task involved silent reading of words and pseudowords as soon as they appeared on the screen. The resting condition involved fixating to a cross in the middle of the screen. The data of each subject contains 360 3D brain images. For more details, reader is kindly referred to Mechelli et al. (2000).

In order to obtain connectivity structure, SPM analysis was performed for activation detection. The time courses of significantly activated regions were extracted by taking the first principal component of time-series of voxels at the peak-activated site and its neighbors at the cluster level for all subjects (Zheng and Rajapakse, 2006). Time-series of all subjects were concatenated together to form the dataset for deriving brain connectivity with BN and DBN. MCMC simulation was applied on discretized data. As the number of regions involved in the dataset is large ($n=10$), we used longer burn-in and sampling phases, and a large sampling interval [50000, 50000, 100] to avoid trapping into local optimum.

Fig. 4 shows learned network structures from BN and DBN. For DBN, edges having probabilities of connectivity greater than 0.5 were selected. As seen, similar connectivity structures were found in both BN and DBN. The diagonal edges (from top left to bottom right) are not compared as every region is assumed to be related to its own history in DBN framework but not in case for BN. As seen from the structure derived from DBN, left hemisphere has more directed connections to the right hemisphere, which is consistent with the previous findings that left hemisphere is more involved in language processing. DBN found several connections previously reported in the literature for language processing: LSPL→RSPL (Honey et al., 2002) and REC→LEC for homologous interconnection (McIntosh et al., 1994), LSPL→RIFG for phonemic decisions (Honey et al., 2002), LMTC→LIFG for semantic phonologic retrieval (McKiernan et al., 2001; Matsumoto et al., 2004), LMTC→REC for memory retention (McIntosh et al., 1994), REC→LMFG and REC→ RMFG for semantic decision and analysis (Krause et al., 1999). Temporal relationships of REC and RMTC with LEC and LMTC were not found in BN, most likely due to bi-directional possibility of connections (Zheng and Rajapakse, 2006). In contrast, DBN found a recurrent network (LEC→LMTC→RMTC→REC→LEC) associated with retaining and recalling words from the memory, in which MTC is in charge of integrating inputs from lower level auditory and visual areas for retaining in the memory.

As the extrastriate cortex (EC) and superior parietal lobe (SPL) are important in visual representation and analysis in word processing, respectively, the functional connection between EC and superior parietal lobe (LSPL→LEC, LSPL→REC) and other connections with prefrontal cortex (LSPL→RIFG, LMFG→RSPL, REC→RMFG, REC→LMFG, RIFG→LEC, LIFG→LEC, LMFG→LEC) enable the perception of visual word form (Kolb and Whishaw, 1996). Connections between MTC and SPL indicate their close relationship in dealing with visual words (LSPL→RMTC, LMTC→RSPL). Within prefrontal cortex, intra- and interhemispheric connections were found similar to BN (LIFG→RMFG, LIFG→RIFG). The connection (LMFG→

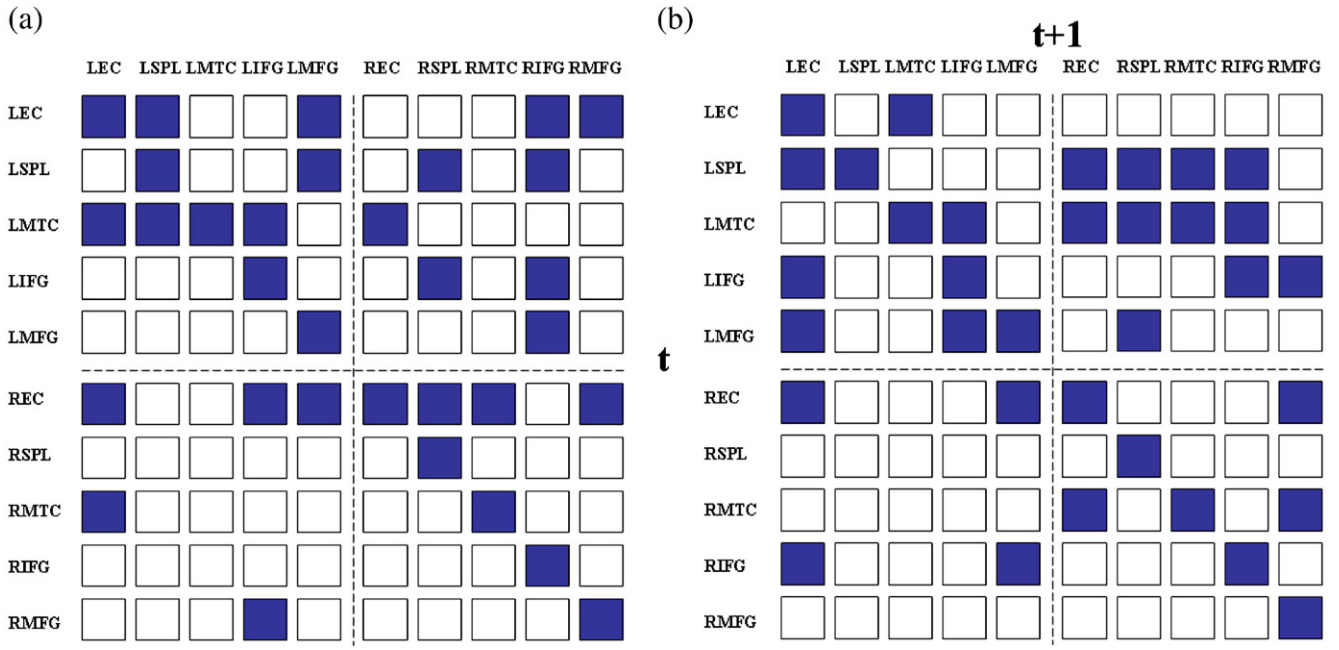

Fig. 4. The neural system learned from fMRI data on silent reading task with (a) BN and (b) DBN. L(R)EC: left(right) extrastriate cortex (BA 18/19), L(R)SPL: left(right) superior parietal lobe (BA 7), L(R)MTC: left(right) middle temporal cortex (BA 21/22), L(R)IFG: left(right) inferior frontal gyrus (BA 44/45), L(R)MFG: left(right) middle frontal gyrus (BA 46/9). A shaded square in the transition diagram of DBN indicates the existence of a connection from a region in the row to a region in the column. Rows and columns in DBN represent brain regions at time instances t and $t+1$, respectively.

LIFG) found in DBN was absent from BN, which has been earlier reported in an experiment demanding semantic categorization and subvocal rehearsal (Bullmore et al., 2000).

Counting Stroop task

Functional MRI data in a counting Stroop task investigating the performance of females with fragile X-syndrome on cognitive interference processing compared to normal subjects were analyzed. There are two conditions in the task: neutral and interference (Please refer to Tamm et al., 2002 for more details about the task and data collection). Using SPM2, time-series were extracted to derive connectivity. We explored the networks involved in neutral and interference counting Stroop condition of normal controls and attempted to make a comparison. Ten regions were found activated in neutral condition while 12 regions were found activated in the interference condition. Time-series of 14 normal control subjects were concatenated together to form the training data; same settings of MCMC simulation for learning the structure of DBN were used as for silent reading task.

Fig. 5 shows the learned structures for neutral and interference conditions by BN (top) and DBN (bottom). Note that the region with the empty diagonal square (from top left to bottom right) means that the region is not activated in the task. Fig. 6 shows the networks with connection strengths derived from DBN. The connectivity structures found by BN and DBN were quite similar: (1) a large amount of connections within medial cortices for both conditions, indicating the responsibility of counting function; and (2) the number of intra-hemispheric connections in left brain, and between left brain and medial cortices were much fewer in neutral condition than interference condition, indicating more involvement of language processing and decision making in the left hemisphere under interference effects.

Anterior cingulate cortex (ACC), connected with the prefrontal cortex and parietal cortex as well as motor system and frontal eye fields, is a central station for processing top-down and bottom-up stimuli and assigning appropriate control to other areas in the brain (Posner and DiGirolamo, 1998). ACC particularly plays an essential role in counting Stroop task (Bush et al., 1998; Hayward et al., 2004; Shin et al., 2001) to resolve competing streams of information in the selection of sensory inputs and responses. This is evidenced by connections in the interference network while absent from neutral network: ACC→MMFG, ACC→SMA, LLIFG→ACC, LLIMFG→ACC, VIFG→ACC. Because more concentration is required for interference task, the connections from ACC to MMFG, which controls the eye movement, are also absent from neutral task. ACC receives the processed input from frontal gyrus (VIFG, LLIMFG, LLIFG) and assigns appropriate motor control to supplementary motor area (SMA). DBN also discovered the connection RSPL→ACC for both neutral and interference condition. This is in consistence with earlier findings about the connection between ACC and parietal cortex (Posner and DiGirolamo, 1998). However, BN could not find this connection for interference condition.

Supplementary motor area (SMA) plays an important role in planning complex and coordinated movements (Kolb and Whishaw, 1996), and primary motor area (PMA) stores motor patterns and responds to voluntary activities (Faw, 2002). VIFG controls voluntary goal directed behavior (Tamm et al., 2002) and a connection chain (VIFG→LPMA→SMA) for voluntary movement planning were found in both conditions with both BN and DBN. AFG is participated in memory retrieval and executive

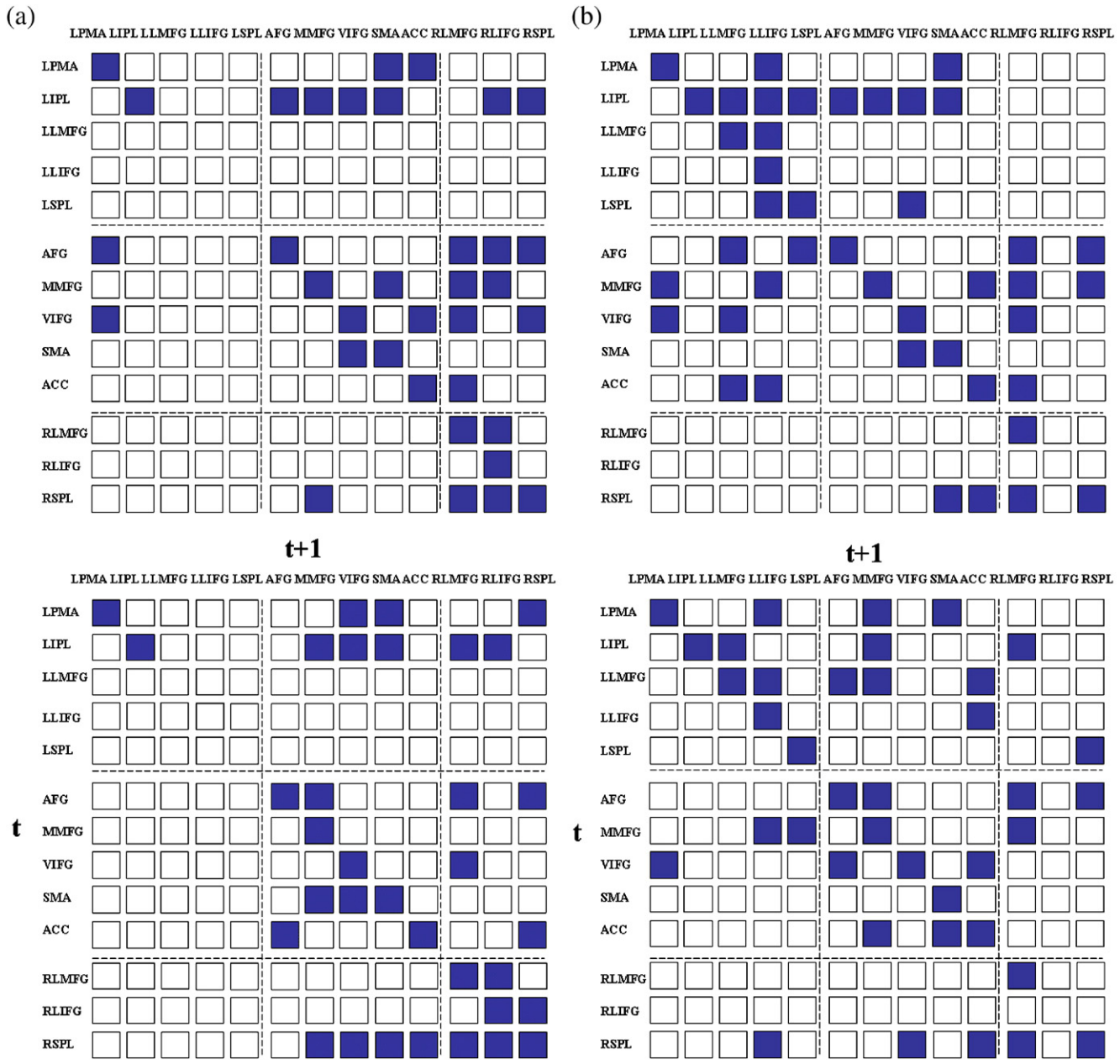

Fig. 5. The neural system learned from fMRI data of counting Stroop task (neutral and interference condition) by BN (top) and DBN (bottom). LPMA: left primary motor area (BA 4); LIPL: left inferior parietal lobe (BA 40); L(R)LMFG: left (right) lateral middle frontal gyrus (BA 9); L(R)LIFG: left (right) lateral inferior frontal gyrus (BA 44); L(R)SPL: left (right) superior parietal lobe (BA 7); AFG: anterior frontal gyrus (BA 10); MMFG: medial middle frontal gyrus (BA 8); VIFG: ventral inferior frontal gyrus (BA 47); SMA: supplementary motor area (BA 6); ACC: anterior cingulate cortex (BA 24). A shaded square indicates the existence of a connection from the region in the row (time t) to the region of column (time $t+1$).

function and LMFG is responsible for the selection of behavior based on the short term memory, in this case, solving conflict, and LIFG is in charge of phonemic decisions (Price, 2000). Both BN and DBN found similar connectivity structures: AFG→RLMFG exists for both neutral and interference conditions, RLMFG→RLIFG exists only in neutral condition while LLMFG→LLIFG is present only in interference condition. MMFG is believed to control eye movements (Faw, 2002) and related to uncertainty management (Volz et al., 2005). Thus, connections related to MMFG express different patterns of connectivity under conflicting situations. In neutral condition, connection SMA→MMFG exists while more involvement of MMFG is found in interference condition such as LPMA→MMFG, LLMFG→MMFG→LLIFG, and MMFG→RLMFG. In addition, DBN discovered an important connection between AFG and MMFG for both conditions, which are missing from networks derived by BN.

The above areas in frontal lobe receive inputs from parietal lobe which usually performs functions of processing and discriminating of sensory inputs. However, the connectivity between frontal gyrus and parietal lobe were different for neutral and interference con-

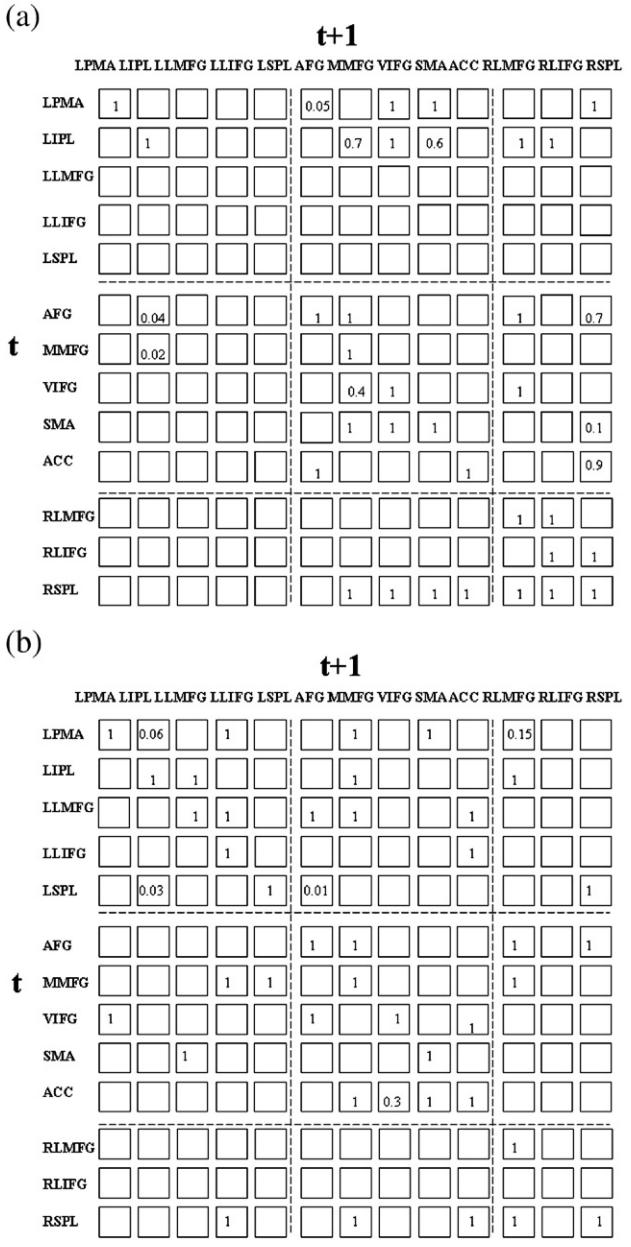

Fig. 6. The neural systems learned from fMRI data of counting Stroop task for neutral and interference conditions by DBN with their connection strengths shown. The rows and columns in DBN represent the regions in time t and $t+1$, respectively. Same abbreviations for brain regions are used as in Fig. 5.

ditions. Both BN and DBN found common connections AFG→RSPL→RLMFG and LIPL→MMFG in both neutral and interference conditions. In addition, DBN found new connections between frontal gyrus and parietal lobe common to both interference and neutral conditions: LIPL→RLMFG and RSPL→VIFG. Besides, the structure derived by DBN shows an important and specific language pathway (LIPL→LLMFG→LLIFG) in the interference counting task for phonetic and semantic analysis and decision. The connections RSPL→LLIFG and MMFG→LSPL are also absent from the neutral condition. Meanwhile, RSPL→RLIFG, LIPL→RLIFG, and RSPL→MMFG only exist in neutral network, which perform a compensatory function in the absence of language pathway to visualize the symbols instead of reading, or automatic speech in right hemisphere (Vanlancker et al., 2003).

Discussion

We proposed DBN to derive effective connectivity of activated brain regions in fMRI experiments. A structure learning algorithm based on MCMC method was introduced to search for the best connectivity structure predicted from data. The accuracy and the robustness of the approach were evaluated on synthetic data emulating true fMRI time-series characteristics. The performance was comparable with GCM in a linearly connected network. The discovered connectivity structures on fMRI data obtained on silent reading and counting Stroop tasks were consistent with the previous literature and showed improvement over those found by BN.

Confirmatory techniques of effective connectivity analysis such as SEM and DCM are only useful when a prior connectivity model is available, so they are not suitable for complex cognitive tasks or functional integration studies for which such prior models of connectivity are unavailable. Bayesian networks in contrast are able to derive the optimal connectivity structure from fMRI data in an exploratory manner without any knowledge of the structure. Moreover, BN and DBN are able to characterize the effective connectivity among brain regions in a complete statistical sense compared to linear VAR models assumed by GCM or second-order connectivity by SEM.

Dynamic Bayesian networks (DBN) are capable of learning the structure more accurately as they explicitly take into account the temporal characteristics of fMRI time-series by using a Markov chain. Our method found temporal and causal relationships among different brain regions more accurately in experiments with real fMRI data. The intrinsic equivalence property of BN results in a loss of information about edge directions because several network structures with the same skeleton but different edge directions can have the same marginal likelihood. Dynamic BN avoids the ambiguity of edge directions by explicitly taking into account the temporal relationships among brain regions in consecutive brain scans. The directions of edges in the DBN framework correctly represent information flow of brain function; for example, if there is a causal relationship from node A to node B, DBN represents this relationship with the existence of an edge $A(t)\rightarrow B(t+1)$ and the absence of an edge $B(t+1)\rightarrow A(t)$ where $A(t)$ and $A(t+1)$ denote the node A in layers representing time instances t and $t+1$, respectively, in the transition network.

Since the structure of BN must satisfy acyclic constraint, recurrent networks of brain connectivity are impossible. As feedback is an essential feature of biological systems, this is a limitation of BN for modelling brain connectivity. However, DBN is capable of modelling recurrent connections; for example, a recurrent network with three causal relationships, say $A\rightarrow B$, $B\rightarrow C$, and $C\rightarrow A$, is represented by the existence of three edges $A(t)\rightarrow B(t+1)$, $B(t)\rightarrow C(t+1)$, and $C(t)\rightarrow A(t+1)$ in the transition network. Similarly, a direct feedback loop with two nodes can be represented by edges $A(t)\rightarrow B(t+1)$ and $B(t)\rightarrow A(t+1)$.

There are some limitations of connectivity patterns derived by the present method. Although intra-scan connectivity is possible due to low temporal sampling in fMRI, the present approach cannot take into account the instantaneous interactions. Hemody-

namic characteristics of the brain are complex, for example, different orders (delay) of connections among regions, and higher-order DBN are required to handle such effects. Instead of simple concatenation of time-series from different subjects together, the inter-subject variability of brain connectivity in a group study could be better treated. The present approach assumes time-invariant connectivity among brain regions while empirical evidences suggest that the interscan connectivity between brain areas vary during scanning. How to overcome the above limitation of the present approach remains as future of this research.

Because of large complexity of human brain networks, functional integration research remains an important area of brain research. The complete validation of effective connectivity models derived from exploratory methods is difficult. We attempted to validate our method with synthetic data and justified the networks derived on real fMRI data with the aid of previous literature. In experiments, DBN showed more promise and accuracy in estimating connectivity structure than earlier methods as it takes into consideration of temporal and causal relationships among brain regions in higher-order statistical sense. We have shown with experiments the robustness of our method to characteristics of fMRI data such as sampling interval, hemodynamic noise, and scanner noise. In addition, we studied the variability of HRF on the accuracy of networks derived.

Whenever prior information about the functional structure is available, taking a hybrid of confirmatory and exploratory techniques should improve the accuracy of the connectivity pattern and render efficient algorithms for deriving connectivity. Though exploratory techniques of deriving connectivity patterns could open a new way of looking at brain functions, due care should be taken into account when interpreting such connectivity patterns. The explored structures from fMRI data could be helpful in distinguishing brain functions of different groups of subjects, such as healthy controls and patients.

Acknowledgments

This work was supported by a grant to J.C. Rajapakse, jointly by the Ministry of Education (MOE) and the Agency of Science, Technology and Research (A*Star), Singapore. The data sets were supported by fMRIDC (The fMRI Data Center, Dartmouth College, http://www.fmridc.org, accession number: 2-2000-11189 and 2-2001-1123b). We thank the anonymous reviewers for their valuable comments which have improved the quality of this manuscript.

References

Bagarinao, E., Sato, S., 2002. Algorithm for vector autoregressive model parameter estimation using an orthogonalization procedure. Ann. Biomed. Eng. 30, 260–271.

Buchel, C., Friston, K.J., 1997. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cereb. Cortex 7, 768–778.

Bullmore, E., Horwitz, B., Honey, G., Brammer, M., Williams, S., Sharma, T., 2000. How good is good enough in path analysis of fMRI data? NeuroImage 11, 289–301.

Bush, G., Whalen, P.J., Rosen, B.R., Jenike, M.A., McInerney, S.C., Rauch, S.L., 1998. The counting Stroop: an interference task specialized for functional neuroimaging validation study with functional MRI. Hum. Brain Mapp. 6, 270–282.

Chickering, D.M., 1995. A transformational characterization of equivalent Bayesian network structures. Int. Conf. Uncertainty Artif. Intell. 11, 87–98.

Eichler, M., 2005. A graphical approach for evaluating effective connectivity in neural systems. Philos. Trans. R. Soc. B 360, 953–967.

Faw, B., 2002. Pre-frontal executive committee for perception, working memory, attention, long-term memory, motor control, and thinking: a tutorial review. Conscious. Cogn. 12, 83–139.

fMRI DC, Dartmouth College, 2004. The fMRI Data Center, http://www.fmridc.org.

Friston, K.J., 2003. Dynamic causal modeling. NeuroImage 19, 1273–1302.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C.D., Frackowiak, R.S.J., 1995. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210.

Goebel, R., Roebroecka, A., Kimb, D., Formisanoa, E., 2003. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modelling and Granger causality mapping. Magn. Reson. Imaging 21, 1251–1261.

Harrison, L., Penny, W.D., Friston, K.J., 2003. Multivariate autoregressive modelling of fMRI time series. NeuroImage 19, 1477–1491.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometika 57, 97–109.

Hayward, G., Goodwin, G.M., Harmer, C.J., McCharley, R.W., 2004. The role of the anterior cingulate cortex in the counting Stroop task. Exp. Brain Res. 154, 355–358.

Heckerman, D., Geiger, D., Chickering, M., 1995. Learning Bayesian networks: the combination of knowledge and statistical data. Mach. Learn. 20, 197–243.

Honey, G.D., Fu, C.H.Y., Kim, J., Brammer, M.J., Groudace, T.J., Suckling, J., Pich, E.M., William, S.C.R., Bullmore, E.T., 2002. Effects of verbal working memory load on corticocortical connectivity modeled by path analysis of functional magnetic resonance imaging data. NeuroImage 17, 573–582.

Husmeier, D., 2003a. Inferring dynamic Bayesian networks with MCMC http://www.bioss.sari.ac.uk/dirk/software/DBmcmc/#Preliminaries.

Husmeier, D., 2003b. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. Bioinformatics 19 (17), 2271–2282.

Kolb, B., Whishaw, I.Q., 1996. Fundamental of Human Neuropsychology. W.H. Freeman and Company.

Krause, B.J., Horwitz, B., Taylor, J.G., Schmidt, D., Mottaghy, F.M., Herzog, H., Halsband, U., Muller-Gartner, H.W., 1999. Network analysis in episodia encoding and retrieval of word-pair associates: a PET study. J. Neurosci. 11, 3293–3301.

Matsumoto, R., Nair, D.R., LaPresto, E., Najm, I., Bingaman, W., Shibasaki, H., Luders, H.O., 2004. Functional connectivity in the human language system: a cortico-cortical evoked potential study. Brain 127, 2316–2330.

McIntosh, A.R., Gonzalez-Lima, F., 1994. Structural equation modeling and its application to network analysis in functional brain imaging. Hum. Brain Mapp. 2, 2–22.

McIntosh, A.R., Grady, C.L., Ungerieider, L.G., Haxby, J.V., Rapoport, S.I., Horwitz, B., 1994. Network analysis of cortical visual pathways mapped with PET. J. Neurosci. 14 (2), 655–666.

McKiernan, K.A., Conant, L.L., Chen, A., Binder, J.R., 2001. Development and cross-validation of a model of linguistic processing using neural network and path analyses with fMRI data. NeuroImage 13 (6), 2.

Mechelli, A., Friston, K.J., Price, C.J., 2000. The effects of presentation rate during word and pseudoword reading: a comparison of PET and fMRI. Cogn. Neurosci. 12, 145–156.

Mechelli, A., Penny, W.D., Price, C.J., Gitelman, D.R., Friston, K.J., 2002. Effective connectivity and intersubject variability: using a multi-subject network to test differences and commonalities. NeuroImage 17, 1459–1469.

Murphy, K.P., 2001. The Bayes net toolbox for Matlab. Comput. Sci. Stat. 33, 331–351.

Posner, M.I., DiGirolamo, G.J., 1998. Executive attention: conflict, target detection, and cognitive control. In: Parasuraman, R. (Ed.), The Attentive Brain. MIT Press, Cambridge, MA, pp. 401–423.

Price, C.J., 2000. The anatomy of language: contributions from functional neuroimaging. J. Anat. 197, 335–359.

Rajapakse, J.C., Piyaratna, J., 2001. Bayesian approach to segmentation of statistical parametric maps. IEEE Trans. Biomed. Eng. 48, 1186–1194.

Rajapakse, J.C., Kruggel, F., Maisog, J.M., von Cramon, D.Y., 1998. Modeling hemodynamic response for analysis of functional MRI time-series. Hum. Brain Mapp. 6, 283–300.

Rajapakse, J.C., Tan, C.L., Zheng, X., Mukhopadyay, S., Yang, K., 2006. Exploratory analysis of brain connectivity with ICA. IEEE Eng. Med. Biol. Mag. 25 (2), 102–111 (March/April Issue).

Roebroeck, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using Granger causality and fMRI. NeuroImage 25, 240–242.

Shin, L.M., Whalen, P.J., Pitman, R.K., Bush, G., Macklin, M.L., Lasko, N.B., Orr, S.P., McInerney, S.C., Rauch, S.L., 2001. An fMRI study of anterior cingulate function in posttraumatic stress disorder. Soc. Biol. Psychiatry 50, 932–942.

Tamm, L., Menon, V., Johnston, C.K., Hessl, D.R., Reiss, A.L., 2002. fMRI study of cognitive interference processing in females with fragile X syndrome. J. Cogn. Neurosci. 14 (2), 160–171.

Vanlancker, D., McIntosh, A.R., Grafton, S., 2003. PET activation studies comparing two speech tasks widely used in surgical mapping. Brain Lang. 85, 245–261.

Volz, K.G., Schubotz, R.I., von Cramon, D.Y., 2005. Variants of uncertainty in decision-making and their neural correlates. Brain Res. Bull. 67 (5), 403–412.

Wang, Y., Rajapakse, J.C., 2006. Contextual modelling of functional MR images with conditional random fields. IEEE Trans. Med. Imag. 25 (6), 804–812.

Zheng, X., Rajapakse, J.C., 2006. Learning functional structure from fMR images. NeuroImage 31 (4), 1601–1613.