The final version of this paper can be obtained from the following link.
「IEEE Xplore - Abstract Page」
http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6033247
Digital Object Identifier: 10.1109/IJCNN.2011.6033247

Recognition Model of Cerebral Cortex
based on Approximate Belief Revision Algorithm

Yuuji Ichisugi

Abstract—We propose a computational model of recognition of the cerebral cortex, based on an approximate belief revision algorithm. The algorithm calculates the MPE (most probable explanation) of Bayesian networks with a linear-sum CPT (conditional probability table) model. Although the proposed algorithm is simple enough to be implemented by a fixed circuit, results of the performance evaluation show that this algorithm does not have bad approximation accuracy. The mean convergence time is not sensitive to the number of nodes if the depth the network is constant. The computation amount is linear to the number of nodes if the number of edges per node is constant. The proposed algorithm can be used as a part of a learning algorithm for a kind of sparse-coding, which reproduces orientation selectivity of the primary visual area. The circuit that executes the algorithm shows better correspondence to the anatomical structure of the cerebral cortex, namely its six-layer and columnar features, than the approximate belief propagation algorithm that has been proposed before. These results suggest that the proposed algorithm is a promising starting point for the model of the recognition mechanism of the cerebral cortex.

I. INTRODUCTION

The cerebral cortex is the part of the brain that is most closely related to humans' high intelligence. It is strongly desired to uncover how the cerebral cortex recognizes and learns the outer world.

Some computational neuroscientists have begun to understand that the Bayesian network[1] is the essential mechanism of the cerebral cortex[2-10]. Models based on Bayesian networks can successfully explain the fundamental mechanism of the cerebral cortex: namely, robust pattern recognition using prediction based on context[3]. Furthermore, previous studies have strongly suggested that the cerebral cortex is a Bayesian network, according to models that reproduce electrophysiological phenomena[4][9] and models that explain the roles of major anatomical characteristics of the cerebral cortex[5][6]. Now, we can say that we have acquired a very important clue for understanding the brain.

Even if the cerebral cortex is a kind of Bayesian network, an essential question still remains: what kind of algorithm is used for recognition? There are at least two candidates, the belief propagation algorithm and the belief revision algorithm.

Most previous models[3-7,9] based on Bayesian networks claim that the recognition of the cerebral cortex is achieved by belief propagation[1], which calculates the posterior probabilities of all unobserved random variables; however, these models could not explain the theoretical relation between the recognition algorithm and the learning algorithm.

On the other hand, some models[8][10] assume that the purpose of recognition is not to calculate posterior probabilities but to find an MPE (most probable explanation), which can be efficiently achieved by the belief revision algorithm[1]. The MPE is the set of values of random variables that most likely explains the given observed data. One study[8] realized the approximate belief revision algorithm for the Markov Random Field (a model similar to Bayesian networks) by a biologically plausible neural circuit of spiking neurons. Another study[10] showed that the calculation of MPE with the hill-climbing method can be used as part of an on-line learning algorithm, which is a kind of sparse-coding algorithm[11] to reproduce the orientation selectivity of the primary visual area.

At this point it has not been determined which algorithm, belief propagation or belief revision (or some other algorithm such as Markov chain Monte Carlo) is the brain's actual recognition algorithm.

In this paper, we show evidence that supports the belief revision model. We propose an approximate belief revision algorithm that shows good performance despite its simplicity and good correspondence to the anatomical structure of the cerebral cortex, namely its six-layer and columnar features.

II. BAYESIAN NETWORKS

Bayesian networks are the model of knowledge representation that expresses causal relationships between random variables using a directed acyclic graph. Random variables are expressed as nodes, and relationships between random variables are expressed as edges. Each node has a table of conditional probability, which denotes the degree of causal relation between the node and the set of its parent nodes.

Belief propagation and belief revision are often used for inferences in Bayesian networks. Both algorithms have some similarities with the cerebral cortex. For example, they perform local and asynchronous communications and transfer only non-negative values. Nevertheless, exact algorithms would be too complex to be implemented by neurons naively.

Yuuji Ichisugi is with National Institute of Advanced Industrial Science and Technology(AIST) Tsukuba Central 2,Tsukuba,Ibaraki 305-8568, Japan (e-mail: y-ichisugi@aist.go.jp).

$$\lambda_{Y_l}^{t+1}(x) = \sum_{y_l}(\rho^t(y_l) + \lambda^t(y_l)P(y_l\,|\,x))$$

$$\lambda^{t+1}(x) = \prod_{l=1}^{n}\lambda_{Y_l}^{t}(x)$$

$$\kappa_{U_k}^{t+1}(x) = \sum_{u_k}P(x\,|\,u_k)BEL^t(u_k)$$

$$\pi^{t+1}(x) = \sum_{k=1}^{m}\kappa_{U_k}^{t+1}(x)$$

$$\rho^{t+1}(x) = \lambda^{t+1}(x)\pi^{t+1}(x)$$

$$Z_X^{t+1} = \sum_{x}\rho^{t+1}(x)$$

$$BEL^{t+1}(x) = \rho^{t+1}(x)\,/\,Z_X^{t+1}$$

Fig. 1. The approximate belief propagation algorithm[5].

$$\lambda_{Y_l}^{t+1}(x) = \max_{y_l}(\rho^t(y_l) + \lambda^t(y_l)P(y_l\,|\,x))$$

$$\lambda^{t+1}(x) = \prod_{l=1}^{n}\lambda_{Y_l}^{t}(x)$$

$$\kappa_{U_k}^{t+1}(x) = \max_{u_k}P(x\,|\,u_k)BEL^t(u_k)$$

$$\pi^{t+1}(x) = \sum_{k=1}^{m}\kappa_{U_k}^{t+1}(x)$$

$$\rho^{t+1}(x) = \lambda^{t+1}(x)\pi^{t+1}(x)$$

$$Z_X^{t+1} = \max_{x}\rho^{t+1}(x)$$

$$BEL^{t+1}(x) = \rho^{t+1}(x)\,/\,Z_X^{t+1}$$

$$x*^{t+1} = \arg\max_{x}BEL^{t+1}(x)$$

Fig. 2. Obtained approximate belief revision algorithm. x* is the estimated value of the node X.

III. Approximate Belief Propagation and Belief Revision Algorithms

The author has derived an approximate belief propagation algorithm[5] based on the following two natural assumptions.

1. Each conditional probability table (CPT) satisfies the following "linear-sum CPT model", which is qualitatively similar to the noisy-OR model[1]:

$$P(x\,|\,u_1,\cdots,u_m) = \frac{1}{m}\sum_{k=1}^{m}P(x\,|\,u_k)$$

Actually, the normalization factor $1/m$ can be ignored because it does not affect the normalized posterior probabilities. (This point has not been mentioned in [5].)

2. Each node has a sufficient number of parent and child nodes to make the following two approximations valid ($\pi_X(u_i)$ and $\lambda_{Y_l}(x)$ are messages from parents and children of a node respectively) :

$$\sum_{u_1,\cdots,u_m/u_k}\sum_{j\neq k}P(x\,|\,u_j)\prod_{i\neq k}\pi_X(u_i)$$
$$\approx \sum_{u_1,\cdots,u_m}\sum_{j}P(x\,|\,u_j)\prod_{i}\pi_X(u_i)$$

$$\prod_{j\neq l}\lambda_{Y_j}(x)$$
$$\approx \prod_{j}\lambda_{Y_j}(x)$$

Although the validity of the approximations based on the second assumption has not been experimentally verified in [5], a similar approximation for Markov Random Fields has been verified by simulation in [8].

The derived algorithm based on these two assumptions is shown in Figure 1. See [5] for more details on the derivation

process. The expression for λ_{Y_l} has been slightly changed but the changes are not essential. These formulas calculate "messages" from a node X to its parent nodes $U_k(k=1,\cdots,m)$ and child nodes $Y_l(l=1,\cdots,n)$, using old messages sent from its parents and children. After some iteration of message sending, the values of *BEL* are expected to converge to the posterior probabilities.

Unfortunately we cannot apply the derivation processes described in [5] to the belief revision algorithm because the expressions in it consist of both max- and sigma-operators.

Therefore, we use an ad-hoc approach. As mentioned in [1], the belief revision algorithm can be obtained by merely replacing sigma-operators with max-operators. So, we hypothesized that the approximate belief revision algorithm may be obtained by replacing "sigma-operators whose origin is not the CPT model" with max-operators. The obtained algorithm is shown in Figure 2. The validity of this algorithm is experimentally evaluated in the following sections.
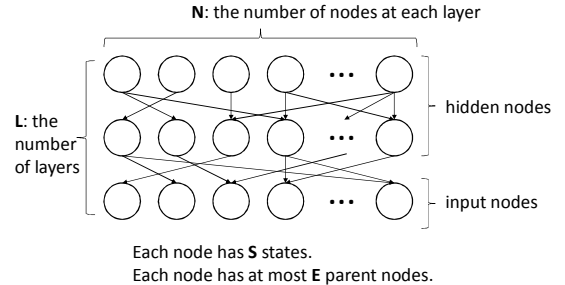

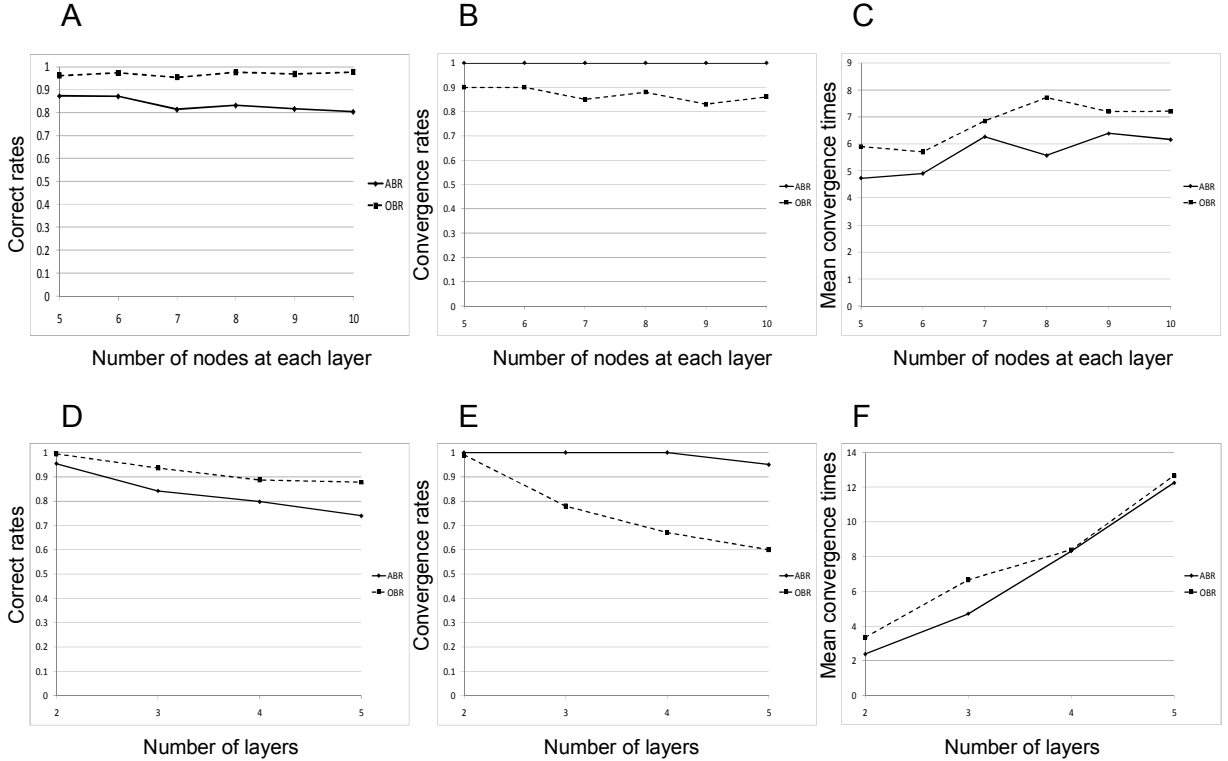Fig. 3. Structure of Bayesian networks used for evaluation.

Fig. 4. Graphs of correct rates (A and D), convergence rates (B and E) and mean convergence times (C and F) for ABR (solid lines) and OBR (broken lines). The upper graphs (A, B and C) plot results on shallow networks with various widths. (L=2, N=5, 6, ..., 10, S=2, E=5) [erratum: L=2 to L=3] The lower graphs (D, E and F) plot results on narrow networks with various depths. (L=2, 3, ..., 5, N=5, S=2, E=full)

IV. EVALUATION METHOD

A. Network Structure

We have executed algorithms on Bayesian networks that have a layered structure similar to that of the brain (Figure 3). The number of layers is denoted by "L" hereafter. There are no edges between the nodes when they are in the same layer.

The observed data are input to the input layer, which is the lowest layer. The other layers are hidden layers, which contain unobserved (hidden) variables. All layers contain the same number (denoted by "N") of nodes. All random variables have the same number (denoted by "S") of states. Elements of the conditional probability tables are randomly set, but values close to zero are avoided.

Two nodes at the neighboring layers are randomly connected and each node has at most "E" parent nodes. More precisely, each node except for nodes at the highest layer randomly chooses "E" parent nodes allowing duplication from the next higher layer.

Some simulations use fully connected neighboring layers (denoted by "E=full").

B. Algorithms

In order to evaluate the proposed algorithm, we have implemented three algorithms, a naive full search algorithm to find strict MPE (STR), the original loopy belief revision (OBR) and the proposed approximate belief revision (ABR). In OBR and ABR, at each iteration step, all messages are sent to the neighboring nodes simultaneously. The initial values of messages are random values. To avoid overflow and underflow, λ-messages are rescaled before using; the rescaling makes the maximum value of the elements of a λ-message be 1.

C. Measures of Performance

We measured three values: correct rates, convergence rates and mean convergence times.

The correct rate is defined as the number of hidden variables whose estimated values are correct MPE values, divided by the number of all hidden nodes. The correct rate can be calculated only for the small networks to which the STR algorithm can be applied.

The convergence rate is defined as the number of recognition trials at which all hidden variables converge within the predetermined number of iteration steps (50 steps,
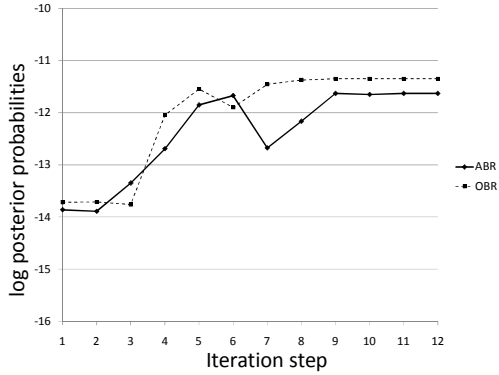
Fig. 5. An example of a transition of log posterior probabilities of estimated values at the time after each iteration step. More accurately, we use **f(t)= log P(*<estimated value at* t *>* | *<input>*) – c** as index, where **t** is a number of iteration. In this case, OBR (broken line) converges to the strict MPE and ABR (solid line) seems to converge to a sub-optimal solution. (L=5, N=5, S=2, E=full)
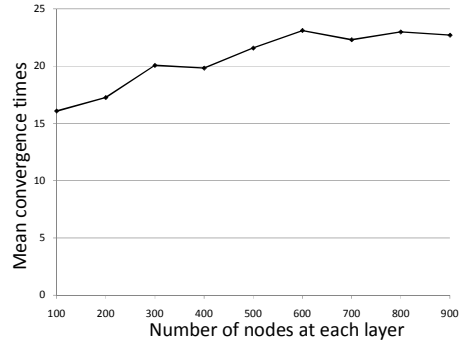

Fig. 6. Mean convergence times for shallow and wide networks. (L=3, N=100, 200, ..., 900, S=4, E=20)
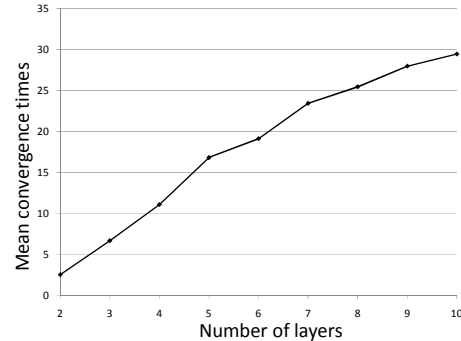

Fig. 7. Mean convergence times for deep networks. (L=2, 3, ..., 6, N=10, S=2, E=5)
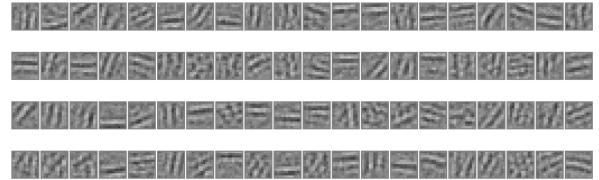

Fig.8. Learning result of natural images. Visualized conditional probability tables have orientation selectivity. 4 rows of 20 images are corresponding to 4 nodes with 20 states. Large values of conditional probabilities of each state are shown by bright pixels.

in all simulations in this paper), divided by the number of all trials.

The mean convergence time is defined as the total number of iteration steps before convergence divided by the number of converged recognition trials. Trials in which some hidden variables did not converge are excluded for calculation of the mean convergence time; therefore, this measure is meaningful only for network configurations that achieve high convergence rates.

All graphs in this paper plot the mean measured values of 100 recognition trials with differently initialized networks.

V. RESULTS

A. Accuracy for Small Networks

Firstly, we show the result of evaluation of ABR and OBR on small networks.

Graphs in Figure 4 plot the correct rates, the convergence rates and the mean convergence times.

The upper graphs plot results on shallow networks with various widths. The lower graphs show results on narrow networks with various depths.

Although ABR is inferior to OBR, its accuracy is not so bad, despite its simplicity (Figure 4 A and D). It seems that ABR is less likely to vibrate compared to OBR, which vibrates easily when the network becomes deeper (Figure 4 B and E). When the network converges, the mean convergence time of ABR is slightly shorter than that of OBR (Figure 4 C and F).

Although the correct rate of ABR is not 100%, the posterior probabilities of estimated values tend to increase as iteration steps progress. And usually, the difference of log posterior probabilities between the estimated values and the strict MPE is not so large. Therefore, it is thought that the estimated values of ABR may be sub-optimal solutions. Figure 5 shows an example of the transition of log posterior probabilities in a recognition trial. In this case, OBR converges to the strict MPE.

B. Scalability

We have also evaluated the scalability of the ABR algorithm.

Figure 6 and Figure 7 plot the mean convergence times of ABR on shallow networks with various widths and narrow networks with various depths, respectively. (We have chosen these parameters carefully so that convergence rates become always nearly 1.)

Figure 6 shows that the mean convergence time is not so sensitive to N if the network depth is constant.

On the other hand, Figure 7 shows that the mean convergence time is almost linear to the depth of the network.
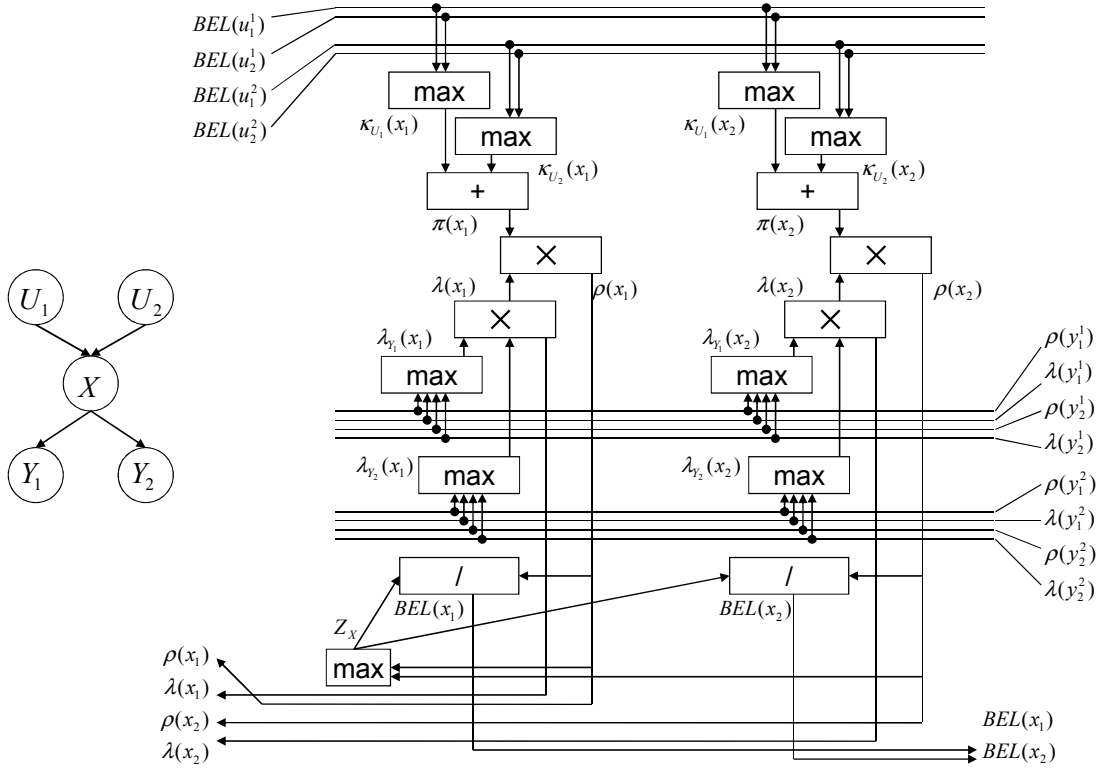
Fig. 9. A biologically plausible circuit that calculates messages from node X to the two parent nodes, U_1 and U_2, and to the two child nodes, Y_1 and Y_2. All nodes are binary, just for making the figure simple. See also Figure 10 which describes the details of the calculation of $\kappa_{U_k}(x_i)$ and $\lambda_{Y_i}(x_i)$. The circuit structure has congruencies with the six-layer and the columnar features of the cerebral cortex structure (see main text).

According to current anatomical knowledge, the depth of the hierarchy of cerebral areas of the macaque monkey is about ten. This means the depth of the Bayesian network represented by the cerebral cortex is not so deep. On the other hand, the number of nodes may be more than one hundred thousand if each macro-column (or hyper-column) of the cerebral cortex is assumed to be a node of a Bayesian network. The scalability of ABR shown in this section leads us to expect that ABR will work well on such large-scale networks. In addition, the characteristics of ABR's scalability might explain why the evolution of the brain makes the number of columns larger but does not makes the network depth deeper.

C. Computation Amount

The computation amount required for each iteration step of ABR is linear to the number of nodes, if the number of edges per node is constant. Indeed, we have verified this characteristic by measuring the computation time during the experiments of Figure 6 and Figure 7.

Because ABR can be easily executed in parallel, massively parallel computers will execute each iteration step of ABR in a constant time. (OBR may also have same characteristic.)

This efficiency might explain a surprising fact: although mammals have very different-size brains, all of them work well in real time.

VI. LEARNING NATURAL IMAGES

In this section, we show that ABR can be used as a part of a learning model of the cerebral cortex.

The used learning algorithm is almost the same as that described in [10], which uses a hill-climbing method instead of ABR, however. The neighborhood learning and sparseness control described in [10] is omitted here. Input images are gray-scale natural images preprocessed by a 3×3 Laplacian filter in order to emphasize the edges. We used a two-layered network that consists of 4 hidden nodes with 20 values and 12×12 input nodes with two values.

Each cycle of the learning process consists of a recognition step and a learning step. When an input image (randomly selected 12x12 pixel image) is given as observed data in the input layer, MPE is calculated (recognition step). Then, the conditional probability tables are updated based on the estimated MPE values (learning step).

Figure 8 is the result of learning, which shows acquired basis images (learned conditional probabilities) that have orientation selectivity like receptive fields of simple cells in the primary visual area[11].

This result indicates that ABR models a recognition part of the actual brain's learning of natural images with sufficient accuracy.
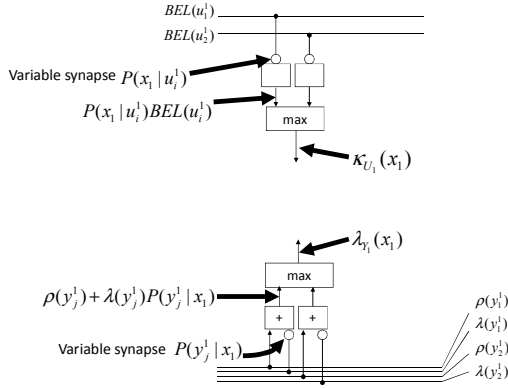
Fig. 10. Detailed structure of input to the calculation units of $\kappa_{U_k}(x_i)$ (upper) and $\lambda_{Y_l}(x_i)$ (lower).

VII. CORRESPONDENCES OF THE ALGORITHM WITH THE ANATOMICAL STRUCTURE OF THE CEREBRAL CORTEX

The circuit that executes the ABR shows better correspondence to the anatomical structure of the brain than the approximate belief propagation algorithm (ABP, Figure 1) proposed before.

Figure 9 is a circuit that consists of simple arithmetic units. The circuit calculates the values of messages to the neighboring nodes. Places of the plastic synapses that represent conditional probabilities are shown in Figure 10.

We have matched arithmetic units that calculate the variables κ, ρ, λ_{Y_l} and BEL with layer 2, 3, 4 and 5 of the cerebral cortex respectively. According to this circuit, ABR has, like ABP[5], the following correspondences with the anatomical structure: A) Bottom-up (feed-forward) connections are mainly from layer 3 to layer 4 and top-down (feed-back) connections are mainly from layers 5/6 to layer 1[12]. B) Information is mainly processed within long and slender structures called cortical columns, C) Many horizontal fibers are seen in layers 1 and 3 and D) Many cells are seen in layers 2 and 4.

ABR has one more correspondence than ABP does. It has been shown that max-operation can be implemented by a biologically plausible recurrent network of excitatory and inhibitory neurons[8]. The circuit of Figure 9 predicts the existence of max-operation units at layers 2, 4 and 5. Indeed, there exist recurrent circuits at layer 2/3, 4, 5 and 6 in the actual cerebral cortex[13]. In addition to this, as mentioned in [8], some electrophysiological experiments found max-like responses in the cerebral cortex. These facts support the validity of belief revision models rather than belief propagation models.

VIII. CONCLUSION

The evidence described in this paper suggests that the proposed approximate belief revision algorithm is a promising starting point for the model of the recognition mechanism of the cerebral cortex.

On the other hand, previous models based on belief propagation have reproduced some important electrophysiological phenomena[4][9]. We think the belief revision with some extension might also reproduce the same phenomena because the meaning of the variable BEL is very similar in both algorithms. We think this is important to pursue in future work.

From the point of view of engineering, because the proposed algorithm efficiently works in large-scale Bayesian networks with a linear-sum CPT model, its improved version may be expected to be used as a part of efficient pattern-recognition mechanism.

REFERENCES

[1] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, 1988.

[2] T.S. Lee, D. Mumford, Hierarchical Bayesian inference in the visual cortex. Journal of Optical Society of America, A. . 20(7): 1434-1448, 2003.

[3] D. George, J. Hawkins, A hierarchical Bayesian model of invariant pattern recognition in the visual cortex, in proc. of IJCNN 2005, vol. 3, pp.1812-1817, 2005.

[4] R. Rao., Bayesian inference and attention in the visual cortex. Neuroreport 16(16), 1843-1848, 2005.

[5] Y. ICHISUGI, The cerebral cortex model that self-organizes conditional probability tables and executes belief propagation, In Proc. of International Joint Conference on Neural Networks (IJCNN2007), pp.1065--1070, Aug 2007.

[6] F. Roehrbein, J. Eggert, and E. Koerner, Bayesian Columnar Networks for Grounded Cognitive Systems, In Proc. of the 30th Annual Conference of the Cognitive Science Society, pp.1423--1428, 2008.

[7] H. Hosoya: A motor learning neural model based on Bayesian network and reinforcement learning, In Proceedings of International Joint Conference on Neural Networks, 2009.

[8] S. Litvak, S. Ullman: Cortical Circuitry Implementing Graphical Models, Neural Computation 21, 3010.3056, 2009.

[9] S. Chikkerur, T. Serre, C. Tan and T. Poggio, What and Where: A Bayesian Inference Theory of Attention, Vision Research, 55(22), pp. 2233–2247, Oct 2010.

[10] Y. Ichisugi, H. Hosoya, Computational Model of the Cerebral Cortex that Performs Sparse Coding Using a Bayesian Network and Self-Organizing Maps, In Proc. of 17th International Conference on Neural Information Processing (ICONIP 2010), Part I, LNCS 6443, pp.33--40, Nov 2010.

[11] B.A. Olshausen, D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, NATURE 381 (6583): 607-609 JUN 13 1996.

[12] D.N. Pandya, and E.H. Yeterian, Architecture and connections of cortical association areas. In: Peters A, Jones EG, eds. Cerebral Cortex (Vol. 4): Association and Auditory Cortices. New York: Plenum Press, 3-61, 1985.

[13] C.D. Gilbert, Microcircuitry of the visual-cortex, Annual review of neuroscience, 6: 217-247, 1983.