Bayesian AutoEncoder: Generation of Bayesian Networks with Hidden Nodes for Features

Kaneharu Nishino and **Mary Inaba**
Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan
{nishino.kaneharu, mary}@ci.i.u-tokyo.ac.jp

Abstract

We propose Bayesian AutoEncoder (BAE) in order to construct a recognition system which uses feedback information. BAE constructs a generative model of input data as a Bayes Net. The network trained by BAE obtains its hidden variables as the features of given data. It can execute inference for each variable through belief propagation, using both feedforward and feedback information. We confirmed that BAE can construct small networks with one hidden layer and extract features as hidden variables from 3x3 and 5x5 pixel input data.

Introduction

In these days, Deep Learning has high performance in object recognition tasks (Le 2013). It trains multi layer neural networks to extract patterns of given data. In trained networks, nodes in the lower layers respond for simpler patterns (like lines), and nodes in the higher layers respond for more complex patterns (like faces). This hierarchical architecture of patterns is also found in our brains.

On the other hand, though the neural nets used in Deep Learning generally have feedforward architectures, brains have not only feedforward connection from the lower level to the higher level but also feedback connections from the higher to the lower. As the reason for these top-down connections, some studies propose that the information processing model of brains is based on a Bayesian Network (Bayes Net) (Lee and Mumford 2002) (Shon and Rao 2005) (Rao 2005) (Kenji Doya and Rao 2007) (Matsumoto and Komatsu 2005) (Hosoya 2012). According to these studies, using predictions and attentions as top-down information, brains can recognize objects along Bayes inference.

Based on these studies, we consider that multi layer Bayes Net is necessary for making brainlike recognition system. In this paper, we propose Bayesian AutoEncoder as a method to construct multi layer Bayes Net as a recognition system.

Bayesian AutoEncoder

BAE is a method that constructs a generative model as a Bayes Net. It decides the network structure and adjusts the conditional probabilities so that the network can predict each part of a given datum from the rest. It obtains hidden variables as the features of given training data. The networks constructed by BAE can behave as a Bayes Net, therefore they can recognize features through belief propagations.

BAE first constructs a two layer Bayes Net as a directed complete bipartite graph that contains a visible layer of child variable nodes and a hidden layer of parent variable nodes. BAE adjusts its parameters and cuts some links so that the network can predict input data. The variables are binary variables, which are either true (T) or false (F) (existent or non existent respectively) of corresponding features. When the network has been trained sufficiently, BAE stacks a new hidden layer over the parent variables, and adjusts parameters between them in the same way. Thus BAE constructs a multi-layer Bayes Net as a generative model, like Figure 2. In this paper, we implemented BAE in order to construct the Bayes Net with one parent layer and one child layer.

The parameters of the network are two types: Link Strength (LS) and Conditional Probability (CP). LSs refer to how each variable has links and how strongly each link affects the inference. CPs represents conditional probabilities of child variables between parent variables. Updating these parameters consists of two stages: updating LSs and updating CPs. BAE carry out these stages alternately, based on messages used in belief propagation. The LS between a parent u and a child v is updated to the absolute value of the correlation coefficient between the message to v from u and input value of v. $p(v|u)$ the CPs between a parent u and a child v are updated so that inference for v by messages only from parents of v is correct.

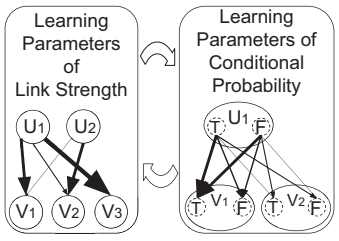


Figure 1: Two stages of learning

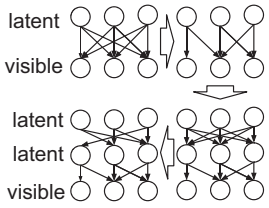
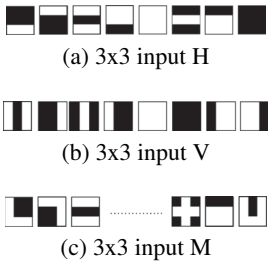


Figure 2: An example of network construction by BAE

0	1	2
3	4	5
6	7	8

Figure 3: Index of pixels

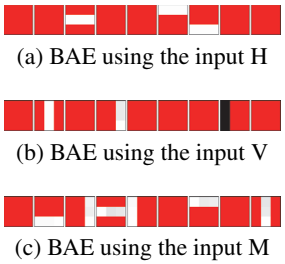


(a) 3x3 input H

(b) 3x3 input V

(c) 3x3 input M

Figure 4: The three groups of input data: (a) input H, (b) input V and (c) input M.

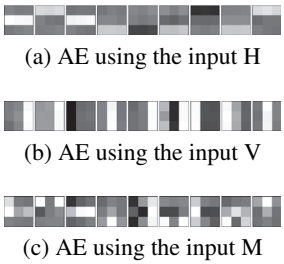


(a) BAE using the input H

(b) BAE using the input V

(c) BAE using the input M

Figure 5: Extracted links of parents by BAE, using (a) the input H, (b) using V and (c) using M.



(a) AE using the input H

(b) AE using the input V

(c) AE using the input M

Figure 6: Weights learned by AE, from (a) the input H, (b) from V and (c) from M.

Experiments for Constructing Networks

To confirm BAE constructs the network with parents representing features of input data, we carried out several experiments. We generated 3x3 or 5x5 pixel input data from generators represented as Bayes Net we prepared. We constructed small networks through BAE using these input data, and confirmed the networks have same structure as the Bayes Net generating input data.

3x3 Experments for Feature Extraction

Examples of generated images for the 3x3 experiments are shown in Figure 4. We prepared three generator and generated three groups of images: Horizontal (H), Vertical (V), Mixed (M). Each generator has some factors to generate images; H: Three factors which whiten each triple of pixels (0, 1, 2), (3, 4, 5) and (6, 7, 8), V: Three factors which whiten (0, 3, 6), (1, 4, 7) and (2, 5, 8), M: Six factors used in generator H and V, where pixels are indexed as Fgiure 3. Each factor has a state T or F randomly, and T state factors whiten corresponding pixels, and else pixels remain black. We constructed the network with nine parents. The children and parents were fully connected initially, and BAE updated LSs and CPs.

We visualized the parameters with colors in Figure 5. It shows how the parents have links; Each box represents a parent, 3x3 tiles in the box does a children (pixels). The color represents to the LS and CPs of the links. If the CP $p(v = T|u = T)$ between the parent u and the child v is near 1, the corresponding tile is brighten, and if near 0, it is dark. Red tiles represent the links has low LS and were cut.

From these results, we can confirm that the parents reflect the factors generating the input data. From the input H, each parent has links to the pixels in only a horizontal row. A case in the input V is the same. Also when the input M, horizontal rows and vertical columns are extracted as its factors.

For comparison, we have tried the same feature extraction through the auto-encoder (AE). Trained weights without any regularizations results in Figure 6. Gray tiles represent zero weights, and the brighter do the greater weights. Figure 6(a), 6(b) and 6(c) are from the input H, V, and M.

5x5 Experiments for Confirmation Scalability

To confirm that BAE can work with large network, we carried out 5x5. In this experiment, BAE constructed the network with 5x5 pixel input and 25 parents. The input is more complex than 3x3, like Figure 7. The obtained links are shown in Figure 8. BAE extracted 16 features, which are same as factors used to generate 5x5 input images.

Figure 7: Examples of 5x5 input.

Figure 8: Weights learned by BAE given 5x5 input.

Concluding Remarks

We proposed Bayesian AutoEncoder as a method to construct a Bayes Net with hidden nodes representing features of given data. Observing Figure 5(c) and Figure 6(c), BAE extracted features correctly in the case AE didn't. It can extract features of input as hidden nodes correctly in 3x3 and also in more difficult case of 5x5. We consider it may construct more complex networks with multi hidden layer.

The top-down messages in the network trained through BAE can be used for filling-in a lack of data. As a future work, we are going to try image completions.

References

Hosoya, H. 2012. Multinomial bayesian learning for modeling classical and nonclassical receptive field properties. *Neural Computation* 24(8):2119–2150.

Kenji Doya, Shin Ishii, A. P., and Rao, R. P. 2007. *Bayesian brain : probabilistic approaches to neural coding*. Computational neuroscience. Cambridge, Mass. MIT Press.

Le, Q. 2013. Building high-level features using large scale unsupervised learning. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 8595–8598.

Lee, T. S., and Mumford, D. 2002. Hierarchical bayesian inference in the visual cortex.

Matsumoto, M., and Komatsu, H. 2005. Neural responses in the macaque v1 to bar stimuli with various lengths presented on the blind spot. *Journal of Neurophysiology* 93(5):2374–2387.

Rao, R. P. N. 2005. Bayesian inference and attentional modulation in the visual cortex. *Neuroreport* 16(16):1843–1848.

Shon, A. P., and Rao, R. P. 2005. Implementing belief propagation in neural circuits. *Neurocomputing* 65-66(0):393–399. Computational Neuroscience: Trends in Research 2005.