Integration of situated prior probability and neural network classifier in a handwriting recognition task

Yoichi MOTOMURA
Email:motomura@etl.go.jp

†Electrotechnical Laboratory
Umezono 1-1-4, Tsukuba, Ibaraki, 305-8568, Japan

ABSTRACT

To achieve robust and flexible information processing ability, we study how contextual information can be used, in addition to direct observation. In order to deal with a concrete example, a handwriting character classification task is considered. We define the framework studied here as an integration of a situated prior probability and conventional classifier, feed-forward neural networks. The situated prior probability can be modeled by a conditional probability described on Bayesian network. The output of the classifier is proportional to likelihood of the input pattern. Then, the integration proposed in this paper can be interpreted as a posterior probability in Bayesian framework. We demonstrate an experiment about ambiguous handwriting character recognition. The result improves better classification performance than an ordinary neural network classifier.

KEYWORDS: Probabilistic and Statistical Methods, Real World Systems, Cognition and Perception, System Integration

1. Introduction

Human-like information processing, such as, recognition task for ambiguous, missing or incomplete observations, is one of the desirable features for an intelligent system working in the real world robustly. In real world problem solving, people use not only direct observation, but also contextual information, previous observations, or experiences which depend on particular situation in the task.

For example, even in a dark room, a person can find a light switch in his room easily, because he can estimate the approximate position of the switch with his belief derived from his daily experiences and other information like walk distance or something. Such internal belief is quite efficient for the cognitive process in the real world. The question is how to model this belief, and how to apply this information to an intelligent system. To make this study clear, we concentrate on a pattern recognition task in this paper.

Let us consider the Selfridge's ambiguous characters[2] shown in Figure 1. The second character alone cannot be uniquely identified. In the upper row in Figure 1, people can recognize the second character as 'H'. In the bottom row, he can recognize second character as 'A'. Why ? People get an additional information from recognizing results of the neighboring characters in each case. This additional information can be formulated as prior probability in Bayesian probabilistic framework. Major pattern recognition schemes get result to maximize a posterior probability in Bayesian framework, but conventional methods treat the prior probability as static or uniform probability distribution. The point emphasized in this context is such prior probability should be context and situation dependent. We call this concept situated prior probability and it is represented by a probability distribution conditional by the particular situation. This probability is also understood as the degree of belief for the interest matter.

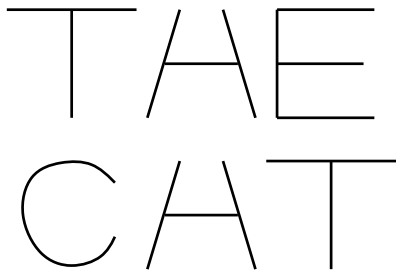

Figure 1: Ambiguous characters

In this paper, we discuss how we can model this situated prior probability, and how we can integrate it with direct information processing as in neural network classifiers. The key idea is that the situated prior probability is considered as the conditional prior probability represented by Bayesian network, and it is combined naturally with the likelihood of observations in Bayesian framework. In order to show the efficiency of the proposed method in a concrete example, we consider an ambiguous pattern handwriting classification task. The advantages of our approach are flexible representation ability and a systematic scheme to realize probabilistic reasoning for various applications, because even if the situated prior probability is complex, conditional probability can be modeled by more sophisticated Bayesian networks.

In the second section, we introduce Bayesian framework, which is the basic element of our study. In section 3, situated prior probability is discussed in pattern recognition problem. In section 4, a integration model to treat situated prior probability and simple classifiers using neural networks. In order to extend Bayesian networks to non-linear and continuous domain, neural networks are used to represent conditional probabilities [8]. Then, integration model can be interpreted as a large neural network combined with modular simple neural networks. In section 5, this integration model is evaluated in an ambiguous handwriting character recognition task. We show that the result obtained with our integration model is better than that obtained with an ordinary neural network classifier learnt by back propagation methods.

2. Bayesian framework

Bayesian pattern classification scheme is executed in the following manner:

A pattern x is classified into the class c_i that maximizes the a posterior probability [1] $P(c_i|x)$ according to Eq.(1).

$$P(c_i|x) = \frac{P(x|c_i)P(c_i)}{\sum_j P(x|c_j)P(c_j)}, \quad (1)$$

where $P(x|c_i)$ is the likelihood that x belongs to c_i. Prior distribution of c_i, $P(c_i)$ can imply the prior belief that a pattern belong to class c_i.

This is a Bayesian classifier. If all prior $P(c_i)$ are set to the same constant value on this framework, it is equivalent to a maximum likelihood method. A feed forward neural network pattern classifier resembles a maximum likelihood method when an output neuron generates output proportional to the likelihood for each class. The key point of effective Bayesian framework is setting prior probability adequately for the target patterns to classify.

3. Situated prior probability

We define prior probability as a degree of belief related to a possibility for each class before observing x. It is obvious that this belief could be affected by a particular situation or a contextual information, which could be derived from information different from direct observation x. We call such a belief *situated prior probability*. This situated prior probability can be described as the conditional probability, $P(c_i|S)$, where S denotes a situation. Then, the Bayes theorem for pattern recognition Eq.(1) is also described as,

$$P(c_i|x,S) = \frac{P(x|c_i)P(c_i|S)}{\sum_i P(x|c_i)P(C_i|S)}. \quad (2)$$

When the likelihood is not enough to distinguish class of x, this prior probability is relevant to evaluate posterior probability $P(c_i|x,S)$. (Figure 2, Figure 3).

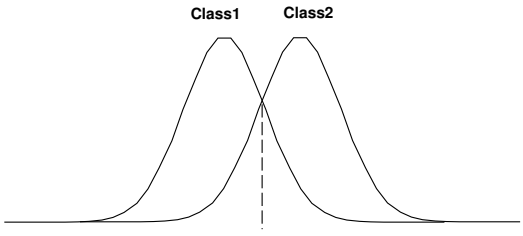

Figure 2: Posterior probability ($P(Class1) = P(Class2)$) Overlap of Gaussians corresponds to classification error rate.

4. Models for integration

4.1. Bayesian network

Bayesian networks (also called belief networks)[1] are probabilistic reasoning models that represent the dependency between variables and give a concise specification of the joint probability distribution. They are defined as directed acyclic graphs (DAGs) which have conditional probabilities in each node. Nodes denote random variables in a domain. Directed links in the graph denote dependency between nodes. Complex dependency among related random variables can be represented in this model. The structure of a DAG depends on each practical application. Because conditional probabilities are not obtained explicitly, each each conditional probability is estimated from a database of simultaneous incidence examples. After constructing a Bayesian network, probabilistic reasoning is executed to estimates a degree of belief about uncertain variables.

[1] In this article, $P(x)$ denotes the probability of x,$P(x|y)$ denotes conditional probability of x, given y.

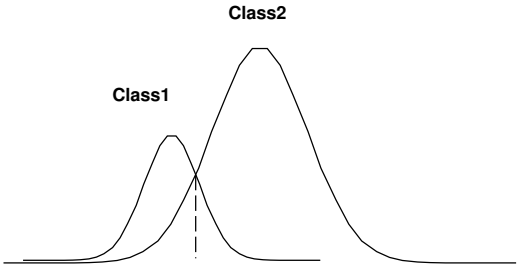

Figure 3: Posterior probability ($P(Class1) < P(Class2)$) Classification error rate decrease than Figure 1.

In order to represent a suitable situated prior probability for a particular task, we use Bayesian network specialized for a domain of the task. Then, we can construct an integration model combined with this situated prior probability and a pattern classifier as,

(1).A situated prior probability $P(c_i|S)$ is approximated by Bayesian network, **(2).**A maximum likelihood classifier is constructed by neural network learning to approximate likelihood, and **(3).** The classifier integrates an estimation result of Bayesian network, $P(c_i|S)$ and output of neural network classifier, $P(x|c_i)$ by Eq.(2).

The situated prior probability for ambiguous character recognition task is implemented using Bayesian network shown in Figure 4. C_n denotes n-th character in a line. This Bayesian network means that C_n depends on its neighbor C_{n-1} and C_{n+1}.

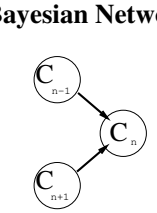

Figure 4: Bayesian network for ambiguous characters

4.2. Neural implementation

In order to handle different kinds of variables, such as multidimensional or continuous information, we implemented Bayesian network using an artificial neural net-

work to represent conditional probability. We The author's group developed a Bayesian net system based on neural implementation(BAYONET)[8].

As shown in Figure 5, the child node of the Bayesian network is represented by an output neuron, and the parent nodes, A and B, correspond to input neurons, so that each child node of the Bayesian network has its own neural network.

Thanks to this neural implementation, the system can handle multidimensional variables and both continuous and discrete variables in the same style. Conditional probabilities are taught by the back propagation algorithm.

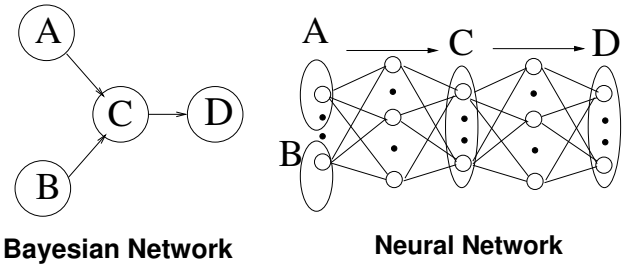

Figure 5: Bayesian Network on Neural Networks

5. Experiments

5.1. Outline

As the followings, a classifier is constructed using a feed-forward neural network that has two hidden layers. The network has ten neurons in each hidden layer, a hundred input-neurons and ten output-neurons. Input image for a handwriting character is a 20×20 dot matrix. Handwriting characters are written on this dot matrix by a mouse device. Each input neuron corresponds to 2×2 receptive field, and it takes value between 0 and 4 according to the density in the receptive field. As a teaching data set for the neural network, ten handwriting characters {A,B,C,D,E,H,I,J,K,T} are given (Figure 6), the network is trained until the mean squared error reaches less than 0.01 by the back propagation.

The Bayesian network for this problem (Figure 4) learned conditional probabilities $P(C_n|C_{n-1}, C_{n+1})$ from an English dictionary containing 43,000 words.

We evaluated the performance of an ordinary neural network classifier constructed by input layer, two hidden layer and output layer(I) and the performance of the integration model shown in Figure 7(II). Test patterns are shown in the second image in the bottom line of Figure 6.

5.2. Results

Classification results (%) for each test pattern of I and II are shown in Table 1.

	A	B	C	D	E	H	I	J	K	T
I	15.6	3.3	17.9	9.1	7.5	22.1	8.6	3.8	9.8	2.1
II	75.2	0.0	0.4	0.0	3.3	4.7	12.7	0.0	3.8	0.0

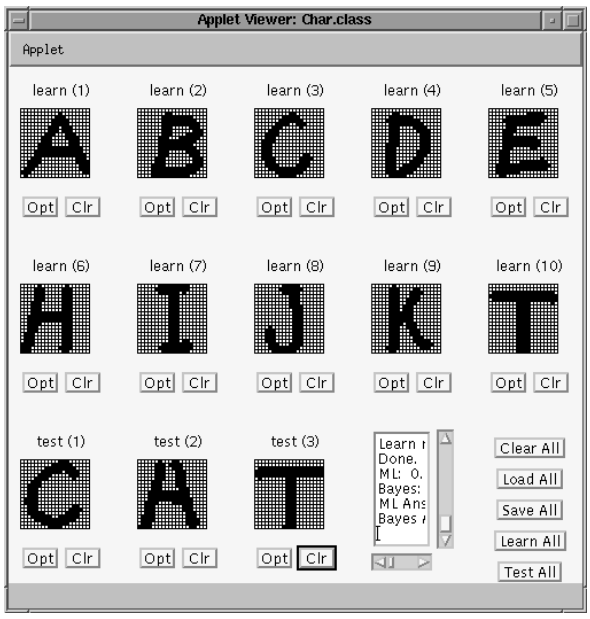

Figure 6: Input window (Upper 2 rows: learning patterns, bottom row: test patterns)

Table 1.

When C_1 ='C', C_3 ='T', highest belief is C_0 ='H' with 22.1% in I and C_0 ='A' with 75.2% in II. Thus, II can recognize 'A' correctly,but in I, 'A' is misclassified as 'H'.

In the other situation, C_1 ='T', C_3 ='E', desirable belief, C_2 ='H' is supported 22.1% in I and 84.0% in II. Thus II can work well also in these cases.

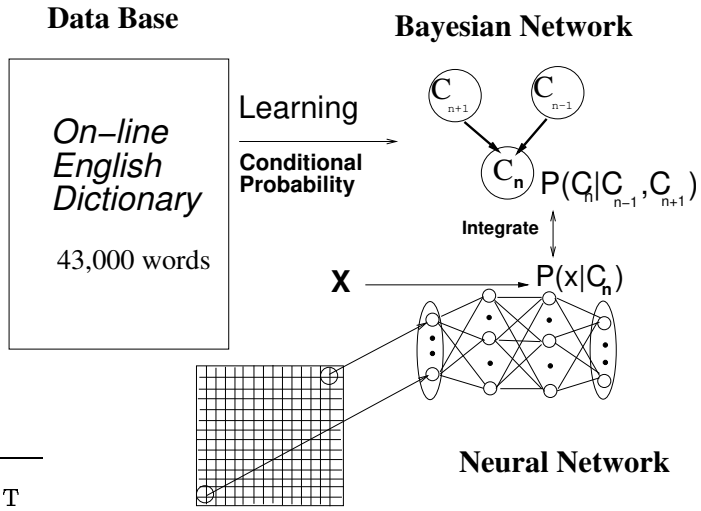

Figure 7: Integration model using Bayesian network

6. Related works

The hidden Markov Model (HMM) for this task [3] is probabilistic character recognition model, using information about neighboring patterns. The main difference is that the situated prior probability and information about classification are not distinguished in a HMM and almost all actual HMM are specialized in time series like voice or words. In the case of the integration method proposed here, the dependency between target pattern and neighboring patterns is expressed and learned independently from the classification part. The situated prior probability can be described more generally by an arbitrary Bayesian network structure. We expect that our integration model can be expanded for other kinds of object recognition tasks. For such general object recognition tasks, human use more complex prior probability obtained from many experiences in the real world environment. Then, learning situated prior probability from large samples is necessary.

7. Conclusion and future work

In this paper, the situated prior probability has been discussed in the context of recognizing ambiguous patterns. The model proposed in this paper demonstrates that our integration method based on Bayesian framework can work well for handwriting pattern recognition tasks. In this paper, a simple example is introduced to show the efficiency of the integration model clearly. However, we can expand the method to more practical applications by extending the Bayesian networks in the same framework. Thus, future work will be demonstrate that this method is a general framework which can be applied to more wide-spread applications like face recognition tasks in an office environment.

8. Acknowledgments

Special thanks are due to Dr. Toshihiro Matsui and Dr.Hideki Asoh for supporting my work and valuable discussions, and Dr.Ulrich Nemzow for helpful comments. The author wishes to express his gratitude to Dr. Nobuyuki Otsu, director of Real World Intelligent Center for affording an opportunity of this study. This work was supported in part by the Real World Computing Program.

References

[1] Pearl,J., "Probabilistic Reasoning in Intelligent Systems", Morgan Kaufmann Publishers(1988).

[2] Rumelhart,D.E., Hinton,G.E., and Williams,R.J., "Learning internal representations by error propagation", *Parallel Distributed Processing* Vol. 1, McCleland,J.L., Rumelhard,D.E. and The PDP Research group, MIT Press.(1986).

[3] Se-Chang Oh, Jin-Young Ha and Jin-H.Kim, "Context Dependent Search in Interconnected Hidden Markov Model for Unconstrained Hand Writing Recognition", *Pattern Recognition*, 28, No.11, 1693–1704(1995).

[4] Friedman,N., Geiger,D. and Goldszmidt,M., "Bayesian network Classifier", *Machine Learning*, 29, No.2/3, 131–164(1997).

[5] Uesaka,Y., Yamagishi,E. and Nozawa,M., "On the Human Being Presuposition Used in Learning", Proc. of Ann. Conference of Japan Neural Network Society'95, p.243–244, (1995).

[6] Shouno,H. and Fukushima,K., "Recognition of Cursive Handwriting with Selective Attention Model", Proc. of Ann. Conference of Japan Neural Network Society'95, p.272–273, (1995).

[7] Nishi,M., Ohzeki,K., Sakurai,N. and Omori,T., "An Architechture for Image Understanding by Symbol and Pattern Integration", Journal of Society of Instrument Control Engineers, Vol.30, No.11, pp.1368–1377, (1994).

[8] Motomura Y., Hara I., Asoh H., and Matsui, T., "Bayesian Network that Learns Conditional Probability by Neural Network", *Progress in Connectionist-based information processing systems (proc. of ICONIP'97)*, 584–587, Springer(1997).

[9] Motomura Y., "Integration model of context-specific and situated prior knowledge by conditional probability distributions", Japanese Society of Artificail Intelligence Technical Report SIG-J-9702-12, 71–76, (1997).