A Gradient BYY Harmony Learning Algorithm on Mixture of Experts for Curve Detection*

Zhiwu Lu, Qiansheng Cheng, and Jinwen Ma**

Department of Information Science, School of Mathematical Sciences
and LMAM, Peking University, Beijing, 100871, China
jwma@math.pku.edu.cn

Abstract. Curve detection is a basic problem in image processing and has been extensively studied in the literature. However, it remains a difficult problem. In this paper, we study this problem from the Bayesian Ying-Yang (BYY) learning theory via the harmony learning principle on a BYY system with the mixture of experts (ME). A gradient BYY harmony learning algorithm is proposed to detect curves (straight lines or circles) from a binary image. It is demonstrated by the simulation and image experiments that this gradient algorithm can not only detect curves against noise, but also automatically determine the number of straight lines or circles during parameter learning.

1 Introduction

Detecting curves (straight line, circle, ellipse, etc.) is one of the basic problems in image processing and computer vision. In the traditional pattern recognition literature, there are two kinds of studies on this problem. The first kind of studies use the generate-and-test paradigm to sequentially generate hypothetical model positions in the data and test the positions (e.g., [1]). However, this kind of methods are sensitive to noise in the data. The second kind of studies are Hough Transform (HT) variations (e.g.,[2]). They are less sensitive to noise, but their implementations for complex problems su er from large time and space requirements and from the detection of false positives, although the Random Hough Transform (RHT) [3] and the constrained Hough Transform [4] were proposed to improve these problems. In the field of neural networks, there have also been some proposed learning algorithms that can detect curves in an image (e.g., [5]-[6]).

Proposed in 1995 [7] and systematically developed in past years [8]-[9], Bayesian Ying-Yang (BYY) harmony learning acts as a general statistical learning framework not only for understanding several existing major learning approaches but also for tackling the learning problem with a new learning mechanism that makes model selection automatically during parameter learning [10].

* This work was supported by the Natural Science Foundation of China for Projects 60471054 and 40035010

** The corresponding author

. , . , . (.): D A 2005, C 3578, . 250 257, 2005.
○ - B 2005

Specifically, the BYY harmony learning has been already applied to detecting the best number k^* of straight lines via a selection criterion $J(k)$ on the mixture of experts (ME) in [8]. However, the process of evaluating the criterion incurs a large computational cost since we need to repeat the entire parameter learning process at a number of di erent values of k.

In this paper, we implement the BYY harmony learning on an architecture of the BYY system with the ME via a gradient harmony learning algorithm so that the curve detection can be made automatically during parameter learning on the data from a binary image, which is demonstrated by the simulation and image experiments for both straight lines and circles.

2 Gradient Harmony Learning Algorithm

A BYY system describes each observation $x \in \mathcal{X} \subset R^n$ and its corresponding inner representation $y \in \mathcal{Y} \subset R^m$ via the two types of Bayesian decomposition of the joint density $p(x,y) = p(x)p(y|x)$ and $q(x,y) = q(x|y)q(y)$, being called Yang machine and Ying machine, respectively. Given a data set of x, the aim of learning on a BYY system is to specify all the aspects of $p(y|x), p(x), q(x|y), q(y)$ with a harmony learning principle implemented by maximizing the harmony functional:

$$H(p||q) = \int p(y|x)p(x)ln[q(x|y)q(y)]dxdy - lnz_q, \quad (1)$$

where z_q is a regularization term. The details are referred to [8]-[9].

The BYY system and harmony learning can also be applied to supervised leaning tasks of mapping $x \to y$ based on a given data set $\{x_t, y_t$

where $\delta(x)$ is the δ-function. Then, we get the following alternative ME model for mapping $x \longrightarrow y$ implied in the BYY system:

$$q(y \mid x) = \sum_l q(y \mid x, l)P(l \mid x), \qquad P(l \mid x) = q(x \mid l)\alpha_l / \sum_{j=1}^{k} q(x \mid j)\alpha_j. \tag{4}$$

Letting the output of expert l be $f_l(x, \theta_l)$, we have the following expected regression equation:

$$E(y \mid x) = \int yq(y \mid x)dy = \sum_l f_l(x, \theta_l)P(l \mid x). \tag{5}$$

That is, $E(y \mid x)$ is a sum of the experts weighted by the gate functions $P(l \mid x)$, respectively.

We now ignore the normalization term (i.e., set $z_q = 1$), substitute these components into Eq.(1), and have

$$H(p\|q) = \frac{1}{N}\sum_{t=1}^{N}\sum_{l=1}^{k} \frac{q(y_t \mid x_t, l)q(x_t \mid l)\alpha_l}{\sum_{j=1}^{k} q(y_t \mid x_t, j)q(x_t \mid j)\alpha_j} \ln(q(y_t \mid x_t, l)q(x_t \mid l)\alpha_l), \tag{6}$$

where

$$q(y \mid x, l) = \frac{1}{\sqrt{2\pi}\tau_l} e^{-\frac{(y - w_l^T x - b_l)^2}{2\tau_l^2}}, \qquad q(x \mid l) = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma_l^n} e^{-\frac{\|x - m_l\|^2}{2\sigma_l^2}},$$

$$\sigma_l = e^{d_l}, \tau_l = e^{r_l}, \quad \alpha_l = e^{\beta_l} / \sum_{j=1}^{k} e^{\beta_j}.$$

By the derivatives of $H(p\|q)$ with respect to the parameters w_l, b_l, r_l, m_l, d_l and β_l, respectively, we have the following gradient learning algorithm:

$$\Delta w_l = \frac{\eta}{N}\sum_{t=1}^{N} U(l \mid x_t, y_t)\frac{(y_t - w_l^T x_t - b_l)}{e^{2r_l}} x_t, \tag{7}$$

$$\Delta b_l = \frac{\eta}{N}\sum_{t=1}^{N} U(l \mid x_t, y_t)\frac{(y_t - w_l^T x_t - b_l)}{e^{2r_l}}, \tag{8}$$

$$\Delta r_l = \frac{\eta}{N}\sum_{t=1}^{N} U(l \mid x_t, y_t)\frac{(y_t - w_l^T x_t - b_l)^2 - e^{2r_l}}{e^{2r_l}}, \tag{9}$$

$$\Delta m_l = \frac{\eta}{N}\sum_{t=1}^{N} U(l \mid x_t, y_t)\frac{(x_t - m_l)}{e^{2d_l}}, \tag{10}$$

$$\Delta d_l = \frac{\eta}{N}\sum_{t=1}^{N} U(l \mid x_t, y_t)\frac{(x_t - m_l)^2 - ne^{2d_l}}{e^{2d_l}}, \tag{11}$$

$$\Delta \beta_l = \frac{\eta}{N}\sum_{t=1}^{N}\sum_{j=1}^{k} U(j \mid x_t, y_t)(\delta_{jl} - \alpha_l), \tag{12}$$

where

$$U(l \mid x_t, y_t) = P(l \mid x_t, y_t)(1 + \sum_{j=1}^{k}(\delta_{jl} - P(j \mid x_t, y_t)) \ln(q(y_t \mid x_t, j)q(x_t \mid j)\alpha_j),$$

δ_{jl} is the Kronecker function, and η is the learning rate which is usually a small positive constant.

The above gradient BYY harmony learning algorithm is designed for straight line detection. Here, a set of black points $\{x_t\}_{t=1}^{N}$ ($x_t = [x_{1t}, x_{2t}]^T$) are collected from a binary image with each point being denoted by its coordinates $[x_1, x_2]$. Suppose that $w_l^T x + b_l = 0, l = 1, \cdots, k$ are the parametric equations of all the straight lines to be detected in the image. For each point x, if $w_l^T x + b_l = 0$, we let $L(x) = l$. Then, the mapping between x and y implemented by the BYY system is just $y = w_{L(x)}^T x + b_{L(x)}$. For each point in $\{x_t\}_{t=1}^{N}$, it is supposed to be on some straight line (at most disturbed by some noise) and we always set $y_t = 0$. We train the ME model implied in the BYY system on the sample set $\{x_t, y_t\}_{t=1}^{N}$ via this gradient BYY harmony learning algorithm and lead to the result that each expert will finally fit a straight line $w_l^T x + b_l = 0$ with the mixing proportion α_l representing the proportion of the number of points on this straight line over N, i.e., the number of all the black points in the image.

As for circle detection, we can use $f_l(x, \theta_l) = (x - c_l)^T (x - c_l) - R_l^2,\ R_l = e^{bl}$ instead of $f_l = w_l x + b_l$ in the above model and derivations for the output of each expert in the ME model. Hence, the gradient BYY harmony learning algorithm is modified by replacing the first three learning rules Eqs (7)-(9) with the following ones:

$$\Delta c_l = -2\frac{\eta}{N}\sum_{t=1}^{N} U(l \mid x_t, y_t)\frac{(y_t - (x_t - c_l)^T (x_t - c_l) + R_l^2)}{e^{2r_l}}(x_t - c_l), \tag{13}$$

$$\Delta b_l = -2\frac{\eta}{N}\sum_{t=1}^{N} U(l \mid x_t, y_t)\frac{(y_t - (x_t - c_l)^T (x_t - c_l) + R_l^2)}{e^{2r_l}} e^{2b_l}, \tag{14}$$

$$\Delta r_l = \frac{\eta}{N}\sum_{t=1}^{N} U(l \mid x_t, y_t)\frac{(y_t - (x_t - c_l)^T (x_t - c_l) + R_l^2)^2 - e^{2r_l}}{e^{2r_l}}. \tag{15}$$

3 Experimental Results

In this section, several experiments are carried out for both straight line and circle detection with the gradient BYY harmony learning algorithm. On the one hand, we make some simulation experiments to demonstrate that the algorithm can detect the straight lines or the circles automatically. On the other hand, we apply the algorithm to the strip line detection and the container recognition.

3.1 Automated Detection on the Straight Lines and Circles

Using k^* to denote the true number of curves in the original image, we implemented the gradient algorithm on data sets from binary images always with

$k \geq k^*$ and η =0.1. Here, k is the number of experts in the ME model. Moreover, the other parameters were initialized randomly within certain intervals. In all the experiments, the learning was stopped when $|\Delta H| < 10^{-6}$.

During the BYY harmony learning process, some mixing proportions of the experts can be reduced to a very small number. In this case, $(-\ln \alpha_l)$ will become very large. However, as shown in the mathematical expressions of the gradient algorithm, it is always regulated by α_l so that $\alpha_l \ln \alpha_l$ will tend to zero. Therefore, the gradient algorithm will always converge to a reasonable solution and cannot diverge to infinity.

The experimental results on the straight line and circle detections are given in Fig.1 (a), (b), respectively, with the parameters listed in Table 1, 2, respectively. From Fig.1(a) and Table 1, we find that the four straight lines in the binary image are successfully detected, with the mixing proportions of the other four straight lines reduced below 0.001, i.e., these straight lines are extra and should be discarded. That is, the correct number of straight lines have been detected from the image. Likewise, from Fig.1(b) and Table 2, we find that the two circles are successfully detected, while the mixing proportions of the other two extra circles become less than 0.001.

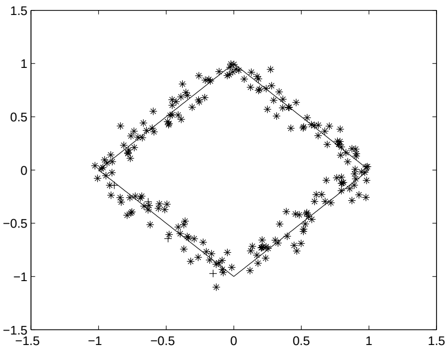
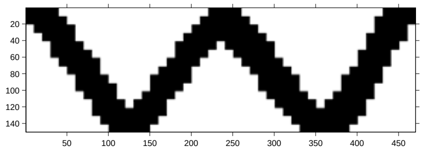
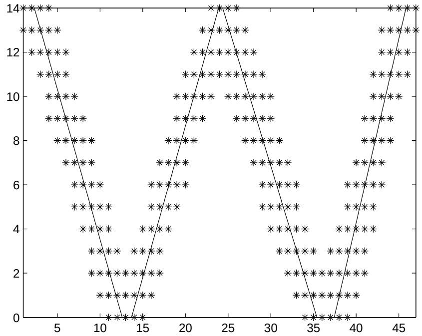


Table 1. The empirical result of the straight line detection on the data set from Figure 1(a), with k=8 and k*=4

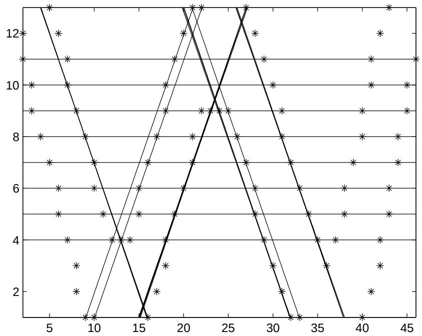
l	α_l	$w_{1l}x_1 + w_{2l}x_2 + b_l = 0$
1	0.0008	$1.2070x_1 - 0.7370x_2 - 0.2235 = 0$
2	0.0007	$-1.0329x_1 + 0.9660x_2 - 0.0455 = 0$
3	0.2319	$-0.9778x_1 - 1.0217x_2 - 0.9814 = 0$
4	0.0009	$-0.8693x_1 + 1.1155x_2 - 0.2002 = 0$
5	0.2369	$1.0181x_1 - 0.9816x_2 + 1.$



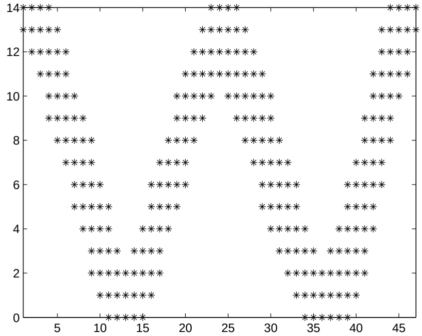
(a)

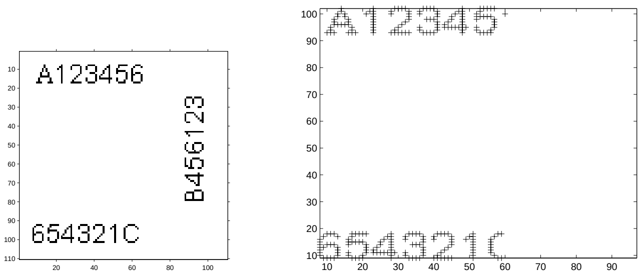


(b)



(c)





(a)