A Supervised Classification Scheme Using Positive Boolean Function*

Chin-Chuan Han
Department of Computer Science and Information Engineering,
Chung-Hua University, Hsinchu, Taiwan, R.O.C.
E-mail: cchan@chu.edu.tw

Abstract

In this paper, a classifier based on the positive Boolean function (PBF) is proposed for the supervised pattern classification. A PBF is exactly represented as one sum-of-product form without any negative components. The PBF possesses the well-known threshold decomposition and stacking properties. The classification errors can be calculated from the summation of the absolute errors incurred at each level. The optimal PBF are found and designed to be a classifier by minimizing the classification error rate along the training samples. The experimental results were given to show the validity of our proposed approaches.

Keywords: pattern recognition, supervised classification, positive Boolean function, stack filter, minimum classification error

1 Introduction

Positive Boolean Function (PBF) has been successfully applied on the design of stack filter. Each stack filter corresponding to a PBF possesses the weak superposition property known as the threshold decomposition and the ordering property known as the stacking property[1]. The class of stack filters includes many familiar filter types, such as median filter, weighted median filter, order statistic filter, weighted order median filter, and so on. The main function of stack filter is to remove the noises, detect the edges, ..., etc. All their applications focus on the image and signal processing. In this paper, we attempt to apply the positive Boolean function to the supervised classification scheme in pattern recognition.

In 1986, Wendt et al.[1] defined the stack filter as the class of all filters. They also pointed out the connection between stack filters and positive Boolean functions. Maragos[2] used mathematical morphology to represent the class of morphological filters and their equivalent classes. He also showed that any positive Boolean function has exactly one sum-of-product form and one product-of-sum form by morphological notation. Besides, Dougherty[3] defined a basis class of filters satisfying the morphological basis criteria and applied the erosion operation as an estimator to find the optimal morphological filter by minimizing the mean square error. Lin and Coyle[4] developed the generalized stack filter, including all rank order filters, stack filters, and digital morphological filters.

Postaire[5] et al. proposed a binary morphological technique to cluster the $N$-dimensional observations. It is an un-supervised pattern clustering approach by using the mathematical morphological operations. In the traditional supervised pattern classification, the classification problem was modelled as a distribution estimation problem based on the Bayes decision theory. Juang et al.[6] discussed the issue of training process from a perspective with root in the classical Bayes decision theory in speech recognition. They summarized the different viewpoints of the classification issues associated with the classical Bayes approaches. First, it is a difficult task to determine the right parametric form of the sample distribution. Besides, a good parameter estimation method should be necessary. Second, the training set must be of sufficient size. They have shown the superiority of the minimum classification error (MCE) method over the distribution estimation method. In this paper, we will apply the PBF to do the supervised classification tasks based on the MCE criterion for pattern classification.

An optimal stack filter $S_f$ is defined as a filter whose mean absolute error (MAE) value between the filter's output and the desired signals is minimum. At the supervised training stage, the identities of the training samples are previously defined and identified. These identities of samples can be defined to be the desired

*This work is supported by National Science Council of Taiwan under grant no. NSC89-2213-E-216-019.
values and the filter’s outputs are considered to be the classification results. Since the stack filter possesses the well-known threshold decomposition and stacking properties, the multi-level MAE value can be decomposed into the summation of binary values incurring at each level. At each level, if the desired value of a sample is identical with the classification result, then no error occurs. Otherwise, one error should be counted. Therefore, the best classifier can be generated from the optimal PBF with the minimal classification error.

2 Review of Stack Filter

Let $X$ be the signal processed via a stack filter $S_f(\cdot)$, where $X$ takes the values in $Q = \{0, 1, \ldots, L - 1\}$. A window $X_n$ of length $n$ slides through the input signal $X$ from left to right at the position $j$ and is defined as $X_n = (x_1, x_2, \ldots, x_n)$. Based on the mean absolute error (MAE) criteria, the optimal stack filter $S_f$ is defined as a filter whose mean absolute error between the filter’s output $S_f(X_n(j))$ and the desired signal $S(j)$ at the point $(j)$ is minimum. Assume the signals $S(j)$ and $X_n(j)$ are jointly stationary, the estimation error of a specified stack filter $S_f$ is defined as $E[|S(j) - S_f(X_n(j))|]$.

The signal $X_n$ of length $n$ can be decomposed into $L - 1$ binary signals $X^l_n = (x^l_1, x^l_2, \ldots, x^l_n) \in \{0, 1\}^n$ by using the threshold function as follows.

$$x^l_i = T_l(x_i) = \begin{cases} 1 & \text{if } x_i \geq l, \\ 0 & \text{else.} \end{cases} \quad (1)$$

Based on the threshold decomposition property, the output value of stack filter $S_f(X)$ can be obtained from the summation of the binary values generated from the positive Boolean function $B_f$ in the following formula:

$$S_f(X_n) = S_f \left( \bigoplus_{l=1}^{L-1} X^l_n \right) = \bigoplus_{l=1}^{L-1} B_f(X^l_n). \quad (2)$$

Wendt et al.[1] have shown that the necessary and sufficient condition for a Boolean function to possess the stacking property is that it be a positive Boolean function. An $n$-input Boolean function $f(\cdot)$ is said to possess the stacking property if

$$f(x^l_1, x^l_2, \ldots, x^l_n) \geq f(y^l_1, y^l_2, \ldots, y^l_n) \quad (3)$$

when $x_i^l \geq y_i^l$ for all $i$, and $l = 1, \ldots, L - 1$.

Since the stack filter possesses the well-known threshold decomposition and stacking properties, the multi-level mean absolute error can be decomposed into the summation of the absolute errors incurred at each level and represented as follows.

$$MAE = E \left[ \left| S(j) - S_f(X_n(j)) \right| \right] = E \left[ \sum_{l=1}^{L-1} s_l(j) - f(T_l(X_n(j))) \right]$$

(by threshold decomposition property)

$$= E \left[ \sum_{l=1}^{L-1} \left| s_l(j) - f(T_l(X_n(j))) \right| \right] = \sum_{l=1}^{L-1} E \left[ \left| s_l(j) - f(T_l(X_n(j))) \right| \right], \quad (4)$$

where the threshold function $T_l(\cdot)$ and the Boolean function $f(\cdot)$ are used at each level, and $s_l(j) = T_l(S(j))$.

Let us demonstrate an example illustrated by Coyle[1] to show the complete filtering process. Example 1: Consider the filter implementation process as shown in Fig. 1, the stack filter $S_f(\cdot)$ (e.g., a median filter) of window’s length three is applied to the input signals $(2, 2, 0, 1, 1, 3, 3, 3, 2)$. The output signals $(2, 2, 1, 1, 3, 3, 3, 2)$ will be generated by this stack filter $S_f(\cdot)$. On the other hand, the binary filter $B_f(\cdot)$ of window’s length three based on the positive Boolean function $f(x_1, x_2, x_3) = x_1 x_2 + x_2 x_3 + x_3 x_1$ slides through the input signals from left to right at each threshold level. The errors only occur at level one during the binary filtering process. Thus, the total MAE value generated by the binary filter $B_f(\cdot)$ as shown at the right hand side of Fig. 1 is also computed by summing the errors occurring at each level. Furthermore, Coyle et al.[7] reformulated the MAE value as the minimal sum over the $2^n$ possible binary vector which are observed from the window of length $n$. That is

$$\min \sum_{j=0}^{2^n-1} \left[ \alpha_j p(0|x_j) + \beta_j p(1|x_j) \right], \quad (5)$$

where $\alpha_j, \beta_j$ are the constant cost coefficients in an observation window of length $n$. $p(0|x_j)$ and $p(1|x_j)$ are the probabilities that the filter outputs a 0 or 1 when it observes a binary vector $x_j$ in its window. The output signals can be obtained by summing the binary filtering results at each threshold level.

3 Pattern Recognition Using Positive Boolean Functions

In the supervised classification scheme, the classification function is defined to be a PBF that satisfies
the threshold decomposition and the stacking properties. According to these properties, the classification errors of the training set are decomposed into the sum of the absolute errors incurred at each level.

Consider \( N \) classes and \( M \) samples in each class. These \( MN \) training samples of length \( n \) can be represented as the feature vectors \( x_{ij} \), where \( i = 1, \ldots, N \), and \( j = 1, \ldots, M \). The center of a specified class \( w_i \) can be obtained by averaging the \( M \) samples in class \( w_i \), i.e., \( \bar{w}_i = \frac{1}{M} \sum_{j=1}^{M} x_{ij} \). The distance between a sample \( x_{ij} \) to a specified class center \( \bar{w}_k \) can be calculated and represented as a vector form \( z_{ij,k} = x_{ij} - \bar{w}_k \), \( k = 1, \ldots, N \). Thus, there are \( MN^2 \) distance vectors to be the occurrences in finding the optimal PBF.

Since the filtering of stack filter is an integer-based operation, the distance vectors \( z_{ij,k} \) should be quantized into \( L \) levels. Based on the threshold decomposition property, an occurrence of vector form \( z_{ij,k} \) can be decomposed into \( L - 1 \) binary vectors \( z_{ij,k} = (z^+_1, z^+_2, \ldots, z^+_L, z^-_1, \ldots, z^-_L) \), and \( l = 1, \ldots, L - 1 \). Next, the desired value for each occurrence at each level has to be determined. The desired value of an occurrence \( z_{ij,k} \) can be considered as the error value of sample \( x_{ij} \) and class \( w_k \). If sample \( x_{ij} \) belongs to class \( w_k \), no error occurs, i.e., \( d(z_{ij,k}) = 0 \). Otherwise, if sample \( x_{ij} \) does not belong to class \( w_k \), assign the desired value of occurrence \( z_{ij,k} \) to be one, i.e., \( d(z_{ij,k}) = 1 \). Thus, the desired value for each occurrence is defined as

\[
d(x_{ij} - \bar{w}_k) = d(z_{ij,k}) = \begin{cases} 0 & \text{if } x_{ij} \in w_k, \text{ i.e., } i = k \\ 1 & \text{if } x_{ij} \notin w_k, \text{ i.e., } i \neq k \end{cases}
\]

Next, we will show the classification problem satisfies the stacking property. Consider two samples \( x \) and \( y \), and a specified class \( w_k \) of center being \( \bar{w}_k \). The binary vectors for these two occurrences \( x - \bar{w}_k \) and \( y - \bar{w}_k \) thresholded at level \( l \) are \( u^l \) and \( v^l \), respectively. We can conclude the relationship of samples \( x \) and \( y \) as follows. If \( u^l \leq v^l \) for all dimensional elements and \( l = 1, 2, \ldots, L - 1 \), sample \( x \) is nearer to the center of class \( w_k \) than sample \( y \). There are two relations along the center \( w_k \) and two samples \( x \) and \( y \).

**C1:** If sample \( x \) does not belong to class \( w_k \) \( (d(x - \bar{w}_k) = 1) \), then sample \( y \) does not belong to class \( w_k \) \( (d(y - \bar{w}_k) = 1) \).

**C2:** If sample \( y \) is an element of class \( w_k \) \( (d(y - \bar{w}_k) = 0) \), then sample \( x \) should be an element of class \( w_k \) \( (d(x - \bar{w}_k) = 0) \).

Next, we define the classification function for samples \( x \) and \( y \) to be a positive Boolean function (PBF) at level \( l \). The classification results for samples \( x \) and \( y \) will be generated as follows. Consider \( u^l \leq v^l \) for two samples \( x \) and \( y \), \( l = 1, 2, \ldots, L - 1 \).

**CE** = \( E \left[ |d(z_{ij,k}) - S_f(z_{ij,k})| \right] \)

\[
= \left[ \sum_{l=1}^{L-1} |d(z_{ij,k}) - f(T_l(z_{ij,k}))| \right]
\]

Furthermore, we can reformulated the CE value as the minimal sum over the \( 2^n \) possible binary vectors of length \( n \) as same as EQ. (5). In EQ. (5), the terms \( \alpha_j p(0|x_j) \) and \( \beta_j p(1|x_j) \) are referred as the type II and type I errors, respectively. The best classifier can be generated from the optimal PBF with the minimal classification errors occurring at all levels. As we know, it is a time-consuming procedure in finding the optimal stack filter. Many researchers[7, 8, 9] have proposed the efficient approaches to find the optimal PBF. In
our previous works[9], the graphic search-based algorithm is further improved the searching process by utilizing the greedy and constraint satisfaction searching algorithms.

4 Experimental Results

In this section, we have conducted an experiment to show the validity of our proposed classification scheme. The well-known database “IRIS” created by Fisher was used in many pattern recognition literatures. The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. The first 25 samples of each class are used to be the training samples, and the other samples(75 instances) are used to evaluate the performance of our PBF-based classifier. The recognition rate is 96% by using the PBF-based classifier.

5 Conclusions

A PBF-based classifier has been proposed for the supervised pattern classification in this paper. The classifier based on the optimal PBF can be found by using our fast searching algorithm under the minimum classification error(MCE) criterion. The experimental results were conducted to show the validity of PBF-based classifier.

References


