Technical University of Denmark

DTU

Quantized, piecewise linear filter network

Sørensen, John Aasted

DTU Library
Technical Information Center of Denmark

Quantized, Piecewise Linear Filter Network.

John Aasted Sørensen, Electronics Institute, Build 349
Technical University of Denmark, Lyngby
DK-2800 Denmark
email: jaas@dthei.ei.dth.dk; fax: +4542880117

Abstract A quantization based piecewise linear filter network is defined. For stationary signals this filter network is a generalization of the classical Wiener filter, with an input signal x_n and a desired response d_n, $n = 1, 2, \ldots$. The object of the piecewise linear filter network, is the approximation of a possible nonlinear function $G(x)$ between x_n and d_n. A method for the training of this network based on local approximation in the input space is devised. The training is carried out by repeatedly alternating between vector quantization of the training set into quantization classes and equalization of the quantization classes linear filter mean square training errors. The equalization of the mean square training errors is carried out by adapting the boundaries between neighbor quantization classes such that the differences in mean square training errors are reduced.

Introduction

A quantization based, piecewise linear filter network is defined. For stationary signals this filter network is a generalization of the classical Wiener filter [1], with an input signal x_n and a desired response d_n, $n = 1, 2, \ldots$.
The object of the piecewise linear filter network [2] is the approximation of a possible nonlinear function $G(x)$ between x_n and d_n, as shown in Fig. 1. The training of the filter network is carried out by repeatedly alternating between vector quantization of the training set $\{x_n, d_n, n = 1, \ldots, N\}$ into quantization classes Q_i and equalization of the quantization classes linear filters mean square training errors J_i, by adapting the boundaries between neighbor quantization classes, such that the differences in mean square errors is reduced.

The optimization criterium used in the training of the filter network is

$$J = \sum_i J_i + \frac{1}{2} \sum_{(i,j)} |J_i - J_j|$$

$i : quantization\ classes$

$(i,j) : neighbor\ quantization\ classes$

The first sum in J reduces the total training error of the linear filters $\mathbf{w}_i$, associated the quantization classes and the second sum equalize the training errors by controlling the boundaries between neighbor quantization classes. Piecewise linear filters are given in [3], [4] and [5]. In [3] the local linear filters are constructed for each partition resulting from a vector quantization. In [4] a piecewise linear filter is devised, based on the "canonical piecewise-linear (PWL) structure" and [3] is based on a scalar quantization of the filter outputs, where the filters are organized in a tree structure.

The Constituent Elements of the Training Algorithm

$\{x_n, d_n, n = 1, \ldots, N\}$: Training set.
M : Dimension of input space.
$\mathbf{x}_n' = (x_n, x_{n-1}, \ldots, x_{n-M+1})$ where $'$ denotes transposed.
$\mathbf{X} = [\mathbf{x}_n'] \in \mathbf{R}^{(N-M+1) \times M}$
$\mathbf{x}d_n' = [\mathbf{x}_n', d_n]$
$\mathbf{X}d = [\mathbf{x}d_n'] \in \mathbf{R}^{(N-M+1) \times (M+1)}$
$dist(\mathbf{x}d_m, \mathbf{x}d_n)$: Distance between $\mathbf{x}d_m$ and $\mathbf{x}d_n$.
s : The number of quantization iterations.
k : The number of equalization iterations at a given number of quantizations.
Q_s: The number of quantization classes after s quantization iterations.
$\mathcal{Q}_i = \{\mathbf{x}d_{ij} \in$ quantization class no. $i\}$, $j = 1, \ldots, q_i$, $i = 1, \ldots, Q_s$
q_i: The number of training elements in quantization class no. i.
$\mathbf{J}^k = [J_i^k], i = 1, \ldots, Q_s$: Mean square error of the linear filter in quantization class no. i at the equalization iteration no. k
$\mathbf{W}^k = [\mathbf{w}_j^k] \in \mathbf{R}^{(M+1) \times Q_s}$: linear filters of the quantization classes.
$\mathbf{C}^k = [\mathbf{c}_j^k] \in \mathbf{R}^{(M+1) \times Q_s}$: Mean values for quantization classes.
$\mathbf{Q}_k = [\mathbf{q}_j^k] \in \{0,1\}^{(N-M+1) \times Q_s}$: Partition matrix for $\mathbf{X}d$.
$\mathbf{q}_j^k = [q_{ij}^k]$ where $q_{ij}^k = 1$ if $\mathbf{x}d_i \in$ quantization class no. j. Otherwise $q_{ij}^k = 0$.
$\mathbf{P}^k = [p_{ij}^k] \in \{0,1\}^{Q_s \times Q_s}$: Neighbor matrix for quantization classes.
$p_{ij}^k = 1$ if the quantization classes i and j are neighbors. Otherwise $p_{ij}^k = 0$.

471

Training Algorithm Overview.

The training algorithm shown in the next section, is based on two nested operations carried out iteratively. The first operation is *vector quantization* of the training set, represented by the matrix $\mathbf{X}d$. The result of this is the mean values of quantization classes $\mathbf{C}^0$, the training set partition matrix $\mathbf{Q}^0$, the neighbor matrix $\mathbf{P}^0$ of the quantization classes and the mean square training errors $\mathbf{J}^0$ of the quantization classes.

In the second operation, denoted equalization, the quantization classes are adapted through the operation *adapt quantization classes* according to the current training errors of the classes and their neighbor classes training errors. The two adaptation rules considered here are:

Rule 1: Adapt the current quantization class in the direction of the largest training error gradient of the neighbors to the current quantization class.

$$\mathbf{c}_i^k = \mathbf{c}_i^{k-1} + \mu_{ij} \max_j \left\{ \frac{h(J_j^{k-1} - J_i^{k-1})}{J_j^{k-1}} p_{ij}^{k-1} \right\} (\mathbf{c}_j^{k-1} - \mathbf{c}_i^{k-1}) \qquad (1)$$

where: $i,j = 1, \ldots, Q_s$, $h(x) = 0.5(|x| + x)$, p_{ij}^{k-1} : neighbor matrix element and μ_{ij} is a constant.

Rule 2: Adapt the current quantization class in the direction of the mean training error gradient to the neighbors of the current quantization class.

$$\mathbf{c}_i^k = \mathbf{c}_i^{k-1} + \frac{1}{\eta_i} \sum_{j=1}^{Q_s} \mu_{ij} \frac{J_j^{k-1} - J_i^{k-1}}{J_j^{k-1}} p_{ij} (\mathbf{c}_j^{k-1} - \mathbf{c}_i^{k-1}) \qquad (2)$$

$$i,j = 1, \ldots, Q_s$$

η_i is the number of neighbor quantization classes of class i.
Both rules leads to the transformation:

$$\mathbf{C}^k = \mathbf{C}^{k-1} \mathbf{\Gamma}^{k-1} \quad \text{where} \quad \mathbf{\Gamma}^{k-1} = [\gamma_{ij}] \in \mathbf{R}^{Q_s \times Q_s}.$$

Rule 1 leads to these properties of $\mathbf{\Gamma}$:

- $0 \leq \gamma_{ij} \leq 1$

- $\gamma_{ij} \gamma_{ji} = 0$.

- If $\gamma_{ii} = 1$ then the quantization class no. i has a local (global) maximum in the training error, compared to its neighbors. At least one γ_{ii} will be equal to 1.

- $\mathbf{\Gamma}$ is a column stochastic matrix. (Each column sums to 1).

472

Rule 2 leads to these properties of $\boldsymbol{\Gamma}$:

- $\gamma_{ij} < 1$

- If $sign(\gamma_{ij}) = -1$ for $i = 1, \ldots, Q_s, i \neq j$ then the quantization class j is a local (global) maximum.

- If $sign(\gamma_{ij}) = 1$ for $i = 1, \ldots, Q_s, i \neq j$ then the quantization class j is a local (global) minimum.

- $\boldsymbol{\Gamma}$ is a column stochastic matrix.

The adaptation is followed by a minimum distance classification of the training set using $\mathbf{C}^k$ and finally determination of the local filter matrix $\mathbf{W}_k$. The results of the training algorithm are the quantization class means $\mathbf{C}^k$ defining the quantization classes and the local linear filters $\mathbf{W}_k$.

The Training Algorithm.

Initialize:
$s = 0$: The number of quantization iterations.
$k = 0$: The number of equalization iterations.
$J^{-1} = \infty, J^0 = 0$.
$\boldsymbol{\Lambda}^0$: Identity partition for the training set.
$\boldsymbol{\Theta}^0$: Mean value of the training set.
while $|J^{s-1} - J^s|/J^{s-1} > Quantize_{thres.}$
    $s = s + 1$
    $(\mathbf{C}^0, \mathbf{Q}^0, \mathbf{P}^0, \mathbf{J}^0) = $ vector quantization $(\boldsymbol{\Lambda}^k, \boldsymbol{\Theta}^k, \mathbf{X}d)$
    $k = 0$
    while $\max_{(i,j)} |J_j^k - J_i^k| p_{ij}^k / J_j^k > Equalize_{thres.}$
        $k = k + 1$
        $(\mathbf{C}^k) = $ adapt quantization classes$(\mathbf{C}^{k-1}, \mathbf{Q}^{k-1}, \mathbf{J}^{k-1})$
        $(\mathbf{Q}^k, \mathbf{P}^k) = $ classify training set$(\mathbf{X}d, \mathbf{C}^k)$
        $(\mathbf{W}^k, \mathbf{J}^k) = $ quantization class filters$(\mathbf{Q}^k, \mathbf{X}d)$
        $J^k = \sum_{i=1}^{Q_s} J_i^k + 0.5 \sum_{i=1}^{Q_s} \sum_{j=1}^{Q_s} |J_i^k - J_j^k| p_{ij}^k$
    endwhile equalization
    $J^s = J^k$
    $\boldsymbol{\Theta}^k = \mathbf{C}^k$
    $\boldsymbol{\Lambda}^k = \mathbf{Q}^k$
endwhile quantization

Result:
    $\mathbf{W}_s^k$: Quantization class filters.
    $\mathbf{C}_k^s$: Quantization class means.

473

A Training Example

A training set is defined as follows:
x_n: White, Gaussian noise, $\sigma_x^2 = 1$.
Prefiltering: $y_n = 0.5y_{n-1} + x_n$.
Scalar, 2-slope nonlinearity: $d_n = f(y_n)$ where:

$$f(x) = \begin{cases} -f(-x) & \\ x & for \quad 0 \le x \le 0.5 \\ 0.666 - 0.333x & - \quad 0.5 \le x \end{cases}$$

A training set with $N = 200$, $wdim = 2$, $\mu_{ij} = 1.0$ leads to the following ratio between the mean square training error of the linear system and a quantized system with 4 quantization classes: 1.6.

References

1. S. Haykin, "Adaptive Filter Theory", 2. Ed, Prentice-Hall, 1991.

2. J.Aa. Sørensen, "A Family of Quantization Based Piecewise Linear Filter Networks". IEEE ICASSP-92, March 23-26., San Francisco, USA.

3. Todd K. Leen, Nandakishore Kambhatla, "Fast Non-Linear Dimension Reduction." Neural Networks for Computing Conference. Snowbird, Utah, USA, April 13-16. 1993.

4. Ji-Nan Lin, Rolf Unbehauen, "Adaptive Nonlinear Digital Filter with Canonical Piecewise-Linear Structure". IEEE Trans. CAS-37, March 1990.

5. Saul B. Gelfand, C.S. Ravishankar, Edward J. Delp, "A Tree-Structured Piecewise Linear Adaptive Filter". IEEE ICASSP-91 pp.2141-2144.
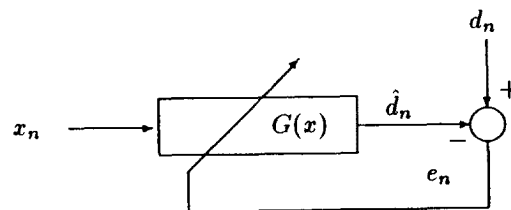
Figure 1: Adaptive filter modelling a nonlinear function $G(x)$

474