IGUALACIÓN MEDIANTE CLASIFICADORES LINEALES Y NO LINEALES DE MAXIMO MARGEN

Rafael G. Ayestarán, Ignacio Santamaría, Carlos Pantaleón *José C. Príncipe*

Departamento de Ingeniería de Comunicaciones
Universidad de Cantabria
{rafa,nacho,carlos}@gtas.dicom.unican.es

Computational NeuroEngineering Lab.
University of Florida
principe@cnel.ufl.edu

ABSTRACT

In this paper we propose the use of maximum margin classifiers in order to minimize the bit error rate in transversal equalizers. We consider both linear ("Optimal Hyperplane" OH), and non-linear classifiers ("Support Vector Machine" SVM), as an alternative to FIR-MMSE and Bayesian equalizers, respectively. Training is performed through a simple and fast adaptive algorithm called Adatron. In this way we avoid the high computational cost of quadratic programming techniques, usually employed to obtain maximum margin classifiers. Some simulation examples show the advantages of the proposed equalizers: better performance in comparison to the FIR-MMSE and simpler structure in comparison to the Bayesian equalizer.

1. INTRODUCCIÓN

La interferencia entre símbolos (ISI) causada por el canal en sistemas de comunicaciones digitales es compensada habitualmente mediante un igualador en el receptor. Las estructuras tradicionales (transversal, con decisiones realimentadas, etc.), emplean el criterio de Mínimo Error Cuadrático Medio (MMSE) para obtener los coeficientes del igualador. Sin embargo, es sabido que el diseño MMSE no conduce a la solución de mínima BER.

Teniendo en cuenta que la igualación es, en definitiva, un problema de clasificación, en esta comunicación diseñamos igualadores transversales como clasificadores de máximo margen aplicando el principio SRM ("Structural Risk Minimization") [1]. El clasificador de máximo margen puede obtenerse en el espacio de entrada, proporcionando un clasificador lineal que llamamos OH ("Optimal Hyperplane"); o en el espacio de características (de dimensión mayor que el espacio de entrada), implementando un clasificador no lineal que denotamos SVM ("Support Vector Machine").

Recientemente algunos autores han comenzado a estudiar este nuevo tipo de igualadores. En [2], por ejemplo, se entrena la parte "feedforward" de un igualador realimentado por decisión (DFE), con este nuevo criterio. El objetivo fundamental de esta comunicación es extender dicho estudio al caso de un igualador transversal. Aspectos novedosos de este estudio son: el empleo de un eficiente algoritmo de entrenamiento conocido como algoritmo Adatrón [3], y el empleo de clasificadores lineales regularizados ("soft margin") como forma de incorporar información *a priori* sobre la varianza de ruido.

Este trabajo ha sido parcialmente financiado con el proyecto FEDER 1FD97-1066-C02-01

2. MÁQUINAS DE VECTORES SOPORTE

2.1. Planteamiento del problema

La señal recibida a la entrada del igualador puede expresarse como

$$y(k) = \sum_{i=0}^{n_c} h_i s(k-i) + e(k) \tag{1}$$

donde la secuencia $s(k)$ es una secuencia binaria equiprobable {+1,-1}, h_i son los coeficientes del canal, y $e(k)$ representa el ruido aditivo, que se asume gaussiano, de media cero y varianza s_n^2.

Un igualador transversal lineal estima el valor del símbolo transmitido como una combinación lineal de las observaciones del canal. Su solución MMSE viene dada por el filtro de Wiener. No obstante, el igualador óptimo Bayesiano define una frontera no lineal de decisión que puede implementarse mediante una Red de Funciones de Base Radial (RBF)

2.2. Igualadores basados en OH y SVM

El problema de igualación puede reformularse como un problema de clasificación de la siguiente manera: obtener la frontera de separación óptima (lineal o no lineal) a partir del conjunto de entrenamiento $I=\{(\mathbf{r}_i,s_i),\ i=1,...,N\}$, donde $\mathbf{r}_i$ son los estados del canal y $s_i \in \{+1,-1\}$ son los símbolos deseados.

Si los estados del canal son separables linealmente, una aproximación razonable al clasificador de mínima BER es construir el hiperplano que maximiza la distancia entre los vectores más cercanos al mismo (el margen). En [4] se demuestra que dicho clasificador puede obtenerse minimizando

$$J(\boldsymbol{w}) = \frac{1}{2}\|\boldsymbol{w}\|_2^2 \tag{2}$$

sujeto a las restricciones $s_i(\mathbf{r}_i\boldsymbol{w}+b)\leq 1$, $i=1,...,N$. En el espacio dual el problema puede resolverse maximizando la expresión matricial

$$W(\mathbf{A}) = \mathbf{A}^T\mathbf{1} - \frac{1}{2}\mathbf{A}^T\mathbf{D}\mathbf{A} \tag{3}$$

donde **1** es un vector unidad de dimensión N, **A** es un vector con elementos α_i, y **D** es una matriz $N \times N$ con elementos dados por $D_{ij}=s_i s_j<\mathbf{r}_i,\mathbf{r}_j>$.

La solución se puede expandir en función de los estados del canal: $\boldsymbol{w}=\sum_i \alpha_i s_i \mathbf{r}_i$ [1, 4]. Tan sólo los $\mathbf{r}_i$ más cercanos al hiperplano tienen $\alpha_i>0$ y se llaman Vectores Soporte. Para todos los demás $\alpha_i=0$.

La generalización a un clasificador no lineal se realiza reemplazando el producto escalar $<\mathbf{r}_i,\mathbf{r}_j>$, por un "kernel" no lineal $k(\mathbf{r}_i,\mathbf{r}_j)$ que satisfaga la condición de Mercer [1]. Para el problema de igualación, un "kernel" apropiado es una función gaussiana

$$k(\mathbf{r}_i,\mathbf{r}_j)=\exp\left(\frac{-\|\mathbf{r}_i-\mathbf{r}_j\|_2^2}{s^2}\right) \qquad (4)$$

donde σ^2 es un parámetro dado. En este caso, la SVM toma la forma de un igualador RBF, con la ventaja de que la SVM selecciona sólo un número reducido de Vectores Soporte lo que simplifica la estructura del igualador.

2.3. Regularización del OH

Los estados del canal que se emplean como Vectores Soporte no tienen en cuenta información sobre el ruido, mientras que el MMSE incorpora información *a priori* sobre la σ_n^2. Por este motivo la solución obtenida resolviendo (3) sólo puede considerarse óptima cuando $SNR\rightarrow\infty$. Esto hace que el OH pueda proporcionar peores resultados que el MMSE para *SNR* bajas.

Una alternativa para incorporar información sobre el ruido es regularizar la función de coste mediante el empleo de un parámetro C

$$W(\mathbf{A})=\mathbf{A}^T\mathbf{1}-\frac{1}{2}\left(\mathbf{A}^T\mathbf{D}\mathbf{A}+\frac{1}{C}\mathbf{A}^T\mathbf{A}\right) \qquad (5)$$

La similitud con la regularización realizada en la solución MMSE sugiere seleccionar un parámetro proporcional a s_n^2.

3. EL ALGORITMO ADATRÓN

El algoritmo Adatrón proporciona la solución para el OH o la SVM de forma más sencilla y rápida que resolviendo el problema de optimización cuadrática. Conocidos los estados del canal $\mathbf{r}_i$ el algoritmo consiste en:

1.- Inicializar $\alpha_i=0$, $i=1,...,N$; y el factor de aprendizaje h.

2.- Hasta que se produzca la convergencia

  2.1.- Escoger un patrón $\mathbf{r}_i$, $i\in\{1,...,N\}$

  2.2.- Calcular $d_i=h(1-s_i f(\mathbf{r}_i))$

  2.3.- Si $(\alpha_i+d_i)>0$ entonces $\alpha_i=\alpha_i+d_i$, si no $\alpha_i=0$

donde $f(\mathbf{r}_j)=<a\otimes\mathbf{s},\mathbf{D}_j>$, $\otimes$ indica multiplicación elemento a elemento, y $\mathbf{D}_j$ es la fila *j-ésima* de la matriz **D**.

4. RESULTADOS

En la Figura 1 se muestran los resultados obtenidos por los siguientes igualadores: FIR-MMSE, OH, "soft margin" OH, SVM (σ^2=1) y Bayesiano, para el canal $H(z)=1+0.5z^{-1}$, con m=2 y retardo d=1, donde el número de estados del canal es 8. En la Figura 2 se muestran los resultados obtenidos para el canal $H(z)=0.6+z^{-1}+0.5z^{-2}+0.2z^{-3}$, con m=7, y d=4, para el que el número de estados del canal es 1024. El OH proporciona, en general, peores resultados que el MMSE para *SNR* bajas (más apreciable en la Fig. 2), aunque el "soft margin" OH evita este problema.

La SVM mejora los resultados obtenidos con los igualadores lineales, permitiendo un compromiso entre coste y prestaciones. Mientras que el estimador Bayesiano emplea todos los estados del canal para obtener la función de decisión (8 y 1024 para cada caso), la SVM utiliza un número limitado de vectores soporte, en estos ejemplos 4 y 303 respectivamente.

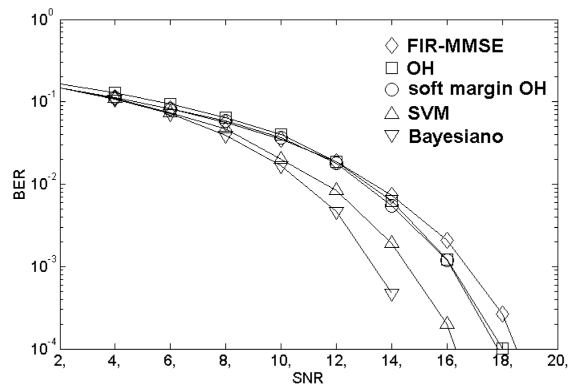

Figura 1. *Probabilidad de error para H(z)=1+0.5z^{-1}*

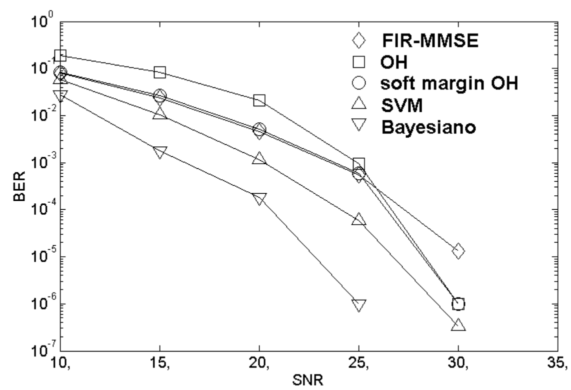

Figura 2. *Probabilidad de error para H(z)=0.6+z^{-1}+0.5z^{-2}+0.2z^{-3}*

5. REFERENCIAS

[1] V. Vapnik [4], *The Nature of Statistical Learning Theory*, New York: Springer Verlag, 1995.

[2] I. Santamaría, C. Pantaleón, J.C. Príncipe, "Minimizing BER in DFE's with the Adatrón Algorithm", enviado para su presentación en *IEEE Int. Workshop on Neural Networks for Signal Processing (NNSP'2001)*.

[3] J.K. Anlauf, M. Biehl, "The AdaTron: an adaptive perceptron algorithm", *Europhys. Lett.*, vol. 10, pp. 687-692, 1989.

[4] C. Cortes, V. Vapnik, "Support-Vector networks", *Machine Learning*, vol. 20, pp. 273-297, 1995.