Cómputo de Pesos Óptimos en Ecuaciones Lineales para Estimación en Procesos Aleatorios (Septiembre 2011)

Iván López Espejo

En este texto se presentan la demostración y definición del cómputo de pesos óptimos en un conjunto de ecuaciones lineales destinadas a la estimación en procesos aleatorios. Dicha herramienta puede resultar de gran utilidad como sencillo sustituto de las redes neuronales en las situaciones que convengan. El criterio de optimalidad está basado en la minimización del error cuadrático.

I. Desarrollo

Definamos un ejemplo concreto a través del cual comprender el ámbito de utilidad de la teoría que a continuación se presenta. Supóngase que se desea construir un VAD (Voice Activity Detector) de tal forma que las tramas de audio se clasifiquen en sonoras o en silenciosas a partir del estudio de los primeros 12 coeficientes cepstrales de la trama en cuestión. Una forma posible de hacerlo sería entrenando mediante retropropagación una red neuronal, de tal forma que se constituya una serie de pesos óptimos tal que, empleados en futuras estimaciones, produzcan una buena clasificación. No obstante, en determinadas ocasiones puede resultar más sencillo e igualmente (o suficientemente) efectivo emplear directamente un conjunto de ecuaciones lineales tal que, alimentado con suficientes datos de entrenamiento, se pueda calcular rápidamente un conjunto de pesos óptimos en base a algún un criterio:

$$
\begin{gathered}
y_1 = x_{11}w_1 + x_{12}w_2 + \cdots + x_{1N}w_N \\
y_2 = x_{21}w_1 + x_{22}w_2 + \cdots + x_{2N}w_N \\
\vdots \\
y_M = x_{M1}w_1 + x_{M2}w_2 + \cdots + x_{MN}w_N
\end{gathered}
$$

Sea M el número de datos de entrenamiento y N el número de variables de entrada en juego, si $M = N$, tendremos un sistema de ecuaciones lineales determinado siempre y cuando no existan ecuaciones linealmente dependientes en él. Sin embargo, si $M > N$, el sistema estaría sobredeterminado, pero no en el ámbito de procesos aleatorios, donde es necesario encontrar el conjunto de coeficientes o pesos, $\{w_n\}$, tal que sea óptimo en función de algún criterio. El criterio seguido en el presente texto es la minimización del error cuadrático. En la situación del ejemplo expuesto, $N = 12$ es el número de coeficientes cepstrales a partir de los cuales se clasificará la trama de audio de entrada al sistema, donde x_{mn} es el coeficiente cepstral *n-ésimo* que participa de la ecuación de entrenamiento *m-ésima*. Por último, y_m es el resultado deseado y conocido durante la fase de entrenamiento que indica si la trama de audio es sonido o silencio. Deberemos poder asignar el valor de salida deseado (al igual que en una red neuronal): por ejemplo, 1 para las tramas sonoras y 0 para las tramas de silencio. Es interesante que $M > N$, es decir, que exista una suficiente cantidad de datos de entrenamiento, si bien este aspecto es crítico (así como los valores de entrada y salida para una ecuación de entrenamiento concreta) y complicado de gestionar, al igual que en una red neuronal.

A continuación calculamos el conjunto de pesos $\{w_n\}$ tal que es óptimo en el sentido de que minimiza el error cuadrático. Defínase el error acaecido de estimación para la ecuación de entrenamiento *m-ésima* como:

$$e(m) = y_m - \hat{y}_m = y_m - \sum_{n=1}^{N} w_n x_{mn}$$

En consecuencia, el error cuadrático se expresa como:

$$\varepsilon = \sum_{m=1}^{M} e^2(m) = \sum_{m=1}^{M} \left[y_m - \sum_{n=1}^{N} w_n x_{mn} \right]^2$$

Minimizamos el error cuadrático para obtener el conjunto de pesos óptimo haciendo:

$$\frac{\partial \varepsilon}{\partial w_n} = 0$$

De lo cual resulta:

$$
\begin{gathered}
-2 \sum_{m=1}^{M} \left[y_m - \sum_{n=1}^{N} w_n x_{mn} \right] x_{ml} = 0 \Rightarrow \\
\Rightarrow \sum_{n=1}^{N} w_n \sum_{m=1}^{M} x_{mn} x_{ml} = \sum_{m=1}^{M} y_m x_{ml}
\end{gathered}
$$

Si disponemos el anterior resultado en formato matricial, este puede resultar de gran utilidad a la hora de calcular los coeficientes óptimos computacionalmente:

$$
\begin{bmatrix}
\sum_{m=1}^{M} x_{m1}x_{m1} & \sum_{m=1}^{M} x_{m2}x_{m1} & \cdots & \sum_{m=1}^{M} x_{mN}x_{m1} \\
\sum_{m=1}^{M} x_{m1}x_{m2} & \sum_{m=1}^{M} x_{m2}x_{m2} & \cdots & \sum_{m=1}^{M} x_{mN}x_{m2} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{m=1}^{M} x_{m1}x_{mN} & \sum_{m=1}^{M} x_{m2}x_{mN} & \cdots & \sum_{m=1}^{M} x_{mN}x_{mN}
\end{bmatrix}
\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix}
=
\begin{bmatrix}
\sum_{m=1}^{M} y_m x_{m1} \\
\sum_{m=1}^{M} y_m x_{m2} \\
\vdots \\
\sum_{m=1}^{M} y_m x_{mN}
\end{bmatrix}
$$

Resolviendo el anterior sistema de ecuaciones llegamos a la solución deseada. En nuestro ejemplo, para nuevos conjuntos de coeficientes cepstrales podemos emplear los anteriores pesos con el fin de clasificar una nueva trama de audio en sonora o silenciosa ($\hat{y}$ más cercano a 1 o 0, respectivamente), habiendo así constituido un VAD con un método alternativo al empleo de una red neuronal. El anterior resultado puede aplicarse a numerosos procesos aleatorios, siendo preciso estudiar a priori su validez, pudiendo suponer una solución sencilla que sustituya otras metodologías más complejas. Para ello, deberemos verificar que el error cuadrático medio de entrenamiento es lo suficientemente bajo, monitorizar el rendimiento posterior y, en caso necesario, volver a realimentar al primer paso para modificar la definición de las ecuaciones de entrenamiento con el fin de reestimar unos coeficientes óptimos.