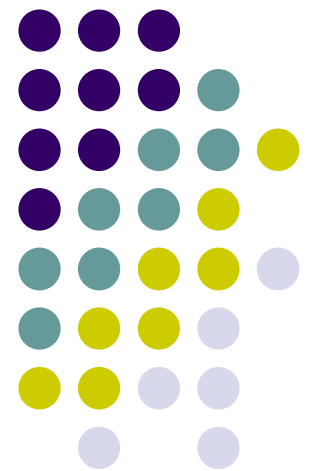


Clasificación de Música por Genero Utilizando Redes Neuronales Artificiales

Elkin García, Germán Mancera,
Jorge Pacheco





Presentación

- Los autores han desarrollado un método de clasificación de música a través de la extracción de parámetros característicos de genero que son utilizados como entradas en el trabajo de la red neuronal.
- La extracción de parámetros se agrupa en:
 - Análisis en tiempo, que hace uso de la correlación
 - Análisis en frecuencia usando la transformada de Fourier aplicada a diferentes intervalos de frecuencia.
 - Extracción de dos aportes rítmicos independientes haciendo uso de la transformada de Wavelet y teniendo como función madre el sombrero mejicano.

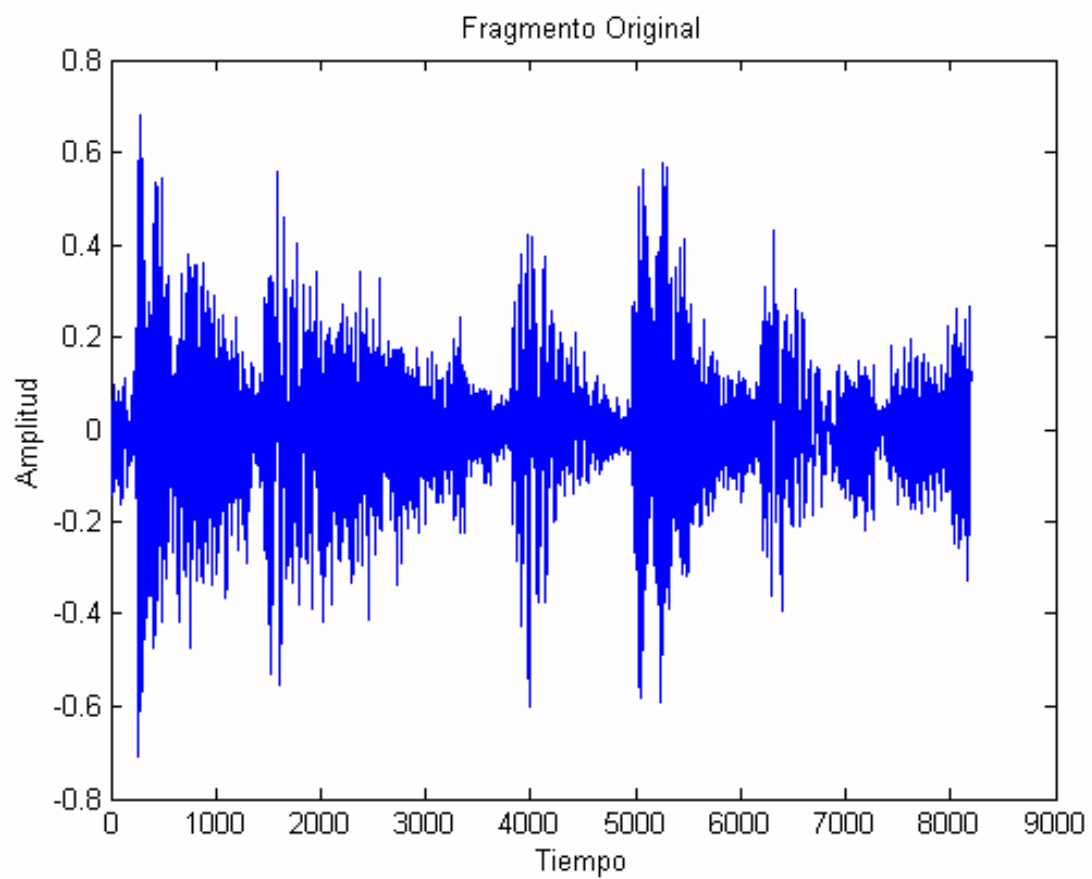


Presentación

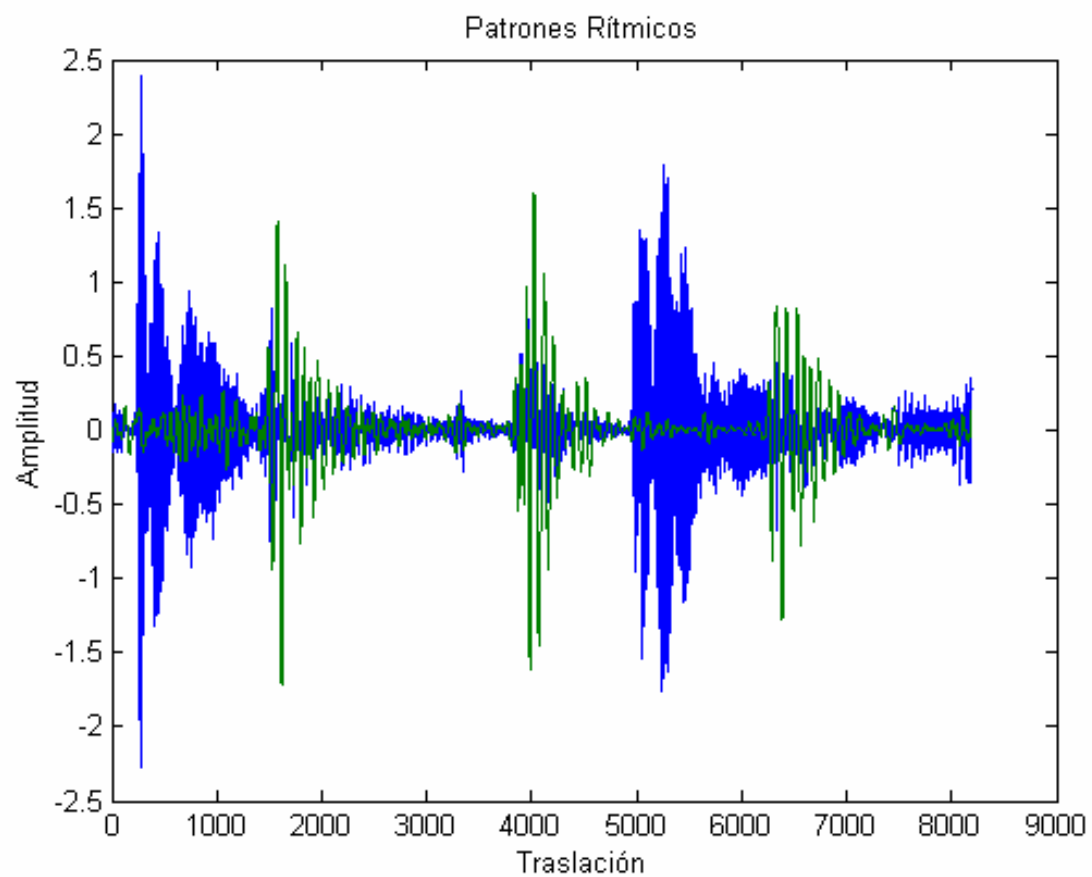
- Los parámetros son filtrados utilizando la transformación de Karhunen-Loève, buscando eliminar el contenido poco informativo
- Parámetros extraídos a partir de las canciones:

Parámetros Basados en Análisis de Tiempo 1 – 5
Parámetros Basados en Análisis de Frecuencia 6 – 10
Parámetros Basados en la Transformada Wavelet 11 – 33

Presentación



Presentación





Presentación

- Topologías de red:
 - Red de propagación inversa al error (backpropagation)
 - Red de funciones de base radial
 - Adaboost

Presentación



- Pseudocódigo del algoritmo de entrenamiento backpropagation

Escoja el número de neuronas de la capa oculta.

Inicialice los pesos W .

Repita

Repita para cada uno de los datos i X

Calcule la salida para la primera capa

$$G_j = f \left(\sum_{i=1}^N X_i W_{hji} + W_{hjB} \right)$$

Calcule la salida total

$$F_j = f \left(\sum_{i=1}^K G_i W_{oji} + W_{ojB} \right)$$

Fin

Repita para cada una de las salidas F_j para todo j desde 1 hasta M

Calcule el error como

$$\mathcal{E}_{oj} = (Y_j - F_j)[F_j(1 - F_j)]$$

Fin

Hasta condición de terminación



Presentación

- Pseudocódigo del algoritmo de entrenamiento red de función de base radial

Escoja el número de neuronas de la capa escondida.

Inicialice ω_i (Centros de las Funciones de Base Radial).

Repita

Para cada uno de los datos x_i

Asigne x_i a θ_i (cluster) tal que $\|x_i - \omega_i\|^2$ sea la mínima entre las posibles

Fin

Para cada uno de los clusters θ_i

$$\omega_i = \frac{1}{\|\theta_i\|} \sum_{j \in \theta_i} x_j$$

Fin

Hasta que ninguna de las x_i cambien de cluster.

Encuentre la varianza de los datos mediante

$$\sigma_i^2 = \frac{1}{|\theta_i| - 1} \sum_{x \in \theta_i} (x - \omega_i)^T (x - \omega_i)$$

Encuentre los pesos de la capa de salida mediante

$$W^T = \Phi^T T$$



Presentación

- Pseudocódigo del algoritmo de entrenamiento Adaboost

Se hace uniforme $D_t(i) = \frac{1}{m}$

Para t desde 1 hasta T

Se ejecuta el clasificador débil con entradas X, Y, D

Se calcula

$$\alpha_t = \frac{1}{2} \left(\log \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right) \right) \in R$$

Se actualiza

$$D_{t+1}(i) = \frac{D_t(i) e^{(-\alpha_t \beta_t(i))}}{Z_t}$$

donde $\beta(i)$ es un factor tal que

$$\beta(i) = \begin{cases} -1 & \text{si } H_t(i) = Y(i) \\ 1 & \text{si } H_t(i) \neq Y(i) \end{cases}$$

Fin

Se calcula la hipótesis final

$$H(i) = p \text{ tal que } \sum_{t=1}^T \alpha_t g_t(i) \text{ es máxima}$$



Criterios Generales

- El contenido del artículo atañe a una importante porción de la comunidad actual.
- A pesar de no ser los primeros en enfrentar el problema, es cierto que han logrado un importante avance en el proceso de reconocimiento de señales y en la solución de la clasificación de archivos de audio.
- Gracias a la eficacia del método desarrollado, el trabajo sobresale entre el reducido grupo de artículos.
- La temática es de interés general y el tratamiento del tema es el apropiado.
- El artículo ofrece una respuesta específica al problema de la clasificación de música para tres géneros específicos: salsa, merengue, vallenato.



Marco Teórico

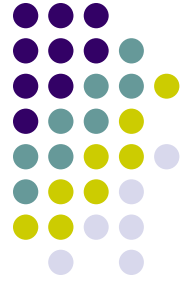
- **Red Neuronal Artificial**

- Una red neural artificial es un sistema de análisis consistente en un grupo de ecuaciones matemáticas interconectadas que aceptan datos de entrada y calculan una respuesta utilizando análisis no lineal a través del cálculo de mínimos cuadrados.
- Esas ecuaciones reciben el nombre de neuronas, nodo, unidad de procesamiento (PE) o celda y funcionan de un modo análogo a las neuronas, en las que un potencial de acción se genera cuando se ha acumulado una cantidad suficiente de neurotransmisor. En el nodo el neurotransmisor es la suma neta del peso de todas las señales entrantes, excitatorias e inhibitorias el cual es comparado con un nivel límite; cuando éste es superado, el nodo dispara, de otro modo permanece silencioso. Así, a diferencia de los sistemas de computación convencionales, las redes neurales tienen múltiples procesadores trabajando en paralelo, que sirven además como elementos de memoria.



Marco Teórico

- Red Neuronal Artificial
 - El sistema es capaz de encontrar relaciones nuevas entre los datos con los que es entrenado; así mismo, a través de retroalimentación la RNA puede modificar sus parámetros para disminuir el error de sus respuestas
 - Una de las características importantes de la RNA es su capacidad para generalizar, es decir, obtener una respuesta adecuada (predicción) a partir de un grupo de datos diferentes a los usados en la fase de entrenamiento.



Marco Teórico

- **Perceptrón Multinivel**
 - El perceptrón multinivel es la forma más común de red neural. Consiste en una serie de unidades procesadoras (nodos) interconectadas por una serie de pesos de un modo que permite a la información viajar en serie o en paralelo.

Marco Teórico

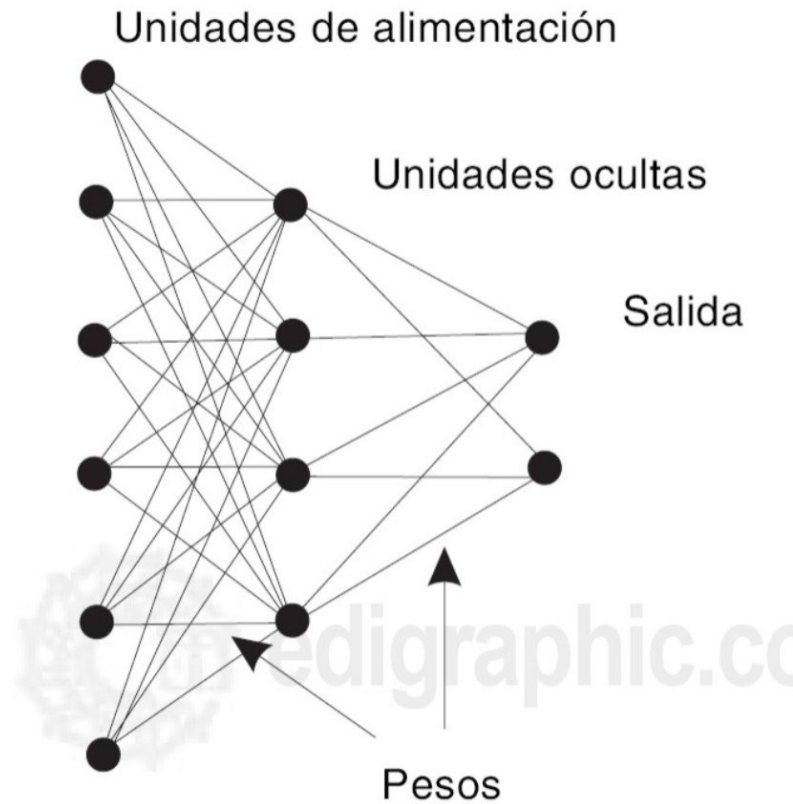
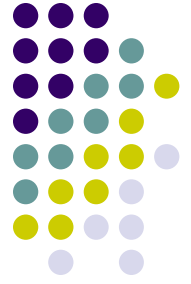


Figura 1. *Perceptrón multinivel.*



Marco Teórico

- **Backpropagation**

- La Backpropagation es un tipo de red de aprendizaje supervisado, que emplea un ciclo propagación – adaptación de dos fases. Una vez que se ha aplicado un patrón a la entrada de la red como estímulo, éste se propaga desde la primera capa a través de las capas superiores de la red, hasta generar una salida. La señal de salida se compara con la salida deseada y se calcula una señal de error para cada una de las salidas.
- Las salidas de error se propagan hacia atrás, partiendo de la capa de salida, hacia todas las neuronas de la capa oculta que contribuyen directamente a la salida. Sin embargo las neuronas de la capa oculta sólo reciben una fracción de la señal total del error, basándose aproximadamente en la contribución relativa que haya aportado cada neurona a la salida original. Este proceso se repite, capa por capa, hasta que todas las neuronas de la red hayan recibido una señal de error que describa su contribución relativa al error total. Basándose en la señal de error percibida, se actualizan los pesos de conexión de cada neurona, para hacer que la red converja hacia un estado que permita clasificar correctamente todos los patrones de entrenamiento.

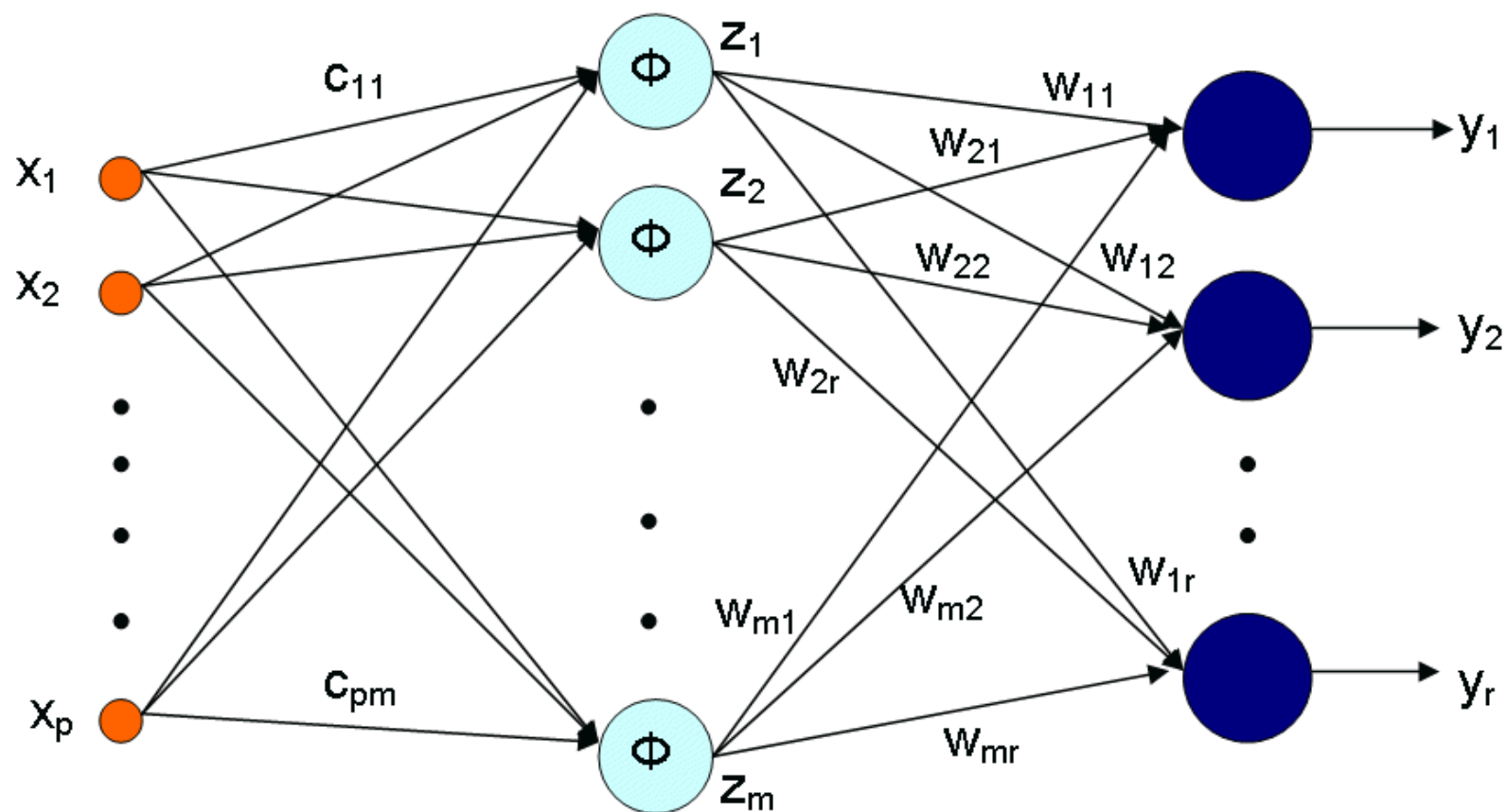


Marco Teórico

- **Red de funciones de base radial**

- Son funciones cuya salida depende de la distancia a un punto denominado ***Centro***.
- **Características:**
 - - Simétricas respecto de $x=0$
 - - Se definen con al menos dos parámetros.
 - o ***Centro***: Punto donde la función posee un extremo.
 - o ***Anchura***: Magnitud de la variación de la función según se aleja del centro.
 - Capa de entrada: reciben las señales del exterior, no realizan ningún preprocesado.
 - Capa Oculta: reciben las señales de la capa de entrada y realizan una ***transformación local*** y no lineal sobre dichas señales
 - Capa de Salida: Se realiza una combinación lineal de las activaciones de las neuronas de la capa oculta y actúa como salida de la red.

Marco Teórico





Metodología

- De los elementos pertenecientes a una base de 500 canciones, 177 salsas, 170 merengues y 163 vallenatos, en formato mp3, son extraídos los parámetros. Se establecieron los valores de las salidas deseadas y se evaluó el desempeño de las redes.
- Metodología apropiada, utilización de recursos físicos y digitales, como bases de archivos de música, computadores, trabajo investigativo.



Resultados

- Los resultados de la extracción de parámetros están claramente consignados, así como los de eficiencia de clasificación y las comparaciones entre redes.
- Los datos expuestos son coherentes y en apariencia sólidos.
- Debido a que no todos los géneros guardan la misma relación de diferencia la efectividad del método podría estimarse de forma errónea.



Resultados

- **Red backpropagation**

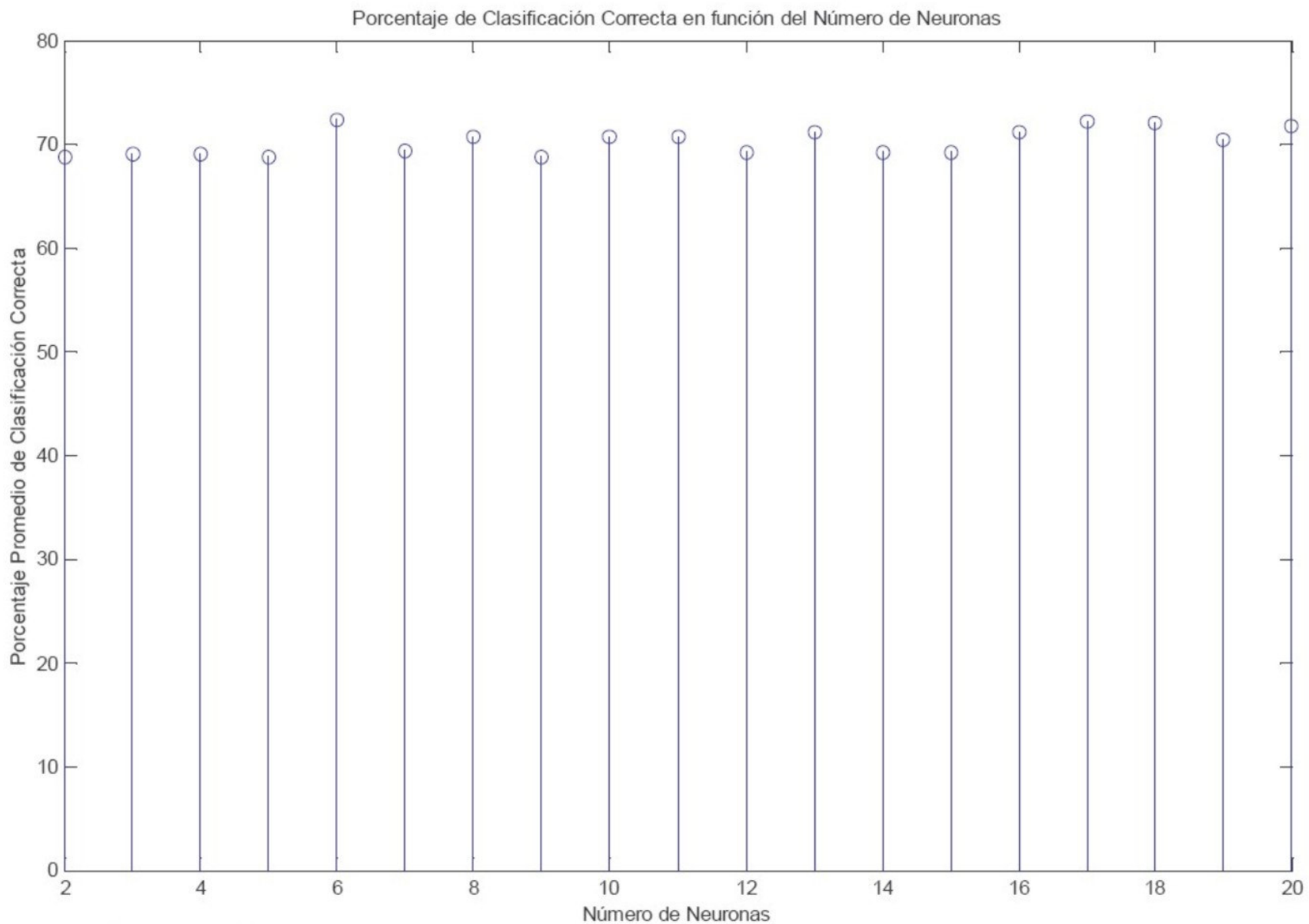
- El error promedio de clasificación para el conjunto de validación es bastante similar para los diferentes ordenes de la red, y no se presentan mejoras significativas al aumentar el número de neuronas de la capa oculta. El porcentaje promedio de clasificación correcta tiene su máximo para la red con 6 neuronas en la capa escondida, con un porcentaje de acierto del 72.25 %.

- **Red de funciones de base radial**

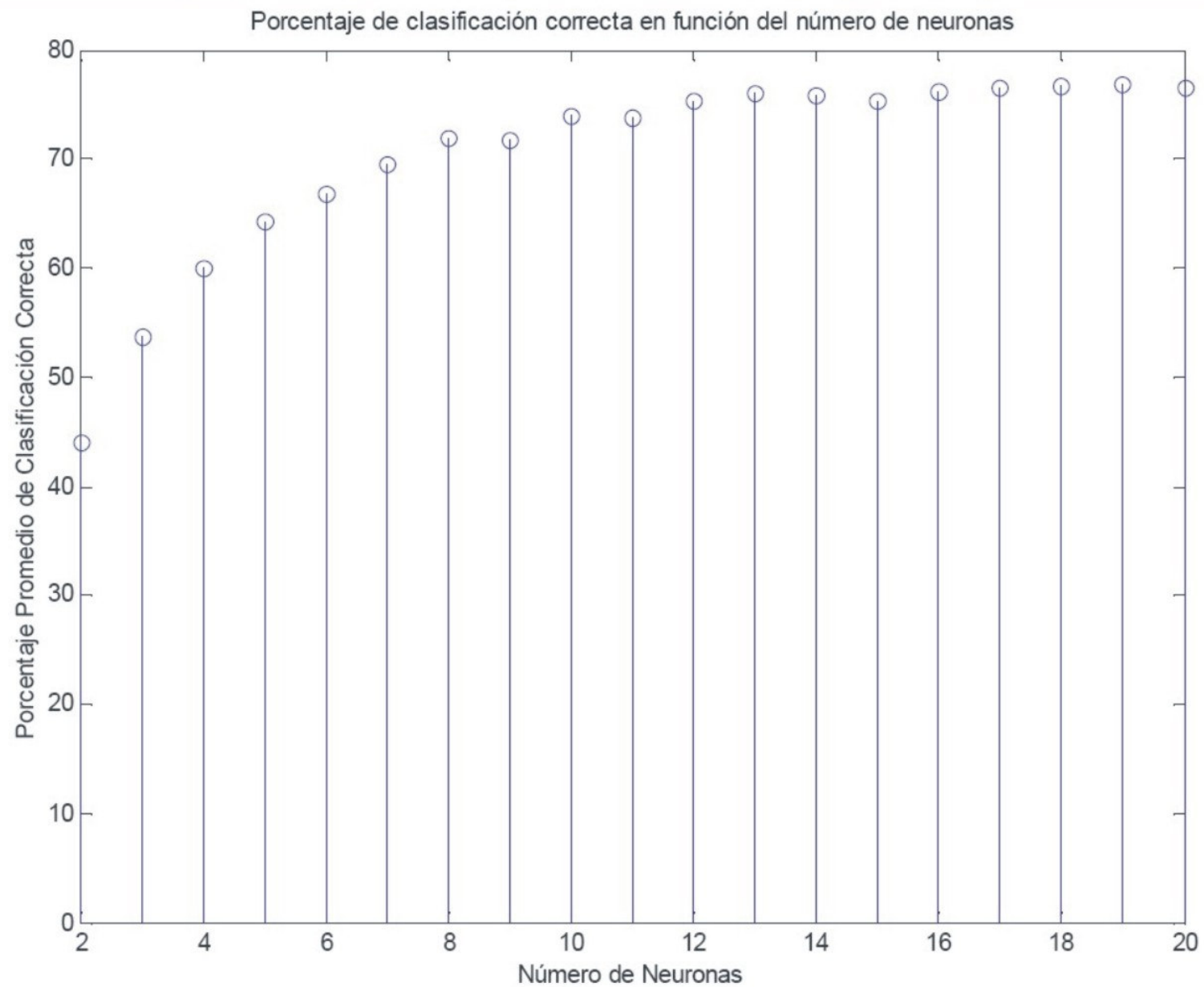
- El porcentaje promedio de clasificación aumenta a medida que se agregan neuronas a la capa escondida, para esta topología. El porcentaje promedio de clasificación correcta tiene su máximo para una red con 19 neuronas en la capa escondida, con un porcentaje promedio de acierto del 76.89 %.

- **Adaboost**

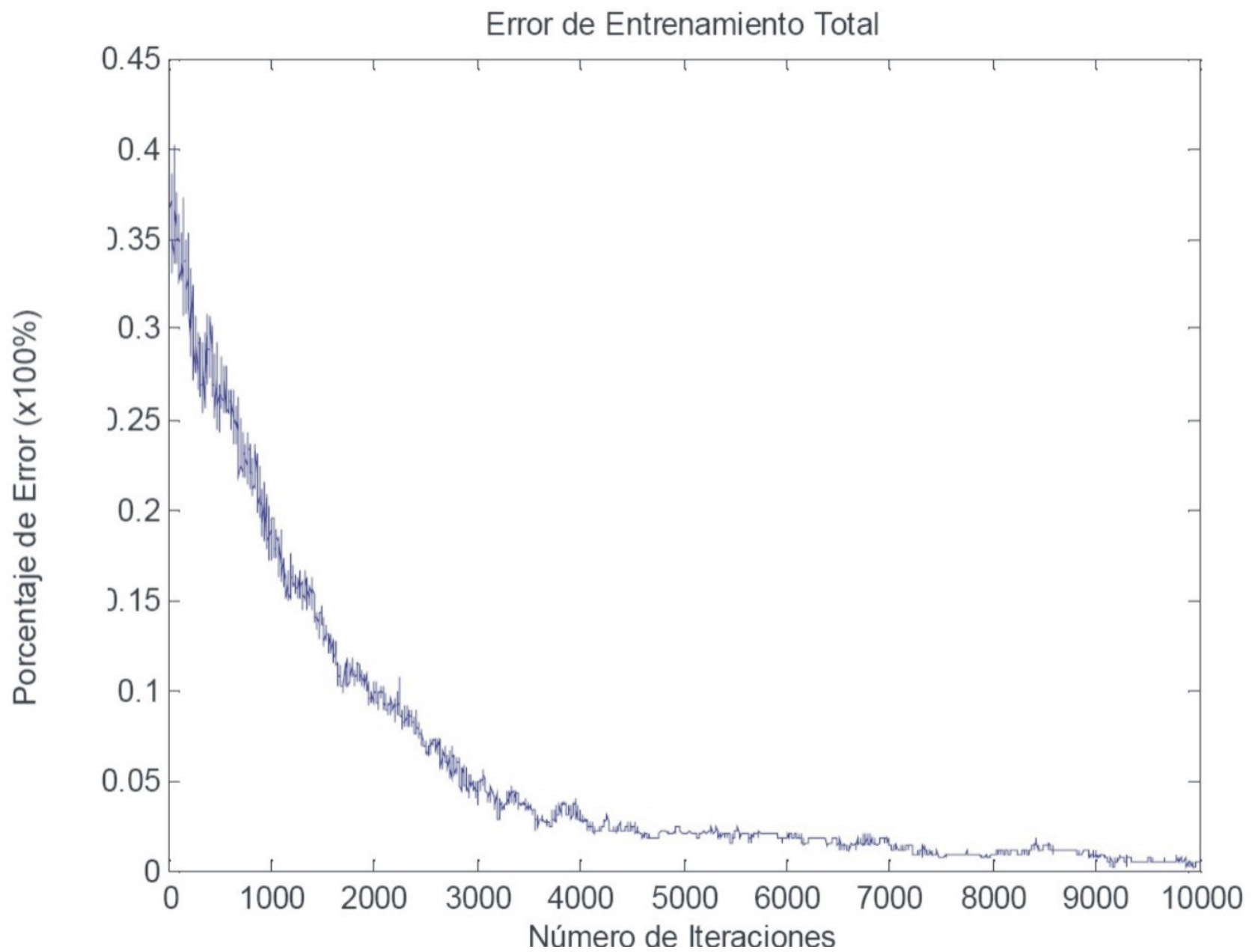
- El error de entrenamiento baja considerablemente hasta la iteración 3000 y continúa bajando lentamente hasta la iteración 10000 a un punto cercano a 0. El error total de entrenamiento alcanzó un valor mínimo de 0.0022, que equivale a una canción del conjunto de entrenamiento mal clasificada.



Porcentaje Promedio de Clasificación Correcta de Backpropagation



Porcentaje Promedio de Clasificación Correcta de RBF



Error de entrenamiento total para clasificador Adaboost



Conclusiones

- La lógica con que se trata el problema es valida, extraer datos para crear generalización por comparación entre intervalos separados es una idea valida.
- Las conclusiones son producto de lo observado durante el desarrollo del tema.
- El tratamiento del tema es preciso, centrado, los autores recurren a la teoría pertinente y exponen los resultados.



Forma

- El artículo está dividido en:
 - Introducción
 - Desarrollo
 - Especificaciones
 - Resultados
 - Conclusiones
- No presenta errores tipográficos, gramaticales u ortográficos.
- Todas las tablas y figuras están correctamente tabuladas y explicadas, las etiquetas indican el contenido y la relación con el texto.