APLICACIÓN DEL MODELO ANFIS A LA SINTETIZACION DE NOTAS MUSICALES Y SEÑALES DE VOZ

Carlos Stalin Alvarado Sánchez
Carlos Jordán Villamar
Facultad de Ingeniería en Electricidad y Computación
Escuela Superior Politécnica del Litoral
Campus Gustavo Galindo, Km. 30.5 Vía Perimetral
Apartado 09-01-5863. Guayaquil-Ecuador
calvarad@fiec.espol.edu.ec
cjordan@fiec.espol.eu.ec

Resumen

El objetivo de este trabajo es presentar una metodología de Redes Neuro-Difusa aplicado a sintetizar notas musicales y señales de voz, cuyo resultado puede emplearse para mejorar la calidad de sonido de los sintetizadores que actualmente existen. Se procuró emplear una herramienta que combine los Sistemas de Inferencia Difusa y las Redes Adaptables; se escogió el modelo ANFIS (Adaptive Neuro-Based Fuzzy Inference System). Para encontrar la mejor arquitectura ANFIS se empleó un método heurístico que combina la cantidad y tipo de funciones de pertenencia de las variables de entrada. Para realizar este trabajo se uso el software CACIQUE, que es una herramienta para sintetizar señales, el cual fue desarrollado por el autor. Con el fin de evaluar el desempeño del ANFIS se comparó los resultados con los obtenidos por Axel Roben en [6].

Palabras Claves: ANFIS, Modelos Neuro-Difusa

Abstract

The aim of this paper is to present a methodology of Neuro-Fuzzy Networks applied to synthesize musical notes and voice signals, the result of which can be used to improve the quality of sound synthesizers currently exist. Sought to use a tool that combines the Fuzzy Inference Systems and Adaptive Networks, the model was chosen ANFIS (Adaptive Neuro-Based Fuzzy Inference System). To find the best ANFIS architecture employed a heuristic method that combines quantity and type of membership functions of input variables. To make this work we use CACIQUE software, which is a tool to synthesize signals, which was developed by the author. In order to evaluate the performance of the ANFIS are compared with results obtained by Axel Röben in [6].

Keywords: ANFIS, Neuro-Fuzzy Models

1. Introducción

Las interfases con el usuario en lenguaje natural serán las interfases del futuro, porque permitirán al usuario comunicarse con el computador de una manera mas “natural”, como lo hacemos normalmente con nuestros congeneres, utilizando la facultad del habla. Al diseñar estas interfases, dos funciones esenciales deben implementarse: el reconocimiento y la sintetización de la voz. En este trabajo queremos aplicar el modelo ANFIS (Sistemas de Inferencias Borrosas Basados en Redes Adaptables) al problema de sintetizar señales de voz y notas musicales.

Existe una gran clase de problemas para los que la inteligencia humana es mucho más rápida y eficiente que el procesamiento de la mejor computadora actual. Tratando de atacar estas deficiencias en la forma de resolver problemas tradicionales surgieron áreas como Lógica Difusa (Fuzzy Logic, FL), Redes Neuronales Artificiales (Artificial Neural Networks, ANN) y otras herramientas que se suelen agrupar en el concepto de Inteligencia Artificial (AI).

En la actualidad existen instrumentos electrónicos diseñados para producir sonido generado artificialmente, usando técnicas como síntesis aditiva, substractiva, de modulación de frecuencia, de modelado físico o modulación de fase, para crear sonidos.

Los más conocidos son los órganos musicales que pueden producir sonidos de cualquier instrumento musical; la gran mayoría utiliza el estándar MIDI que se trata de un protocolo industrial estándar que permite a las computadoras, sintetizadores, secuenciadores, controladores y otros dispositivos musicales electrónicos comunicarse y compartir información para la generación de sonidos.

En lo que compete a la Sintetización de señales de voz, en la actualidad estas tecnologías han alcanzado un importante desarrollo, tanto en el ámbito de la síntesis como en el del reconocimiento. Este progreso está muy relacionado con el conocimiento de la fisiología de la voz y de la audición, a los que la tecnología trata de imitar.

Los sintetizadores de voz están divididos en su forma de trabajar y como fueron realizados para la reproducción de un mejor sonido y no sean tan robotizados, entre ellos tenemos:

- Articulatorios
- Por formantes
- Derivados de las técnicas de predicción lineal
- Por concatenación de forma de onda

Siendo el último de ellos el que proporciona más calidad y un alto grado de naturalidad.

Por otro lado existen investigadores que han aportado a la sintetización de sonidos y de voz humana tal como Axel Röben que con sus trabajos de Redes Neuronales para modelar series de tiempo de instrumentos musicales y sintetizar voz humana ha logrado obtener resultados aceptables en cuanto al error de la sintetización a lo real.

2. Metodología

Los sistemas híbridos que combinan lógica difusa, redes neuronales, algoritmos genéticos y sistemas expertos proporcionan los métodos más eficientes para resolver una gran variedad de problemas. Cada una de esas técnicas tiene propiedades computacionales particulares (por ejemplo: habilidad de aprender) que las hace óptimas para resolver ciertos problemas. Uno de estos sistemas híbridos corresponde a los sistemas Neuro-Difusa, que combinan las técnicas de redes neuronales artificiales y las técnicas de inferencia difusa.

La lógica difusa proporciona un mecanismo de inferencia sobre la incertidumbre y las redes neuronales ofrecen grandes ventajas computacionales, tales como el aprendizaje, adaptación, tolerancia a fallas, el paralelismo y la generalización. Las redes neuronales son usadas para representar los sistemas de inferencia difusa, los mismos que son empleados como sistemas de toma de decisiones. A pesar de que la lógica difusa puede codificar el conocimiento a través de etiquetas lingüísticas, usualmente toma mucho tiempo definir y ajustar las funciones de pertenencia. Las técnicas de aprendizaje de las redes neuronales pueden automatizar este proceso y reducir sustancialmente el tiempo y el costo de desarrollo al mejorar el desempeño del modelo.

Teóricamente las redes neuronales y los sistemas difusos son equivalentes, pero en la práctica cada uno tiene sus propias ventajas y desventajas. En las redes neuronales, el conocimiento se adquiere automáticamente por el algoritmo de backpropagation, pero el proceso de aprendizaje es relativamente lento (gran cantidad de épocas de entrenamiento) y el análisis de la red entrenada es difícil (modelo de caja negra). No es posible extraer el conocimiento estructural (reglas) de la red neuronal ni puede éste integrarse a la información especial sobre el problema en la red neuronal con el fin de simplificar el procedimiento de aprendizaje. Los sistemas difusos son más favorables porque su comportamiento puede ser explicado con base en reglas difusas y, de esta forma, su desempeño puede ser ajustado modificando estas reglas. Sin embargo, la adquisición del conocimiento es difícil, y, además, el universo de discurso de cada variable necesita ser dividido en intervalos, por lo que las aplicaciones de los sistemas difusos se restringen a problemas en los cuales el conocimiento está disponible en un número de variables de entrada pequeño. Para superar el problema de la adquisición del conocimiento, las redes neuronales son extendidas para extraer automáticamente la regla difusa de los datos numéricos.

2.1 Descripción del modelo ANFIS

ANFIS (Sistemas de Inferencia difusos basados en redes adaptivas) es una clase de redes adaptivas el cual es funcionalmente equivalente a sistemas de inferencia difuso.

Un modelo ANFIS es un modelo híbrido donde las reglas se aplican siguiendo una estructura de red tipo neuronal que puede ser interpretado como una red neuronal con parámetros difusos o como un sistema difuso con parámetros o funcionamiento distribuido.

Las capacidades adaptivas de las redes ANFIS las hacen directamente aplicables a una gran cantidad de áreas como la sintonización automatizada de los controladores difusos, en el modelamiento donde se necesita explicar datos pasados y predecir datos futuros, en control adaptativo, en procesamiento y filtrado de señales, en clasificación de datos y extracción de características a partir de ejemplos, entre otros.

Un sistema ANFIS engloba las mejores características de los sistemas difusos y de las redes neuronales. De los primeros utiliza la representación del conocimiento previo en un conjunto de restricciones (que se representan en la topología de la red) para reducir el espacio de búsqueda de optimización, mientras que de las redes neuronales emplean la adaptación de propagación inversa a la red estructurada para automatizar el ajuste de los parámetros.

La parte de la premisa de una regla define un subespacio difuso, mientras que el consecuente especifica la salida dentro de ese subespacio.

La estructura de los sistemas ANFIS permite utilizar métodos cualitativos y cuantitativos en la construcción de modelos. Además permite integrar, a la información incluida dentro de un conjunto de datos, el conocimiento de expertos expresados en forma lingüística y a través de la teoría de conjuntos difusos, expresados con base

La arquitectura ANFIS es la siguiente:

Este sistema híbrido neuro-difuso es funcionalmente equivalente al mecanismo de inferencia Takagi-Sugeno (T-S) de primer orden [5].

Regla 1: Si x es A1 and y es B1, entonces f1 = p1 x + q1 y + r1
Regla 2: Si x es A2 and y es B2, entonces f2 = p2 x + q2 y + r2

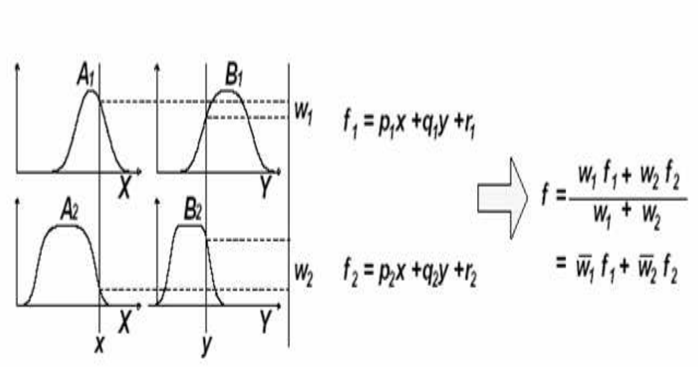

Figura 1 Razonamiento ANFIS

Donde A1, A2, B1, B2 son funciones de pertenencias (Conjuntos difusos)

Los niveles de activación de las reglas se calculan como wi = Ai(x). Bi(y) , i=1,2.., donde el operador lógico and puede ser modelado por una t-norma continua (producto). Las salidas individuales de cada regla son obtenidas como una combinación lineal entre los parámetros del antecedente de cada regla: fi = pi x+qi y+ri, i=1,2..,,. La salida de control del modelo se obtiene por la normalización de los grados de activación de las reglas por la salida individual de cada regla:

W1 y w2 son los valores normalizados de w1 y w2 y con respecto a la suma w1+w2 La red neuronal híbrida que representa este tipo de inferencia es una red adaptable con 5 capas, donde cada capa representa una operación del mecanismo de inferencia difuso. Esta red se muestra en la figura 2.

En esta arquitectura, todos los nodos de una misma capa tienen la misma función (los nodos representados con cuadros son nodos adaptables, es decir, sus parámetros son ajustables). La estructura de la red ANFIS consiste de cinco capas.

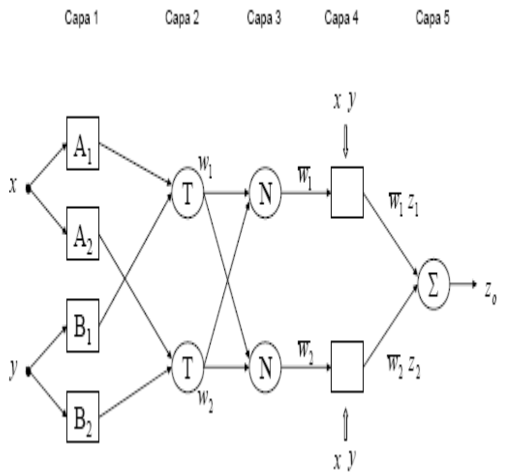

Figura 2 Red adaptiva equivalente ANFIS

Capa 1: Las entradas en esta capa corresponden a las entradas x y y, y la salida del nodo es el grado de pertenencia para el cual la variable de entrada satisface el término lingüístico asociado a este nodo.

$$O_i^1 = A_i(x)$$

Capa 2: Cada nodo calcula el grado de activación de la regla asociada a dicho nodo. Ambos nodos están representados con una T en figura 2, por el hecho de que ellos pueden representar cualquier t-norma para modelar la operación lógica and. Los nodos de esta capa son conocidos como nodos de reglas.

$$O_i^2 = w_i = A_i(x) \cdot B_i(y),\ i = 1,2..$$

Capa 3: Cada nodo en esta capa está representado por una N en la figura 2, para indicar la normalización de los grados de activación. La salida del nodo es el grado de activación normalizado (con respecto a la suma de los grados de activación) de la regla

$$O_i^3 = \overline{w}_i = \frac{w_i}{w_1 + w_2},\ i = 1,2.$$

Capa 4: La salida de los nodos corresponde al producto entre el grado de activación normalizado por la salida individual de cada regla.

$$O_i^4 = \overline{w}_i f_i = \overline{w}_i (p_i x + q_i y + r_i)$$

pi, qi, ,ri y forman el conjunto de parámetros. Los parámetros de esta capa se conocen como parámetros del consecuente.

Esos parámetros, como se puede ver, son los coeficientes de las funciones lineales que forman el consecuente de las reglas. Son parámetros ajustables, como los de la capa 1.

Capa 5: El único nodo de esta capa calcula la salida total del sistema (agregación) como la suma de todas las entradas individuales de este nodo.

$$O_1^5 = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}$$

En resumen, cada una de las capas tiene una misión concreta dentro del sistema:

- La primera capa representa la capa de pertenencia.
- La segunda capa se usa para generar el grado de disparo de la regla (T-norma)
- La tercera capa actúa de normalizador.
- La cuarta capa calcula la salida
- La última capa combina todas las salidas en una en su único nodo.

El modelo ANFIS tiene dos conjuntos de parámetros que deben ser entrenados: los parámetros del antecedente (constantes que caracterizan las funciones de pertenencia) y los parámetros del consecuente (parámetros lineales de la salida del modelo de inferencia).

El paradigma de aprendizaje del modelo ANFIS emplea algoritmos de gradiente descendiente para optimizar los parámetros del antecedente y el algoritmo de mínimos cuadrados para determinar los parámetros lineales del consecuente. Debido a esta combinación se lo conoce como regla de aprendizaje híbrido.

Jang [2] describe que para aplicar el aprendizaje híbrido en grupo, en cada época de entrenamiento debe ejecutarse un paso forward y un paso backward. En el paso forward, los parámetros de las funciones de pertenencia son inicializados y se presenta un vector de entrada-salida, se calculan las salidas del nodo para cada capa de la red y entonces los parámetros del consecuente son calculados usando el método de mínimos cuadrados.

Una vez identificados los parámetros del consecuente, el error es calculado como la diferencia entre la salida de la red y la salida deseada presentada en los pares de entrenamiento. Una de las medidas más usadas para el error de entrenamiento es la suma de errores cuadráticos, definida como:

$$ECM = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$$

Los Yi corresponden a los patrones de entrenamiento proporcionados (salidas deseadas) y Yi sombrero es la correspondiente salida de la red. En el paso backward, las señales de error son propagadas desde la salida, en dirección de las entradas; el vector gradiente es acumulado para cada dato de entrenamiento. Al final del paso backward para todos los datos de entrenamiento, los parámetros en la capa 1 (parámetros de las funciones de pertenencia) son actualizados por el método descendente

Esto constituye la forma más corriente de entrenar un sistema ANFIS, pero de hecho, hay 4 métodos de actualización de parámetros que según sus complejidades son:

- Sólo gradiente descendente: Todos los parámetros se actualizan por esta técnica.
- Gradiente descendente y un paso de mínimos cuadrados: Sólo se aplican una vez al principio los mínimos cuadrados, para obtener los valores iniciales de los parámetros del consecuente y luego se utiliza el gradiente descendente para actualizar todos los parámetros.
- Gradiente descendente y mínimos cuadrados: Es el entrenamiento híbrido que se ha descrito.
- Mínimos cuadrados sólo: Linealiza el ANFIS. Utiliza los parámetros de las premisas y el algoritmo del filtro de Kalman extendido para actualizar los parámetros.

Se debe elegir el método más adecuado en función de la complejidad de computación y de los resultados obtenidos. En general, el método de mínimos cuadrados suele llevar una mayor computación que el gradiente descendente

3. Resultados

Para señales de voz se usara como ejemplo la palabra Hola y para las notas musicales, la nota LA de una guitarra eléctrica

Para identificar la mejor configuración del modelo ANFIS para pronosticar las señales se han tomado dos consideraciones:

a) S construirán un modelo ANFIS por cada cincuenta pares cardinales, determinándose en total n/50 modelos por cada señal
b) Se considerará que el orden autorregresivos del modelo es 1 es decir, la red adaptable tendrá solo una entrada correspondiente a la entrada anterior y una salida correspondiente al valor que se desea sintetizar.

La cantidad de funciones de pertenencia nFP, empleadas para fuzzyficar la entrada al modelo dependerá de la cantidad de información (datos) disponible, ya que debe guardar relación con el número de parámetros no lineales que deberán ser calculados para la salida del modelo.

Cada función de pertenencia presente en la entrada del modelo está relacionada con el número de parámetros a calcularse en la capa 4 de la red de la figura 24, por tanto, nFP deberá variar desde 2 hasta nFPmax. Se tendrá un máximo de 30 funciones de pertenencia para la entrada del modelo ANFIS

Se consideraron además 2 tipos de funciones de pertenencia, tFP, para representar la entrada al modelo; las funciones de pertenencia empleadas son, en orden de utilización, las funciones tipo triangular (trimf), gausiana (gaussmf).

De acuerdo a la metodología heurística para identificación de modelos ANFIS propuesta por Jang en [2] y aplicada por Riyanto en [11], se realizaron 30 x 2 simulaciones, lo que determinó 60 posibles modelos ANFIS.

Para determinar la mejor combinación de nFP x tFP se utilizó como criterio de selección el mínimo Error Cuadrático Medio determinado en cada una de las simulaciones mensuales (entrenamientos de la red adaptable) con el mínimo de funciones de pertenencias. Cada entrenamiento de la red adaptable fue realizado considerando un número máximo de 300 épocas.

Una vez realizadas las simulaciones se obtuvo la configuración final de los modelos ANFIS, caracterizados por el número de funciones de pertenencia por entrada, tipo de funciones de inferencia por entrada y el error cuadrático medio mínimo entre las funciones de pertenencia. Estos resultados se muestran en la tabla 1 y 2.

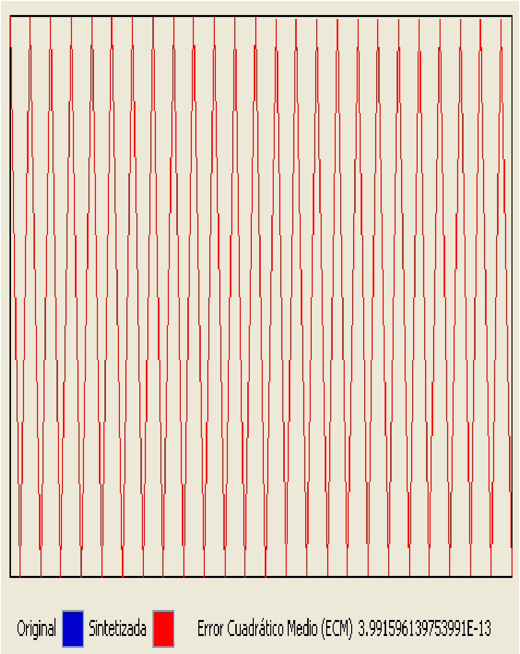

Figura 3 Sexto Segmento de la nota musical LA de una guitarra eléctrica

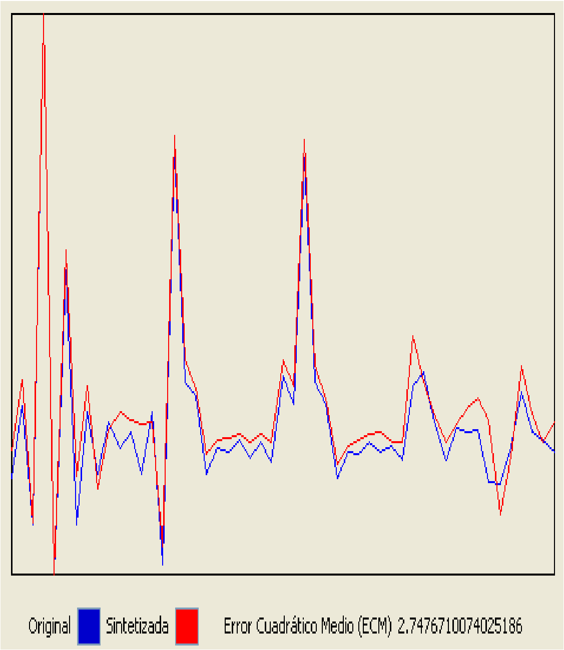

Figura 4 Sexto Segmento de la palabra Hola

nFP	tFP	ECM
25	Gaussiana	1.02 x E-14
25	Gaussiana	4.3 x E-14
25	Gaussiana	5.33 x E-14
25	Gaussiana	1.59 x E-13
25	Gaussiana	2.94 x E-13
25	Gaussiana	3.99 x E-13
25	Gaussiana	1.23 x E-12
25	Gaussiana	2.75 x E-12
25	Triangular	0.16

Tabla 1 Configuración final de los modelos ANFIS de los segmentos de la nota LA de una guitarra eléctrica con sus respectivos errores cuadráticos medios

nFP	tFP	ECM
25	Gaussiana	2.5 x E-16
25	Gaussiana	3.65 x E-16
25	Gaussiana	2.76 x E-15
24	Triangular	0.06
24	Triangular	1.42
22	Triangular	2.75

Tabla 2 Configuración final de los modelos ANFIS de los segmentos de la palabra Hola con sus respectivos errores cuadráticos medios

Finalmente el promedio del ECM para las señales es la siguiente:

$$\overline{ECM} = \sum_{i=1}^{n} ECMi / n$$

ECM de la señal Hola

$$\overline{ECM} = (2.5 \times E\text{-}16 + 3.65 \times E\text{-}16 + 2.76 \times E\text{-}15 + 0.06 + 1.42 + 2.75)/6$$

$$\overline{ECM} = 0.705$$

ECM de la señal LA

$$\overline{ECM} = (1.02 \times E\text{-}14 + 4.3 \times E\text{-}14 + 5.33 \times E\text{-}14 + 1.59 \times E\text{-}13 + 2.94 \times E\text{-}13 + 3.99 \times E\text{-}13 + 1.23 \times E\text{-}12 + 2.75 \times E\text{-}12 + 0.16)/9$$

$$\overline{ECM} = 0.017$$

ECM - ANFIS	ECM – Axel Roben
0.705	0.05

Tabla 3 Errores determinados para señales de voz

ECM - ANFIS	ECM – Axel Roben
0.017	0.205

Tabla 4 Errores determinados para notas musicales

4. Conclusiones

Se ha presentado la aplicación del modelo Neuro-Difuso ANFIS, basado en la combinación entre las redes adaptables y los sistemas de inferencia difusa. Se ajustó un modelo ANFIS para el sintetizar señales de notas musicales y de voz. Para identificar el modelo ANFIS para cada segmento de señal, se aplicó un método heurístico para identificar el número y tipo de funciones de pertenencia que mejor se ajustan a la distribución de los datos.

Se observo que en algunos casos se necesita de 26 funciones de membresías para lograr el aprendizaje con éxito, esto se debe a cada segmento de señal esta conformado por 50 puntos cardinales, y es difícil que el algoritmo converja con pocas funciones de membresía.

El modelo ANFIS aplicado en esta trabajo ha presentado un excelente rendimiento debido a que el error cuadrático medio (ECM) no es mas que 0.71

Al comparar con el método de Redes Neuronales aplicados a series de tiempos de Axel Roben se observo que el ANFIS presenta un mejor desempeño en las señales de notas musicales y en cambio en las señales de voz el método de RNA presenta resultados más óptimos.

Una de las principales características del modelo ANFIS es la rapidez con que alcanza valores aceptables de error de entrenamiento debido al empleo del método de mínimos cuadrados para determinar los parámetros de la salida del modelo de inferencia en el paso forward del algoritmo de entrenamiento, lo que aumenta significativamente el tiempo de ejecución de este tipo de modelos. Modelos de Redes Neuronales Artificiales empleados para previsión de series de tiempo generalmente necesitan alrededor de 5000 épocas para completar su entrenamiento, debido a que solo emplean el algoritmo de backpropagation para actualizar los parámetros asociados a la red que almacenan el conocimiento adquirido por el modelo (pesos sinápticos).

No se recomienda usar el algoritmo con mas de cincuenta pares cardinales ya que no podría converger debido a la cantidad de datos de entrada

5. Referencias

- [1] Jang, J-S. R., “Neurofuzzy Modeling: Architecture, Analyses and Applications”, Tesis de Doctorado, University of California, Berkeley, CA, Estados Unidos, 1992.
- [2] Jang, J-S. R., “ANFIS: Adaptive-network-based fuzzy inference system”, IEEE Transactions on Systems, Man, and Cybernetics, No.23, pp. 665-685, 1993.
- [3] Jang, J-S. R., Sun, C-T., “Neurofuzzy Modeling and Control”, Proceedings of the IEEE, 1995.
- [4] Jang, J-S. R., “Input Selection for ANFIS Learning”, Proceedings of the IEEE International Conference on Fuzzy Systems, New Orleans, 1996.
- [5] Takagi, T. y Sugeno, M., “Fuzzy identification of systems and its applications to modeling and control”, IEEE Transactionson Systems, Man, and Cybernetics, 1985, No. 15, pp. 116-132.

- [6] Axel Robel. Neural networks for modeling time series of musical instruments
- [7] J. Wesley Hines. Fuzzy and Neural Approaches in Engineering MATLAB Supplement
- [8] http://en.wikipedia.org/wiki/Neuro-fuzzy
- [9] Sistemas Neuro-Fuzzy Noemí Moya Alonso
- [10] Andrés Zúñiga. Aplicación de Redes Adaptables y Sistemas de Inferencia Fuzzy para la prevención de caudales afluentes en centrales hidroeléctricas
- [11] Riyanto, B., Febrianto, F., Machbub, C., "Adaptive-Network-Based Fuzzy Inference System for Forecasting Daily Gasoline Demand", Proceedings of the Sixth AEESEAP Triennial Conference, Bali, Indonesia, 2000