CORPORACIÓN MEXICANA DE INVESTIGACIÓN EN MATERIALES

DIVISIÓN DE ESTUDIOS DE POSGRADO

RED NEURONAL DE FUNCIÓN DE BASE RADIAL REDISEÑADA PARA MODELACIÓN Y OPTIMIZACIÓN, CASO DE ESTUDIO: PROCESOS DE SOLDADURA

POR

M.C. ROLANDO JAVIER PRAGA ALEJO

TESIS

DOCTORADO EN INGENIERÍA INDUSTRIAL Y DE MANUFACTURA

SALTILLO, COAH. ABRIL DEL 2011

RED NEURONAL DE FUNCIÓN DE BASE RADIAL REDISEÑADA PARA MODELACIÓN Y OPTIMIZACIÓN, CASO DE ESTUDIO: PROCESOS DE SOLDADURA

por

M.C. Rolando Javier Praga Alejo

Tesis

Presentada al Programa Interinstitucional en Ciencia y Tecnología

Sede

Corporación Mexicana de Investigación en Materiales S.A. de C.V.

como requisito parcial para obtener el grado académico de

Doctor en Ciencia y Tecnología
Especialidad en Ingeniería Industrial y Manufactura

Programa Interinstitucional en Ciencia y Tecnología /Comimsa

Saltillo Coahuila, Abril del 2011

RED NEURONAL DE FUNCIÓN DE BASE RADIAL REDISEÑADA PARA MODELACIÓN Y OPTIMIZACIÓN, CASO DE ESTUDIO: PROCESOS DE SOLDADURA

por

M.C. Rolando Javier Praga Alejo

Comite Revisor

Tutor Académico: Dr. Luis Martin Torres Treviño
CIIDIT UANL

Revisor No.1: Dr. Ismael López Juárez
CINVESTAV

Revisor No.2: Dr. Héctor Mancha Molinar
CIFUNSA DIESEL, S.A. DE C.V.

Revisor No.3: Dr. Mario Cantú Sifuentes
PICYT-Ing. Industrial y Manufactura-COMIMSA

Revisor No.4: Dr. Arturo Berrones Santos
CIIDIT UANL

Tutor de planta: M.C. Armando García Sánchez
Aplicación Tecnológica / Corporación Mexicana de Investigación en Materiales

Saltillo Coahuila, Abril del 2011

A MIS PADRES:

Javier y María Guadalupe

A MIS HERMANOS:

Juan de Dios y Esper

A MI FAMILIA:

A mis Abuelos, Tíos y Primos ...

A MIS AMIGOS ...

Agradecimientos

A Dios por permitirme concluir una de las metas más importantes de mi vida profesional.

A mis padres Javier Praga y María Guadalupe Alejo; y **a mis hermanos Juan de Dios y Esper** por su apoyo incondicional.

A mi familia por su atención y apoyo.

A **David González** por ser mas que un amigo.

A mi tutor, Dr. Luis Martín Torres Treviño por compartir sus conocimientos y principalmente por brindarme su amistad.

A los académicos que de forma directa e indirecta contribuyeron en mi formación **Dr. Luis Martín Torres Treviño**, **Dr. Pedro Pérez Villanueva**, **Dr. Miguel Gastón Cedillo Campos**, **Dr. Arturo Reyes**, **Dr. Ismael López Juárez**, **Dr. Federico Zertuche Luis**, **Dr. Manuel Román Piña Monarrez**; principalmente al **Dr. Mario Cantú Sifuentes** por su apoyo, conocimientos y sobretodo por sus consejos; gracias por su amistad.

Al **Dr. Héctor Mancha Molinar** por sembrar el interés (la semilla) por la investigación.

A todos mis compañeros y amigos, que conocí durante mis estudios y que siempre los llevaré presentes, especialmente: a todos.

A la **M. C. Claudia González**, coordinadora del Posgrado de COMIMSA por el apoyo brindado durante mis estudios y por creer siempre en mi.

Al Jurado: **Dr. Ismael López Juárez**, **Dr. Mario Cantú Sifuentes**, **Dr. Arturo Berrones Santos** y **Dr. Héctor Mancha Molinar**, por sus excelentes comentarios y recomendaciónes, que contribuyeron en la mejora de este trabajo de investigación.

A las personas que día a día hacen posible el desempeño del Posgrado en COMIMSA, **personal administrativo del posgrado**.

Agradezco los apoyos recibidos durante mi formación doctoral a dos importantes Instituciones del ámbito Científico y Tecnológico de este país; sin ellas mi progreso y desarrollo académico, actividades de movilidad e investigación y producción científica no hubieran sido posibles: **El Consejo Nacional en Ciencia y Tecnología** (CONACYT) y a la **Corporación Mexicana de Investigación en Materiales**.

Al tutor de planta, **M.C. Armando García Sánchez** por su apoyo durante el proyecto.

Resumen

En la actualidad las empresas se están apoyando en nuevas técnicas o métodos numéricos para mejorar los procesos, algunos de estos métodos se derivan de los sistemas inteligentes y técnicas estadísticas. Existen diferentes técnicas en sistemas inteligentes para modelar y optimizar los procesos de manufactura; una de estas técnicas es la de Redes Neuronales Artificiales (RNA). Dentro de estás herramientas existen diferentes tipos, por ejemplo la red neuronal artificial de Función de Base Radial.

La construcción de las redes de Funciones de Base Radial (FBR), en su forma básica, incluye tres capas totalmente diferentes.

Capa de entrada: está formada por los nodos fuente x_i (unidades sensoriales).

Capa intermedia: es una capa oculta de gran dimensión en la que las unidades (neuronas) son las funciones base para los datos de entrada $\varphi_i(x_i)$.

Capa de salida: tiene la responsabilidad en la red de la activación de patrones aplicados en la capa de entrada.

En este trabajo de investigación después de utilizar la red de Función de Base Radial, se encontró una dificultad importante en la red de Función de Base Radial, la cual es la forma de calcular la distancia Euclidiana dada por $\|x - t_i\|$, y la aleatoriedad en la variable t_i que son los centroides de la red neuronal, lo cual provoca inestabilidad en la desviación estándar y por consecuencia genera una predicción y una optimización inestables de la red neuronal de FBR. Esto constituye el motivo de esta investigación.

Por otra parte, las características modificadas de la red neuronal de Función de Base Radial ayudaron en este trabajo a mejorar el comportamiento de la red neuronal y comprobando la hipótesis de que la red neuronal de Función de Base Radial Rediseñada con la Distancia de Mahalanobis y el Algoritmo Genético presenta un mejor comportamiento que la red de Función de Base Radial con el método de Clustering k-medias con la distancia Euclidiana, en base a las métricas estadísticas de R^2, $R^2_{Ajustada}$, $R^2_{Predicción}$, $PRESS$ y Cuadrados Medios de los Residuales MS_{Res}. Se consideró así mismo una modificación para determinar los centroides, aplicando un Algoritmo Genético en lugar de los métodos que comúnmente se utilizan para este tipo de red neuronal. También se reemplazó la Distancia Euclidiana dada por $G\left(\|x - t_i\|\right)$ o $\|x - t_i\|^2$ por la Distancia de Mahalanobis en la determinación de la matriz G.

Aunado a esto se aplicaron las métricas R^2 y R^2_{Adj} que fueron las cantidades usadas para indicar si el modelo es adecuado para predecir y optimizar; además se aplicaron los estadísticos $PREMS$, el $PRESS$, los cuadrados medios del error MS_{Res} y el coeficiente de determinación de la predicción $R^2_{Predicción}$, para el análisis residual de la red neuronal modificada; los cuales son muy importantes ya que definen si el modelo o los modelos a analizar son adecuados.

Una de las mejorías de la red FBR Rediseñada se midió con las métricas del análisis de residuales, una de éstas es la $R^2_{Predicción}$ de la predicción, los cuadrados medios de los residuales MS_{Res} y el $PREMS$, los cuales indican que el modelo será un buen predictor si el $PREMS$ es menor que los cuadrados medios de los residuales MS_{Res}. Además se realizó el Análisis de Varianza (ANOVA) para este tipo de Red Neuronal.

Índice

1. Introducción **1**

2. Planteamiento del Problema **10**

2.1. Descripción del Problema 10

2.2. Hipótesis 15

2.2.1. Hipótesis General 16

2.3. Objetivos 16

2.3.1. Objetivo General 16

2.3.2. Objetivos Específicos 16

2.4. Justificación 17

3. Revisión bibliográfica **20**

3.1. Revisión de Literatura 20

3.2. Selección de la Problemática 23

3.2.1. Principales Hallazgos 23

3.3. Para resolver la problemática 24

4. Marco Teórico **26**

4.1. Redes Neuronales. 26

4.1.1. Funciones de Redes Neuronales Artificiales 31

4.1.2. Función de Base Radial 32

4.1.3. Estructuras de las Redes Neuronales Artificiales 39
4.2. Algoritmos Evolutivos 40
4.2.1. Algoritmos Genéticos 41
4.3. Lógica Difusa 47
4.4. Técnicas Estadísticas 49
4.4.1. Análisis de variancia (ANOVA) 50
4.4.2. Diseños de Experimentos (DOE) y Regresión Múltiple 56
4.4.3. Análisis Residual 58
4.4.4. Distancia de Mahalanobis 60

5. Metodología 63
5.1. Metodología Propuesta 63
5.1.1. Plan detallado seguido en esta investigación. 64

6. Desarrollo experimental y resultados 68
6.1. Aplicación y experimentación con la red neuronal de FBR Rediseñada. . . 70
6.1.1. Aplicación del algoritmo genético y distancia de Mahalanobis en FBR. 71
6.1.2. Análisis Residual y de Varianza. 77
6.2. Resultados y Validación en Campo. 80

7. Conclusiones 85

Bibliografía 90

A. ANEXO A 99

B. ANEXO B 102

Índice de Tablas

2.1. Variables del Proceso 13
2.2. Predicciones con $t1$ 14
2.3. Predicciones con $t2$ 14
2.4. Resultados de la comparación entre Redes Neuronales 18

4.1. Ejemplo de tabla de ANOVA 52

6.1. Variables del Primer Experimento 72
6.2. Variables del Segundo Experimento 75
6.3. Variables del Tercer Experimento 76
6.4. Resultados de la comparación entre Redes Neuronales FBR 77
6.5. Resultados del proceso Láser 79
6.6. ANOVA en FBR Rediseñada 80
6.7. Predicciones con FBR Rediseñada 82

Índice de Figuras

4.1. Estructura de una Red Neuronal 30
4.2. Cruce en un algoritmo genético 44
4.3. Mutación en un algoritmo genético 45

5.1. Esquema de la Metodología 64
5.2. Diagrama de Aplicaciones de Ingeniería Industrial 65
5.3. Diagrama de Modelación 66
5.4. Diagrama de Validación 66
5.5. Diagrama de Optimización 67

6.1. Ajuste mediante Red Neuronal 73
6.2. Ajuste mediante Regresión 73
6.3. Ajuste mediante Regresión y FBR Rediseñada 74
6.4. Ajuste de los Residuales 80
6.5. Predicción de Red Neuronal FBR Rediseñada 82
6.6. Penetración en el proceso Láser (Prueba 1) 83
6.7. Penetración en el proceso Láser (Prueba 2) 83
6.8. Penetración en el proceso Láser (Prueba 3) 83

Capítulo 1

Introducción

En la actualidad las empresas se están apoyando en nuevas técnicas y diferentes métodos numéricos para mejorar los procesos productivos; algunos de estos métodos se derivan de los sistemas inteligentes y técnicas estadísticas (Benyounis y Olabi, 2008; Paliwal y Kumar, 2009). Existen diferentes técnicas en sistemas inteligentes para modelar y optimizar los procesos industriales de manufactura; una de estas son las Redes Neuronales Artificiales (RNA) (Meziane et al., 2000). Existen diferentes tipos de RNA como la Retro propagación, Adaline, Madaline, Hopfield, Funciones de Base Radial, entre otras (Tsoukalas y Uhrig, 1998). De igual forma, en las técnicas estadísticas para mejorar procesos existen herramientas como los Diseños Experimentales y Regresiones lineal simple o múltiple, Metodología de Superficie de Respuesta, etc. (Benyounis y Olabi, 2008).

La tendencia mundial de constante cambio, así como la incorporación de nuevos materiales en la manufactura de productos y componentes, ha propiciado que se tengan procesos cada vez más eficientes. La tecnología de los procesos de soldadura ha avanzado considerablemente, sin embargo en la mayoría de los casos el ajuste de los parámetros y variables se realiza a prueba y error, siendo esta la práctica más común.

La soldadura, es una tecnología para la unión de materiales metálicos, que se remonta a la Edad de Bronce, donde se encuentran los primeros vestigios de procesos de soldadura utilizados para fines ornamentales. Por otro lado la soldadura, en ingeniería, se define

como un procedimiento por el cual dos o más piezas de metal se unen por aplicación de calor, presión, o una combinación de ambos, con o sin el aporte de otro metal (Welding handbook, 2001; Jefferson's welding encyclopedia, 1997).

Existen diferentes tipos de procesos de soldadura, uno de ellos es el proceso de soldadura Láser, el cual en los últimos años ha tenido gran auge; en el proceso, el Láser se utiliza para crear una unión que tiene una estrecha zona afectada por el calor, reduce la rugosidad de la superficie de la unión y elimina los efectos mecánicos que causan otros tipos de procesos de soldadura. Los sistemas Láser operan en modo pulsado o continuo. Las principales variables a controlar son: velocidad, ancho y frecuencia, que generan una respuesta o variable de salida que es la penetración o profundidad de la soldadura. (Cepeda, 2010).

Otro de los procesos de soldadura más utilizados es el de gas de arco metal (GMAW por sus siglas en inglés) el cual es un proceso que une los metales aplicando calor con un arco establecido entre un alambre fino de electrodo y los metales. El arco protegido y la soldadura fundida son regularmente obtenidos usando gases inertes como argón y helio, este es llamado proceso de soldadura de metal de gas inerte (MIG). Este es un proceso muy usado para aleaciones de aluminio. Gases de protección: argón, helio y sus mezclas son usados para metales no ferrosos y en aleaciones de acero. El arco de energía es disperso y menos uniforme en un arco de argón que en un arco de helio, por la baja conductividad térmica del argón (Welding handbook, 2001; Jefferson's welding encyclopedia, 1997).

Este tipo de proceso de soldadura tiene diferentes variables que se deben controlar para una mejor calidad en los productos donde se emplea. Algunas de estas variables son las siguientes: amperaje, voltaje y velocidad al soldar. Con estas variables el proceso de soldadura (GMAW) tiene una medida de penetración, que es la variable de respuesta o de salida (García-Sánchez et al., 2008; Praga-Alejo et al., 2008a Anexo B1).

En la actualidad se está trabajando para mejorar la predicción y optimización en los procesos, y sobretodo en el manejo de parámetros, variables y aplicaciones de la ingeniería

industrial; para dar un panorama general de la aplicación de los métodos matemáticos, tanto de sistemas inteligentes y herramientas estadísticas, se presentan enseguida una serie de autores y artículos que aplicaron estas técnicas.

Uno de los primeros trabajos sobre optimización en los procesos, fue el de Andrzej Sluzalec (1989) que desarrolló en su investigación para la optimización de una superficie de soldadura. La optimización se hizo mediante el método del elemento finito.

Algunas investigaciones que se realizaron después, fueron mediante trabajos de modelación y optimización, por ejmplo Olabi et al., en el 2006 propuso un Diseño Experimental Taguchi y Redes Neuronales en la optimización de soldadura Láser CO_2. Otro fue de Song et al., para la optimización del proceso de soldadura 3D mediante un diseño experimental (Song et al., 2005).

Una comparación entre Algoritmos Genéticos y Superficie de Respuesta en la optimización de un proceso de soldadura fue desarrollada por Sampaio et al., (2005). Método de elemento finito para el pronóstico de efectos en el proceso de soldadura de Lindgreen (2006).

Un análisis de sensibilidad para los parámetros del proceso GMAW usando el método de diseño experimental factorial fue realizado por Kim et al. (2003). La modelación, optimización y clasificación de la calidad del proceso de soldadura TIG fue aplicada por Tarng et al. (1999).

La optimización de soldadura Nd: YAG láser en una aleación de magnesio con análisis de diseño de experimentos con el método Taguchi fue aplicado por Pan et al. (2004). Un diseño de un proceso estadístico para un robot del proceso de soldadura GMAW fue implementado por Allen et al. (2002). Usando el método de Taguchi para determinar parámetros del proceso de soldadura de arco sumergido (SAW) en revestimientos duros fue una aplicación realizada por Tarng, et al. (2002).

Las investigaciones que se han hecho en la parte de modelación y optimización, con diferentes técnicas y metodologías fueron realizadas por ejemplo por Von stryk (1996)

aplicando la optimización mediante Dinámica de Sistemas en aplicaciones industriales. La cuantificación simultánea de análisis de mezclas mediante voltametría y genéticamente optimizadas mediante redes neuronales fue hecha por Richards et al. (2004). Un diseño experimental para un diseño robusto de parámetros fue presentado por Romero-Villafranca et al. (2007). Un diagnostico de un proceso de manufactura usando un modelo de regresión fue diseñado por Wei (2007).

Wang et al., (2010) aplicaron dos métodos, uno fue la red neuronal de Función de Base Radial y el otro método fue el Algoritmo Genético los cuales fueron utilizados para optimizar un sistema de equipo eléctrico en una fuente de poder. Otra aplicación con la red neuronal de Función de Base Radial la hicieron Lui et al., (2010) en este trabajo se aplicaron dos modelos un Método Heurístico y la red de Función de Base Radial lo hicieron para determinar las predicciones en el potencial orgánico, inorgánico y medir las propiedades de la relación.

Por otra parte hablando de estas metodologías de las cuales se derivan: la Lógica Difusa, Redes Neuronales, Algoritmos evolutivos y Sistemas híbridos; en general se puede decir que estos sistemas inteligentes tratan de hacer que las computadoras imiten el razonamiento humano para la solución de problemas, tareas o toma de decisiones, en problemas complejos (Yen y Langari, 2000, Rao et al., 1999).

Una Red Neuronal es un sistema compuesto por un número de elementos interconectados y agrupados en capas. Este sistema posee varias entradas y salidas, las cuales son entrenadas y tienen un nivel de aprendizaje; son capaces de procesar una gran cantidad de información mediante modelos matemáticos (Hagan et al., 1996, Tsoukalas y Uhrig, 1998, Nelles, 2001, Arbib, 2003, Martín del Brío y Sanz, 2007).

Existen redes neuronales que trabajan con funciones de base y de activación que consisten en que una neurona suministra un valor a su salida que se propaga a través de la red de conexiones unidireccionales hacia otras células de la red. En este sentido, la salida final se puede expresar como una función de la entrada y de los pesos (Martín del Brío y

Sanz, 2007). Esta función es de activación o transferencia donde se define la salida de la neurona. La forma de hacerlo está determinada matemáticamente por la función de base para dar un valor de activación. Las funciones de activación más comunes son:

- Función escalón.
- Función sigmoidal.
- Función gaussiana.

Existen la Función lineal de base y la Función de Base Radial.

- **Función de Base Radial**

Es de segundo orden o de tipo hiperesférico. El valor de red representa la distancia a un determinado patrón de referencia. Las funciones de base radial o *Radial Basis Functions (RBF por sus siglas en inglés)* son un modelo híbrido ya que utiliza tanto el aprendizaje supervisado como el no supervisado. La construcción de las redes de Funciones de Base Radial (FBR), en su forma básica, incluye tres capas diferentes (Haykin, 1999).

Capa de entrada: está formada por los nodos fuente x_i (unidades sensoriales).

Capa intermedia: es una capa oculta de gran dimensión y en la que las unidades (neuronas) que la forman son las funciones base para los datos de entrada $\varphi_i(x_i)$.

Capa de salida: d_j que tiene la responsabilidad en la red de la activación de patrones aplicados en la capa de entrada.

Uno de los principales problemas y dificultades de las redes neuronales artificiales, por ejemplo en el caso de la red Retro propagación (Backpropagation en inglés) se encuentran en:

- La velocidad de aprendizaje.
- El problema de su lenta convergencia.

- Existe sobreajuste (sobre entrenamiento).
- No alcanza mínimo global, solo un mínimo local.

De aquí surge la siguiente pregunta: ¿Cómo mejorar o solucionar este problema?. Una alternativa es utilizar técnicas y herramientas estadísticas en conjunto con los sistemas inteligentes, para mejorar las predicciones y la optimización de la red neuronal artificial.

En la actualidad también se están realizando estos diferentes tipos de trabajos en áreas y procesos usando la red neuronal de base radial para predicciones principalmente; una descripción completa, detallada y se profundiza sobre la estructura de la red neuronal de función de base radial, es en Simon Haykin (1999), donde se hace una descripción completa del funcionamiento de la red neuronal, así como la demostración matemática de cada una de las formulas utilizadas en cada uno de los pasos de la elaboración y funcionamiento de la red neuronal artificial de función de base radial.

Algunos detalles del funcionamiento de la red neuronal de base radial, se puede verificar en el libro de Oliver Nelles (2001); se hace un análisis de la estructura interna de la red neuronal y sobre todo se profundiza y se encuentran los métodos para calcular los centroides, así como sus ventajas y desventajas de la red neuronal. También se hace énfasis en la estructura matemática aplicada a cada una de las formulas utilizadas en la red neuronal de función de base radial.

Otra descripción de la red neuronal se puedde ver en el libro de Martín del Brío y Sanz (2007), donde se explica los pasos aplicados en la estructura de la red y su funcionamiento. En el Handbook de Arbib (2003) se encuentra la parte teórica de la estructura de la red y su forma básica, así como algunas de sus formulas y aplicaciones.

Por otro lado, después de utilizar la red de Función de Base Radial, se encontraron principalmente dos dificultades; la primera es la forma de calcular la distancia Euclidiana dada por $\|x - t_i\|$, y la segunda es la aleatoriedad en la variable t_i que son los centroides de la red neuronal, lo cual provoca la inestabilidad de la desviación estándar predicha

y por consecuencia dará una predicción y optimización inestable de la red neuronal de función de base radial.

Una de las más interesantes alternativas para resolver estos problemas, los son los recientes desarrollos en Estrategias Evolutivas o Algoritmos Evolutivos. Existen también métodos estadísticos para mejorar los modelos de redes neuronales artificiales, estos métodos han sido exitosamente usados para la selección óptima de parámetros o estructura de las redes neuronales de Función de Base Radial.

Uno de los primeros trabajos en este tipo de aplicaciones fue realizado por Billings y Zheng (1995). Con el fin de superar el problema planteado con la red neuronal de base radial, se han aplicado varios métodos para mejorar los parámetros deseados y las variables mediante el desarrollo de modelos con sistemas inteligentes como lo vemos en Meqin y Jida (1998).

Otro trabajo similar fue realizado por Whitehead y Choate (1995). Por otro lado, Ding et al. (2005) propuso otro Algoritmo Evolutivo en este caso PSO para mejorar la red de Base Radial. Otra investigación mediante una optimización fue hecho por Sun et al. (2006). Finalmente otra aplicación de estos algoritmos se encuentra en el trabajo realizado por Valls et al. (2005) donde aplican la distancia de Mahalanobis utilizando un Algoritmo Genético para encontrar la matriz de varianza y covarianza; este trabajo mejora la precisión de la predicción con matrices desarrolladas simétricamente.

Con todo esto y después de haber encontrado los hallazgos y problemática que presenta la red neuronal de Función de Base Radial (para mas detalles ver el Capítulo 2, 3 y 6), las modificaciones que se realizaron para mejorar el comportamiento de la red neuronal, fueron adecuadas y proporcionaron un mejor modelo de red neuronal. Las herramientas utilizadas, tanto de sistemas inteligentes como técnicas estadísticas (Algoritmos Genéticos y Distancia de Mahalanobis) modifican la red neuronal de Función de Base Radial y lleva a reemplazar los métodos que se aplicaban anteriormente.

Por otra parte aunado a estás modificaciones también se desarrolló todo el análi-

sis y validación de los modelos mediante técnicas estadísticas, por ejemplo, Análisis de Residuales, Análisis de Varianza (ANOVA) entre otras métricas estadísticas, las cuales se mencionan en el capítulo 4 sección 4.4.3. Además se desarrolló experimentación en campo y se logró realizar un diseño central compuesto de un proceso de soldadura que en este caso fue de tipo Láser.

En la presente investigación todo esto se analiza y estructura como parte complementaria de la metodología creada para modelar los procesos de soldadura. Básicamente la metodología tiene los siguientes pasos:

- Estudio del estado del arte en aplicación de los sistemas inteligentes y técnicas estadísticas que se utilizan en los procesos de soldadura.
- Descripción y caracterización del proceso de soldadura y materiales involucrados.
- Desarrollo de una etapa experimental.
- Modelación y cuando es necesario simulación del proceso del cual se esta analizando.
- Validación de los modelos con análisis de varianza y análisis residual.
- Optimización del proceso considerando diversos esquemas como por ejemplo global o multiobjetivo en caso de ser necesario.
- Transferencia de los resultados al proceso de soldadura.

La investigación y estructura del trabajo tiene los siguientes Capítulos; En el capítulo dos se presenta la descripción del problema, los objetivos generales y específicos, las hipótesis, las ventajas y desventajas de la solución del problema y resultados esperados. El Capítulo tres se desarrolla el estado del arte con referente a la red neuronal artificial de Función de Base Radial, su utilización y sus principales hallazgos, para mejorar y modificar este modelo, del cual se desarrolla la presente disertación.

El Capítulo cuatro se hace una introducción al marco de referencia y conceptos de los temas que se utilizaron en la disertación. El Capítulo cinco se plantea la metodología para analizar, modelar y optimizar un proceso, que en este caso son procesos de soldadura; mediante modelos híbridos de sistemas inteligentes combinación con métodos estadísticos.

En el Capítulo seis se muestra la experimentación, aplicación y resultados que se obtuvieron, modificando la red neuronal de Función de Base Radial y donde se aplicó la metodología propuesta. Por último se presenta el Capítulo siete dando las conclusiones de la investigación y comentarios en cuanto al trabajo futuro.

Capítulo 2

Planteamiento del Problema

En este capítulo se presentará la descripción del problema, los objetivos generales y específicos, la hipótesis de la disertación, las ventajas y desventajas de la solución del problema y los resultados esperados.

2.1. Descripción del Problema

Una red neuronal, esta compuesta por: Un vector de n entradas, que tienen una estructura de pesos sinápticos que son utilizados para aumentar o disminuir los valores de entrada. Se compone también de un umbral o bias, para reducir la entrada a la función de activación (Arbib, 2003). Y por último existe la función de activación o transferencia donde se define la salida de la neurona.

Descripción de las redes neuronales del tipo Función de Base Radial: Es una función de segundo orden o de tipo hiperesférico. El valor de la red es la distancia a un determinado patrón de referencia. Las funciones de base radial, o *Radial Basis Functions (RBF por sus siglas en inglés)* es un modelo híbrido ya que utiliza tanto el aprendizaje supervisado como el no supervisado. La construcción de las redes neuronales de Funciones de Base Radial (FBR), en su forma básica, incluye tres capas diferentes (Haykin, 1999).

- Capa de entrada: está formada por los nodos fuente x_i (unidades sensoriales).

- Capa intermedia: es una capa oculta de gran dimensión y en la que las unidades (neuronas) que la forman son las funciones base para los datos de entrada $\varphi_i(x_i)$.
- Capa de salida: d_j tiene la responsabilidad de la activación de patrones aplicados en la capa de entrada.

Las neuronas de la capa de salida son lineales. Las neuronas de la capa oculta calculan la diferencia entre el vector de entradas y los denominados centroides, a esta diferencia se le aplica por lo general una función radial en forma gaussiana. La función de transferencia, radial, de tipo gaussiana adopta la siguiente forma:

$$G(\|x - t_i\|) = \exp\left(-\|x - t_i\|^2\right) \tag{2.1}$$

La cual se simplifica con la ecuación (2.2):

$$\varphi_i(x) = G(\|x - t_i\|) \tag{2.2}$$

Donde x son las variables de entradas y t_i son los centros o centroides.

Cuando la distancia entre t y x decrece, la función crece. Su máximo sucede cuando $t - x = 0$, momento en el cual la función devuelve 1. Cuando se presenta un vector de entrada, cada neurona de la capa oculta, dará una u otra salida dependiendo de lo cerca que está el vector de entrada de los pesos sinápticos.

La entrada de cada neurona de la capa oculta es la distancia entre el vector de entradas y su vector de pesos. Esta distancia será multiplicada por el vector de los umbrales. El resultado se le aplica a la función radial. La relación de entradas y salidas de la red neuronal esta dada por (2.3):

$$y(x) = \sum_{i=1}^{m} wG(\|x - t_i\|) + b \tag{2.3}$$

Donde $y(x) = d_j$; Por lo tanto d_j es la salida o respuesta del proceso, $d = [d_1, d_2, \ldots, d_N]^T$, las ecuaciones son escritas en forma matricial:

$$Gw = d \tag{2.4}$$

Donde $w = [w_1, w_2, \ldots, w_m]^T$ son los pesos sinápticos de la red neuronal artificial y se determinan con (2.5), los cuales se determinan mediante los métodos de Pseudoinversa y Mínimos Cuadrados:

$$\begin{aligned} w &= G^+ d \\ &= (G'G)^{-1} G'd \end{aligned} \tag{2.5}$$

La matriz G está dada por (2.6):

$$G = \begin{bmatrix} G(x_1 - t_1) & G(x_1 - t_2) & G(x_1 - t_m) \\ G(x_2 - t_1) & G(x_2 - t_2) & G(x_2 - t_m) \\ \vdots & \vdots & \vdots \\ G(x_N - t_1) & G(x_N - t_2) & G(x_N - t_m) \end{bmatrix} \tag{2.6}$$

La red neuronal FBR dada por $G(\|x - t_i\|)$ o $\|x - t_i\|^2$, que es llamada distancia Euclidiana, la cual esta compuesta por el vector de entrada x y t_i que son los centros (Haykin, 1999), los cuales se encuentran mediante el método Clustering K-medias, mediante simulación, con redes neuronales o en forma aleatoria. Estos métodos dan un poco de inestabilidad en la red, como se muestra a continuación en el siguiente ejemplo:

La Tabla 2.1 muestra los resultados de un proceso de soldadura con cuatro variables de entrada y una salida (Gunuraj y Murugan, 2000), usando una FBR, con centroides definidos en forma aleatoria. Reportados en la matriz $t1$ con los niveles y rangos que se presentan en la Tabla 2.1.

$$t1 = \begin{bmatrix} 32 & 1.62 & 0.75 & 40 \\ 31.3066 & 1.5403 & 0.7223 & 39.1334 \\ 25.2190 & 0.8402 & 0.4788 & 31.5238 \\ 24 & 0.7 & 0.43 & 30 \end{bmatrix}$$

Otra estimación de los centroides para los datos de la Tabla 2.1 con centroides definidos en forma aleatoria se reporta en la matriz $t2$.

$$t2 = \begin{bmatrix} 32 & 1.62 & 0.75 & 40 \\ 28.2227 & 1.1856 & 0.5989 & 35.2784 \\ 27.2927 & 1.0786 & 0.5617 & 34.1159 \\ 24 & 0.7 & 0.43 & 30 \end{bmatrix}$$

Tabla 2.1: Variables del Proceso

	Factores				Respuesta
Número de Corrida	Voltaje (Volts)	Velocidad de Alambre (m/min)	Velocidad al Soldar (m/min)	Distancia hacia la pieza (mm)	Penetración (mm)
1	26	0.93	0.51	32.5	3.52
2	30	0.93	0.51	32.5	3.4
3	26	1.39	0.51	32.5	4.75
4	30	1.39	0.51	32.5	4.1
5	26	0.93	0.67	32.5	3.25
6	30	0.93	0.67	32.5	3.18
7	26	1.39	0.67	32.5	3.52
8	30	1.39	0.67	32.5	3.33
9	26	0.93	0.51	37.5	3.85
10	30	0.93	0.51	37.5	3.6
11	26	1.39	0.51	37.5	4.1
12	30	1.39	0.51	37.5	3.8
13	26	0.93	0.67	37.5	3.2
14	30	0.93	0.67	37.5	3
15	26	1.39	0.67	37.5	4.1
16	30	1.39	0.67	37.5	3.88
17	24	1.16	0.59	35	4.1
18	32	1.16	0.59	35	3.75
19	28	0.7	0.59	35	3.26
20	28	1.62	0.59	35	4.97
21	28	1.16	0.43	35	4.25
22	28	1.16	0.75	35	3.48
23	28	1.16	0.59	30	3.82
24	28	1.16	0.59	40	3.58
25	28	1.16	0.59	35	3.45
26	28	1.16	0.59	35	3.47
27	28	1.16	0.59	35	3.66
28	28	1.16	0.59	35	3.6
29	28	1.16	0.59	35	3.3
30	28	1.16	0.59	35	3.6
31	28	1.16	0.59	35	3.92

La Tabla 2.2, muestra las predicciones y la desviación estándar de la predicción con la matriz $t1$. La Tabla 2.3, muestra las predicciones y la desviación estándar de la predicción con la matriz $t2$.

Esta aleatoriedad provoca una diferencia de la desviación estándar en la predicción, que depende de los valores aleatorios de los centroides t_i.

Tabla 2.2: Predicciones con $t1$

Predicción				
Número de Corrida	Penetración (mm)	Número de Corrida	Penetración (mm)	Desviación estándar
1	3.37	17	3.72	0.163196
2	3.72	18	3.72	
3	4.07	19	3.72	
4	3.72	20	3.72	
5	3.49	21	3.72	
6	3.72	22	3.72	
7	4.12	23	3.7	
8	3.72	24	3.74	
9	3.72	25	3.72	
10	3.3	26	3.72	
11	3.72	27	3.72	
12	3.78	28	3.72	
13	3.72	29	3.72	
14	3.33	30	3.72	
15	3.72	31	3.72	
16	3.72			

Tabla 2.3: Predicciones con $t2$

Predicción				
Número de Corrida	Penetración (mm)	Número de Corrida	Penetración (mm)	Desviación estándar
1	3.75	17	3.73	0.1857464
2	3.73	18	3.73	
3	3.62	19	3	
4	3.73	20	4.39	
5	3.73	21	3.61	
6	3.73	22	3.81	
7	3.61	23	3.73	
8	3.73	24	3.73	
9	3.73	25	3.71	
10	3.65	26	3.71	
11	3.73	27	3.71	
12	3.61	28	3.71	
13	3.73	29	3.71	
14	3.65	30	3.71	
15	3.73	31	3.71	
16	3.6			

Debido a esta deficiencia mostrada y la inferencia que tienen los centroides sobre la estimación de los datos, de la matriz G, (ec. (2.6)) y la importancia en los valores de w (ec. (2.5)) llamados pesos de la red, la presente investigación se enfoca principalmente a buscar una solución a esta problemática; se considera que esta estimación es determinante para la realización de predicciones y optimización de un proceso, en el cual se busca la reducción de costos.

Esta problemática ha sido evidenciada por diferentes autores Billings y Zheng (1995) en uno de los primeros trabajos en este tipo de problemas hicieron un entrenamiento de la red FBR, basado en un conjunto de datos prácticos, usados para demostrar el comportamiento de un Algoritmo Genético en la estructura de la red neuronal de FBR.

Otro aportación con respecto a este tipo de problemas fue desarrollada mediante una optimización con Particle Swarm Optimization (PSO) la cual fue desarrollada por Sun et al. (2006) en este caso se aplicó una operación de Quantum para mejorar los parámetros de la red neuronal FBR.

Después de haber revisado las aportaciones de los autores para mejorar la red FBR, se presentan tres posibles enfoques:

1.- Encontrar la forma de sustituir la Distancia Euclidiana y/o centroides, para tener una mejor predicción y optimización de la red neuronal FBR.

2.- Verificar si la modificación de la distancia euclidiana y/o centroides, no provoca inestabilidad en los pasos subsecuentes de la red, por ejemplo en la estimación de los pesos dados por w, ecuación (2.5).

3.- Analizar la matriz G dada por (2.6) y el procedimiento de obtener los pesos w, ya que esta matriz podría presentar problemas en la predicción y optimización de los datos.

2.2. Hipótesis

A continuación se plantea la hipótesis de la disertación, la cual se tiene que demostrar con las técnicas y metodologías propuestas en esta investigación.

2.2.1. Hipótesis General

La red neuronal de FBR Rediseñada con la Distancia de Mahalanobis y el Algoritmo Genético es mejor que el modelo de red FBR (Clustering k-medias y distancia Euclidiana), con respecto a las métricas estadísticas de R^2, $R^2_{Ajustada}$, $R^2_{Predicción}$, $PRESS$ y Cuadrados Medios de los Residuales MS_{Res}.

2.3. Objetivos

En la presente disertación se plantean los siguientes objetivos tanto generales como específicos, con el fin de hacer mejores modelos, predicciones y optimizaciones de los procesos que se están analizando.

2.3.1. Objetivo General

Desarrollar un modelo de red neuronal para la modelación y optimización de procesos de soldadura, mediante la utilización de sistemas inteligentes y herramientas estadísticas.

2.3.2. Objetivos Específicos

1. Utilizar metodologías estadísticas para la modelación y validación del modelo generado.

2. Mejorar o cambiar la forma de encontrar la Distancia Euclidiana, para disminuir el problema de la aleatoriedad que provoca la estructura de la red neuronal, para dar una mejor predicción y optimización de la red neuronal de FBR.

3. Mejorar la estimación de los pesos w dados por (2.5), mediante modelos estadísticos y sistemas inteligentes.

4. Verificar los efectos de las modificaciones de la red de FBR, generadas por las distancias y centroides.

5. Crear la metodología para la transferencia del modelo y de los resultados de la predicción y optimización.

2.4. Justificación

Los procesos de soldadura son de gran importancia y de extenso uso en diversos centros de manufactura, sobre todo en las empresas de la región. La tendencia mundial de constante cambio y el aumento de la personalización de los productos, ha propiciado que se tengan procesos cada vez más robustos, así como la incorporación de nuevos materiales en la manufactura de productos y componentes.

En los procesos de soldadura que no son la excepción, se ha buscado el desarrollo de metodologías para modelarlos, predecirlos y optimizarlos. Por este motivo en estos procesos se están aplicando nuevas tecnologías o métodos numéricos, para disminuir costos en sus procesos y hacerlos mas robustos y productivos.

Sin embargo, prueba y error sigue siendo la práctica más común para tratar de hacerlos mas robustos. La aplicación del método de prueba y error en los procesos de soldadura para el ajuste de los parámetros y variables que se involucran, implica pérdidas en material que elevan los costos, tiempo, entre otras consideraciones. (Hamill, 1993; Tarng et al. 1999; Benyounis y Olabi, 2008).

Inclusive las corridas de simulación se realizan también a prueba y error. Por lo tanto una de estas alternativas tecnológicas o métodos para mejorar los procesos, es el uso de los sistemas inteligentes y herramientas estadísticas (Tarng et al. 1999; Tarng et al. 2002; Benyounis y Olabi, 2008).

Existen diferentes técnicas en sistemas inteligentes para modelar y optimizar los procesos industriales; una de estas técnicas es la aplicación de las Redes Neuronales Artificiales (RNA) (Meziane et al., 2000). Dentro de estás herramientas existen diferentes tipos, como la red neuronal artificial Retro propagación, Adaline, Madaline, Hopfield, Funciones de Base Radial, entre otras (Tsoukalas y Uhrig, 1998).

Ya que las redes neuronales han sido utilizadas en muchas situaciones de predicción y optimización en diferentes procesos industriales y han demostrado ser herramientas útiles para el modelado, se consideró interesante realizar un estudio comparativo entre diferentes redes neuronales, para predecir procesos de soldadura (se presenta en el paper Praga-Alejo et al., 2008b Anexo B2).

En este articulo se utilizó una red neuronal multicapa perceptrón Retro propagación, una red neuronal de Máxima Sensibilidad y la red neuronal de FBR y la validación se hizo en base a el análisis de los resultados con métodos estadísticos y la predicción de los modelos de redes neuronales.

La Tabla 2.4 muestra los resultados obtenidos para la comparación. Se puede ver que la FBR tiene un mejor comportamiento.

Tabla 2.4: Resultados de la comparación entre Redes Neuronales

	Evaluación 1		
	Red Neuronal		
Método Estadístico	Retro-Propagación	Máxima Sensibilidad	Función de Base Radial
$\bar{\hat{y}}$	3.43	3.43	3.43
MS_{Res}	0.422	0.3019	0.2792
$PREMS$	0.342	0.2444	0.226
Evaluación 2			
$\bar{\hat{y}}$	3.94	4.01	3.98
MS_{Res}	0.666	0.676	0.5816
$PREMS$	0.533	0.5408	0.4653
Evaluación 3			
$\bar{\hat{y}}$	2.86	2.81	2.86
MS_{Res}	0.082	0.0945	0.0545
$PREMS$	0.07	0.0805	0.0464

Los resultados indican que los métodos estadísticos pueden ser un buen método alternativo para validar la eficacia de los modelos de redes neuronales (para más detalles véase en Praga-Alejo et al., 2008b Anexo B2).

La red neuronal que permitió un ajuste con menos variación fue la FBR que no muestra demasiado sesgo con respecto a la media de las respuestas originales. Como conclusión en este caso, los métodos estadísticos pueden ser un buen método alternativo para validar la eficacia de los modelos de redes neuronales, la red neuronal de FBR es un buen método

para predicción y las otras RNA pueden ser un método alternativo para este propósito (Praga-Alejo et al., 2008b Anexo B2).

Por tanto en base a la conclusión de este estudio comparativo se escogió el modelo de red neuronal de FBR para predecir y optimizar los procesos de soldadura.

Capítulo 3

Revisión bibliográfica

En este capítulo se presenta el estado del arte referente a la utilización y las principales limitaciones de la red neuronal artificial de Función de Base Radial.

3.1. Revisión de Literatura

Las funciones de base radial (RBF por sus siglas en inglés Radial Basis Functions) son un modelo híbrido ya que utiliza tanto el aprendizaje supervisado como el no supervisado (Haykin, 1999). Una descripción completa y detallada donde se profundiza sobre la estructura de la red neuronal de función de base radial, se puede observar en el libro de Simon Haykin (1999), donde se hace una descripción del funcionamiento de la red neuronal, así como la demostración matemática de las formulas utilizadas en la elaboración de una red neuronal artificial de función de base radial.

Algunos detalles del funcionamiento de la red neuronal de base radial, se pueden verificar en el libro de Oliver Nelles (2001) donde se hace un análisis de la estructura interna de la red neuronal y se presentan los métodos para calcular los centroides, así como sus ventajas y desventajas de la red neuronal.

Otra descripción de la red neuronal se puede encontrar en el libro de Martín del Brío y Sanz (2007), donde se explican la estructura de la red y su funcionamiento. En el Handbook de Arbib (2003) se puede observar la parte teórica de la estructura de la red y

su forma básica, así como algunas de sus formulas y aplicaciones.

Una pequeña descripción del funcionamiento de este tipo de red se puede encontrar en las redes neuronales supervisadas de García (2005).

En la actualidad se están aplicando de manera predictiva redes neuronales a diferentes tipos de procesos, por ejemplo en el campo agrícola con un modelo econométrico de exportación de tomate (Pérez et al., 2007).

Otro trabajo realizado es el de Kenné et al. (2006) donde se utilizó una red neuronal de base radial para predecir la resistencia de una pieza llamada rotor y la velocidad de inducción de un motor.

Navarro-Esbrí et al. (2007) aplicaron la red de base radial con modelación Kernel para modelar y predecir procesos de compresión de refrigeración.

Un trabajo para modelar las condiciones de aire en sistemas de tecnología de almacenamiento termal fue realizado por Yao et al. (2006); este tipo de modelación se realizó para obtener un pronóstico mediante la red neuronal de base radial.

En otra investigación realizada por Gudur y Dixit (2008), se desarrolló una red neuronal de base radial entrenada con datos obtenidos por la técnica del elemento finito y se usó para predecir un proceso de rolado. Los resultados fueron muy adecuados y con un tiempo computacional menor con respecto al elemento finito sin la utilización de la red neuronal.

Como se observa hay diferentes tipos de trabajos con la red neuronal de base radial, como se muestra en el articulo realizado por Park et al. (2008), donde se utilizó este tipo de red neuronal complementada con programación heurística, para controlar un generador sincronizado; esta red es utilizada para evitar posible inestabilidad cuando se encuentra este tipo de controlador en diferentes condiciones.

Esta red neuronal de base radial también se a utilizado en trabajos para sistemas de control de tracción para vehículos como se ve en Matusko et al. (2008) para estimar la fuerza de fricción en casos donde hay cambios importantes de los parámetros de la fricción.

Este tipo de modelos también se pueden utilizar para respuestas múltiples por ejemplo lo realizado por Du et al. (2010) usando un algoritmo recursivo para encontrar mejores centroides utilizando mínimos cuadrados ortogonales; la forma de encontrarlos fue reducir el error de la traza de la matriz de covarianza.

Abdalla y Hawileh (2010) hicieron una aplicación con la red de FBR en un modelo para la vida de fatiga en bajo número de ciclos de barras de acero; en este trabajo los autores mostraron que la red de FBR proporciona una buena predicción para dicho modelo.

Otro ejemplo sobre la aplicación de la red de FBR se puede encontrar en Chiddarwar y Babu (2010) donde los autores aplican varias redes neuronales artificiales para ver las relaciones cinemáticas en el problema de cinemática inversa de los robots; los autores determinaron que el modelo de red de FBR es un buen predictor para este propósito.

Wang et al. (2010) aplicaron dos métodos: la red de FBR y un algoritmo genético, para optimizar un sistema de pulverización para los consumidores de energía eléctrica y de equipos en plantas de energía. De acuerdo con este trabajo las condiciones de operación óptimas se determinaron con la red de FBR y el algoritmo genético. Otra aplicación con la red neuronal de Función de Base Radial fue realizada por Liu et al. (2010); en este trabajo se usaron dos modelos: un método heurístico y la red de Función de Base Radial para determinar las predicciones del potencial orgánico e inorgánico y medir las propiedades de la relación.

Por último se menciona la importante investigación realizada en el articulo propuesto por Gregorcic y Lightbody (2008) donde se hace una descripción completa del desarrollo de la red neuronal de función de base radial, presentando la estructura y funcionamiento de la red; es decir el cálculo de las distancias euclidianas, la elección de los centros, el cálculo de los pesos, utilización de las funciones Gaussianas, etc. Este trabajo describe tanto el proceso Gaussiano y su forma gráfica, el modelo local de la red y su aplicación en el proceso Gaussiano.

3.2. Selección de la Problemática

Para realizar una predicción mediante la red neuronal de función de base radial se utiliza la matriz t dentro de la estructura de la red neuronal que contiene los centroides, la cual se encuentran mediante el método Clustering (K-medias), mediante simulación, se encuentran también con redes neuronales o en forma aleatoria.

La manera aleatoria de encontrar los centroides dados por t y la distancia Euclidiana entre el vector de entrada x y los centroides afecta la desviación estándar en la predicción dando como resultado un incremento o disminución en la desviación estándar predicha, según los datos aleatorios de t_i generados por la red neuronal de función de base radial.

Esta deficiencia fue evidenciada en la sección 2.1 del Capítulo 2; dada la influencia que tiene la variable t_i sobre la estimación de valores w (pesos) y la matriz G, dadas por las ecuaciones (2.5) y (2.6), la presente investigación se enfocó principalmente al análisis y solución de esta problemática.

3.2.1. Principales Hallazgos

Cuando la distancia entre los centroides t y el vector de entrada x decrecen, la función crece. El valor máximo ocurre cuando $t-x=0$, momento en el cual la función devuelve 1. Cuando se presenta un vector de entrada, cada neurona de la capa oculta, con funciones de transferencia de base radial, dará una u otra salida dependiendo de lo cerca que está el vector de entrada de los pesos sinápticos.

La forma de calcular la distancia euclidiana dada por $\|x-t_i\|$ y la aleatoriedad en la variable t_i provocan la inestabilidad de la desviación estándar predicha y por consecuencia se tendrán predicciones inestables de la red neuronal de función de base radial. Como se observa en la sección 2.1 del Capítulo 2 la modificación de la distancia euclidiana $\|x-t_i\|$ provoca inestabilidad en los pasos subsecuentes de la red, por ejemplo en la estimación de los pesos de w dados por (2.5).

La matriz G dada por (2.6) y el procedimiento de obtener los pesos w dados por (2.5),

podrían introducir problemas en la predicción y optimización.

3.3. Para resolver la problemática

Una de las interesantes alternativas para resolver este complicado problema puede ser el uso de los recientes desarrollos en algoritmos evolutivos u otros algoritmos inspirados en la naturaleza, por ejemplo los algoritmos genéticos, Particle Swarm Optimization (PSO), entre otros; existen también métodos estadísticos para mejorar los modelos de redes neuronales artificiales; estos métodos también han sido exitosamente usados para la selección óptima de parámetros o la estructura de las redes neuronales de FBR.

En la actualidad se esta trabajando para mejorar los parámetros y variables que están asociadas con el proceso de aprendizaje de las redes de FBR. Enseguida se muestra uno de los primeros trabajos en este tipo de aplicaciones el cual fue realizado por Billings y Zheng (1995) donde los autores hicieron un entrenamiento de la red de FBR, basado en un conjunto de datos experimentales, los cuales fueron usados para demostrar el comportamiento de un algoritmo genético en la estructura de la red neuronal de FBR. El algoritmo genético fue usado para encontrar las neuronas de capa intermedia y los centros en la red neuronal.

Con el fin de superar el problema planteado con la red neuronal de base radial se han aplicado varios métodos para mejorar los parámetros deseados y las variables mediante el desarrollo de modelos con sistemas inteligentes como lo presenta Meqin y Jida (1998) que utilizan una técnica que combina un algoritmo genético y una red de FBR, para determinar los centroides y los anchos en la estructura de red neuronal.

Otro trabajo similar fue el realizado por Whitehead y Choate (1995), los cuales hicieron un estudio teórico del mecanismo de nicho y lo aplicaron al rendimiento de una red de FBR.

Por otra parte, Whitehead (1996) también aplica un método similar con cobertura mediante nichos ortogonales para predecir mejor el rendimiento en series de tiempo caóticas

aplicando la red neuronal de base radial.

Por otro lado, Ding et al. (2005) propusieron otro algoritmo evolutivo en este caso Particle Swarm Optimization que se utilizó para encontrar las neuronas de capas ocultas por medio de la modificación de los centros en la red de FBR; se hizo una comparación de la significancía de los errores al cuadrado.

Otro investigación mediante Particle Swarm Optimization (PSO) fue hecho por Sun et al. (2006); en este caso se aplicó una operación de Quantum para mejorar los parámetros de la red neuronal.

Finalmente vemos otra aplicación en la cual encontramos el trabajo realizado por Valls et al. (2005) donde los autores aplican la distancia de Mahalanobis y un Algoritmo Genético para encontrar la matriz de varianza y covarianza; este trabajo mejora la precisión de la predicción con el uso de matrices desarrolladas simétricamente, elaboradas mediante el algoritmo genético.

Capítulo 4

Marco Teórico

En el presente capítulo se hace una introducción al marco de referencia y conceptos de los temas que se abordarán y utilizarán en la disertación.

4.1. Redes Neuronales.

Las redes neuronales artificiales evolucionaron al observar que el cerebro humano procesa la información en una forma especialmente diferente de la de una computadora digital común. Las neuronas son los elementos arquitectónicos y los módulos de encausar la información básica del cerebro. Por ejemplo, el cerebro humano que se comporta normalmente tiene una gran cantidad de neuronas, agrupadas en una estructura demasiada compleja, no lineal y paralela (esto es aproximadamente más de 10 billones de neuronas y aproximadamente más de 60 trillones del proceso de sinapsis o conexiones entre las neuronas). Por consecuencia, el cerebro humano tiene una arquitectura muy eficaz para el análisis de información, el aprendizaje y el razonamiento.

Una red neuronal artificial es una estructura elaborada para resolver cierta clase de dificultades, tratando de emular la manera en que lo hace el cerebro de los humanos. El diseño o la estructura general de una red neuronal artificial es utilizada como un modelo o técnica, que con frecuencia se usa para modelar datos multidimensionales, complejos y no lineales. Normalmente, la mayor parte de las redes neuronales artificiales se usan

para resolver problemas de predicción o de pronóstico de algún sistema, proceso o de alguna serie de datos, en oposición a la estructura formal de métodos o al desarrollo de conocimientos básicos del cómo trabaja el sistema o proceso.

Las redes neuronales artificiales de diferentes capas y de movimiento directo se podrían llamar modelos estadísticos multivariables, utilizados para relacionar p variables de predicción que se pueden mostrar de la siguiente manera, $x_1, x_2, \ldots, x_p$, con q, variables de salida, $y_1, y_2, \ldots, y_q$.

Este tipo de modelos tienen varias capas, cada una estructurada por las variables reales u otras variables encontradas dentro del sistema. En la arquitectura más común de las redes neuronales intervienen y se conectan en tres capas: las entradas, que son las variables de entrada para la predicción, la capa oculta, diseñada por un conjunto de variables transformadas, y la capa de salida, conformada por las respuestas o salidas del sistema. Cada variable en las capas se le conoce como nodo dentro de la estructura de la red neuronal. A esta estructura y conjunción de nodos se le puede llamar una red neuronal artificial, de tres capas, que es la más utilizada.

Un nodo se apropia de un valor de entrada mediante una combinación lineal ponderada de las salidas de los nodos de la capa inferior del nodo. Después envía, una salida, que se convierte en una de las entradas para uno o más nodos en la capa siguiente. Las funciones que realizan las transformaciones principalmente son funciones sigmoidales (en forma de S) o lineales, y se les llama comúnmente funciones de activación o funciones de transferencia. Por ejemplo en la siguiente ecuación cada uno de los k nodos a_u, de la capa oculta es una combinación lineal de las variables de entrada es: $a_u = \sum_{j=1}^{p} w_{1ju} x_j + \theta_u$.

Donde los w_{1ju} son parámetros que se desconocen que deben ser estimados, estos parámetros son llamados pesos y el θ_u, es un parámetro que juega el papel de ordenada al origen en la regresión lineal este parámetro se llama, a veces, el nodo de sesgo, que también se puede encontrar en la literatura como el bias.

Cada nodo se transforma por una función $g()$. En la mayoría de la literatura sobre las

redes neuronales artificiales representa a esas funciones de activación en la forma $\sigma()$, por su forma sigmoidal, esta opción de notación no es muy recomendable ni adecuada para las personas que se dedican a la estadística. Entonces sea $z_u = g(a_u)$ la salida del nodo a_u. Con esto se forma una combinación lineal de esas salidas, por ejemplo se podría formar lo siguiente: $b_v = \sum_{u=0}^{k} w_{2uv} z_u$, donde $z_0 = 1$. Por lo tanto, la $v - esima$ respuesta y_v es una transformación de las b_v, así se tiene que $y_v = \tilde{g}(b_v)$, siendo $\tilde{g}()$ la función de activación de la respuesta. Con todo esto se pueden combinar las ecuaciones y se tendría lo siguiente: $y_v = \tilde{g}\left[\sum_{n=1}^{k} w_{2uv} g(\sum_{j=1}^{p} w_{1ju} x_j + \theta_{1j}) + \theta_{2u}\right]$.

La respuesta y_v es una combinación lineal transformada de combinaciones lineales transformadas de las variables de entrada predictores originales. Así para determinar la capa oculta, se acostumbra a escoger la función de activación de la función logística $g(x) = \frac{1}{(1+e^{-x})}$ o la función tangente hiperbólica, $g(x) = \tanh(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})}$.

Por otra parte la selección de la función de activación para la capa de salida depende de la naturaleza de la respuesta. Por ejemplo si la respuesta es acotada o dividida, la función de activación de salida comúnmente se supone sigmoidal, mientras que si es continua la respuesta, se utiliza con frecuencia una función identidad.

Ahora si analizamos la ecuación $y_v = \tilde{g}\left[\sum_{n=1}^{k} w_{2uv} g\left(\sum_{j=1}^{p} w_{1ju} x_j + \theta_{1j}\right) + \theta_{2u}\right]$ es una forma muy flexible, que contiene muchos parámetros, y es esta consideración que lleva a una red neuronal a ser una propiedad casi universal de aproximación en los modelos. Esto es, la red neuronal se ajusta a diversas funciones que ocurren en la naturaleza, sistemas o procesos; sin embargo, se deben de encontrar los parámetros de la ecuación $y_v = \tilde{g}\left[\sum_{n=1}^{k} w_{2uv} g\left(\sum_{j=1}^{p} w_{1ju} x_j + \theta_{1j}\right) + \theta_{2u}\right]$, los cuales se podría decir que son muchos. La forma más común de estimarlos es minimizando la suma de cuadrados residual general, la cual se calcula con todas las respuestas y todas las observaciones. Se puede decir que se trata de un problema de mínimos cuadrados no lineales, y para resolverlo se pueden usar diversos algoritmos y métodos.

Muchas veces se usa un procedimiento llamado propagación en reversa o retro propa-

gación, que es una variación del método de descenso más pronunciado o el método de ascenso acelerado, aunque también se han utilizado métodos como el del gradiente, que se realiza a base de derivadas. Como en cualquier procedimiento no lineal de estimación, se deben especificar los valores iniciales de los parámetros, para usar esos algoritmos. Por lo general se realiza una estandarización en todas las variables de entrada, por lo que se escogen valores pequeños, principalmente aleatorios, estos serian los valores iniciales.

Con muchos parámetros que se involucran en una función no lineal compleja existe demasiado riesgo de hacer un sobreajuste, lo que nos indica que una red neuronal artificial permite llegar a un ajuste casi perfecto, llegando un conjunto de datos históricos, o de entrenamiento, pero con frecuencia el modelo de red neuronal pronostica mal nuevos datos. El sobreajuste es un problema demasiado frecuente para las personas que trabajan con la estadística y que son expertos en la formación de modelos empíricos.

Los autores que trabajan con los modelos de redes neuronales artificiales han desarrollado varios métodos para resolver este problema, como por ejemplo minimizando la cantidad de parámetros desconocidos, realizándolo antes de que el proceso de estimación de parámetros llegue a la convergencia completa, y muchas veces aplicando la validación cruzada para determinar la cantidad de iteraciones que se deben de utilizar, complementándolo con una función de penalización a la suma de cuadrados de los residuales, la cual aumenta en función de la suma de cuadrados de los estimadores de los parámetros.

Por otra parte existen muchas diferentes estrategias para escoger la cantidad de capas, así como la cantidad de neuronas, también existen la forma de las funciones de activación, a esto se le llama o es la forma de determinar la estructura de la red neuronal artificial. También se puede utilizar la validación cruzada para seleccionar la cantidad de nodos en la capa oculta.

Los modelos de redes neuronales artificiales son una línea muy activa en la investigación y aplicaciones en los procesos, en especial para analizar problemas grandes, complejos y no lineales.

El asunto de sobreajuste permanentemente se pasa por alto, por parte de muchos usuarios e investigadores de las redes neuronales, y como muchos miembros de esta área de investigación no tienen conocimiento o entrenamiento firme en la formación empírica de modelos, en repetidas ocasiones no se aprecian las dificultades que puede provocar el sobreajuste, además, muchos simuladores y programas de cómputo para implementar redes neuronales no manejan correctamente y adecuadamente el problema de sobreajuste.

Muchos estadísticos y matemáticos creen que las redes neuronales son un complemento de las herramientas estadísticas y matemáticas, por ejemplo para el análisis de regresión y del diseño de experimentos, y muchos de ellos creen que este tipo de modelos no son sustitutos, porque muchas redes neuronales solo pueden representar un modelo de predicción, y no pueden reproducir un panorama elemental del funcionamiento del proceso básico que produjo los datos.

Por otro lado se puede definir una red neuronal artificial como un sistema compuesto por un número de elementos agrupados en capas que se encuentran interconectados. Este sistema posee entradas y salidas, es entrenado para conferirle un grado de aprendizaje según (Figura 4.1) y es capaz de procesar gran cantidad de información mediante modelos matemáticos (Hagan et al., 1996, Tsoukalas y Uhrig, 1998, Nelles, 2001, Arbib, 2003, Martín del Brío y Sanz, 2007).

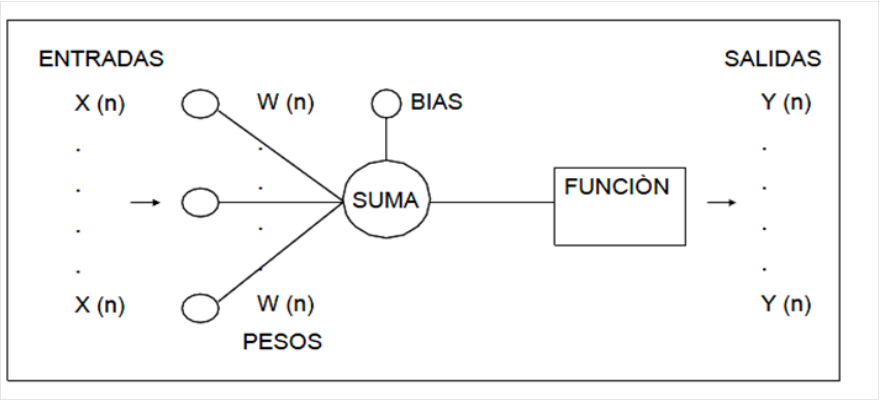

Figura 4.1: Estructura de una Red Neuronal

En el proceso de aprendizaje su funcionamiento es conocido como algoritmo de aprendizaje; su objetivo es modificar los pesos con el objetivo de minimizar la función (Tsoukalas y Uhrig, 1998).

4.1.1. Funciones de Redes Neuronales Artificiales

- Funciones de base y activación:

Una neurona suministra un valor a su salida que se propaga a través de la red de conexiones unidireccionales hacia otras células de la red. Asociada a cada conexión hay un peso sináptico, que determina el efecto de la neurona j sobre la neurona i. Las entradas a la neurona i, que provienen de las otras neuronas son acumuladas junto con el umbral externo, para dar el valor de la red neuronal artificial.

La forma de hacerlo está determinada matemáticamente por la función de base para dar un valor de activación. En este sentido, la salida final se puede expresar como una función de la entrada y pesos (Hagan et al., 1996, Tsoukalas y Uhrig, 1998, Nelles, 2001, Arbib, 2003, Martín del Brío y Sanz, 2007) Figura 4.1.

- Función de activación o función de neurona:

La función de activación se encarga de transformar el valor de la red expresado por la función de base. Las funciones de activación más comunes son:

- Función escalón.
- Función lineal y mixta.
- Función sigmoidal.
- Función gaussiana.
- Función Lineal de Base: La función lineal de base es de primer orden o de tipo hiperplano. El valor de red es una combinación lineal de las entradas.

- **Función de Base Radial.**

4.1.2. Función de Base Radial

La función de base radial es de segundo orden o de tipo hiperesférico. El valor de la red neuronal representa la distancia a un determinado patrón de referencia. Las redes de FBR son un modelo híbrido ya que utiliza tanto el aprendizaje supervisado como el no supervisado.

Una red neuronal FBR (Haykin, 1999) tiene tres capas diferentes: entrada, oculta y salida.

La capa de entrada está formada por los nodos fuente x_i que constituyen las unidades sensoriales.

La capa intermedia es una capa oculta de gran dimensión y en la que las unidades (neuronas) que la forman son las funciones base para los datos de entrada $\varphi_i(x_i)$.

La capa de salida d_j tiene la responsabilidad de la activación de patrones aplicados en la capa de entrada.

Las neuronas de la capa de salida son lineales. Las neuronas de la capa oculta calculan la diferencia entre el vector de entradas y los pesos sinápticos, y a esta diferencia le aplican una función radial con forma gaussiana principalmente, pero se tiene la ventaja que se pueden utilizar otras funciones radiales.

La función de transferencia radial, de tipo gaussiano adopta la forma de la ecuación:

$$G(\|x - t_i\|) = \exp\left(-\|x - t_i\|^2\right)$$

Está formada por la distancia Euclidiana o norma Euclidiana de $\|x - t_i\|^2$. La cual fue simplificada con (2.2) y la estructura dada por $\varphi(x)$ se puede ver de la siguiente manera:

$$\varphi_i(x) = G(\|x - t_i\|)$$

Donde x son las entradas y t_i son los centroides, y la función crece cuando la distancia entre t y x decrece, alcanzando un máximo de 1 cuando $t - x = 0$. Cuando se presenta un vector de entrada, cada neurona de la capa oculta, con funciones de transferencia de base radial, dará una salida que depende de lo cerca que está el vector de entrada de los pesos sinápticos. La entrada de cada neurona de la capa oculta es la distancia entre el vector de entradas y su vector de pesos. Esta distancia será multiplicada por el vector de los umbrales. Al resultado se le aplica la función radial (Haykin, 1999). La relación de entradas y salidas de la red neuronal esta dada por:

$$y(x) = \sum_{i=1}^{m} wG(\|x - t_i\|) + b$$

Donde $y(x_j) = d_j$ fue definida por (2.3). La variable d_j es la salida o respuesta $d = [d_1, d_2, \ldots, d_N]^T$; las ecuaciones fueron escritas en forma matricial por (2.4).

$$Gw = d$$

Donde el vector $w = [w_1, w_2, \ldots, w_m]^T$ representa los pesos y se determinan con (2.5) y el cálculo se realiza por la pseudoinversa o por mínimos cuadrados.

$$\begin{aligned} w &= G^{+}d \\ &= (G'G)^{-1} G'd \end{aligned}$$

La matriz G fue establecida por (2.6):

$$G = \begin{bmatrix} G(x_1 - t_1) & G(x_1 - t_2) & G(x_1 - t_m) \\ G(x_2 - t_1) & G(x_2 - t_2) & G(x_2 - t_m) \\ \vdots & \vdots & \vdots \\ G(x_N - t_1) & G(x_N - t_2) & G(x_N - t_m) \end{bmatrix}$$

Cabe mencionar que la ecuación (2.1) es la más utilizada, aunque existen diferentes tipos de funciones de base radial como:

$$\varphi_i(x) = r^2 \ln(r) \tag{4.1}$$

$$\varphi_i(x) = r^3 \tag{4.2}$$

$$\varphi_i(x) = r^4 \ln(r) \tag{4.3}$$

$$\varphi_i(x) = \exp\left(-r^2\right) \tag{4.4}$$

$$\varphi_i(x) = \exp\left(-\beta r^2\right) \tag{4.5}$$

$$\varphi_i(x) = \sqrt{r^2 + \beta^2} \tag{4.6}$$

$$\varphi_i(x) = r^k \tag{4.7}$$

$$\varphi_i(x) = r^k \ln(r) \tag{4.8}$$

Donde r esta dada por la distancia entre la variable de entrada x y los centroides t_i.

Las redes neuronales de FBR describen un método alternativo para la descomposición de un patrón de referencia en el espacio dentro de una región, describiendo conjuntos de puntos de datos en el espacio como si fueran generados de acuerdo a una función de densidad de probabilidad.

De esta manera, los modelos que se derivan de las redes perceptrón se concentran en los límites de clase, mientras que el enfoque de funciones de base radial se concentra en las regiones donde los datos de densidad son más altos, la construcción de aproximaciones globales para las funciones utilizando combinaciones de funciones de base centradas alrededor de los vectores de los pesos.

Las redes de FBR muestran que este enfoque no sólo tiene un rango de propiedades teóricas útiles sino que también es útil en la práctica o en las aplicaciones, como han sido

aplicadas de manera eficiente a los problemas de clasificación y discriminación de datos, de series de tiempo, de predicción y extracción de características.

Por estas condiciones las redes neuronales de FBR podrían aplicar el método Bayesiano, para los casos donde la aproximación está dada por $y = y(x; w)$ y está basado en una red usando combinaciones de funciones de base radial, cada una de ellas están centradas alrededor del vector de pesos w, para que la respuesta para la entrada x dependa de alguna medida de distancia desde x hasta w, en lugar del producto escalar $w \cdot x = \sum_i w_i x_i$ como en las demás neuronas. La distribución de los pesos w pueden ser determinados por alguna forma de agrupamiento por ejemplo mediante el método de Clustering.

Además el aprendizaje de ajuste de la conexión directa hacia una neurona cuyos resultados dan una estimación, por ejemplo, la probabilidad posteriori $p(c|\, x)$, la variable de clase c está presente dada la observación (entrada de la red) x. Sin embargo, es más fácil para modelar otros aspectos relacionados con los datos, tales como la distribución incondicional de los datos $p(x)$ y la verosimilitud de los datos, $p(x|\, c)$, y recrear la posteriori de acuerdo a estas cantidades para las leyes de Bayes, $p(c_i|\, x) = p(c_i)p(x|\, c_i)/p(x)$.

Estos métodos se pueden aplicar para los casos de Dinámica y aprendizaje de redes neuronales artificiales, para casos especiales de modelos gráficos que incluye esencialmente todos los modelos de aprendizaje no supervisado como es la red FBR. Es aplicable en otras funciones no lineales, como lo es la red FBR, la cual típicamente modificaría el cálculo de los pesos w y las entradas x; estas cosas se aplican en las cuestiones de comunicación y procesamiento digital; en circuitos integrados, entre otros.

Los modelos lineales generalizados tienen la forma $f(x; w) = \sum\limits_i w_i \phi_i(x) \equiv w^T \Phi(x)$. Estos modelos tienen una dependencia del parámetro lineal, sin embargo representan un mapeo no lineal de entrada-salida si las funciones de base $\phi(x)$, $i = 1, \ldots, k$ son no lineales. Las funciones de base radial son un ejemplo de una red.

Una opción popular es el uso de funciones Gaussianas de base $\phi_i(x) = \exp(-(x - \mu^i)^2/(2\lambda^2))$. En esta parte se asume que los centros μ^i son fijos, pero el ancho de las

funciones de base λ es un hiperparámetro que puede ser adaptado. Puesto que la salida es linealmente dependiente de w se puede rechazar a los valores extremos de salida, penalizando los valores de gran peso.

Un enfoque no bayesiano para modelar el ajuste basado en la minimización de un error de entrenamiento regularizado, normalmente utiliza un procedimiento como el de validación cruzada para determinar los parámetros de regularización (hiperparámetros). Tales enfoques requieren el uso de datos de validación. Una ventaja del enfoque bayesiano es que los hiperparámetros se pueden establecer sin la necesidad de la validación de datos, y así todos los datos pueden ser utilizados directamente para el entrenamiento.

En la parte anterior, la red de FBR fue revisada por una interpretación estadística de distribuciones de datos. En ese caso, el generador subyacente de los datos (la función de densidad de probabilidad) fue muestreado estocásticamente. Sin embargo, la formulación original de la red de FBR fue desarrollada con el fin de producir una asignación determinista de datos por explotar los vínculos con la función tradicional de aproximación. Este enfoque ha intentado introducir la noción de que la formación de redes neuronales podría describirse como ajuste de curvas.

Por lo tanto, la generalización en consecuencia tiene una interpretación natural como una interpolación a lo largo de esta superficie de conexión. La idea básica es la siguiente: Se supone que se tiene un conjunto de pares de entrada y salida de los patrones de entrada y la meta o target que representan los datos de una superficie suave pero desconocida en $\Re^n \times \Re^c$. Como un simple ejemplo, se considera la posibilidad de un conjunto de (x, y) pares, generados según $y = x^2$. En este enfoque, el problema es elegir una función $y : \Re^n \to \Re^c$, que reúna las condiciones y la interpolación $y(x_p) = t_p, p = 1, 2, \ldots, P$.

Se trata de una interpolación estricta, en el que la función está restringida a pasar a través de todos los puntos de los datos conocidos. La estrategia en la teoría de la interpolación es construir un espacio de una función lineal definida por un conjunto de funciones de base no ortogonal que dependa de la posición de los puntos de los datos

conocidos.

La asignación de expansión de la función de base radial a una dimensión originalmente se expresó como $y(x) = \sum_{j=1}^{p} \lambda_j \phi(\|x - x_j\|)$. Mediante el uso de las condiciones de la interpolación, los parámetros de conexión λ pueden determinarse mediante métodos de la matriz inversa.

El enfoque se generalizó a asignaciones de dimensiones superiores, para incorporar el sesgo de los términos, y para tener en cuenta el hecho de la interpolación estricta donde no es una buena estrategia para los datos con ruido del mundo real, llevando a la topología de red la neuronal FBR ya analizada.

En el caso del ejemplo de $y = x^2$ antes mencionado, las entradas son x valores, los objetivos son valores y específicos, y se construye la red de FBR para producir una superficie de conexión para la parábola $y = x^2$, que es una superficie en $\Re^1 \times \Re^1 = \Re^2$. Se toma en cuenta que este es el ajuste de curvas a la parábola en el espacio del producto, no solo los datos de la muestra de sí mismos.

Esta parábola podrá interpretarse como el generador de los datos observados, como las ubicaciones de la parábola que produce o genera (x, y) pares de entrada y salida. Siempre que el generador de los datos es suave y no lineal, se debería poder arbitrariamente o estrechamente aproximar con una red neuronal de FBR.

Esto explica por qué las redes neuronales pueden ser exitosas en la predicción, en las series de tiempo caóticas; aunque las series de tiempo mismas pueden exhibir aleatoriedad aparente, el mapa subyacente que produce las muestras es en sí misma por lo general muy suave.

Por otro lado en el aprendizaje competitivo hay una realización de la red de FBR para este modelo de aprendizaje, hay J unidades ocultas de base radial y una unidad de salida lineal que da la función de verosimilitud sobre las condiciones del modelo. Una regla de aprendizaje para la competencia no tan intensa es similar a la de la competencia más severa.

La diferencia es que cada uno de los centros del cluster se modifica en el sentido de la línea que une el cluster del centro y el patrón x_k en proporción a la probabilidad de que pertenece a ese grupo o cluster. Esto se deduce de la formulación de los datos como un modelo Gaussiano-mezclado. El enfoque de la competencia poco intensa está estrechamente relacionado con el método de Clustering k-medias. Esto quiere decir que el modelo de red neuronal de FBR tiene un concepto o modelo de aprendizaje.

Hay diferentes tipos de redes neuronales tales como las funciones de base radial, que se pueden utilizar como parte de las redes neuronales basadas en el conocimiento. Las redes neuronales basadas en el conocimiento son redes neuronales previamente estructuradas que permiten la manipulación de datos y conocimiento, incluyendo el aprendizaje a partir de datos, tienen las reglas de la inserción, la regla de extracción, adaptación, y el razonamiento. Las redes neuronales basadas en el conocimiento se han desarrollado, ya sea como una combinación de los sistemas inteligentes y redes neuronales, o como una combinación de los sistemas de lógica difusa y redes neuronales, o como otros sistemas híbridos.

En el método de ensamble de aprendizaje, se puede utilizar las redes neuronales de FBR como un árbol de decisión con un algoritmo de aprendizaje.

Para el proceso Gaussiano para problemas de regresión también se puede utilizar una red neuronal de FBR. Otro uso es en la generalización y regularización de los sistemas de aprendizaje no lineal. La red de FBR es un buen método para la regresión no paramétrica. En control e identificación también se puede utilizar la red de FBR. Que se puede considerar una red de dos capas en el que la capa oculta realiza una transformación no lineal en las entradas.

Otras formas de utilización de la red de FBR esta en el reconocimiento de objetos y patrones de referencia, puede utilizarse con métodos supervisados y k-medias. Se puede utilizar en el aprendizaje no supervisado para robots.

4.1.3. Estructuras de las Redes Neuronales Artificiales

Estos sistemas capaces de procesar gran cantidad de información mediante modelos matemáticos (Yen y Langari, 2000) poseen varios grados de conectividad cuyo ajuste determina el desempeño de la red neuronal.

Por otra parte, la mayoría de las redes neuronales artificiales están compuestas por una matriz X de vectores de n entradas $X_p = (x_{p1}, x_{p2}, \ldots, x_{pN})^T$ tienen una estructura de pesos sinápticos $\left(w_{ji}^h\right)$ que son utilizados para la ponderación de los valores de entrada en la configuración interna de la red; se componen también de un umbral o bias $\left(\theta_j^h\right)$, como se observa en (4.9), el cual se aplica para reducir la entrada a la función de activación (Babuska, 2001).

$$net_{pj}^h = \sum_{i=1}^{N} w_{ji}^h x_{pi} + \theta_j^h \tag{4.9}$$

Otro componente de una red es el operador matemático suma, que realiza la operación de suma de los componentes de las señales de entrada multiplicados por sus respectivos pesos. Por último existe la función de activación o transferencia (i_{pj}) donde se define la salida de la neurona (net_{pk}^o, o_{pk}). Para el proceso de aprendizaje de la red neuronal es conocido como algoritmo de aprendizaje, como se expresa en (4.10) a (4.12) y el objetivo del algoritmo de aprendizaje es modificar los pesos para minimizar la función (Fullér, 1995).

$$i_{pj} = f_j^h(net_{pj}^h) \tag{4.10}$$

$$net_{pk}^o = \sum_{j=1}^{L} w_{kj}^o i_{pj} + \theta_k^o \tag{4.11}$$

$$o_{pk} = f_k^o(net_{pk}^o) \tag{4.12}$$

Los aspectos más característicos de las redes neuronales artificiales son la estructura de conexión. Una red neuronal está determinada por la neurona y la matriz o vector de pesos. El comportamiento de la red depende en gran medida del desempeño de la matriz o vector de pesos. Hay tres tipos de capas de neuronas: la de entrada, la oculta y las de salida. Entre dos capas de neuronas existe una red de pesos de conexión, que puede ser de los siguientes tipos: Hacia delante, hacia atrás, lateral y de retardo (Fullér, 1995).

- **Redes Neuronales Supervisadas y No Supervisadas**

Las redes neuronales se clasifican comúnmente en términos de sus algoritmos o métodos de entrenamiento en: redes de pesos fijos, redes no supervisadas y redes de entrenamiento supervisado. Para las redes de pesos fijos no existe ningún tipo de entrenamiento.

- Redes con entrenamiento supervisado:

Las redes con entrenamiento supervisado han sido los modelos de redes más desarrolladas desde los inicios de este tipo de diseños. Los datos para el entrenamiento están constituidos por variables de entrada y de salida. El hecho de conocer la salida implica que el entrenamiento se beneficia de la supervisión de una variable.

- Redes con entrenamiento no supervisado:

Las redes con entrenamiento no supervisado utilizan datos de entrenamiento consistentes en sólo variables de entrada. Por lo tanto, la red es entrenada sin el beneficio de una variable de salida. La red aprende a adaptarse basada en las experiencias recogidas de los patrones de entrenamiento anteriores (Hagan et al., 1996, Tsoukalas y Uhrig, 1998, Nelles, 2001, Arbib, 2003, Martín del Brío y Sanz, 2007).

4.2. Algoritmos Evolutivos

La computación evolutiva comprende técnicas de computación que son modelos basados en la teoría de la evolución de Charles Darwin. En la mayoría de la literatura, este tipo

de algoritmos están inspirados en la naturaleza y son conocidos como "Soft Computing" (Computación Suave).

Dentro de este tipo de sistemas inteligentes existen modelos que usan Algoritmos Evolutivos (EAs por sus siglas en inglés), Algoritmos de Estimación de Distribuciones (EDAs por sus siglas en inglés), Programación Evolutiva y Estrategias Evolutivas, entre otros. Algunos tipos de estos modelos de sistemas inteligentes son:

- Algoritmos Genéticos (evolutivos)
- PSO (Particle Swarm Optimization)
- Simulated Annealing
- Colonia de hormigas
- Algoritmos Culturales
- Algoritmo del Diluvio
- PBIL
- BOA
- CGA
- EVONORM (Torres-Treviño, 2007)

4.2.1. Algoritmos Genéticos

Este tipo de algoritmos fueron propuestos por John Holland (1975) y se han desarrollado como uno de los métodos mas utilizados para resolver problemas de optimización. Los Algoritmos Genéticos (AG) están definidos como un procedimiento de búsqueda basado en los mecanismos de genética y selección natural. Estos tipos de algoritmos resuelven problemas difíciles de una manera rápida, adecuada y confiable (Goldberg, 2002).

Los Algoritmos Genéticos (AG) son una técnica de búsqueda adaptiva que emula la teoría biológica evolutiva. La población, un conjunto de individuos (potenciales soluciones al problema de optimización) cambia probabilísticamente en el tiempo debido a la aplicación de operadores inspirados biológicamente, cruza y mutación. Un proceso de selección determina cuales individuos (entre los padres y descendientes) permanecen en la próxima generación. Según las teorías evolutivas, sólo los individuos más aptos de la población tienen mayor probabilidad de sobrevivir y generar descendencia y de esta manera, transmitir su herencia biológica a las nuevas generaciones (Sanhueza et al., 1999).

La búsqueda del óptimo global de un problema de optimización, basada en la evolución natural de las especies, se realiza al pasar de una población (generación) antigua de individuos a otra población nueva mediante la aplicación de operadores genéticos. Cada individuo (cromosoma) representa una solución candidata del problema de optimización; entonces, una población de individuos consiste en un conjunto de soluciones del problema a optimizar.

Un individuo es modelado como un "string" de símbolos de longitud fija; usualmente, los símbolos corresponden al alfabeto binario, donde cada bit o grupo de bits, representa un valor de alguna variable del problema (gen). Una función de evaluación, llamada función aptitud ("fitness"), asigna un valor a cada individuo de la población. El valor "fitness" es una medida de la calidad de un individuo. La idea básica consiste en procesar los individuos más aptos mediante operadores genéticos (los que poseen mayores "fitness") para producir mejores individuos (descendencia de mayor calidad) a medida que el proceso de búsqueda avanza (Tzafestas y Blekas, 1999).

En resumen la mayoría de los Algoritmos Genéticos se compone de, un esquema de Decodificación, una función de evaluación o evaluación "fitness", una selección y operadores de Cruce y Mutación.

- Esquemas de decodificación:

Los esquemas de decodificación transforman los puntos que están en un espacio de

parámetros, que pueden ser representados en código binario. Los cuales forman los cromosomas. (Nelles, 2001)

- Evaluación "fitness":

El primer paso después de crear una generación (padres e hijos) es calcular el valor de "fitness" de cada miembro de la población; aquí es donde es importante la función objetivo.

- Selección:

Después de la evaluación se tiene que crear la nueva población de la actual generación. La operación de selección determina cuales padres participan en producir los siguientes hijos, para la próxima generación. En esta parte se localizan los hijos que son los mejores individuos, o se hace la selección para encontrar los mejores individuos.

- Cruce:

Para mejorar el potencial del actual gen, se necesitan los operadores de Cruce, para generar un nuevo cromosoma, que ayude o pueda retener buenos elementos de la generación previa. El Cruce es comúnmente aplicado para seleccionar el par de padres con una probabilidad igual dada por un indicador o índice de Cruce. El Cruce o recombinación que también se le llama, es un operador genético, que combina los bits y partes de las soluciones de los padres para formar una nueva y posiblemente mejores hijos. La principal idea es entender que los hijos sobre un cruce no deberían ser idénticos y deberían combinar los rasgos de los padres de mejor manera. Digamos que el Cruce crea nuevos individuos por cruce de los rasgos de dos o más padres (Jang et al., 1997).

El Cruce de un punto es el operador más común donde un cruce de un punto en el código genético es seleccionado aleatoriamente y los dos cromosomas de los padres son intercambiados en este punto. Existe también el cruce de dos puntos; donde, dos puntos

de cruce son seleccionados y la parte de la serie o cadena de cromosoma entre estos dos puntos son intercambiados los cromosomas para generar dos hijos. También se puede decir que un operador cuyo trabajo es escoger una posición y cambiarlas antes y después de ese lugar entre los cromosomas, todo esto para crear hijos, por ejemplo, la figura 4.2, donde 01 00 00 11 y 11 10 01 01 se pueden cruzar del cuarto lugar y se obtiene la descendencia 01 00 01 01 y 11 10 00 11.

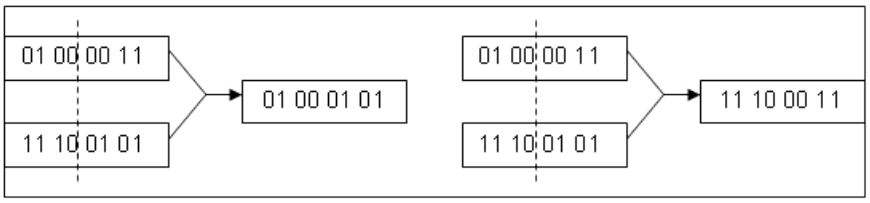

Figura 4.2: Cruce en un algoritmo genético

La recombinación es el mayor operador genético en los algoritmos genéticos. Los bits de dos o más padres son cortados en dos o más piezas y las partes de los bits se cruzan; este tipo de recombinación como se mencionó anteriormente es llamado cruce. Ambos hijos contienen una parte de información de un padre y el resto del otro padre. El índice de cruce o el indicador esta constituido comúnmente por un valor alto de 0.6 a 0.9. Este valor determina la probabilidad con la cual el cruce esta diseñado o se esta comportando. A este dato también se le llama la probabilidad de cruce (Goldberg, 2002).

- Mutación:

La Mutación actúa para modificar simplemente un solo individuo. Hay muchas variaciones de mutación, pero la más significativa es que los hijos son idénticos a los individuos de los padres excepto que uno o más cambios están hechos para un rasgo o rasgos de individuos hechos por el operador. Por si misma la mutación representa un camino aleatorio en la vecindad de una solución en particular; si se hace repetidamente sobre la población de individuos se puede esperar que la población resultante sea indistinguible de una creada aleatoriamente (Jang et al., 1997).

El cruce aprovecha genes actuales potenciales, pero si la población no contiene toda la información decodificada necesita resolver un problema en particular, por mucho que es la mezcla de genes que produce una solución satisfactoria. Por esta razón, un operador de mutación es capaz de generar espontáneamente nuevos cromosomas que son incluidos. La manera más común de implementar la mutación es lanzar un bit con una probabilidad igual para un muy bajo índice de mutación. Un operador de mutación puede impedir a cualquier bit solo de converger a un valor a través de la población de entrada, y aún más importante, puede impedir que la población de convergencia se estacione en cualquier óptimo local (Jang et al., 1997; Goldberg, 2002).

La mutación imita la introducción de errores (no reparados) dentro de un código genético natural, por ejemplo el causado por radiación. El principal propósito de la mutación es evitar al algoritmo genético de quedar estancado en ciertas regiones del espacio de los parámetros. La mutación tiene un efecto similar a la generación de ruido en los parámetros de actualización, es una estrategia para escapar de los óptimos locales, que son conocidas como las técnicas de optimización local (Jang et al., 1997; Nelles, 2001; Goldberg, 2002).

Este operador genético puede hacer movimientos o reemplazos con los bits en un cromosoma, por ejemplo en la figura 4.3 el cromosoma 01 00 00 11 produce una mutación con la quinta posición y así se obtiene 01 00 10 11.

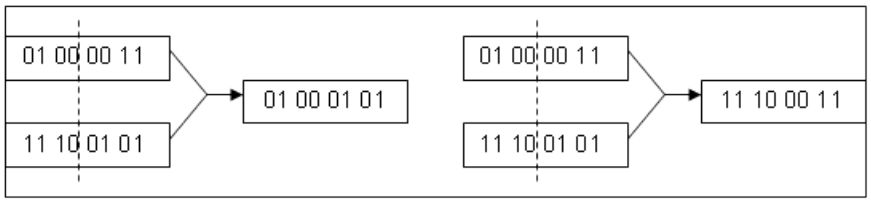

Figura 4.3: Mutación en un algoritmo genético

En un algoritmo genético estándar todos los bits son mutados con igual probabilidad pero no todos los bits tienen la misma significancia. Este índice de mutación o valor de probabilidad de mutación es un valor pequeño que usualmente esta entre 0.001 y 0.01.

El índice de probabilidad de mutación es usualmente bajo pero es bueno para obtener cromosomas del cruce que no están perdidos. Si el índice de probabilidad de mutación es alto (mayor de 0.1), se podría comportar o tener un desempeño como el de los algoritmos de búsqueda aleatoria (Goldberg, 2002).

Por otra parte los algoritmos genéticos son fáciles de programar y pueden ser aplicados a un amplio rango de problemas de optimización; usualmente donde los algoritmos tradicionales no son aplicables o si lo son no es posible encontrar una solución en un tiempo razonable. En relación a otros algoritmos, los algoritmos genéticos difieren en los puntos siguientes (Harnisch et al., 2000):

Trabajan con un código de los parámetros del problema y no con los parámetros en sí mismo. Procesan una población de soluciones candidatas (individuos) y no solamente una solución única. No dependen de la existencia de derivadas, sino solamente del valor "fitness" de las soluciones candidatas. Usan reglas de transición probabilísticas y no utilizan reglas de transición determinísticas. Al resolver un problema mediante algoritmos genéticos se deben considerar tres aspectos importantes:

- Representación
- Operadores genéticos
- Función de aptitud

Los algoritmos genéticos se pueden desarrollar de la siguiente manera:

Paso 1: Generar una población aleatoria.

Paso 2: Evaluar cada solución representada (un conjunto de parámetros) de la población (se podría usar Redes Neuronales).

Paso 3: Seleccionar los mejores individuos de la población.

Paso 4: Generar una nueva población de individuos.

Paso 5: Encontrar una solución óptima (o la mejor de las soluciones).

4.3. Lógica Difusa

La lógica difusa es una metodología que proporciona una manera simple de obtener una conclusión a partir de información de entrada vaga, ambigua, imprecisa o incompleta; permite imitar la manera como los seres humanos razonamos donde la información a manipular contiene un alto grado de imprecisión (Yen y Langari, 2000). En general la lógica difusa imita como una persona toma decisiones basadas en información con las características mencionadas. La técnica principal de la lógica difusa esta fundamentada en cuatro conceptos básicos (Yen y Langari, 2000):

1. Conjuntos Difusos: Conjuntos con límites variables.

2. Variables Lingüísticas: Variables que toman valores que pueden ser cuantitativos o cualitativos, contenidos en un conjunto difuso.

3. Distribuciones de Posibilidad: Restricciones donde el valor de una variable lingüística es asignada a un conjunto difuso.

4. Reglas If-Then: Formula lógica que generaliza una implicación en valores lógicos.

Los primeros tres conceptos son fundamentales para todas las áreas de la lógica difusa; sin embargo, el cuarto concepto es importante debido a que es la base de las aplicaciones industriales de la lógica difusa.

Conjuntos Difusos:

Los conjuntos difusos pueden contener elementos con un grado parcial de membresía, no se tienen límites claramente definidos. Para entender mejor los conjuntos difusos seria conveniente describir primeramente lo que es un conjunto clásico; este es un contenedor que incluye o excluye completamente un elemento dado. Un conjunto clásico no permite elementos con grado parcial de membresía o pertenencia, es decir pertenece o no pertenece al conjunto funcionando como lógica binaria. El razonamiento difuso proporciona la habilidad de responder a una pregunta de respuesta si o no con si, mas o menos, un poco, casi nada o no (Rao et al., 1999).

Funciones de membresía:

Una función de membresía es una curva que define como es representado cada punto del espacio de entrada en un valor de membresía (o grado de membresía entre 0 y 1). La función puede ser una curva arbitraria donde la forma puede ser definida como una función que se ajuste desde el punto de vista de: la simplicidad, la conveniencia, la velocidad y la eficiencia (Yen y Langari, 2000).

Variables lingüísticas:

Una variable lingüística permite describir cualitativa como cuantitativamente un valor; cualitativamente por medio de un término lingüístico y cuantitativamente por su representación en una función de membresía. Los términos lingüísticos son usados para expresar conceptos y conocimientos en el lenguaje que el ser humano utiliza comúnmente (Yen y Langari, 2000).

Reglas If-Then:

Estas reglas If-Then son usadas para formular los enunciados condicionales que conforman la lógica difusa. Una regla difusa es la unidad básica para representar el conocimiento en muchos sistemas difusos. Una regla difusa tiene dos componentes: una parte condicional If (que se refiere al antecedente) y una parte consecuente Then (que se refiere a la consecuencia). El antecedente describe una condición y la consecuencia describe una conclusión. Una regla simple asume la forma:

- Si x es A entonces y es B.

Donde A y B son valores lingüísticos definidos por conjuntos difusos en los rangos de x y de y (Yen y Langari, 2000).

Proceso de inferencia difusa:

El proceso de inferencia difusa consiste en formular la combinación de las entradas para la obtención de valores de salida utilizando la lógica difusa, involucra todas las partes mencionadas anteriormente: reglas difusas, variables lingüísticas, funciones de membresía.

Los sistemas mas populares de lógica difusa que se encuentran en la literatura son, sistemas difusos tipo Mamdani (con fuzzificador y defuzzificador) y sistemas difusos tipo

Takagi-Sugeno (Yen y Langari, 2000). En los sistemas difusos tipo Mamdani la salida del mecanismo de inferencia es un conjunto difuso resultante. Para generar la salida numérica a partir de este conjunto existen varias opciones como el centro de gravedad dado por (4.13), los centros promediados que se muestran en (4.14) entre otros.

$$y = \frac{\sum_i b_i \int \mu(i)}{\sum_i i \int \mu(i)} \tag{4.13}$$

$$y = \frac{\sum_i b_i \mu_{premisa}(i)}{\sum_i \mu_{premisa}(i)} \tag{4.14}$$

Los sistemas difusos tipo Sugeno, para calcular la salida del sistema difuso se ponderan los diferentes consecuentes teniendo en cuenta el valor que se activó, el antecedente de cada una de las reglas, para un sistema con dos reglas la salida del sistema difuso sería como en (4.15) y el cálculo de la salida de un sistema difuso Sugeno está dado por (4.16):

$$y = \frac{w_1 y_1 + w_2 y_2}{w_1 + w_2} \tag{4.15}$$

$$\begin{aligned} y_1 &= f_1(x) \\ y_2 &= f_2(x) \end{aligned} \tag{4.16}$$

4.4. Técnicas Estadísticas

En esta sección se abordan temas sobre las herramientas estadísticas que se emplean para identificar las variables que tienen influencia en un proceso; así como, los métodos que se utilizan para estimar su influencia en el proceso. Aunado a esto se tocan temas estadísticos que se emplean, se fusionan y se conjugan con sistemas inteligentes y sobre todo que pueden ayudar a mejorar este tipo de modelos.

4.4.1. Análisis de variancia (ANOVA)

El análisis de variancia (ANOVA) es un procedimiento para probar la igualdad de varias medias y determinar la significancia estadística de la relación entre una variable dependiente Y una o varias variables independientes X que han sido organizadas dentro de 2 o más grupos discretos o niveles (Montgomery, 1991).

El modelo estadístico lineal representa las observaciones del análisis:

$$y_{ij} = \mu + \tau_i \varepsilon_{ij} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n \end{cases} \tag{4.17}$$

Donde y_{ij} es la (ij)-ésima observación, μ es un parámetro común a todos los tratamientos denominado media global, τ_i es un parámetro único para el i-ésimo tratamiento llamado efecto del tratamiento i-ésimo, y ε_{ij} es la componente aleatoria del error.

El objetivo será probar hipótesis apropiadas con respecto a los efectos del tratamiento y hacer una estimación de ellos. Para probar las hipótesis se supone que los errores del modelo son variables independientes con distribución normal, con media cero y variancia σ^2 conocida. Se supone que ésta última es constante para todos los niveles del factor (Montgomery, 1991).

Descomposición de la suma total de cuadrados:

La denominación análisis de variancia resulta de descomponer la variabilidad total de los datos en sus partes componentes. La suma total de cuadrados corregida (4.18) se usa como una medida de la variabilidad total de los datos.

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}..)^2 \tag{4.18}$$

Intuitivamente esto parece razonable, ya que si se divide SS_T entre el número apropiado de grados de libertad (en este caso entre $an - 1 = N - 1$), se obtiene la variancia muestral de y. La variancia muestral es una medida estándar de la variabilidad.

La suma total de cuadrados puede descomponerse en la suma de cuadrados de las

diferencias entre los promedios de los tratamientos y el promedio general, y en la suma de cuadrados de las diferencias entre las observaciones dentro del tratamiento y el promedio del mismo. La diferencia entre los promedios observados de los tratamientos y el promedio general constituye una medida de la diferencia entre las medias del tratamiento; mientras que la causa de las diferencias de las observaciones dentro de los tratamientos con respecto al promedio del tratamiento puede ser solamente el error aleatorio. Por tanto, simbólicamente la ecuación (4.19) puede ser escrita como:

$$SS_T = SS_{Tratamientos} + SS_E \tag{4.19}$$

Donde $SS_{Tratamientos}$ se denomina suma de cuadrados debida a los tratamientos y SS_E suma de cuadrados debida al error. La SS_T tiene $N-1$ grados de libertad porque hay un total de $an = N$ observaciones. Existen a niveles del factor de manera que $SS_{Tratamientos}$ tiene $a-1$ grados de libertad. Existen n replicas dentro de cada tratamiento, las cuales proporcionan $n-1$ grados de libertad para estimar el error experimental. Como hay tratamientos, se tienen $a(n-1) = an - a = N - a$ grados de libertad para el error. Por otra parte los grados de libertad son todas las combinaciones posibles que se pueden hacer entre la fuente de variación.

Las cantidades $MS_{Tratamientos}$ y MS_E se denominan respectivamente medias de cuadrados (4.20) y (4.21) o cuadrados medios de los tratamientos o del modelo y cuadrados medios del error o residuales (Montgomery, 1991).

$$MS_{Tratamientos} = \frac{SS_{Tratamientos}}{a-1} \tag{4.20}$$

$$MS_E = \frac{SS_E}{N-a} \tag{4.21}$$

Una prueba para la hipótesis de la igualdad en el nivel medio de los tratamientos pueden efectuarse comparando $MS_{Tratamientos}$ y MS_E; a continuación se muestra como puede realizarse la comparación.

Como la suma de los grados de libertad de $SS_{Tratamientos}$ y de SS_E es igual a $N-1$, es decir el total de los grados de libertad, esto implica que $SS_{Tratamientos}/\sigma^2$ y SS_E/σ^2 son variables aleatorias independientes con distribución χ cuadrada. Por tanto, si la hipótesis nula de igualdad de medias de los tratamientos es verdadera, la razón (4.22) tiene una distribución F con $a-1$ y $N-a$ grados de libertad.

$$F_0 = \frac{SS_{Tratamientos}/(a-1)}{SS_E/(N-a)} = \frac{MS_{Tratamientos}}{MS_E} \tag{4.22}$$

La ecuación (4.22) es la fórmula estadística para probar la hipótesis de igualdad de medias de los tratamientos. El valor esperado del numerador en la fórmula estadística de prueba es mayor que el valor esperado del denominador si la hipótesis alterna es verdadera y, en consecuencia debe rechazarse H_0 si el valor de tal estadística es demasiado grande. Esto implica una región critica unilateral superior (Montgomery, 1991). En otras palabras se rechaza H_0 si $F_0 > F_{a,a-1,N-a}$.

La suma de cuadrados son las unidades cuadradas al área bajo la curva que representa cada fuente de variación. Los cuadrados medios son las unidades cuadradas promedio que le corresponde a cada combinación.

La Tabla 4.1 muestra un ejemplo de una tabla de análisis de varianza (ANOVA):

Tabla 4.1: Ejemplo de tabla de ANOVA

Análisis de Varianza				
Fuente de Variación	**Grados de Libertad**	**Suma de Cuadrados**	**Cuadrados Medios**	**F**
Modelo	k	SS_{Modelo}	MS_{Modelo}	$\frac{MS_{Modelo}}{MS_{Error}}$
Residual o Error	$n-k-1$	SS_{Error}	MS_{Error}	
Total	$n-1$	SS_{Total}		

Donde los grados de libertad están dados por k o $p = r-1$ donde r son los coeficientes, por eso p es el número de p variables independientes; n es el número total de observaciones; SS_{Modelo} es la suma de cuadrados del modelo; SS_{Error} es la suma de cuadrados de los

residuales; SS_{Total} es la suma de cuadrados total; MS_{Modelo} son los cuadrados medios del modelo y MS_{Error} son los cuadrados medios del error o residuales (Montgomery, 2006).

- **Terminología de análisis de varianza:**

Factor: Una variable independiente en un experimento estadístico.

Nivel o Tratamiento: Un valor discreto o establecido para un factor.

Datos Balanceados: Número igual de observaciones para cada factor y combinación de nivel.

Datos No Balanceados: Número diferente de observaciones para cada factor y combinación de nivel.

- **Tipos de ANOVA:**

ANOVA 1 Vía: Un solo factor X ocurre en más de dos niveles (Tratamientos) y hay una sola variable dependiente Y, determina que tanto contribuyen los niveles del factor a la variación total de la variable de respuesta.

ANOVA Balanceado: Factores múltiples (2 o más), representados en múltiples niveles con una variable dependiente continua Y. Determina que tanto contribuye cada factor a la variación total de la variable de respuesta.

- **Supuestos básicos para usar ANOVA:**

1. La distribución de los residuales (e) deben ser normales $e \approx N(0, \sigma^2)$ (Montgomery, 2006). Para esto existen pruebas de Normalidad como:

a) Kolmogorov-Smirnov.

b) Anderson-Darling.

a. Kolmogorov-Smirnov.

La prueba de Kolmogorov-Smirnov es de distribución libre en el sentido que los valores críticos no dependen de la distribución específica que es probada. Para realizar la prueba se pueden seguir los siguientes pasos:

- Determinar el número de intervalos y sus rangos de una lista de n datos.
- Considerar todos los datos en cada intervalo y determinar la probabilidad acumulada correspondiente.
- Proponer una función de distribución para calcular su acumulado (en este caso es la función de distribución Normal) considerando como entrada el rango superior de cada intervalo (prueba de similitud).
- Calcular las diferencias absolutas entre las probabilidades acumuladas de los datos reales contra los propuestos.
- Si la diferencia máxima es menor al valor dado por la tabla de Kolmogorov-Smirnov, considerando $1 - alpha$ de confianza con n grados de libertad, entonces no se puede rechazar la prueba.

La prueba Kolmogorov-Smirnov esta definida por:
H_o: Los datos siguen una distribución normal.
H_a: Los datos NO siguen una distribución normal.
La prueba estadística de Kolmogorov-Smirnov se define como:

$$D = \text{máx}\left\{D^+, D^-\right\} \tag{4.23}$$

$$D^+ = \underset{i}{\text{máx}}\left\{\frac{i}{n - Z_{(i)}}\right\} \tag{4.24}$$

$$D^- = \underset{i}{\text{máx}}\left\{Z_{(i)} - \frac{(i-1)}{n}\right\} \tag{4.25}$$

Donde $Z_{(i)} = F\left(X_{(i)}\right)$ y $F\left(X\right)$ es la función de probabilidad de la distribución Normal. $X_{(i)}$ es el i-ésimo estadístico ordenado de una muestra aleatoria, $1 \leq i \leq n$ y n es el tamaño de muestra.

b. Anderson-Darling.

La prueba del Anderson-Darling es una modificación de la prueba de Kolmogorov-Smirnov y da más peso a las colas que la prueba de Kolmogorov-Smirnov. Hace uso de la distribución específica para calcular valores críticos. Mide el área entre la línea ajustada (basándose en la distribución elegida, que en este caso es la distribución Normal) y la función escalón no paramétrica. Esto tiene la ventaja de permitir una prueba más sensible y la desventaja que los valores críticos se deban calcular para cada distribución (D'Agostino y Stephens, 1986).

La estadística es un cuadrado de la distancia que es mayor ponderación en las colas de la distribución. Valores más bajos de la prueba Anderson-Darling indican que la distribución se adapta mejor a los datos. La prueba de normalidad Anderson-Darling está definida como:

H_o: Los datos siguen una distribución normal.

H_a: Los datos NO siguen una distribución normal.

La fórmula de la prueba estadística Anderson-Darling está definida como:

$$A^2 = -N - \left(\frac{1}{N}\right) \sum_{i=1}^{N} (2i-1) \left(\ln\left(F\left(Y_i\right)\right) + \ln\left(1 - \ln\left(F\left(Y_{N+1-i}\right)\right)\right)\right) \tag{4.26}$$

Donde F es la función de distribución acumulada de la distribución Normal. Y_i son las observaciones ordenadas (D'Agostino y Stephens, 1986).

2. Igualdad de Varianzas:

Las varianzas de las poblaciones deben ser iguales (o cercanas a la misma) a través de todos los valores para un factor dado.

Homogeneidad de las varianzas:

- Prueba de Bartlett

- Prueba de Levene

3. Promedios Muestrales Normalmente Distribuidos:

Las medias muestrales deben ser independientes y normalmente distribuidas.

4. Datos Aleatorios:

Los datos muestrales deben representar verdaderamente su variación poblacional. Para cada nivel los datos muestrales deben ser seleccionados al azar en cada población.

4.4.2. Diseños de Experimentos (DOE) y Regresión Múltiple

Para determinar los factores que intervienen en el comportamiento de una variable de respuesta, cuando es posible controlar los niveles de esos factores, el diseño de experimentos y los modelos de regresión han mostrado ser excelentes herramientas para modelar, caracterizar y optimizar los procesos al facilitar la identificación y minimización de sus causas de variación (Montgomery, 2006).

Ya que un diseño experimental es una prueba o una serie de pruebas donde se inducen cambios deliberados en las variables de entrada de un proceso o sistema, de manera que sea posible observar e identificar las causas de los cambios en las respuestas de salida. Variar en forma planeada y deliberada las variables de entrada al sistema para observar el comportamiento de la o las variables de salida bajo diferentes condiciones, es el objetivo del diseño experimental (Montgomery, 2006). Entre los objetivos del experimento puede incluirse:

- Determinar cuales variables tienen mayor influencia en la respuesta y.
- Determinar el mejor valor de las x, que influyen en y, de modo que y tenga casi siempre un valor cercano al valor nominal deseado.
- Determinar el mejor valor de las x que influyen en y, de modo que la variabilidad de y, sea pequeña.
- Determinar el mejor valor de las x que influyen en y, de modo que se minimicen los efectos de las variables incontrolables.

Durante el proceso de optimización, los diseños experimentales hacen uso del método de ascenso acelerado para que en cada iteración los resultados de salida se aproximen al óptimo, identificando (a través del ANOVA) los niveles de los factores y las interrelaciones que determinan el comportamiento de la variable de respuesta.

Los diseños más utilizados para este fin, debido a su bajo costo y efectividad son entre otros, el diseño factorial de la serie 2^k, el diseño factorial fraccionado de resolución III, el diseño central compuesto el cual ayuda a construir la superficie de respuesta, el diseño de experimentos general y el diseño de Taguchi.

El comportamiento de la variable aleatoria Y, analizada a través de un diseño experimental, generalmente se representa a través de un polinomio empírico, lineal, cuadrático y rara vez cúbico, dependiendo del comportamiento de los datos.

La representación matricial por ejemplo de un polinomio cuadrático completo es:

$$\hat{Y} = \hat{\beta}_0 + X^t\hat{\beta} + X^tBX \tag{4.27}$$

Donde: $\hat{\beta}_0$ es la ordenada al origen, $X = (X_1, X_2, ..., X_p)$ es una matriz no estocástica $(n \times p)$ de rango p, $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, ...\hat{\beta}_k)$ es el vector de coeficientes de regresión de primer orden y B, es una matriz simétrica $(k \times k)$ de coeficientes cuadráticos de interacción obtenida de (4.27) representada por (4.28).

$$B = \begin{pmatrix} \hat{\beta}_{11} & \frac{1}{2}\hat{\beta}_{12} & \cdots & \frac{1}{2}\hat{\beta}_{1k} \\ & \hat{\beta}_{22} & \cdots & \frac{1}{2}\hat{\beta}_{2k} \\ & & \ddots & \vdots \\ Sim & & & \hat{\beta}_{kk} \end{pmatrix} \tag{4.28}$$

La estimación de los $\hat{\beta}_j$ del polinomio, si los Factores de Inflación de la Varianza (VIF por sus siglas en inglés) son menores de 10, se realiza mediante mínimos cuadrados según la ecuación (4.29):

$$\hat{\beta}_i = (X^TX)^{-1}X^TY \tag{4.29}$$

Pero si los VIF son mayores de 10 la estimación de los $\hat{\beta}_j$ se realiza a través del modelo de Regresión Ridge (RR) dado por (4.30) (Montgomery et al., 2005):

$$\hat{\beta}_i = (X^T X + Ik)^{-1} X^T Y \tag{4.30}$$

Aquí es donde juega un papel importante el valor de k ya que este modelo de regresión logra que los coeficientes se ajusten y tengan menor varianza, den estabilidad a las predicciones del modelo (Piña et al., 2005).

El método de RR, es útil para detectar la multicolinealidad dentro de un modelo de regresión $Y = X^t\hat{\beta} + \varepsilon$. El método fue propuesto por Hoerl y Kennard en 1970 y es usado para trabajar con modelos que presentan sesgo. La idea del método es simple y consiste en que dado que la matriz $X^T X$ es altamente condicionada o cercana a singular; es posible agregar constantes positivas a los elementos de la diagonal, para asegurar que la matriz resultante tenga un menor condicionamiento.

El vector de coeficientes de RR está dado por (4.30). Dado que el objetivo es un modelo polinomial cuya varianza de predicción sea mínima, se utilizó la función del cuadrado medio de la predicción (CMP) de mínimos cuadrados, para determinar el valor de la constante k donde el resultado final esta dada por (4.31) (Piña et al., 2005).

$$K = \frac{P\hat{\sigma}^2}{\hat{\beta}^t (X^t X)\hat{\beta}} \tag{4.31}$$

Otra manera propuesta por Hoerl et al. (Montgomery et al., 2005) esta dada por (4.32).

$$K = \frac{P\hat{\sigma}^2}{\hat{\beta}^t \hat{\beta}} \tag{4.32}$$

4.4.3. Análisis Residual

Existen otros métodos para el análisis y validación de los modelos mediante técnicas estadísticas, el análisis de residuales, el análisis de varianza (ANOVA) entre otros. Una

métrica una métrica utilizada para ver el comportamiento del modelo o validarlo es R^2, que es el coeficiente de determinación y que representa la proporción de la variación explicada por el regresor x. Los valores cercanos a uno implican que la mayor parte de la variabilidad de y esta explicada por el modelo (Montgomery et al., 2005). La R^2 esta dada por (4.33) y se formula como sigue:

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TT}} \tag{4.33}$$

Donde la SS_T es la suma de cuadrados total y fue definida por (4.18) y también se puede escribir de la siguiente manera, $SS_{TT} = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = y' \cdot y - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$. Y SS_{RES} es la suma de cuadrados de los residuales y se muestra a continuación:

$$SS_{RES} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = e' \cdot e = \sum_{i=1}^{n} e_i^2 \tag{4.34}$$

Existe también la métrica R^2_{Adj} que es el coeficiente de determinación ajustada y se define en (4.35).

$$R^2_{Adj} = 1 - \frac{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{(n-p)}}{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{(n-1)}} \tag{4.35}$$

La métrica R^2_{Adj} penaliza la adición de los términos que no son útiles y es ventajosa para evaluar y comparar modelos. En (4.35) los $(n-p)$ y $(n-1)$ son los grados de libertad, y las y_i son las salidas o respuestas, las $\hat{y}_i$ podrían ser las predicciones de la red neuronal y $\bar{y}$ es la respuesta media.

Nos referimos a las métricas R^2 y R^2_{Adj} como la cantidad de variabilidad en los datos explicados por los modelos.

El análisis residual se realiza principalmente con las siguientes métricas:

Se utiliza el cuadrado medio de los residuales $MS_{Res} = \frac{1}{n-p} \sum_{i=1}^{n} (y_i - \hat{y}_i)$, los cuales fueron definidos por (4.21) en la sección 4.4.1, donde los MS_{Res} pueden ser considerados

como la varianza promedio de los residuales del ajuste.

También existe el calculo del $PREMS$ (Prediction Mean Square) que es el error de predicción promedio al cuadrado.

$$PREMS = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \tag{4.36}$$

Donde n es la cantidad de datos y e son los residuales. El $PREMS$ se utiliza ya que si esta métrica es menor que el MS_{Res}, indica que el modelo será un buen predictor.

La siguiente métrica que se utiliza es el $PRESS$ (Prediction Sum Square) que es la suma de cuadrados del error de predicción, el cual es una medida de lo bien que funciona un modelo para predecir nuevos datos y se define en (4.37).

$$PRESS = \sum_{i=1}^{n}\left(\frac{e_i}{1-h_{ii}}\right) \tag{4.37}$$

Donde e son los residuales y h_{ii} es el i-ésimo elemento de la diagonal de la matriz de sombreo H.

El otro estadístico utilizado en el análisis residual es la $R^2_{Predicción}$, que es el coeficiente de determinación de la predicción.

$$R^2_{Predicción} = 1 - \frac{PRESS}{SS_T} \tag{4.38}$$

Este estadístico se compone del $PRESS$, que fue definido en (4.38) y por SS_T que es la suma de cuadrados total y fue definida por (4.18).

4.4.4. Distancia de Mahalanobis

El concepto de distancia se puede dar dados dos puntos x_i, x_j pertenecientes a $\Re^p$, diremos que hemos establecido una distancia, o una métrica, entre ellos si hemos definido una función d con las propiedades siguientes:

- $d : \Re^p \times \Re^p \rightarrow \Re^+$, es decir, dados dos puntos en el espacio de dimensiones p su distancia con esta función es un número no negativo, $d(x_i, x_j) \geq 0$;
- $d(x_i, x_j) = 0, \forall\, i, j$, la distancia entre un elemento y el mismo es cero.
- $d(x_i, x_j) = d(x_j, x_i)$, la distancia es una función simétrica en sus argumentos.
- $d(x_i, x_j) \leq d(x_j, x_p) + d(x_p, x_j)$, la distancia debe verificar que si tenemos tres puntos, la suma de las longitudes de dos lados cualesquiera del triángulo formado por los tres puntos debe siempre ser mayor que el tercer lado. Esta propiedad se conoce como desigualdad triangular.

Existen diferentes tipos de distancias como la distancia Minkowski, la distancias Promedio, la familia de Métricas Euclídeas Ponderadas, la distancia Euclídea, que es la mas utilizada, por ejemplo en la red neuronal de FBR (véase las ecuaciones (2.1) y (2.2) y la sección 4.1.1).

Otra medida es la distancia de Mahalanobis d_m (4.39), creada por P.C. Mahalanobis (1893-1972), el padre de la Estadística en la India, P.C. Mahalanobis había utilizado la medida de distancia que lleva su nombre para investigar las diferentes razas en la India; la distancia está definida y desarrollada como en (4.39).

$$\begin{aligned} d_2 &= \sqrt{\left(\tfrac{x-\mu}{\sigma}\right)\left(\tfrac{x-\mu}{\sigma}\right)} \\ d_2 &= \sqrt{(x-\mu)\left(\sigma^2\right)^{-1}(x-\mu)} \\ d_m &= \sqrt{(x-\mu)^T S^{-1}(x-\mu)} \\ d_m &= \sqrt{(x-\mu)^T \Sigma^{-1}(x-\mu)} \end{aligned} \tag{4.39}$$

Donde Σ es la matriz de varianzas y covarianzas, es cuadrada $m \times m$ y simétrica. Así mismo esta distancia de Mahalanobis se distribuye como una χ^2 con p grados de libertad. La función de densidad de probabilidad de una χ^2 de una variable aleatoria es

$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2} \qquad x > 0$ donde k es el número de grados de libertad. La media y varianza de la distribución χ^2 son k y $2k$ respectivamente. Nótese que la variable aleatoria χ^2 es no negativa y que la probabilidad de la distribución de probabilidad es sesgada a la derecha. Sin embargo, como k crece, la distribución llega a ser más simétrica. Como $k \to \infty$, la forma límite de la distribución χ^2 es la distribución normal. Se define la χ^2 como el punto de porcentaje o el valor de la variable aleatoria χ^2 con k grados de libertad tal que la probabilidad de que X^2 exceda este valor es α. Esto es $P(X > \chi^2) = \int_{\chi^2}^{\infty} f(u)\, du = \alpha$ (véase la demostración y para mas detalles de la distancia en Peña, 2002 y Härdle y Simar, 2007).

Capítulo 5

Metodología

A continuación se plantea la metodología para analizar, modelar y optimizar un proceso (en esta investigación es un proceso de soldadura); mediante modelos híbridos de sistemas inteligentes en combinación con métodos estadísticos.

5.1. Metodología Propuesta

De manera general la metodología aplicada consistió en:

1. Hacer un análisis estadístico preliminar, ésto quiere decir que la red FBR debe cumplir con las métricas del análisis residual, para determinar si los hallazgos encontrados dentro de la red neuronal de FBR pueden ser solucionados mediante los métodos de sistemas inteligentes y técnicas estadísticas.

2. Proponer soluciones posibles a los hallazgos encontrados en la red neuronal.

3. Sustentar y validar, mediante herramientas estadísticas, las propuestas hechas para solucionar los hallazgos.

4. Comparar y verificar si el modelo hibrido propuesto (compuesto por técnicas estadísticas y sistemas inteligentes) realiza una modelación y una optimización adecuadas en comparación con los modelos ya existentes.

5.1.1. Plan detallado seguido en esta investigación.

Los pasos seguidos en este trabajo de investigación fueron los siguientes:

- se revisó el estado del arte sobre aplicaciones de sistemas inteligentes y técnicas estadísticas en los procesos de soldadura.
- se realizó una descripción del proceso de soldadura y materiales involucrados.
- se planeó y desarrolló la etapa experimental.
- se modeló y cuando fue necesario se simuló el proceso.
- se validaron los modelos mediante análisis de varianza y análisis residual.
- se optimizó el proceso considerando un esquema global.
- se transfirieron los resultados al proceso de soldadura.

En la Figura 5.1 se presentan los pasos generales de la metodología.

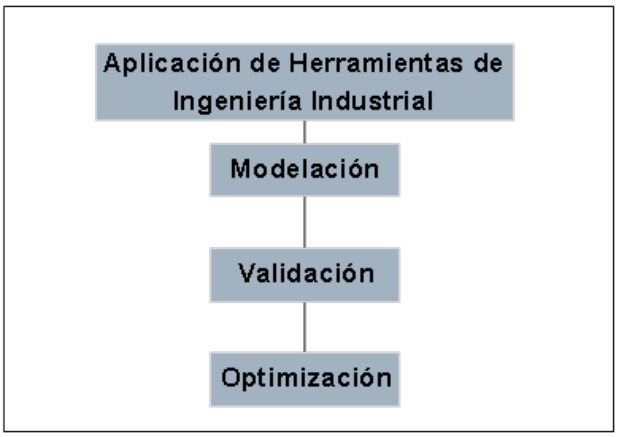

Figura 5.1: Esquema de la Metodología

La primera parte de la metodología consiste en la aplicación de herramientas de ingeniería industrial, que sigue el diagrama de flujo que se muestra en la Figura 5.2. La segunda

parte comprende la modelación, donde se aplicó la modificación a la red neuronal FBR. La tercera parte fue dedicada a la validación del modelo, mediante herramientas estadísticas. La última parte de la metodología es la de optimización de los modelos desarrollados para el proceso de soldadura.

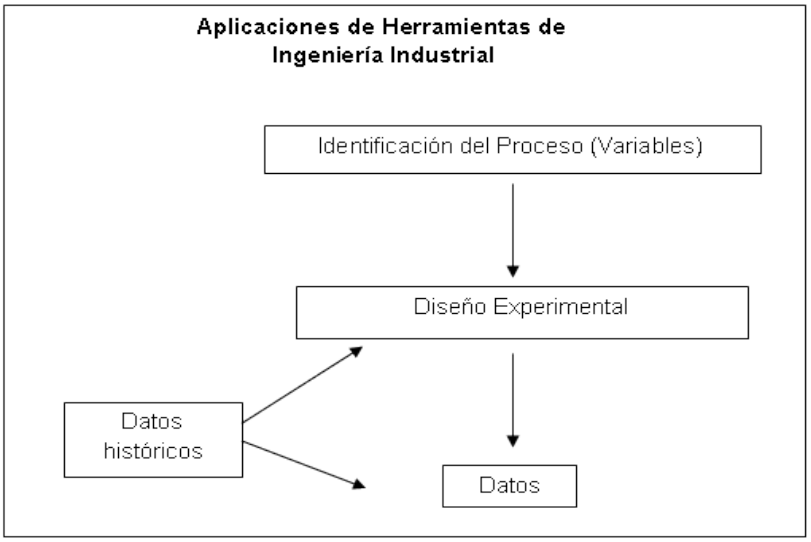

Figura 5.2: Diagrama de Aplicaciones de Ingeniería Industrial

En la Figura 5.2 se desglosa la parte de la aplicación de herramientas de ingeniería industrial; después de haber identificado las variables que se involucran en el proceso de soldadura, se planeó el diseño de experimentos tomando referencia de los datos históricos.

El siguiente paso de la metodología es la modelación, Figura 5.3, donde se presenta el flujo de los datos y algunas de las herramientas que se van a utilizar en el proceso de modelación. Se generó un modelo usando redes neuronales del tipo FBR la cual fue modificada en esta investigación, para ajustar el conjunto de datos y resultados con la red neuronal.

En la tercera etapa de la metodología Figura 5.4, se realizó la validación de los datos y modelos mediante las técnicas estadísticas (Praga-Alejo et al., 2010b Anexo B4), como el análisis de varianza (ANOVA), análisis de los residuales e indicadores estadísticos como $MS_{Residuales}$, R^2, $R^2_{Predicción}$.

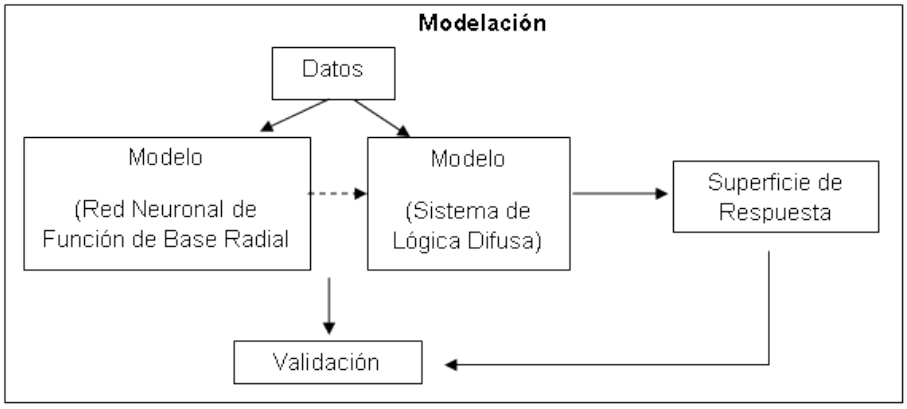

Figura 5.3: Diagrama de Modelación

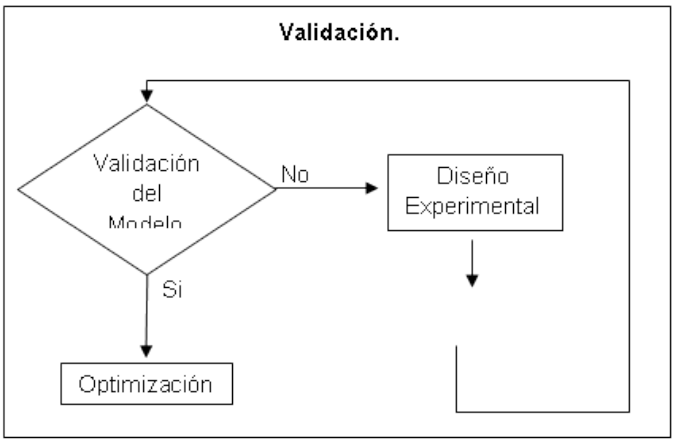

Figura 5.4: Diagrama de Validación

Cuando en la validación del modelo no fue posible obtener el resultado deseado, se planeó un nuevo ciclo iniciando con un nuevo diseño experimental, y una nueva validación, hasta tener una buena representación del proceso de soldadura.

En el cuarto paso de la metodología, Figura 5.5, se realizó la optimización del proceso, mediante algoritmos inspirados en la naturaleza, los cuales son una rama de los sistemas inteligentes, como: algoritmos evolutivos (EAs por sus siglas en inglés), algoritmos de estimación de distribuciones (EDAs por sus siglas en inglés), programación evolutiva, estrategias evolutivas.

El modelo o modelos desarrollados, ya sea la red neuronal de Función de Base Radial modificada (propuesta en esta disertación) o el sistema de Lógica Difusa, se utilizarán como función de evaluación para la optimización (Praga-Alejo et al., 2010a Anexo B3).

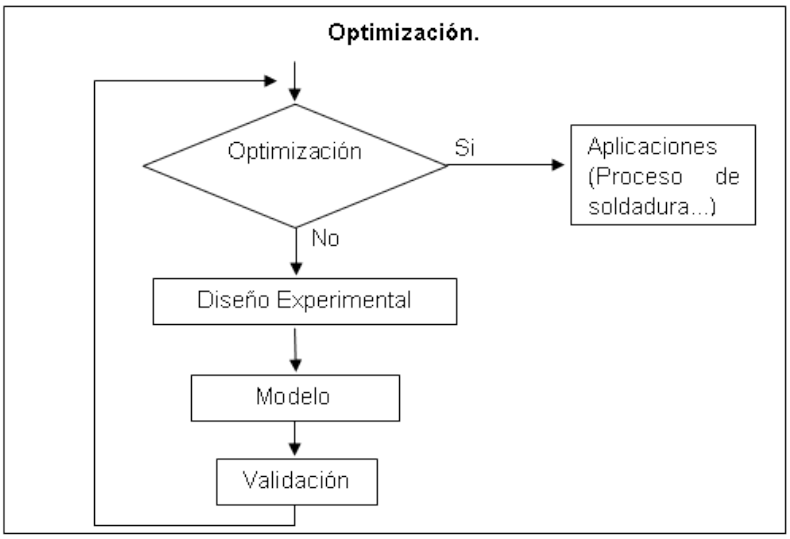

Figura 5.5: Diagrama de Optimización

Capítulo 6

Desarrollo experimental y resultados

En este capítulo se presentan el desarrollo experimental y los resultados.

Después de los hallazgos que presenta la red neuronal de FBR (ver capítulos 2 y 3), las modificaciones que se realizaron para mejorar el comportamiento de la red neuronal, fueron las siguientes:

1. Se reemplazó el método del cálculo de los centroides t_i de la red neuronal de FBR. El cálculo de los centroides se realizaba de forma aleatoria y mediante el método de Clustering K-medias (Haykin, 1999, Nelles, 2001, Arbib, 2003, Martín del Brío y Sanz, 2007). Para suplir estos métodos para el cálculo de los centroides t_i se implementó una técnica derivada de los algoritmos evolutivos en este caso un algoritmo genético.

2. Otra herramienta utilizada para mejorar la red neuronal es la distancia de Mahalanobis, donde se está eliminando el cálculo de la distancia Euclidiana $\|x - t_i\|^2$.

Estas herramientas, tanto de sistemas inteligentes como de técnicas estadísticas (algoritmos genéticos y distancia de Mahalanobis) modifican la red neuronal de FBR y permite reemplazar los métodos que se aplicaban comúnmente.

Los motivos por los cuales se utilizó la distancia de Mahalanobis son sus propiedades y beneficios sobre la distancia Euclidiana (Capítulo 4, sección 4.4.4). La distancia de

Mahalanobis tiene la ventaja sobre la distancia Euclidiana de que toma en cuenta la relación entre variables y encuentra una matriz de varianzas y covarianzas que ayuda por ejemplo en las desviaciones y varianzas tanto en la predicción como en el modelo, cosa que no hace la distancia Euclidiana y por eso esta distancia encuentra dificultades en el modelo. Por otra parte, la distancia de Mahalanobis es una medida de distancia, su utilidad es que es una forma de determinar la similitud entre dos variables aleatorias multidimensionales (Por ejemplo x y t_i en la red de FBR). Además la distancia de Mahalanobis se diferencia de la distancia Euclidiana en que toma en cuenta la correlación entre variables aleatorias (x y t_i); lo cual permite reescribir la ecuación (4.39) aplicada a la red neuronal de la siguiente manera:

$$d_m = \sqrt{(x - A)^T \Sigma^{-1} (x - A)} \tag{6.1}$$

Donde las x son las variables de entrada del proceso, las A son los centros calculados mediante el algoritmo genético y Σ^{-1} es la inversa de la matriz de covarianza. Así reconstruimos las demás ecuaciones quedando como:

$$G_R (d_m (x, A)) = \exp\left(-\sqrt{(x - A)^T \Sigma^{-1} (x - A)}\right) \tag{6.2}$$

$$\varphi_i (x) = G_R (d_m (x, A_i)) \tag{6.3}$$

La relación de entradas y salidas de la red neuronal propuesta está dada como:

$$y (x) = \sum_{i=1}^{n} w G_R (d_m (x, A_i)) + b \tag{6.4}$$

Donde $y (x_j) = d_j$, por ende d_j es la salida o respuesta del proceso a analizar, $d = [d_1, d_2, \ldots, d_N]^T$, las ecuaciones son escritas en forma matricial:

$$G_R w = d \tag{6.5}$$

Donde $w = [w_1, w_2, \ldots, w_n]^T$ son los pesos de la red neuronal que se determinan mediante Pseudoinversa y Mínimos Cuadrados.

$$\begin{aligned} w &= G_R^+ d \\ &= \left(G_R' G_R\right)^{-1} G_R' d \end{aligned} \tag{6.6}$$

Donde la matriz G_R Rediseñada y transformada mediante el algoritmo genético y la distancia de Mahalanobis está dada por (6.7):

$$G_R = \begin{bmatrix} G_R\left(d_m\left(x_1, A_1\right)\right) & G_R\left(d_m\left(x_1, A_2\right)\right) & \ldots & G_R\left(d_m\left(x_1, A_n\right)\right) \\ G_R\left(d_m\left(x_2, A_1\right)\right) & G_R\left(d_m\left(x_2, A_2\right)\right) & \ldots & G_R\left(d_m\left(x_2, A_n\right)\right) \\ \vdots & \vdots & \ddots & \vdots \\ G_R\left(d_m\left(x_N, A_1\right)\right) & G_R\left(d_m\left(x_N, A_2\right)\right) & \ldots & G_R\left(d_m\left(x_N, A_n\right)\right) \end{bmatrix} \tag{6.7}$$

Como se describe en el Capítulo 4, existen diferentes tipos de Funciones de Base Radial, de las cuales se aplica (4.1) en la red neuronal modificada y por lo tanto ésta se transforma y se reescribe de la siguiente manera:

$$\varphi_i\left(x\right) = \left(d_m\left(x, A\right)\right)^2 \cdot \ln\left(d_m\left(x, A\right)\right) \tag{6.8}$$

6.1. Aplicación y experimentación con la red neuronal de FBR Rediseñada.

La experimentación se realizó aplicando diferentes conjuntos de datos que corresponden a distintas corridas del procesos de soldadura; los resultados se compararon con modelos de regresión y modelos de redes neuronales mediante la predicción y el coeficiente de determinación R^2. Se realizó experimentación adicional con diversas tablas de procesos de soldadura.

6.1.1. Aplicación del algoritmo genético y distancia de Mahalanobis en FBR.

Se hizo una optimización de un proceso de aprendizaje hibrido sobre los parámetros de la red neuronal de FBR Rediseñada con un algoritmo genético como complemento.

En el algoritmo genético la función de evaluación para estos experimentos fue la R^2, que es el coeficiente de determinación definido en (4.33) como:

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TT}}$$

Con esto se requirió una solución de $A^* = (t_i \,|i = 1, 2, \ldots, m)$ donde el conjunto de centroides $\{t_i \,|1, 2, \ldots, m\}$ quedan determinados al maximizar la métrica R^2 en el modelo de red neuronal. El algoritmo genético utilizó una representación binaria de 16 de posibles soluciones que es necesaria porque este tipo de algoritmo manipula bits.

La población inicial del algoritmo es un vector de 250 líneas, el cual se representa por cada individuo y la matriz $16 \times (t_i \,|i = 1, 2, \ldots, m) = 16 \times (m \times m)$ es para representar cada bit de la solución binaria codificada. Cada individuo de la población representa un conjunto de $(m \times m)$ soluciones codificadas de 16 bits. Una selección de torneo de tamaño 2 con un solo punto de cruce y una simple mutación fue usada (Goldberg, 1989). Una probabilidad de cruce $Pc = 0.9$ y una mutación $Pm = 0.01$ fueron utilizadas. Como se mencionó anteriormente se utilizó la R^2 como función de evaluación y el algoritmo genético tuvo ciclos durante 50 generaciones.

Después de haber aplicado el algoritmo genético se reconstruyó la red de FBR con la distancia de Mahalanobis, con lo cual se aplicaron las ecuaciones desde (6.1) hasta (6.8).

Los primeros experimentos y comparaciones de la red FBR Rediseñada fueron con regresión lineal múltiple.

El primer experimento se realizó con los datos de la Tabla 6.1, el cual resume los datos y resultados de un proceso de soldadura con tres variables de entrada y una respuesta (Sampaio et al., 2005).

Tabla 6.1: Variables del Primer Experimento

Parámetros			Respuesta
Voltaje (V)	*Velocidad del Alambre (m/min)*	*Velocidad al Soldar (cm/min)*	*Penetración (mm)*
36.8	5.4	55	3.5
38.4	5.4	55	3.8
36.8	8.3	55	5.1
38.4	8.3	55	5.3
36.8	5.4	65	3
38.4	5.4	65	3.1
36.8	8.3	65	4.5
38.4	8.3	65	4.5
36	6.9	60	3.5
39.1	6.9	60	3.5
37.7	3.9	60	3.1
37.7	9.7	60	5.5
37.7	6.9	51.6	4.3
37.7	6.9	68.4	3.8
37.7	6.9	60	4
37.7	6.9	60	3.9
37.7	6.9	60	4
37.7	6.9	60	4
37.7	6.9	60	3.8
37.7	6.9	60	3.4

Los resultados se simularon usando: un modelo de regresión y la red neuronal de FBR con el método de Clustering K-medias, el método original sin modificaciones. Las Figuras 6.1 y 6.2, se presenta el comportamiento de ambos métodos.

Los resultados fueron los siguientes: la R^2 es igual a **84.19** usando el modelo de regresión y R^2 es igual a **67.25** usando la red FBR con Clustering K-medias y distancia Euclidiana. De acuerdo con el valor de R^2 el modelo de regresión es mejor que la red neuronal con el método de Clustering K-medias y distancia Euclidiana.

Los datos de la misma Tabla 6.1, se utilizaron para ajustar con la red neuronal de FBR Rediseñada (con el algoritmo genético y la distancia de Mahalanobis) y comparar con el modelo de regresión. De esto se tiene que la R^2 alcanzó un valor de **94.31**, con la red FBR Rediseñada. La Figura 6.3 muestra la comparación de los ajustes con la red neuronal FBR Rediseñada y con el modelo de regresión.

A continuación se presentan dos experimentos más usando los modelos de regresión y la red neuronal de FBR Rediseñada. La Tabla 6.2 corresponde a un proceso de soldadura

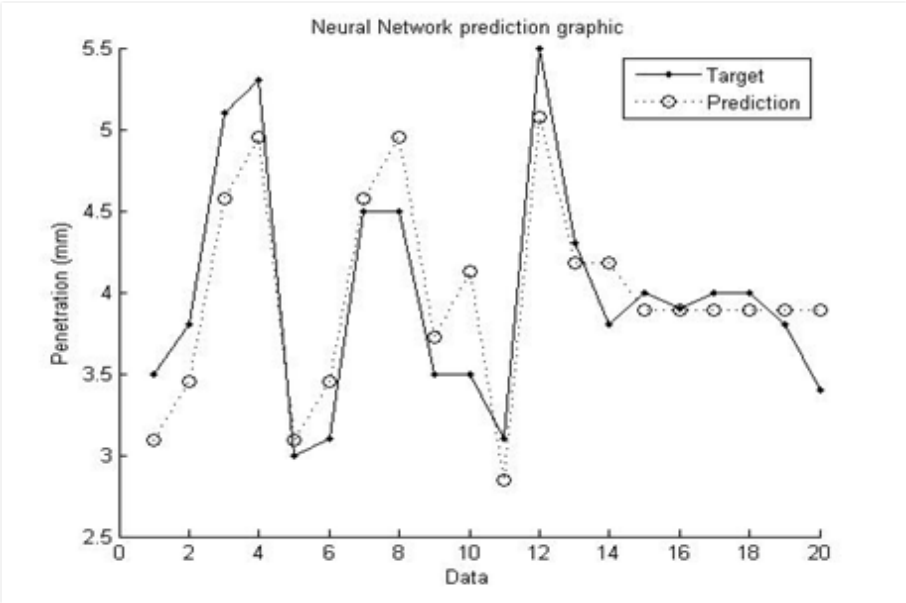

Figura 6.1: Ajuste mediante Red Neuronal

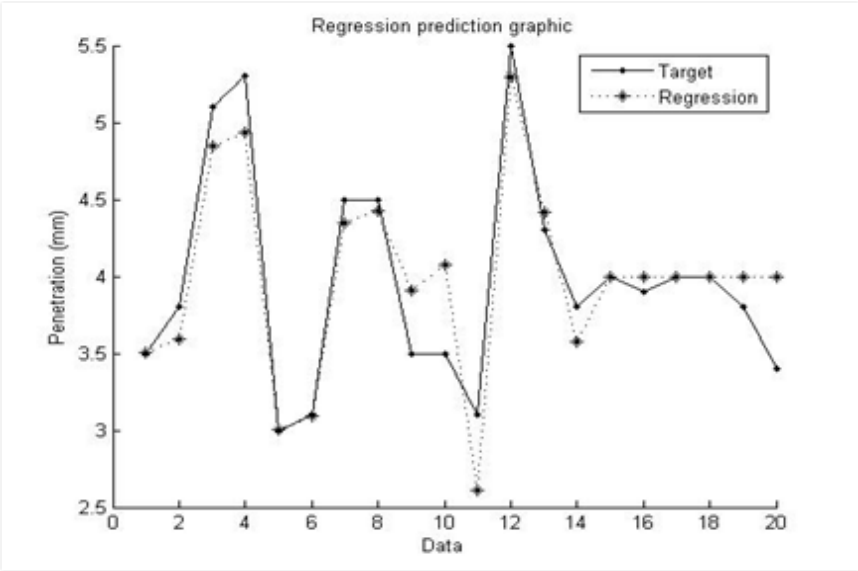

Figura 6.2: Ajuste mediante Regresión

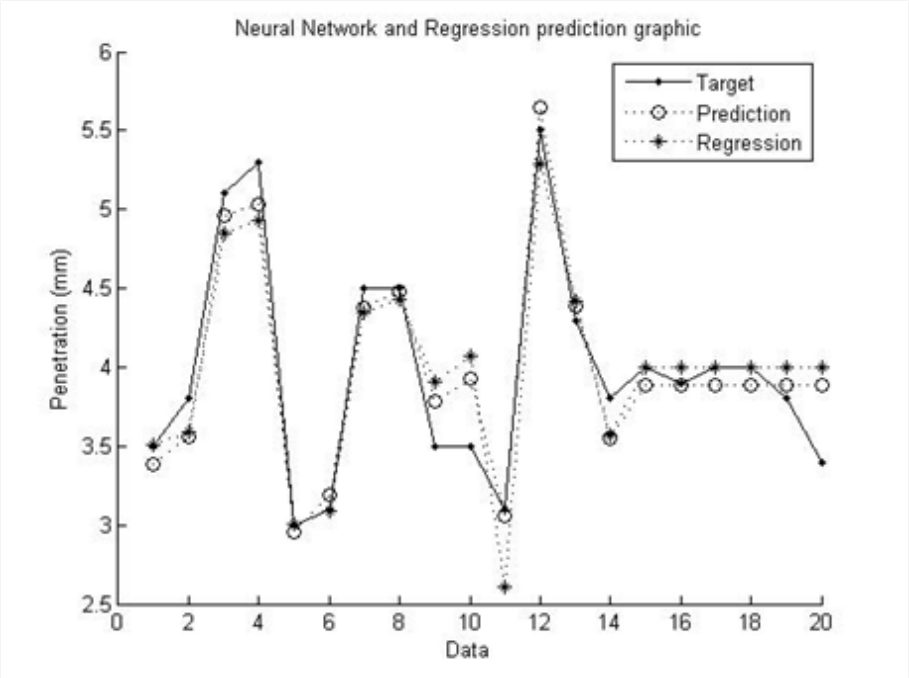

Figura 6.3: Ajuste mediante Regresión y FBR Rediseñada

con cuatro variables de entrada, amperaje, voltaje, velocidad al soldar y una interacción entre la primera variable y la segunda, con una respuesta de salida que es la penetración de la soldadura (Karadeniz et al., 2005).

Los resultados fueron los siguientes, la R^2 es igual a **93.91,** usando la regresión alcanza un valor de **98.33** usando la Red FBR modificada.

El experimento de la Tabla 6.3, tiene cuatro variables de entrada y una respuesta que corresponde a la penetración de un proceso de soldadura (Gunaraj y Murugan, 1999). Los resultados fueron los siguientes: la R^2 es igual a **63.78** mediante regresión y R^2 igual a **94.15** usando la red FBR modificada. Como se puede observar los resultados con la red neuronal de FBR Rediseñada tiene un ajuste adecuado y mejor en comparación con el modelo de regresión.

Con el fin de ilustrar la aplicación y la eficiencia de la red neuronal FBR Rediseñada, se realizó la comparación entre la FBR utilizando el método Clustering K-medias y la

Tabla 6.2: Variables del Segundo Experimento

Parámetros				Respuesta
Amperaje (Amper)	*Voltaje (V)*	*Velocidad al Soldar (cm/mm)*	*Amperaje * Voltaje*	*Penetración (mm)*
95	22	40	2090	2.46
95	22	60	2090	2.51
95	22	80	2090	2.36
95	24	40	2280	2.61
95	24	60	2280	2.67
95	24	80	2280	2.56
95	26	40	2470	2.64
95	26	60	2470	2.69
95	26	80	2470	2.59
105	22	40	2310	2.84
105	22	60	2310	2.92
105	22	80	2310	2.79
105	24	40	2520	2.89
105	24	60	2520	2.97
105	24	80	2520	2.84
105	26	40	2730	2.98
105	26	60	2730	3.01
105	26	80	2730	2.91
115	22	40	2530	3.02
115	22	60	2530	3.06
115	22	80	2530	2.94
115	24	40	2760	3.11
115	24	60	2760	3.14
115	24	80	2760	3.04
115	26	40	2990	3.19
115	26	60	2990	3.24
115	26	80	2990	3.16

Tabla 6.3: Variables del Tercer Experimento

Parámetros				Respuesta
Voltaje (V)	*Velocidad de Alambre (m/min)*	*Velocidad al Soldar (m/min)*	*Distancia hacia la pieza (mm)*	*Penetración (mm)*
32.5	0.93	0.51	32.5	2.7
37.5	0.93	0.51	32.5	3.2
32.5	1.39	0.51	32.5	2.9
37.5	1.39	0.51	32.5	4
32.5	0.93	0.67	32.5	2.2
37.5	0.93	0.67	32.5	2.3
32.5	1.39	0.67	32.5	2.9
37.5	1.39	0.67	32.5	3.3
32.5	0.93	0.51	37.5	2.2
37.5	0.93	0.51	37.5	2.9
32.5	1.39	0.51	37.5	2.7
37.5	1.39	0.51	37.5	3.9
32.5	0.93	0.67	37.5	2.6
37.5	0.93	0.67	37.5	2.5
32.5	1.39	0.67	37.5	3
37.5	1.39	0.67	37.5	3.4
30	1.16	0.59	35	1.9
40	1.16	0.59	35	2.8
35	0.7	0.59	35	2.9
35	1.62	0.59	35	4.5
35	1.16	0.43	35	3.2
35	1.16	0.75	35	2.6
35	1.16	0.59	30	3.1
35	1.16	0.59	40	2.8
35	1.16	0.59	35	3
35	1.16	0.59	35	2.8
35	1.16	0.59	35	2.9
35	1.16	0.59	35	3.1
35	1.16	0.59	35	2.9
35	1.16	0.59	35	2.9
35	1.16	0.59	35	2.8

distancia Euclidiana y la red FBR propuesta.

La Tabla 6.4 presenta los resultados de las métricas R^2 y R^2_{Adj} que muestran que la red de FBR Rediseñada usando el algoritmo evolutivo y la distancia de Mahalanobis tiene mejor comportamiento que la red de FBR con el método de Clustering K-medias y distancia Euclidiana.

Se puede observar que la FBR Rediseñada está por encima del 93 % en R^2 y R^2_{Adj}; lo cual quiere decir que el 93 % de la variabilidad de la respuesta esta explicada por el modelo. Además la red FBR con Clustering K-medias alcanzó valores inferiores al 84 % esto significa que los modelos explican el 10 % menos que la red FBR Rediseñada.

Tabla 6.4: Resultados de la comparación entre Redes Neuronales FBR

	Resultados usando la Tabla 6.1		Resultados usando la Tabla 6.2		Resultados usando la Tabla 6.3	
Métricas Estadísticas	FBR con Clustering K-medias	FBR Rediseñada	FBR con Clustering K-medias	FBR Rediseñada	FBR con Clustering K-medias	FBR Rediseñada
R^2	67.25	***94.31***	83.69	***98.33***	51.16	***94.15***
R^2_{Adj}	61.11	***93.24***	80.73	***98.03***	43.65	***93.26***

6.1.2. Análisis Residual y de Varianza.

Por otra parte aunado a estas comparaciones también se desarrolló todo el análisis y validación de los modelos mediante técnicas estadísticas, como Análisis de Residuales y Análisis de Varianza (ANOVA).

El Análisis Residual se realizó con el cuadrado medio de los residuales $MS_{Res} = \frac{1}{n-p}\sum_{i=1}^{n}(y_i - \hat{y}_i)$ donde los MS_{Res} pueden ser considerados como la varianza promedio de los residuales del ajuste.

Se calculó el $PREMS$ que es el error de predicción promedio al cuadrado y definido en (4.36) como:

$$PREMS = \frac{1}{n}\sum_{i=1}^{n} e_i^2$$

La siguiente métrica utilizada fue el $PRESS$ (Prediction Sum Square) que es la suma de cuadrados del error de predicción, definido en (4.37).

$$PRESS = \sum_{i=1}^{n} \left(\frac{e_i}{1 - h_{ii}} \right)$$

El otro parámetro estadístico utilizado en el análisis residual fue la $R^2_{Predicción}$ cuya ecuación está dada en (4.38) que es el coeficiente de determinación de la predicción.

$$R^2_{Predicción} = 1 - \frac{PRESS}{SS_T}$$

Para la aplicación de los métodos estadísticos del análisis residual y los supuestos del ANOVA y construcción de la misma (ver Capítulo 4 de la sección 4.4) se desarrolló el siguiente experimento en campo.

Se realizó un diseño de experimento central compuesto (DCC) de un proceso de soldadura que en este caso fue del tipo Láser. El cálculo arrojó la Tabla 6.5. En este tipo de procesamiento el rayo láser tipo pulsado es utilizado para crear una unión donde la zona afectada térmicamente sea estrecha, para reducir la rugosidad de la superficie de la unión y eliminar los efectos mecánicos que ocasionan otros procesos de soldadura.

Este tipo de soldadura tiene tres variables de entrada y una de salida; las variables de entrada son: el ancho del spot (el cual involucra la energía por pulso y la densidad de potencia), la frecuencia y la velocidad de barrido del láser. En el experimento éstos parámetros fueron seleccionados para la optimización de la energía consumida por unidad de área tratada. La variable de salida fue la penetración de la soldadura en la pieza.

A los datos de la Tabla 6.5 se les aplicó la red neuronal de FBR Rediseñada, se aplicaron también el ANOVA y los estadísticos de los residuales; los resultados fueron los siguientes.

El modelo de la red FBR modificada aplicado a los datos de la Tabla 6.5 generó una $R^2 = 94.68$ y una $R^2_{Adj} = 93.68$, un $PREMS = 0.0056$, un $PRESS = 0.1623$, un $MS_{Res} = 0.0071$ (los cuadrados medios del error), un $R^2_{Predicción} = 92.42$. Este resultado

Tabla 6.5: Resultados del proceso Láser

Parámetros			Respuesta
Velocidad (mm/seg)	Ancho (mm)	Frecuencia (Hertz)	Penetración Real (mm)
1.1	4	3.5	0.6889
1.1	5.68179283	3.5	0.2953
1.1	4	3.5	0.6186
1.77271713	4	3.5	0.6341
1.5	3	3	0.928
0.7	5	4	0.433
0.7	3	3	1.0053
0.7	5	3	0.5568
1.5	5	3	0.5413
1.5	3	4	1.1599
1.1	4	3.5	0.6186
1.1	2.31820717	3.5	1.5466
1.1	4	3.5	0.5413
1.1	4	3.5	0.433
1.1	4	4.34089642	0.4794
1.1	4	3.5	0.4952
1.1	4	2.65910358	0.4021
0.42728287	4	3.5	0.3557
0.7	3	4	1.2218
1.5	5	4	0.3403

indica que el modelo es un buen predictor ya que el $PREMS$ es menor que los cuadrados medios MS_{Res} (Montgomery et al., 2005). El valor del coeficiente de determinación de la predicción de $R^2_{Predicción} = 92.42$ indica que la red neuronal FBR Rediseñada explica más del 92 % de la variabilidad de los datos durante la predicción.

Así los resultados muestran que el modelo FBR Rediseñada es adecuado para la predicción y optimización. Para validar aún mejor el modelo se construyó la ANOVA y una validación en campo.

La ANOVA de la red FBR Rediseñada se muestra en la Tabla 6.6; cabe mencionar que dicha ANOVA tiene que cumplir con los supuestos de la misma (véase capítulo 4, sección 4.4.1). Las pruebas de normalidad para los residuales, tipo Anderson-Darling y Kolmogorov-Smirnov, arrojaron que los residuales son normales; la prueba de normalidad Anderson-Darling dió un valor de 0.235 y el valor de P de 0.759. Esto significa que los residuales se distribuyen normalmente; una grafica de los residuales se muestra en la figura 6.4.

De acuerdo con estos resultados de la prueba de normalidad de los residuales se con-

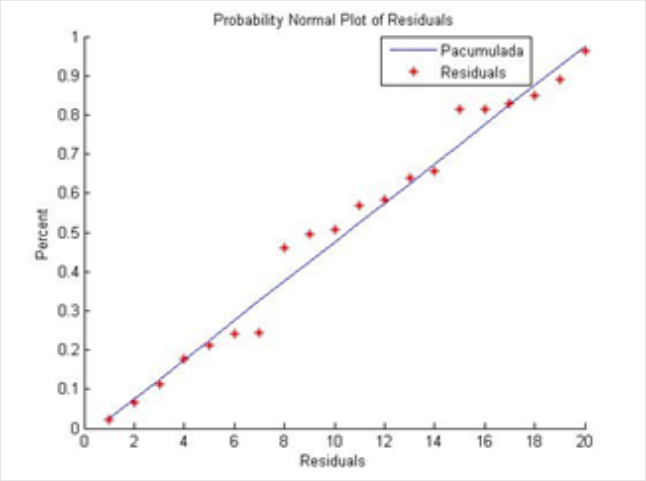

Figura 6.4: Ajuste de los Residuales

struyó el ANOVA que indica y desglosa el comportamiento del modelo de la red neuronal, ya que determina que el modelo es adecuado para predecir. Además el ANOVA permite determinar la significancia estadística, la relación entre una variable y dependiente y una o varias variables independientes x, y cuanto contribuye cada factor a la variación total de la respuesta.

Tabla 6.6: ANOVA en FBR Rediseñada

Análisis de Varianza					
Fuente de Variación	Grados de Libertad	Suma de Cuadrados	Cuadrados Medios	F	P
Modelo	3	2.02914	0.67638	94.96	0.0000
Residual o Error	16	0.11397	0.00712		
Total	19	2.14310			
	S = 0.084397	R-Sq = 94.6%		R-Sq(adj) = 93.6%	

6.2. Resultados y Validación en Campo.

Con el análisis anterior se puede decir que el modelo de red neuronal FBR Rediseñada usado para modelar el proceso que corresponde a los datos de la Tabla 6.5 es adecuado y cumple con todos los supuestos de un análisis de varianza; se puede utilizar para predecir

y optimizar el proceso físico. La predicción de los resultados con el modelo de red FBR Rediseñada se muestra en la Figura 6.5 y en la Tabla 6.7; se aprecia como la predicción presenta un comportamiento muy similar a la de los datos reales.

Además de este análisis se realizó la validación cruzada de la red FBR Rediseñada usando la métrica $PRESS$ (Chen et al., 1999; Montgomery et al., 2005; Chiddawar y Babu, 2010); donde el mejor modelo es el que tiene menor valor residual. Esta validación se realizó comparando el modelo de FBR Rediseñada con la distancia Mahalanobis y el Algoritmo Genético con la red FBR con los métodos de Clustering K-medias y distancia Euclidiana.

Los resultados de comparación mostraron que la FBR Rediseñada tuvo unos residuales $PRESS = 1.7033$; mientras que la red de FBR no modificada tuvo unos residuales $PRESS = 5.2641$. Con esto se puede inferir que la red de FBR Rediseñada es mejor que la red FBR con los métodos de Clustering K-medias y distancia Euclidiana

Para la validación final del modelo, se realizó la comparación de la predicción de la red propuesta contra los resultados de las pruebas de campo que se muestran en la Tabla 6.7. Los parámetros de las corridas 10, 12 y 19 (Tabla 6.7) que corresponden a corridas con mejor penetración y mejor calidad de la soldadura, fueron usados en la simulación con la red neuronal propuesta.

Los parámetros de la prueba 10 fueron: velocidad de 1.5, ancho de 3 y una frecuencia de 4; mientras que los resultados de penetración fueron: la experimental de 1.0826 y la de predicción de 1.1607 (Figura 6.6).

Los parámetros de la prueba 12 fueron: velocidad = 1.1, ancho = 2.31 y frecuencia = 3.5; mientras que los resultados de penetración fueron: la experimental de 1.5468 y la de predicción de 1.16008 (Figura 6.7).

Los parámetros de la prueba 19 fueron: velocidad = 0.7, ancho = 3 y frecuencia = 4; mientras que los resultados de penetración fueron: la experimental de 1.2232 y la de predicción de 1.1904 (Figura 6.8).

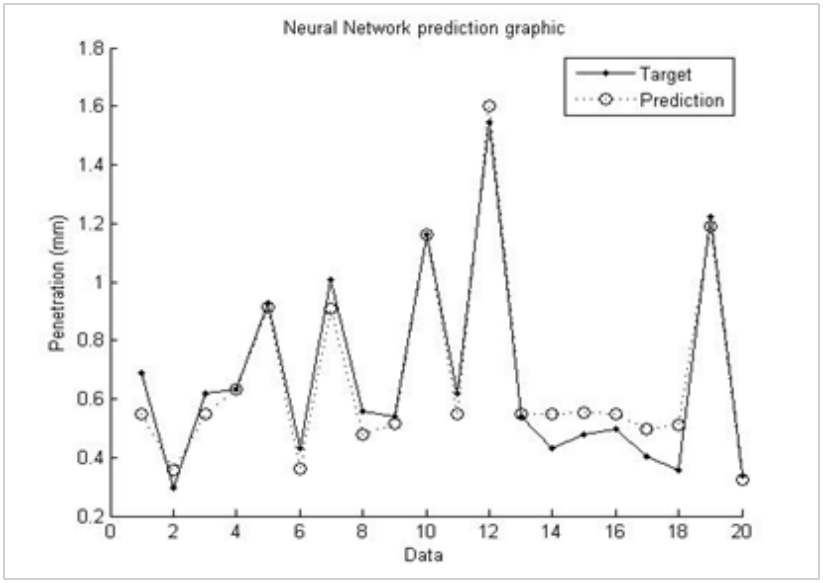

Figura 6.5: Predicción de Red Neuronal FBR Rediseñada

Tabla 6.7: Predicciones con FBR Rediseñada

Número de Corrida	Penetración Predicha por RBF Rediseñada	Velocidad (mm/seg)	Ancho (mm)	Frecuencia (Hertz)
1	0.5491	1.1	4	3.5
2	0.3576	1.1	5.68179283	3.5
3	0.5491	1.1	4	3.5
4	0.6324	1.77271713	4	3.5
5	0.9146	1.5	3	3
6	0.3597	0.7	5	4
7	0.9107	0.7	3	3
8	0.4768	0.7	5	3
9	0.5138	1.5	5	3
10	1.1607	1.5	3	4
11	0.5491	1.1	4	3.5
12	1.6008	1.1	2.31820717	3.5
13	0.5491	1.1	4	3.5
14	0.5491	1.1	4	3.5
15	0.5511	1.1	4	4.34089642
16	0.5491	1.1	4	3.5
17	0.4959	1.1	4	2.65910358
18	0.5126	0.42728287	4	3.5
19	1.1904	0.7	3	4
20	0.3237	1.5	5	4

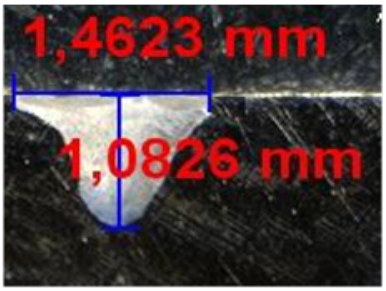

Figura 6.6: Penetración en el proceso Láser (Prueba 1)

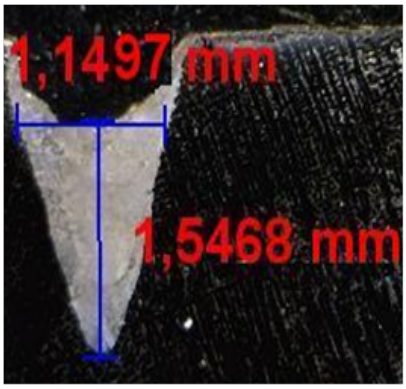

Figura 6.7: Penetración en el proceso Láser (Prueba 2)

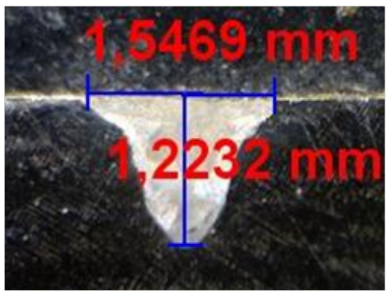

Figura 6.8: Penetración en el proceso Láser (Prueba 3)

Con estos resultados se puede inferir que la red neuronal de FBR Rediseñada se comporta adecuadamente y hace una predicción y optimización adecuada, ya que la tercera prueba presenta una muy buena concordancia con los resultados del proceso de soldadura.

Capítulo 7

Conclusiones

En esta tesis se presenta el trabajo de investigación sobre la utilización de redes neuronales para modelar y optimizar procesos de soldadura láser aplicado a materiales de acero. Los resultados obtenidos permitieron llegar a las siguientes conclusiones:

Se realizó un análisis de la habilidad de la red neuronal del tipo FBR para optimizar procesos tecnológicos como la soldadura de piezas metálicas cuya complejidad los convierte en candidatos muy pobres para intentar modelos matemáticos de naturaleza determinística basados en las leyes de la física y la termodinámica.

Se ha desarrollado un nuevo método numérico para modelar y optimizar el proceso de soldadura láser basado en una red neuronal tipo FBR; en el cual se han mejorado el método comúnmente usado de ensayo y error para realizar el ajuste de los parámetros del modelo heurístico consistente en una red neuronal tipo FBR.

Una red neuronal del tipo FBR fue combinada con métodos estadísticos de distancia de Mahalanobis y análisis residual con sistemas inteligentes como el algoritmo genético, para converger de manera rápida y obtener un resultado más preciso que cuando se aplica una definición aleatoria de los centroides de una red neuronal.

El método híbrido propuesto en comparación con el de una red neuronal resultó mejor.

Por otra parte en esta investigación se pudo encontrar un mejor modelo de red neuronal de Base Radial; ya que el método para calcular la distancia Euclidiana dada por $\|x - t_i\|^2$

y la forma de encontrar la variable t_i (centroides), causan un efecto sobre la desviación estándar de la predicción y como consecuencia una predicción y optimización deficiente. Este problema pudo ser resuelto mediante la modificación de la distancia y los llamados centroides.

Como conclusión se puede decir que la red neuronal rediseñada es mucho mejor que otros métodos que se utilizan actualmente en la red neuronal de FBR. Dado que el método del proceso de aprendizaje híbrido usa un algoritmo genético para calcular la matriz de centros y luego calcula la distancia de Mahalanobis maximizando el coeficiente de determinación R^2; este estadístico se convierte en la función de evaluación del algoritmo genético, lo cual incrementa tanto la precisión de la predicción como la convergencia de la red de FBR.

Con esto se puede ver que los resultados obtenidos por la red neuronal de FBR rediseñada con el algoritmo genético y la distancia de Mahalanobis tienen un ajuste mucho mejor en comparación con el método de Clustering K-medias aplicando la distancia Euclidiana. La red neuronal de FBR rediseñada funciona mejor en el sentido de las métricas R^2 y R^2_{Adj}. Nótese que si la R^2 y R^2_{Adj} son muy cercanas al 100 % el modelo será mejor.

Las métricas R^2 y R^2_{Adj} en la FBR rediseñada son mayores que en el método de FBR con Clustering K-medias y distancia Euclidiana en los tres casos presentados y esto nos indica que el modelo propuesto (FBR Rediseñada) es mejor que los métodos comúnmente aplicados (Tabla 6.4).

Como conclusión para estos casos, los métodos estadísticos son un buen método alternativo para validar la eficacia de los modelos de redes neuronales y en ese sentido la red de FBR rediseñada es una buena herramienta para la predicción y optimización.

Por otra parte en esta investigación se aplicó y se verificó el comportamiento del algoritmo evolutivo para encontrar un valor óptimo de la penetración en un proceso de soldadura, utilizando la red neuronal de FBR rediseñada; a la cual se le pudo incluir para este tipo de problemas una herramienta estadística como el diseño central compuesto que

se usó en campo en un proceso real de soldadura láser.

El modelo de la red FBR modificada aplicando los datos y resultados de la Tabla 6.5 permite explicar más del 92 % de la variabilidad de los datos; por lo tanto con estos resultados el modelo de FBR rediseñada es adecuado para la predicción y optimización del proceso de soldadura láser.

Por otra parte estas métricas R^2 y R^2_{Adj} fueron las magnitudes usadas para indicar si el modelo es adecuado para predecir y optimizar; aunado a estas métricas se aplicaron los estadísticos $PREMS$, el $PRESS$, los cuadrados medios del error MS_{Res} y el coeficiente de determinación de la predicción $R^2_{Predicción}$, para el análisis residual de la red neuronal modificada.

La mejoría de la red FBR rediseñada se midió con las métricas del análisis de residuales; una de estas es la $R^2_{Predicción}$ de la predicción, los cuadrados medios de los residuales MS_{Res} y el $PREMS$, los cuales nos indican que el modelo será un buen predictor si el $PREMS$ es menor que los cuadrados medios de los residuales MS_{Res}.

La aplicación de estos estadísticos permitió ver que el modelo de FBR rediseñada es adecuado para la predicción y optimización. Para validar mejor los datos se aplicaron la ANOVA y una validación en campo (Capítulo 6, sección 6.1.2 y 6.2).

Como se puede observar el valor de P (Tabla 6.6 del ANOVA) para la red FBR rediseñada es muy pequeño; con esto la conclusión es que la penetración en el proceso de soldadura láser se relaciona con las variables dependientes. Esto en conjunto con los estadísticos del análisis residual implica que la relación que se encontró es adecuada para predecir y optimizar la penetración en el proceso de soldadura con el modelo de FBR rediseñada.

Esta aportación es importante ya que se logró medir y validar un modelo de red neuronal con el análisis estadístico. Esto indica que estos métodos de diagnostico (Análisis residual y ANOVA) son útiles para validar los modelos de redes neuronales generados; y con esto se llegó a describir métodos para manejar inadecuaciones en el modelo propuesto

(red neuronal de FBR rediseñada).

Esta aportación importante se logró cumplir ya que la red neuronal de FBR rediseñada cumple con los supuestos básicos para usar el ANOVA por ejemplo la distribución de los residuales deben cumplir lo siguiente $e \approx N(0, \sigma^2)$; quiere decir que los errores o residuales del modelo deben tener $\mu = 0$, σ^2 conocida y deben de estar normalmente distribuidos (Montgomery, 2006). Para verificar las pruebas de Normalidad se utilizaron la prueba Kolmogorov-Smirnov y la Anderson-Darling. De acuerdo con los resultados de las pruebas de normalidad de los residuales se construyó el ANOVA cumpliendo con sus supuestos, lo cual fue muy útil ya que determinó valores estadísticos importantes para saber si el modelo era adecuado para optimización y predicción. La ANOVA que se construyó en la red neuronal rediseñada fue importante porque indicó y desglosó el comportamiento del modelo de red neuronal propuesto.

Por otra parte en caso de que no cumpliese con la normalidad de los residuales se propone una transformación de los mismos. Esta transformación se haría con el desarrollo Box-Cox, en casos atípicos, ya que por la estructura misma de la red de FBR siempre cumple con el supuesto de normalidad en los residuales, y en caso de tener una situación atípica se propone la transformación Box-Cox, para cumplir con la normalidad de los residuales y así construir el ANOVA para la red neuronal de FBR rediseñada.

Aunado a esto, las características modificadas de la red neuronal de Función de Base Radial ayudaron a mejorar el comportamiento de la red neuronal y se comprobó la hipótesis que la red neuronal de FBR rediseñada fue mejor que el modelo de red de FBR con el método de Clustering k-medias con la distancia Euclidiana con respecto a las métricas estadísticas de R^2, $R^2_{Ajustada}$, $R^2_{Predicción}$, $PRESS$ y Cuadrados Medios de los Residuales MS_{Res}.

La mejoría también se puede ver con la validación cruzada. La métrica utilizada para este método fue la de los residuales $PRESS$; el mejor modelo será el que tenga menor residual $PRESS$. Los resultados mostraron que para el caso del proceso de soldadura

Láser la red FBR rediseñada tuvo unos residuales $PRESS = 1.7033$; mientras que, la otra red de FBR tuvo unos residuales $PRESS = 5.2641$.

En cuanto al trabajo futuro se puede aplicar otros algoritmos evolutivos, para encontrar un valor óptimo de la penetración de un proceso de soldadura utilizando la red neuronal de FBR rediseñada como una función de evaluación.

Además, el trabajo futuro podría incluir una comparación para este tipo de problemas usando cualquier otro enfoque de estadística como la metodología de superficie de respuesta, con otros algoritmos de Soft Computing; es decir, algoritmos evolutivos como PSO, Simulated Annealing, etc., algoritmos de estimación de distribuciones EDAs como PBIL, BOA, EVONORM, etc.

La aplicación de estas metodologías y la red neuronal rediseñada podrían mejorar la optimización multiobjetivo, la cual no se exploró en este trabajo. Además de estos experimentos, se puede aplicar otra técnica estadística, como es el Análisis Canónico para el modelo de red neuronal de FBR rediseñada.

Otra investigación interesante para el futuro es aplicar el análisis residual y análisis de varianza (ANOVA), a otro tipo de redes neuronales y en general a las técnicas derivadas de los sistemas inteligentes; así como la creación de intervalos de confianza para este tipo de metodologías.

Bibliografía

[1] Abdalla Jamal A. and Hawileh Rami. (2010). Modeling and simulation of low-cycle fatigue life of steel reinforcing bars using artificial neural network. Journal of the Franklin Institute

[2] Allen T.T., Richardson R.W., Tagliabue D.P. and Maul G.P. (2002). Statistical process design for robotic GMAW of sheet metal. Welding Journal. Vol. May pp. 69-77.

[3] Arbib Michael A. (2003). *The Handbook of Brain Theory and Neural Networks.* The MIT Press. London, England.

[4] Babuska Robert. (2001). Fuzzy and Neural Control. Delft University of Technology, Holanda.

[5] Benyounis K.Y. and Olabi A.G. (2008). Optimization of different welding process using statistical and numerical approaches – A reference guide. Advances in Engineering-Software. Elsevier. 39, 483-496.

[6] Billings Steve A. and Zheng Guang L. (1995). Radial Basis Function Network Configuration Using Genetic Algorithms. Neural Networks. Vol. 8 No. 6, pp. 877-890.

[7] Cepeda Francisco. (2010). Análisis y optimización de parámetros empleados en la unión y tratamientos superficiales mediante laser CO_2 y Nd-YAG en la aleación de cobalto ASTM F-75. Master Thesis. COMIMSA.

[8] Chen S., Wu Y., and Luk B. L. (1999). Combined Genetic Algorithm Optimization and Regularized Orthogonal Least Squares Learning for Radial Basis Function Networks. IEEE Transactions on Neural Networks, Vol. 10, No. 5 pp. 1239-1243.

[9] Chiddarwar Shital S. and Babu N. Ramesh. (2010). Comparison of RBF and MLP neural networks to solve kinematic problem for 6R serial robot by a fusion approach. Engineering Applications of Artificial Intelligence. 23, pp. 1083-1092.

[10] D'Agostino R.B. and Stephens M.A. Eds. (1986). *Goodness-of-Fit Techniques*, Marcel Dekker.

[11] Ding Hongkai, Xiao Yunshi and Yue Jiguang. (2005). Adaptive Training of Radial Basis Function Networks using Particle Swarm Optimization Algorithm. L. Wang, K. Chen , and Y.S. Ong (Eds.): ICNC 2005, LNCS 3610, Springer-Verlag Berlin Heidelberg. pp. 119-128.

[12] Du Dajun, Li Kang and Fei Minrui. (2010). A fast multi-output RBF neural network construction method. Neurocomputing. 73, pp. 2196-2202.

[13] Fullér Robert. (1995). *Neural Fuzzy Systems.* Abo Akademi University. Turquía.

[14] García-Sánchez A., Trejo-Aguirre Mario F., Praga-Alejo R. J., Torres-Treviño L. M., Acevedo-Dávila J.L., García G., and M. Garza-Castañón. (2008). Optimización de los parámetros de soldadura por proceso de arco eléctrico con protección de gas (GMAW) para soldar un acero A-487. *14 Convención Científica de Ingeniería y Arquitectura*, Ministerio de Educación Superior, La Habana, Cuba. V Congreso de Ingeniería Mecánica y Metalúrgica Memorias, Pág. 867-872 ISBN 978-959-261-281-5.

[15] García P. (2005). Las Redes Neuronales Supervisadas. www.telefonica.net/web2/pgestevez/publicaciones.htm

[16] Goldberg David E. (1989). *Genetic algorithms in search, optimization, and machine learning,* Oxford University Press, New York. USA.

[17] _____. (2002). *The Design of Innovation, Genetic Algorithms and Evolutionary Computation.* USA, Kluwer Academic Publishers.

[18] Gregorcic G. and Lightbody G. (2008). Nonlinear system identification: From multiple-model networks to Gaussian processes. Engineering Applications of Artificial Intelligence. 21(7): 1035-1055.

[19] Gudur P. P. and Dixit U. S. (2008). A neural network-assisted finite element analysis of cold flan rolling. Engineering Applications of Artificial Intelligence, No. 21, pp. 43-52.

[20] Gunaraj V. and Murugan N. (1999). Application of response surface methodology for predicting weld bead quality in submerged arc welding of pipes. Journal of Materials Processing Technology, 88 pp. 266–275.

[21] _____. (2000). Prediction and Optimization of Weld Bead Volume for the Submerged Arc Process-Part 1. Welding Journal Magazine, Published by the American Welding Society (AWS). pp.: 286-s to 294-s

[22] Hagan, Demuth and Beale. (1996). *Neural Network Design.* Thomson Publishing Inc. EUA.

[23] Hamill Jack A. (1993). Fusion welding P/M components for automotive applications. SAE technical paper. SAE international congress. Detroit, USA.

[24] Härdle Wolfgang and Simar Léopold. (2007). *Applied Multivariate Statistical Analysis.* Springer-Verlag Berlin Heidelberg 2003, second edition 2007.

[25] Harnisch Ildefonso, Sanhueza H. Raúl, and Díaz R. Horacio. (2000). Despacho económico con unidades de características no convexas empleando algoritmos genéticos,

Revista Facultad de Ingeniería. Enero-junio, vol. 7, Universidad de Tarapacá, Arica, Chile. pp. 13-19.

[26] Haykin Simon. (1999). *Neural Networks: A comprehensive foundation.* Prentice Hall.

[27] Holland John H. (1975). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence.* University of Michigan Press, Ann Arbor.

[28] Jang J.- S. R., Sun C.- T. and Mizutani E. (1997). *Neuro-fuzzy and Soft Computing: A computational approach to learning and machine intelligence.* Prentice Hall, Inc. USA.

[29] *Jefferson's welding encyclopedia.* (1997). 18th edition, American Welding Society.

[30] Karadeniz Erdal, Ozsarac Ugur, and Yildiz Ceyhan. (2005). The effect of process parameters on penetration in gas metal arc welding processes. Materials and Design. Elsevier. Volume 28, Issue 2, Pages 649-656.

[31] Kenné G., Ahmed-Ali T., Lamnabhi-Lagarrigue F., and Nkwawo H. (2006). Nonlinear systems parameters estimation using radial basis function network. Control Engineering Practice. No. 14, pp. 819-832.

[32] Kim I.S., Son K.J., Yang Y.S. and Yaragada P.K.D.V. (2003). Sensitivity analysis for process parameters in GMA welding process using a factorial design. International Journal of Machine Tools & Manufacture. Volume 43, Number 8, June 2003, pp. 763-769(7).

[33] Lindgreen L.-E. (2006). Numerical modeling of welding. Computer Methods in Applied Mechanics and Engineering. Volume 195, Issues 48-49, 1, Pages 6710-6736.

[34] Liu Huitao, Han Ping, Wen Yingying, Luan Feng, Gao Yuan and Li Xiuyong. (2010). Quantitative structure-electrochemistry relationship for variously-substituted

9, 10-anthraquinones using both a heuristic method and a radial basis function neural network. Dyes and Pigments. 84, pp. 148-152.

[35] Martín del Brío Bonifacio and Sanz Alfredo. (2007). *Redes Neuronales y Sistemas Difusos.* Alfaomega. España.

[36] Matusko J., Petrovic I., and Peric N. (2008). Neuronal network based tire/road friction force estimation. Engineering Applications of Artificial Intelligence. No. 21. pp. 442-456.

[37] Meqin Liu and Jida Chen. (1998). Determining the structures and parameters of Radial Basis Function Neural Networks using improved Genetic Algorithms. Journal CSUT. Vol. 5 No. 2, pp. 141-146.

[38] Meziane Farid, Vadera Sunil, Kobbacy Khairy and Proudlove Nathan. (2000). Intelligent systems in manufacturing: current developments and future prospects. Journal Integrated Manufacturing Systems. Volume: 11 Issue: 4 Page: 218 - 238.

[39] Montgomery Douglas. C. (1991). *Diseño y Análisis de Experimentos.* Grupo. Editorial Ibero América. pp. 1-172.

[40] _____, Peck E. A. and Vining G. G. (2005). *Introducción al Análisis de Regresión Lineal.* Editorial Continental. Tercera edición. México.

[41] _____. (2006). *Diseños y análisis de experimentos.* Editorial Limusa Willey segunda edición. México.

[42] Navarro-Esbrí J., Berbegall V., Verdu G. Cabello R., and Llopis R. (2007). A low data requirement model of a variable-speed vapour compression refrigeration system based on neuronal networks. International Journal of Refrigeration. No. 30, pp. 1452-1459.

[43] Nelles Oliver. (2001). *Nonlinear System Identification: From classical approaches to neural networks and fuzzy models.* Springer-Verlag Berlin Heidelberg. Germany.

[44] Olabi A.G., Casalino G, Benyounis K. Y., and Hashmi M. S. J. (2006). An ANN and Taguchi algorithms integrated approach to the optimization of CO_2 laser welding. Advances in Engineering Software. Volume 37, Issue 10, Pages 643-648.

[45] Paliwal Mukta and Kumar Usha A. (2009). Neural networks and statistical techniques: A review of applications. Expert Systems with Applications. 36, pp. 2-17.

[46] Pan Lung Kwang, Wang Che Chung, Hsiao Ying Ching and Ho Kye Chyn. (2004). Optimization of Nd: YAG Laser welding onto magnesium alloy via Taguchi analysis. Optics & Laser Technology. Volume 37, Issue 1, Pages 33-42.

[47] Park J. W., Harley R. G., Venayagamoorthy G. K., and Jang G. (2008). Dual heuristic programming based nonlinear optimal control for a synchronous generator. Engineering Applications of Artificial Intelligence. No. 21, pp. 97-105.

[48] Peña Daniel (2002). *Análisis de Datos Multivariantes.* Mc Graw Hill. ISBN: 84-481-3610-1.

[49] Pérez Mesa J. C., Valenciano J. P., and Levy Manguin J. P. (2007). Empleo de Redes Neuronales de Base Radial a un modelo econométrico de exportación de tomate. Ciencia Ergo Sum, marzo-junio, año/vol. 14, número 001 Universidad Autónoma del Estado de México. Toluca, México, pp. 6-14.

[50] Piña M., Rodríguez M., and Días J. (2005). Superioridad de la regresión general ridge sobre mínimos cuadrados. CULCyT// Enero-Febrero, Año 2, No 6, pp 21-26.

[51] Praga-Alejo R., Torres-Treviño L., and Piña-Monarrez M. (2008a). Prediction in Welding Process Using Multiple Linear Regression and Neural Network. International Journal of Industrial Engineering. ISSN 1072-4761 P. 481-488.

[52] _____. (2008b). Comparison between several neural networks models using statistical methods. Proceedings of the 13th Annual International Conference on Industrial

Engineering Theory, Applications and Practice, Las Vegas, Nevada. USA. September 7-10, 2008.

[53] _____, Torres-Treviño L. M., Piña-Monarrez M. R. and González-González D. S. (2010a). A Direct Fuzzy Logic System building by Design of Experiments. International Journal of Industrial Engineering. ISSN 1072-4761 P. 400-408.

[54] _____, Torres-Treviño Luis M., González David S., Acevedo-Dávila Jorge, and Cepeda Francisco. (2010b). A Radial Basis Function Redesigned for Predicting a Welding Process. G. Sidorov et al. (Eds.): MICAI 2010, Part II, LNAI 6438 pp. 257-268, 2010. © Springer-Verlag Berlin Heidelberg 2010.

[55] Rao Subba, Nahm Abraham, Shi Zhengzhong, Deng Xiaodong and Syamil Ahmad. (1999). Artificial intelligence and expert systems applications in new product development - a survey. Journal of Intelligent Manufacturing. 10, 231-244

[56] Richards Edward, Bessant Conrad, and Saini Selwayan. (2004). Simultaneous quantification of analysis in quaternary mixtures using dual pulse staircase voltammetry and genetically optimized neural networks. The Analyst. No. 129 Vol. 4, 355-358. UK.

[57] Romero-Villafranca R., Zúnica L. and Romero-Zúnica R. (2007). Ds-optimal experimental plans for robust parameter design. Journal of Statistical Planning and Inference Volume 137, Issue 4, Pages 1488-1495.

[58] Sampaio Davi, Vasconcelos Cristine, Simoes Sabastiao, and Valtair Antonio. (2005). Comparison between genetics algorithms and response surface methodology in GMAW welding optimization. Journal of Materials Processing Technology. Volume 160, Issue 1, Pages 70-76.

[59] Sanhueza H. Raúl, Harnisch Ildefonso, Díaz R. Horacio, and Guirriman C. Ramón. (1999). Aplicación de algoritmos genéticos al problema de planificación de sistemas

eléctricos de distribución. Revista Facultad de Ingeniería. enero-diciembre, vol. 6. Universidad de Tarapacá. Arica, Chile. pp. 55-63.

[60] Sluzalec Andrzej. (1989). Shape optimization of weld surface. International Journal Solids and Structures. 25 (1), 23-31.

[61] Song Yong-Ak, Park Sehyung and Chae Soo-Won. (2005). 3D welding and milling: part II, optimization of 3D welding process using an experimental design approach. International Journal of Machine Tools and Manufacture
Volume 45, Issue 9, Pages 1063-1069.

[62] Sun Jun, Xu Wenbo and Liu Jing. (2006). Training RBF Neural Network Via Quantum-Behaved Particle Swarm Optimization. I. King et al. (Eds.): ICONIP 2006, Part II, LNCS 4233, Springer-Verlag Berlin Heidelberg, pp. 1156-1163.

[63] Tarng Y.S., Tsai H.L., and Yen S.S. (1999). Modeling, optimization and classification of weld quality in tungsten inert gas welding. International journal of machine tools & manufacture. Volume 39, Issue 9, Pages 1427-1438.

[64] _____, Juang S.C. and Chang C.H. (2002). The use of grey-based Taguchi methods to determine submerged arc welding process parameters in hard facing. Journal of Materials Processing Technology, Volume 128, Number 1, 6 October 2002, pp. 1-6(6).

[65] The MathWorks, Inc. (2004). *Fuzzy Logic Toolbox for Use with MATLAB.* The Math Works Inc.. COPYRIGHT 1995 — 2004, pp 19-47

[66] Torres-Treviño L. M. (2007). Evonorm: Easy and effective implementation of estimation of distribution algorithms. Research in Computing Science. 23, Pages 75-83.

[67] Tsoukalas Lefteri and Uhrig Robert. (1998). *Fuzzy and neural approaches in engineering.* Wiley-Interscience Publication. EUA.

[68] Tzafestas Spyros G. and Blekas Konstantinos D. (1999). Hybrid Soft Computing systems: a critical survey with engineering applications, National technical University of Athens. Athens, Greece.

[69] Valls J. M., Aler R. and Fernández O. (2005). Using a Mahalanobis-Like Distance to train Radial Basis Neural Networks. J. Cabestany, A. Prieto, and D.F. Sandoval (Eds.): IWANN 2005, LNCS 3512, Springer-Verlag Berlin Heidelberg. pp. 257-263.

[70] Von stryk. (1996). Optimization of dynamic de systems in industrial applications. Proc. 2nd European Congress on Intelligent Techniques and Soft Computing.

[71] Wang Heng, Jia Min-ping, Huang Peng and Chen Zuo-liang. (2010). A study on a new algorithm to optimize ball mill system based on modeling and GA. Energy Conversion and Management. 51, pp. 846-850.

[72] *Welding Handbook.* (2001). volume 1, 19th edition, American Welding Society.

[73] Wei li. (2007). Manufacturing process diagnosis using functional regression. Journal of Materials Processing Technology. Volume 186, Issues 1-3, Pages 323-330.

[74] Whitehead Bruce A. and Choate Timothy D. (1995). Cooperative – Competitive Genetic Evolution of Radial Basis Function Centers and Widths for Time Series Prediction. IEEE Transactions on Neural Networks. Vol. 7, issue 4, pp. 869-880.

[75] _____. (1996). Genetic Evolution of Radial Basis Function Coverage using Orthogonal Niches. IEEE Transactions on Neural Networks. Vol. 7, issue 6, pp. 1525-1528.

[76] Yao Y., Lian Z., Hou Z., and Liu W. (2006). An innovative air-conditioning load forecasting model based on RBF neuronal network and combined residual error correction. International Journal of Refrigeration. No. 29, pp. 528-538.

[77] Yen John and Langari Reza. (2000). *Fuzzy Logic: Intelligence, Control and Information.* Prenhall. EUA.

Anexo A

ANEXO A

A continuación se presenta una lista de publicaciones relacionadas con la presente investigación, en estos trabajos aparece el autor de la presente Tesis.

Praga-Alejo Rolando J., Torres-Treviño Luis M., González David S., Acevedo-Dávila Jorge, and Cepeda Francisco. (2010). A Radial Basis Function Redesigned for Predicting a Welding Process. G. Sidorov et al. (Eds.): MICAI 2010, Part II, LNAI 6438 pp. 257–268, 2010. © Springer-Verlag Berlin Heidelberg 2010.

Praga-Alejo Rolando J., Torres-Treviño Luis M., González David S. andPiña-Monarrez Manuel R. (2010). Residual Analysis in a Radial Basis Function Redesigned. Proceedings of the 15th Annual International Conference on Industrial Engineering Theory, Applications and Practice, México, D.F. México. 2010.

Praga-Alejo R., Torres-Treviño L. M., Piña-Monarrez M. R. and González-González D. S. (2010). A Direct Fuzzy Logic System building by Design of Experiments. International Journal of Industrial Engineering. ISSN 1072-4761 Pp. 461-471.

González-González D. S., Piña-Monarrez M. R., Cantú-Sifuentes Mario and Praga-Alejo R. (2010). Collinearity Reduction in the Cox's Proportional Hazard Model by Applying PCA. International Journal of Industrial Engineering. ISSN 1072-4761 Pp. 417-425.

Praga-Alejo R., Torres-Treviño L. M., Piña-Monarrez M. R. and González-González D.

S. (2009). Development of a Fuzzy Logic System through a Design Of Experiment for a Welding Process. Proceedings of the 14th Annual International Conference on Industrial Engineering Theory, Applications and Practice, Anaheim. USA. 2009.

Praga-Alejo R., Torres-Treviño L., and Piña-Monarrez M. (2008). Prediction in Welding Process Using Multiple Linear Regression and Neural Network. International Journal of Industrial Engineering. ISSN 1072-4761 P. 481-488.

Praga-Alejo R., Torres-Treviño L., and Piña-Monarrez M. (2008). An Optimization using Multiple Linear Regression and Genetic Algorithm in Welding Process. Proceedings of the 13th Annual International Conference on Industrial Engineering Theory, Applications and Practice, Las Vegas, Nevada. USA. September 7-10, 2008.

Praga-Alejo R., Torres-Treviño L., and Piña-Monarrez M. (2008). Comparison between several neural networks models using statistical methods. Proceedings of the 13th Annual International Conference on Industrial Engineering Theory, Applications and Practice, Las Vegas, Nevada. USA. September 7-10, 2008.

Praga-Alejo R., and Torres-Treviño L. (2008). Multi-objective Optimization in a welding process of the automotive industry. IV Congreso Mexicano de Computación Evolutiva COMCEV 2008. Pp. 99-102.

Praga-Alejo R., Torres-Treviño L. M., Piña-Monarrez M. R. and González-González D. S. (2008). Optimization welding process parameters through Response Surface and Neural Network of Maximum Sensibility plus Genetic Algorithm. Procedings of the Second International Conference on Industrial, Mechatronics and Manufacturing Engineering. Pp.177-186. October 23-25, 2008. Ciudad Juárez, Chihuahua. México.

García-Sánchez A., Trejo-Aguirre Mario F., Praga-Alejo R. J., Torres-Treviño L. M., Acevedo-Dávila J.L., García G., and M. Garza-Castañón. (2008). Optimización de los parámetros de soldadura por proceso de arco eléctrico con protección de gas (GMAW) para soldar un acero A-487. 14 Convención Científica de Ingeniería y Arquitectura, Ministerio de Educación Superior, La Habana, Cuba. V Congreso de Ing-

eniería Mecánica y Metalúrgica Memorias, Pág. 867-872 ISBN 978-959-261-281-5.

Praga-Alejo R., Torres-Treviño L. M., and Piña-Monarrez M. R. (2008). Optimal Determination of K Constant of Ridge Regression Using a Simple Genetic Algorithm. Electronics, Robotics, and Automotive Mechanics Conference CERMA 2008. Pp. 39-44. Cuernavaca, Morelos. México September 30 -03 October.

Praga-Alejo R., Torres-Treviño L. M., and Piña-Monarrez M. R.. (2008). Optimization welding process parameters through Response Surface, Neural Network and Genetic Algorithms. Electronics, Robotics, and Automotive Mechanics Conference CERMA 2008. Pp. 393-399. Cuernavaca, Morelos. México September 30 -03 October.

Torres-Treviño L. M., Piña-Monarrez M. R., Praga-Alejo R. J., Reyes-Valdes A., and Acevedo-Dávila J. (2007). Comparison between Multiple Linear Regressions and Neural Networks in Optimization welding process. Proceedings of the 12th Annual International Conference on Industrial Engineering Theory, Applications and Practice. Cancun, Mexico. November 4-7, 2007.