Ministerio de Educación Superior

Universidad Central "Marta Abreu" de Las Villas

TRABAJO DE DIPLOMA

Título:

EXTRACCIÓN DE MODELOS INTELIGENTES A PARTIR DE DATOS CON APLICACIÓN EN SISTEMAS AGRÍCOLAS.

Autor:

Carlos Rafael Herrera Contreras

Tutores:

Dr. C. María M. García Lorenzo

Dr. C. Francisco Herrera Fernández

2009

"Año del 50 Aniversario del Triunfo de la Revolución"

PENSAMIENTO

La realidad es simplemente una ilusión, aunque muy persistente.

Albert Einstein

DEDICATORIA

A mis padres y hermano por su esfuerzo y dedicación en todos estos años.

AGRADECIMIENTOS

Les agradezco a todas las personas que colaboraron de forma científica, material y emocional a realizar este trabajo.

A mi familia que siempre estuvo presente.

A mi padre y tutor Dr. Francisco B. Herrera Fernández por su asesoría en este trabajo y en toda mi vida.

A mis profesores que de una forma u otra ayudaron en mi formación.

A mi tutora Dr. Maria M. Garcia Lorenzo por la asesoría prestada.

RESUMEN

La presente investigación tiene como propósito mostrar el procedimiento utilizado para la predicción y pronóstico de variables agro meteorológicas asociadas a sistemas productivos agrícolas, **tales como el rendimiento en cultivos y sus componentes**. Se emplea los sistemas difusos como la base para determinar los correspondientes modelos predictivos. Estos modelos fueron construidos usando la técnica conocida como agrupamiento difuso. La temperatura y humedad relativa fueron las dos variables de un sistema agro meteorológico usadas en esta investigación para el ensayo y ajuste del algoritmo y programa para el pronóstico de variables. Ambas variables formarán parte de un sistema multiagente. Finalmente se concibe la implementación de sistema multiagente **para el pronóstico del rendimiento.**

ABSTRACT

This research aims to show the procedure used for the prediction and prognosis of agro meteorological variables associated with agricultural production systems, **such as crop yield and its components**. Fuzzy systems is used as the basis for determining appropriate predictive models. These models were constructed using the technique known as fuzzy clustering. The temperature and relative humidity were the variables of an agrometeorological system used in this research for testing and tuning of the algorithm and program to variables forecasting. Both variables are part of a multiagent system. Finally sees the implementation of multiagent system **for yield prediction**

ÍNDICE

INTRODUCCIÓN 1

CAPÍTULO 1 APRENDIZAJE AUTOMATIZADO. APLICACIÓN DE SISTEMAS COMPUTACIONALES EN LA PREDICCION DE PARAMETROS AGROINDUSTRIALES. 6

1.1 Aprendizaje Automatizado. 1

1.2 Series Temporales. 1

1.2.1 Análisis de series temporales. 2

1.2.2 Aplicación de técnicas de Inteligencia Artificial a la predicción de Series Temporales. 4

1.3 Fundamentos de Lógica Difusa. 5

1.3.1 Conjuntos Difusos. 6

1.3.2 Funciones de Pertenencia 6

1.3.3 Tipos de Funciones de Pertenencia 6

1.3.4 Operadores Lógicos Difusos 7

1.3.5 Reglas Difusas 7

1.4 Clasificación difusa. 8

1.5 Sistemas de Inferencia Difusos. 8

1.5.1 Funcionamiento de un Sistema de Inferencia Difuso 9

1.6 Técnica adaptativa Neuro-Difusa ANFIS 10

1.6.1 Funcionamiento del modelo ANFIS 11

1.7 Aplicación de sistemas computacionales en el sector agroindustrial. 11

1.8 Problemas de la agro meteorología. 16

1.9 Conclusiones parciales. 17

CAPITULO 2 ANÁLISIS Y DISEÑO DE LA HERRAMIENTA. SISTEMA DE PREDICCIÓN AGRO METEOROLÓGICO. 19

2.1 Datos de Campo. 19

2.2 Selección de entradas..... 20

2.3 Construcción de sistema Neuro-Fuzzy. 23

2.4 La técnica de agrupamiento. 24

2.4.1 Método Fuzzy C-Means 24

2.4.2 Subtractive Clustering. 26

2.4.3 Aplicando técnicas de agrupamiento al problema. 27

2.5 Construcción del Sistema de Inferencia Borroso. 28

2.6 Aplicando Técnica Adaptativa Neuro-Difusa ANFIS. 32

2.7 Conclusiones parciales. 34

CAPÍTULO 3 IMPLEMENTACIÓN DE AGENTES PARA LA DETERMINACIÓN DE UN SISTEMA MULTIAGENTE DE PRONÓSTICO DE RENDIMIENTO..... 35

3.1 Visión general del sistema. 35

3.2 Descripción del sistema. 35

3.3 Procedimiento para la construcción de agentes. 36

3.4 Conclusiones parciales. 37

REFERENCIAS BIBLIOGRÁFICAS 40

Anexos..... 45

Anexo 1. Transformación de las mediciones. 45

Anexo 2. Selección de Muestras de Entrenamiento y Prueba 46

Anexo 3. Resultado del FIS obtenido por diferentes técnicas de agrupamiento. 48

Anexo 4. Programa de validación usando la muestra de control..... 50

INTRODUCCIÓN

En los últimos años, se ha manifestado un acelerado crecimiento en las capacidades de generar y coleccionar datos, debido básicamente al poder de procesamiento de las máquinas así como su bajo costo de almacenamiento.

Sin embargo, dentro de estas enormes masas de datos existe una gran cantidad de información "oculta", de gran importancia estratégica, a la que no se puede acceder por las técnicas clásicas de recuperación de la información. El descubrimiento de esta información "oculta" es posible gracias a la aplicación de modernas herramientas computacionales, como la conocida por Minería de Datos (*Data Mining*), vinculándose actualmente a sofisticadas técnicas de la inteligencia artificial para encontrar patrones y relaciones dentro de los datos. Es imprescindible convertir estos volúmenes de datos existentes en experiencia, conocimiento y sabiduría, formas que atesora la humanidad para que sea útil en la toma de decisiones, especialmente en las grandes organizaciones y proyectos científicos. El procesamiento y análisis de grandes volúmenes de datos por estas técnicas es importante en ramas como: bioinformática, medicina, economía y finanzas, industria, medio ambiente, entre otras.

En la agro meteorología el conocimiento del medio es fundamental para el desarrollo del sector agrario, siendo importante conocer los efectos que ejerce el tiempo y clima en los sistemas de producción agropecuaria (suelo-planta, ganado, etc.), lo que permite desarrollar planes y estrategias a corto, medio y largo plazo. Partiendo de los datos presentes y pasados en cierta región, se puede programar las actividades futuras con mayor efectividad.

La predicción del tiempo, realizada a escala regional, es base fundamental para la planificación de las labores agrícolas. Esto implica la disponibilidad de predicciones locales, utilizando técnicas de predicción numérica, junto con un profundo conocimiento del medio. No obstante, la información distribuida habitualmente por los medios de comunicación tiene un carácter general y resulta a todas luces insuficiente cuando se trata de gestionar ciertas actividades, como el riego o el seguimiento de plagas o enfermedades.

El anterior análisis apunta la importancia de una adecuada combinación de modelos y datos, para alcanzar una buena gestión en estos sistemas productivos, de ahí la necesidad del permanente trabajo en el desarrollo y aplicación de medios computacionales y particularmente de la inteligencia computacional.

Antecedentes

El Descubrimiento de Conocimiento en Datos (conocido general e internacionalmente por las siglas KDD, del inglés, *Knowledge Discovery in Data*,) se define como "el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y comprensibles a partir de los datos". El KDD se ha desarrollado sobre la base de la intersección de diversas áreas de investigación tales como aprendizaje de máquinas, reconocimiento de patrones, bases de datos, estadística, inteligencia artificial, adquisición de conocimiento, visualización, con el objetivo de procesar automáticamente grandes cantidades de datos, identificando los modelos más relevantes y presentarlos como conocimiento para satisfacer las necesidades de los usuarios finales.

Este proceso comprende diversas etapas, que van desde la obtención de los datos hasta la aplicación del conocimiento adquirido en la toma de decisiones. Entre esas etapas, se encuentra la que puede considerarse como el núcleo del proceso KDD y que consiste en la extracción del conocimiento a partir de los datos. Esta fase es crucial para la obtención de resultados apropiados, y depende del algoritmo de aprendizaje automatizado que se aplique, no obstante, se ve influida en gran medida por la calidad de los datos que llegan desde la fase previa (Ruiz 2006).

En la figura 1.1 se muestra el esquema físico de un proceso de KDD, apreciándose los principales subprocesos que toman parte en el mismo. Aunque es válido aclarar que en ocasiones es necesario alterar este orden, se considera que la etapa de preparación de los datos contribuirá al buen desempeño del proceso en su conjunto.

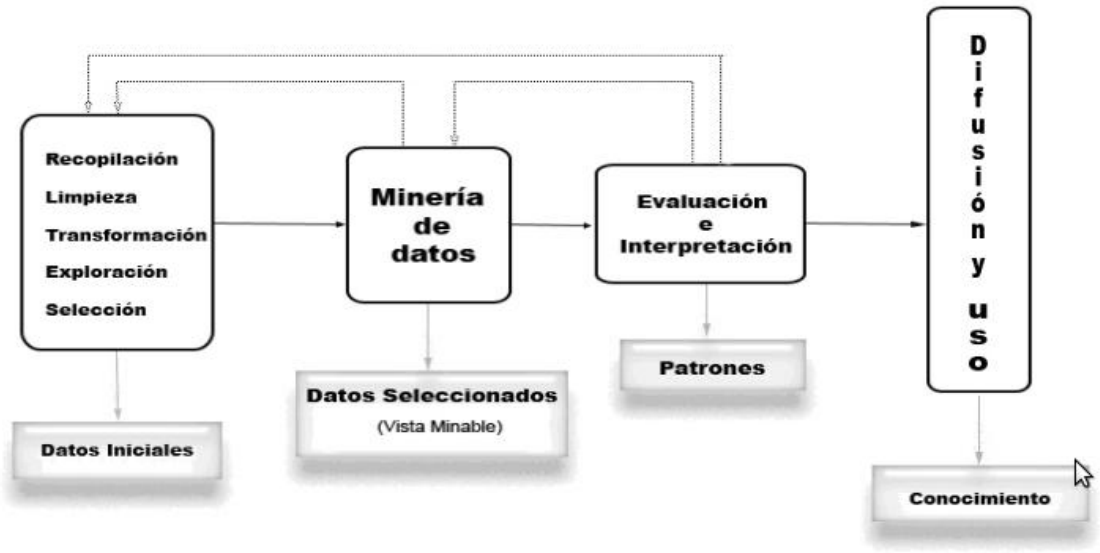

Figura 1**.** Proceso de KDD

Los métodos de aprendizaje constituyen una componente básica en los sistemas inteligentes de procesamiento de datos. Entre estos métodos se reconocen los que se basan en el empleo de Redes Neuronales Artificiales (RNA) (Rosemblatt 1962); el método de los k-Vecinos más Cercanos (K-NN) (Cover and Hart 1967), y el algoritmo C4.5 (Quinlan 1993), entre otros. El aprendizaje supervisado en general y la clasificación en particular, puede resultar en la práctica, tediosa y en ocasiones inaplicable si se dispone de conjuntos de entrenamiento numerosos y con datos de alta dimensionalidad. Estos dos factores determinan el costo computacional de la aplicación de la mayoría de los métodos de clasificación supervisada existentes (Caballero 2005).

Otra estrategia consiste en la aplicación de sistemas *Neuro-Fuzzy*, los que representan un nuevo desarrollo de sistemas inteligentes híbridos, donde se combina las principales características de redes neuronales artificiales con las de los sistemas basados en la lógica difusa. El principal objetivo es eludir las dificultades encontrados en la aplicación de la lógica difusa en la representación del conocimiento por sistemas numéricos (conjuntos de datos), o viceversa, en la aplicación de redes neuronales para representar conocimiento mediante sistemas de información lingüística.

Aunque la lógica difusa y redes neuronales tienen su origen en paradigmas totalmente diferente del formalismo matemático, se ha podido mostrar que ambas metodologías son aproximadores universales para una gran clase de asignaciones no lineal. También

ha sido demostrado que se pueden combinar para formar híbridos de potente estructura para hacer frente a una amplia gama de aplicaciones para la representación y procesamiento de diferentes tipos de conocimientos, expresados tanto en forma numérica como lingüística.

Planteamiento del Problema

El desarrollo de modernas tecnologías se ha reflejado en la agricultura en diferentes formas, aunque por lo general su aplicación implica un costo relativamente alto en el desarrollo de los sistemas, como es el caso de las técnicas comprendidas en la Agricultura de Precisión. No obstante en la implementación de sistemas inteligentes en procesos agrícolas el costo económico que implica, no impide su aplicación; y si, por otro lado, impacta de manera directa en beneficios económicos, además de crear condiciones iníciales para el uso de otras herramientas no solo de Inteligencia Artificial, sino en lo que en general se refiere a los Sistemas Computacionales.

Hoy día hay mucha necesidad de desarrollar sistemas basados en estas herramientas inteligentes, ya que se adolece de sistemas que puedan ser utilizados como predictores o en diagnóstico de diferentes clases de problemas.

Es obligado entonces, pensar en cómo se pueden implementar técnicas factibles en lo económico y en lo técnico-cultural (amen de el aspecto de investigación científica y de generación y transferencia de tecnología), que garanticen la estabilidad ambiental del entorno; y que como objetivo principal identifique un modelo y patrón de comportamiento, a fin de tener información confiable en la que se puedan apoyar decisiones que a su vez, deriven en beneficios económicos haciendo más rentable esta actividad.

Problema de investigación

La necesidad de introducción de modernas técnicas de apoyo a las decisiones en procesos agro meteorológicos, siendo el problema central la determinación de relaciones, parámetros y variables que condicionan una decisión

El **objetivo general** de esta investigación es:

Diseñar sistemas *Neuro-Fuzzy*, para la detección de relaciones y modelos a partir de grandes volúmenes de datos en pronóstico de rendimiento de sistemas agrícolas.

Para lograr este objetivo se plantean los siguientes **objetivos específicos** de la investigación:

1. Desarrollar un método general que permita determinar relaciones entre las variables y parámetros en un sistema multivariable usando sistemas neurodifusos de apoyo a la toma de decisiones en sistemas productivos y en condiciones meteorológicas determinadas.
2. Obtener un modelo de relaciones para las principales variables agrometeorológicas en una región específica haciendo uso de la *Toolbox Fuzzy* de Matlab).

La **hipótesis de la investigación:**
Resulta factible diseñar sistemas neurodifusos para la detección de relaciones de un gran conjunto de datos agro meteorológicos, siendo los resultados adecuados a los que se alcanzan con técnicas similares.

La **tesis** está **estructurada** en tres capítulos:
El Capitulo 1 trata el problema del aprendizaje, se hace énfasis en el análisis de series temporales y de algunos métodos existentes. Se describen aspectos fundamentales de las técnicas Adaptativa Neuro-Difusa ANFIS. Además se dan elementos fundamentales sobre el problema real de sistemas inteligentes en el sector agroindustrial. Planteando la necesidad del preprocesamiento de los datos agro meteorológicos.

En el Capitulo 2 se proponen dos técnicas de agrupamiento para la construcción de sistemas difusos, uno de ellos trabaja con radio de similitud, y el otro con las medias. Se establece correlaciones para la obtención del modelo entre los rasgos significativos. Se muestran los resultados alcanzados por cada uno de los agentes implementados.

En el Capítulo 3 Se describe una visión general de sistema multiagente y el campo de aplicación del mismo.

CAPÍTULO 1 APRENDIZAJE AUTOMATIZADO. APLICACIÓN DE SISTEMAS COMPUTACIONALES EN LA PREDICCIÓN DE PARAMETROS AGROINDUSTRIALES.

Una de las críticas que se oyen más a menudo respecto a la IA, es que las máquinas no se pueden considerar inteligentes hasta que no sean capaces de aprender a hacer cosas nuevas y a adaptarse a las nuevas situaciones, en lugar de limitarse a hacer aquellas actividades para las que fueron programadas. En vez de preguntarse si una computadora es capaz de "aprender", resulta mucho más clarificador intentar describir a qué actividades se refiere exactamente cuando se dice "aprender " y cuáles mecanismos se pueden utilizar para llevar a cabo dichas actividades. Simon en 1983 (Simon 1983) se refiere al aprendizaje como cambios en el sistema para desarrollar tareas a partir de las mismas condiciones, de un modo más eficiente y eficaz cada vez.

La mayoría de los programas de IA se basan principalmente en el conocimiento como fuente de recursos. El conocimiento se adquiere generalmente a través de la experiencia. La adquisición del conocimiento en sí misma incluye muchas actividades diferentes. El simple almacenamiento de información computada, también conocido como aprendizaje memorístico, constituye la actividad de aprendizaje más básica (Kelly Christian and Robert 2001).

El aprendizaje a partir de ejemplos puede requerir o no un experto que ayude a clasificar corrigiendo cuando surge una equivocación. Un caso típico es el problema de clasificación supervisada, donde se aprende a clasificar sin haber recibido previamente reglas explícitas (Knight 1994) .

La clasificación supervisada consiste en el proceso de asignar a una entrada concreta, el nombre de una clase a la que pertenece. Las clases entre las que puede elegir el procedimiento de clasificación supervisada se pueden describir en variadas formas.

La clasificación supervisada constituye una parte importante de muchas de las tareas de resolución de problemas y en su forma más simple, se presenta como una tarea de reconocimiento. Antes de que se pueda hacer este tipo de clasificación, se deben definir las clases que utilizará.

1.1 Aprendizaje Automatizado.

Desde hace más de dos décadas se vienen desarrollando y utilizando complejos algoritmos para la extracción de patrones útiles en grandes conjuntos de datos. Gran parte de esta información representa transacciones o situaciones que se han producido, siendo útil no sólo para explicar el pasado, sino para entender el presente y realizar predicciones en el comportamiento futuro del sistema fuente de la información.

En muchas ocasiones, el método tradicional de convertir los datos en conocimiento consiste en un análisis e interpretación realizada de forma manual por especialistas en la materia estudiada. Esta forma de actuar es lenta, cara y altamente subjetiva. De hecho, la enorme cantidad de datos desborda la capacidad humana de procesarlos y comprenderlos. El análisis manual hace que las decisiones se tomen según la intuición de los especialistas.

A finales de la década de los 80, la creciente necesidad de automatizar todo este proceso inductivo abre una línea de investigación para el análisis inteligente de datos, impulsando las investigaciones en aprendizaje automatizado. ***El Aprendizaje Automatizado es el área de la Inteligencia Artificial que se ocupa de desarrollar técnicas capaces de aprender, es decir, extraer de forma automatizada conocimiento subyacente en la información***. Constituye, junto con la estadística, el corazón del análisis inteligente de los datos (Ruiz 2006).

El principio seguido en el aprendizaje automatizado es que la máquina genera un modelo a partir de ejemplos y lo usa para resolver el problema. Uno de los modelos de aprendizaje más estudiado es el aprendizaje inductivo, que engloba todas aquellas técnicas que aplican inferencias inductivas sobre un conjunto de datos para adquirir conocimiento inherente a ellos.

1.2 Series Temporales.

Resulta difícil imaginar una rama de la ciencia en la que no aparezcan datos que puedan ser considerados como series temporales.

En este epígrafe se describirán los conceptos y análisis fundamentales de las series temporales y su utilidad en la predicción. Dado que el estudio del clima está asociado a

las épocas del año donde se tiene en cuenta el tiempo y un orden cronológico en la obtención y posterior procesamiento de los datos.
Una serie temporal es una colección de datos observados de una variable realizada de forma secuencial en el tiempo, en las que el orden de observación es importante. Los valores de una serie temporal van ligados a instantes de tiempo, de manera que el análisis de una serie implica el manejo conjunto de dos variables; la variable en estudio propiamente dicha y la variable tiempo.

Las series pueden tener una periodicidad anual, semestral, trimestral, mensual, etc., según los periodos de tiempo en los que están recogidos los datos que la componen. Las ventas trimestrales de una empresa, el número de personas de personas afectadas por el SIDA en cada mes, la cantidad de accidentes semanales de tráfico o el número de exportaciones efectuadas cada año por un determinado país, son ejemplos de series temporales con diferentes periodicidades (Catalán 2004).

1.2.1 Análisis de series temporales.

Los objetivos del análisis de series temporales son diversos, pudiéndose destacar entre ellos, la predicción de variables de estado o principales, identificación y control de procesos, incluyendo la simulación, la generación de nuevas teorías físicas o biológicas, etc.

Denominamos **predicción** a la estimación de valores futuros de la variable en función del comportamiento pasado de la serie. Este objetivo se emplea ampliamente prácticamente en todos los campos de la ciencia y la técnica, incluyendo aspectos tan importantes como la sanidad pública y la vigilancia de la salud. Así por ejemplo, la predicción mediante modelos basados en la teoría de series temporales, puede servir para una buena planificación de recursos sanitarios, en función de la demanda que se espera en el futuro. Otro de los campos en los que se aplica la predicción mediante series temporales es en la meteorología o en la predicción de otros fenómenos naturales.

En el **control de procesos**, se trata de seguir la evolución de una variable determinada con el fin de regular o controlar su comportamiento. Otra importante aplicación es en el

campo de la medicina, siendo un buen ejemplo las aplicaciones en los Centros de Control de Enfermedades.

Por otra parte la **simulación**, en tanto una herramienta de análisis de alcance universal, tiene un particular uso en el estudio de procesos temporales mediante el uso de los correspondientes modelos, generalmente en investigaciones aplicadas, cuando el proceso es muy complejo para ser estudiado de forma analítica.

Evidentemente aunque el valor futuro de una serie temporal no sea predecible con total exactitud, para que tenga interés su estudio, el resultado tampoco puede ser completamente aleatorio, existiendo alguna regularidad en cuanto a su comportamiento en el tiempo, lo que hará posible su modelado y por ende, en su caso, la predicción. La búsqueda de regularidades y de patrones ha sido siempre una de las tareas básicas de la ciencia, y muchas veces se descubren simetrías que sirven de fundamento para la predicción del comportamiento de los fenómenos, incluso antes de que se entienda la razón o causa que justifica esa regularidad.

Es posible encontrar patrones de regularidad en diferentes secciones de una serie temporal, podremos también describirlas mediante modelos basados en distribuciones de probabilidad.

Un concepto importante que encontramos en este ámbito, es el de procesos estacionarios. De una manera informal, diremos que una serie es **estacionaria** cuando se encuentra en equilibrio estadístico, en el sentido de que sus propiedades no varían a lo largo del tiempo, y por lo tanto no pueden existir tendencias. Un proceso es no−estacionario si sus propiedades varían con el tiempo, como el clima. Si examinamos por ejemplo la temperatura para un determinado mes a lo largo de los años en una determinada zona geográfica, y se está produciendo un cambio climático, aunque haya fluctuaciones, habrá una tendencia creciente.

Existen tres enfoques diferentes, aunque relacionados, para el análisis de series temporales: Modelado clásico, Modelos ARIMA, Análisis del espectro de frecuencias (Molinero 2004).

1.2.2 Aplicación de técnicas de Inteligencia Artificial a la predicción de Series Temporales.

Las principales herramientas de la Inteligencia Artificial también tienen su espacio de aplicación en la predicción de series temporales. Particularmente las redes neuronales, los sistemas difusos (también conocido por el término *fuzzy*, proveniente del inglés) e híbridos de ambas herramientas han sido ampliamente investigadas. El objetivo en este epígrafe es resumir los principales conceptos relacionados para su aplicación en modelos predictivos que son aplicados posteriormente en este trabajo.

Redes Neuronales Artificiales

En la actualidad se han desarrollado múltiples aplicaciones en diversas ramas donde se han utilizado las Redes Neuronales Artificiales para su solución, tales como: predicción de heladas en la estación de Río Cuarto, Córdoba, Argentina, en el año 2005 (Ovando 2005); construcción de un sistema automatizado para el riesgo de mortalidad hospitalaria en una unidad de cuidados intensivos, en el año 2005 (Cabello 2005); predicción de lluvia-escorrentía en tiempo real en el entorno del Sistema Automático de Información Hidrológica de la Cuenca Sur de España (SAIH-SUR), en el año 2006 (Rueda 2006); construcción de un sistema automatizado de valoración de viviendas en la ciudad de Albacete, España, en el año 2006 (García 2006); entre otras (Borracci and Arribalzaga 2005; Milone 2005; Zúñiga and Jordán 2005; Calderón and Jara 2006; González and Cases 2006; Romero, Machado et al. 2006).

La arquitectura de redes neuronales más ampliamente utilizada es la que se conoce con el nombre de Perceptron Multicapas (MLP) (Rosemblatt 1962); (Witten 2005), la cual se caracteriza por el hecho de que sus neuronas se agrupan en capas por niveles. Cada una de estas capas está constituida por un conjunto de neuronas. Hay tres tipos de capas diferentes: la capa de entrada, las capas ocultas y la capa de salida, como se observa en la figura 3:

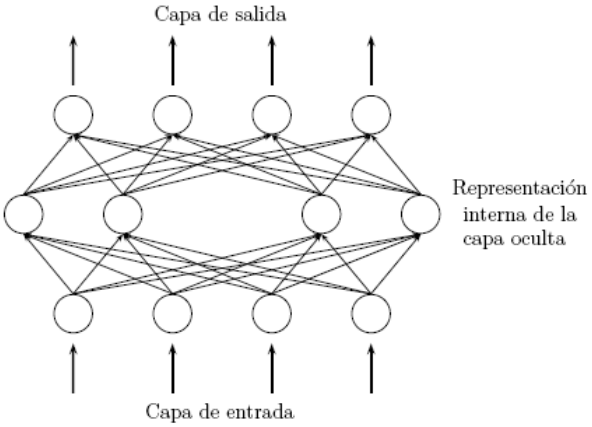

Figura 1.1. Arquitectura redes neuronales MLP.

La principal característica que se le asocia es que se considera como caja negra y no facilita la interpretación de los resultados (Borgelt 2003).

Método de intercambio de modelos autorregresivos en base a cadenas de Markov (MS-AR)

Un modelo autorregresivos (AR) es también conocido como *Infinite Impulse Response Filter* (IIRF). Existe una memoria o realimentación y por lo tanto el sistema puede generar dinámicas internas. El modelo autorregresivo ensayado se expresa mediante la siguiente ecuación MS(K)-AR(p):

$$x_t = \alpha^{S_t} + \sum_{i=1}^{p} a_i^{S_t} x_{t-i} + \sigma^{S_t} \varepsilon_t$$

donde St representa el estado en el periodo t, St=1...K, donde K es el número de estados representados mediante la cadenas de Markov, p es el retardo máximo en la componente autorregresiva, α St es la constante en el estado St, σ St es desviación estándar del modelo en el estado St, aiSt es el parámetro autorregresivo para el lag i en el estado St, ε t es el residuo el cual sigue una distribución particular, en este caso distribución normal N(0,1) (Hamilton 2005).

1.3 Fundamentos de Lógica Difusa.

En la siguiente sección se hace una breve descripción de lo que son los conjuntos difusos y sus propiedades, las funciones de pertenencia y sus tipos.

1.3.1 Conjuntos Difusos.

Un conjunto difuso se define por una función de pertenencia que asocia a cada objeto del universo X un valor en el intervalo [0,1]. Si x es un objeto en el universo X y y=C(x) es el valor asociado a x, se dice que y es el grado de pertenencia del objeto x al conjunto difuso C.

Un conjunto difuso es una clase de objetos con grados de pertenencias continuos. Por lo tanto es un conjunto caracterizado por su función de pertenencia la cual asigna a cada objeto un grado de pertenencia en el rango cero y uno (Zadeh, 1997).

Un término difuso puede ser interpretado como un subconjunto difuso sobre la recta real siempre que éste sea un subconjunto normalizado; es decir, que por lo menos uno de sus valores tenga pertenencia total al subconjunto.

1.3.2 Funciones de Pertenencia

Es una función μA(x), que especifica el grado con el cual una entrada cualquiera pertenece a un conjunto difuso A ó es relativo a ese concepto. Podemos decir que una función de pertenencia es una curva que define como a cada punto en el espacio de entrada se le asigna un grado de pertenencia (valor entre cero y uno).

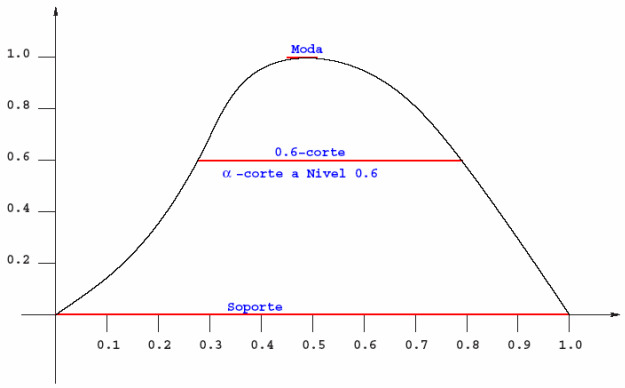

Gráfico 1.1. Representación de las Propiedades de un Conjunto Difuso

1.3.3 Tipos de Funciones de Pertenencia

Generalmente se utilizan ciertas familias de formas estándar, las más frecuentes son: Triangular, Trapezoidal, Gaussiana, Sigmoidal, tipo π y tipo S.

Triangular: Definido por sus límites inferior a y superior b, y el valor modal m, tal que a < m < b.

La cual puede representarse como:

A (x; a, m, b) = max{min {(x-a)/ (m-a), (b-x)/ (b-m)}, 0}

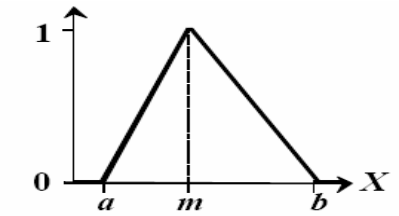

Gráfico 1.2. Función Triangular.

En el caso de la función Gauss la función de pertenencia $f(x;\sigma,c) = e^{\frac{-(x-c)^2}{2\sigma^2}}$ depende de dos parámetros; su valor medio c y $\sigma > 0$.

1.3.4 Operadores Lógicos Difusos

En lógica difusa los operadores son identificados como difusos y su correspondencia con los operadores clásicos de la lógica booleana (unión, intersección y negación), son la función mínimo para intersección, función máximo para unión y función complemento para negación (Gráfico 3).

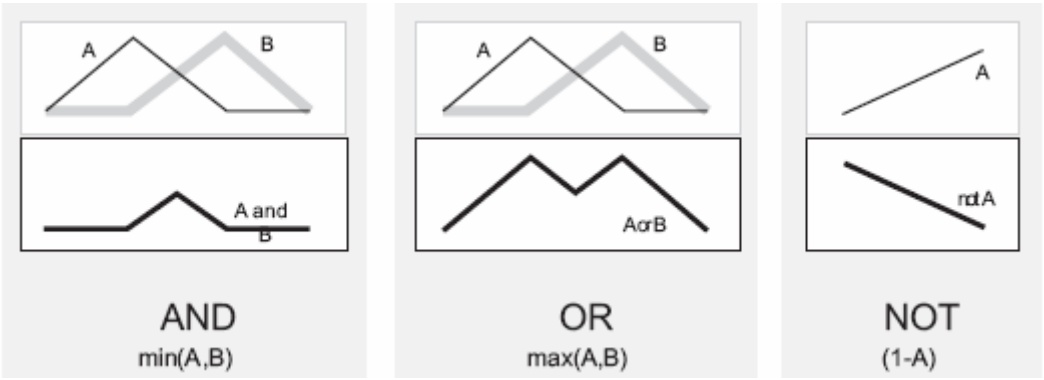

Gráfico 1.3. Representación de los Operadores Lógicos Difusos

1.3.5 Reglas Difusas

Los conjuntos difusos y los operadores difusos forman el sujeto y verbo de la lógica difusa y la forma de comprimirlos, es a través de las reglas difusas. Las reglas difusas son la representación en lenguaje natural del conocimiento (empírico ó no) que se tiene acerca de un problema del mundo real.

Las reglas difusas combinan uno o más conjuntos difusos de entrada, llamados antecedentes o premisas, y las asocian a un conjunto difuso de salida llamado consecuente. Estas reglas son las que permiten expresar el conocimiento que se dispone sobre la relación entre antecedentes y consecuentes [Martín del Brío y Sanz, 2002].
Una regla difusa ampliamente usada, es del tipo "Si-Entonces" por ejemplo:
Si x es A and y es B Entonces C es z
Donde los antecedentes son: x es A and y es B
La consecuencia: C es z (*Y el operador lógico de intersección*).

1.4 Clasificación difusa.

La clasificación difusa tiene sus orígenes en Dunn (Dunn 1973), y posteriormente generalizado por Bezdek (Bezdek 1981), con el algoritmo "Fuzzy C-means". Se define como una técnica diseñada para alcanzar una determinada representación de un espacio vectorial de vectores de entrada, basado en la medición de las distancias euclidianas entre vectores. Esta técnica ha sido útil para determinar las reglas difusas que describen un sistema desconocido o caja negra (Brío and Sanz 2002).

Los algoritmos basados en clasificación difusa son ampliamente utilizados, no solamente para la construcción de los modelos difusos, sino también en aplicaciones de compresión, organización y clasificación de información. Estas técnicas pueden utilizarse también para la obtención de los parámetros iníciales de las reglas en sistemas de inferencia difusos (Bruno 1999).

1.5 Sistemas de Inferencia Difusos.

La inferencia difusa es el proceso de formulación del recorrido que se realiza desde una entrada dada hasta que se genera la salida usando para ello, lógica difusa. Es un método que interpreta los valores en el vector de entrada y basado en un conjunto de reglas, asigna valores al vector de salida. Esta formulación involucra: la selección de los funciones de pertenencia, la selección de los operadores lógicos difusos, el diseño de las reglas difusas, la elección del mecanismo de implicación y agregación de las reglas difusas (mecanismo de inferencia), y finalmente, la selección del método de defuzificación para la obtención de la salida del sistema.

Por lo general, existen dos tipos de sistemas de inferencia difusos (FIS) estos son: FIS tipo Mamdani y FIS tipo Sugeno. Mamdani es un tipo de inferencia en el cual los conjuntos difusos de la consecuencia de cada regla son combinados a través del operador de agregación y el conjunto resultante es defuzificado para obtener la salida del sistema. Sugeno, por su lado, es un tipo de inferencia en el cual la consecuencia de cada regla es una combinación lineal de las variables de entrada y la salida es una combinación lineal ponderada de las consecuencias.

El método Sugeno fue introducido en 1985 por Takagi-Sugeno-Kang, es similar al método Mamdani en muchos aspectos. De hecho, las dos primeras partes del sistema de inferencia, fuzificación de la entrada y aplicación de los operadores difusos son **exactamente** iguales, pero, en el último paso se utiliza como método de defuzificación el promedio de las ponderaciones "weighted average" es decir, un total de "N" reglas se calcula como:

$$l = \frac{\sum_{i-1}^{N} W_i Z_i}{\sum_{i-1}^{N} W_i}$$

Z_i es el nivel de salida de cada regla.

W_iPeso ponderado de cada regla.

1.5.1 Funcionamiento de un Sistema de Inferencia Difuso

En los FIS se realizan cinco pasos hasta la obtención de la salida.

Paso 1. Fuzificación de las variables de entrada. Consiste en tomar las entradas y determinar el grado con el cual ellas pertenecen a cada conjunto difuso a través de las funciones de pertenencia. La entrada al proceso de fuzificación siempre es un valor numérico del mundo real o valor natural limitado al universo de discurso de la variable de entrada. La salida al proceso es un grado de pertenencia el intervalo [0,1].

Paso 2. Aplicación de los operadores lógicos difusos. Una vez que las entradas han sido transformadas en difusas ya es conocido el grado con el cual cada parte de los antecedentes se satisface para cada regla. Si el antecedente de una regla dada tiene más de una parte, entonces el operador difuso es aplicado para obtener un solo número que represente el resultado del antecedente.

Paso 3. Implicación de los antecedentes a la consecuencia. Esta se realiza para modificar el conjunto difuso de la salida en el grado especificado por el antecedente. La entrada al proceso de implicación es el número dado por el antecedente y la salida es un conjunto difuso truncado.

Paso 4. Agregación de las consecuencias de las reglas. La agregación es el proceso en el cual los conjuntos difusos que representan a la salida de cada regla son combinados en una sola área o conjunto. La entrada al proceso de agregación es una lista de conjuntos difusos truncados resultantes del proceso de implicación y la salida es un conjunto difuso por cada variable de salida. Debido a que este proceso es conmutativo el orden en el cual son ejecutadas las reglas no es importante.

Paso 5. Defuzificación. La entrada al proceso de defuzificación es un conjunto difuso (resultante del proceso de agregación) y la salida es un número natural. El método de defuzificación más popular, para el caso Mamdani, es el centroide, el cual retorna el centro de un área bajo la curva. Para el caso Sugeno, es el promedio de las ponderaciones.

1.6 Técnica adaptativa Neuro-Difusa ANFIS

El acrónimo ANFIS derivado de "*Adaptative Neuro Fuzzy Inference System*". Esta técnica es utilizada para crear o mejorar sistemas de inferencia difusos de tipo Sugeno. Estas técnicas, combinan los sistemas de inferencia difusos y las redes neuronales, aprovechando las características sobresalientes de cada modelo. Por un lado, los FIS proveen un mecanismo intuitivo y de alto nivel para representar el conocimiento mediante la utilización de reglas Si-Entonces; y por el otro, las redes neuronales poseen un alto grado de adaptabilidad, capacidad de aprendizaje y generalización.

La construcción de herramientas que se nutren de estas dos áreas constituye un mecanismo eficiente a la hora de modelar sistemas reales. La idea básica detrás de esta técnica es muy simple: proveer al proceso de modelado difuso un método que permita "aprender" de la información contenida en un conjunto de datos. Ahora si ya existe un FIS inicial, se utiliza para optimizar los parámetros de las correspondientes funciones de pertenencia.

1.6.1 Funcionamiento del modelo ANFIS

El aprendizaje en la aplicación de este método se basa en la tecnología usada en las redes neuronales para este fin. Este método funciona similar a una red neuronal. Para usar un modelo ANFIS se necesita el conjunto de datos de entrada-salida, y un modelo previo de inferencia (FIS), a partir del cual se comienza a ajustar los parámetros de las funciones de pertenencia del antecedente y de las funciones lineales del consecuente, basándose en alguna técnica de gradiente y mínimo cuadrados (ejemplo el conocido por "*backpropagation*"). Esto permite que se realice el proceso de aprendizaje a partir de los datos suministrados. El cálculo o ajuste de estos parámetros es efectuado por medio de un vector gradiente, el cual provee una medida de que tan bien el FIS está modelando los datos de entrada-salida.

El método ANFIS se detiene cuando el número de épocas es alcanzado o el error deseado es menor o igual al requerido. ANFIS también puede ser invocado usando el argumento opcional para prueba de modelos. El tipo de prueba que se realiza con esta opción es llamada: chequeo de sobreajuste del modelo, y el argumento necesario es el conjunto de los datos usados para la validación o prueba.

1.7 Aplicación de sistemas computacionales en el sector agroindustrial.

El desarrollo de técnicas computacionales aplicadas en la agricultura ha llegado a ser cada vez más crucial en la provisión de alimentos, el manejo ambiental y el abastecimiento de energía. La naturaleza intrínsecamente compleja, dinámica, y no lineal de los sistemas agropecuarios ha requerido siempre las soluciones basadas en técnicas y tecnologías avanzadas, para proporcionar mayor exactitud, una mayor comprensión y soluciones apropiadas.

Actualmente se realizan diversos esfuerzos a nivel mundial para poder aplicar los conocimientos en IA al sector agropecuario, dando lugar a diversas conferencias periódicas entre las que se pueden citar el "*World Congress of Computers in Agriculture and Natural Resources", "International Workshop on Artificial Intelligence in Agriculture", "EFITA Congresses: European Federation for Information Technology in Agriculture"*, así mismo se han creado diversos Journals con el objetivo principal de divulgar las investigaciones acerca de este tema, entre los más importantes se pueden

citar, "*Agricultural Systems*", "*Computers and Electronics in Agriculture*", "*Biosystems*", "*Biosystems Engineering*".

Algunas de las aplicaciones que usan IA en problema de la agricultura, están expuestas en la Figura 1.2:

A. <u>Ingeniería de Poscosecha</u>

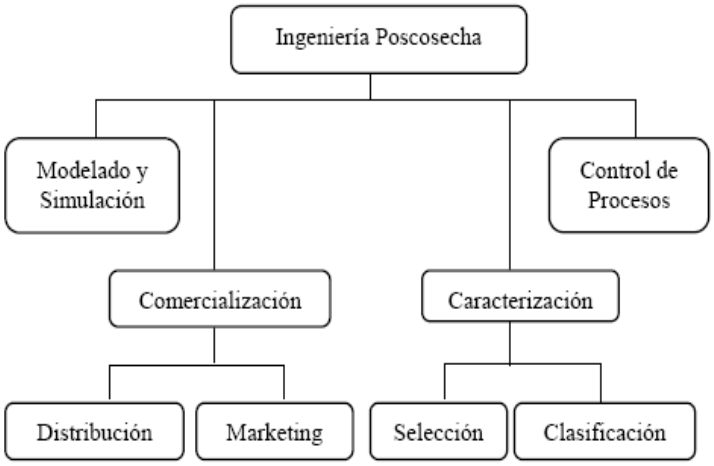

Figura 1.2. Líneas de Trabajo en Ingeniería de Poscosecha.

La aplicación de la IA para el modelado y la simulación de procesos industriales y fisiológicos de productos agropecuarios es un campo que se está empezando a abordar en la actualidad, algunos de los trabajos realizados en esta área son un modelo de control de predicción para procesos industriales de evaporación por medio de redes neuronales (Benne 2000), una red neuronal topológica para el modelado de secado de granos (Farkas 2000), Modelado y optimización de temperatura en procesos térmicos usando redes neuronales y algoritmos genéticos (Chen and Ramaswamy 2002), una simulación de daños a Melocotones en una línea de transporte (Bielza 2003), y una simulación de un proceso de secado de Echinacea Angustifolia con redes neuronales (Erenturk 2004).

La utilización de la IA para la caracterización de productos agropecuarios es uno de los campos con mayor aporte debido a su gran importancia, para la selección y clasificación de productos, entre ellos podemos dictar: sistema de clasificación de Manzanas por tamaño que usa redes Bayesianas, redes neuronales y lógica difusa (Shahin 2001), sistema para la clasificación de carnes en el Reino Unido mediante el uso de

luminosidad y tres algoritmos diferentes de IA (Diez 2003), sistema de inspección automática de calidad de arroz de acuerdo al contenido de proteína mediante el uso de técnicas de visión computacional (Kawamura 2003), 0 Sistema para la clasificación de Manzanas por medio de su consistencia usando diferentes algoritmos de clasificación como por ejemplo redes neuronales (Moshou 2003), Autentificación de jugos fermentados de uvas blancas por medio de sensores de aroma, Transformadas de Fourier (IR) y espectrometría (UV) usando algoritmos genéticos, sistemas expertos y clasificación multivariable (Roussel 2003), Revisión de las diferentes aplicaciones de sistemas de visión de computadora por medio de tomografía para evaluar propiedades físicas de la comida (Abdullah 2004), Base de datos de propiedades físicas de materiales provenientes del sector agropecuario como apoyo a sistemas inteligentes basados en casos (Nesvadba 2004).

B. Control Ambiental

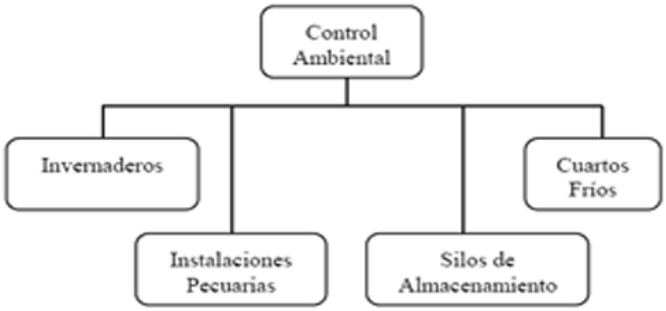

Figura 1.3. Líneas de Trabajo en Control Ambiental

En la actualidad la IA ha obtenido uno de sus mayores impulsos en el sector agropecuario con el manejo de ambientes, y es en estructuras tipo invernadero donde mayor auge se ha presentado, entre los trabajos que se pueden mencionar encontramos diversos esfuerzos que tratan de realizar un control ambiental no lineal con retroalimentación mediante el uso de diferentes técnicas de Inteligencia Artificial (Pasgianos 2003) (Tantau 2003) (Magdi. A. Koutb 2004).

Las investigaciones para el control ambiental de silos de almacenamiento en los últimos años no han sido numerosas, y se puede señalar, un sistema de control ambiental para el almacenamiento de papa usando controladores con lógica difusa (Gottschalk 2003).

C. Maquinaria y Mecanización Agrícola

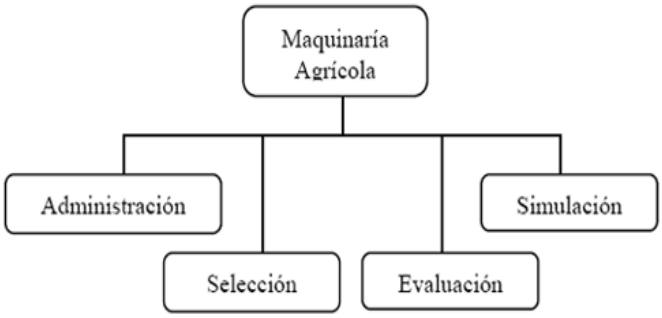

Figura 1.5. Líneas de Trabajo en Ingeniería de Riegos.

También el uso de técnicas de IA en la administración de maquinaría agrícola ha tenido su espacio, donde se pueden nombrar desde años anteriores al año 2000, un sistema experto para el soporte de decisiones en la administración de maquinaría agrícola (Lal et at, 1992) y un sistema basado en conocimiento para el diseño y mantenimiento de maquinaría agrícola (Kusz & Marciniak, 1995).

Para la selección de maquinaría agrícola se han desarrollado en los últimos años diversos sistemas expertos, en época más reciente se puede mencionar, un sistema basado en conocimiento para la selección de juntas en maquinarías de tipo agrícola (Deniz Yilmaz 2004).

D. Ingeniería de Riegos.

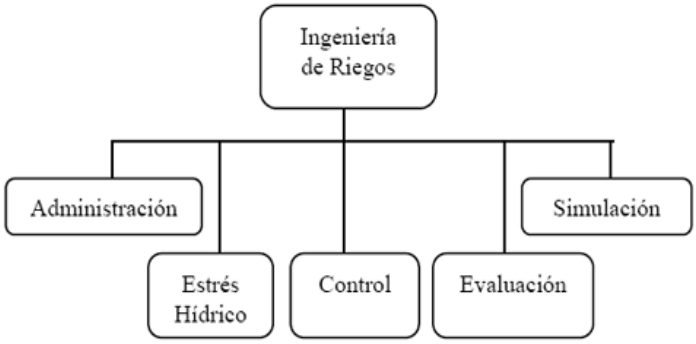

Figura 1.6. Líneas de Trabajo en Ingeniería de Riegos.

Las técnicas de IA para la administración de sistemas de riego han sido ampliamente usadas, logrando su mayor impulso con los sistemas expertos. Entre los trabajos realizados últimamente podemos señalar, un sistema de toma de decisiones para optimizar la programación de riego usando propiedades bio-físicas del sistema (Bergez 2002), un sistema para la adquisición automática de conocimiento para sistemas expertos dedicados a la irrigación y fertilización (Rafea 2003), un método de división jerárquica para optimización de estrategias para la irrigación de cultivos (Bergez 2004), un sistema experto para la configuración de la programación de riego basado en investigación de operaciones (Iman Hassen 2004), una herramienta especializada en la construcción de sistemas expertos para irrigación (Samhaa R. El-Beltagy 2004).

Una de las aplicaciones más interesantes de la IA en el área de la Ingeniería de Riegos es la determinación del estrés hídrico de un cultivo, entre los trabajos realizados en los últimos años se puede nombrar, un sistema de visión de máquina para indicar el nivel de estrés hídrico de un cultivo de Tomate (Farkas I. 2004).

E. Producción Vegetal

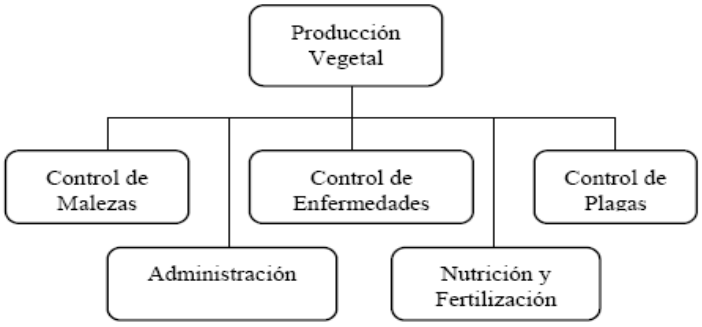

Figura 1.7. Líneas de Trabajo en Producción Vegetal

Como soporte en las diferentes labores de sistemas de producción vegetal ha sido de amplia utilización en el control de malezas, enfermedades y plagas. En los últimos años ha tenido gran acogida debido a su importancia económica y ambiental.

Entre los trabajos realizados en los últimos años para el control de malezas se pueden mencionar, un sistema para discriminar las malezas del cultivo con imágenes por medio de redes neuronales (Aitkenhead 2003), una clasificación de datos híper espectrales por medio de árboles de decisión y redes neuronales para la determinación del estrés ocasionado por malezas al Maíz (Goel 2003), un sistema experto para el control y manejo de malezas en Irán [6], una plataforma robótica autónoma para el descubrimiento de malezas en un cultivo el cual usa un sistema de control reactivo con retroalimentación (Bak 2004).

Existen otras aplicaciones en el sector agrario en las cuales se trabaja arduamente para la obtención de mejores resultados. Todas ellas se ven relacionada por un factor que no solo afecta a la agroindustria: el clima. En la actualidad existen proyectos muy importantes que tratan este tema, ya que cada año que pasa se observen los cambios climáticos en diferentes aéreas del hemisferio.

1.8 Problemas de la agro-meteorología.

Hasta la década del 60, las Ciencias de la Atmósfera tuvieron como eje dominante a la Meteorología y más precisamente, la Dinámica de la Atmósfera como un desarrollo particular de la Hidrodinámica. Ello obedeció a que el desafío tecnológico fundamental era la previsión del tiempo. A partir de los 60, las innovaciones tecnológicas ocurridas en otros campos del saber tuvieron un enorme impacto sobre las Ciencias de la Atmósfera y modificaron los ejes principales de la investigación y del desarrollo tecnológico.

Surge la agro meteorología como una aplicación de la meteorología. En los últimos años ha tomado vital importancia por la preocupación de los cambios climáticos y el crecimiento de la población mundial. Su estudio es necesario porque el tiempo atmosférico desempeña un papel importante en la obtención de buenos rendimientos en las prácticas agrícolas.

Por otra parte, los enfoques estadísticos a la predicción del tiempo incluyen distintos enfoques (WILKS 1995; SANTHANAM M. 2001). Los sistemas inteligentes aparecen como alternativas útiles a las técnicas de modelado estadístico tradicional (HAYKIN 1994).

Recientemente se han desarrollado aplicaciones automatizadas para la predicción y control agro meteorológicas (Bressan 2007). El sistema que se propone en (Bressan 2007), está encaminado a utilizar la predicción de infección de plantas dañinas en determinados cultivo con Lógica Fuzzy y Redes Bayesianas. Otros casos son modelado de la cosecha de un cultivo (Hartog et al, 1995) (Plant, 1995).

En Cuba se ha desarrollado en los últimos años en la predicción meteorológica, sistemas como CubaForecast, por el Centro Meteorológico de Cienfuegos en el año 2003 (Díaz 2003), PROMETEM 1.0 por el Centro Meteorológico de Camagüey en el 2005 (Caballero 2005). Esto ha llevado a un desarrollo también en la aplicación de sistemas agros meteorológicos, dado que unos de los renglones económicos de nuestro país es la agricultura se trabaja arduamente en lograr avances considerados en obtener mejores resultados en la producción.

Uno de los problemas encontrados es regionalización de los sistemas por las características que presenta cada área y tipo de cultivo determinado. Pero lo que se quiere lograr es un sistema que pueda ser aplicado en los principales cultivos del país a nivel nacional.

1.9 Conclusiones parciales.

La elección de las técnicas para el manejo de estos valores dinámicos resulta una tarea compleja, pues un mismo método en determinadas situaciones produce predicciones precisa para determinadas regiones y en otros no.

Los SIF se han utilizado frecuentemente para determinar relaciones entre variables que dependen del tiempo con resultados adecuados.

Los ANFIS constituyen un sistema Hibrido que conjuga efectivamente las ventajas de los SIF y las RNA puesto que las RNA se usa para optimizar la estimación de los parámetros que conforman el SIF.

CAPITULO 2 ANÁLISIS Y DISEÑO DE LA HERRAMIENTA. SISTEMA DE PREDICCIÓN AGRO METEOROLÓGICO.

Se proponen dos técnicas de agrupamiento para la construcción de sistemas difusos. En este capítulo se describen estas tecnicas asi como correlaciones para la obtención del modelo entre los rasgos significativos. Mostrando los resultdos obtenidos en la prediccion de la prediccion de los agentes de humedad y temperatura.

2.1 Datos de Campo.

Los datos obtenidos para la construcción del modelo fueron medidos durante todo un año con un periodo de 15 minutos donde intervienen variables agro-meteorológicas tales como:

1. Fecha [date]
2. Tiempo [time]
3. Voltaje de la batería (V) [real]
4. Velocidad del viento (km/h) [real]
5. Dirección del viento (°) [real]
6. Precipitación (mm) [real]
7. Temperatura (°C) [real]
8. Humedad relativa (%) [real]
9. Humedad en las hojas (U) [real]
10. Radiación global (W/m²) [real]
11. Nivel RF de entrada (µV) [real]
12. Nivel RF de salida (µV) [real]

Estas mediciones durante el año alcanzaron a tener un gran volumen, siendo así se obtuvieron 34 967 muestras. Por la estacionalidad de algunos rasgos como la temperatura, la humedad relativa, la humedad en las hojas, entre otros, no presentaron variación alguna en las mediciones hechas en un día, se trabajó entonces con los promedios diarios de cada día reduciendo la muestra a 345 mediciones hechas en el año (ver anexo 1).

El resultado obtenido se puede observar en la Figura 2.1

	1	2	3	4	5
1	165.8906	54.75	242.0469	17.8891	2.625
2	205.4842	67.1474	159.5368	15.44	2.4632
3	215.7684	70.2316	161.9368	13.8589	2.7474
4	198.6737	74.8947	134.7789	13.4084	2.5684
5	207.3579	73	127.3053	15.3789	2.0737
6	172.3053	70.5263	147.8947	15.9463	1.6737
7	221.2105	66.0211	109.7579	15.8189	2.9579
8	166.1789	66.3895	157.1263	15.1947	3.3579
9	206.5789	66.8316	119.1474	16.8516	2.1158
10	189.3579	66.6947	147.0737	17.0042	1.8316
11	169.7684	83.1789	46.4	15.7463	0.4316

Figura 2.1. Mediciones obtenidas.

2.2 Selección de entradas.

La selección de las variables de entradas de cualquier modelo de predicción es uno de los primeros pasos, para lo cual se aplican diferentes pasos y técnicas. Uno de los pasos en este sentido es la consulta a expertos con vistas a determinar las relaciones entre variables ya históricas y prácticamente comprobadas sus existencia. Estos criterios de expertos aplicado en este caso dio por resultado que en posibles modelos desarrollados para predecir tanto la temperatura como la humedad relativa no es necesario considerar algunas variables, como Nivel RF de entrada, Nivel RF de salida, Voltaje de la batería.

La selección de las variables para la predicción se basó en el análisis de correlación, el análisis de correlación cruzada y la función de auto correlación parcial.

Las variables de entrada se organizaron de tal modo que se pueda predecir la humedad relativa y la temperatura en (t+1) usando la serie de demanda y variables explicadas. Inicialmente se analizó la auto correlación de la serie de humedad y temperatura, así como la correlación cruzada respectivamente para detectar el rezago significativo.

Para el caso de la serie humedad, se encontró que esta presenta una tendencia. Lo cual se puede verificar en la función de auto correlación parcial (PACF) que muestra una correlación significativa entre los valores t-1, t-3, t-4, t-7 Figura 2.2.

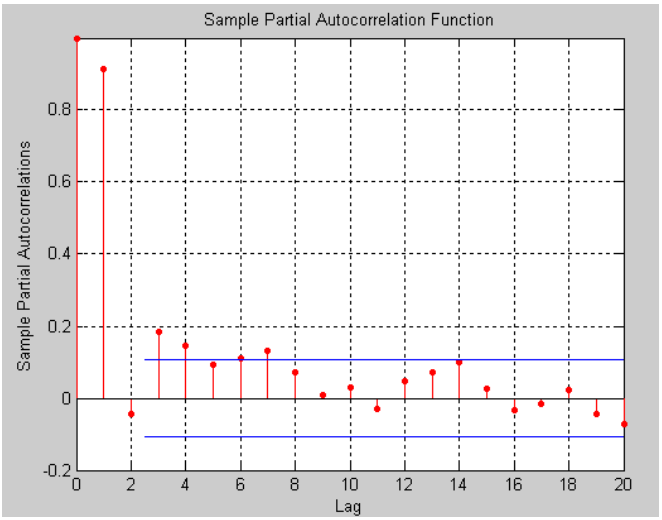

Figura 2.2 Función de auto-correlación de la humedad.

Analizando de igual forma la temperatura como muestra la Figura 2.3 existe una correlación significativa entre los valores t-1, t-2, t-4, t-5, t-10.

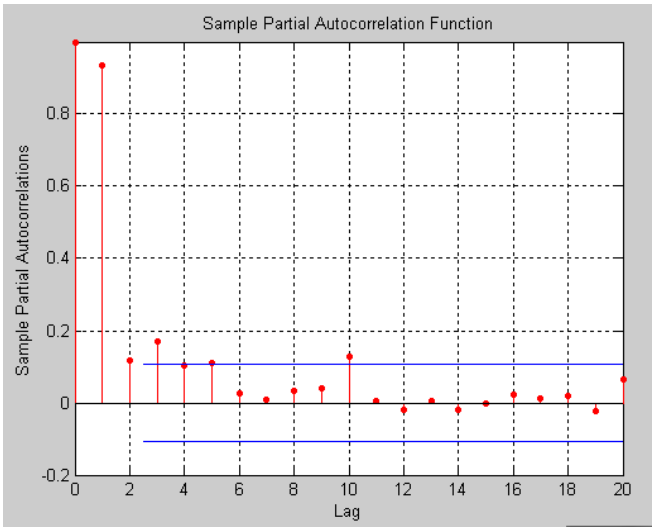

Figura 2.3. Función de auto-correlación de la temperatura.

La correlación cruzada entre la Humedad vs. Temperatura y Humedad vs Velocidad del viento. Los resultados de la Tabla 2.1 y Tabla 2.2 respectivamente, indican la fuerza de relación lineal entre la humedad en el periodo t y la temperatura rezago k (con k entre 1 y 5), como se indica en la tabla existe relación directa entre la variables. El comportamiento de la humedad está relacionada con la temperatura rezagos k = 1.

Humedad vs Temperatura

lang	Cross-Correlation
0	0.2607
1	0.2812
2	0.1875
3	0.1093
4	0.0653
5	0.0583
6	-0.0046
7	-0.0149
8	0.0200
9	-0.0265
10	0.0982
11	0.1107
12	-0.0356
13	-0.0428
14	-0.0479
15	0.0014
16	-0.0355
17	0.0005
18	0.0193
19	0.0789
20	0.0202

Tabla 2.1 Correlación cruzada entre Humedad y Temperatura.

Humedad vs V. Viento

lang	Cross-Correlation
0	-0.0712
1	-0.0120
2	-0.0011
3	0.0379
4	0.0588
5	0.0489
6	0.0703
7	0.0390
8	-0.0022
9	-0.0181
10	0.0337
11	0.0221
12	0.0556
13	0.0741
14	0.0040
15	-0.0185
16	0.0045
17	-0.0027
18	-0.0066
19	-0.0039
20	-0.0481

Tabla 2.2 Correlación cruzada entre Humedad y Velocidad del Viento.

Las variables seleccionadas se muestran en la Tabla 2.3 y se considera al menos dos variables por factor con mayor correlación en la Humedad y la Temperatura.

Variables de Entrada	
Humedad relativa	HR(k-2)
Humedad relativa	HR(k-1)
Dirección del viento	DV(k-1)
Dirección del viento	DV(k)
Velocidad del viento	VV(k)
Velocidad del viento	VV(k-1)
Radiación global	RA(k-1)
Temperatura	TE(k-1)
Temperatura	TE(k-2)

Tabla 3. Variables finales seleccionadas para el modelo.

Teniendo las variables de entrada seleccionadas entonces se forma el conjuto de entrenamiento y control. Para la muestra de 345 mediciones se tomó como cojunto de entrenamiento 254 mediciones y de control 91. Estos conjuntos tomados se utilizaron para predecir tanto la humedad como la temperatura correspondiente con cada una de las entradas seleccionadas (ver anexo 2).

2.3 Construcción de sistema Neuro-Fuzzy.

Por lo general, para construir un sistema neuro-difuso se parte de conjunto de datos numéricos pertenecientes al espacio de entrada-salida. La construcción del sistema consta de dos fases esenciales:

a) Determinación de la estructura del modelo *neuro-fuzzy* (tipo y cantidad de funciones de pertenencias y reglas).
b) Estimación de los parámetros del modelo. Fase de aprendizaje donde se ajustan y optimizan los parámetros de cada regla difusa construida en la primera fase.

Parte importante en la determinación de la estructura del modelo consiste en la división de los espacios de entrada y de salida. Como resultado del proceso de partición, cada partición representa un conjunto difuso. En el sistema que utilizan las normas nítidas, no hay solapamiento entre los límites de las particiones. Sin embargo, normalmente esta superposición existe en sistemas difusos. El particionado del espacio de entrada posibilita construir los antecedentes del sistema de reglas. La estructura y el número de reglas van a estar determinados por los resultados de la división efectuada en el espacio de entrada (Fakherddine O. Karray 2004).

2.4 La técnica de agrupamiento.

La idea del particionado por agrupamiento (*clustering, en inglés)* se basa en clasificar, dividir o agrupar una colección de objetos (vector de puntos de datos) en distintos subconjuntos o *clusters.* Los objetos que pertenecen a un mismo *cluster* tienen propiedades similares respecto a un grupo de determinados criterios. Dentro de esta técnica, hay diferentes formas de realizarla, entre ellas la conocida por agrupamiento clásico donde cada objeto (puntos de datos) es asignado a uno y solo a un grupo sin ningún grado de membresía de otros grupos. Por otra parte en la técnica más reciente denominada agrupamiento difuso (*fuzzy clustering)* cada punto de dato pertenece a un grupo solo con un grado específico determinado por el valor de la función de membresía a un determinado rasgo o conjunto difuso. Por lo tanto, un punto de dato puede pertenecer a más de un grupo con diferentes valores de membresía respecto a cada grupo.

2.4.1 Método Fuzzy C-Means

Una de las variantes más popularmente usadas dentro de esta última técnica es la denominada agrupamiento por *Fuzzy C-mean* (abreviado FCM en MATLAB y en otros sistemas desarrollados y en uso frecuente). Esta técnica fue introducida por Dunn (Dunn 1973) y mejorado por Bezdek (Bezdek 1981). El algoritmo es una extensión del clásico algoritmo c-*mean* donde los puntos de datos son divididos en grupos de acuerdo a la distancia a cada punto, que es llamado el *"centro del grupo".*

Este método se basa en la reducción al mínimo de la siguiente función objetivo:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2, 1 \leq m < \infty \qquad (1)$$

donde *m* es cualquier número real mayor que 1,

u_{ij} es el grado de pertenencia de x_i en el grupo *j*,

x_i es el conjunto i-ésimo vectores que representa los datos medidos,

c_j es el centro d-dimensión de la agrupación.

El particionado difuso se realiza a través de un procedimiento iterativo de optimización de la función objetivo indicado por (1), con la actualización de los valores de pertenencia u_{ij} a cada centro y las coordenadas de los centros de grupo c_j por:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}}, \quad c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} * x_i}{\sum_{i=1}^{N} u_{ij}^{m}} \tag{2}$$

Este procedimiento iterativo se detendrá cuando se alcance la siguiente condición

$$max_{ij}\left\{\left|u_{ij}^{(k+1)} - u_{ij}^{(k)}\right|\right\} < \varepsilon, \tag{3}$$

donde ε es un criterio de terminación con un valor entre 0 y 1. Por otra parte k es un índice representando el número de la iteración en curso. Este procedimiento converge a un mínimo o un punto de ensilladura de J_m.

El algoritmo se compone de los siguientes pasos:

1. Inicializar $U = [u_{ij}]\ matriz, U(0)$.
2. Cálculo de los centros de vectores $C(k) = [c_j] con\ U(k)$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} * x_i}{\sum_{i=1}^{N} u_{ij}^{m}} \tag{4}$$

3. Actualización de $U(k),\ U(k+1)$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}} \tag{5}$$

4. Si $\|U(k+1) - U(k)\| < \varepsilon$ entonces se detiene, en caso contrario retorna al paso 2.

Otra de las técnicas de agrupamiento es el denominado agrupamiento sustractivo (en inglés: *subtractive clustering).* Este método fue propuesto por Chiu (Chiu 1994) como una mejora a un método similar denominado "método Mountain"; Yager (1981). Ambos métodos se plantean como objetivo un algoritmo para determinar c dentro de un conjunto de datos, así como un estimado inicial de los mismos, de tal forma que pueda aplicarse dentro de las técnicas de agrupamiento como el agrupamiento FCM.

2.4.2 Subtractive Clustering.

El método se basa en considerar inicialmente cada punto como un centro de cluster en potencia. Luego, se define una función que mida para cada punto su potencial Pi como centro de grupo. Con ayuda de estos Pi y mediante un procedimiento iterativo se logra una estimación del número P de grupos. A continuación se describe el algoritmo:

Paso 1. Se normalizan los datos en Y en el intervalo [0,1], con lo que se restringen los datos a un hipercubo de igual dimensión (unitaria) en cada dimensión.

Paso 2. Se calcula el potencial de cada punto y_i.

$$P_i = \sum_{i=1}^{m} e^{-\alpha \|y_i - y_1\|^2} \quad (6)$$

$$\alpha = (2/r_a)^2 \quad (7)$$

Donde

P_i representa el potencial del i-ésimo punto

r_a es el radio definiendo una vecindad.

P_i es función de las distancias del punto i al resto de los $(m-1)$ puntos en Y.

Paso 3. Seleccionar el punto con el mayor P_i como el primer centro de grupo y sus coordenadas como las coordenadas del centro de grupo:

$$P_i^* = \max(P_i) \quad (8)$$

Paso 4. Se corrige el potencial de cada punto y_i. Para ello, se resta a cada uno una cantidad que es función de su distancia al primer potencial de punto detectado:

$$P_{i-corr} = P_i - P_1^* e^{-\beta \|y_i - y_1\|^2} \quad (9)$$

$$\beta = (2/r_b)^2 \quad (10)$$

Dónde $r_b = 1.25 * r_a$. Al aplicar (8), los potenciales de los puntos cercanos al primer centro serán los que se vean más fuertemente reducidos.

Paso 5. Se detecta el siguiente centro de grupo como el punto con el máximo potencial corregido:

$$P_i^* = \max\left(P_{i-corr}\right) \qquad (11)$$

Paso 6. Se repiten 3 y 4 de forma continua y según el resultado del siguiente criterio de terminación:

Si $P_i^* > \overline{\varepsilon} \cdot P_1^*$

Se acepta y_i^* como centro de grupo y se vuelve a 3.

Sino si $P_i^* < \underline{\varepsilon} \cdot P_1^*$

Se rechaza y_i^* y termina el proceso de agrupamiento.

Sino

Sea d_{min} la distancia más corta entre y_i^* y todos los centros de grupos previamente encontrados:

Si $\frac{d_{min}}{r_a} + \frac{P_i^*}{P_1^*} \geq 1$

Se acepta y_i^* como un centro de grupo y se vuelve a 3.

Sino

Se rechaza y_i^* y se fija su potencial en 0. Se selecciona el siguiente punto con P_{i-corr} más alto y se vuelve a evaluar todo el criterio de terminación.

Fin

En el criterio anterior ε representa un umbral sobre el cual se acepta (sin lugar a dudas) un punto como centro de agrupamiento. Por el contrario, ε representa un umbral bajo el cual se rechaza (sin lugar a dudas) un punto como centro. La región entre estos 2 valores se evalúa según la parte final del criterio.

2.4.3 Aplicando técnicas de agrupamiento al problema.

El método de agrupamiento difuso FCM por medio de la función “genfis3”, del programa MATLAB 7.0 es utilizado para generar sistemas de inferencia difusos (FIS) automáticamente, basados en los conglomerados creados. Generando un FIS tipo Sugeno de un conjunto de datos de entrada-salida usando para ello la función “fcm”, que es la encargada de generar la cantidad de conglomerados y la ubicación de sus centros.

La función "genfis3" utiliza la información generada por "fcm" y genera los parámetros de las funciones de pertenencia en la entrada y salida de los FIS y construye los sistemas de inferencia difusos. Adicionalmente, "genfis3" determina el número de reglas basándose en el número de conglomerados generados, determina los antecedentes y finalmente usa estimación lineal de mínimos cuadrados para determinar las consecuencias de cada regla.

Otras de las funciones encontradas en MATLAB es "genfis2" que tienen el mismo propósito de generar FIS automáticamente. La diferencia radica en el método de agrupamiento siendo genfis2 por *"subtractive clustering"* obteniendo la información de "subclust" para generar el FIS.

Los resultados obtenidos al aplicar genfis3 fueron satisfactorios en comparación al ser evaluados con genfis2, teniendo en cuenta el coeficiente de error de predicción. (ver anexo 3).

2.5 Construcción del Sistema de Inferencia Borroso.

Cuando se tiene una sola salida, "genfis3" es utilizado para generar un FIS inicial el cual pueda ser mejorado con la técnica de aprendizaje adaptativa neuro-difusa ANFIS.

El propósito es ajustar los valores de los parámetros (funciones de pertenencia, coeficientes lineales, etc.) de manera que el modelo pueda describir en forma satisfactoria al sistema real.

Para ello se sigue un conjunto de pasos los cuales permitieron la obtención del modelo descrito por dicha función.

Obtención del modelo difuso.

Paso 1. Se separan las variables medidas a criterio del experto que influyen directamente en las variables objetivos humedad y temperatura.

Paso 2. Se seleccionan las observaciones que servirían para entrenamiento y prueba de los modelos difusos, basados en las comparaciones hechas en el epígrafe 2.2.

TE(k) = F(HR(k-1),DV(k-1),DV(k),TE(k-2),TE(k-1))
HR(k) = F(RA(k-1),TE(k-2),HR(k-2),VV(k),DV(k))

Paso 3. Para el modelo obtenido donde intervienen las cinco variables descritas anteriormente se obtuvo el FIS aplicando *genfis3.*

Resultados

Como se muestra en la Figura 11 para cada una de las cinco entradas se obtuvieron siete particiones las cuales son descritas por las funciones de membresía Figura 2.4 para la temperatura en instante k HR(k).

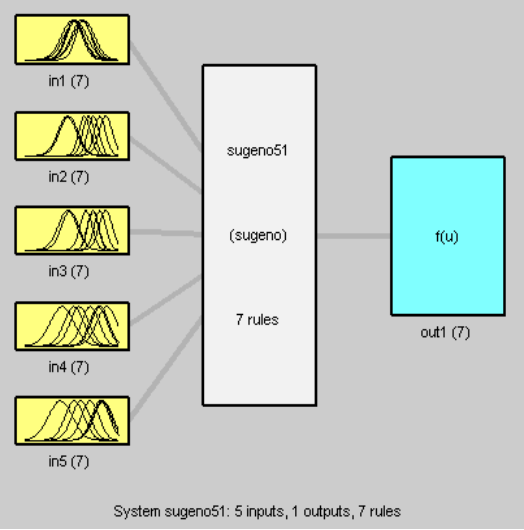

Figura 2.4. FIS tipo T-Sugeno.

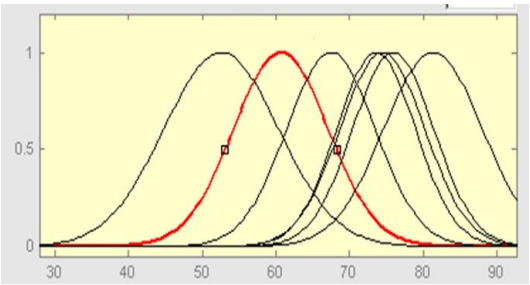

Figura 2.5. Funciones de pertenencia.

Para la representación de la función de membresía se utilizo la gausiana ya que es la que registra mejores resultados en comparación con otros.

Las reglas generadas por FIS tipo *Takagi-Sugeno-Kang* fueron:

1. If (in1 is in1cluster1) and (in2 is in2cluster1) and (in3 is in3cluster1) and (in4 is in4cluster1) and (in5 is in5cluster1) then (out1 is out1cluster1) (1)
2. If (in1 is in1cluster2) and (in2 is in2cluster2) and (in3 is in3cluster2) and (in4 is in4cluster2) and (in5 is in5cluster2) then (out1 is out1cluster2) (1)
3. If (in1 is in1cluster3) and (in2 is in2cluster3) and (in3 is in3cluster3) and (in4 is in4cluster3) and (in5 is in5cluster3) then (out1 is out1cluster3) (1)

4. If (in1 is in1cluster4) and (in2 is in2cluster4) and (in3 is in3cluster4) and (in4 is in4cluster4) and (in5 is in5cluster4) then (out1 is out1cluster4) (1)
5. If (in1 is in1cluster5) and (in2 is in2cluster5) and (in3 is in3cluster5) and (in4 is in4cluster5) and (in5 is in5cluster5) then (out1 is out1cluster5) (1)
6. If (in1 is in1cluster6) and (in2 is in2cluster6) and (in3 is in3cluster6) and (in4 is in4cluster6) and (in5 is in5cluster6) then (out1 is out1cluster6) (1)
7. If (in1 is in1cluster7) and (in2 is in2cluster7) and (in3 is in3cluster7) and (in4 is in4cluster7) and (in5 is in5cluster7) then (out1 is out1cluster7) (1)

Siendo:

In1 = HR(k-1),
In2 = DV(k-1)
In3 = DV(k)
In4 = TE(k-2)
In5 = TE(k-1)
Out1 = HR(k)

Como se aprecia en cada una de las reglas está presente el consecuente out1 que tiene la característica que es lineal para cada uno de las reglas.
Con el fragmento de código mostrado a continuación se logra construir FIS inicial.

```
datin = trn_data(:,1:5); %datos entrada
datout = trn_data(:,6);  %datos salida
plot(trn_data(:,6));
a=genfis3(datin, datout);
figure(1)
plotfis(a);
```

Obteniendo también cómo se comporta la humedad relativa en el año para el modelo seleccionado, como se muestra en la Figura 2.6.

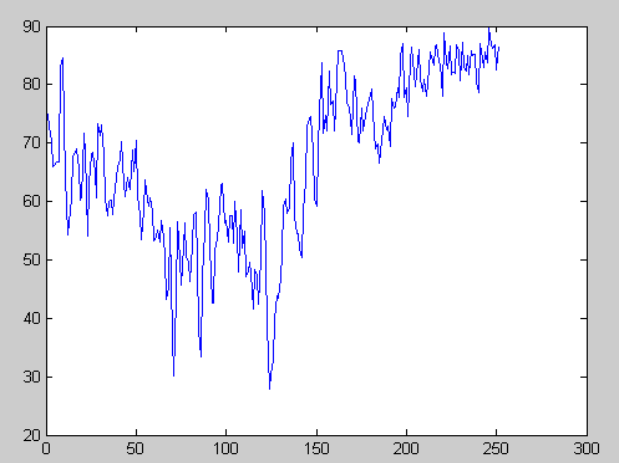

Figura 2.6. Comportamiento de la humedad relativa.

Se visualiza cómo se comporta las cinco entradas y evaluamos el FIS obtenido mediante la función *evalfis* como se muestra:

```
figure(1);
subplot(1,1,1)
fuzout=evalfis(datin,a);
subplot(2,1,1)
plot(datin)
subplot(2,1,2)
plot([datout fuzout])
```

Podemos observar el valor de predicción (rojo) y el valor real (azul) en la parte inferior de la Figura 2.7 (en la Figura 2.8 es una ampliación del valor predecido) y el comportamiento de las cinco entradas.

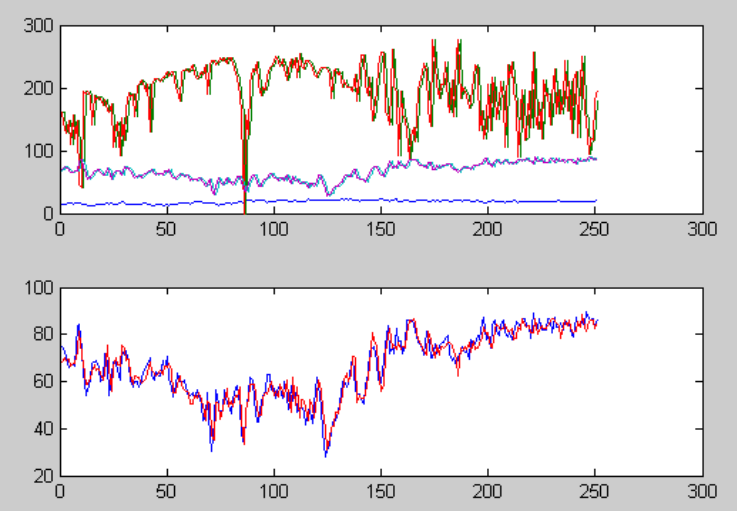

Figura 2.7. Resultados al evaluar el FIS con muestra de entrenamiento.

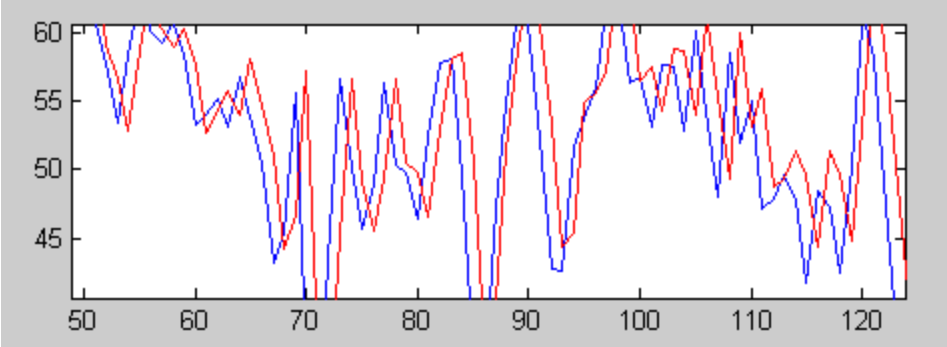

Figura 2.8 Resultados al evaluar el FIS con muestra de entrenamiento.

2.6 Aplicando Técnica Adaptativa Neuro-Difusa ANFIS.

En esta sección se muestran los resultados obtenidos mediante la utilización del modelo ANFIS. Dicho modelos tienen la característica de permitir cambios en las condiciones de entrada como número de neuronas, funciones de activación, algoritmos de entrenamiento, tasa de aprendizaje, numero de funciones de pertenencia, tipo de función de pertenencia, entre otros. Esto le da al proceso de obtención de resultados la característica de ser un proceso dinámico en el sentido que se cambian los parámetros y se observan los resultados hasta que se consigue un nivel de error deseado.

Método de Optimización	Hibrido
Tolerancia de error	0
Épocas	40

Tabla 2.4. Parámetros de entrenamiento del ANFIS.

Obtenido el FIS aplicamos entonces modelo ANFIS como muestra la línea de código siguiente:

%fismat es el FIS inicial

```
[fis,error,stepsize,chkFis,chkErr]=anfis(trn_data,fismat,40
,[],chk_data);
```

Los parámetros obtenidos para el modelo fueron:

Número de nodos: 92
Número de parámetros lineales: 42
Número de parámetros no lineales: 70
Número total de parámetros: 112
Número de patrones de entrenamiento: 251
Número de patrones de validación: 92
Número de reglas borrosas generadas: 7

La Figura 2.9 y Figura 2.10 muestra la serie original y el pronóstico que realiza el FIS para los datos de validación.

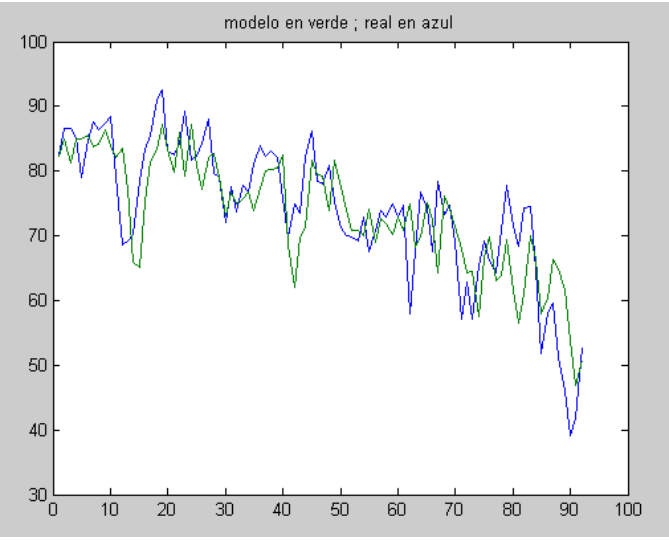

Figura 2.9. Valores de Humedad y salida del modelo ANFIS grupo de validación.

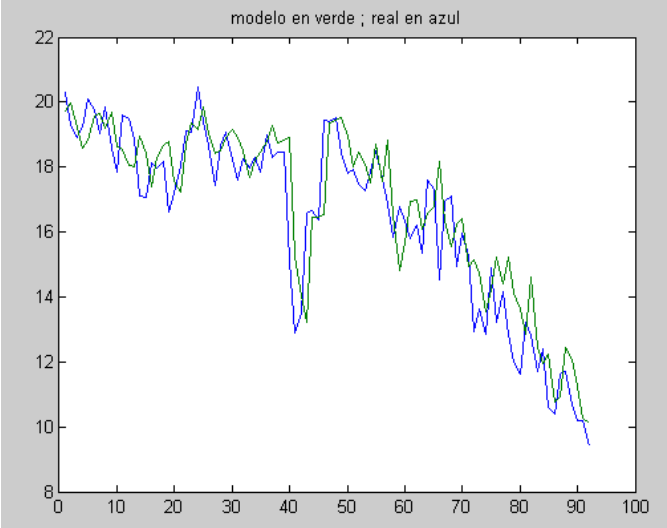

Figura 2.10. Valores de Temperatura y salida del modelo ANFIS grupo de validación.

2.7 Conclusiones parciales.

En la solución al problema de pronósticos automatizados de las temperaturas y la humedad basándose en modelos ANFIS se lograron resultados satisfactorios con error de 0.2135.

Los modelos considerados son fáciles de implementar y producen una adecuada correspondencia entre las variables de entrada y salida a través del entrenamiento del modelo. Los modelos pueden captar la información de los datos como llegan y se reduce el tiempo de análisis.

Los resultados obtenidos presentaron precisión en el pronóstico de la temperatura y la humedad, no siendo así en interpretación de las reglas Si-Entonces obtenidas por el modelo (sin embargo esto depende del dominio del problema). Este resultado posiblemente sea consecuencia del número limitado de casos que se cuenta para el entrenamiento y validación del modelo ANFIS. (El número de parámetros que debe calcular un modelo ANFIS es grande cuando se consideran más de 5 variables de entrada por lo que se requiere una base de datos suficientemente grande). No fue posible correr el modelo ANFIS con más variables de entrada.

CAPÍTULO 3 IMPLEMENTACIÓN DE AGENTES PARA LA DETERMINACIÓN DE UN SISTEMA MULTIAGENTE DE PRONÓSTICO DE RENDIMIENTO.

Este capítulo está dedicado al procedimiento para implementar los agentes que formaran parte del sistema multiagente de pronóstico de rendimiento.

3.1 Visión general del sistema.

La estimación o predicción del rendimiento trata el análisis de las variables agro meteorológicas que afectan directamente al cultivo. Particularmente la predicción del rendimiento partiendo de un gran volumen de datos dados pretende determinar cuáles son las variables que afectan el desarrollo del cultivo. Proporcionando información útil al usuario para la toma de decisión.

Una solución prometedora para el problema de predicción del rendimiento es el uso de agentes, dado la cantidad de análisis de factores que intervienen en el mismo. En este sentido el concepto de agente aplicado al desarrollo de sistemas de predicción logran una comunicación más rápida y segura entre los usuarios y el propio sistema.

3.2 Descripción del sistema.

El sistema se ha concebido para que forme parte de redes agroclimáticas en las que actividades vinculadas a la agricultura como son: planificación a largo plazo de sistemas agrícolas, utilización de las tierras, selección de los cultivos, determinación de épocas de siembra y de cosecha, control de plagas y enfermedades y programaciones de riego dependen del tiempo y del clima. Magnificando estas prácticas agronómicas mediante redes de información nos permitirá ser más eficientes y productivos.

Una red de estaciones agro-climáticas automáticas está diseñada para medir y registrar de modo automático y continuo diferentes variables climáticas y los requerimientos del cultivo. La optimización de la información radica en la unificación de los datos de toda la red, dependiendo del sistema de comunicación con el cual se recopilen los datos.

El sistema estará basado en la tecnología Cliente-Servidor, lo cual permite una mayor velocidad en el funcionamiento de la aplicación. Los clientes no tendrán necesidad de instalar ningún programa adicional, ni deberán contar con un procesador de alta tecnología para utilizar el sistema de manera óptima.

3.3 Procedimiento para la construcción de agentes.

Haciendo uso de métodos de inferencia borrosa contenido en la "*Toolbox Fuzzy*" de Matlab se logro obtener dos agentes, humedad relativa y temperatura. Partiendo de que las mediciones hechas por los sensores que fueron almacenados en archivos Excel.

Para el desarrollo de los agentes el primer aspecto a tener en cuenta es la naturaleza de los datos. Debido a que estos conjuntos de datos de agro meteorología provienen de diferentes fuentes y se han presentado problemas en el almacenamiento de los mismos, existen para varios atributos, valores ausentes y valores ruidosos. Estos valores con ruido pueden provocar que el algoritmo de aprendizaje obtenga soluciones erróneas, afectando la precisión de cada agente.

Los resultados obtenido mediante el análisis grafico y analítico pueden ser compartidos con lenguajes de alto nivel como Java, C++, C# entre otros, dando la posibilidad de integración de otros agentes mediantes varias técnicas de predicción.

3.4 Conclusiones parciales.

La construcción del sistema multiagente para el pronóstico de rendimiento no fue posible su implementación por el dominio de los datos, por la falta de rasgo significativos y de variable objetivo rendimiento. Se concibe por tanto las características generales que tendrá el sistema y como integrar nuevos agentes a los ya implementados Humedad y Temperatura.

CONCLUSIONES

Se diseña un sistema multiagente que permite el pronóstico de rendimiento de sistemas agrícolas diseñando e implementando algunos de los agentes principales tales como, a partir de estos se logra:

Un procedimiento general que establece como concebir los restantes agentes a partir de la experiencia presentada en los agentes humedad relativa y temperatura.

Se obtienen los pronostico de la variables humedad y temperatura a partir del diseño de un sistemas ANFIS el cual fue validado usando un método *Hold-Out*. Los resultados que se alcanza con un error cuadrático medio de 0.2135. Se emplearon dos métodos de *clustering Subtractive Clustering* alcanzándose mejores resultados *Fuzzy C-Means*, superando con un coeficiente de error al otro método empleado.

RECOMENDACIONES

- Desarrollar el sistema multiagente.
- Analizar como influye en la solución final los errores de prediccion de cada uno de los agentes.

REFERENCIAS BIBLIOGRÁFICAS

Abdullah, M. Z. G., L.C.; Lim, K.C. & Karim, A.A. (2004). "The applications of computer vision system and tomographic radar imaging for assessing physical properties of food." Journal of Food Engineering 61(1), 125--135.

Aitkenhead, M. J. D., I.A.; Mullins, C.E.; McDonald, A.J.S. & Strachan, N.J.C. (2003). "Weed and crop discrimination using image analysis and artificial intelligence methods." Computers and Electronics in Agriculture: 39(3), 157--171.

Bak, T. J., H. (2004). "Agricultural Robotic Platform with Four Wheel Steering for Weed Detection." Biosystems Engineering: 87(2), 125--136.

Benne, M. G.-P., B.; Chabriat, J.-. & Herve, P. (2000). "Artificial neural networks for modelling and predictive control of an industrial evaporation process." Journal of Food Engineering 46(4) 227--234.

Bergez, J.-D., J.-.; Lacroix, B.; Leroy, P. & Wallach, D. (2002). "Improving irrigation schedules by using a biophysical and a decisional model." European Journal of Agronomy: 16(2), 123-135.

Bergez, J.-G., F. & Lapasse, L. (2004). "A hierarchical partitioning method for optimizing irrigation strategies." Agricultural Systems: 80(3), 235--253.

Bezdek, J. (1981). "Pattern Recognition with Fuzzy Objective Function Algorithms." Plenum Press.

Bielza, C. B., P.; Rodriguez-Galiano, M.I. & Martin, J. (2003). "Logistic regression for simulating damage occurrence on a fruit grading line." Computers and Electronics in Agriculture: 39(2), 95-113.

Borgelt, C. a. R. K. (2003). Neural Networks, Working Group Neural Networks and Fuzzy Systems. Otto-von-Guericke. University of Magdeburg.

Bressan, G. M. (2007). Clasificación de riesgo de infeccion de regiones por plantas dañinas utilizando la Logica Fuzzy y Redes Bayesianas. Brasil, Universidad de Sau Paulo.

Brío, M. d. and Sanz (2002). "Redes neuronales y sistemas difusos. Segunda edición." 399.

Bruno, N. (1999). "Sistemas de Inferencia Difusos basados en Redes Neuroadaptativas Facultad de Informática." Universidad de Murcia. España.

Caballero, Y. (2005). Conjuntos Aproximados. Santa Clara, Universidad Central de las Villas.

Cabello, J. T., M. B. Castello (2005). "Redes neuronales artificiales en Medicina Intensiva. Ejemplo de aplicación con las variables del MPM II." Medicina Intensiva: **29**(1): 13 - 20.

Catalán, C. E. (2004). Series Temporales, Laboratorio de Estadistica.

Cordon, O. and F. Herrera (1997). "A tree stage evolutionary proces for learning descriptive and approximate Fuzzy-Logic controller knowledge Bases from examples." International Journal of Approsimate Reasoning: 17:369-407.

Chen, C. R. and H. S. Ramaswamy (2002). "Modeling and optimization of variable retort temperature (VRT) thermal processing using coupled neural networks and genetic algorithms." Journal of Food Engineering: 53(3), 209--220.

Chiu, S. L. (1994). "Fuzzy model identification based on cluster estimation." J. Intelligent Fuzzy Systems 2: 267-278.

Deniz Yilmaz, I. A. (2004). Kbs-selection of universal joint used in agricultural machinery, in '5th IFAC Workshop On Artificial Intelligence In Agriculture Cairo, Egypt.

Díaz, Y. a. A. F. (2003). CubaForecast para el pronóstico de variables meteorológicas. C. M. d. Cienfuegos. Cuba.

Diez, J. B., A.; Alonso, J.; Lopez, S.; del Coz, J.J.; Quevedo, J.R.; Ranilla, J.; Luaces, O.; Alvarez, I.; Royo, L.J. & Goyache, F. (2003). "Artificial intelligence techniques point out differences in classification performance between light and standard bovine carcasses." Meat Science: 64(3), 249--258.

Dunn, J. C. (1973). "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters." Journal of Cybernetics 3: 32-57.

Erenturk, K. E., S. & Tabil, L.G. (2004). "A comparative study for the estimation of dynamical drying behavior of Echinacea angustifolia: regression analysis and neural network." Computers and Electronics in Agriculture: 45(1-3), 71-90.

Fakherddine O. Karray, C. d. S., Ed. (2004). Soft Computing and Intelilgent Systems Design.

Farkas I., F. L. (2004). Machine Vision System For Water Shortage Indication Of Tomato Plants Using Local Orientation, in '5th IFAC Workshop On Artificial Intelligence In Agriculture. Cairo, Egypt.

Farkas, I. R., P. & Biro, A. (2000). "A neural network topology for modelling grain drying." Computers and Electronics in Agriculture: 26(2), 147--158.

García, N., M. Gámez (2006). RNA+SIG: Sistema automático de valoración de viviendas. . C.-. Campus Multidisciplinar en Percepción e Inteligencia por los 50 Años de la Inteligencia Artificial. Albacete, España.

Goel, P. K. P., S.O.; Patel, R.M.; Landry, J.A.; Bonnell, R.B. & Viau, A.A. (2003). "Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn." Computers and Electronics in Agriculture: 39(2), 67--93.

Gottschalk, K. N., L. & Farkas, I. (2003). "Improved climate control for potato stores by fuzzy controllers." Computers and Electronics in Agriculture: 40(1-3), 127--140.

Hamilton, J. D. (2005). Regime Switching Models. Palgrave Dictionary of Economics.

HAYKIN, S. (1994). Neural networks: A comprehensive foundation. Prentice-Hall, Englewood Cliffs, NJ.

Iman Hassen, M. R. (2004). Configuration Irrigation Schedule Based On Expert Systems And Operations Research, in '5th IFAC Workshop On Artificial Intelligence In Agriculture. Cairo, Egypt.

Kawamura, S. N., M.; Takekura, K. & Itoh, K. (2003). "Development of an automatic rice-quality inspection system." Computers and Electronics in Agriculture: 40(1-3), 115--126.

Kelly Christian and H. Robert (2001). Case-Based Reasoning.

Knight, R. E. K. (1994). Artificial Intelligence. Madrid, España.

Magdi. A. Koutb, N. M. E. (2004). Environmental Control For Plants Using Intelligent Control Systems, in '5th IFAC Workshop On Artificial Intelligence In Agriculture. Cairo, Egypt.

Männle, M. (1999). "Identifying Rule-Based TSK Fuzzy Models." 286.

Molinero, L. M. (2004). "Análisis de series temporales."

Moshou, D. W., S.; Strasser, R.; Schenk, A. & Ramon, H. (2003). "Apple mealiness detection using fluorescence and selforganising maps." Computers and Electronics in Agriculture: 40(1-3), 103--114.

Nesvadba, P. H., M.; Wolf, W.; Gekas, V.; Jarvis, D.; Sadd, P.A. & Johns, A.I. (2004). "Database of physical properties of agrofood materials." Journal of Food Engineering: 61(4), 497--503.

Ovando, G., M. Bocco (2005). "Redes neuronales para modelar predicción de heladas." Agricultura Técnica: **65**(1).

Pasgianos, G. D. A., K.G.; Polycarpou, P. & Sigrimis, N. (2003). "A nonlinear feedback technique for greenhouse environmental control." Computers and Electronics in Agriculture: 40(1-3), 153--177.

Rafea, A. H., H. & Hazman, M. (2003). "Automatic knowledge acquisition tool for irrigation and fertilization expert systems." Expert Systems with Applications 24(1), 49-57.

Rosemblatt, F., Ed. (1962). Principles of Neurodynamics. New York.

Roussel, S. B.-M., V.; Roger, J. & Grenier, P. (2003). "Authenticating white grape must variety with classification models based on aroma sensors, FT-IR and UV spectrometry." Journal of Food Engineering: 60(4), 407--419.

Rueda, R. E. (2006). "Modelos lluvia-escorrentía basados en redes neuronales artificiales." Ingeniería Civil:: 127-136.

Ruiz, R. (2006). Heurísticas de selección de atributos para datos de gran dimensionalidad. Sevilla, Universidad de Sevilla., Departamento de Lenguajes y Sistemas Informáicos

Samhaa R. El-Beltagy, G. A. H., H. (2004). Towards Specialized Expert System Building Tools: A Tool For Building Irrigation Expert Systems, in '5th IFAC Workshop On Artificial Intelligence In Agriculture. Cairo, Egypt.

SANTHANAM M., a. P., P. (2001). "Statistics of atmospheric correlations. Physical Review." E 64: 016102-1-1-7.

Shahin, M. A. T., E.W. & McClendon, R.W. (2001). "AE--Automation and Emerging Technologies: Artificial Intelligence Classifiers for sorting Apples based on Watercore." Journal of Agricultural Engineering Research: 79(3), 265--274.

Simon, H. A. (1983). Why should machines learn? Machine Learning, An Artificial Intelligence approach. R. S. M. J. G. C. T. M. Mitchel. Palo Alto, CA.

Tantau, H. L., D. (2003). "Greenhouse climate control: an approach for integrated pest management." Computers and Electronics in Agriculture: 40(1-3), 141--152.

WILKS, D. (1995). Statistical methods in Atmospheric Sciences. Academic Press, London.

Witten, I. a. E. F. (2005). Implementations: Real machine learning schemes. Data Mining. Practical Machine Learning Tools and Techniques. I. Witten and E. Frank, Department of Computer Science. University of Waikato:**:** 187-283.

ANEXOS

Anexo 1. Transformación de las mediciones.

```
%Este programa transforma las mediciones cada 15 min
%en promedios diario de las mismas.

clear all;
dianno = 0;
mes                                                                    =
char('enero','febrero','marzo','abril','mayo','junio','julio','agosto'
,'septiembre','octubre','noviembre','diciembre');

for j=1:12
[num, txt, raw] = xlsread(strcat(mes(j,1:10),'.xls'));

fecha = datenum(txt(:,1),'dd/mm/yyyy');
%Variables principales
VV = num(:,2);
DV = num(:,3);
TE = num(:,5);
HR = num(:,6);
RA = num(:,8);

k = size(fecha);
%Calcular la cantidad de dia por mes
cambio = fecha(1);
numero = 1;
for i=2:k
    if fecha(i) > cambio
        cambio = fecha(i);
        numero = numero + 1;
    end;
end;
dianno = dianno + numero;
%numero guarda la cantidad de dia del mes
cantidad = 0;
DVs = 0; TEs = 0; HRs = 0; RAs = 0; VVs = 0;
```

```
DVp = 0; TEp = 0; HRp = 0; RAp = 0; VVp = 0;
cambio = fecha(1);

for i=1:k
    if fecha(i) > cambio
        cambio = fecha(i);
        pro= TEs / cantidad;  TEp = [TEp pro];
        pro= HRs / cantidad;  HRp = [HRp pro];
        pro= DVs / cantidad;  DVp = [DVp pro];
        pro= RAs / cantidad;  RAp = [RAp pro];
        pro= VVs / cantidad;  VVp = [VVp pro];
        cantidad = 0;
        TEs = 0; HRs = 0; RAs = 0; DVs = 0; VVs = 0;
    else
        TEs = TEs + TE(i);
        HRs = HRs + HR(i);
        DVs = DVs + DV(i);
        RAs = RAs + RA(i);
        VVs = VVs + VV(i);
        cantidad = cantidad +1;
    end;
end;

dia = zeros(1,28);
for i=1:numero
    dia(i) = i;
end;

dia = dia'; TEp = TEp'; DVp =DVp' ; HRp = HRp'; RAp = RAp'; VVp =
VVp';
%Guarda los promedios de las mediciones de todo el mes
save(strcat(mes(j,1:10),'.mat'),'dia','HRp','DVp','TEp','RAp','VVp');
end;
```

Anexo 2. Selección de Muestras de Entrenamiento y Prueba

```
%Esta función crea las muestra de entrenamiento y control con las
%variables de entradas seleccionadas.
```

```
function [chk_data, trn_data] = preparadatos(varp)

mes                                                                    =
char('enero','febrero','marzo','abril','mayo','junio','julio','agosto'
,'septiembre','octubre','noviembre','diciembre');

for i=1:12
    mest = load_mes(strcat(mes(i,1:10)),varp);
    if i == 1
        ano = [mest(2:size(mest),:)];
    else
     ano = [ano ; mest(2:size(mest),:)];
    end
end

%ano tiene la siguiente estructura:
% columna 1: un indice del dia que corresponde al dia del año, desde
el uno al 352
% columna 2,3,4: son los parametros

%datos auxialares
tama= size(ano);
diaano = (1:tama(1,1))';
HRano = ano(:,4);RAano = ano(:,3);TEano = ano(:,2);

trn_data = zeros(251, 11);
chk_data = zeros(92,11);



% juego 1: HR(k) = F( HR(k-1),HR(k-2),RA(k),RA(k-1),TE(k))
chk_data                                                               =
[ano(252:343,2),ano(251:342,3),ano(252:343,3),ano(251:342,4),ano(252:3
43,4),ano(253:344,4)];
trn_data                                                               =
[ano(3:253,2),ano(2:252,3),ano(3:253,3),ano(2:252,4),ano(3:253,4),ano(
4:254,4)];
```

```
save datosano ano;
save datosEC trn_data chk_data;

% clear ano TEp RAp HRp
```

Anexo 3. Resultado del FIS obtenido por diferentes técnicas de agrupamiento.

Subtractive Clustering

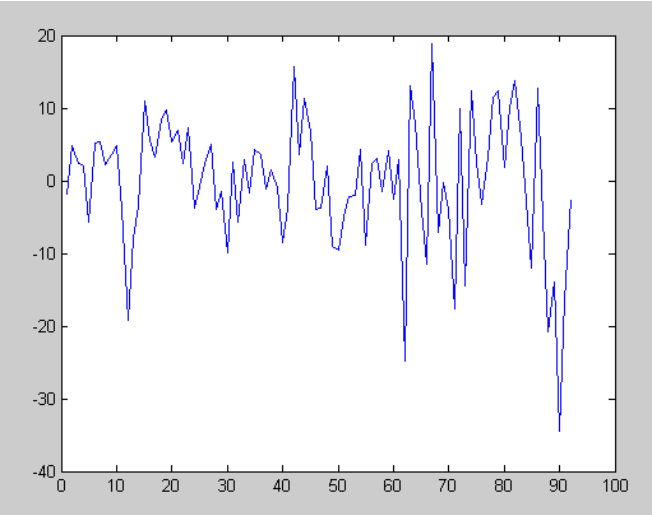

Figura 1. Coeficiente de error de predicción.

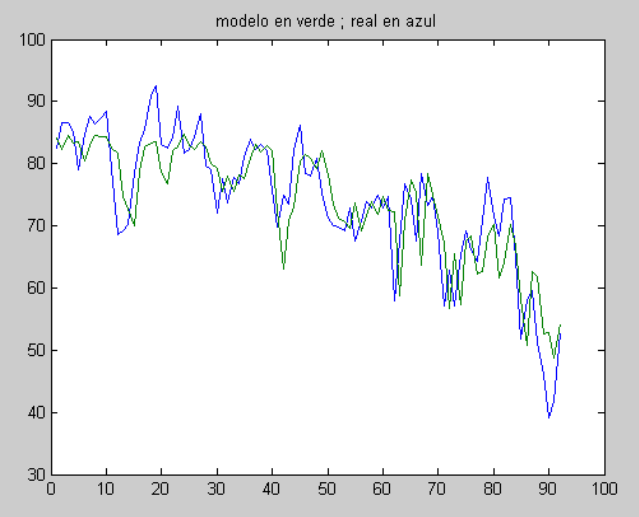

Figura 2. Resultado al evaluar el FIS.

Fuzzy C-Means

Figura 3. Resultado al evaluar el FIS.

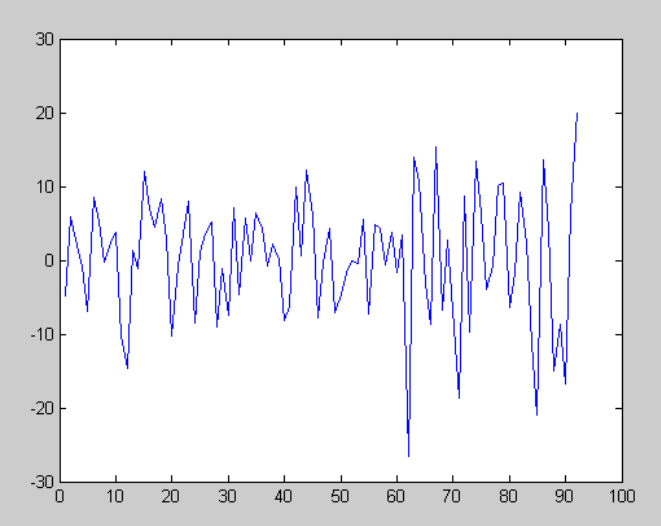

Figura 4. Coeficiente de error de predicción.

Anexo 4. Programa de validación usando la muestra de control.

```
chkdatin = chk_data(:,1:5);
chkdatout = chk_data(:,6);
figure(3);
chkfuzout=evalfis(chkdatin,a);
plot(chkdatout,chkfuzout,'bx',[0 1],[0 1],'r:')
axis square
xlabel('Actual Value')
ylabel('Predicted Value')

figure(4);
plot([chkdatout chkfuzout]);
title('modelo en verde ; real en azul');
error = 100*(chkdatout-chkfuzout)./chkdatout;
figure(5);
plot(error);
```