

Red neuro-difusa para el relleno de datos faltantes en la estación meteorológica Chapingo

Neuro-fuzzy para o preenchimento de rede de dados em falta estação de Chapingo tempo

Juan Daniel Peña Durán

Centro Universitario UAEM Texcoco
texsmallville@hotmail.com

Irene Aguilar Juárez

Centro Universitario UAEM Texcoco
ireneico@gmail.com

Joel Ayala de la Vega

Centro Universitario UAEM Texcoco
joelayala2001@yahoo.com.mx

Resumen

Esta investigación presenta la aplicación de un modelo de red neurodifusa llamado ANFIS para el problema de estimación de datos faltantes meteorológicos: temperatura, velocidad del viento, humedad relativa y radiación solar en la estimación de la Evapotranspiración de referencia ET_o. ANFIS es un método que permite crear la base de reglas de un sistema difuso, utilizando el algoritmo de retro propagación a partir de los datos de un proceso. La estructura de la red neuro-difusa para cada variable meteorológica consiste en dos entradas y una salida. La evaluación del relleno de datos faltantes se realiza mediante la Raíz Cuadrada del Error Cuadrático Medio (RMSE). Los resultados muestran que al usar un mayor número de iteraciones y variación de datos en el entrenamiento puede ayudar a la ANFIS a obtener resultados más precisos.

Palabras clave: ANFIS, datos faltantes, evapotranspiración de referencia.

Resumo

Esta pesquisa apresenta a aplicação de um modelo chamado rede neuro ANFIS para o problema da falta de tempo de estimativa de dados: temperatura, velocidade do vento, umidade relativa e

radiação solar na estimativa da evapotranspiração de referência ETo. ANFIS é um método para criar a base de regras de um sistema fuzzy utilizando o algoritmo de propagação retro com base nos dados de um processo. A estrutura da variável meteorológica neuro-fuzzy para cada rede consiste em duas entradas e uma saída. A avaliação do preenchimento dos dados em falta é feito por erro Square Root Mean Square (RMSE). Os resultados mostram que o uso de um maior número de iterações e dados de alteração na formação do ANFIS pode ajudá-lo a obter resultados mais precisos.

Palavras-chave: ANFIS, dados em falta, de evapotranspiração de referência.

Fecha recepción: Agosto 2014

Fecha aceptación: Diciembre 2014

Introdução

A Comissão Nacional da Água (CONAGUA), através do Serviço Nacional de Meteorologia é a fonte oficial de dados de tempo e clima no México. Monitoramento e gravação desses dados é realizada por meio de estações meteorológicas automáticas (EMAS), que são distribuídos em torno de toda a República Mexicana. No entanto, informações de registros meteorológicos são muitas vezes incompletos devido a vários fatores, incluindo: mau funcionamento e calibração instrumentais na manutenção da estação e seus instrumentos. Por exemplo, a estação meteorológica Chapingo, administrado pela Agência Cuenca Aguas del Valle de Mexico (OCAVM, localizado no município de Texcoco, Estado do México, com as coordenadas geográficas de latitude norte $19^{\circ} 50'$ de longitude oeste e $98^{\circ} 88'$) está monitorando diferentes variáveis meteorológicas, mas ocasionalmente tem falta de registros de dados. Isso afeta a precisão dos resultados ao fazer cálculos importantes, tais como a estimativa de necessidades de água das culturas em áreas irrigadas. No entanto, para satisfazer as necessidades de água das culturas é necessário calcular a perda de água causada pela evaporação e transpiração das culturas, por esta razão a Organização Mundial das Nações Unidas para a Alimentação ea Agricultura (FAO) em seu Guia as necessidades hídricas das culturas (Rivera, 2008) introduz o conceito de cultura da evapotranspiração de referência (ET), que estuda a taxa de evapotranspiração, independentemente do tipo de colheita e características do solo (Doorenbos & Pruitt, 1997).

Dada a sua definição, os fatores que afetam ETo são fatores climáticos podem ser calculados parâmetros climáticos como temperatura, velocidade do vento, umidade relativa e radiação solar. Estes dados são fornecidos pela EMA, mas, como mencionado acima sofrem de perda de dados, cujo efeito pode ser insignificante, mas quando o tempo aumenta com o banco de dados de acompanhamento a perda de dados torna-se pouco fiável. Para resolver o problema da falta de registros, a literatura sugere que o uso de diferentes técnicas, do tradicional, tal como a regressão linear, chamadas redes neurais artificiais; mas determinar qual modelo é mais eficiente, alguns estudos não encontraram nenhuma diferença entre os resultados. Por sua vez, tendem a apoiar os outros com ligeira superioridade de redes neurais artificiais (Pitarque & Roy, 1998). Assim, a implementação de uma técnica deve ser considerada a situação do meio ambiente no qual ele vai desenvolver o estudo para se obter os resultados desejados.

Vale ressaltar que a Estação Meteorológica Automática Chapingo carece de um registro de dados de história período de uma hora em anos anteriores, os dados da estação pode ter diferentes porcentagens de falta de dados; Você não tem qualquer outro parâmetro que se relaciona diretamente com os dados medidos. Além disso, você não pode extrapolar dados de uma estação meteorológica, uma vez que são piores, limitando as opções de solução. Portanto, considera-se a usar a técnica da lógica fuzzy, porque você pode trabalhar com todos os registros completos e incompletos, e criar diferentes cenários de acordo com o comportamento de cada variável meteorológica durante o ano e durante o dia. Isso nos permite formular as regras de inferência e conjuntos fuzzy para cada variável. O objetivo deste trabalho é projetar e implementar o modelo ANFIS para um modelo aproximado do comportamento das variáveis meteorológicas dos dados gravados ou obtidos a partir da Estação Meteorológica Chapingo, e estimar os dados faltantes.

O trabalho está organizado da seguinte forma: a seção onde a história relacionada com a recuperação de dados em falta são trabalhos de revista; neurofuzzy seção Foco, esta seção elementos e relações que compõem a estrutura de ANFIS show de rede; a proposta através da seção ANFIS Modelo apresenta a modelagem de conjuntos fuzzy, o intervalo de valores e funções da nossa rede neuro; Seção de experiência e resultados mostra a implementação da rede neuro no cálculo da evapotranspiração e estima RSME; Finalmente, as Conclusões e Trabalhos Futuros seção mostra uma visão geral dos benefícios do modelo ANFIS em estimativa satisfatória de dados em falta.

Fundo

Na literatura tem havido abordagens diferentes para o problema da falta de dados de enchimento, dependendo da variável a ser completada em registros; pesquisa (Campos, Quispe, e Tatiana, 2012) metodologia de gerenciamento para o trabalho em dados de precipitação meteorológicos e da aplicação do método de preenchimento da equação inversa da distância euclidiana com a matriz de correlação é proposta se É que os espaços são apresentados. Os resultados foram satisfatórios com a equação da inversa da distância Euclidiana com a matriz de correlação, o que permite uma variação de ambas as médias e as variações que ocorrem durante o período de tempo manuseado. No entanto, o estudo conclui que, durante a selecção de estações devem ser postos de selecção cuidadosas com coeficientes de correlação superiores a 0,75 por cento e inferior a 20% vazio, o que garante que os dados preenchidos fiáveis uma vez que são percentagens de variação varia de 0 a 15%. Neste escrita (Ferreira, 2003) foram estudados diversos métodos de análise e imputação de dados incompletos, a partir do ponto de vista da sua aplicação para completar valores em falta em série da velocidade do vento; seus resultados são interessantes para a velocidade do vento. No trabalho descrito (R. Alfaro, Alfaro & Pacheco, 2000) métodos diferentes para preencher as lacunas na série de precipitação anual aplicada aos registros das estações meteorológicas localizadas em diferentes regiões da Costa Rica é. Preenchendo os métodos usados para jogar uma série de dados que eles estimados usando pelo menos uma estação perto da estação em estudo são: regressão simples, a razão pela qual o conjunto, de regressão múltipla e razão normal. Os resultados do estudo mostram diferenças máximas absolutas entre os valores reais e estimados de cerca de 30%, o que sugere o uso de mais complexo dos apresentados neste estudo se você quiser fazer estimativas mais precisas dos métodos de dados precipitação anual. Referência (Valesani & Quintana, 2009) diz respeito à aplicação de Redes Neurais Artificiais (RNA) como método de imputação simulando ausência de taxas de dados usando o MCR (perdendo completamente a Random) técnica. A sua eficácia foi então avaliada em situações diferentes, a fim de avaliar o desempenho com parâmetros diferentes, tais como MAE (o erro médio absoluto), RMSE (Mean Square Error), e regressão, a fim de determinar se o ARN são adequados para a imputação de dados neste caso particular. Os resultados mostrados na pesquisa são satisfatórios e considerada aceitável em termos estatísticos. Da mesma forma, a proposta (Solana & Boat, 1998) destaca as Redes Neurais Artificiais citam

conceitos técnicos, bem como aplicações específicas no domínio da recuperação de informação. A proposta (Cruz, 2012) uma metodologia para o preenchimento de dados faltantes meteorológicos desenvolvidos no ambiente de programação Matlab usando a técnica de interpolação Piecewise Cubic Hermite polinómio de interpolação (PCHIP) com software é estabelecida. Para avaliar o desempenho de interpolação raiz quadrada média do erro (RMSE) é usado. Note-se que os resultados sobre o preenchimento de dados em falta não são apresentados neste trabalho. Para resolver o problema da falta de dados meteorológicos têm sido usadas algumas técnicas tradicionais, como a regressão, a homogeneidade das estações próximas e utilizar as mais recentes técnicas de Inteligência Artificial, baseado em lógica fuzzy (aprendizagem indutiva), algoritmos genéticos e redes neurais (Saba & Ortega, 2008) (Alfaro & Soley, 2009), (Chen, 1995); no entanto, o problema surge quando encontramos resultados conflitantes quando se determina quais modelos são mais eficientes para resolver o problema da falta de dados.

Este trabalho de pesquisa tem considerado o uso como uma técnica para preencher os dados em falta na estação meteorológica automática Chapingo para redes neuro-fuzzy. Assim, o estudo da inteligência artificial em sistemas não-lineares complexas é reforçada.

Neurofuzzy Foco

As redes são sistemas neurodifusas que tiram proveito de recursos de redes neurais como a capacidade de aprender ou ajustar-se e generalizar, somada às características da lógica fuzzy, que trabalha com o raciocínio lógico baseado em funções de pertinência que permitem trabalhar com variáveis lingüísticas, muito natural para os seres humanos. Os sistemas de inferência fuzzy pode representar o conhecimento baseado em if-then regras, mas não tem a capacidade de se adaptar quando as condições externas mudança. Por esta razão, o conceito de redes neurais de reconhecimento estão incluídos.

Com base nos sistemas de neuro-fuzzy eles estão em neurónios difusas, com base emular morfologia neuronal biológica, seguido por um sistema de aprendizagem característica mais difusa. Nós podemos classificar neurónios difundem em duas classes: em primeiro lugar, a característica fuzzy é encontrado na descrição dos pesos sinápticos, e os segundos sinais são transmitidos em conjunto com os pesos sinápticos difusas.

Devemos levar em consideração as ações muito importantes para o desenvolvimento de sistemas de difusão neurais: a) definição dos valores de entrada e de saída, b) definição de conjuntos fuzzy que são necessárias para usar, c) a definição de regras difusas, d) estruturação a rede e e) modelagem neural de conexões sinápticas que pode incorporar interpretação difusa.

Nas últimas décadas, os sistemas de difusão neurais foram posicionados em aplicações importantes em diferentes áreas, tais como: controle (na maioria dos sistemas); análise quantitativa (operações, gerenciamento de dados); inferência (sistemas especialistas para diagnóstico, planejamento e previsão, processamento de linguagem natural, robótica, engenharia de software) e recuperação de informação (base de dados), entre outras aplicações (Lin, Lee, e CS, 1996).

Portanto, para este estudo foi decidido para modelar a dinâmica do processo através do primeiro sistema neuro conhecido e mais estabelecido, ANFIS; para além de ser um dos trabalhos pioneiros, é um dos mais simples computacionalmente (Jang., 1993). ANFIS implementa o modelo Takagi-Sugeno para a estrutura de if-then regras do sistema difuso. Um modelo ANFIS é composto de cinco camadas, em que todos os nós de uma mesma camada tem uma função semelhante. A primeira camada é usado para as entradas. A última camada de saída e tem 3 camadas intermediárias ocultos. O número de camadas ocultas permanece constante em todos os tipos de ANFIS de implementar, independentemente das entradas que têm o sistema e tem apenas uma saída. Na Figura 1, as cinco camadas de nós de rede ANFIS e a relação entre as variáveis de entrada, regras linguísticas, nodos regras normalizadas e os parâmetros são apresentados, em seguida, a função da camada de rede por camadas é explicado.

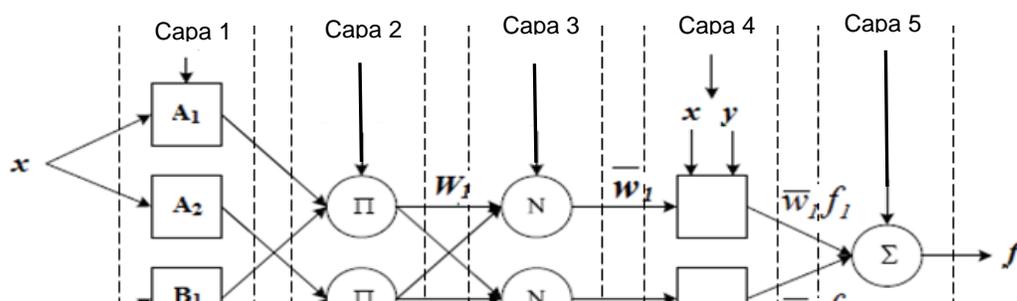


Figura 1. Arquitectura del sistema ANFIS [16].

O comportamento de cada camada ANFIS descrito abaixo:

Camada 1: Cada nó i desta camada é adaptável, isto é, tem os parâmetros ajustáveis e descrito pela equação 1.

$$O_i^1 = \mu_{A_i}(x) \quad (1)$$

Onde X é o nó de entrada i , A_i é uma variável linguística associada com a função deste nó. Em outras palavras, O_i^1 é a função de membro de A_i e especifica o grau de pertinência de X para A_i .

Camada 2: Cada nó nesta camada é marcada com Π (Figura 1). Nesta camada de sinais de entrada são multiplicados e o produto à saída, isto é, quando vários sinais de entrada são para este nó como um resultado enviar o produto para cada exemplo I é enviado. Por um momento, $O_i^2 = \mu_{A_i}(x) \mu_{B_i}(y)$, $i=1,2$ (2)

Cada nó de saída representa o grau de activação de uma regra. Eles também representam o T-regra ou T-conorm para modelar a lógica AND e OR operações. São muitas vezes referidos como nós de regras.

Camada 3: Cada nó N rotulado (Figura 1), indicando a normalização do grau de activação. O nó N calcula as regras normais de activação com a soma de todas as regras ativadas de acordo com a equação 3. As saídas desta camada pode ser chamado de regras de activação padrão.

$$O_i^3 = \bar{w}_i = w_i / (w_1 + w_2) \quad (3)$$

Camada 4: Cada nó i nesta camada é quadrado e tem um nó de recurso:

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i), \quad i=1,2 \quad (4)$$

Onde w_i é a camada de saída 3 e o conjunto de parâmetros $\{p_i, q_i, r_i\}$ são referidos os parâmetros em conformidade.

Camada 5: circular apresenta um único nó marcado Σ , Aqui, a saída é calculada a partir dos sinais de entrada (Equação 5):

$$O_i^5 = f = \sum \bar{w}_i f_i = \sum_i \bar{w}_i f_i / \sum_i \bar{w}_i = w_1 f_1 + w_2 f_2 \quad (5)$$

O processo de treinamento é realizado com dois conjuntos de parâmetros: o antecedente (caracterizadores constantes as funções de pertinência) eo resultado (coeficientes de funções lineares da consequente das regras). As ligações entre os nós apenas indicam a direção em que sinaliza pesos fluxo, já associados (F & Garcia Villada).

Modelo proposto

Para implementar foram utilizados os ANFIS conjuntos de dados climáticos reais, que começa a partir da estação meteorológica Chapingo. As variáveis meteorológicas foram registradas a cada 10 minutos durante os períodos diários de setembro 2013 ao dia atual. Variáveis modelados são a velocidade do vento (VELS), temperatura (TEMP), umidade relativa (UR) ea radiação solar (RAD-SOL). Na neuro variáveis do modelo são atribuídos H (Time) e da Estação (EA) como variáveis de entrada e variáveis meteorológicas velocidade do vento, temperatura, umidade relativa e radiação solar como variáveis de saída. Variáveis de tempo e são dependentes Estação do ano, enquanto que a velocidade do vento, temperatura, umidade relativa e radiação solar são dependentes dos dois primeiros. Como valores lingüísticos de H (horas) são propostas: de manhã cedo (M1), Tomorrow (M2), meio-dia (M3), Sunset (A1), Anoitecer (A2) e Noite (N). Conjunto fuzzy para hora do dia indicado na secção A da Figura 2.

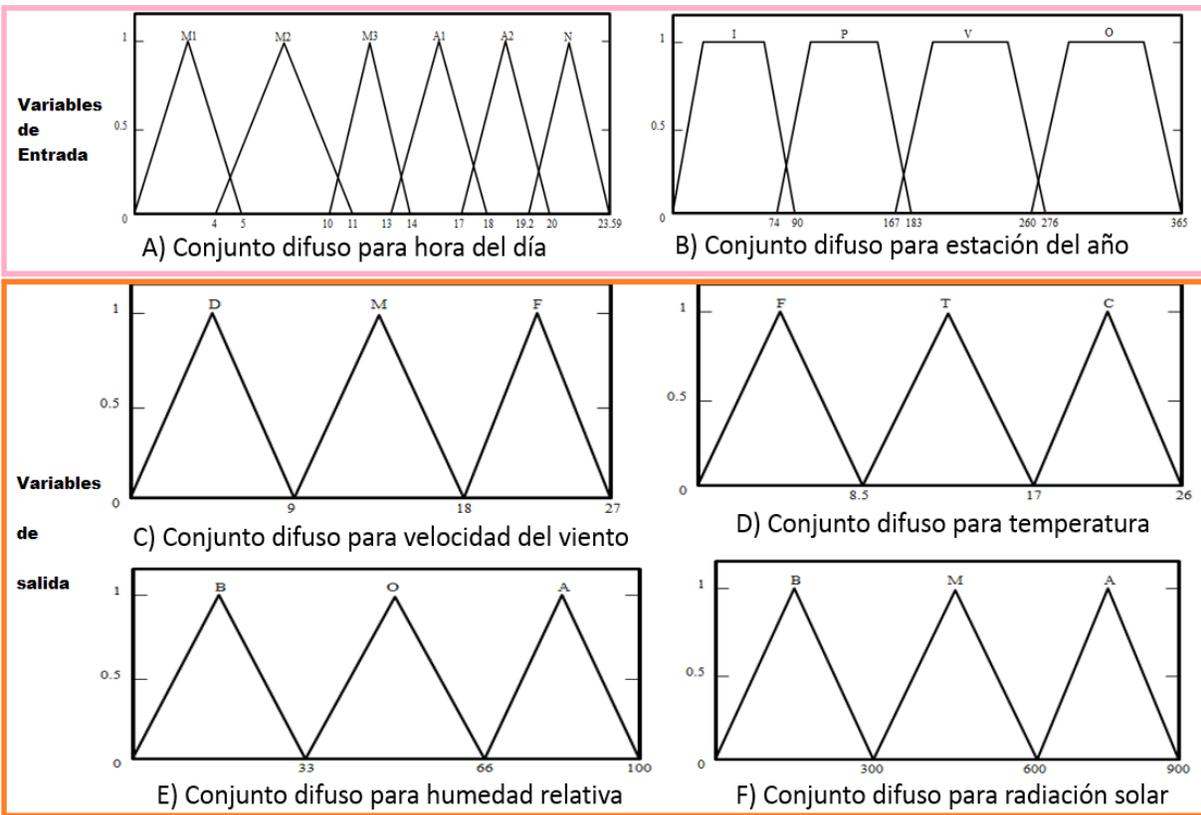


Figura 2: Conjuntos fuzzy para: a) hora) do dia, b) temporada, c) a velocidade do vento, d) temperatura, e) e umidade relativa f) a radiação solar

Para não incorrer em um mínimo estados mal definidas de reticulação é utilizado, considerando-se o comportamento das variáveis meteorológicas durante o dia, de modo que cada elemento permanece representado em pelo menos duas funções de pertinência. A função triangular para especificar a altura de cada etiqueta é utilizado linguística.

Por causa da dificuldade de estabelecer uma escala uniforme no comprimento dos meses do ano, foi decidido distribuir os dias do ano, com base nas temporadas. Assim, neste estudo, o primeiro do ano foi atribuído ao primeiro dia do inverno, que é em 21 de dezembro e 90 é o dia 21 de março último dia de inverno.

Como conjuntos fuzzy e linguística rótulo variável proposta temporada: Inverno (I), Primavera (P), Verão (V) e Outono (O). A função de membro da temporada variável mostrada na seção B da Figura 2. A ligação cruzada foi realizada considerando o comportamento do tempo, porque em a mudança das estações do ano, nos dias de início de uma variável eo último os outros dias do comportamento variável é semelhante. Além disso, a função trapezoidal foi utilizado como nas

estações existe um intervalo em que o comportamento das variáveis é estável e mantida por um período.

Os intervalos para as quatro variáveis lingüísticas são: velocidade do vento [0, 27] km / h, temperatura [0, 26] ° C, umidade relativa [0, 100] e radiação solar em% [0, 900] em wm2. Os intervalos são valores médios obtidos em cada época. Conjuntos fuzzy com os respectivos rótulos lingüísticos que compõem cada uma das variáveis foram: velocidade do vento [Weak (D), moderado (M) e Strong (F)] ver subseção C da Figura 2; para temperatura [Frio (F), moderado (t), morna (C)] ver subseção D da Figura 2; de umidade relativa [Baixa (B), moderado (M) e Alta (A)] ver subseção E na Figura 2 e radiação solar [Baixa (B), moderado (M) e Alta (A)] ver subseção F Figura 2.

A forma das regras utilizadas para a velocidade do vento é a seguinte: se a estação é X e depois VELs Hora e é z. Ver Tabela I.

Tabla I. Reglas de inferencia para velocidad del viento.

# Regla	Estación	Hora	VELS	# Regla	Estación	Hora	VELS
1	I	M1	D	14	O	M1	D
2	I	M1	M	15	O	M1	M
3	I	M2	D	16	O	M2	D
4	I	M2	M	17	O	M2	M
5	I	M3	D	18	O	M3	D
6	I	M3	M	19	O	M3	M
7	I	A1	D	20	O	A1	D
8	I	A1	M	21	O	A1	M
9	I	A2	D	22	O	A2	D
10	I	A2	M	23	O	A2	M
11	I	N	D	24	O	N	D
12	I	N	M	25	O	N	M
13	I	N	F	26	O	N	F

Para temperatura é a forma Se a estação é X e depois TEMP é Tempo e Z é tal como mostrado na Tabela II.

Tabla II. Reglas de inferencia para Temperatura.

# Regla	Estación	Hora	TEMP	# Regla	Estación	Hora	TEMP
1	I	M1	F	15	O	M1	T
2	I	M1	T	16	O	M1	C
3	I	M1	C	17	O	M2	T
4	I	M2	F	18	O	M2	F
5	I	M2	T	19	O	M3	T
6	I	M3	F	20	O	M3	F
7	I	M3	T	21	O	A1	F
8	I	A1	F	22	O	A1	T
9	I	A1	T	23	O	A1	C
10	I	A1	C	24	O	A2	C
11	I	A2	T	25	O	N	C
12	I	A2	C	26	O	N	T
13	I	N	T				
14	I	N	F				

Da mesma forma, para a humidade relativa é como se segue: IF Estación es x AND Hora es y THEN HR es z. Ver tabla III.

Tabla III. Reglas de inferencia para Humedad Relativa

# Regla	Estación	Hora	HR	# Regla	Estación	Hora	HR
1	I	M1	B	16	O	M1	O
2	I	M1	O	17	O	M1	A
3	I	M1	A	18	O	M2	O
4	I	M2	B	19	O	M2	A
5	I	M2	O	20	O	M3	A
6	I	M2	A	21	O	A1	O
7	I	M3	O	22	O	A1	A
8	I	M3	A	23	O	A2	B
9	I	A1	B	24	O	A2	O
10	I	A1	O	25	O	A2	A
11	I	A1	A	26	O	N	A
12	I	A2	B	27	O	N	O
13	I	A2	O	28	O	N	B
14	I	N	B				
15	I	N	O				

Por fim, a radiação solar é como se segue: IF Estación es x AND Hora es y THEN RAD-SOL es z. Ver tabla IV.

Tabla IV. Reglas de inferencia para Radiación Solar

# Regla	Estación	Hora	RAD-SOL	# Regla	Estación	Hora	RAD-SOL
1	I	M1	B	12	O	M1	B
2	I	M2	B	13	O	M1	B
3	I	M3	B	14	O	M3	B
4	I	A1	B	15	O	A1	B
5	I	A1	M	16	O	A1	M
6	I	A1	A	17	O	A1	A
7	I	A2	M	18	O	A2	B
8	I	A2	A	19	O	A2	M
9	I	N	B	20	O	A2	A
10	I	N	M	21	O	N	B
11	I	N	A	22	O	N	M
				23	O	N	A

As regras são armazenadas no motor de inferência também conhecida como um centro de controle, pois nela são ordens que devem operar. Para modelar ANFIS usou o Toolbox distorcido Matlab. Selecione os valores que serão processadas no momento entradas e estação do ano, passando o fuzzifier, isso se traduz em uma linguagem compreensível para o motor de inferência, que associa a cada entrada para um conjunto difuso de voltar como uma saída de um valor numérico e não difusa.

Experiência e resultados

Para as funções de pertinência dos conjuntos fuzzy de velocidade do vento, temperatura, umidade relativa e radiação solar procedeu à recolha de dados diários sobre as variáveis acima em um período dado pela CONAGUA, através do Serviço Nacional de Meteorologia e da utilização de indicadores de tempo da área de estudo onde o clima predominante é o semi-seco temperado. Esta média anual também produziu de cada um dos valores de variáveis para gerar os intervalos de cada conjunto difuso, e, assim, descartar valores extremos.

AME realizar a coleta e monitoramento de variáveis meteorológicas para gerar arquivos a cada 10 minutos em média para todas as variáveis, essa informação é enviada via satélite em intervalos de 1-3 horas por temporada. Os dados de estação são apresentados em três formatos de acordo com as estações de recolha de dados de frequência. O relatório corresponde a uma hora e 10 minutos período, o período de 24 horas 60 minutos relatório de 90 dias é o período de 24 horas correspondente, que recebem 144 registra cada variável meteorológica.

Para a pesquisa é considerada resolver o problema da falta de relatórios em uma hora de dados meteorológicos, uma vez que são eles que têm maior ausência de dados, e, em seguida, para calcular a ETo para períodos de vezes.

A formação da rede neuro foi realizada para cada uma das variáveis de tempo, que consistem de um mínimo de vezes (número de iterações) para processar as amostras seleccionadas. Estas amostras variam em tamanho de acordo com o número de regras para cada variável de tempo, que foram mencionados na secção anterior, bem como da estação em questão. A avaliação de cada uma das redes são baseadas em dados de um dia para a temporada de inverno (8) correspondente a 28 de dezembro e para o Outono (dia 279), correspondendo a 25 de setembro. Os dados para a formação foram seleccionados de acordo com a formulação de cada regra avaliada durante o dia.

Tabelas V e VI mostrar algumas peculiaridades que estavam assistindo o treinamento da rede neuro. Os dados no decorrer do dia são 144 por variável meteorológica; no entanto, este número pode variar de acordo com as falhas da estação; regras difusas como mencionadas acima variam no que diz respeito ao curso de cada época. Os dados totais de formação para obter o número de vezes pelo número de regras nebulosas para cada variável meteorológica. Os elementos de formação corresponde a essas duplicatas ou de enchimento antes ou após o dia que nos permitiu satisfazer as regras fuzzy em registros avaliados. Os dados reais indicam o número de dados precisos que satisfaz a formação da rede neuro no número correspondente de iterações, também pode ser visto que não estão uniformemente distribuídos no interior de cada conjunto fuzzy da variável tempo.

Tabla V. Entrenamiento de la red neurodifusa con 5 épocas

5 Épocas								
Variable Meteorológica	Día 279 Otoño				Día 8 Invierno			
	VELS	TEMP	HR	RAD-SOL	VELS	TEMP	HR	RAD-SOL
Datos en el transcurso del día	144	144	144	144	144	144	144	144
Reglas Difusas	13	12	13	12	13	14	15	11
Elementos en el Entrenamiento (datos faltantes rellenados por duplicidad o relleno de datos anterior o posterior)	17	27	25	18	25	35	45	15
Datos reales (satisficieron las reglas)	48	33	39	42	38	35	30	40
Total de Datos de entrenamiento (época * n de reglas)	65	60	65	60	65	70	75	55

Tabla VI. Entrenamiento de la red neurodifusa con 10 Épocas

10 Épocas								
Variable Meteorológica	Día 279 Otoño				Día 8 Invierno			
	VELS	TEMP	HR	RAD-SOL	VELS	TEMP	HR	RAD-SOL
Datos en el transcurso del día	144	144	144	144	144	144	144	144
Reglas Difusas	13	12	13	12	13	14	15	11
Elementos en el Entrenamiento (datos faltantes rellenados con relleno de datos anterior o posterior)	47	55	61	51	60	72	92	39
Datos reales (satisficieron las reglas)	83	65	69	69	70	68	58	71
Total de Datos de entrenamiento (época * n de reglas)	130	120	130	120	130	140	150	110

Como mostrado nas tabelas acima, o número de elementos de formação não se compara com os dados totais de formação em ambas as tabelas, isto significa que, embora o número de dados de formação é aumentada, uma fração correspondente a dados reais . Da mesma forma, ainda que passou de 5-10 vezes os dados reais na tabela V não VI correspondente a 10 vezes o dobro na tabela, isto devido a dois fatores principais: o dia não satisfaz plenamente as regras difusas em adição à gama compreendendo cada conjunto tempo variável nebulosa (alvorecer, manhã, meio-dia, sol, Anoitecer e Noite) é diferente, resultando numa alteração dos dados reais sobre a formação em cada iteração. Portanto, enchimentos de espaço que não são cobertos em cada iteração do treinamento são feitas levando-se em consideração os dados que estão mais perto

avaliadas diariamente, que podem ser dias mais tarde ou mais cedo, conforme o caso, ou igualmente um fato de repetição em uma iteração.

Portanto, o treinamento e validação da rede neuro levou 279 dias para a temporada de outono e dia 8 para a temporada de inverno. O desempenho de cada rede foi feito com a raiz quadrada do erro médio quadrático (RMSE, ver Tabela VII) calculada pela seguinte equação 6.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - x_i)^2} \tag{6}$$

Onde n é o número de observações consideradas, XI é o valor real e o valor de YI é estimado pelo modelo. Tabela VII pode observar os dados obtidos com o treinamento da rede neuro por 5 a 10 vezes.

Tabla VII: resultados obtenidos con la red neurodifusa

	VELS	TEMP	HR	RAD-SOL
Datos evaluados en el transcurso del día	144	144	144	144
Rango de Hora de día	0:00-23:50	0:00-23:50	0:00-23:50	0:00-23:50
Valor para estación del año	otoño	otoño	otoño	otoño
Rango de valores para la variable	0-27 Km/h	0-26 ° C	0-100 %	0-900Wm ²
Día 279 Otoño				
RMSE con 5 épocas de entrenamiento	2.06	0.95	1.97	381.61
RMSE con 10 épocas de entrenamiento	2.05	0.53	1.89	92.53
Diferencia en precisión	0.01	0.42	0.08	289.08
Día 8 Invierno				
RMSE con 5 épocas de entrenamiento	2.37	2.98	3.60	46.48
RMSE con 10 épocas de entrenamiento	1.43	0.88	2.35	44.84
Diferencia en precisión	0.94	2.1	1.25	1.64

De acordo com os resultados obtidos para as variáveis de tempo melhoradas mostrados para aumentar o número de vezes com 10 vezes o erro é reduzido RMSE. Em relação ao comportamento das variáveis, você pode ver claramente uma melhoria no ajuste dos cerca de 10 vezes em relação aos dados reais como as linhas no gráfico mostram várias secções de dados coincidência. Veja a Figura 3.

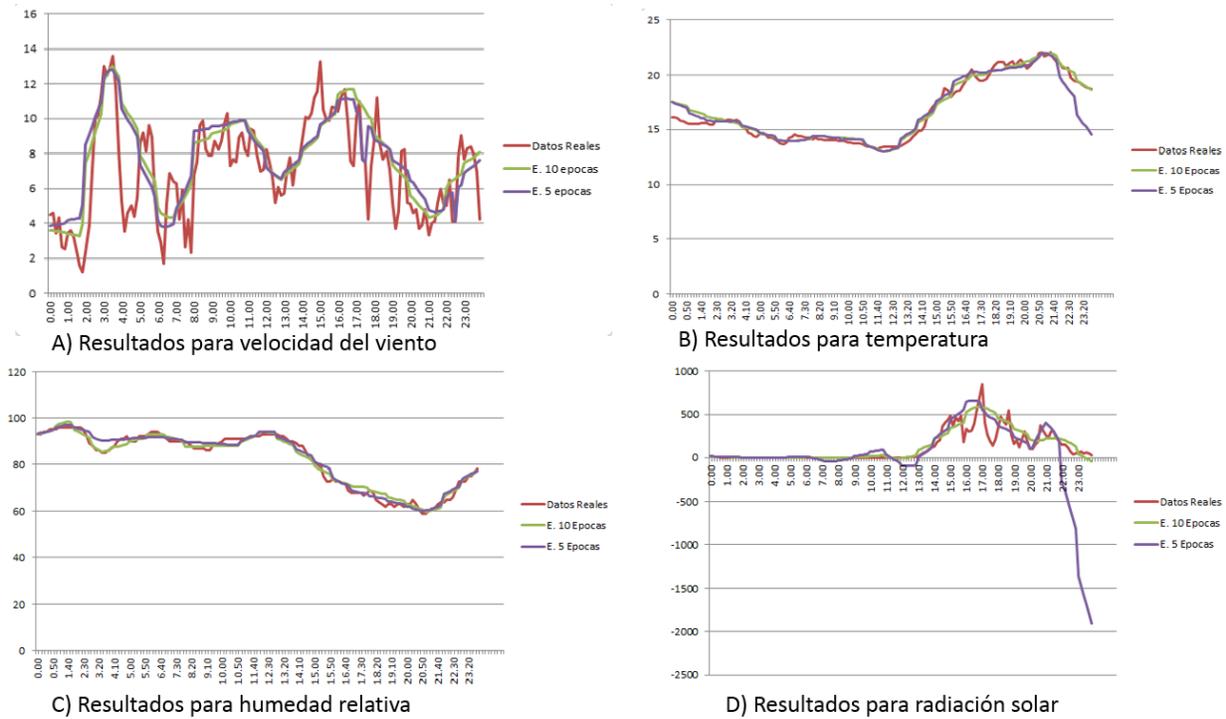


Figura 3: Resultados para: A) velocidad del viento, B) temperatura, C) humedad relativa, D) radiación solar

Para avaliar a utilidade da rede neuro na estimativa da evapotranspiração de referência passou a sua estimativa seguindo os procedimentos para períodos de vezes vezes apontando Allen (2006), que explica que na implementação da FAO Penman-Monteith por períodos horas de tempo ou menos, a equação e alguns procedimentos para o cálculo de dados meteorológicos deve ser ajustada para esses períodos, então a equação de Penman-Monteith FAO para cálculos horários são alterados da seguinte forma:

$$ET_o = (0.408 \cdot \Delta \cdot (R_n - G) + \gamma \cdot 37 / (T_{hr} + 273)) \cdot u_2 \cdot (e^o(T_{hr}) - e_a) / (\Delta + \gamma \cdot (1 + 0.34 \cdot u_2)) \quad (7)$$

Donde:

- ET_o evapotranspiración de referencia [mm hora⁻¹]
- R_n radiación neta en la superficie del cultivo [MJ m⁻² hora⁻¹],
- G flujo del calor de suelo [MJ m⁻² hora⁻¹]
- T_{hr} temperatura media del aire a cada hora [°C]
- Δ pendiente de la curva de presión de saturación de vapor en T_{hr},
- γ constante psicrométrica [kPa °C⁻¹],

$e^{\circ}(T_{hr})$ presión de saturación de vapor a temperatura del aire T_{hr} ,
 e_a promedio horario de la presión real de vapor [kPa],
 u^2 promedio horario de la velocidad del viento [$m\ s^{-1}$]

A fórmula modificada para períodos horários ETo com dados horários é calculado em períodos de luz a partir de 8:00 às 19:00 horas de 25 de Setembro (279 Outono) e 28 de dezembro de 2013 (8 inverno), localizado em Chapingo 19 ° 50 'de latitude norte e 98 ° 88' de longitude oeste e 2800 metros acima do nível do mar.

Tabela VIII os resultados obtidos com dados reais e dados estimados em 25 de setembro, cobrindo o período de 8:00 às 18:00 am no dia intervalos de uma hora mostrados.

Tabla VIII. Cálculo del RMSE de la ETo real y estimada del 25 de septiembre y 8 de diciembre

Hora del día	Eto Real 25 sep	Eto Estimada 25 sep	RMSE	Eto Real 8 dic	Eto Estimada 8 dic	RMSE
08:00	0.029	0.023	2.6807E-06	0.031	0.044	1.3919E-05
09:00	0.027	0.039	1.288E-05	0.030	0.068	0.00013122
10:00	0.030	0.053	4.6105E-05	0.038	0.412	0.01269089
11:00	0.029	0.024	2.2721E-06	0.036	0.027	7.598E-06
12:00	0.031	0.028	6.0344E-07	0.047	0.079	9.5851E-05
13:00	0.087	0.154	0.00040424	0.221	0.383	0.00240965
14:00	0.296	0.293	1.2983E-06	0.702	0.679	4.9402E-05
15:00	0.340	0.431	0.0007513	0.927	0.990	0.00035727
16:00	0.677	0.707	8.1985E-05	1.198	1.254	0.00028123
17:00	0.260	0.288	7.042E-05	0.734	0.942	0.00395536
18:00	3.654	3.034	0.03500723	9.098	9.518	0.01601642
			RMSE = 0.19073808			RMSE = 0.18975991

Este cálculo é importante porque com a sua ajuda você pode obter outras estimativas importantes para a gestão da água, tais como o cálculo da irrigação líquida para diferentes culturas.

Por exemplo, é a estimativa do espinafre irrigação rede com dados de 25 de Setembro (279 testes) e uma precipitação de 110,95 mm. A irrigação líquida é estimada com a fórmula 8:

$$ET_c = ETo * Kc \tag{8}$$

Para calcular os requisitos de irrigação líquidas (NN) usam a seguinte fórmula:

$$(N_h) = ET_c - P_e \tag{9}$$

öndende:

K_c= Coeficiente de cultivo (determinado por la fase de desarrollo del cultivo)

ET_o= Evapotraspiración de referencia

ET_c = Necesidades diarias de riego de cultivo.

Precipitación efectiva (P_e)= 0.8 P – 25

P_e = 0.8 (110.95) -25 = 63.76 mm

Tabela IX mostra a ET_o estimada e etc com K_{c2} porque espinafre para 25 de setembro está em sua segunda fase de desenvolvimento e tem um valor de 1 e Tabela X mostra a estimativa da necessidade de em conjunto com o líquido de irrigação RSME.

Tabla IX. Cálculo de la evapotranspiración de cultivo

Datos Reales				Datos Estimados		
Hora	ET _o	K _{c2}	ET _c	ET _o	K _{c2}	ET _c
08:00	0.031	1	0.031	0.044	1	0.044
09:00	0.03	1	0.03	0.068	1	0.068
10:00	0.038	1	0.038	0.412	1	0.412
11:00	0.036	1	0.036	0.027	1	0.027
12:00	0.047	1	0.047	0.079	1	0.079
13:00	0.221	1	0.221	0.383	1	0.383
14:00	0.702	1	0.702	0.679	1	0.679
15:00	0.927	1	0.927	0.99	1	0.99
16:00	1.198	1	1.198	1.254	1	1.254
17:00	0.734	1	0.734	0.942	1	0.942
18:00	9.098	1	9.098	9.518	1	9.518

Hora	Nn con datos reales	Nn con datos estimados con la Red Neurodifusa	RSME
08:00	0.031	0.044	1.5364E-05
09:00	0.03	0.068	0.00013127
10:00	0.038	0.412	0.012716
11:00	0.036	0.027	7.3636E-06
12:00	0.047	0.079	9.3091E-05
13:00	0.221	0.383	0.00238582
14:00	0.702	0.679	4.8091E-05
15:00	0.927	0.99	0.00036082
16:00	1.198	1.254	0.00028509
17:00	0.734	0.942	0.00393309
18:00	9.098	9.518	0.01603636
		RSME	0.03601236

Conclusões

Foram obtidos resultados satisfatórios na estimativa de dados em falta para a estação de Chapingo com a implementação de neuro-fuzzy, rede com a qual vários cenários foram concebidos; Por exemplo, o desenho de uma rede neuro-difusa formulação conjuntos fuzzy para o período de um dia, um mês e um ano. As últimas produzidos melhores resultados (apresentados na pesquisa) porque o número de conjuntos fuzzy e dados para o treinamento pode ser adaptado com dias anteriores e seguintes é reduzida. Para o período de um mês conjuntos mais difusas deve ser criada, resultando no problema de ordenação de dados para atender cada uma das regras também deve considerar que dentro de meses qualquer semelhança entre o comportamento do variáveis.

Após a implementação dos resultados da rede neuro-fuzzy precisos foram obtidos com o aumento de iterações na formação da rede; No entanto, é evidente que existe uma possibilidade de que não foi apenas o número de iterações de cada iteração, mas passar a formação de entrada de dados satisfazer cada uma das regras fuzzy. Desta forma, a duplicação de dados em iterações subsequentes são evitados, o que leva a uma boa distribuição de dados no interior da formação.

A proposta deste trabalho é testar com base neuro-fuzzy para o preenchimento de redes de dados, devido à sua capacidade de resolver problemas relacionados com a incerteza das informações ou

o modelo de conocimiento especializado. Assim, a rede neuro-fuzzy é considerado uma boa opção dependendo do tipo de dados disponíveis eo formato em que (formato de ficheiro Excel) são publicados.

Nosso modelo pode ser melhorada se tem uma história de registros de anos anteriores, bem como nova entrada seria anexar o ano em análise na rede, o que comparar os dados de anos anteriores ou prever dados futuros.

Você também pode melhorar o desempenho da rede através de um sistema que pode gerar os dados de treinamento de forma automatizada e que a investigação foi realizada manualmente, o que facilitaria o feedback para a rede e estabelecer um ponto em que ele pode overtraining ocorrer. Embora a lógica fuzzy tem uma história curta, é uma técnica promissora na área de recuperação de informação meteorológica.

Bibliografía

- Alfaro, J., & Javier, F. (2008). Descripción de dos métodos de rellenado de datos ausentes en series de tiempo meteorológicas. *Revista de Matemática: Teoría y Aplicaciones*, 60-758.
- Alfaro, J., & Soley, J. (2009). Descripción de dos métodos de rellenado de datos ausentes en series de tiempo meteorológicas. *Revista de Matemática: Teoría y Aplicaciones*.
- Arca, B. B. (2001). Evaluation of neural network techniques for estimating evapotranspiration. En *Evolving Solution with Neural Networks*, Italia: Baratti, R. y De Canete, pp. 62-97.
- Campos, A., Quispe, S., & Tatiana, S. (2012). Gestión de datos meteorológicos. *XXII Congreso Nacional de Hidráulica*. Acapulco Guerrero.
- Chen, H. (1995). *Machine learning for information retrieval: neural networks, symbolic learning, and genetic algorithms*. *Journal of the American Society for Information Science*.
- Cruz, H. (2012). *Estimación de la evapotranspiración de referencia en regiones con datos climáticos limitados*. Texcoco.
- Doorenbos, J., & Pruitt, W. (1997). *Guidelines for predicting crop water requirements*. Rome: Irrigation and Drainage Paper.
- Elizondo, D. G. (1994). Development of a neural network model to predict daily solar radiation. *Agricultural and Forest Meteorology*, pp. 115-132.
- Enke W, S. A. (1997). Downscaling climate model outputs into local and regional weather elements by classification and regression. *Climate Research*, pp. 195-207.
- Ferreira, M. (2003). Metodologías de análisis e imputación de datos faltantes en series de velocidad del viento. *VI Congreso Galego de Estatística e Investigación de Operaciónes*, pp. 5-7.

- Jang, J. S. (1993). ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems*, pp. 665-685.
- JE, K. (2001). Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over North America. *Cagliari*.
- Lin, C., Lee, T., & CS, G. (1996). *Neural Fuzzy System: A Neuro-Fuzzy Synergism to Intelligent Systems*. Prentice Hall.
- M. del Brio Bonifacio, S. M. (2010). *Redes Neuronales y Sistemas Borrosos*. Madrid, España: Alfaomega.
- Pitarque, A., & Roy, J. (1998). Redes neuronales vs modelos estadísticos: Simulaciones sobre tareas de predicción y clasificación, Valencia, España: Universidad de Valencia, pp. 387-400..
- R. Alfaro, R. P., Alfaro, R., & Pacheco, R. (2000). Aplicación de algunos métodos de relleno a series anuales de lluvia de diferentes regiones de Costa Rica. *Redalyc*.
- Rivera, M. (2008). *Estimación estadística de valores faltantes en series históricas de lluvia*. Pereira: Facultad De Ingeniería Industrial, Escuela de Posgrados.
- Saba, I., & Ortega, J. (2008). Estimación de datos faltantes en estaciones meteorológicas de Venezuela vía un modelo de redes neuronales.
- Solana, H., & Bote, G. (1998). *La aplicación de Redes Neuronales Artificiales (RNA): a la recuperación de la información*.
- Valesani, E., & Quintana, P. (2009). Imputación de datos con redes neuronales. *XI Workshop de Investigadores en Ciencias de la Computación*, Argentina, pp. 281-285.
- Villada F, & García, E. (s.f.). Pronóstico del Precio de la Energía Eléctrica usando Redes Neuro-Difusas. *Redalyc*.
- Wilby R, W. T. (1997). Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography*, pp. 530-548.