정량적인 월강수량 예측에 관한 연구

A Study of Quantitative Monthly Precipitation Forecast

신주영*, 허준행**

Ju-Young Shin, Jun-Haeng Heo

요 지

효율적인 장기 수자원 운영을 위하여 미래의 강수량을 예측하는 것은 중요하다. 특히 월 또는 계절단위의 강수량의 정량적인 예측이 필요하다. 우리나라에서는 기상청에서 향후 3개월의 강수량과 온도에 대하여 정성적으로 예측을 하고 있다. 정성적인 예측은 적음과 많음만을 나타내어 정보를 활용하기에 많은 제약이 있다. 기상수치모형을 통한 예측의 경우 월간과 같은 시간스케일에서 정량적인 예측이 가능하나 예측 정확도가 떨어지는 문제로 인하여, 일반적으로 정성적인 예측을 하고 있다. 이런 문제점을 극복하고자 본 연구에서는 기상수치모형을 이용하지 않고 시계열 모형을 이용하여 월강수량을 예측하고자 한다. 기존의 통계학에서 사용되는 시계열 모형과 자기학습모형 등을 이용하여 정량적인 월 강수량을 예측하는 다양한 모형을 구성하고, 각 모형의 적용성을 평가하고자 한다.

핵심용어 : ANFIS, 뉴로-퍼지, 강수량 예측, 월강수량

1. 서론

효율적인 수자원 운영 및 확보를 위해서는 미래의 다가올 강수량에 대한 정확한 예측이 동반되어야 한다. 특히 정량적이고, 장기간의 미래에 대한 예보가 일정 수준이상의 정확도를 넘어 설 경우 보다 효율적인 수자원 운영 및 확보가 가능하다. 현재 우리나라에서는 기상청에서 장기기상전망이라는 예보로 1~3개월 후 월강수량과 기온을 예보하고 있다. 장기기상전망은 정량적인 강수량을 나타내지 않고 정성적인 강수량을 예측 하고 있다. 정성적인 예보의 한계점을 극복하고 자 본 연구는 정량적인 월강수량 예측기법에 대하여 연구하였다. 본 연구는 장기간의 강수 예측을 위하여 적응형 뉴로-퍼지 시스템(Adaptive Nuero-Fuzzy Inference System, ANFIS)을 이용한 월강수량 예측 방법을 제안하였다. ANFIS 모형은 인공신경망과 퍼지모형의 특징을 가지는 모형으로 자료간의 관계가 선형이 아닌 비선형관계를 가질 경우 매우 정확한 예측모형을 구축할 수 있는 특징이 있다. 월강수량 예측이 관측된 기상자료들과 비선형 관계에 있다고 생각되어 ANFIS 모형을 이용하여 월강수량을 예측하였다. 본 연구의 대상 지점으로는 금강유역의 대전지점으로 선정하였다. ANFIS 모형의 입력자료로는 월 관측 기상 자료를 사용하였다. 1변수 입력자료 집합과 2변수 입력자료 집합을 이용하여 모형을 학습하였고, 그 결과를 분석하였다.

2. Adaptive Nuero-Fuzzy Inference System (ANFIS) 모형의 개요

2.1 ANFIS 모형

ANFIS모형은 Jang(1993)이 처음으로 제안하여 비선형 과정을 모의하는데 널리 사용되고 있다. ANFIS 모형은 인공신경망모형과 퍼지로직모형의 복합적인 모형으로 두 모형의 장점을 가지고 있다. 입력값과 출력

* 정회원・관동대학교 첨단해양공간개발연구센터 연구원・E-mail: hyjyshin@gmail.com
** 정회원・연세대학교 사회환경시스템공학부 토목환경공학전공 교수・E-mail: jhheo@yonsei.ac.kr

값이 존재하는 자료 구조간의 서로의 관계를 알고자 할 때 인공신경망 모형은 널리 쓰이고 있다. 또한 자료가 정량적이지 않은 정성적인 자료로 주어졌을 때나 룰을 가지고 있을 경우 퍼지로직모형이 널리 사용되고 있다. ANFIS 모형은 인공신경망모형의 학습능력과 퍼지로직모형의 자료의 유연한 지식제공 능력을 모두 가지고 있는 모형으로 비선형 모형 구축 시 좋은 성능을 보인다.

2.2 ANFIS 모형의 구조

ANFIS 모형은 Sugeno 퍼지 모형을 기반으로 개발되었다. ANFIS 모형은 총 5개의 층으로 이루어져 있으며 그림 1은 ANFIS 모형의 구조를 나타낸 그림이다.

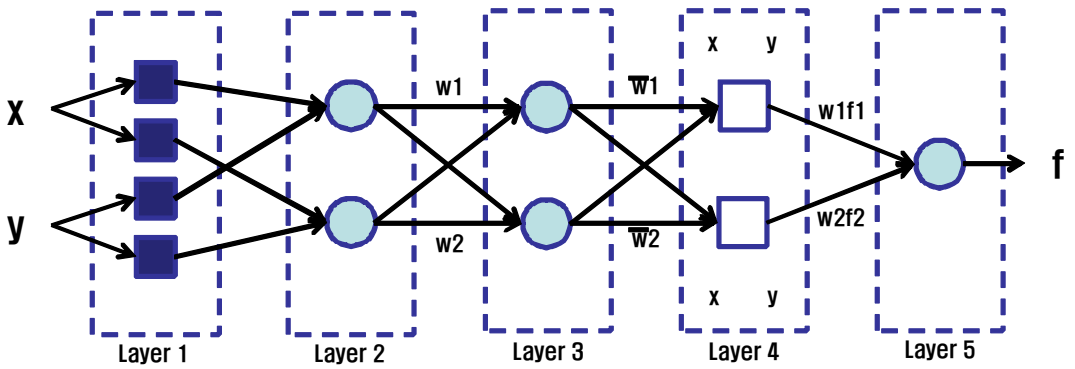

그림 1 Structure of ANFIS

2.2.1 1번 layer

이 층안에 있는 드는 입력값 x(또는 y)를 받아 membership 함수를 이용하여 입력값의 정성적인 자료를 정량화 시킨다. 본 연구에서는 gaussian 2 변수 함수를 이용하여 정성적인 자료를 정량화 시켰으며 식 (1)은 gaussian 2변수 함수를 나타낸 것이다. 또한 이 층에서 사용한 변수를 premise 매개변수라고 한다.

$$\mu_A(x) = \exp(-(\frac{x-c_i}{a_i})^2) \tag{1}$$

2.2.2 2번 layer

이번층의 모든 모드들은 고정된 노드로써 membership 함수를 통하여 계산된 정량화된 값들을 합산한다. 일반적으로 T-norm 실행자가 사용되며 이것은 fuzzy AND 연산자를 적용한 것과 같은 결과를 나타낸다. 식 (2)는 일련의 과정을 수식으로 나타낸 것이다.

$$O_{2,i} = w_i = \mu A_i(x)\mu B_i(y), \qquad i=1,\ 2. \tag{2}$$

2.2.3 3번 layer

이번 층에 있는 모든 모드는 고정된 노드로써 각 룰들의 강도를 합산하여 총 룰강도를 계산하게 된다.

$$O_{3,i} = \overline{w_i} = \frac{w_i}{w_1 + w_2}, \qquad i=1,\ 2. \tag{3}$$

2.2.4 4번 layer

4번 층은 1번~3층에서 fuzzy화 된 자료를 defuzzification을 하는 층이다. 식 (4)은 defuzzification 시켜주는 함수로 $\overline{w_i}$는 정규화된 강도를 나타낸다. 일반적인 fuzzy 모형에서는 defuzzification시 여러 경험적이고 정성적인 방법을 이용한다. 그러나 ANFIS 모형에서는 이러한 정성적인 defuzzification 방법을 정량화 하여 인공신경망과 같은 학습능력을 갖을 수 있도록 하였다. 4번 층에서 사용된 매개변수 p_i, q_i, r_i를 consequent 매개변수라 한다.

$$O_{4,i} = \overline{w_i} f_i = \overline{w_i}(p_i x + q_i y + r_i) \tag{4}$$

2.2.5 5번 Layer

5번 층에서는 노드는 고정된 노드로써 4번 층에서 계산 결과를 합산하여 결과값을 도출한다.

$$overall\ output = O_{5,i} = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (5)$$

3. 적용된 ANFIS 모형 구조

본 연구에서는 한 변수에 세 개의 rulse을 적용하였으며, 적용된 membership 함수는 gaussian 2변수 함수를 적용하였다. 4번 layer에서 적용되는 defuzzification 방법으로는 linear 방법이 적용되었다.

3.1 입력자료

ANFIS 모형은 결과를 도출하는 큰 틀은 fuzzy-system이나, fuzzy-system의 매개변수들을 도출해 내는 방법은 neural network와 같이 역전파(backpropagation) 방법을 적용한다. 이런 구조적인 특성은 입력변수들을 모형에 학습시키기 전 자료특성에 맞는 전처리를 거쳐야 한다. 또한 전처리된 입력변수는 적절한 입력변수 집합을 구성되어야 한다.

월강수량 예측 모형의 입력변수로 기상청에서 제공하는 대전 지점 관측 월기후자료를 사용하였다. 관측 월기후자료를 사용한 목적은 관측 기후변수들에 비하여 사용자가 구하기 쉽고 비교적 연관성이 크다고 생각되어 적용하였다. 아래 표 1은 월강수량 예측 모형에 사용된 관측 기상인자 즉 입력자료들을 나타내었다.

표 1 Input data

번호	관측 기상인자	영문 코드	번호	관측 기상인자	영문 코드
1	평균 기온 (℃)	AVTEM	16	평균 현지기압 (hPa)	LVPRS
2	최고 기온 (℃)	MXTEM	17	평균 해면기압 (hPa)	AvSVP
3	최저 기온 (℃)	MNTEM	18	최고 해면기압 (hPa)	MXSVP
4	월강수 (mm)	MTPCP	19	최저 해면기압 (hPa)	MNSVP
5	평균 풍속 (m/s)	AVWVT	20	평균 지면온도 (℃)	AVEGT
6	최대 풍속 (m/s)	MXWVT	21	0.05m 지중온도 (℃)	005GT
7	최대 순간풍속 (m/s)	MNWVT	22	0.1m 지중온도 (℃)	010GT
8	평균 습도 (%)	AVERH	23	0.2m 지중온도 (℃)	020GT
9	최소 습도 (%)	MINRH	24	0.3m 지중온도 (℃)	030GT
10	평균 이슬점 (℃)	DEWPT	25	0.5m 지중온도 (℃)	050GT
11	평균 증기압 (hPa)	AVPRS	26	1.0m 지중온도 (℃)	100GT
12	일조합 (hr)	SUNLT	27	1.5m 지중온도 (℃)	150GT
13	일조율 (%)	SNLTR	28	3.0m 지중온도 (℃)	300GT
14	증발량 (mm)	EVAFO	29	5.0m 지중온도 (℃)	500GT
15	평균 전운량 (10)	CLDQT			

3.2 입력자료 전처리 방법

입력자료 전처리는 모형을 잘 학습시키시 위하여 매우 중요한 절차이다. ANFIS 모형은 전처리를 통하지 않고도 학습이 가능하나, 전처리를 하였을 경우 보다 좋은 결과를 도출한다. 입력자료로 사용되는 관측 월기후자료들은 서로간의 특성이 매우 상이하며, 또한 관측값들의 크기에 따른

민감도가 달라 자료 그대로를 사용하기에는 무리가 있다. 본 연구에서는 관측값들의 크기를 일정한 수준의 크기로 맞추어 주기 위하여 표준화 방법을 적용하여 입력자료들을 전처리 하였다. 식(6)은 표준화 방법을 나타낸 식이다.

$$Y_{i,j}=\frac{\widehat{Y_{i,j}}-\hat{\mu_j}}{\sigma_j} \quad (6)$$

$\widehat{Y_{i,j}}$=i번째 j계열 월기후자료, $\hat{\mu_j}$=j계열 월기후자료 평균 σ_j=j계열 월기후자료 표준편차

3.3 입력자료 집합 구축

입력자료로 사용된 관측 월기후자료는 총 29개이다. 목표값을 다음 월 월강수량으로 하였다. 입력자료의 집합으로는 1변수 집합, 2변수 집합을 구성하였다. 1변수 집합은 1~29번 변수를 1개씩 입력변수로 사용한 입력자료 집합을 나타낸다. 1~29번 변수를 중복되지 않게 2개씩 입력변수로 사용한 입력자료 집합을 나타낸다. 사용된 관측 자료는 총 41년치 자료를 사용하였으며, 모형을 학습시키는 데 1969~2005년까지 자료를 사용하였고 2006~2009년까지 자료를 사용하여 모형의 성능을 검증하였다.

4. 결과분석

4.1 1변수 입력자료 집합 결과 분석

1변수 입력자료 집합을 입력하여 ANFIS 모형을 학습시켰다. 총 29개의 모형을 구축하였으며, 표 2는 29개의 모형의 예측 정확도를 RMSE로 나타낸 것이다. 학습에 사용한 자료의 오차는 지면이 한정되어 있어 발표 포스터에서 자세히 다루도록 하겠다. 그림 2는 RMSE 값이 가장 작은 16번 변수를 이용하여 학습한 결과를 나타낸 것이다. 예측 결과를 보면 7월, 8월 9월 자료가 잘 안 맞는 것으로 나타났다. 이러한 이유는 장마기간의 강수량 발생 요인으로 적용된 전월 기후자료와 연관성이 적은 사건의 발생하여 ANFIS 모형이 전 월 관측자료에서 그 영향을 파악하지 못하였기 때문에 일어나는 것으로 생각된다.

표 2 Checking errors(RMSE)

입력변수	1	2	3	4	5	6	7	8	9	10
Checking	100.39	101.36	109.70	106.33	123.73	131.55	130.66	130.49	118.83	113.98
입력변수	11	12	13	14	15	16	17	18	19	20
Checking	110.84	152.67	114.10	117.01	113.43	93.03	97.11	98.70	97.23	105.04
입력변수	21	22	23	24	25	26	27	28	29	
Checking	99.90	107.61	115.75	121.70	121.35	119.63	120.37	132.76	126.93	

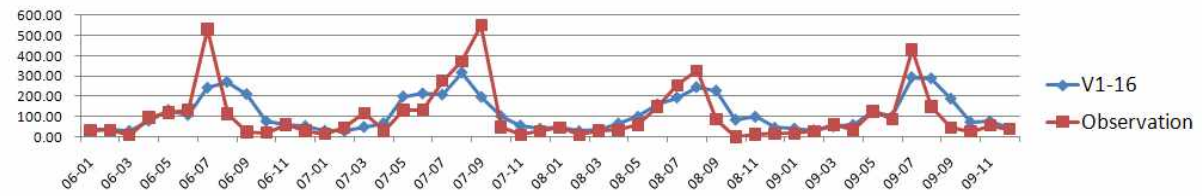

그림 2 Plot of forecast(16) and observation

4.2 2변수 입력자료 집합 결과 분석

2변수 입력자료 집합을 입력하여 ANFIS 모형을 학습시켰다. 총 406개의 모형을 구축하였다.

RMSE를 기준으로 하였을 때 몇몇 모형의 경우 1변수 입력자료 집합을 사용한 것 보다 작은 값을 보였고 몇몇 모형의 경우는 큰 값을 보였다. 아래 그림 3과 그림 4는 2변수 입력자료 집합 중 20번과 16번을 입력변수로 사용하여 학습 시킨 결과의 예측값과 관측값을 도시한 그래프와 28번과 20번을 입력변수로 사용하여 학습 시킨 결과의 예측값과 관측값을 도시한 그래프이다.

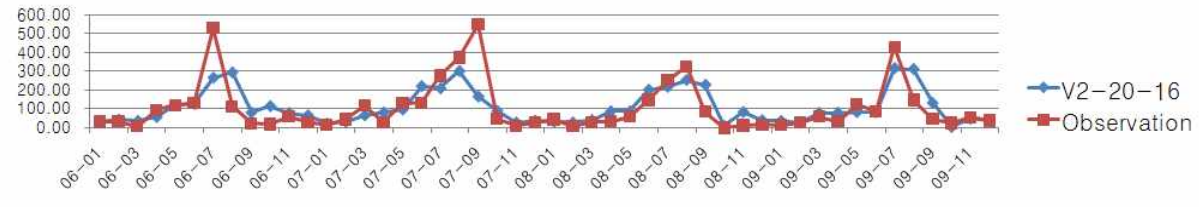

그림 3 Plot of forecast(20-16) and observation

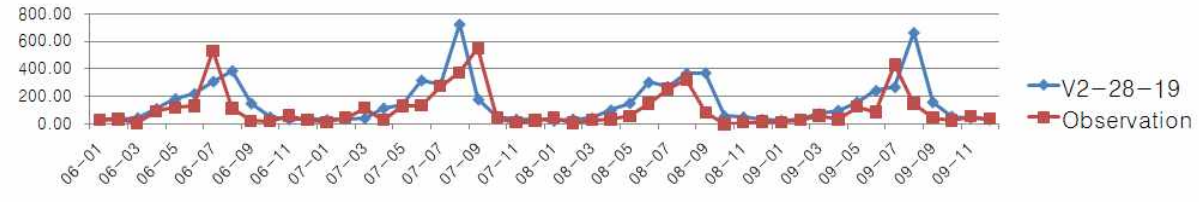

그림 4 Plot of forecast(28-19) and observation

5. 결과

대전 지점의 관측 월 기후자료를 이용하여 다음 월강수량을 예측하는 ANFIS 모형을 구축하였다. 1변수 입력자료 집합과 2변수 입력자료 집합을 입력변수로 사용하여 총 435개의 월강수량 예측 모형을 구축하였다. 각 모형의 정확도를 비교해본 결과 2변수 입력변수 집합을 사용하였을 때 1변수를 사용한 것보다 RMSE를 기준으로 낮은 값을 보이는 입력자료 집합을 찾아 낼 수 있었다. 예측 결과 갈수기에는 일정수준의 정확도를 보이나 7~10월의 결과가 잘 안 맞는 것으로 나타났다. 향후 7~10월에 강수량에 영향을 주는 인자를 찾아 학습할 경우 보다 정확한 예측이 가능할 것으로 생각된다.

감 사 의 글

본 연구는 21세기 프론티어연구개발사업인 수자원의 지속적 확보기술개발사업단의 연구비지원(과제번호:1-6-3)에 의해 수행되었습니다.

참 고 문 헌

1. Jang, J. S. R. (1993). ANFIS : Adaptive-Network-Based Fuzzy Inference System, *IEEE Transactions on Systems, Man and Cybermetics*, Vol. 23(3), pp. 665-685.