

단기 용수수요예측을 위한 신경망모델에 관한 연구

A study on neural network model for a short term water demand forecasting

○황석환* · 신현석** · 김종훈***

1. 서론

용수수요는 일별주기형태를 나타내기 때문에 전형적인 운영계획은 1일단위로 계획되어진다. Wolff와 Loos(1956)등에 의한 최고치 용수수요량 분석과 같은 용수수요와 예측방법을 다룬 많은 보고자들이 제출되어 왔다. Howe(1968)는 여름철에 있어서의 용수수요와 최고 용수수요량을 고려하여 용수수요예측을 위한 방정식을 발표한바 있다. 용수사용 시계열모형의 주기적인 구성요소 들은 Fourier 급수변환을 통한 sine 과 cosine 함수의 조합으로 나타내어 질 수 있다. 이 방법은 영국에서 용수수요 시계열모형을 만들기 위해 Perry(1981)에 의해서도 사용된바 있다. 그는 모형의 변수를 결정하기 위해 Kalman filter 의 sinusoidal 함수를 사용하여 계절적 구성요소 들을 나타내었다.

Higgins 와 Brubach(1971)는 용수수요 시계열의 계절적이고 일반적 성향을 분석하기 위해서 계절보정 방법을 확장하여 사용하였다. 그들은 모형에서 관측치의 시간지체를 일주일간의 시간에 해당하는 168 시간일 때까지를 선택하였다. 그 기술은 비록 단순하지만 잘 구성되어진 이동평균방법 이다. 본 연구에서 제안된 모델의 주말 특성을 나타내기 위하여 한 주간에 있어서의 계절성분을 사용하는 이 방법을 적용하였다.

2. 신경망 모델

전방향 역전파 네트워크(feedforward backpropagation network)는 신경망에 있어서 매우 인기 있는 모델로서 피드백(Feedback: 출력의 일부를 입력으로 반환하는 조작) 연결부분이 없다. 그러나 오차는 학습되는 동안 역전파되고 최소평균자승오차(least mean squared error)가 사용된다. 층들간의 연결가중치(connection weight)의 보정치를 결정하는데 기반이 되는 출력층에서의 오차는 은닉층 출력오차의 주정치를 결정한다. 그리고 출력 값의 개계산은 오차가 기대치 이하로 내려갈 때까지 반복적으로 수행된다. 모델링 인자는 전 반복 계산치로부터 보정치를 추정하는데 사용되고 보정치에 더해지게 된다. 신경망

* 고려대학교 토목환경공학과 석사과정

** 부산대학교 토목공학과 조교수

*** 고려대학교 토목환경공학과 부교수

모델은 다음과 같이 구성된다. 입력, 은닉, 그리고 출력층의 단위 수를 각각 k, j , 그리고 i 라고 놓으면 각층에서의 단위들은 다음과 같이 표현되어진다.

$$\begin{aligned}
 \text{Hidden Unit } j \text{ receives } h_j^u &= \sum_k w_{jk} I_k^u \\
 \text{outputs } a_j^u &= g(h_j^u) \\
 &= g\left(\sum_k w_{jk} I_k^u\right) \\
 \text{Output Unit } i \text{ receives } h_i^u &= \sum_j w_{ij} a_j^u \\
 &= \sum_j w_{ij} g\left(\sum_k w_{jk} I_k^u\right) \\
 \text{outputs } O_i^u &= g(h_i^u) \\
 &= g\left(\sum_j w_{ij} a_j^u\right) \\
 &= g\left(\sum_j w_{ij} g\left(\sum_k w_{jk} I_k^u\right)\right)
 \end{aligned}$$

본 모델에서는 최적화 문제를 풀기 위해 그리고 최적의 연결가중치를 얻기 위해서 gradient descent method 가 사용되었다.

3. 전방향 역전파 신경망 모델(Feedforward Backpropagation Neural Network Model)

신경망 이론을 이용한 용수수요 예측에 영향을 주는 입력인자 들은 매우 많다. 그러나 용수수요와 기후인자 그리고 생활습관에서 오는 영향을 제외한 기타 변수들은 단기 용수수요 예측에 있어 그리 큰 영향을 주지 못한다. 이런 이유에서 이런 외부변수 들은 고려 하지 않았다. 본 연구에서 구성된 신경망 모델들이 Table 1.에 나타나 있다.

TABLE 1. Neural Network Models Tested in This Study

| Input | Output | Model | Input | Output | Model |
|-----------------|--------|---------|-----------------------|--------|----------|
| (A) | Demand | Model 1 | (A) & (C) & (D) | Demand | Model 10 |
| (B) | Demand | Model 2 | (A) & (C) & (E) | Demand | Model 11 |
| (C) | Demand | Model 3 | (D) & (G) | Demand | Model 12 |
| (A) & (B) | Demand | Model 4 | (E) & (G) | Demand | Model 13 |
| (A) & (C) | Demand | Model 5 | (H) | Demand | Model 14 |
| (F) | Demand | Model 6 | (I) | Demand | Model 15 |
| (G) | Demand | Model 7 | (A) & (B) & (D) & (E) | Demand | Model 16 |
| (A) & (B) & (D) | Demand | Model 8 | (A) & (C) & (D) & (E) | Demand | Model 17 |
| (A) & (B) & (E) | Demand | Model 9 | (D) & (E) & (G) | Demand | Model 18 |

(A): previous day demands, (B): forecasted temperatures, (C): forecasted high temperatures, (D): the duration of sunshine, (E): week patterns (Sunday or weekday), (F): input type which has past two demands and temperatures, (G): input type which has past two demands and high temperatures, (H): input type which has past three demands and temperatures, (I) input type which has past seven (1-week) demands and high temperatures.

4. 일년예측

TABLE 2. Comparison of Results between Neural Network and Multiple Linear Regression Model

| TEST 1 | Model | AMB | RMSE | RRMSE | MAPE(%) |
|--------|----------|----------|----------|----------|---------|
| | NN | 40900.24 | 52249.04 | 0.010447 | 0.81245 |
| MLR | 44793.85 | 56523.32 | 0.011302 | 0.89540 | |

Table 2 의 모든 수치들을 비교해 볼 때 신경망 모델이 좀더 정확히 예측함을 볼 수 있다. 결과치들로부터 신경망 모델이 다중선형회귀분석(Multiple Linear Regression Model)에 비해 신뢰성 있고 적당한 모델임을 알 수 있다. 다시 말해 입력변수와 출력 변수들간의 관계는 비선형적임이 증명된 셈이다.

TABLE 3. Comparison of Results among Neural Network Models for Different Input Type

| | Model | AMB | RMSE | RRMSE | MAPE(%) |
|---------------|---------|----------|----------|----------|---------|
| TEST 2 (1) | Model 1 | 44369.00 | 59286.29 | 0.011854 | 0.8885 |
| | Model 2 | 75616.35 | 95560.29 | 0.019107 | 1.5178 |
| | Model 3 | 77040.19 | 96970.43 | 0.019389 | 1.5352 |
| | Model 4 | 44860.89 | 59099.43 | 0.011817 | 0.8946 |
| | Model 5 | 43721.11 | 56362.59 | 0.011264 | 0.8705 |
| | Model 6 | 43320.89 | 58295.04 | 0.011656 | 0.8636 |
| | Model 7 | 42805.56 | 57093.6 | 0.011416 | 0.8546 |

Table 3 에 열거된 각각의 신경망 모델들의 결과치를 비교해 보면 Model 5, Model 6, Model 7 은 다른 것들에 비해 좀더 정확한 값을 보이고 있다. 이것은 일년간의 예측에 있어서 기온이 신경망 모델의 중요한 입력변수로 작용한다는 것을 증명하는 것이다. 특히 Model 6 과 Model 7 의 결과값들은 지난 이 분간의 용수요구량과 기온이 오늘의 용수요구량에 영향을 줄을 명백히 한다. 그러나 Model 6 과 Model 7 은 똑같은 입력구조를 가지고 있기 때문에 최종적으로는 Model 5 와 Model 7 을 가장 적당한 모델로 택하였다.

TABLE 4. Comparison of Results in case of Adding Duration of Sunshine to Model 5 and Model 7

| | Model | AMB | RMSE | RRMSE | MAPE(%) |
|--------|-------------|-----------|----------|----------|---------|
| TEST 3 | Model 10 | 42694.74 | 54891.68 | 0.010976 | 0.8526 |
| | Model 12 | 43425.75 | 57765.30 | 0.011550 | 0.8688 |
| | *Model 12-1 | 42596.114 | 56484.96 | 0.011294 | 0.8504 |

Model 10 은 Model 5 의 입력인자에 일조량을 추가한 것으로 Table 4 에서 보듯이 상당히 결과치가 개선되었음을 알 수 있다. 그 반면에 Model 7 은 Model 12 와 비교해 볼 때 결과가 오히려 더 안 좋아졌다. 이 결과로 부터 하루전의 일조량은 오늘의 용수수요에 영향을 미치지만 이틀 전 이상의 일조량은 용수수요에 영향을 주지 못함을 알 수 있다. *Model 12-1 의 구조는 Model 12 와 입력인자에서 약간 다르다. *Model 12-1 은 Model 12 의 결점을 개선한 것으로 일조량에 있어서 단지 전날자료만을 사용한 것으로 Model 12 에 비해 결과가 좋아졌음을 볼 수 있다. 그러나 여기서 주목해 보아야 할 점은 AMB 와 MAPE 의 경우에는 Model 10 이나 Model 12 보다 더 좋지만 RMSE 와 RRMSE 에 있어서는 더 큰 오차를 보인다. 이것은 비록 Model 12-1 이 예측에 있어 더 정확하긴 하지만 더 큰 오차범위를 가짐을 의미한다. 즉 다시 말해 Model 12-1 은 Model 10 이나 Model 12 보다 비교적 안정하지 못하고 믿을 수 없는 구조임을 말한다.

TABLE 5. Comparison of Result in case of Adding Week Patterns to Model 5 and Model 7

| | Model | AMB | RMSE | RRMSE | MAPE(%) |
|--------|-------------|----------|----------|----------|---------|
| TEST 4 | Model 11 | 41163.29 | 53272.46 | 0.010652 | 0.8203 |
| | Model 13 | 42230.20 | 56463.55 | 0.011290 | 0.8432 |
| | *Model 13-1 | 42707.24 | 57111.72 | 0.011420 | 0.8527 |

Table 5 는 Model 5 와 Model 7 에 일별특성을 입력인자로 추가한 것이다. Model 11 과 Model 13 의 결과치에서 보는 것처럼 오차가 감소했음을 볼 수 있다. 그리고 TEST 3(Table 4)과 TEST 4(Table 5)을 비

고에 보면 일별특성이 일조량보다 더 중요한 입력요소임을 알 수 있다. *Model 13-1은 입력인자에 있어서 이틀전의 일별특성은 사용하지 않고 하루전의 일별특성만을 사용한 것으로 Model 13과 비교해 볼 때 일조량과는 달리 오차가 증가 했음을 알 수 있다. 따라서 Model 11을 최적의 모델로 선정하였다.

TABLE 6. Comparison of Results in case of Adding Duration of Sunshine and Week Patterns to Model 5, 7

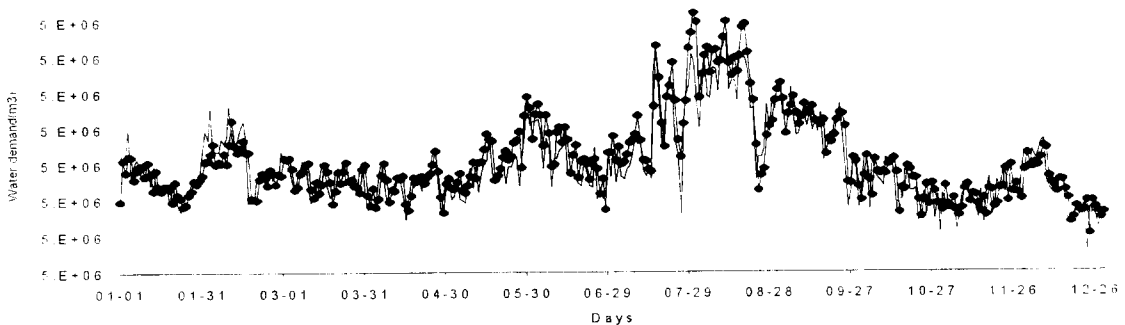
| TEST 5 | Model | AMB | RMSE | RRMSE | MAPE(%) |
|--------|-------------|----------|----------|----------|---------|
| | Model 17 | 40900.24 | 52249.04 | 0.010447 | 0.8124 |
| | Model 18 | 42436.22 | 56328.44 | 0.011263 | 0.8504 |
| | *Model 18-1 | 41242.91 | 55117.74 | 0.011021 | 0.8234 |

Table 6은 앞에서 언급한 모든 입력 변수들(기온, 하루전의 용수수요량, 일조량, 일별특성)을 가진 모델의 결과치들이다. 그리고 Model 17은 Model 5, 10, 11과 같은 맥락으로, 네 결과치들을 비교해 보면 Model 17이 가장 정확한 예측을 한다는 것을 알 수 있다. *Model 18-1은 이틀전의 일조량과 일별특성이 빠진 모델로 Model 18에 비해 더 나은 예측결과를 보여주긴 하지만 Model 17에 비해 결과치가 좋지 못하다. 따라서 Model 17이 일년예측의 실용적 측면에서 볼 때 가장 정확하고, 좀더 신뢰성 있고 적합한 모형이라는 결론을 내렸다.

TABLE 7. Results for Model 14 and Model 15

| TEST 6 | Model | AMB | RMSE | RRMSE | MAPE(%) |
|--------|----------|----------|----------|----------|---------|
| | Model 14 | 44649.45 | 59741.48 | 0.011946 | 0.8910 |
| | Model 15 | 44511.90 | 60876.54 | 0.012173 | 0.8876 |

Model 14는 과거 3일간의 용수수요와 최고기온이 입력자료로 사용된 형태이다. 그러나 결과적 측면에서 Model 7과 비교해 볼 때 더 부정확한 예측을 하고 있다. 이 결과는 오늘의 용수수요는 3일전 이상의 용수수요와 기온의 영향을 받지 않는다는 것을 알 수 있다. Model 15는 일주일 동안의 용수수요와 기온이 사용된 모델로 Model 14보다는 결과치가 좋다. 즉 용수수요는 일주간의 주기적 특성을 가진다는 것을 알 수 있다.



<FIG. 1. Comparison of Result between Estimated and Forecasted Data for One-year >

5. 7, 8 월 예측

한국의 경우 1년 중 갈수기이면서 고온인 7.8월의 경우엔 기온과 용수수요가 높은 상관관계를 보인

다. 그리고 특히 이 시기에 관심을 두는 이유는 용수사용량이 가장 많고 갈수기인 관계로 정확한 용수 수요량의 예측이 중요시되기 때문이다. 그래서 이 시기의 일별 용수수요량에 대한 정확한 예측은 도시의 급수시스템을 운영하는데 필수조건이라 하겠다.

TABLE 8. Comparison of Results among Neural Network Models for Different Input Type

| | Model | AMB | RMSE | RRMSE | MAPE(%) |
|--------|---------|-----------|-----------|----------|---------|
| TEST 7 | Model 1 | 68927.45 | 88863.57 | 0.017303 | 1.3478 |
| | Model 2 | 108785.70 | 127386.80 | 0.024803 | 2.1320 |
| | Model 3 | 111560.10 | 136148.70 | 0.026509 | 2.1923 |
| | Model 4 | 62653.03 | 80033.43 | 0.015583 | 1.2262 |
| | Model 5 | 64177.83 | 80456.66 | 0.015666 | 1.2589 |
| | Model 6 | 70574.93 | 90337.15 | 0.017574 | 1.3763 |
| | Model 7 | 71259.15 | 92162.19 | 0.017930 | 1.4488 |

Table 8 에 열거된 각각의 신경망 모델들의 결과치를 비교해 보면 Model 4, Model 5 은 다른 것들에 비해 좀더 정확한 값을 보이고 있다. 이 결과는 7, 8 월의 경우도 일년예측과 마찬가지로 기온이 용수수요예측에 있어 중요한 설명변수로 작용한다는 것을 보여주는 것이다. 그러나 특히 Model 4 의 결과에서 주목할 점은 이 시기의 경우에는 일년예측의 경우와 다르게 평균기온이 중요한 영향인자로 작용한다는 것이다. 7, 8 월의 경우는 온도가 높기 때문에 평균 기온역시 고온의 추세를 잘 따라가지만, 용수수요는 7, 8 월의 고온변화를 따라갈 만큼 민감하게 반응하지 못하기 때문에 생기는 결과로 본다.

TABLE 9. Comparison of Results in case of Adding Duration of Sunshine to Model 5 and Model 7

| | Model | AMB | RMSE | RRMSE | MAPE(%) |
|--------|----------|----------|----------|----------|---------|
| TEST 8 | Model 8 | 63867.65 | 76918.80 | 0.014977 | 1.2507 |
| | Model 10 | 64111.19 | 80316.54 | 0.015638 | 1.2576 |

Table 9 의 Model 8 의 결과에서 Model 4 에 일조량 입력인자를 추가하여 예측률이 향상된 것을 볼 수 있다. 그러나 Model 10 의 경우엔 Model 5 에 비해서 더 낮아졌다고 보기 힘들다. 이 결과로부터 앞에서 언급했듯이 일조량을 사용했을 경우 불안정하고 신뢰성이 떨어지는 모형이 된다는 것을 다시 한번 확인할 수 있다. 하여간 7, 8 월 예측의 경우가 일년예측의 경우보다 일조량을 입력인자로 사용했을 경우 좀더 안정한 모형이 되었다.

TABLE 10. Comparison of Results in case of Adding Week Patterns to Model 5 and Model 7

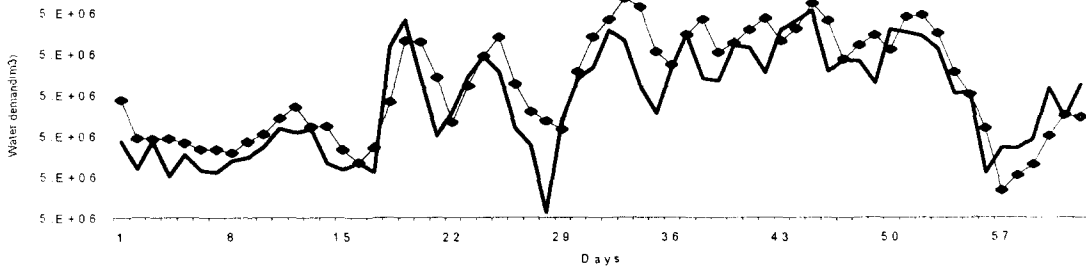
| | Model | AMB | RMSE | RRMSE | MAPE(%) |
|--------|----------|----------|----------|----------|---------|
| TEST 9 | Model 9 | 57995.53 | 71795.91 | 0.013979 | 1.1326 |
| | Model 11 | 58174.3 | 72555.05 | 0.014127 | 1.1390 |

TEST 8 과 비교해 볼 때 TEST 9 의 예측률이 더 높다. 이 결과에서 보듯이 일별특성이 일조량보다 좀더 중요한 입력인자임을 알 수 있다. 그리고 일용수수요량 예측에 있어서 일별특성은 중요한 설명변수로 작용한다는 것을 보여준다.

TABLE 11. Comparison of Results in case of Adding Duration of Sunshine and Week Patterns to Model 5,7

| | Model | AMB | RMSE | RRMSE | MAPE(%) |
|---------|----------|----------|----------|----------|---------|
| TEST 10 | Model 16 | 56538.53 | 68799.82 | 0.013396 | 1.1053 |
| | Model 17 | 58228.23 | 72616.93 | 0.014139 | 1.1400 |

Table 11 은 위에서 언급된 모든 입력인자 들을 사용하여 7, 8 월의 경우에 용수수요량을 예측한 결과를 보여준다. 결과들을 비교하여 볼 때 비록 일조량 같은 불안정한 인자가 있긴 하지만 가장 정확한 예측을 하는 모델임을 알 수 있다. 그리고 입력인자로 평균기온, 하루전의 용수수요량, 일조량 그리고 일별특성을 가진 Model 17 이 7, 8 월 용수수요 예측의 경우에 가장 예측률이 높았고 실무에 적용하기에 가장 안정하고 적합한 모델이라 말할 수 있다. 일년측의 경우와 다른 점은 일년 용수수요 예측의 경우엔 최고 기온이 중요한 역할을 하였지만 7, 8 월 용수수요 예측의 경우엔 평균기온이 중요한 입력인자로 작용한다는 것이다.



<FIG. 2. Comparison of Result between Estimated and Forecasted Data for July and August>

6. 결론

FFNN 을 이용하여 서울의 일별 용수수요량예측을 하여 보았다. 실제 일별용수수요자료, 기후인자, 그리고 일별특성을 적용하여 모형을 구성하였고 기존의 예측 방법인 다중선형회기분석과의 비교를 통하여 신경망 모델이 더 정확하고 안정적인 예측을 한다는 결론을 얻을 수 있었다. 일별 용수수요량에 영향을 주는 모든 인자들 가운데 기온, 하루전의 용수수요량, 일조량 그리고 일별특성은 중요한 설명 변수임이 밝혀졌다. 그리고 또 다양한 입력형태를 가진 신경망모형을 구성하여 각각의 신경망모델들 간의 상호 비교에 의하여 일년예측의 경우에는 최고기온, 전날의 용수수요량, 일조량 그리고 일별특성을 입력인자로 사용한 Model 17 이 가장 최적의 신경망 예측 모델이었고 7, 8 월 예측의 경우에는 평균기온, 전날의 용수수요량, 일조량 그리고 일별특성을 사용한 Model 16 이 적합한 모델이었다. 그러나 일조량의 경우는 불안정한 인자이므로 실무에서 적용할 때는 세심한 주의와 연구가 필요하다. 이들 전날까지의 용수수요와 기온자료는 오늘의 용수수요량에 영향을 주지만 그 이전자료는 거의 영향을 주지 못했다.

7. 참고문헌

- Higgins, L., and Brubach, S.(1976). "New method of forecasting hourly demands." 8th World Energy Conf. Howe C.W., Municipal Water Demands, Chapter 3 of Forecasting the Demands for Water by Swwell, W.R.D., B.Bower, et.al., Policy and Planning Branch, Dept. Of Energy, Mines and Resources, Ottawa, Canada, 1968.
- Perry, F. P. (1981). "Demand forecasting in water supply networks." J. Hydr. Div., ASCE, 107(9), 1077-1087. in Supply, Water Resources Research, vol.5, No.2, pp.350-361, 1969.
- Wolff, T.B. and Loos, J.F., Analysis of Peak Water Demands, Public Works, September, 1956.