신경망을 이용한 상수도 응집제 PAC의 주입률 산정 (Determination of PAC Injection Rate in Water Treatment Process using Neural Network)

이경훈[1] · 문병석[2] · ○이병조[3]

1. 서론(Introduction)

응집제 주입률은 원수의 탁도와 수온의 변화에 따라 수시로 실시하는 Jar-Test결과에 의해 결정하고 있으며 이 결과를 참고하여 조건표를 만들어 활용하고 있다. 그러나, 이와 같은 방법에 의한 응집제 주입률 결정은 2시간 이상이 소요되어 수시로 변하는 수질변동에 능동적으로 대처할 수 없다. 이 때문에 수질상태에 의해 약품투입률을 실시간으로 결정할 수 있는 방법이 요청된다.

김강석(1997)은 통계적 결정방법인 다중회귀 분석방법을 이용하여 상수도 응집제 PAC과 LAS의 주입률 결정을 하였으며 또한 공식을 제안하였다. 수자원공사(1997)에서는 응집제 투입량 자동결정 시스템 개발연구를 하였다. 여기서는 최적 약품투입률의 결정에 Fuzzy Neural Network(FNN) 알고리즘을 도입하였으며 스스로 학습이가능한 자기조직형 FNN을 제시하였고 비선형 다변수적인 특성을 가진 응집제 투입률 결정에 적용하여 수질변화에 능동적으로 대처할 수 있음을 확인하였다.

본 연구는 광주광역시 물 공급원인 주암댐 원수를 이용하여 응집공정에 투입되는 상수도 응집제인 PAC(폴리염화알루미늄; Poly Aluminum Chloride)의 주입률을 신경망(Neural Network)을 이용하여 결정하였다. 또한, 신경망 모델의 상수도 응집제 주입률 결정에 대한 적용 가능성 및 신경망모델과 통계적 방법인 다중회귀분석(Multi-Regression)과 비교 검토하였다.

2. 신경망 이론(Theory of Neural Networks)

신경망 이론은 신경세포가 외부로 부터오는 자극에 동적반응 과정을 묘사하여 개발된 것으로 뉴로컴퓨팅, 인공신경망, 신경회로망등의 여러 가지 용어로써 혼용되고 있다. 신경망은 의학계, 이공계, 자연계등 여러 분야에서 연구되고 있으며 음성인식, 문자인식, 영상처리, 제어, 예측, 의료진

1) 전남대학교 토목공학과 부교수
2) 서남대학교 토목공학과 전임강사
3) 전남대학교 공업기술연구소 연구원

단시스템, 최적화등 그 응용분야가 매우 넓다.

본 연구에서는 신경망을 상수도 응집제 주입률 예측모형의 개발에 적용하였다. PAC 주입률 예측 신경망은 입력층(Input Layer), 은닉층(Hidden Layer) 그리고 출력층(Output Layer)으로 구성된 다층 신경망을 사용하였다. 학습방법은 모멘텀 방법이며 기울기 강하 방법(Gradient Descent Method)에 의하여 수행되었다. 기울기 강하 방법은 응집제 주입률을 예측하기 위하여 다층 신경망을 학습시켜 모형의 최적 연결강도를 구하는 방법이다. 이것은 예측 주입률과 Jar-Test에 의한 응집제 주입률의 제곱오차로 표현되는 목적함수의 기울기를 구하여 매개변수 공간에서 탐색이동 과정을 반복함으로써 오차의 최소점에 도달하고 이때의 매개변수를 최적해로 하는 학습알고리즘이다.

3. 모형의 구성(Composition of Models)

PAC 주입률 결정은 원수 pH, 원수 탁도, 원수 알칼리도, 응집후 탁도, 원수 온도를 PAC과의 상관정도에 따라 모델식을 구성하였다. 또한, 신경망의 활성화 함수는 비선형함수인 시그모이드 함수를 사용하였다.

은닉층의 수가 많으면 당연히 연결강도(Weight)가 많아져서 신경망의 학습이 매우 느려지게 되는데, 다행스럽게도 특수한 경우를 제외하고는 은닉층을 1개로 하여도 거의 대부분 응용에 적합한 것으로 알려져 있어, 본 연구에서도 은닉층 수를 1개로 하였다. 또한 은닉층 노드수(Nodes)는 정확한 수를 계산할 수 있는 방법이 없지만 일반적으로 노드수가 많은 것은 모델이 좋은 예측력을 보이는 것으로 알려져 있다. 그러나 반드시 노드수가 많은 좋은 예측력을 보이는 것이 아니다. 또한 노드수가 많으면 과다접합(Overfitting)으로 인하여 예측력 및 일반화 능력이 떨어질 수 있다. 본 연구에서 구성한 모형은 Model 1과 Model 8까지 이며 N-N-1형태의 신경망을 구성하였다. 목표값은 PAC 주입률이며 각 모형에 대한 입력변수와 은닉층 수는 표 1과 같다.

표 1 입력변수에 따른 PAC 주입률 결정 모델

모델명	입력변수	은닉층 노드수
Model 1	원수탁도, 원수 pH, 원수 온도 원수 알칼리도, 응집후 탁도	5
Model 2	원수탁도, 원수 pH, 원수 온도, 원수 알칼리도,	4
Model 3	원수탁도, 원수 pH, 원수 온도, 응집후 탁도	4
Model 4	원수탁도, 원수 pH, 원수 알칼리도, 응집후 탁도	4
Model 5	원수탁도, 원수 pH, 원수 알칼리도	4
Model 6	원수탁도, 원수 pH, 원수 온도	3
Model 7	원수탁도, 원수 pH, 응집후 탁도	3
Model 8	원수탁도, 원수 pH	2

PAC 주입률 결정은 학습과정을 통하여 추정된 연결강도를 이용하여 계산된다. 연결강도는 각 층의 상태를 매개변수로 저장시킨 값으로서 처리 노드간의 입출력의 연결정도를 나타내며 연결되어 있는 처리 노드간의 영향력 강도를 나타낸다. 따라서 PAC 주입률 결정은 추정된 연결강도 행렬을 이용하여 계산된다.

4. 모의결과 및 예측(Results of Simulation & Forecasting)

4.1 분석방법(Analysis Methods)

자료 분석에서 모형의 식별에 의해 잠정적으로 선택된 모형의 모수추정에 주로 이용되고 있는 자료에 대한 적용타당성 문제를 다루기 위하여 본 연구에서는 가장 일반적으로 사용되고 있는 평균제곱오차(MSE; Mean Square Error), 절대평균오차(MAE; Mean Absolute Error), 절대오차율(Absolute Error Rate), 상관계수(Correlation) 및 결정계수를 사용하였으며, 계산식은 다음과 같다.

$$\text{평균제곱오차} = \frac{1}{N}\sum_{i=1}^{N}(y_i - o_i)^2 \qquad \text{절대평균오차} = \frac{1}{N}\sum_{i=1}^{N}|y_i - o_i|$$

$$\text{오차율} = \left|\frac{(y_i - o_i)}{o_i}\right| \times 100 \qquad R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - o_i)^2}{\sum_{i=1}^{N}(y_i - \overline{o_i})^2}$$

여기서, N은 자료의 수, y_i는 목표값, o_i는 출력값, $\overline{o_i}$는 출력값의 평균이다. 그리고 예측결과와 학습결과의 정확성 및 적합성은 위 식들로 계산한 값 중에서 평균제곱오차, 절대평균오차, 절대오차율은 적을수록 상관계수 및 결정계수는 클수록 좋은 것으로 판단하였다.

4.2 훈련결과(Results of Train)

훈련과정은 최적의 연결강도(Weight)를 얻는 과정이다. 본 연구에서는 N-N-1을 신경망의 구조로, 훈련방법은 모멘텀방법, 활성화 함수는 시그모이드(Sigmoid)함수로 훈련을 실시 하였다.

훈련결과 최종적인 각 모형의 오차는 표 2와 같다. 표2에서 보듯이 Model 4가 MAE는 0.99NTU, MSE는 1.51NTU로 다른 모형보다 좋은 결과를 나타내고 있다. 또한 최대 및 평균오차율이 11.77%, 4.25%로 나타났으며 상관계수는 0.988, 결정계수는 0.976으로 가장 훈련이 잘된 모형임을 알 수 있다. 그러나 Model 6의 경우 MAE는 1.456110이고 MSE는 3.528503, 최대오차율 21.89%, 평균오차율 6.32%로 매우 높게 나타났다. 또한 Model 3의 평균오차율이 7%로 다른 모형보다 가장 높은 오차율을 나타내었다. 그러나 모든 모형의 상관계수가 최고 0.988, 최저 0.972로

매우 높은 상관관계를 나타냈으며, 결정계수 또한 최고 0.976, 최저0.944로 나타났다. 이는 모든 모형이 대체적으로 만족할 만한 훈련결과를 나타냈다고 판단된다.

4.3 예측결과(Results of Forecasting)

표 3은 각 모형별 PAC 주입률 예측결과이다. 가장 좋은 예측결과를 보인 Model 4의 경우 MAE가 0.94 NTU , MSE는 1.33 NTU 그리고 최대오차율은 9.82%이고 평균오차율은 3.50%로 나타났다. Model 3의 경우 최대오차율이 20%이고 평균오차율이 7.99%로 가장 나쁜 결과를 보여주고 있다. 훈련결과와 비교해 볼 때 거의 비슷한 예측결과를 보여주고 있다. 대체적으로 훈련이 잘된 모형이 예측 또한 잘 된 것으로 판단 되지만, 반드시 그렇지는 않다. 표 3을 보면 Model 7은 MAE 1.07과 최대오차율 7.82%로 Model 4보다 더 좋은 결과를 나타내고 있다. 하지만 MSE나 평균오차율, 상관계수 및 결정계수는 Model 4가 더 좋은 결과를 보여준다.

표 2 각 모델 훈련결과

Model	MAE	MSE	오차율		상관계수	결정계수
			최 대	평 균		
Model 1	1.169961	2.201218	13.60127	4.575415	0.982	0.965
Model 2	1.273455	2.647793	16.48591	5.558970	0.979	0.959
Model 3	1.825606	4.776318	20.44540	7.640494	0.972	0.945
Model 4	0.988081	1.512722	11.77558	4.245577	0.988	0.976
Model 5	1.078347	1.715075	13.48428	4.724978	0.987	0.974
Model 6	1.456110	3.528503	21.89245	6.326182	0.972	0.944
Model 7	1.118218	1.791962	12.74717	4.769489	0.986	0.972
Model 8	1.116097	1.812242	14.56275	4.895508	0.987	0.973

표 4은 다중회귀분석(Multi-Regression)으로 구한 식으로 예측한 결과이다. 각 모형의 인자는 신경망의 모형과 동일하다. 다중회귀식은 훈련자료로 만들었으며 예측은 검증자료를 사용하였다. 다중회귀분석에 의한 예측은 신경망과 달리 MR 3 모델이 가장 좋은 결과를 나타내었는데 MAE는 0.99 NTU, MSE는 1.58 NTU로 나타났으며, 최대오차율은 13.70%이고 평균오차율은 3.51%이다. 신경망 모형인 Model 4와 비교해 보면 평균오차율은 비슷하나 최대오차율에서 차이가 나고 있다. 이것은 신경망이 다중회귀분석보다 예측력이 우수함을 나타내고 있다. 예측결과는 신경망의 일반화능력을 평가해보는 것으로서 신경망을 구성하는 최종목표라 하겠다. 훈련결과가 좋다고 하여 반드시 예측결과가 좋은 것은 아니다. 이때는 훈련이 과다적합(Overfitting) 되었는지를 점거해

볼 필요가 있다. 마찬가지로 훈련결과가 나쁘다고 하여 예측결과가 반드시 나쁜 것은 아니다. 하지만 대체적으로 훈련결과가 좋을 때는 예측결과도 좋다.

신경망을 이용하여 결정한 모형 Model 4의 경우 예측결과의 평균오차율이 3.5%로 비교적 정확한 결과를 나타냈다. 다중회귀분석에 의한 MR3의 평균오차율은 3.51%이다. 그러나 최대오차율은 신경망 모형은 9.83%이고 다중회귀 분석은 13.70%로 나타났다. 이것은 평균오차율은 비슷하나 최대오차율에서 신경망 모형이 더 우수함을 나타내고 있다.

표 3 각 모델의 예측결과

Model	MAE	MSE	오차율		상관계수	결정계수
			최 대	평 균		
Model 1	1.033869	1.521394	9.386000	3.981424	0.983	0.967
Model 2	1.253857	2.709865	20.32555	4.659673	0.975	0.950
Model 3	2.192515	6.084341	20.61518	7.995329	0.962	0.926
Model 4	0.940531	1.335403	9.827942	3.500431	0.986	0.972
Model 5	1.021444	2.018464	16.84515	4.184173	0.978	0.957
Model 6	1.609718	3.695880	19.64638	5.920892	0.970	0.942
Model 7	1.072822	1.575825	7.826481	3.766397	0.984	0.969
Model 8	1.024904	1.984260	16.71058	4.062662	0.978	0.956

표 4 각 모델의 다중회귀에 의한 예측결과

Model	MAE	MSE	오차율		상관계수	결정계수
			최 대	평 균		
MR 1	1.050590	1.773294	14.41038	3.932151	0.986	0.973
MR 2	1.221740	2.492311	15.74423	4.906031	0.979	0.958
MR 3	0.958155	1.582792	13.69885	3.511849	0.987	0.975
MR 4	1.888250	2.234434	14.94115	4.632539	0.982	0.964
MR 5	1.245985	2.549386	15.53846	4.985446	0.978	0.957
MR 6	1.070825	2.062571	16.01300	4.264247	0.981	0.962
MR 7	2.544850	7.454924	21.89400	9.089055	0.984	0.969
MR 8	1.095735	2.018304	15.13800	4.314465	0.981	0.962

5. 결론(Conclusions)

본 연구에서는 PAC 주입률 결정을 위하여 원수 탁도, 원수 pH, 원수 온도, 응집후 탁도, 원수 알칼리도를 이용한 신경망을 구성하여 훈련 및 예측을 하였다. 또한 PAC 응집제 주입률을 결정에 신경망 모델의 적용 타당성을 보였으며, 다음과 같은 결론을 얻었다.

먼저, 비선형 다변수적인 특성을 가진 PAC 주입률 결정에 신경망(Neural Network)을 이용하여 결정할 수 있음을 확인하였고, Jar-Test결과와 비교할 때 평균오차율이 5%내외로 낮은 오차율을 나타내었다. 또한, 신경망에 의한 예측과 다중회귀(Multi-Regression)로 예측한 결과 신경망에 의한 예측이 다소 우수한 것으로 판단된다.

인공지능의 한 종류인 신경망(Neural Network)을 상수처리 과정 중 하나인 응집공정에 적용하였다. 하지만 많은 불확실성과 인위적인 문제점을 해결하기 위해서는 신경망 뿐만 아니라 퍼지(Fuzzy)알고리즘과 유전(Genetics)알고리즘에 대한 심도 있는 연구가 있어야 한다고 판단된다.

참고문헌(References)

1. 문병석, "신경망을 이용한 상수도 일일 급수량 예측", 전남대학교 토목공학과 박사학위 논문, 1998. 8.
2. 김강석, "상수처리용 응집제 PAC와 LAS의 특성비교 및 주입률에 관한연구", 전남대학교 토목공학과 석사학위 논문, 1997. 8.
3. 한국수자원공사, "응집제 투입량 자동결정 시스템 개발연구(2차년도)", 1997. 12.
4. Sekita, I., Kuraita, T., Asoh. H. an Chiu, D., "Reconfigurating Feed forward Networks with Fewer Hidden Nodes", Proc. of SPIE INF. on Adaptive and Learning System, 1993.
5. Warren S. Sarle, "Stopped Training and Other Remedies for Overfitting", To appear in Proceedings of the 27th Symposium on the Interface, pp.1~10, 1995.
6. Tin-Yau Kwok and Dit-Yan Yeung, "Use of Bias Term in Projection Pursuit Earning Improves Approximation and Convergence Properties", IEEE Transaction on Neural Networks, Vol. 7, No.5, pp. 1168~1182, 1996.