

# 원수조건에 따른 상수도 응집제 종류와 주입량 결정을 위한 데이터 마이닝 적용

## Application of Data Mining for Coagulant Dosage of Water Treatment Plants Corresponding to Input Conditions

배현, 김성신, 최대원, 이승태, 김예진\*

부산대학교 전기공학과

\*부산대학교 환경공학과

Hyeon Bae, Sungshin Kim, Dae-Won Choi, Seung-Tae Lee, and Yejin Kim\*

Department of Electrical Engineering, Pusan National University

\*Department of Environmental Engineering, Pusan National University

E-mail : sskim@pusan.ac.kr

### 요 약

본 논문은 정수장에서 사용하는 응집제의 종류와 주입량을 결정하기 위한 시스템 개발에 관한 내용이다. 정수장은 여러 단위 처리장으로 구성되며, 탁도와 색도를 제거하기 위하여 혼화지에서 응집제를 주입하여 침전을 시킨다. 현재까지 응집제 결정을 위해 Jar-test를 이용하였는데, 이 방법은 사람의 주관적인 판단에 의존하므로 실험 오차가 발생할 수 있다. 특히 정수장의 자동화를 위한 시스템 개발에서 가장 큰 걸림돌로 작용하고 있다. 본 논문은 이러한 문제점을 해결하기 위하여 데이터 마이닝 기법 등을 이용한 응집제 종류와 양을 결정하는 제어기를 개발하였다.

### 1. 서론

일반적인 정수처리의 공정은 응집, 침전, 여과, 살균소독 처리 과정을 거치며, 약품 주입에 의한 응집, 침전 및 살균소독 처리는 상수처리 시스템의 가장 핵심 부분을 이룬다. 응집공정에서 쓰이는 응집제의 종류로는 여러 가지가 있지만, 현재 덕산 정수장에서는 PAC, PASS, PSO-M의 3가지 약품을 주로 쓰고 있다. 응집제의 선택과 주입을 결정은 원수를 Jar-test를 토대로 전문가의 경험적 지식에 의존한다. 그러나 실험을 위해 2시간 이상이 소요되므로 수시로 변하는 원수의 상황에 적절하게 대처하기 어렵다[1].

실제 정수처리에 있어서의 응집효과는 탁도, 유발물질의 양과 종류, 입자의 크기, 특정 이온의 존재여부, pH, 알칼리도 등에 영향을 받게 된다[2]. 이에 본 연구에서는 덕산 정수장의 일별수질 데이터를 참조하여 응집제 선택에 영향을 미치는 입력인자(pH, 탁도, 알칼리도, Chl-a, 수온)에 대한 규칙을 의사결정나무(decision tree)를 이용하여 추출하고, 응집제 양을 신경망으로 구현하여 응집제 주입량을 제어하고자 하였다.

응집제 주입에 대한 연구로 Sugeno 등은 운전자의 경험을 바탕으로 퍼지 규칙을 구성하여 제어하였다[3]. Baba 등은 운전지원시스템에 퍼지 제어를 위한 입력으로 응집결과를 영상으로 분석한 기능을 추가하였으며[4], Enbutsu 등은 퍼지 제어규칙을 만드는 과정에 신경망의 학습이론을 도입하여 개선된 퍼지 규칙 추출방법을 제안하였다[5]. 국내의 응집제 주입에 관한 연구를 보면 뉴로-퍼지(neuro-fuzzy)모델이나 신경망 모델을 적용한 사례들이 있다[6].

본 논문에서는 응집제 제어에 데이터 마이닝 기법을 적용하여 원수수질에 대한 대처능력을 높이고, 응집공정을 자동화함으로써 전체적인 정수공정의 경제적 효율을 높이고자 하였다.

### 2. 정수처리공정

#### 2.1 정수공정

상수처리설비는 하천수의 취수구로부터 공급지로 급수하기까지의 정수공정을 포함한다. 정수처리 시설은 원수의 수질이 악화되더라도 음용수의 수질기준에 적합한 정수를 생산, 공급할 수 있는

기능을 갖춰야 한다.

그림 1은 정수장의 공정그림으로 취수, 오존처리, 혼화, 침전, 모래여과, 활성탄여과 순으로 처리하게 된다.

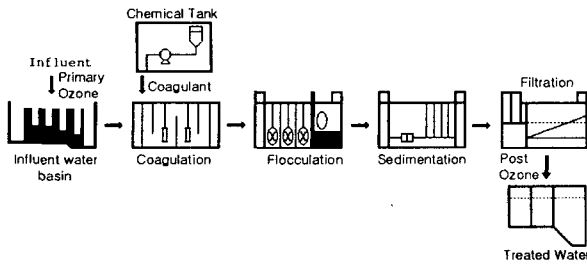


그림 1. 정수처리 공정 개요.

### 2.2 응집제 종류

응집제로는 제 2세대의 응집제로 황산반토의 뒤를 이어 현재 세계적으로 앞서가는 제품 중의 하나인 PAC(Poly Aluminium Chloride), 추운겨울에 응집성능이 뛰어난 PASS(Poly Aluminium Sulfate Silicate), 그리고 갈수기 조류로 인한 pH 상승 시 pH 강화효과가 있는 PSO-M(Poly Organic Aluminum Magnesium Sulfate) 등이 널리 사용된다. 각 응집제가 가지고 있는 장·단점이 있으므로 원수의 성상에 따라 선택하여 사용한다.

### 3. 데이터 마이닝

데이터분석 기법들은 회귀분석, 다변량분석 등과 같은 전통적인 통계기법, 의사결정나무, 의사결정규칙 등과 같은 기계학습 기법, 역전파, 코호넨 신경망 등의 인공신경망기법 등 크게 세 그룹으로 나뉜다. 이러한 다양한 데이터마이닝 기법 중 적요 데이터에 적절한 방법을 찾는 것은 노하우이며 데이터의 특성을 잘 분석하여 결정하여야 한다. 본 논문에서는 사전분석을 토대로 하여 분류를 위해 많이 사용되는 신경망과 의사결정나무를 적용하여 응집제의 종류와 양을 결정하는 제어를 개발하였다.

### 4. 실험 및 결과

#### 4.1 원수데이터 특성

본 연구에서 사용한 데이터는 덕산 정수장에서 취득한 것으로 응집제 선택은 수질데이터의 pH, 탁도(Tu), 알칼리도(Al)에 따라 결정되며 pH와 알칼리도(Al)는 참조자료로 쓰인다. 탁도가 투입량에 큰 비중을 차지하는데 50NTU 이상에서는 PAC를 사용하나 이하에서는 조류가 심할 때는 PASS를 사용하고, 그렇지 않을 때는 PSO-M을 사용한다. PASS와 PSO-M은 가격차이는 거의

없고, 조류나 알칼리도에 따라서 차이가 있을 수 는 있다.

#### 4.2 규칙도출을 위한 의사결정나무 적용

현장에서 응집제의 선택에 중요한 요소로 작용하는 pH, 탁도, 알칼리도, 클로로필, 수온을 의사결정나무의 입력으로 사용하였다. 그림 2는 원수 데이터를 보여주고 있으며 계절에 따라 특성이 변함을 알 수 있다. 그림 3은 실제 선택 투입된 약품의 종류와 양을 보여주고 있으며 계절에 따라 선택되는 응집제 종류가 다른 것을 알 수 있다. 특히 100일과 200일 부근에서 응집제 종류의 변화를 확연하게 볼 수 있다. 그림 4는 본 논문에서 제안한 응집제 종류와 양을 결정하는 시스템의 구성도를 보여주고 있다.

본 논문에서는 입력변수 5개와 출력변수 3개로 구성하였다. 출력변수 3가지는 덕산 정수장에서 사용하는 응집제의 종류로서 PSO-M, PASS, PAC로 출력 변수로 분류하였다.

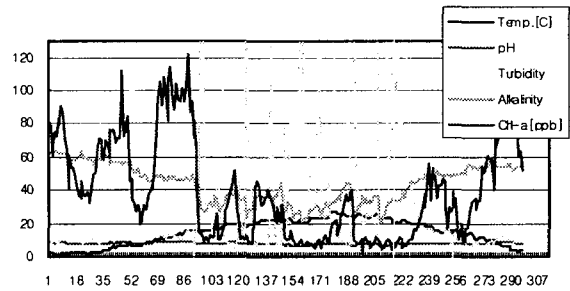


그림 2. 원수 입력데이터.

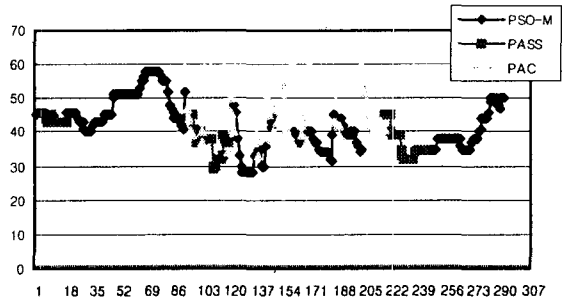


그림 3. 응집제 사용 특성.

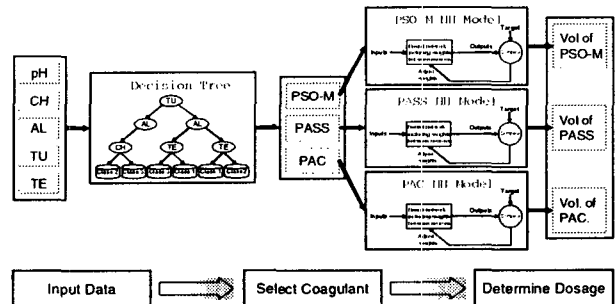


그림 4. 시스템 구성도.

4.3 의사결정나무 적용결과

의사결정나무의 가지치기(pruning)를 각각 25%와 70%를 주었고 트레이닝과 테스트의 데이터도 전체 데이터의 70%와 50%를 각각 사용하였다.

4.3.1 학습: 50%, 테스트: 50%, 가지치기: 25%

전체 데이터 중 50%의 데이터를 학습에 사용하여 실험하였다. 나머지 50%의 데이터로 구성된 모델의 성능을 평가하였다. 결과를 보면 학습보다 테스트 성능이 나쁜데 학습데이터가 충분하지 않음으로서 생기는 저학습된 사례다.

Read 147 cases (5 attributes) from water\_data\_50\_1.data

Decision tree:

```
TU > 29.5:
... AL <= 34: 3 (25/2)
... AL > 34:
... CH <= 27: 2 (6)
... CH > 27: 3 (2)
TU <= 29.5:
... TE > 20: 1 (24/1)
... TE <= 20:
... TE > 15:
... CH <= 61.5: 2 (16/1)
... CH > 61.5: 1 (2)
TE <= 15:
... TE > 2: 1 (58/2)
TE <= 2:
... TU <= 9.71: 1 (4)
... TU > 9.71: 2 (10/4)
```

그림 5. 데이터를 통해 추출한 규칙(가지치기 25%).

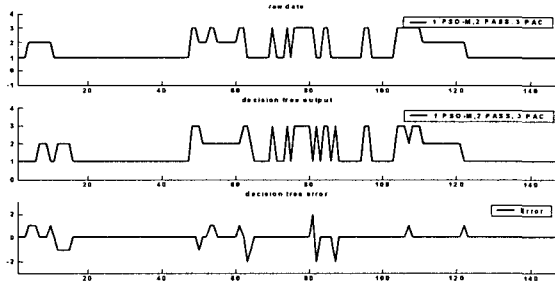


그림 6. 데이터, 결과, 오차 비교 그래프.

4.3.2 학습: 70%, 테스트: 70%, 가지치기: 70%

전체 데이터 중 학습데이터와 테스트데이터를 각각 70%:30%로 나누는 경우 성능이 더 우수하였다. 반면, 과도 학습이나 모델의 성능에 미칠 수 있는 다른 변수를 고려하기 위하여 가지치기의 비율을 함께 고려하여 실험을 수행하였다. 이전의 사례와 달리 70%의 가지치기 비율을 사용함으로써 보다 간소화된 규칙을 도출하였다. 결과에서 25%의 가지치기보다 70%의 가지치기 결과가 더 우수하였다. 이러한 결과는 모델이 과도 학습을 통해 성능이 나빠질 수 있음을 의미한다. 그리고 전체 변수를 규칙에 모두 사용하는 것보다 경우에 따라서는 특정 변수만을 사용하여 만든 규칙이 데이터의 특성을 잘 표현할 수 있다.

Read 206 cases (5 attributes) from water\_data\_all data

Decision tree:

```
TU > 29.5:
... AL > 36: 2 (7)
... AL <= 36:
... TU > 53: 3 (27)
... TU <= 53:
... CH > 11: 2 (3/1)
... CH <= 11:
... AL <= 27: 3 (3)
... AL > 27: 1 (7/3)
TU <= 29.5:
... AL <= 29:
... TE <= 22: 3 (4/1)
... TE > 22: 1 (2)
AL > 29:
... TE > 20: 1 (29)
... TE <= 20:
... TE > 16: 2 (18/2)
... TE <= 16:
... TE > 2:
... TE <= 15: 1 (80/3)
... TE > 15:
... PH <= 8.36: 2 (4)
... PH > 8.36: 1 (5)
TE <= 2:
... TU <= 9.02: 1 (3)
... TU > 9.02:
... TE <= 1: 2 (5/1)
... TE > 1:
... CH > 51.4: 2 (3)
... CH <= 51.4:
... PH <= 7.9: 2 (2)
... PH > 7.9: 1 (4)
```

그림 7. 데이터를 통해 추출한 규칙(가지치기 70%).

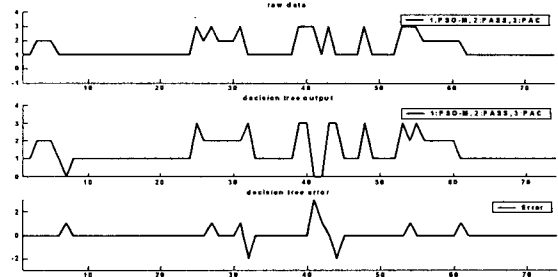


그림 8. 데이터, 결과, 오차 비교 그래프.

표 1. 규칙도출 및 실험 결과.

학습: 테스트	가지치기	오차	
		학습	테스트
70:30	25%	8.3	9.1
	<b>70%</b>	<b>5.3</b>	<b>8.0</b>
50:50	25%	6.8	17.0
	70%	5.4	13.6

4.4 신경망 모델구성 및 테스트 결과

의사결정나무(decision tree)에 의해 선택된 응집제에 따라 각 응집제의 주입율을 결정하기 위해 응집제 별로 신경회로망을 구성하였다.

4.4.1 응집제 PSO-M의 신경회로망구성 및 결과

응집제 PSO-M의 주입율을 결정하기 위한 신경회로망은 3층으로 구성하였고, 입력층, 은닉층, 그리고 출력층을 각각 3개, 4개, 1개의 노드로 구성하였다. 응집제 PSO-M이 사용된 총 186개의 데이터 중에 학습과 테스트를 위해 70%:30%로 나눠서 70%는 학습에 나머지 30%는 테스트에

사용하였다. 또한 186개의 데이터를 50%:50%로 나누어서 50%는 학습에 나머지 50%는 시뮬레이션을 해봄으로써 학습과 시뮬레이션 비율에 따른 정확성을 비교해 보았다. 그림 9는 추정한 결과와 실제 값을 비교하여 보여주고 있다.

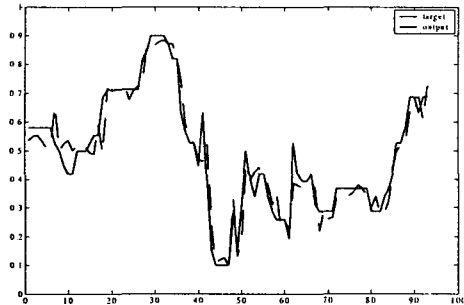


그림 9. PSO-M 50%학습, 50% 실험 결과.

#### 4.4.2 응집제 PASS의 신경회로망구성 및 결과

응집제 PASS의 주입율을 결정하기 위한 신경회로망은 2개의 입력층 노드, 1개의 출력층 노드로 구성하였다. 그림 10은 테스트한 결과를 실제 값과 비교하여 보여주고 있다.

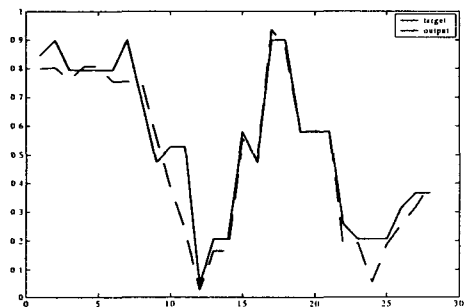


그림 10. PASS 50%학습, 50% 실험 결과.

#### 4.4.3 응집제 PAC의 신경회로망구성 및 결과

응집제 PAC의 주입율 결정에 대해서는 입력층, 출력층의 2계층으로 구성하였고, 입력층은 2개의 노드를 출력층은 1개의 노드로 구성하였다. 그림 11은 테스트한 결과를 실제 값과 비교한 것이다.

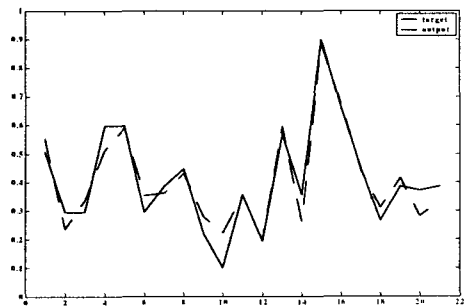


그림 12. PAC 50%학습, 50% 실험 결과.

표 3. 데이터 사용에 따른 최종 예측 결과 정리.

응집제	학습:테스트 데이터	RMSE
PSO-M	50%:50%	0.0058
	70%:30%	0.0100
PASS	50%:50%	0.0157
	70%:30%	0.0153
PAC	50%:50%	0.0118
	70%:30%	0.0156

## 5. 결론 및 향후과제

적절한 응집제와 주입율의 결정은 처리효율과 수질의 개선을 가져온다. 따라서 응집반응에 영향을 미치는 주요 수질인자의 변화에 신속하고 정확하게 응집제의 선택과 주입율 결정을 위한 방법으로 본 연구에서는 응집제 선택 시에는 의사결정나무를 이용하였으며, 응집제 주입율 결정을 위해서는 신경회로망을 사용하였다. 본 연구에서 얻어진 신경망 모델이 실제 정수처리 시설에서 응집제 주입율에 대한 현장 검증이 필요하다고 판단된다.

## 6. 참고문헌

- [1] 수자원연구소, 응집제 투입량 자동결정 시스템 개발연구, 한국수자원공사 기술보고서, pp. 14-17, 1977.
- [2] A. P. Black *et al.*, "Electrophoretic Studies of Turbidity removal Coagulant with Aluminum Sulfate," J. AWWA, vol. 53, p. 438, 1961.
- [3] Yagishita, O., Itoh *et al.*, "Application of fuzzy reasoning to the water purification process," Industrial Application of Fuzzy Control, pp. 19-39, 1985.
- [4] K. Baba *et al.*, "Intelligent support system for water and sewage treatment plants which includes a past history learning function-coagulant injection guidance system using neuralnet algorithm," IAWPRC, pp. 227-234, 1990.
- [5] I. Enbutsu, K *et al.*, "Integration of multi AI paradigms for intelligent operation support systems-fuzzy rule extraction from a neural network," IAWQ, pp. 333-340, 1993.
- [6] 이병조, "신경망 이론을 이용한 응집제 PAC의 주입율 결정 모델에 관한 연구," 전남대학교 석사학위논문, pp. 34-41, 1992.