

## XAI를 이용한 하수처리 활성슬러지공정 지능형 모델링에 관한 연구

**남익석\***  
극동대학교\*

### A Study on Intelligent Modeling of Activated Sludge Process in Wastewater Treatment System using XAI(eXplainable AI)

Eui-Seok Nahm\*  
Far East University\*

**Abstract** - 본 논문에서는 하수처리 활성오니공정의 신뢰성이 있는 데이터와 XAI(eXplainable AI) 기법 중 LRP(Layer-wise Relevance Propagation)를 적용하여 하수처리 활성오니공정 최적 모델링을 하고자 한다. 즉, 하수처리장 운전자들이 일상의 운영에서 정확하다고 인정하는 센서 데이터와 부정확하다고 하는 센서 데이터를 분류하고, 정확한 센서 데이터와 부정확한 센서 데이터 사이의 상관관계를 eXplainable Neural Network를 사용하여 학습하여 부정확한 센서 데이터에 영향을 미치는(설명 가능한) 정확한 센서 데이터를 찾아낸다. 이를 기반으로 Neural Network의 입출력을 정의하여 딥러닝을 수행하여 모델링을 하고자 한다.

중심의 접근 방법으로 원인에 기반한 제어보다는 단순 예측 결과에 기반하는 것으로, 이 경우 데이터 의존도가 높아 한계가 있다. 본 논문에서는 이러한 기존의 딥러닝을 활용한 공정 예측에 설명가능한 AI를 적용함으로써 단순히 예측 결과에만 의존하는 것이 아니라 특정 성능(KPI : Key Performance Index)을 달성하지 못한 공정에 대한 원인 분석을 통하여 공정을 최적으로 제어할 수 있는 모델링 기법을 제시하고자 한다.

## 1. 서 론

하수처리장은 주로 생물학적인 처리 방법인 활성슬러지공정을 주로 사용한다. 이 방법이 주로 사용되는 이유는 처리 비용이 싸고, 효율이 좋기 때문이다. 이 방법은 하수와 활성슬러지를 혼합시키고, 공기를 불어 넣음으로써 활성슬러지가 하수 속의 유기물을 섭취 분해하는 방식이다. 여기서 공기는 미생물인 활성슬러지를 번식시키는 요소로 공기의 양을 제어하여 투명한 처리수를 얻을 수 있다. 공기의 양을 제어하는 방법은 주로 송풍기를 사용하여 하수 속에 공기를 주입하는 방식으로, 송풍기에 소모되는 전력량이 하수처리장 운영비에 큰 비중을 차지한다.(하수처리 총 운영유지비의 약 21%가 전력비, 전력비 중 약 54~60%가 송풍기 운영에 사용).[1]

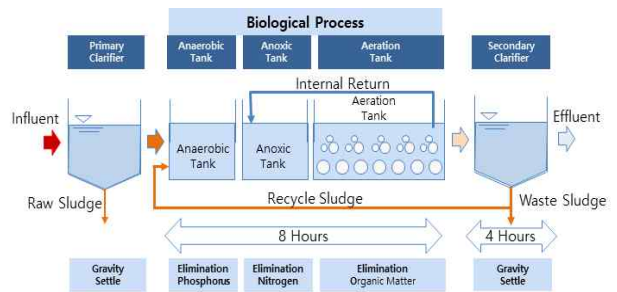
최근에는 수질 기준 강화에 따라 효과적인 공정제어를 통한 수질 개선과 더불어 경제적인 이유로 공정의 효율 개선을 위한 다양한 노력들이 시도되어 운영되고 있다. 즉, 하수처리장의 운영비 개선을 위한 전력비 절감에 대한 많은 연구가 진행되고 있다. 특히, 해외에서는 이를 위하여 다양한 지능형(AI)기법들이 개발되어 적용되고 있다. 그러나, 국내에서는 하수처리장 공정제어시스템의 1) 하수처리장의 특성(토목 구조와 크기, 유입수 수질, 센서 설치 위치 등)에 따라 최적화 방안이 달라서 운전자의 경험에 근거하여 운영하고, 2) 생물학적 공정의 특성으로 피드백의 시간 지연이(보통 1~6시간의 반응 시간) 하수처리장 별로 달라서 일반적인 제어 이론의 적용이 어려워 수질과 경제적 운영에 영향 미치는 요소(제어량)의 판단이 어려운 문제점 등으로 이러한 일반화된 선진 AI기법 적용이 어려운 실정이다. 즉, 유출수 수질 개선을 위해 유입수의 상태에 따라 적절한 공기의 양을 결정하고 이에 따라 송풍기를 가동하여야 하나, 보통의 경우 수질의 개선을 위해 과도한 공기의 양을 주입함으로 전력이 낭비되는 경우가 많다. 따라서, 활성슬러지공정의 최적화를 위해서는 반드시 해당 하수처리장의 특성을 반영하여 유출수 수질에 영향을 미치는 유입수의 수질 항목의 결정이 중요하고 그에 따른 최적의 제어량의 결정을 위해 모델링이 중요하다.[1]

따라서, 본 논문에서는 이러한 하수처리 활성슬러지공정의 유출수 수질에 영향을 미치는 유입수의 수질과 제어량을 구하기 위해 XAI(eXplainable AI) 기법 중 LRP(Layer-wise Relevance Propagation) 기법을 적용 비교하여 최적의 모델링을 하고자 한다. 기존의 공정을 분석하여 그 원인을 파악 방법은 주로 예측

## 2. 제안된 기법

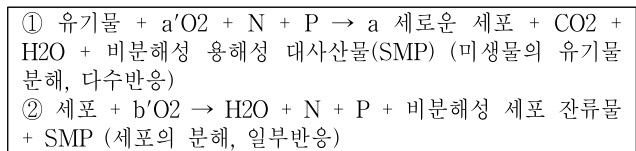
### 2.1 활성슬러지 공정

하수처리 활성슬러지공정 과정은 그림 1과 같다.



**<그림 1> 하수처리시스템의 활성슬러지공정**

하수처리시스템의 활성슬러지공정 일차침전에서는 침강성 고형물을 제거하는 과정으로 물의 체류시간을 길게 늘여서 침강성 고형물이 아래로 퇴적되게 한 뒤, 상등수(위에 뜬 물)를 분리하여 다음 공정으로 보낸다. 반응조(포기조)에서는 활성슬러지(미생물로 구성된 슬러지)가 물 속의 유기물질을 섭취하며 저분자 물질로 분해하는 과정이다. 이 과정에서 분해될 물질 중의 탄소 성분은 대부분 이산화탄소가 되거나 세포에 잔류하게 되며, 질소와 인이 일부 소비된다. 반응식은 다음과 같다.



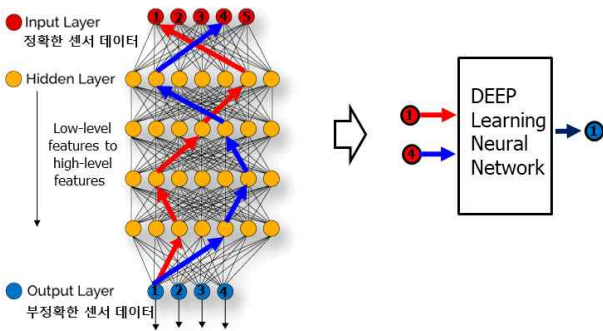
이 반응은 자연계에서의 자정 작용이다. 공정에서는 단지 인위적으로 포기(공기를 물 속에 주입)해주는 방식으로 활성슬러지를 만들고, 이 활성슬러지가 용존산소를 이용해서 유기물을 분해한다. 즉, 인위적 공간에서 짧은 시간 안에 다량의 물질을 처리하게 만드는 것이다. 즉, 공기를 적절하게 주입하는 방식으로 활성슬러지의 공정을 제어한다. 이 공정의 감시와 제어를 위해 측정되는 보편적인 수질 센서 데이터와 신뢰도는 표 1과 같다.

**[표 1] 활성슬러지공정 수질 센서 데이터 및 신뢰도**

센서 데이터	센서 및 신뢰도			
	유입수	반응조	유출수	신뢰도
COD(화학적산소요구량)	○		○	상
TN(총 질소)	○		○	상
TP(총 인)	○		○	상
PH	○	○	○	상
SS(Suspended Solid)	○		○	상
DO(용존 산소)		○		하
NO <sub>3</sub> (질산염)		○		중
NH <sub>4</sub> (암모늄)		○		중
MLSS(혼합부유물)		○		중

여기에서, 가장 정확한 데이터는 이차침전지 방류수의 COD, TN, TP, SS이다. 이는 법적으로 SS, COD, TN, TP, PH, 온도, 방류량 7가지 센서 데이터를 실시간으로 한국환경관리공단 관계 시스템으로 전송되어지며 기준치 초과시 벌금을 부과하기 때문이다. 특히, 이중 DO의 경우 실제적인 제어 데이터이고, 하수처리장 운영비와 직결되는 송풍기와 연관된 데이터로 매우 중요하다. 방류수의 SS, COD, TN, TP를 기준치 이하로 유지하기 위해 일반적으로 DO를 높여야 하는데 이 경우 송풍기의 송풍량을 증가시키는데 불필요한 전력이 낭비되어 실제적인 운전에서는 매우 민감한 제어 항목이다. 또한, 하나의 대응량 송풍기에 다수의 반응조(포기조)가 연결되어 있고, 어느 하나의 반응조의 DO값을 올리기 위해 송풍량을 증가하면 모든 반응조가 동일하게 DO값이 상승해야 하나, 실제적으로는 각 반응조마다 DO값의 상승이 다르게 나타난다. 따라서, 불필요하게 과도하게 송풍기를 가동하는 경우가 자주 발생한다.[2]

따라서, 이러한 내용들을 기반으로 방류수의 SS, COD, TN, TP를 사용하여 DO에 영향을 미치는 입력 센서 항목을 LRP XAI 알고리즘으로 분석하여 DO에 영향 미치는 정도가 큰 입력 변수를 구하고, 이를 다시 모델의 입력변수로 하여 모델링 한 후 성능을 비교한다. 센서 데이터의 유효성의 범위는 법적 기준치인 ±3%이고, 머신러닝 모델의 최대 오차가 ±5% 점을 고려하면 최대 ±8% 오차가 발생할 수 있으므로, 사용되는 머신러닝 공정 모델의 유효성(수처리 공정의 경우) ±10% 오차 범위가 되어야 한다.[3] 전체적인 모델링 방법은 그림2와 같다.



**<그림 2> Neural Network Modeling Scheme**

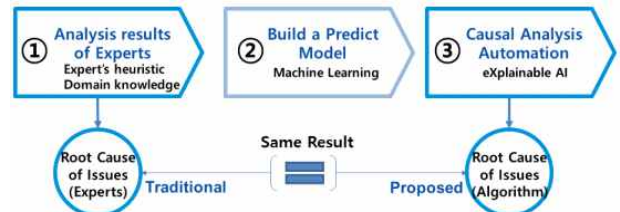
**2.2 XAI를 이용한 모델링**

XAI는 예측된 결과에 대해 원인을 제공하는 것으로 머신러닝 모델 예측에 영향을 미치는 요소를 찾아 예측 결과에 대해 이해할 수 있게 하는 것으로 모델 자체를 해석하는 Interpreting Model과 원인에 대해 데이터로 설명하는 Explaining Decision으로 구분된다.

본 논문에서는 LRP(Layer-wise Relevance Propagation : 뉴런의 가중치를 이용하는 방법으로 딥러닝에 특화된 방식) 알고리즘을 수도권 하수처리장 3년치 시간데이터에 적용하여 최적 모델링을 하고자 한다.

그림 3은 LRP 적용하는 전체적인 방법을 보여준다. 본 논문은 머신러닝 알고리즘을 이용하여 공정 결과를 예측하고, 이 모

델을 XAI(LRP) 알고리즘으로 분석하여 영향력이 큰 입력변수를 선정하여 다시 모델링하여 오차를 비교 분석한다.



**<그림 3> LRP 적용 모델링**

**3. 시뮬레이션 결과**

DO에 영향력이 큰 입력 변수를 구하기 위하여 활성슬러지공정의 유입수 수질값을 입력으로 하고 출력을 DO로 하는 설명가능한 신경회로망을 구현한다. 입력으로 RRSP(Return Sludge Ratio Set-Point), WSR(Waste Sludge Ratio), COD(Chemical Oxygen Demand) SS(Suspended Solid), MLSS(Mixed Liquid SS), NH<sub>3</sub>-N, 온도, pH를 입력으로 하고 DO를 출력으로 한다. DO에 변화에 크게 영향 미치는 입력변수를 찾기 위해 본 논문에서는 이를 해결하기 위해 모델의 오차의 변화율이 10%되는 시점부터의 LRP를 적용하여 각각의 입력변수에 대해 순위를 정하여 순위가 높은 것이 DO에 미치는 영향이 큰 것으로 판단하였다.

수도권 하수처리장의 3년치(3년×356일×24시=26,280 Set) 시간 데이터 세트를 사용하여 시뮬레이션한 결과 랭킹이 COD=1, SS=2, NH<sub>3</sub>-N=3, 온도=4, pH=5, RRSP=6, WSR=7로 산출되었다. 따라서, 본 논문에서는 COD, SS, NH<sub>3</sub>-N, 온도 4개만 DO에 미치는 영향이 큰 것으로 하여 신경회로망 모델을 구축하여 모델링을 수행하였다. 26,280개 데이터 중 절반으로 모델의 성능을 평가한 결과 최소과승오차는 0.25, 나머지 절반의 데이터를 이용하여 모델을 평가한 결과 오차는 0.35로 계산되었다. 기존의 방식대로 입력변수를 모두 사용한 경우의 모델 오차는 0.44로 제안된 기법이 효율적임을 알 수 있었다.

**4. 결 론**

본 논문에서는 포기조 미생물에 필요한 최적 용존산소농도를 산출하여 최적의 수질 유지와 공기 과도 주입으로 인한 송풍기 전력비를 최소화할 수 있는 최적 용존산소 제어(DO Control) 모델을 제안하였다. 최적 제어를 위하여 실시간 미생물 호흡률에 영향 미치는 주요 입력 수질 항목을 설명가능한 신경회로망 기법을 이용하여 찾아내고, 찾아낸 입력 수질을 입력으로 한 활성슬러지공정 신경회로망 모델을 구현하고, 이로부터 최적의 용존산소 농도를 산출한다.

시뮬레이션 결과 COD, SS, NH<sub>3</sub>-N, 온도 4개가 DO에 미치는 영향이 큰 것으로 결정되어, 이를 입력변수로 하여 신경회로망 모델을 구축하여 모델링을 수행한 결과 모델링 오차 0.25, 평가용 데이터를 이용한 경우 오차는 0.35, 기존 방법에 의한 모델링 오차는 0.44로 제안된 기법이 효율적임을 알 수 있었다.

**[참 고 문 헌]**

[1] 남의석, “보정 기법을 이용한 하수처리 활성슬러지 공정의 입출력 데이터 유효성 검증 및 전력절감에 관한 연구”, 전기학회논문지, 69권1호, 177-183, 2020  
 [2] Ahn S. Y, “Analysis of electric consumption cost by dissolved oxygen control in aeration tank of sewage treatment plant”, Monthly Environmental Engineer, 40-42, 2002  
 [3] 남의석, “신경회로망을 이용한 하수처리시스템 활성슬러지공정 센서 데이터 상호작용 분석”, 대한전기학회 학술대회 논문집, 2020년도 제 51회 대한전기학회 하계학술대회, 1,850 - 1,851, 2020