

AI 기반 복합센서 사전검증시스템을 활용한 약품투입량 분석 및 예측모델 개발

성민석* · 김국일 · 안상병 · 홍성택
한국수자원공사

Development of Drug Input Analysis and Prediction Model Using AI-based Composite Sensors Pre-Verification System

Min-Seok Seong · Kuk-Il Kim · Sang-Byung An · Sung-Taek Hong
K-water Research Institute

E-mail : shaelryn@kwater.or.kr / kikim@kwater.or.kr / sban11@kwater.or.kr / sthong@kwater.or.kr

요 약

수돗물 생산·공급의 안정성을 확보하기 위하여 AI 기반 복합센서를 실증지인 정수장에 적용하기 전에 사전검증할 수 있는 시스템을 구축하였으며, 2019년 12월부터 2021년 12월까지 약 2년간의 고령정수장 생활용수의 약품투입량 관련된 데이터를 수집·분석하여 약품투입량 예측모델을 개발하고자 하였다. 이상치 제거와 파생변수 생성 등 데이터 전처리를 통해 각 Tag의 이상치를 제거하고 1분 주기 데이터를 60분 구간 평균 데이터로 주기를 설정하고 PLS 모형을 사용하여 모델을 학습시켰으며, 각 모델들의 예측 정도를 비교·검증하였다.

ABSTRACT

In order to secure the stability of tap water production and supply, we have built a system that can be pre-verified before applying AI-based composite sensors to the water purification plant, which is a demonstration site. We have collected and analyzed data related to the drug input of the GO-RYEONG water purification plant for about two years from December 2019 to December 2021. The outliers of each tag were removed through data preprocessing such as outliers and derived variable, and the cycle was set as average data for 60 minutes of each one-minute period, and the model was learned using the PLS model.

키워드

정수장, 복합센서, 인공지능 사전검증 (WTP, Complex Sensor, Artificial Intelligence, Pre-verification)

I. 서 론

AI 기반 복합센서를 실증지인 고령정수장에 적용하기 전 단계로 시뮬레이터를 구축하였으며, 2019년 12월부터 2021년 12월까지 약 2년간의 고령정수장 생활용수의 약품투입량 관련된 7개의 태그에 대한 데이터를 수집하고 분석하였으며, 이를 토대로 약품투입량 예측모델을 개발하고자 한다.

II. 데이터 전처리

데이터 전처리를 위해 표 1과 같이 이상치 제거와 파생변수 생성, 분석용 데이터 셋 수집주기 설정, 실 사용 데이터를 확인하며, 각 Tag의 이상치를 제거하고 1분 주기 데이터를 60분 구간 평균 데이터로 변경하여 데이터 통계량 비교 및 이를 바탕으로 데이터 주기를 설정하였으며, 트렌드 차트와 매트릭스 GL 차트를 확인하여 여타 수치와 눈에 보일 정도로 거리가 먼 경우 또는 상식적으

* speaker

로 불가능한 수치를 나타내는 경우 이상치로 판단하여 제거하였다.

표 1. 데이터 전처리

No	전처리 내용	상세 업무 내용	
1	이상치 제거		
	A	PAC생활약품유량	이상치 제거 및 정상 범위 설정
	B	유입수_탁도_생활_원수_탁도	
	C	유입수_pH	
	D	유입수_온도	
	E	유입수_알칼리도	
	F	유입수_전기전도도	
	G	원수유입량_순시	
2	파생변수 생성		
	A	편차에 따라 구간 분류	트렌드 차트 내에서 그래프 편차에 따라 시간 별 구분
	B	10분 구간 평균 데이터, 60분 구간 평균 데이터 생성	1분 주기 데이터의 주기 재설정
3	분석용 데이터 셋 수집주기 설정		
	A	주기 별 평균 데이터 통계량 비교	각 주기 데이터를 비교하여 분산, 표준편차 등을 비교
	B	분석용 데이터 셋 수집 주기 설정	비교한 데이터를 바탕으로 데이터 주기 설정
4	실 사용 데이터 확인		
	A	단위 통일	생활원과 PAC생활약품유량 (ml/min) 단위를 로 통일

III. 탐색적 분석

약품투입량 트렌드는 그림 1과 같으며, 평균 약품 투입량은 27.28이며, 2020년(투입량 24.95)에 비해 2021년(투입량 30.09)에 더 투입되었으며, 하절기에 약품 투입량 변동이 심함을 알 수 있다.

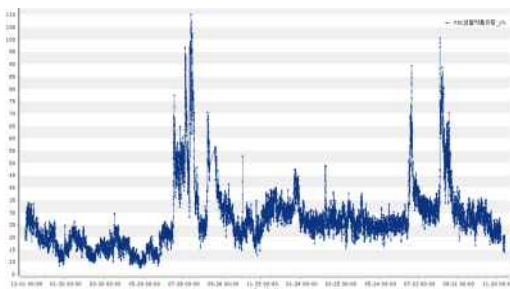


그림 1. 약품투입량 트렌드

약품투입량과 타 인자 연관성 분석은 그림 2와 같으며, 약품투입량은 원수유입량_순시와 유입수_탁도_생활_원수_탁도, 유입수_온도, 원수 유입량_순시는 양의 관계를 나타내며, 유입수_전기전도도, 유입수_알칼리도와는 음의 관계를 나타낸다.

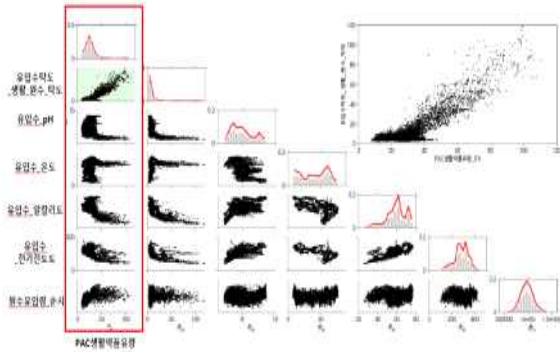


그림 2. 약품투입량 vs. 타 인자 연관성 분석

분석 전체 기간에 대한 약품투입량 평행좌표 GL chart는 그림 3과 같으며, PAC 약품 투입과 관련하여 탁도, 원수유입량이 양의 상관관계를, 전기전도도가 음의 상관관계를 보인다.

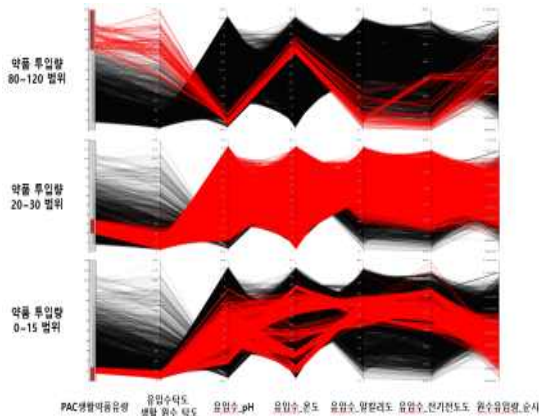


그림 3. 약품투입량 평행좌표 GL chart(전체기간)

분석 전체 기간에 대한 유입수_탁도와 원수 유입량_순시를 비교해 보면 그림 4와 같으며, 약품투입량과 유입수_탁도의 경우 유사한 트렌드를 보이거나 유입량의 경우 두 태그와 비교할 시 뚜렷한 유사점이 보이지 않는 것으로 분석되었다.

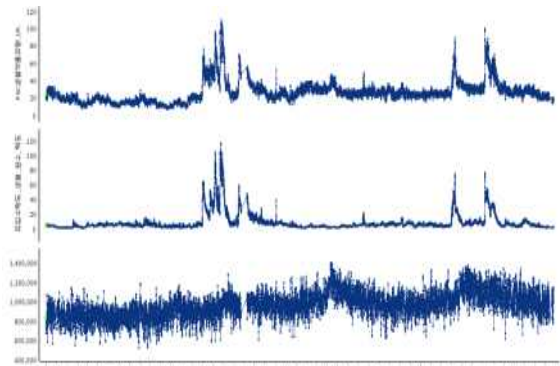


그림 4. 유입수_탁도와 원수 유입량_순시를 비교

III. 예측 모델

약품유량을 예측하기 위하여 표 2와 같이 PLS 모델을 사용하였으며, 각 모델들의 예측 정도를 비교 검증하였다[1].

표 2. 24시간 PLS 모델링 개요

	PLS 모델링	상세 업무 내용
A	데이터 24시간 데이터 셋 및 PLS용 데이터셋 생성	24시간 데이터 셋과 이를 활용하여 PLS 분석용 데이터 셋 생성
B	PLS분석으로 익일 수요 예측량 계산	PLS모델을 활용하여 익일 수요 예측량을 계산
C	예측값 검증	관측 데이터와 수요예측값비교

기존 데이터 셋을 그림 5와 같이 하루 전체 수요예측을 위해 24시간으로 하루 전체 값을 더하여 활용하고, 24시간 주기로 변경한 데이터 셋에서 직전 변를 종매칭하여 미래값을 도출하도록 학습시키고 학습한 모델을 바탕으로 Y_HAT 예측값을 도출하였다.

부식량	부식량	부식량	부식량	부식량	부식량	부식량	부식량	부식량	부식량
4차정수장생수(10, 30, 50)	부식량	부식량	부식량	부식량	부식량	부식량	부식량	부식량	부식량
공급수(10, 30, 50)	부식량	부식량	부식량	부식량	부식량	부식량	부식량	부식량	부식량
공급수(10, 30, 50)	부식량	부식량	부식량	부식량	부식량	부식량	부식량	부식량	부식량
공급수(10, 30, 50)	부식량	부식량	부식량	부식량	부식량	부식량	부식량	부식량	부식량
공급수(10, 30, 50)	부식량	부식량	부식량	부식량	부식량	부식량	부식량	부식량	부식량

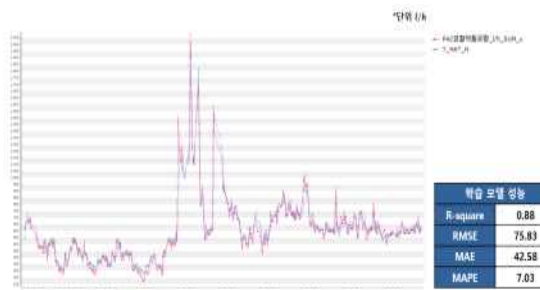
24시간 주기 기술통계

시간 t-1	공급변수 Y _t	독립변수 X _{t-1}	부식량	부식량	부식량	부식량	부식량	부식량	부식량
1	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
2	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
3	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
4	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
5	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
6	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
7	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
8	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
9	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
10	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
11	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
12	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
13	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
14	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
15	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
16	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
17	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
18	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
19	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
20	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
21	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
22	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
23	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000
24	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000	100.0000

생성된 PLS 예측용 데이터 셋

그림 5. 분석 데이터 셋 생성

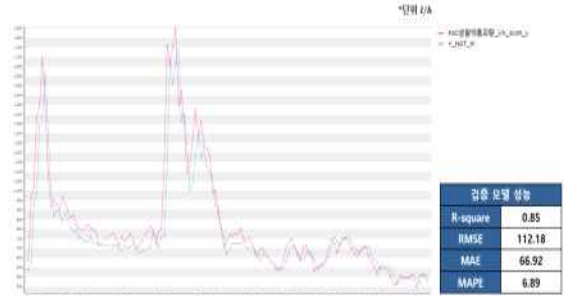
데이터 전체 수량 중 80%를 선택하여 모델을 학습시킨 결과는 그림 6과 같으며, 2019년 12월 1일~2021년 4월 27일 데이터를 추출하여 PLS 모델을 학습한 결과이다.



모델 학습 예측값과 실측값 비교 그래프

그림 6. 모델 학습 : 24시간

데이터 전체 수량 중 20%는 학습시켜 나온 결과 모델 노드로 정확도를 그림 7과 같이 확인하였으며, 2021년 4월 28일~2021년 11월 24일 데이터를 추출하여 PLS 모델을 검증한 결과이다.



검증 모델 예측값과 실측값 비교 그래프

그림 7. 모델 검증 : 24시간

IV. 결론

본 논문에서는 AI 기반 복합센서 사전검증시스템을 활용하여 고령정수장 생활용수의 약품투입량 관련된 데이터 수집 및 분석을 통해, PAC생활·공업약품 유량, 공업·생활 급수 유량, 취·정수장 전력량 등을 예측하는 데이터 모델을 개발하였다. 각 분석 모델별 통계 데이터를 시각화하여 효율적인 운영 및 관리로 인한 수돗물 생산·공급의 안정성을 확보하고자 하였다.

Acknowledgement

이 논문은 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP)의 지원에 의한 No. 2020200000010.

References

[1] S. Y. Kim, and Bomsock Lee, "Development of the Adaptive Predictive Model for product qualities in chemical processes using PLS method," *Theories and Applications of Chem. Eng.*, Vol. 11, No. 2, pp. 1813-1816, 2005.