

AI 기반 분리 모형을 활용한 구례취수장의 용수수요량 예측에 관한 연구

노희성¹·최시중¹·강성규¹·김동현^{2*}¹한국건설기술연구원 수자원하천연구본부 수석연구원^{2*}인하대학교 수자원시스템연구소 연구원

A Study on Water Demand Prediction of Gurye Intake Station Using AI-Based Classification Models

Hui-Sung Noh¹ · Si-Jung Choi¹ · Sung-Kyu Kang¹ · Dong-Hyun Kim^{2*}¹Senior Researcher, Department of Hydro Science and Engineering Research, Korea Institute of Civil Engineering and Building Technology, Goyang, Korea^{2*}Researcher, Department of Water Resources System, Inha University, Incheon, Korea

[요 약]

가뭄은 사회 및 환경에 심각한 영향을 끼치고 있지만, 발생 지역이 포괄적이고 점진적으로 진행된다. 이에 가뭄으로 인한 피해를 사전에 대비 및 대응 차원에서 가뭄 단계에 따라 탄력적으로 적용할 수 있는 비구조적인 대책의 수립이 필요한 실정이다. 본 연구에서는 위기경보 단계별 기준을 마련하기 위해 용수수요량 자료의 확률밀도함수와 구간에 확률변수가 포함될 확률을 산출하였다. 설정된 기준을 바탕으로 가까운 미래의 용수수요량의 규모를 파악하고자 AI 기반의 분류 모형인 의사결정나무(decision tree, DT), 랜덤포레스트(random forest, RF) 모형을 이용하였다. 모형의 학습 구간은 '04년부터 '15년까지 검증 구간은 '16년부터 '21년까지 자료를 이용하여 모형의 정확성을 평가하였다. 모형의 정확성 평가 결과, 랜덤포레스트 모형의 F1-score 값이 0.81로 우수한 성능을 나타냈다.

[Abstract]

Drought has a serious impact on society and the environment, but it is a comprehensive and gradual process in the affected area. It is necessary to establish non-structural measures that can be flexibly applied according to the drought stage in order to prepare for and respond to damage caused by drought in advance. In this study, the probability density function of water demand data and the probability that a random variable is included in the interval are calculated to prepare the standard for each stage of crisis alert was established. AI-based classification models such as decision tree (DT) and random forest (RF) models were used to figure out the size of water demand in the near future based on the established criteria. The model's learning section was from 2004 to 2015, and the validation section was from 2016 to 2021, and the accuracy of the model was evaluated. As a result of evaluating the accuracy of the model, the F1-score of the random forest model was 0.81, indicating excellent performance.

색인어 : 가뭄, 랜덤포레스트, 비구조적 대책, 용수수요량, 의사결정나무**Keyword** : Drought, Random forest, Non-structural measures, Water demand, Decision tree<http://dx.doi.org/10.9728/dcs.2022.23.12.2545>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 24 November 2022; Revised 15 December 2022

Accepted 15 December 2022

*Corresponding Author; Dong-Hyun Kim

Tel: +82-32-876-9783

E-mail: yesdktpdi@naver.com

1. 서론

기후변화 및 이상기후 등으로 인해 증가하는 가뭄 위협에 대비하고자, 기존 다목적댐 및 용수전용댐 뿐만 아니라 취수장의 이·취수 활용에 대한 기대가 높아지고 있다(Jeong et al., 2022). 특히, 영상·점진강 수계에서는 용수수요량으로 인해 가뭄시에도 안정적으로 용수를 공급하기 위한 다양한 노력이 계속되고 있다. 장기적으로 국가 차원의 가뭄 및 물 부족 대응을 위해 「수자원장기종합계획」을 수립하며, 전국 유역을 대상으로 수요량 및 공급량을 예측해 물 부족량을 산정하고 있다. 그러나, 이는 국가하천 및 다목적 댐 등에 초점이 맞춰져 있으며, 지방하천 및 소하천 등은 상대적으로 미진한 실정이다(Kim et al., 2021; Jang et al., 2021).

용수수요량 예측은 정수장의 정수 생산계획과 송수펌프의 운전계획 및 배수지의 운영계획 등을 결정하는데 중요한 정보를 제공한다. 따라서, 예측된 값을 운영에 적절하게 이용한다면 생산 및 송수비용을 절감할 수 있다. 하지만, 정확도 높은 용수수요량 예측이 이루어지지 않을 경우, 정수장에서 배수지로 과량의 물을 송수하여 비효율적인 펌프운전이 이루어지게 된다. 이로 인해, 전력을 과다하게 사용하게 되고 과다공급으로 인한 배수지 내의 수위조절이 적절하게 이루어지지 않는 문제를 초래할 수 있다(Global Water Intelligence, 2011; Kwon et al., 2012).

용수수요량 예측을 위한 선행연구를 살펴보면, Atsalakis, et al. (2007)은 대규모 물 공급 관리는 최적 비용으로 양수 작업을 계획하기 위해 용수수요량을 예측의 필요성을 제안하였다. 용수수요량 예측을 위해, adaptive neuro-fuzzy inferences system(ANFIS)을 개발하였고 auto regressive(AR) 모형과 예측 성능을 비교했을 때 더 좋은 성능을 제공하였다. Alvisi, et al. (2007)은 단기적 의미의 미래 용수수요량을 예측하기 위해, AR 모형을 적용하였고, 연간, 주간 및 일일 단위의 일반적인 경향성과 주기성을 파악하였다. 하지만, AR 모형은 비교적 단순한 주기 패턴에서 적합하지만, 다양한 패턴의 주기성분이 복합적으로 결합된 용수수요량을 예측하기에는 무리가 있다.

이러한 문제를 해결하기 위해, Tabesh and Dini. (2009)은 기상 및 수위 자료 등의 외적 요인의 영향을 함께 고려하고 비선형적 모형으로 알려진 인공신경망(artificial neural networks, ANN) 모형을 적용하여 용수수요량 예측 모형을 제안하였다. 그리고 Choi, et al. (2009)은 다층 신경망(multi layer perceptron)인 AI 기반의 모형이 적합하다고 언급하였다. Firat, et al. (2010)은 용수수요량을 예측하기 위해, eneralized regression neural networks(GRNN), cascade correlation neural network(CCNN) 등신경망 계열 모형을 적용하였다. 머신러닝 및 딥러닝 기반의 예측 모형과 AR 모형을 비교하여 보다 개선된 예측 결과를 확인하였다(Kwon et al., 2012; Altunkaynak et al., 2017; Choi and Kim., 2018). 이처럼 용수수요량 예측은 최적의 수자원 관리

측면에 필수적이기 때문에 많은 연구들에서 정확한 예측을 위해 다양한 방법을 적용하고 발전되어져 왔다. 그러나 용수 수요량의 정량적인 값을 예측하는 것 보다, 단기적인 측면에서 용수수요량이 어느정도가 부족하고 충족하는지에 대한 연구가 필요한 실정이다.

우리나라의 가뭄상황관리 현황은 「가뭄 재난」 위기관리 표준매뉴얼을 바탕으로 파악할 수 있다. 그리고 가뭄 위기와 관련된 징후를 포착하거나 위기 발생이 예상되는 경우 위협의 수준을 평가하기 위해, 위기경보 발령한다. 위기관리 4단계는 관심(blue)→주의(yellow)→경계(orange)→심각(red)이다. 가뭄 재난의 경우 위기경보 수준별 판단기준이 존재하며, 이를 고려한 상황관리가 이루어진다(Kwon et al., 2018; Jung et al., 2020).

따라서, 본 연구의 목적은 가뭄재난 위기관리 기본방향의 목표와 부합하게, 신속하고 효율적인 대응으로 가뭄 피해 최소화 하고자 한다. 용수수요량 자료에 대한 확률 밀도 함수($f(x)$)와 구간에 대해 확률 변수가 포함될 확률을 산정하였다. 이를 바탕으로 주어진 확률변수가 특정 값보다 작거나 같은 확률을 산정하고자 누적 분포 함수를 이용하였고 가뭄 위기경보 단계에 맞추어 기준을 설정하였다. 설정된 기준을 바탕으로 가까운 미래의 용수수요량의 규모(scale)를 개략적으로 파악하고자 의사결정나무(decision tree, DT), 랜덤포레스트(random forest, RF) 모형을 이용하였다.

II. 연구 방법

2-1 확률 밀도 함수 및 누적 분포 함수

대부분의 홍수나 가뭄을 완전히 방어한다는 것은 불가능함에 특정 발생 확률(probability of occurrence)을 가지는 수문량에 대하여 계획하게 된다. 이를 초과할 경우, 피해를 감수할 수 밖에 없으나, 장기적으로 피해복구에 소비되는 비용, 과다설계로 인한 사업비의 증분 등을 고려하여 양에 관한 설계(quantitative design)를 설정한다. 이러한 관점에서 볼 때, 특정 크기의 발생확률을 정확하게 결정한다는 것은 수자원 시스템의 설계에 필수적이라 할 수 있다(Kim, 2010; Kim et al., 2020).

확률법칙(law of probability)은 반복되는 관측값의 통계학적 특성분석의 근거가 된다. 특정 사상 E_1 의 확률 $P(E_1)$ 이란 반복적인 시도(trials) 끝에 그 사상이 발생할 도수(度數)로 정의된다.

$$P(E_1) = \frac{n_1}{N} \tag{1}$$

여기서, n_1 은 E_1 이라는 사상의 발생도수(frequency)이고, N 은 시도횟수로서 충분히 큰 값, $\frac{n_1}{N}$ 은 상대도수

(relative frequency) 또는 확률(probability)이라 명한다.

연속형(continuous) 또는 이산형(discrete) 모두 무작위 변량(random variable)은 각 변량이 가지는 특정 값의 확률 분포에 의하여 그 특성이 표기된다. 확률 밀도 함수(probability density function, PDF)는 확률 변수의 분포를 나타내는 함수이다. 확률 밀도 함수 $f(x)$ 와 구간 $[a, b]$ 에 대해서 확률 변수 X 가 구간에 포함될 확률 $P(a \leq X \leq b)$ 는 다음과 같다.

$$\int_a^b f(x)dx \tag{2}$$

누적 분포 함수(cumulative distribution function, CDF)는 주어진 확률 변수가 특정 값보다 작거나 같은 확률을 나타내는 함수이다. 즉, 누적 분포 함수 $F(x)$ 는 어떤 변량 X 가 특정변수 x 보다 크지 않을 확률을 의미한다.

$$F(x) = P(X \leq x) \tag{3}$$

따라서, $F(x)$ 는 0에서부터 1까지 증가하는 함수이고 각 계급구간(class interval)으로 나누어 각 구간에 속하는 자료를 표기한다.

2-2 의사결정나무

의사결정나무(decision tree, Tree)는 자료들의 규칙을 바탕으로 자료를 나무형태의 그래프(graph)로 표현하여, 유사한 자료들을 세분화하고 항목별로 분류할 수 있도록 규칙을 도출하는 모형이다(Breiman et al., 1984).

의사결정나무는 동적 계획(dynamic programming)의 하양식 귀납법 방법을 바탕으로 상위마디에서 분리된 자료들은 기준에 의해 서로 유사한 자료들로 세분화된다. 그리고 반복적인 세분화 작업을 통해 최종적으로 항목별 분류가 완전하게 이루어질 때 까지 반복한다. 의사결정나무는 뿌리마디(root node), 중간마디(internal node), 끝마디(leaf node) 및 가지(branch)들로 구성된다.

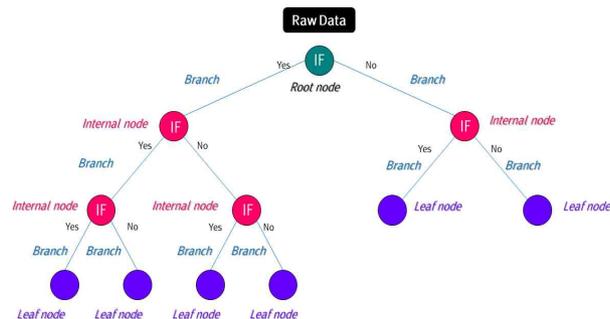


그림 1. 의사결정나무 모형의 개념도
Fig. 1. Conceptual diagram of decision tree model

여기서, 끝마디를 제외한 모든 마디에서는 분류 기준을 바탕으로 조건문(conditional)을 통해 만족하는 경우와 만족하지 않는 경우를 학습하여 예측 결과를 도출한다. 모형의 학습 과정에서 가지치기 정도에 따라 예측 결과를 보다 정확하게 구축할 수 있다. 즉, 가지치기를 위해 매개변수 별 정확도를 파악하고 최적의 매개변수를 바탕으로 예측 결과를 표현할 수 있도록 한다(그림. 1).

2-3 랜덤포레스트

랜덤포레스트(random forest, RF)는 앙상블(ensemble) 기반의 모형으로 여러 개의 의사결정나무 모형에 샘플을 여러 번 뽑아(bootstrap) 모형을 학습시켜 집계(aggregation)하는 방법인 배깅(bootstrap aggregation, bagging)의 기본 원리와 임의성(voluntariness)을 더한 분류 모형이다. 랜덤포레스트는 분류 모형 중에서도 정확도가 높은 편이다(Choi, 2017).

랜덤포레스트는 배깅의 기본 원리를 바탕으로 학습 자료를 무작위로 추출하여, 의사결정나무를 독립적으로 구성하고 총 n-tree만큼 생성한다. 일반적으로 약 500개의 의사결정나무를 생성하고 생성된 500개의 의사결정나무는 예측 결과를 투표(vote)함으로써 최종 결과를 도출할 수 있다. 여기서, 출력 결과를 도출할 때 의사결정나무를 무작위로 결정하여 결과를 도출할 수 있도록 한다. 이는 분류기의 개수(mtry)로 정의한다. 학습 과정에서 최적의 매개변수를 선정하여 가장 우수한 성능의 예측 결과를 도출하도록 모형을 반복적으로 학습한다(그림. 2).

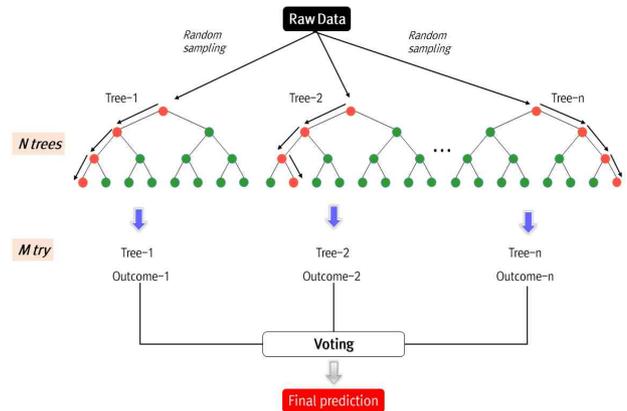


그림 2. 랜덤포레스트 모형의 개념도
Fig. 2. Conceptual diagram of the random forest model

2-4 용수수량 계급구간 분류 모형의 성능 평가 방법

분류 모형의 정확성 검증은 혼동 행렬(confusion matrix)를 바탕으로 검증을 수행한다. 혼동 행렬은 관측값을 1, 모형의 결과값을 1로 예측한 것을 true positive (TP), 관측값을 1, 모형의 결과값을 0으로 예측한 것을 false negative (FN), 관측값을 0, 모형의 결과값을 1로 예측한 것을 false positive (FP), 관측값을 0, 모형의 결과값을 0으로 예측한

것을 true negative (TN)이라 한다. 혼동 행렬은 다음 표 1과 같이 표현할 수 있다(Fawcett, 2006; Kim et al., 2021; Lee et al., 2021).

표 1. 혼동 행렬
Table 1. Confusion matrix

Classification		Predicted	
		Negative	Positive
Observed	Negative	TN	FP
	Positive	FN	TP

산정된 혼동 행렬을 바탕으로 정확도(accuracy), 오류율(error rate), 민감도(sensitivity), 정밀성(precision), 특이도(specificity)를 산정할 수 있다(표 2).

표 2. 각 평가 지표에 대한 방정식

Table 2. Equation for each evaluation index

Classification	Equation
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Error Rate	$\frac{FN + FP}{TP + TN + FP + FN}$
Sensitivity	$\frac{TP}{TP + FN}$
Precision	$\frac{TP}{TP + FP}$
Specificity	$\frac{TN}{TN + FP}$

F1-score는 정밀성과 민감도를 이용하여 아래와 같이 산정할 수 있으며, 일반적으로 β는 1로 표기한다.

$$F1-score = \frac{(1 + \beta^2)(Prec. \times Sens.)}{(\beta^2 Prec. + Sens.)} \tag{4}$$

III. 연구 대상 및 결과

3-1 대상지역 현황

섬진강 유역에는 곡성군과 구례군 그리고 광양시가 대부분을 차지하며, 순천시와 화순군 그리고 보성군은 일부를 포함하며, 경상남도 남원시와 하동군 일부와 남해군이 여기에 속한다. 섬진강의 수질은 대표지점인 구례를 비롯해서 남원시, 하동군 등 모든 지점이 거의 1급수 수준에 가깝다. 또한, 도수터널을 이용하여 동진강수계 하류의 계화간척지까지 농업용

수를 공급하고 있어, 가뭄 및 수질의 지속적인 모니터링이 필수적인 지점이다(그림. 3).

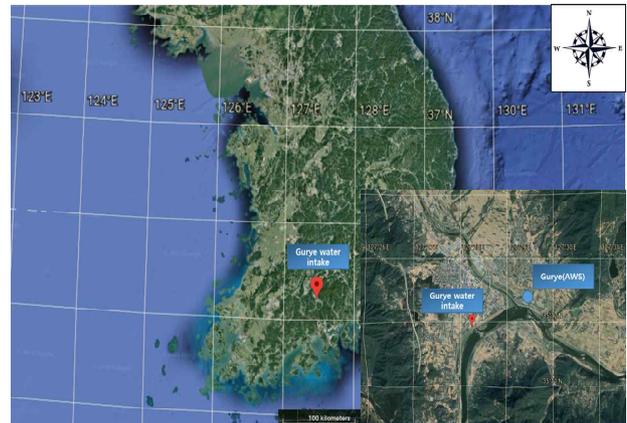


그림 3. 연구 유역의 위치
Fig. 3. Location of study basin

구례취수장 지점은 섬진강 댐에서 섬진강 하구까지의 거의 중간지점이라는 적정성과 보성강이 합류되어 풍부한 유량을 형성하고 있다. 구례취수장의 취수능력은 11,000(m³/day)이고 허가량은 9,098(m³/day), 급수지역은 구례읍으로 급수인구는 9,881명이다(그림. 3).

3-2 위기경보 기준 설정

가뭄의 상황관리에서 고려하는 위기의 유형은 생활, 농업 및 공업용수의 부족으로 인해, 농작물 피해, 하천유지유량 감소, 지하수 고갈 등이다. 이러한 위기의 전개 양상은 단계별로 1단계는 기후변화에 따른 강수부족으로 가뭄 발생, 2단계는 생활, 농업 및 공업용수 부족현상과 농작물 피해가 일부 지역에서 발생, 3단계는 생활, 농업 및 공업용수 부족현상과 농작물 피해가 대규모 지역에서의 확대이다.

가뭄예경보 기준에 근거하여 관심(blue)→주의(yellow)→경계(orange)→심각(red)별 판단 기준을 살펴보면 다음과 같다. 관심 단계는 하천(river) 및 수자원 시설(water resource facility)의 수위(water level)가 평년(normal year)에 비해 낮아 생활 및 공업용수의 가뭄대비가 필요한 경우이다. 주의 단계는 하천의 하천유지유량이 부족하거나 댐(저수지)에서 하천유지용수 공급 등의 제한이 필요한 경우이다. 경계 단계는 생활 및 공업용수 부족이 일부 발생하거나 발생이 우려되어 용수공급의 제한이 필요한 경우이다. 심각 단계는 생활 및 공업용수 부족이 확대되어 하천 및 댐(저수지) 등에서 공급 제한이 발생하였거나 필요한 경우이다. 이처럼 생활 및 공업용수의 경우 정량적으로 위기경보 기준이 설정되어 있지 않아, 정성적인 판단에 근거하여 단계별 행동에 제약이 있다.

따라서, 정량적인 기준을 수립하기 위해, 구례취수장의 용수수요량 자료를 2004년 1월 1일부터 2021년 12월 31일까

지 도수 분포를 파악하기 위해, 히스토그램(histogram)을 작성하여 파악하였다. 구레취수장의 용수수요량은 5,000부터 6,000까지에서 사용량이 증가하였고 6,000부터 7,000까지의 사용량이 가장 많이 분포하였다(표 3, 그림. 4).

표 3. 용수수요량 자료의 빈도 및 축적

Table 3. Frequency and accumulation of water demand data

Water demand (m ³)	Frequency (the number)	Accumulation rate (%)
0	0	0.00
1,000	0	0.00
2,000	2	0.03
3,000	2	0.06
4,000	19	0.35
5,000	670	10.54
6,000	4,209	74.56
7,000	891	88.11
8,000	722	99.09
9,000	52	99.88
10,000	8	100.00

물 수요관리 종합계획 작성지침에 따르면, 생활용수, 공업용수 및 농업용수의 수요량 산정에서 시설기준상 일최대 급수량의 70%에서 80%를 기준으로 산정하고 있다. 구레취수장의 허가량은 9,098(m³/day)이고 일 최대 급수량의 70%는 6,368.6(m³/day)이고 80%는 7,278.4(m³/day)이다.

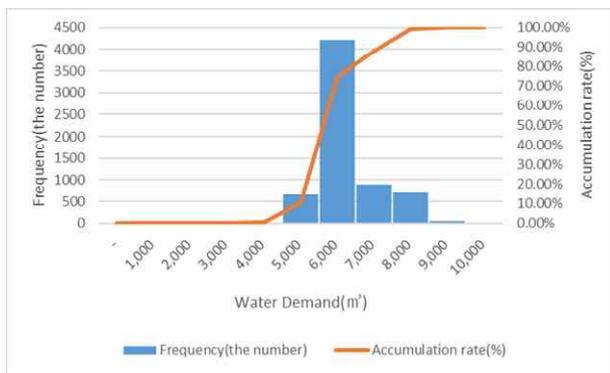


그림 4. 용수수요량 자료의 히스토그램 생성
Fig. 4. Create a histogram of water demand data

따라서, 본 연구에서는 가뭄예정보 기준, 물 수요관리 종합계획 작성지침 및 선행 연구 등을 참고하여 구레취수장의 위기경보 기준을 다음과 같이 설정하였다. 구레취수장의 허가량 및 일 최대 급수량을 고려하여, 심각 단계의 기준은 구레취수장의

최대 허가량을 기준으로 설정하였고 기준값은 9,098.0 (m³/day)이다. 경계 단계의 기준은 일 최대 급수량의 75%를 기준으로 설정하였고 기준값은 6,823.5(m³/day)이다. 주의 단계는 50%를 기준으로 기준값은 4,549.0(m³/day)이다. 관심 단계는 25%를 기준으로 기준값은 2274.5(m³/day)이다. 설정된 위기경보 기준은 다음 표 4와 같이 나타났다.

표 4. 구레 취수장의 위기경보 기준 설정

Table 4. Setting the crisis alert standards at Gurye water intake

Class	Classification	Water demand (the number)
Serious	9,098.0 ≤ 11,000.0	38
Alert	6,823.5 ≤ 9,097.0	2,102
Caution	4,549.0 ≤ 6,823.0	4,313
Attention	2,274.5 ≤ 4,548.0	8

3-3 용수수요량 계급구간 분류 예측 모형 개발

분류(classification)는 독립변수에 따라 가장 상관성(correlation)이 큰 종속변수(계급 구간, class interval)를 예측하는 것이다. 즉, 용수수요량 표본에 대한 자료가 어떠한 계급 구간에 속하는지를 파악하는 방법이다. 분류 모형(classification model)은 두 가지 항목으로 구분할 수 있다. 자료를 계급 구간에 따라 서로 다른 영역으로 나누는 경계면(decision boundary)을 찾아내고 이 경계면으로부터 어느 구간에 분포하는지를 계산하는 판별함수 모형이 있다. 또한, 입력된 자료에 대해 계급 구간에 분포할 확률을 계산하는 확률적 모형이 있다. 본 연구에서는 가까운 미래의 용수수요량의 규모를 파악하고자 의사결정나무, 랜덤포레스트 모형을 적용하고자 하였다.

구레취수장의 용수수요량은 1일전 용수수요량에 영향을 받으며, 주변 강우 관측소의 강우량 및 최고기온, 평균기온에 영향을 받는다. 이러한 영향을 고려하여, 구레취수장의 관측 용수수요량 및 강우 관측소의 1일전 강우량, 최고기온, 평균기온 자료를 2004년부터 2021년까지 1일 단위로 구축하였다.

1) 의사결정나무를 활용한 용수수요량 계급구간 분류 예측 모형 개발

학습 구간은 2004년부터 2015년까지 4,383개 자료를 교차검증(k-fold cross validation)을 적용하여, 학습 및 평가를 수행하였다. 그리고 2016년부터 2021년까지 2,192개 자료를 검증 구간으로 설정하여 모형의 정확성을 평가하였다. 의사결정나무는 1개의 매개변수(cp)의 값을 이용하여 모형을 학습하고 cp는 의사결정나무 모형의 가지치기(tree pruning)를 위한 매개변수를 의미한다. 가지치기는 의사결정나무의 과

적합을 줄이고 일반화 가능성을 증대시키는 과정이다. 다음 표 5는 의사결정나무 모형의 매개변수에 따른 결과를 나타냈고 최적의 매개변수를 선정하여 모형을 개발하였다.

표 5. 의사결정나무 모형의 매개변수 도출

Table 5. Deriving the parameters of the decision tree model

Classification	CP	Accuracy
1	0.00102	0.8920
2	0.00204	0.8932
3	0.00512	0.8952
4	0.00614	0.8953
5	0.05635	0.8909

다음 표 6은 의사결정나무 모형을 활용한 용수수요량 계급 구간 모형의 학습구간에 대한 결과이다.

표 6. 의사결정나무를 이용한 용수수요량 분류 예측 모형 성능평가(학습구간)

Table 6. Water demand classification prediction model performance evaluation using decision tree (learning section)

Tree		Obs.			
		1	2	3	4
Pre.	1	0	6	0	0
	2	0	3767	128	0
	3	0	298	183	0
	4	0	1	0	0

입력자료에 맞추어 의사결정나무 모형의 매개변수를 최적화하고 모형의 예측 성능을 평가하였다. 다음 표 7은 의사결정나무 모형을 활용한 용수수요량 계급구간 분류 모형의 평가구간에 대한 결과이다.

표 7. 의사결정나무를 이용한 용수수요량 분류 예측 모형 성능평가(평가구간)

Table 7. Water demand classification prediction model performance evaluation using decision tree (evaluation section)

Tree		Obs.			
		1	2	3	4
Pre.	1	0	0	0	0
	2	1	351	801	12
	3	1	23	624	14
	4	0	0	0	0

혼동 행렬을 살펴보면, Class 1, Class 2, Class 3, Class 4에서 낮은 예측력을 보였다. 모든 Class에 대한 예측력 평가 결과, F1-score가 0.43으로 예측력이 낮음을 알 수 있다.

표 8. 의사결정나무 모형의 F1-score

Table 8. F1-score of decision tree model

Class	1	2	3	4
precision	0.00	0.94	0.44	0.04
sensitivity	0.00	0.30	0.94	1.00
F1-score	0.43			

2) 랜덤포레스트를 활용한 용수수요량 계급구간 분류 예측 모형 개발

학습 구간은 2004년부터 2015년까지 4,383개 자료를 교차검증(k-fold cross validation)을 적용하여, 학습 및 평가를 수행하였다. 그리고 2016년부터 2021년까지 2,192개 자료를 검증 구간으로 설정하여 모형의 정확성을 평가하였다. 랜덤포레스트는 1개의 매개변수(mtry)의 값을 이용하여 모형을 학습하고 mtry는 독립변수들에서 각 트리에 사용될 변수 후보를 의미한다. 다음 표 9는 랜덤포레스트 모형의 매개변수에 따른 결과를 나타냈고 최적의 매개변수를 선정하여 모형을 개발하였다.

표 9. 랜덤포레스트 모형의 매개변수 도출

Table 9. Deriving the parameters of the random forest model

Classification	mtry	accuracy
1	1	0.8991
2	2	0.8915
3	3	0.8821
4	4	0.8785

다음 표 10은 랜덤포레스트 모형을 활용한 용수수요량 계급구간 모형의 학습구간에 대한 결과이다.

표 10. 랜덤포레스트를 이용한 용수수요량 분류 예측 모형 성능평가(학습구간)

Table 10. Water demand classification prediction model performance evaluation using random forest (learning section)

Tree		Obs.			
		1	2	3	4
Pre.	1	2	4	0	0
	2	0	3823	72	0
	3	0	248	233	0
	4	0	1	0	0

입력자료에 맞추어 랜덤포레스트 모형의 매개변수를 최적화하고 모형의 예측 성능을 평가하였다. 다음 표 11은 랜덤포레스트 모형을 활용한 용수수요량 계급구간 분류 모형의 평가구간에 대한 결과이다.

표 11. 랜덤포레스트를 이용한 용수수요량 분류 예측 모형 성능평가(평가구간)

Table 11. Water demand classification prediction model performance evaluation using random forest (evaluation section)

Tree		Obs.			
		1	2	3	4
Pre.	1	1	0	0	0
	2	1	782	77	0
	3	0	37	1283	7
	4	0	0	0	4

혼동 행렬을 살펴보면, Class 2, Class 3, Class 4에서 높은 예측력을 보였다. 모든 Class에 대한 예측력 평가 결과, F1-score가 0.81로 예측력이 높음을 알 수 있다(표 12).

표 12. 랜덤포레스트 모형의 F1-score

Table 12. F1-score of random forest model

Class	1	2	3	4
precision	0.50	0.95	0.94	0.36
sensitivity	1.00	0.91	0.97	1.00
F1-score	0.81			

IV. 결론

가까운 미래의 용수수요량의 규모를 개략적으로 파악하는 것은 용수수요량 관리에 대한 불확실성을 감소하기 위해 항구적인 물 공급확대/수요관리 해소방안이라는 것이다. 따라서, 위기관리 및 적응형 가뭄대책으로 접근하는 것이 필요하다. 위기관리 차원의 용수수요량 관리란 가뭄 예·경보 기준에 대한 이해를 바탕으로 적절한 가뭄 예방, 대비 및 대응 조치를 취하는 것이라 할 수 있다. 이를 위해, 중요한 것은 위험을 파악하는 조기경보이며, 예·경보 능력 향상이 우선적으로 진행되어야 한다.

현재 가뭄 예·경보 기준은 기상가뭄(기상청), 생활 및 농업용수(환경부), 농업용수(농림식품축산부) 별로 기준이 설정되어 있다. 기상가뭄 및 농업용수의 가뭄 예·경보 기준은 정량적으로 제시가 되어있어, 이러한 정보를 토대로 단계별 국민행동요령을 이행하기에 불확실성이 적다고 할 수 있다. 그러

나, 생활 및 농업용수의 가뭄 예·경보 기준은 정량적으로 제시가 되어있지 않아, 정성적인 판단에 의거하여 단계별 국민행동요령을 이행하기에 불확실성이 상대적으로 크다고 할 수 있다.

용수수요량 적응형 관리란 불확실성이 존재하는 상황에서 가까운 미래의 정보를 반복적인 학습 과정을 통해 추진하는 것이라 할 수 있다. 본 연구에서는 공급대책 및 수요관리 정책을 중점적으로 적응형 관리 차원에서 접근하여, 신속하고 정확도 높은 최선의 정보를 바탕으로 관리대책이 수행되도록 사전 정보를 알려주는 것이다. 이러한 정보를 활용하여 정책의 구성과 이행과정에서 유연성과 신속성이 확보되도록 해야 한다.

감사의 글

본 결과물은 환경부의 재원으로 한국환경산업기술원의 가뭄대응 물관리혁신기술개발사업의 지원을 받아 연구되었습니다.(2022003610004)

참고문헌

- [1] A. Altunkaynak, and T. A. Nigusie, "Monthly water consumption prediction using season algorithm and wavelet transform-based models" *The Journal of Water Resources Planning and Management*, Vol. 143, No. 6, June 2017. 04017011. DOI: 10.1061/(ASCE)WR.1943-5452.0000761
- [2] S. Alvisi, M. Franchini, and A. Marinelli, "A short-term, pattern-based model for water-demand forecasting" *Journal of hydroinformatics*, Vol. 9, No. 1, pp. 39-50, September 2007. <https://doi.org/10.2166/hydro.2006.016>
- [3] G. Atsalakis, C. Minoudaki, N. Markatos, A. Stamou, J. Beltrao, and T. Panagopoulos, "Daily irrigation water demand prediction using adaptive neuro-fuzzy inferences systems (anfis)" In Proc. 3rd IASME/WSEAS International Conference on Energy, Environment, Ecosystems and Sustainable Development, July 2007.
- [4] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. "Classification and Regression Trees" *CRC press*, October 1984. <https://doi.org/10.1201/9781315139470>
- [5] C. K. Choi, "Evaluation of Flood Impact Variables and Development of Flood Damage Function - Case Study for Residential Buildings and Contents -" Ph.D. dissertation, *Inha University*, February 2017.
- [6] G. S. Choi, C. Yu, R. M. Jin, S. K. Yu, and M. G. Chun, "Short-term water demand forecasting algorithm using AR model and MLP" *Journal of Korean Institute of Intelligent Systems*, Vol. 19, No. 5, pp.713-719, October 2009. <https://doi.org/10.5391/JKIS.2009.19.5.713>
- [7] J. H. Choi, and J. B. Kim, "Analysis of Water Consumption

- Data from Smart Water Meter using Machine Learning and Deep Learning Algorithms” *Journal of the Institute of Electronics and Information Engineers*, Vol. 55, No. 7, pp. 31-39, July 2018. <https://doi.org/10.5573/ieie.2018.55.7.31>
- [8] T. Fawcett, “An introduction to ROC analysis” *Pattern recognition letters*, Vol. 27, No. 8, pp. 861-874, June 2006. <https://doi.org/10.1016/j.patrec.2005.10.010>
- [9] M. Firat, M. E. Turan, and M. A. Yurdusev, “Comparative analysis of neural network techniques for predicting water consumption time series” *Journal of hydrology*, Vol. 384, No. 1-2, pp. 46-51, April 2010. <https://doi.org/10.1016/j.jhydrol.2010.01.005>
- [10] G. W. Intelligence, I. D. Yearbook, G. W. Summit, and R. Card, “Global Water Intelligence” *Global Water Intelligence*, Vol. 12, No. 10, pp. 1-72, December 2011.
- [11] O. J. Jang, Y. I. Moon, and H. T. Moon, “Methodology for assessment and forecast of drought severity based on the water balance analysis” *Journal of Korea Water Resources Association*, Vol. 54, pp. 241-254, April 2021. <https://doi.org/10.3741/JKWRA.2021.54.4.241>
- [12] G. Jeong, D. Kang, and T. Kim, “Evaluation of hydropower dam water supply capacity (III): development and application of drought operation rule for hydropower dams in Han river” *Journal of Korea Water Resources Association*, Vol. 55, pp. 531-543, July 2022. <https://doi.org/10.3741/JKWRA.2022.55.7.531>
- [13] J. Jung, D. H. Park, and J. Ahn, “Drought evaluation using unstructured data: A case study for Boryeong area” *Journal of Korea Water Resources Association*, Vol. 53, No. 12, pp. 1203-1210, December 2020. <https://doi.org/10.3741/JKWRA.2020.53.12.1203>
- [14] D. H. Kim, J. S. Kim, W. J. Wang, J. S. Lee, J. W. Jung, and H. S. Kim, “Analysis of Morphological Characteristics of Collapsed Reservoirs in Korea” *Journal of the Korean Society of Hazard Mitigation*, Vol. 20, No. 5, pp. 207-216, October 2020. <https://doi.org/10.9798/KOSHAM.2020.20.5.207>
- [15] H. S. Kim, *Hydrologic*, Paju: Donghwa Technology Publishing Co, January 2010.
- [16] J. Kim, D. Kim, W. Wang, H. Lee, M. Lee, and H. S. Kim, “Comparative analysis of linear model and deep learning algorithm for water usage prediction” *Journal of Korea Water Resources Association*, Vol. 54, pp. 1083-1093, December 2021. <https://doi.org/10.3741/JKWRA.2021.54.S-1.1083>
- [17] H. H. Kwon, M. J. Kim, and O. D. Kim, “A development of water demand forecasting model based on Wavelet transform and Support vector machine”. *Journal of Korea Water Resources Association*, Vol. 45, No. 11, pp. 1187-1199, December 2012. <https://doi.org/10.3741/JKWRA.2012.45.11.1187>
- [18] M. Kwon, J. H. Sung, T. W. Kim, and J. Ahn, “Drought assessment by bivariate frequency analysis using standardized precipitation index and precipitation deficit: Focused on Han river basin”. *Journal of Korea Water Resources Association*, Vol. 51, No. 10, pp. 875-886, October 2018. <https://doi.org/10.3741/JKWRA.2018.51.10.875>
- [19] H. Lee, H. S. Kim, S. Kim, D. Kim, and J. Kim, “Development of a method for urban flooding detection using unstructured data and deep learning” *Journal of Korea Water Resources Association*, Vol. 54, No. 12, pp. 1233-1242, December 2021. <https://doi.org/10.3741/JKWRA.2021.54.12.1233>
- [21] M. Tabesh, and M. Dini, “Fuzzy and Neuro- Fuzzy Models for Short-Term Water Demand forecasting in Tehran” *Iranian Journal of Science & Technology*, Vol. 33, No. B1, pp. 61-77, August 2009.

노희성(Hui-Seong Noh)



2011년 : 한경대학교 대학원 (공학석사)
2015년 : 인하대학교 대학원 (공학박사-수자원/지형정보)

2007년~2008년: 한국시설안전공단 시설안전네트워크연구단
2015년~현 재: 한국건설기술연구원 수자원하천연구본부
※ 관심분야 : 수재해, 레이더영상처리, 머신러닝(딥러닝), 데이터마이닝, 사물인터넷(IoT) 등

최시중(Si-Jung Choi)



1999년 : 고려대학교 대학원 (공학석사)
2016년 : 고려대학교 대학원 (공학박사-수공학)

1999년~2001년: 금호엔지니어링(주) 건설기술연구소
2001년~현 재: 한국건설기술연구원 수자원하천연구본부 수석연구원
※ 관심분야 : 물수지 분석, 수자원평가, 수자원계획 등

강성규(Seong-Kyu Kang)



2002년 : 고려대학교 대학원 (공학석사)
2012년 : 고려대학교 대학원 (공학박사-수공학)

2006년~현 재: 한국건설기술연구원 수자원하천연구본부 수석연구원
※ 관심분야 : 물배분, 하천수관리, 하천유지유량 등

김동현(Dong-Hyun Kim)



2019년 : 인하대학교 대학원 (공학석사)
2022년 : 인하대학교 대학원 (공학박사-수공학)

2022년~현 재: 인하대학교 수자원시스템연구소
※ 관심분야 : 수재해, 머신러닝(딥러닝), 데이터마이닝, 사물인터넷(IoT) 등