

〈研究発表〉

AIを活用した運転管理設定値予測ガイダンスシステムの下水処理場への適用

藤原 翔¹⁾, 平林 和也²⁾, 大場 正隆³⁾, 綿引 綾一郎⁴⁾
 石川 進⁵⁾, 張 亮⁶⁾, グェン タン フォン⁷⁾, 糸川 浩紀⁸⁾
 橋本 敏一⁹⁾, 松橋 学¹⁰⁾

¹⁾ (株)安川電機 環境・社会システム事業部
 (〒105-6891 東京都港区海岸1-16-1 ニューピアサウスタワー E-mail: Sho.Fujiwara@yaskawa.co.jp)

²⁾ (株)安川電機 環境・社会システム事業部
 (〒105-6891 東京都港区海岸1-16-1 ニューピアサウスタワー E-mail: Kazuya.hirabayashi@yaskawa.co.jp)

³⁾ (株)安川電機 環境・社会システム事業部
 (〒824-8511 福岡県行橋市西宮市2-13-1 E-mail: Masataka.Oba@yaskawa.co.jp)

⁴⁾ 前澤工業(株) 環境事業本部
 (〒332-8556 埼玉県川口市仲町5番11号 アクアテクノセンター E-mail: ryouichirou_watabiki@maezawa.co.jp)

⁵⁾ 前澤工業(株) 環境事業本部
 (〒332-8556 埼玉県川口市仲町5番11号 アクアテクノセンター E-mail: susumu_ishikawa@maezawa.co.jp)

⁶⁾ 前澤工業(株) 環境事業本部
 (〒332-8556 埼玉県川口市仲町5番11号 アクアテクノセンター E-mail: liang_zhang@maezawa.co.jp)

⁷⁾ 前澤工業(株) 環境事業本部
 (〒332-8556 埼玉県川口市仲町5番11号 アクアテクノセンター E-mail: nguyen_thanh_phong@maezawa.co.jp)

⁸⁾ 日本下水道事業団 技術戦略部
 (〒113-0034 東京都文京区湯島二丁目31番27号 湯島台ビル E-mail: ItokawaH@jswa.go.jp)

⁹⁾ 日本下水道事業団 技術戦略部
 (〒113-0034 東京都文京区湯島二丁目31番27号 湯島台ビル E-mail: Hashimotot@jswa.go.jp)

¹⁰⁾ 国土技術政策総合研究所 下水道研究部
 (〒305-0804 茨城県つくば市旭1番地 E-mail: matsuhashi-m92ta@milit.go.jp)

概要

筆者らは、平成30～令和元年度の下水道革新的技術実証事業「AIを活用した下水処理運転管理支援技術に関する研究」において、AI技術（ランダムフォレスト）を用いて水処理の運転管理設定値の予測・ガイダンスを行う技術の開発・検証を行った。本技術を規模の異なる複数の下水処理場にオフラインで適用し、MAPE（平均絶対パーセント誤差）等により予測性能を評価した。その結果、予測対象ごとにAIの設定パラメータ値等を調整することでMAPEの目標値（10%以下）を達成し、本技術の汎用性が確認された。

キーワード：人工知能（AI）、運転管理設定値、運転管理支援、ランダムフォレスト

原稿受付 2020.6.26

EICA: 25(2・3) 26-28

1. はじめに

下水道事業においては、職員数の減少や技術力の低下が喫緊の課題である。特に、下水処理場では生物処理プロセスである活性汚泥法が主流であり、その運転管理設定値は、ベテラン職員が計測データや水質などの定量項目と自身の知識・経験から総合的に判断し決定していることが多い。しかしながら、高齢化等によるこのような技術を持つベテラン職員数の減少が懸念されている。

これに対応するため、筆者らは国土交通省の下水道革新的技術実証事業（B-DASH プロジェクト）のFS

調査の「AIを活用した下水処理運転管理支援技術に関する研究」において平成30～令和元年度採択され、AI技術（ランダムフォレスト）を用いた水処理の運転管理設定値の予測・ガイダンスを行う技術を開発した。今回はこのB-DASHプロジェクトの中でもFS調査として、規模の異なる複数の下水処理場にオフラインで適用して予測性能、運用方法、経済性等を検討した。本稿では、本技術の汎用性を検証する目的で、計3箇所の下水処理場を対象としてMAPE（平均絶対パーセント誤差；詳細は後述）を指標として予測性能を確認した結果を報告する。

2. 検討方法

2.1 AIを活用した予測手法

運転管理設定値の予測手法としてランダムフォレスト法を用いた¹⁾。本手法は、条件分岐を繰り返すことでデータを分割し、その領域で回帰式によって予測する決定木を、複数作成・統合した結果を用いて予測する手法のことである。本手法の特長として、比較的性能の良い結果が得られることと、モデル生成時に「重要度」という係数により計測データ項目ごとの予測値への影響度が確認できることがある。

2.2 使用データ

(1) 対象処理場

本検討では標準活性汚泥法が採用されている下水処理場3箇所(A/B/C処理場)の計測データをオフラインで使用した(表1)。これらの対象処理場は、様々な規模での同技術の適用性を確認するために、処理能力が異なる処理場として選定したものである。

Table 1 Outline of target WWTPs

	A 処理場	B 処理場	C 処理場
規模 (現有処理能力)	約 4.4 万 m ³ /d	約 26 万 m ³ /d	約 1.3 万 m ³ /d
実施年度	2019	2018	2019

(2) 収集データ

使用した各処理場の期間は、A 処理場については 2015/4/1~2018/3/31 までの3年分、B/C 処理場については 2015/4/1~2019/3/31 までの4年分である。学習用データについて、A 処理場では 2015/4/1~2017/3/31、B/C 処理場では 2015/4/1~2018/3/31 の範囲で対象期間(日数)を変更して予測誤差への影響を検討した一方、検証用データは、A 処理場については 2017/4/1~2018/3/31、B/C 処理場については 2018/4/1~2019/3/31 の各 365 日分で固定した。また、各処理場から入手したデータは、(a) ランダムフォレストの学習(予測モデルの生成)に使用する「学習用データ」、(b) 予測性能の検証に使用する「検証用データ」に、期間で分割して使用した。

(3) 予測対象項目

予測対象項目として、活性汚泥法における重要な操作因子である曝気風量と余剰汚泥流量を選択した。曝気風量については自動計測データが存在する系列から代表で1系列を対象とした。一方、余剰汚泥流量については、A 処理場では全4系列のうち1, 2系、B/C 処理場では全系列合計の余剰汚泥流量を対象としたのは、3, 4系では反応槽構造等の水処理施設としての条件が1, 2系とは大きく異なるためである。

なお、使用する計測データ項目としては、各処理場の全項目を収集したうえでデータが不足しているものを除外し、A 処理場:168項目、B 処理場:371項目、C 処理場:131項目を使用した。

2.3 予測性能の評価方法および手順

(1) 性能指標および評価方法

本検討では、検証用データの1年間の予測結果(予測値)を、同一期間の運転実績(設定値)と比較することで予測性能を検証した。このときの性能指標として、平均絶対誤差率(MAPE; Mean Absolute Percentage Error)を用いた。これは、運転管理設定値の予測値と実績値(設定値)の乖離を予測誤差と見なし、通年での平均的な予測誤差を評価するための指標である。MAPEの定義式は以下のとおりである。

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{f_i - y_i}{y_i} \right| \quad (1)$$

すなわち、MAPEは、各時間の設定値(y_i)に対する予測値(f_i)の偏差(誤差)を設定値で除し、その絶対値を検証用データ期間で平均したものである。また、今回の検討では各予測対象でのMAPEの目標値を10%以下に設定した。この10%以下とは筆者らによる下水処理場関係者へのヒアリングに基づき、実用的な誤差率の限界値(最大値)として設定したものである。

(2) 検討手順

予測誤差(MAPE)が小さくなるような予測モデルの生成条件を見いだすために、まず、(a)学習用データに係る条件として項目数および日数、(b)ランダムフォレストのパラメータとして決定木本数および階層を選定し、これらを変化させた際のMAPEへの影響を順に確認した(表2)。ここで、学習用データの項目については、全計測項目×2年間または3年間のデータに対して、パラメータ初期値にてモデル構築を行った結果に基づき全データ項目の重要度の順位付けを行ったうえで、重要度の上位から所定数を選択した。また、上述の手順で検討実施後、目標MAPE(10%以下)に未達の場合には、(c)モデル更新の頻度を毎年から毎月、毎週へと順次増加させた。これは、最新のデータを用いて予測モデルを構築する

Table 2 Model creation conditions for validations

	学習用 データ項目数	学習用 データ日数	ランダムフォレスト パラメータ	
			決定木本数	階層
A 処理場	1~168	30~731	10~500	3~20
B 処理場	1~371	30~1096		
C 処理場	1~131			

ことで、精度が向上するケースが存在すると考えられた。上述の手順でも目標 MAPE に達しない場合には、検証用データに特殊条件が含まれないかを確認し、必要に応じて除外した。これらの検討を経て、最終的に MAPE が最小となった条件を各ケースのモデル生成条件とし、その下での MAPE を各ケースの予測誤差とした。

3. 結果と考察

3.1 曝気風量予測のモデル生成条件の検討

モデル生成条件の検討結果から (Table 3, Table 4), 学習用データの日数は、30~1096日 (A 処理場では 30~731日) の範囲での検討において、長期間のデータを使用するほど予測性能が向上することが確認された。また、データ仕様やその他のモデル生成条件等については、ケースにより最終値が異なっており、対象とする処理場や項目に応じて個別に最適値を検討する必要があると判断された。

また、A/B/C 各処理場の曝気風量の予測について、最終的に MAPE の最小値は各々 4.20%, 8.29%, 9.93% となり、目標値 (10% 以下) を達成した。ただし、C 処理場については、(a) 学習用データ、(b) モデルパラメータの調整のみでは目標 MAPE を達成できず、(c) モデル更新頻度を毎週にまで短縮することで達成した。また、余剰汚泥流量については、A/B/C 各処理場について MAPE が各々 8.09%, 4.88%, 8.38% となったが、C 処理場では検証用データに約 3 か月間の工事期間を含んでおり、これを除外した期間での目標値達成となった。

Table 3 Final validations of prediction of air flow rate

		A 処理場	B 処理場	C 処理場
学習用データ	項目数	20	2	200
	日数	731 (2年間)	1096 (3年間)	1096 (3年間)
ランダムフォレスト パラメータ	決定木本数	500	500	500
	階層	10	5	10
モデル更新頻度		毎年	毎年	毎週
MAPE		4.20%	8.29%	9.93%

Table 4 Final validations of prediction of waste activated sludge rate

		A 処理場	B 処理場	C 処理場
学習用データ	項目数	5	3	5
	日数	731 (2年間)	1096 (3年間)	1096 (3年間)
ランダムフォレスト パラメータ	決定木本数	100	100	50
	階層	5	3	5
モデル更新頻度		毎年	毎年	毎年
MAPE		8.09%	4.88%	8.38% (工事期間含む: 14.81%)

4. まとめ

AI (ランダムフォレスト法) を用いた水処理施設の運転管理設定値を予測・ガイダンスする支援技術について、下水処理場 3 箇所の過去の運転データを使用して曝気風量および余剰汚泥流量の予測性能の検証を行った。主たる成果は以下のとおりである。

- ・処理場 3 箇所×予測項目 2 点の各ケースについて、予測誤差 (MAPE) の目標値 (10%) が達成され、各種規模の下水処理場を対象に本技術が適用可能である点を確認した。
- ・学習用データやモデルパラメータ値等のモデル生成条件については、予測対象 (処理場, 制御項目) ごとに個別に検討する必要があると判断された。

今後は検討事例の更なる蓄積 (寒冷地, ほかの処理方式等) やリアルタイムでの予測検証等, 本技術の実用化に向けた更なる検討を実施していく。

謝辞

本研究の実施にあたりデータ提供を頂いた A/B/C 処理場の関係者および本研究の関係者各位に謝意を表す。

参考文献

- 1) 藤原翔, 平林和也, 大場正隆, 綿引綾一郎, 石川進, 張亮, グェン タン フォン, 糸川浩紀, 山下洋正, 矢本貴俊: AI を活用した水処理制御支援技術による処理場運転管理設定値の予測誤差の検証, 第 56 回下水道研究発表会講演集, pp. 1016-1018 (2019)