

浄水プロセスにおける正常・異常データに着目した凝集後濁度の回帰予測 CNN CNN-Based Regression of Post-Flocculation Turbidity with a Focus on Normal and Abnormal Data in the Water Purification Process

中井 優希[†] 鈴木 昭弘[‡] 松崎 博季[‡] 稲垣 潤[‡] 山村 寛[§] 根本 雄一[¶]
Yuki Nakai Akihiro Suzuki Hiroki Matsuzaki Jun Inagaki Hiroshi Yamamura Yuichi Nemoto

1. はじめに

浄水場の水処理において、凝集プロセスは後続処理に大きな影響を与える重要な工程である。このプロセスでは、原水中の懸濁質を凝集剤 (PAC) で集塊させた凝集塊(フロック)を形成するが、その凝集剤の注入量の決定には課題がある。日本の浄水場では、原水の pH、濁度、アルカリ度などを用いた近似式で注入量を決定しているが、安全性のため必要量以上に注入される傾向にあり、コスト増加に繋がっている。さらに近年、水道事業における職員数は減少傾向にあり[1]、熟練者の経験に頼る現状では、人手不足が今後の大きな課題となる。このことから凝集剤の注入量を最適化し、さらに自動化することが求められている。そのための一助として、筆者らのこれまでの研究で、凝集中に生成される凝集剤の塊である「フロック」の画像から凝集沈澱後の濁度を回帰予測する試みをしてきた。その結果、多くのデータは予測可能であったが、約 1 割程度の予測困難なデータがあることがわかっている[2]。

そこで本研究では、予測困難なデータを識別し、これに特化したモデルで凝集沈澱後の上澄み水の濁度 (凝集後濁度) を予測することで、全体の予測精度の向上を目指す。

2. 提案手法

本研究では、予測が困難な残り 1 割のデータを高い精度で予測可能にするために次のようなアプローチをとった。

まず、凝集沈澱後の濁度を予測する回帰予測モデルを学習させ、その予測結果から ± 0.5 度以内の誤差で予測できたデータ (予測可能データ) とそうでないデータ (予測困難データ) に分ける。そしてこの分けたデータを用いて、予測可能データと予測困難データを分類するモデルを学習する。さらに予測困難データのみを高精度に回帰予測するモデルを学習させ、分類モデルの予測結果が「予測困難データ」であった場合のみ、このモデルを利用して予測する。つまり、残り 1 割の予測困難データの場合のみ、それに特化した回帰予測モデルを利用することで全体としての精度を向上させるアプローチである。本研究の最終的な目標は、全てのデータに対して予測誤差 ± 0.5 度に収めることであり、本手法はその達成を目標としている。

2.1 データセット

本研究では、埼玉県吉見浄水場に設置した小型凝集装置で撮影した画像と、同時に計測された原水データを使用する。画像データは、凝集剤注入直後の画像を 1 秒ごとに 15

[†] 北海道科学大学大学院工学研究科

Hokkaido University of Science Graduate School

[‡] 北海道科学大学 Hokkaido University of Science

[§] 中央大学 Chuo University

[¶] 前澤工業株式会社 Maezawa Industries

枚合成し、1 枚として扱った。データセットの画像の一例を図 1 に示す。

原水データは、小型凝集装置内で 1 秒ごとに計測される凝集剤注入前の濁度、pH、水温であり、これらを画像データと組み合わせてモデルの入力とする。また、「凝集後濁度」は回帰予測モデルの教師信号として使用する。

データセットの総数は 4,467,433 個であり、使用した期間は、2021 年 10 月～11 月および 2022 年 9 月～11 月のデータである。本研究では凝集沈澱後の濁度が 2 度以下のデータのみを使用した。

これらのデータは、予測誤差に応じて「予測可能データ」と「予測困難データ」に分類した。具体的には、異なるエポックで学習した 2 つの回帰予測モデルを用い、予測可能データと予測困難データに分類するデータを予測した。

その結果、片方のモデルでも予測誤差が ± 0.5 度を越えたデータを「予測困難データ」、両方のモデルで予測誤差が ± 0.5 度以内のデータを「予測可能データ」とした。なお、分類器の学習に用いるため、予測可能データと予測困難データの比率を 6:4 に調整した。さらにテストデータは、2023 年の 9 月～11 月のデータからランダムに 1,000,000 個抽出した。本研究で使用したデータ数を表 1 に示す。



図 1 小型凝集装置内で撮影したフロック画像

表 1 データの内訳

	学習データ	検証データ	テストデータ
ベースライン手法	-	600,000	1,000,000
予測可能データモデル	2,000,000	600,000	-
予測困難データモデル	190,867	38,456	-
2 分類モデル	409,225	164,082	-
性能検証	-	-	1,000,000

2.2 モデル概要

本研究で学習するモデルとそのハイパーパラメータを以下に示す。

- ① 予測困難データモデル：分類モデルが困難と判断したデータを予測するモデルである。ハイパーパラメータは、エポック：200，バッチサイズ:100，最適化関数:Adam，損失関数:MSE，評価関数:MAE，学習率:0.0001 である。
- ② 2 分類モデル：データが予測可能か予測困難かを分類するモデルである。クラス 0 を「予測可能データ」，クラス 1 を「予測困難データ」とした。ハイパーパラメータは、エポック数:200，バッチサイズ:100，最適化関数:Adam，損失関数:Binary Crossentropy，学習率:0.0001 である。

3. 実験

予測可能データモデル，予測困難データモデルおよび 2 分類モデルの各モデルの学習を行った。そして、これらのモデルを組み合わせた提案手法による実験を行った。

3.1 予測可能データの回帰予測モデル

予測可能データの回帰モデルにはベースライン手法と同一のモデルを利用した。このモデルで検証データを予測した。その結果，誤差が ± 0.5 度以内に収まったデータの割合は 99.65%であった。このモデルを使用し，テストデータを予測した。予測の結果は，誤差が ± 0.5 度以内に収まったデータの割合は 67.63%であった。

3.2 予測困難データの回帰予測モデル

予測困難データを用いて学習を行った。この学習したモデルで検証データを予測した。その結果，誤差が ± 0.5 度以内に収まったデータの割合は 89.13%であった。このモデルを使用し，テストデータを予測した。予測の結果は，誤差が ± 0.5 度以内に収まったデータの割合は 55.83%であった。

3.3 2 分類モデル

予測可能データと予測困難データを分類する 2 分類モデルを学習した。この学習したモデルで検証データを予測した。2 分類モデルの混同行列を図 2 に示す。その結果，正解率は 0.73 であった。

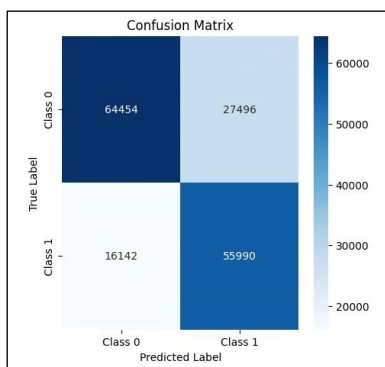


図 2 2 分類の混同行列

3.4 ベースライン手法と提案手法の比較

ベースライン手法でテストデータを予測した。その結果，誤差が ± 0.5 度以内に収まったデータの割合は 69.48%であった。次に提案手法でテストデータを予測した。提案手法でテストデータの予測した結果を図 3 に示す。このグラフは，提案手法で予測したテストデータの誤差のヒストグラムである。その結果，誤差が ± 0.5 度以内に収まったデータの割合は 66.87%であった。このことから，ベースライン手法と提案手法の精度を比較すると，提案手法の精度が 2.61 ポイント低下した。

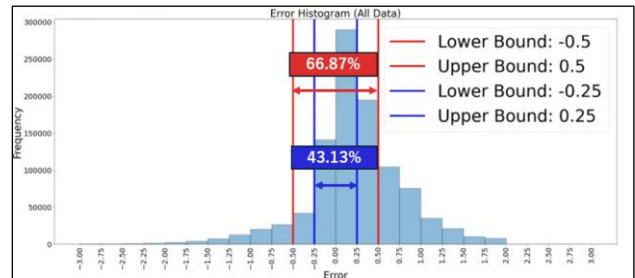


図 3 提案手法のヒストグラム

4. 考察

提案手法の予測精度の低下の原因として，分類器が正しく機能していないと考えられる。これは現在のデータセットでは，年ごとの特徴を学習しきれていないことに起因すると考えられる。ベースライン手法で，テストデータ(2023 年)を予測した場合の予測誤差が ± 0.5 度以内に収まったデータの割合が 66.87%に対して，検証データ(2021 年，2022 年)を予測した場合は予測誤差 ± 0.5 度以内に収まったデータの割合が 93.22%であった。この結果より，テストデータの予測結果が 26.35 ポイント低下した。検証データとテストデータでは使用している年が異なることから，現在のデータセットでは年ごとの特徴を学習しきれていないことが示唆された。

5. おわりに

本研究では予測困難なデータを識別し，これに特化したモデルで凝集後濁度を予測することで，全体の予測精度の向上を目指した。

今後の展望は，新たに取得する複数年のデータを用いて学習データセットを増強することで年ごとの特徴を捉え，未知の年のデータに対する予測精度向上を目指す。また今回学習した予測可能データを回帰予測するモデルと予測困難データを回帰予測するモデルを用いたアンサンブル学習により，分類を行わずに両データを予測できるモデルの学習することができるのかについて実験する。

参考文献

- [1] 厚生労働省:水道の現状について，厚生労働省，入手先 < <https://www.mhlw.go.jp/content/10601000/000476640.pdf> > (参照 2024-10-6)
- [2] 鈴木昭弘，川上敬，山村寛，根本雄一，松川瞬. 小型凝集プラントから取得したフロック画像からの季節性を考慮した凝集後濁度予測. FIT2023(第 22 回情報科学技術フォーラム)講演論文集, pp. 455-456, 2023.