

機械学習を用いたダムの浸透量推定における特徴量評価 Feature Evaluation of Dam Seepage Rate Using Machine Learning

○尾下智都*・鈴木麻里子*・井上一哉*
Tomofumi Oshita, Mariko Suzuki and Kazuya Inoue

1. 緒論

我が国の農業生産基盤を支える農業用ダムの管理は、自然災害の増加や施設の老朽化に伴い、より効率的・効果的な維持管理が重要視されている。ダム管理においては観測浸透量を管理指標と比較することにより、ダムの健全状態を判断している¹⁾。しかしながら、ダムごとに浸透特性は異なるため、浸透量により健全度を評価する基準は明確に定められていない。本研究では、健全度の評価基準として機械学習によりダムの浸透量を推定する。浸透量推定に用いる特徴量を機械学習のモデルから求められる重要度に基づいて絞った後、推定精度により評価した。

2. 推定手法

本研究では、機械学習の中でも学習の概念を理解しやすく精度の良い、Extra Trees (以下、ETと称する) を使用してダムの浸透量を推定する。ET²⁾ は、ランダム性を導入することにより個々に独立して学習した複数の決定木の平均により値を求めるモデルである。軸となる決定木は図1に示すように、分岐に用いられる特徴量 X_j における分岐基準 S_t の大小により、データ t を t_L と t_R に分岐する。ETでは各特徴量からランダムに抽出した値の中から分散を最小とする値が分岐基準に選ばれる。

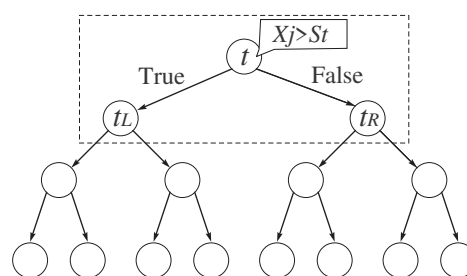


図1：決定木
Decision tree

対象とするダムは、2004年4月から2013年の3月にかけて建設された中心遮水ゾーン型ロックフィルダムである徳之島ダムとする。図2に示す2014年2月1日から2017年の10月31日にかけて1時間間隔で計測された、32,856時間分の1時間降水量、貯水位、浸透量のデータを使用した。ここに、図の上側は浸透量の推定するための特徴量に用いられる1時間降水量と貯水位、下側は目的変数である浸透量の時間に対する推移を示す。

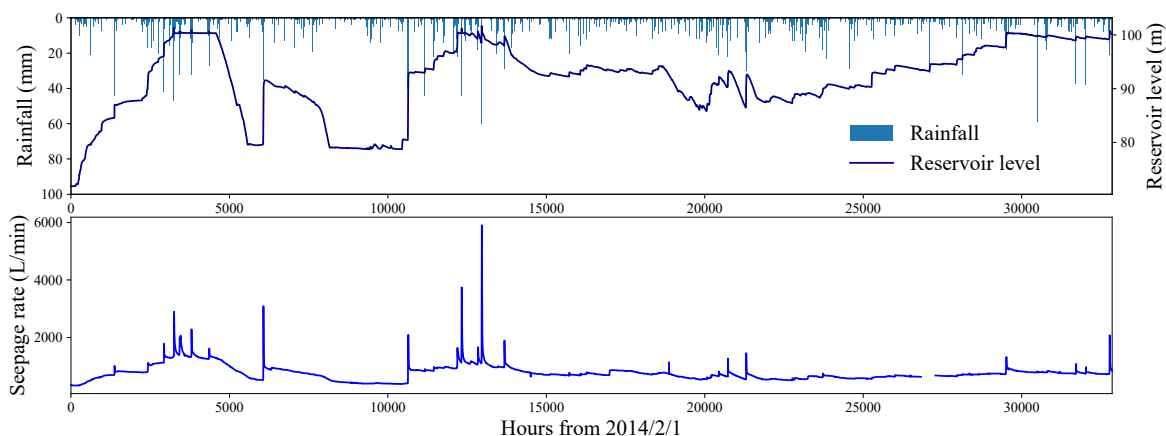


図2：徳之島ダムの観測データ（降水量、貯水位、浸透量）

Observed data of Tokunoshima Dam (rainfall, reservoir level and seepage rate)

* 神戸大学大学院農学研究科：Graduate School of Agricultural Science, Kobe University.

キーワード：浸透量，機械学習，特徴量

3. 特徴量重要度

特徴量をETから求められる重要度により評価した。当該時刻から n 時間前までの降水量を示す r_n 、 n 時間前の貯水位を示す l_n 、当該時刻貯水位と n 時間前貯水位の比を示す gl_n の3種類を評価対象とした。特徴量は n が168までの降水量 r_n 、貯水位 l_n 、貯水位勾配 gl_n 、ならびに、供用時間 $time$ を加えた計506個として、構築した回帰モデルにおける重要度を求めた。各特徴量における重要度を図3に示す。重要度は l_n, r_n, gl_n の順に高い値を示し、貯水位は1日前まで、降雨は2日前まで推定に大きく影響することが示された。

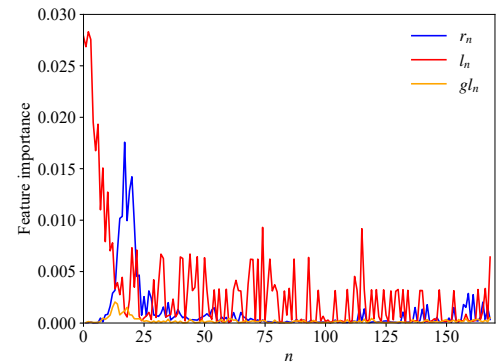


図3：各特徴量における重要度

4. 推定精度による特徴量評価

図4に示すように観測データを時系列に並べ、テストデータよりも古いデータを学習データとして、1つのテストデータを推定する。この動作を繰り返して、8760時間分のテストデータにおける推定精度を平均絶対誤差（以下、MAEと称する）と平均二乗誤差（以下、MSEと称する）により評価した。

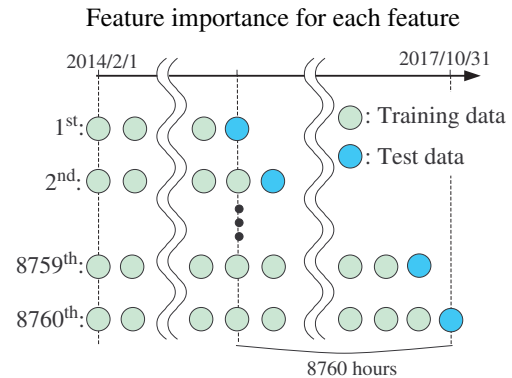


図4：推定精度による評価方法

推定精度を評価するにあたり、特徴量として、供用時間 $time$ と l_0 に加えて、 n 時間降水量 r_n を n の小さい順に増やした。降水量 r_n の総数に対する推定精度を図5に示す。誤差は $r_1 \sim r_7$ を加えることで増加、 $r_8 \sim r_{36}$ は低減する結果となった。そのため、短期的な降雨より長期的な降雨が推定精度の向上に寄与することが示された。

Evaluation method for estimation accuracy

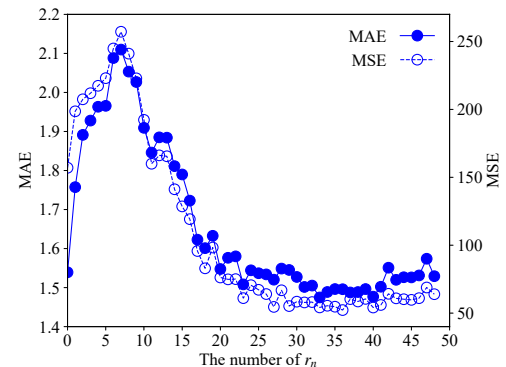


図5：降水量 r_n の総数に対する推定精度

図5より最もMSEが低い値を示した、 $time, l_0$ と $r_1 \sim r_{36}$ に貯水位 l_n 及び貯水位勾配 gl_n を n の小さい順に加え、推定精度を評価した。特徴量数に対する推定精度を図6に示す。特徴量の増加による誤差の低減は見られないため、貯水位の変動やタイムラグが推定精度の向上に与える影響は小さいと考えられる。

Estimation accuracy for the number of r_n

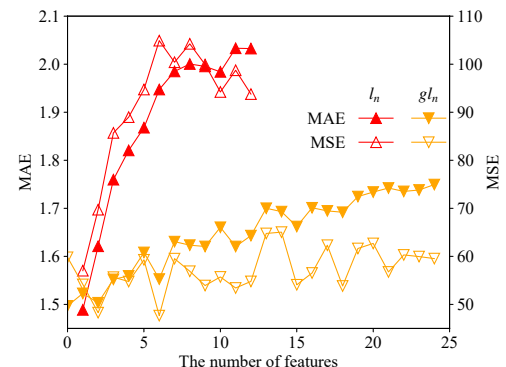


図6：特徴量数に対する推定精度

Estimation accuracy for the number of features

5. 結論

ロックフィルダムの観測データに機械学習を適用して浸透量を推定するための特徴量を重要度と推定精度により評価した。結果として、貯水位、降水量、貯水位勾配の順に重要度は高い値を示し、2日前までの降雨と当該時刻貯水位が推定精度の向上に寄与することが示された。

参考文献：1) 国土交通省: 国土交通省河川砂防技術基準維持管理編（ダム編），2016. 2) Geurts, P., Ernst, D. and Wehenkel, L.: Extremely randomized trees, *Machine Learning*, 63, pp.3-42, 2006.