ダム流入量予測への Gradient Boosting の適用

八千代エンジニヤリング株式会社　技術創発研究所　　*天方　匡純　AMAKATA Masazumi
八千代エンジニヤリング株式会社　技術創発研究所　AI 解析研究室　藤井　純一郎　FUJII Junichiro
八千代エンジニヤリング株式会社　技術創発研究所　AI 解析研究室　安野　貴人　YASUNO Takato
八千代エンジニヤリング株式会社　技術創発研究所　AI 解析研究室　大久保　順一　OKUBO Junichi
八千代エンジニヤリング株式会社　技術創発研究所　AI 解析研究室　吉田　龍人　YOSHIDA Ryuto

1. はじめに

平成 25 年 10 月の大島町土石流災害，平成 26 年 8 月の広島土砂災害，平成 27 年 9 月の関東・東北豪雨，平成 29 年 7 月の九州北部豪雨，平成 30 年 7 月の西日本豪雨と，近年，従来の河川計画で想定していなかった外力現象の多発により，多くの水災害・土砂災害が発生している．

高水流量と低水流量に大きな差がある我が国において，水運用の平滑化を目指し建設されたのがダムである．ダムの最大貯水容量は柔軟に変えられないが，運用の工夫により平滑化のポイントを変えることができる．現在も事前放流操作や異常洪水時防災操作といった運用規定により私達の生活に安全・安心をもたらしている．しかし，短時間強雨の発生頻度増大等の降雨特性変化に伴い，運用上の工夫のみでは十分な被害軽減対応が困難となっている．

代表的なダム運用支援ツールとして，ダム流入量を予測するモデルが存在する．ダム流入量予測モデルは，雨量を入力条件とし，流量を算定する流出解析モデルから構成される（図 1 上段）．これらのモデルは，既往洪水に対してモデル検証した上でパラメータ確定するが，各過程やモデルには入力・モデル条件等の不確定性を数多く抱え，リアルタイムでの流入量再現も難しい[1]．

そこで，上流水文量から下流水文量（本稿ではダム流入量）のデータ関係のみでダム流入量を予測する手法（図 1 下段）として，ブースティング法を活用する．ブースティング法は，幾つかの弱識別器を組み合わせ，より強い識別器とするアンサンブル学習手法の一つである．ブースティング法は，図 2 に示すように識別器を継続的に順番に学習し，前に学習した識別器が残した教師データとの残差を次の識別器で補正しながら学習精度を高めていく方法である．本稿ではブースティング法の一つである Gradient Boosting を取り扱い，弱識別器に決定木を用いる（当該手法を Gradient Boosting Decision Tree（以下，GBDT と記す）と呼ぶ）．合わせて，前後データの繋がりを重視する自然言語処理の分野で開発された RNN（Recurrent Neural Network）の改良版である LSTM（Long Short-Term Memory）を用いた結果と比較し，GBDT のダム流入量予測への適用性を考察する．

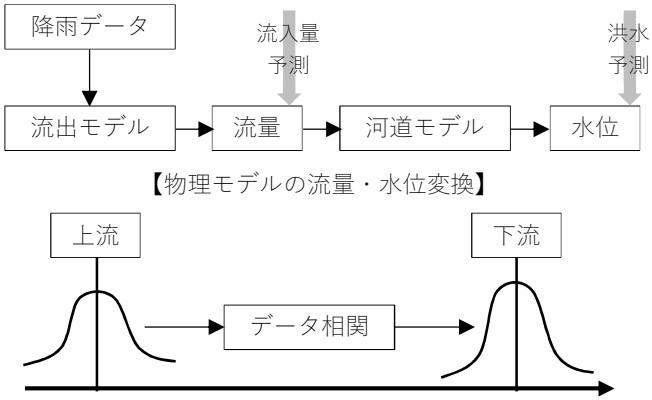

図 1　ダム流入量予測モデルの仕組み

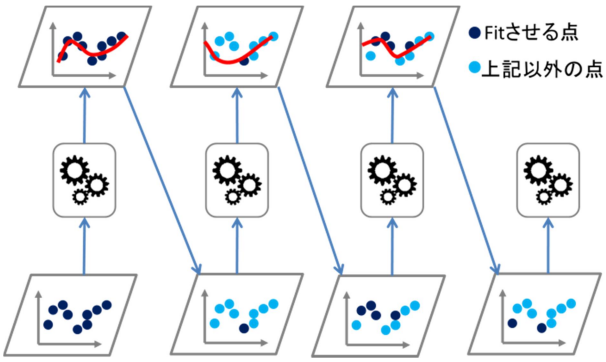

図 2　Boosting の順次 training のイメージ図
（Hands on Machine Learning with scikit-learn and Tensorflow に追記）

2. GBDT (Gradient Boosting Decision Tree)

k個目の決定木を関数f_kとする．そこで，入力x_iに対する出力値$\hat{y}_i$との関係は以下で記される．

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i)$$

・・・(1)

また，損失関数$l(a,b)$（$a = y_i(観測値), b = \hat{y}_i(予測値)$）を導入して(2)式を定式化する．これを最小化するf_tを構築しモデルを確定する．

$$\min_{f_t} \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)}) \quad \min_{f_t} \sum_{i=1}^{n} l\left(y_i, \sum_{k=1}^{t} f_k(x_i)\right)$$

$$= \min_{f_t} \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right)$$

・・・(2)

しかし，(2)式のみでは過学習が発生し予測精度が低下する可能性が高まるため正則化項を追加する（(3)式）．ここで，Tは決定木を構築したときの最終ノード数（木の大きさ），ωは決定木f_tが返す最適化された重みベクトルである．また，γ，λは，木の大きさ，重みの大きさを制御する値である．

$$\min_{f_t} L^{(t)}(f_t) =$$

$$\min_{f_t} \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \gamma T + \frac{1}{2}\lambda\|\omega\|^2$$

・・・(3)

3. LSTM (Long Short-Term Memory)

LSTMは，再帰機能を有するRNN（図4）の勾配消失問題を解決するべく，LSTMブロックと呼ばれる3つのゲート（忘却ゲート，入力ゲート，出力ゲート）とメモリセルを保持するネットワーク構造（図5）を抱え，Long-Term Memoryを実現する．

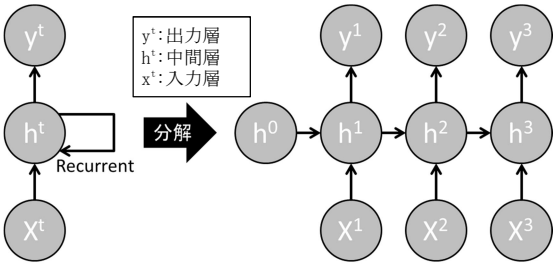

図4　RNNの特徴

4. 適用流域

本稿にて対象とする道志ダム上流域は，図3の通り，神奈川県相模川水系道志川に位置し，流域面積が112.5km^{2}であり，上流域に降った雨量が流入量として道志ダムに到達するまでの時間（図1下段）は概ね60～90分程度である[2]．

5. おわりに

発表ではGBDTとLSTMによるダム流入量予測モデルの精度の違いを報告する．

参考文献

[1] 天方匡純，渡邉優：ダム流入量予測技術の一つの考え方，第72回年次学術講演会，(2017)．

[2] 天方匡純，梁田信河，田代克浩：SOINN（自己増殖型ニューラルネットワーク）を活用したダム流入量予測に関する研究，電力土木，No.339,pp.9-15，(2017)．

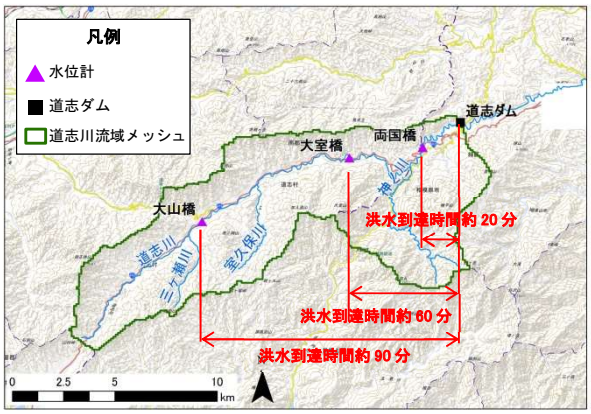

図3　道志ダム流域の概要

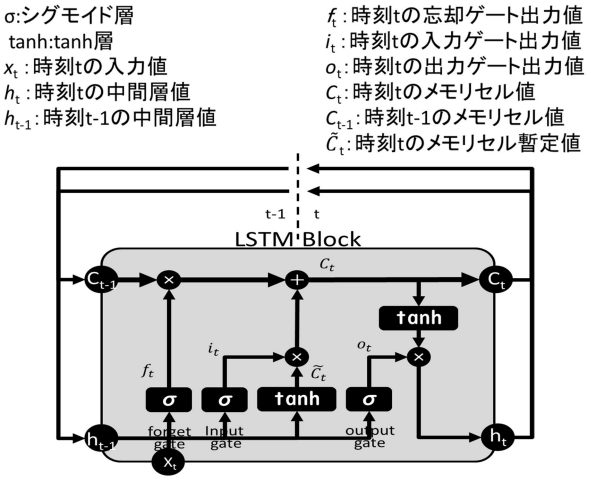

図5　LSTMの特徴