

水位推定誤差の確率分布に基づく 河川水位観測データのリアルタイム異常検知

一言 正之¹, 川越 典子², 橋田 創³, 清 雄一⁴, 房前 和朋⁵

¹正会員 日本工営株式会社 技術本部 (〒300-1259 茨城県つくば市稲荷原 2304)

E-mail: hitokoto-ms@n-koei.jp

²正会員 日本工営株式会社 福岡支店 (〒812-0007 福岡県福岡市東比恵 1-2-12)

E-mail: a7078@n-koei.co.jp

³正会員 日本工営株式会社 福岡支店 (〒812-0007 福岡県福岡市東比恵 1-2-12)

E-mail: a7913@n-koei.co.jp

⁴電気通信大学 大学院情報理工学研究科 (〒182-0021 東京都調布市調布ヶ丘 1-5-1)

E-mail: sei@is.ucc.ac.jp

⁵国土交通省 九州地方整備局 (〒812-0013 福岡県福岡市博多区博多駅東 2-10-7)

E-mail: fusamac-k8910@mlit.go.jp

水位計による河川水位のオリジナル観測データには、各種の異常値が含まれる。観測水位の異常値は、防災対応の判断や洪水予測システムに致命的なエラーを引き起こす可能性があるが、リアルタイムでの異常検知は十分に行われていない。本研究では、10分毎に観測所から配信される河川水位データを対象として、リアルタイムに異常値を検出する技術を開発した。機械学習による水位推定手法の技術を用いて、周辺の水位・雨量状況から対象とする観測地点の現時刻の水位を推定し、実観測データとの乖離度合いから異常度を算出した。さらにルールベースによる異常検知と組み合わせ、検知性能の向上を図った。九州管内の実データを用いて提案手法の精度検証を行い、既存手法と比較して高い検知性能を確認した。

Key Words: anomaly detection, river water level, machine learning, deep learning, flood prediction

1. はじめに

河川水位観測データ（以後、観測水位）の品質は、洪水時のリアルタイム防災情報としての実況把握、洪水予測の精度を左右する入力データとして重要である。

観測水位をリアルタイムでインターネット配信している水文水質データベース¹⁾や川の防災情報²⁾では、迅速な情報提供を優先し、異常値の含まれる観測所からの送信データを暫定値としてそのまま公表している。また国土交通省と放送局との連携により、地上デジタル放送による河川情報のリアルタイム提供においても、基本的に観測所のデータをそのまま配信している。なお九州地方整備局管内の地上デジタル放送では、簡易な異常検知の仕組みがあるが、事後的に管理者などへの通知メールにより人間の判断を促すものにとどまっている。

また国土交通省の管轄する水位計については、水文観測業務規程に則った年2回の事後的なデータ照査が行わ

れている。照査要領³⁾や、照査要領に基づく手引きも作成されており、異常データのスクリーニングに用いられているが、最終的に行われる異常値の修正は専門家による分析・判断にゆだねられている。

一方、近年は機械学習を用いた異常検知手法の研究が進められている⁴⁾。例えば海洋観測データについては、決定木を用いてリアルタイムでの品質管理が検討されている⁵⁾。また地盤の常時微動記録データについて、深層学習を用いて事後的にノイズ区間を取り除く検討が行われている⁶⁾。

本研究は、10分ごとに配信される観測水位データを対象に、リアルタイムに異常値を検知することを目的とした。具体的には、機械学習による水位推定器を用いて、観測水位の異常度を算出する仕組みとした。また補助機能として、ルールベースによる異常検知を組合せ、異常検知性能の向上を図った。構築した手法を用いて、九州地方の代表5地点において精度検証を行った。



図-1 水位観測所の位置図

表-1 水位観測所と検討対象データ期間の一覧

対象地点	水系	データ数 (データ年)
恵蘇ノ宿	筑後川	472465 (2008.1~2016.12)
東名	筑後川	135691 (2014.6~2016.12)
片ノ瀬	筑後川	372667 (2009.12~2016.12)
宮田橋	遠賀川	210379 (2013.1~2016.12)
春日橋	遠賀川	684799 (2003.12~2016.12)

2. 対象データ

(1) 対象観測所

対象とした観測所は、筑後川水系の恵蘇ノ宿、東名、片ノ瀬、遠賀川水系の宮田橋、春日橋の5地点とした。図-1に観測所の位置を示す。本検討での10分ごとの観測データを異常検知の対象とした。用いたデータの期間およびデータ数を表-1に示す。観測地点・データ期間の選定基準として、データ品質照査の有無、異常値の要因の明確さ・多様さ、テレメータ地点であることを考慮した。なお、本検討では欠測データは対象外とした。

(2) 異常データの概要

河川の水位観測における代表的な異常値として、スパイクノイズ(瞬間的な異常値)、水位の頭打ち・底打ち、正常値から平行にスライドする異常などがある。異常値を含んだリアルタイムデータの例を図-2に示す。異常の要因として、電気系統的異常、水位計内のゴミ詰まり、工事や水位計の付け替え・メンテナンスに伴う観測機器本体のズレなどがありうる。

なお図-2下(東名)では、何らかの異常により欠測が続いた後、リアルタイムデータが一律で正しい値よりも大きくなっている。また、照査済みの正常なデータについても、一見して異常値のような急激な水位の上昇・下

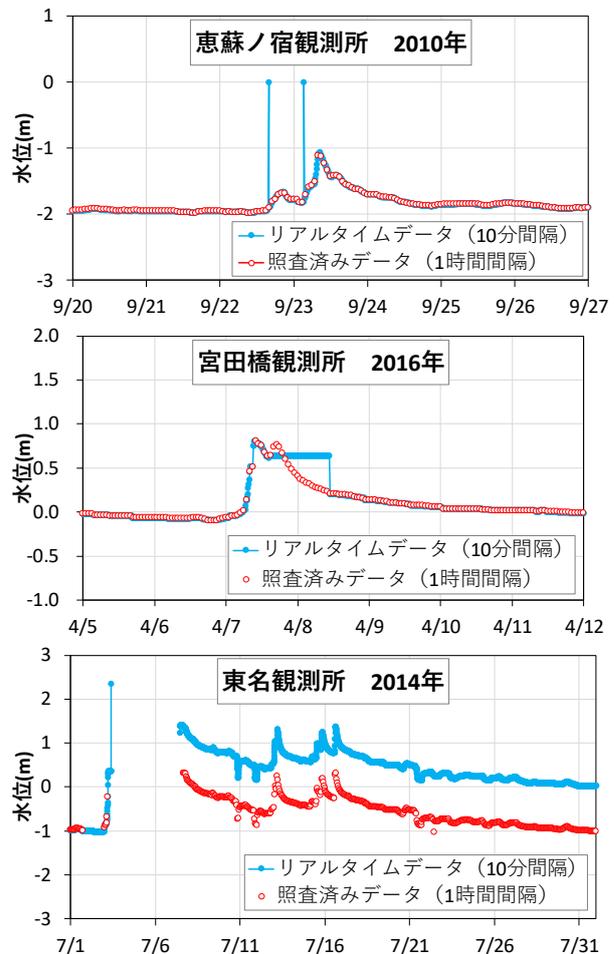


図-2 河川水位観測データの異常値の例(上から;スパイクノイズ, 水位の頭打ち, スライド)

表-2 リアルタイム観測水位に含まれる異常データの集計

異常の種類別	データ数 (全データに対する割合)	本検討での 検知対象
スパイクノイズ	31 (0.002%)	○
水位の頭打ち・底打ち	703 (0.037%)	○
スライドずれ	3787 (0.202%)	×
合計 (上記以外も含む)	4525 (0.241%)	

降がみられる。これらは実際には堰の操作による水位変動であり、異常値ではない。他にも集中豪雨の際などは中小河川で急激に水位が変動する場合もあり、見た目や閾値だけでは異常検知が難しい場合も多い。

(3) 異常データの集計、ラベル付け

検討に用いたデータは、リアルタイムで観測された10分間隔の水位データと、事後に照査された1時間間隔の水位データの2種類である。リアルタイムデータには異常値が含まれており、照査済みデータには異常値は含まれないものとした。異常検知手法の構築および精度検

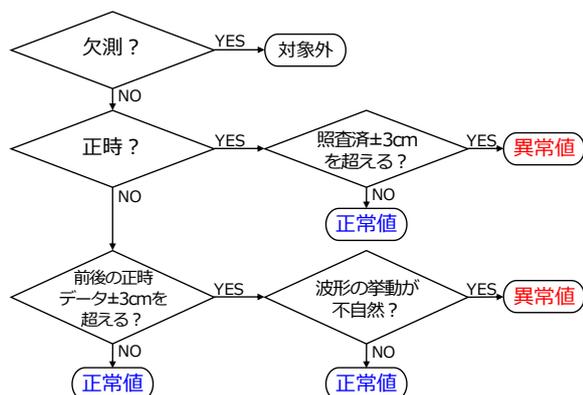


図-3 リアルタイムデータの異常値・正常値のラベル付け手順

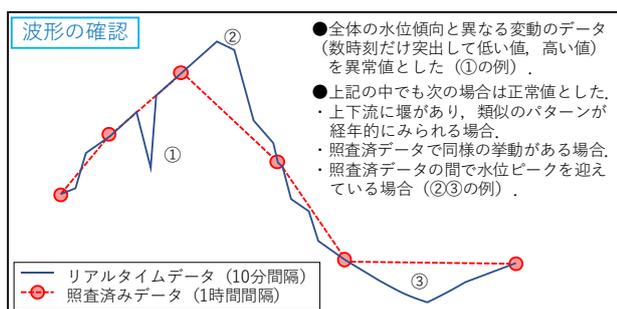


図-4 水位時系列による異常値・正常値のラベル付け基準

証のため、過去の観測データについて、異常値・正常値のラベル付けを行った。ラベル付け手順を図-3に示す。まず、リアルタイムデータもしくは照査済みデータが欠測の場合は検討対象外とした。次に正時データについては、照査済みデータと3cmより大きな乖離がある場合を異常値とした。非正時データについては、前後の正時の照査済みデータと3cmより大きな乖離があり、なおかつ時系列的な水位の波形が不自然な場合を異常値とした。具体的な波形による判断基準は図-4に示す通りとした。

観測水位に含まれる異常値の割合を表-2に示す。異常値の割合は、全体で0.241%となった。これは平均で3日に1個の異常値が含まれる割合である。異常値の内訳は、ほぼ全てがスパイクノイズ、頭打ち・底打ち、スライドのいずれかである。このうちスライド異常が最も多いが、図-2下(東名)の後半部分のようなスライド発生後に異常を検知するのは困難であり、本検討の対象外とした。なお、水位計の付け替えや何らかの異常を発端としてスライド異常が生じるケースがほとんどであるため、実用上は異常が始まった時点を検知できれば良い。

3. 異常検知手法

(1) 異常検知の考え方

異常検知は様々な分野で行われている。近年は機械学習を用いた異常検知の研究が進められており、データ群から直接的に異常値を検出することが検討されており、

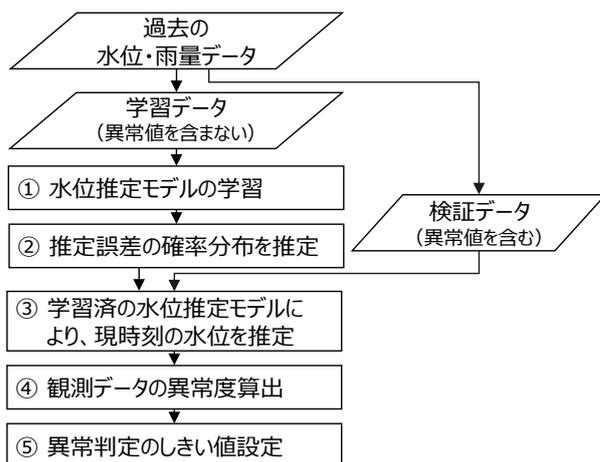


図-5 水位推定誤差の確率分布に基づく異常検知器の作成手順

表-3 水位推定器の入出力項目

出力項目	入力データ項目 (使用した時系列)
恵蘇ノ宿 ΔH	恵蘇ノ宿 $H(-1, -2)$, 恵蘇ノ宿 $\Delta H(-1, -2, -3)$, 荒瀬 $\Delta H(-2, -3, -4, -5, -6)$, 片ノ瀬 $H(0, -1)$
東名 ΔH	東名 $H(-1, -2)$, 東名 $\Delta H(-1, -2, -3)$, 流域平均 $R_{12}(0, -1)$, 馬責馬場 $R_{12}(0, -1)$
片ノ瀬 ΔH	片ノ瀬 $H(-1, -2)$, 片ノ瀬 $\Delta H(-1, -2, -3)$, 恵蘇ノ宿 $H(-6, -7, -8)$, 荒瀬 $H(-12)$, 流域平均 $R_{18}(-3, -4, -5)$,
宮田橋 ΔH	宮田橋 $H(-1, -2)$, 宮田橋 $\Delta H(-1, -2, -3)$, 流域平均 $R_{6}(-8, -9, -10)$
春日橋 ΔH	春日橋 $H(-1, -2)$, 春日橋 $\Delta H(-1, -2, -3)$, 小柳 $R_{1}(0, -1)$, 流域平均 $R_{6}(0, -1)$

表-4 学習条件 (恵蘇ノ宿, 東名, 片ノ瀬, 宮田橋, 春日橋)

種別	設定内容
学習データ期間	過去6年間の28, 13, 10, 11, 9の出水データセット数
学習モデル	ニューラルネットワーク (深層学習)
ミニバッチサイズ	100
モーメンタム	0.5
学習率の設定方法	AdaGrad(初期値0.1)

有力な手法と考えられる。一方で、本検討では直接的なデータ分析に依るのではなく、既往の水理・水文的な知見を利用する方が有利だと判断し、既往の水位予測手法をベースにした異常検知手法を試みた。

まず対象地点や周辺の雨量・河川水位の時系列挙動から、現時刻の対象地点の河川水位を推定するモデルを構築した。実際に観測された値と、周辺情報から推定した値とのずれに着目して、観測値の異常度を算出するものとした。具体的には図-5に示す通り、以下のような手順とした。①まず異常値を含まない過去の水位・雨量データを用いて河川水位の推定器を作成し、②作成した水位推定器の推定誤差の確率分布を求めた。③次に異常値を

含んだ検証データを用いて、学習済みの水位推定器により水位を算出し、④推定水位とのずれ具合から観測水位の異常度を算出した。⑤最後に異常度の閾値を設定し、異常値を精度よく分類できるようにした。

(2) 河川水位の推定器の構築

河川水位の推定器としてニューラルネットワークを用いた。推定器の入力データは上流の雨量・水位とし、出力は 10 分前から現時刻までの対象地点の水位変化とした。入出力データの組み合わせは、各推定器で表-3の通りとした。表中で H は水位、 ΔH は 10 分あたりの水位変化、 R_x は X 分雨量を表す。モデルの基本的な構築手順は既往研究 [7] に倣い、入力データは出力データと関連の高いものを用いた。学習期間は表-1 に示す期間のうち 2015 年 12 月までとし、期間中の連続的に 10 分値が存在する出水時データを学習対象とした。なお異常値を含む出水は検出目標となるため学習データから除外した。学習の諸条件は表-4 の通りとし、モデルのネットワーク素子数や dropout 率は交差検証により設定した。

(3) 河川水位推定器による異常度算出

ある時刻までに得られる水位計周辺の様々な観測値 x を用いて、ある時刻の水位 y を予測する推定器を $f(x)$ とする。学習期間における推定結果から、推定器の予測結果と観測値との誤差 $\Delta y = (f(x) - y)$ を集計する。ここで推定誤差が正規分布と仮定すると、推定誤差の確率分布は次のように表される。

$$P(\Delta y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(\Delta y - \mu)\right\} \quad (1)$$

ここで、 μ は推定誤差の平均、 σ は推定誤差の標準偏差である。これらの値は最尤推定により求める。

以上の手順により、学習期間での水位推定結果から、推定誤差の頻度分布と誤差の確率密度関数を求めた。このように事前に推定誤差分布が求められた推定器がある場合、新たな時刻で観測された水位の異常度 $a(y)$ は、推定器による推定値を用いて下記のように表される。

$$a(y) = \left(\frac{\Delta y - \mu}{\sigma}\right)^2 \quad (2)$$

以上のように算出される異常度に対し、適切な閾値を設定することで異常検知が可能である。なお本検討では異常値の事例に限られるため、検知性能が高くなるように閾値を手動で設定した。

(4) ルールベースによる異常検知

水位の頭打ち・底打ちにより同じ値が連続するような異常については、前後の時系列挙動から検知する必要があるため、リアルタイムでの検知は難しい。このような異常については、前後数十分から数時間の時系列挙動を

表-5 東名におけるルールベースでの異常検知基準

条件	ルール 1	ルール 2
1	直前の 60 分で 水位変動 25cm 以上	直前の 60 分で 水位変動 20cm 以上
2	直後の 50 分で 水位変動 0cm	直後の 180 分で 水位変動 1cm 以下
3	上記の検査期間で 平均水位が -0.7m 以下	上記の検査期間の前半で 平均水位が -0.7m 以下

表-6 宮田橋におけるルールベースでの異常検知基準

条件	ルール 1	ルール 2
1	直前の 120 分で 水位変動 40cm 以上	直前の 120 分で 水位変動 18cm 以上
2	直後の 50 分で 水位変動 0cm	直後の 210 分で 水位変動 0cm
3	上記の検査期間で 平均水位が 0.0m 以下	上記の検査期間の前半で 平均水位が 0.0m 以下

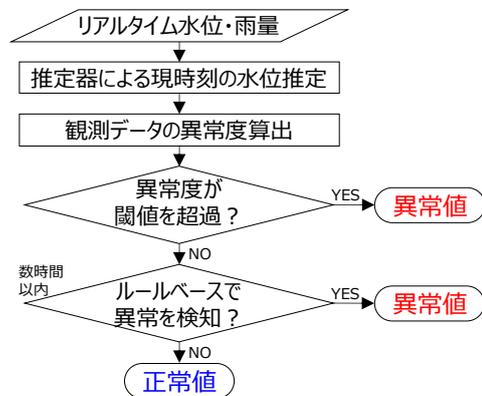


図-6 水位推定器とルールベースを組み合わせた異常検知手順

踏まえてルールベースによる検知基準を設定した。なお、本検討では頭打ち・底打ちの異常が見られた宮田橋、東名のみでルールベース検知を適用した。設定したルールを表-5, 6 に示す。表のように、各観測所でルールを 2 つずつ設定し、どちらかのルールで条件を全て満たした場合に異常値と判定した。設定した 2 つのルールは「大きな水位変化直後の短時間の水位静止」および「小さな水位変化直後の長時間の水位静止」という組み合わせで、様々な頭打ち・底打ちパターンに対応させた。各ルールの数値は、検知性能が高くなるように手動で設定した。

以上のように、水位推定器の推定誤差に基づく異常検知と、ルールベース検知との 2 段階による異常検知 (図-6) を、本検討における異常検知の提案手法とした。

4. 実河川データを用いた異常検知の性能評価

構築した異常検知手法の性能評価を行った。評価データは 2016 年の全期間および 2015 年以前の異常値を含んだ出水期間とした。なお、性能評価用のデータは河川水位推定器の学習には用いていない。

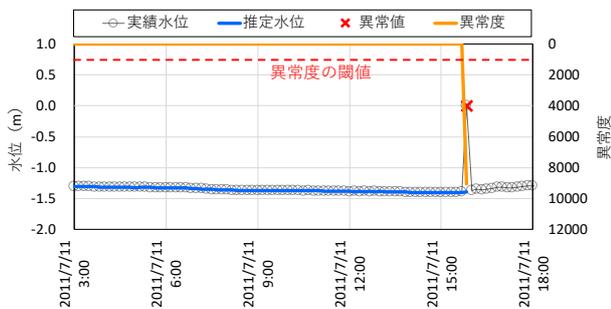


図-7 提案する異常検知の適用例 (恵蘇ノ宿)

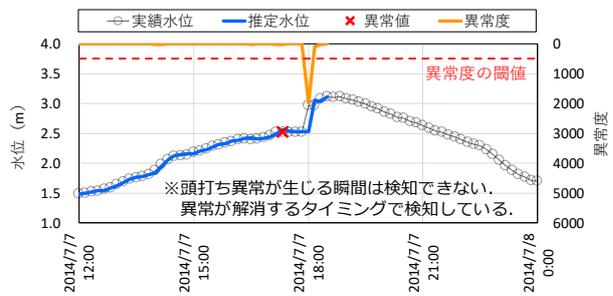


図-9 提案する異常検知の適用例 (宮田橋 1)

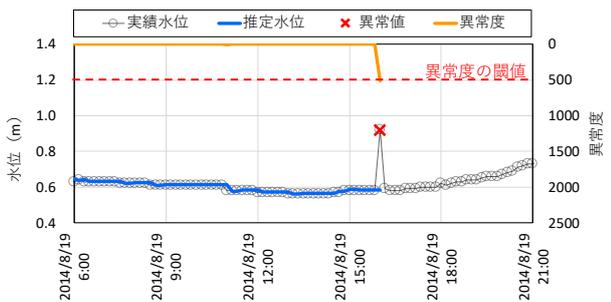


図-8 提案する異常検知の適用例 (片ノ瀬)

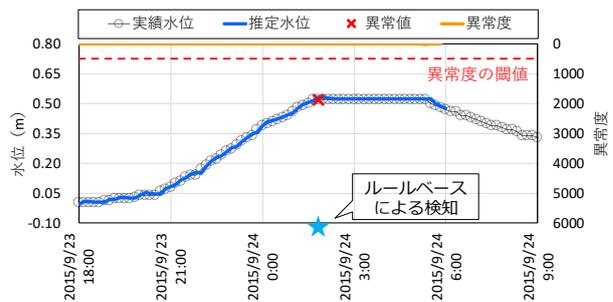


図-10 提案する異常検知の適用例 (宮田橋 2)

(1) 異常検知の結果概況

検証対象とした水位データと、推定器による水位推定結果および算出した異常度の例を図-7 から図-12 に示す。

図-7, 8 のようなスパイクノイズについては、推定器によって急激な異常度の上昇が生じるため、即座に異常値の検知が可能であった。図-9, 10 のような水位の頭打ち異常については、ルールベースにより検知できた。推定器による検知では、即座の検知はできていないが、頭打ちから正常値に回復する際の急激な水位変化を異常値として検知している。図-11, 12 の例では異常検知ができていない。図-11 は数 cm 程度のわずかな異常であり、図-12 も自然な水位挙動との区別が難しい。これらのように人間の目でも異常検知が困難なケースについては、提案手法による検知も難しいと考えられる。

(2) 異常検知の性能評価

評価期間における提案手法の精度を図-13, 表-7 に示す。異常値を正しく検出した場合を真陽性 (TP), 異常値の検出に失敗した場合を偽陰性 (FN), 正常データを誤って異常値として検出した場合を偽陽性 (FP), 正常データを正しく検出した場合を真陰性 (TN) とし、適合率 (Precision), 再現率 (Recall), F 値を算出した。

また提案手法の評価のため、九州地方整備局管内での地上デジタル放送における管理者向けの異常通知基準に基づく手法 (以下、既存手法) との比較を行った。既存手法では、観測下限値より低い値、左右岸堤防を超える値、10 分間で±50cm を超える変化があった値、同一の値が7日間以上続いた場合に異常と検知する。

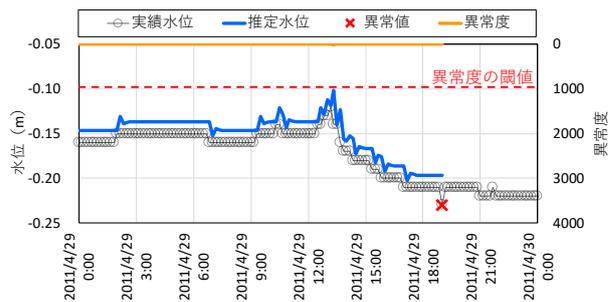


図-11 提案する異常検知の適用例 (春日橋 1)

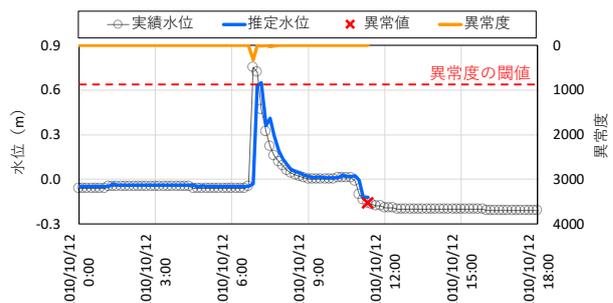


図-12 提案する異常検知の適用例 (春日橋 2)

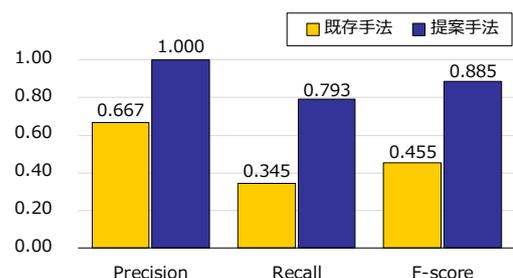


図-13 異常検知手法 (既存手法・提案手法) の精度評価

表-7 異常検知の検証結果

観測所	異常度 閾値	水位推定器によるリアルタイム検知				ルール ベースに よる検知	エラー検知数 (提案手法)		エラー検知数 (既存手法)	
		真陽性 TP	偽陰性 FN	偽陽性 FP	真陰性 TN		検知可	検知不可	検知可	検知不可
恵蘇の宿	1000	3	0	0	52916	-	3	0	3	0
東名	5000	5	11	0	50733	7	12	4	5	11
片ノ瀬	500	1	0	0	52801	-	1	0	0	1
宮田橋	500	6	1	0	52768	1	7	0	1	6
春日橋	1000	0	2	0	52030	-	0	2	1	1

※ルールベース検知は、リアルタイム検知の陰性データのみ適用

提案手法では、既存手法で見逃している異常値も検知できている場合があり、再現率が向上している。なお正常値に対する誤検知は提案手法ではゼロ、既存手法では5つ（いずれも春日橋）であった。

検知を行ったが、関連データからダイレクトに異常検知を行うような機械学習技術の適用性についても検討の余地が残されている。

5. おわりに

機械学習による水位推定技術と、ルールベースとを組み合わせた、河川水位観測データの異常検知手法を構築した。提案手法を九州管内の5観測所に適用し、従来手法を上回る異常検知性能を確認した。なお今後の課題として、次のような点が挙げられる。

- 入手できるリアルタイムデータ、照査済みデータが限られており、異常値自体もわずかであった。そのため、本検討では異常度の閾値やルールベースの条件を手動で設定しており、検証結果には精査の余地が残されている。今後はデータの蓄積・共有を進めて行くことが技術的発展のために有用である。
- 異常値・正常値のラベル付けに多大な労力を必要とした。検知モデルは地点ごとに構築・検証する必要があるため、全国に多数ある水位観測所へ適用するためには、異常値・正常値のラベル付けを始めとした各種作業の自動化、省力化に取り組む必要がある。
- 本検討では従来からの水位予測技術を援用して異常

参考文献

- 1) 国土交通省, 水文水質データベース web ページ, <http://www1.river.go.jp/> (最終閲覧日: 2019年5月20日)
- 2) 国土交通省, 川の防災情報 web ページ, <http://www.river.go.jp/kawabou/ipTopGaikyo.do> (最終閲覧日: 2019年5月20日)
- 3) 国土交通省, 水文観測データ品質照査要領, http://www1.river.go.jp/hinsitu_syosa.pdf (最終閲覧日: 2019年5月20日)
- 4) Raghavendra Chalapathy, Sanjay Chawla: Deep Learning for Anomaly Detection: A Survey, arXiv:1901.03407, 2019.
- 5) 上川路洋介, 松山開, 福井健一, 細田滋毅, 小野智司: 条件付確率場を用いた海洋観測データの品質管理, 人工知能学会誌, Vol. 33(3)G, pp. 1-11, 2018.
- 6) 宮本崇, 古谷貴彦, 森川仁: 深層学習を用いた常時微動記録からの解析対象区間の自動抽出, 土木学会論文集 A2(応用力学), Vol. 73, No. 2, pp. 321-331, 2018.
- 7) 一言正之, 櫻庭雅明, 清雄一: 深層学習を用いた河川水位予測手法の開発, 土木学会論文集 B1(水工学), Vol. 72, No. 4, pp. 187-192, 2016.

(Received May 31, 2019)

(Accepted August 7, 2019)

REAL-TIME ANORMALY DETECTION OF RIVER WATER LEVEL OBSERVATION BASED ON PROBABILITY DISTRIBUTION OF WATER LEVEL ESTIMATION ERROR

Masayuki HITOKOTO, Noriko KAWAGOE, Hajime HASHIDA, Yuichi SEI and
Kazutomo FUSAMAE

Real-time observation data of the river water level includes various anomalies. Such anomalies may cause fatal errors in judgments on disaster prevention activity and flood forecasting systems, but real-time anomaly detection has not been sufficiently implemented. In this study, we developed the model to detect anomalies in real-time for river water level data sent from observation stations every 10 minutes. By using machine learning, the water level at the current time of the objective observation station was estimated from the neighboring water level and rainfall. Then the anomaly score was calculated from the degree of deviation between the estimated water level and actual observation. Furthermore, the model was combined with the rule-based anomaly detection model. The proposed method was verified using actual observation data, and better performance was confirmed compared to the existing method.