2ZK-08

グローバル水文モデルのための堤防自動検知 -米国を対象とした事例研究-

池側 正人 [1]　Tristan Hascoet[1]　Victor Pellet[2]　渡辺 恵 [2]　Xudong Zhou[2]
田中 良明 [2]　滝口 哲也 [1]　山崎 大 [2]

[1] 神戸大学, [2] 東京大学

1　はじめに

現状の水文モデルでは Digital Elevation Models(DEM) から得られる地形情報をもとに地表水や河川氾濫原をシミュレーションするが、DEM は堤防やダムを詳細に表現していない [1]。近年、そのような洪水防護施設を全球水文モデル上で考慮する取り組みが進められており、モデルの精度向上に伴う洪水リスク算定や気候研究への大きな影響が期待される。しかし、全球で利用可能な堤防データは少なく、稀有な例として USACE が公開する National Levee Database(NLD) と呼ばれる堤防データベースが全米で利用可能である [1]。そこで本稿では米国大陸全域を対象に、機械学習を用いて堤防の存在を予測する手法を提案する。

2　データと問題定義

本研究では、堤防は人間活動と洪水リスクが重なる場所に予測されるという仮説を立てる。

2.1　出力データ

本研究では、田中らによって提供された堤防データ [1] をラベルとして用いる。NLD から得られた堤防位置情報を 10km 解像度の集水域単位で表したものを、同スケールのグリッドに変換した。図 1 に示すように、堤防位置は 2 値で表され、1 ピクセルが 1 つの集水域に対応する。

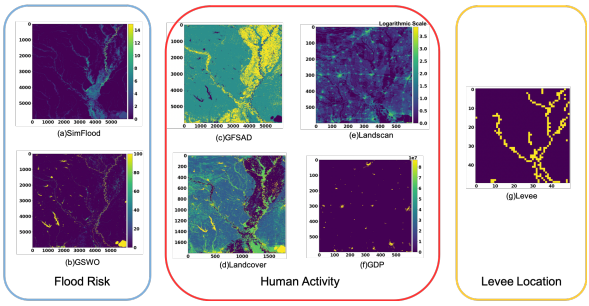

図 1　入出力データ：それぞれ、(a)SimFlood、(b)GSWO、(c)GFSAD、(d)Landcover、(e)Landscan、(f)GDP、(g) 堤防、を示す。

2.2　入力データ

仮説を検証するため、洪水リスクと人間活動の空間的な分布を定量化するデータセットを構築した。図 1 に、ミシシッピ・デルタ周辺における、前処理を行う前のデータを示す。

洪水リスク: 洪水リスクを定量化するために、図 1(a)、(b) に示す SimFlood[2] と Global Surface Water Occurrence(GSWO)[3] を用いた。Simflood は水文モデルのシミュレーションにより得られた 20、50、100 年の再現期間での洪水深度 [m] の平均を表す。GSWO は衛星により洪水を観測しており、本研究では過去 30 年にわたる地表水の頻度 [%] データを用いている。Simflood、GSWO は共に 90m 解像度である。

人間活動: 人間活動を定量化するために、図 1(c)、(d)、(e)、(f) に示す土地利用 (GFSAD[4]、Landcover[5]) と人口 (LandScan)[6]、そして GDP[7] のデータを用いた。GFSAD は農耕地、水域、非農耕地の 3 クラスを持つ質的データで、90m 解像度である。Landcover も同様に質的データであり、林、草地など全 22 クラスに分類され、300m 解像度である。LandScan は人口分布を表し、GDP は田口らによって推定され

Automatic Levee Detection for Global Hydrological Models -A Case Study for the United States-
M.Ikegawa, T.Hascoet, V.Pellet, M.Watanabe, X.Zhou, Y.Tanaka, T.Takiguchi, D.Yamazaki, Kobe University, The University of Tokyo

Copyright ©2022 Information Processing Society of Japan.
All Rights Reserved.

た [7] データであり、共に 1km 解像度である。

最終的な入力特徴量は、これらのデータセットにいくつかの前処理を施すことで得られる。まず、質的データはワンホットエンコーディングにより数値的なデータに変換する。また、0 から 100% の範囲の値を持つ GSWO は 0、0-10、10-90、90-100、100% の 5 クラスの質的データとしてまとめる。最後に、出力と対応する堤防データセットは 10km 解像度と最も粗いため、平均を取ることで調整する。

2.3 問題定義

本研究では、以下の式

$$F(X) = P(levee|X), \quad F : R^{14} \longmapsto [0,1] \quad (1)$$

に示すような 2 値クラス分類問題を考えることで、堤防予測モデルを構築する。

3 実験

堤防データセットに対して安定した評価を行うため、本実験では ECE とカリブレーション曲線を評価指標に用いる。本実験では 2 つの調査を行う。まず、仮説を検証するため、洪水リスクと人間活動の特徴量の直積をとったものを入力特徴量に加える特徴量エンジニアリングを行う。ロジスティック回帰、MLP、Random Forest、XGBoost、LightGBM の 5 つのモデルの評価を行なった。そして、学習、テストデータを、*i.i.d.* と地域差を保った (regional)2 つの方式でサンプリングし、地域差がモデルに大きな影響を与えることを示す。

表 1 は、特徴量エンジニアリングによってほとんどのモデルの精度が向上し、仮説が堤防の存在を適切にモデル化できることを示す。また、図 2 は、regional なサンプリングがモデルのカリブレーションを悪化させることを示し、モデルの予測は地域差に大きく影響を受けることが確認できた。

4 まとめ

米国全域を対象に、堤防位置を予測するモデルの構築を行なった。洪水リスクと人間活動を定量化し、複数の実験を行うことで、堤防に対する良い事前知識を得ることができた。

表 1 *i.i.d.* サンプリングでの ECE[%] の結果。元々の特徴量 (OR.) と特徴量エンジニアリング (FE.) の 2 つの特徴量 (Fe.) を比較。

	Models				
Fe.	LR	RF	Lgbm	XgB	MLP
OR.	0.16	0.26	0.22	0.05	0.21
FE.	0.15	0.13	0.16	0.06	0.13

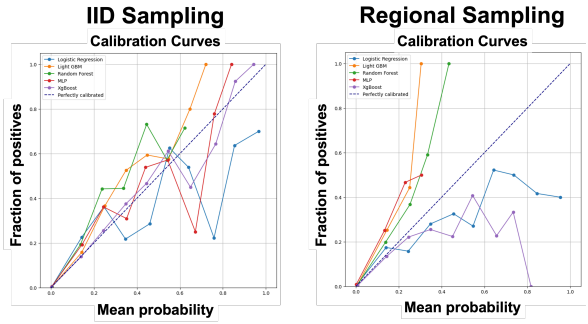

図 2 *i.i.d.*(左) と regional(右) なサンプリングでのカリブレーション曲線。それぞれの色はモデルの違いを表し、対角線 (点線) に近づくほど良いモデルとなる。

参考文献

[1] Y. Tanaka and D. Yamazaki: The Automatic extraction of physical flood protection parameters for global river models. Journal of Japan Society of Civil Engineers, Ser. B1(Hydraulic Engineering), Vol. 75, No.2, 1099-1104 (2019)

[2] X. Zhou, *et al.*: The uncertainty of flood frequency analyses in hydrodynamic model simulations. Nat. Hazards Earth Syst. Sci., 21, 1071–1085 (2021)

[3] JF. Pekel, *et al.*: High-resolution mapping of global surface water and its long-term changes. Nature 540, 418–422 (2016)

[4] https://www.usgs.gov/centers/wgsc/science/global-food-security-support-analysis-data-30-m-gfsad

[5] http://www.cec.org/north-american-environmental-atlas/land-cover-2010-landsat-30m/

[6] https://landscan.ornl.gov

[7] (Ryo Taguchi)The Importance of Business Interruption Losses in Global Scale Flood Economic Damage Analysis (Unpublished)

Copyright ©2022 Information Processing Society of Japan.
All Rights Reserved.