

表 1 データセットの詳細

説明変数	環境データ	温度[°C], 湿度[%], 飽差[kPa], CO ₂ [ppm], PPFD[$\mu\text{mol m}^{-2} \text{s}^{-1}$]
	植物データ	NIR/VR (葉面積)
目的変数	光合成速度[$\mu\text{mol p}^{-1} \text{s}^{-1}$], 蒸発散速度[$\text{g p}^{-1} \text{min}^{-1}$]	
データ数	11874 (光合成速度) 10689 (蒸発散速度)	
データ周期間隔	10分 (目的変数), 1分 (説明変数)	
データ収集期間	2021/01/06 ~ 2021/07/31	
データ収集時間	0:00 ~ 24:00	

4 検証実験

4.1 データセット

イチゴを対象とし、2021年1月6日から2021年7月31日にかけて静岡県農業技術産学官連携研究開発センターの温室内でデータセット収集を行った。表 1 にデータセットの詳細を示す。

説明変数は環境データ 5 変数、植物データ 1 変数の計 6 変数を持ち、各変数に当該時点から 5 時点前までの 6 時点の時間情報を含めた 6x6 の 2次元データを 1つのシーケンシャルデータとして入力とする

4.2 実験方法

リサンプリング処理を行わないデータ (以下, RAW) と既存手法の RU, RO, SMT, WRS を適用したデータ, CREAMER を適用したデータを用いて、機械学習の推定精度を比較した。CREAMER のリサンプリング部は任意の手法を選択可能なため、RU と RO を組込んだ CRM_RU_RO と、RU と SMT を組込んだ CRM_RU_SMT を用いた。各モデルの推定値について平均絶対誤差 (MAE)、二乗平均平方根誤差 (RMSE) で精度比較を行った。

4.3 実験結果

図 3 に光合成速度、図 4 に蒸発散速度推定におけるリサンプリング手法ごとの各評価値、図 5 に各リサンプリング手法適用後の目的変数の分布を示す。

光合成速度推定においては、RAW の MAE (0.2223)、RMSE (0.3414) を基準として、CRM_RU_RO は MAE (0.2246)、RMSE (0.34)、CRM_RU_SMT は MAE (0.2224)、RMSE (0.3332) となっており、CREAMER を用いることで、MAE は劣るものの、RMSE は最大で 2.4% ほどの誤差を削減できていることが分かる。MAE よりも RMSE のほうが大きな誤差値に敏感に反応するため、希少データへの追従性が確認できる。

蒸発散速度推定においては、RAW の MAE (0.0466)、RMSE (0.0708) を基準として、CRM_RU_RO は MAE (0.0437)、RMSE (0.0664)、CRM_RU_SMT は MAE (0.0445)、RMSE (0.0664) となり、CREAMER を用いることで MAE, RMSE とともに最大 6.2% ほどの誤差削減に成功

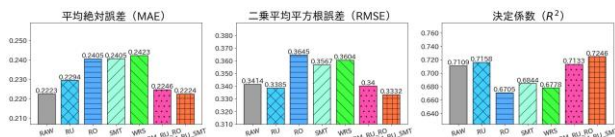


図 3 光合成速度の推定結果

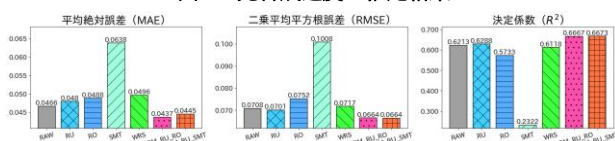


図 4 蒸発散速度の推定結果

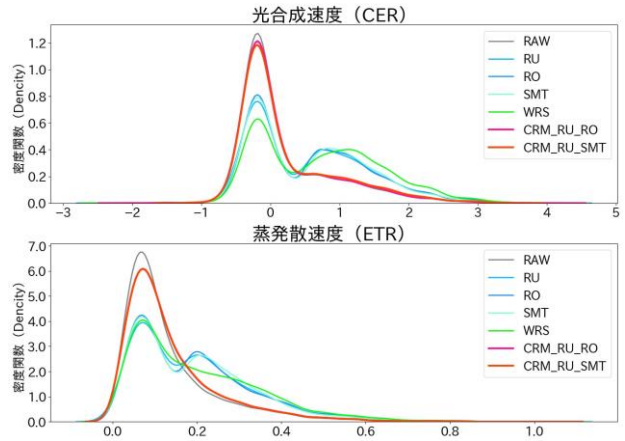


図 5 リサンプリング後の目的変数の分布

し、推定精度の向上を確認した。また、光合成速度、蒸発散速度推定の両方で、CREAMER が他の既存手法よりも精度が高いことが分かる。

各リサンプリング手法を適用することで、目的変数の分布が変動し、不均衡性が低減していることが分かる。RU, RO, SMT (以下, 3 手法) は二峰性の分布になっており、閾値を用いない WRS は満遍なく分布していることが分かる。一方 CREAMER は k 個のクラスタに分割しているため、 k 峰性の分布であると考えられる。これは RAW と 3 手法の中間の分布といえる。

5 おわりに

本研究では、光合成速度と蒸発散速度の推定において、データの不均衡性を解消するリサンプリング手法である CREAMER を改良し、1D-CAE によって時系列性を保ちながらリサンプリングする学習手法を用いることで、他の既存手法よりも高精度に推定することを確認した。

今後の方針として、引続き時系列を考慮した学習手法の検討を進めるとともに、農業データのみならず、他分野のデータに対する CREAMER の有効性を検証していく。

参考文献

- [1] Torgo, L., Branco, P. and Ribeiro, R.P.: Resampling strategies for regression, Expert systems, Vol.32, No.3, pp.465-476 (2015).
- [2] Branco, P., Torgo, L. and Ribeiro, R.P.: Pre-processing approaches for imbalanced distributions in regression, Vol.343, pp.76-99 (2019).
- [3] Torgo, L., Branco, P., Pfahringer, B., et al.: Smote for Regression, Portuguese conference on artificial intelligence, pp.378-389 (2013).
- [4] Chawla, N.V., Bowyer, K.W., Hall, L.O., et al.: SMOTE: Synthetic Minority Over-sampling TEchnique, Journal of artificial intelligence research, No.16, pp.341-378 (2002).
- [5] 藤浪一輝, 大石直記, 二俣翔, 峰野博史: 栽培データの分布不均衡性を考慮した植物生理状態の推定: 施設栽培イチゴデータにて, 情報処理学会論文誌, Vol.63, No.11, pp.1653-1664 (2022).
- [6] Pelleg, D. and Moore, A.: X-means: Extending K-means with Efficient Estimation of the Number of Clusters, ICML, Vol.1, pp.727-734 (2000).
- [7] Zhao, Z., Guo, S., Xu, Q., et al.: G-Means: A Clustering Algorithm for Intrusion Detection, ICONIP 2008: Advances in Neuro-Information Processing, pp 563-570 (2008).
- [8] Corduneanu, A., and Bishop, C.M.: Variational Bayesian Model Selection for Mixture Distributions, Artificial Intelligence and Statistics, Vol.2001, pp.27-34 (2001).
- [9] Aslam, J.A., Pelehov, E. and Rus, D.: The Star Clustering Algorithm for Static and Dynamic Information Organization, J. Graph Algorithm Appl., Vol.8, No.1, pp.95-129 (2004).
- [10] Cho, K., Van Merriënboer, B., Gulcehre, C., et al.: Learning phrase representations using RNN encoder-decoder for statistical machine translation, arXiv preprint arXiv:1406.1078 (2014).