Noname manuscript No.
(will be inserted by the editor)

An evolutionary approach with feature engineering for drought prediction

Leonardo Goliatt *

Received: date / Accepted: date

Abstract **Keywords** Drought prediction · differential evolution · calibration · Heston model · Monte Carlo simulation

1 Introduction

Mehr et and collaborators [2] presented a new hybrid model, called ENN-SA, for spatiotemporal drought prediction. In ENN-SA, an Elman neural network (ENN) is conjugated with simulated annealing (SA) optimization and support vector machine (SVM) classification algorithms for the standardized precipitation index (SPI) modeling at multiple stations.

Mehr et al [1] presented a tree-based model, namely Fuzzy Random Forest (FRF), for one month ahead Standardized Precipitation Evapotranspiration Index (SPEI) classification and prediction with a noteworthy application in ungauged catchments.

2 Material and Methods

2.1 Study area and dataset

2.2 Feature engineering

2.3 Lasso linear model

The Lasso [3] minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant. Because of the nature of this constraint it tends to produce some coefficients that are exactly 0 and hence gives interpretable models.

* L. Goliatt
Computational Modelling Program, Federal University of Juiz de Fora, Brazil
E-mail: leonardo.goliatt@ufjf.edu.br

2.4 Time-series cross-validation

The use of machine learning models in data from time series requires special care due to their peculiarities. One must choose the test set within a period of time after the training set. Otherwise, some information may leak from the training set to the training set, compromising the learning process of the machine learning model. A more robust solution is to operate similarly to k-fold cross-validation but in a time-ordered way. The figure below illustrates the cross-validation procedure for time series. By training and adjusting the model in training set for each fold and averaging the errors in the test sets, we can obtain an unbiased estimate of the model's performance.

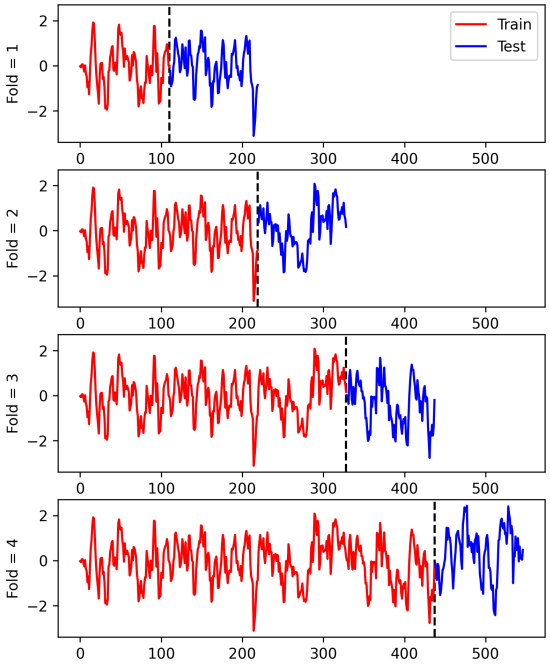

Fig. 1: Time series cross validation.

2.5 Differential evolution

Differential Evolution (DE) is a stochastic optimization algorithm, based on a population of solutions, which operates through computational steps similar to those employed by most Evolutionary Algorithms. DE is simple and easy to implement, and it has some features such as superior performance in relation to accuracy, convergence speed and robustness, and few control parameters.

The objective function, to be minimized by the differential evolution, is written as follows:

$$f(x) = \text{RMSE}(x) * \left(1 + \frac{\beta}{N_{feat}} \sum_{i=1}^{N_{feat}} x_i^{MB}\right) \quad (1)$$

The parameter β indirectly controls the complexity of the expressions generated for the linear model. From this equation, we observe when $\beta = 0$ the resulting expression from the candidate solution is not penalized by its length. As β increases, expressions with a greater number of terms are more penalized than expressions with few terms. A proper choice of β is crucial to balance the accuracy of the expression and the interpretability of the associated prediction model.

3 Computational Experiments

4 Conclusion

References

1. Danandeh Mehr, A., Tur, R., Çalışkan, C., Tas, E.: A novel fuzzy random forest model for meteorological drought classification and prediction in ungauged catchments. Pure and Applied Geophysics **177**(12), 5993–6006 (2020). DOI 10.1007/s00024-020-02609-7. URL https://doi.org/10.1007/s00024-020-02609-7
2. Danandeh Mehr, A., Vaheddoost, B., Mohammadi, B.: ENN-SA: a novel neuro-annealing model for multi-station drought prediction. Computers & Geosciences **145**, 104,622 (2020). DOI https://doi.org/10.1016/j.cageo.2020.104622. URL https://www.sciencedirect.com/science/article/pii/S0098300420306002
3. Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) **58**(1), 267–288 (1996). URL http://www.jstor.org/stable/2346178