Review tc-2021-194

TITLE

Convolutional Neural Network and Long Short-Term Memory Models for Ice-Jam Prediction

RECOMMENDATION

Major revision

REVIEWER

John Quilty

GENERAL COMMENTS

This paper explores for the first time CNN, LSTM, and CN-LSTM deep learning models for river ice jam classification (jam or no jam). The authors test their framework on a dataset stemming from Quebec, Canada where various meteorological variables are used as explanatory variables in a multivariate time series classification problem. The authors show that the CN-LSTM model provides the best performance.

In general, the paper is well-written (although careful editing is still needed) but the motivation for adopting these very complex models is not clear. Extremely complex deep learning models (with over 100, 000 parameters, see section 3.3.2) were developed but not compared to much simpler and widely used machine learning methods (e.g., 3 layer multilayer perceptrons, decision trees, support vector machine) that may be more appropriate for the dataset (small and structured). Further, most of the paper is spent on describing the deep learning model details but very little space is devoted to exploring why the deep learning models perform well on the given dataset or how these results can be interpreted with respect to the physical processes under study.

Further, the structure of the paper needs to be re-worked as model development details are mixed in with the Results and Discussion section.

Other important points are raised in the list of SPECIFIC COMMENTS and TECHNICAL CORRECTIONS sections below, where recommendations for improving the paper are given.

I think the paper needs a major revision before it is considered for publication and I would be happy to re-evaluate the paper should the authors wish to submit a revised manuscript.

SPECIFIC COMMENTS

1. Abstract: define terms CNN, LSTM, etc. at first use in the abstract.

2. Introduction: First, why are deep learning methods needed for this problem more so than traditional machine learning methods (e.g., 3-layer multilayer perceptron, decision trees, support vector machine). Second, there are numerous applications of CNN, LSTM, and their hybrid versions applied in hydrology (Althoff et al., 2021; Apaydin et al., 2020; Barzegar et al., 2021, 2020; Kratzert et al., 2018; Wunsch et al., 2020; Zhang et al., 2018). It would be good to mention the application of such models in hydrology to show that their use is well established

within this domain and to highlight that none of these methods have been explored for ice jam prediction.

3. L183: '…and multiplying it by tanh…' Is this fully correct? tanh is an activation function and therefore needs an input to evaluate, would it not be more correct to write '…and multiplying it by tanh(something)'?

4. L217-219: do you mean that earlier experimentation showed that MinMaxScaler lead to the most accurate results?

5. Figures: the vast majority of figures in this paper are taken from other sources. While the figures appear to be properly cited it may be worthwhile to consider creating some new figures specific to the dataset and models employed in this work.

6. The authors refer to 'loss' (i.e., a loss function) without defining it until section 2.5.1.5, where they then switch to the term 'cost function'. It would be good for the authors to: a) clarify early on that a loss function for neural networks is similar to an objective function for process-based hydrological models (to make this term more approachable for a wide audience) and b) use consistent terminology (i.e., choose loss or cost function).

7. Eqs. 2 and 3: specifically mention which equation pertains to the L1 and L2 regularization. The authors should also explicitly state the cost function (see also comment 6 above to ensure consistent terminology is adopted) or at least point to where it is discussed in more detail within the text.

8. Section 2.5.2: what do you use for identifying the optimal architecture? Grid-search, random-search, Bayesian optimization, …?

9. Eqs. 4-6: why are RELU and Sigmoid in italic font but not tanh? I believe the Sigmoid function is referred to earlier as σ, consistency should be maintained here (and elsewhere in the text, e.g., see comment 6 above).

10. L330: I do not think 'mini-batch' has been discussed yet, nor should it be assumed that most readers will be familiar with the term. What is a 'mini-batch' and what is its purpose?

11. 2.5.3: see comments 6 and 7 above.

12. Much of sections 3.1 – 3.2 (and their sub-sections) do not belong in a Results and Discussion section, these sections are more related to Methodology or Model Development. Some sections such as 3.1.8 include results, but the majority of these sub-sections in 3.1 and 3.2 do not.

13. L517-518: the authors should include at least a single reference that corroborate this statement.

14. L547-551: no new information should be provided in the Conclusion section, why is project 'DAVE' not mentioned earlier in the paper? It would seem best to mention this information in the introduction of the paper to better motivate its goals and objectives.

15. The authors spend most of the paper describing all the components of the different models and spend little time focussing on the significance of the results. The authors may wish to only describe the main components of the models and shift non-essential information to a supplementary material file.

16. Why did the authors not compare these much more complicated models (with, as noted in 3.3.2, 100's of thousands of parameters!) to a simple 3-layered multilayer perceptron model, or a decision tree model (e.g., random forests or eXtreme Gradient Boosting), or a support vector machine (which is ideal for small datasets)? Based on the type of dataset (small and structured) it seems the previously mentioned methods might be more appropriate, may result in better performance with much less complexity (e.g., fewer parameters and hyper-parameters), and have lower training times (e.g., likely minutes rather than hours). Without lack of a meaningful benchmark it is difficult to justify the use of these very complicated models (CNN, LSTM, CN-LSTM) that took nearly the whole paper to describe.

17. There is little emphasis placed on exploring *why* certain models performed better than others and how this relates to the physical system under study. It's great if a new model provides high accuracy for modelling a physical system but understanding why the model may work better than others is also important to explore.

TECHNICAL CORRECTIONS

- L12 (and elsewhere in the text, e.g., L13): remove 'the' before 'ice-jam'.
- L19: 'validation and generalization **sets**'? Why not use test set instead of generalization set, as it is more common in the ML community?
- L35: remove 'to' before 'jam'.
- L49: 'carefully' instead of 'wisely'.
- L50: include 'classifier' after 'kNN'.
- I will mostly stop providing editorial remarks at this point…the paper should be carefully edited.
- L57: '…that use multiple layers where nonlinear transformation is used to extract…'
- L219: I suppose the brackets around the scaled variable should be black instead of red.
- L266: 'covariate **shift**'?
- L310: is 'drown' the right word to use here?
- L336: 'over-training'.

REFERENCES

Althoff, D., Rodrigues, L.N., Bazame, H.C., 2021. Uncertainty quantification for hydrological models based on neural networks: the dropout ensemble. Stoch. Environ. Res. Risk Assess. 35, 1051–1067. https://doi.org/10.1007/s00477-021-01980-8

Apaydin, H., Feizi, H., Sattari, M.T., Colak, M.S., Shamshirband, S., Chau, K.-W., 2020. Comparative Analysis of Recurrent Neural Network Architectures for Reservoir Inflow Forecasting. Water 12. https://doi.org/10.3390/w12051500

Barzegar, R., Aalami, M.T., Adamowski, J., 2020. Short-term water quality variable prediction using a hybrid CNN--LSTM deep learning model. Stoch. Environ. Res. Risk Assess. 1–19.

Barzegar, R., Taghi Aalami, M., Adamowski, J., 2021. Coupling a Hybrid CNN-LSTM Deep Learning Model with a Boundary Corrected Maximal Overlap Discrete Wavelet Transform for Multiscale Lake Water Level Forecasting. J. Hydrol. 126196. https://doi.org/https://doi.org/10.1016/j.jhydrol.2021.126196

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., Herrnegger, M., 2018. Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. Hydrol. Earth Syst. Sci. 22, 6005–6022. https://doi.org/10.5194/hess-22-6005-2018

Wunsch, A., Liesch, T., Broda, S., 2020. Groundwater Level Forecasting with Artificial Neural Networks: A Comparison of LSTM, CNN and NARX. Hydrol. Earth Syst. Sci. Discuss. 1–23. https://doi.org/10.5194/hess-2020-552

Zhang, D., Lin, J., Peng, Q., Wang, D., Yang, T., Sorooshian, S., Liu, X., Zhuang, J., 2018. Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. J. Hydrol. 565, 720–736. https://doi.org/10.1016/j.jhydrol.2018.08.050