Application of Artificial Neural Network in Streamflow Forecasting

Mian Xiao(mxiao18@stanford.edu) & Shanni You(shanni30@stanford.edu)

Introduction

Streamflow forecast is a complicated but important problem in hydrology. Traditional forecast approaches for this problem are the conceptual physical models and the statistical empirical models. However, both physical models and the empirical models could fail due to extreme conditions or lack of necessary data. To resolve these technical issues, neural network models are widely studied recently as an important alternative approach, since they could represent a more complicated nonlinear process.

Dataset & Features

- Runoff to be predicted is assumed to be a nonlinear function of the historical runoff and rainfall several steps earlier.
- Rainfall and runoff series
  - Time span: 2003 - 2012
  - Location: Leaf river basin, Collins, LA
  - Validate: 2008 - 2012
- Normalize the input into range (0,1)

Evaluation Setting

\[
RMSE = \sqrt{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}
\]

RMSE for selected validation points by a threshold:
- 1500: peak RMSE (Qvalid > 1500)
- non-peak RMSE (Qvalid <= 1500)

Framework based on tensorflow, and also utilized the scikit-learn package.

Results

- The overall accuracy for the Standard network is pretty good.
- Standard network performs better than the LSTM network in terms of all types of RMSE. In general, the Standard network exhibits a better accuracy.
- The prediction error at the peaks is more significant, particularly for the LSTM model. This indicates that a potential improvement to the model is to customize the design of LSTM cell architecture so that it could better represent the time-dependency mechanism for this type of problem.
- The difference of the best parameter combination for both models are relatively close to each other, except the learning rate.

Parameter studies

- Fix the maximum epoch at 200, and best learning rate for each model. By retraining the model with selected hyperparameters, use the minimum RMSE for on the validation set as a criterion for best model selection.
- Input Features: m: lay-days for rainfall m: lay-days for runoff
- Patterns with m-n = 1,2,3 are constructed to better generalized.
- Hyper-parameters:
  - Learning rate: Crucial for model convergence. Considered as fixed value in this case, to further study other parameters.
  - Best range: (0.008,0.01) for Standard network; 0.001 for LSTM network.
  - hidden size: too large or too small is not applicable. Once the best lag-day feature determined, it will allow larger range of hidden size.
- In conclusion, the dominant parameter is lay-day feature.

Conclusion and future work

- Satisfactory performance in Standard feed-forward network prediction. LSTM network is not very accurate compared to its standard counterpart.
- Fewer features needed than conceptual physical formula. Useful if the knowledge for prediction is very limited.
- Majority of RMSE was contributed by the errors at peak flows.
- More features should be investigated in future to adapted models to a relatively larger region.

Reference