Ensembled River Level Forecasting

Orion J Junkins

Faculty Advisor: Dr. Patrick Donnelly

Abstract

This project proposes an ensemble of Recurrent Neural Network models to forecast river flow rate using weather data from throughout the catchment. Small, localized models are trained for each weather data source and predictions are combined to predict the flow rate at a chosen site. This ensemble approach allows for a larger input dataset while avoiding the pitfalls of large dimensionality. The final ensemble model yields a MAPE of 8.57%, which is 10.02% lower than the best performing localized model.

Introduction

Although river flow rate is directly correlated to weather in the catchment, these systems are too complex to model with traditional statistical methods. Recurrent Neural networks are well-suited for this task, however, the amount of data required for accurate forecasting is prohibitively restrictive. The accuracy and real-world utility of prior approaches have been constrained by issues that arise from large dimensionality. This project employs an ensemble approach to mitigate the issues imposed by “big data” and maximize both accuracy and utility.

Dataset

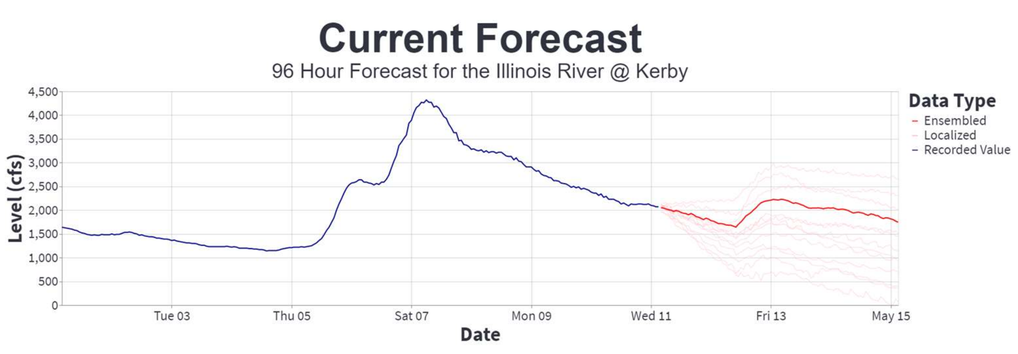
Modeling river flow rate requires access to flow data for the river itself, and weather data for locations throughout the drainage. Flow data (measured in $\mathrm{ft}^3/\mathrm{second}$) is available at hourly precision through USGS river gauges. Historical and forecasted weather data (including temperature, rainfall, snowfall, etc.) is available at hourly precision through the Open Weather Map API. For this project, the Illinois river in southwestern Oregon is used as a case study. The level at the Kerby USGS gauge is forecasted with weather data from 12 locations.

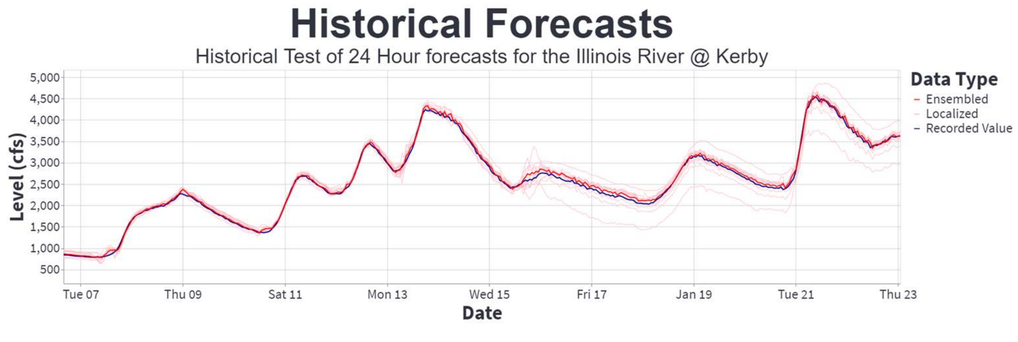
Methods

The first stage of this project implemented a baseline model based on previous studies. Specifically, it implemented a grid-search optimized gated recurrent unit (GRU) model using 12 weather data features from a single location. Although accuracy was reasonable, achieving a mean absolute percentage error (MAPE) of 18.60%, localized weather systems limited performance; varying the source location used for weather data caused massive variation in both the accuracy and consistency of generated forecasts.

The second stage sought to address this variation by leveraging all available weather data across 12 distinct locations, rather than a single site. However, directly using a dataset of this size was deemed ineffective and impractical; complex relationships are harder to identify, and extensive time and resources are needed for training.

An ensemble approach was adopted to handle this dimensionality. Rather than modeling all available data at once, each weather location is modeled independently in a small, localized model with only 12 features. A secondary ensemble model takes in predictions from each localized model and combines them using Linear Regression into a single master prediction. Both the localized models and the regression model are lightweight, but the full ensemble considers all available data to provide more accurate forecasts.





Results

As expected, the final ensemble model outperforms all localized models. For 24-hour advance forecasts generated by localized models, the average MAPE is 42.05%, while the lowest MAPE is 18.60%. When these localized models are ensembled, the MAPE decreases to 8.57%. This is a 33.48% improvement over the average localized model and a 10.02% improvement over the best localized model.

Discussion

This research demonstrates the effectiveness of ensemble models for forecasting river flow rate. Despite large dimensionality, high accuracy has been achieved without compromising real-world utility. However, these results are limited; this architecture has only been applied to a single river. Future extensions of this work will explore performance across more drainages. Future extensions will also explore the domain generalizability of this approach by exploring other weather-related modeling problems, such as avalanche prediction or forest fire monitoring.

References

Fu, M., Fan, T., Ding, Z., Salih, S. Q., Al-Ansari, N., & Yaseen, Z. M. (2020). Deep Learning Data-intelligence model based on adjusted forecasting window scale: Application in daily streamflow simulation. IEEE Access, 8, 32632–32651.

Lejeune, O. (2020, September 19). Using machine learning to predict Rhine water levels. Medium.

Zealand, C. M., Burn, D. H., & Simonovic, S. P. (1999). Short term streamflow forecasting using Artificial Neural Networks. Journal of Hydrology, 214(1-4), 32–48.

Oregon State University Cascades

COLLEGE OF ENGINEERING - COMPUTER SCIENCE

Funding and Support from the Layman Fellowship