Deployable Ensembles for River Forecasting

Zachary Bochanski, Doug Dennis, Orion Junkins, Melissa Swearingen

Faculty Advisor: Patrick Donnelly

Introduction

Anticipation of rapid changes in river flow, especially floods, is a critical task for many stakeholders. While river flow rate is directly correlated to weather, these systems are difficult to model with traditional statistical methods. Neural networks are well-suited for this task, but the amount of computation required for accurate forecasting prohibits real-world deployment. We hypothesize that an ensemble of networks can outperform single network solutions while and will be easier to train and deploy [1]. Here, we demonstrate this in practice and at scale. We present a containerized ensemble of neural network that run scheduled inference on AWS with results published on a public front end.

Methods

Flow data from USGS river gauges serve as a target variable. Weather data, sampled from several locations across the catchment area above the gauge, serve as features. Sampling sites are selected as a grid with 0.1-degree separation for both latitude and longitude.

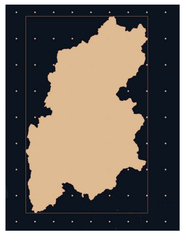

Figure 1. A catchment map with the corresponding grid of weather sampling sites for a single river system (Middle Fork of the Salmon)

A single "contributing model" is trained on a variety of weather features (e.g., rain, snow, soil saturation, temperature, etc.) from each weather data sampling site. A second stage `Linear Regression Ensemble` model is trained on the outputs of these contributing models in order to produce a more accurate prediction. Each individual model in this system is small and lightweight, but the entire ensemble accurately models the system both temporally and spatially.

Results

Three separate models were trained and deployed to a cloud host (AWS). These models generated inferences daily since late February 2023, with only a single disruption in service, caused by a fluctuation in an upstream data provider. The deployed ensemble collections have been able to maintain up to twenty-one individual contributing models and the financial costs of the daily inference process on AWS have been manageable.

Individual contributing models are showing mean percentage errors as low as 22%.

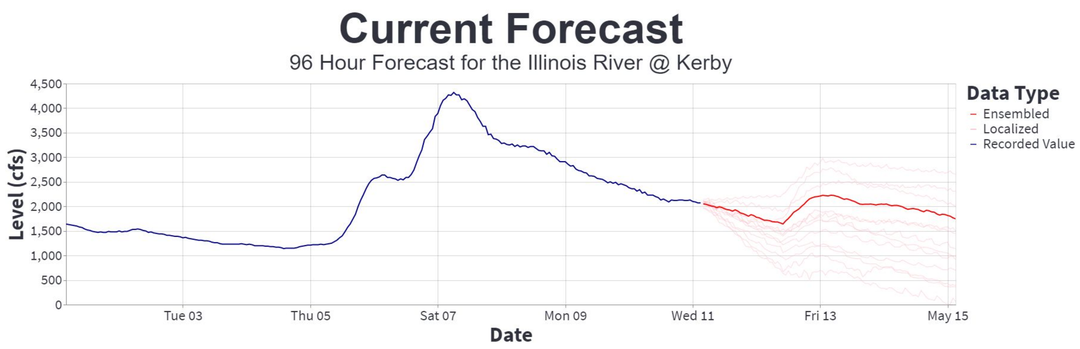

Figure 2. A sample forecast for the Illinois River near Kerby Oregon. Individual localized contributing models are shown in light red, while the ensembled forecast is shown in dark red. True level up to the time the forecast was issued is shown in blue.



Figure 3. The home page for our current public-facing front end is hosted on AWS Amplify. A chart shows forecast data for the currently selected river. The forecast data shown represents the most recently issued forecast.

Currently, our forecasts update every 24 hours. The website also provides usage information, a summary of the project, as well as information about the contributing students, faculty, and organizations.

Other forecasted rivers can be searched for and selected from the table below. Users have the ability to "favorite" certain rivers, pinning them to the top of the list.

Cloud Deploy

Deploying a production-level machine learning model requires the automation of data collection, processing, inference, and delivery. We automated this entire pipeline using Amazon Web Services (AWS). Custom-trained models are deployed as Docker images to the cloud. At scheduled intervals, serverless functions fetch real-time weather and river level data and run inferences with trained models. Generated forecasts are stored in the cloud and made accessible via a RESTful API endpoint.

Front End

`www.level-sage.com` allows users to interact with the river flow forecasting system from a computer or mobile device. The website is built with React and deployed using AWS web hosting tools. Users can easily view real-time predicted river flow data with interactive graphical visualizations. The front end improves the usability and accessibility of the river flow forecasting system, by providing a streamlined user experience for non-technical end users.

References & Acknowledgments

[1] Donnelly, P., & Junkins, O (2022). Short-Term River Forecasting with a Stacked Ensemble of Tributary Models, ICFSP. pp. 189-193.

[2] A. Mosavi, P. Ozturk, and K. Chau (2018). "Flood prediction using machine learning models: Literature review," Water, 10(11), p. 1536.

[3] T. C. Pagano et al. (2014)., "Challenges of operational river forecasting," Journal of Hydrometeorology, 15(4), pp. 1692–1707.

[4] V. Nourani, "An emotional ANN approach to modeling rainfall runoff process," (2017). Journal of Hydrology, vol. 544, pp. 267–277.

This work is supported with funding by Layman Fellowship Program.

Oregon State University Cascades