A Meta-Learning Approach for River Water Level Prediction

Authors: Asma Slaimi, Michael Scriney, Susan Hegarty, Fiona Regan, Noel O'Connor
Institution: Dublin City University
Email: asma.slaimi2@mail.dcu.ie

- River water level prediction is critical for water resource management and flood control. Our research proposes a meta-learning approach to automatically select the best machine learning model for a given dataset, improving prediction accuracy and aiding decision-making.

- Our approach involves three stages:
  - We prepare and integrate relevant datasets.
  - We train candidate models (e.g., linear models, dense neural networks) on a training set and evaluate the performance of each model on a validation set.
  - The meta-learner uses this information as a an input and selects the best-performing model for the given dataset.

Data Preparation and Integration

Spatiotemporal Integration of Heterogeneous Environmental Data

Data Sourcing

Extract Transform and Load (ETL)

Data Extraction

Spatio Temporal Data Integration (STDI)

Mapping

Integration

Data Transformation

Data Loading

- The proposed data integration architecture provides a comprehensive solution for integrating heterogeneous environmental data from multiple sources and leveraging it for prediction.

Model Training and Evaluation

Prediction Models

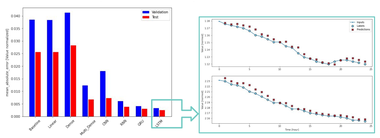
- Our study involved the use of eight different models to predict water levels at 70 locations throughout Ireland. These models included the baseline model, linear regression, dense neural network, multidense neural network, convolutional neural network (CNN), recurrent neural network (RNN), gated recurrent unit (GRU), and long short-term memory (LSTM).
- Additionally, we conducted two main prediction scenarios to evaluate the performance of each model.

Prediction Scenarios

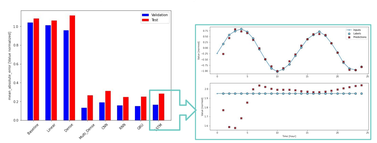
- Both single-time step predictions (1 hour ahead) and multiple-time step predictions (24 hours ahead) were employed
- Two prediction scenarios were considered:
  - Single-feature prediction using only historical water level data
  - Multiple-feature prediction using water and weather data integration
    - Precipitation (mm), temperature (C), water level (meter), and other features were included

Results

Example 1: Banagher hydrometric station results (Best LSTM Model performance)



Example 2: Giles Quay hydrometric station results (Worst LSTM Model performance)



- The validation results were confirmed with the test.
- The predictive performance of the NN-based models is better than the baseline and linear, reflecting NN's ability to follow the general pattern of the water-level.
- Experiment results indicate that although some models outperformed others, there is no overall best model.

Model	Best Validation	Best Test
LSTM	32	29
GRU	24	26
MultiDense	7	6
RNN	5	6
CNN	2	3

Meta-Learning and Model Selection

- ❑ The meta-learner takes as input the performance of each candidate model on the validation set and outputs the best-performing model.
- ❑ It can be seen as a "learning-to-learn" algorithm that learns the optimal way to select the best model for a given dataset.

Current Experiment and Preliminary Results

Model Selection Method	Accuracy	Weighted Avg F1-Score
Naive Bayes	59%	0.50
Logistic Regression	69%	0.40

- Our meta-learning approach using Naive Bayes and Logistic Regression achieved an overall accuracy of 59% and 69%, respectively, in selecting the best-performing machine learning model for river water level prediction. Naive Bayes performed slightly worse than Logistic Regression, which selected the Dense model as the best-performing model.

- These results demonstrate the potential of our approach in automating model selection and improving prediction accuracy in river water level forecasting.

Next Steps

➔ Next steps include further experiments to evaluate the proposed meta-learning approach on larger datasets and with more complex models.

➔ Advanced techniques such as active learning and transfer learning can also be incorporated to further enhance the model selection process.

In summary, our proposed meta-learning approach for model selection can help improve the accuracy of river water level prediction and aid decision-makers in making informed decisions for water resource management and flood control.