Towards an Integrated Problem-Solving Environment for Hybrid Numerical Models with Statistical Learning Components

Lei Jiang1,3, Qi Fan2,3, Gabrielle Allen1,3, and Q. Jim Chen2,3

Departments of Computer Science1 and Civil and Environmental Engineering2 and Center for Computation and Technology3, Louisiana State University, Baton Rouge, LA

Abstract

The motivation of our work originates from an engineering problem. In Louisiana, tropical cyclones (hurricanes), especially Hurricane Katrina in 2005 and Hurricane Gustav in 2008, have caused severe loss of life and property damage. New dynamic storm surge prediction is being used operationally to predict the effect of hurricanes, as accurate and timely predictions are essential for decision-makers to deploy appropriate evacuation plans. However, the dynamic and multi-physics nature of this problem presents new challenges for research in coastal science and engineering.

Artificial intelligence techniques are more and more frequently being used by domain scientists in solving domain-specific problems. Especially, as arrays of networked sensors and experimental apparatus are deployed, an important issue arises of how to utilize this real-time observational data to improve the accuracy and efficiency of conventional numerical solutions. So, statistical learning models are involved in this scenario. The need of an integrated problem-solving environment for hybrid models also arises.

Introduction

Numerical models play an essential role in solving real-life problems in many areas of engineering, environment sciences and physics, where simulations are used to understand and predict natural phenomena. However, for complex systems such as forecasting the impacts of hurricanes in coastal science, the large number of physical factors involved leads to time-consuming numerical solutions and model errors from the uncertainty in practical problem settings.

Statistical learning models, trained from historical data, have already proved to be useful for accelerating and verifying numerical models. This poster describes our first step to develop hybrid numerical models that include statistical learning components that lead to a more adaptive methodology, where information learned from real-time data supplies dynamic feedback to the deterministic numerical model.

Methods

The problem-solving environment will integrate various parameters and models involved in the full problem setting and help users select an optimal combination using artificial intelligence techniques, instead of domain experts manually selecting and experimenting with a number of factors themselves, which is otherwise difficult due to the need of both domain knowledge and learning models.

The environment is developed using the Cactus framework (https://www.cactuscode.org), with the advantage that outputs of each model in each stage can be dynamically integrated with other models. It contains several thorns of numerical and statistical learning models involved: different types and settings of artificial neural network (MultiANN), genetic algorithm (GeneticAlg), and existing numerical models in Cartesian or standalone numerical outputs. MultiANN generates a series of neural network emulations interactively with GeneticAlg, with the provision of myriad input combinations. The verified output of neural network emulations, as the statistical learning component, along with results from numerical solutions, could provide with valuable information in improving current prediction methods.

Results

The experiment results above are for Scenario I. The figure on the left shows the results for time-series prediction using an individual back-propagation neural network. There is a time lag between the observed peak value and predicted peak value due to the rapid increase when hurricane comes very near. Then, another neural network is trained for error prediction, where the outputs of Holland wind model are part of the neural network inputs. The figure on the right, where the predicted error is compared with actual error, clearly illustrates that the time lag could be counteracted. The RMSE of the synthesis of ensemble learning is 0.07 m, and the error at the point of peak value is then reduced to less than 0.35 m.

Conclusion

The poster describes the composition of an integrated problem-solving environment (PSE) for hybrid numerical models with statistical learning components using Cactus framework and two scenarios of application using the PSE. The experiment results, as our progress in the first stage, show the robustness in predicting short-term storm surge level in different modes. In the near future, we are going to integrate more data from numerical simulation (e.g. using ADCIRC model) into the PSE as training datasets and perform more in-depth error analysis. As our research goes deeper, more numerical and learning models are to be coupled, opening new possibilities in both domain-specific research and the area of artificial intelligence.

Acknowledgements

The research work was supported by NSF RII (Louisiana’s Research Infrastructure Improvement Strategy) CyberTools project, NSF ALPACA (Cactus Tools for Application Level Profiling and Correctness Analysis) and COMI (Coastal & Ocean Modeling Infrastructure) project from DoD and the Office of Naval Research contributed to the establishment of experimental testbed. The thanks also go to LONI (Louisiana Optical Network Initiative) for providing computational resources.