Prediction Science:
the 5th Paradigm Fusing the Computational Science and Data Science

by
Takemasa Miyoshi
RIKEN Center for Computational Science, Kobe, Japan
takemasa.miyoshi@riken.jp

High-performance computation (HPC) has consistently been extending the capability of more realistic simulations in various application fields. As a simulation becomes more precise and accurate, real-world data play an increasingly important role in improving the simulation. Moreover, real-time use of data will synchronize the simulation with the real world and enable predicting the future.

Numerical Weather Prediction (NWP) is a successful example of fusing the HPC-based “Big Simulation” with real-world “Big Data” through a method known as data assimilation. As the NWP model became more and more accurate, the relative importance of data assimilation has increased. In the past two decades, data assimilation was considered as important as the NWP model itself. As a result, data assimilation grew rapidly to be a major field in meteorology.

In the past BDEC meetings, I have presented our work on Big Data Assimilation in NWP. We assimilated roughly two orders of magnitude more data from a new radar system with a 100-m-mesh NWP model, 100 times more grid points per area than a typical high-resolution NWP system at 1-km resolution. We examined the feasibility of local severe storm prediction refreshed every 30 seconds, 120 times more rapidly than a typical hourly-update system. This will bring a revolution to weather forecasting, while such a revolutionary NWP system requires intense FLOPS and I/O speed to meet the strict real-time requirement for the 30-second update frequency.

The general concept of data assimilation is to fuse an HPC-based simulation in the cyberworld with the real-world data. This is a realization of a broader concept of cyber-physical system with HPC and Big Data. Here we propose to generalize the success of NWP and data assimilation to broader simulation fields. We have a high-precision simulation synchronized with real world and select the best future scenario based on the simulation with quantified uncertainties. This is what we call the Prediction Science, the 5th paradigm fusing the Computational Science (3rd science) and Data Science (4th science), extending the capability of prediction and control to large and complex systems with which traditional approaches have difficulties.

To create the new Prediction Science, it is essential to bring together the process-based model simulation and data science approaches such as AI and machine learning techniques through modern advanced mathematics including the uncertainty quantification (UQ). This requires new-generation computing resources with high computing capacity with fast networking and large storage access. Simulations will use more FLOPS with more I/O with increasing variety and volume of real-world data. AI and machine learning techniques will be integrated with data assimilation, but these have very different computational requirements. Namely, AI and machine learning techniques are efficient with

GPUs, while simulation and data assimilation codes are efficient with CPUs most of the time. It would not be easy to use GPUs for high-precision simulation and data assimilation. For integrated Prediction Science applications, we will need to have access to both architectures in a seamless manner.

I would like to bring to light the potential future directions of “Prediction Science” fusing Computational Science and Data Science and to discuss what computational needs are expected to enable prediction in broader areas. To make this happen, it will be essential to have synergistic interactions among computer scientists, experts in broad application areas, and theoretical and mathematical scientists.