Calibration of Building Energy Models: Supercomputing, Big-Data and Machine-Learning

Jibonananda Sanyal, Joshua New, Richard Edwards

Energy and Environmental Sciences Directorate

Building Energy Modeling (BEM) is an approach to model the energy usage in buildings for design and retrofit purposes. EnergyPlus is the flagship Department of Energy software that performs BEM for different types of buildings. The input to EnergyPlus can often extend in the order of a few thousand parameters which have to be calibrated manually by an expert for realistic energy modeling. This makes it challenging and expensive thereby making building energy modeling unfeasible for smaller projects. In this poster, we describe the "Autotune"' research which employs machine learning algorithms to generate agents for the different kinds of standard reference buildings in the U.S. building stock. The parametric space and the variety of building locations and types make this a challenging computational problem necessitating the use of supercomputers. Millions of EnergyPlus simulations are run on supercomputers which are subsequently used to train machine learning algorithms to generate agents. These agents, once created, can then run in a fraction of the time thereby allowing cost-effective calibration of building models.

The individual trained machine agents are generated by performing machine learning on large parametric simulations for the different classes of standard DOE buildings across different climate zones. At the core of the Autotune methodology is a set of multi-objective machine learning algorithms that characterize the effect of individual variable perturbations on EnergyPlus simulations and adapts the given model to match its output to the supplied sensor data. Once machine learning agents are tuned and available, the computational cost of tuning a typical user's building model is reduced to matter of a few hours using widely available desktop computational resources.

The system is currently being demonstrated to match a subset of 250 sensors of 15-minute resolution data in a heavily instrumented residential building in addition to DOE's standard reference building models for a medium sized office, a warehouse, and a stand-alone retail building. Further, the simulations comprise of three vintages (old, new, and recent) of the DOE commercial reference buildings across 16 different cities representing the different ASHRAE climate zones and sub-zones. The anticipated data set is about 270 TB.

The project team has been awarded computational time (both competitive and discretionary) on supercomputing resources at the National Institute of Computational Sciences (*Nautilus*, and *Kraken*), San Diego Supercomputer Center (*Gordon*), and the Oak Ridge Leadership Computational Facility (*Titan* and *Frost*). Good scalability results have been observed on both shared-memory supercomputers and the massively parallel *Titan* supercomputer.