Research Statement

By Jie Li

Objective. The objective of my thesis research is to design and develop a machine learning method to build an accurate energy efficiency model for data centers and to automate data center management and optimize energy usage efficiency based on the model.

Background and Motivation. Arguably, we human are creating data at a much faster rate than ever before, and the demand for processing and storing data is rapidly growing. Any entity that generates or uses data has the need of data centers at some level, including government, telecommunication companies and social networking service providers such as Google and Facebook. At the same time, more and more information are moving into the cloud instead of running or storing them on our local computers. The cloud is not out of nowhere, and it is the data center that maintains the hardware and software at remote locations where the clients and their customers can access them via the Internet. However, **a growingly critical challenge for the data center is its power consumption**. According to Arman Shehabi's report [1], data centers in the U.S. consumed an estimated 70 billion kWh in 2014, representing about 1.8% of total U.S. electricity consumption. Data centers in U.S. are projected to consume approximately 73 billion kWh in 2020. **Growing energy costs and environment responsibility require the data center industry to improve its operational efficiency**.

Challenges and Proposed Approach. The modern data center is a complex system and involves interaction of multiple mechanical, electrical, and control systems. It is difficult to use traditional engineering formulas to build the Power Usage Effectiveness (PUE) model accurately because these systems interact with each other and the interdependencies are nonlinear. Furthermore, the number of possible operating configurations make it even more difficult to understand and optimize the PUE model. There is a growing desire to investigate and develop a powerful and automated solution to characterize data center energy usage and to further optimize the configuration and operation dynamically.

In this research, we propose to design and develop a machine learning method to build an accurate energy efficiency model for data centers and, based on the model, we will also design and develop strategies to optimize energy usage efficiency of data centers and to automate data center management on the fly. Machine learning methods have been proven effective in numerous applications and scenarios, such as in self-driving cars, online recommendation systems, fraud detection systems, etc. However, few studies exist yet for investigating and developing a machine learning based method for data center energy efficiency models and management solution.

Figure 1 shows an overview of the proposed research. This study is composed of three major research tasks and an additional task of summarizing all findings for dissemination. The first task is to collect data sets to feed into machine learning algorithms. Our solution is to collect data from sensors in data center servers via a Redfish protocol and API [2] [3]. Redfish is an open source, open industry standard specification for hardware management. It provides system managers with an interface to the hardware status data and it is a powerful way to centrally manage and automate server configuration, deployment, update and monitoring. With a comprehensive data collection through Redfish API, we can train an accurate model for PUE, and more importantly we provide a portable solution across data centers with different vendors.

The second task of this research is to build an accurate PUE model for the complex data centers. We plan to investigate whether a multi-layer perceptron (MLP) [4] model can be used for this situation and produce expected results. As shown in Figure 1, data are collected and fed into MLP, based on these features (neurons) X = x_1, x_2, ..., x_n. MPL can learn a non-linear function approximator. In our case, the input X includes the workloads, number of fans, spinning speed of fans, electrical power used, temperature in data center, etc. We will identify a comprehensive list of input parameters for this specific problem. The f(x) is the output containing the variable of interest, and it can represent a range of metrics like fan speed, HVAC power etc. that we wish to optimize.

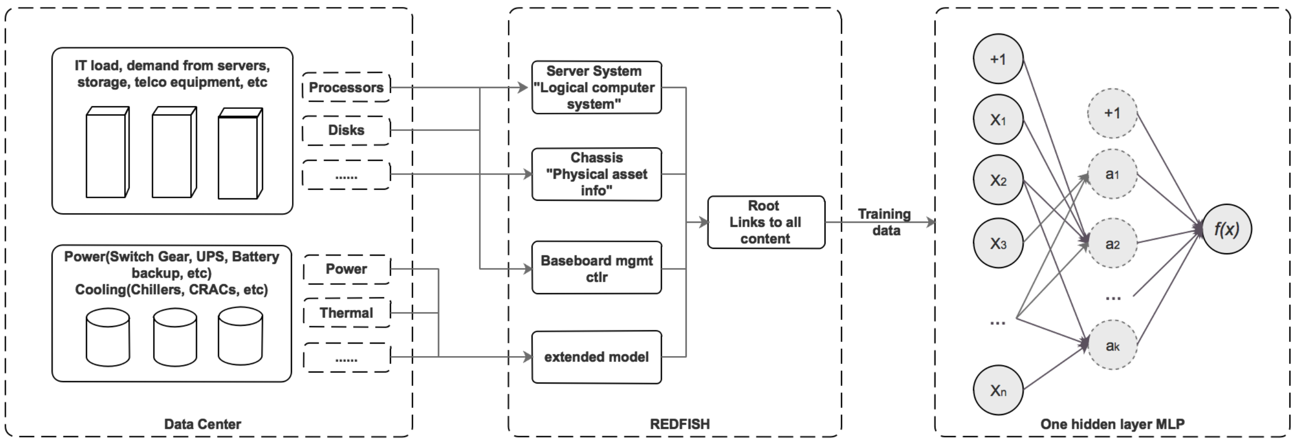

Figure1. Automated Data Center Power Management and Optimization Framework

The third task of this research is to develop strategies to optimize data center power management and improve energy usage efficiency. These strategies can be lowering fan speed on the fly if the monitored thermal values are below a threshold, or putting components of data centers, e.g. disks into a sleeping mode to conserve energy (and waking up them when needed). We plan to investigate these potentially viable solutions and will study the efficacy of each possible solution and a combination of these strategies as a collective solution.

Preliminary Work. We have developed in-depth understandings of the Redfish protocol and conducted initial research in gathering data points via Redfish API. We have also developed understandings of various machine learning models in the context of this research study. These preliminary studies have provided us a solid foundation for carrying out the proposed research.

Deliverables and Timeline. A major deliverable of this research is a machine learning based PUE model for data centers and strategies of optimizing data center power consumption. The timeline of the proposed research is as follows. **Oct.**: Task 1, training data sets collection; **Nov.**: Task 1, preprocess the training data; **Jan.**: Task 2, use mockup data to build a machine learning model; **Feb.**: Task 2, tune machine learning model based on the training data and test on the real data center environment; **Mar**: Task 3, experiment strategies of optimizing power consumption; **Apr.**: Task 4, summarize and disseminate research results.

References

[1] A. Shehabi, S.J. Smith, N. Horner, I. Azevedo, R. Brown, J. Koomey, E. Masanet, D. Sartor, M. Herrlin, W. Lintner. "United States Data Center Energy Usage Report". Lawrence Berkeley National Laboratory, Berkeley, California. LBNL-1005775.
[2] Data-Intensive Scalable Computing Laboratory, [Online] https://discl.cs.ttu.edu/doku.php
[3] Redfish Protocol and API, [Online] https://redfish.dmtf.org/
[4] Multi-layer Perceptron Model, [Online] http://scikit-learn.org/stable/modules/neural_networks_supervised.html