

# 다중 기계 학습을 활용한 데이터 센터의 냉방 에너지 절감 시스템에 관한 연구

장현철\*

\*고려대학교 컴퓨터정보통신대학원 빅데이터 융합학과

e-mail : [justinjang87@korea.ac.kr](mailto:justinjang87@korea.ac.kr)

## A Study on the Cooling Energy Saving System for Data Centers Using Multi-Machine Learning

Hyun-Cheol Jang\*

\*Dept. of Big Data Convergence, Korea University Graduate School of Computer and Information Technology

### 요 약

최근 클라우드 시스템 환경이 점차 늘어남에 따라 데이터 센터(IDC) 구축이 점차 늘어나고 있다. 데이터 센터는 최근 부각하고 있는 4 차 산업 영역에서 사물 인터넷(IoT), 자율주행차 등에서 처리될 대용량 데이터로 인한 이를 처리하는 중요한 역할을 담당하고 있다. 데이터센터 운영에는 대량의 에너지가 필요하다. 수 많은 컴퓨터에서 발생하는 열에너지를 처리하기 위하여 대량의 전력 냉방 에너지를 소비하고 있다. 냉방 공조 운영은 데이터 센터 운영에 중요한 역할을 한다. 이유는 많은 컴퓨터를 가동하는 비용보다 부대 시설로 운영되는 냉방 에너지를 보다 많이 소비하는 현상까지 발생하고 있다. 이에 최근 데이터 센터 냉방 공조 운영을 효율화하는 것에 연구를 맞추고 있다. 본 논문에서는 냉방 공조 운영 효율화 하도록 하기 위해서 다중 기계 학습을 활용한 데이터 센터의 냉방 에너지 절감 시스템을 제안하고자 한다. 기존의 단순 알고리즘을 활용하여 머신 러닝의 모델 구현 방식이 아닌 다중의 기계 학습을 통하여 최적화된 모델을 일일 배치로 생성하여 예측을 하는 시스템이다. 본 시스템을 통하여 사전에 최적화된 냉방 운영을 하여 기존 데이터 센터의 운영되는 과다 냉방을 감축 시켜 에너지를 절감해주는 기능을 제공한다. 본 논문 시스템 연구 결과는 폭발적으로 늘어나고 있는 데이터 센터의 에너지 효율화에 기여할 수 있고, 클라우드 사업에서 경쟁력을 줄 수 있는 운영 시스템 방안을 제시한다.

### 1. 서론 및 선행 연구 소개

최근 4 차 산업 혁명의 발전으로 인해 사물 인터넷과 자율 주행 자동차 등에서 처리될 대용량 데이터로 인하여 이를 처리할 데이터 센터 요구 증대되고 있다. 하지만 데이터 센터에서는 막대한 에너지가 생성되므로 이를 효율적으로 관리할 필요가 있다. 이에 산업계와 학계에서는 이와 관련하여 다양한 프로젝트와 사업을 진행하고 있다.

산업계에서는 구글이 선도적으로 머신 러닝 기술을 활용하여 데이터 센터 최적화 운영에 기술을 활용하고 있다. Machine Learning Applications for Data Center Optimization[1] 프로젝트에서는 “Generic Three-layered neural network”을 활용하여 PUE (Power Usage Effectiveness)[2]을 효율화하기 위해 미래 기온과 압력을 예측하고 이에 따른 필요한 냉각 작업을 추천하는 방식으로 에너지를 효율화 하고 있다. 그리고 또한 슈나이더 일렉트릭에서는 “Data Center Operation: Cooling Optimize” 라는 서비스를 통해 에너지 효율화를 진행하고 있다. 슈나이더 일렉트릭 서비스에서는 핫존(HOT-Zone), 쿨존(Cool-Zone) 2 개의 구역을 나누

어서 Hot-Zone 을 효율적으로 관리를 목표하고 있다. 슈나이더 일렉트릭에서는 서버에서 온도 센서를 설치하여 온도 정보를 획득하고, 항온항습기 정보도 수집하여 이를 통하여 인공지능 엔진에 입력하여 최적화된 항온항습기 조정 값을 결정하여 원격 자동 조정을 통하여 Hot-Zone 의 온도를 낮추어 에너지를 관리하며, 학습에 필요한 데이터를 1 주일 간격으로 학습하여 관리하는 것으로 소개하고 있다.

그리고 학계에서는 이와 관련한 내역으로 [3-5] 등의 연구에서 머신 러닝을 활용하여 데이터 센터의 온도 관리를 최적화하여 에너지를 효율적으로 관리하는 내용을 제시하고 있다.

본 프로젝트의 기술적인 운영 방식은 머신 러닝을 활용하여 최적의 운영을 위하여 효율적인 권장 온도를 예측하여 PUE 을 최적화 시키며, 동적 최적 모델 선정 방법으로 모델을 생성하여 데이터 수집에 따라 일일 단위로 모델을 학습하여 생성하는 방식으로 진행되고 있다는 점이 기존 산업계와 학계의 차이점이다. 이를 활용하면 알고리즘 하나로 한번에 모델을 만들어 내는 방식으로 학습시키는 것과 달리 매일

매일 학습을 진행하여 모델을 업데이트 시키며 다중 알고리즘 중에서 효과적인 알고리즘을 선택하여 모델을 일일 단위로 갱신하여 선정한다는 점에서 자동화 특징 구현을 포함하고 있다. 일 단위로 배치하여 데이터를 수집하여 모델을 업데이트하여 학습시키는 기능은 최신의 정보를 새롭게 수집하여 모델을 학습하도록 하여 데이터가 부족한 상황에서 데이터와 함께 모델을 업데이트 시킬 수 있는 점에서 유연하게 대처할 수 있다.

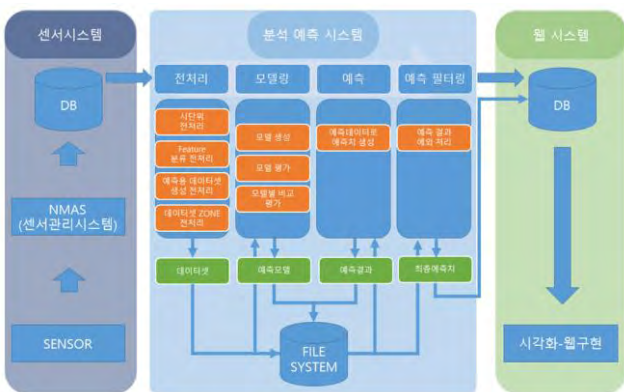
**2. 시스템 소개**

본 시스템은 PUE 최적 관리를 위하여 AI 기반의 기술인 머신 러닝을 활용하여 냉방 에너지 절약하는 시스템 구축을 제안한다. 이 시스템 운영 방식은 그림 1 과 같이 진행이 된다. IDC 에서 수집 데이터를 정제하여 모델을 학습하고 향온향습기의 전력사용량을 예측하고 이에 따른 전력량 기반 하에 예측 기온을 예측하고 이를 통하여 운영 효율적인 에너지 관리를 위한 권장 설정온도를 제시하여 전력소비량을 감축 시켜서 PUE 를 효율적으로 운영하는 방식으로 시스템 진행되는 개요를 가진다.



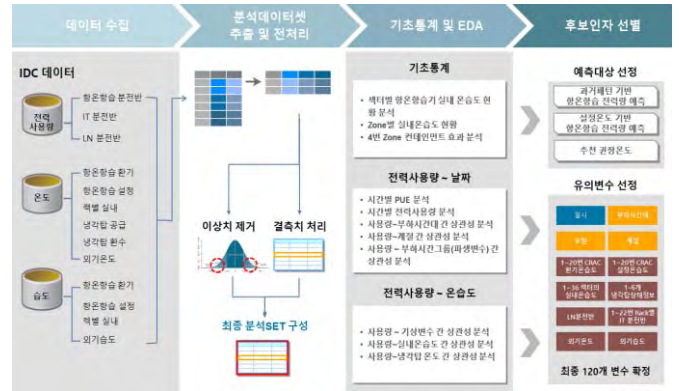
(그림 1) 제안 시스템의 전반적인 개요

본 시스템 구현의 구성도는 그림 2 와 같다. 시스템에서 주요 개요는 사전 단계에 센서 시스템 부분은 데이터를 획득하는 단계이고, 사후 단계인 웹 시스템은 서비스를 시각화하여 제공하는 시스템의 단계가 있다. 그리고 마지막으로 분석 예측 시스템은 최종 결과물을 제공하기 위한 예측 시스템 부분이 있다.



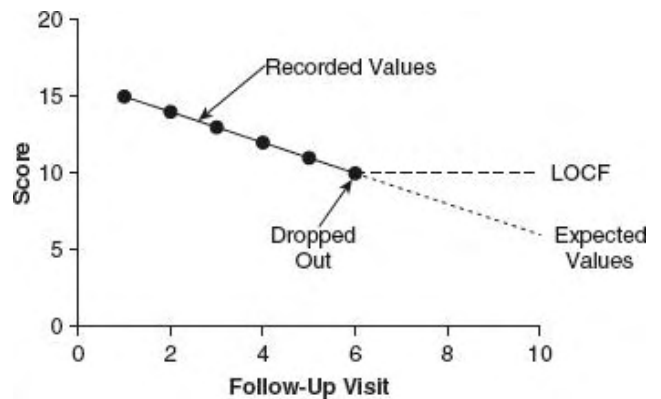
(그림 2) 시스템 구현의 구성도

**3. 데이터 전처리 단계**



(그림 3) 데이터 전처리 과정

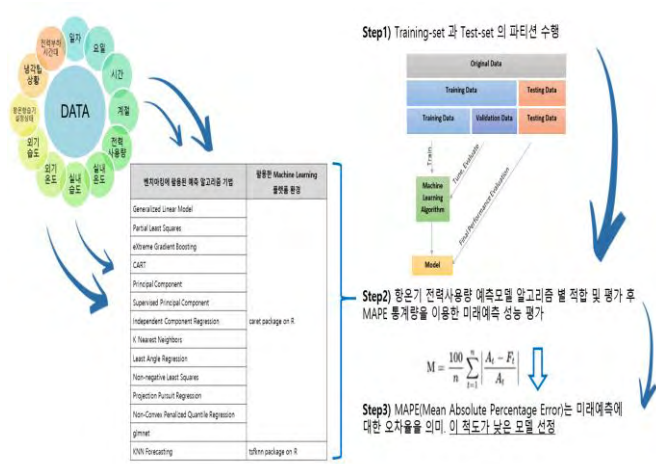
그림 3 은 데이터 전처리 단계에 대한 설명으로 장비 센서로부터 수집 되는 데이터 “전력 사용량”(향온향습기 분전반, IT 분전반, LN 분전반), “온도”(향온향습기 환기 온도, 향온향습기 설정 온도, 서버 랙별 실내 온도, 냉각탑 공급 온도, 냉각탑 환수 온도, 외기 온도),”습도”(향온향습기 환기 습도, 향온향습기 설정 습도, 서버 랙별 실내 습도, 외기 습도) 과 같은 크게 3 가지 형태의 정보를 받아 이상치 제거, 결측 정보를 처리하여 최종 분석 SET 형태로 구성한다. 이를 통하여 분석 SET 을 기초 통계를 통하여 현황 분석을 진행하였고 날짜, 온 습도와 전력사용량 간의 상관 분석을 통하여 데이터간의 상관 관계를 분석하여 최종적으로 120 개 변수를 선정하여 모델의 학습 정보로 구성하였다. 분석 데이터 셋에서 미래 예측 데이터 셋을 만드는 과정은 이월 대체(LOCF-Last Observation Carried Forward)[6] 기법을 활용하여 미래 예측 데이터 셋을 만든다. LOCF 는 최종 MISSING 시점에서 마지막 관측치를 앞으로 보낸다 라는 의미로 다른 말로 마지막 값으로 결측되는 값을 대체한다라는 의미로 현재 프로젝트 시스템에서는 최초 예측 전력량을 예측하기 위해서는 온습도의 변화도 중요한 요소이므로 이를 확인하기 위해서는 예측 데이터 셋이 필요하여 이 방법을 선택하여 미래 예측 데이터 셋을 사용하였다.



(그림 4) Last Observation Carried Forward

#### 4. 데이터 모델 학습 및 선정 단계

본 시스템에서 강조하는 내역인 동적 최적 모델 선정 방법으로 모델을 생성하여 데이터 수집에 따라 일 단위로 모델을 학습하여 생성하는 방식으로 그림 5 와 같은 방식으로 진행이 된다.



(그림 5) 데이터 모델 학습 및 선정 단계

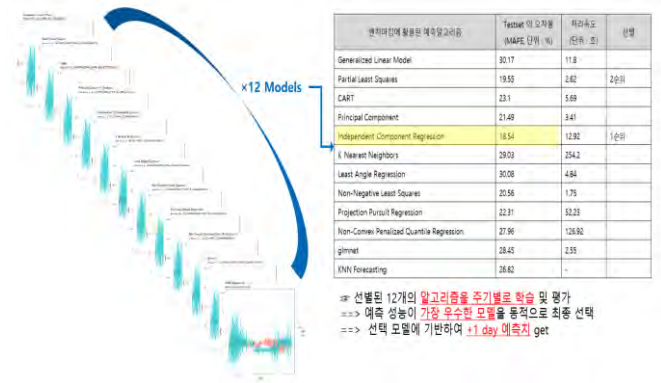
최종 모델 선정까지는 3 단계의 구조를 가지고 있다. 처음에는 전처리 하여 얻어진 Dataset 을 Train-set, Test-set 으로 구성한다. 이때 훈련 데이터를 훈련 시키는 과정에 Caret 패키지를 활용하여 다중의 알고리즘을 훈련시킨다. 이때 훈련 과정에서 MAPE (Mean Absolute Percentage Error)를 통하여 성능의 정확도를 측정한다.

(도식 1) MAPE 공식

$$M = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

MAPE 는 실제 값과 예측 값의 상대 오차의 비율을 모두 더한 다음 기간 수로 나눈 값이다. 이를 활용한 이유는 직관적으로 백분율(%)로 표현이 되기에 오차 정도를 시각적으로 시스템적으로 계산이 효율적이며 해석하기가 쉽다. 따라서 모델의 성능평가에는 MAPE 를 활용하여 진행하였다.

자세한 모델 선정방식은 그림 6 와 같은 방식으로 가장 우수한 알고리즘 모델을 선정한다. 예를 들면 12 가지의 알고리즘 방식으로 훈련 데이터를 통하여 학습시켜서 12 가지의 알고리즘 방식으로 모델을 생성하고 생성된 모델을 시험 데이터에 MAPE 방식으로 계산하여 오차율을 확인하고 이 중에서 가장 효과적인 모델을 선정하는 방식으로 시스템은 구현되고 있다. 그리고 또한 일일 배치 방식으로 매일 같이 모델을 학습을 진행하여 새로 생성된 정보를 획득하여 정보 모델을 생성하여 진행한다.



(그림 6) 모델 선정 방식 과정

#### 5. 결론 및 향후 연구 과제

본 프로젝트 시스템은 다중의 모델을 통하여 우수한 모델을 자동화하여 선정하는 방식을 제안하여 하나의 단수 알고리즘만을 제공하는 방식이 아닌 다중의 모델 중 우수한 모델을 선정하여 예측 모델을 선정하는 기능을 제시한다. 그리고 또한 일일 배치 방식으로 학습 진행하여 모델을 생성하여 최신의 센서 정보를 바탕으로 데이터를 수집하여 최종적으로 모델을 학습시키는 방식을 취하고 있다. 2 가지 방식을 통하여 최종적으로 데이터 센터에서 발생하는 열 에너지를 효율화하기 위한 시스템을 제시한다.

향후 과제로써 데이터 셋을 만드는 과정에서 머신러닝을 활용하여 미래 예측 정보를 대치하는 방식과 보다 빠른 시스템 구현을 위하여 시스템 메모리 관리 부분을 추가적으로 연구하며 다양한 구조의 데이터 센터 구조에서도 바로 도입을 할 수 있는 방안에 대하여 향후 연구하여 본 연구 시스템에서 보안 점들을 향후 과제에서는 향상 시킬 예정이다.

#### 참고문헌

- [1] Jim Gao, "Machine Learning Applications for Data Center Optimization"
- [2] ISO/IEC 30134-2-Information Technology-Data centres-Key performance indicators -Part 2, 2016
- [3] Georgios Smpokos, "On the Energy Consumption Forecasting of Data Centers Based on Weather Conditions: Remote Sensing and Machine Learning Approach"
- [4] Nevena Lazic, "Data Center Cooling using Model-predictive Control"
- [5] Jun Yang, "AI-Powered Green Cloud and Data Center"
- [6] Neil J. Salkind, "Encyclopedia of Research Design"