

# 탄소중립이행을 위한 기계학습 기반의 전력수요예측 방법론 고도화 연구

박창대, 고한결, 신우영\*  
한국과학기술기획평가원 R&D예산정책센터  
\* (교신) e-mail : sinwy93@kistep.re.kr

## Research on advancing electricity demand forecasting method using machine learning to implement carbon neutrality

Wooyoung Shin  
Korea Institute of Science&Technology Evaluation and Planing(KISTEP)

### 요 약

본 연구는 전력 수요 예측을 통해 탄소 중립 이행에 이바지하는 것이다. 전력 데이터의 특징상, 변동성이 크고 계절성, 요일 요인 등 다양한 외부 요인에 의해 예측이 어렵기 때문에, 정교한 수요예측 중요성이 점차 부각 되고 있다. 특히 본 연구에서 주목한 부분은 학습데이터의 평균을 설명하는 회귀식을 예측할 때 나타나는 편향이다. 데이터의 평균을 설명하는 회귀식은 전력 수요 예측에 있어서 상대적으로 적게 쓰거나 많이 쓰는 패턴을 반영하기 어렵다는 것이다. 그러므로 전력 수요의 분위를 설명하는 분위수 회귀분석(Quantile Regression)을 활용하여, 여러 분위 회귀식을 기존 평균 회귀식과 함께 재평균화하는 Composite Variation Method(CVM)을 제안하였다. 해당 방법론을 사용할 경우, 편향을 상충하게 되어 기존 모형 대비 약 10% 정도의 성능 향상을 이끌어낼 수 있었다.

### 1. 서론

마치도록 한다.

현재 우리 사회는 기후변화와 온실가스 배출에 대한 우려가 증대되면서, 탄소 중립이 중요한 과제로 인식되고 있다[4, 6]. 탄소 중립이란 우리의 활동으로 인한 탄소 배출량을 줄이고, 발생한 탄소 배출을 제로로 만들어 환경에 미치는 영향을 최소화하는 일련의 행위를 의미한다[2]. 여기에 전력 부문은 이러한 탄소 중립화의 핵심 요소 중 하나로, 전력 수요 예측과 효율적인 에너지 사용은 전체적인 탄소 중립화 노력의 핵심 요소가 되겠다. 최근 기계학습 기술의 발전으로 인해 더 정확하고 효율적인 전력 수요 예측이 가능해짐에 따라[13], 더욱 정교한 수요예측 모형을 요구하고 있다. 따라서 본 연구에서는 기존의 전력 수요 예측 방법론을 고도화하는 방법론을 제안하고자 한다. 이를 통해 정책 수립 방안에 이바지함으로써 탄소 중립 이행에 보탬이 되고자 한다.

본 연구의 구성은 다음과 같다. 2장에서는 새로운 방법론 제안을 위한 기초 모형까지 함께 설명하고, 3장에서는 기존 방법론 대비 새로운 방법론의 성능 향상을 설명한다. 그리고 마지막 4장에서는 해당 고도화된 수요예측 방법론을 통해 이바지하는 방안을 제시하고 한계점 및 추후 계획을 설명하며

### 2. 방법론

본 절에서는 새로운 전력 수요 예측 방법론 제안을 위해 필요한 방법론을 설명하도록 한다.

#### 2.1 분위수 회귀분석(Quantile Regression)

분위수 회귀분석은 오차의 분포에 대한 분위수( $\tau$ )를 설명하는 회귀식을 찾는 방법론으로[9], 아래의 식을 최소화하는 회귀식  $f(x_i)$ 를 찾게 된다.

$$\tau \cdot \max(y_i - f(x_i), 0) + (1 - \tau) \cdot \max(f(x_i) - y_i, 0)$$

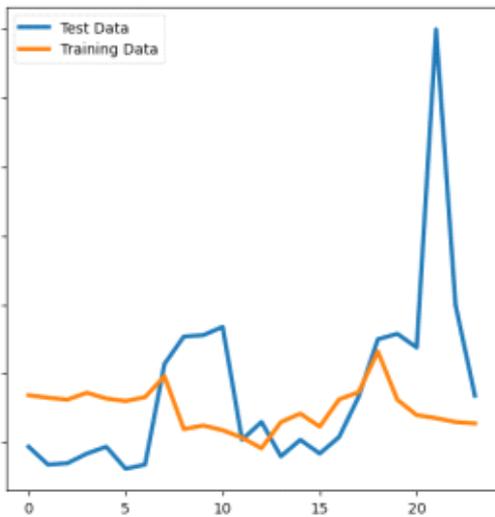
해당 모형의 장점은 L2 손실함수를 사용하여 회귀식을 찾는 것에 반해, 이상치에 영향을 적게 받게 된다. 이상치에 대한 손실함수가 제곱으로 커지므로, 상대적으로 이를 최소화 하려다 보니 본래 추정해야 할 회귀식에 편향이 발생하게 된

1) 초기 제안된 방식은 선형이므로  $x_i, \beta$ 로 표현하여  $\hat{\beta}(\tau)$ 를 추정하게 되나, 본 연구에서 사용할 모형이 비선형이므로  $f(x_i)$ 로 표현함

다. 그에 반해 분위수 손실함수는 선형적으로 그 크기가 증가하므로, 영향을 적게 받는다고 할 수 있겠다. 이에 더해 분위수 회귀분석은 오차에 대한 가정을 하지 않으므로, 정규성 위배에 따른 영향을 받지 않게 된다. 다시 말해 변동성이 지나치게 커지거나, 이분산성(Heteroscedasticity)이 존재할 때 적합하다고 할 수 있겠다. 이러한 점으로 미루어 보았을 때, 전력 수요량이 변동성이 크고 예측하기 어려운 이상치가 종종 나타나는 경우가 있으므로[5], 분위수 회귀분석을 사용하는 것이 적절하다고 할 수 있겠다. 그리고 이를 활용한 새로운 추정 방법을 다음 소절에서 소개하도록 한다.

## 2.2 Composite Variation Method(CVM)

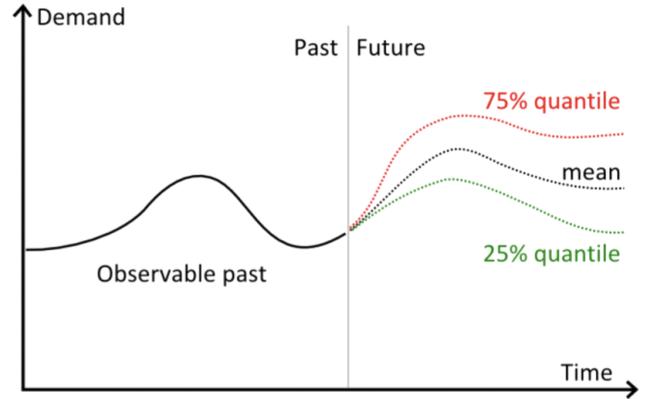
본 소절에서는 분위수 회귀분석에 기반한 새로운 방법론을 설명하도록 한다. 전력 수요는 본래 다양한 요인들로 인해 영향을 받게 되므로 예측이 어렵다고 알려져 있다[8, 10]. 그리고 단적으로 [그림 1]에서와 같이, 예측의 어려움을 간접적으로 알 수 있다.



[그림 1] 학습데이터의 평균(주황색) 및 예측 데이터의 전력 수요(파란색) 차이

[그림 1]에서 모형을 통해 학습에 사용되는 전력 수요량의 평균값과 예측해야 하는 전력 수요량의 평균 간의 차이가 크다는 것을 확인할 수 있다. 다시 말해, 학습데이터로 고도의 방법론을 적용하여 예측할지라도 차이가 클 수 있다는 것을 암시한다. 낮 시간대와 밤에는 적게 사용하는 패턴을 학습하지만, 반대되는 패턴이 나타나는 것처럼, 평균을 활용한 데이터 학습을 통해 예측하기는 어렵다고 볼 수 있겠다. 그에 따라 학습데이터에서 전력 수요가 많은 패턴과 적은 패턴을 동시에 고려할 필요가 있고, 이에 대한 도식도는 [그림 2]와 같다. 학습을 통해 평균 회귀식을 추정하는 다음, 25분위 회귀식과

75분위 회귀식을 추정하여 평균화한 값을 예측값으로 사용하게 된다. 이를 통해 적은 전력 수요와 많은 전력 수요를 함께 고려함으로써 예측에 대한 변동성을 반영할 수 있게 되며, 평균 회귀식을 추정할 때 발생하는 이상치로부터의 편향을 방지할 수 있게 된다.



[그림 2] Composite Variation Method 도식도

## 3. 데이터 분석

### 3.1 활용 데이터

본 분석에 활용한 데이터는 Statistics Society of Canada에서 주취한 온타리오주의 시간별 전력 수요를 예측하는 데이터를 활용하였다.<sup>2)</sup> 해당 데이터는 Ontario from the Independent Electricity System Operator에서 제공하는 주거용 + 산업용 + 상업/기관용 + 농업 + 교통에 해당하는 시간별 전력 수요 데이터와 Natural Resources Canada에서 제공하는 Comprehensive Energy Use Database에서 연간 전력 수요 데이터를 결합하였다. 그리고 보조 변수로 ETH Zurich와 Imperial College London에서 제공하는 날씨 정보 데이터를 결합하여 제공한다. 주어진 데이터와 함께, 전력 수요량이 요일별, 월별, 휴일 여부 등에 따라 달라지므로 이를 결합해 분석 데이터를 구축하였다. 구성된 데이터는 다음과 같다.

- Total Energy Use from Electricity(MW) [Dependent Variable]
- Day(Mon, Tue, Wed, Thu, Fri, Sat, Sun)
- DayType(0: Weekday, 1: Weekend)
- Season(Spring, Summer, Autumn, Winter)
- Precipitation millimeter per hour (mm/hour)
- Temperature (°C)
- Amount of power per area from the Sun on the Earth's surface( $W/m^2$ )
- Amount of power per area from the Sun at the top of the atmosphere( $W/m^2$ )
- Total amount of snow on exposed ground(mm/H)

2) 보다 자세한 데이터에 관한 설명과 구성은 "<https://ssc.ca/en/cas-e-study/predicting-hourly-electricity-demand-ontario>"를 참조

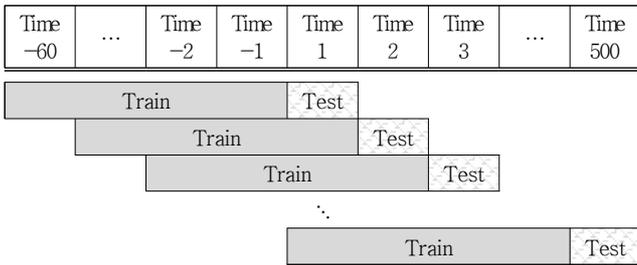
- Total amount of snow on exposed ground(cm)
- Portion of sky that is covered in clouds
- Mass per unit volume of Earth's atmosphere(kg/m<sup>3</sup>)
- Canada Holiday(0: Regular day, 1: Holiday)

2003년 12월 31일 15시부터 2016년 12월 31일 23시까지 구성된 113,972개의 관측값을 지닌다. 본래 가져야 하는 관측값은 113,985개이므로, 범주형 변수와 수치형 변수를 고려할 수 있는 MissForest[12]를 이용해 낱짜 정보를 제외한 수치형 결측치를 추정하였다.

### 3.2 분석 결과

본 연구에서 제안하는 방법론의 성능 확인을 위해, 2016년 12월 31일에 대한 24시간 예측을 기준으로 학습데이터와 예측 데이터의 window를 옮기는 검증 방식을 500번 시행하였으며, 이는 [표 1]로 설명할 수 있다.

[표 1] 방법론 검증 방식



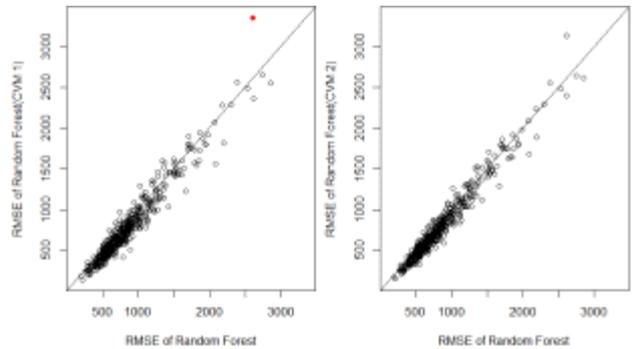
학습데이터의 기간은 60일을 기준으로 하였으며, 학습데이터의 크기는 1,440개(24H×60D)이며 예측 데이터는 24개가 되겠다. 그리고 성능 확인을 위한 기초(BaseLine) 모형은 ARIMA[11]를 선택하였다. 계절성을 고려하였으며, 6개의 모수((p,d,q), 계절성 3개)에 [0,1,2]를 반영하여 총 729개의 조합 모수를 고려하였다. 이후 AIC가 최소가 되는 최적의 조합 모수를 선택하여 예측에 사용하였다. 그리고 본 방법론을 적용할 모형은 랜덤 포레스트[3]를 선택하였다. 결정트리[14]에서

$\sum_{m=1}^M \sum_{x_j \in R_m} (y_j - \bar{y}_{R_m})^2$  을 최소화하는  $R_k(x_j < s)$ 를 찾는 만큼 CVM 효과 검증에 적절하다고 볼 수 있겠다. 그리고 CVM을 적용할 때 분위수 랜덤 포레스트 모형[7]을 사용하였으며, 25분위 회귀식, 평균 회귀식, 75분위 회귀식을 고려한 것을 CVM Random Forest 1이라 정의하고, 여기에 50분위 회귀식까지 고려한 것을 CVM Random Forest 2라고 정의한다. 분석 결과는 [표 2]와 같다.

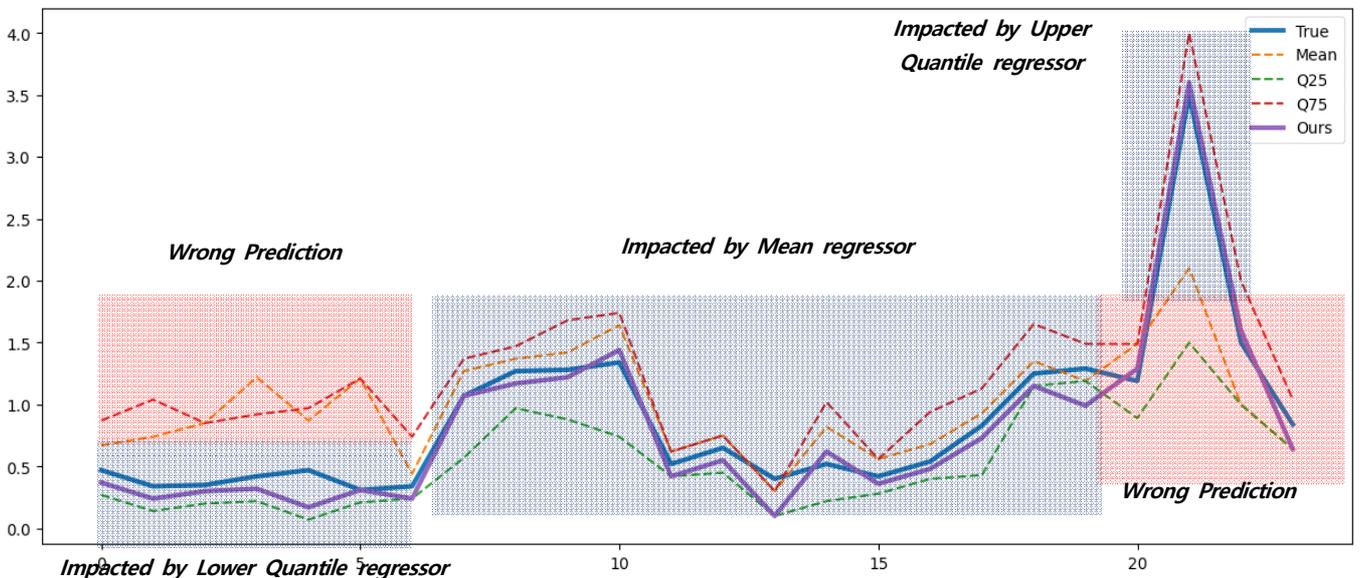
[표 2] 분석결과 요약(평균(표준오차))

Model	RMSE	MAE	MAPE (%)	Reduce (%)*
BaseLine(ARIMA)	1,583 (28.9)	1,295 (26.3)	8.21 (0.14)	
Random Forest	826.7 (18.6)	707.7 (17.3)	4.61 (0.11)	
CVM Random Forest 1	765.6 (19.1)	653.4 (17.7)	4.23 (0.12)	<b>7.77</b>
CVM Random Forest 2	750.4 (19.3)	639.8 (17.8)	4.13 (0.12)	<b>9.75</b>

\* 기존 Random Forest 대비 지표 평균 감축률(%)



[표 2]의 상단에서 확인할 수 있는 것은 기초 모형에 비해 랜덤 포레스트 모형이 RMSE의 관점에서 약 48% 성능 향상



[그림 3] CVM을 활용한 전력 수요 예측 예시

## 참고문헌

을 가져온 것으로 미루어봐, 전력 수요 예측에 비선형 모형을 사용하는 것이 적절하다는 것을 알 수 있었다. 그리고 CVM을 적용할 경우, 7.8~9.8% 정도의 성능 향상이 보임을 확인할 수 있었다. 그리고 하단 그래프에서 확인할 수 있는 사실은 CVM에서 50분위 회귀식을 함께 반영할 경우, RMSE의 관점에서 상대적으로 CVM으로부터 성능 향상이 이루어지지 않았을 때(빨간 점)의 단점을 감축시켜 보다 강건하게 성능 향상을 끌어낼 수 있다. 이러한 성능 향상이 나오는 이유는 [그림 3]을 통해 확인할 수 있는데, 학습데이터의 평균을 사용함으로써 나타나는 편향 및 변동성을 줄일 수 있다는 것이다. 평균 회귀식은 새벽 시간대와 밤 시간대에 편향된 예측을 하게 된다. 학습데이터 상의 전체 평균을 설명하려다 보니, 24시간의 평균 추세를 설명할 수 있는 회귀식을 예측하기 때문이다. 그러므로 25분위, 75분위 회귀식을 통해 이러한 편향을 상충하여 더욱 정교한 예측을 할 수 있게 된다.

## 4. 활용 방안 제안 및 결론

본 연구를 통해 탄소 중립 이행에 대한 활용 방안은 다음과 같다. 첫째, 정교한 수요예측을 통해 불필요한 전력 생산의 방식을 통해 탄소 중립에 이바지하는 것이다. 둘째, 앞선 활용 방안에서 에너지 저장 장치(ESS) 유지 비용 및 신규 전력 발전기 설치 절감을 통해 탄소 중립에 이바지하는 것이다. 마지막으로 국민 DR 사업 활성화를 위한 고객 정산금 산정 정교화 활용이다. 국민 DR 사업은 전력 수요 관리의 한 형태로, 전력 수요 측정을 통해 전기 사용량을 조절하고, 고객기준부하(CBL) 대비 실제 사용량의 차분을 kWh당 1,600원 정도를 보상하여 전력 계통 안정성을 유지하는 사업 모형이다. 이에 대한 필요성을 정부에서는 인식하고 2023년 제3차 지능형전력망 기본계획[1]에서 국민 DR 사업 활성화를 위한 추진 체계를 마련하였다. 이러한 국민의 관심을 끌어내기 위해선 올바른 고객기준부하를 설정해야 한다. 현재의 제도하에서는 아파트 단위 고객을 합쳐 CBL을 산정하므로, 아파트 개인 단위의 CBL 산정에 편향이 발생할 수 있게 된다. 그러므로 이러한 고도의 수요예측 모형을 활용한다면, 올바르게 정산금이 돌아가게 되어 국민 DR 참여 유도를 끌어낼 수 있을 것이다.

하지만 본 연구의 한계는 다양한 데이터에 대한 검증 부족과 다양한 방법론에 적용된다는 것을 검증하지 못했다는 것이다. 추후 연구에서는 이를 고려하여 연구를 진행할 예정이며, 개인 단위의 전력 수요 예측 방안을 모색할 것이다. 그리고 인공지능에 대한 수요와 관심이 증대된 만큼, 고도화된 인공지능 예측 모형을 통해 저변을 확대할 예정이다.

- [1] 산업통상자원부, “2023년도 지능형전력망 시행계획”
- [2] Allen, Myles R., et al. “Net zero: science, origins, and implications.” *Annual Review of Environment and Resources* 47 (2022): 849–887.
- [3] Breiman, Leo. “Random forests.” *Machine learning* 45 (2001): 5–32.
- [4] Caineng, Z. O. U., et al. “The role of new energy in carbon neutral.” *Petroleum exploration and development* 48.2 (2021): 480–491.
- [5] Chakhchoukh, Yacine, Patrick Panciatici, and Lamine Mili. “Electric load forecasting based on statistical robust methods.” *IEEE Transactions on Power Systems* 26.3 (2010): 982–991.
- [6] Cheon, Youngho. “Review of global carbon neutral strategies and technologies.” *Journal of the Korean Society of Mineral and Energy Resources Engineers* 59.1 (2022): 99–112.
- [7] Meinshausen, Nicolai, and Greg Ridgeway. “Quantile regression forests.” *Journal of machine learning research* 7.6 (2006).
- [8] Nti, Isaac Kofi, et al. “Electricity load forecasting: a systematic review.” *Journal of Electrical Systems and Information Technology* 7 (2020): 1–19.
- [9] Koenker, Roger, and Gilbert Bassett Jr. “Regression quantiles.” *Econometrica: journal of the Econometric Society* (1978): 33–50.
- [10] Rhys, J. M. W. “Techniques for forecasting electricity demand.” *Journal of the Royal Statistical Society. Series D (The Statistician)* 33.1 (1984): 23–33.
- [11] Said, Said E., and David A. Dickey. “Testing for unit roots in autoregressive-moving average models of unknown order.” *Biometrika* 71.3 (1984): 599–607.
- [12] Stekhoven, Daniel J., and Peter Bühlmann. “MissForest – non-parametric missing value imputation for mixed-type data.” *Bioinformatics* 28.1 (2012): 112–118.
- [13] Suganthi, L., and Anand A. Samuel. “Energy models for demand forecasting – A review.” *Renewable and sustainable energy reviews* 16.2 (2012): 1223–1240.
- [14] Von Winterfeldt, Detlof, and Ward Edwards. “Decision analysis and behavioral research.” (1993).