



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

2022년 2월
박사학위 논문

딥러닝 기반 전력 계량데이터 결측 보정 모델에 대한 연구

조선대학교 대학원

컴퓨터공학과

권혁록

딥러닝 기반 전력 계량데이터 결측 보정 모델에 대한 연구

A Study on the Power Metering Data Missing
Correction Model Based on Deep Learning

2022년 2월 25일

조선대학교 대학원

컴퓨터공학과

권혁록

딥러닝 기반 전력 계량데이터 결측 보정 모델에 대한 연구

지도교수 김 판 구

이 논문을 공학박사학위신청 논문으로 제출함

2021년 10월

조선대학교 대학원

컴퓨터공학과

권 혁 록

권혁록의 박사학위논문을 인준함

심사위원장 조선대학교 교수 양 희 덕



심사위원 조선대학교 교수 신 주 현



심사위원 조선대학교 교수 최 준 호



심사위원 UST 교수 황 명 권



심사위원 조선대학교 교수 김 판 구



2022년 1월

조선대학교 대학원

목 차

ABSTRACT

I. 서론	1
A. 연구 배경 및 목적	1
B. 연구 내용 및 구성	8
II. 선행연구 및 고찰	9
A. 지능형 계량 인프라	9
1. 스마트미터	9
2. AMI 통신 네트워크	12
3. AMI시스템	17
4. 계량데이터관리시스템(MDMS)	18
B. 결측 데이터 보정에 대한 선행연구	19
1. 선형보정법	19
2. 유사도 기반 보정법	23
3. ARIMA 예측 보정법	27
4. 장단기메모리순환신경망(LSTM) 예측 보정법	29
C. 선행 연구결과 고찰	31
1. 계량데이터 수집 관련 시사점	31
2. 결측 데이터 보정기법 관련 시사점	33
3. 계량데이터 결측 관련 종합 시사점	35

III. 연구 방법론	37
A. 연구 방법론 개요	37
B. 연구 방법론 설정	39
1. 데이터 분석 방법	39
2. 데이터 보정모델 개발 방법	39
3. 데이터 보정모델 평가 방법	39
IV. CNN-LSTM결합 가중치적용 보정법 제안	42
A. 제안한 모델의 개념	42
1. 모델 필요성	42
2. 모델 이론	43
3. 모델의 처리 흐름	46
B. 제안 모델 개발	48
1. 데이터 분석결과	48
2. 모델 개발	56
V. 실험 및 평가	64
A. 실험과정 및 결과	64
1. 자료 수집 및 대상 범위 선정	64
2. 실험 환경	66
3. 실험 절차	67
4. 실험 결과	77
B. 평가 분석 결과	100

1. 시간별 구간 전력사용량 비교 분석 결과	100
2. 선행 보정법과 제안된 보정법의 비교 분석 결과	102
3. 결측 길이에 따른 분석 결과	107
C. 연구 결과 활용 가능성	109
1. 연구 결과 적용 가능성	109
2. 연구 결과 적용시 고려사항	110
VI. 결론 및 제언	113
참고문헌	116
[부록 1] 계량기별 가장 유사한 전력사용량 패턴 탐색	126
[부록 2] CNN-LSTM결합 딥러닝 모델의 학습과정 및 예측결과	131
[부록 3] 계량기별 결측 보정 방법별 예측결과와 실제데이터 비교	141
국문초록	146

표 목 차

[표 1-1] 원격검침 부가서비스	1
[표 1-2] AMI 구축 계획 및 실적	2
[표 1-3] 국외 AMI 정책	4
[표 2-1] 전자식 전력량계 종류	10
[표 2-2] DCU 종류	13
[표 2-3] 검침용 모델에 사용되는 주요 통신기술	13
[표 2-4] 고속PLC 모델	14
[표 4-1] 일반적인 누적 전력사용량 패턴	49
[표 4-2] 사용량 변경 없는 케이스의 누적 전력사용량 패턴	50
[표 4-3] 사용량이 특이한 케이스의 누적 전력사용량 패턴	51
[표 4-4] 일반적인 시간별 구간 전력사용량 패턴	52
[표 4-5] 여름철 특징인 시간별 구간 전력사용량 패턴	53
[표 4-6] 불규칙 사용량인 시간별 구간 전력사용량 패턴	54
[표 4-7] 특이한 케이스의 시간별 구간 전력사용량 패턴	55
[표 4-8] optimizer 종류별 MSE, MAE	61
[표 4-9] 에포크(epoch)에 따른 정확도	62
[표 5-1] H/W 및 S/W 환경	66
[표 5-2] 계량기별 ARIMA 모델의 예측 결과	81
[표 5-3] 계량기별 결측 이전일자와 가장 유사한 전력사용량 패턴 탐색	84
[표 5-4] 계량기별 CNN-LSTM결합 모델의 예측결과(시간별 구간 전력사용량)	91
[표 5-5] 계량기별 CNN-LSTM결합 모델의 학습과정(MAE, R-Square)	93
[표 5-6] 계량기별 CNN-LSTM결합 가중치적용 예측결과와 실제데이터 비교	98
[표 5-7] 계량기별 결측 보정 방법별 예측결과와 실제데이터 비교	101
[표 5-8] 전체 계량기에 대한 각 보정 방법별 실험 평가결과 비교	106

그림 목 차

[그림 1-1] 한국판 뉴딜 중 디지털뉴딜과 그린뉴딜 세부과제	3
[그림 1-2] 중간 결측이 발생한 전력사용량 데이터 예시	7
[그림 2-1] PLC 통신 원리	15
[그림 2-2] AMI시스템 구성도	17
[그림 2-3] MDMS 개념도	18
[그림 2-4] Interpolation과 extrapolation 개념도	19
[그림 2-5] 1차원 선형보정법 개념도	20
[그림 2-6] 2차원 선형보정법 개념도	21
[그림 2-7] 3차원 선형보정법 개념도	22
[그림 2-8] 유사도 기반 보정법 개념도	25
[그림 2-9] 장단기메모리순환신경망(LSTM; Long Short-Term Memory)의 구조	29
[그림 2-10] 통신방식별 AMI구축 비율	32
[그림 2-11] 계량데이터 수집현황	32
[그림 2-12] ACF (Autocorrelation Function)	34
[그림 2-13] PACF (Partial Autocorrelation Function)	34
[그림 3-1] 연구 절차	38
[그림 4-1] CNN의 합성곱 개념	44
[그림 4-2] LSTM(Long Short-Term Memory)	45
[그림 4-3] DNN(Deep Neural Network)	46
[그림 4-4] CNN-LSTM결합 가중치적용 모델 개념도 및 처리과정	47
[그림 4-5] 시간별 구간 전력사용량 데이터 산점도	48
[그림 4-6] 모델의 전체적인 처리 절차	56
[그림 4-7] 딥러닝 모델의 입력 데이터 세트	57
[그림 4-8] 구간 전력사용량 1차 예측 네트워크 구조	59
[그림 5-1] 정보공개포털 청구신청내역	65

[그림 5-2] 선형보정법 실험 절차 68

[그림 5-3] ARIMA 예측 보정법 실험 절차 69

[그림 5-4] 유사도 기반 보정법 실험 절차 71

[그림 5-5] 유사도 가중치적용 보정법 실험 절차 73

[그림 5-6] CNN-LSTM결합 예측 보정법 실험 절차 74

[그림 5-7] CNN-LSTM결합 가중치적용 보정법 실험 절차 76

[그림 5-8] 선형보정법 실험 결과 데이터 77

[그림 5-9] 선형보정법 계량기별 시간별 AE(Absolute Error) 78

[그림 5-10] 선형보정법 시간별 MAE, RMSE 78

[그림 5-11] 선형보정법 시간별, 결측 개수별 MAE 데이터 79

[그림 5-12] 선형보정법 시간별, 결측 개수별 MAE 79

[그림 5-13] ARIMA 모델 학습과정 80

[그림 5-14] ARIMA 예측 보정법 계량기별 시간별 AE(Absolute Error) 82

[그림 5-15] ARIMA 예측 보정법 시간별 MAE, RMSE 83

[그림 5-16] 유사도 기반 보정법 실험 결과 데이터 85

[그림 5-17] 유사도 기반 보정법 계량기별 시간별 AE(Absolute Error) 86

[그림 5-18] 유사도 기반 보정법 시간별 MAE, RMSE 86

[그림 5-19] 유사도 가중치적용 보정법 1차 예측결과 데이터 87

[그림 5-20] 유사도 가중치적용 보정법 최종결과 데이터 88

[그림 5-21] 유사도 가중치적용 보정법 계량기별 시간별 AE(Absolute Error) 89

[그림 5-22] 유사도 가중치적용 보정법 시간별 MAE, RMSE 89

[그림 5-23] 특정 계량기의 전처리 후 결과 데이터 90

[그림 5-24] 특정 계량기의 데이터 학습과정 92

[그림 5-25] CNN-LSTM결합 모델의 예측결과 데이터 94

[그림 5-26] CNN-LSTM결합 예측 계량기별 시간별 AE(Absolute Error) 95

[그림 5-27] CNN-LSTM결합 예측 보정법 시간별 MAE, RMSE 95

[그림 5-28] CNN-LSTM결합 모델의 실험 1차 예측결과 데이터 96

[그림 5-29] CNN-LSTM결합 가중치적용 최종결과 데이터 97

[그림 5-30] CNN-LSTM결합 가중치적용 계량기별 시간별 AE(Absolute Error) 99

[그림 5-31] CNN-LSTM결합 가중치적용 보정법 시간별 MAE, RMSE 99

[그림 5-32] 시간별 구간 전력사용량 평균 100

[그림 5-33] 전체 실험 결과 데이터 102

[그림 5-34] 전체 실험 결과 요약 데이터 102

[그림 5-35] 전체 실험 결과의 보정법별 MAPE 103

[그림 5-36] 전체 실험 결과의 보정법별 시간별 평균 MAPE 103

[그림 5-37] 전체 보정 방법별 MAPE 104

[그림 5-38] MAPE 요약 데이터 105

[그림 5-39] 전체 실험결과 시간별 RMSE 105

[그림 5-40] 전체 데이터의 AE(Absolute Error) boxPlot 106

[그림 5-41] 결측 길이에 따른 MSE, MAE, RMSE 데이터 107

[그림 5-42] 결측 길이에 따른 RMSE 108

[그림 5-43] 시스템 적용 방안 109

[그림 5-44] 시스템 지연수집 처리 과정 111

[그림 5-45] 시스템 결측 보정 배치처리 과정 112

ABSTRACT

A Study on the Power Metering Data Missing Correction Model Based on Deep Learning

HyukRok Kwon

Advisor : Prof. PanKoo Kim, Ph.D.

Department of Computer Engineering

Graduate School of Chosun University

Due to global warming, abnormal climates such as heavy snow, heatwaves, forest fires, and typhoons are occurring in many places around the world. Recognizing the seriousness of the climate change problem, the international panel on climate change (IPCC) general meeting in Songdo, Incheon in 2018 suggested that carbon dioxide emissions should be reduced by at least 45% compared to 2010 and achieved by 2050. Korea also prepared the "2050 Carbon Neutral Promotion Strategy" in December 2020. Five basic carbon-neutral directions: ①Expanding the use of clean electricity and hydrogen ②Improving innovative energy efficiency in connection with digital technology ③Promoting the development and commercialization of future carbon-free technologies ④ Sustainable industrial innovation with a circular economy Promotion ⑤ strengthening the carbon absorption function of nature and ecology such as forests, tidal flats, and wetlands were suggested. After all, the most basic thing to reduce carbon emissions is to reduce energy consumption. In order to save energy, it is possible to quickly build AMI and to reduce the active energy of consumers by supporting advanced time-based rate plans such as Time Of Use (TOU), Critical Peak Pricing (CPP), and Real Time Pricing (RTP) which are various additional services through two-way communication between AMI infrastructure.

KEPCO has also tried to build an AMI infrastructure across 22.5 million units by 2020, but about 10 million units have been built and operated, and the construction will be completed in the next few years. In addition, the government announced in 2020 that it would distribute AMI to 5 million apartments in its detailed task, “Building an Intelligent Smart Grid for Energy Management Efficiency”, through the announcement of the “Korean Version of the New Deal”. As the spread of AMI expands, various new services are emerging, and through this, they are taking a step further toward the goal of carbon neutrality, a government policy.

In order to provide additional services using power metering data, it is essential to acquire measurement data well from the electricity meter. However, due to the limitations and various factors of the communication network that makes up the AMI, there are about 2-5% missing values. In order to improve the quality of AMI additional services, it is necessary to predict and provide missing data. Therefore, various algorithms are being studied and applied to predict missing values of time series data generated by smart meters. However, if the accuracy is not high and the missing section is prolonged, the error increases and quality service cannot be provided.

The power usage data is not a general time series data prediction, but has a continuously increasing cumulative power usage value, so the cumulative power usage value should be predicted and corrected. Among the methods frequently used in the time series field so far, this paper identified their strengths and weaknesses through previous studies such as linear correction, similarity-based correction, Autoregressive Integrated Moving Average (ARIMA) prediction correction, and LSTM (Long Short-Term Memory) prediction correction. However, it was not appropriate to use the previously studied correction methods to predict cumulative power usage data. In particular, while simply predicting and correcting the data, there was an error that some data decreased

the cumulative power usage value that appeared after correction.

To solve these problems, this paper proposes a model that applies weights to a deep learning-based CNN-LSTM combination model as a hybrid method that combines the advantages of linear correction method and correction method using LSTM deep learning algorithms superior to general methods used in time series.

To conduct the study, the study was conducted in the order of Business Understanding, Data Understanding, Data Preparation, Model Development, and Evaluation according to the standards of CRISP-DM(Cross Industry Standard Process for Data Mining) methodology. Through previous studies, the AMI infrastructure was first understood, the cause and ratio of data missing were identified, the characteristics and patterns of power usage data were identified through data analysis, and data was prepared through preprocessing. In order to improve the previously studied long short-term memory deep learning algorithm (LSTM), a deep learning model was created that combines a convolutional neural network (CNN) and a short-term memory circulating neural network (LSTM). The results predicted by the CNN-LSTM combined deep learning model were predicted by applying weights, which are the ratio of each section to the total amount of missing sections. And finally, the final cumulative power usage data was predicted by summing the cumulative power usage value before missing and the hourly power usage value.

When comparing and analyzing the experimental results of the preceding correction method and the proposed correction method, the correction method proposed in this paper was '0.218447', which was about 40 times better than the ARIMA predictive correction method, and the square root mean square error (RMSE) was 6 times better. Even when compared with Mean Absolute Percentage Error(MAPE) value, the correction method proposed in this paper was '0.009899', which was the best at about 45 times more than the ARIMA

prediction correction method '0.448682'. In addition, the correction method proposed in this paper did not reduce the cumulative power usage value because the weight was applied to the total amount of missing sections, and even if there were many errors in predicting power usage over time, stability was very high.

When analyzing the experimental results according to the length of missing, the linear correction method was generally simpler and better than various methods with good performance in the time series field. The linear correction method had the highest accuracy until the number of data in the missing section was less than 7. Therefore, when applying to the AMI system, it is effective to determine the missing length first and use the linear correction method if the missing length is less than 7, and in the above case, it would be better to use the CNN-LSTM combination weighting correction method proposed in this paper.

In the future, if the proposed model was predicted and corrected with deep learning algorithm simply with time series data of power usage, research to increase its accuracy using more input values is needed to learn deep learning. In particular, since power usage has a very high correlation with climate, it is necessary to further increase accuracy by adding weather information.

I. 서론

A. 연구 배경 및 목적

지능형 계량 인프라(AMI; Advanced Metering Infrastructure)는 스마트그리드를 구현하기 위해 필요한 핵심 기반으로 스마트미터, 통신망, 계량데이터 관리시스템(MDMS; Meter Data Management System)과 AMI시스템으로 구성되고 스마트미터 내에 모뎀을 설치하여 양방향 통신이 가능한 지능형 전력계량 인프라를 말한다[1,2]. 또한, 다양한 유형의 분산전원 체계, 배전지능화 시스템 등과 정보연계 등 미래 지능형 전력망 운용을 위해 요구되는 최우선적으로 구축해야 할 지능화 전력망 인프라이다[3]. 소비자와 전력회사 간 양방향통신으로 다양한 부가서비스인 계시별 요금제(TOU; Time Of Use)[4,5], 수요관리형 선택요금(CPP; Critical Peak Pricing)[6], 실시간 요금제(RTP; Real Time Pricing)[7] 등 고도화된 Time-based 요금제 지원을 하게 된다. 이를 통해 수용가 측 수요반응(DR; Demand Response)[8,9]을 통하여 능동적인 에너지 절감 참여 유도가 가능하다. 전력회사 측면에서는 부하예상, 부하제어, 정전관리[10], 전력품질 모니터링 등 효율적인 전력 공급을 위한 부가서비스 제공이 가능하다[11]. [표 1-1]은 원격검침 기술발전에 따른 부가서비스 제공 종류이다.

[표 1-1] 원격검침 부가서비스

범주	AMR(자동계량시스템)	AMI(지능형계량인프라)
요금	에너지 총 소비량	에너지 총 소비량 계시별 요금제(TOU) 수요관리형 선택요금(CPP) 실시간 요금제(RTP)
DR	-	부하제어 수요 예측 소비자 입찰 임계 피크 리베이트제
소비자 피드백	월별 요금	월별 요금

		월별 상세 내역 Web 디스플레이 In Home Display(IHD)
소비자 요금 절약	수동적인 가전기기 turn off	수동/자동 제어 가전기기 turn off 최대 부하이전
고장	고객 알림	자동 검출 개별적 가정에서의 복구 확인
배전 운영	Engineering model 사용	동적, 실시간 운영

[표 1-2]와 같이 한국전력은 에너지 신산업 가속화 정책에 따라 2013년 200만호 AMI구축 1차 사업을 시작으로 2020년까지 2,250만호 고객 전체에 AMI를 구축한다는 목표로 2018년까지 약 680만호를 구축 완료했으며 2019년에도 400만호 설치를 완료하여 약 1,000만호를 운영하고 있다[12]. 당초 계획이었던 2020년까지 2,250만호 전체에 AMI를 구축하는 것은 실현되지 않았지만 지금까지의 노력을 통해 AMI 구축 기반기술은 이미 준비돼 있기 때문에 앞으로 수년 내에 구축이 완료될 것이다.

[표 1-2] AMI 구축 계획 및 실적 (단위: 10,000)

연도	~2015	2016	2017	2018	2019	2020
계획	730	1,000	1,250	1,500	1,830	2,250
실적	250	435	520	680	980	-

※산업통상자원부 보도자료 (2018.7.18.)

정부는 2020년 7월 “한국판 뉴딜” 종합계획을 발표했다[13]. [그림 1-1]과 같이 디지털뉴딜과 그린뉴딜 분야의 10대 대표과제와 28개 세부과제를 발표했다. 이 중 그린뉴딜분야 “에너지관리 효율화 지능형 스마트그리드 구축” 세부과제로 아파트 500만호를 대상으로 AMI를 보급하겠다고 발표했다.

(단위: 국비(조원), 일자리(만개))

분야	과제	'20추~'22	'20추~'25	일자리
총계		49.0	114.1	190.1
합계		18.6	44.8	90.3
디지털뉴딜	소계	12.5	31.9	56.7
	1. D.N.A. 생태계 강화			
	① 국민생활과 밀접한 분야 데이터 구축·개방·활용	3.1	6.4	29.5
	② 1·2·3차 쏠산업으로 5G·AI 융합 확산	6.5	14.8	17.2
	③ 5G·AI 기반 지능형 정부	2.5	9.7	9.1
	④ K-사이버 방역체계 구축	0.4	1.0	0.9
	2. 교육 인프라 디지털 전환			
	소계	0.6	0.8	0.9
	⑤ 모든 초중고에 디지털 기반 교육 인프라 조성	0.3	0.3	0.4
	⑥ 전국 대학·직업훈련기관 온라인 교육 강화	0.3	0.5	0.5
	3. 비대면 산업 육성			
	소계	1.1	2.1	13.4
⑦ 스마트 의료 및 돌봄 인프라 구축	0.2	0.4	0.5	
⑧ 중소기업 원격근무 확산	0.6	0.7	0.9	
⑨ 소상공인 온라인 비즈니스 지원	0.3	1.0	12.0	
4. SOC 디지털화				
소계	4.4	10.0	19.3	
⑩ 4대 분야 핵심인프라 디지털 관리체계 구축	3.7	8.5	12.4	
⑪ 도시·산단의 공간 디지털 혁신	0.6	1.2	1.4	
⑫ 스마트 물류체계 구축	0.1	0.3	5.5	
합계		19.6	42.7	65.9
그린뉴딜	소계	6.1	12.1	38.7
	5. 도시·공간·생활 인프라 녹색 전환			
	⑬ 국민생활과 밀접한 공공시설 제로에너지화	2.6	6.2	24.3
	⑭ 국토·해양·도시의 녹색 생태계 회복	1.2	2.5	10.5
	⑮ 깨끗하고 안전한 물 관리체계 구축	2.3	3.4	3.9
	6. 저탄소·분산형 에너지 확산			
	소계	10.3	24.3	20.9
	⑯ 에너지관리 효율화 지능형 스마트 그리드 구축	1.1	2.0	2.0
	⑰ 신재생에너지 확산기반 구축 및 공정한 전환 지원	3.6	9.2	3.8
	⑱ 전기차·수소차 등 그린 모빌리티 보급 확대	5.6	13.1	15.1
7. 녹색산업 혁신 생태계 구축				
소계	3.2	6.3	6.3	
⑲ 녹색 선도 유망기업 육성 및 저탄소·녹색산단 조성	2.0	3.6	4.7	
⑳ R&D·금융 등 녹색혁신 기반 조성	1.2	2.7	1.6	

[그림 1-1] 한국판 뉴딜 중 디지털뉴딜과 그린뉴딜 세부과제

또한 [표 1-3]과 같이 국외에서도 여러 가지 정책들을 발표하고 AMI구축을 활발하게 추진하고 있다.

[표 1-3] 국외 AMI 정책[14]

국가	AMI정책
미국	<ul style="list-style-type: none"> - 미국 에너지국(DOE; Department Of Energy)은 2003년에 ‘스마트 그리드 2030 비전’ 발표 - 정부 차원에서 에너지 위기에 대비한 신규 발전소 건설을 피하고 온실가스 방출 감소를 지원하기 위한 AMI의 구축을 권장하고, 해당 기업에 인센티브 법안 검토
캐나다	<ul style="list-style-type: none"> - 2009년 2월에 스마트그리드에 대한 포괄적인 정책을 포함하는 그린 에너지법(Green Energy Act) 채택 - British Columbia, Alberta 주에서는 모든 가정이 2010년까지 스마트미터를 의무적 설치 공표
호주	<ul style="list-style-type: none"> - 호주 연방 정부는 7,600만 달러 규모의 예산을 스마트그리드 기술 개발에 투입, 지원할 계획. 2009년 9월부터 스마트미터 구축에 착수한 호주는 빅토리아 주를 중심으로 스마트미터 구축이 확산 되고 있음
영국	<ul style="list-style-type: none"> - 스마트미터는 에너지 기후 변화부(DECC; Department of Energy and Climate Change)에서 주관하며 스마트미터의 보급을 위하여 70억~90억 파운드를 예산으로 책정
덴마크	<ul style="list-style-type: none"> - 2005년부터 연간 사용량이 100MWh 이상 수용가의 경우 시간대별 검침이 가능하도록 스마트미터 설치를 의무화
프랑스	<ul style="list-style-type: none"> - 2008년 7월부터 스마트미터 보급 사업을 통해 신재생 에너지원의 통합 및 고객 수요를 반영한 부하관리, 전기차 충전 관리 등의 기술 개발 추진
스웨덴	<ul style="list-style-type: none"> - EU 회원국 중 가장 빠른 시기인 2009년을 목표로 스마트미터 설치를 완료하고자 하였음. EU 내 가장 빠른 시일 내에 100% 스마트미터 설치가 가능할 것으로 예상
스페인	<ul style="list-style-type: none"> - 스페인의 주요 전력 공급업체인 Endesa는 자사 LV 배전 그리드상의 1,300만 여고객을 관리하기 위해 AMM(Advanced Metering Management) 솔루션보급 결정 - 미터 교체는 6년간(2010년~2015년) 지속되어 스페인 정부가 지정한 시한보다 3년 앞서 완료
오스트리아	<ul style="list-style-type: none"> - 능동식 배전 그리드, 전기차시스템 통합, 스마트미터링, 수요대응 및 기타 기술 분야의 스마트그리드 기술을 망라하는 4~6곳의 오스트리아 시범 지역 개발 예정

칠레	- Chilectra는 AMPLA의 동일한 기술 제공업체와 함께 소비자 약 65,000명에게 스마트미터링 시스템 설치
중국	- 중국에서는 2010년까지 약 1억 대의 전자식 계량기가 구축될 예정이며 양방향을 지원하는 스마트미터는 2014년부터 본격적으로 설치 예정
인도	- 자국 최대의 전력회사인 Reliance Energy가 스마트미터 구축에 착수한 상태로 뭄바이나 델리 등과 같은 일부 대도시에서 운용되고 있으며 2009년 5월부터 Google의 '파워 미터' 도 이용 되었음
일본	- 일본 정부는 국가 스마트미터링 이니셔티브를 밝히며 대형 공익 회사에서 2010년 하반기부터 스마트그리드 프로그램 시작

[표 1-3]과 같이 국내뿐만 아니라 국외에서도 AMI구축이 활발하게 진행되고 있다. AMI 보급이 확대되면서 여러 가지 새로운 서비스들이 생겨나서 실생활에 도움이 되고 많은 변화가 일어날 것이다. 예로 전기 사용량 패턴분석[15]을 통한 상점의 영업시간 예측서비스, 독거노인을 위한 생활안전서비스, 실시간 요금제(RTP; Real Time Pricing)[16], 수요관리형 선택요금(CPP; Critical Peak Pricing)[17], 수요반응(DR; Demand Response) 등 다양한 서비스들이 출현 할 것으로 예상된다[18].

이런 서비스들을 제공하기 위해선 필수적으로 전력량계로부터 계량데이터들을 잘 취득하여야 한다. AMI시스템 구성이 변태주에 설치되어 저압고객의 전력량을 수집하는 데이터 집중장치(DCU; Data Concentration Unit), 통신모뎀 및 전력사용량 계측을 담당하는 스마트미터로 구성되는데, DCU와 모뎀구간의 사용되는 통신방식이 HPGP[19], IEEE 802.15.4g WiSUN[20], IEEE 802.15.4m Television White Space(TVWS)[21], 고속PLC, 저속PLC, Zigbee[22], LTE 등 다양한 통신방식으로 구축돼 있다. 이중에서 한국전력은 고속PLC방식을 가장 많이 사용하고 있다. 국산 PLC 기술과 검침절차의 지속적인 보완을 통해 현재의 AMI시스템은 가공선로에서는 안정적인 성능을 확보했으나, 잡음과 감쇠가 심한 지중선로에서는 안정적인 검침 성능을 확보하는데 어려움을 겪고 있다[23,24]. 이런 이유로 월간 검침성공률은 약 98%, 일간 적시검침 성공률은 95% 정도로 낮은 편이다.

현재까지의 AMI시스템은 ISO/IEC 12139-1[25] 규격의 전력선 통신(PLC) 방식을

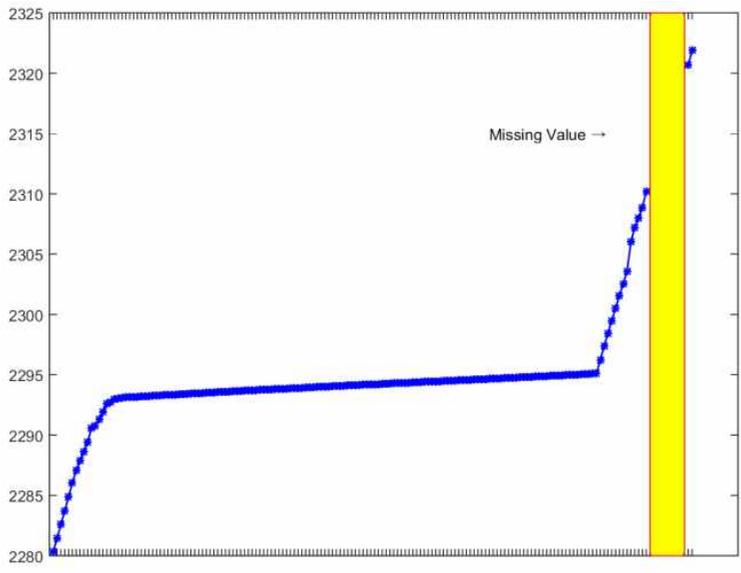
주로 이용하여 저압 AMI 네트워크를 구성하고 있는데[26], 전력선 통신(PLC)의 특
 징인 신호감쇠 등 환경적인 영향 때문에 통신 불량 또는 오작동 등 데이터를 서버
 로 전송하는 통신 과정에서의 오류로 인해 데이터의 결측(Missing Value)이 발생
 할 수 있으며 그에 따라 데이터의 품질이 낮아지는 문제점이 발생한다[27,28]. 이와
 같은 기술적 배경에서 계량데이터의 오검침 검증 및 결측값 추정 알고리즘은 AMI
 계량데이터의 신뢰성을 좌우하는 매우 중요한 부분으로 현장 데이터의 특성을 반
 영한 정교한 알고리즘이 필요하다[29]. 즉, 스마트미터로부터 수집된 시계열 데이
 터를 분석하여 스마트미터 데이터의 결측을 판별하고 해당 부분을 특정 값으로 대체
 하는 데이터 전처리 방법이 필요하다[30-35]. 전력량계의 데이터는 시간에 따라 해
 당 시간대에서의 누적 전력사용량 값이 나타나는 1차원의 시계열 데이터이다. 시계
 열 데이터에 결측이 발생한다면 결측 상황은 결측이 일어난 시점, 결측의 길이, 결
 측이 일어나기 이전 시간대에서의 값, 그리고 결측이 끝난 후 데이터가 처음으로
 나타나는 시점 및 그 때의 값으로 정의 할 수 있다[36].

[그림 1-2]와 같이 전력사용량 데이터는 전자식 전력량계에서 계속 증가하는 누
 적 전력사용량 값을 생성하게 되고, 이를 AMI시스템에서 가공 없이 취득하게 된
 다. 누적 전력사용량 데이터를 이용해서 1개월, 1일, 주간, 야간 등 특정 날짜, 시간
 의 데이터 차를 계산하여 요금을 부과 하게 된다. 또한 실시간적인 예측이 아니고,
 AMI시스템에서 데이터를 수집 시에 직전 시간의 결측 데이터를 보정하여 시스템
 의 신뢰도를 향상하고자 한다. 따라서 본 논문에서는 누적 전력사용량 데이터를 보
 정하는데, 결측이 끝나고 다음 데이터가 나타난 후에 보정하는 방법 즉 결측 구간
 의 앞뒤의 데이터가 존재하고 중간에 데이터가 빠진 상태의 데이터를 보정하는 방
 법에 대해서 연구하고자 한다. 중간에 빠진 데이터를 보정하는 이유는 실시간적인
 서비스보다는 요금계산이나 과거사용량 패턴 제공 등 다양한 서비스를 위해서 반
 드시 필요하다.

데이터를 보정하기 위한 추정 알고리즘에는 가장 기본적인 선형보정법[37,38]이
 있으며, 유사도 기반 보정법[39,40], ARIMA(AutoRegressive Integrated Moving
 Average) 예측 보정법[41,42], 모수 회귀식 기반 결측 데이터 추정기법, B-Spline,
 비모수 회귀식 기반 결측 데이터 추정기법[43], Least square기법을 적용한 결측 데
 이터 추정기법[44], 인공신경망을 이용한 추정기법[45] 등 결측 데이터의 형태에 따

라 다양한 알고리즘 기법을 적용하여 보정을 한다. 그러나 본 논문에서 연구하고자 하는 중간에 데이터가 빠진 경우를 보정하기 위해서는 앞서 서술한 알고리즘들 중에 선형보정법을 제외하고는 시간에 따른 값이 계속 증가하는 누적 전력사용량을 예측하기에는 부적합하다. 이유는 결측이 끝난 후 데이터가 처음으로 나타나는 시점의 값을 고려하지 않고 예측으로만 추정하기 때문에 보정한 값이 결측 후 나타나는 누적 전력사용량 보다 커지는 오류가 발생하게 된다.

이러한 문제들을 해결하기 위해 시계열 데이터 예측분야에서 좋은 성능을 보인 합성곱신경망(CNN; Convolutional Neural Network)과 장단기메모리순환신경망(LSTM; Long Short-Term Memory)[46,47,48]을 결합한 모형[49]으로 1차 시간별 구간 전력사용량을 예측하고, 여기에 결측 전·후 누적 전력사용량 데이터 범위를 벗어나지 않도록 가중치를 적용하는 모델을 제안하고자 한다. 그래서 각 계량기의 시간별 구간 전력사용량 패턴을 최대한 반영하고, 결측 구간 전체 총 사용량 범위를 벗어나지 않도록 누적 전력사용량을 보정함으로써 양질의 서비스를 제공할 수 있다.



[그림 1-2] 중간 결측이 발생한 전력사용량 데이터 예시

B. 연구 내용 및 구성

본 연구는 전력 계량데이터의 결측이 발생하는 다양한 원인을 분석하고 이러한 결측 데이터를 보정하기 위한 여러 가지 선행 알고리즘을 분석하고 전력 계량데이터 보정에 최적인 딥러닝 예측 후 가중치를 적용한 알고리즘을 제안하는 것이다. 본 연구 수행을 위한 주요 내용은 다음과 같다.

제1장에서는 본 연구의 배경 및 목적을 설명하고 제안하고자 하는 연구의 방향을 설정한다.

제2장에서는 AMI 계량인프라인 전자식 전력량계, 통신 네트워크, AMI시스템, MDMS시스템 등 전력 계량데이터가 생성되고 서버에까지 수집돼서 활용되는 과정을 설명하고, 결측 데이터를 보정하기 위한 여러 가지 선행연구들에 대해서 검토하고, 결측 데이터가 발생하는 이유, 알고리즘별 결측 데이터 보정법의 장·단점 등 전력 계량데이터 결측 보정과 관련된 선행연구 결과의 고찰과 이에 대한 시사점을 검토한다.

제3장에서는 선행연구 고찰 결과를 통해 전력 계량데이터 보정을 위해 연구에서 사용되는 방법론과 세부 연구 방향을 제시한다.

제4장에서는 본 논문에서 제안하는 딥러닝 기반의 CNN-LSTM결합 모형에 가중치를 적용한 개념을 설명하고, 제안한 모형의 구성 및 처리 과정 대해서 설명한다.

제5장에서는 딥러닝 기반의 CNN-LSTM결합 모형에 가중치를 적용한 알고리즘을 실험하고, 선행연구 되었던 알고리즘들과 비교분석하고 설명한다. 또한 본 연구 결과를 실제 AMI시스템 적용가능성에 대해서 검토한다.

제6장에서는 본 논문의 종합 결론 및 향후 추가 연구 내용과 필요성 등을 제시한다.

II. 선행 연구 및 고찰

전력 계량시스템의 배경이 되는 검침 인프라를 이해하여 결측이 발생하는 원인을 분석하고, 이를 보정하기 위한 선행 연구 결과들을 고찰하고, 이에 따른 시사점을 파악하고자 한다.

A. 지능형 계량 인프라(AMI; Advanced Metering Infrastructure)

본 절에서는 먼저 전력 계량 인프라 전체 과정을 이해하기 위해 최종말단 스마트미터부터 통신네트워크, AMI시스템 및 계량데이터관리시스템(MDMS)까지의 각 과정에서의 특징 및 데이터 처리과정을 살펴보고자 한다.

1. 스마트미터(Smart Meter)

기존의 전력량계는 적산전력량을 계측, 표시하는 기능만 했다. 그러나 스마트미터는 15분, 60분 등 설정에 따라서 해당 시간에 적산치를 계측해 데이터를 저장, 전송한다. 때문에 전력 수요의 세분화된 데이터 활용이 가능하다. 기존에는 맥내에서 검침이나 계약 폐지 시 육안으로 전력량을 판독했다면 스마트미터 검침에서는 전력회사에서 전력량 데이터를 자동으로 통합 관리한다. 또한 일부 스마트미터는 개폐기 기능을 갖고 있으며, 원격에서의 차단도 가능하다.

스마트미터 즉 전자식 전력량계는 새로운 통신 기술의 적용과 함께 양방향 통신을 지원하고 사용자에게 전력 사용 정보를 제공하여 수요반응(DR; Demand Response)을 통한 에너지 효율 향상을 촉진시키는 AMI 핵심장치 중 하나이고 프로그램이 가능한 전자식 전력량계로써 다음의 기능을 포함한다[50].

- 계시별 요금제(TOU; Time Of Use), 전기 사용량 데이터 수집
- 양방향 계량(Net-Metering), 정전 및 복전 알림
- 원격 부하 차단/복귀 동작, 선불요금제
- 전력 품질 감시, 도전 감지
- 프로그램 원격 변경
- 자율검침

[표 2-1] 전자식 전력량계 종류

구분	E-Type	Standard-Type	AE-Type	G-Type
외 형				
도 입	2009년	2006년	2015년	2014년
대 상	단상 5kW이하	단상5kW초과, 삼상	단상 10kW이하	단상 10kW초과, 삼상
용 도	주택, 농사용	상가, 소규모 공장 등	주택, 농사용	상가, 소규모 공장 등
기 능	양방향검침(×)	양방향검침(×)	양방향검침(○)	양방향검침(○)
	원격제어(×)	원격제어(△)	원격제어(○)	원격제어(○)
			CoverOpen 감시	CoverOpen 감시
				전력품질감시
검침주기	60분	15분	60분	15분
TOU	×	○	○	○
모뎀연결	외장, RS485	내장, IrDA(적외선)	내장, RS485	내장, RS485

저압전력에 설치돼 있는 계량기 종류는 [표 2-1]과 같이 4가지를 사용하고 있고, 표준형(Standard-Type)은 G-Type으로, E-Type은 AE-Type으로 바뀌고 있는 추세이다. 그래서 G-Type과 AE-Type를 간략하게 설명하면 모두 원격제어 및 양방향 검침이 가능하도록 돼 있고, 내장모뎀을 꼽을 수 있고, 모뎀과 RS485통신으로 연결된다. 또한 전자식 전력량계가 측정해서 저장하고 있는 데이터의 종류는 200가

지가 넘는다. 이중에서 몇 가지만 언급하면 매월 1회 전기요금 계산에 사용되는 정기검침 데이터는 계량기별 정기검침일이 설정돼 있어서 정기검침일 당일 0시 정각에 정기검침 데이터를 별도로 저장 관리한다. 서버에서는 데이터를 수집하여 과금하는데 사용하고 있다. 다음은 LP(Load Profile)[51] 데이터가 있다. LP데이터는 G-Type은 15분마다 생성되고, AE-Type은 60분마다 생성되도록 미리 설정돼 있어서 매시 정각이나, 15분마다 저장 관리되고 있다. 물론 생성주기는 원격에서 바꿀 수 있다. LP데이터가 15분이나 60분마다 계량기의 누적 전기사용량을 기록하고 있어서 LP데이터 전·후 시간의 값을 차분하면 구간 전력사용량이 된다. 한국전력은 파워플래너 서비스를 통해서 서비스 가입자에게 LP데이터를 제공해서 실시간적인 계량기의 전기 사용량 패턴을 모바일을 통해서 제공하고 있다. E-Type 계량기의 경우 정기검침 데이터가 존재하지 않기 때문에 LP데이터를 이용해서 정기검침 일자에 서버에서 정기검침 데이터를 생성하는데 사용하고 있다.

한국전력에서 2021년 9월부터 제주를 시범으로 저압전력 이용자도 TOU(Time Of Use) 요금제를 도입하고 있다. TOU요금제는 경부하, 중부하, 최대부하 시간대에 따라서 요금 단가를 다르게 함으로써 사용자가 자발적으로 최대부하 시간대 전기사용을 회피하게 하도록 유도하는 요금제이다. TOU요금제 확대를 통해서 한국전력의 전력 최대 피크를 줄이는데 도움을 줄 수 있어서 경제적인 손실을 막을 수 있다. TOU요금제를 시행하면서 시간대별로 전기사용량이 필요하게 됐다. 기존에는 전기요금 계산 시에 정기검침 데이터만 필요했으나, 이제는 시간별로 전기사용량 데이터 즉, LP데이터가 반드시 필요하고 그 중요도가 높아졌다.

AMI를 구축하기 전에는 1개월에 한번 검침원이 각 주택을 방문하여 검침을 했기 때문에 정기검침 데이터로 전기요금 계산하는 용도로만 사용했다. 하지만 AMI 구축이 진행되면서부터 AMI를 활용한 각종 서비스들이 늘어나게 되고, 실시간적인 서비스들을 하게 되면서부터 LP데이터가 반드시 필요하게 되었다. 따라서 본 논문에서도 LP데이터를 서버에서 수집했을 때 결측이 발생하는 데이터를 보정하여 전력 서비스의 품질을 높이고자 한다.

2. AMI 통신 네트워크

a. 데이터 집중장치(DCU; Data Concentration Unit)

DCU는 E-Type, Standard-Type, G-Type, AE-Type 스마트미터 등의 다양한 전자식 전력량계를 국제 표준 규격인 IEC 62056 DLMS/COSEM(Device Language Message Specification/Companion Specification for Energy Metering)[52]에 기반으로 계량데이터를 취득하여 상위시스템인 AMI시스템으로 전송하는 통신 중계 장치이다.

DCU는 WAN(Wide Area Network) 구간을 통해 서버들과 통신하고, NAN(Neighborhood Area Network) 연결을 통해 스마트미터, IHD(In Home Display) 등의 장치들과 통신하면서 소비자의 에너지 사용정보와 수요반응(DR; Demand Response) 정보를 전송하는 등의 다양한 전력서비스를 소비자에게 전달하는 양방향 기능을 수행한다. WAN영역의 통신기술은 HFC(Hybrid Fiber Coaxial), 광통신망, LTE(Long-Term Evolution) 등을 사용하고 있으며, NAN영역의 통신기술은 PLC, HPGP(HomePlug Green Phy), LTE, Zigbee, Wi-SUN 등을 사용하고 있다[53].

DCU는 대당 최대 200대의 전자식 전력량계를 수용할 수 있는 처리능력을 보유하고 있고, 하위 전자식 전력량계로부터 취득하는 주요 정보는 정기검침, 현재검침, 계량기 구성정보(Meter ID, 현재시각, 계기정수 등), LP, 전력품질(전류, 전압, 역률 등) 등이 있다[55]. 그리고 검침서버로부터 명령을 수신 받아 처리하고, 검침데이터를 전달하는 기능을 수행한다. 또한 DCU는 PLC, HPGP 등의 하위 모델들을 등록하고 통신 할 수 있도록 최적의 경로를 설정하고 네트워크 망 관리를 수행한다[54].

한국전력공사는 AMI 인프라를 보급하기 위해 수년전부터 장비를 설치하고 통신 인프라를 확충하고 있다. [표 2-2]는 한국전력공사가 2008년 시범사업, 2013년 본 사업부터 현재까지의 DCU를 보급한 종류들이다. 특히 2021년부터는 DCU 하위 모델들의 통신방식을 혼용해서 사용할 수 있도록 여러 가지 통신모델들을 슬롯타입으로 제작하여 현장 조건에 맞게 필요한 통신방식을 장착하는 방식을 사용하고 있다.

[표 2-2] DCU 종류

구분	'08 ~ 09년	'10 ~ 12년	'13 ~ 20년	'21년~
외형				

b. 검침용 통신 모델

모델은 전자식 전력량계에 따라서 전력량계 내부 또는 외부에 설치되며, DCU와 전자식 전력량계 사이에서 DLMS/COSEM(Device Language Message Specification / Companion Specification for Energy Metering) 프로토콜을 중계한다[55]. 검침용 모델에 적용되는 유무선 통신방식은 [표 2-3]과 같다. 한국전력의 경우에는 [표 2-4]와 같이 DCU와 모델간 통신방식으로는 국제표준 ISO/IEC 12139-1 방식을 적용한 한국형 고속PLC 방식을 주로 사용하고 있으며, 전자식 전력량계와 모델 간에는 전자식 전력량계의 종류에 따라 G-Type, E-Type, AE-Type 전력량계는 RS485를 사용하고 Standard-Type 은 IrDA 방식을 적용하고 있다[56].

[표 2-3] 검침용 모델에 사용되는 주요 통신기술

구분	DCU ↔ 모델	모델 ↔ 전자식 전력량계	비고
유선	저속PLC, 고속PLC	RS232, RS485	
무선	B-CDMA, Zigbee, Wi-SUN, TVWS, Lora, etc	IrDA(적외선통신)	

[표 2-4] 고속PLC 모뎀

내장형 모뎀	외장형 모뎀
	

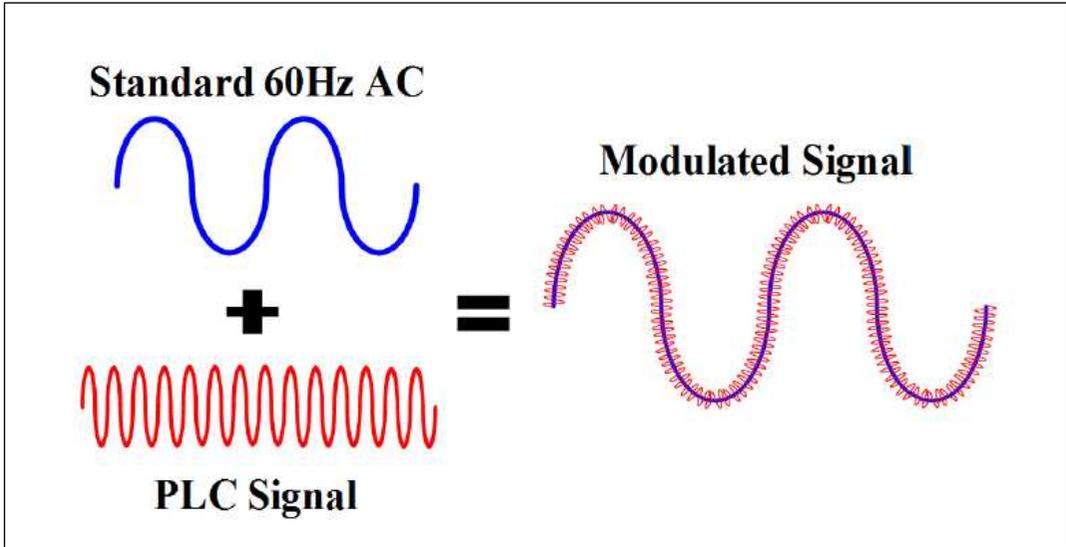
c. 통신방식

데이터 집중장치(DCU)와 모뎀간의 통신방식은 여러 가지가 있으나 국내에서 가장 많이 사용하는 것이 전력선 통신(PLC) 방식이다. 한국전력은 AMI를 구축하는데 전력선 통신(PLC) 방식을 90%이상 적용하였다. 그래서 가장 많이 사용되고 있는 전력선 통신(PLC) 방식에 대해서만 언급하고자 한다.

전력선통신(PLC; Power-Line Communications)은 전력선 즉 전기선을 통신 매체로 하여 통신하는 방식으로 신호전달에서 부터 초고속 데이터 통신 네트워크까지 여러 분야에 활용될 수 있는 유선 통신 기술이다[56]. 전력선이라 하면 철탑 등을 상상하기 마련이지만 우리가 일상적으로 전기에너지를 사용하고 있다면 이미 그곳에는 전력선이 존재하고 있으며 이렇듯 전기에너지를 사용하기 위해 필수적으로 이미 확보되어 있는 전력선을 정보 전달에도 활용하고자 하는 것이 전력선 통신의 기본 개념이다[55].

[그림 2-1]과 같이 전력선 통신의 원리는 교류전류(50~60Hz)를 공급하는 전력선을 통해 수십 Mhz의 고주파 통신신호를 함께 보내 전용 접속장비로 고주파 신호만을 수신하여 통신하는 것이다[55]. 전력선 통신은 가전제품에 영향을 미치지 않는 저출력의 신호를 사용하고, 통신 신호는 고주파 신호로 바꿔 전력선에 실어 보내고 이를 고주파 필터를 이용하여 별도로 분리해 신호를 수신하는 것이 전력선

통신의 핵심이다[55].



[그림 2-1] PLC 통신 원리

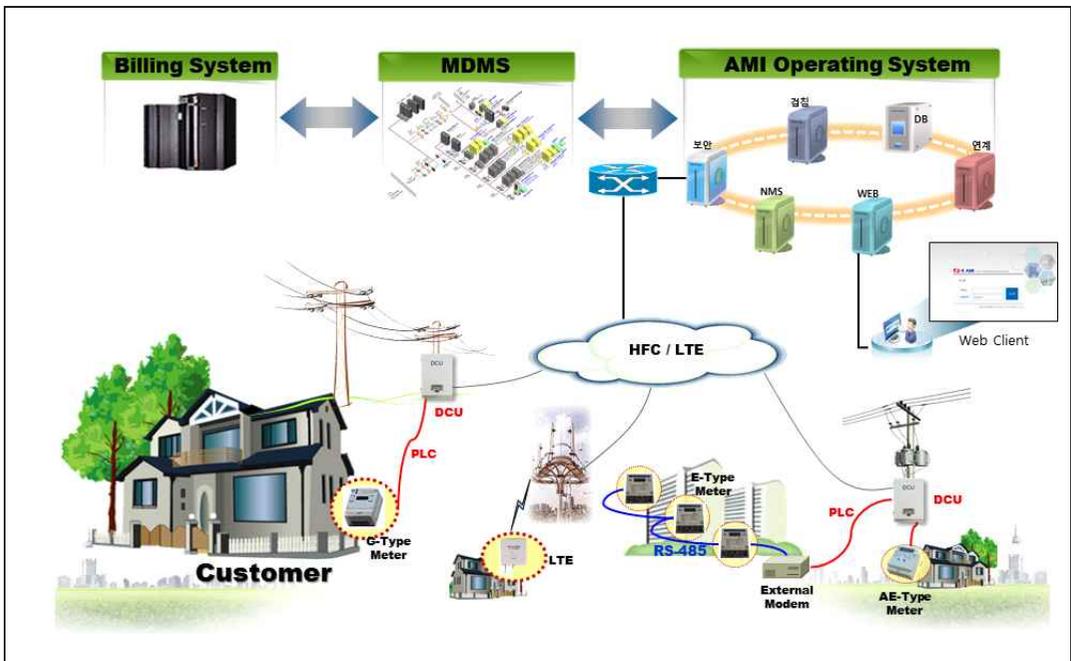
전력선 통신의 장점으로는 별도의 통신선이 없이 기존의 전력시설을 활용하기 때문에 비용이 적게 들고 설치기간이 매우 짧아서 새로운 통신망을 구축하는 것보다 비용측면에서 효율적이다[55]. 또한, 단일 인프라를 통해 음성, 영상, 데이터 및 기타 부가서비스를 보다 쉽게 통합하여 서비스를 제공할 수 있기 때문에 통신, 전력 및 부가서비스 등을 효과적이고 일괄적으로 제공할 수 있는 장점을 가지고 있다[57]. 즉, 통합서비스를 통한 마케팅 및 고객 유지관리를 위한 비용 절감, 일괄적인 요금 청구 및 이용자 보호 등을 동시에 달성할 수 있다[55].

전력선 통신의 단점으로는 홈 네트워크나 인터넷 기반의 접속과 제어 서비스의 대안으로 등장하기 위해서는 아직도 많은 장애요인들이 존재하며, 고비용의 문제점을 가지고 있어서 현재로서는 기술의 상용화와 비용 우위를 동시에 달성하기는 어렵다[55]. 고비용의 주된 원인은 기술적 불안정성과 그것을 보완하기 위한 추가적인 소요비용과 높은 부하간섭과 잡음현상이 기술의 상용화를 더디게 하고 개발비용 측면에서 불리하다[55]. 기술적으로는 가입자접속을 위한 통신선로로써 제한된 전송능력으로 인해 통신가능 거리와 속도에 대한 제약이 존재하고 가변적이고 높은 감쇄

현상, 가변 임피던스 레벨잡음, PLC는 분기가 많아 신호감쇄가 심하며 분기로 인한 주파수 선택적 페이딩 현상이 나타나고 전력선 배치의 구조적 문제로 인하여 가입자 증가시의 폭주 문제로 처리 능력 미비 등 보완이 필요하다[57].

3. AMI시스템

AMI시스템은 데이터 집중장치(DCU)에 수집된 각종 데이터(정기검침, 현재검침, LP, 정.복전 등)를 HFC(Hybrid Fiber Coaxial)나 LTE(Long-Term Evolution) 통신방식으로 수집하여 데이터베이스에 저장 관리하고, DCU나 전력망계의 설정이 필요할 때 원격에서 관리 기능을 수행할 수 있도록 기능을 제공한다[58]. 또한 수집된 검침데이터를 MDMS(Meter Data Management Server)에 전달하는 역할을 수행한다. AMI시스템의 전체적인 구성도는 [그림 2-2]과 같다.



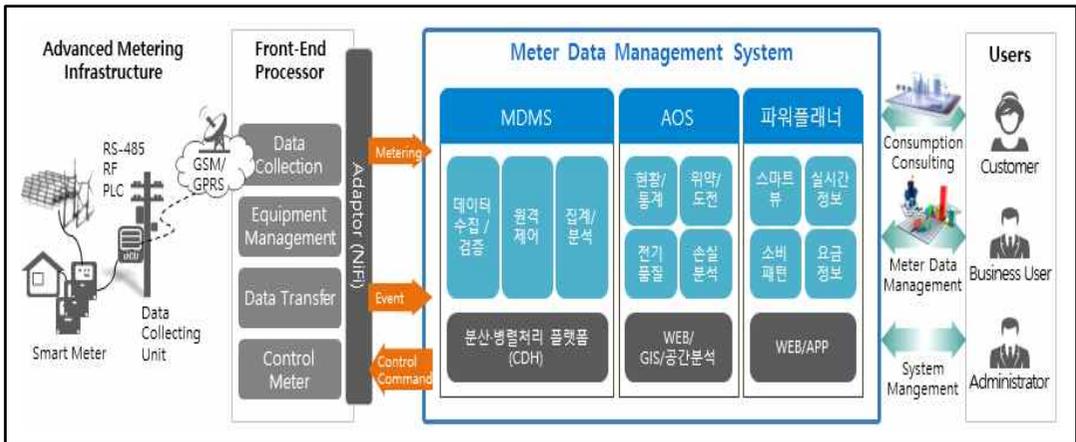
[그림 2-2] AMI시스템 구성도

AMI시스템의 역할은 DCU나 모뎀의 인증서 기반 설비 인증처리 및 키관리, 인증이 완료된 뒤에는 SNMP 프로토콜을 이용하여 DCU 및 모뎀을 등록 시키고 설비를 관리한다. 그 다음 검침 프로토콜을 이용하여 검침 데이터를 수집 관리한다. DCU없이 LTE 모뎀을 통해서 계량기와 직접 연결된 경우도 있으며, LTE 모뎀은 Push와 Polling 두 가지 기능을 다 가지고 있어서 계량데이터를 서버로 Push 전송하기도 한다.

4. 계량데이터관리시스템(MDMS; Meter Data Management System)

계량데이터 관리시스템인 MDMS는 검침서버로부터 수집된 계량데이터를 유효성 검증, 오검침/결측값 추정, 원격제어 관리, 집계 및 분석 등 계량데이터를 통합 관리하는 시스템이다[55]. 빅데이터 처리기술(오픈소스) 활용한 실시간 대용량 계량데이터 처리기술(Scale-Out)과 수집/VEE(Validation Estimation, Editing)/프로파일링/집계/저장관리 등 데이터 처리 전 과정을 분산·병렬 처리 기술을 사용한다[55]. MDMS에서 사용하는 대용량 처리를 위한 미들웨어들은 Nifi, Kafka, Spark, MongoDB, Hadoop 등이 사용된다.

[그림 2-3]는 MDMS 개념도를 표현했다. AMI시스템으로부터 수신 받은 계량데이터들을 ESB(Enterprise Service Bus) 서버의 Nifi 미들웨어를 통해서 수집하여 Hadoop, Spark, MongoDB 등 분산병렬처리 과정을 거쳐 저장하며, 계량데이터의 유효성 검증/추정, 원격제어 관리, 집계 및 분석 등의 기능을 수행한다[59]. AOS (Ami Operating System)에서는 AMI 보급/검침 현황, 위약/도전, 전기품질, 분산전원, 배전GIS/공간분석 등의 기능을 수행한다. 고객의 부가 서비스인 파워플래너에서는 실시간 사용량, 소비패턴, 전기요금 등의 기능을 스마트폰 App 이나 Web 서비스를 통해서 제공한다.



[그림 2-3] MDMS 개념도

B. 결측 데이터 보정에 대한 선행 연구

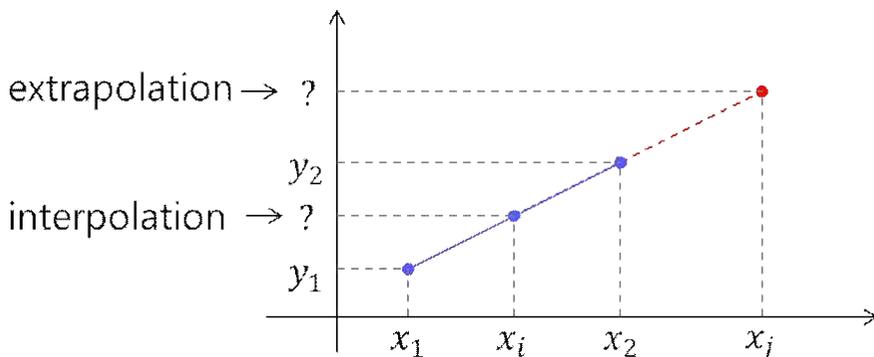
본 절에서는 전력량계에서 1시간 마다 생성된 계량데이터가 AMI시스템까지 수집되는데 이렇게 수집된 전력사용량이 시간의 흐름에 따라 계속 증가하는 시계열 형태의 데이터이다. 결측이 발생한 경우 이를 보정하기 위해서 선행으로 연구된 일반적인 기술들에 대해 자세히 설명한다.

1. 선형보정법

전력 계량시스템에서 데이터 전처리에 대한 연구는 국내외에서 활발하게 진행되어 왔다[60,61]. 그 중에서 가장 기본적이며 많이 사용되는 방법이 선형보정법[62]이다.

선형보정법에는 크게 Interpolation과 extrapolation이 있는데, extrapolation은 알려진 값들 사이의 값이 아닌 범위를 벗어난 외부의 위치에서의 값을 추정하는 것을 말한다.

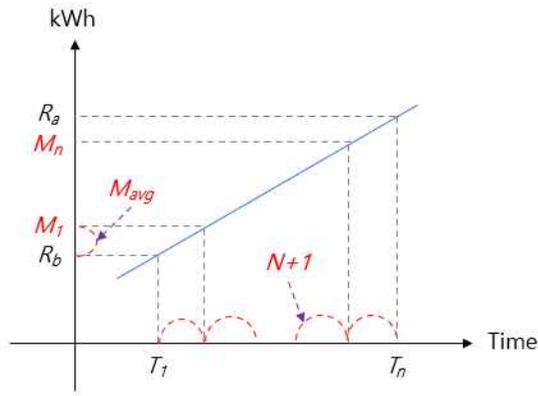
[그림 2-4]에서 x_1, x_2 에서 데이터 값을 알고 있을 때 $x_1 \leq x_i \leq x_2$ 에서 값을 추정하는 것은 interpolation이고 범위 밖인 x_j 에서 값을 추정하는 것은 extrapolation이다.



[그림 2-4] Interpolation과 extrapolation 개념도

1차원 선형보정법(1D Linear Interpolation)은 [그림 2-5]과 같이 두 점의 값이 주

어졌을 때 그 사이에 위치한 값을 추정하기 위하여 직선 거리에 따라 선형적으로 계산하는 방법이다.



[그림 2-5] 1차원 선형보정법 개념도

전력 계량데이터는 [그림 2-5]과 같이 X축은 시간축이고, Y축은 누적 전력사용량이다. $M_1 \sim M_n$ 은 결측 데이터, N 은 결측 데이터의 개수이다. M_{avg} 는 결측 구간별 평균 사용량이 된다. 식(2-1)처럼 결측 후 나타나는 첫 번째 데이터(R_a)에서 결측 전 마지막 데이터(R_b)를 빼면 결측 구간의 총 사용량이 되고 이를 결측 구간 개수 ($N+1$)로 나누면 평균 사용량이 된다.

$$M_{avg} = \frac{R_a - R_b}{N+1} \quad (2-1)$$

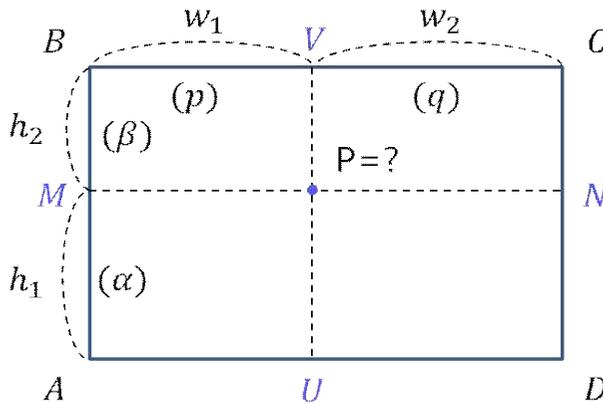
그리고 식(2-2)처럼 평균사용량을 기반으로 결측 전 마지막 데이터(R_b)에 결측 구간의 평균사용량(M_{avg})을 더하면 첫 번째 결측 시간의 누적 전력사용량 데이터 (M_1)가 된다.

$$\begin{aligned}
 M_1 &= R_b + M_{avg} \\
 M_2 &= M_1 + M_{avg} \\
 &\vdots \\
 M_n &= M_{n-1} + M_{avg}
 \end{aligned} \quad (2-2)$$

2차원 선형보정법(Bilinear interpolation)은 이중 선형보정법으로 1차원에서의 선형보정법을 2차원으로 확장한 것이다.

[그림 2-6]과 같이 직사각형의 4개의 꼭짓점 값이 주어져 있을 때, 사각형의 변 및 내부의 임의의 점에서의 값을 추정할 경우 그림과 같이 점 P에서 x축 방향으로 사각형의 변까지의 거리를 w_1, w_2 , y축 방향으로 거리를 h_1, h_2 라 하고, 알려진 네 점에서의 데이터 값을 A, B, C, D라 할 때, P에서의 데이터 값은 bilinear interpolation에 의해 다음과 같이 계산된다.

$$\left(\text{단, } \alpha = \frac{h_1}{(h_1 + h_2)}, \beta = \frac{h_2}{(h_1 + h_2)}, p = \frac{w_1}{(w_1 + w_2)}, q = \frac{w_2}{(w_1 + w_2)} \right)$$



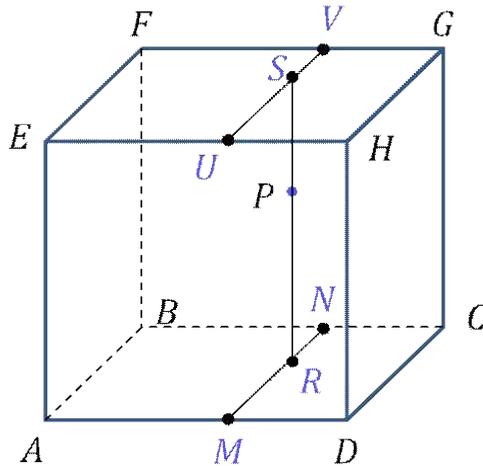
[그림 2-6] 2차원 선형보정법 개념도

계산 원리는 A, B를 보정하여 M을 얻고 C, D를 보정하여 N을 얻은 후에 M, N을 보정하여 P를 얻는 방식이다. 또는 그 순서를 바꾸어 U와 V를 먼저 얻은 후에 U, V를 보정해도 동일한 결과를 얻을 수 있다. 계산식은 식(2-3)과 같다.

$$\begin{aligned}
 P &= q(\beta A + \alpha B) + p(\beta D + \alpha C) \\
 &= q\beta A + q\alpha B + p\beta D + p\alpha C
 \end{aligned}
 \tag{2-3}$$

3차원 선형보정법(Trilinear Interpolation)은 삼 선형보정법으로 1차원에서의 선형 보정법을 3차원으로 확장한 것이다.

Trilinear interpolation 방법은 3차원 공간에서 8개의 꼭짓점으로 이루어진 육면체의 변 및 내부의 임의의 점에서의 데이터 값을 선형적으로 보정하는 방법을 일컫는 말이다.



[그림 2-7] 3차원 선형보정법 개념도

Trilinear interpolation 방법은 [그림 2-7]과 같이 bilinear interpolation과 원리는 동일하며 P에서의 보정값을 구하기 위해 먼저 M, N에서의 값을 보정하고 이로부터 R에서의 값을 보정한다. 마찬가지로 U, V에서의 값을 보정한 후 이로부터 S에서의 값을 보정한다. 마지막으로 R, S로부터 P를 보정한다.

2. 유사도 기반 보정법

전력사용량 데이터는 다른 데이터와 다르게 관성을 가지고 있는 특징이 있다. 이것은 특정 시점의 데이터는 그와 가까운 과거 시점의 데이터와 굉장히 유사하며 영향을 많이 받고 있다는 뜻이다. 예로 일반적인 주택의 경우 주중에는 아침에 출근해서 저녁에 퇴근하는 삶을 살아간다. 그래서 시간에 따라 전력사용량 패턴이 유사하다. 이점을 착안하여 유사한 과거 전력사용량 패턴을 이용해서 결측된 데이터를 보정하는 방법이다. 유사도 측정 방법에는 유클리드 거리(Euclidean Distance)[63,64], 코사인 거리(Cosine Distance)[65,66], 맨해튼 거리(Manhattan Distance)[67], 피어슨 상관계수(Pearson Correlation Coefficient)[68], 자카드 거리(Jaccard Distance)[69] 측정법 등이 있으며 가장 보편적으로 유클리드 거리와 코사인 거리, 피어슨 상관계수, 자카드 거리 측정법이 활용되고 있다.

유클리드 거리는 소비자 또는 아이템 간의 유사도를 계산하는 가장 쉬운 방법으로 가장 직관적이고 일반적인 거리의 개념이다. 일반적으로 2차원 상에서 두 점간의 거리는 피타고라스 정리에 따라 계산된다. 하지만 n차원의 공간에서 두 점간의 거리를 구하기 위해 피타고라스의 정리를 조금 더 확장시킨 것이 유클리드 거리로 해당 수식은 식(2-4)와 같다.

$$EuclideanD(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2-4)$$

여기서 (x, y)는 두 개의 연속적인 데이터이며, n은 데이터 세트의 자료의 개수를 의미한다. 식(2-4)를 통해 구해진 유클리드 거리는 거리의 최댓값이 존재하지 않아 해당 거리를 비교할 수가 없다. 따라서 해당 거리 법을 쓰기 위해선 0과 1사이의 값으로 데이터의 정규화가 우선적으로 이루어져야 하며, 정규화 된 거리에서 두 벡터(Vector)가 가까울수록 0에 가깝고 멀수록 1에 가까워지게 된다. 하지만 유클리드 거리를 통한 유사도 계산 방식은 두 벡터간의 단순한 거리를 계산한다는 점에서 해당 벡터가 같은 방향성을 지니고 있는지를 확인할 수가 없다. 따라서 해당 유사도를 활용할 경우 두 벡터 간의 유클리드 거리가 같다면 다른 방향성을 갖

더라도 유사한 정도가 큰 것으로 나타날 수 있다는 한계점이 존재한다[70].

코사인 유사도는 내적공간에서 두 벡터 간의 각도를 코사인(Cosine)방식을 이용하여 측정한 값이다. 두 벡터 간의 각도가 0° 로 그 방향이 완전하게 같다면 코사인 값은 1, 90° 의 각도로 서로 관계가 없다면 0, 180° 로 두 벡터간의 방향이 완전히 반대일 때는 -1과 같이 -1과 1의 사이 값을 갖게 되는데, 이때 코사인 유사도의 결과 값은 0과 1 사이의 양수 공간에서 표현되며 이 값은 벡터의 크기가 아닌 두 벡터간의 유사한 정도를 나타낸다[70]. 코사인 거리의 계산은 식(2-5)와 같다.

$$CosSim(x, y) = \frac{\sum_{i=1}^n x_i \times y_i}{\sqrt{\sum_{i=1}^n (x_i)^2} \times \sqrt{\sum_{i=1}^n (y_i)^2}} \quad (2-5)$$

여러 가지 유사도 측정 방법 중 코사인 유사도는 각도 값을 이용하여 유사한 방향으로 뻗어나가는지를 찾기 때문에 모든 벡터가 양수만을 가지고 있다고 간주하며 0에서 1사이의 값만 추출되고, 이는 정규화가 되어 있는 것으로 볼 수 있어 데이터를 별도로 정규화 시킬 필요가 없다. 또한 벡터 간의 양적 값을 이용해 거리를 계산하는 유클리드 거리보다 비슷한 성향의 것을 찾아낼 수 있다는 점에서 유사도 측정에 많이 활용되고 있다. 하지만 A, B, C 벡터간의 방향성이 서로 같아 코사인 거리가 0일 때, A가 B와 C중에 어느 벡터와 근접하고 있는지를 판별할 수는 없다는 한계점이 있다.

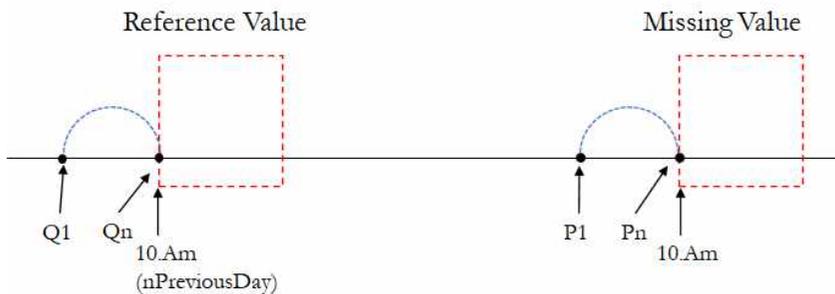
피어슨 상관 계수를 통한 유사도는 유클리드 거리, 코사인 거리 측정방법과 다르게 두 변수 간의 상관관계를 통해 얻어진다. 따라서 사용자 기반 협업 시스템에서 유사도 계산이 되는 대상은 사용자가 되며, 아이템 기반 협업 시스템에서는 두 개의 아이템이 계산 대상이 된다. 피어슨 상관계수는 두 변수(벡터)간의 공분산 값을 변수들의 표준편차의 곱으로 나눈 값으로 그 공식은 식(2-6)과 같다.

$$\text{Pearson}D(r) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2-6)$$

피어슨 상관계수를 통한 유사도의 경우 두 변수간의 상관관계, 즉 두 사용자가 따로 변하는 정도를 두 사용자가 함께 변하는 정도로 나눈 값으로 나타나기 때문에 사용자가 어떠한 아이템에 대해 평점을 높게 측정하여도 해당 점수의 영향을 덜 받으며 해당 값이 1이라면 두 변수가 완전히 동일하다고 보며, 전혀 다르다면 0, 반대방향으로 완전히 동일하면 -1의 값을 갖는다[70].

이렇게 유사도 측정(Similarity Measure)하는 방법이 여러 가지가 있다. 그 중에서 전력사용량 데이터에서의 유사도 기반 보정법으로 가장 좋은 것은 선행 논문에서 검토한 결과에 따라 유클리드 거리 측정방법이다[12].

유클리드 거리를 이용한 유사 과거 상황 대체법의 원리는 [그림 2-8]과 같다.



[그림 2-8] 유사도 기반 보정법 개념도

결측 구간의 데이터(Missing Value)의 바로 앞 데이터($P_1 \sim P_n$)와 이전 일자 중 가장 유사한 구간의 데이터($Q_1 \sim Q_n$)를 찾는다. 그런 다음 Q_n 이후의 데이터를 참조 데이터(Reference Value)로 하여 결측 구간의 데이터를 보정하는 방법이다. 보정방법을 수식으로 나타내면 식(2-7)과 같다.

$$\begin{aligned}
 M_1 &= P_n + R_1 \\
 M_2 &= M_1 + R_2 \\
 &\vdots \\
 M_n &= M_{n-1} + R_n
 \end{aligned}
 \tag{2-7}$$

M_1 은 첫 번째 결측 데이터이고, R_1 은 과거 유사상황의 첫 번째 구간 전력사용량이 된다. 결국 M_1 의 값은 결측 직전의 누적 전력사용량(P_n)과 과거 유사상황의 첫 번째 참조사용량(R_1)값을 더하여 보정하고, 두 번째 결측 데이터(M_2)는 첫 번째 보정데이터(M_1)에 과거 유사상황의 두 번째 참조사용량(R_2)값을 더하여 보정한다.

3. ARIMA 예측 보정법

시계열 데이터는 크게 규칙성을 가지는 패턴과 그렇지 않은 패턴으로 나눌 수 있다. 또한 규칙성을 가지는 패턴에는 이전의 결과가 이후에 결과에 영향을 끼치는 자기상관성(Auto Correlativeness)과 이전에 생긴 불규칙한 사건이 이후에 결과에 편향성을 초래하는 이동평균(Moving Average)현상으로 구분된다[71]. 익히 알려진 대표적인 시계열 모델로는 AR모형, MA모형, ARMA모형, ARIMA모형이 있다[71].

자기상관(Autocorrelation)-AR모형은 직전 데이터가 다음 데이터에 영향을 준다고 가정한 모형이며, 식(2-8)과 같다.

$$X_t = c + \sum_{i=1}^p \varphi_i X_{t-1} + \varepsilon_t \quad (2-8)$$

시점 t에서 얻게 될 X(t)의 평균값은 시점 t-1에서 얻었던 X(t-1)의 값에 a를 곱하고 c를 더한 것과 같다는 뜻이다. 여기서 e(t)항은 white noise라고 부르며 평균이 0이고 분산이 1인 정규분포에서 도출된 random한 값이다. 즉, X(t) 값은 평균이 a*X(t-1)+c이며 분산이 u인 정규분포에서 도출되는 임의의 값이라는 뜻이다[71].

이동평균(Moving Average)-MA모형은 시간이 지날수록 어떠한 Random Variable의 평균값이 지속적으로 증가하거나 감소하는 경향이 생길 수 있다[71]. 예를 들어 봄에서 여름이 될수록 일반적으로 가계 전기 수요량은 증가하는 경향이 있고, 여름에서 겨울로 갈수록 감소하는 경향이 있다. 전월의 전기 사용량이 다음 월 전기 사용량에 상관을 주지는 않을 것이라고 가정할 수 있고, 전기 사용량이 얼마가 될 것인지는 확실히 말하기 어렵기 때문에 이러한 경우 평균이동이 있는 시계열 데이터가 될 것이라고 생각할 수 있다[71]. 하지만 본 논문에서는 이전 데이터가 이후 데이터에 영향을 미치는 세밀한 시간에 대한 데이터를 다루므로 MA모형보다는 AR모형이 맞다고 할 수 있다.

이동평군을 시계열 모형으로 구성한 것이 MA모형이라고 부르는데, 데이터의 평균값 자체가 시간에 따라 변화하는 경향이 Moving Average이다.

ARMA모형은 Autoregressive Moving Average라는 뜻으로 AR모형과 MA모형

을 합친 것으로, 가장 단순한 형태인 ARMA(1,1)모형은 식(2-9)와 같다.

$$X(t) = \{a \times X(t-1)\} + \{b \times e(t-1)\} + c + u \times e(t) \quad (2-9)$$

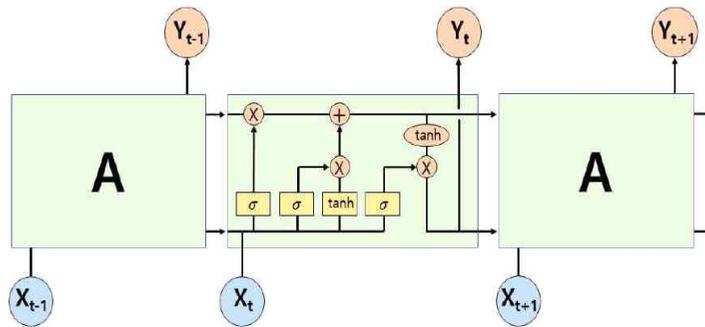
ARIMA(Autoregressive Integrated Moving Average)모형은 이전의 관측 값과 오차를 사용해서 현재의 시계열 값을 설명하는 ARMA (Autoregressive Moving Average)모형을 일반화 한 것이다[71]. ARMA(Autoregressive Moving Average)모형은 안정적인 시계열 데이터에만 적용 가능하고, 과거의 데이터들만을 사용하는 것에 반해, ARIMA(Autoregressive Integrated Moving Average)모형은 분석 대상이 불안정한 시계열 데이터라도 적용이 가능하며[71]. 과거의 데이터가 지니고 있던 추세(Momentum)까지 반영 할 수 있다. ARIMA모형은 자기 자신의 추세만 고려하며, 화이트 노이즈의 추세는 고려하지 않는다. 올바른 모형의 화이트 노이즈는 추세가 존재하면 안 되기 때문이다.

ARIMA 예측 보정법에서는 ARIMA 알고리즘을 통해서 사용량을 예측하여 결측 구간의 데이터를 보정하는 방법을 사용했다. 차분 데이터를 사용하지 않고 누적 전력사용량을 그대로 입력 값으로 사용 가능하므로 사용법이 편리하다.

4. 장단기메모리순환신경망(LSTM) 예측 보정법

순환신경망의 한 종류인 장단기메모리순환신경망(LSTM; Long Short-Term Memory)은 Hochreiter와 Schmidhuber(1997)에 의해 소개되었으며, 최근에 장기 의존성 문제를 가지는 다양한 분야에서 잘 작동하는 것으로 알려져 있어 많이 사용되는 구조이다[72]. 장단기메모리순환신경망(LSTM)은 은닉층의 메모리 셀에 입력 게이트, 삭제 게이트, 출력 게이트를 추가하여 불필요한 기억을 지우고, 기억해야 할 것들을 정한다. 이를 통해서 장기 의존성 문제를 해결하였다. 은닉 상태와 셀 상태를 구하기 위해 입력 게이트, 출력 게이트, 삭제 게이트를 가지고 있으며, 이 3개의 게이트는 공통적으로 시그모이드 함수가 존재한다. 시그모이드 함수를 통해서 각 게이트를 조정한다[73,74].

입력 게이트는 정보를 기억하기 위한 게이트로 현재 입력 값과 이전 셀에서의 출력 값을 가중치를 적용하여 시그모이드와 하이퍼볼릭 탄젠트 함수를 통과시켜 기억할 정보의 양을 결정한다. 삭제 게이트는 기억을 삭제하기 위한 게이트로 현재 입력 값과 이전 셀에서의 출력 값을 가중치를 적용하여 시그모이드를 통과시켜 삭제할 정보의 양을 결정한다. 출력 게이트는 단기상태와 장기상태 값을 출력하는 게이트로 현재 입력 값과 이전 셀에서의 출력 값을 가중치를 적용하여 시그모이드를 통과시켜 단기상태를 출력하고, 입력 게이트와 출력 게이트의 결과 값을 하이퍼볼릭 탄젠트 함수를 통과시켜 다음 셀에 사용할 장기상태를 출력한다.



[그림 2-9] 장단기메모리순환신경망(LSTM; Long Short-Term Memory)의 구조

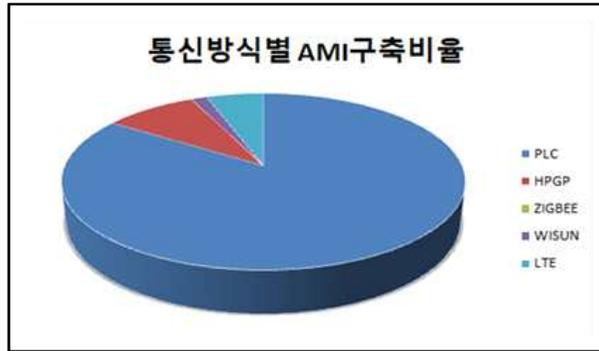
장단기메모리순환신경망(LSTM; Long Short-Term Memory)을 통해서 결측 구간
간의 구간 전력사용량을 예측하고, 예측된 구간 전력사용량에 결측 직전의 누적 전
력사용량에 예측 구간 전력사용량을 합산하여 첫 번째 결측 데이터를 보정한다. 다
음으로 첫 번째 결측 보정데이터에 두 번째 예측 데이터를 합산하여 두 번째 결측
데이터를 보정한다. 이렇게 순차적으로 결측 구간의 데이터를 보정한다.

C. 선행 연구결과 고찰

본 절에서는 전력량계 계량데이터의 결측이 발생하는 이유와 결측 데이터를 보정하기 위한 선행연구 보정법들의 장·단점 및 전력데이터의 보정 시에 발생하는 문제점과 오류들에 대해서 자세히 설명한다.

1. 계량데이터 수집 관련 시사점

어떠한 통신방식을 사용하더라도 완벽할 순 없을 것이다. 특히 AMI 인프라에서는 통신 네트워크의 취약점이 많다. 한국전력에서는 90%이상을 전력선 통신(PLC) 방식을 사용하고 있다. 그래서 전력선 통신망에 대해서만 간략하게 언급하고자 한다. AMI의 특성상 계량데이터를 취득하기 위해서 전기선의 연결이 필수적으로 되어있는 전력선 통신방식이 유리하기 때문에 가장 많이 사용하고 있다. 그래서 한국 전력은 국산 고속PLC 통신방식을 사용해서 인프라를 구축하고 있다. 그런데 전력선 통신의 장점에도 불구하고 여러 가지 단점들이 존재한다. 특히 전기선을 이용하다보니 전력선의 노이즈에 취약하다. 첫 번째, 요즘 태양광 셀 부품가격이 저렴해지고 효율이 좋아지면서 주택이나 아파트 등 소규모의 태양광 패널 설치가 증가하고 있다. 이렇게 태양광 설치 확대가 정부정책인 탄소 배출량 감소에 도움이 많이 되고 있는 것은 사실이다. 하지만 태양광 패널에서 전기를 많이 생산하는 시간대에는 인버터가 작동하게 되고, 이 때문에 전력선에 노이즈가 많이 생기게 된다. 두 번째 일반적으로 변압기를 통해 나온 하나의 전기선이 여러 집으로 분기돼서 들어가게 되는 경우가 있는데, 이런 경우에도 전력선 통신은 신호 감쇄로 인해 통신 성능이 떨어지게 된다. 세 번째 대도시나 신도시 등에는 미관상 전력선을 노출하지 않고 지중으로 설치하고 있는 곳이 늘어나고 있다. 또 하나의 전력선 통신의 약점이 지중구간에서는 통신이 잘 되지 않는다는 점이다. 그래서 통신이 잘 되지 않는 지점에는 LTE등 무선 통신방식을 병행해서 구축하고 있다.



[그림 2-10] 통신방식별 AMI구축 비율

전력망계가 AMI시스템 즉 서버와 연결되기 위해서는 중간에 데이터 집중장치(DCU)가 있어야 한다. 물론 일부 LTE 통신모뎀을 통해서 직접 연결되는 곳도 있지만 LTE 통신모뎀을 사용해서 서버와 직접 연결하는 것은 서버 입장에는 효율이 좋지 못하다. DCU는 약 200대의 전력망계의 데이터를 모아서 한꺼번에 서버와 중계하므로 LTE 모뎀에 비해 약 200배의 효율이 있다고 할 수 있다. 이렇게 서버와 연결되는 중간에 통신장비인 DCU나 LTE 모뎀 등의 장애로 인해서 데이터가 유실 또는 결측이 발생하는 경우도 많이 있다. 특히나 DCU나 LTE 모뎀도 옥외 변압기 주변이나 주택에 설치되므로 습도나 온도에 영향을 많이 받게 되고, 낙뢰 등 전력선의 과전압, 과전류에 영향을 많이 받아서 고장이 발생할 수 있다. 고장 발생하면 장애복구 이후부터 계량데이터를 수집하게 되고 고장발생 시점부터 복구시점까지는 계량데이터가 결측이 발생하게 된다. 단, DCU나 LTE 모뎀 장비의 문제가 아니고 통신상의 문제라면 결측 데이터를 수집하여 서버로 전송하게 된다. 하지만 DCU나 LTE 모뎀이 최신의 데이터를 먼저 전송하고 결측된 데이터는 천천히 수집하여 서버로 전송하게 되므로 서버에는 지연돼서 수집되는 현상이 발생한다.



[그림 2-11] 계량데이터 수집현황

[그림 2-11]은 LP적시 성공률을 보여주고 있다. 본 논문에서 다루고자 하는 것이 LP적시에 대한 내용이며, 계량기가 데이터를 생성해서 지연 없이 적시에 수집되는 비율을 나타낸다. 이 적시 성공률이 약 95% 정도이고, 1일 이내 지연돼서 수집되는 지연 수집률 즉 일간성공률이 약 97%이기 때문에 지연돼서 수집되는 비율이 약 2%가량 발생한다.

2. 결측 데이터 보정기법 관련 시사점

결측 데이터 보정을 위한 여러 가지 선행 연구한 보정법들의 특징과 장·단점을 분석하였다.

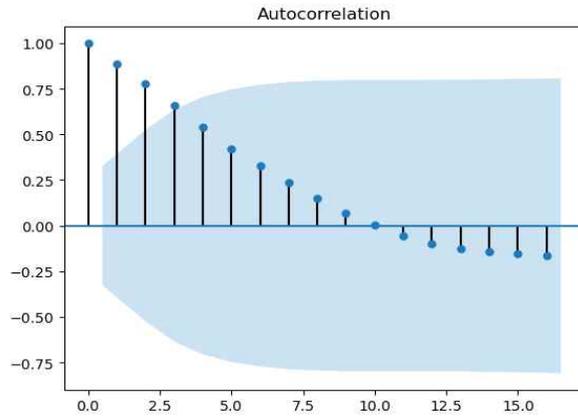
첫 번째 선형보정법의 경우 전력사용량 데이터가 계속 증가하는 누적 값을 가지는데 다른 보정법들처럼 차분을 하여 구간 전력사용량을 계산할 필요 없고 결측 이후 나타나는 첫 번째 값에서 결측 이전 마지막 데이터를 빼면 전체 결측 구간의 사용량이 되며, 이를 결측 구간의 개수로 일정하게 나누어 등분하게 되므로 빠르게 계산이 가능하며, 결측 구간의 개수가 적을 경우 매우 좋은 성능을 보였다.

두 번째 유사도 기반 보정법의 경우 전력사용량이 특정 패턴을 가지고 있다는 가정에서 좋은 아이디어임에 틀림없다. 그러나 휴일이나 날씨에 영향을 받아 불규칙하게 전력사용량이 변할 경우 좋지 못한 결과가 나온다. 또한 전력사용량이 누적 전력사용량 데이터이므로 차분하여 사용량 패턴을 구하고 유사도를 판별한다. 이에 따라 유사구간을 찾고, 그 이후 시간인 참조구간의 시간별 구간 전력사용량 값을 결측 구간에 적용하게 된다. 결측 이전 마지막 존재하는 데이터에 참조구간 시간별 구간 전력사용량을 합산하는 방식이므로 계속 더하여 나가면 마지막으로 결측 데이터를 보정한 값이 결측이 끝나고 첫 번째로 존재하는 실제 데이터보다 커지게 되는 오류가 발생하게 된다.

세 번째 ARIMA 예측 보정법의 경우 ARIMA 모형을 적용하기 위해서 구간 전력사용량을 사용하지 않고 누적 전력사용량을 그대로 입력으로 하고 ARIMA 모형의 d 를 '1'로 하여 1차분 시키면 데이터가 정상성을 만족하게 된다[75].

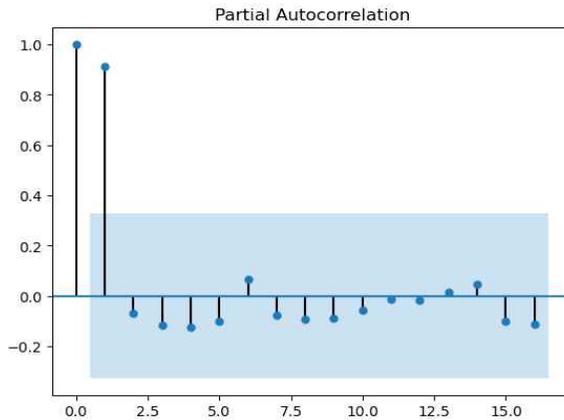
ARIMA 모델을 결정하기 위해서 acf (Autocorrelation Function)함수와 pacf (Partial Autocorrelation Function)함수를 사용하여 p , d , q 값을 찾는 과정을 수행하고[76], ACF 결과 [그림 2-12]과 같이 지수 적으로 감소하는 형태를 보여 AR모형

을 선택하였다[77].



[그림 2-12] ACF (Autocorrelation Function)

또한 PACF 결과 아래 [그림 2-13]와 같이 2차 이후 절단 형태를 보여 AR모형의 p값을 '2'로 결정하였다.



[그림 2-13] PACF (Partial Autocorrelation Function)

이렇게 실험결과 여러 가지 보정법 중에 가장 좋지 못한 결과가 나타났다. p, d, q값을 결정하기 위해 특정 한 계량기의 샘플 데이터로만 p, d, q값을 결정했기 때문에 다른 계량기의 데이터 특성이 전혀 반영되지 않아서 좋지 못한 결과가 나왔다. 또한 유사도 기반 보정법처럼 마지막 결측 보정된 데이터가 결측 이후 첫 번째로 나타나는 실제 데이터보다 커지는 오류가 발생하게 됐다.

네 번째 장단기메모리순환신경망(LSTM) 예측 보정법의 경우 장단기메모리순환신경망(LSTM) 딥러닝 알고리즘을 적용하기 위해 먼저 전처리과정을 통해 누적 전력사용량을 차분하여 시간별 사용량으로 변환하여야 한다. 시계열 데이터 예측분야에서는 좋은 성능을 보여주는 장단기메모리순환신경망(LSTM) 딥러닝 알고리즘을 사용한 결과 역시 가장 좋은 성능을 보여줬다. 하지만 유사도 기반 보정법, ARIMA 예측 보정법처럼 마지막 보정된 누적 전력사용량 데이터가 결측 이후 첫 번째로 존재하는 실제 데이터보다 커지는 현상, 즉 시간에 따라 누적 전력사용량이 증가하여야 하나 감소되는 오류가 발생했다.

3. 계량데이터 결측 관련 종합 시사점

AMI 인프라의 특성 때문에 결측 데이터가 무시 할 수 없을 만큼 상당히 발생한다는 것을 알 수 있고, 이런 이유로 전력 계량데이터를 활용해서 다양한 서비스를 제공하기에는 어려움이 따를 것이다. 다양한 분야에서 결측 데이터를 취급하는 연구는 많이 되어 왔으나, 전력 계량데이터처럼 누적되는 전력사용량 데이터를 보정하는 방법에 대한 연구는 많지 않았다.

결측 데이터를 보정하는 선행 연구 결과를 보았을 때 선형보정법이 간단하면서도 꽤 좋은 성능을 보였다. 하지만 결측 구간이 길어졌을 때 일정한 양으로 등분하므로 결측 중간부분에서 상당한 오차가 발생했다. 유사도 기반 보정법은 과거 전력사용량 패턴 중 유사도 구간을 찾아서 보정하는 방법으로 아이디어는 좋았으나 사용량 패턴이 불규칙 할 경우, 봄철과 여름철 전력사용량이 확연히 다를 경우 패턴은 유사하나 사용량의 크기의 차이가 많이 날 경우 상당한 오차를 보이게 된다. 또한 유사구간의 시간별 구간 전력사용량을 참조해서 결측 구간의 누적 전력사용량에 합산하는 방식이여서 결측 구간의 길이가 길어지면 길어질수록 오차가 커지고, 시간흐름에 따른 누적 전력사용량이 감소되는 현상도 발생할 수 있다. ARIMA 예측 보정법은 시계열 분야에서 전통적으로 많이 사용해 오던 방식이며, 차분할 필요 없이 누적 전력사용량을 바로 입력하여 결과를 얻어 낼 수 있는데, 실험한 여러 보정법 중에서 가장 좋지 못한 성능을 보였다. 장단기메모리순환신경망(LSTM) 예측 보정법은 시계열 분야에서 요즘에 가장 많이 사용하고 있고, 성능이 좋다고 알려져 있다. 실제로 여러 가지 보정법 중에 가장 좋은 성능을 보였다. 하지만 시간별 구간 전력사용량을 예측하여 결측 구간의 누적 전력사용량에 합산하는 방식을 적용

하다보니 결측 구간이 길어질수록 오차가 커지고, 누적 전력사용량 값이 감소되는 현상도 발생하였다.

선행 연구가 10시간 결측 데이터에 대한 실험이었으며, 이를 실제 AMI시스템에 적용하기에는 어려움이 있다. 선행 연구 결과만 봤을 때도 결측 구간이 짧은 경우에는 선형보정법이 빠르고, 간편하며 매우 좋은 성능을 보였다. 이처럼 실시간으로 수집되는 대량의 데이터를 처리하는 AMI시스템에 적용하기 위해서는 결측 구간의 길이에 따라서 데이터를 보정하는 알고리즘을 선택할 필요가 있고, 결측 구간의 오차를 줄이고, 누적 전력사용량 값이 감소되는 오류를 범하지 않도록 연구가 필요하다.

본 논문에서는 결측 구간에 길이 따른 보정법별 성능을 비교하고, 장단기메모리 순환신경망(LSTM) 딥러닝 알고리즘을 개선한 CNN-LSTM결합한 딥러닝 모형을 만들고 여기에 가중치를 적용하여 최종 예측 모델을 만들고자 한다.

Ⅲ. 연구 방법론

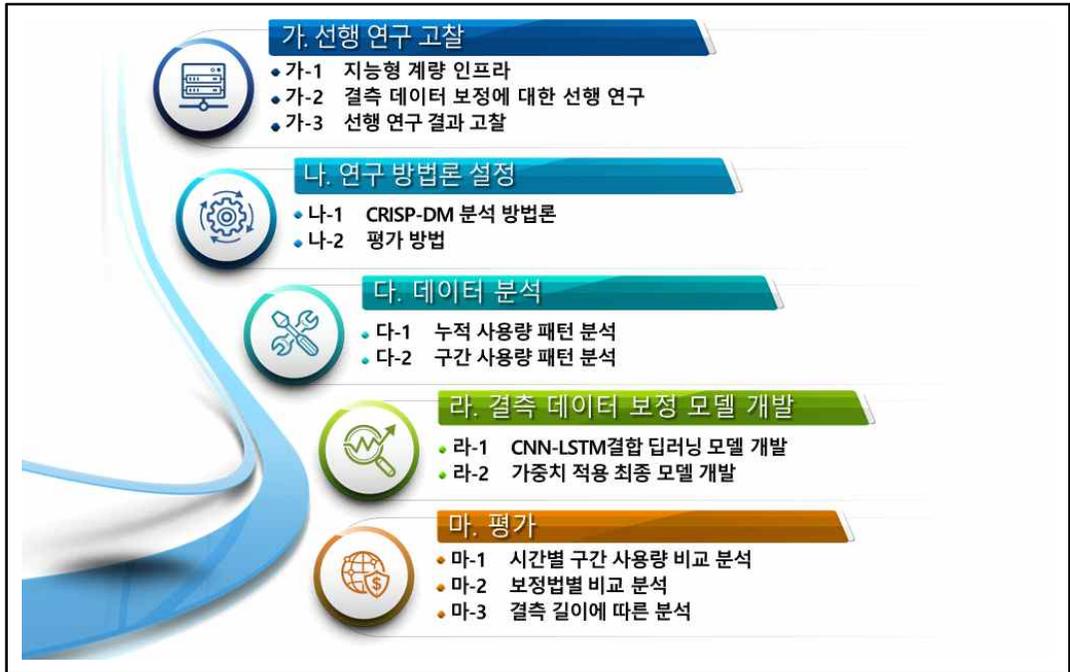
A. 연구방법론 개요

본 연구의 목적은 전력 계량데이터의 결측 발생 시에 보정하는 방법으로 딥러닝 기반의 CNN-LSTM결합 모형에 가중치를 적용한 알고리즘을 개발하여 전력 계량데이터의 결측치를 보정하는 것이다. 이를 위해서 제2장에서 결측 데이터 보정에 대한 선행 연구들을 검토하고 분석하였다. 그 결과를 반영하여 전력 계량데이터의 결측 데이터 보정을 위해 본 연구는 [그림 3-1]과 같이 수행되며, 주로 고려해야 할 내용은 다음과 같다.

첫 번째, AMI시스템에 수집되어있는 계량데이터를 분석한다. 전자식 전력량계의 종류에 따른 데이터 생성주기 및 데이터의 특성 등을 분석하고 그에 따라 알고리즘을 선택하고 적용하기 위한 방법을 도출한다. 또한 결측이 발생하는 원인으로 통신 네트워크나 장비에 따라 지연 수집되는 현상 및 통신방식에 따른 연속으로 결측되는 길이를 분석하고 결측 길이에 따른 알고리즘을 선택하고 적용하기 위한 방법을 도출한다.

두 번째, 전력 계량데이터의 특성들을 분석하고, 전력 계량데이터의 결측 데이터 보정을 위해서 적용 가능한 최적의 알고리즘을 개발한다. 이때 고려해야할 사항들을 검토하고 선행 연구 결과들의 장점들을 분석해서 정확도를 향상시킬 수 있는 개발 방법론을 도출한다.

세 번째, 전력 계량데이터의 결측 데이터 보정을 위해서 딥러닝 예측 후 가중치를 적용한 최적의 알고리즘을 제안하고, 제안모델이 기존 보정법들과 어떤 차이점이 있는지 실험을 통해 논하고, 전력 계량데이터의 결측 데이터 보정 모델의 성능 결과를 도출한다.



[그림 3-1] 연구 절차

A. 연구방법론 설정

1. 데이터 분석 방법

일반적으로 데이터 분석방법론에는 CRISP-DM(Cross Industry Standard Process for Data Mining) 방법론[78,79]이 전 세계에서 가장 많이 사용되는 데이터 마이닝 표준 방법론이다. 본 논문에서는 CRISP-DM 방법론의 표준에 따라서 업무 이해(Business Understanding), 데이터 이해(Data Understanding), 데이터 준비(Data Preparation), 모델 개발(Modeling), 평가(Evaluation), 적용(Deployment) 등 6단계 중에서 마지막 단계인 적용단계는 현업에서 수행하는 단계이므로 제외하고 나머지 5단계까지는 순차적으로 진행한다. 먼저 업무 이해를 위해서는 선행 연구 고찰을 통해서 AMI 계량인프라 및 결측 발생원인 등을 파악했고, 데이터 이해 단계에서는 통신 네트워크 방식에 따른 결측 발생 길이 분석이나 전자식 전력량계의 종류에 따른 계량데이터의 생성주기 및 데이터 기록 형태 등을 분석한다. 데이터의 형태에 따라 데이터 준비 단계 즉 전처리를 수행한다.

2. 데이터 보정모델 개발 방법

본 연구에서는 전력 계량데이터의 결측 데이터 보정 모델 개발을 위해서 기존의 시계열 데이터 보정법 등의 선행 연구 결과를 검토하고, 일반적인 시계열 데이터 예측이 아닌 결측 전, 결측 후의 실제 데이터가 존재하는 즉 중간에 데이터가 결측된 경우를 보정하는 모델을 개발한다. 또한 전력사용량 데이터가 누적 값으로 계속 증가하는 형태이므로 결측 데이터를 보정 한 후에 보정 된 마지막 데이터가 결측 이후 존재하는 실제데이터보다 크면 안 되기 때문에 이점을 고려해서 개발하여야 한다. 선행 연구 중 실제 업무에 적용 가능한 모델이 선형보정법이다. 본 연구는 선형보정법보다 예측 결과의 정확도를 향상시키고 결측 길이에 따라 알고리즘별 정확도를 비교하여 실제 업무에 적용할 때 유용한 정보를 제공한다.

3. 데이터 보정모델 평가 방법

일반적으로 회귀모델의 평가방법에는 MSE, MAE, RMSE, MAPE 등이 있다.

식(3-1) 평균제곱오차(MSE; Mean Squared Error)는 예측값과 실제값의 차이인

오차들의 제곱 평균으로, 제곱을 하기 때문에 특이치(아웃라이어)에 민감하다[80].

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3-1)$$

식(3-2) 평균절대오차(MAE; Mean Absolute Error)는 실제값과 예측값 오차들의 절댓값 평균으로, MSE보다는 특이치에 덜 민감하다[81].

$$MAE = \frac{1}{n} \sum_{i=1}^n (|y_i - \hat{y}_i|) \quad (3-2)$$

식(3-3) 제곱근 평균제곱오차(RMSE; Root Mean Squared Error)는 MSE에 root를 씌운 값으로, 오류 값을 실제 값과 유사한 단위로 다시 변환하여 직관적으로 해석할 수 있다[82].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3-3)$$

식(3-4)은 평균절대비오차(MAPE; Mean Absolute Percentage Error)로 오차가 예측값에서 차지하는 정도를 나타내는 지표이다. MAE는 normalizing을 해주지 않는 반면 MAPE는 예측값으로 normalizing을 해준다[83]. 기존 MAE에서는 outlier에 너무나도 취약하다. 예를 들어 예측값과 실제값 차의 평균이 작은 값들 사이에 갑자기 아주 큰 오차들인 outlier가 발생한다면 MAE의 값은 아주 큰 오차가 발생하게 되고, 불안정한 모델로 평가된다. 이를 보완하기 위해 나온 것이 MAPE이다. 오차의 비율로 계산되기 때문에 Outlier에는 강하다. 하지만 MAE에 비해 직관적으로 오차를 확인하기 어려운 단점이 있다.

$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (3-4)$$

전력 계량데이터의 결측 데이터 보정 모델을 개발하고 정확도를 비교함에 있어서 실제 데이터를 기준으로 평가한다. 이를 위해 존재하는 실제 데이터에서 일부 데이터를 제거하여 결측 보정 실험을 진행한다. 선행 연구된 알고리즘과 비교 평가를 위해서 알고리즘별 같은 데이터와 같은 결측 구간을 선정해서 실험을 진행한다. 선행 연구에서는 정확도 평가를 위해서 평균절대오차(MAE; Mean Absolute Error)를 사용해서 비교를 수행했다. 각 개별 계량기의 절대 오차값을 전체 계량기의 평균을 취해서 시간대별로 보정법간 결과를 비교했다. 이 평가방법에 문제점을 발견하였다. 일반적인 데이터 분석의 경우는 데이터 전처리를 통해서 scaling을 수행하므로 MSE, MAE, RMSE 등의 평가는 문제가 없다. 하지만 계량기별로 최종 누적 전력사용량을 비교하기 때문에 실제 데이터의 Scale 이 다를 경우 모델의 평가 결과가 차이가 많이 발생할 수 있다. 예로 전력사용량이 아주 큰 하나의 계량기 때문에 전력사용량이 미미한 999개 계량기의 정확도가 높더라도 낮게 평가될 수가 있다. 그래서 본 연구에서는 평균절대비오차(MAPE; Mean Absolute Percentage Error)를 계산하여 추가적으로 모델별 평가 시에 비교한다. 평균절대비오차(MAPE; Mean Absolute Percentage Error) 계산 방법을 사용하면 계량기별 누적 전력사용량 데이터의 크기에 따른 오차 비율로 표현됨으로써 정확한 평가가 될 수 있다.

IV. CNN-LSTM결합 가중치적용 보정법 제안

A. 제안한 모델의 개념

1. 모델 필요성

전력사용량 데이터는 휴일이나 날씨 등에 많이 영향을 받게 된다. 특히 요즘 같은 코로나 팬데믹 시대처럼 사람들의 이동자제와 집합금지 등 바이러스 확산방지를 위해서 정부에서 강력하게 규제를 하고 있어서 전력사용량이 일반적이지 않고 상점이나 음식점 보다는 가정에서 전력사용량이 많이 늘고 있다. 이런 배경에서 전력사용량을 선행연구처럼 선형보정법, 유사도 기반 보정법[84], ARIMA 예측 보정법 등은 정확도가 좋지 못하다.

전력사용량은 누적된 계속 증가하는 사용량 값을 가지고 있기 때문에 선형보정법을 제외하고는 일반적으로 사용량 패턴을 기반으로 예측을 하여 구간 전력사용량을 1차적으로 구한다. 그리고 결측 전 마지막 누적 전력사용량에 합산하여 더해 나가게 된다. 이렇게 하면 전력사용량 오차가 계속 증가하게 되고 결국 결측 이후 존재하는 실제 데이터보다 커지는 현상이 발생하게 된다.

전력사용량 데이터는 전자식 전력량계에서 계속 증가하는 누적 전력사용량 값을 생성하고 있다. AMI시스템에서는 가공 없이 취득하여 요금계산 및 서비스를 할 수 있도록 누적 전력사용량 값을 그대로 제공하게 된다. 만약 누적 전력사용량이 아니고 시간별 구간 전력사용량을 이용하게 된다면 데이터 누락이 발생했을 경우 요금계산을 할 수 없는 치명적인 문제가 발생한다. 그래서 본 논문에서도 AMI시스템에서 누적 전력사용량을 예측하여 보정하는 것을 목표로 하고 있다. 또한 결측 전·후 데이터가 존재하는 경우를 가정하고 있기 때문에 1차로 시간별 구간 전력사용량을 먼저 예측하고 여기에 가중치를 적용하여 결측 전·후 총 사용량 범위를 벗어나지 않도록 하여야 한다.

1차 시간별 구간 전력사용량 예측을 위해서 시계열 데이터를 예측하는 딥러닝 알고리즘을 사용하고자 한다. 선행연구 사례에서 유사도 기반 보정법, ARIMA 예측 보정법, LSTM 예측 보정법 등을 연구하였다. 연구결과 전통적으로 많이 사용

하던 ARIMA 예측 보정법 보다 최근 시계열 예측분야에서 우수한 LSTM 딥러닝 모델을 이용한 예측 방법이 우수하였다. 최근 LSTM 모델보다 정확도를 높인 연구 결과들이 많이 발표되고 있다. 시계열 데이터 예측분야에서 CNN, LSTM, CNN+LSTM 모델의 성능을 비교한 연구가 있는데 CNN+LSTM 모델을 적용하였을 경우에 정확도가 가장 높게 나타난 바 있다[85,86]. 또한 전력수요 예측 등에도 CNN-LSTM 결합 모델을 많이 사용하고 있다. 따라서 본 논문에서도 1차 시간별 구간 전력사용량 예측을 위해서 CNN-LSTM을 결합한 딥러닝 모델을 사용하고자 한다.

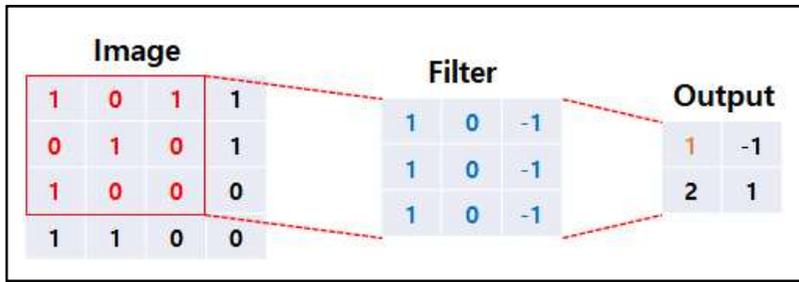
그리고 CNN-LSTM결합 딥러닝 모델로 예측한 시간별 구간 전력사용량에 결측 구간의 전체 사용량 합을 기준으로 가중치를 적용하여 결측 이후 존재하는 실제 데이터보다 커지는 현상을 제거 할 필요가 있다. 또한 이 모델을 적용했을 경우 딥러닝 기반 알고리즘의 예측 오차가 많이 발생하더라도 해당 계량기의 전체 사용량에는 차이가 없이 정확하므로 안전하게 사용할 수 있다.

2. 모델 이론

본 논문에서 제안하는 전력사용량 결측 구간의 데이터를 보정하는 모델은 합성곱신경망(CNN; Convolutional Neural Network)과 장단기메모리순환신경망(LSTM; Long Short-Term Memory)을 결합하고 여기에 가중치를 적용하는 모델이다. 주된 아이디어는 시계열 데이터 예측에 많이 사용되는 장단기메모리순환신경망(LSTM)에 정확도를 높이고자 합성곱신경망(CNN)을 결합하였다. 그리고 예측된 결과에 가중치를 적용함으로써 결측 구간의 길이에 따라 오차가 증가하는 단점을 보완하고 결측의 마지막 보정 데이터가 결측 후 첫 번째 존재하는 실제 데이터보다 커지는 현상을 제거 한다.

일반적으로 딥러닝 모델에서 적용하는 신경망 이론은 크게 세 가지를 많이 사용한다.

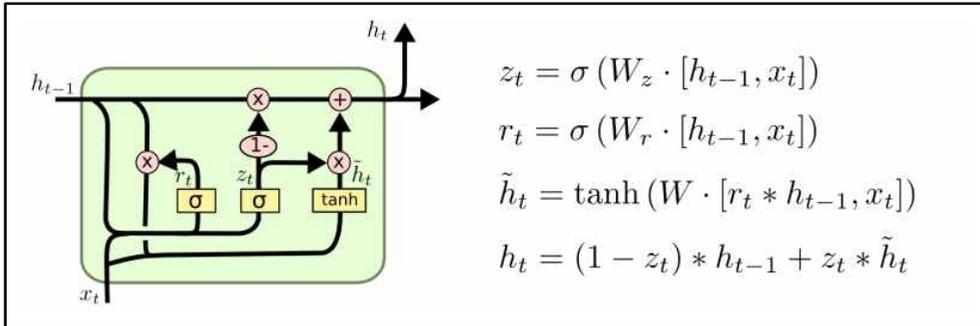
첫 번째 이론은 합성곱신경망(CNN)으로 주로 사진 및 문자 인식 등에 사용되며, 아래 [그림 4-1]과 같이 입력 이미지 행렬(4×4)과 필터 행렬(3×3)을 합성곱하여 출력(2×2) 값을 만들어 낸다. 합성곱신경망(CNN)은 필터를 사용해서 이미지의 특징을 추출해내는 것이 특징이다. 본 논문에서는 값이 하나인 시계열 데이터이므로 1차원 합성곱신경망(CNN)을 사용하여 모델은 만든다.



[그림 4-1] CNN의 합성곱 개념

두 번째 이론은 순환신경망(RNN; Recurrent Neural Network)으로 은닉층의 노드에서 활성화 함수를 통해 나온 결과값을 출력층 방향으로 보내면서, 동시에 다시 은닉층 노드의 다음 계산의 입력으로 보내는 특징이 있다. 결국 이전 셀의 값을 기억하고 있다. 이를 메모리 셀이라고 한다. 은닉층의 메모리 셀은 각각의 시점에서 바로 이전 시점의 은닉층의 메모리 셀에서 나온 결과를 입력으로 재귀적으로 사용하고 있다. 즉, 현재 시점에서의 메모리 셀은 과거의 메모리 셀들의 값에 영향을 받고 있는 것이다. 그래서 앞의 입력이 뒤에 영향을 주는 시계열 예측이나, 자연어 처리에서 많이 사용되고 있다. 특히, 딥러닝 기반의 시계열 자료 예측모형인 기본 순환신경망(RNN), GRU(Gate Recurrent Unit) 그리고 장단기메모리순환신경망(LSTM)의 성능을 비교한 연구가 있는데 장단기메모리순환신경망(LSTM) 모델을 적용하였을 경우에 예측 정확도가 가장 높게 나타난 바 있다[87]. RNN은 시점(time step)의 길이가 짧은 경우는 효과적이거나, 시점의 길이가 길 경우 앞의 정보를 뒤로 잘 전달하지 못하는 현상이 발생한다. 뒤로 갈수록 정보의 손실이 생기고, 영향력이 거의 없어진다. 이렇게 RNN은 충분한 기억력을 가지고 있지 못해서 예측 정확도가 떨어진다. 이를 장기 의존성 문제(the problem of Long-Term Dependencies)라고 한다.[88]. 또한, 수치형 값으로 구성된 전력사용량 자료와 같이 시간의 변화에 따라 사용량의 차이가 클 경우 과거 데이터의 학습 과정을 거치면서 오차가 증가하는 현상이 발생할 수 있다. 이러한 문제를 해결하기 위해 장단기메모리순환신경망(LSTM)을 적용하는 것이 바람직하다[89]. [그림 4-2]와 같이 장단기메모리순환신경망(LSTM)은 RNN의 이러한 단점을 보완한 RNN의 장단기 메모리(Long Short-Term Memory)라고 하며, 줄여서 LSTM이라고 한다. LSTM은 은닉층의 메모리 셀에 입력 게이트, 삭제 게이트, 출력 게이트를 추가하여 불필요한 기억을 지

우고, 기억해야할 것들을 정한다. 이를 통해서 장기 의존성 문제를 해결하였다. 은닉 상태와 셀 상태를 구하기 위해 입력 게이트, 출력 게이트, 삭제 게이트를 가지고 있으며, 3개의 게이트는 공통적으로 시그모이드 함수가 존재한다. 이 시그모이드 함수를 통해서 각 게이트를 조정한다.



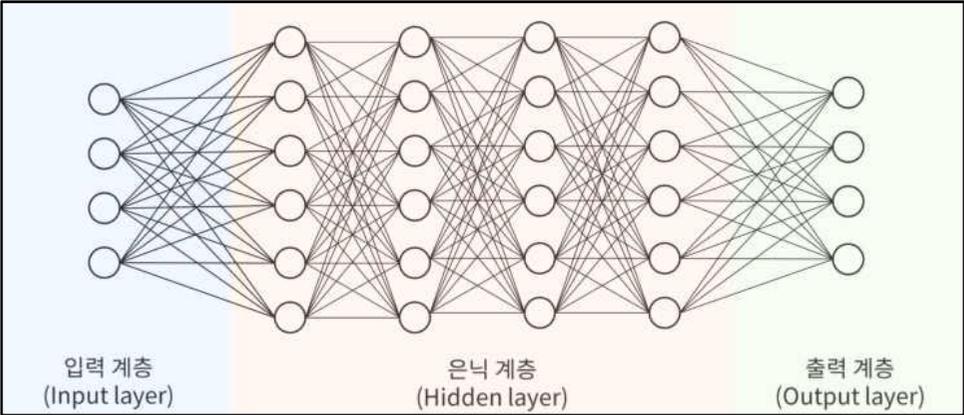
[그림 4-2] LSTM(Long Short-Term Memory)

입력 게이트는 정보를 기억하기 위한 게이트로 현재 입력 값과 이전 셀에서의 출력 값을 가중치를 적용하여 시그모이드와 하이퍼볼릭 탄젠트 함수를 통과시켜 기억할 정보의 양을 결정한다. 삭제 게이트는 기억을 삭제하기 위한 게이트로 현재 입력 값과 이전 셀에서의 출력 값을 가중치를 적용하여 시그모이드를 통과시켜 삭제할 정보의 양을 결정한다. 출력 게이트는 단기상태와 장기상태 값을 출력하는 게이트로 현재 입력 값과 이전 셀에서의 출력 값을 가중치를 적용하여 시그모이드를 통과시켜 단기상태를 출력하고, 입력 게이트와 출력 게이트의 결과 값을 하이퍼볼릭 탄젠트 함수를 통과시켜 다음 셀에 사용할 장기상태를 출력한다.

세 번째는 심층신경망(DNN; Deep Neural Network)으로 심층신경망은 입력층과 여러 개의 은닉층, 출력층으로 이뤄진 신경망이다. 일반적인 인공신경망처럼 복잡한 비선형 관계들을 모델링하는데 사용되며, 분류 및 수치 예측에 많이 사용된다.

무엇보다도 본 논문에서 제안한 CNN-LSTM결합 가중치적용 모델이 일반적인 시계열 데이터 예측과 가장 큰 차이점은 미래를 예측하는 것이 아니고 실제 데이터의 앞, 뒤 시점의 데이터가 존재하고 중간에 결측이 발생한 경우만을 가정하고 있으며, 그 데이터의 형태가 계속 증가하는 값을 가져야 하며, 누적 전력사용량 값

이 이전 시간대보다 적어지는 경우가 발생하지 않아야 한다. 또한 일반적인 시계열 데이터 예측 알고리즘인 장단기메모리순환신경망(LSTM) 모델에 합성곱신경망(CNN) 알고리즘을 결합하여 24시간의 전력사용량 패턴을 입력하여 그 특징을 추출하여 이용한다는 것이 차이점이며, 예측한 결과 값에 가중치를 적용하여 본 논문에서 고려해야하는 누적 전력사용량의 중간 결측 보정법에 최적화된 기법을 제안한다.



[그림 4-3] DNN(Deep Neural Network)

3. 모델의 처리 흐름

본 연구에서 제안하고자 하는 CNN-LSTM결합 가중치적용 모델의 개념도 및 그 처리 과정은 [그림 4-4]와 같으며, 처리 흐름은 데이터 전처리 → 딥러닝(CNN → LSTM)모델 학습 및 예측 → 가중치적용 → 최종 계산 순으로 학습하고 예측을 수행한다. 다음은 처리 단계별 구성과 역할에 대해 설명한다.

- 데이터 전처리 : 데이터 전처리는 입력된 실험데이터를 딥러닝 모델에서 사용할 수 있는 형태로 전환하는 과정이다. 특히 전력사용량 데이터는 계속 증가하는 누적 전력사용량 데이터의 형태이므로 이를 (t)시간 누적 전력사용량에서 (t-1)시간의 누적 전력사용량을 차분하여 (t)시간의 구간 전력사용량을 계산한다. 준비된 시간별 구간 전력사용량을 시계열 데이터 입력으로 만들기 위해서 24개+1개(Label)씩 하나

의 세트르 다시 만들어 준다. 데이터 세트는 1씩 쉬프트하며 만들어 준다.

- 딥러닝 학습 : 딥러닝 기반 신경망에 학습 데이터를 입력하여 예측을 수행하고 실제 데이터와 비교를 통해 신경망의 가중치를 조절하면서 학습을 수행한다. 딥러닝 기반 신경망 구성이 각각 계량기별로 되었기 때문에 신경망에 대한 학습도 계량기별로 수행한다.

- 가중치적용 : 딥러닝 예측을 통해 얻어진 결과를 기반으로 전체 합을 1로 하는 비율로 다시 계산한 구간예측비율을 구한다. 그리고 실제 결측 전·후 누적 전력 사용량 차분을 통해서 결측 구간의 총 사용량을 계산하여 각 구간예측비율과 곱하여 2차 시간별 구간 예측사용량을 만든다.

- 최종 예측값 계산 : 2차 예측된 시간별 구간 전력사용량을 기반으로 결측 이전 마지막 누적 전력사용량에 합산하여 최종 예측값을 계산한다.

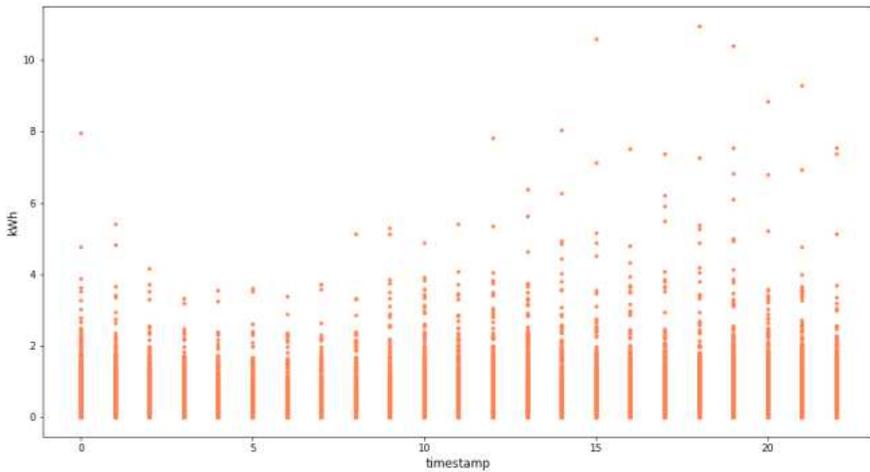


[그림 4-4] CNN-LSTM결합 가중치적용 모델 개념도 및 처리과정

B. 제안 모델 개발

1. 데이터 분석결과

이 절에서는 실제 전력사용량 패턴 분석을 통해서 CNN-LSTM결합 딥러닝 모델 적용 시 성공 가능성을 확인하고자 한다. [그림 4-5]는 실험 대상 1000개 계량기의 시간별 구간 전력사용량을 산점도로 표시하였다. 대부분의 계량기 데이터가 1시간 동안 전력사용량이 4 kWh를 넘지 않는다는 것을 알 수 있고, 또한 최댓값도 10 kWh가 넘지 않았다. 이것으로 실험용 데이터의 특이 값이 없음을 알 수 있다.



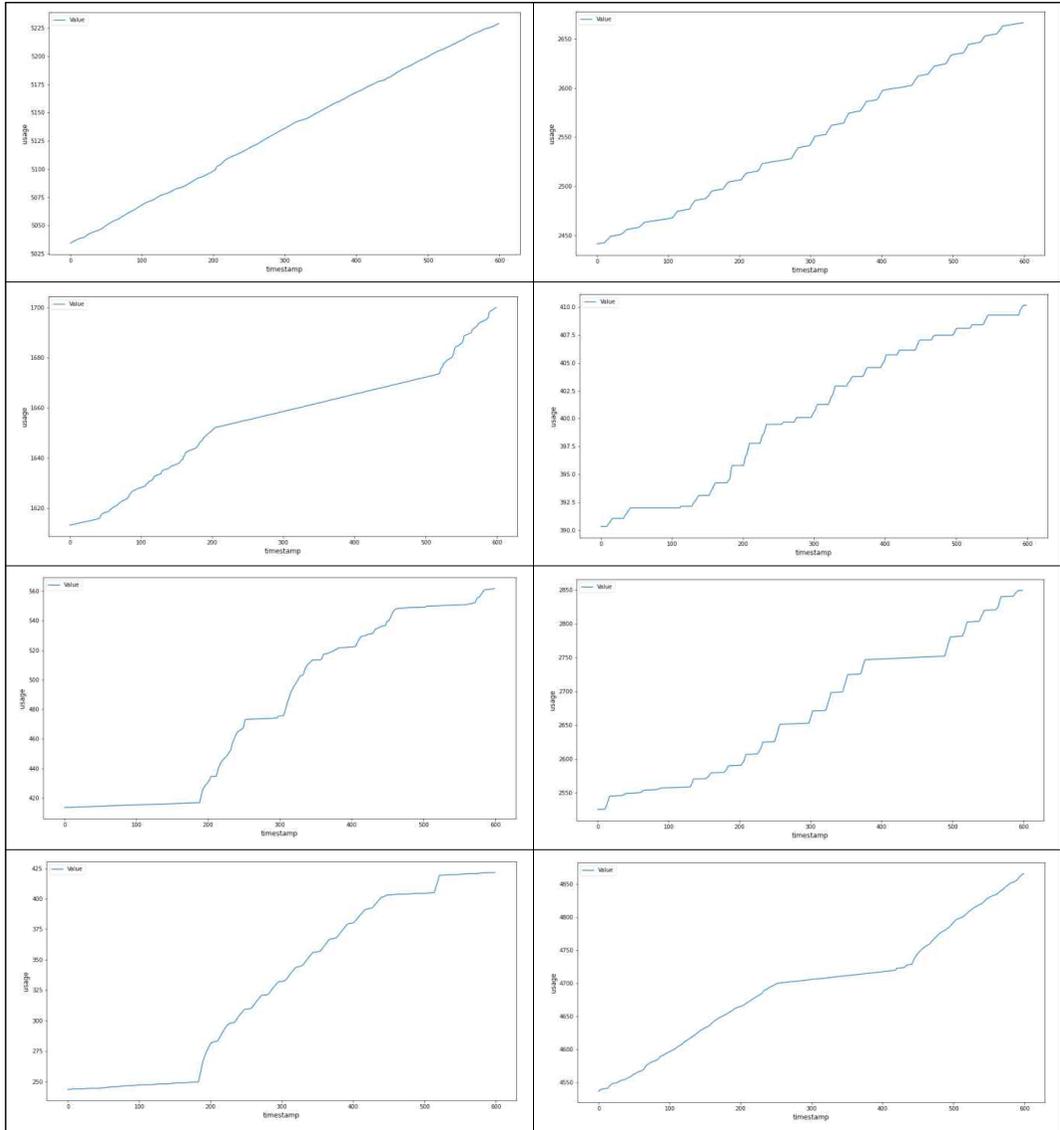
[그림 4-5] 시간별 구간 전력사용량 데이터 산점도

데이터 분석은 실제 누적 전력사용량과 차분을 한 후 구간별 전력사용량 패턴을 분석하고 그 의미를 파악하고자 한다.

먼저 아래 [표 4-1], [표 4-2], [표 4-3]은 누적 전력사용량 패턴을 표현하였다.

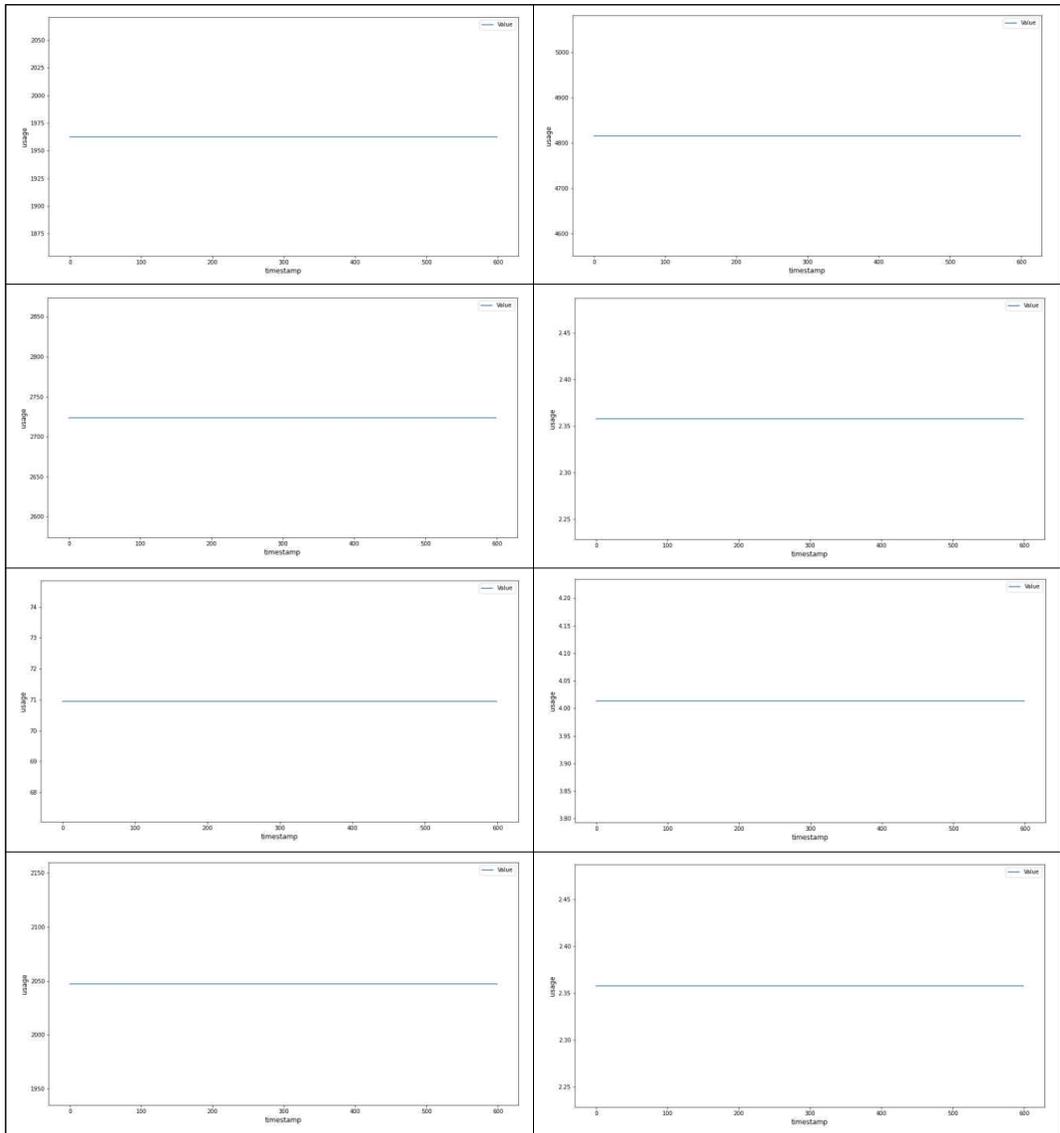
[표 4-1]은 일반적인 누적 전력사용량 패턴을 보여주고 있다. 7월 1일 00:00분 데이터부터 7월 25일 23:00까지의 시간별로 계속 증가하는 값을 가지고 있음을 알 수 있다. 특히 1000개의 계량기 중 976개가 [표 4-1]과 같은 패턴을 보였다.

[표 4-1] 일반적인 누적 전력사용량 패턴



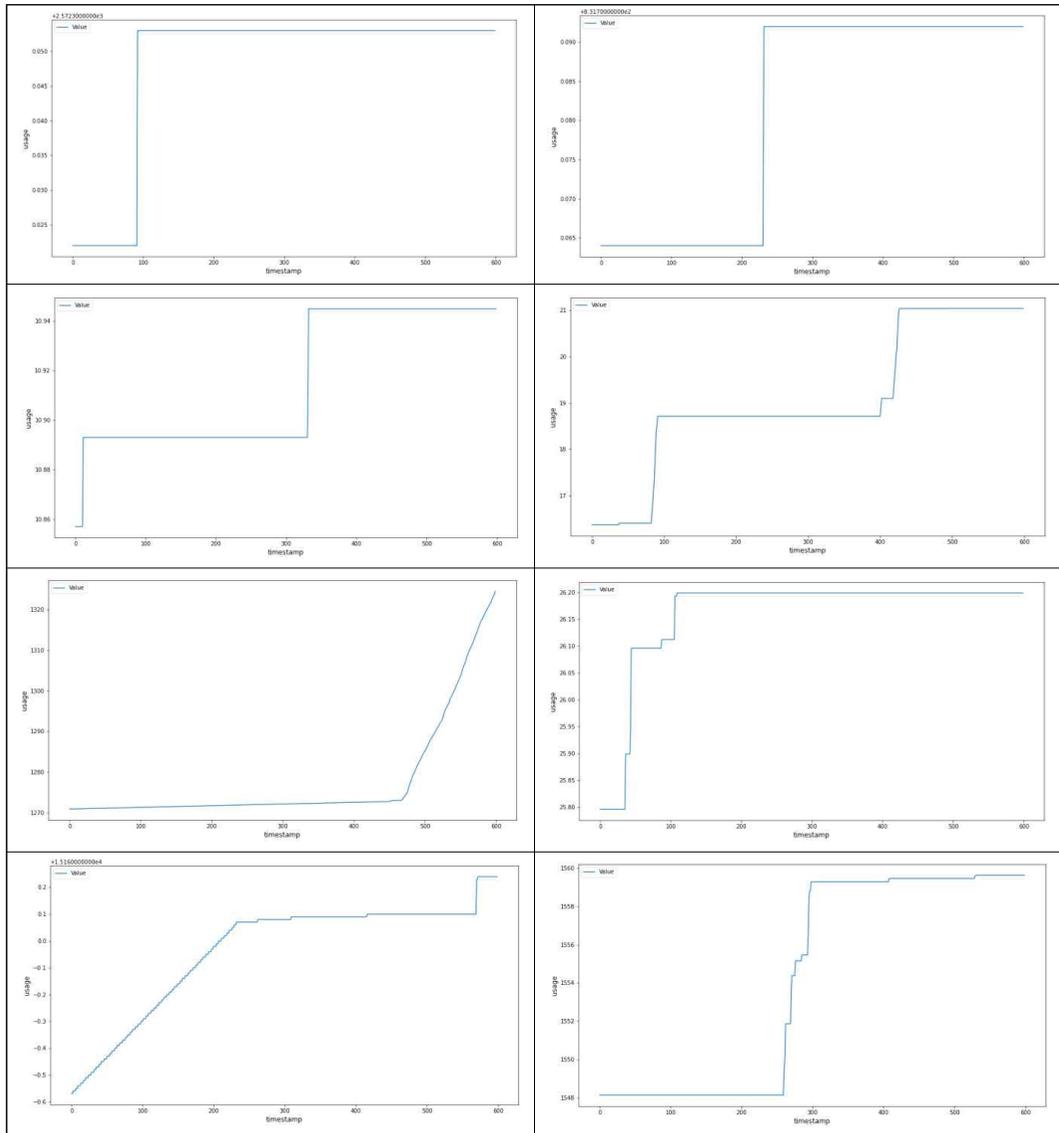
아래 [표 4-2]는 일반적인 누적 전력사용량 패턴과 다르게 사용량이 1개월 동안 전혀 변하지 않는 데이터들이 11개가 존재했다. 이 경우는 농촌에서 사용하는 전기 중에 논두렁 옆에 펌프를 설치하여 논에 물을 공급하기 위해 설치된 계량기의 경우 전기를 실제로 사용하지 않았을 수도 있고, 또는 전자식 전력량계나 데이터집중 장치의 오류로 인해 데이터 오류가 발생했을 수도 있다. 하지만 이런 데이터의 경우 예측이 실패하더라도 최종 가중치 보정을 적용하기 때문에 문제가 되지 않는다.

[표 4-2] 사용량 변경 없는 케이스의 누적 전력사용량 패턴



아래 [표 4-3]은 누적 전력사용량이 계속 변경되지 않고 1개월 중에 갑자기 특정 시간에만 급격하게 사용되고, 나머지 시간대에는 변경이 없는 특이사용 계량기가 23개 존재했다. 이 계량기들의 사용량을 분석해보니 평상시 전력사용량이 거의 없는 계량기들이며 그 사용량이 0.001 ~ 0.2 kWh 이내로 미미하여 그래프로 표현시 특이 케이스로 분류되었다. 이런 분류의 데이터가 딥러닝 알고리즘으로 예측 시에 그 정확도가 떨어질 수 있다.

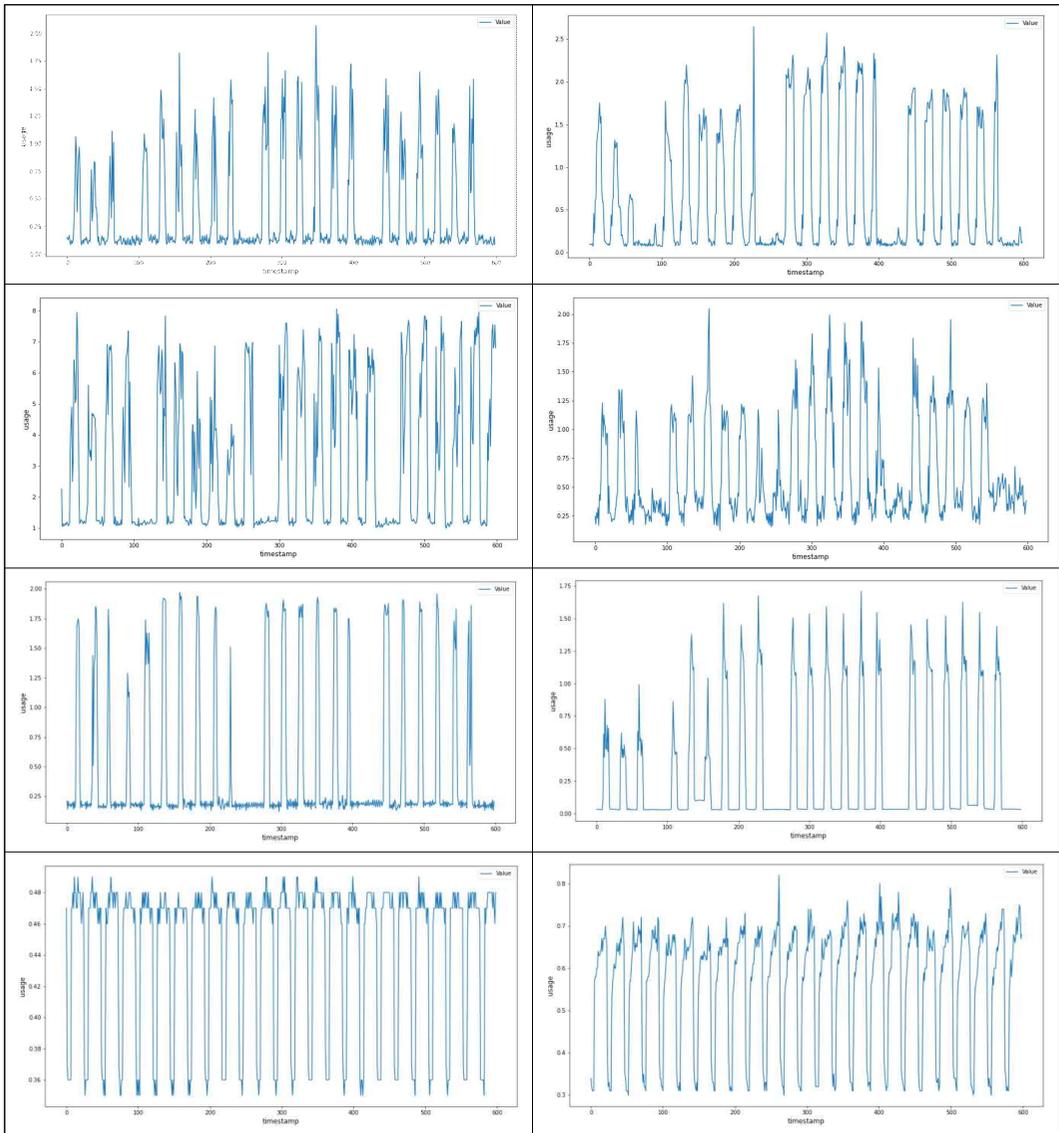
[표 4-3] 사용량이 특이한 케이스의 누적 전력사용량 패턴



다음은 아래 [표 4-4], [표 4-5], [표 4-6]은 시간대별 구간 전력사용량 패턴을 표현하였다.

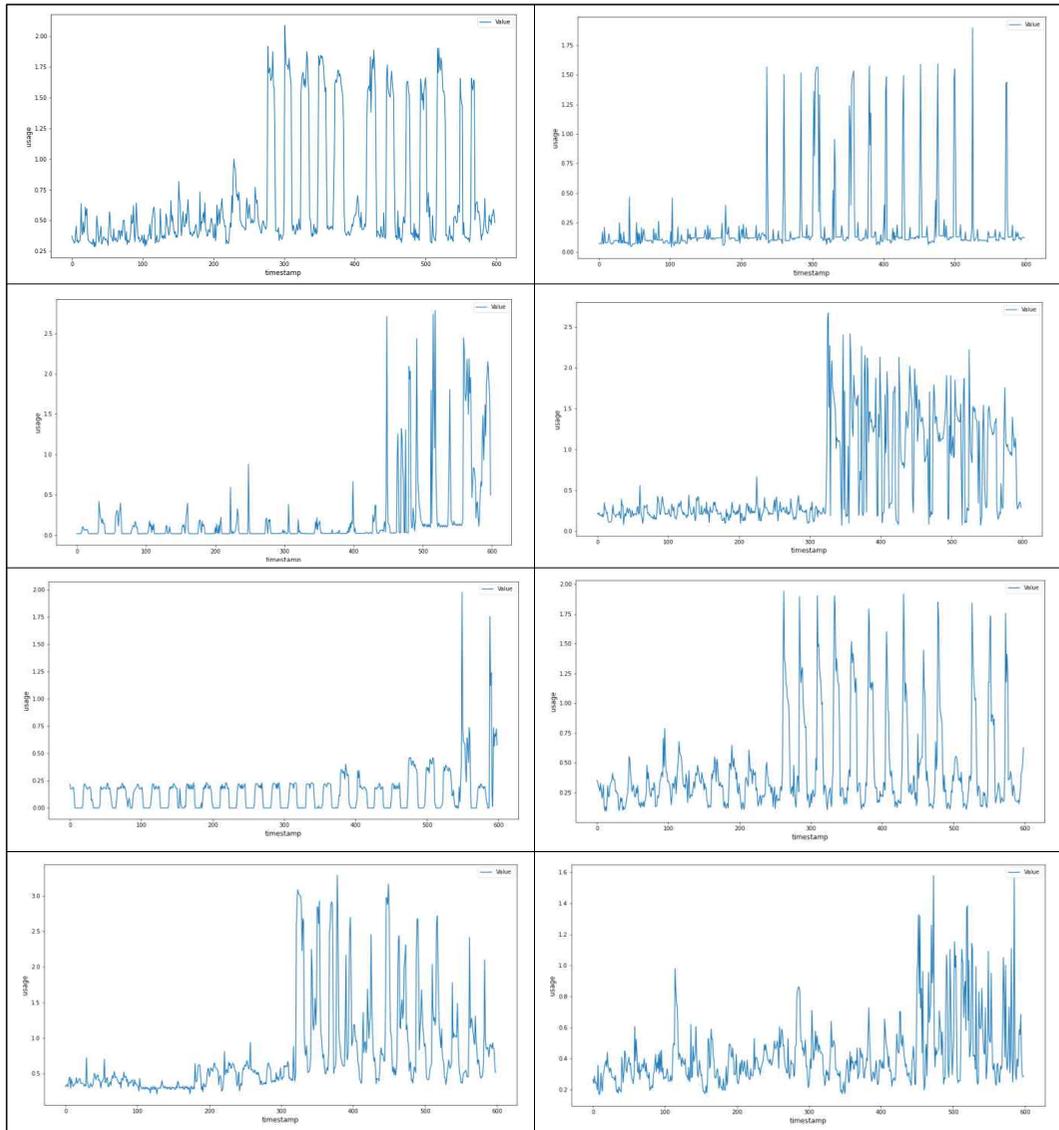
아래 [표 4-4]는 일반적인 시간별 구간 전력사용량 패턴을 보여주고 있다. 아래 계량기들은 월요일부터 토요일까지 아침 9시부터 18시까지 전기를 사용하고 있다. 저녁 시간 때와 일요일에는 전력사용량이 적은 것으로 봐서 이 계량기들은 주택이 아니고 회사인 것으로 추측된다.

[표 4-4] 일반적인 시간별 구간 전력사용량 패턴



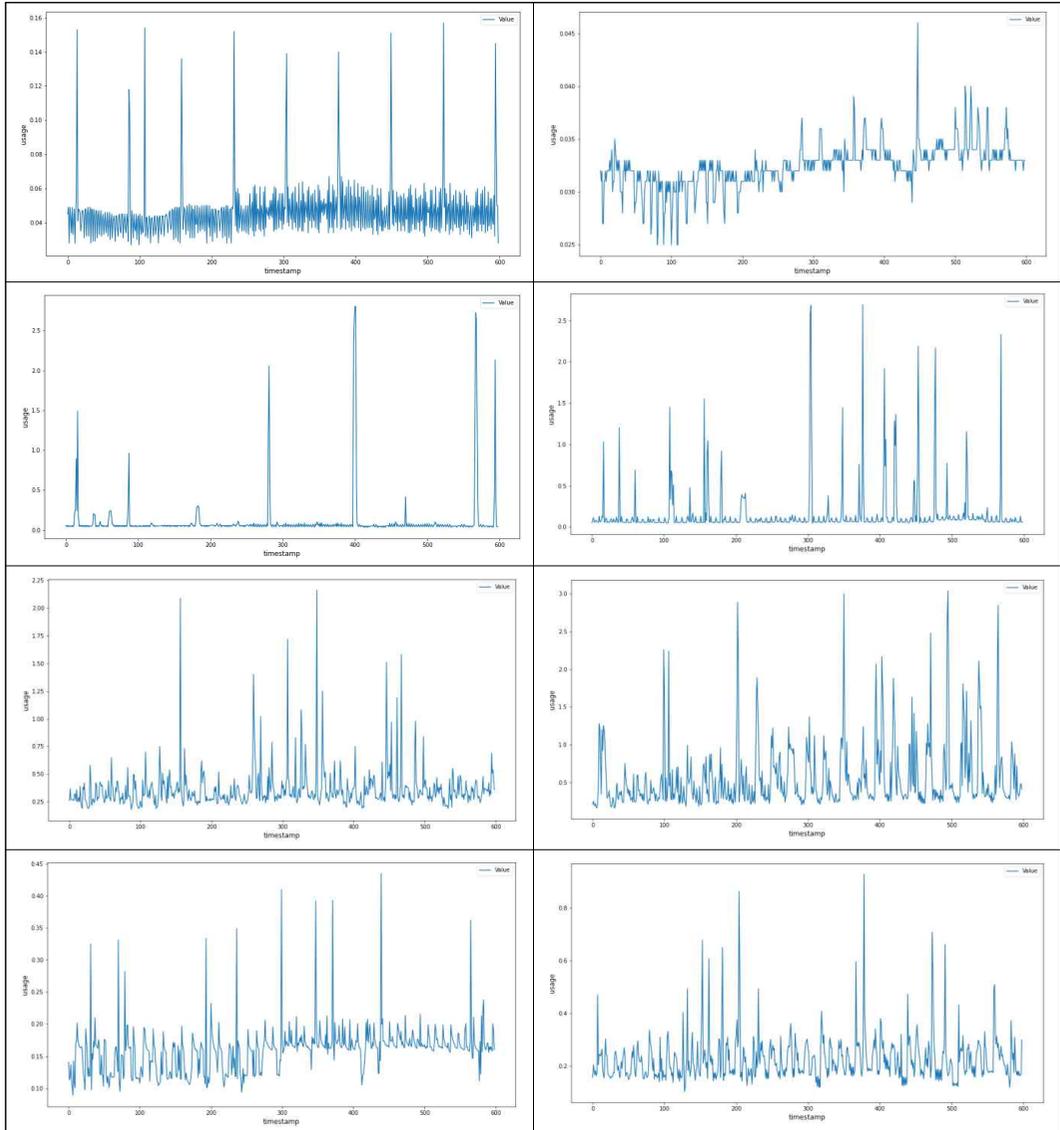
아래 [표 4-5]는 7월 중순경부터 갑자기 전력사용량이 늘어나는 패턴으로 보아 여름철에 에어컨 가동 등으로 갑자기 전력소비가 늘어났을 것으로 판단된다. 특정 일자에 갑자기 사용량이 크게 증가하였지만 연속성이 있으므로 일반적인 패턴으로 볼 수 있다.

[표 4-5] 여름철 특징인 시간별 구간 전력사용량 패턴



아래 [표 4-6]은 시간별 구간 전력사용량이 불규칙적인 패턴을 보이는 계량기들이다. 특별한 규칙이 없이 갑자기 전력사용량의 변화량이 크게 나타났다. 이런 데이터들은 딥러닝으로 예측하더라도 그 정확도가 현저히 떨어질 것으로 예상된다.

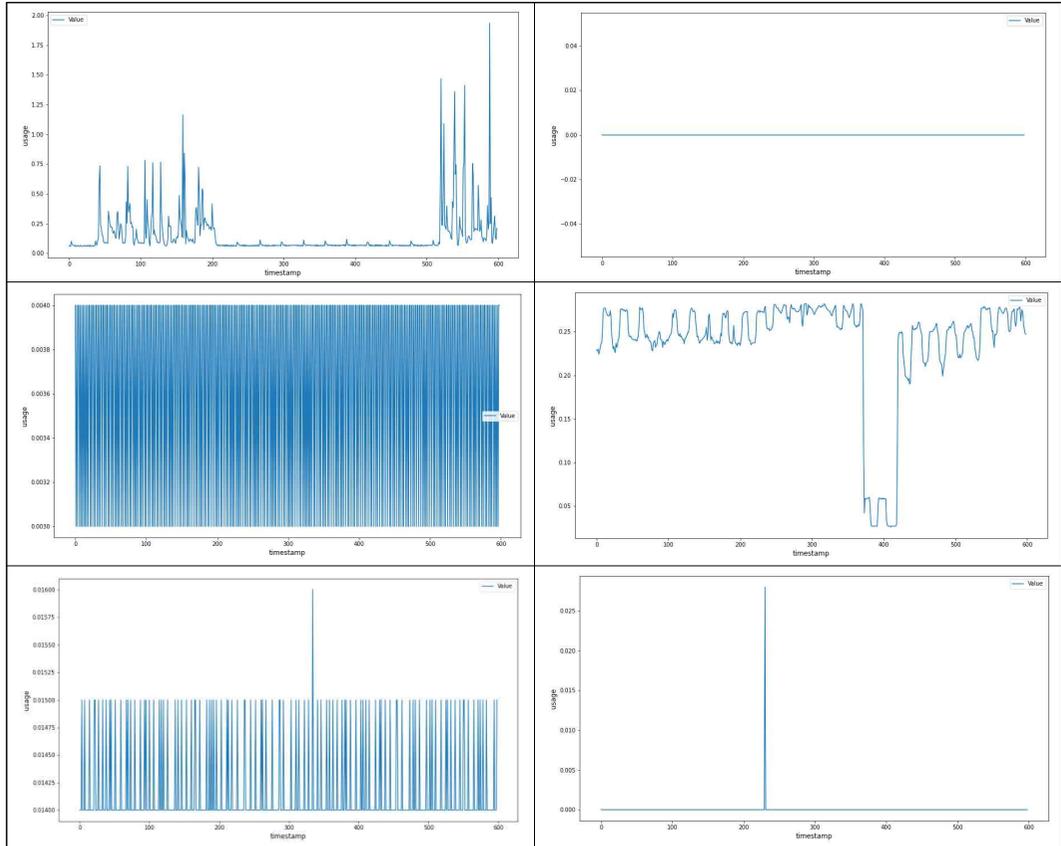
[표 4-6] 불규칙 사용량인 시간별 구간 전력사용량 패턴



아래 [표 4-7]은 누적 전력사용량 분석에서도 이상 값을 보인 것처럼 중간에 사용량 변화량이 없거나 1개월 동안 사용량이 전혀 없는 경우, 사용량이 미미하여 변화

지 않고 있다가 일시적으로 조금 증가한 경우, 또는 구간 전력사용량이 0.003, 0.004로 그 값이 너무 적어서 그래프로 표현 시 아래처럼 촘촘한 화살촉 모양으로 표현되었다.

[표 4-7] 특이한 케이스의 시간별 구간 전력사용량 패턴

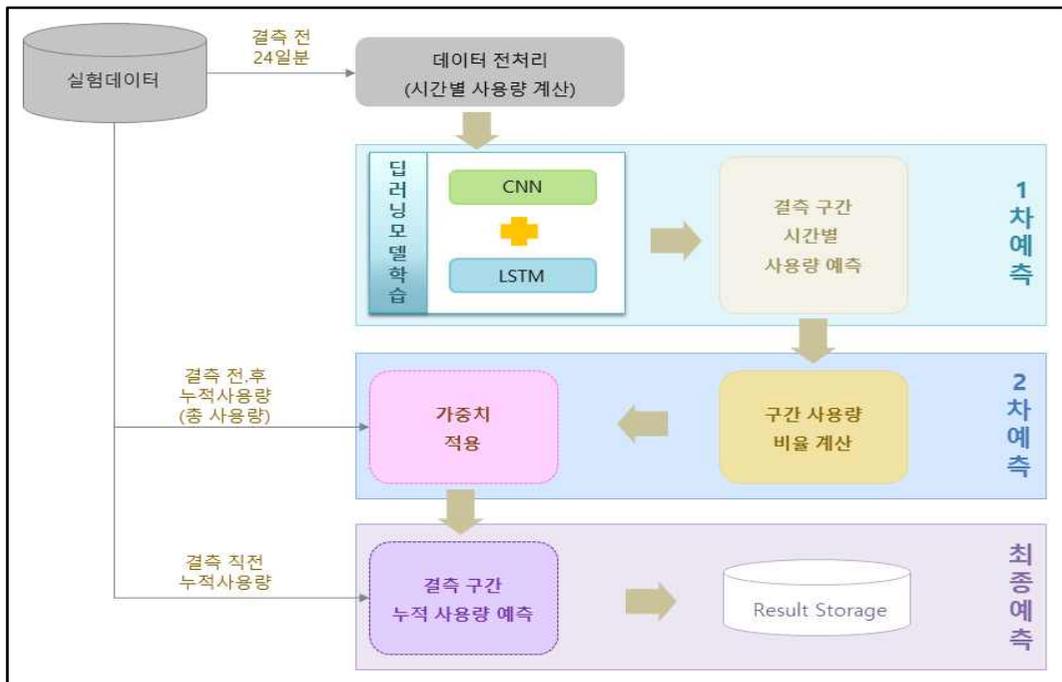


이처럼 다양한 형태의 전력사용량 패턴들이 존재하는 것을 알 수 있었고, 이것들이 전자식 전력량계나 데이터 집중장치의 오류로 인해서 발생할 수도 있다. 하지만 실제로 존재하는 데이터일 가능성이 더 크고 AMI시스템 또한 이러한 데이터 들이 수집되고 운영되고 있기 때문에 이렇게 비정상적인 데이터들도 고려해서 알고리즘을 연구해야 한다. 그래서 총 1000개의 샘플 데이터들을 제거하지 않고 모두 테스트하고 평가를 하려고 한다.

2. 모델 개발

본 연구는 딥러닝 기반 CNN-LSTM결합 가중치적용 모델 개발을 위해 텐서플로우(Tensorflow 2.5.0) 패키지를 이용해서 개발을 수행하였다.

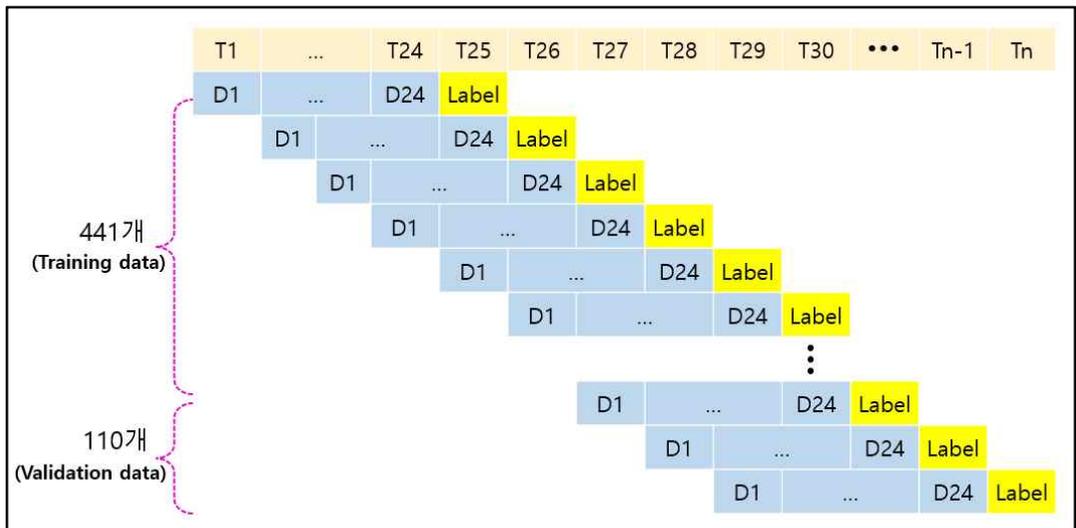
사전 데이터 분석을 통해서 각 계량기별로 전력 사용량 패턴들이 매우 다르게 나타나는 것을 알 수 있었다. 예로 상가, 주택, 공장, 학교 등 그 용도와 사용방식 때문에 전력사용 패턴들이 각기 다르다. 그래서 본 연구에서도 제안모델을 계량기별로 학습시키고 테스트를 진행하였다. [그림 4-6]은 모델을 개발하기 위한 전체적인 절차를 나타낸다. 모델 개발을 위해서 각 계량기별 1개월분 데이터를 가지고 실험을 진행하였다. 그래서 계량기별 누적 전력사용량 데이터는 1시간에 1개의 데이터를 가지고 있으므로 1일(24개) × 31일로 744개 데이터가 된다.



[그림 4-6] 모델의 전체적인 처리 절차

먼저 전처리 대상 데이터로는 7월 1일 0시 데이터부터 7월 24일 23시까지 데이터를 준비한다. 입력된 데이터는 누적 전력사용량 데이터이므로 (t)시간의 누적 전

력사용량에서 (t-1)시간의 누적 전력사용량을 차분하여 (t)시간의 구간 전력사용량을 계산하는 데이터 전처리를 수행한다. 그리고 전처리가 끝난 데이터를 이용하여 CNN-LSTM결합 딥러닝 모델에 입력을 위한 데이터 세트를 만든다. [그림 4-7]처럼 하나의 데이터 세트는 24시간 구간 전력사용량 데이터와 Label 데이터 1개로 구성하고, 1시간씩 쉬프트하면서 총 551개의 딥러닝 모델의 입력 데이터 세트를 만들었다. 본 논문에서 7월 25일 24개의 데이터가 결측 되었다고 가정하고 실험을 진행하기 때문에 학습데이터는 551개, 테스트 데이터는 24개로 실험을 진행하였다. CNN-LSTM결합 딥러닝 모델의 하이퍼파라미터 최적화를 위해서 학습데이터 551개 중 검증 데이터로 110개를 사용하였다. 검증 데이터는 모델 개발 시에만 사용하고, 전체 1000개 계량기 실험 시에는 사용하지 않는다.



[그림 4-7] 딥러닝 모델의 입력 데이터 세트

전처리된 데이터 세트를 입력하여 CNN-LSTM결합 딥러닝 모델에 입력하여 모델을 학습시키고 7월 25일 0시부터 7월 25일 23시까지의 구간 전력사용량 데이터를 예측한다. 이것을 1차 예측이라고 한다. 그리고 1차로 예측한 데이터에 가중치를 적용하여 2차 시간별 구간 전력사용량을 예측한다. 마지막으로 결측 이전 마지막 데이터인 7월 24일 23시 누적 전력사용량 데이터에 예측된 시간별 구간 전력사용량 데이터인 25일 0시 사용량을 합산하고, 다시 합산된 7월 25일 0시 누적 전력사용량에 예측된 구간 전력사용량인 7월 25일 1시 사용량을 합산한다. 이렇게 계속

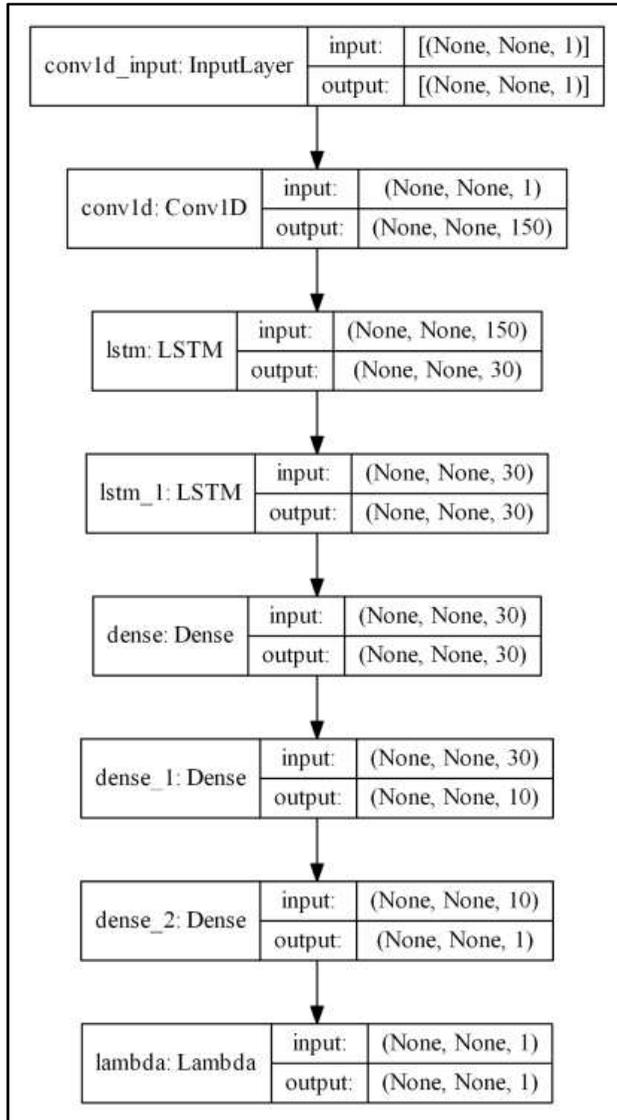
반복하여 결측 구간의 누적 전력사용량을 최종 예측한다.

본 논문에서는 딥러닝 기반 CNN-LSTM결합 가중치적용 모델 개발하기 위해서 우선적으로 CNN-LSTM결합한 딥러닝 모델의 정확도를 향상시킬 필요가 있고, 이를 위해 여러 가지 실험을 수행하고, 최적의 모델을 만들어서 가중치를 적용하고자 한다.

시간별 구간 전력사용량을 예측하는 모델을 아래 [그림 4-8]과 같이 만들었다. 계량기에서 생성되어 수집되는 데이터 종류는 여러 가지가 있으나 모두 누적 전력 사용량과 무관하며, 결측이 발생하게 되면 동시간대 모든 데이터가 결측 된다. 그래서 입력 데이터로 사용할 수 있는 데이터가 누적 전력사용량 데이터만 있으므로 1차원 CNN모델을 사용했다.

시계열 데이터에 적용하기 위해서 1차원 CNN(1D convolution)[90,91]을 첫 번째 레이어에 배치하여 전력사용량 패턴을 인식하여 성능을 향상시키고자 한다. Kernel size를 5, filter를 150개, padding은 'same', strides는 '1', activation function은 relu를 사용했다.

두 번째, 세 번째 레이어에는 LSTM을 배치하였고, 출력 레이어는 30개씩 하였다. 그리고 네 번째 레이어는 Dense Layer[92]를 출력 레이어 30개로 배치, 다섯 번째 레이어는 Dense Layer를 출력 레이어 10개로 배치, 그 다음 출력결과를 하나만 얻기 위해 Dense Layer를 출력 레이어 하나로 설정한다. 그리고 마지막으로 딥러닝 정확도 향상을 위해서 람다(Lambda) 레이어를 추가하여 scale-up을 150배 하였다.



[그림 4-8] 구간 전력사용량 1차 예측 네트워크 구조

딥러닝 모델을 만들고 모델을 학습시키기 위해서 loss function과 optimizer가 중요하다. 그래서 정확도 향상을 위해서 여러 종류의 loss function과 optimizer를 실험을 통해서 결정하고자 한다.

손실함수(loss function)는 머신러닝 혹은 딥러닝 모델의 예측값과 실제값(레이블)의 차이를 계산하는 함수이며, 차이(loss)를 최대한 줄이는 방향으로 학습이 진행된다. 여러 종류의 손실함수가 있으며 이항 교차 엔트로피(binary_crossentropy)

나 범주형 교차 엔트로피(categorical_crossentropy) 등은 분류 문제에 사용되고, 연속 값인 회귀 문제에 널리 사용되는 손실함수로써는 평균제곱오차(MSE; Means Squared Error)와 MSE에 루트($\sqrt{\quad}$)를 씌운 제곱근 평균제곱오차(RMSE; Root Mean Squared Error)가 있다.

회귀 모델을 학습시키는데 잡음 데이터가 조금 있을 경우 평균 제곱 오차는 큰 오차에 너무 과한 벌칙을 가하기 때문에 정확하지 않은 모델이 만들어질 수 있다. 평균 절댓값 오차는 이상치에 관대해서 학습이 수렴되기까지 시간이 걸린다. 그리고 모델이 잘 학습되지 않는다. 이런 경우 평균 제곱 오차대신 후버(Huber) 손실을 사용한다[93]. 식(4-5)처럼 후버(Huber) 손실은 오차가 1보다 작은 경우는 Square Error를 사용하고, 오차가 1보다 큰 경우는 오차값에서 0.5를 빼서 사용한다. 본 논문에서는 사전 데이터 분석결과에서 보듯이 이상치 데이터가 많이 존재한다. 그래서 손실함수로 후버(Huber)손실을 사용한다.

$$\begin{aligned}
 \text{SmallErrorLoss} &= \sqrt{\text{error}} & (4-5) \\
 \text{BigErrorLoss} &= |\text{error}| - 0.5 \\
 \text{HuberLoss} &= IF(|\text{error}| \leq 1) \text{SmallErrorLoss} : \text{BigErrorLoss}
 \end{aligned}$$

optimizer는 여러 종류가 있는데 그 중에서 본 연구에서는 몇 가지를 실험하여 최적의 optimizer를 선정하고자 한다.

첫 번째로 확률적 경사 하강법(SGD; Stochastic Gradient Descent)은 학습할 때 전체 데이터에 가중치를 조절하지 않고, 일부 추출한 랜덤한 데이터에 가중치를 조절한다[94]. 결과적으로 속도는 개선되지만 정확도는 낮다.

두 번째로 모멘텀(Momentum)은 경사 하강법처럼 매회 기울기를 계산하지만, 가중치를 수정하기 전에 수정 전 방향(+,-)을 참고하여 계속 같은 방향으로 일정한 비율만 수정하는 방법이다. 그래서 이전에 같은 값을 고려해서 다음 값을 결정하므로 관성의 효과가 나타날 수 있다[95].

세 번째로 아다그라드(Adagrad)는 변수의 갱신 횟수에 따라 학습률을 수정하는 옵션이 추가된 방법이며, 크게 변화하지 않은 변수는 학습률을 크게하고, 크게 변화한 변수들은 학습률을 적게 한다[96]. 이는 크게 변화한 변수는 최적값에 근접했을 것이라는 가정하에 작은 크기로 이동하면서 세밀한 값을 조정하고, 적게 변화한 변수들은 학습률을 크게 하여 빠르게 loss값을 줄인다. 또한 같은 입력 데이터가

여러번 학습되는 학습모델에 유용하게 쓰인다.

네 번째로 알엠에스프롭(RMSprop)은 지수 이동평균으로, 아다그라드의 $G(t)$ 의 값이 무한정 커지는 것을 방지할 수 있는 방법이다[97].

다섯 번째 아담(Adam; Adaptive Moment Estimation)은 Momentum과 RMSprop를 조합한 경사하강법이다. RMSprop의 gradient 제곱을 지수 평균한 값을 사용하는 것이 Momentum의 특징, gradient를 제공하지 않은 값을 사용하며 지수평균을 구하고 수식에 활용한다[98].

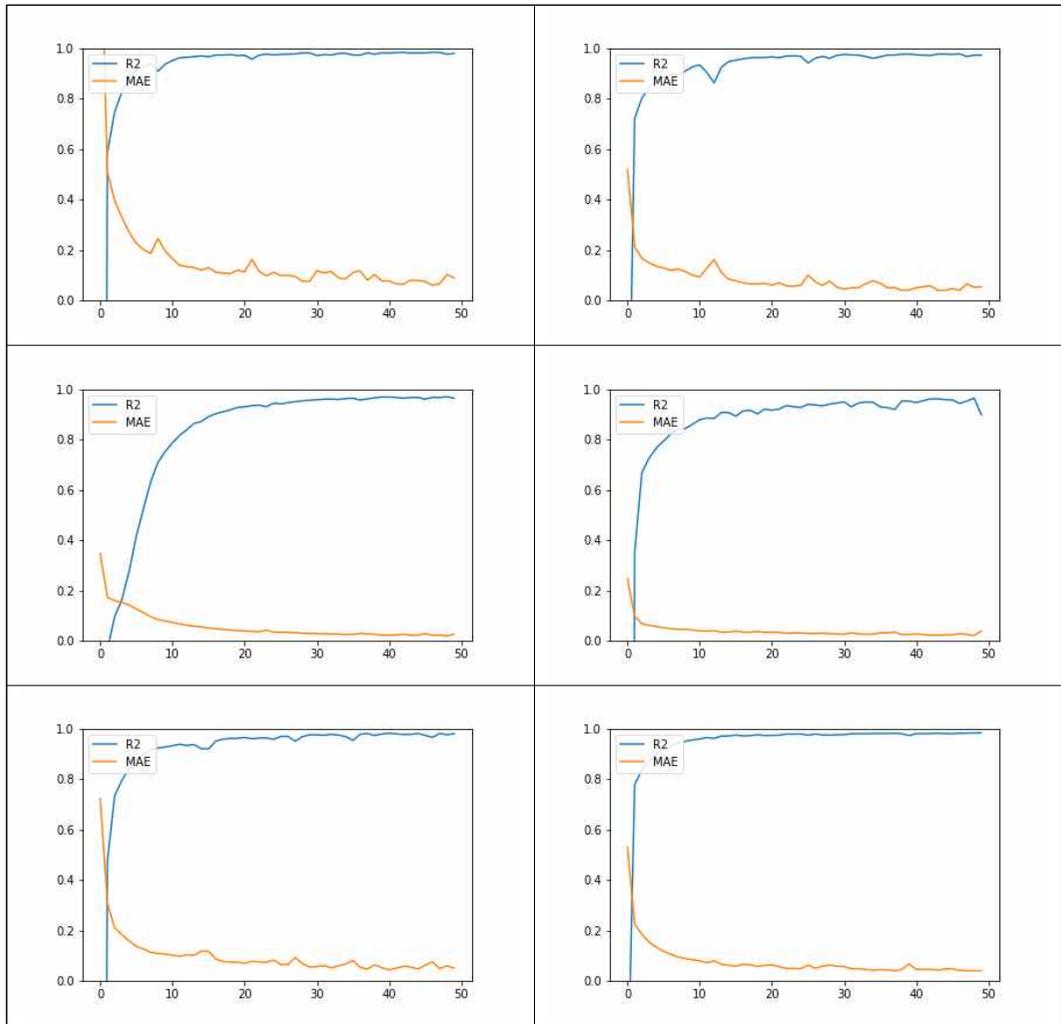
아래 [표 4-8]는 optimizer를 선정하기 위해 샘플 데이터를 대상으로 실험을 진행하였다. 그 결과 Adam 이 본 논문에서 제안한 예측 모델에는 가장 좋은 성능을 보였다. 그래서 본 연구의 실험에도 Adam을 사용하여 모델을 학습한다.

[표 4-8] optimizer 종류별 MSE, MAE

optimizer	SGD & Momentum	Adagrad	RMSprop	Adam
MSE	0.001924465	0.002298609	0.0047534	0.000537506
MAE	0.036171153	0.036102742	0.0540142	0.019394478

다음은 에포크(epoch)의 수에 따른 모델의 성능을 비교해보고 적절한 반복횟수를 선정한다[99]. x축은 에포크(epoch) 개수이다. 계량기마다 반복학습 횟수에 따라 정확도가 달라지기 때문에 에포크는 크게 하면 좋을 수도 있지만 오버피팅(overfitting)으로 정확도가 떨어질 수도 있고[100], 실제 업무에 적용 시에 서버 자원의 사용률도 고려해야 함으로 적정수준을 찾는 것이 바람직하다. 물론 1000개 계량기의 모든 모델을 비교평가 할 수는 없으므로 몇 개의 샘플 데이터에 대해서 에포크 개수별 정확도를 비교했다. [표 4-9]에서 보듯이 에포크 수가 증가함에 따라 오차가 줄어들고 R-Squared 값이 증가하는 것을 알 수 있다. 그 결과 에포크 개수가 30 이상에서는 크게 좋아지지 않는다. 그래서 에포크 개수는 30개로 선정했다.

[표 4-9] 에포크(epoch)에 따른 정확도



1차 예측을 위한 CNN-LSTM결합 딥러닝 모델을 만들었고, 여기에 가중치를 적용한 2차 예측을 수행한다. 결측 전·후 총 사용량 범위를 벗어나지 않도록 보정하기 위해서 1차 예측한 시간별 구간 전력사용량을 총 합이 1이 되도록 각 구간별 비율을 계산하고, 각 구간별 비율에 결측 구간의 총 사용량 값을 곱하여 2차 예측한다.

그리고 마지막으로 결측 직전 누적 전력사용량 데이터에 2차로 예측한 각 구간별 가중치가 적용된 시간별 구간 전력사용량을 합산하여 최종 예측을 수행한다.

본 논문에서 제안한 모델을 적용한 Algorithm 1을 설명하고자 한다. 먼저 입력데이터로는 실험 데이터로 준비한 계량기번호 목록이다. 계량기 번호를 먼저 DB에서 조회한다. 계량기별 누적 전력사용량 데이터를 조회하고, 차분을 통해 시간별 구간 전력사용량을 계산하고, 입력 데이터셋을 준비하는 전처리 과정을 수행한다. 그리고 전처리된 데이터를 딥러닝 모델에 입력하여 학습시키고 결측 구간의 구간 전력사용량을 1차 예측한다. 결측 전·후 누적 전력사용량 차분을 통해 전체 구간 전력사용량을 계산하고 이를 기반으로 1차 예측된 사용량의 가중치를 계산하여 2차 예측을 한다. 2차 예측된 데이터를 이용해서 최종 누적 전력사용량을 예측하고, 결과를 저장한다.

Algorithm 1. Weighted CNN-LSTM Processing

Input: MeterList

Output: ResultDataPool

Definition: *Ra* - After real data(first real data after missing termination)

Rb - Before real data(real data just before missing)

```

1  begin
2    Initialize list TmpRetPool;
3    READ MeterList with MissingValues;
4    for each MID in MeterList do
5      AccumulatedUsage = AccumulatedUsageDB.getvalue(MID);
6      IntervalUsage = COMPUTE IntervalUsage by time using AccumulatedUsage;
7      Model = DeepModel.fit(IntervalUsage);
8      DeepModel_IntervalUsage = Model.predict;
9      Sum_DeepModel_IntervalUsage =  $\sum$  DeepModel_IntervalUsage;
10     TmpRet = Rb
11     for each EIU in DeepModel_IntervalUsage do
12       Weighted_Usage = (Ra - Rb) x EIU / Sum_DeepModel_IntervalUsage ;
13       TmpRet = TmpRet + Weighted_Usage;
14       TmpRetPool = TmpRetPool.putvalue(TmpRet);
15     end for
16     ResultDataPool = ResultDataPool.putvalue(TmpRetPool);
17     ClearTmpRetPool;
18   end for
19 end
    
```

V. 실험 및 평가

본 장에서는 제안된 모델을 실험절차에 따라 실험하고 그 결과를 선행연구 결과들과 비교 분석을 통해 실제 업무시스템에 적용가능성 여부를 검토한다.

A. 실험 과정 및 결과

1. 자료수집 및 대상 범위 선정

본 연구를 위해 실제 주택의 전력사용량 데이터를 취득하고자 정보공개포털[그림 5-1]을 통해 한국전력공사로부터 연구용 목적으로 임의의 데이터 1000개 계량기의 2021년 7월분 전력사용량 데이터를 요청하여 실험을 진행하였다.

이 데이터는 전자식 전력량계의 누적 전력사용량 LP(Load Profile) 데이터이며 계량기 번호는 비식별화 처리하였고, 임의의 1000개 계량기 데이터는 실험결과를 비교분석하기 위해서 결측이 없이 모두 존재하는 데이터를 대상으로 하였다.

전자식 전력량계의 종류에 따라 LP기록방식이 15분, 60분 등으로 분류되는데 본 연구에서는 국내 점유율 85% 이상 가장 많이 사용하고 있는 전자식 전력량계 종류인 E-Type 과 AE-Type을 대상으로 실험을 진행한다. 따라서 해당 종류는 LP 기록방식은 60분이며 하루에 24개의 레코드가 존재한다. 15분마다 기록되는 G-Type 계량기의 데이터를 사용하더라도 본 연구의 알고리즘은 크게 변경되지 않는다.

실험용 데이터가 사용량이 전혀 없거나 특정 일자에만 사용량이 발생하는 등의 특이치가 있어도 실험대상에 그대로 포함시킨다. 실제 데이터도 이러한 특이치의 데이터가 존재할 것이기 때문에 실제 AMI시스템과 똑 같은 조건으로 실험을 하기 위함이다. 만약 실험대상 데이터 중에 특이치 외에 누적 전력사용량이 감소되는 데이터는 실험에서 제외시킨다. 전자식 전력량계는 LP데이터를 unsigned integer 형식의 pulse 값으로 관리되므로 그 범위가 0~4,294,967,295 pulse이며, 사용량 값은 계기정수 약 1000으로 나눈 0~4,294,967 이내이기 때문에 최댓값에 도달했을 경우 다시 0으로 회전하게 된다. 이 경우를 제외하고는 누적 전력사용량이 감소되는 현상이 발생하면 안 된다. 이렇게 회전하는 계량기는 시험에서 제외시킨다.

실험에서 사용한 데이터는 1000개 계량기의 2021년도 7월 1일~31일까지 1개월분

데이터를 사용하여 실험을 진행하였다. 본 연구의 목적이 중간에 결측이 발생한 것
 을 가정하고 실험하기 때문에 1000개 계량기의 샘플 데이터는 결측이 없이 100%
 실제 존재하는 계량기 데이터를 대상으로 했다. 계량기별로 1시간마다 생성된 누적
 전력사용량 레코드이므로 전체 레코드 수는 744,000개 이다. [1000(계량기) × 31(일)
 × 24(시간) = 744,000]

■ 공공기관의 정보공개에 관한 법률 시행규칙 [별지 제1호외2서식] <개정 2021. 6. 28.> 정보공개시스템(www.open.go.kr)		인쇄일자 : 2021. 10. 11. 12:42:15 정보공개시스템(www.open.go.kr)
정보공개 청구서		
※ 색상이 어두운 칸은 신청인(대리인)이 작성하지 않습니다.		
접수번호 8172316	접수일 2021. 08. 26.	처리기간 10일(2021. 09. 13.)
청구인	성명(법인·단체명 및 대표자 성명) 권혁복	생년월일(성별) 1974. 09. 14 (남)
	여권·외국인등록번호(외국인의 경우 작성)	사업자(법인·단체)등록번호
	주소(소재지) 전라남도 나주시 우정로 77 106-1602	전화번호(또는 휴대전화번호) 010-3818-1237
	전자우편주소 sephia22@nate.com	팩스번호 (—)
청구 내용	[AMI계량데이터의 누락시 데이터 보정 연구를 위한 연구용 데이터 협조 요청] 1.목적 - AMI 계량데이터의 누락이 발생한 경우 이를 보정하기 위한 연구 2.요청 데이터 - 시간별 생성되는 계량데이터 (계기번호(비식별), 날짜, 시간, 검침값) - 기간 : 계량기 1000대 * 1개월분(2021년 7월분) - 조건 : 1개월 동안 한번도 누락이 없는 계량기 대상 ※ 계기번호는 비식별화 처리(예:M0001, M0002, ...)	
공개 방법	<input type="checkbox"/> 열람·시청 <input type="checkbox"/> 사본·출력물 <input checked="" type="checkbox"/> 전자파일 <input type="checkbox"/> 복제·인화물 <input type="checkbox"/> 기타 ()	
수령 방법	<input type="checkbox"/> 직접방문 <input type="checkbox"/> 우편 <input type="checkbox"/> 팩스전송 <input checked="" type="checkbox"/> 정보통신망 <input type="checkbox"/> 전자우편 등 ()	
수수료	<input type="checkbox"/> 감면 대상임 감면 사유	<input checked="" type="checkbox"/> 감면 대상 아님
	※ 「공공기관의 정보공개에 관한 법률 시행령」 제17조제3항에 따라 수수료 감면 대상에 해당하는 경우에만 적용되며, 감면 사유를 증명할 수 있는 서류를 첨부하시기 바랍니다.	
「공공기관의 정보공개에 관한 법률」 제10조제1항 및 같은 법 시행령 제6조제1항에 따라 위와같이 정보의 공개를 청구합니다.		
청구인		2021년 08월 26일 권혁복 (서명 또는 인)
(한국전력공사사장) 귀하		

[그림 5-1] 정보공개포털 청구신청내역

2. 실험 환경

전력 사용량 데이터 결측 보정 연구의 실험을 위한 H/W와 S/W 환경은 [표 5-1]과 같다.

[표 5-1] H/W 및 S/W 환경

구분	상세 내용
Hardware 환경	<ul style="list-style-type: none"> - OS : Windows 10 64bit - CPU : Intel(R) Core(TM) i5-8265U CPU @ 1.60GHz 1.80 GHz - MEM : 8 GB
Software 환경	<ul style="list-style-type: none"> - Database : Oracle Express Edition (XE) Release 18.4.0.0.0 (18c) - Language : Python 3.8.8 (Anaconda 4.10.1 64-bit) - Deep learning : Tensorflow 2.5.0 (with keras)

위 [표 5-1]에서 제시한 바와 같이 실험환경이 개인용 컴퓨터에서 수행되었다. 그래서 예측 보정 실험을 하나의 계량기씩 순차적으로 처리하였다. 실제 업무에 적용 한다면 H/W성능에 맞춰 병렬로 처리하면 좋을 것이다. 실험 데이터가 1000개 계량기의 데이터로 실험하기 때문에 중간 결과 저장 및 결과 비교를 위해서 오라클 데이터베이스를 개인용 컴퓨터에 설치하여 데이터를 저장 관리하였다.

3. 실험 절차

본 연구를 실험하기 위해 준비된 1000개 계량기의 1개월분의 데이터를 실험대상으로 한다. 또한 결측 실험을 위해서 결측이 없는 원본 데이터에서 임의로 데이터를 제거하여 실험한다. 이때 실험은 결측 길이에 따른 비교실험과 선행 연구된 보정법과의 비교실험으로 진행된다. 선행 연구된 데이터 보정법인 선형보정법, ARIMA 예측 보정법, 유사도 기반 보정법, CNN-LSTM결합 예측 보정법 등을 실험한다. 또한 가중치적용의 장점을 확인하기 위해 유사도 기반 보정법에 가중치를 적용한 유사도 가중치적용 보정법과 마지막으로 본 논문에서 제안하는 CNN-LSTM결합 가중치적용 보정법을 모두 실험하고 그 결과를 DB에 저장하고 비교 분석한다.

a. 선형보정법

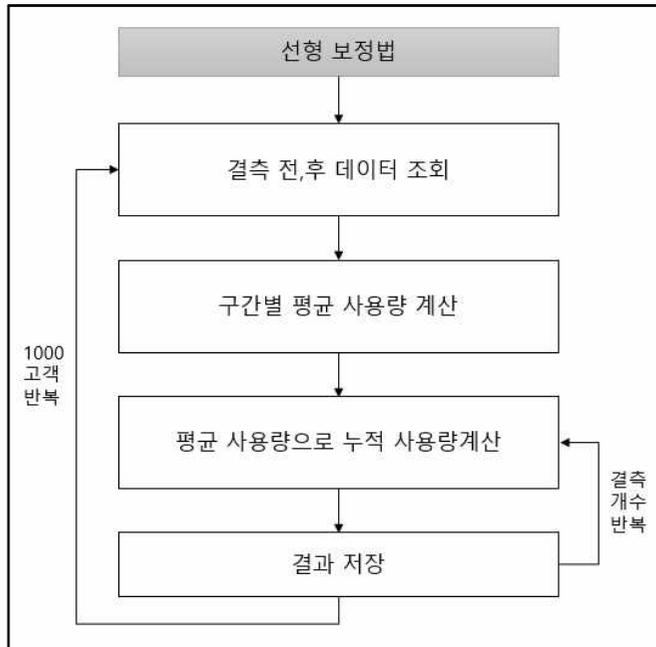
[그림 5-2]는 선형보정법의 실험 절차이다. 먼저 첫 번째 계량기의 결측 전·후 데이터, 즉 결측 전 데이터인 7월 24일 23:00분 누적 전력사용량 데이터와 결측 후 데이터인 7월 26일 00:00분 누적 전력사용량 데이터를 조회한다. 그리고 식(5-1)처럼 결측 후 나타나는 첫 번째 데이터(R_a)에서 결측 전 마지막 데이터(R_b)를 빼면 결측 구간의 총 사용량이 되고 이를 결측 구간 개수($N+1$)로 나누면 평균 사용량이 된다.

$$M_{avg} = \frac{R_a - R_b}{N+1} \quad (5-1)$$

그리고 식(5-2)처럼 평균사용량을 기반으로 결측 전 마지막 데이터(R_b)에 결측 구간의 평균사용량(M_{avg})을 더하면 첫 번째 결측 시간의 누적 전력사용량 데이터(M_1)가 된다. 이렇게 결측 개수만큼 반복 수행하여 그 결과를 DB에 저장한다.

$$\begin{aligned}
 M_1 &= R_b + M_{avg} \\
 M_2 &= M_1 + M_{avg} \\
 &\vdots \\
 M_n &= M_{n-1} + M_{avg}
 \end{aligned} \quad (5-2)$$

실험대상 1000개 계량기를 반복수행하여 각 계량기별 결측 구간의 예측 결과값을 저장한다.



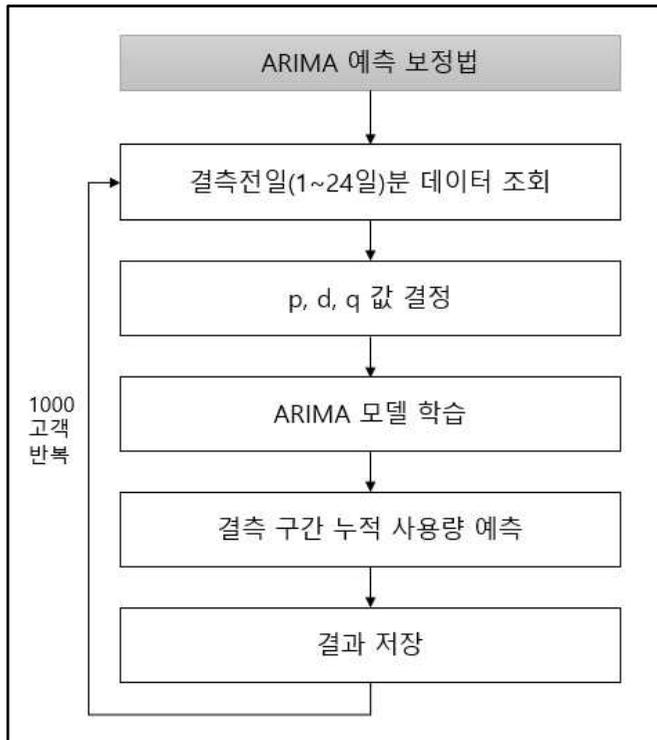
[그림 5-2] 선형보정법 실험 절차

b. ARIMA 예측 보정법

[그림 5-3]은 ARIMA(Autoregressive Integrated Moving Average) 모델을 활용해서 결측 구간의 누적 전력사용량을 예측하는 실험 절차이다. 먼저 결측 직전 7월 1일 0시부터 7월 24일 23시까지의 데이터를 조회한다. 그리고 선행 연구에서는 p , d , q 값을 하나의 특정 계량기를 대상으로 결정했다면 본 논문에서는 더 좋은 결과를 얻고자 `auto_arima`를 이용해서 p , d , q 값을 각 계량기별로 결정을 하고 ARIMA 모델을 학습시켜 사용량을 예측하려고 한다. `auto_arima` 알고리즘은 AIC(Akaike's Information Criterion)가 최소가 되는 모델을 자동으로 선택하게 되는데[101,102], 이 AIC는 식(5-3)로 나타낼 수 있다.

$$AIC = -2\text{Log}(L) + 2(p + q + k + 1) \quad (5-3)$$

여기에 L 은 데이터의 가능도, $c \neq 0$ 이면 $k=1$ 이고, $c=0$ 이면 $k=0$ 이다. 또한 ARIMA 모델은 전처리를 하지 않고 바로 누적 전력사용량을 예측하게 되므로 다른 보정법 보다는 빠르게 수행하고 결과를 얻을 수 있다. 먼저 결측 직전 7월 1일 0시부터 7월 24일 23시까지의 데이터를 조회한다. auto_arma 알고리즘을 이용하여 p, d, q 값을 자동으로 결정하고 결측 구간의 누적 전력사용량을 예측하여 저장한다. 이렇게 실험대상 1000개 계량기를 반복수행하여 각 계량기별 결측 구간의 예측 결과값을 저장한다.



[그림 5-3] ARIMA 예측 보정법 실험 절차

c. 유사도 기반 보정법

[그림 5-4]는 유사도 기반 보정법의 실험 절차이다. 해당 계량기의 전력사용량

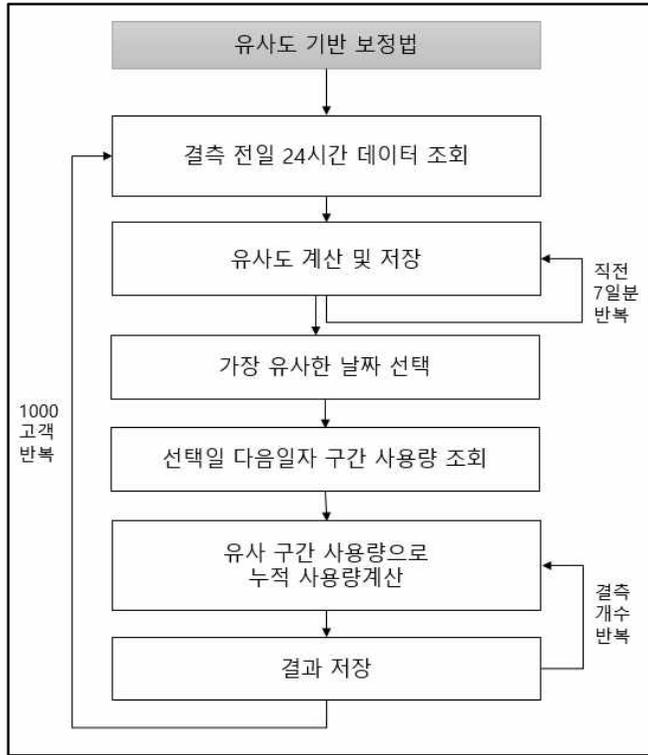
패턴이 결측 구간과 가장 유사한 날짜를 찾기 위해 하루 전 일자(P)와 유사한 날짜(Q)를 찾아서 유사일 하루 뒤 날짜의 시간별 구간 전력사용량을 기반으로 데이터를 보정 하려고 한다. 먼저 첫 번째 계량기의 전일 24시간 데이터를 조회한다. 전일을 기준으로 앞으로 나가면서 1일분씩 같은 시간별 구간 전력사용량으로 식(5-4)처럼 유클리드 거리 유사도를 계산한다.

$$\sqrt{(P_1 - Q_1)^2 + (P_2 - Q_2)^2 + \dots + (P_n - Q_n)^2} = \sqrt{\sum_{i=1}^n (P_i - Q_i)^2} \quad (5-4)$$

그래서 가장 유사한 날짜를 찾고, 유사한 날짜 다음날(참조일)의 시간별 구간 전력사용량을 조회한다. 그리고 식(5-5)처럼 유사한 날짜 다음날(참조일)의 시간별 구간 전력사용량을 기반으로 결측 전 마지막 데이터(R_b)에 참조일의 같은 시간 구간 전력사용량(S_1)을 합산하면 첫 번째 결측 시간의 누적 전력사용량 데이터(M_1)가 된다. 이렇게 결측 개수만큼 반복 수행하여 그 결과를 DB에 저장한다.

$$\begin{aligned} M_1 &= R_b + S_1 \\ M_2 &= M_1 + S_2 \\ &\vdots \\ M_n &= M_{n-1} + S_n \end{aligned} \quad (5-5)$$

실험대상 1000개 계량기를 반복수행하여 각 계량기별 결측 구간의 예측 결괏값을 저장한다.



[그림 5-4] 유사도 기반 보정법 실험 절차

d. 유사도 가중치적용 보정법

[그림 5-5]는 유사도 가중치적용 보정법의 실험 절차이다. 이 실험은 유사도 기반 보정법과 같이 과거 유사한 날짜를 찾아서 참조일을 선택한다. 이를 1차 예측이라고 하며, 1차 예측한 데이터에 가중치를 적용한 2차 예측, 그리고 2차 예측한 시간별 구간 전력사용량 기반으로 누적 전력사용량을 계산한다.

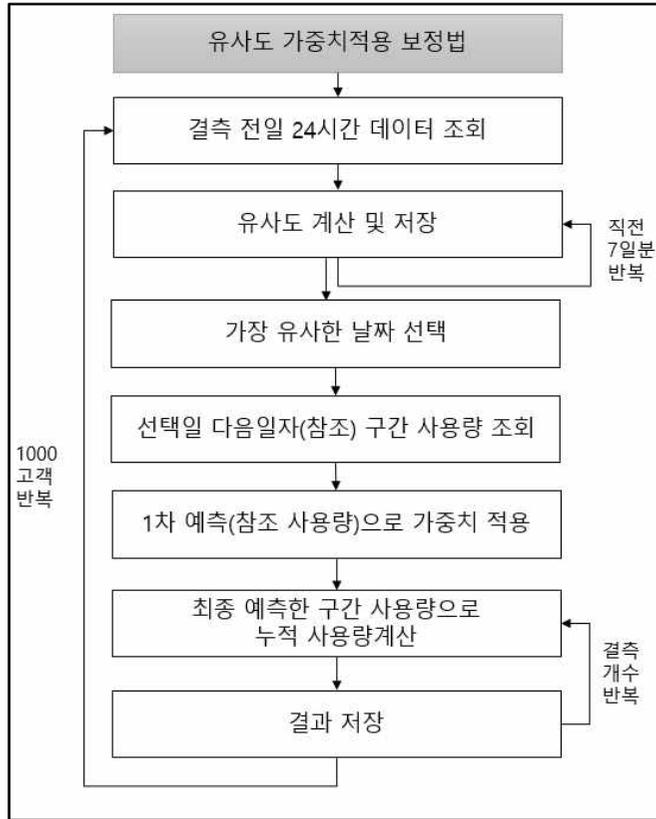
결측 구간의 시간별 구간 전력사용량 예측은 실험(c)에서와 같은 방법으로 진행한다. 그리고 1차 예측된 결과를 기반으로 가중치를 적용해야하기 때문에 결측 전·후 데이터를 조회한다. 식(5-6)처럼 결측 후 데이터(R_a)에서 결측 전 데이터(R_b)를 빼면 결측 구간의 총 사용량이 된다. 여기에 예측된 구간 전력사용량의 각 시간별 비율을 곱하면 가중치가 적용된 2차 예측결과가 된다.

$$S_w(x) = (R_a - R_b) \times \frac{S_x}{\sum_{x=1}^n S_x} \quad (5-6)$$

그리고 식(5-7)처럼 최종 예측된 시간별 구간 전력사용량($S_w n$)을 기반으로 결측 전 마지막 데이터(R_b)에 예측된 같은 시간 구간 전력사용량($S_w 1$)을 더하면 첫 번째 결측 시간의 누적 전력사용량 데이터(M_1)가 된다. 이렇게 결측 개수만큼 반복 수행하여 그 결과를 DB에 저장한다.

$$\begin{aligned} M_1 &= R_b + S_w 1 \\ M_2 &= M_1 + S_w 2 \\ &\vdots \\ M_n &= M_{n-1} + S_w n \end{aligned} \quad (5-7)$$

실험대상 1000개 계량기를 반복수행하여 각 계량기별 결측 구간의 예측 결괏값을 저장한다.



[그림 5-5] 유사도 가중치적용 보정법 실험 절차

e. CNN-LSTM결합 예측 보정법

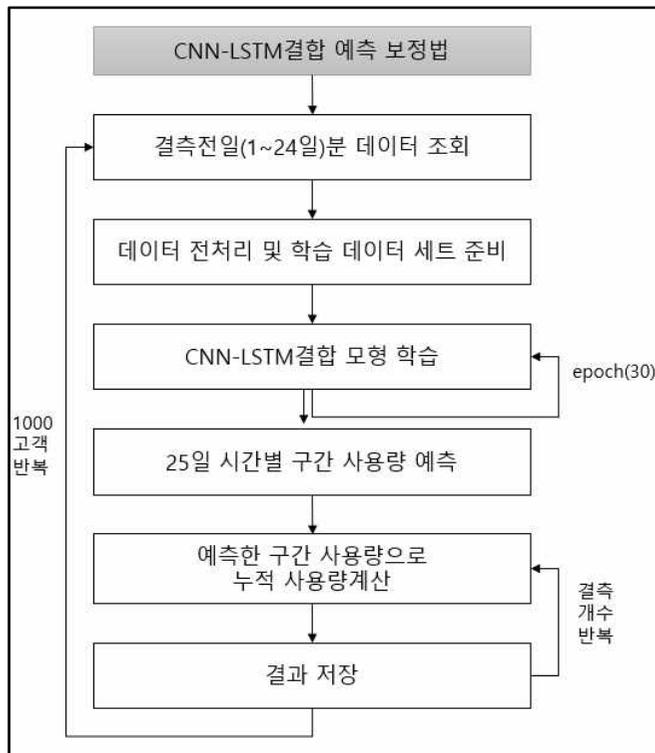
선행 연구에서는 장단기메모리순환신경망(LSTM; Long Short-Term Memory) 예측 보정법을 사용했으나 본 논문에서는 선행 딥러닝 알고리즘보다 개선시킨 CNN-LSTM결합한 모델을 사용해서 실험하고자 한다. 딥러닝 모델의 학습 파라미터인 loss function, optimizer, epoch개수는 데이터 분석 및 모델 개발 시에 결정하였으므로 결정된 하이퍼파라미터를 기준으로 실험을 진행한다.

[그림 5-6]은 CNN-LSTM결합 모델을 만들어서 결측 구간의 누적 전력사용량을 예측하는 실험 절차이다. 먼저 결측 직전 7월 1일 0시부터 7월 24일 23시까지의 데이터를 조회한다. 그리고 전처리를 통해 구간 전력사용량을 계산하여 딥러닝 모델에 입력하여 학습시킨다. 학습된 모델을 기반으로 25일 0시부터 23시까지의 시간별

구간 전력사용량을 예측한다. 그리고 식(5-8)처럼 시간별 구간 전력사용량을 기반으로 결측 전 마지막 데이터(R_b)에 예측한 같은 시간 구간 전력사용량(D_1)을 더하면 첫 번째 결측 시간의 누적 전력사용량 데이터(M_1)가 된다. 이렇게 결측 개수만큼 반복 수행하여 그 결과를 DB에 저장한다.

$$\begin{aligned}
 M_1 &= R_b + D_1 \\
 M_2 &= M_1 + D_2 \\
 &\vdots \\
 M_n &= M_{n-1} + D_n
 \end{aligned}
 \tag{5-8}$$

실험대상 1000개 계량기를 반복수행하여 각 계량기별 결측 구간의 예측 결과값을 저장한다.



[그림 5-6] CNN-LSTM결합 예측 보정법 실험 절차

f. CNN-LSTM결합 가중치적용 보정법

[그림 5-7]은 본 논문에서 제안하는 CNN-LSTM결합 가중치적용 보정법의 실험 절차이다. 이 실험은 크게 딥러닝 모델로 1차 예측하고 가중치를 적용한 2차 예측, 그리고 2차 예측한 시간별 구간 전력사용량 기반으로 누적 전력사용량 계산한다.

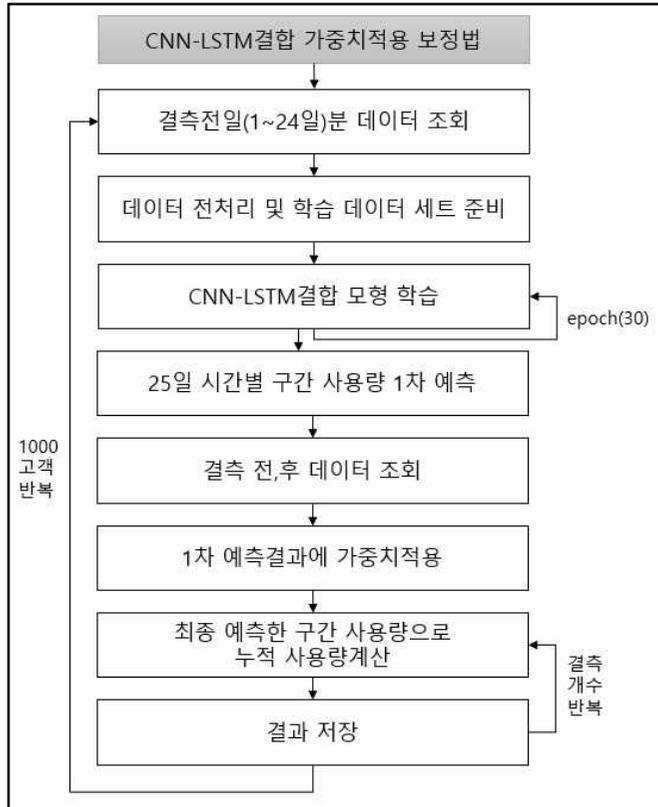
결측 구간의 시간별 구간 전력사용량 예측은 실험(e)에서와 같은 방법으로 진행한다. 그리고 1차 예측된 결과를 기반으로 가중치를 적용해야하기 때문에 결측 전·후 데이터를 조회한다. 식(5-9)처럼 결측 후 데이터(R_a)에서 결측 전 데이터(R_b)를 빼면 결측 구간의 총 사용량이 된다. 여기에 예측된 구간 전력사용량의 각 시간별 비율을 곱하면 가중치가 적용된 2차 예측결과가 된다.

$$D_w(x) = (R_a - R_b) \times \frac{D_x}{\sum_{x=1}^n D_x} \quad (5-9)$$

그리고 식(5-10)처럼 최종 예측된 시간별 구간 전력사용량($D_w n$)을 기반으로 결측 전 마지막 데이터(R_b)에 예측된 같은 시간 구간 전력사용량($D_w 1$)을 더하면 첫 번째 결측 시간의 누적 전력사용량 데이터(M_1)가 된다. 이렇게 결측 개수만큼 반복 수행하여 그 결과를 DB에 저장한다.

$$\begin{aligned} M_1 &= R_b + D_w 1 \\ M_2 &= M_1 + D_w 2 \\ &\vdots \\ M_n &= M_{n-1} + D_w n \end{aligned} \quad (5-10)$$

실험대상 1000개 계량기를 반복수행하여 각 계량기별 결측 구간의 예측 결과값을 저장한다.



[그림 5-7] CNN-LSTM결합 가중치적용 보정법 실험 절차

4. 실험 결과

a. 선형보정법의 실험 결과

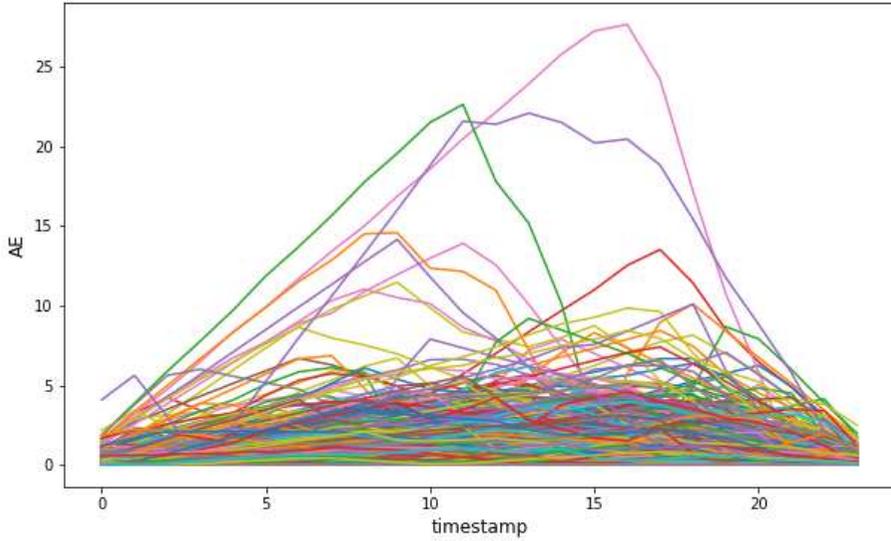
[그림 5-8]은 1000개 계량기에 대한 선형보정법을 실험한 후 그 결과를 데이터 프레임 형태로 표시하였다. 해당 결과는 7월 25일 0시부터 23시까지의 실제 누적 전력사용량과 선형보정법으로 예측한 누적 전력사용량을 표시하였다.

	Meter_id	Hour	Real	Linear
0	M0001	0000	555.578	555.6709
1	M0002	0000	4791.182	4791.1090
2	M0003	0000	4450.127	4450.2320
3	M0004	0000	2250.931	2250.5870
4	M0005	0000	3878.120	3878.0980
...
23995	M0996	2300	37736.230	37736.1100
23996	M0997	2300	1191.424	1191.4240
23997	M0998	2300	9.591	9.5910
23998	M0999	2300	662.786	662.9062
23999	M1000	2300	1799.463	1799.5480

24000 rows × 4 columns

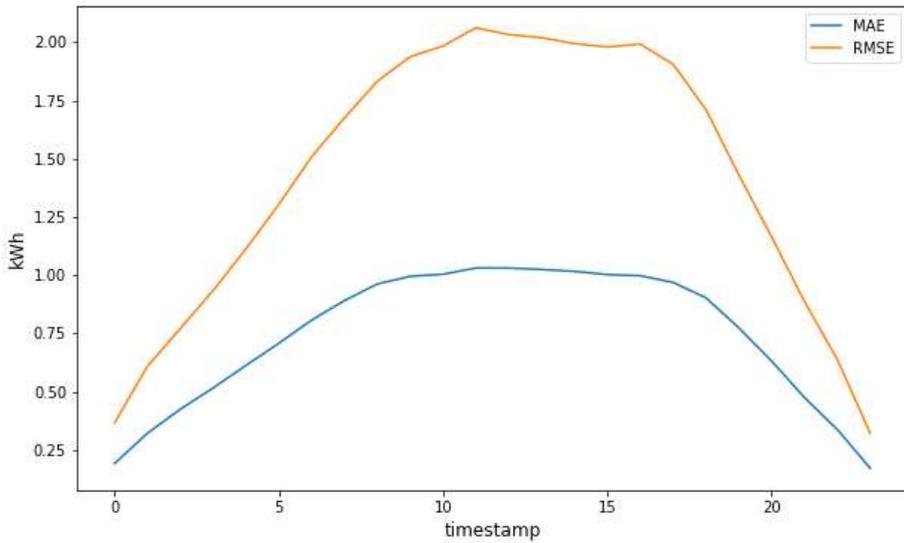
[그림 5-8] 선형보정법 실험 결과 데이터

[그림 5-9]는 선형보정법으로 예측한 결과와 실제 데이터간의 절대 오차(AE; Absolute Error)를 1000개 계량기 전체를 표시하였다. X축은 결측 구간의 시간축(0시~23시)이고, Y축은 절대 오차값이다.



[그림 5-9] 선형보정법 계량기별 시간별 AE(Absolute Error)

[그림 5-10]은 1000개 계량기에 대한 시간별 MAE와 RMSE를 표시하였다. MAE는 특이치에 덜 민감한 반면, RMSE는 특이치에 민감하게 표현됐다. 선형보정법은 전체 구간을 평균값으로 일정하게 적용하여 예측하였기 때문에 결측 중간부분에서 오차가 크게 나타남을 알 수 있다.



[그림 5-10] 선형보정법 시간별 MAE, RMSE

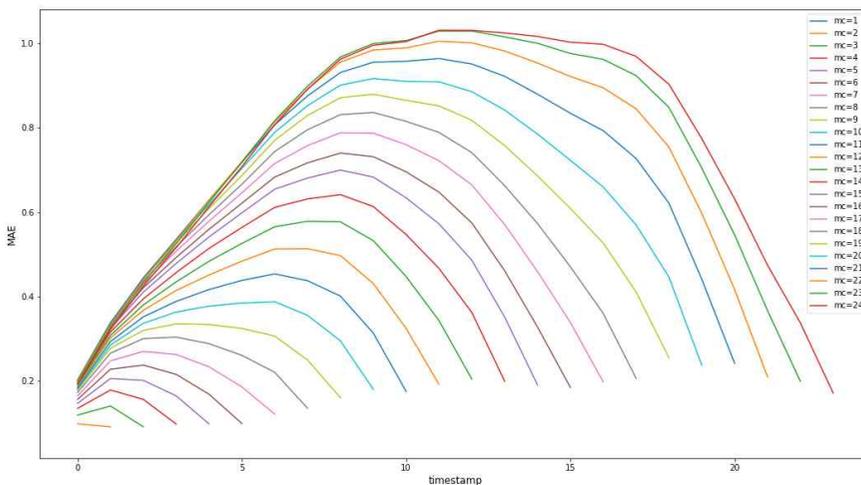
[그림 5-11]은 선형보정법의 시간별 결측 길이별 전체 계량기의 MAE를 계산한 결과 데이터이다.

Hour	Missing_Count	MAE
0	0	1 0.064996
1	0	2 0.098327
2	0	3 0.118986
3	0	4 0.134950
4	0	5 0.146835
...
295	21	23 0.366811
296	21	24 0.474475
297	22	23 0.198961
298	22	24 0.337511
299	23	24 0.171122

300 rows × 3 columns

[그림 5-11] 선형보정법 시간별, 결측 개수별 MAE 데이터

[그림 5-12]는 선형보정법의 시간별, 결측 길이별로 전체 계량기의 MAE를 계산하여 그래프로 표현했다. 결측 길이가 증가 할수록 오차가 커지는 것을 알 수 있다.



[그림 5-12] 선형보정법 시간별, 결측 개수별 MAE

b. ARIMA 예측 보정법의 실험 결과

선행 연구된 arima 알고리즘을 이용하여 예측하는 방법은 arima 알고리즘의 p, d, q값을 특정 계량기의 데이터로만 수작업 검토하여 결정하였고, 결정된 p, d, q값을 모든 계량기의 결측 데이터를 예측하는데 똑같이 적용하였다. 하지만 본 논문에서는 auto arima 알고리즘을 사용하여 자동으로 p, d, q값을 결정하고 예측하는 방법을 사용하여 정확도를 높였다. [그림 5-13]은 auto arima를 이용하여 실험을 수행하고 있는 과정을 표현하였다. 계량기별 데이터를 학습하여 AIC가 최소가 되는 Best model을 선택하고 이 모델을 이용해서 결측 구간을 예측하고 있다.

```

Target Count => 1000

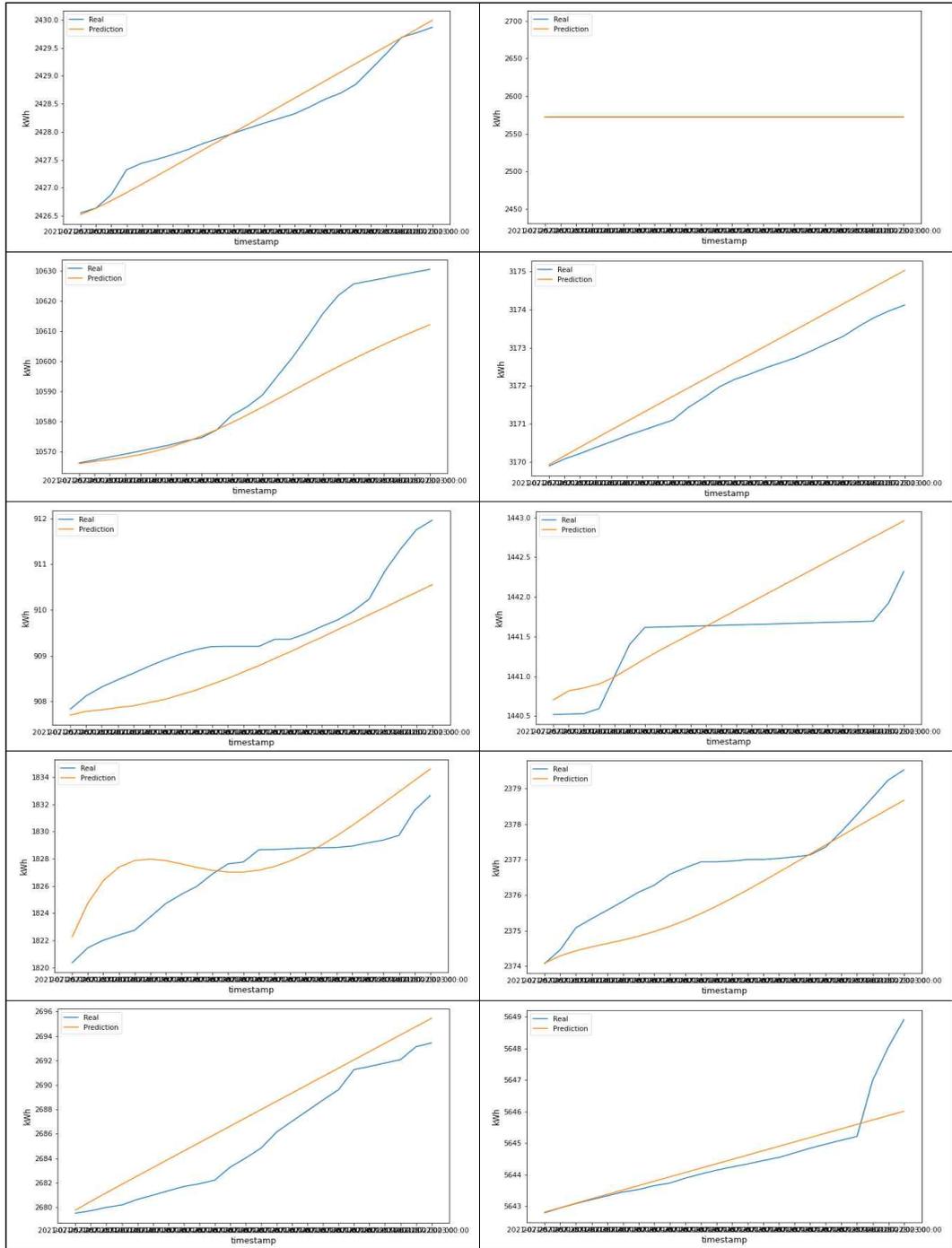
Start Meter => M0001
Performing stepwise search to minimize aic
ARIMA(2,1,2)(0,0,0)[0] intercept : AIC=-1538.715, Time=0.58 sec
ARIMA(0,1,0)(0,0,0)[0] intercept : AIC=-1301.744, Time=0.08 sec
ARIMA(1,1,0)(0,0,0)[0] intercept : AIC=-1539.291, Time=0.05 sec
ARIMA(0,1,1)(0,0,0)[0] intercept : AIC=-1467.861, Time=0.18 sec
ARIMA(0,1,0)(0,0,0)[0] : AIC=539.566, Time=0.03 sec
ARIMA(2,1,0)(0,0,0)[0] intercept : AIC=-1540.933, Time=0.08 sec
ARIMA(3,1,0)(0,0,0)[0] intercept : AIC=-1539.280, Time=0.15 sec
ARIMA(2,1,1)(0,0,0)[0] intercept : AIC=-1540.810, Time=0.52 sec
ARIMA(1,1,1)(0,0,0)[0] intercept : AIC=-1541.294, Time=0.31 sec
ARIMA(1,1,2)(0,0,0)[0] intercept : AIC=-1540.147, Time=0.40 sec
ARIMA(0,1,2)(0,0,0)[0] intercept : AIC=-1513.328, Time=0.22 sec
ARIMA(1,1,1)(0,0,0)[0] : AIC=-1475.593, Time=0.13 sec
Best model: ARIMA(1,1,1)(0,0,0)[0] intercept
Total fit time: 2.728 seconds
Complete => M0001

Start Meter => M0002
Performing stepwise search to minimize aic
ARIMA(2,2,2)(0,0,0)[0] : AIC=inf, Time=0.53 sec
ARIMA(0,2,0)(0,0,0)[0] : AIC=-310.168, Time=0.02 sec
ARIMA(1,2,0)(0,0,0)[0] : AIC=-317.592, Time=0.03 sec
ARIMA(0,2,1)(0,0,0)[0] : AIC=-327.948, Time=0.05 sec
ARIMA(1,2,1)(0,0,0)[0] : AIC=inf, Time=0.24 sec
ARIMA(0,2,2)(0,0,0)[0] : AIC=inf, Time=0.16 sec
ARIMA(1,2,2)(0,0,0)[0] : AIC=inf, Time=0.25 sec
ARIMA(0,2,1)(0,0,0)[0] intercept : AIC=-325.960, Time=0.08 sec
Best model: ARIMA(0,2,1)(0,0,0)[0]
Total fit time: 1.335 seconds
Complete => M0002
  
```

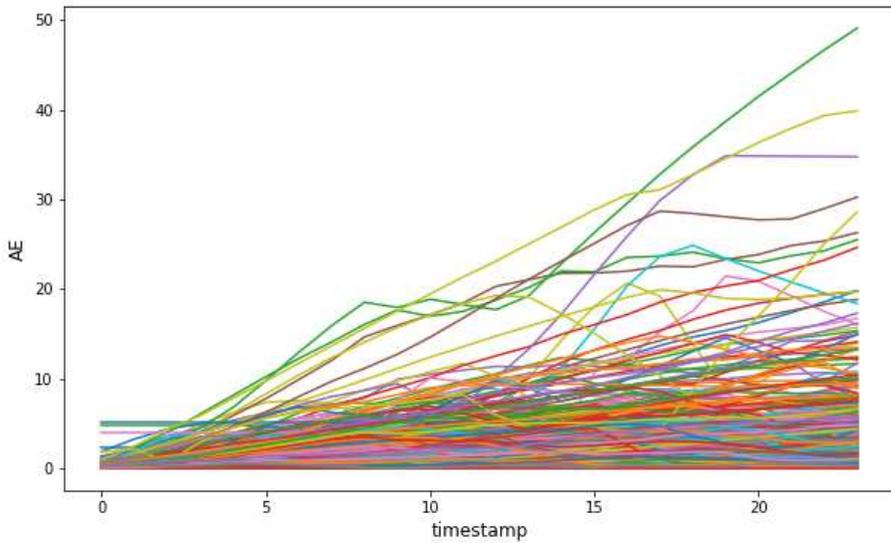
[그림 5-13] ARIMA 모델 학습과정

[표 5-2]는 계량기별 ARIMA 모델을 이용해서 최종 누적 전력사용량을 예측하고 실제 데이터와 비교한 결과이다. ARIMA 모델은 구간 전력사용량을 예측하는 것이 아니고 최종 누적 전력사용량을 한 번에 예측하기 때문에 추가 작업이 필요 없다. 표에서 보듯이 전력사용량이 변동이 없는 계량기에 대해서도 잘 예측이 된 것을 알 수 있고, 전반적으로 예측은 큰 변동성을 있는 경우에는 반영이 잘 안 되는 것을 알 수 있다. 아래 표는 샘플로 10개만 표시하였다.

[표 5-2] 계량기별 ARIMA 모델의 예측 결과

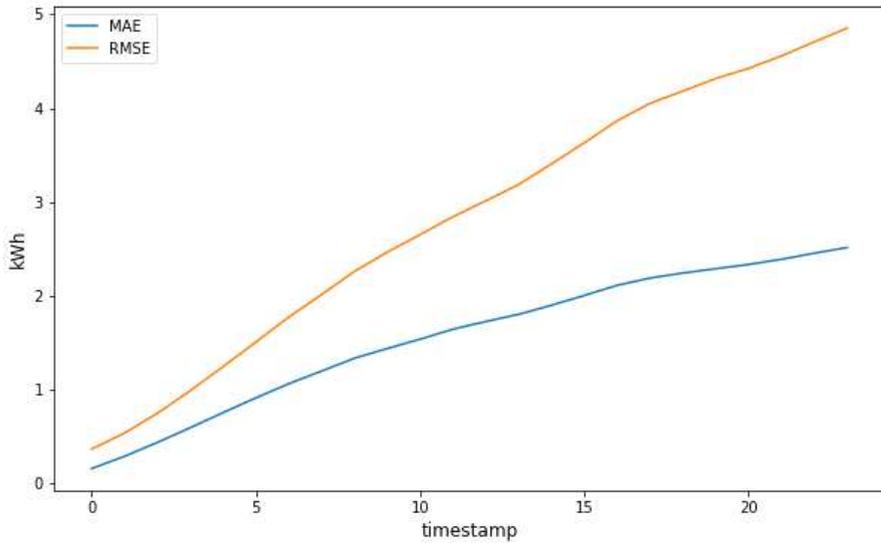


[그림 5-14]는 ARIMA 예측 보정법으로 예측한 결과와 실제 데이터간의 절대 오차(AE; Absolute Error)를 1000개 계량기 전체를 표시하였다. X축은 결측 구간의 시간축(0시~23시)이고, Y축은 절대 오차값이다. 아래 그림과 같이 시간에 따라 오차가 계속 증가함을 알 수 있다.



[그림 5-14] ARIMA 예측 보정법 계량기별 시간별 AE(Absolute Error)

[그림 5-15]는 1000개 계량기에 대한 시간별 MAE와 RMSE를 표시하였다. ARIMA 예측 보정법은 ARIMA 알고리즘을 이용하여 데이터를 전처리 즉 차분을 통한 구간 전력사용량을 계산하지 않고, 실제 누적 전력사용량을 가공 없이 입력으로 사용하기 때문에 사용이 편리하고 빠르게 계산결과가 나온다. 하지만 예측 결과 오차가 상당히 발생하였다.

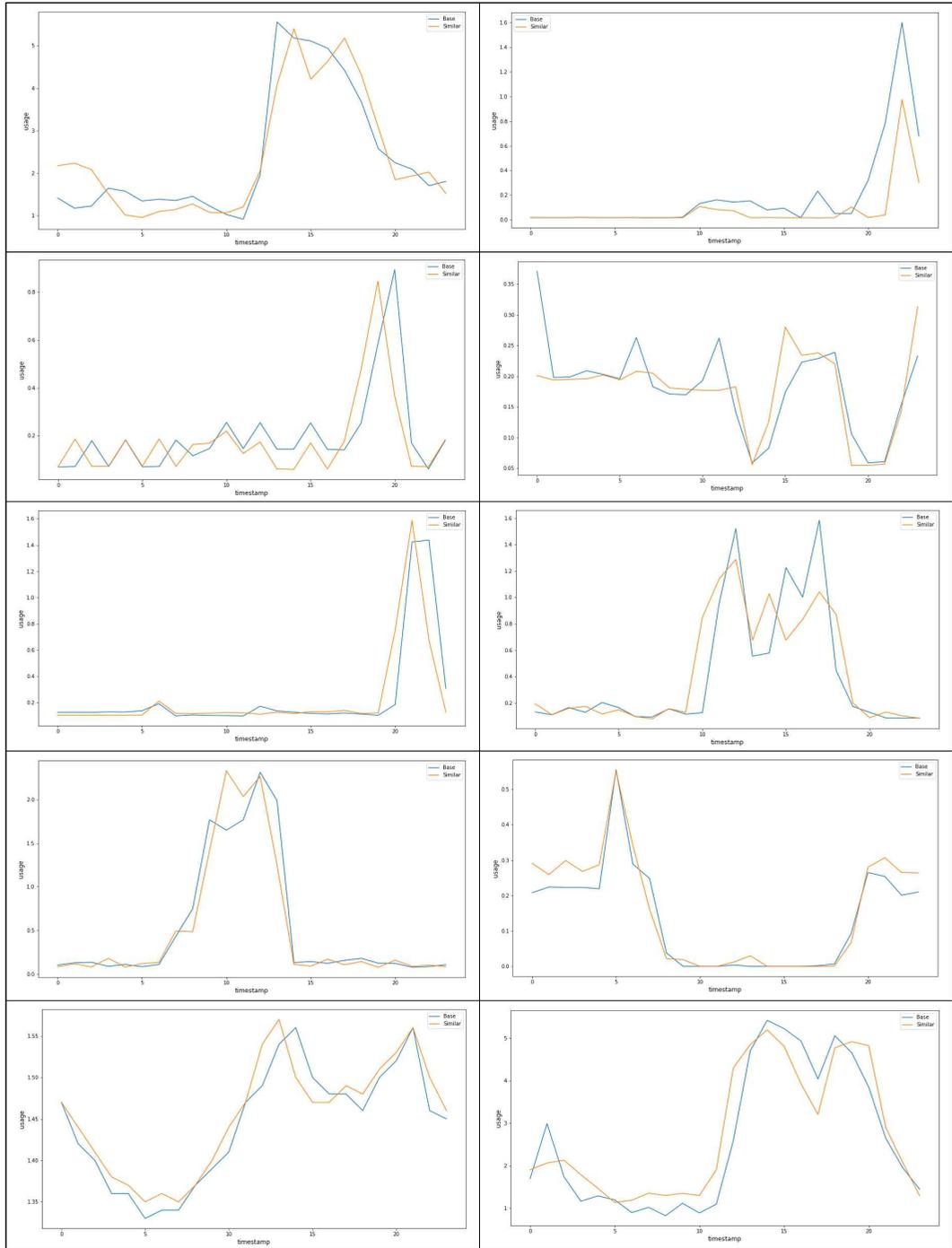


[그림 5-15] ARIMA 예측 보정법 시간별 MAE, RMSE

c. 유사도 기반 보정법의 실험 결과

아래 [표 5-3]은 계량기별로 결측 이전 날짜인 7월 24일 0시부터 23시까지를 기준으로 가장 유사한 전력사용량 패턴을 탐색한 결과이다. 7월 24일을 기준으로 7월 23일부터 7월 17일까지 같은 시간대별 구간 전력사용량을 비교하여 유클리드 거리 유사도 기법으로 가장 유사한 날짜를 탐색하였고, 그 결과 아래 [표 5-2]와 같이 사용량이 유사한 패턴을 보였다. 아래 표는 샘플로 10개만 표시하였고, 추가로 더 많은 자료를 [부록 1]에 표시하였다.

[표 5-3] 계량기별 결측 이전일자와 가장 유사한 전력사용량 패턴 탐색



7월 24일과 가장 유사한 날짜를 탐색한 후 결측된 날짜와 유사할 것으로 판단되는 유사한 날짜의 다음날을 참조일로 선정한다. 예로 7월 20일이 7월 24일과 가장 유사하다면 결측 보정을 위한 참조일은 7월 21일이 된다.

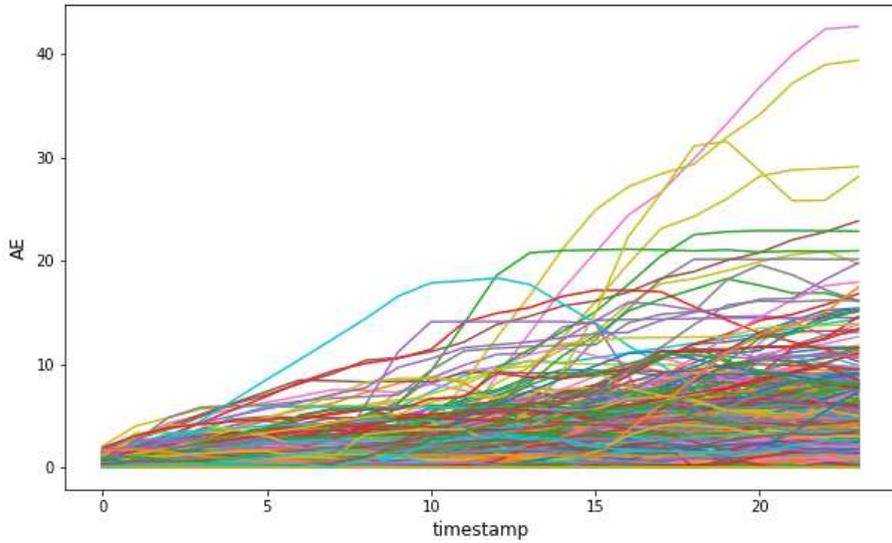
[그림 5-16]은 유사도 기반으로 기준일 같은 시각의 구간 전력사용량을 이용하여 결측 구간의 0시부터 23시까지를 예측하였고, 그 실험 결과 데이터이다.

	Meter_id	Hour	Real	Similar
0	M0001	0000	555.578	556.932
1	M0002	0000	4791.182	4791.140
2	M0003	0000	4450.127	4450.136
3	M0004	0000	2250.931	2250.416
4	M0005	0000	3878.120	3878.070
...
23995	M0996	2300	37736.230	37739.740
23996	M0997	2300	1191.424	1191.424
23997	M0998	2300	9.591	9.591
23998	M0999	2300	662.786	662.791
23999	M1000	2300	1799.463	1802.214

24000 rows × 4 columns

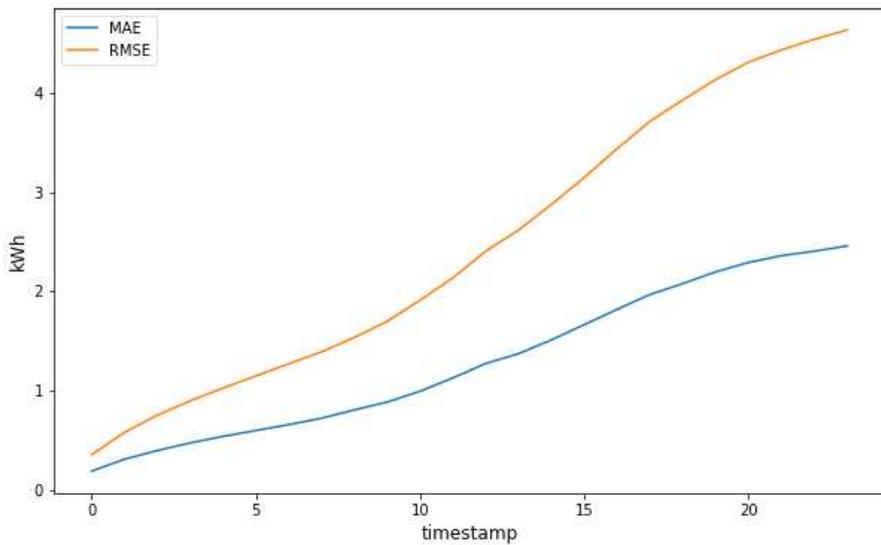
[그림 5-16] 유사도 기반 보정법 실험 결과 데이터

[그림 5-17]은 유사도 기반 보정법으로 예측한 결과와 실제 데이터간의 절대 오차(AE; Absolute Error)를 1000개 계량기 전체를 표시하였다. X축은 결측 구간의 시간축(0시~23시)이고, Y축은 절대 오차값이다. 그림처럼 시간에 따라 오차가 계속 증가함을 알 수 있다.



[그림 5-17] 유사도 기반 보정법 계량기별 시간별 AE(Absolute Error)

[그림 5-18]은 1000개 계량기에 대한 시간별 MAE와 RMSE를 표시하였다. 유사도 기반 보정법은 유사일 기준으로 구간 전력사용량을 결측 구간에 누적 합산하였기 때문에 결측 구간이 길어질수록 오차가 커짐을 알 수 있다.



[그림 5-18] 유사도 기반 보정법 시간별 MAE, RMSE

d. 유사도 가중치적용 보정법의 실험 결과

[그림 5-19]는 위 실험에서 수행한 내용과 같은 유사도 기반 보정법으로 결측일의 시간별 구간 전력사용량을 1차 예측하였고, 1차 예측결과 컬럼은 Similar-Predict 컬럼이다. 그 다음 결측 전체 구간의 합을 1로 하는 각 시간별 구간 전력사용량을 비율로 다시 계산하였고, Similar-TermRate 컬럼이다. 그리고 결측 전체 구간 전력사용량인 결측 전·후 시간의 누적 전력사용량 값의 차를 계산하여 각 시간별 구간 전력사용량 비율을 곱하면 가중치를 적용한 2차 예측사용량이 된다. 해당 컬럼은 Similar-Weight 이다.

	Meter_id	Hour	Similar_Predict	Similar_TermRate	Similar_Weight
0	M0001	0000	1.544	0.173835	1.132186
1	M0001	0100	0.765	0.086129	0.560960
2	M0001	0200	1.035	0.116528	0.758946
3	M0001	0300	0.873	0.098289	0.640154
4	M0001	0400	0.406	0.045710	0.297712
...
23995	M1000	1900	0.180	0.016064	0.135807
23996	M1000	2000	0.595	0.053101	0.448918
23997	M1000	2100	1.121	0.100045	0.845777
23998	M1000	2200	0.841	0.075056	0.634522
23999	M1000	2300	0.673	0.060062	0.507768

24000 rows × 5 columns

[그림 5-19] 유사도 가중치적용 보정법 1차 예측결과 데이터

[그림 5-20]은 유사도 기반 보정법에 가중치를 적용한 2차 예측결과에 결측 이전 마지막 누적 전력사용량 데이터에 결측 각 시간대 구간 전력사용량 합산하여 최종 예측한 누적 전력사용량 결과 데이터와 실제 데이터를 표현했다.

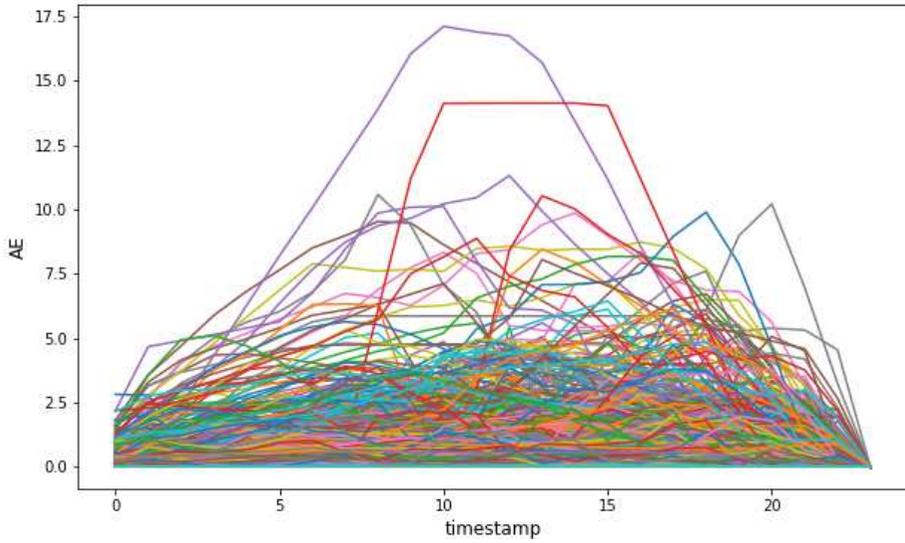
	Meter_id	Hour	Real	Similar_Weight
0	M0001	0000	555.578	556.5202
1	M0002	0000	4791.182	4791.0777
2	M0003	0000	4450.127	4450.1243
3	M0004	0000	2250.931	2250.3725
4	M0005	0000	3878.120	3878.1144
...
23995	M0996	2300	37736.230	37736.2302
23996	M0997	2300	1191.424	1191.4240
23997	M0998	2300	9.591	9.5910
23998	M0999	2300	662.786	662.7863
23999	M1000	2300	1799.463	1799.4629

24000 rows × 4 columns

[그림 5-20] 유사도 가중치적용 보정법 최종결과 데이터

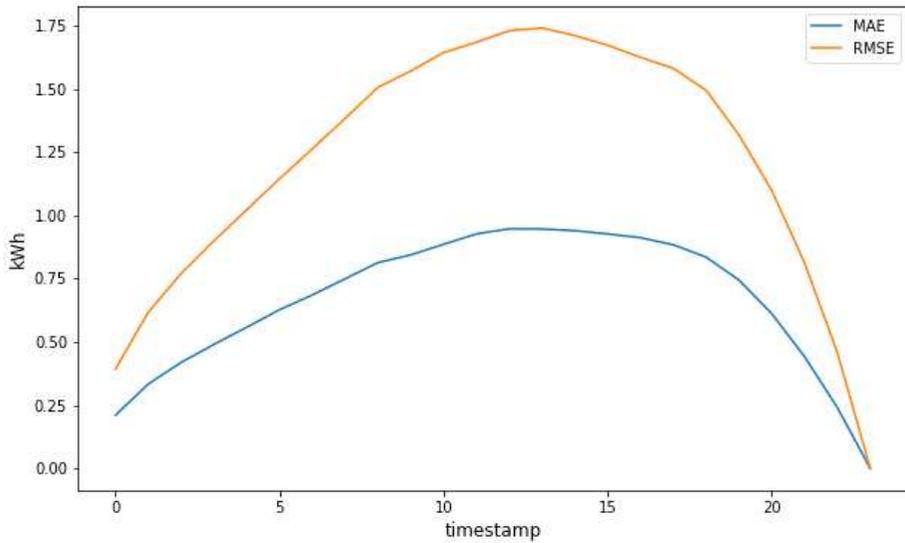
유사도 기반 보정법에 가중치를 적용한 방법으로 기존 유사도 기반 보정법에서는 결측 구간 개수가 증가 할수록 오차도 증가하는 형태를 보였으나 결측 구간 예측 총 합계를 1로하는 비율을 적용함으로써 결측 전·후 총 누적 전력사용량 범위를 벗어나지 않도록 하여 안정성이 우수하다. MAE, RMSE 그래프로 볼때도 결측 양 끝부분으로 갈수록 오차가 줄어든다.

[그림 5-21]은 유사도 기반 보정법에 가중치를 적용한 1000개 계량기의 시간별 절대오차를 그래프로 나타냈다. 유사도 기반 보정법도다 오차가 현저히 감소하고 결측 양 끝부분으로 갈수록 오차가 '0'에 수렴한다.



[그림 5-21] 유사도 가중치적용 보정법 계량기별 시간별 AE(Absolute Error)

[그림 5-22]는 1000개 계량기에 대한 시간별 MAE와 RMSE를 표시하였다. 유사도 가중치적용 보정법은 유사도 기반 보정법과 다르게 결측 구간의 양 끝부분이 '0'에 수렴하는 것을 알 수 있다.



[그림 5-22] 유사도 가중치적용 보정법 시간별 MAE, RMSE

e. CNN-LSTM결합 예측 보정법의 실험 결과

[그림 5-23]은 첫 번째 계량기의 누적 전력사용량 데이터를 전처리하여 시간별 구간 전력사용량으로 계산한 후 데이터이며, 이 데이터를 CNN-LSTM결합 딥러닝 모델에 입력하여 결측 구간인 7월 25일 0시부터 23시까지의 시간별 구간 전력사용량을 예측한다.

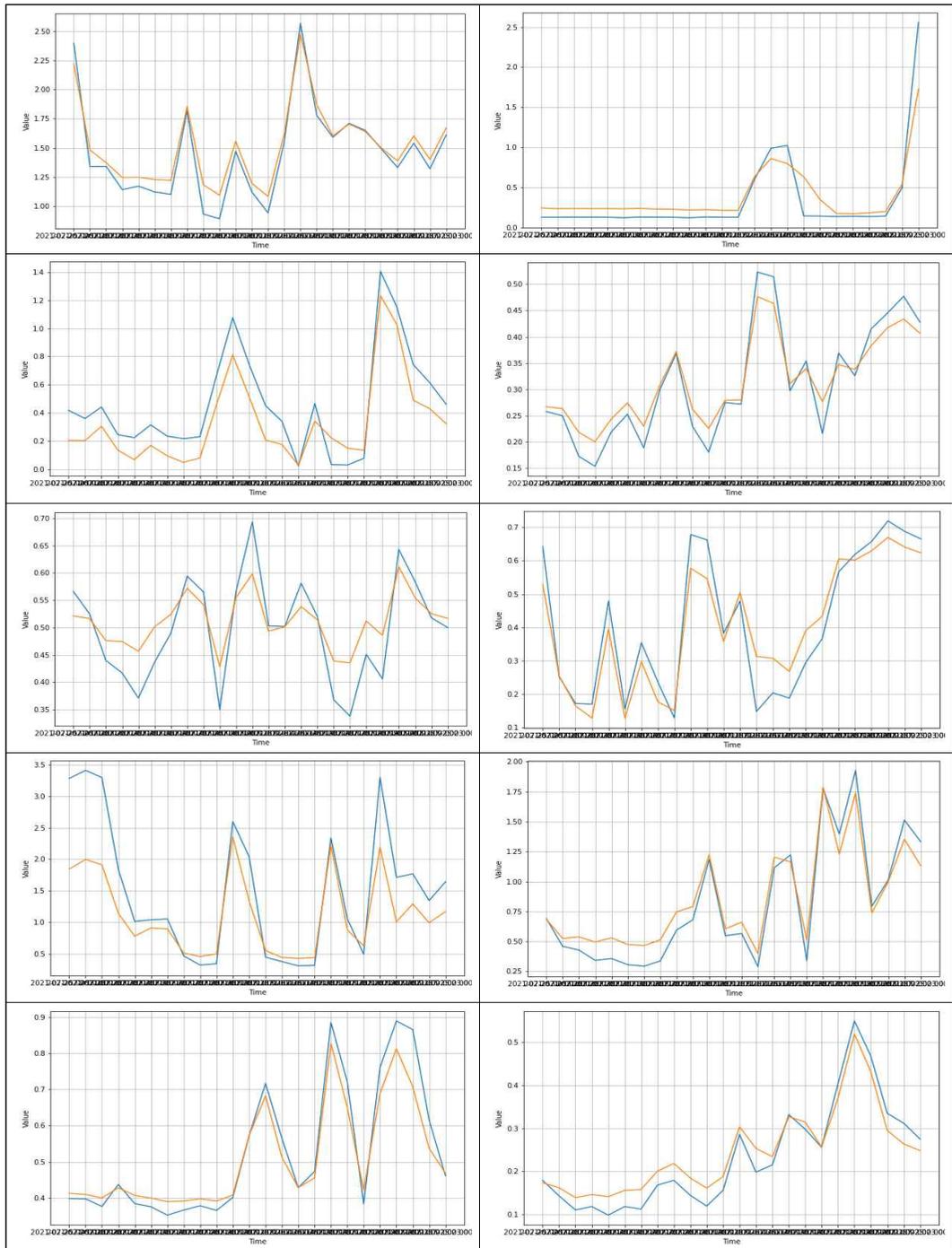
	Time	Value
0	2021-07-01 01:00:00	1.12
1	2021-07-01 02:00:00	1.03
2	2021-07-01 03:00:00	0.94
3	2021-07-01 04:00:00	1.27
4	2021-07-01 05:00:00	1.00
...
594	2021-07-25 19:00:00	1.49
595	2021-07-25 20:00:00	1.33
596	2021-07-25 21:00:00	1.54
597	2021-07-25 22:00:00	1.32
598	2021-07-25 23:00:00	1.61

599 rows × 2 columns

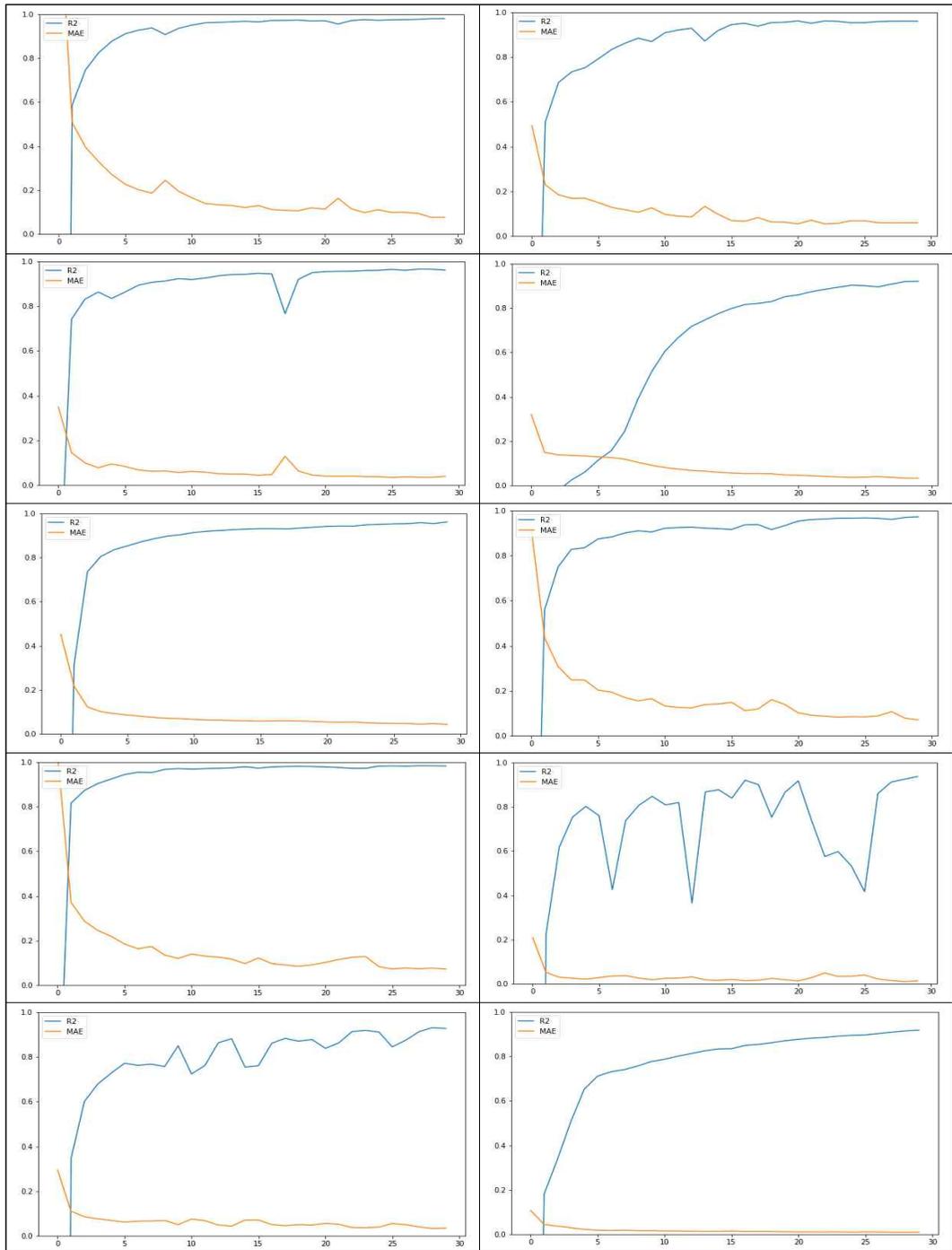
[그림 5-23] 특정 계량기의 전처리 후 결과 데이터

[표 5-4]는 각 계량기별로 7월 1일 0시부터 7월 24일 23시 데이터를 전처리 후 CNN-LSTM결합 모델을 학습 및 테스트를 진행하여 결측 당일 7월 25일 0시부터 23시까지 24개의 시간별 구간 전력사용량을 예측하였다. 시계열 분야에서 많이 사용하고 있는 LSTM을 CNN과 결합하였으며 그 결과가 만족스럽게 나왔다. 아래 표는 샘플로 10개만 표시하였고, 추가로 더 많은 자료를 [부록 2]에 표시하였다.

[표 5-4] 계량기별 CNN-LSTM결합 모델의 예측결과(시간별 구간 전력사용량)



[표 5-5] 계량기별 CNN-LSTM결합 모델의 학습과정(MAE, R-Square)



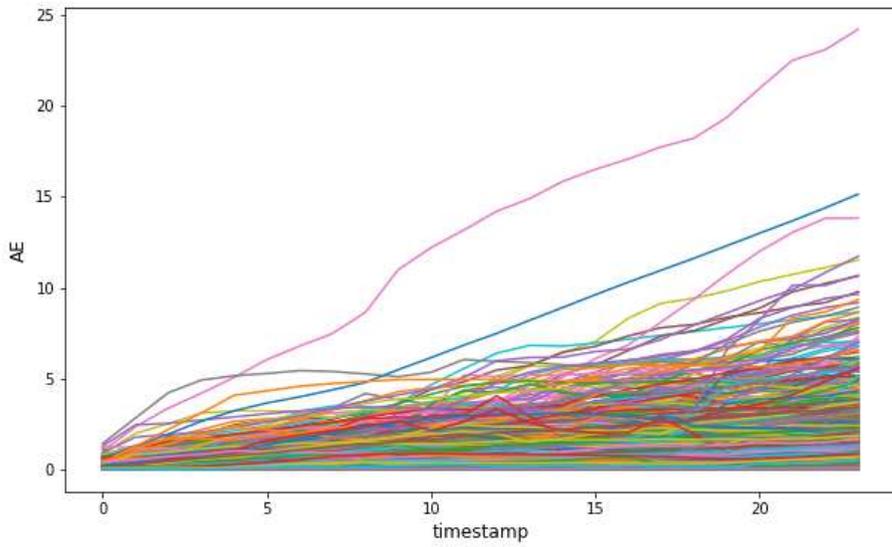
[그림 5-25]는 CNN-LSTM결합 모델의 최종 예측 결과 데이터이다.

	Meter_id	Hour	Real	Real_Usage	CNN-LSTM	CNN-LSTM_Usage
0	M0001	0000	555.578	0.190	555.4855	0.0975
1	M0002	0000	4791.182	0.566	4791.1374	0.5214
2	M0003	0000	4450.127	0.059	4450.1789	0.1109
3	M0004	0000	2250.931	0.603	2250.7780	0.4500
4	M0005	0000	3878.120	0.452	3878.0999	0.4319
...
23995	M0996	2300	37736.230	0.670	37739.6852	0.8534
23996	M0997	2300	1191.424	0.010	1191.3558	0.0072
23997	M0998	2300	9.591	0.000	9.5934	0.0001
23998	M0999	2300	662.786	0.313	662.0591	0.2275
23999	M1000	2300	1799.463	0.478	1798.9197	0.3552

24000 rows × 6 columns

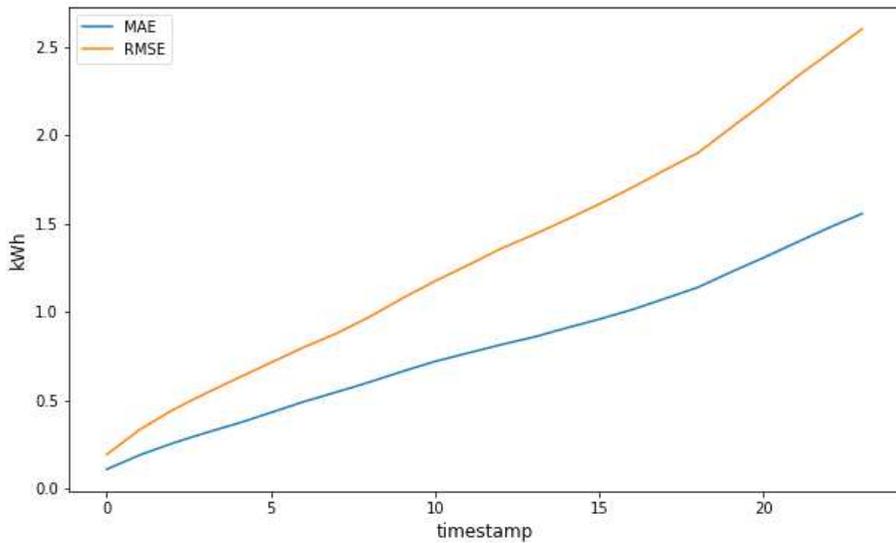
[그림 5-25] CNN-LSTM결합 모델의 예측결과 데이터

[그림 5-26]은 CNN-LSTM결합 모델의 예측 결과를 기반으로 누적 전력사용량을 계산하여 실제 데이터와 절대오차를 비교하였다. 결측 길이가 증가할수록 오차가 계속 증가하는 것을 알 수 있다. 이런 이유로 본 논문에서는 여기에 가중치를 적용하여 개선하고자 제안했다.



[그림 5-26] CNN-LSTM결합 예측 계량기별 시간별 AE(Absolute Error)

[그림 5-27]은 1000개 계량기에 대한 시간별 MAE와 RMSE를 표시하였다. CNN-LSTM 예측 보정법은 구간 전력사용량을 예측하여 결측 구간에 누적 합산하였기 때문에 결측 구간이 길어질수록 오차가 커짐을 알 수 있다.



[그림 5-27] CNN-LSTM결합 예측 보정법 시간별 MAE, RMSE

f. CNN-LSTM결합 가중치적용 보정법의 실험 결과

[그림 5-28]은 위 실험에서 수행한 내용과 같은 CNN-LSTM결합 모델로 결측일의 시간별 구간 전력사용량을 1차 예측하였고, 1차 예측결과 컬럼은 CNN-LSTM-Predict 컬럼이다. 그 다음 결측 전체 구간의 합을 1로 하는 각 시간별 구간 전력사용량을 비율로 다시 계산하였고, CNN-LSTM-TermRate 컬럼이다. 그리고 결측 전체 구간 전력사용량인 결측 전·후 시간의 누적 전력사용량 값의 차를 계산하여 각 시간별 구간 전력사용량 비율을 곱하면 가중치를 적용한 2차 예측사용량이 된다. 해당 컬럼은 CNN-LSTM-Weight 이다.

	Meter_id	Hour	CNN-LSTM-Predict	CNN-LSTM-TermRate	CNN-LSTM-Weight
0	M0001	0000	0.0975	0.0289	0.1884
1	M0002	0000	0.5214	0.0424	0.5059
2	M0003	0000	0.1109	0.0451	0.1778
3	M0004	0000	0.4500	0.0659	0.3794
4	M0005	0000	0.4319	0.0420	0.4345
...
23995	M0996	2300	0.8534	0.0366	0.7268
23996	M0997	2300	0.0072	0.0417	0.0100
23997	M0998	2300	0.0001	0.0417	0.0000
23998	M0999	2300	0.2275	0.0606	0.2715
23999	M1000	2300	0.3552	0.0449	0.3796

24000 rows × 5 columns

[그림 5-28] CNN-LSTM결합 모델의 실험 1차 예측결과 데이터

[그림 5-29]는 CNN-LSTM결합 가중치를 적용한 2차 예측결과에 결측 이전 마지막 누적 전력사용량 데이터에 결측 각 시간대 구간 전력사용량 합산하여 최종 예측한 누적 전력사용량 결과 데이터와 실제 데이터를 표현했다.

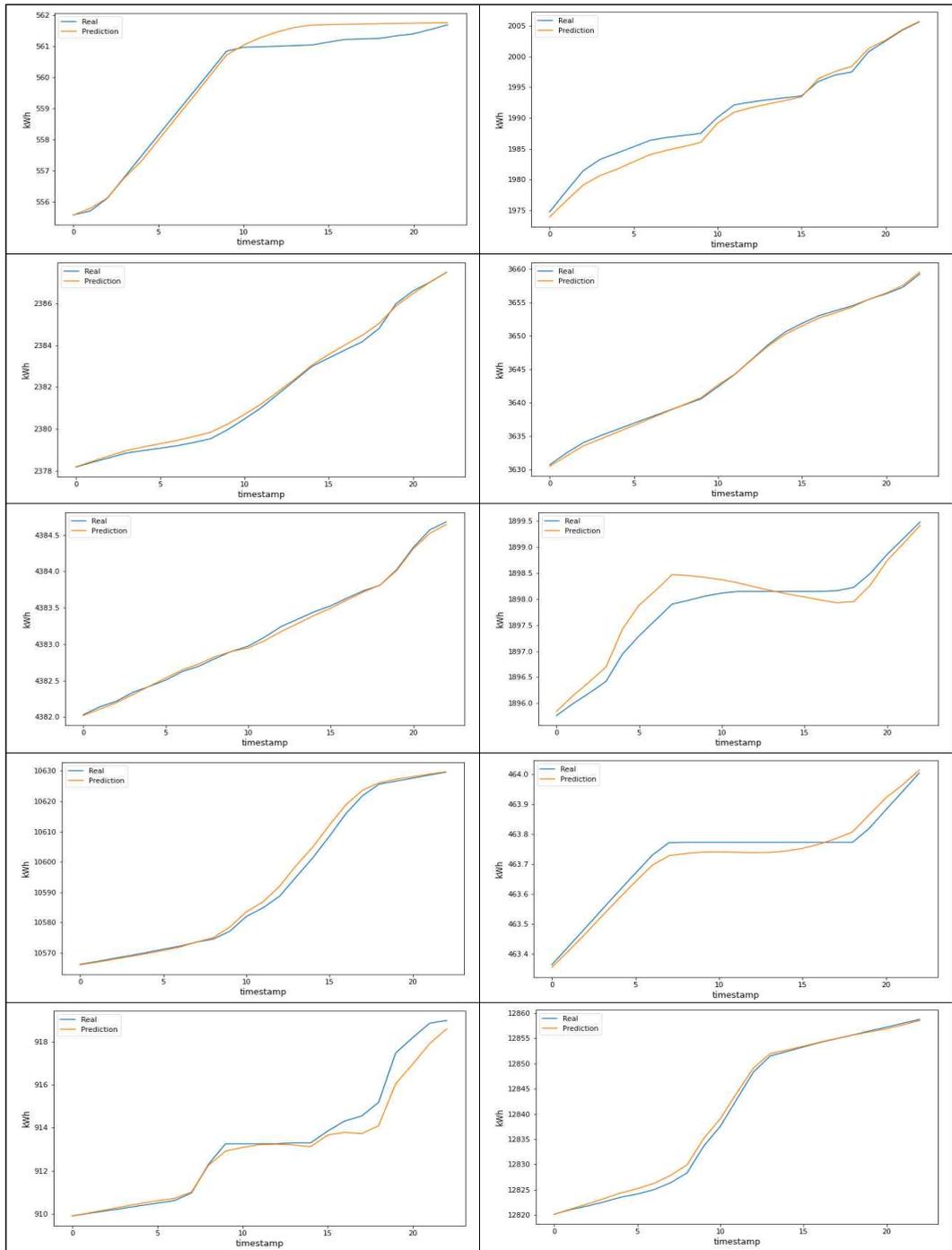
	Meter_id	Hour	Real	CNN-LSTM-Weight
0	M0001	0000	555.578	555.5764
1	M0002	0000	4791.182	4791.1219
2	M0003	0000	4450.127	4450.2458
3	M0004	0000	2250.931	2250.7074
4	M0005	0000	3878.120	3878.1025
...
23995	M0996	2300	37736.230	37736.2300
23996	M0997	2300	1191.424	1191.4230
23997	M0998	2300	9.591	9.5910
23998	M0999	2300	662.786	662.7860
23999	M1000	2300	1799.463	1799.4631

24000 rows × 4 columns

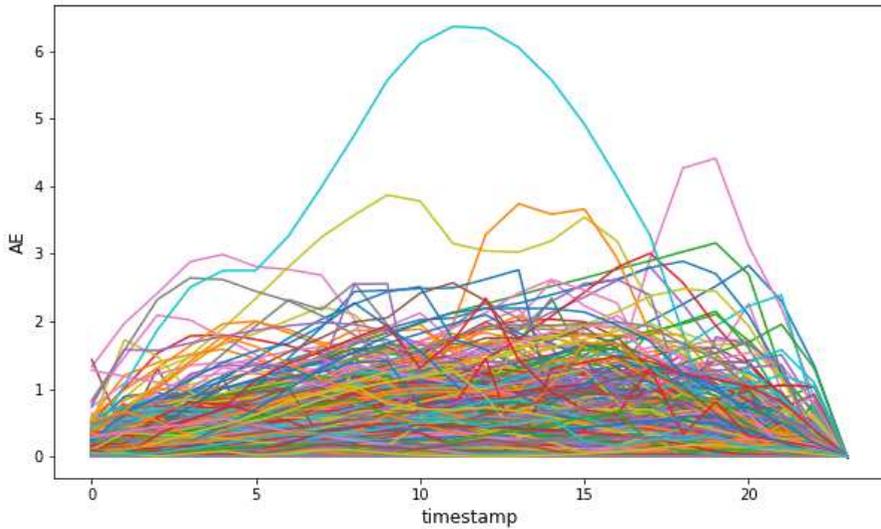
[그림 5-29] CNN-LSTM결합 가중치적용 최종결과 데이터

[표 5-6] 계량기별 CNN-LSTM결합 가중치적용 예측결과와 실제데이터를 비교한 그래프들이다. 실제 데이터 대비 매우 근사하게 잘 맞는 것을 볼 수 있다. 아래 표는 샘플로 10개만 표시하였다.

[표 5-6] 계량기별 CNN-LSTM결합 가중치적용 예측결과와 실제데이터 비교

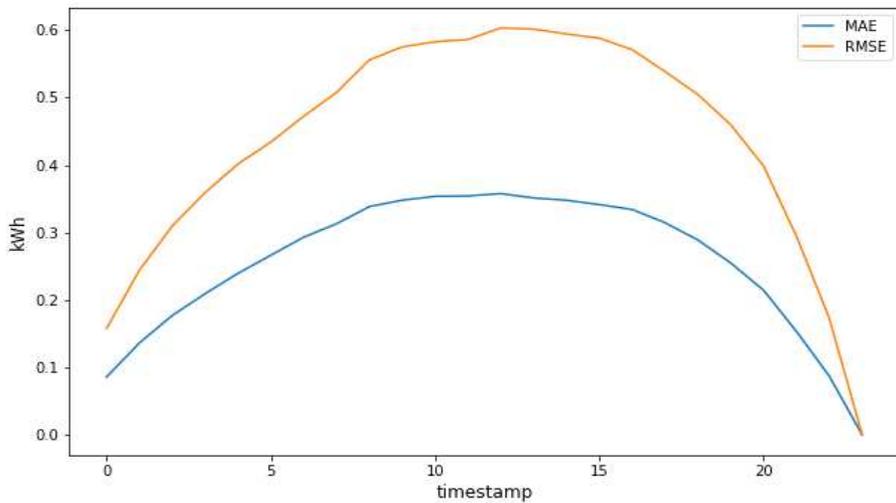


[그림 5-30]은 CNN-LSTM결합 가중치를 적용한 1000개 계량기의 시간별 절대 오차를 그래프로 나타냈다. 오차가 현저히 감소하고 결측 양 끝부분으로 갈수록 오차가 '0'에 수렴한다.



[그림 5-30] CNN-LSTM결합 가중치적용 계량기별 시간별 AE(Absolute Error)

[그림 5-31]은 1000개 계량기에 대한 시간별 MAE와 RMSE를 표시하였다. CNN-LSTM결합 가중치적용 보정법은 CNN-LSTM결합 보정법과 다르게 결측 구간의 양 끝부분이 '0'에 수렴하는 것을 알 수 있다.

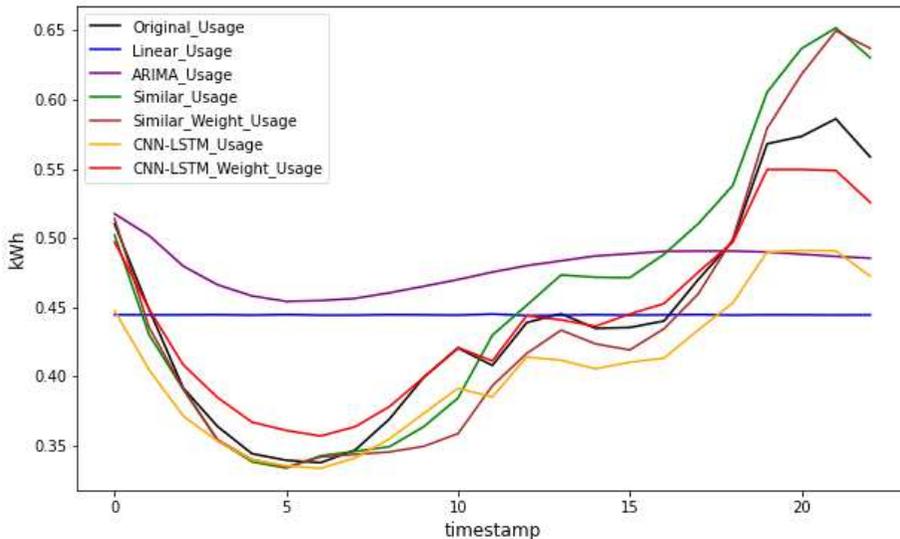


[그림 5-31] CNN-LSTM결합 가중치적용 보정법 시간별 MAE, RMSE

B. 평가 분석 결과

1. 시간별 구간 전력사용량 비교 분석 결과

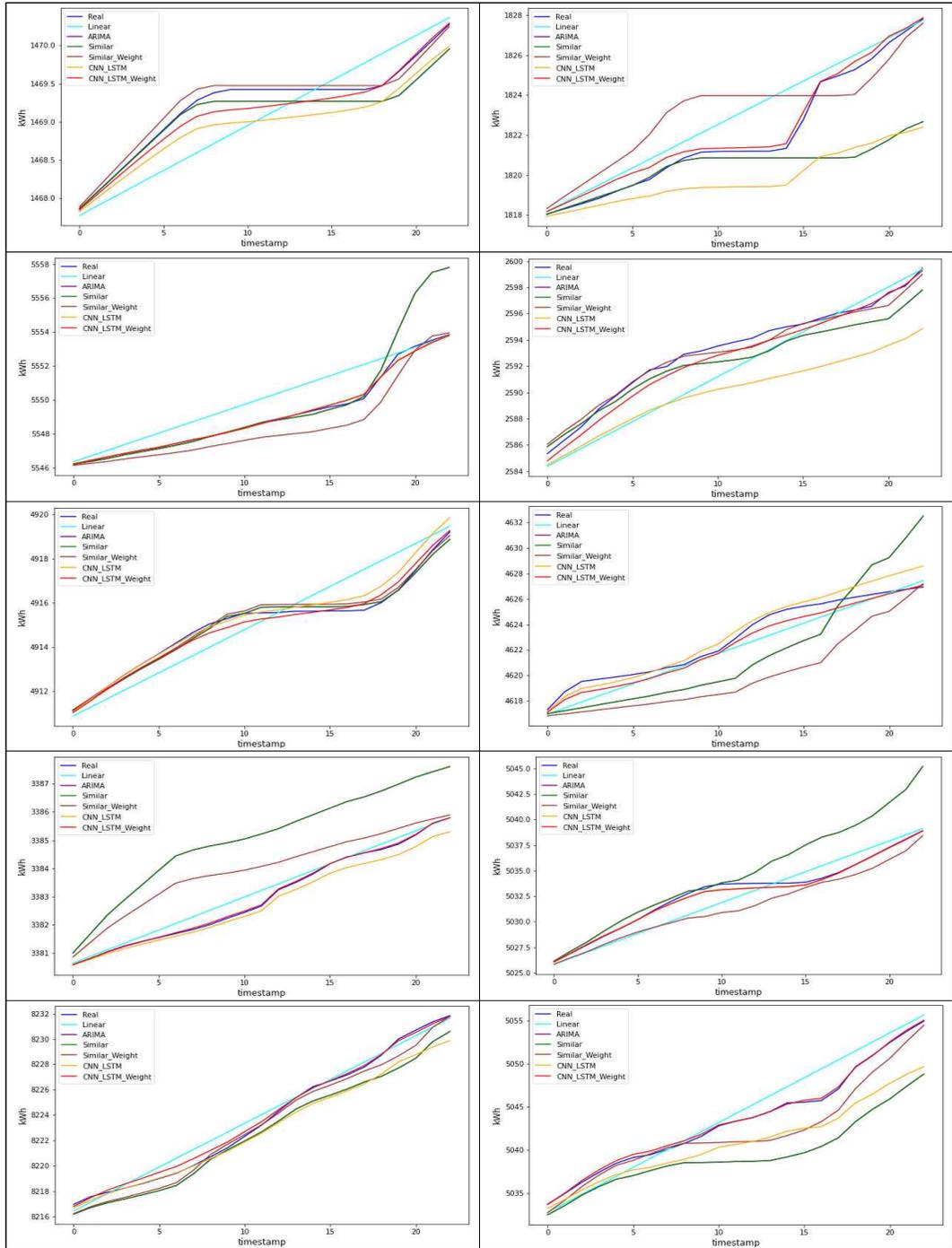
[그림 5-32]는 결측 구간의 데이터를 예측한 후 1000개 계량기의 시간별 구간 전력사용량 평균을 표현했다. 선형보정법은 결측 구간 전체를 같은 사용량으로 예측했기 때문에 직선으로 나타났다. 그리고 전체 계량기의 평균 사용량은 새벽 시간대에서 전력 사용량이 가장 적게 나타나고 21시경이 가장 높게 나타났다. 시간별 구간 전력사용량의 전체 계량기의 평균으로 표현했기 때문에 보정 방법별로 정확도가 확연히 들어나지는 않지만 본 논문에서 제안한 CNN-LSTM결합 가중치적용 보정법이 새벽시간대를 제외하고는 실제 사용량과 매우 유사하게 나타났다. 특히 시간별 구간 전력사용량으로 볼 때 실제 사용량과 비교했을 때 다른 형태는 선형보정법과 ARIMA 예측 보정법이 전혀 다른 패턴을 보였다.



[그림 5-32] 시간별 구간 전력사용량 평균

[표 5-7]은 실험한 6가지 보정법에 대해서 계량기별로 실제 데이터와 비교한 그래프를 표현하였다. 제안한 모델이 타 보정법에 비해서 월등히 우수함을 알 수 있다. 아래 표는 샘플로 10개만 표시하였고, 추가로 더 많은 자료를 [부록 3]에 표시하였다.

[표 5-7] 계량기별 결측 보정 방법별 예측결과와 실제데이터 비교



2. 선행 보정법과 제안된 보정법의 비교 분석 결과

[그림 5-33]은 결측 구간의 시간별 누적 전력사용량을 예측한 최종 결과 데이터이다.

	Meter_id	Hour	Real	Linear	ARIMA	Similar	Similar_Weight	CNN_LSTM	CNN_LSTM_Weight
0	M0001	0000	555.578	555.6594	555.7607	556.932	556.5202	555.4855	555.5764
1	M0001	0100	555.707	555.9308	556.1681	557.697	557.0812	555.5971	555.7921
2	M0001	0200	556.113	556.2021	556.4842	558.732	557.8401	555.7693	556.1249
3	M0001	0300	556.793	556.4735	556.7911	559.605	558.4803	556.1010	556.7659
4	M0001	0400	557.469	556.7449	557.0664	560.011	558.7780	556.3801	557.3052
...
22995	M1000	1800	1797.288	1797.7020	1797.7030	1798.804	1796.8901	1797.1134	1797.5327
22996	M1000	1900	1797.467	1798.0540	1798.0520	1798.984	1797.0259	1797.3374	1797.7721
22997	M1000	2000	1798.141	1798.4060	1798.4020	1799.579	1797.4748	1797.8015	1798.2681
22998	M1000	2100	1798.573	1798.7590	1798.7510	1800.700	1798.3206	1798.2022	1798.6963
22999	M1000	2200	1798.985	1799.1110	1799.1000	1801.541	1798.9551	1798.5645	1799.0835

23000 rows × 9 columns

[그림 5-33] 전체 실험 결과 데이터

[그림 5-34]는 1000개 계량기 전체 실험 결과 데이터를 요약한 정보이다. 최댓값과 최솟값의 차이가 많이 발생함으로 MSE, MAE, RMSE 등의 평가지표로 오차를 비교하는 것은 정확하지 않을 수도 있다. 그래서 추가적으로 평균절대비오차(MAPE; Mean Absolute Percentage Error)도 계산하여 비교를 수행하였다.

	Real	Linear	ARIMA	Similar	Similar_Weight	CNN_LSTM	CNN_LSTM_Weight
count	23000.000000	23000.000000	22724.000000	23000.000000	23000.000000	23000.000000	23000.000000
mean	6164.029763	6164.435855	6218.109679	6164.047446	6163.904251	6163.720038	6164.118128
std	6660.493715	6660.728606	6681.569134	6660.637622	6660.410366	6660.607773	6660.537885
min	1.045000	1.045000	0.000000	1.045000	1.045000	1.045000	1.045000
25%	2074.145250	2074.170000	2120.460000	2074.144000	2074.145050	2074.071825	2074.147450
50%	4336.715000	4336.424500	4369.331500	4337.145000	4336.395100	4336.853100	4336.562050
75%	8600.557250	8600.544500	8779.334000	8601.038000	8600.654850	8600.149550	8600.586725
max	82950.020000	82950.110000	82956.270000	82958.080000	82950.639000	82952.991400	82949.984200

[그림 5-34] 전체 실험 결과 요약 데이터

[그림 5-35]는 계량기별 시간별로 평균절대비오차(MAPE; Mean Absolute Percentage

Error)를 계산하였다.

Meter_id	Hour	MAPE_Linear	MAPE_ARIMA	MAPE_Similar	MAPE_Similar_Weight	MAPE_CNN_LSTM	MAPE_CNN_LSTM_Weight	
0	M0001	0000	0.014651	0.032885	0.243710	0.169589	0.016649	0.000288
1	M0001	0100	0.040273	0.082975	0.358102	0.247289	0.019777	0.015314
2	M0001	0200	0.016022	0.066749	0.470947	0.310566	0.061804	0.002140
3	M0001	0300	0.057382	0.000341	0.505035	0.303039	0.124283	0.004867
4	M0001	0400	0.129891	0.072219	0.455989	0.234811	0.195329	0.029383
...
22995	M1000	1800	0.023035	0.023090	0.084349	0.022139	0.009715	0.013615
22996	M1000	1900	0.032657	0.032546	0.084397	0.024540	0.007210	0.016974
22997	M1000	2000	0.014737	0.014515	0.079971	0.037049	0.018881	0.007068
22998	M1000	2100	0.010342	0.009897	0.118260	0.014033	0.020616	0.006855
22999	M1000	2200	0.007004	0.006392	0.142080	0.001662	0.023374	0.005475

23000 rows × 8 columns

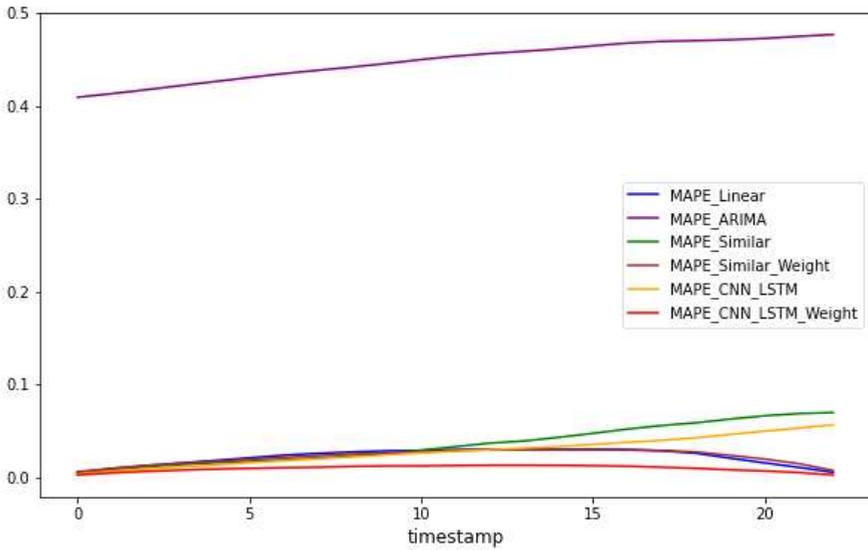
[그림 5-35] 전체 실험 결과의 보정법별 MAPE

[그림 5-36]은 평균절대비오차(MAPE; Mean Absolute Percentage Error)를 시간별로 계산한 결과이다.

Hour	MAPE_Linear	MAPE_ARIMA	MAPE_Similar	MAPE_Similar_Weight	MAPE_CNN_LSTM	MAPE_CNN_LSTM_Weight	
0	0000	0.005791	0.409093	0.005474	0.006546	0.004000	0.003080
1	0100	0.009842	0.412935	0.008972	0.010141	0.007023	0.005269
2	0200	0.012995	0.417186	0.011801	0.013112	0.009485	0.006952
3	0300	0.015767	0.421605	0.014154	0.015426	0.011733	0.008251
4	0400	0.018481	0.425957	0.016013	0.017416	0.013915	0.009322
5	0500	0.021261	0.430289	0.017748	0.019723	0.016083	0.010105
6	0600	0.024085	0.434457	0.019340	0.021804	0.018201	0.010872
7	0700	0.026027	0.438054	0.021242	0.023756	0.020201	0.011341
8	0800	0.027571	0.441595	0.023518	0.025699	0.022183	0.012161
9	0900	0.028647	0.445339	0.025731	0.026219	0.024375	0.012705
10	1000	0.029040	0.449585	0.029408	0.027409	0.026578	0.012745
11	1100	0.029937	0.453275	0.033168	0.029078	0.028136	0.013105
12	1200	0.030354	0.456141	0.036888	0.029936	0.029869	0.013409
13	1300	0.030274	0.458442	0.039213	0.029708	0.031615	0.013446
14	1400	0.030177	0.461043	0.043037	0.030046	0.033354	0.013319
15	1500	0.030105	0.464290	0.047283	0.030583	0.035523	0.012963
16	1600	0.029798	0.467294	0.051765	0.030144	0.037787	0.012400
17	1700	0.028805	0.469120	0.055597	0.029059	0.039885	0.011338
18	1800	0.026212	0.469887	0.058510	0.027542	0.042543	0.010194
19	1900	0.020954	0.470873	0.062436	0.023927	0.046154	0.008537
20	2000	0.015975	0.472309	0.065962	0.020025	0.049628	0.007402
21	2100	0.011161	0.474490	0.068280	0.015098	0.052988	0.005629
22	2200	0.005949	0.476444	0.069426	0.007876	0.056275	0.003155

[그림 5-36] 전체 실험 결과의 보정법별 시간별 평균 MAPE

[그림 5-37]은 계량기별 시간별로 MAPE를 계산하였고, 0시부터 22시까지의 각 보정법의 MAPE값을 그래프로 표현하였다. 선형보정법과 가중치를 적용한 보정법을 제외하고는 결측 길이가 커질수록 MAPE값이 점점 커지는 것을 알 수 있다. 그리고 본 논문에서 제안한 보정법이 선형보정법보다도 결측 중간시간대에서도 정확도가 명확히 높다는 것을 알 수 있다.



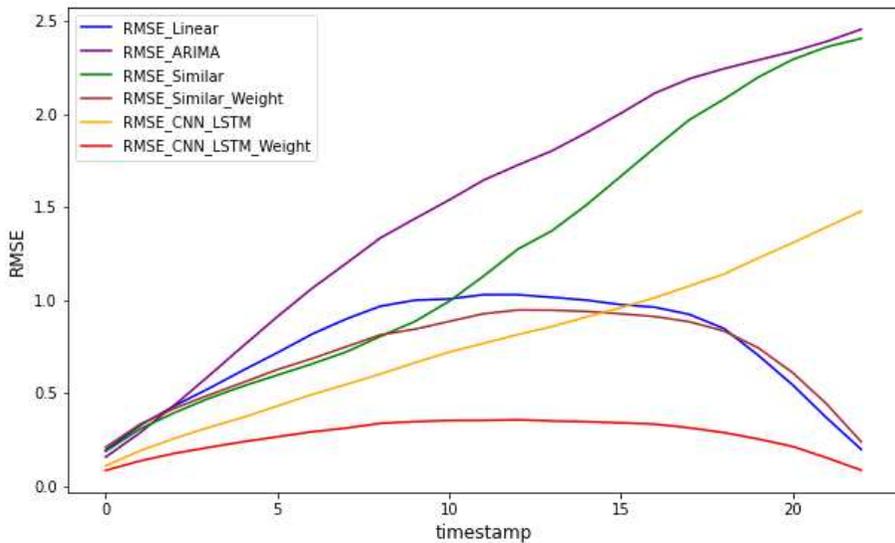
[그림 5-37] 전체 보정 방법별 MAPE

[그림 5-38]은 전체 MAPE값의 요약정보이다. 아래 요약정보로 보았을 때 전체 1000개 계량기의 24시간 평균 MAPE값은 유사도 기반 보정법이 '0.448683' 으로 정확도가 가장 낮았고, 본 논문에서 제안한 보정법이 '0.010020' 으로 가장 좋은 결과가 나왔다.

	Time	MAPE_Linear	MAPE_ARIMA	MAPE_Similar	MAPE_Similar_Weight	MAPE_CNN_LSTM	MAPE_CNN_LSTM_Weight
count	23.00000	23.000000	23.000000	23.000000	23.000000	23.000000	23.000000
mean	11.00000	0.022409	0.448683	0.036304	0.022455	0.028936	0.010020
std	6.78233	0.008522	0.021397	0.020864	0.007791	0.015157	0.003345
min	0.00000	0.005862	0.409093	0.005541	0.006626	0.004048	0.003117
25%	5.50000	0.016064	0.432373	0.018769	0.016621	0.017350	0.007921
50%	11.00000	0.026343	0.453275	0.033571	0.024218	0.028478	0.011004
75%	16.50000	0.029777	0.468207	0.054333	0.029421	0.039308	0.012880
max	22.00000	0.030723	0.476444	0.070269	0.030955	0.056958	0.013609

[그림 5-38] MAPE 요약 데이터

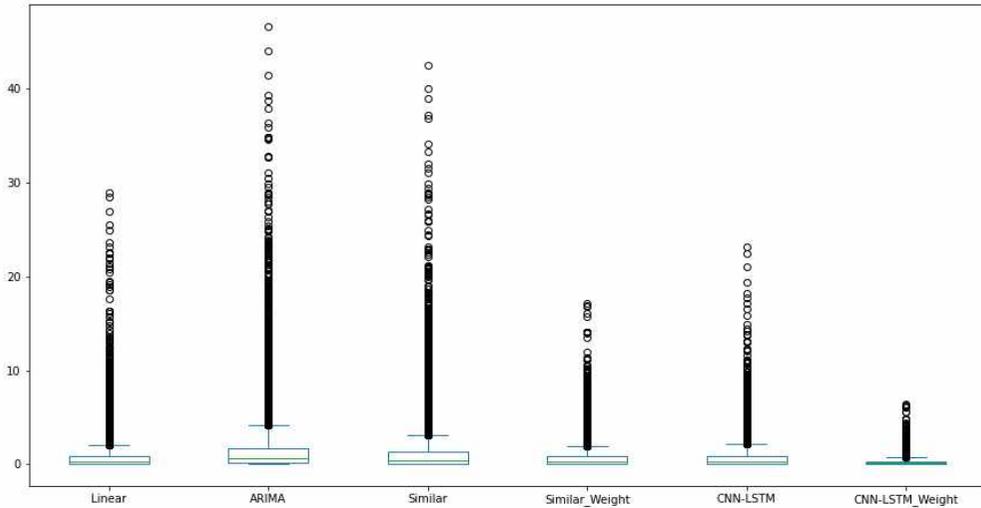
[그림 5-39]는 각 보정 방법별 RMSE(Root Mean Square Error) 오차를 그래프로 표현하였다. 선형보정법과 가중치를 적용한 보정법만이 결측 구간의 양쪽 끝 부분 오차가 '0'에 수렴하고 있다. 가중치를 적용함으로써 다른 보정법도 성능이 좋아졌다.



[그림 5-39] 전체 실험결과 시간별 RMSE

[그림 5-40]은 전체 데이터에 대한 AE(Absolute Error)를 박스 플롯(box plot)으로 표현했다. 박스 플롯으로만 봤을 때 절대 오차에 대한 순서는 ARIMA 예측 보정법 > 유사도 기반 보정법 > 선형보정법 > CNN-LSTM결합 예측 보정법 > 유사도 가중치적용 보정법 > CNN_LSTM결합 가중치적용 보정법 순이다. 요즘 시계열 분야에서 좋은 성능을 보이는 딥러닝 보정법이 상당히 좋은 결과를 보였으나,

앞서 얘기한 여러 가지 이유로 본 논문에서는 딥러닝 보정법에 가중치를 적용한 것이 더 좋은 결과가 나온 것을 확인 할 수 있다.



[그림 5-40] 전체 데이터의 AE(Absolute Error) boxPlot

[표 5-8]은 전체 계량기에 대한 각 보정 방법별 실험 평가결과 비교자료를 표현했다. 본 논문에서 제안한 보정법이 모든 평가지표에서 가장 우수한 결과가 나왔다. 정확도가 가장 나쁜 보정법은 ARIMA 예측 보정법이었다. 평균제곱오차(MSE)는 본 논문에서 제안한 보정법이 '0.218447'으로 ARIMA 예측 보정법보다 약 40배 이상 좋았고, 제공근 평균제곱오차(RMSE)도 6배 이상 좋았다. MAPE 값으로 비교 시에도 본 논문에서 제안한 보정법이 '0.009899' 으로 ARIMA 예측 보정법 '0.448682' 보다 약 45배 이상으로 정확도가 가장 우수했다.

[표 5-8] 전체 계량기에 대한 각 보정 방법별 실험 평가결과 비교

구 분	MSE	MAE	RMSE	MAPE
선형보정법	2.320586	0.712757	1.523347	0.022139
ARIMA 예측 보정법	8.883553	1.448703	2.980529	0.448682
유사도 기반 보정법	7.044095	1.193031	2.654071	0.035867
유사도 가중치적용 보정법	1.782093	0.694318	0.694318	0.022185
CNN-LSTM결합 예측 보정법	1.960150	0.735364	1.400053	0.028588
CNN-LSTM결합 가중치적용 보정법	0.218447	0.256664	0.467383	0.009899

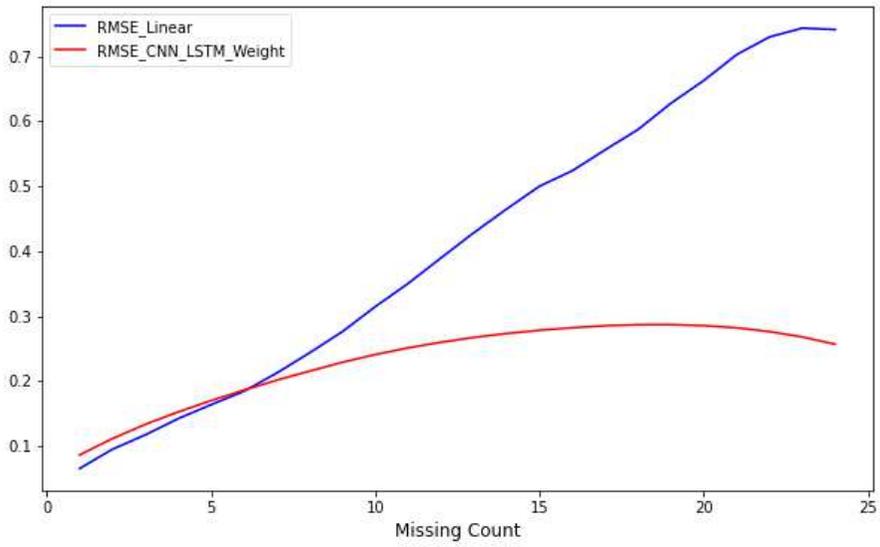
3. 결측 길이에 따른 분석 결과

앞서 실험한 것은 6가지 보정법에 대해서 24시간의 결측이 발생한 경우를 실험을 하였고, 이에 따라 각 보정법 비교분석을 진행하였다. 실험한 보정법 중에서 결측 길이에 따라 예측이 달라지는 보정법은 선형보정법만 존재한다. 그래서 본 논문에서 제안한 보정법과 선형보정법의 결측 길이에 따른 비교 분석을 하였다. [그림 5-41]은 선형보정법과 CNN-LSTM결합 가중치적용 보정법의 MSE, MAE, RMSE 값의 계산 결과 데이터이다. MAE, RMSE의 값이 결측 길이 7개부터 CNN-LSTM결합 가중치적용 보정법의 오차가 더 작아지기 시작한다.

Missing_Count	MSE_Linear	MSE_CNN_LSTM_Weight	MAE_Linear	MAE_CNN_LSTM_Weight	RMSE_Linear	RMSE_CNN_LSTM_Weight
1	0.0168033420000000	0.0249462786000000	0.0649960000000000	0.0859336000000000	0.0649960000000000	0.0859336000000000
2	0.0440353209400000	0.0423017898300000	0.0947450000000000	0.1111762000000000	0.0947450000000000	0.1111762000000000
3	0.0683654381433333	0.0601882220133333	0.1169449666666667	0.1331463333333333	0.1169449666666667	0.1331463333333333
4	0.0968202389500000	0.0773833194500000	0.1418713500000000	0.1522039500000000	0.1418713500000000	0.1522039500000000
5	0.1235558162080000	0.0941552736260000	0.1633384000000000	0.1696807400000000	0.1633384000000000	0.1696807400000000
6	0.1487911861533333	0.1098718242083333	0.1840566333333333	0.1857887833333333	0.1840566333333333	0.1857887833333333
7	0.1840644658371429	0.1260320015414286	0.2121744857142857	0.2011095857142857	0.2121744857142857	0.2011095857142857
8	0.2368897018375000	0.1424808950525000	0.2433038250000000	0.2150913500000000	0.2433038250000000	0.2150913500000000
9	0.3157423802111111	0.1609746376333333	0.2759906222222222	0.2287913555555555	0.2759906222222222	0.2287913555555555
10	0.4055508228260000	0.1779392437050000	0.3144923000000000	0.2407007500000000	0.3144923000000000	0.2407007500000000
11	0.4985359819545455	0.1926190428154546	0.3501588909090909	0.2509660272727273	0.3501588909090909	0.2509660272727273
12	0.6157012936483334	0.2051796015383333	0.3896057666666667	0.2595682166666667	0.3896057666666667	0.2595682166666667
13	0.7549528620146154	0.2173756841469231	0.4284464076923077	0.2671112076923077	0.4284464076923077	0.2671112076923077
14	0.9301835302492857	0.2276746403400000	0.4650136214285714	0.2731185000000000	0.4650136214285714	0.2731185000000000
15	1.1662590557146667	0.2360288698760000	0.5001157466666667	0.2780949600000000	0.5001157466666667	0.2780949600000000
16	1.3150993480956250	0.2428762356218750	0.5240205562500000	0.2820449437500000	0.5240205562500000	0.2820449437500000
17	1.4932633724247060	0.2477609589476470	0.5562338000000000	0.2851057705882353	0.5562338000000000	0.2851057705882353
18	1.6788330064588890	0.2501023813883334	0.5875298555555556	0.2867367833333334	0.5875298555555556	0.2867367833333334
19	1.9140072496910527	0.2503297030363158	0.6279943526315790	0.2868519315789474	0.6279943526315790	0.2868519315789474
20	2.0585520625729998	0.2483710375875000	0.6629494000000000	0.2852492950000000	0.6629494000000000	0.2852492950000000
21	2.2492338291414287	0.2441020514395238	0.7029026904761905	0.2818753952380952	0.7029026904761905	0.2818753952380952
22	2.3930479056636362	0.2369481089736364	0.7301635636363636	0.2760390090909091	0.7301635636363636	0.2760390090909091
23	2.4214812538717392	0.2279449451321739	0.7437464130434782	0.2678234086956522	0.7437464130434782	0.2678234086956522
24	2.3594392225208334	0.2184472414116667	0.7415496833333334	0.2566702333333333	0.7415496833333334	0.2566702333333333

[그림 5-41] 결측 길이에 따른 MSE, MAE, RMSE 데이터

[그림 5-42]는 위 데이터를 기반으로 RMSE를 그래프로 표현하였다. 결측 길이 1개 일때는 선형보정법이 더 좋았으나, 7개부터 역전되기 시작하면서 CNN-LSTM결합 가중치적용 보정법이 오차가 훨씬 더 적게 나타났다.



[그림 5-42] 결측 길이에 따른 RMSE

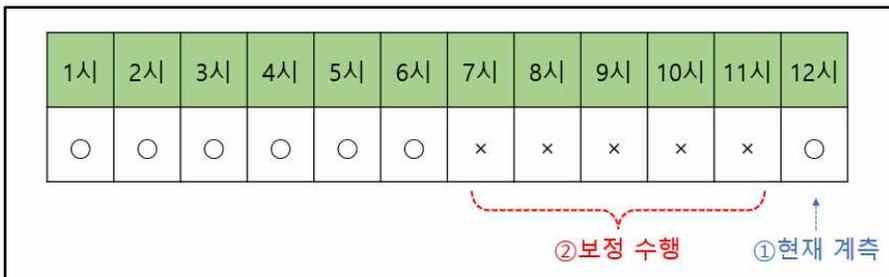
C. 연구 결과 활용 가능성

본 연구는 AMI시스템에서 수집된 전자식 전력량계의 결측 데이터를 보정하기 위한 방법으로 CNN-LSTM결합 가중치적용 모델을 제안하였다. 그리고 1000개 계량기의 1개월분 실험 데이터를 대상으로 결측 데이터를 보정하기 위한 선행 연구들의 방법과 비교했을 때 정확도가 우수한 것으로 평가되었다. 이와 같은 분석 결과를 바탕으로 본 연구결과의 활용방안으로 제안한 CNN-LSTM결합 가중치적용 모델의 적용 가능성과 적용 했을 때 고려사항 등에 대해서 제안하고자 한다.

1. 연구결과 적용 가능성

본 연구에서 제안한 CNN-LSTM결합 가중치적용 모델은 전력 계량데이터를 수집하고 관리하는 서버시스템 분야인 AMI시스템이나 MDMS시스템에 적용이 가능할 것이다.

첫 번째는 제안모델을 시스템에 적용하는 방법은 [그림 5-43]처럼 데이터를 수집할 때, 예로 12시 데이터가 수신되었다면 먼저 DB에 저장한 후에 직전 시간의 데이터가 존재하지 않는다면 제안모델을 통해서 결측 구간을 예측하여 DB에 실제 데이터와 구별하기 위해서 Flag와 함께 저장한다. 앞 뒤 데이터가 존재하고 중간에만 결측된 경우에 본 논문에서 제안한 모델로 데이터를 예측하여 보정했기 때문에 누적 전력사용량 데이터가 감소되는 현상이 발생하지 않고, 또한 누적 전력사용량을 보정했기 때문에 일간 또는 월간 등 전체 사용량에는 영향을 끼치지 않는다.



[그림 5-43] 시스템 적용 방안

두 번째 본 논문에서는 중간 결측 부분만 연구를 시행하였지만 결측 앞·뒤 데이터가 존재하는 경우가 아닌 뒤에 데이터가 없는 경우 즉, 현재 시점에도 데이터가

수집되지 않고 있는 경우에는 가중치를 적용하기 전인 CNN-LSTM결합 딥러닝 모델의 예측을 사용하면 가능하다. CNN-LSTM결합 딥러닝 모델의 예측 또한 정확도가 매우 높았기 때문에 다른 예측보다는 효과가 높을 것이다.

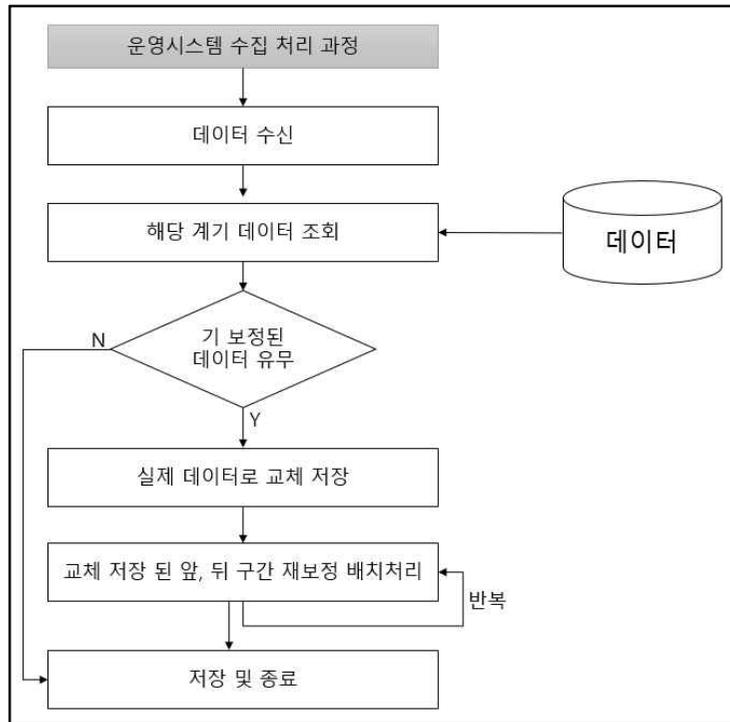
세 번째 전력 계량데이터 분야 외에 타 시스템에도 적용가능하다. 누적 전력사용량처럼 데이터가 계속 증가하는 누적 증가 데이터를 관리하는 시스템에서 결측이 발생했을 때 사용하면 효과적일 것이다.

2. 연구결과 적용시 고려사항

본 연구에서 제안한 CNN-LSTM결합 가중치적용 모델을 적용하기 위해서는 몇 가지 고려해야 할 사항들이 있다.

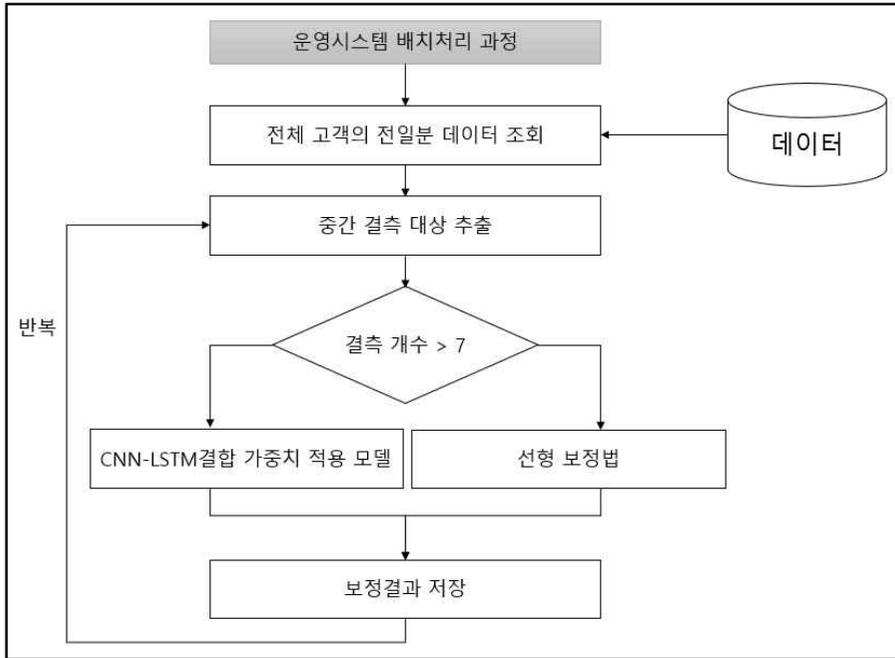
첫 번째 결측 데이터를 예측하여 보정을 할 때 실제 데이터와 구별할 수 있도록 데이터에 Flag 표시를 해 두어야 한다. 이유는 DCU나 모뎀으로부터 계량데이터가 수집될 때 중복된 데이터나 지연된 데이터가 수집되고 있기 때문에 중복 데이터라면 저장할 필요가 없지만 지연된 데이터가 수집됐다면 실제 데이터가 아니고 예측된 데이터라면 데이터를 교체해야 한다. 그리고 전력 사용량 데이터를 제공하는 서비스를 할 때에도 예측된 데이터라면 사용자에게 반드시 알릴 필요가 있다.

두 번째 지연수집에 따른 대체처리 방안을 고려해야 한다. 계량데이터 결측 관련 선행연구 결과의 시사점에서 언급했듯이 통신 회선 품질 문제로 DCU나 모뎀 등이 최신 데이터부터 먼저 전송하고 과거 데이터를 나중에 서버로 전송하고 있다. [그림 5-44]는 AMI시스템에서 결측 데이터를 먼저 예측하여 보정을 한 후에 실제 데이터가 지연돼서 수집된다면 실제 데이터로 대체처리를 해야 하고 대체된 데이터의 앞·뒤에 결측 예측된 데이터가 존재한다면 해당 구간에 대해서는 다시 예측을 하여 수정 보정을 수행해야 한다.



[그림 5-44] 시스템 지연수집 처리 과정

세 번째 결측 길이에 따른 분석결과에서도 보았듯이 결측 길이가 7개 미만일 경우는 선형보정법을 사용하는 것이 효과적이고 정확도로 높을 것이다. 선형보정법은 계산방법도 쉽고 컴퓨팅 리소스를 적게 필요로 한다. 그러므로 AMI시스템에 적용 시에는 결측 길이를 먼저 판단하고, 결측 길이가 Threshold 값 이하인 경우는 선형보정법을 사용하고, 이상인 경우에는 본 논문에서 제안한 CNN-LSTM결합 가중치 적용 보정법을 사용하는 것이 좋을 것이다. 자세한 처리과정은 [그림 5-45]에 표시하였다.



[그림 5-45] 시스템 결측 보정 배치처리 과정

VI. 결론 및 제언

지구 온난화, 이상기후 등 국제사회는 기후변화의 심각성을 인식하고 탄소배출량을 감축하기 위해 노력하고 있다. 이에 탄소배출량을 줄이기 위해서 가장 기본이 되는 에너지를 절약 하는데 초점을 맞추고 있다. 에너지 절약을 위해서 AMI구축을 서두르고 AMI 인프라를 활용해 소비자와 전력회사 간 양방향통신으로 다양한 부가서비스, 계시별 요금제(TOU; Time Of Use), 수요관리형 선택요금(CPP; Critical Peak Pricing), 실시간 요금제(RTP; Real Time Pricing) 등 고도화된 Time-based 요금제 지원을 통해 수용가의 능동적인 에너지 절감 참여를 유도하고 있다.

이러한 계량데이터를 활용한 부가 서비스들을 제공하기 위해선 필수적으로 전력량계로부터 계량데이터들을 잘 취득하여야 한다. 그러나 AMI를 구성하고 있는 통신망의 한계 및 여러 가지 요인으로 인해서 계량데이터의 결측이 2~5% 가량 발생하고 있다. AMI 부가 서비스의 품질을 높이기 위해서는 결측 데이터를 예측하여 제공하는 것이 필요하다. 그래서 스마트미터에서 생성한 전력량 데이터의 값을 예측하기 위해 여러 가지 알고리즘들을 연구하고 적용하고 있다. 하지만 그 정확도가 높지 않고 결측 구간이 길어지면 오차가 커져 양질의 서비스를 제공할 수가 없다.

본 논문에서는 누적 전력사용량 데이터에 대해서 결측이 끝나고 다음 데이터가 나타난 후에 보정하는 방법 즉 결측 구간의 앞·뒤의 데이터가 존재하고 중간에 데이터가 빠진 상태의 데이터를 보정하는 방법에 대해서 연구를 하였다.

다양한 분야에서 결측 데이터를 취급하는 연구는 많이 되어 왔으나, AMI 계량데이터처럼 누적되는 전력사용량 데이터를 보정하는 방법에 대한 연구는 많지 않았다. 결측 데이터를 보정하기 위해 그 동안 수행돼 왔던 연구 결과를 보았을 때 선형보정법은 간단하면서도 꽤 좋은 성능을 보였다. 하지만 결측 구간이 길어졌을 때 일정한 양으로 등분하므로 결측 중간부분에서 상당한 오차가 발생했다. 유사도 기반 보정법은 과거 전력사용량 패턴 중 유사도 구간을 찾아서 보정하는 방법으로 아이디어는 좋았으나 사용량 패턴이 불규칙 할 경우, 봄철과 여름철 전력사용량이 확연히 다를 경우 패턴은 유사하나 사용량의 크기가 차이가 많이 날 경우 상당한

오차를 보이게 되었다. 또한 유사구간의 시간별 구간 전력사용량을 참조해서 결측 구간의 누적 전력사용량에 합산하는 방식이므로 결측 길이가 길어지면 길어질수록 오차가 커지고, 누적 전력사용량이 감소되는 현상도 발생하였다. ARIMA 예측 보정법은 시계열 분야에서 전통적으로 많이 사용해오던 방식이며, 차분할 필요 없이 누적 전력사용량을 바로 입력하여 결과를 얻어 낼 수 있는데, 하지만 실험한 여러 가지 보정법 중에서 가장 좋지 못한 성능을 보였다. 장단기메모리순환신경망(LSTM; Long Short-Term Memory) 예측 보정법은 시계열 분야에서 요즘에 가장 많이 사용하고 있고, 성능이 좋다고 알려져 있다. 실제로 여러 가지 보정법 중에 좋은 성능을 보였다. 하지만 시간별 전력사용량을 예측하여 결측 구간의 누적 전력사용량에 합산하는 방식을 적용하다보니 결측 구간이 길어질수록 오차가 커지고, 누적 전력사용량이 감소되는 현상이 발생하였다.

본 논문에서는 이러한 문제들을 해결하기 위해 시계열에 사용되는 일반적인 방법들보다 선형보정법의 장점과 시계열 분야에서 우수한 장단기메모리순환신경망(LSTM; Long Short-Term Memory) 예측을 이용한 보정법의 장점을 결합한 하이브리드 방법으로 딥러닝 기반 CNN-LSTM결합 모형에 가중치를 적용한 모델을 제시하였다. 본 연구에서 제안한 CNN-LSTM결합 가중치적용 모델의 가장 큰 특징은 딥러닝 모델로 시간별 구간 전력사용량을 1차 예측하고, 예측한 결괏값에 가중치 비율을 적용한 2차 예측을 수행하고, 결측 구간에 합산과정을 통해 누적 전력사용량을 최종 예측하게 되므로 누적 전력사용량이 감소되는 오류를 방지할 수 있다. 이러한 이유로 누적 전력사용량의 중간 결측 보정법에 최적화된 기법을 제시했다. 세부적인 주요 연구 결과는 다음과 같다.

첫째, 데이터를 분석한 결과 전체 실험대상 1000개 계량기 중 976개가 일반적인 누적 전력사용량 패턴인 시간별로 계속 증가하는 형태를 보였다. 하지만 2.4%인 24개는 사용량이 없거나 특이치를 보였다. 또한 시간별 구간 전력사용량 패턴을 분석한 결과 시간과 요일에 따라 사용량이 일정한 패턴을 보인 계량기들이 많았고, 여기에 7월 중순경부터 갑자기 전력소비가 늘어나는 패턴을 보인 계량기들도 있었다. 하지만 규칙적인 패턴이 아닌 계량기들도 일부 있었다. 이런 불규칙적인 패턴을 보인 데이터는 딥러닝 예측 시에도 좋지 않은 결과가 나왔다.

둘째, 전체 계량기에 대한 선행 보정법과 제안된 보정법의 실험 결과를 비교 분석했을 때 평균제곱오차(MSE)는 본 논문에서 제안한 보정법이 '0.218447'로 ARIMA 예측 보정법 '8.883553' 보다 약 40배 이상 좋았고, 제곱근 평균제곱오차(RMSE)도 6배 이상 좋았다. 평균절대비오차(MAPE) 값으로 비교 시에도 본 논문에서 제안한 보정법이 '0.009899' 으로 ARIMA 예측 보정법 '0.448682' 보다 약 45배 이상으로 정확도가 가장 우수했다. 또한 본 논문에서 제안한 보정법은 결측 구간의 총량 대비 구간별 비율인 가중치를 적용하였기 때문에 누적 전력사용량 값이 감소되는 현상도 발생하지 않았고, 시간별 구간 전력사용량 예측 오차가 많이 발생하더라도 총 사용량 내에서 예측을 하게 되므로 안정성이 매우 높다.

셋째, 선행 보정법 또한 가중치를 적용함으로써 성능이 좋아졌고, 누적 전력사용량 값이 감소되는 현상도 발생하지 않기 때문에 본 논문에서 주장하는 누적 전력사용량의 중간 결측 값을 보정하기 위해서는 가중치 적용이 반드시 필요하다는 것을 알 수 있다.

네째, 결측 길이에 따른 실험결과를 분석했을 때 일반적으로 시계열 분야에서 성능이 좋은 여러 가지 방법들보다 선형보정법이 간단하면서도 성능이 좋았다. 결측 구간의 데이터가 개수가 7개 미만 일 때 까지는 선형보정법이 정확도가 가장 높았다. 그래서 AMI시스템에 적용 시에는 결측 길이를 먼저 판단하여 결측 길이가 7개 미만일 경우는 선형보정법을 사용하는 것이 효과적이고, 이상인 경우에는 본 논문에서 제안한 CNN-LSTM결합 가중치적용 보정법을 사용하는 것이 좋을 것이다.

본 논문의 실험결과가 시사하는 점은 결측 데이터의 보정이 전기에너지에 대해서만 유효한 것은 아니다. 기타 다른 에너지원에 대한 수요/공급에서도 동일할 것이다. 또한 본 논문의 시사점을 통해서 계량데이터를 수집하여 서비스를 하는 시스템에서는 여러 가지 방법을 혼합하여 시스템을 구축하는 것이 효과적이라는 결과를 제공하였다.

향후, 제안된 모델이 단순히 전력사용량의 시계열 데이터만을 가지고 딥러닝 알고리즘으로 예측하여 보정하였다면, 딥러닝을 학습시키기 위해서 더 많은 입력값을 사용하여 그 정확도를 높이는 연구가 필요하다. 특히 전력사용량은 기후와 그 연관성이 매우 높기 때문에 날씨정보를 추가하여 정확도를 더 높일 필요가 있다.

참고문헌

- [1] Jung, J. Seo, C. “An Efficient Method for Meter Data Collection in AMI System.” J. Korean Inst. Commun. Inf. Sci, Vol.43, pp.1311 - 1320, 2018.
- [2] Dusa, P. Novac, C. et al. “Configuration a Meter Data Management System using Axiomatic Design.” Procedia CIRP, Vol.34, pp.174 - 179, 2015.
- [3] HJ Lee. “Visible Light Communication System For Optical Wireless Sensor Networks Based on PLC-Ethernet Interface.” Master’s Thesis, Korea Maritime & Ocean University, Busan, Korea, 2012.
- [4] Kusakana, Kanzumba. “Energy management of a grid-connected hydrokinetic system under Time of Use tariff.” Renewable Energy 101, pp.1325-1333, 2017.
- [5] Venizelou, Venizelos, et al. “Development of a novel time-of-use tariff algorithm for residential prosumer price-based demand side management.” Energy 142, pp.633-646, 2018.
- [6] Naz, Muqaddas, et al. “Efficient power scheduling in smart homes using hybrid grey wolf differential evolution optimization technique with real time and critical peak pricing schemes.” Energies 11.2, pp.384, 2018.
- [7] Steriotis, Konstantinos, et al. “A novel behavioral real time pricing scheme for the active energy consumers’ participation in emerging flexibility markets.” Sustainable Energy, Grids and Networks 16, pp.14-27, 2018.
- [8] Jordehi, A. Rezaee. “Optimisation of demand response in electric power systems, a review.” Renewable and sustainable energy reviews 103, pp.308-319, 2019.
- [9] Yan, Xing, et al. “A review on price-driven residential demand response.” Renewable and Sustainable Energy Reviews 96, pp.411-419, 2018.
- [10] Qi, Chensen, Chen-Ching Liu. “Integrated Outage Management with Feeder Restoration for Distribution Systems with DERs.” IEEE Access 9, pp.112978 -112993, 2021.
- [11] Kang, H-J. “A Study on the AMI Communication Method Combining High-Rate PLC of ISO/IEC 12139-1 and IEEE 802.15.4g Based Wi-SUN.” Ph.D. Thesis, Chonnam National University, Gwangju, Korea, 2018.

- [12] Kwon, H-R. Hong, T-E. “Estimate method of missing data using Similarity in AMI system.” *Smart Media J*, Vol.8, pp.80 - 84, 2019.
- [13] 성욱준. “한국판 뉴딜 종합계획과 데이터 시대의 진화: 한국판뉴딜종합계획, 데이터 3 법, 데이터기반행정법의 주요 이슈를 중심으로.” 건강보험심사평가원, 2020.
- [14] 김현제, 박찬국 “세계 스마트그리드 정책 및 시장 변화 연구.” 에너지경제연구원 수시연구보고서, pp.13-14, 2013.
- [15] Hyuk-Rok Kwon and Pankoo Kim. “Analysis of Regional Characteristics by Analyzing the Patterns of Electricity Usage.” *The Journal of Information Technology and Architecture*, Vol.16, No.3, pp.303-309. 2019.
- [16] Qian, X. Yang, Y. Li, C. Tan, S.C. “Economic Dispatch of DC Microgrids Under Real-Time Pricing Using Adaptive Differential Evolution Algorithm.” *Proceedings of the 2020 IEEE 9th International Power Electronics and Motion Control Conference (IPEMC2020-ECCE Asia)*, Nanjing, China, 29 November - 2 December, 2020.
- [17] Song, H. Yoon, Y. Kwon, S. “Optimal scheduling of critical peak pricing considering photovoltaic generation and electric vehicle load.” In *Proceedings of the 2019 IEEE Transportation Electrification Conference and Expo, Asia-Pacific (ITEC Asia-Pacific)*, Seogwipo, Korea, 8 - 10 May, 2019.
- [18] Lv, H. Wang, Y. Dong, X. Jiang, et al. “Optimization Scheduling of Integrated Energy System Considering Demand Response and Coupling Degree.” In *Proceedings of the 2021 IEEE/IAS 57th Industrial and Commercial Power Systems Technical Conference (I&CPS)*, Las Vegas, NV, USA, 27 - 30 April, 2021.
- [19] HomePlug Alliance. “HomePlug Green PHY Specification Release Version 1.1.1.”, 2013.
- [20] Junjalearnvong, Thidarut, et al. “Performance evaluation of multi-hop network configuration for Wi-SUN FAN systems.” *2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2019.
- [21] Kassem, Mohamed M. et al. “On the potential of TVWS spectrum to enable a low cost middle mile network infrastructure.” *2018 10th International*

- Conference on Communication Systems & Networks (COMSNETS). IEEE, 2018.
- [22] Bianchi, Valentina, et al. "RSSI-based indoor localization and identification for ZigBee wireless sensor networks in smart homes." IEEE Transactions on Instrumentation and Measurement, pp.566-575, 2018.
- [23] Choi, M.-S. "Development and Performance Analysis of Hybrid Communication Technology for Advanced Metering Infrastructure System." KIEE, Vol.69, pp. 610 - 616, 2020.
- [24] Chung, Y.M. Chung, B.J. Son, S.-Y. "Performance Analysis of a PLC for an Advanced Metering Infrastructure (AMI)." J. Electron. Inf. Eng. (IEIE), pp. 332 - 334, 2016.
- [25] International Organization for Standardization. "ISO/IEC12139-1." 2009(E).
- [26] C.Pirak, T.Cngsuwan, S.Buyairaksa. "Recent Advances in Communication technologies for Smart Grid Application: A Review." In Proceedings of the International Electrical Engineering Congress, Chonburi, Thailand, 19-21 March, 2014.
- [27] Chen,W. Zhou, K. Yang, S. Wu, C. "Data quality of electricity consumption data in a smart grid environment." Renew. Sustain. Energy Rev, Vol.75, pp. 98 - 105, 2017.
- [28] Choi, Y.J. Kim, S.Y. "Analysis on The Change of Power Consumption Pattern According to Single-Households." In Proceedings of the 2014 Conference on The Korean Institute of Electrical Engineers, Jeju, Korea, 15 - 19 June, pp.1 53 - 154, 2014.
- [29] Lee, J. Shin, J. et al. "A VEE Algorithm Improvement Research for Improving Estimation Accuracy and Verification Responsibility of The AMI Meter Data." KEPCO J. Electr. Power Energy, Vol.2, pp.557 - 562, 2016.
- [30] Jang, M. Nam, K. Lee, Y. "Analysis and Application of Power Consumption Patterns for Changing the Power Consumption Behaviors." J. Korea Inst. Inf. Commun. Eng, Vol.25, pp.603 - 610, 2021.
- [31] Kim, J.-O. "A Study on the Prediction of Short Term Electric Power Load by Deep Learning System." Master's Thesis, Dankook University, Yongin-si, Korea, 2019.

- [32] Ryu, S. “Deep Learning for Electric Load Data Analytics.” Master’s Thesis, Sogang University, Seoul, Korea, 2020.
- [33] Choi, H. “Short-Term Load Forecasting Based on ResNet and LSTM.” Master’s Thesis, Sogang University, Seoul, Korea, 2018.
- [34] Kim, D. “Short-Term Load Forecasting Based on LSTM and CNN.” Master’s Thesis, Konkuk University, Seoul, Korea, 2019.
- [35] Kwon, B.-S. Park, R.-J. Song, K.-B. “Analysis of Short-Term Load Forecasting Accuracy Based on Various Normalization Methods.” J. Korean Inst. Illum. Electr. Install. Eng, Vol.32, pp.30 - 33, 2018.
- [36] Koh, S. “Outlier Detection and Imputation Method for Smart Meter Data Using Pattern Analysis.” Master’s Thesis, Korea University, Seoul, Korea, 2019.
- [37] Timofey, S. Antonio, N. “Fraction-of- Time Density Estimation Based on Linear Interpolation of Time Series.” In Proceedings of the 2021 Systems of Signals Generating and Processing in the Field of on Board Communications Signals Generating and Processing in the Field of on Board Communications, Moscow, Russia, 16 - 18 March, pp.1 - 4, 2021.
- [38] Seo, S.-W. Kim, D.-H. Kim, S.J. “A Study on the Linear Compensation Method of Ideal Surface Roughness to Actual Roughness in Milling.” Korean Soc. Manuf. Process. Eng, Vol.15, pp.15 - 20, 2016.
- [39] Pejić, N. Cvetanović, M. Radivojević, Z. “Estimating similarity between differently compiled procedures using neural networks.” In Proceedings of the 2019 27th Telecommunications Forum (TELFOR), Serbia, Belgrade, 26 - 27 November, pp.26 - 27, 2019.
- [40] Lee, S. “Applying Different Similarity Measures based on Jaccard Index in Collaborative Filtering.” J. Korea Soc. Comput. Inf, Vol.26, pp.47 - 53, 2021.
- [41] Behera, A.P. Gaurisaria, M.K. et al. “Predicting Future Call Volume Using ARIMA Models.” In Proceedings of the 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS) Intelligent Computing and Control Systems (ICICCS), Madurai, India , 6 - 8 May, pp.1351 - 1354, 2021.
- [42] Garlapati, A. Krishna, D.R. et al. “Stock Price Prediction Using Facebook Prophet and Arima Models.” In Proceedings of the 2021 6th International

- Conference for Convergence in Technology (I2CT) Convergence in Technology (I2CT), Maharashtra, India, 2 - 4 April, pp.1 - 7, 2021.
- [43] Chang, H. Park, D. Lee, Y. Yoon, B. “Multiple time period imputation technique for multiple missing traffic variables: Nonparametric regression approach.” *Can. J. Civ. Eng.*, Vol.39, pp.448 - 459, 2012.
- [44] Asif, M.T. Mitrovic, N. Dauwels, J. Jaillet, P. “Matrix and Tensor Based Methods for Missing Data Estimation in Large Traffic Networks.” *IEEE Trans. Intell. Transp. Syst.*, Vol.17, pp.1816 - 1825, 2016.
- [45] Shakir, M. Marwala, T. “Neural network based techniques for estimating missing data in databases.” In *Proceedings of the 16th Annual Symposium of the Recognition Association of South Africa*, Langebaan, South Africa, 23 - 25 November, 2005.
- [46] Yu, Yong, et al. “A review of recurrent neural networks: LSTM cells and network architectures.” *Neural computation*, Vol.31, No.7, pp.1235-1270, 2019.
- [47] Il-Taeck Joo, Seung-Ho Choi. “Stock Prediction Model based on Bidirectional LSTM Recurrent Neural Network.” *Korea Institute of Science and Technology Information (KISTI)*, Vol.11, No.2, 2018.
- [48] Hyungeun Cho, et al. “Short-term Load Forecasting based on ResNet and LSTM.” *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*, Vol.1, pp.392-393, 2019.
- [49] Kim Han Gyung, “A Study on CNN-LSTM Coupled Model For Prediction of Waterworks Operation Data.” *Ph.D. Thesis, PaiChai University, Daejeon, Korea*, 2018.
- [50] Ha Seung woo. “A study on AMI Deployment Case and Analysis of the Effect for Smart Grid Materialization.” *Master’s Thesis, Kongkuk University, Seoul, Korea*, 2014.
- [51] Jin Ha Kim, Myong Ho Choi, et al. “Study of gathering Load Profile on Smart Meter efficiently.” *The Korean Institute of Electrical Engineers*, Vol. 7. pp.15-17, 2015.
- [52] Sahu, Abhijeet, and Ana Goulart. “Implementation of a C-UNB Module for NS-3 and Validation for DLMS-COSEM Application Layer Protocol.” 2019

- IEEE ComSoc International Communications Quality and Reliability Workshop (CQR). IEEE, 2019.
- [53] Joonhong Jung, Joungok Lee, “An Efficient Method for Meter Data Gathering in Data Concentration Unit.” The Korean Institute of Electrical Engineers, p p.68-69, 2016.
- [54] Seo, Chung-Ki. “An Efficient Routing Path Search Technique in Power Line Communication.” The Transactions of The Korean Institute of Electrical Engineers, Vol.67, No.9, pp.1216-1223, 2018.
- [55] Kwon, H-R. “Implementation dual-role network management agent for wired / wireless hybrid AMI system.” Master’s Thesis, Chosun University, Kwangju, Korea, 2018.
- [56] Kepco KDN. “Electric Power Information and Communications Technology Principle (Ver.3)”, Kepco KDN, 2014.
- [57] Seo, Chung-Ki. “A study on the power line communication-based network management agent for automatic meter reading with the routing optimization algorithm.” Ph.D. Thesis, Inha University, Inchun, Korea, 2015.
- [58] Ahn Joong Seok. “Risk Analysis for Advanced Metering Infrastructure Cybersecurity.” Master’s Thesis, Chungbuk National University, Cheongju, Korea, 2016.
- [59] Nam-Joon Jung. “A Design and Implementation of MDMS for AMI based Integrated Services of Smart Grid Application.” Master’s Thesis, Hanbat National University, Daejeon, Korea, 2013.
- [60] Huang, Z. Zhu, T. “Real-time data and energy management in microgrids.” In Proceedings of the 2016 IEEE Real-Time Systems Symposium (RTSS), Porto, Portugal, 29 November - 22 December, pp.79 - 88, 2016.
- [61] Peppanen, J. Zhang, X. Grijalva, S. Reno, M.J. “Handling bad or missing smart meter data through advanced data imputation.” In Proceedings of the 2016 IEEE Power & Energy Society, Innovative Smart Grid Technologies Conference (ISGT), Minneapolis, MN, USA, 6 - 9 September, pp.1 - 5, 2016.
- [62] Hye-Lym Lee, Han-Soo Kim. “Optimum Design of Reinforced Concrete Outrigger Wall Opening Using Piecewise Linear Interpolation.” Journal of

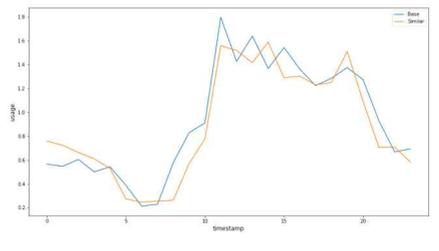
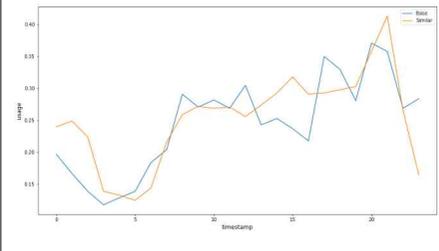
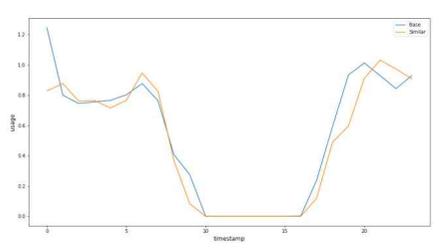
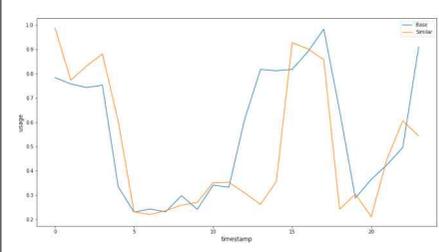
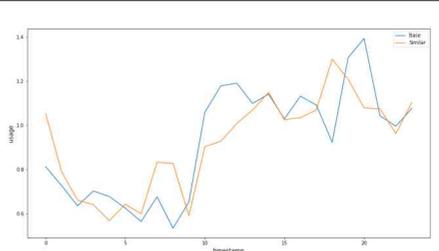
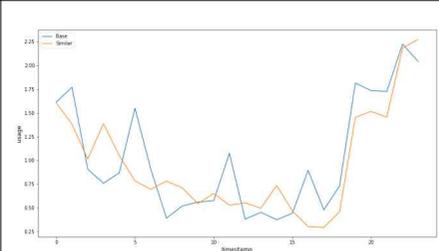
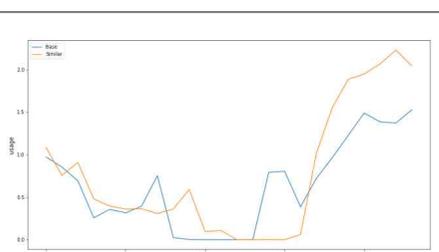
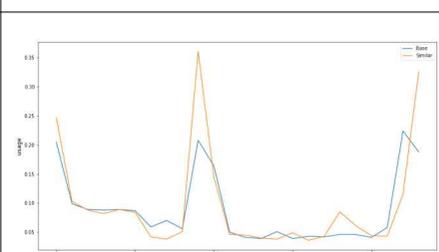
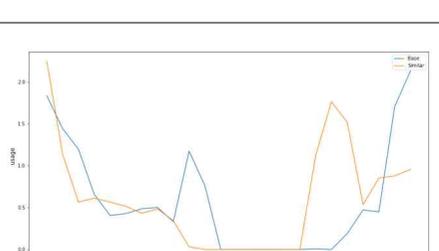
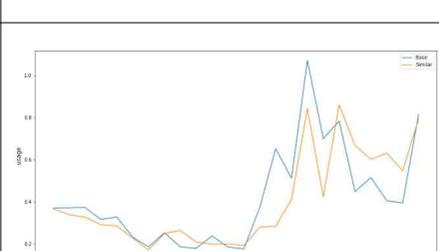
- the Computational Structural Engineering Institute of Korea, Vol.33, No.4, p p.217-224, 2020.
- [63] Yu, K. Guo, G.-D. Li, J. Lin, S. “Quantum Algorithms for Similarity Measurement Based on Euclidean Distance.” *Int. J. Theor. Phys*, Vol.59, pp.3134 - 3144, 2020.
- [64] Iglesias, F. Kastner, W. “Analysis of Similarity Measures in Times Series Clustering for the Discovery of Building Energy Patterns.” *Energies* 2013, p p.579 - 597, 2013.
- [65] Hwang, Chi-gon, Chang-Pyo Yoon, Dai Yeol Yun. “Sentence Similarity Analysis using Ontology Based on Cosine Similarity.” In *Proceedings of the Korean Institute of Information and Commucation Sciences Conference*, 2021.
- [66] Liu, Donghai, Xiaohong Chen, Dan Peng. “Some cosine similarity measures and distance measures between q rung orthopair fuzzy sets.” *International Journal of Intelligent Systems*, Vol.34, No.7, pp.1572-1587, 2019.
- [67] Nishom, M. “Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square.” *J. Inform*, Vol.4, No.01, 2019.
- [68] Afyouni, Soroosh, Stephen M. et al. “Effective degrees of freedom of the Pearson’s correlation coefficient under autocorrelation.” *NeuroImage* 199, pp. 609-625, 2019.
- [69] Temma, Seiya, Manabu Sugii, et al. “The document similarity index based on the Jaccard distance for mail filtering.” 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC). IEEE, 2019.
- [70] Jo, Jun-Ho. “Development of wine recommendation algorithm using similarity algorithm : Focus on Bigdata analysis techniques.” Master’s Thesis, Namseoul University, Seoul, Korea, 2018.
- [71] “Time Series Analysis-Model.” *blog.naver.com*. last modified n.d., accessed September 2, 2021, <https://m.blog.naver.com/bluefish850/220749045909>.
- [72] Park, Hye Won. “A Study on LSTM Performance in the Prediction of Time Series Model.” Master’s Thesis, Keimyung University, Daegu, Korea, 2019.
- [73] Do-Hyun Kim, et al. “A Study of Shot-Term Load Forecasting Using LSTM

- CNN based on Deep Learning.” The Korean Institute of Electrical Engineers, pp.281-282, 2018.
- [74] Ji-Eun Park, et al. “A Comparison of Growth Curve Models for Forensically Important Flies Using Sigmoid Functions.” Korean J Leg Med, Vol.44, pp.84-91, 2020.
- [75] Hyugwoo Kim, et al. “Estimation of Crematory Facility Demand in Suwon City Using Seasonal ARIMA Model.” J Health Info Stat, Vol.42, No.4, pp.346-354, 2017.
- [76] Benvenuto, Domenico, et al. “Application of the ARIMA model on the COVID-2019 epidemic dataset.” Data in brief 29, 105340, 2020.
- [77] Alghamdi, Taghreed, et al. “Forecasting traffic congestion using ARIMA modeling.” 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC). IEEE, 2019.
- [78] Kristoffersen, Eivind, et al. “Exploring the relationship between data science and circular economy: An enhanced CRISP-DM Process Model.” Conference on e-Business, e-Services and e-Society. Springer, Cham, 2019.
- [79] Exenberger, Emil, Jozef Bucko. “Analysis of online consumer behavior-Design of CRISP-DM process model.” AGRIS on-line Papers in Economics and Informatics 10.665-2021-559, pp.13-22, 2020.
- [80] Prayudani, S., et al. “Analysis Accuracy Of Forecasting Measurement Technique On Random K-Nearest Neighbor (RKNN) Using MAPE And MSE.” Journal of Physics: Conference Series, Vol.1361, No.1, 2019.
- [81] Qi, Jun, et al. “On mean absolute error for deep neural network based vector-to-vector regression.” IEEE Signal Processing Letters 27, pp.1485-1489, 2020.
- [82] Wang, Weijie, Yanmin Lu. “Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model.” IOP conference series: materials science and engineering, Vol.324, No.1, 2018.
- [83] Lin, Jintang, Wenzhong Shi. “Statistical correlation between monthly electric power consumption and VIIRS nighttime light.” ISPRS International Journal of Geo-Information, Vol.9, No.1, pp.32, 2020.
- [84] Saehyun Koh, “Outlier Detection and Imputation Method for Smart Meter

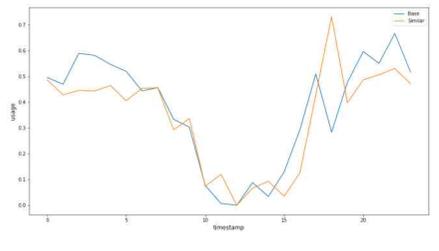
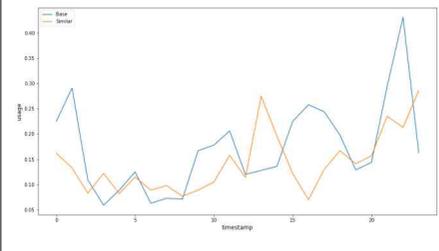
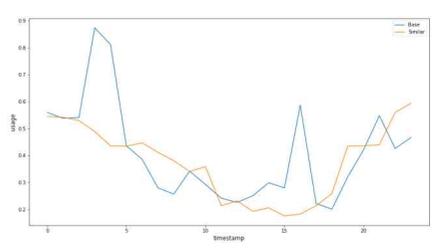
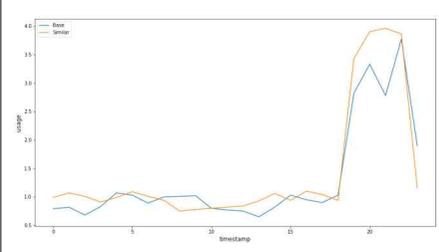
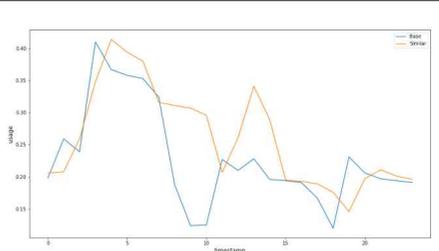
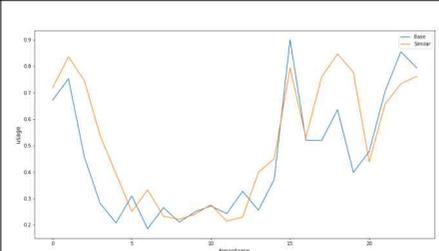
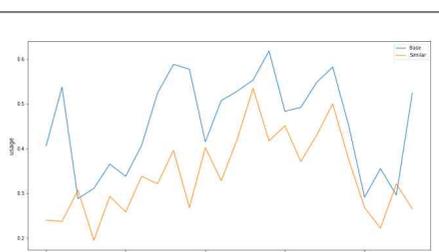
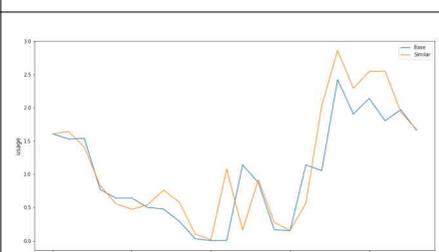
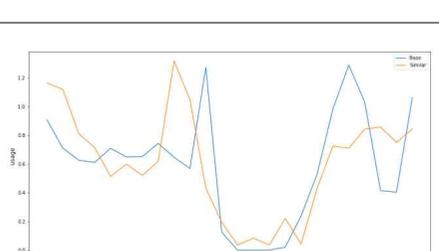
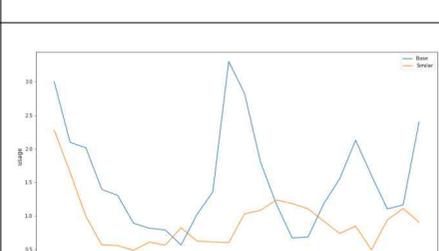
- Data Using Pattern analysis.” Master’s Thesis, Korea University, Seoul, Korea, 2018.
- [85] Dong-Won Shin. “Emotion Classification in Dialogue Utterances Using CNN-LSTM Complex Model.” Master’s Thesis, Korea University, Seoul, Korea, 2017.
- [86] Hwang, Chul-Hyun, “CNN-LSTM Combination Method for Improving Particular Matter Contamination (PM2.5) Prediction Accuracy.” Journal of the Korea Institute of Information and Communication Engineering, Vol 24, pp.57-64, 2020
- [87] Zadid Khan, Sakib Mahmud Khan, et al. “Development and Evaluation of Recurrent Neural Network-Based Models for Hourly Traffic Volume and Annual Average Daily Traffic Prediction.” Transportation Research Record, Vol.2673, No.7, pp.489-503, 2019.
- [88] Donggyun Ku. “Taxi Passengers Demand Prediction using Deep Learning.” Master’s Thesis, The University Of Seoul, Seoul, Korea, 2018.
- [89] Ahn, SungMahn. “Deep Learning Architectures and Applications.” Journal of Intelligence and Information Systems, Vol.22, No.2, pp.127 - 42, 2016.
- [90] Huang, Deqing, et al. “Fault diagnosis of high-speed train bogie based on the improved-CEEMDAN and 1-D CNN algorithms.” IEEE Transactions on Instrumentation and Measurement, Vol.70, pp.1-11, 2021.
- [91] Ji-Wook Kim, et al. “A Study on Fault Classification of Machining Center using Acceleration Data Based on 1D CNN Algorithm.” Journal of the Korean Society of Manufacturing Process Engineers, Vol.18, No.9, pp.29-35, 2019.
- [92] Dongyeon, et al. “A Digital Twin Software Development Framework based on Computing Load Estimati.” JOURNAL OF BROADCAST ENGINEERING, Vol.26, No.4, pp.368-376, 2021.
- [93] Jae-Lyong Jang, et al. “Super-Resolution Reconstruction Algorithm using MAP estimation and Huber function.” The Institute of Electronics Engineers of Korea - Semiconductor and Devices, Vol.46, No.5, pp.39-48, 2009.
- [94] Min-Jae Kang. “Comparison of Gradient Descent for Deep Learning.” Korea Academy Industrial Cooperation Society, Vol.21, No.2, pp.189-194, 2020.
- [95] Gihun Joo. “Performance Evaluation of Machine Learning Optimizers.” urnal

- of IKEEE, Vol.24, No.3, pp.766-776, 2020.
- [96] Lydia, Agnes, Sagayaraj Francis. “Adagrad – an optimizer for stochastic gradient descent.” Int. J. Inf. Comput. Sci, Vol.6, No.5, 2019.
- [97] Kurbiel, Thomas, Shahrzad Khaleghian. “Training of deep neural networks based on distance measures using RMSProp.” arXiv preprint arXiv:1708.01911, 2017.
- [98] Zhang, Zijun. “Improved adam optimizer for deep neural networks.” 2018 IEEE /ACM 26th International Symposium on Quality of Service (IWQoS). IEEE, 2018.
- [99] Chun Yang Zhang. “Analysis of Impact Position based on Deep Learning CNN Algorithm.” Trans.Korean Soc.Mech. Eng.A, Vol.44, No.6, pp.405-412, 2020.
- [100] Junho Jeong, et al. “Avoid Overfitting of R3D Networks using Dropout.” The Korea Institute of Information and Communication Engineering, pp.12-13, 2020.
- [101] Taegon Kim, et al. “Short-term Residential Power Consumption Forecasting using Daily SARIMA(Seasonal Autoregressive Integrated Moving Average) Algorithm.” The Korean Institute of Electrical Engineers, Vol.4, pp.337-338, 2021.
- [102] KIM Jae Hwan, JANG Ki Tae. “Subway Demand Forecast using Seasonal Autoregressive Integrated Moving Average.” Proceedings of the KOR-KST Conference, pp.450-453, 2017.

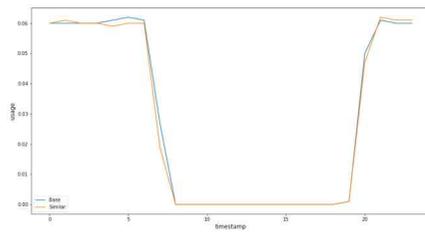
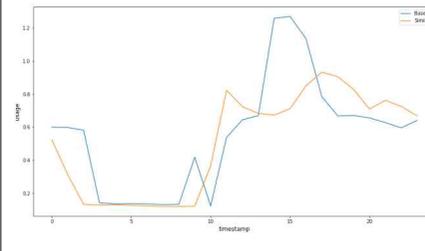
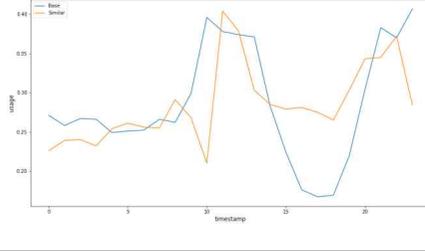
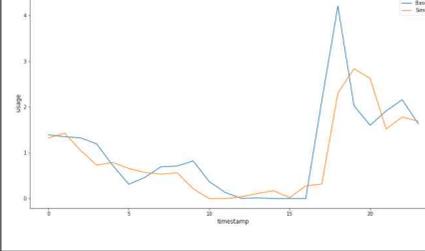
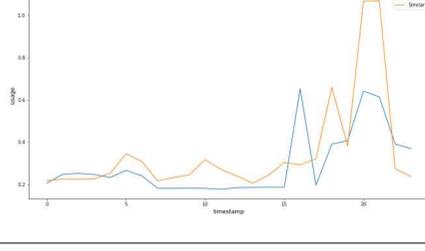
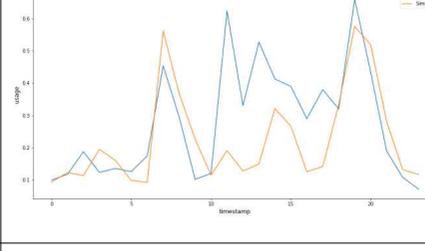
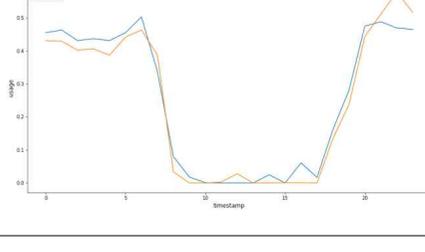
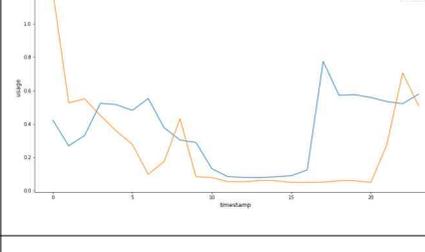
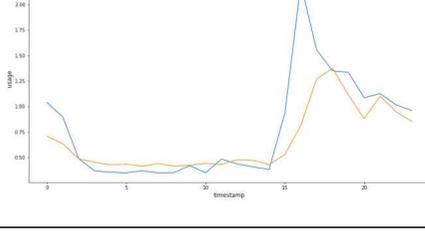
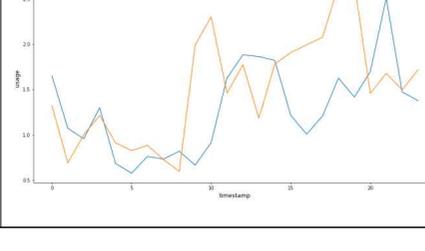
[부록 1] 계량기별 가장 유사한 전력사용량 패턴 탐색 #1

NO	과거 유사패턴	NO	과거 유사패턴
M00001		M00006	
M00002		M00007	
M00003		M00008	
M00004		M00009	
M00005		M00010	

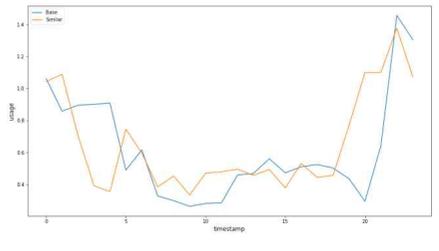
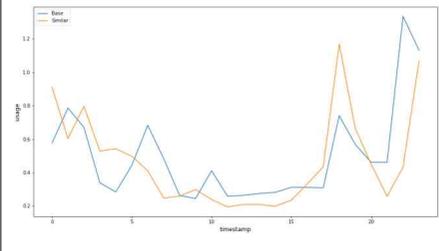
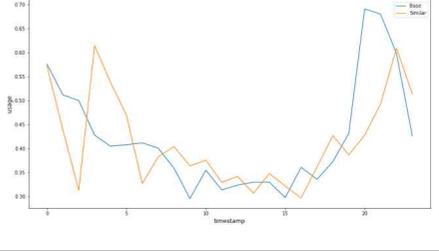
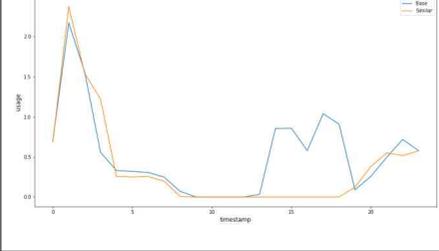
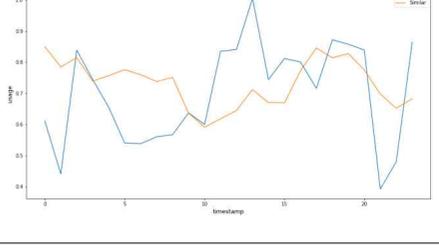
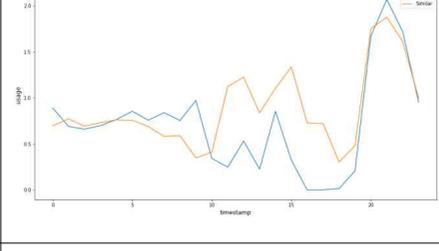
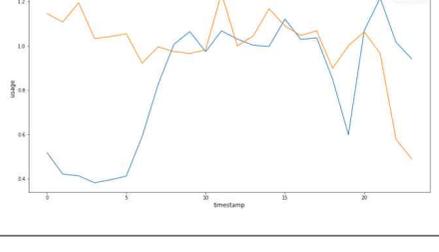
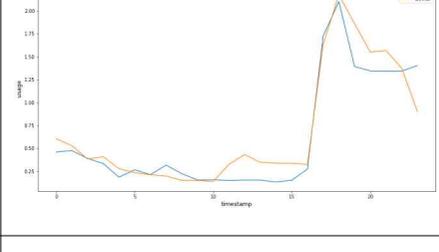
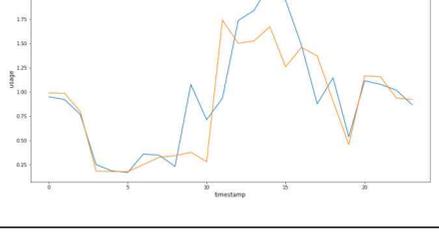
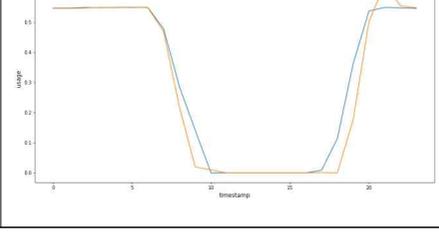
[부록 1] 계량기별 가장 유사한 전력사용량 패턴 탐색 #2

NO	과거 유사패턴	NO	과거 유사패턴
M0011		M0016	
M0012		M0017	
M0013		M0018	
M0014		M0019	
M0015		M0020	

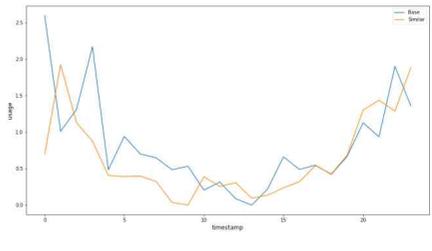
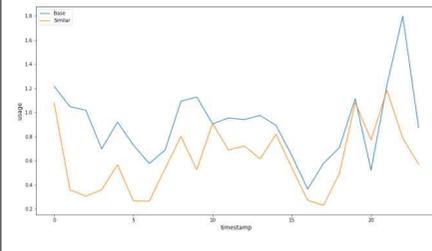
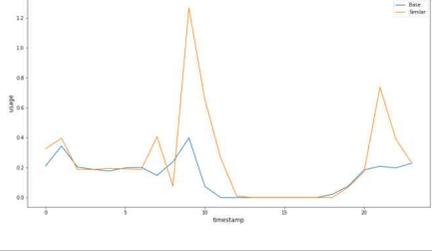
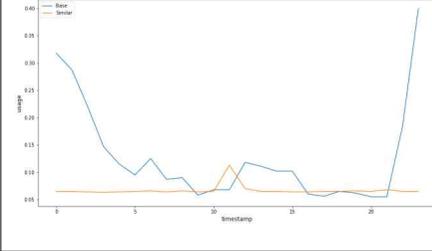
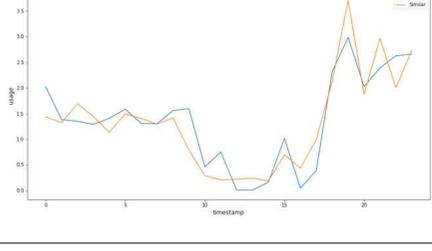
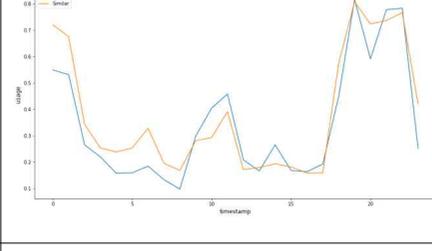
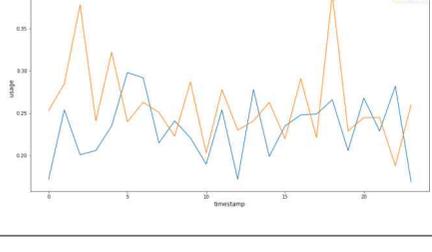
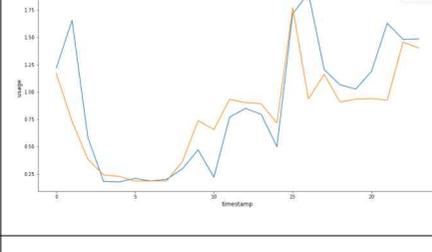
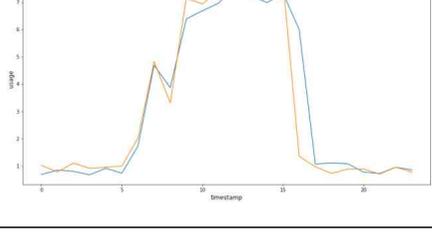
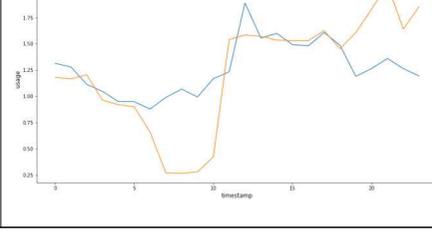
[부록 1] 계량기별 가장 유사한 전력사용량 패턴 탐색 #3

NO	과거 유사패턴	NO	과거 유사패턴
M00221		M00226	
M00222		M00227	
M00223		M00228	
M00224		M00229	
M00225		M00230	

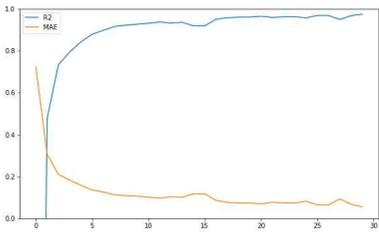
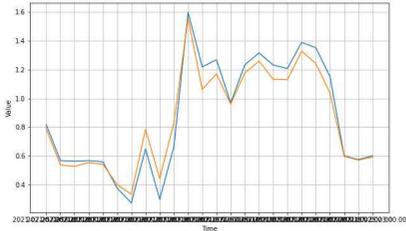
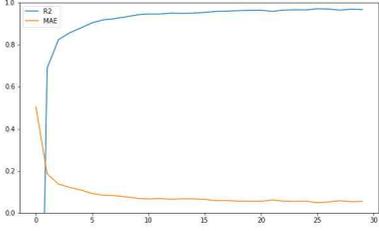
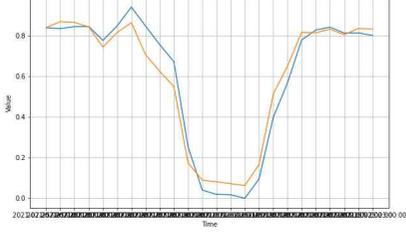
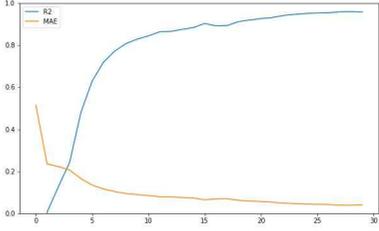
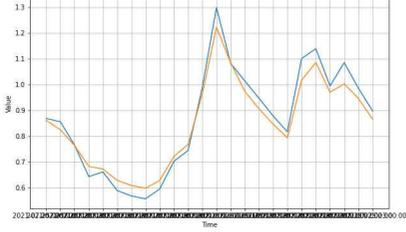
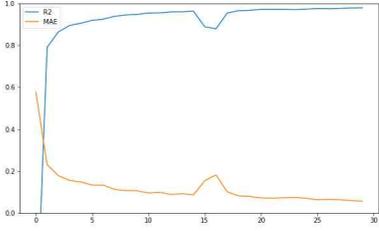
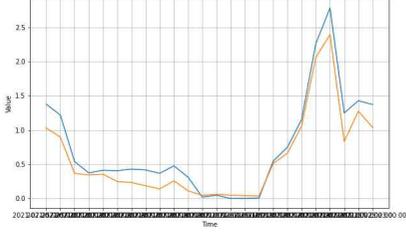
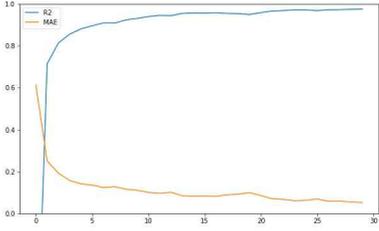
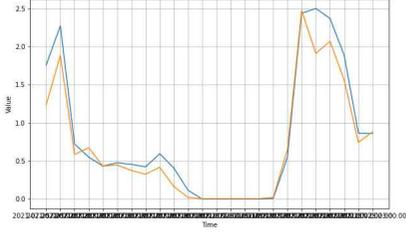
[부록 1] 계량기별 가장 유사한 전력사용량 패턴 탐색 #4

NO	과거 유사패턴	NO	과거 유사패턴
M0031		M0036	
M0032		M0037	
M0033		M0038	
M0034		M0039	
M0035		M0040	

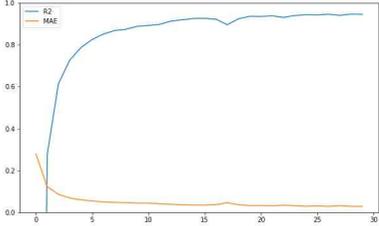
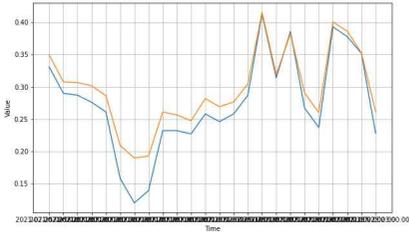
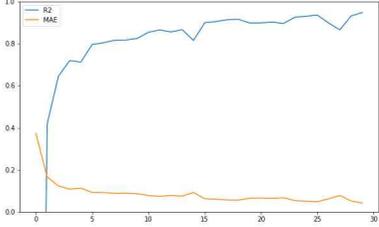
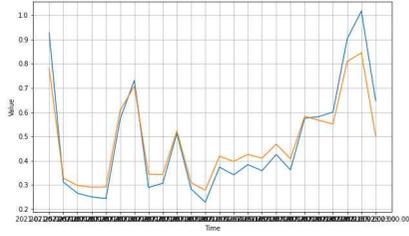
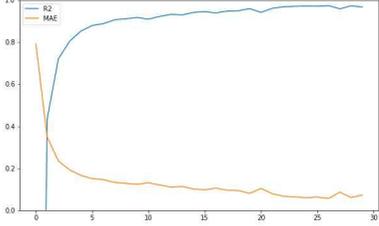
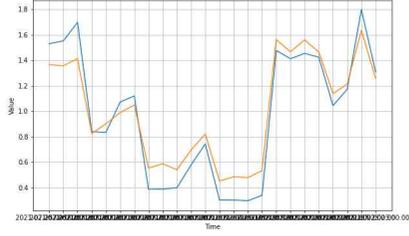
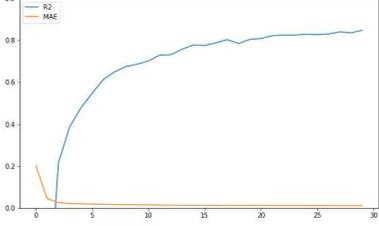
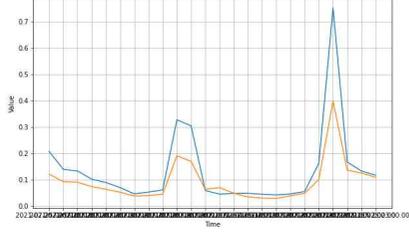
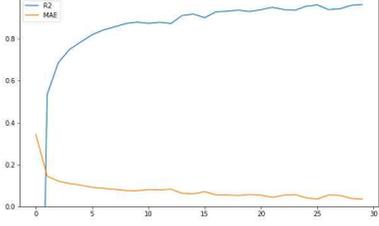
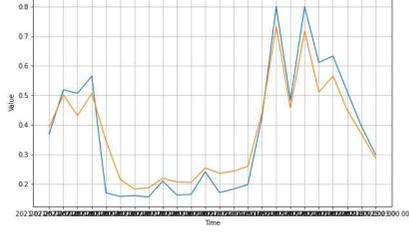
[부록 1] 계량기별 가장 유사한 전력사용량 패턴 탐색 #5

NO	과거 유사패턴	NO	과거 유사패턴
M0041		M0046	
M0042		M0047	
M0043		M0048	
M0044		M0049	
M0045		M0050	

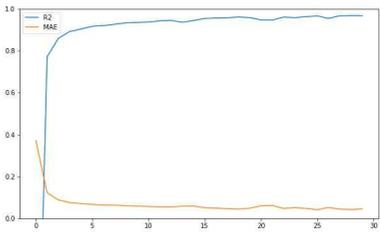
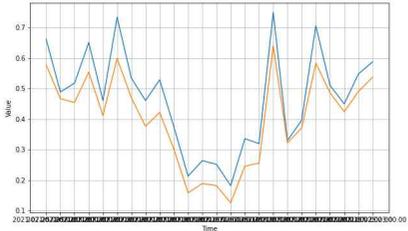
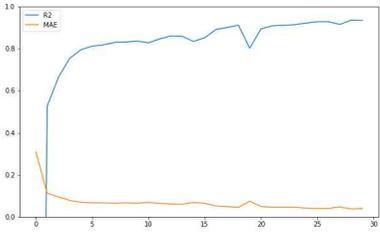
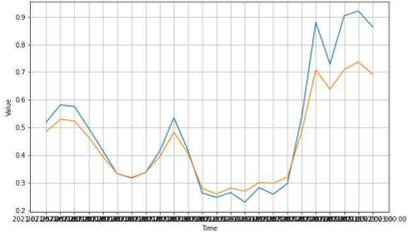
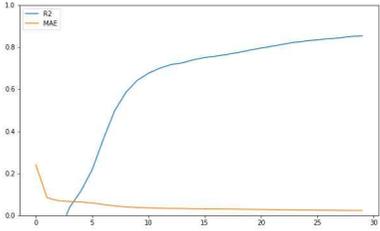
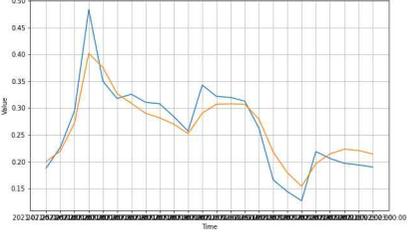
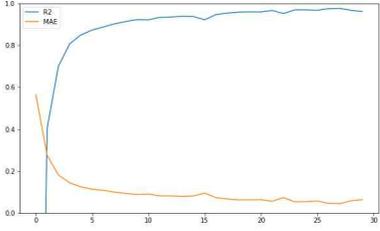
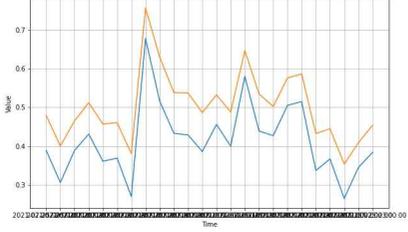
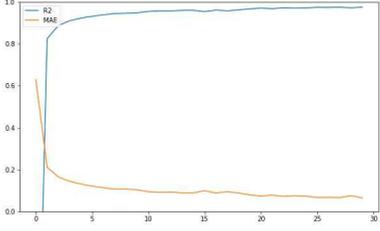
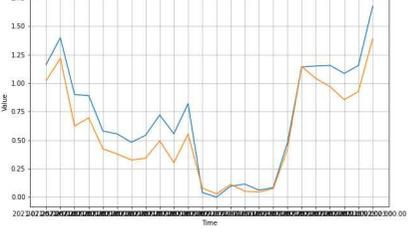
[부록 2] CNN-LSTM결합 딥러닝 모델의 학습과정 및 예측결과 #1

NO	학습과정	시간별 구간 전력사용량 예측 결과
M0001		
M0002		
M0003		
M0004		
M0005		

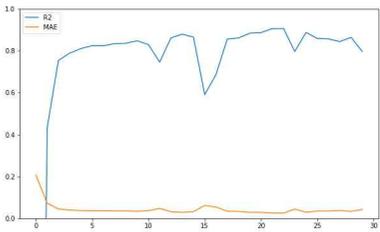
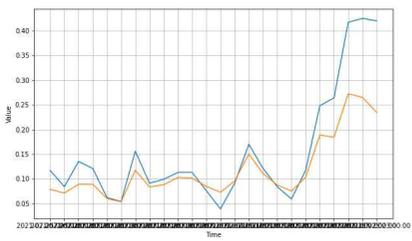
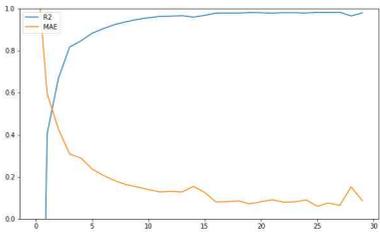
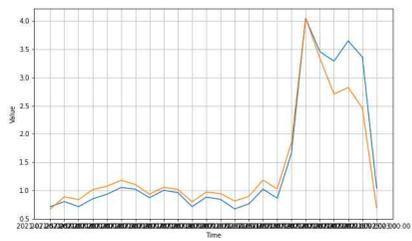
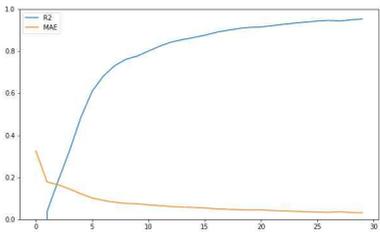
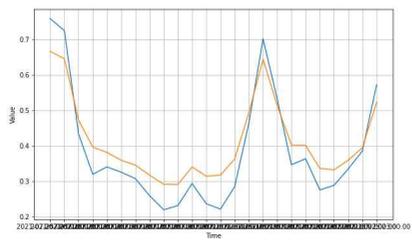
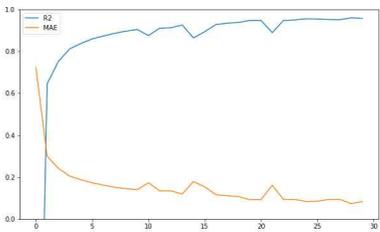
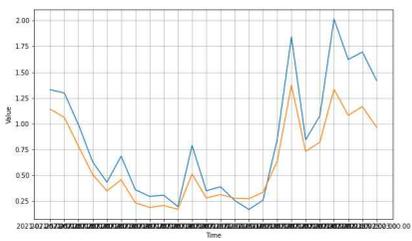
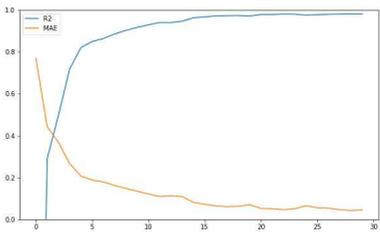
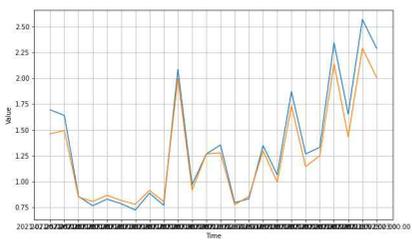
[부록 2] CNN-LSTM결합 딥러닝 모델의 학습과정 및 예측결과 #2

NO	학습과정	시간별 구간 전력사용량 예측 결과
M0006		
M0007		
M0008		
M0009		
M0010		

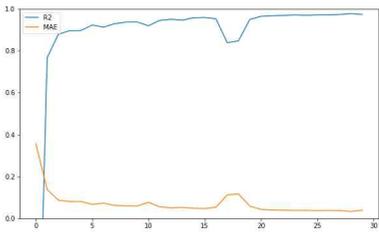
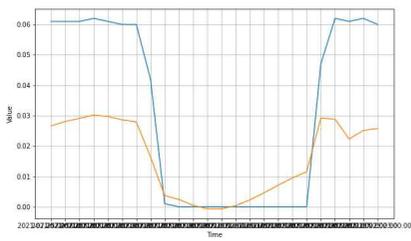
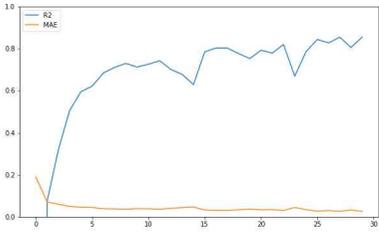
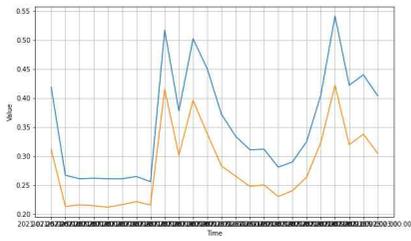
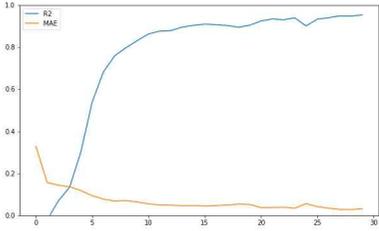
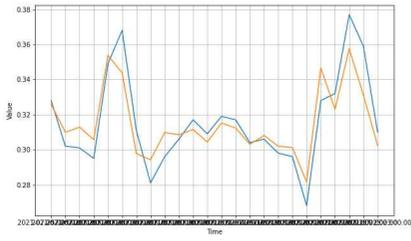
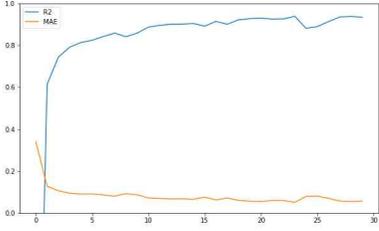
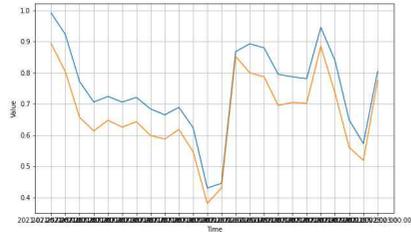
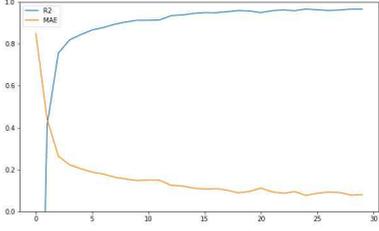
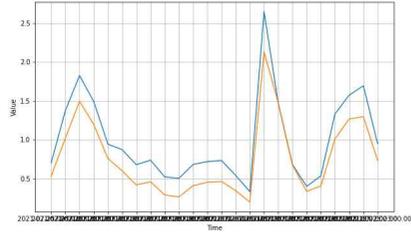
[부록 2] CNN-LSTM결합 딥러닝 모델의 학습과정 및 예측결과 #3

NO	학습과정	시간별 구간 전력사용량 예측 결과
M0011		
M0012		
M0013		
M0014		
M0015		

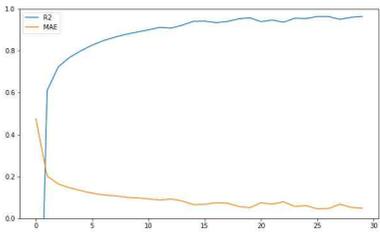
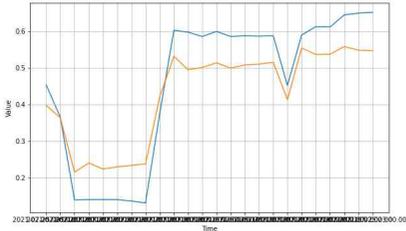
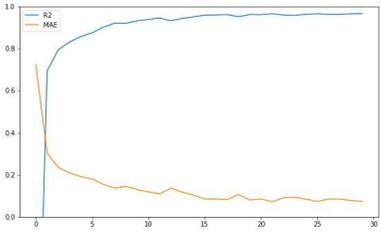
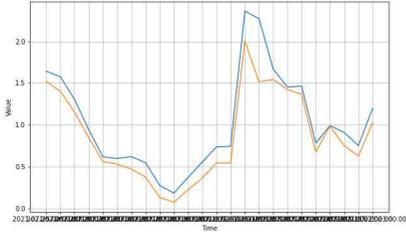
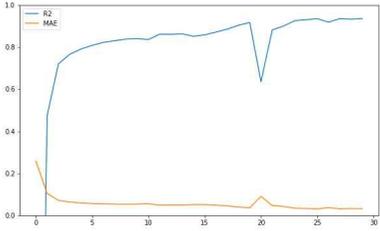
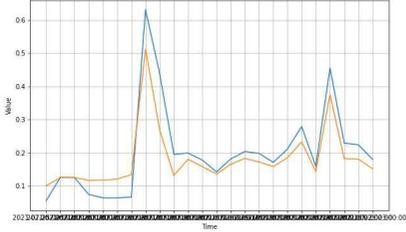
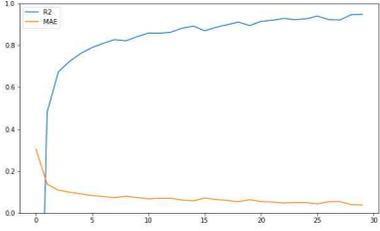
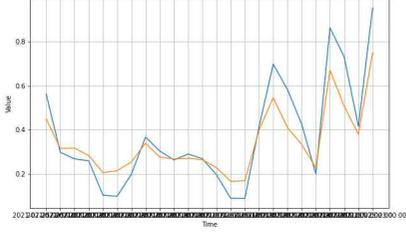
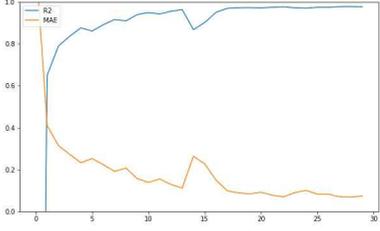
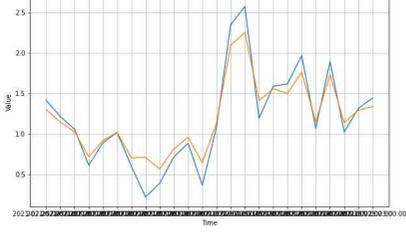
[부록 2] CNN-LSTM결합 딥러닝 모델의 학습과정 및 예측결과 #4

NO	학습과정	시간별 구간 전력사용량 예측 결과
M0016		
M0017		
M0018		
M0019		
M0020		

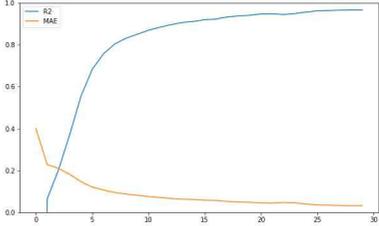
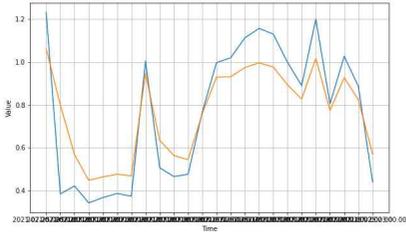
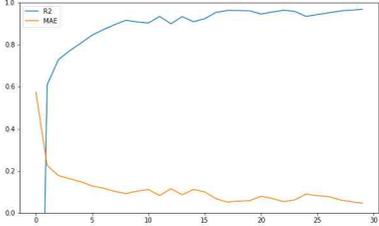
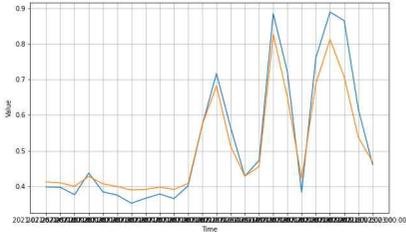
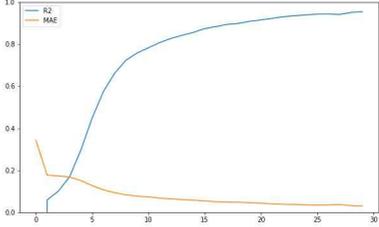
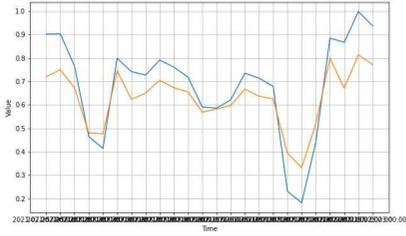
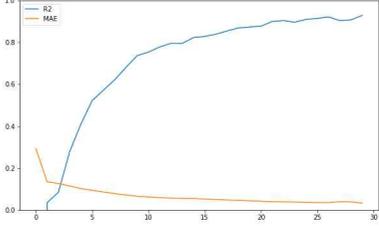
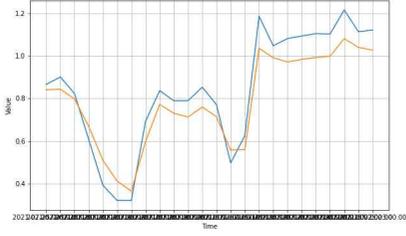
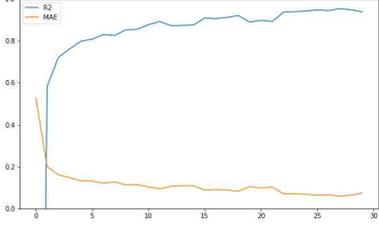
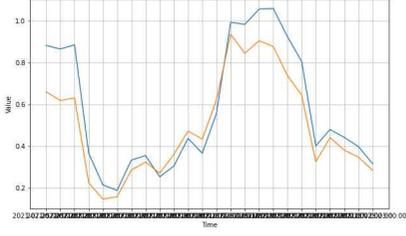
[부록 2] CNN-LSTM결합 딥러닝 모델의 학습과정 및 예측결과 #5

NO	학습과정	시간별 구간 전력사용량 예측 결과
M0021		
M0022		
M0023		
M0024		
M0025		

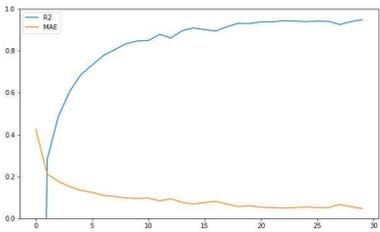
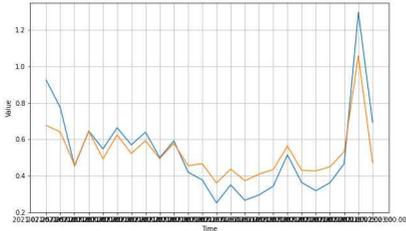
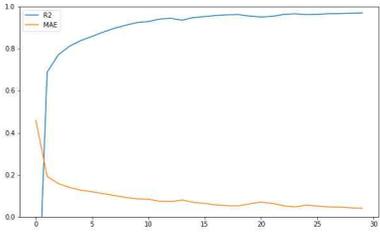
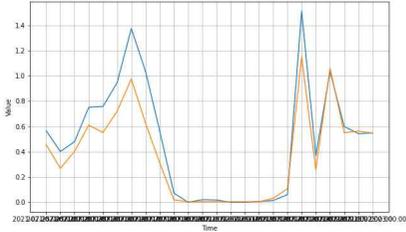
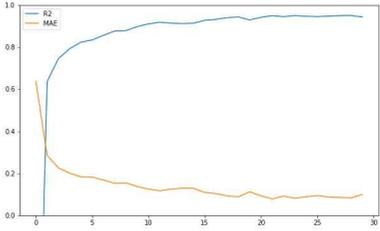
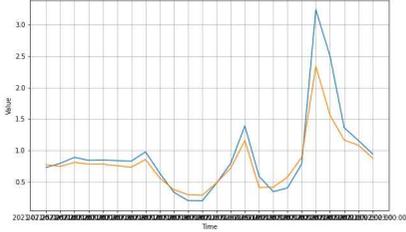
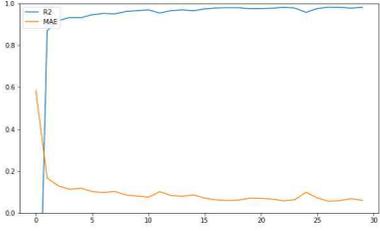
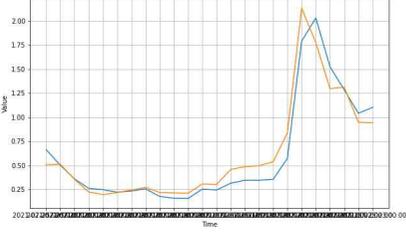
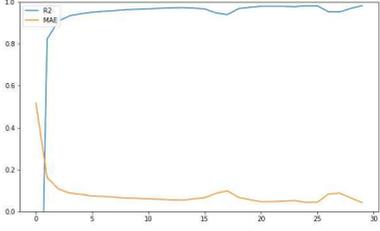
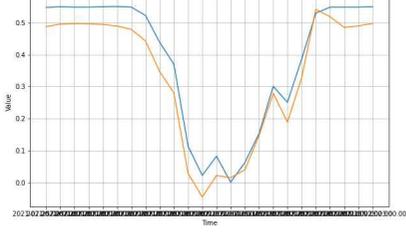
[부록 2] CNN-LSTM결합 딥러닝 모델의 학습과정 및 예측결과 #6

NO	학습과정	시간별 구간 전력사용량 예측 결과
M0026		
M0027		
M0028		
M0029		
M0030		

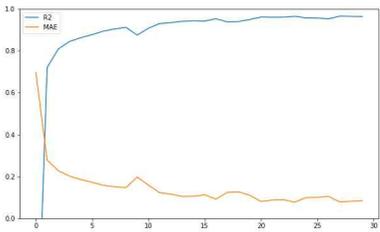
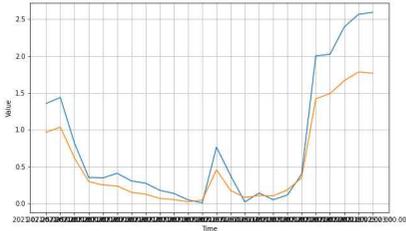
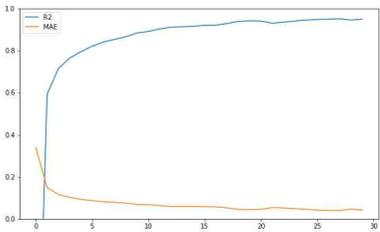
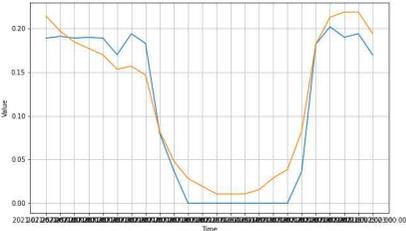
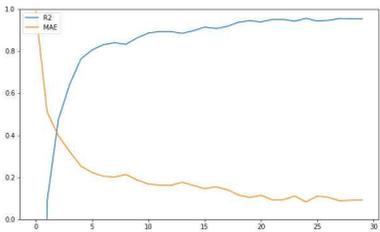
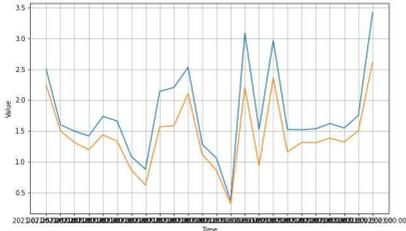
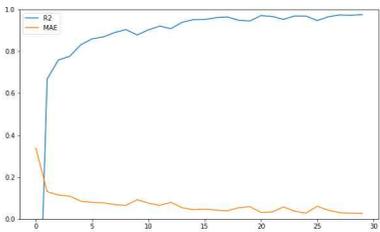
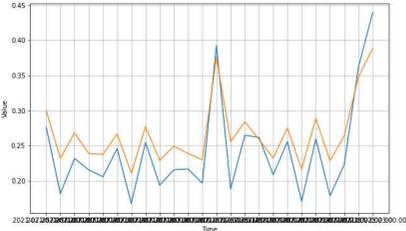
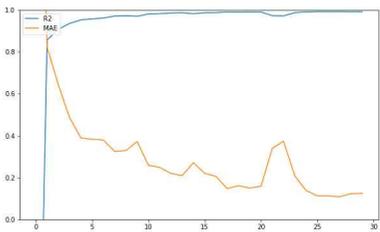
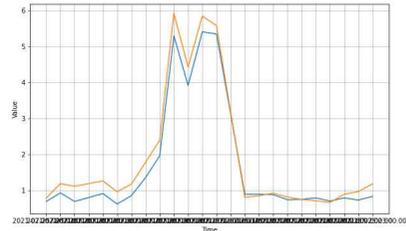
[부록 2] CNN-LSTM결합 딥러닝 모델의 학습과정 및 예측결과 #7

NO	학습과정	시간별 구간 전력사용량 예측 결과
M0031		
M0032		
M0033		
M0034		
M0035		

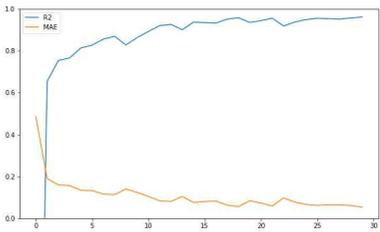
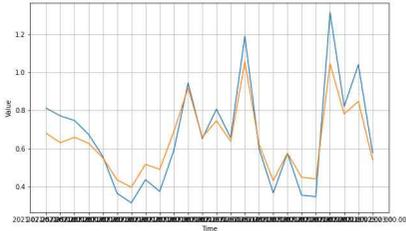
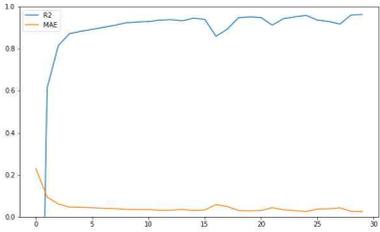
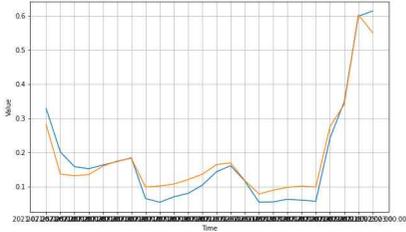
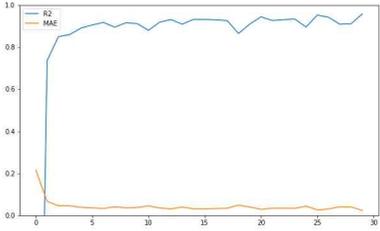
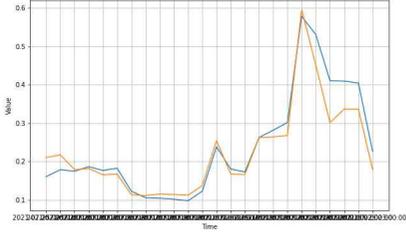
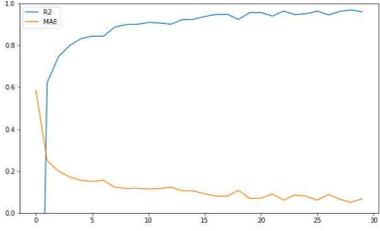
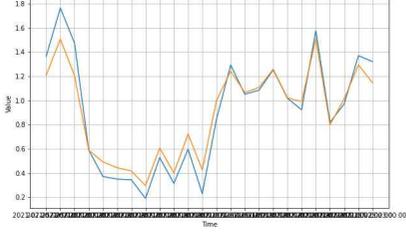
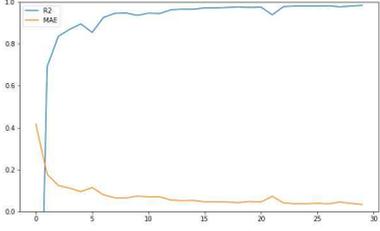
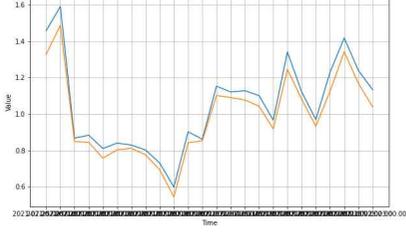
[부록 2] CNN-LSTM결합 딥러닝 모델의 학습과정 및 예측결과 #8

NO	학습과정	시간별 구간 전력사용량 예측 결과
M0036		
M0037		
M0038		
M0039		
M0040		

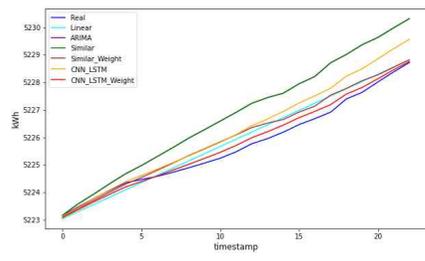
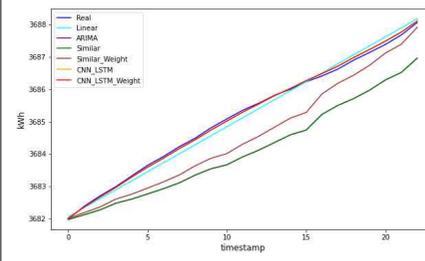
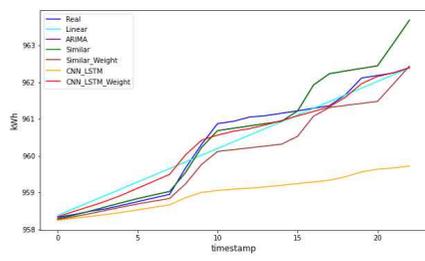
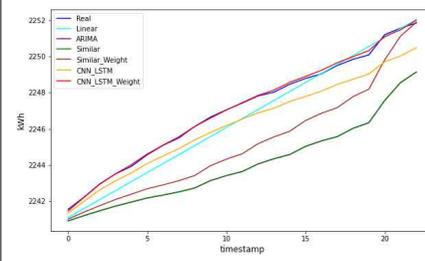
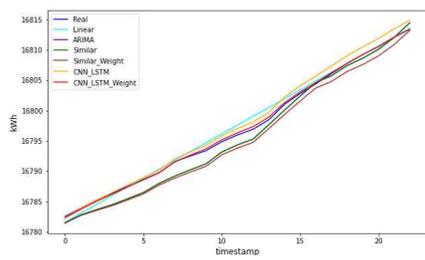
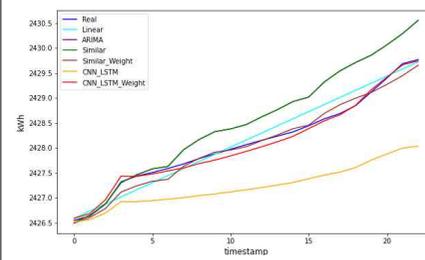
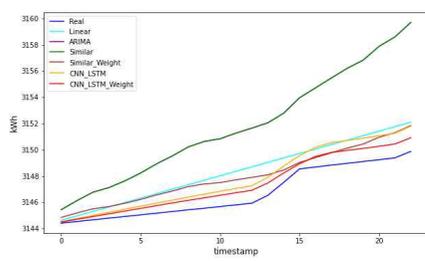
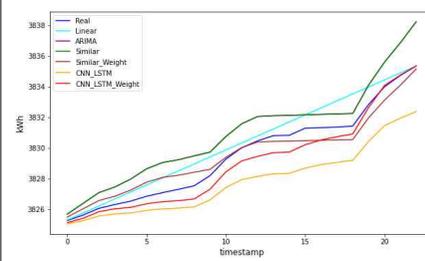
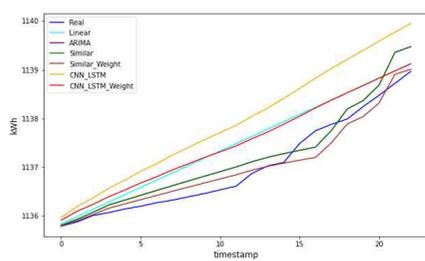
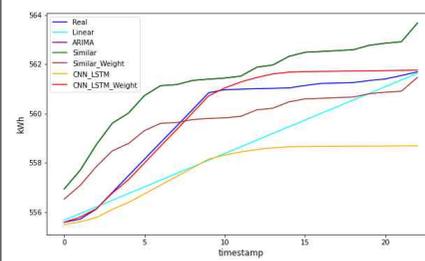
[부록 2] CNN-LSTM결합 딥러닝 모델의 학습과정 및 예측결과 #9

NO	학습과정	시간별 구간 전력사용량 예측 결과
M0041		
M0042		
M0043		
M0044		
M0045		

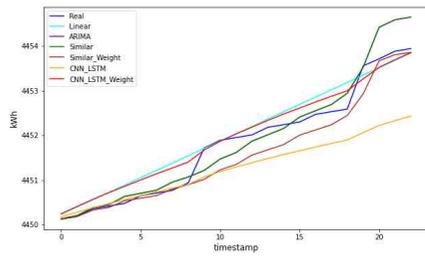
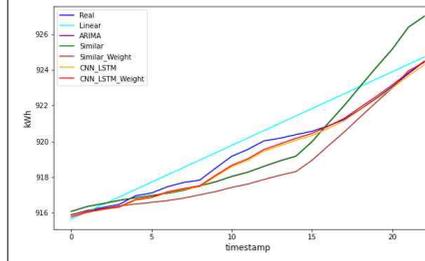
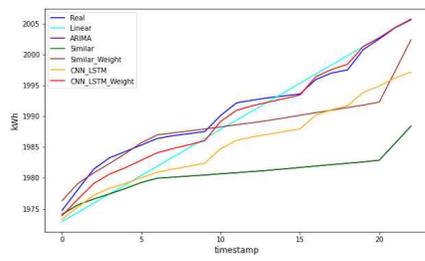
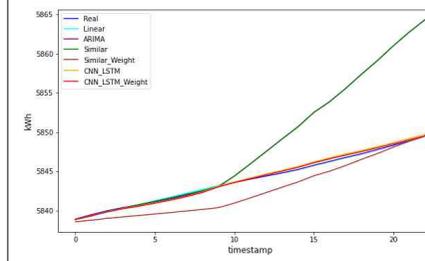
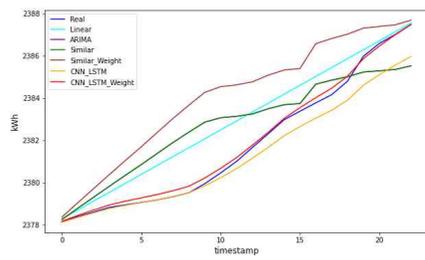
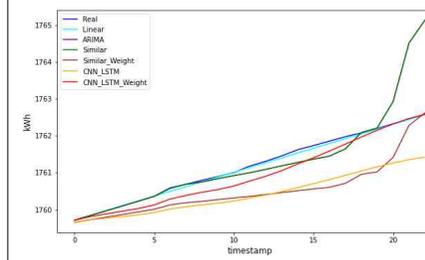
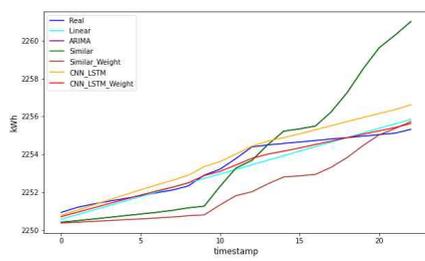
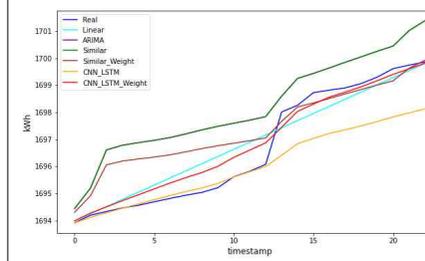
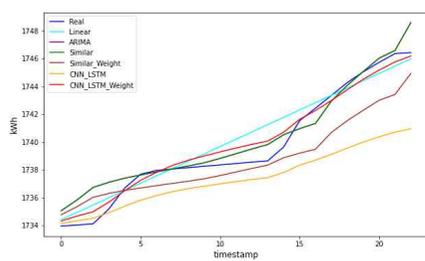
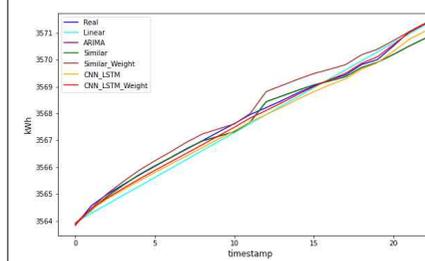
[부록 2] CNN-LSTM결합 딥러닝 모델의 학습과정 및 예측결과 #10

NO	학습과정	시간별 구간 전력사용량 예측 결과
M0046		
M0047		
M0048		
M0049		
M0050		

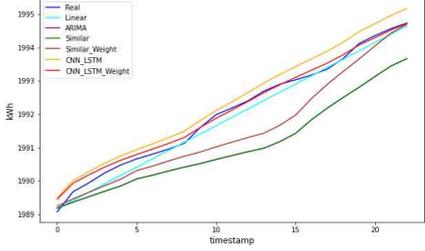
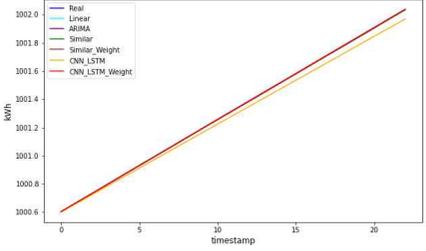
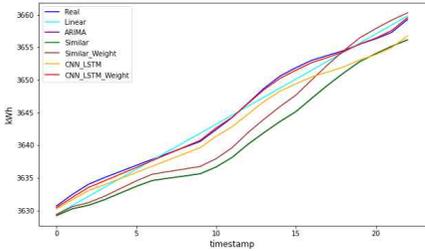
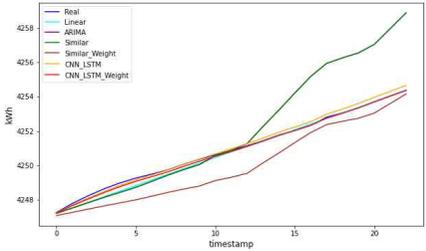
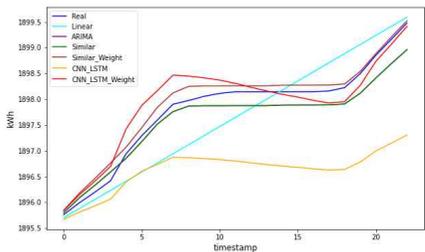
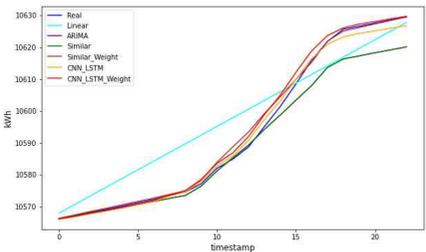
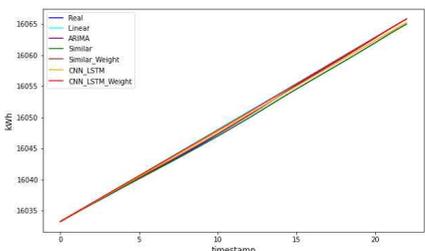
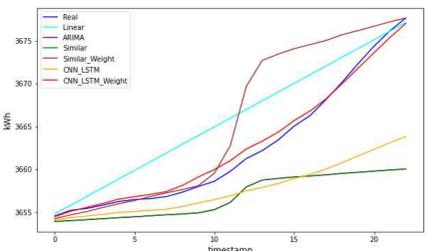
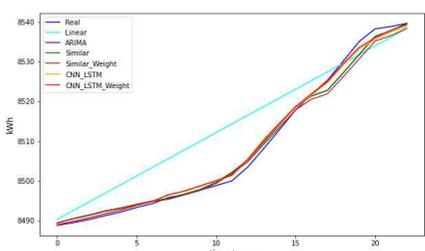
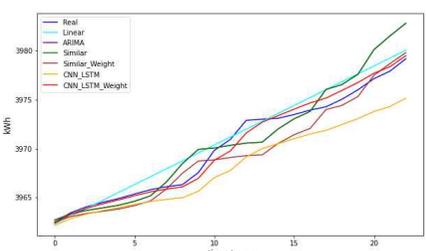
[부록 3] 계량기별 결측 보정 방법별 예측결과와 실제데이터 비교 #1

NO	보정 방법별 예측결과 비교	NO	보정 방법별 예측결과 비교
M0001		M0006	
M0002		M0007	
M0003		M0008	
M0004		M0009	
M0005		M0010	

[부록 3] 계량기별 결측 보정 방법별 예측결과와 실제데이터 비교 #2

NO	보정 방법별 예측결과 비교	NO	보정 방법별 예측결과 비교
M0011		M0016	
M0012		M0017	
M0013		M0018	
M0014		M0019	
M0015		M0020	

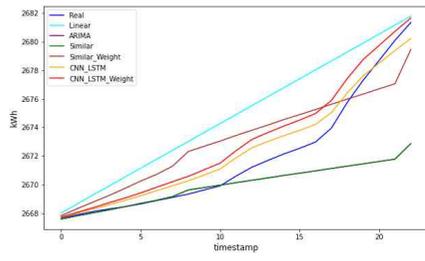
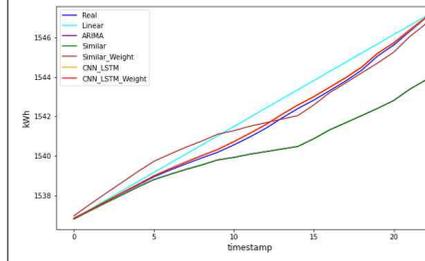
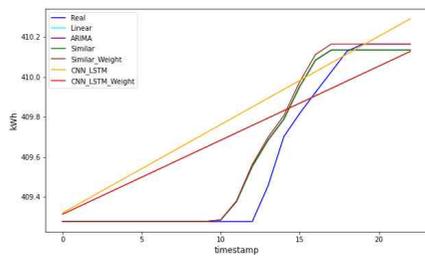
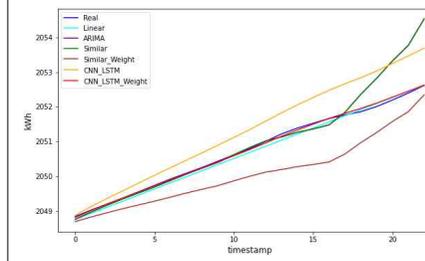
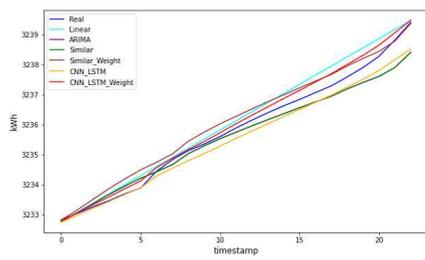
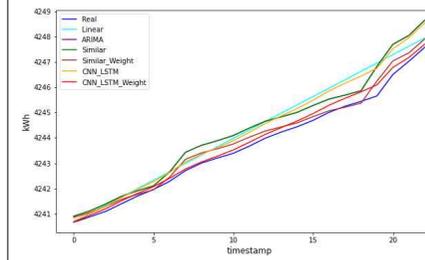
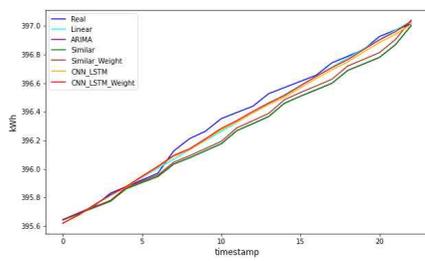
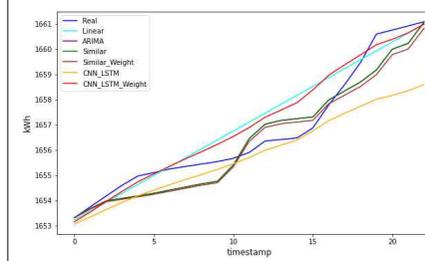
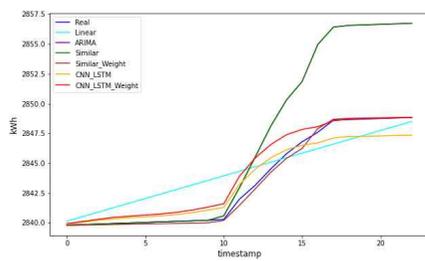
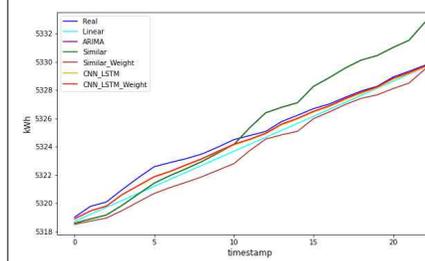
[부록 3] 계량기별 결측 보정 방법별 예측결과와 실제데이터 비교 #3

NO	보정 방법별 예측결과 비교	NO	보정 방법별 예측결과 비교
M0021		M0026	
M0022		M0027	
M0023		M0028	
M0024		M0029	
M0025		M0030	

[부록 3] 계량기별 결측 보정 방법별 예측결과와 실제데이터 비교 #4

NO	보정 방법별 예측결과 비교	NO	보정 방법별 예측결과 비교
M0031		M0036	
M0032		M0037	
M0033		M0038	
M0034		M0039	
M0035		M0040	

[부록 3] 계량기별 결측 보정 방법별 예측결과와 실제데이터 비교 #5

NO	보정 방법별 예측결과 비교	NO	보정 방법별 예측결과 비교
M0041		M0046	
M0042		M0047	
M0043		M0048	
M0044		M0049	
M0045		M0050	

국문초록

딥러닝 기반 전력 계량데이터 결측 보정 모델에 대한 연구

권혁록

지도교수 : 김판구

컴퓨터공학과

조선대학교 대학원

지구 온난화로 폭설, 폭염, 산불, 태풍 등 이상기후 현상이 세계 여러 곳에서 나타나고 있다. 이에 국제사회는 기후변화 문제의 심각성을 인식하고 이를 해결하기 위해 2015년 파리기후변화협정, 2018년 인천 송도에서 개최된 IPCC(Intergovernmental Panel on Climate Change) 총회에서 지구온난화 1.5℃ 목표의 과학적 근거마련 등 2030년까지 이산화탄소 배출량을 2010년 대비 최소 45% 이상 감축하여야 하고, 2050년경에는 탄소중립(Netzero)을 달성하여야 한다는 경로를 제시했다. 우리나라도 2020년 12월 '2050 탄소중립 추진전략'을 마련했다. 여기에 탄소중립 5대 기본방향으로 ①깨끗하게 생산된 전기·수소의 활용 확대 ②디지털 기술과 연계한 혁신적인 에너지 효율 향상 ③탈 탄소 미래기술 개발 및 상용화 촉진 ④순환경제로 지속 가능한 산업 혁신 촉진 ⑤산림, 갯벌, 습지 등 자연·생태의 탄소 흡수 기능 강화 등이 제시됐다. 결국 탄소배출량을 줄이기 위해서 가장 기본이 되는 것이 에너지 사용량을 줄이는 것이다. 에너지 절약을 위해서 AMI구축을 서두르고 AMI 인프라를 통해 소비자와 전력회사 간 양방향통신으로 다양한 부가서비스인 계시별 요금제(TOU; Time Of Use), 수요관리형 선택요금(CPP; Critical Peak Pricing), 실시간 요금제(RTP; Real Time Pricing) 등 고도화된 Time-based 요금제 지원을 통해 수용가의 능동적인 에너지 절감 참여 유도가 가능하다.

한국전력도 2020년까지 2,250만호 전체에 AMI 인프라를 구축하기 위해 노력했지만 약 1,000만호를 구축 운영 상태이며, 앞으로 수년 내에 구축이 완료될 것이다. 또한 정부는 2020년 “한국판 뉴딜” 발표를 통해 그 세부과제인 “에너지관리 효율화

지능형 스마트그리드 구축”에서 아파트 500만호를 대상으로 AMI를 보급하겠다고 발표했다. AMI 보급이 확대되면서 여러 가지 새로운 서비스들이 생겨나고 있고, 이를 통해 정부정책인 탄소중립 목표에 한걸음 더 나아가고 있다.

전력 계량데이터를 활용한 부가 서비스들을 제공하기 위해선 필수적으로 전력량 계로부터 계량데이터들을 잘 취득하여야 한다. 그러나 AMI를 구성하고 있는 통신망의 한계 및 여러 가지 요인으로 인해서 계량데이터의 결측이 2~5% 가량 발생하고 있다. AMI 부가 서비스의 품질을 높이기 위해서는 결측 데이터를 예측하여 제공하는 것이 필요하다. 따라서 스마트미터에서 생성한 시계열 데이터의 값을 예측하기 위해 여러 가지 알고리즘들을 연구하고 적용하고 있다. 하지만 그 정확도가 높지 않고 결측 구간이 길어지면 오차가 커져 양질의 서비스를 제공할 수가 없다.

전력 사용량은 일반적인 시계열 데이터 예측이 아니라 계속 증가하는 누적 전력 사용량 값을 가지고 있고, 따라서 누적 전력사용량 값을 예측해서 보정해야 한다. 지금까지 시계열 분야에서 자주 이용되는 방법들 중에 본 논문에서는 선형보정법, 유사도 기반 보정법, ARIMA(AutoRegressive Integrated Moving Average) 예측 보정법, 장단기메모리순환신경망(LSTM; Long Short-Term Memory) 예측 보정법 등을 선행 연구를 통해 그 장·단점을 파악하였다. 하지만 선행 연구된 보정법들을 누적 전력사용량 예측에 사용하는 것은 적합하지 않았다. 특히 단순히 데이터를 예측 후 보정하면서 일부 데이터들이 결측 후 나타나는 누적 전력사용량 값이 감소되는 오류가 발생했다.

본 논문에서는 이러한 문제들을 해결하기 위해 시계열에 사용되는 일반적인 방법들보다 선형보정법의 장점과 시계열 분야에서 우수한 LSTM 딥러닝 알고리즘을 이용한 보정법의 장점을 결합한 하이브리드 방법으로 딥러닝 기반 CNN-LSTM결합 모형에 가중치를 적용한 모델을 제안하였다.

연구를 수행하기 위해 CRISP-DM(Cross Industry Standard Process for Data Mining) 방법론의 표준에 따라서 업무 이해(Business Understanding), 데이터 이해(Data Understanding), 데이터 준비(Data Preparation), 모델 개발(Modeling), 평가(Evaluation) 순으로 연구를 진행하였다. 선행 연구를 통해서 AMI 인프라를 먼저

이해하고, 데이터 결측이 발생하는 원인, 비율 등을 파악했으며, 데이터 분석을 통해서 전력사용량 데이터들의 특징 및 패턴을 파악하고, 전처리를 통해서 데이터를 준비하였다. 선행 연구된 장단기메모리순환신경망(LSTM; Long Short-Term Memory) 딥러닝 알고리즘을 개선하고자 합성곱신경망(CNN; Convolutional Neural Network)과 장단기메모리순환신경망(LSTM)를 결합한 딥러닝 모형을 만들었고, CNN-LSTM 결합 딥러닝 모형의 예측 결과에 결측 구간의 총량 대비 구간별 비율인 가중치를 적용하여 시간별 구간 전력사용량을 예측하였고, 마지막으로 결측 이전 누적 전력사용량 값에 시간별 구간 전력사용량을 합산하여 최종 누적 전력사용량을 예측하였다.

선행 보정법과 제안된 보정법의 실험 결과를 비교 분석했을 때 평균제곱오차(MSE)는 본 논문에서 제안한 보정법이 '0.218447'로 ARIMA 예측 보정법보다 약 40배 이상 좋았고, 제공된 평균제곱오차(RMSE)도 6배 이상 좋았다. 평균절대비오차(MAPE) 값으로 비교 시에도 본 논문에서 제안한 보정법이 '0.009899'으로 ARIMA 예측 보정법 '0.448682'보다 약 45배 이상으로 정확도가 가장 우수했다. 또한 본 논문에서 제안한 보정법은 결측 구간의 총량 대비 구간별 비율인 가중치를 적용하였기 때문에 누적 전력사용량 값이 감소되는 현상도 발생하지 않았고, 시간별 구간 전력사용량 예측 오차가 많이 발생하더라도 총 사용량 내에서 예측을 하게 되므로 안정성이 매우 높았다.

결측 길이에 따른 실험결과를 분석했을 때 일반적으로 시계열 분야에서 성능이 좋은 여러 가지 방법들보다 선형보정법이 간단하면서도 성능이 좋았다. 결측 구간의 데이터 개수가 7개 미만 일 때 까지는 선형보정법이 정확도가 가장 높았다. 그래서 AMI시스템에 적용 시에는 결측 길이를 먼저 판단하여 결측 길이가 7개 미만 일 경우는 선형보정법을 사용하는 것이 효과적이고, 이상일 경우에는 본 논문에서 제안한 CNN-LSTM결합 가중치적용 보정법을 사용하는 것이 좋을 것이다.

향후, 제안된 모델이 단순히 전력사용량의 시계열 데이터만을 가지고 딥러닝 알고리즘으로 예측하여 보정하였다면, 딥러닝을 학습시키기 위해서 더 많은 입력값을 사용하여 그 정확도를 높이는 연구가 필요하다. 특히 전력사용량은 기후와 그 연관성이 매우 높기 때문에 날씨정보를 추가하여 정확도를 더 높일 필요가 있다.