

논문 2022-59-2-16

공동주택에서 5종 검침데이터의 오류 보정 및 특성 비교

(Error Correction and Comparison of Characteristics for Five Types of Metering Data in Apartment Houses)

김 동 식*, 강 송 희**, 홍 은 영***, 정 영 모****, 정 범 진*****

(Dong Sik Kim, Songhee Kang, Eunyeong Hong, Young Mo Chung, and Beom Jin Chung[©])

요 약

공동주택 환경에서는 데이터집중장치를 통하여 정기적으로 전기에너지, 수도, 온수, 가스, 그리고 난방 사용량의 5종 계량 데이터를 검침한다. 이러한 다양한 검침 데이터를 활용하여 각 검침과 예측의 정확도를 개선할 수 있다. 그러나 각 5종 데이터를 수집하는 과정에서 다양한 통신 오류 등으로 인하여 수집된 데이터에 오류가 발생한다. 본 논문에서는 실제 3개 단지의 공동주택에서 수집한 5종 데이터를 사용하여, 수집하는 과정에서 발생한 오류를 형태별로 분석하고 그러한 오류의 보정 방법을 제시하였다. 또한 보정된 데이터를 사용하여 5종 데이터 각각의 24시간 사용 형태와 상관도를 군집화 방법으로 분석하였다. 결과로부터, 전기데이터의 24시간 패턴은 수도 사용패턴과 비슷한 형태를 가지고 있는데 반하여 난방데이터는 아주 다른 형태를 가지고 있음을 관찰할 수 있었다.

Abstract

In apartment housing environments, five metering data such as the electric energy, water, hot water, gas, and heating data, are periodically collected through a data concentrator unit. These metering data can be used to improve each metering and prediction accuracy. The collected data, however, are usually corrupted by various errors during the data collection process. In this paper, by using the five types of metering data collected from the field of three apartment complexes, we analyzed the errors that occurred during the collection process by types and proposed methods for correcting such errors. In addition, using the corrected data, 24-hour pattern and correlation for each of the five types of data were analyzed through clustering methods. From the results, it is observed that the electricity energy usage has a similar 24-hour pattern with those of water, while the heating usage pattern is quite different from those of the others.

Keywords : Advanced metering infrastructure (AMI), 전기검침, 5종 검침 데이터, 공동주택

I. 서 론

기계식 전력량계를 대체하는 전자식 전력량계가 본격적으로 보급되기 시작하면서 국내 계량산업구조의 변화도 자연적으로 발생하게 되었다. 현재 전자식 전력량

계는 전력회사 공급용 전력량계와 건설사 공급용(또는 민수용) 전력량계로 양분되어 있는데, 건설사 공급용 전력량계는 분양가 가격경쟁력에서 중요한 요구사항이어서 계량기 크기를 최소화하고, 원격검침 기능이 제공되는 계량 시스템이 주로 공급되기 시작하였다. 또한

*정회원, 한국외국어대학교 전자공학과(Department of Electronics Engineering, Hankuk University of Foreign Studies)

**비회원, 스타라드비전(StradVision)

***비회원, 몬드리안에이아이(Mondrian AI)

****정회원, 한성대학교 전자정보공학과(Department of Electronics and Information Engineering, Hansung University)

*****정회원, 서울과학기술대학교 전기정보기술연구소(Research Center of Electrical and Information Technology, Seoul National University of Science & Technology)

© Corresponding Author(E-mail : bjchung@seoultech.ac.kr)

※ 이 논문은 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NO. NRF-2019R1A6A1A03032119).

Received : October 18, 2021

Revised : November 16, 2021

Accepted : December 15, 2021

전기 외에 수도, 온수, 난방, 가스의 5종 검침이 가능한 통합검침 시스템으로 계량 및 검침 환경이 확대되었다. 현재 국내에서 분양되는 대부분의 아파트는 이러한 5종 또는 3종(전기, 수도, 가스)통합검침 시스템이 적용되고 있다.

최근 들어, 데이터의 중요성이 부각되면서 데이터 중심 사회로의 진입을 위해서 에너지 분야도 에너지 데이터의 수집, 저장, 관리, 공유 및 폐기 등의 에너지 데이터 플랫폼화 연구가 진행되고 있다^[1, 2]. 또한 이와 함께 탄소중립 및 에너지 전환을 위해서 가정용 에너지에 대한 수요예측 및 최적화 관리 기술에 대한 정책적·사회적 요구도 증대되고 있다^[3]. 현재 아파트에서 수집되는 5종 에너지 데이터는 사용자에게 에너지 사용정보 및 요금정보를 제공하고 있으나, 에너지 데이터 공유를 위한 체계는 매우 미비한 실정이다. 이는 기본적으로 아파트에서 수집, 관리되는 계량데이터는 거의 에너지 요금부과용으로 아파트 내부적으로만 활용되기 때문이다.

에너지 데이터 시대 진입을 위해서 지난 10여 년 이상 아파트에서 내부적으로만 수집됐던 5종의 에너지 데이터를 활용할 수 있을 경우, 기술적, 산업적, 사회적으로 매우 유용한 데이터 자원화가 가능할 것이다.

한편, 그동안 아파트 원격검침과 관련한 연구로는 아파트 현장에서의 통신 방법과 관련된 사항들이 주로 연구됐다^[4~7]. 아파트의 원격검침 데이터를 데이터 자원화하기 위해서는 아파트 원격검침 데이터의 품질 수준을 조사·분석하여 데이터 자원화가 가능하도록 오류를 보정해주어야 한다. 최근 들어서 에너지 데이터 활용의 중요성이 증대됨에 따라, 공동주택에서의 데이터 오류 보정 및 데이터 활용과 관련한 연구가 수행된 바 있다^[8~12]. 이를 위해서 본 논문에서는 세 곳의 실거주 아파트에서의 세대별 5종 계량데이터를 수집·저장·관리하는 과정에 발생하는 데이터 처리 오류의 유형을 분석하였다. 5종 계량데이터의 오류들을 관찰 분석하여 총 여섯 가지의 오류 형태로 분류하였다. 이를 통해서 오류 형태에 따른 보정을 수행하였으며, 군집화 알고리즘을 적용하여 5종 검침 데이터의 평균 소비 형태를 도시하고 시간별 소비패턴 변화에 대해 관찰·분석하였다.

본 논문의 구성은 다음과 같다. 제 II장은 공동주택에서 원격검침으로 수집된 5종 데이터에서 발생하는 오류들을 정의하고 그에 따른 보정 방법을 서술한다. 제 III장은 오류 유형별 보정 작업이 완료된 5종 데이터의 평균 소비패턴을 관찰하기 위하여 군집화 방법으로 24시간 소비패턴을 분석한다. 마지막 장에서는 논문의 결론

을 정리한다.

II. 데이터 오류 형태 및 보정

본 장에서는 공동주택에서 검침한 5종 데이터에서 발생하는 오류들을 분석하고 보정 방법에 관해 기술한다.

실험을 위하여 김포, 광고, 동탄의 세 공동주택 단지에서 검침한 5종 계량데이터를 사용하였다. 김포 공동주택은 15개 동 총 1,007호 수를 가지며 평균 126m²의 면적을 가지는 공동주택이다. 여기서 획득한 데이터는 1시간 단위로 검침한 5종 데이터로, 수집 기간은 2017년 4월 1일 0시부터 2019년 11월 6일 13시까지 총 31개월이다. 해당 데이터에는 날짜, 시간, 동, 호수, 데이터의 적산량, 시간별 사용량에 대한 정보가 있다. 광고의 실증현장 아파트는 5개 동 총 375개 호수를 가진 평균 39.9m²의 면적을 가지는 공동주택이다. 여기서 획득한 데이터는 1시간 단위로 검침된 5종 데이터(다만, 가스 사용량은 모두 0으로 검침됨)로, 2019년 8월 30일 14시부터 2021년 05월 20일 10시까지 총 21개월까지의 데이터이다. 동탄 실증현장 아파트는 6개 동 총 608개 호수를 가진 평균 25.2m²의 면적을 가지는 공동주택이다. 여기서 획득한 데이터는 1시간 단위로 검침한 5종 데이터(다만, 가스 사용량은 모두 0으로 검침 됨)로, 2019년 1월 1일 0시부터 2021년 05월 20일 11시까지 총 29개월까지의 데이터이다.

A. 검침데이터오류 및 보정

유선 또는 무선으로 데이터를 원격 검침하는 과정에서 데이터를 수신했지만 수신된 데이터에 오류가 발생하여 데이터의 기본 형태가 어긋나게 된다. 본 논문에서는 이러한 오류를 검침데이터오류라 부르기로 한다. 김포 데이터에서 발생하는 해당 오류에는 검침 누적량 앞자리의 숫자 1이 9로 바뀌는 경우가 존재한다. 표 1과 같이 2019년 9월 23일 22시의 전기 누적량 앞자리 수는 21시의 누적량과 관계없이 9로 바뀐 것을 확인할 수 있으며 23시에는 다시 1로 바뀌어 누적되는 것을 확인할 수 있다. 해당 오류는 유선통신규격 ISO488 등을 통해 공동주택에서 데이터를 수집할 때 발생하는 오류로 이진 데이터 0001에서 1bit 오류 등이 발생하여 1001로 바뀌는 경우로 추정된다. 일반적으로 이러한 1bit 오류는 전송 과정에서 패리티 비트 체크로 확인이 가능하다. 또한, 이러한 오류는 이전 수신한 누적량 데이터를 사용하여 검출 및 복원이 가능하며, 김포 데이터에서

약 0.008ppm 정도의 빈도로 매우 드문 경우이다. 보정 방법은 표 1과 같이 9로 잘못 표기된 맨 앞자리 수를 1로 수정하고 큰 음수 값으로 잘못 표기된 사용량의 값을 0으로 1차 보정하였다. 사용량의 값에 대한 오류는 추후에 다시 보정한다 (E 오류 참조).

표 1. 검침데이터오류 예와 1차 보정(김포)
Table 1. Example of metering data error and correction (Gimpo).

Before correction			
Y-M-D	hour	electricity accumulation	electricity usage
2019-09-23	21	19930.7	0.7
2019-09-23	22	99932.2	80001.5
2019-09-23	23	19932.5	-79999.7
2019-09-24	0	19932.8	0.3
After correction			
Y-M-D	hour	electricity accumulation	electricity usage
2019-09-23	21	19930.7	0.7
2019-09-23	22	19932.2	80001.5
2019-09-23	23	19932.5	0
2019-09-24	0	19932.8	0.3

광고 데이터의 검침데이터오류에서는 누적량 데이터의 마지막 자리 숫자가 오른쪽으로 이동하여 그 다음 시간대의 사용량이 마이너스 값으로 나오는 경우로, 표 2의 예와 같이 2021년 3월 30일 16시-2021년 3월 30일 20시의 데이터는 21년 3월 30일 15시의 전기 누적량 4478.7의 자릿수가 이동하여 검침 되는 것을 확인할 수 있다. 오류 보정 방법은 오류가 발생한 해당 구간의 마이너스 사용량을 0으로 보정하고 자리 이동된 누적량의 데이터를 보정된 사용량에 맞춰 수정하였고, 오류가 발생하기 전 데이터(2021년 3월 30일 15시)의 누적량과 오류가 발생한 기간의 다음 시간 데이터(2021년 3월 30일 21시) 누적량의 차이를 구하여 2021년 3월 30일 21시의 전기 사용량을 보정하였다.

동탄 데이터에서는 검침데이터오류로 사용량이 마이너스 값이 나타나는 경우가 있는데, 마이너스 사용량 값을 0으로 보정해주었다. 해당 오류의 발생 빈도는 0.0001%로, 해당 오류로 인해 발생하는 2020년 6월 26일 15시의 누적량 예러는 누적량역전오류에서 수정하는 방법으로 보정하였다.

표 2. 검침데이터오류 예와 보정(광고)
Table 2. Example of metering data error and correction (Gwanggyo).

Before correction			
Y-M-D	hour	electricity accumulation	electricity usage
2021-03-30	15	4478.7	0
2021-03-30	16	44.7	-4434
2021-03-30	17	4.4	-40.3
2021-03-30	18	0.4	-4
2021-03-30	19	0	-0.4
2021-03-30	20	0	0
2021-03-30	21	4480.2	4480.2
2021-03-30	22	4480.7	0.5
After correction			
Y-M-D	hour	electricity accumulation	electricity usage
2021-03-30	15	4478.7	0
2021-03-30	16	4478.7	0
2021-03-30	17	4478.7	0
2021-03-30	18	4478.7	0
2021-03-30	19	4478.7	0
2021-03-30	20	4478.7	0
2021-03-30	21	4480.2	1.5
2021-03-30	22	4480.7	0.5

B. 신호누적오류 및 보정

김포 공동주택의 검침 데이터를 관찰하면 데이터를 해당 시간에 검침하여 수신하였지만, 누적량에 변화가 없고 따라서 계산된 사용량이 연속해서 0인 경우가 존재한다. 이러한 오류를 신호누적오류라고 부른다. 표 3에서와같이 2019년 3월 24일 오전 7시 이후 데이터의 사용량은 0으로 유지되었다가 2019년 4월 17일 오전 9시에는 444.6이라는 큰 값이 나오는 것을 확인할 수 있다. 이러한 오류는 어느 기간 동안 누적량 데이터를 수신하지 못하는 경우 발생할 수 있다. 이러한 오류는 약 0.001% 정도로 발생한다. 이에 따른 보정 방법은 해당 기간의 지난달과 다음 달의 같은 날짜의 시간대의 사용량 평균을 비율로 고려해 누적된 사용량을 분배하였다.

신호누적오류의 다른 형태도 존재하는데, 검침된 누적량은 변화가 없이 일정한데 계산된 사용량에 이상이 있는 경우로, 해당 오류는 사용량을 보정하지 않고 삭제하였다. 표 4에서와 같이 2017년 11월 24일 12시부터 2017년 11월 27일 13시까지 가스 사용량이 0의 값을 가지고 있다가 2017년 11월 27일 14시부터 2017년 11월

표 3. 신호누적오류 예 1과 보정(김포)
Table 3. Example 1 of signal accumulation error (Gimpo).

Before correction				
Y-M-D	hour	electricity accumulation	electricity usage	
2019-03-24	7	32060.2	0.6	
2019-03-24	8	32060.2	0	
2019-03-24	9	32060.2	0	
...				
2019-04-17	8	32060.2	0	
2019-04-17	9	32504.8	444.6	
After correction				
Y-M-D	hour	electricity accumulation	electricity usage	average usage ratio
2019-03-24	7	32060.2	0.6	
2019-03-24	8	32061.4866	0.7624	0.00171
2019-03-24	9	32062.0107	0.52415	0.00117
...				
2019-04-17	8	32504.1074	0.6671	0.0015
2019-04-17	9	32504.8	0.6926	0.00155

27일 17시까지 가스 사용량의 평균값보다 훨씬 큰 값이 나타나는 것을 확인할 수 있다. 해당 오류의 발생 빈도는 한 개의 호수에서 발생하여 약 0.004ppm 정도로 발생한다. 보정 방법은 표 2의 예에서 기술한 신호누적오류 보정 방법과 동일하다.

광고 데이터의 경우 이러한 형태의 오류는 발견되지 않았다. 동탄 데이터에서 발생하는 신호누적오류는 김포의 경우와 다르게 특정 날짜에서 다양하게 발생하였다. 보정 방법은 앞에서 서술한 김포 공동주택의 신호누적오류 보정 방법과 동일하며 해당 오류의 발생 빈도는 10ppm이다.

표 4. 신호누적오류 예 2와 보정(김포)
Table 4. Example 2 of signal accumulation error and correction (Gimpo).

Y-M-D	hour	electricity accumulation	electricity usage
2017-11-24	11	21707.3	0.1
2017-11-24	12	21707.3	0.0
2017-11-24	13	21707.3	0.0
...			
2017-11-27	14	21707.3	7.1
2017-11-27	15	21707.3	11.1
2017-11-27	16	21707.3	12.1
2017-11-27	17	21707.3	10.4

C. 누적량초기화오류 및 보정

전자식 계량기는 실시간으로 사용량을 측정하고 사용량을 내부 기억장치에 누적하여 저장하는 장치이다. 해당 오류는 일정 검침 주기마다 누적량을 수집할 때 계량기의 오류로 내부 기억장치에 저장된 누적량이 초기화되어 0으로 바뀌는 경우를 의미한다. 본 논문에서는 이러한 오류를 누적량초기화오류라고 부르기로 한다. 표 5와 같이 2019년 6월 25일 16시의 누적량은 0이 되고 따라서 사용량이 -652의 음수 값이 되었음을 확인할 수 있다. 해당 오류의 발생 빈도는 약 0.18%로 빈도수가 상대적으로 높다. 해당 오류의 보정 방법은 음수 값이 발생한 시간대 사용량 값을 0으로 보정하고 누적량은 이전 시간대의 누적량을 더하여 보정하였다. 이때 누적량이 적어지는 경우가 발생하지 않도록 보정하였다. 광고와 동탄 공동주택 데이터의 경우 이러한 형태의 오류는 발견되지 않았다.

D. 누적량역전오류 및 보정

원격검침에서 전자식 계량기는 현재까지 사용한 양의 누적량을 송신하므로, 최소한 검침 누적량은 같거나 이전 값보다 커야 한다. 그러나 검침 결과를 보면, 현재 누적량이 이전 누적량의 값보다 작은 경우가 있다. 표 6의 보정 전 내용을 확인해보면 2018년 3월 7일 13시 사용량(784.5)이 14시 사용량(784)보다 큰 경우를 관찰

표 5. 누적량초기화오류 예와 보정(김포)
Table 5. Example of accumulation reset error (Gimpo).

Before correction			
Y-M-D	hour	gas accumulation	gas usage
2019-06-25	15	652	0
2019-06-25	16	0	-652
2019-06-25	17	0	0
...			
2019-06-27	18	0	0
2019-06-27	19	1	1
After correction			
Y-M-D	hour	gas accumulation	gas usage
2019-06-25	15	652	0
2019-06-25	16	652	0
2019-06-25	17	652	0
...			
2019-06-27	18	652	0
2019-06-27	19	653	1

표 6. 누적량역전오류 예와 보정(김포)
Table 6. Example of accumulation reverse error (Gimpo).

Before correction			
Y-M-D	hour	water accumulation	water usage
2018-03-07	13	784.5	0
2019-06-25	14	784	0
...			
2018-03-09	20	784	0
2018-03-09	21	785	1
After correction			
Y-M-D	hour	water accumulation	water usage
2018-03-07	13	784.5	0
2019-06-25	14	784.5	0
...			
2018-03-09	20	784.5	0
2018-03-09	21	785	1

할 수 있다. 해당 오류는 약 0.17%의 비율로 발생하였다. 보정 방법은 k-1 시의 누적량과 k 시의 누적량을 비교하여 k-1 시의 누적량의 값이 큰 경우 k 시의 누적량을 k-1 시의 누적량으로 보정하였다. 이때 이전 시간의 누적량으로 계속 보정하다가 k 시의 누적량이 k-1 시의 누적량보다 커질 때 더 이상 보정을 진행하지 않는다.

광고와 동탄 데이터의 경우 해당 오류의 발생 빈도는 각각 0.02%와 1.32%이며 보정 방법은 김포 공동주택의 누적량역전오류에서의 보정 방법과 동일하다.

E. 사용량계산오류 및 보정

사용량은 계산에 의해 채워지는 값으로 'k-1 시의 누적량 + k 시의 사용량 = k 시의 누적량'의 규칙을 따른다. 해당 오류는 이전 오류 보정 과정에서 발생하는 변화로 인해 규칙에 맞지 않은 사용량 값을 의미한다. 이러한 오류의 발생 빈도는 C절의 오류와 비슷하였다. 보정 방법은 모든 사용량에 대해서 'k-1

시의 누적량 + k 시의 사용량 = k 시의 누적량' 규칙에 의해 값을 새로 계산하여 채워 넣는 것으로 하였다. 표 7에는 사용량계산오류의 예와 보정을 나타내었다.

광고와 동탄 데이터의 경우 해당 오류의 발생 빈도는 각각 0.1%와 0.03%로 빈번하게 발생하였으며 보정 방법은 김포 공동주택의 경우와 동일하다.

표 7. 사용량계산 오류 예와 보정(김포)
Table 7. Example of usage calculation error (Gimpo).

Before correction			
Y-M-D	hour	water accumulation	water usage
2018-03-09	20	784.5	0
2018-03-09	21	785	1
After correction			
Y-M-D	hour	water accumulation	water usage
2018-03-09	20	784.5	0
2018-03-09	21	785	0.5

F. 데이터누락오류 및 보정

해당 오류는 특정 시간대의 데이터 값이 누락되는 것으로, 모든 세대에서 같은 날짜의 동일한 시간대에 3번 발생하는 형태가 관찰된다. 이러한 오류를 데이터누락 오류라고 부르기로 한다. 해당 오류의 발생 비율은 약 0.0001%이다. 보정 방법은 누적량초기화오류와 유사하게 이전 주와 다음 주의 동일한 기간의 사용량 평균을 비율로 고려하여 누적된 사용량을 분배하였다. 비율 적용에는 다음과 같은 3가지 방법이 있다.

지난주와 다음 주 사용량의 평균이 특정 비율을 가지는 경우, 누락된 사용량을 특정 비율을 기반으로 누적량을 배분하여 누적량을 다시 계산하여 보정한다. 표 8의 수정 전 12시의 전기 사용량이 비율에 맞게 보정할 경우 0.162로 계산되는데 이때 누적량이 20547.236으로 20547.2보다 크기 때문에 증가된 값을 맨 마지막 사용량에서 조절하였다. 표 8의 보정 전 가스 사용량을 확인해보면 누락된 시간의 누적량이 0인데, 이 경우 비어있는 시간의 누적량은 누락 이전 시간의 누적량으로 대체하고 사용량을 0으로 보정한다. 표 8의 보정 전 수도 사용량에서, 지난주와 다음 주 사용량이 모든 시간에서 0인데, 이 경우에는 1/n으로 누적량을 배분하여 계산해 보정한다.

광고와 동탄 데이터의 경우 해당 오류의 발생 빈도는 각각 3ppm과 0.4ppm이며, 보정 방법은 김포 공동주택의 경우와 동일하다.

G. 그 밖의 오류

A부터 F까지의 6종류의 오류는 기본적인 검침 데이터의 형식에 어긋나므로 검출이 용이하다. 그러나 이러한 기본 형식을 준수하고 있지만, 데이터의 내용에 이상이 있는 경우가 있다. 예를 들면, 김포 1,007개 세대

표 8. 데이터누락 오류 예와 보정(김포)
Table 8. Example of data missing error (Gimpo).

Before correction										
Y-M-D	hour	electricity accumulation	electricity usage	gas accumulation	gas usage	water accumulation	water usage			
2018-03-07	9	20546.3	0.3	376.6	0	1021.5	0			
2018-03-07	12	20547.2	0.9	376.6	0	1021.6	0.1			
After correction										
Y-M-D	hour	electricity accumulation	electricity usage	average usage ratio	gas accumulation	gas usage	average usage ratio	water accumulation	water usage	average usage ratio
2018-03-07	9	20546.3	0.3		376.6	0		1021.5	0	
2018-03-07	10	20546.912	0.612	0.64	376.6	0		1021.53	0.03	0.3
2018-03-07	11	20547.074	0.162	0.18	376.6	0		1021.56	0.03	0.3
2018-03-07	12	20547.2	0.126	0.18	376.6	0		1021.6	0.04	0.4

에 대한 시간당 평균 전력사용량을 구하였을 때, 그림 1에서 보는 바와 같이 12시경에 갑자기 평균 전력사용량이 급증하였다가 14시경에 0으로 급감하고 15시경에는 다시 급증하는 이상 패턴이 관찰된다. 이는 전체 세대의 계량데이터를 수집하는 장치인 DCU(data concentrator unit)에서 데이터를 검침하는 과정에서 문제가 생기는 경우로, 이 시간대에서는 모든 세대의 5종 데이터에서 이상이 발생한다. 이러한 오류를 세대별 에너지원별 요일별 특성을 고려하지 않고 임의로 처리하면 이전 데이터 패턴과 상이한 형태의 사용 패턴을 얻게 된다. 이러한 오류는 보다 심도 있게 관찰하여 보정해줄 필요가 있다.

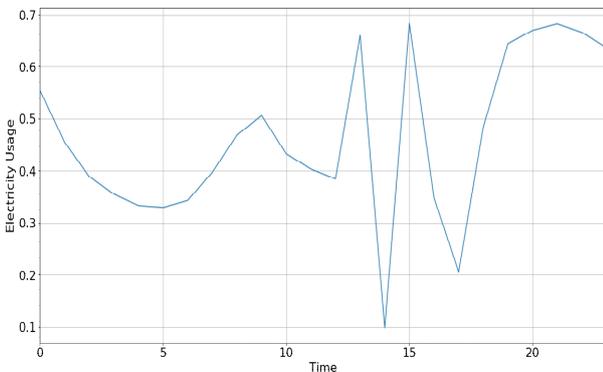


그림 1. 1,007 세대의 시간대별 평균 전기사용량(김포)
Fig. 1. Average electricity usage by hour of 1,007 households (Gimpo).

III. 검침 데이터의 패턴 분석

본 장에서는 제 II장에서 서술한 오류들을 보정하여 각 5종 데이터의 시간별 전체 평균 사용량을 구하여 에너지 사용 패턴을 간단히 비교하였다. 그림 2의 전기 사용 곡선을 보면, 대체로 출근 시간인 오전 5-7시 경에 전기 사용량이 증가하는 패턴을 보이고 이후에는 점차 전기 사용량이 감소하다가 퇴근 시간인 20시 이후 급격하게 전기 사용량이 증가하는 경향을 보이고 있다. 이러한 패턴은 전기에너지뿐만 아니라 수도, 온수, 가스 에너지의 경우에도 비슷한 사용 패턴을 보여주고 있다.

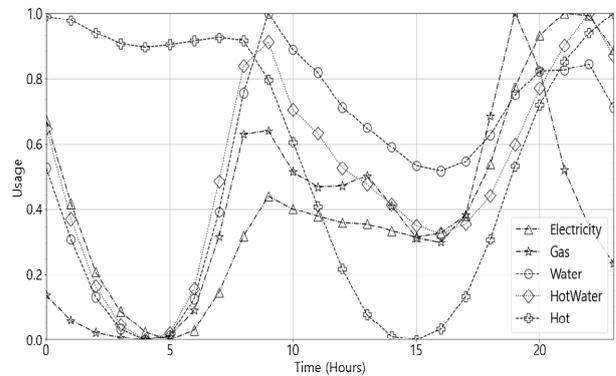


그림 2. 시간별 평균 5종 데이터 사용량(김포)
Fig. 2. Average of 5 types of data usage per hour (Gimpo).

반면에 난방데이터의 경우 계절에 의한 영향이 뚜렷하기 때문에 전기, 수도, 온수, 가스의 4종 데이터와 다른 패턴을 보여주고 있다. 또한 난방데이터는 10시-15시에 급격히 감소하는 경향을 보이다가 20시 이후에 급

격하게 증가하는 형태를 보여주고 있는데, 이 역시 나머지 데이터와는 다른 사용 패턴임을 알 수 있다.

이제 k-means 군집화 알고리즘을 이용하여 각 5종 데이터를 k개 군집으로 분류하여 24시간 사용 패턴을 각 군집별로 관찰한다. 분류된 데이터는 사용량 기준으로 5개의 군집으로 대부분의 군집에서 서로 유사한 사용 패턴이 있음을 알 수 있었다.

A. 전기 데이터

김포 전기 데이터는 그림 3에서 보듯이 4개의 군집에서 비슷한 24시간 사용 패턴을 보여주고 있으나 나머진 개의 군집은 사용량이 매우 적은 경향을 보여주고 있다. 주로 출근 시간대인 오전 6시경에 사용량이 증가하다가 12시경에 감소하는 패턴을 보여주고 있고 퇴근 시간에 다시 증가하는 경향이 4개의 군집에서 동일하게 관찰되었으며 사용량이 많은 시간대는 20시-24시로 관찰되었다. 광고와 동탄 공동주택 전기 데이터도 김포의 경우와 매우 유사한 사용 형태를 보여주고 있다.

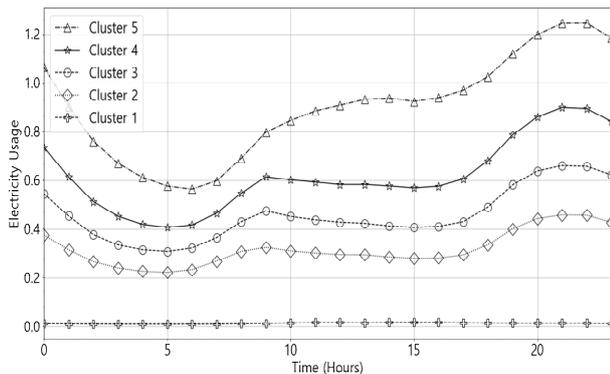


그림 3. 전기사용량의 호수별 군집화 분석(김포)
Fig. 3. Clustering analysis of electricity usage (Gimpo).

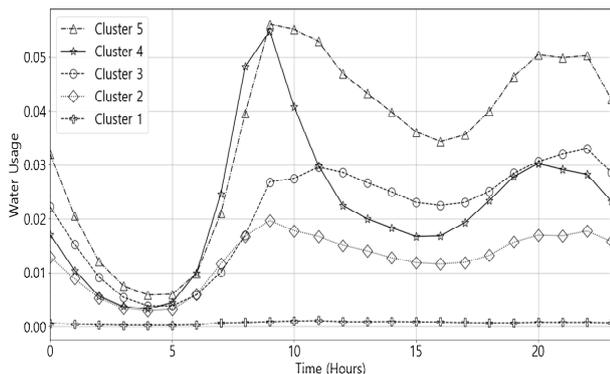


그림 4. 수도사용량의 호수별 군집화 분석(김포)
Fig. 4. Clustering analysis of water usage (Gimpo).

B. 수도 데이터

김포 공동주택의 수도 데이터는 그림 4에서 보는 바와 같이 전기 데이터와 유사하게 4개의 군집이 서로 비슷한 경향을 보여주었고 출퇴근 시간대에 사용량이 증가하다가 그 외 시간에 감소하는 패턴을 보여주었다. 다만 전기 데이터에 비해 시간대별 사용량의 패턴 변화 폭이 큰 경향이 있는 것으로 관찰되었으며, 수도 데이터의 사용량이 많은 시간대는 8시-9시와 22시-23시로 관찰되었다.

광고 공동주택의 수도 데이터는 그림 5에서 보는 바와 같이 1개의 군집을 제외하고 나머지 4개의 군집이 서로 비슷한 경향을 보여주었다. 다른 형태를 가지는 'Cluster 4' 군집의 경우 10시경에 수도 사용량이 크게 증가하는 군집으로 관찰되며 나머지 3개의 군집은 전기 데이터와 동일하게 출퇴근 시간의 사용량이 증가하는 형태를 보여주었으며, 광고의 수도 데이터는 대체로 20시에서 사용량이 많은 것으로 관찰된다.

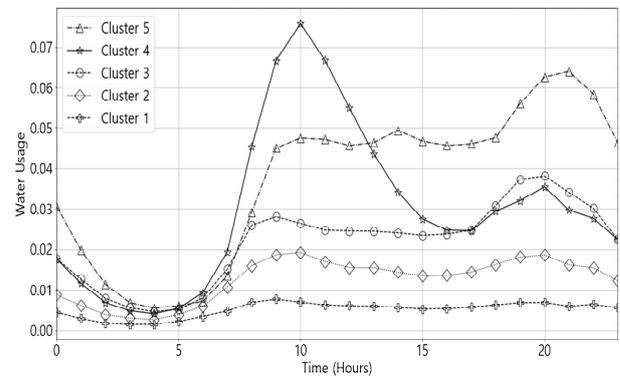


그림 5. 수도사용량의 호수별 군집화 분석(광고)
Fig. 5. Clustering analysis of water usage (Gwanggyo).

동탄 공동주택의 수도 데이터의 군집은 그림 6에서 보는 바와 같이 한 개의 군집을 제외하고 나머지 군집들은 상대적으로 적은 사용량을 보여주는데, 이는 동탄의 경우 크기가 작은 주택으로 1인이 주거하는 경우가 많고, 이런 경우 물의 사용량이 상대적으로 적은 것을 의미한다. 사용량이 큰 'Cluster 5' 군집은 11시에 급증하여 감소하다가 20시에 다시 증가하는 형태를 나타낸다. 이와 유사한 사용량 형태는 그림 5의 광고의 경우에도 관찰된다.

C. 온수 데이터

김포 공동주택 온수 데이터의 경우 그림 7에서 보는 바와 같이 출퇴근 시간대에 다소 증가하는 경향을 보여

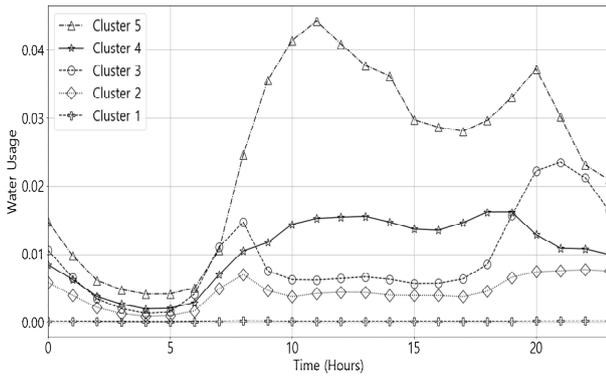


그림 6. 수도사용량의 호수별 군집화 분석(동탄)
Fig. 6. Clustering analysis of water usage (Dongtan).

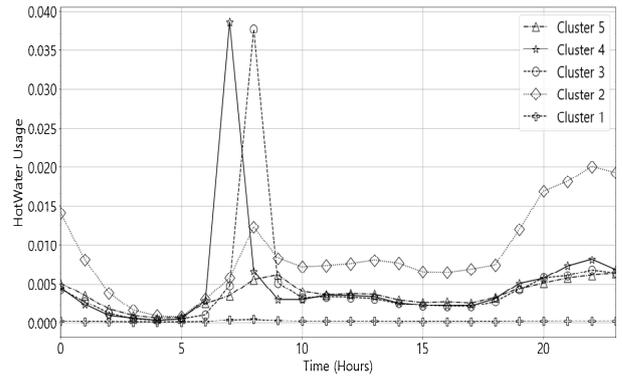


그림 8. 온수사용량의 호수별 군집화 분석(동탄)
Fig. 8. Clustering analysis of hot water usage (Dongtan).

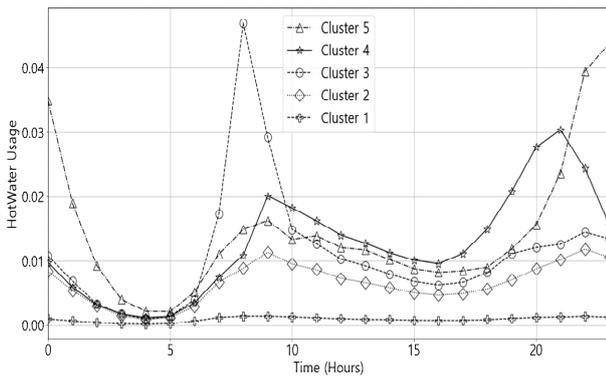


그림 7. 온수사용량의 호수별 군집화 분석(김포)
Fig. 7. Clustering analysis of hot water usage (Gimpo).

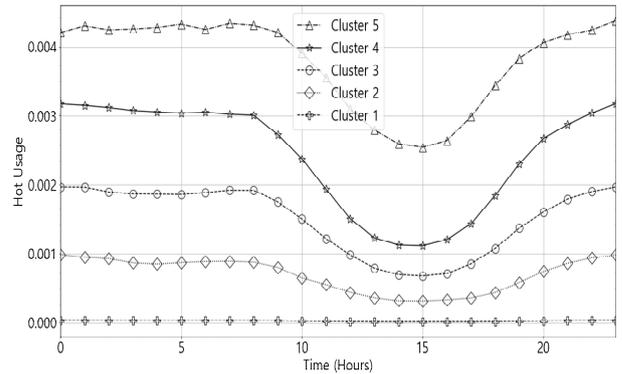


그림 9. 난방데이터의 호수별 군집화 분석(김포)
Fig. 9. Clustering analysis of heat usage (Gimpo).

주었지만 ‘Cluster 3’의 군집에서는 새벽 시간대에 사용량이 급증하는 이상 패턴을 확인할 수 있다. 온수 데이터의 경우 다른 데이터와 달리 사용량이 많은 구간은 출근 시간대인 오전 6시-8시와 새벽 시간인 22시-24시로 관찰되었다.

광교 공동주택의 경우는 김포 데이터와 달리 각 군집의 사용 패턴이 서로 비슷한 경향을 보여주었으며 사용량이 많은 시간대는 20시로 관찰된다.

동탄 공동주택의 경우 그림 8에서 보는 바와 같이 오전 시간대에 크게 사용량이 증가하는 군집이 관찰되었고, ‘Cluster 3’과 ‘Cluster 4’ 군집은 다른 군집과는 다르게 6시-10시 사이에 크게 사용량이 증가하다가 급격히 감소하는 형태를 나타낸다.

D. 난방 데이터

난방 데이터의 경우 그림 9의 김포 데이터에서 보는 바와 같이 4개의 군집이 기온이 낮은 시간대(17시 이후)에 사용량이 증가하다가 오전 시간대(10시 이후)에 사용량이 감소하는 패턴을 보여주었다. 반면에 난방 데

이터는 전기 데이터와 동일하게 서로 겹치는 구간 없이 군집마다 사용량 패턴이 고르게 분배되어 있는 것을 확인할 수 있었다. 난방데이터는 22시-24시 구간에서 가장 많은 사용량 패턴을 보여주었다.

광교 공동주택의 경우 그림 10에서 보는 바와 같이 김포의 경우처럼 15시에 사용량이 급격하게 줄어드는 형태는 하나의 군집에서만 보이고 나머지는 일정한 소비 특성을 보인다. 이러한 일정한 형태는 동탄의 경우

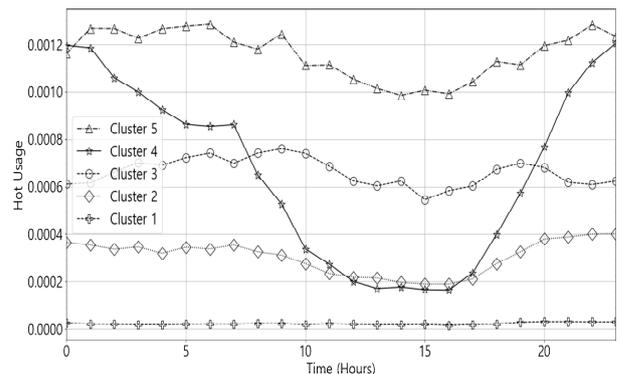


그림 10. 난방데이터의 호수별 군집화 분석(광교)
Fig. 10. Clustering analysis of heat usage (Gwanggyo).

에도 관찰이 된다. 다만 그림 11의 동탄의 경우를 보면 대부분이 사용량이 적고 한 군집만 상대적으로 큰 값을 가진다. 이러한 특성은 광고와 동탄의 경우 김포의 경우에 비하여 작은 면적의 주거 형태에서 기인하는 것으로 판단된다.

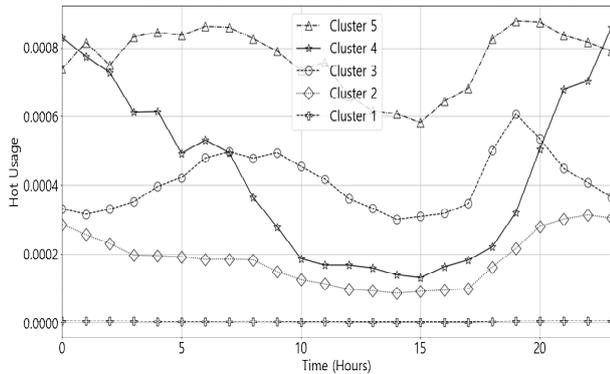


그림 11. 난방데이터의 호수별 군집화 분석(동탄)
Fig. 11. Clustering analysis of heat usage (Dongtan).

E. 가스 데이터

김포 공동주택에서 수집된 가스 데이터의 경우 그림 12에서 보는 바와 같다.

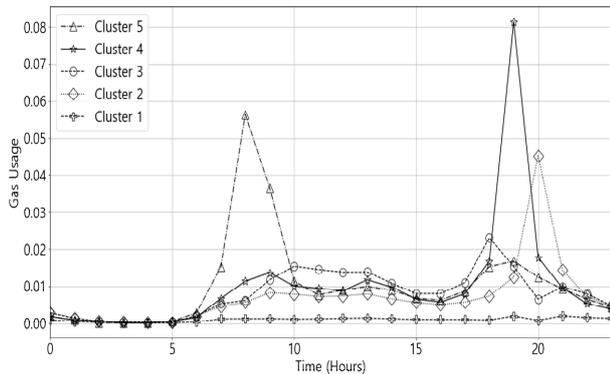


그림 12. 가스 데이터의 호수별 군집화 분석(김포)
Fig. 12. Clustering analysis of gas usage (Gimpo).

가스 사용패턴은 전기, 수도, 온수와 비슷하게 출퇴근 시간에 사용량이 증가하는 경향을 보여주었으며 주로 저녁 시간대에서 가장 많은 사용량을 소비하는 것을 확인할 수 있었다. 가스 데이터의 사용량이 많은 구간은 18시-19시 구간으로 관찰되었다.

IV. 5종 검침 데이터의 상관도

본 장에서는 김포 공동주택에서의 5종 검침 데이터 간의 상관도에 대하여 관찰 분석한다.

각 검침 데이터의 24시간 사용 형태는 각 집에 거주하는 사람의 행동 형태와 밀접한 관계를 가지고 있다. 일반적으로 전기, 수도, 온수의 사용 형태는 서로 비슷한 형태를 가지고 있다. 즉 24시간 중에 전기를 많이 쓸 때는 사람의 활동량이 많은 시간대로 수도나 온수도 많이 소모하게 된다. 이러한 특성은 앞의 그림 2에서 관찰된다. 이러한 특성을 정량적으로 분석하기 위하여 Pearson 상관도와 Spearman 상관도를 구하여 그림 13과 14에 각각 도시하였다. 이때 각 세대별, 24시간대별로 한 달간의 사용량에 대한 평균을 구하고 이들의 상관도를 구하였다.

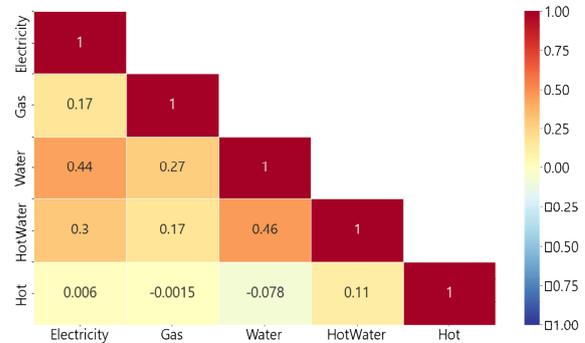


그림 13. 5종 검침 데이터의 Pearson 상관도 관찰
Fig. 13. Observation of the Pearson correlation coefficients for the 5 types of metering data.

그림 13의 Pearson 상관도를 보면, 전기 데이터와 가장 큰 상관도를 보이는 것은 수도 데이터이다. 반면에 난방 데이터와는 가장 상관도가 없음을 알 수 있다. 그림 13에서 가장 높은 상관도를 보이는 경우는 수도와 온수로, 이는 수도와 온수가 같이 사용되는 경우가 많다는 것을 의미한다. 이러한 추세는 Spearman 상관도의 경우에 더 분명해진다. 5종 검침 데이터의 경우 공가나 사람이 없는 경우 일반적인 사용 형태가 아닌 다른 특성이 보이므로 outlier가 존재할 때 보다 강인한 상관도를 보여주는 그림 14의 Spearman 상관도를 관찰하는 것도 의미가 있다고 판단된다. 참고로, 상관도 실험에서 사용량이 0에 가까운 세대들은 아직 입주자가 되지 않은 경우로 상관도 분석에서 제외하였다.

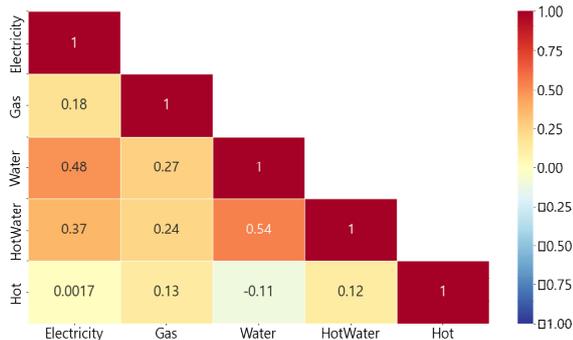


그림 14. 5종 검침 데이터의 Spearman 상관도 관찰

Fig. 14. Observation of the Spearman correlation coefficients for the 5 types of metering data.

V. 결 론

본 논문에서는 공동주택에서 수집된 5종 원격검침 데이터를 관찰하여 다양한 오류 형태를 분류하고 오류를 보정하였다. 보정된 5종 데이터를 분석한 결과, 시간별 평균 전기, 수도, 온수, 가스 데이터 사용량은 세대의 출퇴근 생활패턴에 맞추어서 에너지 사용량이 변화하는 패턴을 보여주었고, 계절의 영향을 받는 난방 에너지 사용 패턴은 그 외 4종 에너지 데이터와는 다른 형태를 보여줄 수 있었다. 또한 주택의 면적이 좁은 경우에는 각 계량데이터의 군집화 분석에서 서로 상이한 형태를 보이고 있음을 관찰할 수 있었으나, 면적이 넓어지면 모든 군집들이 서로 유사한 사용 패턴을 보였다. 향후 연구로는 이렇게 보정된 5종 에너지 검침 데이터를 사용하여 다양한 공동주택 단지를 대상으로 에너지원별 상관도를 보다 면밀히 관찰하고, 에너지원별 세분화 및 딥러닝 기반 에너지 사용량 예측 기술 등을 개발하여 에너지 데이터의 활용도를 확대하는 것이 필요하다고 판단된다.

REFERENCES

[1] H. Hwang, Y. Seo, "A development of real-time energy usage data collection and analysis system based on the IoT", *Jour. of Korea Multimedia Society*, vol. 22, no. 3, pp. 366-373, Mar. 2019.

[2] M. H. Park, Y. H. Kim, and S. B. Lee, "A study on the development of energy IoT platform," *Korea Information Processing Society Transactions on Computer and Communication Systems*, vol. 5, no. 10, pp. 311-318, 2016.

[3] Ministry of Trade, Industry and Energy, *The*

3rd Energy Master Plan, 2019.

- [4] B. Shim, K. Lim, H. Kim, Y. Yun, S. Choi, "AMI system based on LoRa IoT" in *Proc. KICS Summer Conference*, pp. 1302-1303, Jeju, Korea, Aug. 2017.
- [5] M. Choi, S. Ju, Y. H. Lim, "Integrated metering system based on power-line communication" in *Proc. KIEE Summer Conference*, pp. 596-597, Yongpyong, Korea, July 2007.
- [6] S. Lee, S. Lee, M. Song, and Y. Kwon, "An empirical research on the IoT basis gas AMI platform and smart metering services," *Jour. of the Korean Institute of Gas*, vol. 24, no. 3, pp. 1 - 10, Jun. 2020.
- [7] J. Lee, D. Lee, H. Lee, "R&D trends and prospects of automatic meter reading system" in *Proc. ITFE Summer Conference*, pp. 150-154, Korea, Aug. 2008.
- [8] N. Yoo, E. Lee, B. J. Chung, and D. S. Kim, "Analysis of apartment power consumption and forecast of power consumption based on deep learning," *Jour. IKEEE*, vol. 23, no. 4, pp. 1373-1380, Dec. 2019.
- [9] H. Yang, B. J. Chung, D. S. Kim, "Correlation analysis between the electricity energy usage and various metering data in apartment housing environment," in *Proc. IEIE Summer Conference*, pp. 1397-1400, Jeju, Korea, Aug. 2020.
- [10] D. Oh, J. Park, C. Kim, "A method for duplicated data removal considering data processing efficiency," in *Proc. KISM Autumn Conference*, Gwangju, Korea, Nov. 2020.
- [11] W. Jung, B. J. Chung, and D. S. Kim, "Analysis of the single and general contracts in electricity supply for high-voltage apartments," *Jour. IEIE*, vol. 57, no. 10, pp. 87-95, Oct. 2020.
- [12] J. Jung, D. S. Kim, B. J. Chung, and Y. M. Chung, "Analysis on the residential progressive and time-of-use rates based on AMI data," *Jour. IEIE*, vol. 58, no. 9, pp. 66-77, Sep. 2021.

저 자 소 개



김 동 식(정회원)
1986년 서울대학교 제어계측공학과
학사 졸업.
1988년 서울대학교 제어계측공학과
석사 졸업.
1994년 서울대학교 제어계측공학과
박사 졸업.

2000년~현재 한국외국어대학교 전자공학과 교수
<주관심분야: 신호처리, 의료영상시스템, 의료영
상처리, 무선센서네트워크, 스마트미터링>



정 영 모(정회원)
1986년 서울대학교 제어계측공학과
학사 졸업.
1988년 서울대학교 제어계측공학과
석사 졸업.
1993년 서울대학교 제어계측공학과
박사 졸업.

1995년~현재 한성대학교 전자정보공학과 교수
<주관심분야: 디지털통신시스템, 신호처리, 전력
선통신, 스마트미터링>



강 송 희(비회원)
2018년 한국외국어대학교 수학과
컴퓨터공학과 이중 졸업
2020년 한국외국어대학교
컴퓨터전자시스템공학과
석사 졸업
현재 스트라드비전 재직중

<주관심분야: 데이터통계분석, 영상처리, 머신러
닝, 딥러닝>



정 범 진(정회원)
1986년 서울대학교 제어계측공학과
학사 졸업.
1988년 서울대학교 제어계측공학과
석사 졸업.
2014년 한국외국어대학교 전자공학
박사 졸업.

현재 서울과학기술대 전기정보기술연구소 연구교수
<주관심분야: 스마트미터링, 스마트그리드, AMI,
전기요금, 전력정책>



홍 은 영(비회원)
2020년 한국외국어대학교
전자공학과 석사졸업
현재 몬드리안에이아이 재직 중

<주관심분야: 딥러닝, 영상처리, 신호처리>