

# IoT 센서 데이터를 이용한 단위실의 재실추정을 위한 Decision Tree 알고리즘 성능분석

김석호 · 서동현<sup>†</sup>

충북대학교 건축공학과

## A Study on Occupancy Estimation Method of a Private Room Using IoT Sensor Data Based Decision Tree Algorithm

Kim Seok-Ho · Seo Dong-Hyun<sup>†</sup>

Department of Architectural Engineering, Chungbuk National University

<sup>†</sup>Corresponding author: seodh@cbnu.ac.kr

### Abstract

Accurate prediction of stochastic behavior of occupants is a well known problem for improving prediction performance of building energy use. Many researchers have been tried various sensors that have information on the status of occupant such as CO<sub>2</sub> sensor, infrared motion detector, RFID etc. to predict occupants, while others have been developed some algorithm to find occupancy probability with those sensors or some indirect monitoring data such as energy consumption in spaces. In this research, various sensor data and energy consumption data are utilized for decision tree algorithms (C4.5 & CART) for estimation of sub-hourly occupancy status. Although the experiment is limited by space (private room) and period (cooling season), the prediction result shows good agreement of above 95% accuracy when energy consumption data are used instead of measured CO<sub>2</sub> value. This result indicates potential of IoT data for awareness of indoor environmental status.

**Keywords:** 재실정보(Occupancy Information), 재실추정(Occupancy estimation), 사물 인터넷 데이터(IoT data), 의사결정나무(Decision tree), 데이터 마이닝(Data mining)

### 기호설명

- D : 데이터 집합 (또는 그 집합 내의 데이터 수)
- C : 출력변수
- A : 입력변수
- m : 출력변수(C)의 가짓수
- v : 입력변수 값(a)의 가짓수
- $C_{i,D}$  : 데이터 집합(D)내에 존재하는 i번째 출력변수(C)의 개수

 OPEN ACCESS



Journal of the Korean Solar Energy Society  
Vol.37, No.2, pp.23-33, April 2017  
<https://doi.org/10.7836/kses.2017.37.2.023>

pISSN : 1598-6411

eISSN : 2508-3562

Received: 06 February 2017

Revised: 21 April 2017

Accepted: 25 April 2017

Copyright © Korean Solar Energy Society

This is an Open-Access article distributed under the terms of the Creative Commons Attribution NonCommercial License which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

$p_i$  : 데이터 집합(D) 중 임의의 출력변수(C)가  $C_{i,D}$  일 확률

$D_j$  : 데이터 집합(D) 중 입력변수(A)가 j번째 값(a)인 데이터의 집합 (또는 그 집합 내의 데이터 수)

## 1. 서론

### 1.1 연구의 배경

건축공간에서 재실여부와 재실인원 등의 재실정보는 건물 외피의 단열성능, 설비시스템의 종류, 효율, 제어방식 등 효율적 설계와 관리와는 관계없이 건물의 에너지 사용량을 결정하는 주된 요인 중 하나이다.

상세 건물에너지 사용량 예측에 사용되는 건물에너지 모델링 도구들은 재실정보를 ASHRAE (American Society of Heating, Refrigerating, and Air Conditioning Engineers)에서 제시한 건물용도별 표준재실스케줄<sup>1)</sup>을 활용하는 것이 보편적이다. 표준재실스케줄은 해당 용도건물의 전형적인 재실스케줄의 의미를 지니지만 개별 건물의 재실특성을 반영할 수는 없으므로 특정 건물의 에너지 수요예측에 표준재실스케줄을 사용할 경우, 그 결과는 실제 값과 상이할 수 있다. 이러한 이유로 건물에너지 모델링 도구를 이용하여 에너지 사용량을 예측함에 있어 재실스케줄의 정확성에 따른 에너지 사용량의 편차에 관한 연구가 진행되고 있다<sup>2)3)</sup>.

최근에는 표준재실스케줄의 단점을 극복하고자, 실제 건물의 재실정보 데이터를 이용하여 기계학습을 바탕으로 특정 건물의 실제 재실특성을 추정하여 건물에너지 모델에 반영하는 연구들이 진행되고 있다<sup>4)5)</sup>. 특히 Simona and Tianzhen (2015)의 연구에서는 기계학습에 의해 추정된 재실스케줄이 기존의 단순 재실확률에 기반하여 추정된 재실스케줄에 비해 정확도가 약 30% 정도 향상되었다고 보고하였다.

한편 재실정보의 수집을 위해 직접적으로 재실정보를 측정하는 적외선센서나 이미지센서 등은 정확도 및 사생활 침해의 문제가 있다. 최근 IoT기술의 발달로 재실자로 인해 변화하는 실내 환경 데이터(CO<sub>2</sub>농도, 온습도 등)나 에너지수요 데이터(조명, 사무기기전력 등)의 측정이 용이해지고 있고, 이에 따라 IoT데이터를 이용한 재실정보추정에 대한 관심도 높아지고 있다. 이러한 방식의 재실정보추정은 재실과 실내 환경 및 에너지 수요의 관계를 고려하여 간접적으로 재실을 추정하므로, 별도의 재실센서가 요구되지 않으며 사생활 침해의 문제가 완화된다는 장점이 있다.

본 연구에서는 냉방기간 동안 실제 사용 중인 단위실에서 수집된 온도, 습도, CO<sub>2</sub>농도, 조도, A/C전력, 조명전력 등의 IoT 센서 데이터를 이용한 재실정보 추정에 있어, 데이터마이닝의 분류기법 중 하나인 Decision Tree 기법인 CART 와 C4.5 알고리즘을 이용하여 재실추정성능을 분석하고자 한다. 최종적으로 이를 통해 간접적으로 재실을 추정할 수 있는 IoT 센서 데이터의 단위실 재실추정에 대한 활용성능을 분석하고자 한다.

### 1.2 문헌 고찰

재실정보의 추정에 대한 연구들은 추정의 정확도에 주요한 영향을 미치는 입력데이터의 종류와 사용한 추정기법에 따라 다양한 연구가 진행되고 있다.

한화택 외(2012)<sup>6)</sup>는 실내외 CO<sub>2</sub>농도 데이터를 바탕으로 동적신경망을 이용하여 재실인원을 예측하였다. 동적 신경망은 Matlab의 Neural Network Toolbox를 통해 구현하였고, CO<sub>2</sub>농도에 기반한 재실인원 추정에는 공간의 환기 및 침기성능에 따른 CO<sub>2</sub>농도 변화특성에 시간상수의 고려가 필요하다는 결론을 얻었다.

배우빈 외(2015)<sup>7)</sup>는 Wi-Fi 접속인원 데이터를 바탕으로 선형회귀분석을 이용해 재실인원을 추정하였다. Wi-Fi 접속인원 데이터는 공유기에 접속된 MAC 주소의 개수로 산정하였고, 추정의 결과는 FEMP(Federal Energy Management Program)의 M&V Guideline<sup>8)</sup>에서 제시하는 기준에 만족하는 결론을 얻었다.

Seung Ho Ryu와 Hyeun Jun Moon (2016)<sup>9)</sup>은 실내외 CO<sub>2</sub>농도와 조명 및 사무기기전력사용량 데이터를 바탕으로 Decision Tree 알고리즘인 CART (Classification And Regression Tree)를 이용해 어떠한 입력변수를 활용하여 재실인원을 추정하는 것이 정확성을 보장하는지 분석하였다. Decision Tree 알고리즘은 Matlab으로 구현하였고 CART 알고리즘을 이용하여 실내 CO<sub>2</sub>농도, 실내외 CO<sub>2</sub> 농도비율, 조명 및 사무기기전력 데이터를 바탕으로 생성된 추정모델이 가장 성능이 좋았으며, 이때의 RMSE 값은 약 0.2202 명으로 보고되었다.

위와 같은 문헌들을 통해 입력변수로 단일한 변수를 사용하는 것보다 여러 변수들의 조합을 사용하는 것이 재실 정보 추정에 효과적이며, Decision Tree의 CART 알고리즘은 이러한 변수조합을 이용해 재실정보를 추정함에 있어 적합함을 알 수 있었다. 본 문헌에서는 Seung Ho Ryu와 Hyeun Jun Moon의 연구에서 더 나아가 재실정보추정에 있어, Decision Tree의 가장 대표적인 알고리즘인 C4.5와 CART의 성능을 분석하였다. 그리고 실험을 위한 Test-bed가 아닌 실제 사용되는 공간을 대상공간으로 선정하였으며, 실제 대상공간으로부터 수집되고 간접적으로 재실정보를 나타내는 다양한 실내 환경 및 에너지수요 데이터의 재실추정에 대한 활용 가능성을 검토하였다.

## 2. 본론

### 2.1 대상 공간

본 연구의 대상 공간은 데이터의 수집과 분석이 용이해야 하고 불필요한 변수들의 발생을 예상하거나 통제할 수 있어야 하므로 연구자의 활동범위 안에 있는 1개의 개별실로 선정하였다. 대상 공간의 정보는 Table 1에서 확인할 수 있다.

**Table 1** Status of target experiment room

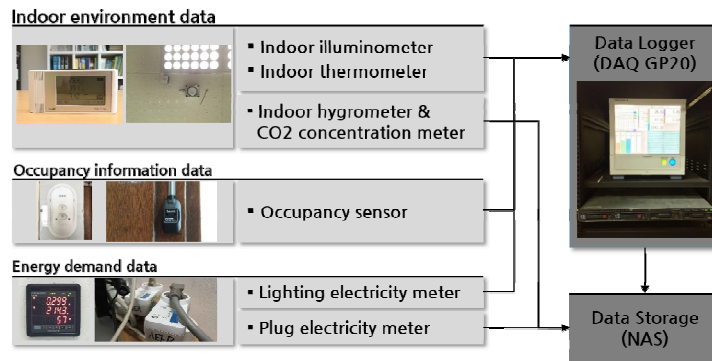
Location	Cheongju, Chungbuk
Room area	22.51m <sup>2</sup>
Room purpose	Private office
Occupant number	1 person
Lighting equipment & control	LED, Auto dimming control with On/Off switch
Heating and cooling equipment & control	EHP, Auxiliary heating(Individual control), Steam radiator(Central control)

## 2.2 데이터의 개요 및 전처리

본 연구에 사용된 데이터는 대상공간으로부터 수집되었고, 수집기간은 2016년 7월 1일 ~ 21일까지 3주이다. 수집데이터의 항목 및 세부사항은 Table 2와 Fig. 1에서 확인할 수 있다. Table 2의 온도, 상대습도, CO<sub>2</sub>농도, 조도, 조명전력 그리고 A/C전력 데이터는 재실추정을 위한 입력변수로 사용되었고, 적외선 센서로 수집된 재실데이터는 추정된 재실여부의 정확성을 평가할 비교데이터로 사용되었다. 각 로거로부터 수집된 데이터의 전처리에는 R 소프트웨어가 사용되었고 시간형식은 ISO 8601<sup>10)</sup>에서 제시하는 형식으로, 시간간격은 15분으로 통일하였다.

**Table 2** Overview of measured data of target experiment room

Logger	Measuring element	Time interval	Time format	Usage
Logger 1	Temperature	1 min	yyyy/mm/dd hh:mm:ss	Predictor
	Illuminance			Predictor
	Lighting power			Predictor
	Occupancy presence			Reference data
Logger 2	Relative humidity	15 mins	yyyy-mm-dd hh:mm	Predictor
	CO <sub>2</sub> concentration			Predictor
Logger 3	A/C power	1 sec	yyyy-mm-dd hh:mm:ss	Predictor



**Fig.1** Configuration of sensors and data collecting system

## 2.3 Decision Tree의 이론적 배경<sup>1)12)</sup>

Decision Tree 기법은 다양한 입력변수에 기반하여 최종 출력변수의 값을 예측하는 모델을 구축하는 기계학습 분류기법의 일종이다. Decision Tree 알고리즘은 분류 후 데이터들의 출력변수의 동질성을 최대화한다는 분류 규칙에 따라, 데이터의 입력변수 및 입력변수 값을 기준으로 출력변수를 분류한다. 각 분류의 단계에서 분류규칙에 따라 데이터를 분류하는 과정을 반복하고, 이를 마치면 추정모델이 생성된다. 이 때 분류 전후 출력변수의 동질성을 판단하는 척도를 불순도 척도라 하며 불순도 척도는 낮을수록 출력변수의 동질성이 높음을 의미한다. C4.5알고리즘<sup>13)</sup>은 불순도 척도로 Entropy를 사용하고 CART알고리즘<sup>14)</sup>은 Gini Index를 사용한다. Entropy와 Gini Index는

식(1), (2)로 구할 수 있다.

$$Entropy = - \sum_{i=1}^m p_i \log_2(p_i) \quad (1)$$

$$Gini Index = 1 - \sum_{i=1}^m p_i^2 \quad (2)$$

Decision Tree의 분류규칙은 입력변수 선택척도라고도 하며, 입력변수 선택척도에 의해 분류기준으로 가장 적합한 입력변수와 입력변수 값이 선택된다. C4.5알고리즘은 입력변수 선택척도로 Gain Ratio를 사용하고 CART알고리즘은  $\Delta Gini$ 을 사용한다. 두 입력선택 선택척도는 각각 다음의 식(3), (4)를 통해 구할 수 있다.

$$Gain Ratio(A) = \frac{Entropy(D) - \sum_{j=1}^v \frac{|D_j|}{|D|} \times Entropy(D_j)}{- \sum_{j=1}^v \frac{|D_j|}{|D|} \times \log_2\left(\frac{|D_j|}{D}\right)} \quad (3)$$

$$\Delta Gini(A) = Gini Index(D) - \sum_{j=1}^v \frac{|D_j|}{|D|} \times Gini Index(D_j) \quad (4)$$

## 2.4 검증지표

재실정보 추정을 위한 Decision Tree 추정모델의 검증에는 정확도와 오차의 분포 및 편향성을 검토하는 RMSE와 MBE, 그리고 추정모델의 통계적인 유의성을 검증하는 t-통계량이 사용되었다. 정확도는 관측된 데이터 수에 대한 옳게 추정된 데이터 수의 비율로 산정하였다. t-통계량은 추정모델의 통계적인 유의성을 검토하는 지표로 t-통계량 값이 임계값(critical t-value)보다 작은 추정모델은 통계적인 의미를 지니는 것으로 고려할 수 있고, R. J. Stone(1993)<sup>15)</sup>에 의해 기존의 t-통계량 수식을 바탕으로 RMSE와 MBE로 정리된 식(5)를 통해 계산되었으며 n은 데이터 수를 의미한다.

$$t - statistics = \sqrt{\frac{(n-1) \times MBE^2}{RMSE^2 - MBE^2}} \quad (5)$$

## 2.5 Decision Tree를 이용한 재실정보 추정

재실정보의 추정에는 Decision Tree 알고리즘 중 가장 대표적인 알고리즘인 C4.5와 CART를 사용하였고, 일련의 과정은 R을 통해 구현하였다. 재실 추정에 사용된 데이터는 실내 환경 데이터(온도, 상대습도, CO<sub>2</sub>농도, 조도)와

전력수요 데이터(조명전력, A/C전력), 재실정보 데이터로 실내 환경 및 전력수요 데이터는 입력변수로 사용되었고, 재실정보데이터는 출력변수로 사용되었다. 입력변수 중 실내 조도와 조명전력은 디밍제어를 사용하는 조명기구로 실내 조도와 조명전력은 단순한 선형관계가 아니며, 자연채광만 사용하는 경우가 있을 수 있어 두 변수를 모두 사용하였다. 데이터의 시간간격은 15분으로 2016년 7월 1일 ~ 21일까지 3주간의 데이터가 사용되었고, 그 중 초기 2주간의 학습데이터로 사용되었으며 나머지 데이터는 추정 및 결과검증을 위한 검증데이터로 사용되었다. 학습데이터의 경우 편향된 결과가 도출되는 것을 예방하기 위해 임의적으로 섞는 과정을 거친 뒤 사용하였다. 재실정보의 추정은 다음과 같이 3단계로 진행하였다.

(1) 주요 입력변수의 선별

Rui Zhang et al.(2012)<sup>16)</sup>이 제시한 방법을 바탕으로 Decision Tree의 입력변수 선택척도를 이용해 6개의 입력변수(온도, 상대습도, CO<sub>2</sub>농도, 조도, 조명전력, A/C전력) 중 재실여부에 주요한 영향을 미치는 4개의 입력변수를 선별하였다. 입력변수 선택척도로는 Gain Ratio를 사용하였고 6개의 입력변수에 대한 Gain Ratio값은 Table 3과 같다.

**Table 3** Gain ratio of all considered input variables

Variables	Gain ratio
Lighting Power (LP)	0.4565
Indoor Illuminance (Illu)	0.4031
A/C Power (ACP)	0.2978
Indoor CO <sub>2</sub> Concentration (CO <sub>2</sub> )	0.2045
Indoor Relative Humidity (Humi)	0.1036
Temperature (Temp)	0.1022

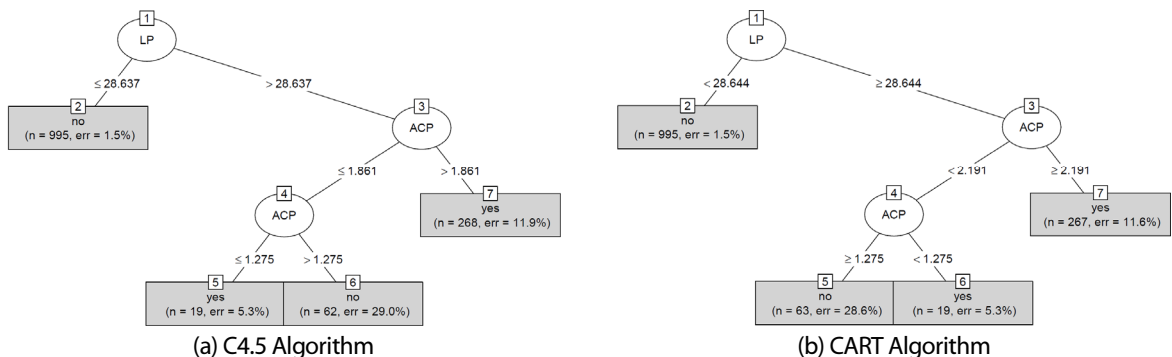
6개의 입력변수 중 Gain Ratio 값이 가장 높은 단일변수는 조명전력으로, 이는 조명전력이 재실정보의 추정에 가장 큰 영향을 미칠 것이라 추정할 수 있음을 의미한다. 같은 맥락으로 실내 조도, A/C전력, 실내 CO<sub>2</sub>농도, 실내 상대습도, 실내 온도 순으로 재실추정에 높은 영향을 미치고 있는 것으로 판단되며, 6개의 단일변수 중 Gain Ratio가 높은 4개의 입력변수(조명전력, A/C전력, 조도, CO<sub>2</sub>농도)를 선정하였다. Decision Tree 추정모델의 학습에는 선정된 4개의 입력변수들의 모든 조합이 고려되어야 하며, 각 변수들을 조합으로 생성된 15개의 입력변수조합은 Table 4에서 확인할 수 있다.

**Table 4** Combinations of selected input variables

Combination Number	The number of variables	Combinations of variables
1	1	CO <sub>2</sub>
2	1	Illu
3	1	LP
4	1	ACP
5	2	CO <sub>2</sub> +Illu
6	2	CO <sub>2</sub> +LP
7	2	CO <sub>2</sub> +ACP
8	2	Illu+LP
9	2	Illu+ACP
10	2	LP+ACP
11	3	CO <sub>2</sub> +Illu+LP
12	3	CO <sub>2</sub> +Illu+ACP
13	3	CO <sub>2</sub> +LP+ACP
14	3	Illu+LP+ACP
15	4	CO <sub>2</sub> +Illu+LP+ACP

## (2) Decision Tree 알고리즘의 적용

학습데이터를 바탕으로 선정된 입력변수들의 조합을 고려해 Decision Tree 알고리즘을 적용하였다. Fig. 2는 C4.5와 CART 알고리즘을 이용하여 학습된 추정모델 중 정확도가 가장 높은 알고리즘별 재실추정모델을 보여주고 있다. Fig. 2를 살펴보면 C4.5와 CART 알고리즘을 통해 생성된 두 재실추정모델은 각 분기점마다 분류 기준으로 사용된 입력변수는 동일하지만, 입력변수의 값에는 차이가 있다. 예로 C4.5의 두 번째 분기점에서는 A/C전력이 1.861W를 초과할 경우 재실로 분류하고 1.861W 이하의 추가적인 분류로 이어졌다. 반면 CART의 두 번째 분기점에서는 A/C전력이 2.191 W이상인 경우는 재실로 분류하고 2.191 W미만은 추가적인 분류로 이어졌다. 이러한 차이는 C4.5와 CART가 각 분기점에서 입력변수와 분류기준값을 선정함에 있어, Gain Ratio와  $\Delta$ Gini라는 서로 다른 분류규칙을 사용함으로 인해 발생하는 것으로 나타났다.

**Fig. 2** Example of learned decision tree models

(3) 학습된 추정모델을 이용한 재실여부 추정과 검증

검증데이터를 바탕으로 학습된 추정모델을 이용해 재실여부를 추정하고 Table 5와 같이 정확도, RMSE, MBE, t-통계량을 이용해 결과를 검증하였다. C4.5와 CART 모두 정확도가 가장 높은 입력변수는 조명전력과 A/C전력 두 가지를 동시에 사용할 경우였으며 정확도는 94.5% (C4.5), 94.4% (CART)로 큰 차이는 없었다. 입력변수에 따른 정확도를 살펴보면 동일한 Decision Tree 알고리즘으로 생성된 재실추정모델이라도 사용된 입력변수에 따라 정확도에는 큰 차이가 있음을 알 수 있고, 이와 같은 간접데이터의 재실추정에 대한 영향은 Fig. 3과 Fig. 4를 통해 가시적으로 확인할 수 있다. Table 5에서 Selected는 t-통계량 값이 임계값(critical value)보다 작아 통계적인 의미를 지니는 추정모델은 표시한 것이며 RMSE와 MBE는 재실여부에 대한 Fraction으로 이에 따라 t-통계량도 단위가 없는 척도이다.

**Table 5** Accuracy and error of estimation result

Algorithm	Combination of variables	Accuracy	RMSE	MBE	t-statistics*	Selected
C4.5	LP+ACP	0.945	0.235	-0.005	0.493	✓
	Illu+LP+ACP	0.944	0.238	-0.003	0.324	✓
	ACP	0.932	0.262	0.015	1.476	✓
	CO <sub>2</sub> +ACP	0.932	0.262	0.033	3.267	
	CO <sub>2</sub> +Illu+ACP	0.927	0.270	-0.007	0.714	✓
	CO <sub>2</sub> +LP	0.918	0.286	-0.064	5.944	
	Illu+ACP	0.908	0.304	0.048	4.111	
	CO <sub>2</sub> +Illu+LP+ACP	0.908	0.304	-0.024	2.037	
	CO <sub>2</sub> +LP+ACP	0.903	0.311	-0.016	1.365	✓
	CO <sub>2</sub> +Illu+LP	0.862	0.372	0.061	4.307	
	LP	0.860	0.374	-0.140	10.446	
	CO <sub>2</sub> +Illu	0.851	0.386	0.071	4.881	
	Illu	0.848	0.390	-0.131	9.245	
	Illu+LP	0.847	0.392	-0.124	8.612	
	CO <sub>2</sub>	0.766	0.483	0.123	6.847	
Average	0.890	0.324	-0.011	4.271		
CART	LP+ACP	0.944	0.238	-0.003	0.324	✓
	Illu+LP+ACP	0.933	0.259	0.010	1.044	✓
	ACP	0.932	0.262	0.015	1.476	✓
	CO <sub>2</sub> +ACP	0.932	0.262	0.033	3.267	
	CO <sub>2</sub> +Illu+ACP	0.929	0.267	0.006	0.577	✓
	CO <sub>2</sub> +LP	0.919	0.286	-0.055	5.080	
	CO <sub>2</sub> +LP+ACP	0.919	0.286	-0.055	5.080	
	CO <sub>2</sub> +Illu+LP+ACP	0.919	0.286	-0.055	5.080	
	Illu+ACP	0.915	0.291	0.022	1.991	
	CO <sub>2</sub> +Illu+LP	0.877	0.351	0.002	0.110	✓
	LP	0.860	0.374	-0.140	10.446	
	Illu+LP	0.853	0.384	-0.109	7.644	
	CO <sub>2</sub> +Illu	0.850	0.388	0.073	4.960	
	Illu	0.801	0.447	-0.021	1.210	✓
	CO <sub>2</sub>	0.769	0.480	0.106	5.841	
Average	0.890	0.324	-0.011	3.609		

\*critical t-value=1.962 (DF: 671, confidence level: 95%)



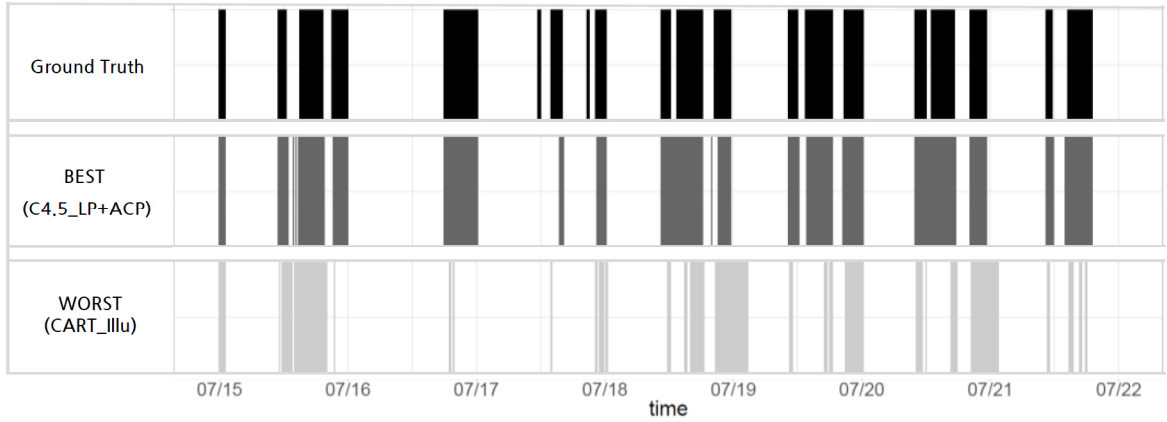


Fig. 3 Comparison of ground truth with best and worst models (for all predicted period)

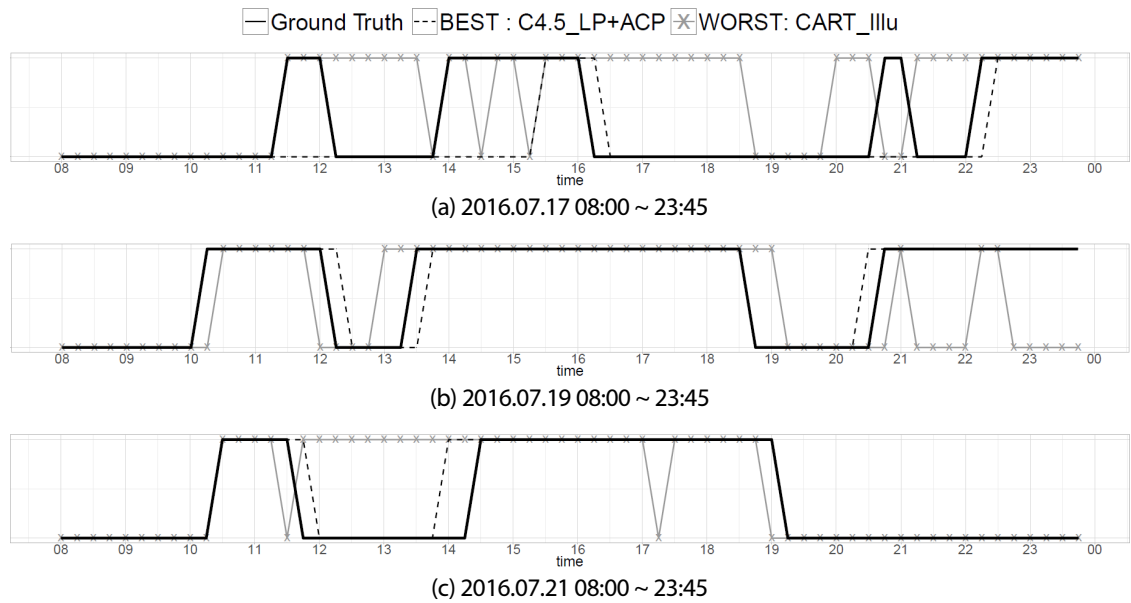


Fig. 4 Comparison of ground truth with best and worst models

Fig. 3과 Fig. 4는 정확도가 가장 높은 알고리즘(C4.5, LP+IIIU)과 가장 낮은 알고리즘(CART, IIIU)의 재실여부 추정결과를 적외선재실센서를 이용하여 측정한 재실데이터(ground truth)와 비교한 그래프이다. Fig. 3은 전체 시험기간에 대해 재실인 경우 어두운 색으로 표기하여 비교하였으며, Fig. 4는 하루 중 재실시간을 15분 단위로 추정한 상세 결과를 나타낸 것이다. 알고리즘이 사용하는 변수에 따라 재실추정 정확도의 차이는 상당히 크며, 실의 사용특성에 맞는 적절한 간접재실센서데이터를 사용할 경우 전체 추정기간에 걸쳐 약 95%의 정확성을 가지고 재실을 예측할 수 있는 것으로 나타났다.

### 3. 결론

본 연구에서는 실제로 사용되는 단위실에서 실내 환경 데이터(온도, 습도, CO<sub>2</sub>농도, 조도)와 에너지수요 데이터(조명, A/C전력)를 수집하였고 이를 Decision Tree 중 C4.5와 CART 알고리즘을 이용하여 재실정보를 추정하였고, 추정결과를 적외선재실센서를 이용하여 측정한 재실데이터(ground truth)와 비교하였다. 이를 통해 단위실의 재실추정에 대한 두 알고리즘의 성능을 분석하였고, 실제 사용 중인 공간에서 수집된 IoT 센서 데이터의 재실추정에 대한 활용성을 확인하였다.

- (1) 간접적으로 재실정보를 나타내는 데이터를 Decision Tree의 대표적인 알고리즘인 C4.5와 CART 알고리즘을 이용하여 단위실의 재실정보를 추정한 결과는 Table 5와 같다. 두 알고리즘을 이용한 재실추정결과, 전반적인 정확도는 유사하였으나 상위 3개의 추정모델에 대해서는 동일한 입력변수를 사용함에도 정확도는 C4.5가 CART보다 미세하지만 약간 높은 경향을 보였다.
- (2) 냉방기간 동안 재실추정에 대한 영향력은 사용되는 입력변수에 따라 큰 차이를 보였으며 에너지 수요데이터(조명전력, A/C전력)가 실내 환경 데이터보다(온도, 습도, CO<sub>2</sub>농도, 조도) 재실추정에 큰 영향력을 보였다. 단일변수로는 A/C전력 데이터가 정확도 가장 높은 5개의 추정모델에서 모두 사용된 입력변수로 가장 영향력이 큰 것으로 나타났다. 변수의 조합으로는 “조명전력+A/C전력”, “실내 조도+조명전력+A/C전력”, “A/C전력”이 정확도가 가장 높은 3개의 추정모델에 사용된 입력변수로 가장 영향력이 큰 것으로 나타나 단위실에서 서브미터링이 가능할 경우 재실 추정에 충분히 활용 가능성을 보여주었다.
- (3) 실내 CO<sub>2</sub>농도는 개별실의 재실정보추정에 비교적 영향력이 크지 않았다. 이는 창문의 개폐정도 및 외부풍속에 따른 불규칙적인 침기로 인해, 재실정보가 실내 CO<sub>2</sub>농도에 미친 영향력이 줄어들었기 때문으로 나타났다.

본 연구는 단위실의 재실추정에만 유효하다는 한계를 가지고 있으며 이러한 기반 연구를 바탕으로 보다 복잡한 Open Office 공간이나 다중이용공간의 재실예측 기술로 확장할 계획이다.

### 후기

이 논문은 2015년 정부(미래창조과학부)의 재원으로 국가과학기술연구회 융합연구단 사업(No. CRC-15-05-ETRI)의 지원을 받아 수행된 연구임.

### REFERENCES

1. User's Manual for ANSI/ASHRAE/IESNA Standard 90.1-2004, Appendix G. Building Performance Rating Method, pp. G39-G47.
2. Kim, H. I. and Yoon, G. Y., Effect of Occupancy and Lighting System Use Patterns on Building Energy Consumption, Journal of the Architectural Institute of Korea, Vol. 27, No. 5, pp. 229-236, 2011.

3. Park, S. L., Kim, J. H., Kim, D. W., and Park, C. S., Control Strategies of Cooling Systems in a Smart Home, *Journal of Korean Institute of Architectural Sustainable Environment and Building Systems*, pp. 257-260, 2011.
4. Simona D'Oca, Tianzhen Hong, Occupancy schedules learning process through a data mining framework, *Energy and Buildings*, Volume 88, pp.395-408, 2015.
5. Xin Liang, Liang X., Hong T., and Shen G. Q., Occupancy data analytics and prediction: A case study, *Building and Environment*, Volume 102, pp.179-192, 2016.
6. Han, H. T., Han, C. H., and Baek, C. I., Occupancy Estimation Based on Carbon Dioxide Concentration Using Dynamic Neural Network, *Korean Journal of Air-Conditioning and Refrigeration Engineering*, pp. 269-272, 2012.
7. Bae, W. B., Kim, Y. J., Mun, S. H., and Huh, J. H. , Prediction of Occupants based on Existing Wi-Fi Infrastructure, *Journal of the Architectural Institute of Korea*, Vol. 31, No. 11, pp. 211-219, 2015.
8. DOE, M&V Guidelines: Measurement and Verification for Performance-Based Contracts Version 4.0, Federal Energy Management Program, 2015.
9. Ryu, S. Ho. and Moon, H. J., Development of an Occupancy Prediction Model Using Indoor Environmental Data Based on Machine Learning Techniques, *Building and Environment*, Vol. 107, pp.1-9, 2016.
10. ISO, ISO 8601 Data Elements and Interchange Formats – Information Interchange – Representation of dates and times, 2004.
11. Jiawei Han, Micheline Kamber, Jian Pei, *Data Mining Concept and Techniques*, The Third Edition, Morgan Kaufmann, pp. 330-350, 2012.
12. Kevin P. Murphy, *Machine Learning A probabilistic Perspective*, The MIT Press, pp.544-551, 2012.
13. Quinlan, J. R. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
14. Breiman, Leo, Friedman, J. H., Olshen, R. A., Stone, C. J. (1984). *Classification and Regression Trees*. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software. ISBN 978-0-412-04841-8.
15. Stone, R. J., Improved Statistical Procedure for the Evaluation of Solar Radiation Estimation Models, *Solar Energy*, Vol. 51, No. 4, pp. 289-291, 1993.
16. Zhang, R., Lam, K. P., Chiou, Y. S., and Dong, B., Information-theoretic Environment Features Selection for Occupancy Detection in Open Office Spaces, *Building Simulation*, 5, pp. 179-188, 2012.