

지열 히트펌프 시스템의 데이터 마이닝 기반 성능 예측 기술

황민혜* · 박명규* · 전인기* · 손병후**†

* (주)브이피코리아, ** 한국건설기술연구원 건축도시연구소

Data Mining-Based Performance Prediction Technology of Geothermal Heat Pump System

Min Hye Hwang*, Myung Kyu Park*, In Ki Jun* and Byonghu Sohn**†

* VP Korea, Inc.

** Building and Urban Research Institute, Korea Institute of Civil Engineering and Building Technology

(Received December 21, 2015 ; Revised January 18, 2016 ; Accepted February 3, 2016)

Key Words: Geothermal Heat Pump System(지열 히트펌프 시스템), Data Mining(데이터 마이닝), Decision Tree Model(의사결정나무 모델), Performance Prediction(성능 예측), KDD(지식 발견 절차)

초록: 지열 시스템을 대상으로 데이터 마이닝 기반 성능 예측 모델을 구축하였다. 지열 시스템의 실시간 성능 분석과 예측에 필요한 데이터의 기본 조건을 검토한 후, 데이터베이스의 구조를 설계하였다. 먼저 시스템 성능계수(COP)와 전력 소비량을 분석 대상으로 설정한 후, 이들 물리량의 추출 주기(1분·5분·10분·30분·60분 간격)가 예측 결과에 미치는 영향을 분석하였다. 이어서 범주형과 수치형 의사결정나무 모델을 적용하여 시스템의 성능을 예측하였다. 범주형 의사결정나무 모델을 적용했을 때, 10분 주기의 예측 결과의 정확도는 97.7%로 가장 높았다. 또한 수치형 의사결정나무 분석 결과를 통해 COP가 변하는 순간의 임계값을 찾을 수 있었다. 본 논문에서 제안한 방법은 지열 시스템의 실시간 성능 분석과 운전 상태 등에 적용할 수 있을 것으로 판단된다.

Abstract: This preliminary study investigated data mining-based methods to assess and predict the performance of geothermal heat pump(GHP) system. Data mining is a key process of the knowledge discovery in database (KDD), which includes five steps: 1) Selection; 2) Pre-processing; 3) Transformation; 4) Analysis(data mining); and 5) Interpretation/Evaluation. We used two analysis models, categorical and numerical decision tree models to ascertain the patterns of performance(COP) and electrical consumption of the GHP system. Prior to applying the decision tree models, we statistically analyzed measurement database to determine the effect of sampling intervals on the system performance. Analysis results showed that 10-min sampling data for the performance analysis had highest accuracy of 97.7% over the actual dataset of the GHP system.

1. 서론

지열 히트펌프(geothermal heat pump) 시스템은 지열에너지(geothermal energy)를 활용하는 여러 분야 중 가장 큰 비중을 차지한다.⁽¹⁾ 지열 히트펌프 시스템(이하, ‘지열 시스템’)은 연중 거의 일정한 온도를 보이는 지반(ground)을 열원(heat source, 난방 운전) 또는 히트싱크(heat sink, 냉방 운전)로 활용한다. 따라서 화석연료를 사용하는 기존 건축 설비나 공기열원 히트펌프에 비해 건물 냉난방 에너지를 크게 줄일 수 있다. 또한 기존 설비와 비교했을 때 온실효과 가스를 적게 배출한다.

† Corresponding Author, byonghu@kict.re.kr

지열 시스템의 핵심은 지열 히트펌프 유닛과 지중열교환기(ground heat exchanger)다. 통상 지열 히트펌프로 불리지만, 시스템과 구분하기 위해 지열 히트펌프 유닛으로 표현하였다. 지중열교환기를 활용한다는 점이 공기열원 히트펌프나 기존 설비와 크게 다른 점이다. 여기서 지중열교환기는 지열 시스템의 난방 성능과 경제성에 영향을 미치는 중요한 요소이며, 아울러 초기 투자비를 높이는 요소이기도 하다. 하지만 지열 시스템은 초기 투자비가 큰 대신 운전비를 절감하여 경제성을 확보할 수 있다.

지열 시스템의 에너지(전력) 소비량을 최소화하는 최적 운전제어는 경제성 확보에 필수적이다. 나아가 시스템 상태를 실시간(real time)으로 진단하고 예측하는 일도 유지관리 측면에서 매우 중요하다. 최근 제어 방법이 시스템 성능에 미치는 영향을 분석한 결과들이 발표되었다.^(2,3) Cervera-Vazquez 등⁽²⁾은 지중열교환기 순환수 펌프를 다단 기동/정지(Multi-stage ON/OFF) 방법으로 제어하면 단단(single-stage) 제어일 때보다 에너지 소비량을 33% 줄일 수 있다고 밝혔다. Jung 등⁽³⁾은 순차 제어(sequential control)가 지열 시스템의 성능을 높이는 데 효과적이라고 밝혔다. 최근 들어 정보통신기술(ICT)을 난방 시스템에 접목하여 시스템의 에너지 소비량과 운전 상태(pattern) 등을 쉽게 파악할 수 있는 기술이 주목받고 있다.^(4,5) Ren 등⁽⁵⁾은 지식 발견(Knowledge Discovery in Database, KDD) 절차와 데이터 마이닝(data mining) 기법⁽⁶⁾을 활용하여 주택용 난방 설비의 성능을 분석하였다.

지식 발견 절차(이하 ‘KDD’)와 데이터 마이닝 기법(이하 ‘데이터 마이닝’)을 지열 시스템에 적용하면, 시스템의 에너지 소비량, 정상 또는 이상(failure) 운전 상태 등을 실시간으로 얻을 수 있다. 통상 지열 시스템에서는 최소 수준의 센서와 모니터링 장비로 기본적인 데이터만 측정한다. 따라서 분석에 필요한 최소 데이터 수준과 분석 대상 설정, 분석용 데이터의 추출 간격, 분석 모델 등 데이터 마이닝에 필수적인 사항들을 사전에 검증해야 한다. Fig 1에서 보듯이 KDD는 추출(selection), 전처리(pre-processing), 변환(transformation), 분석(data mining), 해석/평가(interpretation/evaluation) 과정을 거친다. 여기서 분석, 즉 데이터 마이닝은 KDD의 핵심이라고 할 수 있다. 전처리 과정이나 해석/평가도 넓은 의미에서 데이터 분석에 해당하기 때문에, 데이터 마이닝을 KDD의 전 과정을 포괄하는 개념으로 보는 이들도 있다.⁽⁷⁾

본 연구는 지열 시스템의 난방 측정 데이터를 지식 발견 절차⁽⁶⁾에 적용하여 성능 예측 모델을 구축하였다. 또한 데이터 마이닝 모델 중, 범주형 의사결정나무(categorical decision tree) 모델과 수치형 의사결정나무(numerical decision tree) 모델을 이용하여 시스템의 성능 저하 규칙을 도출하였다. 마지막으로 데이터 마이닝 모델로 예측한 결과와 실측 결과를 비교하였다.

2. 데이터 마이닝 기반 성능 예측

2.1 통계 분석을 통한 전처리

지식 발견 절차(KDD)는 대규모 데이터베이스를 구축한 후, 이를 다양한 데이터 마이닝 모델에 적용하

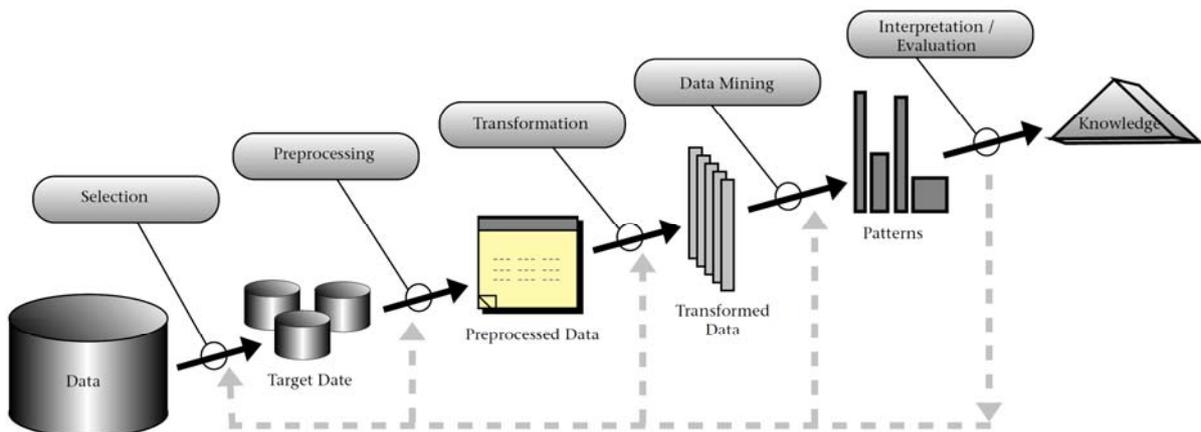


Fig. 1 Graphical representation of the Knowledge Discovery in Databases(KDD) process⁽⁶⁾

여 특정 데이터의 추세를 분석하거나 예측한다. 분석에 앞서 특정 데이터(분석 대상)를 데이터베이스에서 추출한 후 분석 주기(interval)를 결정하는데, 이는 전처리 과정으로 불린다. 데이터에 따라 결과가 달라지기 때문에, 전처리 과정은 매우 중요하다. 전처리는 잡음(noise)이 포함된 데이터, 측정에서 누락된 데이터, 일관성이 부족한 데이터 등을 분석이 가능한 상태로 변환하는 작업이다. 이러한 변환 작업에는 정제, 통합, 변환, 축소 등이 있다.

2.2 의사결정나무 분석(decision tree analysis)

의사결정나무 분석은 의사결정 규칙(decision rule)을 나무(tree) 구조로 도표화하여 분류하거나 예측을 수행하는 계량적 분석법이다. 의사결정 규칙에 따라 데이터를 나무 구조로 표현하기 때문에, 다른 데이터 마이닝 방법에 비해 그 과정을 쉽게 이해할 수 있다. 뿌리 마디는 의사결정나무가 시작되는 마디로 전체 데이터로 구성되며, 자식 마디는 한 가지 마디에서 2개 이상의 마디로 구분된 마디다. 잎은 각각의 줄기에 최종적으로 위치한 마디로, 잎의 수만큼 규칙을 만든다. 가지는 하나의 뿌리마디에서 잎까지 연결된 마디들이며 가지를 이루는 마디의 수를 깊이라고 한다.

의사결정나무를 생성할 때 분리 기준(split criterion)과 정지 규칙(stopping rule)을 적절하게 지정해야 한다. 이때 세분화와 분류, 예측, 차원 축소와 변수 선택, 변수 간 상호 영향 파악 등 분석 목적에 따라 지정한다. 다음으로 분류 오류(classification error)를 높일 수 있는 가지나 추론 규칙이 부적절한 가지를 제거한다. 마지막으로 의사결정나무 모델의 적합 정도를 검증한 후 최종 모델을 구축한다. 지열 시스템의 실측 데이터는 서로 영향을 주고받으며 분포도 일정하지 않다. 따라서 데이터 간 상호 작용 효과와 비모수적 모형 해석이 쉬운 의사결정나무 분석을 실측 데이터에 적용할 수 있다.

2.2.1 범주형 의사결정나무 분석(categorical decision tree analysis)

지열 시스템의 성능 변화를 예측하기 위해, 데이터 마이닝 알고리즘의 하나인 범주형 의사결정나무 모델을 적용하여 분석하였다. 이 분석법은 목표 변수(분석 대상)가 이산형일 때 효과적이며, 목표 변수를 몇 개의 등급(범주)으로 분류한다. 다음으로 규칙을 찾은 후, 이 규칙을 이용하여 성능을 예측한다. 범주형 의사결정나무 모델은 분류 예측 정확도(%)와 평균절대오차(mean absolute error, MAE)로 평가할 수 있다. 분류 예측 정확도는 숫자가 클수록 모델의 적합도가 우수한 반면, 평균제곱오차는 반대다. 평균절대오차는 식 (1)로 계산하며, e_i 는 실제 값과 예측 값의 차이이다.

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (1)$$

2.2.2 수치형 의사결정나무 분석(numerical decision tree analysis)

수치형 의사결정나무 분석은 목표 변수가 연속적일 때 효과적이다. 목표 변수를 대상으로 분산 분석의 F 통계량과 분산의 감소량 등을 적용하여 데이터를 분리한다. 여기서 F 통계량은 집단 간 분산 대 집단 내 분산의 비율을 의미한다. 또한 각 마디들을 형성할 때 노드가 만들어지는데, 이때 노드에 속하는 자료의 순수도(purity)가 가장 크게 증가하도록 재귀적으로 진행된다.

본 연구에서는 지열 시스템의 성능 변화를 예측하기 위해 수치형 의사결정나무 분석을 수행하였다. 분석을 통해 시스템의 성능(COP) 변화 원인에 대한 주요 규칙을 도출하였다. 수치형 의사결정나무 분석에서 주요 규칙은 구체적인 수치(임계값)로 표현된다. 일반적으로 시스템의 전력 소비량이 성능에 가장 크게 영향을 미치기 때문에, 이를 종속변수로 설정하였다. 측정 데이터에서 규칙을 찾아내고, 이를 이용하여 시스템의 성능을 예측하는 것이 본 분석의 목적이다. 이때 예측 모델에 의한 결과와 측정 데이터의 적합 정도는 식 (2)의 결정계수(R^2)로 판단하였으며, 결정계수는 0.6이었다. 식 (2)에서 y_i 는 측정값이고 \bar{y}_i 는 측정값의 평균이다. 또한 결정계수를 이용하여 분할 횟수(number of splits)를 5로 설정하였다.

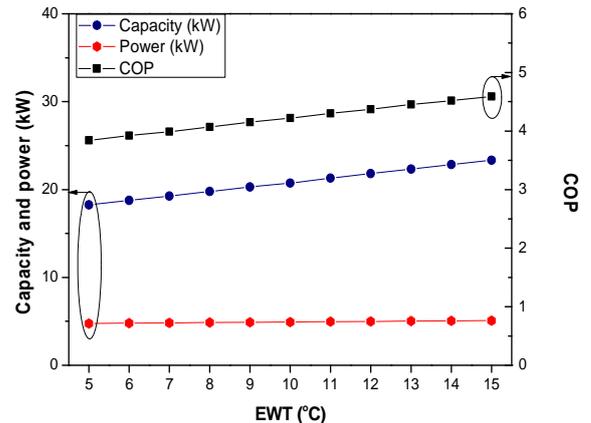
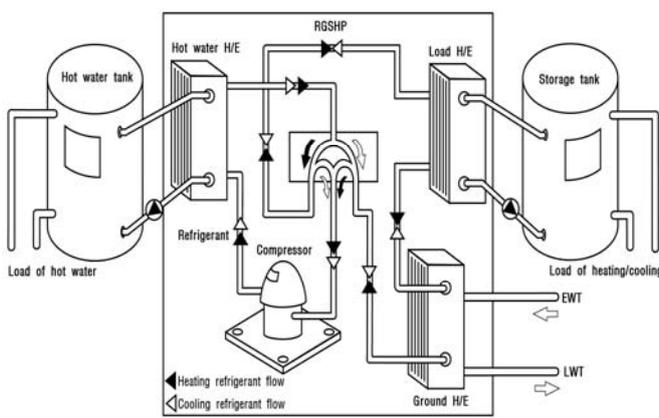
$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y}_i)^2} \tag{2}$$

3. 결과 및 고찰

3.1 통계 분석을 통한 전처리 결과

앞서 기술한 데이터 마이닝 기법을 지열 시스템에 적용하기 위해, H대학 기숙사(충남 아산시)의 난방 실측 데이터⁽⁸⁾를 이용하였다. Fig. 2는 실측용 시스템의 구성도와 지열 히트펌프의 난방 성능 데이터를 보여준다. 시스템 운전 중 지중열교환기 순환수의 온도와 유량, 부하측 순환수의 온도와 유량, 히트펌프와 전체 시스템의 전력 소비량 등을 1분 간격으로 측정하였다. Fig. 3은 측정 결과를 이용하여 시스템의 난방 성능을 분석한 결과다. 실내외 온도차는 평균 33.7°C이었으며, 난방 성능계수는 3.30~4.64(일평균 3.83)이었다. 난방 용량은 평균 21.7 kW, 지중 열교환기가 취득하는 열량은 평균 16.2 kW이었다. 실측 데이터를 데이터베이스로 구축한 후, 난방 성능계수와 전력 소비량을 분석 대상(목표)으로 선정하였다.

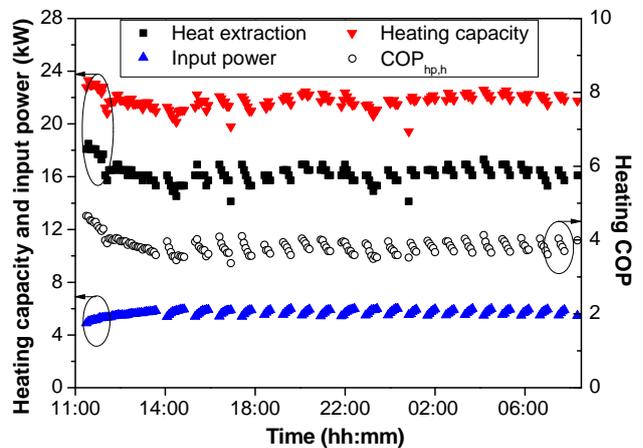
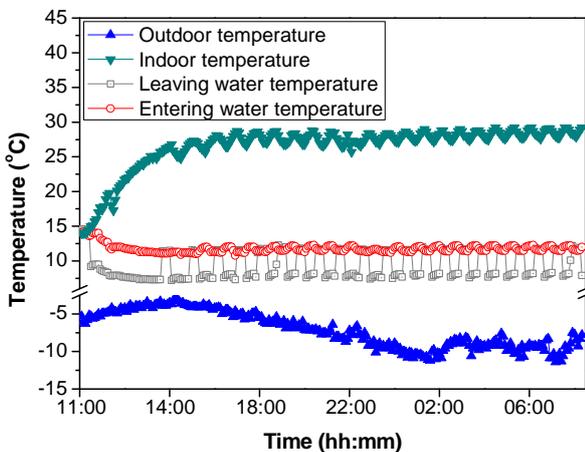
의사결정나무 분석 결과는 선정된 데이터의 품질, 특히 분석용 데이터의 추출 주기(sampling interval)에 크게 영향을 받는다. 따라서 본 연구에서는 원 데이터(raw data)에서 추출 주기(1분, 5분, 10분, 30분, 60분 간격)를 다르게 했을 때, 시스템의 난방 COP와 전력 소비량 변화를 비교하였다. 추출 주기에 상관없



(a) System diagram

(b) Heating performance of heat pump

Fig. 2 Schematic diagram of GHP system and heating performance of heat pump⁽⁸⁾



(a) Outdoor/indoor temperature, EWT, and LWT

(b) Heating performance

Fig. 3 Temperature variation and heating performance of GHP system on Jan. 25~Jan. 26, 2013⁽⁸⁾

Table 1 Statistical analysis results from the various dataset of different sampling interval

Sampling intervals	Values	Power consumption		COP	
		Heat pump	Circulation pump	Heat pump	System
1-min.	Average	0.4999579	0.9112452	0.3249556	1.3103149
	Standard deviation	1.6870079	0.0564369	1.9322985	3.9568772
	Min. value	0	0	0	0
	Max. value	6.98566	1.03746	74.9466	19.4099
5-min.	Average	0.4988184	0.9110747	0.3377358	1.3025885
	Standard deviation	1.6855154	0.0572185	2.1762941	3.9455068
	Min. value	0	0	0	0
	Max. value	6.93	1.04	74.95	19.41
10-min.	Average	0.4954133	0.9112266	0.3648983	1.3114286
	Standard deviation	1.6790868	0.0573449	2.5524101	3.961531
	Min. value	0	0	0	0
	Max. value	6.91	1.04	71.59	19.01
30-min.	Average	0.4859612	0.912065	0.3198573	1.3346373
	Standard deviation	1.6659274	0.056777	2.1269555	3.9928431
	Min. value	0	0.75	0	0
	Max. value	4.91	1.04	71.59	18.29
60-min.	Average	0.4939937	0.911149	0.3257845	1.3457924
	Standard deviation	1.67252617	0.0557341	2.1944776	4.01265
	Min. value	0	0.75	0	0
	Max. value	6.83	1.04	71.59	17.85

이, COP와 전력 소비량 데이터의 분포는 유사하였다. 이때 평균(average)과 표준편차(standard deviation)로 데이터 분포의 유사도를 판단하였다.(Table 1)

3.2 범주형 의사결정나무 분석 결과

전처리 과정에서 추출한 각각의 데이터 세트(1분·5분·10분·30분·60분 간격)를 범주형 의사결정나무 모델에 적용하여 지열 시스템의 난방 COP를 예측하였다. 먼저 실측 데이터의 변화 정도를 감안하여 범주형 의사결정나무 모델을 생성하였다. 이때 수준에 따라 난방 COP를 Table 2와 같이 구분하였다. 시스템 정지 상태(OFF)는 COP를 0으로 처리하였고, 운전(ON) 중에는 평균과 표준편차를 이용하여 구간(Too Low, Medium, High)을 나누었다. 마지막으로 COP가 5보다 큰 경우는 예외 상황(Too High)으로 분류하였다.

분석 결과에 대한 예측 값의 정확도를 Table 3에 정리하였다. 전체적으로 정확도는 95% 이상이었으며, 데이터 추출 주기에 따라 다소 차이를 보였다. 표에서 보듯이, 추출 주기가 10분일 때 정확도가 97.7%로 가장 높았으며, 이때 평균절대오차(MAE)는 0.0138이었다. 이는 매우 작은 값으로 간주되며, 따라서 범주형 의사결정나무 모델이 실측 데이터를 잘 예측한 것으로 판단된다.

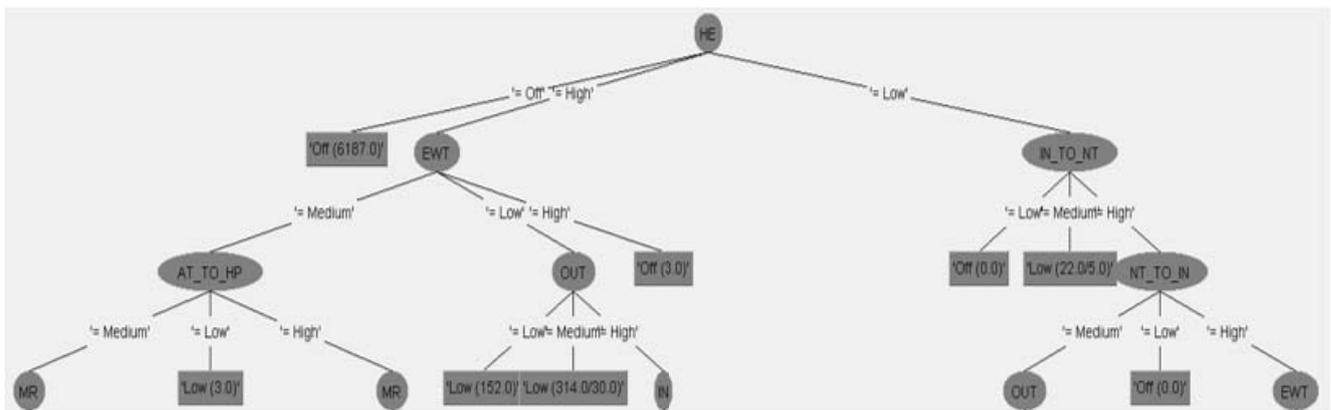
Table 2 Criteria for classification of category

Level	Range
Off	COP = 0
Too Low	$0 < COP < \bar{x} - \sigma_x$ *
Medium	$\bar{x} - \sigma_x < COP < \bar{x} + \sigma_x$
High	$COP > \bar{x} + \sigma_x$
Too High	COP > 5

* \bar{x} is average and σ_x is standard deviation.

Table 3 Prediction rate of decision tree

Intervals	Prediction rate
1-min.	95.8%
5-min.	95.8%
10-min.	97.7%
30-min.	96.7%
60-min.	94.6%



=== Summary ===

Correctly Classified Instances	7397	97.6631 %	=== Confusion Matrix ===				
Incorrectly Classified Instances	177	2.3369 %					
Kappa statistic	0.8593		a	b	c	d	e
Mean absolute error	0.0138		6807	68	0	1	0
Root mean squared error	0.0867		82	577	4	1	0
Relative absolute error	20.4616 %		0	13	7	0	0
Root relative squared error	47.2751 %		4	3	0	6	0
Total Number of Instances	7574		0	1	0	0	0

<-- classified as
 a = Off
 b = Low
 c = Medium
 d = Extreme
 e = High

Fig. 4 Categorical decision tree model using a dataset of 10-min. sampling interval

각각의 데이터 세트를 이용하여 범주형 의사결정나무 모델을 생성하였다. Fig. 4는 예측 정확도가 가장 높은 주기(10분 간격)의 데이터 세트로 만든 모델이다. 의사결정나무의 총 가지 수(total number of branches)는 66개이었으며 이를 다시 정리하여 45개 규칙으로 표현할 수 있었다. 마지막으로 45개 규칙에서 추론 규칙이 적절치 않은 가지들을 제거한 다음 Table 4의 주요 규칙을 도출하였다. 표에서 보듯이, 규칙에 적용된 데이터는 실내의 온도, 지중 순환수의 히트펌프 입구 온도(EWT), 히트펌프와 순환펌프의 전력 소비량 등이다.

3.3 수치형 의사결정나무 분석 결과

지열 시스템의 성능 저하 원인을 탐지하기 위해 범주형 의사결정나무 분석 결과를 토대로 COP 이상을 보이는 데이터(Too Low와 Too High)만 가지고 수치형 의사결정나무 분석을 진행하였다. Fig. 5에서 빨간색 막대로 표시한 부분은 히트펌프의 전력 소비량이 큰 데이터이며, 파란색 막대로 표시한 부분은 전력 소비량이 작은 데이터의 수를 의미한다. 특정 조건에서 히트펌프의 전력 소비량이 모두 높거나 낮

Table 4 Primary decision rules for COP prediction

	Decision rules
1	If AHP is High and EWT is Too Low and OAT is Too Low or OAT is Medium then COP is Too Low
2	If AHP is High and EWT is Too Low and OAT is High and IAT is Too Low or IAT is Medium then COP is Too Low
3	If AHP is Too Low and Room to Buffer tank is Medium then COP is Too Low
4	If AHP is Too Low and Room to Buffer tank is Medium and IAT is High then COP is Too Low
5	If AHP is Too Low and Room to Buffer tank is High and OAT is Medium and EWT is Too Low and Heat pump to Buffer tank is Medium and IAT is Medium and MRT is High then COP is Too Low
6	If AHP is Too Low and Room to Buffer tank is High and OAT is Medium and EWT is Too Low and Heat pump to Buffer tank is Medium and IAT is High and MRT is Medium then COP is Too Low
7	If AHP is Too Low and Room to Buffer tank is High and EWT is Too Low then COP is Too High
8	If AHP is High and EWT is Too Low and OAT is Medium or OAT is Too Low then COP is Too Low
9	If AHP is High and EWT is Too Low and OAT is High and IAT is High or IAT is Too Low then COP is Too Low
10	If AHP is High and EWT is Medium and Hot water tank to Heat pump is Too Low then COP is Too Low
11	If AHP is High and EWT is Medium and Hot water tank to Heat pump is High and MRT is Medium and VGHEP is Medium and OAT is Too Low or OAT is Medium then COP is Too Low
12	If AHP is High and EWT is Medium and Hot water tank to Heat pump is Medium and MRT is High then COP is Too Low
13	If AHP is High and EWT is Medium and Hot water tank to Heat pump is Medium and MRT is Medium and VGHEP is Too Low or VGHEP is High then COP is Too Low
14	If AHP is High and EWT is Medium and Hot water tank to Heat pump is Medium and MRT is Too Low and Room to Buffer tank is Too Low then COP is Too Low

* AHP: Ampere of heat pump, EWT: entering water temperature, OAT: outdoor air temperature, IAT: indoor air temperature, MRT: mechanical room temperature, VGHEP: voltage of circulation pump of ground heat exchanger

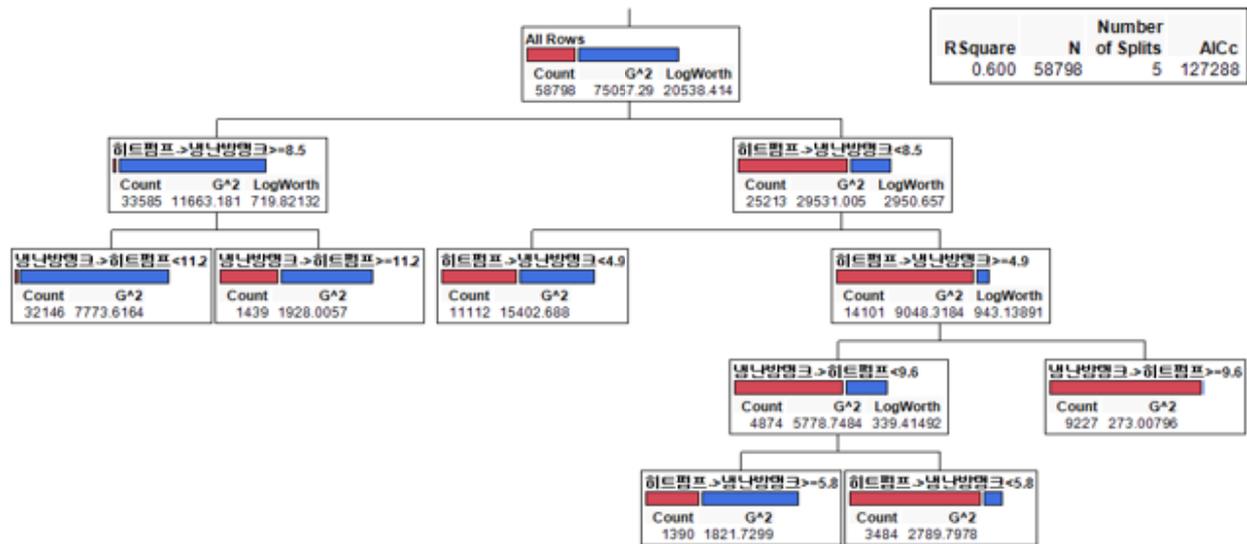


Fig. 5 Data mining process using numerical decision tree model and 10-min sampling data

음을 알 수 있다.

이를 바탕으로 수치형 의사결정나무 모델에서 임계값을 찾아 Table 5의 규칙을 정리하였다. 여기서 임계값은 규칙에서 나타나는 수치를 의미한다. 앞선 범주형 의사결정나무 분석에서는 Too Low와 Too

Table 5 Primary numerical rules for COP prediction

	Numerical rules
1	If heat pump to buffer tank $\geq 8.5^{\circ}\text{C}$ and buffer tank to heat pump $< 11.2^{\circ}\text{C}$ then COP is Too Low
2	If heat pump to buffer tank $< 8.5^{\circ}\text{C}$ and heat pump to buffer tank $\geq 4.9^{\circ}\text{C}$ and buffer tank to heat pump $\geq 9.6^{\circ}\text{C}$ then COP is optimized
3	If heat pump to buffer tank $\geq 47.6^{\circ}\text{C}$ and EWT $< 13.2^{\circ}\text{C}$ then COP is Too High

High 등으로 나타나는 반면, 수치형 의사결정나무 분석에서는 해당 규칙의 결과가 구체적인 수치로 나타난다. 즉, 규칙에 의해 COP가 변하는 순간의 임계값을 찾을 수 있다는 의미와 같다.

4. 결 론

지열 시스템을 대상으로 데이터 마이닝 기반 성능 예측 모델을 구축하였다. 지열 시스템의 실시간 성능 분석과 예측에 필요한 데이터의 기본 조건을 검토한 후, 데이터베이스의 구조를 설계하였다. 먼저 시스템 성능계수(COP)와 전력 소비량을 분석 대상으로 설정한 후, 이들 물리량의 추출 주기(1분-5분-10분-30분-60분 간격)가 예측 결과에 미치는 영향을 분석하였다. 이어서 범주형과 수치형 의사결정나무 모델을 적용하여 시스템의 성능을 예측하였다. 범주형 의사결정나무 모델을 적용했을 때, 10분 단위의 예측 결과의 정확도는 97.7%로 가장 높았다. 또한 수치형 의사결정나무 분석을 통해 COP 분석 결과가 변하는 순간의 임계값을 찾을 수 있었다. 본 논문에서 제안한 방법은 지열 시스템의 실시간 성능 분석과 이상(failure) 운전 예측 등에 적용할 수 있을 것으로 판단된다. 다양한 현장의 장기간 측정 데이터를 이용하여 분석 기법을 검증할 예정이다.

후 기

본 연구는 2013년도 산업통상자원부의 재원으로 한국에너지기술평가원의 지원을 받아 수행한 연구과제입니다.(No. 20133010111750)

참고문헌 (References)

- (1) Lund, J. W. and Boyd, T. L., 2015, "Direct Utilization of Geothermal Energy 2015 Worldwide Review," *Proceedings World Geothermal Congress 2015*, Melbourne, Australia, pp. pp. 1~31.
- (2) Cervera-Vazquez, J., Montagud, C. and Corberan J. M., 2015, "In Situ Optimization Methodology for Ground Source Heat Pump Systems: Upgrade to Ensure User Comfort," *Energy and Buildings*, Vol. 109, pp. 195~208.
- (3) Jung, Y., Jo, J., Kim, Y. and Cho, Y., 2014, "A Study on the Geothermal Heat Pump System Performance Analysis According to Water Flow Rate of the Geothermal Water Circulation Pump," *Journal of the Korean Solar Energy Society*, Vol. 34, No. 6, pp. 103~109.
- (4) He, X., Zhang, Z. and Kusiak, A., 2014, "Performance Optimization of HVAC Systems with Computational Intelligence Algorithms," *Energy and Buildings*, Vol. 81, pp. 371~380.
- (5) Ren, X., Yan, D. and Hong, T., 2015, "Data Mining of Space Heating System Performance in Affordable Housing," *Building and Environment*, Vol. 89, pp. 1~13.
- (6) Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P., 1996, "From Data Mining to Knowledge Discovery in Databases," *AI Magazine*, Vol. 17, No. 3, pp. 37~54.
- (7) Han, J., Kamber, M. and Pei, J., 2011, *Data Mining: Concepts and Techniques*(Third Edition), Morgan Kaufmann, Burlington, MA.
- (8) Kong, H. J., Kang, S. J., Yun, K. S. and Lim, H. J., 2013, "An Experimental Study on the Cooling and Heating Performance of a Residential Ground Source Heat Pump System," *Korean Journal of Air-Conditioning and Refrigeration Engineering*, Vol. 25, No. 3, pp. 156~163.