

# An Evaluation of the Suitability of Data Mining Algorithms for Smart-Home Intelligent-Service Platforms

Kilhwan Kim\* · Changsup Keum\*\*† · Ki-Sook Chung\*\*

\*Department of Management Engineering, Sangmyung University

\*\*Hyper-connected Communication Research Laboratory, ETRI

## 스마트홈 지능형 서비스 플랫폼을 위한 데이터 마이닝 기법에 대한 적합도 평가

김길환\* · 금창섭\*\*† · 정기숙\*\*

\*상명대학교 경영공학과

\*\*한국전자통신연구원 초연결통신연구소

In order to implement the smart home environment, we need an intelligence service platform that learns the user's life style and behavioral patterns, and recommends appropriate services to the user. The intelligence service platform should embed a couple of effective and efficient data mining algorithms for learning from the data that is gathered from the smart home environment. In this study, we evaluate the suitability of data mining algorithms for smart home intelligent service platforms. In order to do this, we first develop an intelligent service scenario for smart home environment, which is utilized to derive functional and technical requirements for data mining algorithms that is equipped in the smart home intelligent service platform. We then evaluate the suitability of several data mining algorithms by employing the analytic hierarchy process technique. Applying the analytical hierarchy process technique, we first score the importance of functional and technical requirements through a hierarchical structure of pairwise comparisons made by experts, and then assess the suitability of data mining algorithms for each functional and technical requirements. There are several studies for smart home service and platforms, but most of the study have focused on a certain smart home service or a certain service platform implementation. In this study, we focus on the general requirements and suitability of data mining algorithms themselves that are equipped in smart home intelligent service platform. As a result, we provide a general guideline to choose appropriate data mining techniques when building a smart home intelligent service platform.

**Keywords** : Smart Home, Service Platform, Intelligent Service, Data Mining, AHP Technique

### 1. 서 론

스마트홈 환경에서 거주자의 라이프스타일과 행동 특성을 학습하여 지능형 서비스를 추천하기 위한 다양한 연구

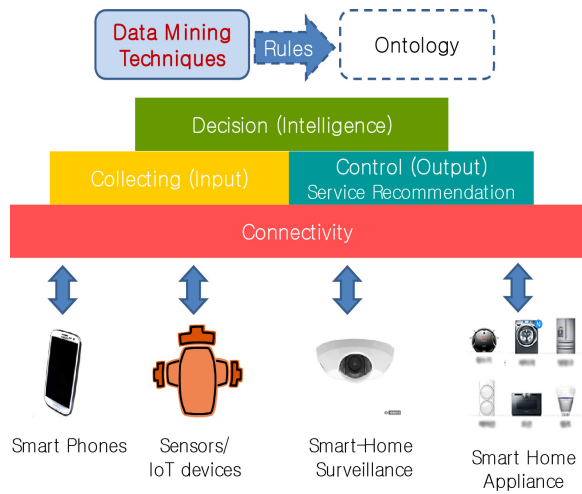
가 지속되고 있다[1-4, 6, 7, 13, 14, 17]. 이러한 연구는 스마트홈의 공기 질의 관리나 온도 조절[3, 13], 거주자의 이상 활동 감지[7], 스마트홈의 에너지 사용 조절[5, 8], 노인을 위한 지능형 서비스[14]처럼 스마트홈 관련 특정 서비스나 특정 종류의 센서 정보에 한정된 연구가 대부분이다.

그런데 스마트홈 환경에서 지능형 서비스를 수행하기 위해서는 스마트 단말, 센서, 가전제품 등에서 수많은 데

Received 12 April 2017; Finally Revised 14 June 2017;

Accepted 15 June 2017

† Corresponding Author : cskeum@etri.re.kr



<Figure 1> Conceptual Architecture of Smart-Home Service Platform

이 데이터를 수집해야 하며, 거주자의 행동 패턴의 특성과 라이프스타일 등을 학습하는 복잡한 과정이 공통적으로 요구된다. 따라서 스마트홈 환경에서 특정 서비스를 제공하기 위해 별도의 시스템을 개발하는 것보다는, 거주자의 필요에 따라 다양한 서비스가 추가될 수 있는 스마트홈 지능형 서비스 플랫폼이 구축하는 것이 바람직하다. 왜냐하면 스마트홈을 위한 서비스 플랫폼이 구축되면 플랫폼에서 거주자에 대한 폭넓은 데이터가 동일한 방식으로 수집되어 공통된 행동 특성과 라이프스타일을 예측한 후, 이를 개별 서비스에서 공통적으로 활용할 수 있기 때문이다. 이러한 이유로 스마트홈 지능형 서비스 플랫폼을 개발하고자 하는 연구들이 있어 왔다[1, 2, 6].

<Figure 1>은 스마트홈 지능형 서비스 플랫폼을 개념적으로 모형화한 것이다. <Figure 1>에서 보듯이 서비스 플랫폼은 스마트홈 환경 내의 다양한 스마트 단말, 센서 및 각종 IoT(Internet of Things) 기기, 스마트홈 관제 기기(방법 카메라 등), 스마트 가전 등과 연결될 수 있어야 한다. 또한 스마트홈 환경에서 지능형 서비스 플랫폼은 다양한 기기가 추가되고 사라지는 동적인 환경에서 끊임 없는 연결성을 유지할 수 있어야 한다. 또한 연결된 기기로부터 데이터를 수집하고, 적절한 방식으로 이러한 기기를 조정할 수 있는 메커니즘을 확보해야 할 것이다.

또한 스마트홈 지능형 서비스 플랫폼은 수집된 데이터를 종합적으로 분석하여 사용자의 행동 패턴을 예측하여 적절하게 서비스를 추천할 수 있는 서비스 추천 모형을 생성할 수 있는 기능이 필요하다. 그리고 이를 위해서는 스마트홈 환경에서 발생할 수 있는 다양한 서비스 추천 모형을 생성할 수 있는 범용형 데이터 마이닝 기법의 플랫폼 탑재가 필요하다.

지금까지 스마트홈 지능형 서비스 플랫폼에 대한 연구

는 주로 기기와의 연결성, 데이터 수집, 서비스 조절 메커니즘 측면에서 주로 연구되었고[1, 2, 6], 서비스 플랫폼의 지능이라고 할 수 있는 서비스 추천 모형을 생성하는 방식에 대한 연구는 적었다. 특히 서비스 추천 모형을 생성하는데 적합한 데이터 마이닝 기법에 대한 연구는 없었다. 유사한 연구로는 IoT 환경 전반이나 이상탐지에 적합한 데이터 마이닝 기법에 대한 연구가 있었다[8, 10, 15]. 그러나 스마트홈 지능형 서비스 플랫폼에 한정된 연구는 없었다.

본 연구에서는 스마트홈 지능형 서비스 플랫폼에서 사용하기에 적합한 데이터 마이닝 기법을 평가하는 것을 연구목표로 한다. 이를 위해 본 연구는 다음 4가지 내용을 수행한다.

첫째, 기존의 스마트홈 환경에서 지능형 서비스를 수행하는 시나리오를 개발한다.

둘째, 개발된 시나리오를 기반으로 스마트홈 지능형 서비스 플랫폼에 탑재될 데이터 마이닝 기법에 대한 기능 요구사항과 기술 요구사항을 도출한다.

셋째, AHP(Analytic Hierarchy Process) 기법을 이용하여 데이터 마이닝 기법에 대해 스마트홈 지능형 서비스 플랫폼에서의 적합도에 대한 평가를 수행한다.

이를 통해 스마트홈 지능형 서비스 플랫폼을 구축할 때 적합한 데이터 마이닝 기법을 선택하는 문제에 대한 가이드라인을 제공하고자 한다.

## 2. 스마트홈 지능형 서비스 시나리오

### 2.1 스마트홈 지능형 서비스 플랫폼

스마트홈은 가정 내 기기들을 네트워크로 연동해 스마트 단말이나 PC 등을 통해 원격으로 모니터링, 제어하고, 지능형 서비스를 제공하는 서비스와 솔루션을 총칭하는 말이다. 사실 스마트홈은 2000년대 초반 인터넷이 가정에 보급되면서 시작된 시장이지만 매우 더디게 성장하였다. 그러나 최근 몇 년 가정에서 스마트 단말 이용이 확대되고, 아울러 무선통신 기술, 센서 기술, IoT 관련 기술이 발전함에 따라, 서비스와 기기 사이의 연결을 위한 구축비용이 대폭 감소하였다. 따라서 스마트홈 시장은 얼리 어답터를 위한 니치 시장에서 매스 마켓으로 급격히 변화하는 조짐을 보이고 있다.

스마트홈의 관련 기술은 크게 다음 6가지 분야로 나눌 수 있다[9].

- 유무선 네트워크 : 가정 내의 단말, 가전 기기, 센서 등을 연결할 수 있는 유무선 네트워크 인프라 관련 기술
- 스마트 기기 : IoT 통신이 가능한 가전, 기기 및 센서 관련 기술

- IoT 통신 표준 : 스마트 기기 간의 원활한 연결을 위한 표준화 기술
- 스마트홈 서비스 플랫폼 : 통신 인프라, 기기 등을 운영하고 제어할 수 있는 홈허브 역할을 하는 플랫폼 기술
- 제어 단말 : 사용자의 편의성과 UX를 고려한 제어 단말 관련 기술
- 서비스 및 콘텐츠 : 사용자의 니즈에 맞는 킬러 서비스 및 콘텐츠 관련 기술

위의 6가지 기술 중 하나인 스마트홈 서비스 플랫폼은 스마트 기기와 서비스를 제어하는 중앙서버이자 기기와 서비스를 이어주는 통로 역할을 수행한다. 플랫폼 기술이 안착화되면 해당 기술을 중심으로 다른 부가적인 서비스와 기술이 개발되어 연동되는 구조이므로 글로벌 IT 회사, 가전업체, 통신 업체들이 관련된 기술의 개발에 박차를 가하고 있다[9].

## 2.2 스마트홈 지능형 서비스 시나리오

현재 스마트홈 서비스를 표방하고 있는 사업자는 다양한 관점에서 스마트홈 서비스를 표방하고 있다. 본 연구에서는 이러한 서비스를 거주자가 집이라는 공간에서 기대하는 핵심 기대가치를 중심으로 5가지 분야로 나누어 보았다.

- 안전한 스마트홈 서비스 분야 : 보안 서비스, 노인 돌봄 서비스, 어린이 돌봄 서비스 등의 안전을 위한 서비스
- 편안한 스마트홈 서비스 분야 : 쾌적한 실내 환경 유지를 위한 사용자 상태와 환경을 고려한 자동 조도, 습도, 온도, 공기청정도 조절 서비스
- 즐거운 스마트홈 서비스(엔터테인먼트) : 사용자의 즐거운 휴식을 위한 멀티미디어 및 콘텐츠 추천 서비스
- 에너지 효율적인 스마트홈 서비스 : 냉난방, 조명 등을 자동 조절하여 에너지 절감
- 우아한 스마트홈 서비스 : 음식 보관 및 조리, 식기 세척 및 보관의 지능화, 의류 세탁 및 보관의 지능화 등의 가사일의 자동화를 넘어선 지능화 서비스

본 연구에서는 이러한 서비스 영역 중 ‘편안한 스마트홈’과 ‘에너지 효율적인 스마트홈’ 서비스를 결합한 서비스 시나리오를 구체화해 본다. 서비스의 범위로는 집의 온도, 습도, 청정도, 조도 등을 적절히 조절하는 지능형 추천 서비스로 한정한다.

이를 위해 서비스 플랫폼은 온도, 습도, 청정도, 조도 등을 조절하기 위해서는 다음과 같은 기기에 대한 제어를 수행한다고 가정한다.

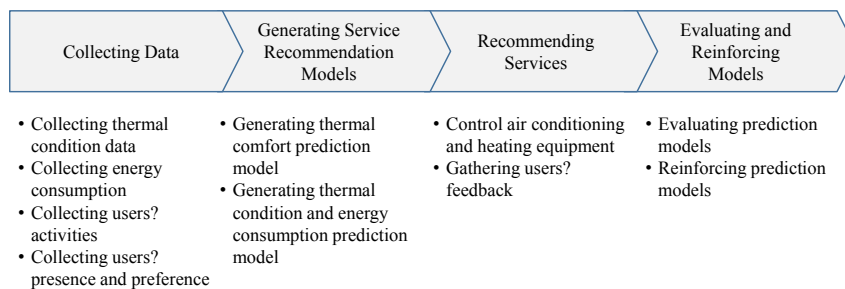
- 온/습도/청정도 가전제품 : 에어컨, 공기청정기, 가습기 등
- 온도 열기구 : 보일러 등
- 온도/습도/청정도 출입 장치 : 창문 개폐 등
- 조도 가전제품 : 스마트 등
- 조도 기구 : 커튼 및 차양막 제어
- 조도 출입장치 : 창문 유리 투명도 조절

<Figure 2>는 본 연구에서 제시하는 서비스 시나리오의 개요를 보여준다. 스마트홈 환경에서 지능형 서비스는 데이터 수집, 서비스 추천 모형의 수립, 서비스 추천 수행, 모형의 평가 및 강화의 4단계의 사이클을 지속적으로 순환하게 될 것이다. 다음은 시나리오의 각 단계에 대한 설명이다.

### 2.2.1 데이터 수집 단계

편안하고 에너지 효율적인 스마트홈을 구현하기 위해서는 먼저 다음의 4가지 정보를 수집할 수 있어야 한다.

- 센서를 이용한 환경 데이터 수집 : 실내의 온도, 습도, 청정도, 조도 정보뿐 아니라, 외부 환경 데이터의 수집이 필요하다. 이를 위해 필요시 외부 서버에 대한 연결도 필요하다.
- 전력 이용량 데이터 수집 : 에너지 효율적인 서비스 제어를 위해서는 에너지 사용원별로 전력 이용량 데이터가 수집되어야 한다.
- 제어기기 활동 이력 수집 : 사용자가 환경 변화를 위해 수행한 제어 이력을 수집해야 한다. 에어컨이나 공기청정기를 켜거나, 커튼의 개폐 등의 이력을 수집한다.
- 사용자 프레즌스 정보 수집 : 집의 구성원별 프레즌스



<Figure 2> Service Scenario

정보를 파악하여 개인별 온습도 환경 등에 대한 선호를 파악한다.

### 2.2.2 서비스 추천 모형 생성 단계

이전 단계에서 데이터가 수집되었다면 스마트홈 지능형 서비스 플랫폼은 서비스 추천을 위한 모형을 생성해야 한다. 서비스 플랫폼은 일정 기간 동안 집의 내외부 환경 데이터와 사용자 행동 데이터를 수집하여 다음 모형을 수립한다.

- 사용자 선호 예측 모형 : 사람의 개입 없이 쾌적한 실내 환경 상태를 능동적으로 구현하기 위해서는 사람의 습관 및 선호를 학습할 수 있어야 한다. 특히 개별 사용자의 프레즌스를 파악하여 사용자별 선호를 분별해 낼 수 있어야 한다.
- 사용자 의도 예측 모형 : 거주자의 선호도 예측뿐 아니라 거주자가 처한 상황을 능동적으로 인지하여 서비스를 제공하기 위해서는 사람의 행동을 통해 의도를 파악할 수 있어야 한다. 예를 들어, 사용자 행위를 파악하여 독서를 하려고 하면 조도를 밝히거나, 자기 전에 조도를 낮추는 등의 서비스를 한다.
- 에너지 사용 최적화 모형 : 적절한 에너지 절감을 위해서는 사람의 선호뿐 아니라 에너지 사용 패턴을 예측하여 에너지 사용을 최적화하도록 서비스를 운영하여야 한다.

### 2.2.3 서비스 자동 제어 수행 단계

스마트홈 지능형 서비스 플랫폼은 생성된 모형을 이용하여 온습도, 조도 등의 자동 제어 서비스를 수행한다. 수행하는 동안 사용자의 피드백을 수집하여 사용자의 의도에 맞추어 자동 조절을 조정해 나간다.

### 2.2.4 서비스 추천 모형의 평가 및 진화 단계

스마트홈 지능형 서비스 플랫폼은 수집된 사용자의 피드백에 근거하여 현재의 모형을 주기적으로 평가한다. 모형은 항상 어느 정도의 부정확성을 가지게 되고, 사용자의 주위환경에 대한 선호도 바뀔 수 있으므로 서비스 자동 제어 단계 동안 수집된 사용자의 피드백에 기반하여 현재의 모형이 적합한지를 평가하고 현재의 추천 모형의 개선이 필요하면 이를 수행하게 된다.

## 3. 스마트홈 지능형 서비스 플랫폼을 위한 데이터 마이닝 기법 요구사항 도출

스마트홈 환경에서 장착될 지능형 서비스 플랫폼에서 적용될 데이터 마이닝 기법을 선정하기 위해 데이터 마

이닝 기법에게 요구되는 기능 요구사항과 기술 요구사항을 도출한다. 도출된 요구사항에 대한 요약은 <Table 1>에 제시되어 있다.

### 3.1 기능 요구사항

앞서 개발된 서비스 시나리오의 각 단계를 분석하여 데이터 마이닝 기법에 요구되는 기능 요구사항을 도출한다. 각 요구사항을 FR(functional requirements) 접두어로 일련 번호를 붙인다.

#### 3.1.1 데이터 수집 단계의 기능 요구사항

- FR1 수치 및 범주형 필드 처리  
데이터 수집 단계에서는 센서를 이용한 다양한 환경 정보 수집이 필요하다. 따라서 다양한 정보 근원으로 부터 들어오는 데이터를 처리하기 위해서는 정량 예측 변수뿐 아니라 정성 예측변수도 다룰 수 있는 데이터 마이닝 기법이어야 한다.
- FR2 변수 선택  
매우 많은 종류의 데이터가 수집될 수 있으므로 이 중 관련된 변수만 추출할 수 있는 데이터 마이닝 기법이어야 한다. 그렇지 않으면 관련성이 적으나 개수가 많은 변수들에 의해 데이터 분석의 정확성이 훼손될 수 있다.
- FR3 결측치 처리  
센서 데이터는 센서의 오작동 등으로 결측치가 발생할 수 있으므로 결측치 데이터를 적절히 다룰 수 있는 데이터 마이닝 기법이어야 한다.

#### 3.1.2 서비스 추천 모형 생성 단계의 기능 요구사항

- FR4 예측의 정확도  
데이터 마이닝 기법을 통해서 산출된 모형은 크게 두 가지 관점에서 평가된다. 하나의 모형의 예측의 정확도가 얼마나 높은가이고, 다른 하나는 모형이 얼마나 해석 가능한 가이다. 본 연구의 대상인 서비스 추천 모형은 모형이 사용자에게 자동 조절 서비스에 대한 새로운 직관을 부여하는 것 보다는 정확한 서비스 추천을 제공하는 것이 가장 중요하다.
- FR5 가중치 적용 학습  
서비스 추천 모형 수립 단계에서 고려해야 할 또 다른 사항은 사람이 환경을 조절하는 활동은 발생 빈도가 낮은 사건이라는 것이다. 매 10분간 사용자의 조절 노력의 여부를 조사한다고 할 때 사용자가 조절을 수행한 구간의 개수는 그렇지 않은 구간에 비해 현저히 적은 개수일 것이다. 일반적으로 예측 모형은 예측하고자 하는 사례가 비슷한 비중으로 존재할 때 모형의 정

확도가 향상된다. 따라서 서비스 추천에 사용되는 데이터 마이닝 기법은 발생 빈도가 낮은 사례에 대해 적절한 가중치를 부여하는 학습이 가능해야 할 것이다.

3.1.3 서비스 자동 제어 수행 단계의 기능 요구사항

- FR6 추정 및 분류가 가능한 학습  
데이터 마이닝 기법이 도출한 모형의 출력은 서비스 조절 명령이 된다. 서비스 조절 명령은 범주형 명령일 수도 있고, 수치적 명령일 수 있다. 예를 들어 온도 조절 서비스는 두 가지 방식으로 추천될 수 있다. 첫 번째 방식은 온도를 올리지, 유지할지, 내릴지에 대한 정성적 판단을 하여 온도 조절기기를 제어하는 것이다. 두 번째 방식은 온도를 몇 도록 할지 정량적인 판단을 내려서 해당 온도로 온도 조절기기를 제어하는 것이다. 이 두 가지 방식에 모두 적용이 가능하려면 목표 변수의 형태가 정성인 분류 모형과, 목표 변수의 형태가 정량인 수치형 예측 모형을 둘 다 생성할 수 있는 데이터 마이닝 기법이 좋다.

3.1.4 서비스 추천 모형의 평가 및 진화 단계의 기능 요구사항

- FR7 모형의 부분적 갱신  
실내 환경에 대한 자동 조절 서비스는 주사용자, 계절, 건강 상태의 변화 등으로 사용자의 선호가 달라질 수 있기 때문에, 현재의 모형의 정확성을 주기적으로 평가하여 정확성이 낮으면 새로운 데이터를 점진적으로 모형에 반영할 수 있는 메커니즘을 가져야 한다. 즉, 지속적으로 발생하는 데이터를 증분적으로 반영하여 모형의 갱신이 가능한 학습방법이어야 한다.
- FR8 모형의 검증 및 평가 용이성  
주기적으로 모형을 개선하기 위해서는 모형의 검증 및 평가가 용이한 데이터 마이닝 기법이어야 한다. 모형을 최적화하기 위한 매개변수의 조절이 단순하지 않으면 모형의 검증 및 평가를 자동화하기 어렵다.

3.2 기술적 요구사항

스마트홈 지능형 서비스 플랫폼이 구현될 기술적 사항을 고려하여 기술 요구사항을 도출한다. 각 기술 요구사항을 TR(technical requirements) 접두어로 일련 번호를 붙인다.

- TR1 모형 학습 시간  
사용자 선호의 급격한 변화의 가능성은 크지 않으므로 스마트홈 환경에서의 모형의 생성 및 평가/재학습의 처리 시간은 일 단위여도 충분할 것이다. 그러면 사용자의 취침 시간 동안 학습이 완료되어 낮 시간 동안의 서비스에 영향을 미치지 않을 것이다.

- TR2 제한된 계산 복잡도  
스마트홈 지능형 서비스 플랫폼은 경량의 스마트홈 게이트웨이 등에 장착될 가능성이 높다. 따라서 데이터 마이닝 기법은 낮은 컴퓨팅 능력의 기기가 감당할 수 없는 매우 높은 계산량을 요구하지 않아야 한다.
- TR3 훈련 집합의 저장 공간  
다양한 센서를 통해 수집된 데이터는 용량이 매우 클 것이다. TR2와 같은 이유로 학습된 데이터는 삭제되거나 요약되어 보관될 수 있어야 한다. 모형이 많은 데이터의 보관을 요구하지 않아야 한다.
- TR4 서비스 추천 처리 시간  
상황의 변화에 따라 적절히 서비스 추천이 가능하려면 초 단위의 처리 시간 안에 모형이 입력을 출력으로 변환할 수 있어야 할 것이다. 따라서 모형이 생성된 후 추정 또는 분류를 위해 과도한 계산량이나 처리 시간을 요구하지 않아야 할 것이다.
- TR5 모형의 저장 공간  
TR3과 같은 이유로 데이터나 모형의 저장을 위한 과도한 저장 공간을 요구해서는 안 된다.

<Table 1> Functional and Technical Equirements

Type	ID	Requirement
Functional Requirement	FR1	Numeric and categorical predictors
	FR2	Variable selection
	FR3	Missing values
	FR4	Prediction accuracy
	FR5	Weighted learning
	FR6	Classification and numeric estimation
	FR7	Incremental learning
	FR8	Ease of model evaluation
Technical Requirement	TR1	Training time
	TR2	Limited computing complexity
	TR3	Limited data storage
	TR4	Predicting time
	TR5	Limited model storage

4. AHP 기법을 이용한 데이터 마이닝 기법에 대한 적합성 평가

4.1 평가 대상 데이터 마이닝 기법 선정

스마트홈 지능형 서비스 추천 플랫폼에 적합한 데이터 마이닝 기법을 찾기 위해서는 먼저 적합도 평가 대상이 될 데이터 마이닝 기법 후보군의 선정이 필요하다. 적합도 평가 후보군이 될 데이터 마이닝 기법 선정에는 다음

기준을 사용하였다.

첫째, 서비스 조절을 위한 추천에는 목표 변수의 예측이 필요하므로 지도 학습(supervised learning) 데이터 마이닝 기법에서 대상을 선정하였다.

둘째, 후보군은 의사결정나무, 배깅 등 데이터 마이닝 기법 수준에서 대상을 선정하였다. 데이터 마이닝 기법이 구현된 세부 알고리즘 수준으로 선정하면 대상이 너무 많아질 뿐 아니라, 각 알고리즘의 세부 단계는 실제 분석 과정에서 필요에 따라 개선될 수 있으므로 세부적인 알고리즘 차이를 평가하는 것은 실효성이 적기 때문이다.

셋째, 다만 C4.5나 CART처럼 알고리즘 구현 사례가 데이터 마이닝 기법 수준으로 매우 널리 알려져 있고, 서로 중요 부분의 차이가 어느 정도 큰 경우에는 하나의 데이터 마이닝 기법으로 분류하였다.

후보군의 선정은 통계적 학습 분야의 저명한 저서로 뽑히고 있는 Hastie et al.[5], 최근의 데이터 마이닝 기법을 훌륭히 반영하고 있는 저서로 뽑히는 Witten et al.[16], 그리고 IoT 분야에 적용되는 데이터 마이닝 기법에 대한 조사분석 논문인 Tsai et al.[15]를 기초 문헌으로 참조하였다.

제시된 세 개의 문헌을 기준으로 <Table 2>와 같은 총 18개의 데이터 분류기법을 추출하였다.

## 4.2 적합도 평가 방법론

### 4.2.1 AHP 기법

AHP 기법은 다기준 의사결정 기법으로 1977년 Satty [11]에 의해 처음 도입되었다. AHP는 좋은 수학적 성질과 의사결정에 필요한 입력 데이터를 산출하기가 비교적 쉽기 때문에 많은 연구자에 의해 널리 받아들여졌다.

AHP는 본질적으로 의사결정의 계층적 구조를 구성한 후, 각 계층의 요소를 쌍대 비교하고, 이를 기반으로 상대적 중요도를 결정하여 최종 의사결정을 수행하는 방법이다. AHP 방법의 개략적 단계는 다음과 같다.

1. 해결해야 할 의사결정 문제를 정의한다.  
복잡한 의사결정 문제를 해결하기 위하여 목적(objectives), 기준(criteria), 세부기준(subcriteria), 대안(alternatives) 등으로 구성된 다계층 구조를 이용한다.
2. 기준의 중요도를 평가하기 위해 전문가에 의한 기준 간의 쌍대 비교(pairwise comparisons)를 수행하고 이를 이용하여 각 의사결정 기준의 상대적 중요도를 측정한다.
3. 각 기준별로 대안을 전문가가 쌍대 비교하고, 이를 이용하여 기준별로 대안의 상대적인 적합도를 측정한다. 쌍대 비교 결과를 이용하여 각 기준 또는 대안의 중요도에 대한 상대적 값을 구한다. 상대적 중요도는 고유

<Table 2> Data Mining Technique Candidates for Evaluation

Type	Techniques	Implementation
rules	IR	OneR, HuperPipes
	covering rules	Prism
probability	naive Bayes	NaiveBayesMulltinomial
	Bayesian networks	BayesNet, AODE, WAODE
trees	C4.5	Id3, C4.5, J48
	CART	SimpleCART, REPTree
	bagging	Bagging, MetaCost
	random forest	RandomSubSpace, RandomForest, RotationForest
	boosting	AdaBoostM1
(non)linear	logistic regression	Various implementations
	LDA	Various implementations
	neural networks	MultilayerPerceptron
	RBF networks	RBFNetwork
	SVM	SMO, LibSVM, SPegasos
	additive regression	AdditiveRegression, LogitBoost
	model trees	M5P
instances	k-nearest neighbours	Ibk, Kstar, NNge
	locaaly weighted linear regression	LWL

치의 계산을 통해 계산된다.

4. 쌍대 비교의 일관성을 검증하기 위해 일관성 지수(CI; consistency index)와 일관성 비율(CR : consistency ratio)를 계산한다. 쌍대 비교의 일관성이 떨어지면 일관성이 떨어지는 쌍대 비교에 대해 전문가에게 재검토를 요청한다.
5. 각 계층에서 계산된 기준의 중요도, 기준별 대안의 적합도를 이용하여 대안별 최종 점수를 계산하고 이를 이용하여 최적 대안을 선정한다.

이렇듯 AHP 기법은 대안이 다수이고, 대안을 선택하는 데 고려해야 하는 기준도 다수인 경우, 전문가의 판단을 체계적으로 반영하여 신속하게 최적 대안을 선택할 수 있다는 장점이 있다.

본 연구의 대상인 스마트홈 지능형 서비스 추천 플랫폼에 적합한 데이터 마이닝 기법은 <Table 2>에서 확인할 수 있듯이 매우 많다. 이러한 모든 기법을 스마트홈 지능형 서비스의 대표적 시나리오 하에서 직접 구현하여 테스트해 본다면 가장 적합한 기법을 선택할 수 있을 것이다. 그러나 이러한 방법은 시나리오 개발과 시나리오별 모형 구축이 필요하므로 현실적으로 많은 비용이 발생할 수 밖에 없다. 따라서 본 연구에서는 AHP 기법을 이용하여 전문가의 의견을 기반으로 적합한 데이터 마이닝 기법을 선별하는 방법을 사용하였다.



4.2.2 본 연구의 AHP 기법 적용 체계

AHP 기법은 쌍대 비교의 대상이 너무 많으면 전문가들이 정확한 평가를 하기 어렵다. 일반적으로 쌍대 비교의 대상이 7, 8개 이하인 경우에 좋은 결과를 준다고 알려져 있다[12]. 그런데 현재 도출된 요구사항은 기능과 기술 영역을 모두 포함하여 모두 13개이고, 데이터 마이닝 기법 대안의 수도 18개이다. 따라서 이러한 점을 고려하여 본 연구에서는 <Figure 3>과 같은 AHP 평가 계층을 단계적으로 구성하여 적합한 데이터 마이닝 기법을 선정한다.

먼저 데이터 적합도 평가의 기준이 되는 요구사항을 기능 요구사항과 기술 요구사항으로 묶어서 먼저 기능 요구사항 그룹과 기술 요구사항 그룹의 중요도를 먼저 쌍대 비교하여 두 그룹의 상대적 중요도를 먼저 도출한다. 각 그룹의 상대적 중요도의 합은 1이 되도록 한다. 이를 수식으로 표현하면 다음과 같다.

$$c_{FR} + c_{TR} = 1 \tag{1}$$

단,  $c_{FR}$  과  $c_{TR}$  은 기능 요구사항 그룹과 기술 요구사항 그룹의 상대적 중요도.

기능 요구사항 그룹의 각 요구사항에 대하여 쌍대 비교를 통해 기능 요구사항 안에서의 상대적 중요도를 도출한다. 마찬가지로 기술 요구사항 그룹의 각 요구사항에 대하여 쌍대 비교를 하여 기술 요구사항 안에서의 상

대적 중요도를 도출한다. 각 그룹 내에서의 요구사항의 중요도의 합은 1이 되도록 한다. 이를 수식으로 표현하면 다음과 같다.

$$\sum_{k \in FRs} d_k = 1, \sum_{k \in TRs} d_k = 1 \tag{2}$$

단,  $d_i$  는 해당 그룹 내에서의 요구사항  $i$  의 상대적 중요도.

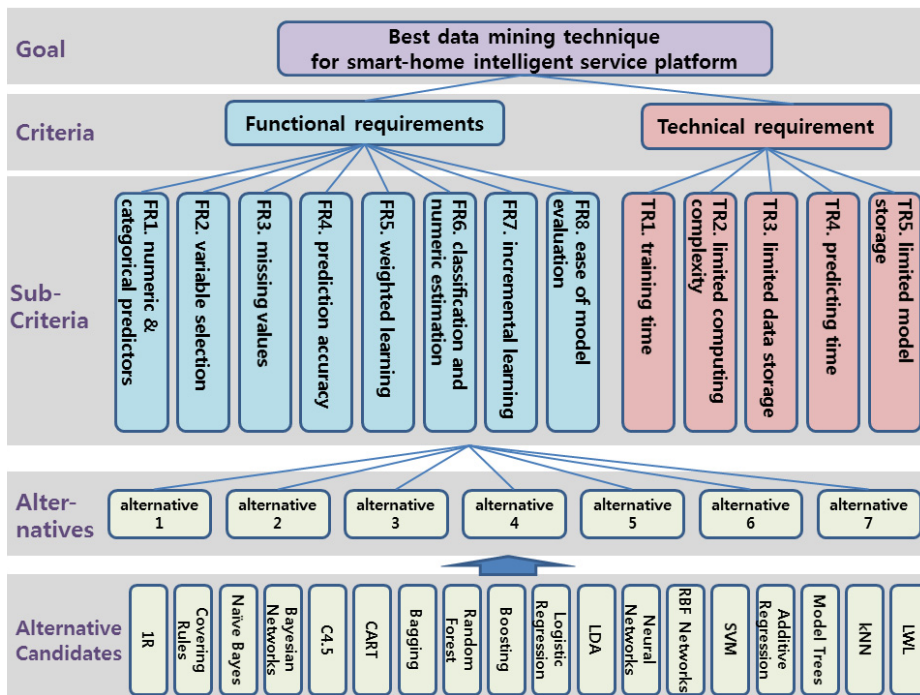
요구사항 그룹에 대한 상대적 중요도와 각 그룹 내에서의 상대적 중요도를 곱하여 최종적으로 모든 기능적이거나 기술 요구사항의 상대적 중요도  $c_k$  를 구한다.

$$c_k = \begin{cases} c_{FR} \cdot d_k, & \text{if } k \in FRs \\ c_{TR} \cdot d_k, & \text{if } k \in TRs \end{cases} \tag{3}$$

이 경우  $\sum_k c_k = 1$  이 된다.

그 다음 후보로 선택된 데이터 마이닝 기법을 AHP 평가의 대안으로 하여, AHP 기준인 요구사항 별로 쌍대 비교를 수행한다. 그리고 쌍대 비교 행렬의 고유치 벡터를 구하여 데이터 마이닝 기법의 상대적 적합도  $w_{ik}$  를 구한다. 각 요구사항별 상대적 적합도의 합은 1이 되도록 한다.

$$\sum_i w_{ik} = 1, \text{ for all } k. \tag{4}$$



<Figure 3> AHP Framework for Evaluating the Techniques

최종적으로 후보 데이터 마이닝 기법의 적합도  $a_i$ 는 요구 사항별 상대적 중요도를 가중치로 다음과 같이 산출한다.

$$a_i = \sum_k w_{ik} \cdot c_k. \quad (5)$$

### 4.3 적합도 평가 결과

데이터 마이닝 기법에 대한 적합도 평가는 연구팀 전문가 2인, 외부 데이터 마이닝 기법 전문가 3인으로 총 5명의 전문가에 의해 이루어졌다. 외부 데이터 분석전문가는 현재 대학교에 관련 분야 교수로 재직하는 사람들이다. 요구사항의 중요도에 대한 평가는 5명의 전문가의 토론을 통해 하나의 평가가 이루어졌고, 데이터 마이닝 기법에 대한 평가는 개별적으로 수행된 후 기하평균을 이용하여 쌍대 비교 행렬을 계산하였다.

#### 4.3.1 요구사항 그룹 간의 상대적 중요도 평가

요구사항 그룹은 기능 요구사항 그룹(FR)과 기술 요구사항 그룹(TR) 두 가지만 존재하므로 둘 사이의 쌍대 비교만 수행하면 된다. 최종 결과로 <Table 3>과 같은 결과를 얻었다.

<Table 3> Relative Impotance of the Criteria

Item	FR	TR
Importance	0.6	0.4

#### 4.3.2 요구사항 그룹 내의 상대적 중요도 평가

기능 요구사항의 쌍대 비교 결과를 통해 얻어진 중요도 결과는 <Table 4>와 같다. 기능 요구사항의 쌍대 비교 행렬의 일관성 비율(CR; consistency ratio)은 6.5%로 10% 미만의 값이 나왔으므로 쌍대 비교에 일관성이 있었다고 판단된다.

<Table 4> Relative Impotance of the FRs

Item	FR1	FR2	FR3	FR4	FR5	FR6	FR7	FR8
Importance	0.06	0.04	0.06	0.36	0.13	0.23	0.07	0.06

기술 요구사항에 대한 쌍대 비교 결과를 통해 얻어진 중요도 결과는 <Table 5>와 같다. 기술 요구사항의 쌍대 비교 행렬의 일관성 비율(CR)은 3.7%로 10% 미만의 값이 나왔으므로 쌍대 비교에 일관성이 있었다고 판단된다.

<Table 5> Relative Impotance of the TRs

Item	TR1	TR2	TR3	TR4	TR5
Importance	0.07	0.17	0.21	0.44	0.12

#### 4.3.3 요구사항 전체 항목의 상대적 중요도 평가

<Table 3>~<Table 5>의 결과는 각각 식 (1)과 식 (2)에 해당하는 결과이고, 이 결과에 식 (3)을 적용하면 이용하면 모든 요구사항의 상대적 중요도를 <Table 6> 같이 계산할 수 있다.

결과에서도 알 수 있듯이 기능 요구사항인 FR4 예측의 정확도에 대한 요구사항이 가장 큰 중요도를 가졌고, 그 다음으로는 기술 요구사항인 TR4 예측 시간 등이 중요한 요구사항으로 평가되었다.

<Table 6> Relative Impotance of the Requirements

Item	Requirement	Relative importance
FR4	prediction accuracy	0.22
TR4	predicting time	0.18
FR6	classification and numeric estimation	0.14
TR3	limited data storage	0.08
FR5	weighted learning	0.08
TR2	limited computing complexity	0.07
TR5	limited model storage	0.05
FR7	incremental learning	0.04
FR8	ease of model evaluation	0.04
FR1	numeric and categorical predictors	0.03
FR3	missing values	0.03
TR1	training time	0.03
FR2	vraiable selection	0.03

#### 4.3.4 데이터 마이닝 기법 사전 적합도 평가

AHP 기법을 이용하여 최적의 데이터 마이닝 기법을 평가하기에는 <Table 2>에 나열된 기법은 너무 많다. 따라서 다음의 점을 고려하여 데이터 마이닝 기법에 대한 토론을 통해 사전 정성 평가를 실시하였다.

첫째, 기술 요구사항의 대부분은 데이터 마이닝 기법이 너무 많은 저장, 컴퓨팅 용량을 요구해서는 안 되며, 예측 시간이 짧아야 한다는 것이다. k-근접 접근 방법이나 지역적으로 가중치된 선형 회귀 모형은 훈련 데이터의 모든 사례를 보관해야 한다. 그리고 모형의 예측 시점에도 가장 가까운 이웃 데이터를 찾아야 하므로 많은 계산을 요구한다. 따라서 제한된 컴퓨팅 환경을 가지는 서비스 추천 플랫폼에 적절한 데이터 마이닝 기법으로 판단되지 않으므로 두 분석기법을 후보에서 제외하였다.

둘째, 요구사항의 중요도 평가에서도 밝혀졌듯이 모형의 해석가능성보다 모형의 정확도가 매우 중요하다. 따라서 모형의 해석 가능성은 좋지만 일반적으로 정확도가 떨어진다고 평가되는 1R, covering rules, 나이브 베이즈, 단순 의사결정나무(C4.5, CART), 로지스틱 회귀분석, LDA,



모형 나무 기법을 후보에서 제외하였다.

셋째, RBF 네트워크는 사실 인공 신경망에서 활성화 함수(activation function)가 radial basis인 경우와 같으므로 인공 신경망 기법의 하나로 간주하여 평가에서 제외하였다.

위와 같은 이유로 최종적으로 <Table 7>의 7개의 후보 데이터 마이닝 기법이 선정되었다.

<Table 7> Data Mining Technique Candidates for the Final Evaluation Stage

Type	Techniques	ID
probability	Bayesian networks	BN
trees	bagging	bag
	random forest	RF
	boosting	boost
(non)linear	neural networks	NN
	SVM	SVM
	additive regression	AR

4.3.5 요구사항별 적합도 평가

요구사항별로 데이터 마이닝 기법의 적합도를 쌍대 비교한 후, 쌍대 비교 행렬의 고유 벡터를 구하여 데이터 마이닝 기법의 상대적 적합성을 구했다. <Table 8>은 요구사항별로 데이터 마이닝 기법의 상대적 적합도를 구한 결과이다.

요구사항별로 쌍대 비교 행렬의 일치성 비율 CR은 <Table 8>의 CR 열에서 확인할 수 있다. 일부 10%를 조금 상회하는 쌍대 비교도 있었지만 대체적은 10% 미만의 CR을 보이고 있어서 쌍대 비교 결과의 신뢰성이 있는 것으로 판단된다.

<Table 8> Adequacy of the Techniques for Each Requirements

Requirement	Adequacy of Techniques							CR
	BN	bag	RF	boost	NN	SVM	AR	
FR1	0.18	0.20	0.20	0.18	0.08	0.08	0.07	0.02
FR2	0.18	0.15	0.15	0.14	0.06	0.05	0.27	0.04
FR3	0.12	0.26	0.23	0.24	0.03	0.04	0.08	0.05
FR4	0.04	0.11	0.17	0.26	0.13	0.15	0.16	0.02
FR5	0.20	0.18	0.21	0.20	0.06	0.06	0.09	0.01
FR6	0.04	0.09	0.25	0.24	0.19	0.12	0.07	0.16
FR7	0.32	0.05	0.06	0.05	0.25	0.16	0.11	0.03
FR8	0.23	0.16	0.14	0.11	0.13	0.11	0.12	0.01
TR1	0.29	0.11	0.10	0.10	0.09	0.17	0.14	0.01
TR2	0.29	0.09	0.09	0.09	0.09	0.17	0.17	0.01
TR3	0.11	0.12	0.19	0.09	0.10	0.17	0.10	0.00
TR4	0.13	0.07	0.13	0.13	0.16	0.26	0.09	0.02
TR5	0.13	0.07	0.07	0.07	0.27	0.16	0.13	0.02

4.3.6 데이터 마이닝 기법 최종 적합도 점수

<Table 6>는 요구사항의 상대적 중요도를 나타내며 식 (3)의 값이 된다. <Table 8>은 요구사항별 데이터 마이닝 기법의 적합도를 나타내며 식 (4)의 값이 된다. 식 (5)를 이용하면 데이터 마이닝 기법의 최종 적합도를 <Table 9> 같이 구할 수 있다.

<Table 9>는 데이터 마이닝 기법을 최종 적합도 점수가 높은 것에서 낮은 것 순으로 보여준다. 가장 적합도가 높은 3개의 데이터 마이닝 기법으로는 부스팅, 랜덤 포레스트, SVM이 선택되었다. 부스팅은 FR4의 예측 정확도에서 높은 적합도 점수를 얻었고, 랜덤 포레스트는 기능 요구사항 전반에서 높은 점수를 얻었다. 반면 SVM은 기술 요구사항 전반에서 높은 점수를 얻었다.

<Table 9> Suitability of the Data Mining Techniques for the Smart-Home Service Platform

Techniques	Adequacy
boosting	0.18
random forest	0.17
SVM	0.15
neural networks	0.14
Bayesian networks	0.14
additive regression	0.12
bagging	0.11

5. 결론

본 연구는 스마트홈 지능형 서비스 플랫폼에 적용될 데이터 마이닝 기법의 적합도를 평가하는 연구로서 다음 사항을 수행하였다.

첫째, 스마트홈 환경에서의 지능형 서비스에 대한 시나리오를 개발하였다.

둘째, 서비스 시나리오를 바탕으로 데이터 마이닝 기법에 요구되는 8개 기능 요구사항을 발굴하고, 제한된 컴퓨팅 용량 하의 플랫폼 구동 환경에서 데이터 마이닝 기법이 갖추어야 할 5개 기술 요구사항을 도출하였다.

셋째, 총 18개 데이터 마이닝 기법을 사전 평가를 통해 7개의 유력한 데이터 마이닝 기법 후보로 추려내었다. 그 후 도출된 요구사항에 적합한 정도를 AHP 기법을 이용하여 적합도 평가를 수행하였다. 이를 통해 최종 3개의 데이터 마이닝 기법을 선정하였다.

기존에 특정 스마트홈 서비스를 위해 데이터 기법을 적용한 연구는 있었으나, 스마트홈 서비스 플랫폼에 장착될 범용형 데이터 마이닝 기법의 적합도에 대한 연구는 없었다. 본 연구는 스마트홈 지능형 서비스 플랫폼을

구축할 때 가이드라인으로 사용될 수 있으므로, 기존의 세부적인 데이터 마이닝 기법 적용 연구와 실제 플랫폼 구현 연구 사이의 상호간극을 줄이는데 기여하였다는 데 의의가 있다.

마지막으로 본 연구의 한계와 발전 가능성을 제시하며 마무리 하고자 한다. 본 연구에서는 전문가의 평가를 통해 스마트홈 지능형 서비스 플랫폼에 적합한 데이터 마이닝 기법을 선정하였지만, 이를 실제적으로 플랫폼에 구현하여 각 기법의 실제 성능에 대한 분석은 수행하지 않았다. 추후 프로토타이핑을 통한 각 기법의 성능에 대한 다양한 분석이 이루어진다면 더 완전한 평가 결과도 도출되리라 기대된다.

## Acknowledgement

This work was supported by ETRI R&D Program funded by the Government of Korea(17Z11510, Development of Technologies for Proximity, Real-time, and Smart Service Recommendation Platform).

## References

- [1] Cook, D.J., Crandall, A.S., Thomas, B.L., and Krishnan, N.C., CASAS : A smart home in a box, *Computer*, 2013, Vol. 46, No. 7, pp. 62-69.
- [2] Cook, D.J., MavHome : An agent-based smart home, *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications*, 2003, pp. 521-524.
- [3] Deleawe, S., Kuszniir, J., Lamb, B., and Cook, D.J., Predicting air quality in smart environments, *Journal of Ambient Intelligence and Smart Environments*, 2010, Vol. 2, No. 2, pp. 145-154.
- [4] Han, D.-M. and Lim, J.-H., Design and implementation of smart home energy management systems based on zigbee, *IEEE Transactions on Consumer Electronics*, 2010, Vol. 56, No. 3, pp. 1417-1425.
- [5] Hastie, T., Tibshirani, R. and Jerome, F., The elements of statistical learning, 2nd ed, Berlin : Springer, 2009.
- [6] Helal, S., Mann, W., El-Zabadani, H., King, J., Kaddoura, Y., and Jansen, E., The gator tech smart house : A programmable pervasive space, *Computer*, 2005, Vol. 38, No. 3, pp. 50-60.
- [7] Jakkula, V.R. and Cook, D.J., Detecting anomalous sensor events in smart home data for enhancing the living experience, *Proceedings of the 7<sup>th</sup> AAI Conference on Artificial Intelligence and Smarter Living*, 2011, pp. 33-37.
- [8] Jung, H. and Kim, J.-W., A Machine Learning Approach for Mechanical Motor Fault Diagnosis, *Journal of the Society of Korea Industrial and Systems Engineering*, 2017, Vol. 40, No. 1, pp. 57-64.
- [9] Kim, Y.K., Six elements of Smart-home ecosystem, Digieco, 2014. 11. 26.
- [10] Park, S.H., Kim, J.S., Park, C.S., Kim, S.S., and Baek, J.G., A Fault Detection of Cyclic Signals Using Support Vector Machine-Regression, *Journal of Korean Society for Quality Management*, 2011, Vol. 39, No. 2, pp. 234-243.
- [11] Saaty, T.L., A scaling method for priorities in hierarchical structures, *Journal of Mathematical Psychology*, 1977, Vol. 15, No. 3, pp. 234-281.
- [12] Saaty, T.L., The Analytic Hierarchy Process : Planning, Priority Setting, Resource Allocation, Advanced book program, McGraw-Hill, 1980.
- [13] Spencer, B. and Al-Obeidat, F., Temperature Forecasts with Stable Accuracy in a Smart Home, *Procedia Computer Science*, 2016, Vol. 83, pp. 726-733.
- [14] Suryadevara, N.K., Mukhopadhyay, S.C., Wang, R., and Rayudu, R., Forecasting the behavior of an elderly using wireless sensors data in a smart home, *Engineering Applications of Artificial Intelligence*, 2013, Vol. 26, No. 10, pp. 2641-2652.
- [15] Tsai, C.-W., Lai, C.-F., Chiang, M.-C., and Yang, L., Data mining for internet of things : A survey, *Communications Surveys Tutorials, IEEE*, 2014, Vol. 16, No. 1, pp. 77-97.
- [16] Witten, I.H., Frank, E., and Hall, M.A., Data Mining : Practical machine learning tools and techniques, 3rd ed, Morgan Kaufmann, 2011.
- [17] Zhou, B., Li, W., Chan, K.W., Cao, Y., Kuang, Y., Liu, X., and Wang, X., Smart home energy management systems : Concept, configurations, and scheduling strategies, *Renewable and Sustainable Energy Reviews*, 2016, Vol. 61, pp. 30-40.

## ORCID

Kilhwon Kim | <http://orcid.org/0000-0002-0577-7906>

Changsup Keum | <http://orcid.org/0000-0002-3412-4007>

Ki-Sook Chung | <http://orcid.org/0000-0001-9780-7094>