

스마트 그리드 환경에서 가정의 소비전력 생성 시뮬레이터 개발 및 데이터 마이닝 기법을 이용한 가족 유형 분류

김지현*, 이윤진*, 김호원^o

Development of the Power Consumption Simulator and Classification of the Types of Household by Using Data Mining Over Smart Grid

Ji-hyun Kim*, Yun-jin Lee*, Ho-won Kim^o

요약

최근 불규칙적인 전력 수요에 따른 수차례의 전력난에 의해, 양방향 통신을 이용하여 전력의 공급과 수요 정보를 실시간으로 교환함으로써 변동하는 전력 수요에 맞추어 효율적으로 전력을 공급할 수 있게 해주는 스마트 그리드의 도입의 필요성이 급속하게 높아지고 있다. 향후 국내에 스마트 그리드 환경이 구축되면, 스마트 그리드 환경에서 얻을 수 있는 정보를 이용하여 서비스를 제공하는 third-party들 역시 활성화될 것이다. 본 논문에서는 스마트 그리드 환경이 구축된 가정에서 발생하는 가전기기 소비전력 패턴을 데이터 마이닝 기법을 이용하여 분석하고, 가족 유형을 분류하는 방법을 제안한다. 가족 유형을 분류하는 분류기를 만들기 위해서는 소비전력이 어떤 가족 유형에서 생성되었는지에 대한 데이터가 필요한데, 현실적으로 그러한 데이터를 수집하기는 힘들다. 따라서 본 연구에서는 가상의 데이터를 생성하기 위해 가족 유형에 따른 소비전력을 생성하는 시뮬레이터를 개발하고, 시뮬레이터에서 생성되는 소비전력 정보를 이용하여 가족 유형을 분류한다. 또한 본 논문에서 제안하는 기법이 향후 각 가구의 특성에 따른 맞춤형 서비스 또는 마케팅에 활용될 수 있는 가능성을 제시한다.

Key Words : smart grid, power simulator, family classifier, data mining, power consumption

ABSTRACT

Recently, because of irregular power demand, we have suffered from an electric power shortage. The necessity of the adoption of smart grid which makes effective supply of power by using the two-way communication across the grid between the customers and electric energy providers is growing more and more. If smart grid set up in our country, the third-parties which provide services to customer using the information acquired from smart grid, might be revved up. In this paper, we suggest a methodology how classify the types of family by analysing an power consumption pattern using data mining technique. To make a classifier for categorizing the household types, we need power consumption data and their family type. However, it is hard to get both of them. Therefore we develop the simulator that generates power consumption patterns of the household and classify the types of family. Also, we present a potential for application services such as customized services for a specific family or goods marketing.

※ 본 연구는 지식경제부 산업융합원천기술개발사업으로 지원된 연구결과입니다(No.10043907)

• First Author : 부산대학교 전자전기컴퓨터공학과 정보보호 및 IoT 연구실, jihyunkim@pusan.ac.kr, 학생회원

◦ Corresponding Author : 부산대학교 정보컴퓨터공학과 정보보호 및 IoT 연구실, hownkim@pusan.ac.kr, 종신회원

* 부산대학교 전자전기컴퓨터공학과 정보보호 및 IoT 연구실

논문번호 : KICS2013-11-496, 접수일자 : 2013년 11월 15일, 심사일자 : 2013년 12월 23일, 최종논문접수일자 : 2014년 1월 8일

I. 서 론

최근 신재생 에너지 개발에 대해 높아진 관심과 올해 발생한 수차례의 전력난으로 인해 스마트 그리드는 현실적인 해결책으로 더욱 주목받고 있다. 스마트 그리드는 전력의 수요측와 공급측이 서로 정보를 주고받을 수 있는 양방향성을 특징으로 하는 IT가 접목된 전력망 구조로, 공급 측의 적절한 수요예측과 수요 측의 효율적인 전력관리를 가능하게 한다^{1,2)}. 이에 국내외에서 스마트 그리드와 관련된 기술 개발 및 연구가 활발히 진행되고 있으며, 제주 실증단지를 통한 검증과 도입이 이루어지고 있다³⁾.

NIST 표준에 따르면 스마트 그리드 환경은 크게 마켓, 오퍼레이션, 서비스공급자, 전력생산자, 전송, 분배, 사용자의 일곱 개 도메인으로 분류된다⁴⁾. 본 연구에서는 이들 도메인 중 사용자 도메인에 있는 다양한 형태의 가정에서 발생하는 방대한 양의 전력 정보를 중점적으로 다룬다. 그러나 일부 몇 가지 형태의 가정에서 일정 주기에 따른 소비전력 정보를 얻는 것은 모니터링을 통해서 가능하나 다양한 형태의 가족들의 실측 데이터를 모두 얻는 것은 현실적으로 어려움이 있다. 스마트 미터에서 발생하는 소비전력을 이용하는 몇몇 연구들에서는^{5,6)}, 특정 지역의 각 가구에서 발생하는 소비전력 정보를 이용하지만 구체적인 가족의 유형은 알 수가 없다. 따라서 본 연구에서는 발생하는 다양한 가족 유형에 따른 소비 전력 정보를 생성하는 시뮬레이터를 개발하였다. 시뮬레이터의 개발은 두 가지 측면에서 이득이 있다. 첫 번째는 여러 가지 유형의 가족 형태에 맞는 소비전력 생성이 가능해지고, 두 번째는 데이터마이닝 기법을 이용하여 가족 유형을 분류하기 위해서는 대량의 데이터가 필요한데 시뮬레이터를 이용하면 원하는 기간만큼의 데이터를 얻을 수 있다는 장점이 있다.

본 논문에서는 시뮬레이터에서 생성되는 소비전력 정보를 기반으로 데이터마이닝 기법을 이용하여 가족 유형을 분류하는 기법을 제시한다. 또한 다양한 분야에서 본 연구의 활용가능성을 제시한다.

논문의 구성은 다음과 같다. 2장에서는 스마트 그리드 환경에서 소비 전력 분석과 관련된 연구들을 소개하면서 본 연구의 필요성을 논한다. 3장에서는 소비 전력데이터 생성을 위한 시뮬레이터와 이를 활용한 시뮬레이션 방법에 대해 설명하고, 4장에서는 실험을 통하여 가족 유형 분류 결과를 알아본다. 마지막으로 5장에서는 본 연구의 의미와 활용가능성에 대한 제시 함께 결론을 내림으로써 글을 맺는다.

II. 관련 연구

가정에서 발생하는 소비전력을 분석하여 여러 가지 의미 있는 데이터를 추출할 수 있다. 본 장에서는 스마트 그리드 환경에서 가정의 소비 전력 분석과 관련된 몇 가지 연구들을 소개한다. Adrian Albert는 갑작스러운 전력 수요 증가에 따른 정전을 막기 위해서 사용자의 하루 생활패턴이 전력 소비에 미치는 영향에 대하여 연구하였다⁷⁾. Adrian의 연구는 미국의 캘리포니아 북부지방의 30,000여 가구들에서 발생했던 소비전력을 기반으로 k-Means 클러스터링 기법⁸⁾을 이용하여 가정의 전력 소비패턴을 분류하였다. Asma Dachraoui는 특정 시간대에 전력 수요가 몰리는 점에 주목하여 전력 수요가 몰리는 것을 미리 예측하는 방법을 제안하여 전력회사로 하여금 소비전력 가격 조절을 유도할 수 있도록 하였다⁹⁾. Asma의 연구에서는 아일랜드의 6000가구의 소비전력 데이터를 500일 동안 30분마다 수집한 데이터를 기반으로 MODL기법¹⁰⁾을 이용하여 전력 수요를 예측하였다. Hongliang Fei는 소비전력을 분석하여 가정에서 열 펌프 사용유무를 알아내는 방법을 제안하고, 이 결과를 열 펌프 판매회사의 마케팅 용도로 사용할 수 있다는 가능성을 제시하고 있다¹¹⁾. Hongliang은 2011년 1월 1일부터 2012년 9월 23일까지 6개 지역에 있는 300,000가구의 소비전력 데이터와 열 펌프 판매 회사의 제품 판매실적을 기반으로 BSVM(biased Support Vector Machine)을 이용하여 열 펌프 사용 유무를 구분하였다.

이처럼 소비전력 데이터에서 의미 있는 데이터를 추출하여 다양한 형태로 활용이 가능하다. 본 논문에서는 소비전력을 분석하여 구체적인 가족 유형을 분류하고자 한다. 앞서 소개된 연구들은 모두 라벨링 되지 않은 데이터를 수집하여 연구에 사용하였다. 그러나 가족 유형을 분류하는 분류기를 학습시키기 위해서는 소비전력 데이터가 어떠한 형태의 가족에서 발생한 정보인지 구분할 수 있는 라벨링된 데이터가 필요하다. 다양한 유형의 가족에서 발생하는 소비전력 데이터를 실제로 얻기란 현실적으로 매우 어려운 일이기 때문에, 본 연구에서는 가족 유형에 따른 소비전력 데이터를 생성하는 시뮬레이터를 개발하여 가족 유형을 분류하였다.

III. 제안하는 전력 패턴 생성 시뮬레이터 구조

본 연구에서는 가족 유형에 대한 전력 소비 시뮬레이션을 통하여 소비 전력 패턴을 얻는다. 시뮬레이션

은 두 단계로 진행된다. 첫 번째는 가족 유형을 입력 받아 가전기기 사용을 스케줄링하는 단계이고, 두 번째는 스케줄링 정보를 이용하여 정해진 기간 동안 소비전력을 생성하는 단계이다.

3.1 스케줄링

스케줄링 단계에서는 가족을 구성하기 위하여 일반적인 범위에서 가족 구성원 유형을 정의하여 기본 생활 패턴(기상, 출근, 퇴근, 취침)을 스케줄링 하고, 구성원 유형별로 옵션을 취서 기본 생활 패턴에 변화를 주었다. 이 때 옵션이란, 구성원 유형에 종속되는 예외상황을 말한다. 예를 들어, 초등학생의 경우 방과 후 수업, 고등학생의 경우 독서실, 회사원의 경우 야근 등의 옵션이 있을 수 있다. 이러한 옵션을 추가함으로써 같은 유형의 구성원이라고 하더라도 시뮬레이션을 진행할 때마다 다른 생활 패턴을 생성할 수 있다.

3.1.1 가족 유형 정의

본 연구에서 구현한 시뮬레이터는 가족 유형에 따라서 소비 전력 데이터를 생성한다. 가족 유형에 대해서 간단하게는 1인, 2인, 3인등과 같이 가족 구성원의 수로 유형을 구분할 수 있지만 가족 구성원의 수가 같다고 하더라도 구성원의 직업에 따라서 생성되는 소비전력 패턴은 달라질 수 있다. 예를 들어 2인 가정일 경우 맞벌이 부부일 수도 있고 남편은 직장인 아내는 전업주부일 수도 있다. 이 때 두 가정의 소비전력 패턴은 다르다.

따라서 보다 정확하게 가족유형을 정의하기 위해서는 가족 구성원의 수가 아니라 직업에 따라서 가족을 구성하여야 한다. 본 연구에서는 가족 구성원을 8가지로 분류하였다.

표 1과 같은 방식으로 가족 구성원을 조합하여 하나의 가정을 만들면 단순히 인원수를 기준으로 가족 유형을 구분하는 것보다 구체적인 형태의 가족을 구

표 1. 가족 구성원 유형
Table 1. Family member type

Type	ID
Infants	1
Kndergartener	2
Elementary School Student	3
Middle School Student	4
High School Student	5
University Student	6
Office Worker	7
Housewife	8

성할 수 있다.

3.1.2 가전기기 정의

일반적인 가정에서 사용하는 가전기기 및 전력을 발생시키는 장치들을 30가지로 분류하였다. 표 2는 가전기기의 분류를 나타낸다. 항목 중 유형이란, 가전기기의 유형을 나타내는데 24시간 사용하면서 스스로 동작하는 가전기기를 2로 분류하고, 사용시간에 비례하여 소비전력을 생성하는 가전기기들을 1로 분류하였다.

표 2. 가전기기 종류
Table 2. Appliance list

Appliance	ID	Type	Power(W)
Refrigerator	1	2	100
Air conditioner	2	1	1300
Kimchi fridge	3	2	130
Washing machine	4	1	130
Toaster	5	1	720
Dishwasher	6	1	1850
Coffee maker	7	1	1270
Vacuum cleaner	8	1	1900
Hair dryer	9	1	1237
Electric oven	10	1	1500
Microwave	11	1	1250
Air cleaner	12	2	40
Fan	13	1	55
TV	14	1	120
Steam iron	15	1	600
Printer	16	1	430
Wireless AP	17	2	6
Telephone	18	1	1
Audio	19	1	40
Kettle	20	1	2000
Rice cooker	21	1	500
1st-bathroom lamp	22	1	10.8
2nd-bathroom lamp	23	1	10.8
1st-room lamp	24	1	54
2nd-room lamp	25	1	36
3rd-room lamp	26	1	18
4th-room lamp	27	1	18
Livingroom lamp	28	1	80
Computer	29	1	170
Kitchen lamp	30	1	36

3.1.3. 스케줄링 모델

I 를 가족 구성원의 집합이라 하고 $i(i \in I)$ 는 가족 구성원의 ID라고 하자. $bs_i = \langle t_i^w, t_i^o, t_i^c, t_i^s \rangle$ 는 튜플(tuple)로 가족 구성원의 기본 생활 패턴을 나타낸다. t_i^w 는 기상시간, t_i^o 는 외출시간, t_i^c 는 귀가시간

그리고 t_i^s 는 취침시간이다. 가족 구성원의 ID마다 기본 생활 패턴은 고정되어 있지만 같은 ID를 가진 가족 구성원이라도 기본 생활 패턴은 달라질 수 있기 때문에 스케줄링을 할 때마다 균등 분포(Uniform Distribution)을^[12] 사용하여 일정 편차 안에서 시간을 선택할 수 있도록 하였다. $op_i = \langle o_i^1, o_i^2, \dots, o_i^n \rangle$ 는 튜플로 o_i^n 는 ID가 i 인 가족 구성원의 옵션을 나타낸다. 옵션은 가족 구성원 유형의 특성을 반영하여 기본 생활 패턴에 영향을 준다. 다음은 가족 구성원의 하루 스케줄을 나타낸다.

$$member_i = bs_i + op_i \quad (1 \leq i \leq 8) \quad (1)$$

식 (2)는 3인가족(직장인 아버지, 전업주부 어머니, 초등학교 자녀)의 표현이다.

$$family = member_3 + member_7 + member_8 \quad (if \ i = 3, 7, 8) \quad (2)$$

A 를 가전기기의 집합이라 하고 $a(a \in A)$ 는 가전 기기 ID라고 하자. $app_a = \langle t_a^{st}, d, \lambda \rangle$ 는 튜플로 가전기기의 기본정보를 나타낸다. t_a^{st} 는 가전기기 a 의 사용시작시간, d 는 사용시간, λ 는 가전기기 a 가 정해진 시간동안 사용되는 빈도를 나타내며, 소비전력 생성 단계에서 가전기기 사용시간을 정해진 시간 내에서 임의적으로 부과하기 위해서 사용되는 poisson inter-arrival time의^[13] 입력으로 사용된다. 식 (3)은 한 가정의 가전기기(냉장고, 에어컨, 토스터기, TV, 컴퓨터)의 기본 스케줄을 나타낸다.

$$home = app_1 + app_2 + app_5 + app_{14} + app_{29} \quad (3)$$

마지막으로 한 가족의 최종 가전기기 사용 스케줄은 식 (4)와 같이 나타낸다.

$$S = Schedule(family, home) \quad (4)$$

$Schedule$ 은 가족의 기본 스케줄과 가전기기의 기본 스케줄을 고려하여 최종 가전기기 사용 스케줄 S 를 출력하는 함수이다.

3.1.4 스케줄러 구조

스케줄러는 가족 및 기기사용 명세서와 기간을 입

력받아 해당 기간 동안의 기기 동작 스케줄을 생성한다. 기기 동작 스케줄은 가족 구성원의 스케줄과 배치된 기기, 기기의 소유 및 사용주기에 따라 생성된다. 기기의 사용은 주기적이거나 확률적일 수 있으며, 따라서 기기의 특성에 따라 명세하고, 명세에 따라 스케줄을 생성한다.

스케줄러의 전체적인 구조는 그림 1과 같다. 스케줄러는 크게 scheduler, home, family 패키지로 구성된다. 이때, scheduler 패키지는 Scheduler 클래스를 가지고, home 패키지는 Home 클래스, Appliance와 Appliance를 상속하는 각종 기기 클래스, Task 클래스를 가지며, family 패키지는 Family 클래스와 User 클래스를 가진다. Scheduler 클래스는 가족 및 기기사용 명세를 입력으로 받아 Home과 Family를 생성하고 Home에는 Appliance를, Family에는 User를 추가하여 전력 사용량 정보를 생성할 가구를 구성한다. 기기의 사용은 사용자의 스케줄에 영향을 받으므로 Home의 Appliance들은 Family의 스케줄의 참조하여 자신들의 운용 스케줄을 생성한다. 생성된 기기 운용 스케줄은 Scheduler를 통해 파일로 저장된다.

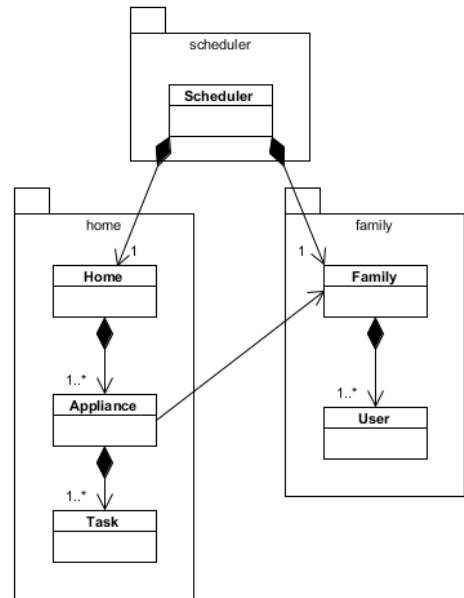


그림 1. 스케줄러 구조
Fig. 1. Architecture of Scheduler

3.2 소비전력 생성

소비전력 생성 단계는 시뮬레이션의 두 번째 단계로 첫 번째 단계에서 생성된 가전기기 사용 스케줄 정보를 받아서 최종적으로 해당 가정의 소비전력 정보

를 지정된 기간 동안 생성한다.

3.2.1 데이터 생성 모델

스케줄링 단계에서 생성된 스케줄 S 에 있는 가전기기의 개수를 n , A 를 가전기기의 집합이라 하고 $a(a \in A)$ 는 가전기기 ID라고 하자. t_a^{st} 는 최종적으로 스케줄된 가전기기 a 의 사용시작시간이고 t_a^e 는 사용종료시간으로 정의한다. $time_a = \langle p_1, p_2, \dots, p_{1440} \rangle$ 은 가전기기 a 의 하루 소비전력을 매분마다 나타낸다. 이 때 분당 소비전력은 다음과 같은 조건에 의해 결정된다.

$$p_i = \begin{cases} \text{electric power (if } t_a^{st} \leq i \leq t_a^e) \\ \text{standby power (otherwise)} \\ , (1 \leq i \leq 1440) \end{cases} \quad (5)$$

이 때 j 번째 날의 가족 소비전력 패턴은 식 (6)과 같고, 데이터 생성 기간을 p 라 할 때 정해진 기간 동안의 소비전력 패턴은 식 (7)과 같다.

$$Day_j = \sum_{a=1}^n time_a \quad (6)$$

$$Power = \text{concat}(Power, Day_j), (1 \leq j \leq p) \quad (7)$$

한 가정의 하루 소비전력 패턴은 각 가전기기의 하루 소비전력 패턴의 합으로 나타낼 수 있고 특정 기간 동안의 소비전력 패턴은 하루 소비전력 패턴들의 연결(concatenation)로 나타낼 수 있다.

3.2.2 데이터 생성 흐름

그림 2는 데이터 생성 흐름을 나타낸다. 첫 번째 단계에서 생성한 스케줄 정보를 읽어서 가전기기의 개수와 데이터 생성 기간을 n 과 p 에 입력한다. j 는 생성 기간을 나타내는 인덱스이고 a 는 가전기기의 개수를 나타내는 인덱스이다. 가전기기의 유형을 확인한 다음 1이면 해당 가전기기의 전력(electric power)이 소비전력 생성함수 $func$ 의 입력으로 들어가서 분당 소비전력을 생성한다. 가전기기의 유형이 2이면 24시간을 사용하면서 스스로 동작을 하는 가전기기이므로 전력과 대기전력(standby power)이 $func$ 함수의 입력으로 들어가 분당 소비전력을 생성한다. 이러한 방식으로 n 번 반복되면서 가전기기별로 생성한 하루 소비전력 패턴을 합산하여 한 가정의 소비전력 패턴으로 만든다. 이 과정을 처음에 입력한 기간인 p 만큼 반복하여 전체 소비전력 패턴을 생성한다.

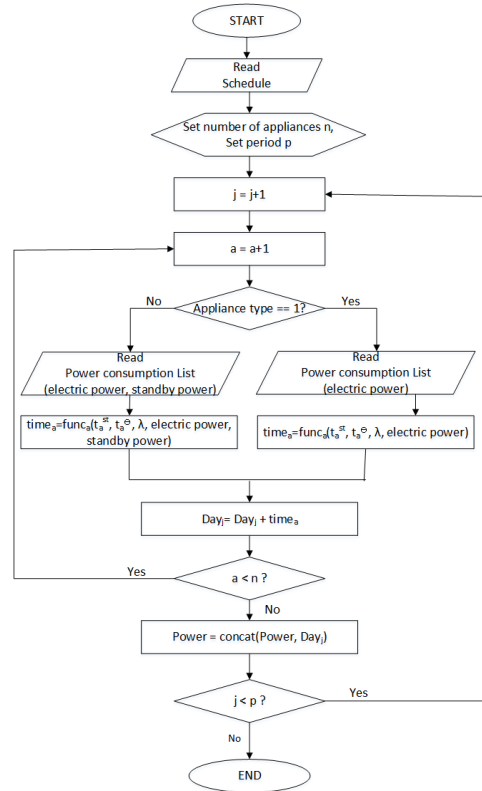


그림 2. 소비전력 생성 흐름 (j 는 생성기간 p 의 인덱스, a 는 가전기기 개수 n 의 인덱스 그리고 $func$ 는 소비전력 생성 함수를 나타냄)

Fig. 2. Data generation flow (j is an index of period p , a is an index of the number of appliances n and $func$)

3.2.3 가족 유형 별 데이터 생성

지금까지 설명한 시뮬레이터를 이용하여 가정 유형별 시뮬레이션 데이터를 생성하였다. 그림 3은 가족 인원수에 따른 분당 소비전력 패턴을 생성한 것이다. 좌측 상단부터 시계방향으로 1인(직장인) 가정, 2인(맞벌이 부부) 가정, 3인(맞벌이 부부와 초등학교 자녀 1명) 가정, 4인(맞벌이 부부와 초등학교 자녀 2명) 가정이다. x축은 분(Minute) y축은 전력(Watt)를 의미한다. 가족 인원수에 따라서 소비전력 패턴의 차이가 있음을 알 수 있다.

그림 4는 같은 수의 가족이라고 하더라도 가족의 형태에 따라 소비전력 패턴이 달라질 수 있음을 나타낸 것이다. 1열은 2인 가정이지만 왼쪽은 맞벌이 부부이고 오른쪽은 남편은 직장인이고 아내가 가정주부일 때의 분당 소비전력 패턴이다. 아내가 직장인일 때보다 가정주부일 때 낮 시간 동안의 전력 소비가 많음을 알 수 있다. 2열은 3인 가정으로 왼쪽은 맞벌이 부부와 초등학교 자녀 1명으로 구성된 가족이고, 오른쪽

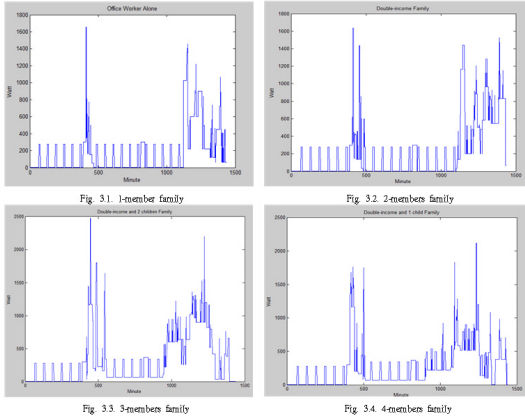


그림 3. 가족 구성원 수에 따른 하루 소비전력 패턴
Fig. 3. Power consumption pattern of families which consist of different number of members

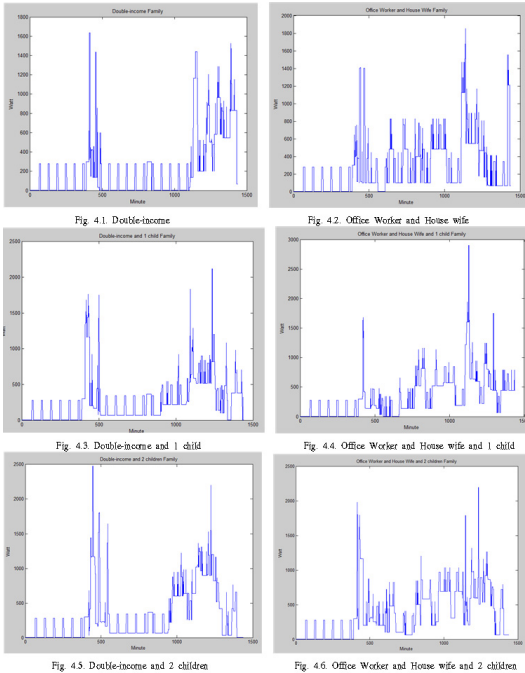


그림 4. 가족 유형에 따른 소비전력 패턴의 변화
Fig. 4. Changing power consumption pattern according to family types

은 남편은 직장인 아내는 가정주부 그리고 초등학교 1 명으로 구성된 가족의 소비전력 패턴이다. 3열은 4인 가정으로 왼쪽은 맞벌이 부부와 초등학교 자녀 2명으로 구성된 가족이고, 오른쪽은 남편은 직장인 아내는 가정주부 그리고 초등학교 자녀 2명으로 구성된 가족의 소비전력 패턴이다. 2인 가정의 경우와 마찬가지로 3인과 4인 가정도 가족의 형태에 따라서 소비전력 패턴이 차이가 있음을 알 수 있다.

턴이 차이가 있음을 알 수 있다.

3.3 특성 추출

시뮬레이터에서는 분당 소비전력 데이터를 생성한다. 이 데이터를 이용하여 시간을 이용한 feature를 생성한다.

3.3.1 용어 정의

가족 유형 분류 실험에서 사용할 특성들을 도출하기 위한 용어들을 표 3에 정의하였다.

표 3. 용어
Table 3. Terminology

Term	Description
$x(j)$	power consumption of j-th minute
$P(i)$	average power consumption of i-th hour
$D(k)$	average power consumption of k-th day
$MAX(i)$	maximum power consumption within i-th hour
$MIN(i)$	minimum power consumption within i-th hour

3.3.2 특성

본 논문에서는 총 7가지 유형의 특성을 도출하였으며 표 4에 정의하였다.

첫 번째 특성은 시간대 별 평균 소비전력으로 i 는 시간, j 는 분을 나타낸다.

$$P(i) = \frac{\sum_{j=1}^{60} x(60 \cdot i + j)}{60} \quad (i = 0, 1, \dots, 23) \quad (8)$$

두 번째 특성은 시간대 별 최고/최소값으로 i 는 시간, j 는 분을 나타낸다.

$$\begin{aligned} MAX(i) &= \max(x(60 \cdot i + j)) \\ MIN(i) &= \min(x(60 \cdot i + j)) \end{aligned} \quad (i = 0, 1, \dots, 23, j = 1, 2, \dots, 60) \quad (9)$$

세 번째 특성은 하루 평균 소비전력으로 다음과 같으며 j 는 분, k 는 날을 나타낸다.

$$D(k) = \frac{\sum_{j=1}^{1440} x(j)}{1440} \quad (k = 1, 2, \dots) \quad (10)$$

표 4. 특성 종류와 설명
Table 4. The feature list and description

Feature	Description
①	Average power consumption per hour
②	Maximum/Minimum power consumption per hour
③	Average power consumption of a day
④	Average power consumption per hour and a day
⑤	Maximum/minimum power consumption per hour and average power consumption of a day
⑥	Maximum/minimum power consumption per hour and average power consumption per hour
⑦	Maximum/minimum power consumption per hour and average power consumption per hour and a day

네 번째 특성은 첫 번째와 세 번째 특성의 조합이며, 다섯 번째 특성은 두 번째와 세 번째 특성의 조합이다. 여섯 번째 특성은 첫 번째와 두 번째 특성의 조합이고 마지막으로 일곱 번째 특성은 첫 번째부터 세 번째까지 특성의 조합으로 표현할 수 있다.

IV. 실험

4.1 실험 데이터 및 제약사항

실험을 위한 데이터는 보편적인 가족의 형태인 맞벌이 부부, 맞벌이 부부와 자녀 1명, 직장인 남편과 가정주부 아내 그리고 자녀 1명, 직장인 남편과 가정주부 아내 그리고 자녀 2명의 4가지 형태의 가족 유형으로 제한한다. 하나의 가족 유형이 한 가지 형태로 정형화 되는 것을 피하기 위해, 가족 구성원의 옵션과 가전기기의 조합을 바꿔가며 4가지 세부유형의 조합으로 하나의 가족 유형을 구성한다. 예를 들어 맞벌이 부부의 경우 직장인의 옵션으로 야근과 회식이 발생할 확률을 조절하고, 표 2에 있는 가전기기의 조합을 바꿔가며 4가지 각기 다른 유형의 데이터를 생성하고 이 데이터를 합쳐서 맞벌이 부부의 데이터를 이룬다. 데이터는 유형별로 400주씩 1600주의 소비전력 패턴을 생성하였으며 가족 유형 분류기를 학습시키기 위한 데이터는 1200주를 사용하였으며 400주는 분류기를 검증하기 위한 샘플로 사용하였다.

4.2 분류기 학습

3.3절에서 추출한 특성을 이용하여 분류 오차를 최소로 하는 분류기를 찾기 위해 k-묶음 교차 검증법

(k-fold Cross validation)을^[14] 사용하였다. k-묶음 교차 검증법은 수집된 샘플을 k개의 묶음으로 나눠서 하나의 묶음은 분류기를 검증하기 위한 샘플로 사용하고 k-1개의 묶음은 분류기 학습 샘플로 사용한다. 본 논문에서는 4-묶음 교차 검증법을 사용하였다. 분류기 학습을 위한 알고리즘으로는 kNN(k-Nearest Neighbors)를 사용하였다^[15]. 표 5는 4-묶음 교차 검증법을 이용하여 분류기의 오류율을 비교한 것이다. 분류기의 오류율이란, 전체 샘플 데이터 중에서 잘못 분류된 샘플 데이터의 비율을 의미한다. 각 유형별로 300주씩 1200주의 샘플 데이터를 4개의 묶음으로 나누어서 실험을 진행하였다.

표 5. 4-묶음 교차 검증법을 이용한 오류율 비교
Table 5. Comparison of error rate using 4-fold cross validation

Feature	fold-1 (%)	fold-2 (%)	fold-3 (%)	fold-4 (%)	Average (%)
①	31.57	21.55	23.05	25.56	25.43
②	25.56	7.26	31.57	19.04	20.86
③	54.38	11.27	25.56	59.89	37.77
④	31.57	21.55	23.05	25.81	25.50
⑤	25.06	7.26	31.57	19.04	20.73
⑥	25.06	13.28	30.32	19.29	21.99
⑦	25.06	13.28	29.82	19.29	21.86

4.3 가족유형 분류 실험결과

4.2절에서 만든 분류기를 이용하여 분류기가 예측한 가족 유형과 실제 가족 유형을 비교하였다. 3번의 분류실험을 하였으며, 테스트 데이터는 매 실험마다 총 4가지 가족 유형별로 25주씩 전력소비패턴을 생성하여 사용했다. 실험 결과는 표 6과 같다. 표 6의 Type 1~Type 4는 4.1절에서 정의한 가족 유형을 나타내며, Type1은 맞벌이 부부, Type 2는 맞벌이 부부와 자녀 1명, Type 3은 직장인 남편과 가정주부 아내 그리고 자녀 1명, Type 4는 직장인 남편과 가정주부 아내 그리고 자녀 2명으로 구성된 가족을 의미한다. 첫 번째 실험에서 총 100개의 테스트 데이터 중 22개의 데이터가 잘못 분류되었으므로 22%의 오류율을 보였다. 두 번째 실험과 세 번째 실험도 똑같은 방법으로 오류율은 각각 14%와 20%이며 세 번의 실험결과 분류기의 평균 오류율은 18.67%로 학습단계에서의 오류율보다 2.06% 낮아졌다.

그림 5는 실험 결과를 토대로 가족 유형별 분류 정

표 6. 분류 실험 결과
Table 6. The result of classification test

Test 1	Real Family Type				
		Type 1	Type 2	Type 3	Type 4
Predicted Family Type by Classifier	Type 1	9	5	-	-
	Type 2	16	20	-	1
	Type 3	-	-	25	-
	Type 4	-	-	-	24
Error Rate	22%				
Test 2	Real Family Type				
		Type 1	Type 2	Type 3	Type 4
Predicted Family Type by Classifier	Type 1	18	1	-	-
	Type 2	7	24	-	6
	Type 3	-	-	25	-
	Type 4	-	-	-	19
Error Rate	14%				
Test 3	Real Family Type				
		Type 1	Type 2	Type 3	Type 4
Predicted Family Type by Classifier	Type 1	16	-	-	10
	Type 2	9	25	-	1
	Type 3	-	-	25	-
	Type 4	-	-	-	14
Error Rate	20%				

확도를 나타낸 것이다. 전체적으로 Type 1의 분류 정확도가 세 번의 실험에서 평균 57.33%로 가장 낮았고 가족 유형 3의 분류 정확도가 평균 100%로 가장 높았다. 결과적으로 시간별 최대/최소 전력과 하루 소비전력을 조합한 특성을 이용하여 실험데이터를 분류한 결과 약 81%의 정확도로 가족 유형을 분류하였다.

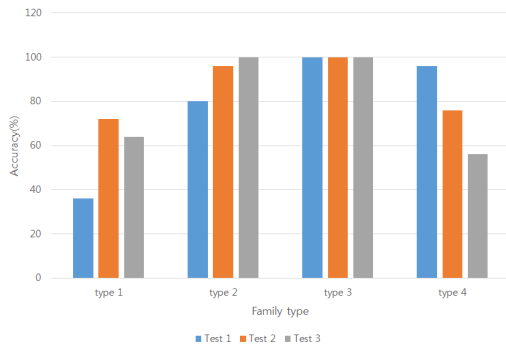


그림 5. 가족 유형 별 분류 정확도 (type1~type4는 실험에서 정의한 가족 유형, Test1~Test3은 실험 순서를 나타냄)
Fig. 5. Accuracy of classification for each family type (type1~type4 represent the family types and Test1~Test3 stand for an order of experiments)

V. 결론 및 고찰

본 논문에서는 스마트 그리드 환경에서 가정의 소

비전력 패턴과 데이터 마이닝 기법을 이용하여 가족 유형을 구분하였다. 다양한 가족 유형의 소비전력 데이터를 생성하기 위하여, 가족 유형 별로 분당 소비전력 패턴을 생성하는 시뮬레이터를 구현하였다. 시뮬레이터에서 생성된 데이터에서 가족 유형을 분류할 수 있는 몇 가지 특성들을 도출하여 kNN 알고리즘을 기반으로 하는 가족 유형 분류기를 학습시켰다. k-묶음 교차 검증법(k-fold Cross validation)을 이용하여 가장 성능이 좋은 분류기를 선택하여 가족 유형을 분류해본 결과 약 81%의 정확도로 가족 유형을 분류할 수 있었다.

본 연구는 향후 국내에 스마트 그리드 환경이 활성화 되었을 때, 서비스를 하는 third-party들이 가족 유형을 분류하고 자신의 제품을 홍보하는 수단으로 사용할 수 있다. 예를 들어 학습지를 홍보할 때 학생이 있는 집임을 알고 전화를 하는 것이 무작위로 홍보전화를 하는 것보다 훨씬 효율적인 것이다. Third-party들의 전력 데이터 활용이 자칫 사생활 침해로 이어질 수 있을지도 모른다는 논란이 있을 수 있지만, 지능형 전력망의 구축 및 이용촉진에 관한 법률 제 23조 1항에 따르면 ‘지능형전력망 사업자는 지능형전력망 서비스를 원활하게 제공하기 위하여 필요한 경우에는 다른 지능형전력망 사업자에게 지능형전력망 정보의 제공 또는 공동 활용을 요청할 수 있다.’라고 명시되

어 있으므로 third-party들의 전력 데이터 활용은 충분히 가능하다. 단, 제공되는 정보에 개인정보가 포함되어 있었으면 정보주체의 동의를 받아야하지만 주체가 명시되지 않은 전력 사용정보로 개인을 식별할 수는 없으므로 추가적인 동의는 없어도 될 것으로 보인다.

이번 연구에서는 단순히 소비전력 패턴을 kNN 알고리즘을 사용하여 분석하였지만, 향후 연구는 가족 유형 분류에 소비전력 뿐만 아니라 날씨, 소득 수준, 거주 지역과 같은 제 3의 데이터들을 반영하여 분류 정확도를 더욱 높이는 방법으로 연구를 진행할 예정이다.

References

[1] F. Hassan, "The path of the smart grid," *Power and Energy Mag., IEEE*, vol. 8, no. 1, pp. 18-28, 2010.

[2] S. Mun, "The concept of smart grid," *Information & Communications Mag.*, vol. 27, no. 4, pp. 3-9, Mar. 2010.

[3] N.-K. Won, B.-N. Ha, S.-W. Lee, I.-Y. Seo, and M.-J. Jang. "A large scale smart-grid field-test in Jeju," *J. KIEE*, vol. 61, no. 1, pp. 1-6, 2012.

[4] NIST, *Nist framework and roadmap for smart grid interoperability standards, release 2.0*, Jan. 2012.

[5] S. C. Lee and Y. Kim. "A consumer level simulation model for demand response analysis on smart grid," 2012.

[6] S. Yoshiyuki et al. "Development of residential energy end-use simulation model at city scale," in *Proc. Eighth Int'l IBPSA Conf.*, pp. 1201-1208, Eindhoven, Netherlands, Aug. 2003.

[7] A. Adrian, et al. "Drivers of variability in energy consumption," *Submitted for review to European Conf. on Machine Learning*, 2013.

[8] C. Ding and X. He, "K-means clustering via principal component analysis", in *Proc. Int'l Conf. Machine Learning(ICML)*, pp. 225-232, 2004.

[9] D. Asma, A. Bondu, and A. Cornuejols, "Early classification of individual electricity consumptions," *RealStream*, 2013.

[10] B. Marc, "MODL: A bayes optimal discret-

ization method for continuous attributes," *Machine learning*, vol. 65, no. 1, pp. 131-165, 2006.

[11] F. Hongliang, et al., "Heat pump detection from coarse grained smart meter data with positive and unlabeled learning," in *Proc. 19th ACM SIGKDD Int'l conf. on Knowledge discovery and data mining, ACM*, 2013.

[12] B. R. Barmish and C. M. Lagoa, "The uniform distribution: A rigorous justification for its use in robustness analysis," *Mathematics of Control, Signals and Systems*, vol. 10, no. 3, pp. 203-222, 1997.

[13] M. Thomas, *Non-life insurance mathematics: an introduction with the Poisson process*, Springer, 2009.

[14] K. Ron, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *IJCAI*, vol. 14, no. 2, 1995.

[15] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175-185, 1992.

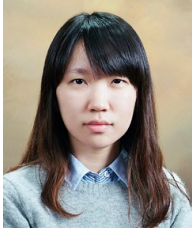
김 지 현 (Ji-hyun Kim)



2010년 2월 : 부산대학교 정보 컴퓨터공학부 학사
 2012년 8월 : 부산대학교 전자 전기 컴퓨터공학과 석사
 2012년 8월~현재 : 부산대학교 전자전기 컴퓨터공학과 박사 과정

<관심분야> 정보보안, 스마트그리드, 암호이론, 데이터마이닝, IoT

이 윤 진 (Yun-jin Lee)



2012년 2월 : 부산대학교 정보
컴퓨터공학부 학사

2012년 2월~현재 : 부산대학교
전자전기 컴퓨터공학과 석사
과정

<관심분야> 스마트그리드 보안,
IoT

김 호 원 (Ho-won Kim)



1995년 2월 : 경북대학교 전자공
학과 학사

1995년 2월 : 포항공과대학교 전
자전기공학과 석사

1999년 2월 : 포항공과대학교 전
자전기공학과 박사

1998년 12월~2008년 2월 : 한
국전자통신연구원(ETRI) 정보보호연구단 선임연
구원 / 팀장

2008년 3월~현재 : 부산대학교 정보컴퓨터공학부 부
교수

<관심분야> 스마트그리드 보안, RFID/USN 정보보
호 기술, PKC 암호, VLSI 설계, embedded
system 보안, IoT