

<https://doi.org/10.7236/JIIBC.2021.21.3.7>
JIIBC 2021-3-2

기계학습 기반 비선형 전력수요 패턴 GP 모델링

GP Modeling of Nonlinear Electricity Demand Pattern based on Machine Learning

김용길*

Yong-Gil Kim*

요약 자동화된 스마트 그리드의 등장은 이러한 문제에 대응을 위한 필수적인 장치가 되고 있으며 스마트 그리드 기반 사회로의 진전을 가져오고 있다. 스마트 그리드는 전기 공급 업체와 소비자 간의 양방향 통신을 가능하게 하는 새로운 패러다임이다. 스마트 그리드는 전력 그리드를 보다 안정적이고 신뢰할 수 있으며 효율적이고 안전하게 만들기 위한 엔지니어의 이니셔티브로 인해 등장했다. 스마트 그리드는 전력 소비자가 전력 사용에서 더 큰 역할을 할 수 있는 기회를 창출하고 전력을 현명하고 효율적으로 사용하도록 동기를 부여한다. 이에 본 연구에서는 기계 학습을 통한 전력 수요 관리에 중점을 둔다. 기계 학습을 사용한 수요 예측과 관련하여 현재 다양한 기계 학습 모델이 소개되어 적용되고 있는데 이에 관한 체계적인 접근이 요구되고 있다. 특히 GP 학습 모델의 경우에 일반 소비 예측 및 데이터의 가시화와 관련하여 다른 학습 모델보다 장점이 있지만, 스마트 미터 데이터의 예측과 관련해서는 데이터 독립성에 강한 영향을 받는다.

Abstract The emergence of the automated smart grid has become an essential device for responding to these problems and is bringing progress toward a smart grid-based society. Smart grid is a new paradigm that enables two-way communication between electricity suppliers and consumers. Smart grids have emerged due to engineers' initiatives to make the power grid more stable, reliable, efficient and safe. Smart grids create opportunities for electricity consumers to play a greater role in electricity use and motivate them to use electricity wisely and efficiently. Therefore, this study focuses on power demand management through machine learning. In relation to demand forecasting using machine learning, various machine learning models are currently introduced and applied, and a systematic approach is required. In particular, the GP learning model has advantages over other learning models in terms of general consumption prediction and data visualization, but is strongly influenced by data independence when it comes to prediction of smart meter data.

Key Words : Guess Process, SVM, interigent network, BPL, AMR

*정회원, 조선이공대학교 컴퓨터보안과
접수일자 2021년 3월 24일, 수정완료 2021년 4월 24일
게재확정일자 2021년 6월 4일

Received: 24 March, 2021 / Revised: 24 April, 2021 /
Accepted: 4 June, 2021

*Corresponding Author: ygkim@cst.ac.kr

Department of computer security, chosun college of science & technology, Korea .

I. 서론

전력 수요 예측은 전력 시스템 계획 및 운영에 대한 의사 결정을 위한 기반을 제공하므로 전력 산업에서 필수적인 역할을 한다. 전기 회사는 단기, 중기 또는 장기 예측에 적용되는 매우 다양한 전력 수요 예측 방법을 사용하고 있다. 그러나 전기의 사용은 기상학적 요인과 사회 경제적 요인 간의 복잡한 상호 작용에서 비롯된다. 이러한 동적인 환경에서는 일반적인 예측 기술로는 충분하지 않으며 보다 정교한 방법이 필요하다. 목표는 수요 변화로 이어지는 모든 요인을 효과적으로 풀고 근본적인 원인을 파악하는 것이다. 데이터 세트는 입력이 과거의 데이터를 나타내고 대상이 미래의 정보를 나타내는 인스턴스 그룹으로 변환된다. 미래 수요와 관련된 온도는 미리 알 수 있으므로 입력변수로도 사용된다. 이 경우 이전 이들 전 데이터에서 다음 날을 예측하도록 프로젝트 구성을 설정할 수 있다. 그 결과 예측 정확도가 향상되어 최선의 조치를 결정하는 데 더 나은 정보를 제공한다. 모델 선택과 관련하여 순서 선택 알고리즘은 신경망에서 최적의 뉴런 수를 얻는 데 사용되며, 이 경우에 증가 순서 방식이 사용된다. 뉴런 증가 과정 중에 여러 하위 집합에 대한 손실 내력을 얻을 수 있는데, 훈련 손실 및 선택 오류를 검토할 필요가 있다. 손실 값을 살펴보면 예측 모델이 손실 값이 낮음으로 정확도가 높는데, 이를 바탕으로 신경망의 아키텍처를 작성할 수 있다. 다음으로 시험 분석의 목적은 신경망의 예측력 평가하는 과정이다. 이 과정은 예측 모델이 배포 단계로 이동하기에 충분한지 여부를 결정한다. 먼저, 스케일링된 신경망 출력과 독립 테스트 하위 집합에 대한 해당 대상 간의 선형 회귀 분석을 수행한다. 전체적으로 평균 백분율 오류가 모델 배포의 척도가 된다. 예측 모델을 테스트하고 나면 이를 적용하여 미래 전력 수요를 예측할 수 있다.

II. 기술적 배경

1. 스마트 미터 데이터 전송 방식

기존의 배전 시스템은 전압 레벨을 증가시킨 다음 전압 레벨을 점진적으로 줄여 최종 사용자에게 전달함으로써 중앙 발전소에서 생성된 전기 에너지를 운송하는 데 사용되었다. 그러나 오늘날 전력 시장에는 스마트 전력망이 필요하다. 스마트 그리드는 원격 관리를 통해 분산 에너지 생성을 모니터링하고 비상 상황에서 부하를 차단

할 수 있는 기회를 제공한다[1]. 데이터 레이트, 데이터 전송 범위, 비용 및 신뢰성은 사용자에게 의한 상이한 데이터 전송 방법의 선호가 고려되었다. 데이터 전송 방법의 사용 영역은 전기 측정의 원격 모니터링 및 자동화 제작 장치, 기계/시스템의 주기적 제어로 요약할 수 있다[2]. 스마트 그리드 환경에서 데이터 전송 방법 외에도 통신 기술과 보안 문제가 다른 관점에서 조사되었다. 스마트 미터기로 모든 네트워크 문제를 해결할 수 있다는 것은 명백하다[3]. 양방향 통신과 함께 스마트 계량 및 데이터 관리는 에너지의 생성, 전송 및 분배에 많은 수의 축진을 제공한다[4]. 예를 들어, 에너지 부족시 에너지 사용에 대한 소비자 임계치를 적용하여 정전을 피할 수 있으며 과부하로 인한 다운 타임을 줄임으로써 가입자가 전력이 없는 것을 방지할 수 있다. 이 외에도 에너지 저장은 마이크로 그리드 네트워크의 각 계층적 주파수 제어에서 에너지 공급 및 부하 수요를 조절함으로써 네트워크 균형의 또 다른 형태이다[5]. 에너지 저장 애플리케이션은 또한 전기 자동차, 재생 가능 발전 시스템, 전기 시장, 에너지 정책 및 전력 시스템 관리에 의해 상이한 스마트 그리드 목적을 위해 수행되었다[6]. 또한, 효율적인 에너지 보호를 제공하기 위해 지능형 변압기를 효과적인 방식으로 작동시킬 수 있으며 스마트 그리드 환경에서 정전을 유발하는 사이버 공격 탐지를 위해 알고리즘을 개선할 수 있다[7]. 데이터 전송 방법은 주로 모바일 컴퓨터, 휴대폰, 컴퓨터의 추가 하드웨어, 원격 모니터링 및 주기적 측정에 사용된다. 차량 추적 및 자동화 장치, 기계 또는 시스템(스마트 홈 애플리케이션)의 원격 제어와 같은 많은 분야에서 사용된다. 집이나 직장에서 데이터 전송 방법을 사용하면 데이터 전송 강도가 증가한다. 이 과정은 안전하고 효율적이며 신뢰할 수 있는 통신 문제에 대한 새로운 연구이다[8]. BPL(Broadband over Powerline) 기술은 데이터 전송률 및 데이터 전송 거리 측면에서 AMR(Automated Meter Reading)의 데이터 전송 방법 중의 하나로 간주될 수 있다. 인터넷 네트워크가 확산되면서 WiFi 기반 WSN(무선 센서 네트워크)과 스마트 그리드의 AMR 시스템을 인터넷에 연결하면 많은 이점이 있다. 그러나 실시간으로 값을 읽는 무선 신호를 통한 측정은 실제 값에 더 가까운 것으로 밝혀졌다. 지능형 커뮤니티에서 전력 스마트 그리드를 향한 WiFi WSN을 기반으로 한 차세대 AMR의 네트워크 아키텍처가 일반화된 형태이다. 원격 데이터 관측의 중요성을 이해하면 자동 계량기 관측 방법의 이점이 점점 일반화되고 있다.

2. 지능형 네트워크 인프라

스마트 그리드 인프라에는 관리, 보호, 정보 및 통신 시스템, 에너지 효율, 배출, 전력 품질 및 공급 보안 문제가 있다. 이외에 광범위한 주제가 있지만, 당사자가 네트워크의 각 지점에서 변경 사항을 인식할 수 있도록 네트워크 자동화를 잘 관리해야 한다. 네트워크에서 얻은 데이터는 다를 수 있지만 동일한 하드웨어에서 다른 데이터를 얻을 수 있으며, 데이터를 얻기 위해 공통 장비를 개발하여 다른 단위로 데이터를 평가할 수 있다. 청구 부서가 색인을 읽는 동안 장애를 탐지 지속적으로 주전원 전압을 모니터링한다. 두 개의 다른 단위가 카운터에서 이 값을 얻을 수 있다. 사용자에게 스마트 그리드 응용 프로그램의 장점 외에도 발생한 문제도 중요하다. 특히, 장비를 전체적으로 작동하는 동안 발생할 수 있는 하드웨어 문제를 고려해야 한다. 또한 하드웨어 외에도 원격 스위칭 시스템을 사용하여 조작할 필요도 있다. 핵심 세그먼트 및 구성 요소로 스마트 그리드 통신 아키텍처 체계를 정의한다.

에너지 품질 기준의 전달은 전류, 전압, 주파수 및 고조파로 제한될 수 있다. 다른 물리량은 그 파생물이므로 센서의 구조를 제한할 수 있다. 또한, 점점의 위치 또는 점 온도가 필요할 수도 있다. 다양한 응용 분야가 있지만, 물리적 크기를 측정하는 데 사용되는 센서 구조는 분배 시스템의 분석 도구를 참조하여 설계해야 한다[9]. 센서 응용은 광범위하게 적용되고 있지만, 지능형 센서를 사용하여 현재 상황에서 발생하는 진동의 관찰에 사용되는 브리지 도체로 더 많은 문제가 해결된다. 이 경우 전력선과 극을 스마트 그리드를 사용하여 애플리케이션으로 변환할 수 있다. 동시에 이것은 건물에 대한 아이디어를 줄 수 있다. 무선 센서로 모든 기계를 실시간으로 관찰할 수 있으며 수신된 데이터를 사용하여 에너지 소비 및 비용을 줄일 수 있다. 프로그램 내에서 수행될 때 기계에 대한 변경 및 생산된 제품의 이점은 표준 방식으로 이루어진다. 에너지 생산의 양방향 흐름의 적용, 전송 및 분배 연속성 및 저렴하고, 방법 데이터 관리 응답은 주로 자체적으로 치유할 수 있는 구조의 형성에서 마이크로 그리드 네트워크를 기반으로 측정되었다. 조정 작업을 수행하는 시스템은 많은 요소를 고려할 것이다[10].

마이크로 그리드는 전송 및 분배 손실을 방지하기 위해 소비자 근처에 풍력 단지 및 태양열 발전소와 같은 소규모 분산 소스에서 전기를 생성한다. 재생 가능 에너지와 분산 에너지 소스의 대규모 통합은 세계 모든 시장 모델에 대한 도전 과제이다. 장기적으로 재생 에너지원은

화석 연료 에너지 생산을 대체할 것이다. 동시에 에너지 생산 추세는 고전압 직류 전송 라인 또는 소위 슈퍼 그리드이다. 슈퍼 스마트 그리드 개념은 배전 시스템보다 리소스의 그리드 ICT 상호 연결의 필요성을 강조한다. 그림 1은 제어 흐름과 분리된 데이터 및 정보 흐름을 도시한다. 이러한 분리의 근거는 합법적인 당사자들이 actuator를 작동시킬 것이라는 데에 있다. 클라우드 컴퓨팅의 특정 특성에 중점을 두어 인터넷 규모의 플랫폼이 되며 스마트 그리드 사용 사례의 데이터 집약적 요구를 촉진할 수 있다. 스마트 그리드 데이터 클라우드에 대한 데이터 스토리지 및 데이터 액세스 모델을 나타낸다 [11]. 분배망에서 많은 문제를 식별하기 위해 정보가 필요할 수 있으므로 사용되는 망 장치로부터 필요한 정보를 얻기 위해 서로 협력하여 작업하는 것이 중요하다. 특히 스킬과 경험이 배후 동인이라는 점에서 인공 지능을 통한 기계 학습이 점차 활발해질 것으로 예측된다.

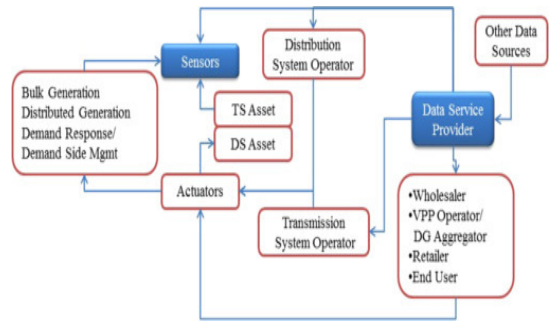


그림 1. 제어와 분리된 데이터 및 정보 흐름
 Fig. 1. Data and information flows separated from control

기술 통합 활동에는 두 가지 주요 과제와 요구 사항이 포함된다. 첫 번째는 실시간 알림, 제어 및 프로세스 통합과 같은 하드웨어이다[12]. 두 번째는 다른 응용 프로그램 간의 대량 및 트랜잭션 기반 데이터 교환과 같은 데이터이다. 네트워크 모니터링은 5단계로 수행할 수 있다: 소재, 초전도성, 분산 발전, 에너지 저장, 전력 전자. 마이크로 일렉트로닉스(그리드 친화형 기기)는 높은 전력 밀도, 높은 신뢰성 및 전력 품질, 향상된 전기 효율을 제공하여 주요 환경 이득 및 실시간 진단 기능을 향상시킨다. 사이버 보안 및 정보 보안 연구를 통해 지속성과 전력 공급의 질을 확보하는 동시에 점점 더 분산되고 재생 가능한 에너지원이 그리드에 연결되도록 하고 있다.[13][14].

3. GP(Guess Process) 모델

지도학습과 관련하여 종종 매개 변수 모델 $p(y | X, \theta)$ 을 사용하여 데이터를 설명하고 모수의 최적값을 최우(maximum likelihood) 또는 최대 사후 추정을 사용하여 추론한다. 필요한 경우에 전체 사후 분포 $p(\theta | X, y)$ 를 점 추정치 $\hat{\theta}$ 대신에 사용할 수 있다. 데이터 복잡성이 증가함에 따라 일반적으로 데이터를 합리적으로 잘 설명하려면 더 많은 수의 모수를 가진 모델이 필요하다. 고정된 수의 모수가 있는 모델을 사용하는 방법을 모수적 추론이라 한다. 비모수적 추론에서는 모수 수가 데이터 세트 크기에 따라 다르다. 예를 들어, Nadaraya-Watson 커널 회귀에서 가중치 w_i 은 각 관찰 대상 y_i 에 할당되고, 새로운 점 x 에서 대상 값을 예측하기 위해 가중 평균이 계산된다.

$$f(x) = \sum_{i=1}^N w_i(x) y_i, \quad w_i(x) = \frac{\kappa(x, x_i)}{\sum_{j=1}^N \kappa(x, x_j)} \quad (1)$$

x 에 더 가까운 관측치들은 더 멀리 떨어진 관측치보다 가중치가 높다. 가중치는 x 와 관측된 x_i 로부터 커널 κ 을 사용하여 계산된다. 특수한 경우는 KNN이다(k 개의 가장 가까운 관측치들은 $1/k$ 의 가중치를 갖고 그렇지 않은 것은 모두 0의 가중치를 갖는다). 종종 비모수적 방법이 예측을 위한 모든 훈련 데이터를 처리하기 위해 필요하기 때문에 모수적인 추론 방법보다 추론 시간이 더딜 수 있다. 반면에 비모수 모델은 훈련 데이터만 기억하면 되므로 훈련은 일반적으로 더 빠르다.

비모수적 방법의 또 다른 예는 GP이다. 모수 함수의 모수들에 대한 분포를 추론하는 대신에 가우스 과정을 사용하여 함수에 대한 분포를 직접 추론할 수 있다. 가우스 과정은 사전 함수를 요구한다. 일부 함수의 값을 관찰한 후에는 사후 함수로 변환할 수 있다. 이러한 맥락에서 연속 함수의 값 추론을 GP 회귀라 하지만 GP는 분류에 사용될 수 있다. 가우스 과정은 어떤 점 $x \in R^d$ 가 확률 변수 $f(x)$ 을 갖고 이들 유한 개수의 변수들에 관한 결합 분포 $p(f(x_1), \dots, f(x_N))$ 가 가우스 분포를 따르는 경우에 무작위 확률 과정이 된다.

$$p(f | X) \equiv N(f | \mu, K) \quad (2)$$

위 식에서 $f = (f(x_1), \dots, f(x_N))$ 이고, $\mu = (m(x_1), \dots, m(x_N))$, $K_{ij} = \kappa(x_i, x_j)$ 이다. 여기서 m 은 평균 함수로 보통 $m(x) = 0$ 이 사용된다.

κ 은 양정치 커널 함수 또는 공분산 함수이다. 따라서 가우스 과정은 모양이 K 에 의해 정의되는 함수들에 관한 분포를 나타낸다. 만일 x_i 와 x_j 가 커널에 의해 유사한 것으로 간주되면, 이들 점에서 함수의 값들 또한 유사한 것으로 기대할 수 있다.

하나의 GP 사전 함수 $p(f | X)$ 는 어떠한 데이터 y 의 관측 후에 GP 사후 함수 $p(f | X, y)$ 로 변환될 수 있다. 그런 다음 이러한 사후 함수를 사용하여 예측할 수 있다. 즉, 새로운 입력 X^* 에 대한 예측 f^* 는 다음과 같이 주어진다.

$$p(f^* | X^*, X, y) = \int p(f^* | X^*, f) p(f | X, y) df = N(f^* | \mu^*, \Sigma^*) \quad (3)$$

위 식에서 사후 예측 분포는 평균 μ^* , 공분 행렬 Σ^* 을 갖는 가우스 분포이다. GP의 정의에 따라 관찰된 데이터 y 와 예측 f^* 의 결합 분포는 N 개의 훈련 데이터와 N^* 개의 새로운 입력 데이터가 사용될 경우에 다음과 같다.

$$\begin{pmatrix} y \\ f^* \end{pmatrix} \sim N\left(0, \begin{pmatrix} K_y & K^* \\ (K^*)^T & K^{**} \end{pmatrix}\right) \quad (4)$$

여기서

$$\begin{aligned} K_y &= \kappa(X, X) + \sigma_y^2 I (\equiv N \times N) \\ K^* &= \kappa(X, X^*) (\equiv N \times N^*) \\ K^{**} &= \kappa(X^*, X^*) (\equiv N^* \times N^*) \end{aligned} \quad (5)$$

σ_y^2 은 K_y 의 대각선 항들의 잡음이다. 훈련 대상에 잡음이 없으면 0으로 설정되고 관측에 잡음이 있으면 0보다 큰 값으로 설정된다. 평균은 편의상 영 벡터로 설정된다. 사후 예측 분포에 관한 충분 통계량 μ^* 와 Σ^* 은 다음과 같다.

$$\begin{aligned} \mu^* &= (K^*)^T K_y^{-1} y \\ \Sigma^* &= K^{**} - (K^*)^T K_y^{-1} K^* \end{aligned} \quad (6)$$

지금까지 언급한 내용은 가우스 과정을 구현하고 회귀 문제에 적용하기 위해 알아야 할 최소 사항이다. 이제 일반 NumPy로 GP를 구현하는 방법과 scikit-learn 및 GPy에서 GP 구현을 나타내기로 한다. 먼저 NumPy로의 구현은 다음과 같다. 여기서는 가우스 커널 또는 RBF 커널이라는 제곱 지수 커널을 사용한다.

$$\kappa(x_i, x_j) = \sigma_f^2 \exp\left(-\frac{1}{2l^2}(x_i - x_j)^T(x_i - x_j)\right) \quad (7)$$

길이 모수 l 은 함수의 완만함을 제어하고 σ_f 은 종적인 변동을 나타낸다. 편이상 모든 입력 차원에 관해 같은 길이의 모수 l 을 고려한다.

가우스 과정에 사용할 수 있는 다른 많은 커널이 있다. 사전 확률 함수의 계산은 다음과 같이 진행 한다. 먼저 평균이 0인 사전 함수와 커널 모수 $l=1$ 과 $\sigma_f=1$ 을 갖는 공분산 행렬을 정의한다. 해당 GP에서 무작위 함수를 그리기 위해서는 대응하는 다변량 정규 분포에서 무작위 샘플을 추출한다. 그림 2는 다변량 정규 분포에서 다음 코드는 3개의 랜덤 샘플을 추출하여 평균 0과 95% 신뢰 구간(공분산 행렬의 대각선에서 계산)을 그래프로 나타낸다.

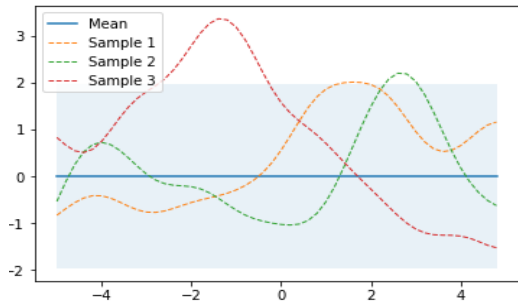


그림 2. 다변량 정규분포
 Fig. 2. Multivariate normal distribution

그림 3은 다변량 정규분포에서 잡음이 없는 훈련 데이터에 X_{train} 과 Y_{train} 을 적용한다. 다음은 사후 예측에서 세 개의 샘플을 가져와 평균, 신뢰 구간 및 훈련 데이터와 함께 가시화시키기 위한 코드이다. 잡음이 없는 모델에서 훈련 지점의 분산은 0이고 사후에서 가져온 모든 임의의 함수는 훈련 지점을 통과한다.

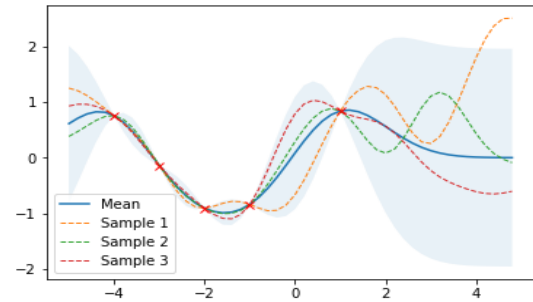


그림 3. 다변량 정규분포에서 잡음이 없는 모델
 Fig. 3. Noise-free model in multivariate normal distribution

그림 4는 다변량 정규분포에서 대부분의 관측 데이터에는 잡음이 있다. 모델에 일부 잡음이 포함되는 경우에

학습 포인트는 근사치만 계산되며 학습 포인트의 분산은 0이 아니다.

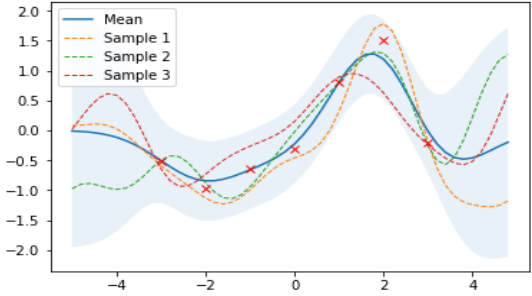


그림 4. 다변량 정규분포에서 잡음이 있는 모델
 Fig. 4. Model with noisy in multivariate normal distribution

커널 및 잡음 모수의 영향을 분석할 필요가 있다. 다음은 커널 모수 ρ 와 잡음 모수로서 σ_y 뿐만 아니라 σ_f 을 분석하기 위한 예 코드이다. 보다 큰 l 값은 보다 부드러운 함수로 이어지므로 훈련 데이터의 대략적인 근사치가 된다. 이에 반해 낮은 l 값은 훈련 데이터 점 간에 넓은 신뢰 구간으로 함수를 더 흔들리게 만든다. σ_f 은 GP를 통해 그려진 함수들의 종적 변화를 제어한다. 이는 훈련 데이터 영역 외부의 넓은 신뢰 구간에서 확인할 수 있다. σ_y 은 훈련 데이터의 잡음 양을 나타낸다. 더 큰 σ_y 값은 잡음이 많은 데이터에 대한 과적합을 방지하는 보다 대략적인 근사치를 만든다.

최적화된 커널 모수를 사용하면 훈련 데이터가 95% 신뢰 구간에 합리적으로 포함되며 사후 예측의 평균은 좋은 근사치이다. 위의 알고리즘은 더 높은 입력 데이터 차원에도 사용할 수 있다. 여기서 GP는 원점에서 xy -평면으로 확장되는 사인파의 잡음이 있는 샘플을 적합시키는데 사용된다. 다음 코드는 커널 모수의 최적화 전후의 잡음이 있는 샘플과 사후 예측 평균을 얻기 위해 사용될 수 있다. GP를 구현하는 라이브러리로 Scikit-learn과 GPy가 보통 사용된다. 먼저 Scikit-

learn은 GP 회귀(Gaussian Process Regressor) 모델을 구현하는데, 미리 정의된 커널 및 사용자 정의 커널로 구성할 수 있다. 커널도 구성할 수 있는데, 제공 지수 커널은 scikit-learn에서 RBF 커널을 사용하는데, RBF 커널만 $length_scale$ 을 갖는다(위 모수 l 에 대응). 모수 σ_f 을 사용하기 위해서는 ConstantKernel을 갖는 RBF 커널을 구성해야 한다. 이에 본 연구에서는 하나의 공분산과 서로 다른 공분산 커널 세트를 사용하는 가우스 프로세스 회귀 알고리즘을 제안한다.

III. GP 학습에 대한 구현

1. GP 학습에 의한 예측

사실 GP와 관련해서는 3장에서 제안된 자바스크립트 모듈보다는 matlab 패키지가 보다 신뢰성이 있다. 본 연구의 구현에서 예측과 관련해서는 SKlearn을 사용했다. GP와 관련하여 훈련될 2개 집합의 매개 변수들이 요구되는데, 첫 번째는 공변량에 관한 것이고, 두 번째는 대상 y에 포함되는 잡음에 관한 것이다. matlab 패키지를 사용하는 경우에 이들 두 집합은 자동 훈련 및 최적화된다. 그렇지만, 3장에서 제안된 스크립트를 사용할 경우에 잡음이 상술되어야 한다. 다행히도 잡음은 결과에 유의한 영향을 미치지 않는다. 참고로 Sklearn과 Matlab 패키지를 사용하는 경우에 초기 공변량 입력이 요구된다. 바람직하지 못한 초기화는 지역적인 점에 수렴할 수 있다. Matlab 패키지의 경우에 예측 결과는 초기 공변량에 그렇게 민감하지 않는 것으로 파악되었다. 그렇지만, Sklearn의 경우에 예측 결과는 초기 공변량에 매우 민감함이 확인되었다. 최적의 초기 공변량 및 잡음을 주기 위해 훈련 집합 상에서 교차 검증이 필요하다. 시간당 예측과 관련하여 시간당 모델 훈련 및 테스트를 위해 같은 방식들을 사용할 수 있다. 데이터 개수가 증가함에 따라 계산 복잡도는 유의하게 증가한다. 대규모 데이터 집합 상에서 교차 검증은 사실상 어려워 매개 변수들을 취하기 위한 작은 훈련/검증 데이터 집합으로 훈련을 하고 전체 데이터 집합에서 테스트하였다. 시간당 전력 소비와 관련해서는 훈련/검토 및 테스트 집합을 취한 표 1은 작은 샘플 데이터에서 실험하였다.

표 1. 작은 훈련.검증 데이터

Table 1. Small training. Verification data

hour	weekday	day	week	cosHour	occupancy	electricity-kWh	
2012-01-01 01:00:00	1	6	1	52	0.866025	0	111.479277
2012-01-01 02:00:00	2	6	1	52	0.965926	0	117.989395
2012-01-01 03:00:00	3	6	1	52	1.000000	0	119.010131
2012-01-01 04:00:00	4	6	1	52	0.965926	0	116.005587
2012-01-01 05:00:00	5	6	1	52	0.866025	0	111.132977

데이터 프레임은 특징들과 대상을 나타낸다.

입력 매개 변수들을 취하기 위한 교차 검증은 그림 5와 같다.

```
nugget = 0.008
```

```
theta = np.arange(0.05, 0.5, 0.05)
```

```
crossValidation(theta, nugget, 10, trainX_
hourlyElectricity, trainY_hourlyElectricity)
```

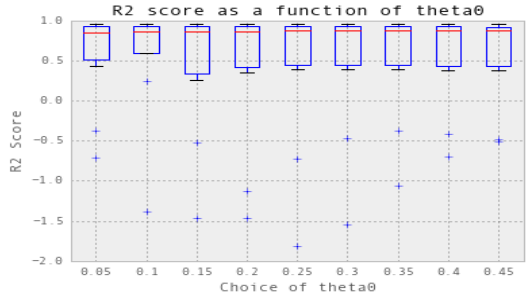


그림 5. 교차 검증

Fig 5. Cross validation

예측, 정확도 계산 및 가시화는 그림 6과 같다.

Train score R2: 0.957912601164

Test score R2: 0.893873175566

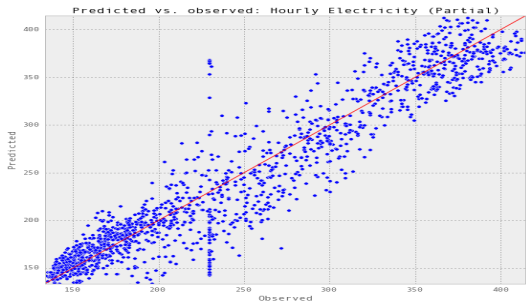


그림 6. 예측 정확도 계산

Fig. 6. Calculate prediction accuracy

그림 7은 예측과 관련된 일부 가시화 부분을 나타낸다. 보다 빠른 비교를 위해 Matlab 코드 출력을 사용했다. 그림 8은 시간당 전력 소비 예측을 위한 것이다.

```
<img src = "Pics/results_hourlyElectricity.png",
style="width:100%">
```

```
trainStart = '2012-01'
```

```
trainEnd = '2013-06'
```

```
testStart = '2013-07'
```

```
testEnd = '2014-10'
```

```
results_allHourlyElectricity
```

```
=
```

```
pd.read_excel('Data/results_allHourlyElectricity.xlsx')
```



```
def plotR2(df, energyType, title):
    testY = df[energyType]
    predictedY = df['predictedY']
    print "Test score R2:", sklearn.metrics.r2_score
        (testY, predictedY)
    plt.figure(figsize = (9,8))
    plt.scatter(testY, predictedY)
    plt.plot([min(testY), max(testY)], [min(testY),
        max(testY)], 'r')
    plt.xlim([min(testY), max(testY)])
    plt.ylim([min(testY), max(testY)])
    plt.title('Predicted vs. observed: ' + title)
    plt.xlabel('Observed')
    plt.ylabel('Predicted')
    plt.show()

plotR2(results_allHourlyElectricity,
'electricity-kWh', 'All Hourly Electricity')
Test score R2: 0.882986662109
```

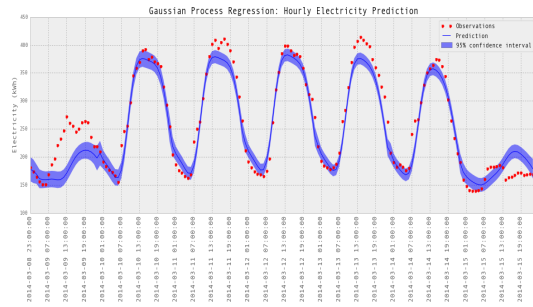


그림 7. 예측 가능 부분 가시화
 Fig.. 7. Visualization of predictable parts

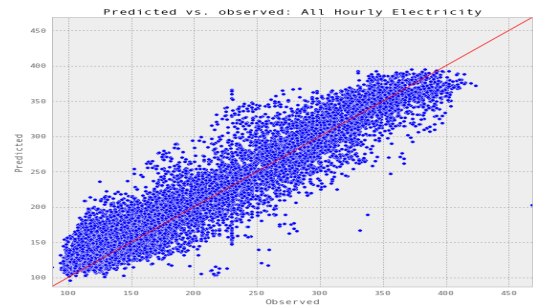


그림 8. 시간당 전력 소비 예측
 Fig. 8. Estimated power consumption per hour

가우스 기계 학습의 전체적인 성능이 전력 소비 예측과 관련하여 매우 우수한 것으로 나타났다. 가우스 과정

의 장점은 불확실한 예측값들에 대한 범위를 제공한다는 점이다. 그렇지만 장기간에 걸친 전력 소비 데이터의 경우에 계산적인 부담이 높다.

IV. 결론

본 연구에서는 기계 학습 알고리즘을 적용하여 전력 수요 관리를 용이하게 하기 위한 연구를 진행했다. 기계 학습을 사용한 수요 예측과 관련하여 가장 큰 문제점은 기계 학습을 사용한 수요 예측과 관련하여 현재 다양한 기계 학습 모델이 소개되어 적용되고 있는데 이에 관한 체계적이고 통합적인 고찰이 부족하다는 점이다. 이에 본 연구에서는 두 가지 기본 모델로서 하나는 모수 통계의 대표적인 회귀 학습 모델과 다른 하나는 전형적인 비모수 통계적 접근이라 할 수 있는 GP 학습 모델에 관한 접근 방식을 나타냈다. 다음으로 일반적인 수요 예측 관련 기계 학습과 관련하여 정적인 관점에서 GP 모델이 비교적 우수한 성능을 보이는 것으로 소개하고, 스마트 미터 데이터와 같은 시계열 데이터의 경우에 샘플 데이터의 독립성이라는 측면이 고려되어야 한다. 여기서 한 가지 중요한 사항이 발견되는데, 예를 들어 동물과 인간 학습의 관점에서 볼 때 몇 번의 학습은 매우 자연스럽다. 보통 동물은 중독 여부를 파악하기 위해 한 번 이상 독성이 있는 과일을 먹지 않는다는 점이다(심지어 유사한 과일이어도). 그렇지만 샘플이 아주 많은 경우에는 상황이 달라진다는 점이다. 심층 신경망 및 RF와 같은 인기가 있는 지도 학습 모델 중 다수는 훈련 데이터 수가 수천 개 미만일 경우에 충분히 신뢰할 수 있는 예측을 제공하지 못하기 때문에 다수의 유사 데이터를 많이 사용하게 된다. 이로 인해 데이터에 불확실성이 빈번하게 반영된다. 본 연구에서는 이를 방지한 보다 현실적인 결과를 얻기 위해 최적화 관련하여 GP의 손실 방식을 사용하였다. 본 연구에서는 웹 사이트에서 제공된 2011년 7월부터 2014년 10월까지 에너지 소비 데이터를 사용하여 기계 학습 모델 적용을 나타냈다. 기계 학습 실험 결과를 비교 분석해보면 순수 전력 소비만을 고려하는 경우에 SVM 모델이 전력 수요를 정확하게 예측할 수 있는 가장 신뢰할 수 있는 기계 학습 도구로 분석된다. 여기에서는 소비자의 정확한 에너지 사용 패턴이 필수적이며 여기서부터 수요 예측이 시작된다. 순수 전력 소비 외에 날씨 및 냉온수와 같은 기타 전기 수요가 요구되는 특징들이 반영될 경우에 가우스 회귀 또는 가우스 혼합 회귀 모델

과 같은 가중치 적용이 보다 좋은 성능을 보이는 것으로 분석되었다.

References

- [1] Bruno S, Lamonaca S, La Scala M, Rotondo G, Stecchi U. "Load control through smart-metering on distribution networks". In: Proceedings of the 2009 IEEE Bucharest PowerTech Innov Ideas Towar. Electr. Grid Futur.:2009. p.1-8.
DOI: <https://doi.org/10.1049/iet-cps.2018.5054>
- [2] Angelos K. Marnerides, Paul Smith, Alberto Schaeffer-Filho, Andreas Mauthe. "Power consumption profiling using energy time-frequency distributions in smart grids". IEEE Commun Lett, 19 (1) (2015), pp. 46-49
DOI : 10.1109 / LCOMM.2014.2371035
- [3] Khan MF, Jain A, Aranuchalam V, Paventhan A. "Communication technologies for smart metering infrastructure". In: Proceedings of the 2014 IEEE students' conf. electr. electron. comput. sci.:2014. p.1-5.
DOI : 10.1109 / SCECS.2014.6804427
- [4] Gert Rietveld, Jean-Pierre Braun, Ricardo Martin, Paul Wright, Wiebke Heins, Nikola Ell, Paul Clarkson, Norbert Zisky. "Measurement infrastructure to support the reliable operation of smart electrical grids". IEEE Trans Instrum Meas, 64 (6) (2015), pp. 1355-1363.
DOI : 10.1109 / TIM.2015.2406056
- [5] O. Palizban, K. Kauhaniemi, J.M. Guerrero. "Microgrids in active network management - Part I: hierarchical control, energy storage, virtual power plants, and market participation", Renew Sustain Energy Rev, 36 (2014), pp. 428-439
DOI : 10.1016 / j.rser.2014.01.016
- [6] Quang Duy La, Yiu Wing Edwin Chan, Boon-Hee Soong. "Power management of intelligent buildings facilitated by smart grid: a market approach". IEEE Trans Smart Grid, 7 (3) (2016), pp. 1389-1400.
DOI : 10.1109 / TSG.2015.2477852
- [7] Neetesh Saxena, Bong Jun Choi, Rongxing Lu. "Authentication and authorization scheme for various user roles and devices in smart grid", IEEE Trans Inf Forensics Secur, 11 (5) (2016), pp. 907-921
DOI : 10.1109 / TIFS.2015.2512525
- [8] Yichi Zhang, Lingfeng Wang, Weiqing Sun. "Trust system design optimization in smart grid network infrastructure". IEEE Trans Smart Grid, 4 (1) (2013), pp. 184-195.
DOI : 10.1109 / TSG.2012.2224390
- [9] Arritt RF, Dugan RC. "Distribution system analysis and the future Smart Grid". In: Proceedings of the 2011 rural electr. power conf., vol. 47 (6): 2011. p. B2-1-B2
- [10] A. Bari, J. Jiang, W. Saad, A. Jaekel. "Challenges in the smart grid applications : an overview". Int J Distrib Sens Netw, 2014 (2014)
DOI: <https://doi.org/10.1155/2014/974682>
- [11] Rusitschka S, Eger K, Gerdes C. "Smart Grid Data Cloud: A Model for Utilizing Cloud Computing in the Smart Grid Domain". In: Proceedings of the 2010 first IEEE int. conf. smart grid commun.: 2010. p. 483-488.
DOI : 10.1109 / SMARTGRID.2010.5622089
- [12] Mets Kevin, Aparicio Ojea Juan, Devellder Chris. "Combining power and communication network simulation for cost-effective smart grid analysis". In: Proceedings of IEEE communication surveys & tutorials. vol. 16(3). Third Quarter: 2014. p. 1771-1796.
DOI : 10.1109 / SURV.2014.021414.00116
- [13] Yong-Gil Kim, Kyung-Il Moon. "Implementation of Secure System for Blockchain-based Smart Meter Aggregation". The Journal of The Institute of Internet, Broadcasting and Communication (IIBC). Vol. 20, No. 2, pp.1-11, Apr. 30, 2020.
DOI: <https://doi.org/10.7236/IIBC.2020.20.2.1>
- [14] Yong-Gil Kim, Kyung-Il Moon. "Clustering of Smart Meter Big Data Based on KNIME Analytic Platform". The Journal of The Institute of Internet, Broadcasting and Communication (IIBC). Vol. 20, No. 2, pp.13-20, Apr. 30, 2020.
DOI: <https://doi.org/10.7236/IIBC.2020.20.2.13>

저 자 소 개

김 용 길(정회원)



- 1990년 호남대학교 전산통계학과 졸업(이학사)
- 1992년 광주대학교 대학원 컴퓨터 공학과 졸업(공학석사)
- 2021년 호남대학교 대학원 컴퓨터 공학과 졸업(공학박사)
- 2014년~현재 조선이공대학교 컴퓨터보안과 부교수

• 관심분야 : 네트워크보안, 통신시스템, 지능시스템