

## 유실 형태별 스마트 미터 데이터 복원 방법

백승준\*, 윤성국\*\*  
 송실대학교 전기공학부

### Interpolation Methods by Loss Types for Missing Smart Meter Data

Sung-Jun Baek\*, Sung-Guk Yoon\*\*  
 Soongsil University

**Abstract** - 현대 전력계통에서 데이터는 필수불가결한 존재이다. 신뢰성이 떨어지는 데이터는 계통 운영 및 관리 측면에서 성능 저하를 야기할 수 있다. 본 논문에서는 통신 불량, 정전 등 여러 요인으로 인한 전력사용량 데이터 누락을 효과적으로 복원하는 방법을 제안한다. 데이터 복원을 위한 다양한 기법들 중 다항식 보간법과 인접일 활용 복원 기법 그리고 새로 제안하는 유사일 기반 인공신경망 활용 복원 기법의 성능을 비교 분석 후 누락 시점 및 기간에 따른 적절한 복원 기법을 제안하였다.

#### 1. 서 론

오늘날 전력계통의 운영 및 관리는 계통 내 수많은 설비들의 데이터를 바탕으로 수행된다. 예를 들어 전력수요예측의 경우 과거 데이터를 바탕으로 다음날 또는 특정기간의 수요를 예측한다. 따라서 안정적이고 효과적인 전력계통의 운영을 위해서는 품질 좋은 데이터의 획득이 중요하다.

스마트그리드의 핵심 기술 중 하나인 AMI(Advanced Metering Infrastructure)를 통해 실시간으로 전력 사용 데이터를 수집할 수 있고 AMI는 전세계 많은 국가에서 보급을 확대하는 추세이다.[1] 미국의 경우 스마트미터 보급률은 약 45%이고, AMI 데이터 유실율은 2.7 ~ 9.4%이다.[2,3] 이렇게 유실되는 데이터는 인접 데이터의 신뢰도를 떨어트려 데이터 셋 전체를 사용하지 못하게 만들고, 그 결과 계통운영 및 관리 능력을 저하시킨다.

데이터가 유실되는 원인은 통신 불량, 서버 정전, 천재지변, 관리자의 부주의 등 수많은 원인이 있기에 유실 없는 완벽한 데이터는 거의 존재하지 않는다. 따라서 유실된 데이터의 복원 기법이 필요하며, 선행 연구들을 통하여 다항식 보간법, 인접일 사용 방법, 기계학습 기법을 사용한 복원 방법 등이 제안되었다.[4]

본 연구에서는 유실된 데이터를 복구하기 위한 인공신경망을 활용한 유사일 기반 복원기법을 제안한다. 제안하는 기법의 성능을 평가하기 위해 2013년 1월 1일부터 2016년 9월 9일까지 취득된 송실대학교 전체 전력사용량 데이터를 활용해 임의로 데이터를 누락시켜 요일별, 누락기간별 케이스를 나눠 실험을 진행하였다. 실험을 통하여 각각의 복원 기법들이 어떤 상황에서 좋은 성능을 보이는지는 확인하였고 제안하는 기법은 높은 어느 상황에서나 무난한 성능을 보이는 것을 확인하였다.

#### 2. 본 론

##### 2.1 데이터 구성 및 전력 사용량 패턴 분석

본 연구에서 사용되는 송실대학교의 전력사용량 데이터는 한국전력공사의 i-smart를 통해 취득되었다. 측정 간격은 1시간 단위로 하루 24시간의 시간별 전력사용량을 계측한다.

송실대학교 캠퍼스의 2015년 10월 4일부터 24일까지의 3주간 일별 전력 사용량 패턴을 그림 1에서 보여준다. 좌측에서부터 일요일부터 토요일을 나타낸다. 그림 1을 통해 평일과 주말의 전력 사용량 패턴은 큰 차이를 보임을 확인할 수 있다.

평일의 전력 사용량 패턴을 보면 대부분 유사한 패턴을 보인다. 일반적으로 수업이 시작하는 8시~9시를 기점으로 전력 사용량이 급격히 상승하고 점차 증가하다가 평균적으로 14시~16시

사이에 최대치를 나타낸다. 대부분의 수업이 끝나는 오후 6시부터는 점차 전력 사용량이 감소하고 새벽에는 전력의 사용량이 거의 변하지 않는다.

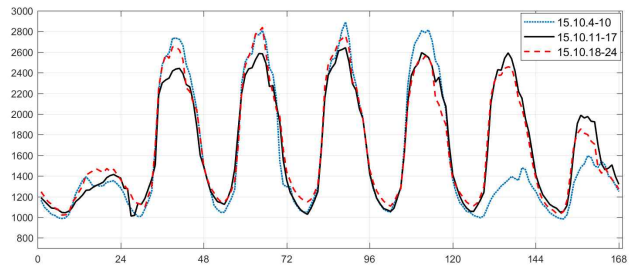


그림 1. 송실대학교 주간 부하 곡선

##### 2.2 데이터 복원 기법

###### 2.2.1 다항식 보간법

보간법이란 측정하지 않았거나, 누락된 데이터를 구해야 할 때 사용하는 기법으로 누락된 데이터를 기준으로 데이터가 존재하는 인접 시간의 데이터를 사용해 누락된 값을 찾아내는 방법이다. 대표적인 방법으로는 선형 보간법, 다항식 보간법이 있다.

본 논문에서는 송실대학교 전력사용량의 패턴에 적합한 Spline 보간법을 사용한다. Spline 함수는 데이터들의 부분집합에 저차 다항식을 적용시켜나가는 연결 다항식을 말한다.[5]

###### 2.2.2 인접일 활용 복원 기법

그림 1에서 보는 것처럼 평일은 대부분 일별 전력사용량 패턴이 유사하다. 주말의 경우 평일과 전력사용량이 크게 차이나지만 지난주 주말과는 유사한 패턴을 갖는다. 인접일 활용 복원 기법은 이러한 부하곡선의 특징을 이용해 값을 복원하는 방법으로 누락된 시간의 i) 전날, ii) 다음날, iii) 전날과 다음날의 평균 값을 사용해 복원을 진행한다. 단 월, 금요일과 주말의 경우 요일특성이 변하기 때문에 이 경우 평일과 주말의 데이터를 분리해 월요일의 경우 지난주 금요일을 전날로 사용하고 토요일의 경우 지난주 일요일을 전날로 사용한다.

$$A_i = A_{d-1,i}$$

$$B_i = B_{d+1,i}$$

$$C_i = \frac{1}{2}(C_{d-1,i} + C_{d+1,i})$$

해당기법은 많은 선행 연구에서 데이터가 누락된 경우 사용하는 방법으로 가장 단순하고 빠른 방법이다.

###### 2.2.3 인공신경망 활용 복원 기법

인공신경망(Artificial Neural Network)은 인간의 두뇌를 모사하여, 다양한 데이터로 기계를 학습시키기 위해 고안된 알고리즘이다. 인체의 신경세포가 자극을 받아 자극의 크기가 역치 이상이 되면 뉴런이 활성화되어 다음 신경세포로 자극이 전달되는 원리를 수학적으로 모델링해 학습을 진행한다. 학습과정은 역전파 알고리즘을 이용하여 최종적으로 계산된 출력층의 값과 학습

벡터의 목표값을 비교, 오차가 크면 가중치를 학습 규칙에 따라 적절한 값에 도달할 때까지 조정해 예측의 정확도를 높인다. 은닉층의 뉴런들의 출력값은 아래의 식(1)에 의해 결정된다.

$$y(x, w) = \Psi \left( \sum_{k=1}^n w_k x_k + b \right) \quad (1)$$

$x$ 는 각 뉴런의 입력,  $w$ 는 각 뉴런 사이의 연결강도,  $b$ 는 바이어스를 나타낸다.  $\Psi$ 는 활성화 함수로 본 논문에서는 ReLU (Rectified Linear Unit) 함수를 사용한다.

식(1)에 의해 계산된 출력값은 역전파 알고리즘을 통해 에러를 최소화하는 방향으로 가중치가 조정된다. 에러( $E(w)$ )는 다음과 같은 수식으로 표현할 수 있다.  $t_i$ 는 참값을 나타낸다.

$$E(w) = \frac{1}{N} \sum_{i=1}^N |y(x_i, w) - t_i|^2 \quad (2-1)$$

역전파 알고리즘은 식(2-2)와 같이 경사하강법으로 오차를 감소시키는 방향으로 가중치를 조정한다.  $\eta$ 는 학습비율이다.

$$w(t+1) = w(t) - \eta \frac{dE}{dw} \quad (2-2)$$

본 연구에서 제안하는 복원 기법은 누락된 데이터가 발생한 시간을 기준으로 누락발생 전, 후 12시간의 데이터를 사용해 유효한 데이터 24개로 데이터 셋을 만들고 식(3)에 의해 전체 데이터 중 해당 데이터 셋과 유사도가 높은 날짜를 찾는다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (s_j - m_j)^2} \quad (3)$$

$n$ 은 유효한 데이터가 존재하는 시간,  $s$ 는 임의의 날짜,  $m$ 은 누락이 발생한 날짜를 나타낸다. RMSE는 평균 제곱근 오차로 누락일과 가장 유사한 날짜가 작은 값을 갖는다. 인공신경망의 학습을 그림 2와 같이 진행된다. 누락된 시간과 동일한 시간의 유사일 데이터를 목표값으로 설정하고 나머지 데이터는 입력데이터로 사용해 유효한 데이터가 입력되었을 때 누락된 시간을 예측할 수 있게 훈련을 진행한다. 최종적으로 누락일의 유효 데이터가 입력되었을 때 누락된 시간의 데이터를 복원한다.

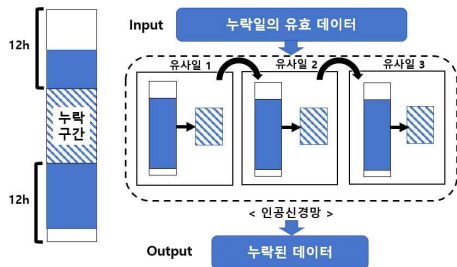


그림 2. 인공신경망 학습과정

### 2.3 모의 실험 결과 및 분석

각 복원 기법의 성능 비교를 위해 요일별, 특정시간별 복원 오차를 측정한다. 복원 성능평가는 식(4)의 MAPE(Mean Absolute Percentage Error)를 사용해 진행한다.

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{A_t - I_t}{A_t} \right| \quad (4)$$

위 수식에서  $A_t$ 는 실제값,  $I_t$ 는 복원값을 뜻한다.

그림 1에서 볼 수 있듯이 부하곡선은 시간에 따라 변동성이 달라진다. 새벽시간은 전력사용량이 작아 값의 변동이 크지 않지만 주간시간의 경우 변동이 크다. 이러한 특징을 반영한 실험을 진행하기 위해 누락시점과 기간을 다양하게 설정한다. 누락기간은 크게 단편 누락과 연속 누락으로 구분하였다. 단편적인 누락은 한 시간의 누락이 여러 개 있는 것이고, 연속적인 누락은 누락시점이 연속해서 발생하는 경우이다. (단편누락 : 1, 3, 5 시간, 연속누락 : 7, 12, 15, 24시간) 누락시점은 새벽, 아침, 점심, 저녁

으로 구분해 위의 누락기간에 해당되는 시간대를 누락시킨다.

요일에 따른 변동성을 확인하기 위해 동일한 케이스를 월, 수, 토에 적용해 실험을 진행했다.

표 1. 요일별 누락기간에 따른 평균 오차. (단위 %)

요일	복원 기법	1시간	3시간	5시간	7시간	12시간	15시간	24시간
월	Spline	0.489	1.494	1.062	6.104	16.24	17.76	18.99
	D-1	6.401	6.542	3.469	5.482	6.312	6.155	5.795
	D+1	8.448	7.61	10.53	8.49	7.708	8.275	8.797
	평균	6.291	7.001	6.001	5.076	5.883	5.992	5.606
	ANN	2.474	2.237	3.031	2.682	2.747	3.745	6.157
수	Spline	1.111	2.484	2.256	8.028	20.03	17.42	24.49
	D-1	3.154	4.459	3.545	3.679	4.081	4.179	3.662
	D+1	2.479	2.912	4.013	2.636	2.683	2.757	2.882
	평균	1.899	2.361	3.122	1.996	1.875	2.082	1.985
	ANN	2.442	1.874	2.638	1.983	4.372	3.732	3.282
토	Spline	2.487	1.306	1.158	5.364	11.95	15.51	12.24
	D-1	16.04	19.2	15.29	16.12	16.23	17.83	15.77
	D+1	15.35	20.48	11.69	13.21	16.13	16.5	12.96
	평균	15.07	19.21	12.03	13.69	15.51	16.45	13.36
	ANN	1.791	1.832	2.822	4.655	3.835	5.433	5.977

표 1은 누락기간에 따른 요일별 복원 오차를 보여준다. 오차는 누락시점을 다르게 적용해 얻은 결과의 평균으로 계산되었다. 실험 결과 단편누락 시 누락 발생 전, 후 데이터를 사용할 수 있는 Spline 보간법이 모든 요일에서 가장 우수한 성능을 보였다. 평일 중 화, 수, 목의 경우 모든 누락기간에서 전날과 다음날의 평균값을 사용할 때 가장 복원 성능이 우수했다. 그림 1에서 볼 수 있듯이 평일의 경우 전력사용량이 유사하기 때문에 다음과 같은 결과를 보였다. 인공신경망을 활용한 복원방법은 주말과 대부분의 경우 안정적인 성능을 보였다.

### 3. 결론

본 논문에서는 스마트미터로 취득한 데이터가 누락될 때 누락시점, 기간 별 적절한 복원기법을 확인하였다. 복원 기법의 성능 비교를 위해 요일별 누락 시점, 구간을 다르게 설정해 실험을 진행했고 최종적으로 단편누락 시 Spline 기법, 평일 화, 수, 목의 경우 평균값 복원, 이외에는 제안하는 유사일 기반 인공신경망활용 복원 기법의 성능이 가장 우수함을 밝혔다.

후속 연구로 대학교 캠퍼스 외 다른 부하들의 경우에도 제한한 결과가 동일한 지 검증하고, 인공신경망의 성능 향상을 위해 유사일을 정확히 찾아낼 수 있는 방법을 연구할 계획이다.

### 감사의 글

본 논문은 한국전력공사의 2018년 착수 에너지 거점대학 클러스터 사업의 지원을 받아 수행된 연구임.

(Grant number. R18XA04)

### 참고 문헌

- [1] 에너지경제연구원, “2015년 세계 스마트 그리드 투자 실적 및 향후 전망”, 세계 에너지시장 인사이트, 제 16-15호, 2016.
- [2] L. Alejandro, C. Blair, L. Bloodgood, M. Khan, M. Lawless, “Global Market for Smart Electricity Meters : Government Policie Driving Strong Growth,” Office of Industries U.S. International Trade Commission, 2014.
- [3] J. Peppanen, M. J. Reno, M. Thakkar, S. Grijalva, R. G. Harley, “Leveraging AMI data for Distribution System Model Calibration and Situational Awareness,” IEEE Trans. Smart Grid, 2015.
- [4] X. Wang, X. Deng, Y. Liu, J. Cao, S. Gao, “A Method for Missing data interpolation by SVR,” IEEE Symposium on Electrical & Electronics Engineering, 2012.
- [5] Hazewinkel, Michiel, “Spline Interpolation,” Encyclopedia of Mathematics, Springer, 2001, ISBN 978-1-55608-010-4.