

데이터 전처리와 퍼지 논리 시스템을 이용한 전력 부하 예측

Electric Load Forecasting using Data Preprocessing and Fuzzy Logic System

방 영 근* · 이 철 희*

(Young-Keun Bang · Chul-Heui Lee)

Abstract - This paper presents a fuzzy logic system with data preprocessing to make the accurate electric power load prediction system. The fuzzy logic system acceptably treats the hidden characteristic of the nonlinear data. The data preprocessing processes the original data to provide more information of its characteristics. Thus the combination of two methods can predict the given data more accurately. The former uses TSK fuzzy logic system to apply the linguistic rule base and the linear regression model while the latter uses the linear interpolation method. Finally, four regional electric power load data in taiwan are used to evaluate the performance of the proposed prediction system.

Key Words : TSK fuzzy logic system, data preprocessing, linear interpolation, electric power load data

1. 서 론

전력 에너지는 산업 성장률이나 인구 규모, 또는 성장 잠재력 등과 같이 한 국가의 경제 규모를 대표할 수 있는 매우 중요한 지표 중 하나이다. 또한 전력은 일반 화석 연료에 비해 안전하고 가격이 저렴하며, 공해로부터 자유로워 그 활용도가 높아지고 있다[1-2].

이러한 전력 에너지 생산의 최근 동향은 화력이나 원자력 보다는 신재생 에너지를 이용하여 생산하려는 추세를 나타내지만, 급성장하는 경제 규모에 따라 소비되는 전력량도 급격하게 증가되고 있으며, 이러한 증가량을 신재생 에너지원으로 충당하기에는 신재생에너지원의 전력 변화 효율이 좋지 않아 제한적이다. 또한, 이상 기온으로 인한 급격한 전력 소모량은 불필요한 전력 설비 투자 및 경제적 손실을 발생시키기도 하며, 전력 분배를 효율적으로 하지 못해 정작 많은 전력을 필요로 하는 곳은 전력 공급이 원활치 못하고 전력 소비가 적은 곳은 전력 예비율을 높여 비효율적인 운영 체제가 발생하기도 한다.

따라서 이러한 문제를 해결하는 기본적인 방법은 전력의 사용량을 정확히 예측하고 이를 토대로 원활한 분산 계획과 설비의 운영을 조합하는 것이 바람직할 것이다[3].

보통 전력부하를 예측 하는 방법으로는 수학적 모델을 이용하는 방법(AR: Auto-regression model, MA: Moving average model, ARMA: Auto-regression moving average model,

ARIMA: Box-Jenkins model, etc)이나 또는 지능형 모델(ANN: Artificial neural network, FLS: Fuzzy logic system, SVM: Support vector machines, etc)을 이용하는 방법들이 많이 사용된다[4-7].

통계에 기반한 수학적 모델은 전력예측을 위한 다양한 변수들을 효과적으로 반영하기 위해선 수학적 모델 구조가 복잡해지거나 변수들을 분석하기 위한 많은 연산량을 요구한다. 반면, 지능형 모델들은 원형의 데이터로부터 데이터 속에 내재된 다양한 변수들의 특성들을 학습을 통해 분석하게 되므로 통계에 기반한 수학적 모델 보단 쉽게 분석할 수 있으며, 최근의 동향에서는 이러한 지능형 시스템을 이용한 예측 모델이 빈번하게 연구되고 있는 중이다. 하지만 이런 지능형 시스템들도 시스템의 학습을 정확히 하지 않을 경우 그 성능이 매우 저조해지는 단점으로 인해 시스템 학습을 위한 많은 양의 학습 데이터를 요구하거나 비감독 학습에서는 많은 시행을 통해 모델이 구축되는 단점이 야기되기도 한다 [8-9].

이러한 지능형 시스템 중에서도 FLS는 언어적 규칙기반과 데이터의 소속 정도에 기반한 추론을 통해 결과를 얻을 수 있어, 구조 설명이 어려운 ANN보다는 시스템의 설명이 용이하며, 또한 TSKFLS(TSK Fuzzy logic system)는 단순 회귀 모델과 같은 수학적 모델에 따른 연산의 편리함과 비선형적 데이터의 특성을 언어적 퍼지 규칙기반을 통해 취급할 수 있어 예측 시스템의 구현에 매우 효과적인 특징을 나타 낸다[10]. 하지만 실제 예측시스템의 성능은 기본적으로 주어진 데이터의 구조에 민감할 수밖에 없으며, 데이터의 구조가 충분한 정보를 포함하거나 시스템 학습에 용이한 구조를 가진다면 이는 분명 시스템의 성능에 긍정적 효과를 미칠 수 있다. 논문 [11-12]에서 원형의 데이터를 분석하여 차분 계열의 데이터를 이용하거나 이동평균 증가율을 이용함으로써 효과적인 시스템을 설계하고 좋은 예측 성능을 얻었다.

† Corresponding Author : Dept. of Electrical and Electronic Engineering, Kangwon National University, Korea.
E-mail: chlee@kangwon.ac.kr

* Dept. of Electrical Engineering, Kangwon National University, Korea.

Received : July 25, 2017; Accepted : November 7, 2017

하지만 두 방법 모두 데이터의 길이가 짧아지는 단점으로 데이터 양이 작을 경우에는 적용에 문제점이 존재한다.

본 논문에서는 주어진 데이터의 길이가 시스템의 학습을 위해 충분하지 못한 경우에도 전력 데이터의 예측을 효과적으로 수행할 수 있는 예측 시스템의 설계 법을 다룬다. 물론 시스템의 설계를 위한 기본 틀은 앞서 설명된 TSKFLS를 이용함으로써 단순 수학적 모델에 따른 연산의 이점과 퍼지 규칙기반을 이용한 비선형 데이터 구조해석의 이점을 동시에 취할 수 있도록 하였으며, 원형의 데이터를 전처리(Preprocessing)하는 과정에서는 데이터의 분석을 통해 정보를 포함하고 있는 보간 데이터를 생성하여 사용함으로써 시스템 설계에 사용될 학습 데이터양을 증가시키고, 생성된 보간 데이터들의 정보를 학습하여 보다 정확한 예측이 수행될 수 있는 구조의 시스템 설계를 다룬다.

마지막으로 시뮬레이션에서는 비교적 데이터의 길이가 짧은 대만의 4개 지역 전력부하 데이터를 이용하여 제안된 시스템의 예측 성능을 검증하고, 다른 시스템들과의 비교를 통해 제안된 설계 방법의 효용성을 기술한다.

2. 제안된 시스템의 개념

2.1 시스템 구현 절차

그림 1은 제안된 시스템의 전체 구조로서 제안된 시스템은 크게 3개의 메인 설계 절차에 따라 전력 부하 데이터를 예측하게 된다. 먼저 그림에서 순서 ①은 제안된 퍼지 예측 시스템의 기본 구조를 만드는 과정으로 크게 원형 데이터의 전처리, 퍼지 규칙 및 파라미터 식별을 통해 퍼지 추론엔진을 만드는 과정이다. 여기에서 데이터 전처리 과정에는 데이터의 추세에 따른 선형 보간 데이터를 삽입하게 되며, 보간 데이터들은 구간 평균값들을 사용한다. 두 번째 퍼지 군집화 과정에서는 구조가 간단하면서도 효과적으로 사용가능한 K-평균 군집화 기법을 이용하여 퍼지 분할을 수행하고 이를 통해 퍼지 규칙들을 생성하게 된다 [12-13]. 마지막으로 규칙의 파라미터 식별 과정에서는 빈번하게 사용되는 최소 자승법을 통해 각 규칙의 파라미터를 식별하게 된다.

과정 ②는 과정 ①을 통해 생성된 시스템을 이용해 실제 예측에 사용되는 보간 데이터를 생성하는 절차이다. 과정 ①에서 사

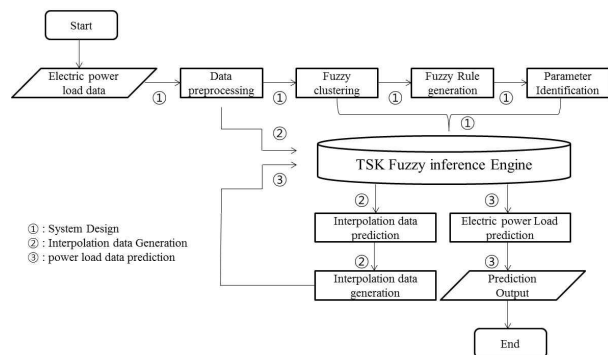


그림 1 제안된 퍼지 시스템의 전체 순서도

Fig. 1 The flow chart of the proposed fuzzy system

용된 전처리 데이터는 데이터 사이의 평균값들을 사용하여 시스템을 학습하기 위한 것이지만, 실제로 예측 과정에서는 예측 값은 미래의 값이므로 그 사이의 보간 값들은 알 수 없는 미지량으로 주어진다.

따라서 실제 예측에 사용되는 보간 데이터들은 예측을 통해 그 값들을 추정해야 하며, 과정 ②는 실제 예측에 사용될 보간 데이터를 예측하는 과정이다. 마지막으로 과정 ③은 예측된 보간 데이터를 이용하여 전력 부하 데이터를 예측하는 과정이다.

비록 제안된 시스템이 1차 예측을 통해 보간 데이터를 생성하고 이를 이용하여 실제 전력 부하데이터를 예측하는 구조를 띠는 2차 예측 과정을 거치지만, 과정 ①에서 관측된 과거 자료의 평균값들을 이용하여 시스템을 설계 하였고, 따라서 예측되는 보간 데이터들 또한 이러한 평균값들과 유사한 특성을 띠는 형태로 예측 될 것이므로 반복적인 시스템 학습 없이 초기 설계된 시스템을 그대로 사용함으로써 예측 구조의 복잡성과 시간을 단축할 수 있도록 하였다.

2.2 데이터의 전처리

전력부하 데이터의 주어진 양이 충분하지 못하면 시스템 설계 시 데이터에 내재된 비선형적 특성을 충분히 고려할 수 없으므로, 추정된 파라미터의 부정확성이 증가하게 되고 결국 시스템의 예측 결과가 좋지 않게 된다. 따라서 이러한 문제를 극복하기 위해선 정보를 포함한 데이터를 증가시키는 방법이 가장 단순하고 효율적일 것이다. 일반적으로 전력 부하 변동을 살펴보면, 증가 모양이 지수적이거나, 포물선 형태로 증가 할 수는 있겠지만, 급등락을 반복하는 형태로 변화되지는 않을 것이다. 따라서 지수적이거나 포물선 형태의 증가나 감소를 보인다면 이러한 구조는 선형 근사화 할 수 있을 것이다. 선형 근사화 과정에 삽입되는 보간 데이터의 양은 데이터 상호간의 상관성에 영향을 미칠 수도 있지만 시스템의 구조, 즉, 연산량이나 규칙의 구조에도 영향을 미칠 수 있다.

따라서 시점에 따른 데이터의 유사성을 분석하고 이들의 관계를 찾을 수 있다면 보간 데이터의 개수를 결정하는데 보다 효과적일 수 있다. 시점에 따른 데이터의 상관성 분석에는 일반적으로 식 (1)과 같은 자기 상관 함수(ACF: Auto-Correlation Function)를 사용할 수 있으며 시간 t_1, t_2, \dots, t_N 에서 관측된 데이터가 x_1, x_2, \dots, x_N 이라면 자기상관함수는 다음처럼 구하게 된다.

$$r_k = \frac{C_k}{C_0} = \frac{\frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{X})(x_{t+k} - \bar{X})}{\frac{1}{N} \sum_{t=1}^N (x_t - \bar{X})^2}, \quad k = 0, 1, \dots, N-1 \quad (1)$$

여기서 분자 C_k 는 자기 공분산 함수(Autocovariance function)이며, 분모 C_0 는 분산(Variance function)을 나타낸다. 또한 x_t 는 t 시점에서의 데이터를 의미하고, \bar{X} 는 모든 데이터의 평균을 의미한다.

식 (1)의 자기 상관함수는 시차에 따라 -1에서 1사이의 값으로 나타나며, 그 모양이 선형적 변화를 나타냄과 동시에 1에 근

접하면 강 상관성, 변화가 지수적 감쇠를 나타내면 약상관성, 시차 0을 제외하고 모든 구간에서 매우 낮은 값으로 나타나면 잡음성 계열로 볼 수 있다. 따라서 삽입된 보간데이터에 따라 자기상관함수의 변화가 강상관성을 가지게 되면, 서로 이웃하는 데이터들 사이의 높은 상관성을 의미하므로 예측의 성능을 높일 수 있을 것이다.

2.3 TSK 퍼지 모델

T. Takagi, M. Sugeno 그리고 K. T. Kang에 의해 1984년에 소개된 TSK 퍼지 모델의 j 번째 퍼지 규칙은 일반적으로 식 (2)와 같이 주어진다 [10].

$$R^j : IF x_1 \text{ is } F_1^j \text{ and } x_2 \text{ is } F_2^j \text{ and } \dots \text{ and } x_p \text{ is } F_p^j \quad (2)$$

$$THEN y^j(x) = c_0^j + c_1^j x_1 + c_2^j x_2 + \dots + c_p^j x_p$$

여기서 R^j 는 j 번째 퍼지 규칙을 의미하고, x 는 그 규칙에 해당하는 입력 값들이며, F 는 입력들을 포함하는 퍼지 집합을 의미한다. 또한 y^j 는 j 번째 퍼지 규칙의 국부 출력이며, c 는 국부 출력을 얻기 위한 모수들을 의미한다.

식(2)에서 보이듯 TSK 퍼지 모델은 퍼지 규칙을 형성하는 전건부(Antecedent part)와 상응하는 규칙의 국부 출력인 후건부(Consequent part)로 구성되어 있다. 전건부는 유사한 입력들을 동일한 집합으로 분류하는 과정과 각 입력에 대하여 상응하는 퍼지 집합에 소속된 정도를 표현하는 퍼지 소속함수(Fuzzy membership function)로 구성되며, 후건부는 규칙의 국부 출력을 산출하는 선형 회귀모형으로 주어지게 된다.

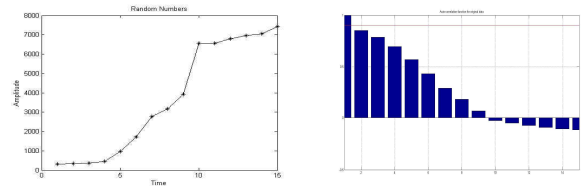
3. 제안된 시스템의 설계

3.1 전처리를 위한 보간데이터의 분석

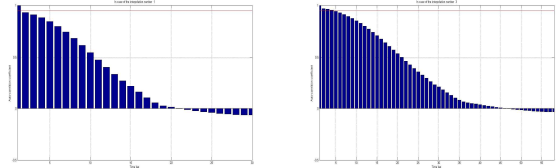
보간데이터의 양은 시스템의 성능이나 능력(학습량과 성능과의 관계)에 많은 영향을 미친다. 보간데이터의 삽입양이 많아지면 정보를 담고 있는 데이터의 양도 증가하고, 이를 통해 시스템의 성능을 개선할 순 있겠지만, 데이터를 무한정 증가시킨다고 해서 시스템의 성능이 이와 비례적으로 좋아진다고는 할 수는 없을 것이며, 무한정 증가한 데이터의 양은 오히려 시스템의 연산양만 가중시키는 결과를 초래할 수 있다. 따라서 적절한 개수의 보간데이터는 시스템의 성능이나 설계의 효율성을 높일 수 있을 것이다.

본 논문에서는 실 예측에 사용된 전력부하 데이터처럼 데이터량이 짧은 랜덤 데이터를 생성하고 이를 분석하였으며 그림 2는 생성된 랜덤 계열의 모습과 보간데이터를 삽입하였을 경우 자기상관함수의 변화를 보여준다.

그림 2의 a)는 랜덤 전력부하 데이터를 보여주며, 그림 b)는 랜덤 계열에 대한 자기상관함수를 보여주고 있으며 시차 0에서 1로 넘어 갈 경우 데이터의 상관성이 확연히 떨어짐을 알 수 있다. 그림 c)는 원형의 랜덤 시계열의 데이터 사이에 그들의 평균



a) 랜덤 시계열 모양 b) 원형 시계열의 자기 상관함수
a) The shape of random time series b) ACF of the original time series



c) 1개의 평균값 보간 d) 3개의 평균값 보간
c) ACF by interpolating one average data d) ACF by interpolating three average data

그림 2 랜덤 시계열과 보간 데이터에 따른 자기상관함수

Fig. 2 Random time series and ACF corresponding to the number of interpolating data

에 해당하는 하나의 보간데이터를 삽입한 후 자기상관함수를 분석한 결과로 그림 b)에 비하여 시차에 따른 상관성이 높아졌으며, 데이터의 양도 증가하였음을 보여주고 있지만, 인접하는 데이터 사이의 상관성이 강 상관성을 만족할 수준은 아니다. 마지막으로 그림 d)는 3개의 보간데이터를 삽입한 경우의 자기상관함수 모양으로 시차 3에 해당하는 자기상관함수의 값이 강상관성에 포함되며, 이는 인접하는 데이터들 간의 상관성이 높은 것으로 간주할 수 있다. 따라서 인접하는 데이터들을 이용하여 다음 단계의 미래 값을 예측하는 데이터 구조로 적합 할 것으로 판단할 수 있다. 물론 보간 데이터의 양을 더 증가하면 인접하는 데이터들 사이의 상관성이 높아 질 수 있지만 분석 결과 두드러지게 변화되지는 않을 뿐만 아니라 그림에서 알 수 있듯 데이터의 양이 상당히 증가하게 되므로 시스템의 학습에 부담을 증가시키게 된다.

따라서 시스템 구현의 복잡성이나 데이터의 상호간의 연관성을 살펴 볼 때 적절한 보간 데이터의 개수는 3개로 선정 하였으며, 원형의 데이터로부터 보간데이터의 생성은 t 시점의 원형 데이터 x_t 와 $t+1$ 번째 시점의 원형 데이터 x_{t+1} 의 사이에 다음과 같이 정의된다.

$$m_t(p) = \frac{p}{4}(x_t + x_{t+1}), \quad p = 1, 2, 3 \quad (3)$$

여기서 $m_t(p)$ 는 t 시점에서의 원형 데이터와 $t+1$ 번째 시점에서의 원형 데이터 사이에 삽입되는 p 번째 보간 데이터를 의미한다.

따라서 제안된 시스템의 데이터들은 다음과 같은 구조를 나타

내게 된다.

$$\begin{matrix}
 x_1, & m_1(1), & m_1(2), & m_1(3), \\
 x_2, & m_2(1), & m_2(2), & m_2(3), \\
 \vdots & \vdots & \vdots & \vdots \\
 x_i, & m_i(1), & m_i(2), & m_i(3), \\
 \vdots & \vdots & \vdots & \vdots \\
 x_{N-1}, & m_{N-1}(1), & m_{N-1}(2), & m_{N-1}(3), & x_N
 \end{matrix} \quad (4)$$

3.2 퍼지 균집화 및 퍼지 규칙 생성

식 (2)에서처럼 TSK 퍼지 논리 시스템은 각 퍼지 규칙의 전건부 언어적 규칙기반과 후건부의 파라미터를 이용한 국부 출력 함수로 주어진다. 따라서 먼저, 전건부의 언어적 규칙기반을 표현하기 위한 입력공간의 퍼지 분할이 요구된다. 본 논문에서는 최소 규칙기반을 이용하여 효과적인 추론을 유도하기 위해 3입력을 기반으로 각 입력 차원별 2개의 퍼지 집합을 이용하여 퍼지 분할을 유도하였다. 이러한 분할 과정은 기존의 논문에서도 충분히 그 성능이 입증되어 있다[13]. 3입력 데이터 공간에 대하여 각 2개의 퍼지 집합의 생성은 구조가 단순하면서도 효과적인 K-평균 균집화 기법을 적용하였다. 먼저 k 차원 입력에 대하여 임의의 c 번째 퍼지집합의 중심 F_c^k 가 주어진다면, 퍼지집합의 중심의 이동 과정은 다음과 같이 정의된다.

$$F_c^k(i+1) = \frac{1}{n_c^k} \sum_{t=1}^{n_c^k} m_c^k(t) \quad (5)$$

여기서 n_c^k 는 k 차원 입력 공간에서 c 번째 퍼지집합에 포함되는 데이터의 개수를 의미하며, $m_c^k(t)$ 는 이때의 각 데이터들을 의미한다.

이는 결국 각 차원 입력에 대하여 퍼지집합의 중심을 임의로 지정한 후, 퍼지집합의 중심과 각 데이터의 유클리드 거리(Euclidean distance)를 계산하여 가까운 퍼지집합으로 각 데이터를 분류하고 이들의 평균을 이용하여 퍼지집합의 중심을 갱신하는 것이다. 여기서 본 논문은 3차원 입력 공간과 각 2개의 퍼지분할을 수행하므로 $k=1,2,3$ 으로 정의되고 $c=1,2$ 로 정의된다. 이러한 퍼지집합의 중심은 다음의 종료 조건을 만족할 때까지 반복 수행된다.

$$\left| \frac{F_c^k(i+1) - F_c^k(i)}{F_c^k(i)} \right| < 10^{-4} \quad (6)$$

위와 같이 각 입력 차원별 두 개의 퍼지집합의 중심이 정의되면 식 (2)의 TSK 퍼지규칙의 j 번째 일반식은 본 논문에서 다음과 같이 수정된다.

$$R^j: \text{IF } D_1 \text{ is } F_{c=1,2}^{1,j} \text{ and } D_2 \text{ is } F_{c=1,2}^{2,j} \text{ and } D_3 \text{ is } F_{c=1,2}^{3,j} \quad (7)$$

$$\text{THEN } y^j(D) = \alpha_0^j + \alpha_1^j D_1 + \alpha_2^j D_2 + \alpha_3^j D_3$$

여기서 $D_1 = [x_1, m_1(1), \dots, m_{N-1}(1)]$, $D_2 = [m_1(1), m_1(2), \dots, m_{N-1}(2)]$, $D_3 = [m_1(2), m_1(3), \dots, m_{N-1}(3)]$ 으로 정의되고 D_k 는 이 입력들 중 j 번째 규칙을 만족하는 데이터를 의미하며, $F_{c=1,2}^{k,j}$ 는 이들이 j 번째 규칙에서 두 개의 퍼지집합 $F_{c=1,2}$ 중 어느 하나를 만족하였을 때를 의미한다.

다음으로 입력 데이터가 각각의 퍼지집합에 소속되는 정도인 멤버십함수(Membership function)는 가우시안 소속함수(Gaussian membership function)를 이용하여 다음과 같이 정의된다.

$$\text{IF } D_k(i) \leq F_1^k \text{ or } D_k(i) \geq F_2^k$$

$$\mu_L\{D_k(i)\} \text{ or } \mu_R\{D_k(i)\} = 1$$

else

$$\mu_L\{D_k(i)\} = \exp\left(-\frac{1}{2} \frac{D_k(i) - F_1^k}{\{F_2^k - \min(D_k)\}^2}\right)$$

$$\text{or } \exp\left(-\frac{1}{2} \frac{D_k(i) - F_1^k}{\{\max(D_k) - F_1^k\}^2}\right)$$

$$\mu_R\{D_k(i)\} = \exp\left(-\frac{1}{2} \frac{D_k(i) - F_2^k}{\{F_2^k - \min(D_k)\}^2}\right) \quad (8)$$

$$\text{or } \exp\left(-\frac{1}{2} \frac{D_k(i) - F_2^k}{\{\max(D_k) - F_1^k\}^2}\right)$$

여기서 μ_L 은 데이터 $D_k(i)$ 가 첫 번째 퍼지집합을 만족하는 소속 정도를 의미하고 μ_R 은 같은 데이터가 두 번째 퍼지집합을 만족하는 소속 정도를 의미한다.

이렇게 정의된 3차원 입력 데이터 쌍은 각 차원별 2개의 퍼지 집합을 사용하므로 총 8개의 퍼지 규칙들 중 전부 또는 일부를 만족하게 될 것이다.

3.3 최소자승법에 의한 파라미터 식별

식 (7)의 후건부는 차원별 입력데이터 D_k 와 국부출력 $y^j(D)$ 및 추정이 필요한 파라미터 $A^j = [a_0^j, a_1^j, a_2^j, a_3^j]$ 로 구성되어 있다. 또한 국부출력은 $Y = [m_1(3), x_2, \dots, x_N]$ 으로 정의될 수 있으며 따라서 입력데이터쌍이 $[D_1(i), D_2(i), D_3(i)]$ 로 주어지면 이때 상응하는 국부 출력 값은 $Y(i)$ 가 된다.

하나의 퍼지규칙은 다수의 입력데이터쌍이 만족할 수 있으므로 j 번째 퍼지규칙을 만족하는 입력데이터의 개수가 q 개라면 식 (7)의 후건부는 다음과 같은 연립방정식으로 표현될 수 있다.

$$y^j(1) = a_0^j + a_1^j D_1^j(1) + a_2^j D_2^j(1) + a_3^j D_3^j(1)$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$

$$y^j(i) = a_0^j + a_1^j D_1^j(i) + a_2^j D_2^j(i) + a_3^j D_3^j(i) \quad (9)$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$

$$y^j(q) = a_0^j + a_1^j D_1^j(q) + a_2^j D_2^j(q) + a_3^j D_3^j(q)$$

여기서 $D_k^j(i)$ 는 j 번째 퍼지규칙을 만족하는 q 개의 입력데이터

집합 중 i 번째 값을 의미하며, $y^j(i)$ 는 이때에 상응하는 국부 출력 값을 의미한다.

식 (9)는 최소자승법을 적용하기 위해 다음과 같이 행렬식과 벡터의 형태로 표현할 수 있다.

$$\begin{bmatrix} y^j(1) \\ \vdots \\ y^j(i) \\ \vdots \\ y^j(q) \end{bmatrix} = \begin{bmatrix} D_1^j(1) & D_2^j(1) & D_3^j(1) \\ \vdots & \vdots & \vdots \\ D_1^j(i) & D_2^j(i) & D_3^j(i) \\ \vdots & \vdots & \vdots \\ D_1^j(q) & D_2^j(q) & D_3^j(q) \end{bmatrix} \begin{bmatrix} a_0^j \\ a_1^j \\ a_2^j \\ a_3^j \end{bmatrix} \quad (10)$$

$$Y^j = D^j A^j \quad (11)$$

따라서 파라미터 A^j 는 최소자승법을 이용하여 다음처럼 추정할 수 있게 된다.

$$\hat{A}^j = [\{D^j\}^T D^j]^{-1} D^j Y^j \quad (12)$$

식 (12)의 최소자승법으로 추정된 파라미터 \hat{A}^j 는 다음과 같이 j 번째 규칙에 대한 오차 자승을 최소화하는 방향으로 탐색되어진다.

$$E^j = (Y^j - D^j \hat{A}^j)^T (Y^j - D^j \hat{A}^j) \quad (13)$$

3.4 시스템의 출력

제안된 시스템의 출력을 얻기 위해선 먼저 보간데이터를 예측하여야만 한다. 즉, $t+1$ 시점에서의 데이터 x_{t+1} 을 예측하기 위해선 식 (4)의 데이터 구조와 식 (7)의 3입력 퍼지 규칙기반을 사용하기 때문에 시스템의 구조상 $m_t(1)$, $m_t(2)$, $m_t(3)$ 데이터를 요구한다. 하지만 이 보간 데이터들은 x_t 와 x_{t+1} 데이터가 주어졌을 경우에 구해진 값이며, 실제로 x_{t+1} 의 값은 예측을 해야 되는 미지량이므로 관측된 자료와 학습을 통해 보간 데이터 또한 예측하여 사용하여야 한다. 만약 보간 데이터 예측을 위해 사용된 입력 데이터 쌍이 M_j 개의 퍼지 규칙을 만족하였다면, 이러한 보간 데이터의 예측 값은 다음과 같이 구해진다.

$$\hat{m}_t(1) = \frac{\sum_{j=1}^{M_1} f_1^j (\hat{a}_0^j + \hat{a}_1^j m_{t-1}(2) + \hat{a}_2^j m_{t-1}(3) + \hat{a}_3^j x_t)}{\sum_{j=1}^{M_1} f_1^j} \quad (14)$$

$$\hat{m}_t(2) = \frac{\sum_{j=1}^{M_2} f_2^j (\hat{a}_0^j + \hat{a}_1^j m_{t-1}(3) + \hat{a}_2^j x_t + \hat{a}_3^j m_t(1))}{\sum_{j=1}^{M_2} f_2^j} \quad (15)$$

$$\hat{m}_t(3) = \frac{\sum_{j=1}^{M_3} f_3^j (\hat{a}_0^j + \hat{a}_1^j x_t + \hat{a}_2^j m_t(1) + \hat{a}_3^j m_t(2))}{\sum_{j=1}^{M_3} f_3^j} \quad (16)$$

식 (14)~(16)은 관측된 데이터로부터 보간데이터를 순차적으로 예측하여 가는 것을 의미하며, 여기서 f^j 는 j 번째 규칙의 점화 강도로(Firing Strength) 다음과 같이 정의된다.

$$\begin{aligned} f_1^j &= \min[\mu^j\{m_{t-1}(2)\}, \mu^j\{m_{t-1}(3)\}, \mu^j\{x_t\}] \\ f_2^j &= \min[\mu^j\{m_{t-1}(3)\}, \mu^j\{x_t\}, \mu^j\{m_t(1)\}] \\ f_3^j &= \min[\mu^j\{x_t\}, \mu^j\{m_t(1)\}, \mu^j\{m_t(2)\}] \end{aligned} \quad (17)$$

식 (14)~(17)을 이용하여 보간데이터에 대한 예측 생성이 수행되면, 이 예측 값들은 $t+1$ 시점에서 실제 예측을 하고자하는 \hat{x}_{t+1} 을 위한 입력으로 사용된다. 앞선 경우와 마찬가지로 실제 예측을 위해 사용되는 1차 예측된 보간데이터들 또한 여러 개의 퍼지 규칙을 만족할 수 있으며, 이들이 만족한 퍼지 규칙의 수가 M_2 개일 때의 출력과 점화강도는 다음과 같다.

$$\hat{x}_{t+1} = \frac{\sum_{j=1}^{M_2} f_t^j (\hat{a}_0^j + \hat{a}_1^j \hat{m}_t(1) + \hat{a}_2^j \hat{m}_t(2) + \hat{a}_3^j \hat{m}_t(3))}{\sum_{j=1}^{M_2} f_t^j} \quad (18)$$

$$f_3^j = \min[\mu^j\{\hat{m}_t(1)\}, \mu^j\{\hat{m}_t(2)\}, \mu^j\{\hat{m}_t(3)\}] \quad (19)$$

식 (19)는 보간 입력 데이터들이 각 규칙을 점화시키는 점화 강도를 의미하며, 이후의 모든 데이터들의 예측 또한 시점 t 를 이동하면서 구할 수 있게 된다.

4. 시뮬레이션

본 논문에서는 제안된 시스템의 성능 검증을 위해 1981년부터 2000년까지 수집된 대만의 4개 지역 전력 피크부하 데이터를 사용 하였으며, 각 지역별 20개의 데이터들 중에서 다른 논문들에서와 같이 16개의 학습 데이터와 4개의 예측 데이터로 적용하였다[2].

그림 3과 4는 각각 대만 서부지역의 원형과 보간된 전력부하 데이터를 보여준다. 그림 4를 살펴보면 학습할 수 있는 데이터의 양이 증가 되어 있으며, 보간데이터의 구조는 선형 보간으로 이루어져 있음을 알 수 있다. 또한 각 원형의 전력부하데이터 사이에 본 논문의 조건에 따라 3개의 보간 데이터들이 삽입되었음을 알 수 있다.

그림 5는 3입력 차원에 따른 K-평균 군집의 변화를 보여주며, 군집의 중심이 초기의 위치에서 반복을 통해 이동됨을 알 수 있다. 이러한 동조 과정은 결국 규칙의 점화 강도를 제어하는 요소로 작용한다.

그림 6은 보간 데이터가 삽입된 전력부하에 대한 예측 결과를 보여준다.

그림 6에서 빨간색에 *마크의 선이 실제 데이터의 모양, 파란색의 o마크의 선이 시스템의 예측 결과를 나타낸다. 결과를 살펴보면 제안된 시스템이 전 구간에서 원형의 데이터를 비교적 정확하게 예측하였음을 알 수 있다.

그림 7은 데이터 추출과정을 통해 서부 지역 원형의 데이터에 상응하는 제안된 시스템의 예측 결과를 보여주는 것으로, 제안된 시스템이 학습 데이터에 대하여 거의 정확하게 예측을 수행하였

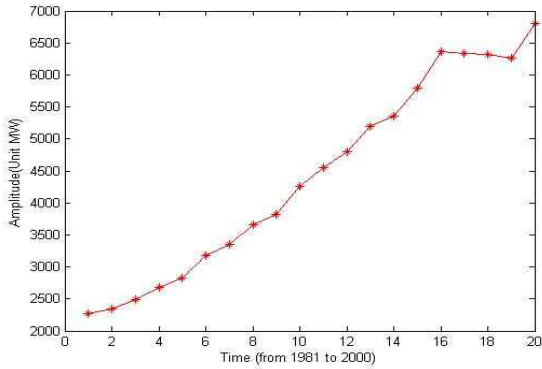


그림 3 대만 서부지역 전력부하 데이터 (1981~2000년)
Fig. 3 Electric power load data of the southern region in Taiwan(from 1981 to 2000)

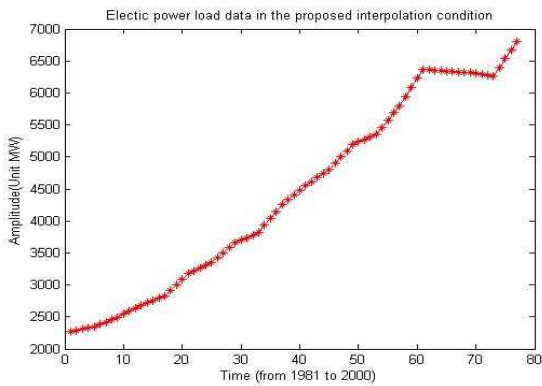


그림 4 제안된 보간 조건에 따른 전력부하 데이터
Fig. 4 Electric power load data in the proposed interpolation condition

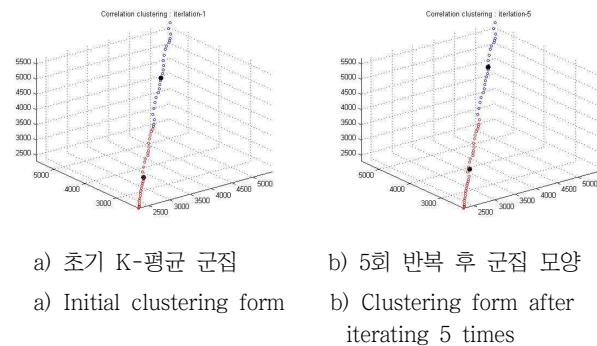


그림 5 제안된 조건에 따른 3입력차원 K-평균 군집화
Fig. 5 K-means clustering algorithm in the proposed condition

으며, 예측 데이터에 대해서도 좋은 예측을 수행하였음을 알 수 있다. 그림 8에서 10까지는 각각 동부, 북부 및 중부지역의 전력 부하 데이터에 대한 제안된 시스템의 예측 결과를 보여준다.

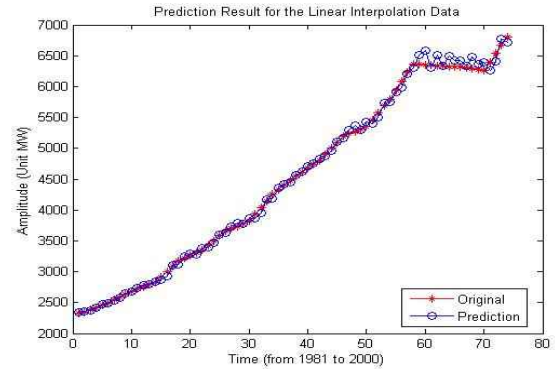


그림 6 보간데이터가 삽입된 전력부하 데이터의 예측 결과
Fig. 6 Prediction result of the electric power load data with the three interpolation data

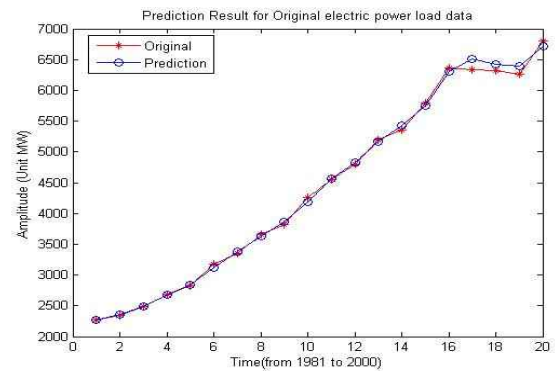


그림 7 제안된 시스템의 예측 결과(서부 지역)
Fig. 7 Prediction result of the proposed system (Southern region)

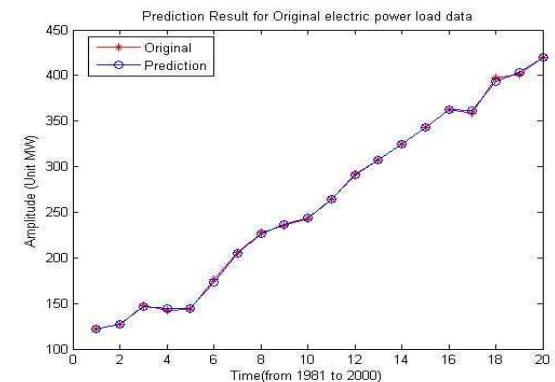


그림 8 제안된 시스템의 예측 결과(동부 지역)
Fig. 8 Prediction result of the proposed system (Eastern region)

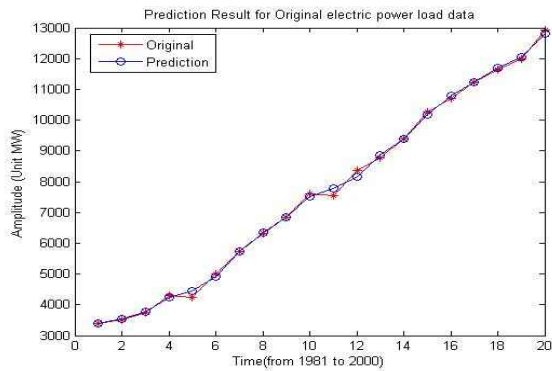


그림 9 제안된 시스템의 예측 결과(북부 지역)
 Fig. 9 Prediction result of the proposed system (Northern region)

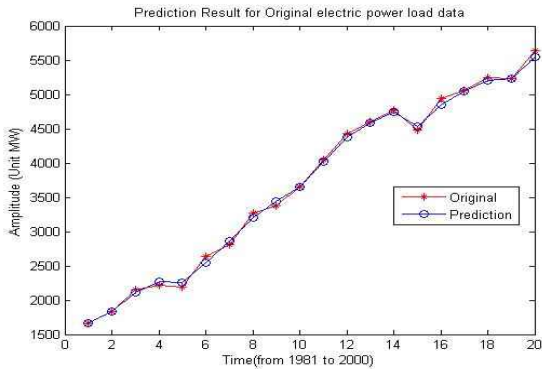


그림 10 제안된 시스템의 예측 결과(중부 지역)
 Fig. 10 Prediction result of the proposed system (Central region)

그림 8에서 10을 살펴보아도 제안된 예측 시스템이 지역별 전력부하 데이터들에 대하여 모든 구간에서 비교적 정확하게 예측을 수행하였음을 알 수 있다.

표 1은 제안된 시스템의 예측 성능을 보다 정확히 분석하기 위한 것으로 다른 시스템들과의 예측 성능을 비교한 것이다 [2,12].

표 1 제안된 시스템과 다른 시스템들과의 성능 비교
 (지표 : MAPE(%))

Table 1 Performance comparison between the proposed system and other systems(index: MAPE(%))

| System | Region | | | | Average |
|------------|----------|---------|----------|---------|---------|
| | Southern | Eastern | Northern | Central | |
| Regression | 8.29 | 4.10 | 2.42 | 8.52 | 5.83 |
| ANN | 2.48 | 3.62 | 1.06 | 1.73 | 2.22 |
| IRARMA | 1.62 | 0.77 | 1.47 | 1.62 | 1.37 |
| Proposed | 1.81 | 0.51 | 0.54 | 0.87 | 0.93 |

표 1의 예측 결과를 살펴보면 회귀 시스템(Regression)이나 인공 신경망 시스템(ANN)의 경우 지역에 따른 예측 편차가 비교적 크게 나타남을 알 수 있으며, IRARMA시스템의 경우는 지역별 예측 편차가 비교적 작으면서도 예측 성능이 우수함을 보여준다. 제안된 예측 시스템의 경우 서부 지역의 예측 성능을 제외하곤 지역별 예측 편차도 작을 뿐만 아니라 예측의 정확도가 매우 높음을 알 수 있다. 또한 서부 지역의 경우에도 IRARMA 시스템에 비해 조금 낮은 예측 성능을 나타내었지만 전 지역에 대한 평균 예측 성능을 비교해 보면, 제안된 예측 시스템의 예측 성능이 보다 평균적으로 더 우수함을 알 수 있다.

5. 결 론

본 논문에서는 선형 보간법을 통한 시스템 학습과 이를 통해 미지량으로 주어지는 보간데이터를 1차 예측하고, 예측된 보간 데이터를 이용하여 원형의 전련부하 데이터를 예측하는 시스템의 설계법을 다루었다. 표 1의 결과를 살펴보면 제안된 시스템의 예측 성능이 대만의 서부, 동부, 북부, 중부 지역의 전력부하의 변화를 모든 지역에서 비교적 정확하게 예측하였음을 알 수 있다. 또한 4개 지역의 예측 성능의 평균을 살펴보면 제안된 시스템의 예측 평균이 0.93%의 MAPE로 다른 시스템들에 비해 매우 우수하였음을 알 수 있다. 이러한 결과는 제안된 시스템의 설계 기법들이 비교적 정보량이 적은 데이터의 예측 분야에서 매우 효과적일 수 있음을 보여준다.

마지막으로 보간 데이터를 이용하여 학습 정보를 보다 정확히 분석할 수는 있었지만 주어진 데이터들이 지속적으로 증가하는 추세를 보이므로 이러한 데이터들에 대한 정규화를 통해 규칙기반의 편중현상을 완화할 수 있는 방법이 연구되어야 할 것으로 생각된다.

감사의 글

본 연구는 2013년도 강원대학교 학술연구조성비로 연구하였음(과제번호-120131222)

References

[1] C. M. Pereira, N. N. Almeida, M. L. F. Velloso, "Fuzzy modeling to forecaste an electric load time series", *Procedia Computer Science*, vol. 55, pp. 395-404, 2015.
 [2] C. C. Hsu, C. Y. Chen, "Regional load forecasting in Taiwan-applications of artificial neural networks", *Energy Conversion and Management*, vol. 44, pp. 1941-1949, 2003.
 [3] J. F. Chen, W. M. Wang, C. M. Huang, "Analysis of and adaptive time-series autoregressive moving-average

- (ARMA) model for short-term load forecasting", Electric Power Systems Research, vol. 34, pp. 187-196, 1995.
- [4] J. H. Puiar, "Fuzzy Ideology based Long Term Load Forecasting", World Academy of Science, Engineering and Technology, vol. 40, pp. 640-645, 2010.
- [5] D. Ali, M. Yohanna, M. I. Puwu, B. M. Garkida, "Long-term load forecast modelling using a fuzzy logic approach", Pacific Science Review A: Natural Science and Engineering, vol. 18, pp. 123-127, 2016.
- [6] J. Dhillon, S. A. Rahman, S. U. Ahmad, M. J. Hossain, "Peak Electricity Load Forecasting Using Online Support Vector Regression", 2016 IEEE Canadian Conference on Electrical and Computer Engineering, pp. 1-5, 2016.
- [7] E. B. UI, Islam, "Comparison of Conventional and Modern Load Forecasting Techniques Based on Artificial Intelligence and Expert Systems", IJCSI, vol. 8, pp. 504-513, 2011.
- [8] G. Mahalakshmi, S. Sridevi, S. Rajaram, "A Survey on Forecasting of Time Series Data", 2016 International Conference on Computing Technologies and Intelligent Data Engineering, pp. 1-8, 2016.
- [9] H. K. Alfares, M. Nazeeruddin, "Electric load forecasting: literature survey and classification of methods", vol. 33, pp. 23-34, 2002.
- [10] J. M. Mendal, Uncertain Rule-based Fuzzy Logic System: Introduction and New Directions, Prentice-Hall, Upper Saddle River, Nj 07458, 2001.
- [11] Y. K. Bang, C. H. Lee, "Fuzzy Time Series Prediction using Hierarchical Clustering Algorithms", Expert Systems with Applications, vol. 38, pp. 4312-4325, 2011.
- [12] Y. K. Bang, J. S. Shim, "Electric Power Load Forecasting using Fuzzy Prediction System", KIEE, vol. 62, pp. 1590-1597, 2013.
- [13] S. Aghabozorgi, A. S. Shirkhorshidi, T. Y. Wah, "Time-Series clustering-A decade review", vol. 53, pp. 16-38, 2015.

저 자 소 개



방 영 근(Young-Keun Bang)

1974년 4월 7일생. 2003년 강원대(삼척) 전기공학과 졸업(석사). 2010년 강원대(춘천) 전기전자공학과 졸업(박사)

E-mail: b2y2c1@hanmail.net



이 철 희(Chul-Heui Lee)

1959년 7월 4일생. 1985년 서울대학교 전기공학과 졸업(석사). 1989년 서울대학교 전기공학과 졸업(박사). 현재 강원대학교 전기전자공학과 교수

E-mail: chlee@kangwon.ac.kr