

준지도 학습 및 신경망 알고리즘을 이용한 전기가격 예측

김항석 · 신현정[†]

아주대학교 산업공학과

Electricity Price Prediction Based on Semi-Supervised Learning and Neural Network Algorithms

Hang Seok Kim · Hyun Jung Shin

Department of Industrial Engineering, Ajou University

Predicting monthly electricity price has been a significant factor of decision-making for plant resource management, fuel purchase plan, plans to plant, operating plan budget, and so on. In this paper, we propose a sophisticated prediction model in terms of the technique of modeling and the variety of the collected variables. The proposed model hybridizes the semi-supervised learning and the artificial neural network algorithms. The former is the most recent and a spotlighted algorithm in data mining and machine learning fields, and the latter is known as one of the well-established algorithms in the fields. Diverse economic/financial indexes such as the crude oil prices, LNG prices, exchange rates, composite indexes of representative global stock markets, etc. are collected and used for the semi-supervised learning which predicts the up-down movement of the price. Whereas various climatic indexes such as temperature, rainfall, sunlight, air pressure, etc. are used for the artificial neural network which predicts the real-values of the price. The resulting values are hybridized in the proposed model. The excellency of the model was empirically verified with the monthly data of electricity price provided by the Korea Energy Economics Institute.

Keywords: Electricity Price Forecasting, Semi-Supervised Learning, Artificial Neural Network, Hybrid Model

1. 서 론

전기가격의 예측은 그 목적과 예측기간에 따라 <Table 1>과 같이 크게 단기예측, 중기예측, 장기예측, 세 분야로 나누어 볼 수 있다. 단기예측은 매일 실시간 거래되는 시간대별/일별 전기가격의 예측을 통해 전력시장에 참여하는 경쟁자들의 이익을 극대화하는데 그 목적이 있다고 할 수 있다(Nogales *et al.*, 2002; Amjady and Keynia, 2008; Pino *et al.*, 2008; Vahidinasab *et al.*, 2008; Amjady and Daraeepour, 2009; Amjady and Keynia, 2009; Tan *et al.*, 2010). 중기예측은 향후 1년간의 월별 전기가격을 예측함으로써 공장을 운영하고 자원 수급을 계획하며 전기생산량을 조절하는데 도움을 준다(Ghiassi *et al.*, 2006; Torghaban *et al.*, 2010). 장기예측은 향후 몇 년간의 연간 전기가격

예측을 통해 에너지 사업의 전망과 투자 계획에 활용된다(Ercan and Soto, 2011).

지금까지 전기가격 예측에 관한 많은 연구가 진행되어 왔는데, 그 연구들의 대부분은 단기 전기가격 예측(STEPF : Short-term electricity price forecasting)에 비중을 두고 있다. 시간대별 및 일별 전기가격 예측을 위해서는 주로 과거 몇 시점까지의 전기가격을 입력으로 사용하는 시계열 분석 방식을 따르고 있다. 주로 사용되는 예측 모델로는 선형회귀 모델(Kim *et al.*, 2006), 다양한 시계열 모델의 조합(Nogales *et al.*, 2002; Weron and Mišiorek, 2008), 인공신경망(Yamin *et al.*, 2004; Gareta *et al.*, 2006; Pao, 2007; Azadeh *et al.*, 2008; Pino *et al.*, 2008; Amjady and Keynia, 2009) 등이 있다. 특히 신경망 모델은 전기가격을 예측하는데 있어 널리 이용되고 있다.

[†] 연락저자 : 신현정 교수, 151-7441 경기도 수원시 영통구 원천동 산 5번지 443-749 아주대학교 산업공학과, Tel : 031-219-2417,

Fax : 031-219-1610, E-mail : shin@ajou.ac.kr

2012년 9월 6일 접수; 2012년 11월 8일 수정본 접수; 2013년 1월 3일 게재 확정.

Table 1. Category of electricity price prediction by forecasting period

	Short-term	mid-term	long-term
period	24 hours ~ 1 week	3 ~ 12 months	2 ~ 30 years
goal	Energy management optimization, Bidding strategy	Power system planning, Scheduling for generation companies, Fuel contracting	Planning of utility investment

기존연구들을 살펴볼 때 몇 가지 아쉬운 점을 찾아볼 수 있다. 우선 대부분의 연구가 전기가격의 단기 예측에 집중되어 있다는 것이다. 현재까지 전기시장은 아직 시장 자유화의 초기단계이기 때문에 중/장기 전기가격 예측 모델을 만들기에 충분한 가격 데이터가 마련되어 있지 못하다는 게 그 이유가 될 수 있다(Aggarwal *et al.*, 2009). 또 다른 문제점으로는 전기가격 예측에 투입되는 입력변수의 범위가 제한적이라는 것이다. 전기가격에 영향을 미치는 요소들을 고려해 보면 전기 생산에 영향을 미치는 요소로서 유가, 환율, LNG 가격 및 각종 경제/금융지표 등을 고려해 볼 수 있겠고, 전기 수요 및 사용에 영향을 미치는 요소로서는 기온이나 습도와 같은 기상관련 요소들을 생각해 볼 수 있다. 그러나 기존 연구들에서는 전기가격 자체만을 입력으로 사용하거나 일부 기상 요소들만을 고려하고 있다(Deb *et al.*, 2000; Chan *et al.*, 2008; Lee and Shin, 2011). 이러한 이유는 대부분의 연구가 시간대별 단기예측에 중점을 두었기 때문에, 일단위로 변동하는 유가나 환율 등을 입력변수로 고려하지 않아도 큰 무리가 없었을 수도 있겠다. 마지막으로 예측모델의 다양화 측면이다. 근래 데이터마이닝 및 기계학습 분야에서는 신경망의 출현 이후에도, 성능이 우수하고 정교한 모델들이 다수 개발되어 왔다. 최신 모델을 예측에 도입한다면 보다 정확한 성능을 기대해 볼 수 있다.

앞서 언급한 기존 연구들의 한계점과 관련하여 본 연구는 다음과 같은 측면에 초점을 둔다. 본 연구에서는 전기가격의 중기 예측을 위한 모델을 제안한다. 월별 전기가격을 예측함으로써, 공장을 운영하는 데 있어서 가동시간 조정, 생산량 조절, 연료 구매 전략 등을 세울 때 중요한 근거자료로 활용할 수 있다. 이와 관련하여 중기 예측에 투입되는 입력변수도 다양화한다. 단기 예측에 주로 사용되었던 기상관련 변수 외에도 경제/금융관련 지표들 즉, 국제 유가, 환율, LNG 가격, 또한 이와 관련된 파생 금융 상품들의 가격, 관련 경제지표 등을 수집하여 입력으로 사용한다. 입력변수가 다양해졌으므로 그에 맞는 예측방법을 설계한다.

전기가격 결정에 있어 전기 생산 측면에 주로 영향을 미치는 경제/금융관련 변수들은, 그들간의 상호관계가 밀접하고 영향을 주고받는 방향도 복잡하고 순환적이므로 이들을 하나의 경제/금융 네트워크로 표현한다. 예측모델로는 최근 등장한 새로운 기계학습 모델인 준지도학습(Semi-Supervised Learning)을 도입한다.

준지도학습 알고리즘은 기존의 감독학습(Supervised Learning)과 비지감독학습(Unsupervised Learning)의 장점을 취하는 새로운

방법으로, 주어진 문제가 데이터 간의 긴밀한 네트워크로 표현되는 경우에는 예측의 성능과 그 결과에 대한 해석 측면에서 매우 우수한 알고리즘으로 알려져 있다(Nigam *et al.*, 1999; Zhou *et al.*, 2004).

전기가격 결정의 수요 측면에는, 주로 전기사용에 영향을 미치는 기온이나 습도 등 기상관련 변수와 과거 전기가격을 사용하고, 여기에는 기존연구에서 안정성이 입증된 신경망을 사용한다(Reich and Barai, 2000; Yamin *et al.*, 2004; Lee *et al.*, 2005; Gareta *et al.*, 2006; Catalao *et al.*, 2007; Pao, 2007; Pino *et al.*, 2008; Vahidinasab *et al.*, 2008).

두 모듈은 각 모듈의 특성에 맞게 분류문제(Classification)와 예측문제(Regression)로 접근한다. 경제/금융 관련 변수들은 원인이 불분명한 오실레이션(Oscillation), 잡음(Noise) 등으로 추세, 주기 등에 불규칙성이 심하여 이로부터 전기가격의 값을 직접 찾는다는 것은 쉽지 않다. 따라서 문제를 단순화하여 전기가격의 등락에 대한 예측값만을 얻는다. 한편, 기상관련 변수들은 추세, 주기, 계절성 등의 패턴이 비교적 규칙적이므로, 이로부터는 전기가격의 실제 예측값을 얻어낸다. 전기가격에 대한 최종 예측 값은, 준지도학습에 의한 등락 예측값과 신경망에 의한 실제가격 예측 값을 일정 기준에 의해 조합하여 산출한다. 제안하는 방법은, 입력변수와 예측모델, 예측방식이 서로 다른 종류의 두 모듈을 결합하는 방식을 취하므로 이를 하이브리드(hybrid) 모델이라 칭한다.

본 논문의 구성은 다음과 같다. 제 2장에서는 하이브리드 모델의 기본 알고리즘인 준지도학습과 인공신경망에 대한 기초 내용을 소개한다. 제 3장에서는 본 연구에서 제안하는 각 모듈의 구성방법과 작동원리, 결합방법에 대해 설명한다. 제 4장에서는 실제 월별 전기가격 예측문제에 대하여 제안한 하이브리드 모델의 성능을 실험, 검증한 후, 제 5장에서는 결론을 짓는다.

2. 관련 알고리즘

본 연구에서 제안하는 전기가격 예측 모델은 최근 기계학습(Machine Learning) 분야에서 각광을 받고 있는 준지도 학습(Semi-Supervised Learning)(Zhu, 2008) 알고리즘과 그 동안 데이터마이닝 관련 다양한 응용 분야에서 성능의 정확도와 안정성이 인정된 인공신경망(Artificial Neural Network)(Bishop and Christopher, 1995) 알고리즘에 기반한다. 다음의 세부 절에서는 두 알고리즘에 대하여 간략히 소개한다.

2.1 준지도 학습 (Semi-Supervised Learning : SSL) 알고리즘

준지도 학습(Semi-Supervised Learning, 이후 SSL로 표기함) 알고리즘은 입력값(X_i)과 함께 목표값(y_i)이 있는 데이터, 즉 레이블이 있는 데이터($\{(X_i, y_i)\}$)와 레이블이 없는 데이터($\{X_i\}$)를 모델을 만들 때 함께 사용할 수 있는 방법이다. 이는 기존의 지도학습(Supervised Learning)에서는 레이블이 있는 데이터만을 사용하고 비 지도학습(Unsupervised Learning)에서는 레이블이 없는 데이터만을 사용하던 방식을 중용한 방법이라 할 수 있다(Chapelle *et al.*, 2006).

근래 등장한 새로운 응용 분야에서는 레이블이 있는 데이터를 얻기가 쉽지 않은 경우가 비일비재하다. 예를 들어, 스팸 메일 분류(Spam Filtering)에서는 이메일에 일일이 스팸 메일인지의 여부를 레이블링 하는 작업에 많은 시간과 비용이 소모되며(Zhu, 2005), 유전자의 기능 분류문제(Gene Expression Data Classification)에서는 발견된 모든 유전자에 대한 기능을 알아내어 레이블링을 하는 데에 고도로 훈련된 전문인력의 수고와 비용이 동반된다(Bair and Tibshirani, 2004; Gong and Chen, 2008). 이외에도 문서 세분화(Document Categorization)(Shin and Tsuda, 2006), 비디오 감시(Video Surveillance)(Shin *et al.*, 2007), 문서 분류(Text Classification)(Subramanya and Bilmes, 2008), 텍스트 단위화 학습(Text Chunking)(Ando and Zhang, 2005), 웹페이지 분류(Webpage Classification)(Liu *et al.*, 2006) 등 많은 신종 응용 분야에서도 앞서 언급한 유사한 어려움이 있다. 이러한 문제들의 공통점은 레이블이 없는 데이터는 쉽게 구할 수 있는 반면 레이블이 있는 데이터는 그 수가 매우 적다는 데에 있다.

모델링 관점에서 볼 때, 지도학습에서는 레이블이 있는 데이터의 수가 적을 때에는 모델이 편기(Bias)되거나 과적합(Overfitting) 또는 과소적합(Underfitting) 등의 오류를 범할 소지가 많다. 이러한 경우, 레이블이 없는 데이터를 모델링에 포함하면 입력 데이터의 분포를 보다 명확하게 만들 수 있다. 이는 클러스터링과 같은 비지도학습에서 데이터의 수가 많아지면 입력 데이터의 구조가 명확해져서 결과가 좋게 나오는 것과 같은 이치이다. 이러한 두 학습방법의 조합 결과로서 모델의 예측 성능을 향상시키고자 하는 것이 SSL의 기본 아이디어이다. 최근에는 SSL의 기본 아이디어가 레이블이 있는 데이터의 수가 적은 경우 뿐만 아니라 많은 경우에도 모델의 정확도와 안정성을 향상시키는 데 기여한다는 내용의 연구가 이루어지고 있다(Shin *et al.*, 2010). 현재까지 알려진 SSL의 기법으로는, 기대값-최대화 기반 방법(Expectation-Maximization Based Approach)(Nigam *et al.*, 1999), Self Training(Yarowsky, 1995), Co-training(Blum and Mitchell, 1998), Transductive Support Vector Machine(Joachims, 1999), Graph Mincuts(Blum and Chawla, 2001), Harmonic Approach(Zhu *et al.*, 2003), Local and Global Consistency(Zhou *et al.*, 2004) 등이 있다.

SSL 알고리즘의 여러 기법들 중 본 연구에서는 그래프에 기반한 SSL(Graph-based Semi-Supervised Learning)을 사용하므로

이를 중심으로 설명한다(Shin *et al.*, 2007; Shin *et al.*, 2010). 그 그래프에 기반한 SSL은 <Figure 1>에서 보듯이 주어진 데이터를 노드와 엣지로 표현한다. 노드는 각 데이터를 나타내고 데이터 간의 연결선인 엣지는 데이터 간의 유사도(Similarity)를 나타낸다. n 개의 데이터($i = 1, l, \dots, n$)가 주어졌을 때, 레이블이 있는 데이터의 수가 l 개라고 하면 전체 데이터는 다음 식과 같이 레이블 $y_i \in \{+1, -1\}$ 이 있는 데이터 셋 D_L 과 레이블이 없는 데이터 셋 D_U 로 나누어 표현할 수 있다. 레이블이 없는 데이터의 목표값 y_i 는 0으로 설정한다.

$$D_L = \{(X_i, y_i) | X_i \in \mathbb{R}^d, y_i \in \{+1, -1\}\}, \quad i = 1, \dots, l$$

$$D_U = \{(X_i, y_i) | X_i \in \mathbb{R}^d, y_i \in \{0\}\}, \quad i = l+1, \dots, n$$

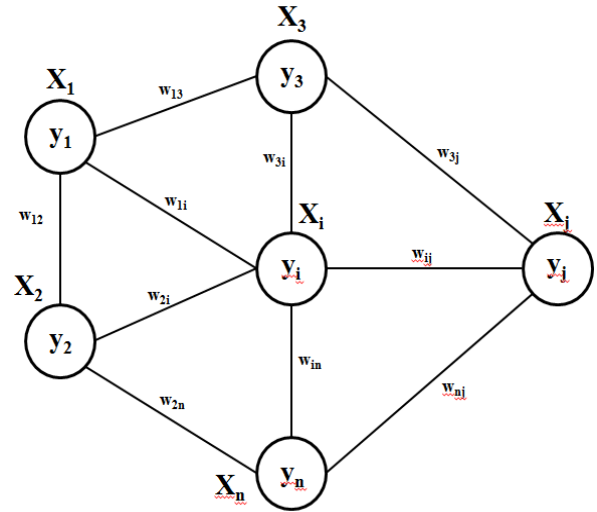


Figure 1. Graph-based semi-supervised learning(SSL)

<Figure 1>에서 보듯이 데이터 간의 연결강도 w_{ij} 는 유사도를 나타내는데 일반적으로 가우시안(Gaussian) 함수가 많이 사용되고 있다.

$$w_{ij} = \begin{cases} \exp\left(-\frac{(X_i - X_j)^T(X_i - X_j)}{\sigma^2}\right) & \text{if } X_i \sim X_j \\ 0 & \text{otherwise} \end{cases}$$

최적화 과정에서는 다음 목적함수의 값을 최소화하는 예측값(\hat{y})을 찾는다.

$$\min_{\hat{y}} (\hat{y} - y)^T (\hat{y} - y) + \mu \hat{y}^T L \hat{y} \quad (1)$$

여기서 \hat{y} 와 y 는 모두 n 차원 열 벡터(column vector)이고, L 은 아래와 같이 정의된 n 차원 정방행렬인 라플라시안(Laplacian) 행렬이다.

$$L = D - W, D = \text{diag}(d_i), d_i = \sum_j w_{ij}.$$

식 (1)의 이차 목적함수를 미분하여 얻어진 최적해, 즉 레이블 예측값 (\hat{y})는 다음과 같다.

$$\hat{y} = \{I + \mu L\}^{-1} y. \quad (2)$$

2.2 인공신경망(Artificial Neural Network : ANN) 알고리즘

인공신경망(Artificial Neural Networks 혹은 신경망, 이후 ANN으로도 표기함)은 분류와 예측을 위한 모형이다. 신경망의 가장 큰 특성은 입력변수와 출력변수간의 관계가 잘 정의되어 있지 않거나 매우 복잡한 관계를 갖더라도, 신경망은 이들 간의 비선형적인 함수 관계를 표현해 낼 수 있는 구조를 가지고 있다는 데에 있다. 이러한 구조적 특성은 종종 다른 분류기 또는 예측기에서는 볼 수 없으며 이로 인한 높은 예측 성능은 신경망의 주요 강점으로 알려져 있다(Gareta *et al.*, 2006).

가장 성공적으로 응용된 신경망은 다층 전방향 신경망(Multilayer Feedforward Networks)이다. <Figure 2>는 일반적인 3개 층의 다층 전방향 신경망 구조를 나타낸다. 각 층은 입력층(Input layer), 은닉층(Hidden layer), 출력층(Output layer)의 순서로 구성되며, 바로 직전 층에서의 노드의 출력이 다음 층 노드들의 입력으로 연결된다. <Figure 2>에서는 $d+1$ 개의 입력노드와 p 개의 은닉노드, 그리고 1개의 출력노드를 갖는 구조를 묘사한 것이다. 전방향 신경망은 입력층에서 출력층 방향으로 노드들이 완전히 연결되어 있으며 순환이 없는 구조이다. 입력층은 단순히 데이터로부터 입력값을 받는 노드들과 한 개의 바이어스(Bias) 노드로 구성되어 있다. 신경망의 특성인 비선형 변환 등의 정보처리 과정은 은닉층과 출력층에서 이루어진다. 은닉층 및 출력층의 노드 개수 및 전이함수(Transfer function)가 적절히 선택되면, 신경망은 간단하지만 강력하고 유연한 분류기/예측기가 될 수 있다고 알려져 있다(Bishop and Christopher, 1995).

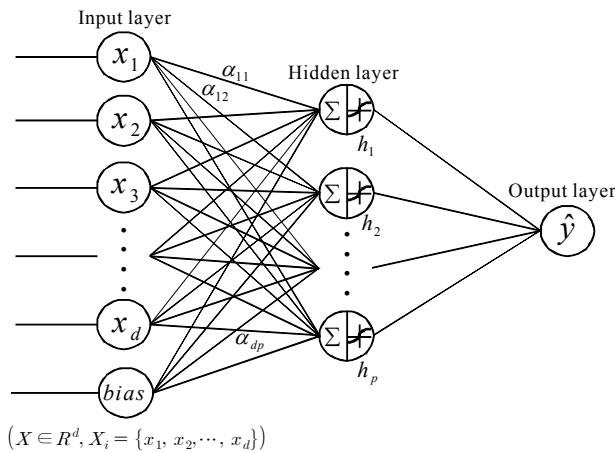


Figure 2. Architecture of the three-layered feed forward artificial neural network

만약 주어진 문제가 실제값을 예측하는 예측문제라면 신경망의 출력노드 수는 한 개로 설정하고 전이함수는 선형함수로 두는 게 일반적이다. 은닉층의 전이함수는 주로 시그모이드 함수를 쓴다. 최적의 신경망 구조는 은닉층의 노드 수와 서로 다른 형태의 전이함수의 조합 중 가장 좋은 것을 선택하여 결정한다.

신경망의 구조가 결정되면 주어진 데이터로 신경망을 학습한다. 신경망 학습이란 최적의 예측결과를 도출하는 노드 간 연결강도 α_{ij} 를 추정하는 것을 의미한다. 최적화 과정은 신경망의 오류(즉, 출력값(\hat{y})과 목표값(y)의 차이)가 사용자가 정한 수준에 다다를 때까지 모든 학습 데이터 셋의 모든 데이터에 대해서 반복적으로 수행된다. 최적화 문제의 목적함수는 오류 제곱 합을 최소화하는 다음의 식을 따른다.

$$\min (y - \hat{y})^T (y - \hat{y}).$$

이 목적함수의 값을 최소화하도록 신경망을 훈련시킨다. 충분한 학습 데이터와 충분한 반복에 의하여 학습이 끝나면 신경망 각 층의 연결강도가 고정된다. 학습을 마친 신경망의 출력노드에서는 새로운 데이터(Test Data)의 입력에 대하여 예측된 출력값을 생성한다.

3. 제안하는 방법론

전기판매단가는 전력생산에 관련된 요소들에 의해 영향을 많이 받는다. 특히, 생산에 직접 투입되는 천연가스 및 원유의 가격은 전기판매단가 결정에 직접적인 영향을 미치며, 이들 수입 원료의 가격은 환율 및 금리 등에 의해 영향을 받는다. 즉, 전기판매단가, 전기수요량, 연료가격, 환율, 선물가격 등 '경제/금융관련 요소'들은 서로 간의 연관성이 긴밀한 네트워크 구조이다. 네트워크 내에서의 한 요소의 변화는 다른 요소에 직접적인 영향을 주기도 하지만 네트워크 구조를 따라 더 많은 요소들에게 간접적으로도 영향을 미친다. 또한, 이러한 영향은 최초 요소로 다시 전달되기도 하는 등 복잡하고 순환적인 관계를 갖는다. 원료를 수입에 의존하는 국내 전기 생산체계의 현실상, 전기판매단가의 가격 형성에 있어서 상당부분은 경제/금융 요소들에 의해 이루어진다고 볼 수 있다. 전기판매단가 결정에 영향을 미치는 또 다른 측면의 요소들로는 '기상 관련 요소'들이 있다. 기온, 습도, 강수량 등의 요소들은 전력수요에 영향을 미치고 궁극적으로는 전기판매단가를 움직이게 한다.

본 연구에서는 전기판매단가를 예측하기 위하여 이에 영향을 미치는 요소들을 '경제/금융관련' 및 '기상관련', 두 측면으로 나누어서 모델링하는 방법을 제안한다. 요소들간의 관계가 복잡, 긴밀한 경제/금융관련 요소들로는 제 2.1절에서 소개한 SSL 네트워크를 구성한다. 경제/금융관련 요소들은, 계절성

및 주기 등의 패턴이 없거나 불규칙하므로, 전기판매단가의 실제값을 예측하기 보다는 다른 요소들로부터 받은 영향에 의해 가격이 오를지 내릴지의 등락을 예측하는 분류문제(Classification)로 접근한다. 한편, 기상관련 요소들로는 제 2.2절에서 소개한 ANN 네트워크를 구성한다. 기상관련 요소들은 계절성 및 주기 등이 비교적 규칙적인 패턴을 가지므로, 전기판매단가의 실제값을 예측하는 예측문제(Regression)로 접근한다. 이후, SSL 네트워크에 의한 가격의 등락값과 ANN 네트워크에 의한 실제값이 결합되어 최종 예측값을 도출하게 된다. 본 연구에서 제안한 SSL과 ANN 하이브리드 모델의 방법론을 도식화하면 <Figure 3>과 같다.

3.1 SSL을 이용한 전기가격의 등락 예측

전기판매단가는 천연가스 가격 및 유가와 밀접한 관련이 있고, 이 두 요소는 각종 금융/경제지표, 즉, 환율, 각종 대표유가(SAUDI, WTI, OPEC, KUWAIT 등) 및 원유 재고량, 세계석유 수요량, 물가지수 등과 연관이 있으므로 이들 간의 관계를 그래프기반 SSL을 활용하여 네트워크로 표현한다. 이를 간략히 도식화하면 <Figure 3>의 SSL 부분과 같다. 여기서 노드들은 전기판매단가에 영향을 미치는 전력수요량 및 각종 금융/경제

관련 변수로 구성한다. 각 노드의 레이블은 '+1' 또는 '-1'로 표현하며 이는 해당변수 값이 t 시점에서 상승했는지 또는 하락했는지의 등락을 의미한다. 예측해야 할 목표변수의 레이블은 '?'로 표기한다. 각 노드 간의 연결강도, 즉 엣지값은 각 변수간의 유사도를 나타낸다. 노드의 레이블링과 노드 간 유사도 설정, 최종 예측값에 대한 해석은 다음 각 절에서 설명한다.

3.1.1 노드의 레이블 y_i 값 설정

경제/금융 관련 요소들로 구성된 네트워크는 각 요소가 네트워크의 노드에 해당하고 각노드 별로 월별 시계열 값을 갖는다. 여기에 SSL을 적용하려면 노드의 레이블을 <Figure 4>와 같이 등/락을 구분하는 이진값으로 변형하여야 한다. 따라서 '등/락'에 대한 판단기준을 정하여, 값이 오르면 '+1', 내리면 '-1'로 표기한다. 본 연구에서는 다음의 식과 같이 레이블을 설정하며 그 의미는 <Figure 5>와 함께 설명한다.

$$y = \text{sign}(x_t - \text{MA}_3(x_t)) \quad (4)$$

t 시점의 데이터를 x_t 라 할 때, t 시점의 실제값(x_t)과 3일 이동평균값($\text{MA}_3(x_t)$)을 비교하여 실제값이 이동평균값 보다 클 때는 '+1', 작을 때는 '-1'로 설정한다. 예를 들어, 이번 달의

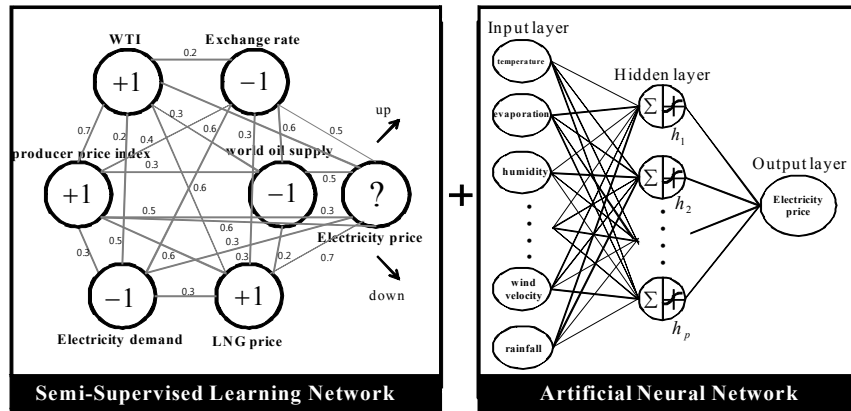


Figure 3. Structure of the hybrid Model

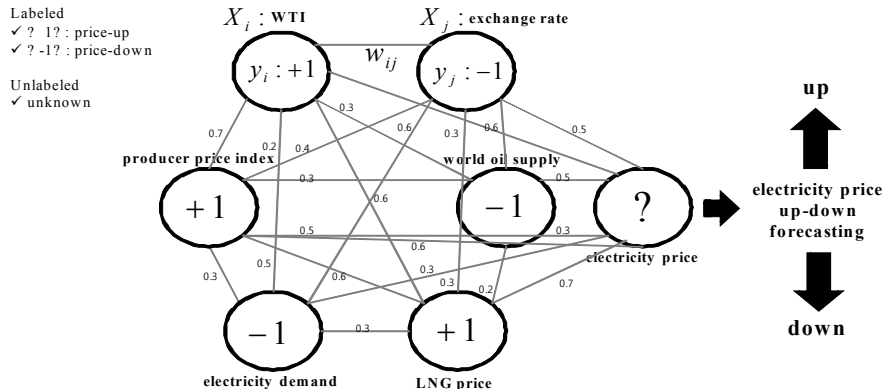


Figure 4. An example of the SSL network composed of economic/financial indexes

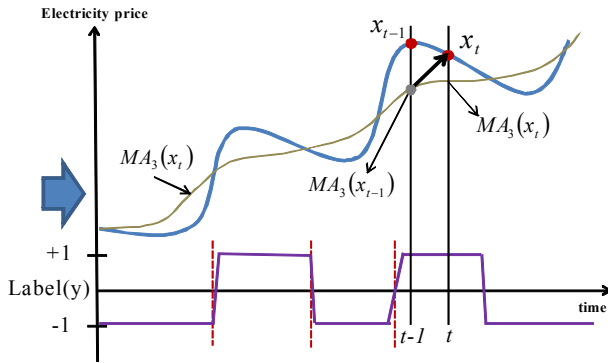


Figure 5. Schematic description for labeling: conversion time-series values into binary values(+1/-1)

LNG 도입가격이 지난 석 달 동안의 이동평균값(MA₃(LNG))보다 상승했으면 LNG도입가격의 이번 달 레이블은 가격이 올랐다는 의미로 ‘+1’로 설정한다. 이러한 레이블링은 <Figure 4>의 모든 노드들에 대하여 매월 갱신된다. 이동평균의 도입이 주는 민감도와 네트워크의 성능을 대비해서 판단하기 위한 이동평균값을 3개월에서 5개월로 확장시켰을 때의 레이블링의 변화 및 그에 따른 예측정확도의 비교는 <부록>에 정리되어 있다.

3.1.2 노드간 유사도 w_{ij} 값 설정

SSL은 어떻게 유사도 매트릭스를 만드느냐가 모델의 예측 성능을 좌우할 수 있다(Shin *et al.*, 2007). 유사도 매트릭스는 <Figure 4>의 SSL 네트워크에서 노드간 엣지에 대한 연결강도(w_{ij})를 나타내며 그 값이 클수록 두 노드 사이의 유사 정도가 크다는 의미를 갖는다. 본 연구에서는 SSL 네트워크를 구성하는 시계열 변수들을 일정기간 동안 각 시점마다 앞장에서 기술한 바와 같이 (+1/-1)로 레이블링하고, 두 노드간의 유사도를 결정하기 위하여 쌍부호검정(Pairwise Sign Test)을 실시한다. 검정 결과 산출된 Kendal 상수의 절대값을 두 노드간의 유사도, 즉 w_{ij} 로 사용한다. 다음 <Table 2>는 LNG 도입가와 전기 판매단가의 레이블을 구하고 이에 대하여 부호검정을 실시한 후 이에 대한 Kendal 상수의 절대값 0.7을 두 노드 간의 유사도로 사용한다.

3.1.3 SSL 네트워크의 예측값 \hat{y} 에 대한 해석

노드의 레이블링과 노드간 유사도가 결정되면, 식 (3)에서와 같이 SSL 네트워크를 통해 전기판매단가의 등락에 대한 예측값(\hat{y})을 얻을 수 있다. 이에 대한 해석은 다음과 같다. <Figure 5>에서 만약 시점 t-1에서 t시점에 대한 전기가격 예측값이 임의의 양수라면, 이는 식 (4)에 의해 ($X_t - MA_3(X_t)$)가 양수라

는 의미이다.

$$\hat{y}_t \Leftrightarrow \text{sign}(X_t - MA_3(X_t)) > 0.$$

이로부터 다음의 부등식이 성립한다.

$$X_t - MA_3(X_t).$$

즉, 다음 달의 전기가격은 과거 3개월의 이동평균보다 높게 상승한다는 의미를 갖는다. 부등식의 우측은 이동평균의 정의

$$MA_k(X_t) = \frac{1}{k}(X_t) + \frac{k-1}{k}MA_k(X_{t-1})$$

에 의해 다음과 같이 표현될 수 있는데

$$X_t > \frac{1}{3}X_t + \frac{2}{3}MA_3(X_{t-1}),$$

이 부등식을 정리하면 다음과 같다.

$$X_t - MA_3(X_t).$$

즉, 시점 t-1에서 t시점에 대한 예측값이 양수이면 다음 달의 전기가격은 과거 3개월의 이동평균보다 높게 상승한다는 의미를 갖는다.

3.2 ANN을 이용한 전기가격의 값 예측

기상관련 변수들 및 과거 몇 시점 이전까지의 전기가격, 즉 자기회귀변수들로는 <Figure 6>과 같이 신경망을 통하여 실제 값을 예측한다. 신경망은 시계열 예측에 많이 사용되어 왔는데 이는 그 구조가 시계열 데이터를 적합시키는데 유연하기 때문이다. 특히, 신경망의 일종인 시간 지연 네트워크에서는 시계열 데이터가 갖는 time-lag를 반영하기도 한다. 본 연구에서는 가장 일반적인 신경망인 feed-forward neural network를 사용하였다.

신경망 입력층의 노드 수는 주어진 입력변수의 수와 바이어스, 그리고 실험에 의해 결정될 자기회귀차수의 수를 합한 수로 결정한다. 마찬가지로 은닉 노드의 수도 실험에 의해 정해진다. 출력층의 노드의 수는 한 개로, 신경망 학습이 끝나면 이로부터 전기판매단가의 예측된 출력값을 얻는다.

Table 2. An example of similarity calculation: LNG price and electricity price

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	Kendal constant
LNG price	+1	-1	-1	+1	+1	+1	+1	-1	-1	-1	+1	-1	0.7
electricity price	+1	+1	-1	-1	+1	+1	+1	+1	-1	-1	+1	+1	

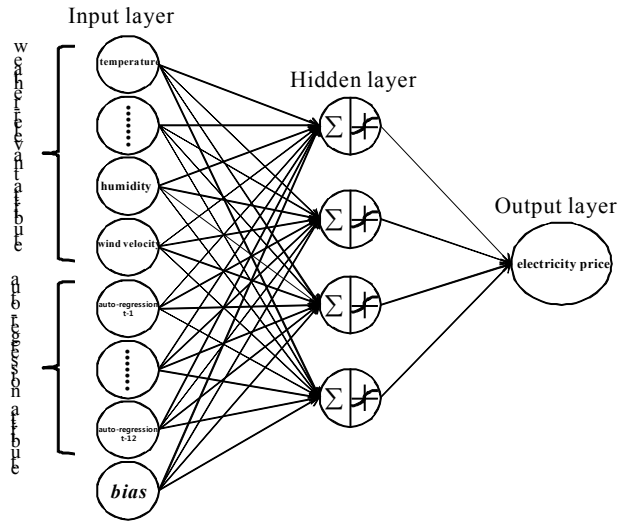


Figure 6. An example of the ANN network composed of weather related indexes

3.3 전기가격의 최종 예측값 도출 기준

<Figure 3>에서 보듯이 본 연구에서 제안한 하이브리드 모델은 SSL 네트워크에서 얻어진 전기가격의 등락 신호와 ANN 모델에서 예측된 전기가격에 대한 실제값에 결합되어 최종 예측값이 도출된다. SSL 출력값과 ANN 출력값에 의한 하이브리드 모델의 최종 예측값 도출과정을 간단히 그림으로 표현하면 <Figure 7>과 같다.

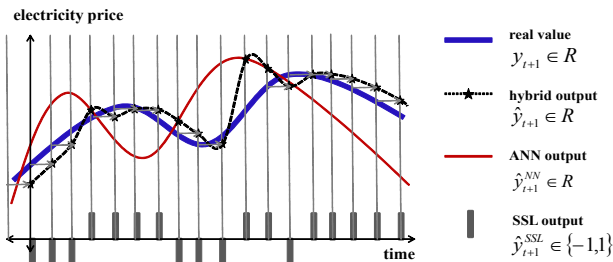


Figure 7. Procedure of generating the final output values from those of SSL and ANN

$y_t \in R$ 를 t 시점에서의 실제 전기가격, $\hat{y}_{t+1}^{SSL} \in [-1, +1]$ 를 $t+1$ 시점에 대한 SSL의 예측값, $\hat{y}_{t+1}^{NN} \in R$ 을 $t+1$ 시점에 대한 ANN의 예측값이라고 했을 때, 최종 예측값(\hat{y}_{t+1})은 다음 식에 의하여 결정된다.

$$\hat{y}_{t+1} = \begin{cases} \hat{y}_{t+1}^{NN} & \text{if } (\hat{y}_{t+1}^{SSL})(\hat{y}_{t+1}^{NN} - y_t) \geq 0 \\ y_t & \text{otherwise} \end{cases} \quad (5)$$

즉, 현재시점 t 에서 다음 시점 $t+1$ 의 전기가격을 예측할 때, SSL과 ANN, 두 모델간의 예측이 일치하는 경우, 즉 (\hat{y}_{t+1}^{SSL})

($\hat{y}_{t+1}^{NN} - y_t$) ≥ 0 인 경우에는 최종 예측값은 ANN의 예측값을 그대로 따르고, 불일치하는 경우에는 현재 시점의 실제값을 그대로 따른다. 두 모델간의 예측이 일치하는 경우는 SSL이 $t+1$ 시점에서 전기가격이 오를 것이라고 예측했을 때 ANN도 역시 전기가격이 t 시점보다 높을 것이라고 예측한 경우거나, 반대로 SSL이 하락을 예측했을 때, ANN도 하락을 예측한 경우이다. 제안한 결합방법은 두 모델의 총론이 일치하는 경우에만 최종 예측값을 변경함으로써, 그림에서 보듯이 어느 한 모델이 잘못된 예측을 하더라도 최종 예측값은 목표값을 비교적 잘 적합하여 안정적인 예측을 가능하게 한다. 특히, ANN은 주로 기상관련 변수를 기반으로 전기가격을 예측하는 것에 반해, SSL은 전기가격에 큰 영향을 미치는 금융/경제 변수들간의 연관관계를 통하여 등락 예측을 하므로, 최종 예측값 결정에는 SSL의 영향력이 더 큰 비중을 갖도록 설계한다.

4. 실험

전기가격 예측에 필요한 원데이터를 실험에 적합하도록 전처리한 후 각 예측모델에 맞도록 변환하여 SSL과 ANN 두 모델에 적용하였다. 이후 하이브리드 모델을 통해 얻어진 최종 예측값에 대한 정확도는 MAPE(Mean Absolute Percentage Error)를 통하여 측정하였다. MAPE는 백분율로 표현하며 그 값이 작을수록 정확도가 좋은 것을 의미하며 수식은 다음과 같이 표현된다.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100$$

(n : 전체 데이터 개수, y_i : 실제값, \hat{y}_i : 예측값)

전체 실험 절차를 정리하면 다음 <Figure 8>과 같다.

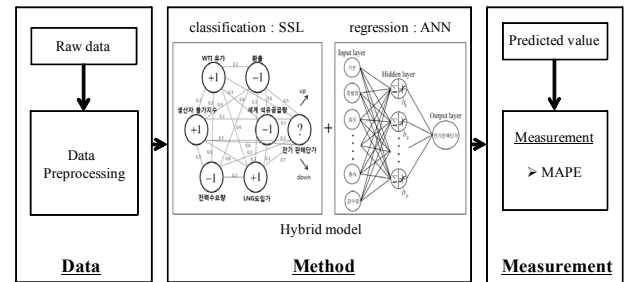


Figure 8. Experimental procedure

4.1 데이터

본 실험에서는 에너지 경제연구원에서 제공한 2000년 1월부터 2008년 7월까지의 월별 전기에너지 관련 데이터와 해당기간의 경제/금융 관련 데이터 및 기상관련 데이터를 사용하였

다. 변수들의 내용은 다음의 <Table 3>과 같고 그 세부내용은 다음의 <Table 4>와 같다.

제안한 알고리즘은 2000년 3월부터 2004년 7월까지의 데이터를 사용하여 학습시키고 2004년 8월부터 2006년 7월까지의 데이터로 검증한 후 2006년 8월부터 2008년 7월까지의 데이터를 사용하여 평가하였다. 해당기간 동안의 전기판매단가를 살펴보면 다음 <Figure 9>와 같다.

4.2 전처리 및 후처리

신경망의 입출력변수의 값은 증가율 $\frac{x_t}{x_{t-1}}$ 의 형태로 변환한 후 다음과 같이 스케일링(Scaling)을 수행하였다.

$$X_t = \frac{(x - \min(x))}{(\max(x) - \min(x))}$$

또한 하이브리드 모델로부터 얻어진 최종 출력값(\hat{y}_t) 또한 전기판매단가의 증가율 $\frac{y_t}{y_{t-1}}$ 에 대한 예측값이므로, 실제 사용을 위해서는 스케일을 재 변환하여 다음과 같이 실제값에 대한 예측값(\hat{y}_t')으로 후처리하였다.

$$\hat{y}_t' = \hat{y}_t \times y_{t-1}.$$

4.3 SSL 네트워크 설정

<Figure 10>은 전기판매단가의 등/락 예측을 위하여, 전력수요와 전기생산에 영향을 미치는 연료가격 및 환율, 그들과 관련 있는 경제/금융 지표, 29개로 이루어진 SSL 네트워크이다. 각 노드 즉, 변수는 <Table 3>에 표기하였다. 그림에서는 임의의 한 시점에서의 네트워크의 노드별 레이블과 노드 간 유사

Table 3. Attributes for the hybrid model

Name Field	Attribute	Count
Electricity-related	Electricity demand, Electricity price	2
Economic/financial	WTI price, LNG price, producer price index, exchange rate, world demand, OECD demand, USA demand, NonOECD demand, China demand, world produce_TOTAL, OPEC produce_TOTAL, SAUDI produce_TOTAL, IRAN_TOTAL, IRAQ_TOTAL, KUWAIT_TOTAL, NonOPEC produce_TOTAL, USA produce_TOTAL, Russia produce_TOTAL, OECD commerce stock_CRUDE, USA commerce stock_TOT, USA commerce stock_CRUDE, OPEC surplus produce, NYMEX_OI, Non commercial real purchase short, Non commercial real purchase long, Commercial volume short, Commercial volume long	27
Weather-related	temperature, evaporation, humidity, wind velocity, sunshine amount, rainfall, cloud amount, air pressure	8

Table 4. Detailed information for the attributes used in electricity price forecasting

year	month	Electricity related attribute		Weather-related attribute				Economic/financial attribute			
		Electricity demand	Electricity price	temperature	evaporation	...	air pressure	LNG price	WTI	...	Exchange rate
		MWh	won/kWh	Celsius	mm	...	hPa	\$/ton	\$/barrel	...	unit/won
2000	1	3,844,623	90.1	-2.0223	0.9194	...	1015.1565	224.5	27.26	...	96.0647
	2	4,088,291	91.48	-1.7062	1.5034	...	1012.0578	237.5	29.37	...	98.2739
	:	:	:	:	:	:	:	:	:	:	:
2001	3	4,774,940	78.72	4.913	2.883	...	1004.7577	262.9	27.24	...	107.3166
	4	4,608,883	82.3	13.2171	4.6069	...	1005.2349	256.1	27.49	...	108.4627
	:	:	:	:	:	:	:	:	:	:	:
2008	1	8,133,650	93.72	-1.6851	1.5913	...	1015.4504	587.9	92.97	...	73.0565
	2	8,428,679	94.8	-1.1769	1.8138	...	1014.1948	603.7	95.39	...	72.5747
	:	:	:	:	:	:	:	:	:	:	:
	7	5,297,262	105.43	25.1101	3.0839	...	997.2516	776.7	133.37	...	70.9195

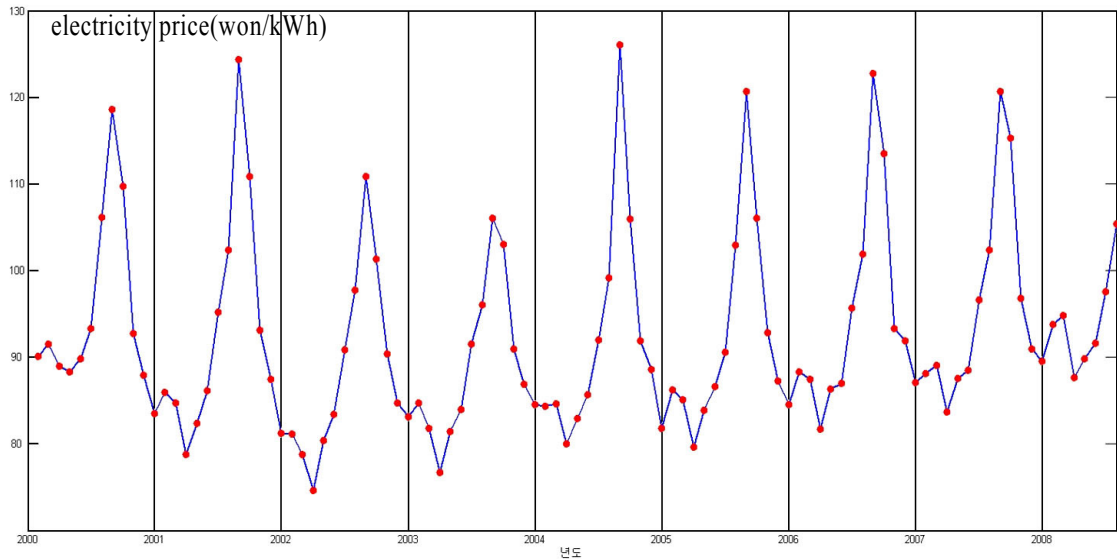


Figure 9. Electricity prices from JAN 2000 to JUL 2008

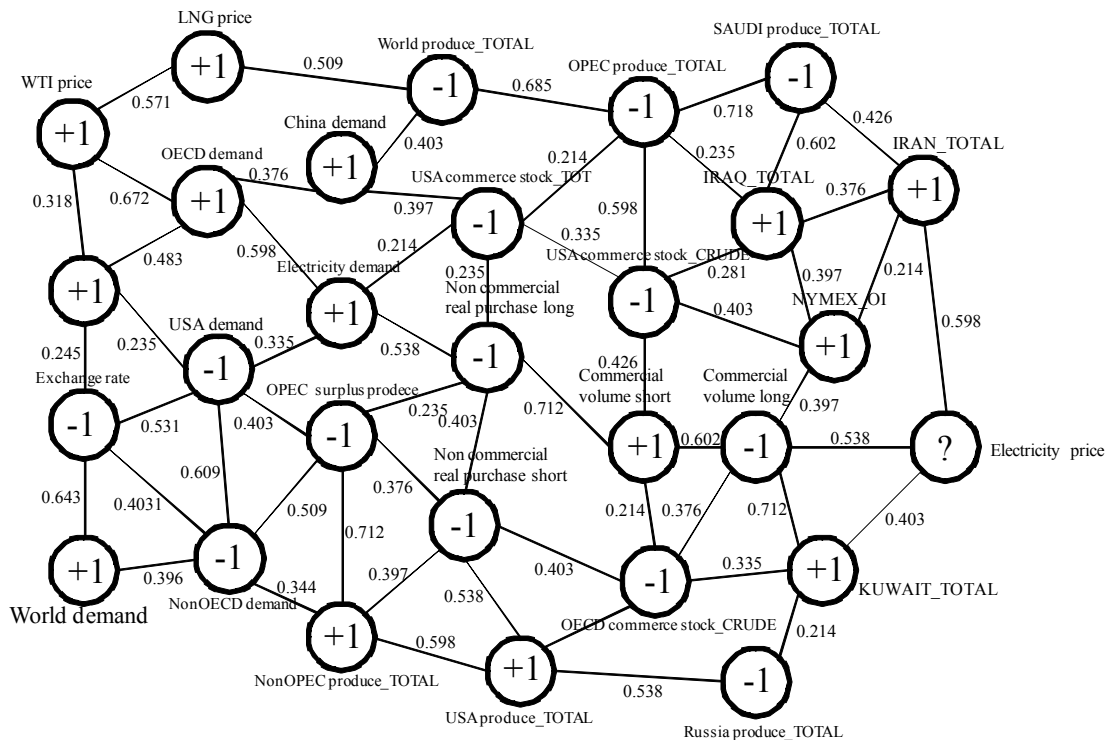


Figure 10. A snapshot of the SSL network at an arbitrary time point: Prediction of the up-down movement of the electricity price

도 값을 보여준다.

4.4 ANN 네트워크 설정

<Table 3>의 8개의 기상관련 지표들은 전기판매단가의 실제 값 예측을 위하여 <Figure 6>의 신경망의 입력변수로 사용하였다. 또한 이전 몇 시점까지의 전기판매단가도 입력변수로 사용되었다. 이를 위하여 자기회귀차수에 대한 결정이 필요하

였다. 실험에서는 t-1시점에서 시작하여 t-16시점까지의 전기 판매단가를 신경망의 입력노드에 추가하면서 검증 데이터 셋에 대하여 가장 작은 MAPE 값을 낸 시점을 자기회귀차수로 결정하였다. <Figure 11>은 t-12시점이 최적의 자기회귀차수로 결정된 것을 보여준다. 마찬가지로 은닉층의 노드 수 또한 검증 데이터 셋을 이용하여 결정하였다. 노드 수를 1부터 입력변수의 수인 최대 20까지 증가시키면서 30회씩 반복 실험을 한 결과, 은닉 노드의 수는 4로 결정되었다. 은닉 노드의 전이함

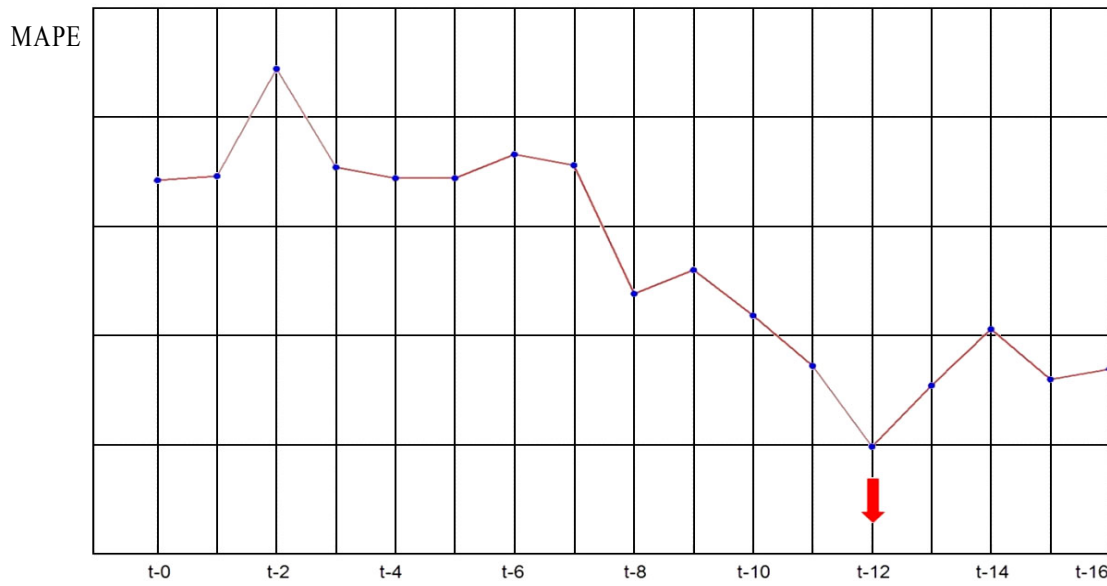


Figure 11. The time-lag selection for the auto-regression of electricity price

수로는 시그모이드 함수를 사용하였고 출력노드의 전이함수로는 선형함수를 사용하였다. 신경망 학습은 Levenberg-Marquardt 방법을 사용하였다.

4.5 비교실험 설정

본 연구에서 제안한 하이브리드 모델(이하 hybrid로도 표기)의 효용성을 검증하기 위하여 다음과 같이 비교 실험을 설정하였다. 첫 번째 비교 모델은, 하이브리드 모델이 사용하는 총 49개의 변수, 즉 전력관련 변수(2개), 경제/금융관련 변수(27개), 기상관련 변수(8개), 자기회귀차수(12개)를 모두 사용하되 신경망만을 통해 예측을 수행한 ANN_A 모델이다. 두 번째 비교 모델은, 기상관련변수(8개) 및 자기회귀차수(12개)만으로 예측을 수행한 신경망 모델 ANN_B이다. 이 비교실험을 통해 알아보고자 하는 효과는 다음과 같다. 각 모델에서 사용하는 변수들을 측면 별로 표시해 보면 <Table 5>와 같다. ANN_A 모델과의

비교를 통해서, 동일한 입력변수들을 사용하되 이들을 SSL 모델과 ANN 모델을 통해 분리해서 사용하는 hybrid 모델의 효과를 검증하고자 한다. ANN_B 모델과의 비교를 통해서, hybrid 모델의 경제/금융 네트워크가 기상관련 변수들로 구성된 신경망의 정확도를 얼마나 향상시키는지에 대한 효과를 검증할 수 있다. 추가적으로, ANN_A 모델과 ANN_B 모델을 비교해 보면, 신경망 모델에 경제/금융 관련변수와 기상관련 변수들을 함께 고려했을 때의 효과를 알아 볼 수 있다. 또한 기계학습과 통계기법의 대표적인 방법인 Support Vector Regression(SVR)과 Auto-regression(AR) 모델이 함께 비교되었다.

4.6 실험 결과

테스트 기간인 2006년 8월부터 2008년 7월까지의 ANN_A, ANN_B, hybrid, SVR, AR의 모델들의 성능을 정리하면 <Figure 12>와 같다.

ANN_A 모델과 hybrid 모델의 MAPE를 비교해 보면, ANN_A가 5.35, hybrid가 3.83으로 제안한 모델의 성능이 더 우수함을 보였다. 두 모델은 입력변수를 동일하게 사용하였으므로, 이러한 성능 차이는 hybrid 모델의 구조로부터 얻어졌음을 알 수 있다. 즉, 경제/금융 변수들로는 전기판매단가의 등락을 예측하고, 기상관련 변수들로는 실제값을 예측한 접근방법이 효과적이었음을 보여준다.

ANN_B 모델과 hybrid 모델을 비교해 보면, ANN_B가 4.64로 제안한 hybrid의 성능이 여전히 보다 우수함을 알 수 있다. 이로부터 hybrid 내의 SSL 모델부분이 ANN 모델 부분의 정확도를 얼마나 향상시켰는지에 대한 정도를 알 수 있었다.

ANN_A 모델과 ANN_B 모델을 비교해 보면, 후자의 성능이 전자보다 더 좋음을 알 수 있다. 두 모델은 동일하게 신경망을 사

Table 5. Setting for comparative experiments between ANN_A, ANN_B and hybrid models

Model Input(count)	ANNA	ANNB	hybrid		SVR	AR
			SSL	ANN		
Electricity-related attribute(2)	●		●		●	●
Economic/financial attribute(27)	●		●		●	●
Weather-related attribute(8)	●	●		●	●	●
Auto-regression attribute(12)	●	●		●	●	●

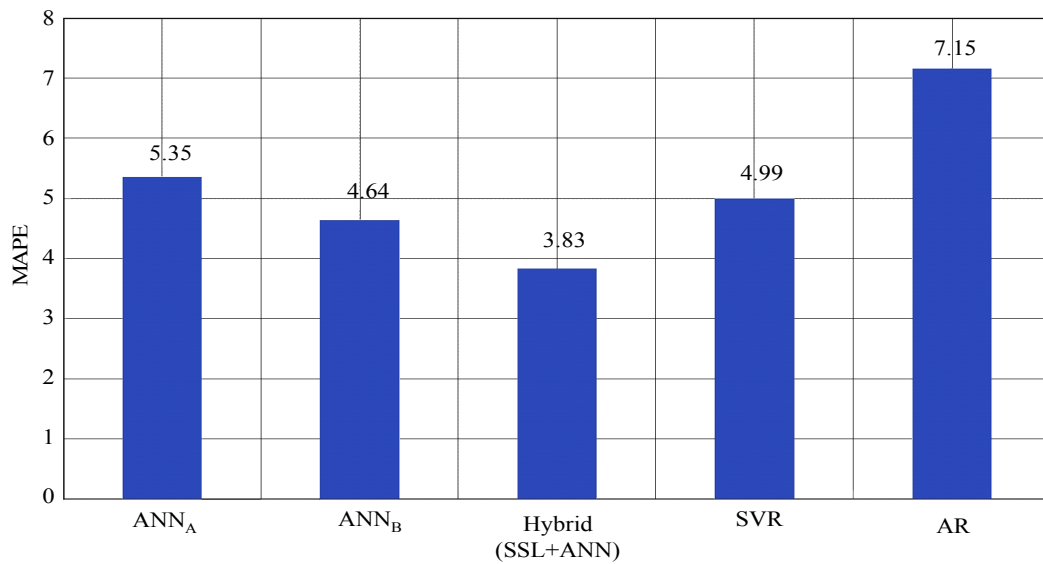


Figure 12. Comparison results(MAPE)

용하므로 두 모델의 성능차이는 입력변수 구성에서 비롯되는 결과라고 볼 수 있다. 이는 경제/금융 관련 변수와 기상관련 변수를 함께 사용하는 경우, 신경망의 성능이 오히려 떨어질 수 있음을 보여준다. 앞서 언급한 바와 같이, 전기판매단가는 기상관련 변수들보다 경제/금융 관련 변수들에 의해 영향을 크게 받는다. 따라서 신경망에서는 기상관련 변수들의 역할이 축소되었을 것으로 보인다. 그런데 문제는 경제/금융 관련 변수들 간에는 상당히 강한 상관관계가 존재한다는 것이다. 신경망 학습 시 입력변수들의 상관성이 강하면 성능의 저하가 초래될 수 있는 데, 이러한 이유로 얻은 결과라고 할 수 있다. 따라서 이들 변수들을 hybrid 모델에서와 같이 분리 사용하는

것이 바람직할 수 있음을 시사하는 결과이다. 또한 기계학습과 통계분야에서 대표적으로 쓰이는 예측모델인 SVR과 AR 역시 동일한 입력변수로 실험한 결과 MAPE값이 각각 4.99, 7.15로 SVR은 신경망과 비슷한 성능을 보였으나 두 모델보다는 hybrid 모델의 성능이 더 우수하였다.

다음의 <Figure 13>~<Figure 15>는 테스트 기간에 대하여 ANN_A, ANN_B, hybrid 세 모델들의 예측값들이 실제 전기판매단가를 얼마나 잘 적합했는지를 보여준다. 세 그림을 비교해 보면, <Figure 15>의 hybrid 모델이 다른 두 모델에 비하여 실제 전기판매단가와 가까운 예측값을 내고 있음을 알 수 있다. SSL에서 노드의 레이블 y_i 값을 설정할 때, 이동평균값을 다르게 설

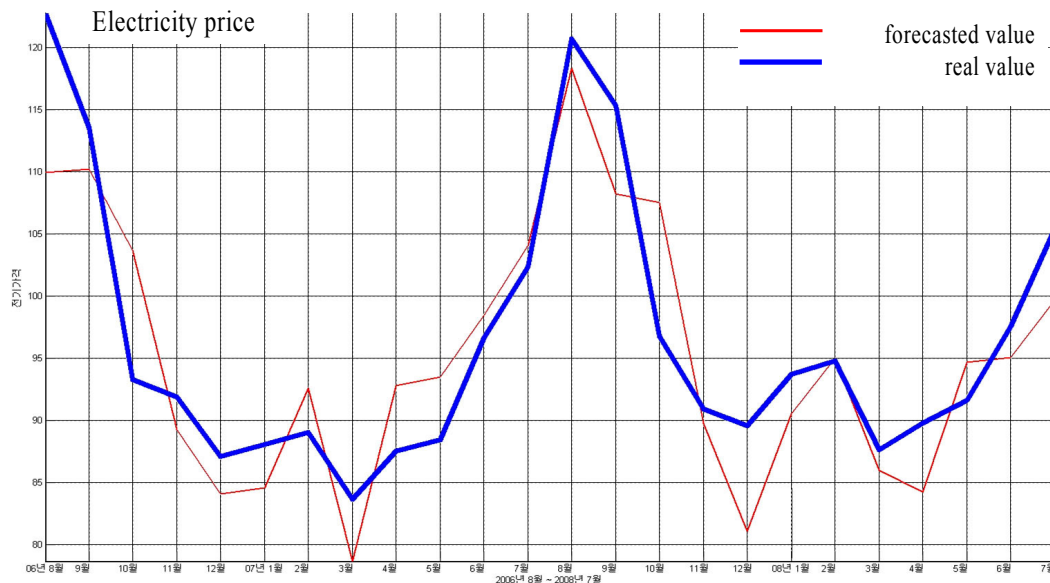


Figure 13. Graph fitting during the test period with the forecasted values by ANNA

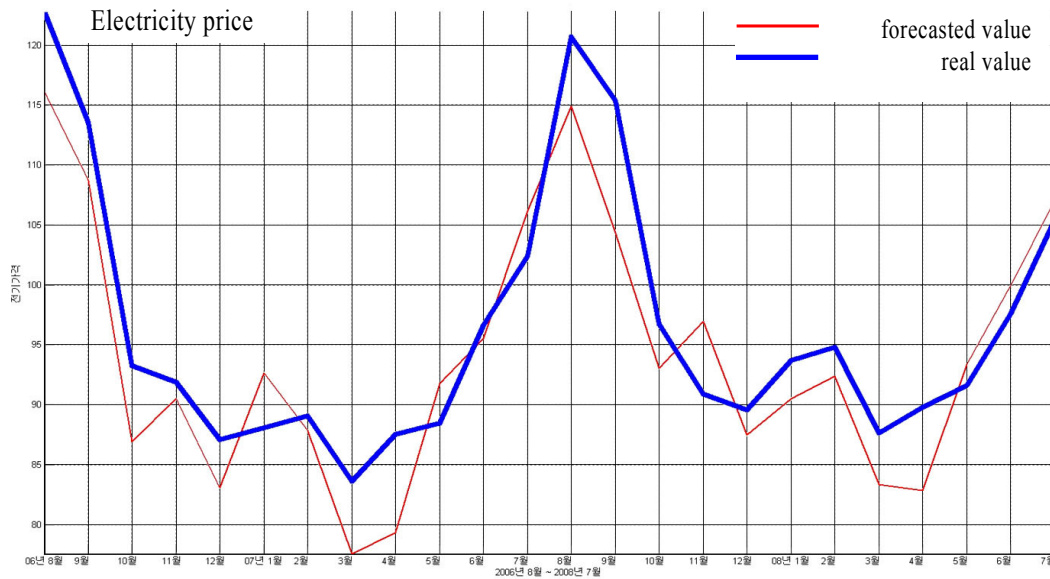


Figure 14. Graph fitting during the test period with the forecasted values by ANNB

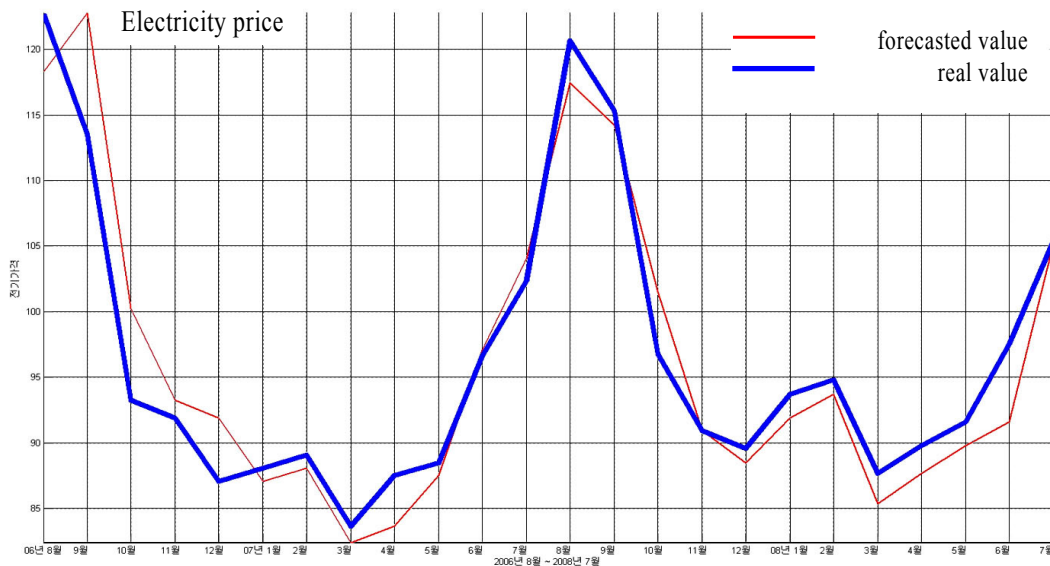


Figure 15. Graph fitting during the test period with the forecasted values by the hybrid model

정했을 때의 추가 실험 결과는 <부록>에 정리되어 있다.

5. 결 론

전기가격의 중기 예측은 월별 전기가격을 예측함으로써, 공장을 운영하는 데 있어서 가동시간 조정, 생산량 조절, 연료 구매 전략 등을 세울 때 중요한 근거자료로 사용된다. 본 연구에서는 중기 전기가격 예측을 위하여 준지도학습 알고리즘(SSL) 및 신경망 알고리즘(ANN)을 조합한 하이브리드 모델을 제안하였다. 제안된 모델의 SSL 모듈에서는 전기 생산과 연관된

경제/금융 관련 변수들을 수집하여 SSL 네트워크를 구성하고 이를 통해 다음 달 전기가격이 상승할지 하락할지를 결정하는 분류문제로 접근하였다. 반면, ANN 모듈에서는 전기 수요에 영향을 미치는 기상관련 변수들로 ANN 네트워크를 구성하고 다음 달 전기가격의 값을 직접 예측하는 예측문제로 접근하였다. 두 모듈로부터 도출된 결과들은 제안한 하이브리드 공식에 의해 결합되어 다음 달 전기가격의 최종 예측값을 산출한다. 이 때에는 국내 전기가격 형성이 주로 원유가격이나 환율 같은 경제/금융 관련 변수에 의해 좌우된다는 현실을 고려하여 SSL 모듈의 결과값에 보다 비중을 주도록 설계하였다. 본 연구는 에너지 경제 연구원이 제공한 2000년 1월부터 2008년 7

월까지의 한국전력거래소 월별 전기가격에 대하여 예측을 수행하였고 그 결과 제안한 알고리즘의 우수성이 검증되었다.

이러한 성과는 제안한 알고리즘이 갖는 다음과 같은 특성들에 기인한다고 요약할 수 있다. 첫째, 전기가격 예측을 수행한 기존 연구들에 비해 사용하는 입력변수들의 수가 다양하고 많다는 점에 있다. 둘째, 제안한 알고리즘은 전기가격을 결정하는 두 측면, 즉 전기생산과 전기수요를 구분하여 고안되었다는 점에 있다. 셋째, 제안한 방법은 입력변수가 갖는 특성에 따라 모듈화 한 후, 각 모듈 별로 적합한 알고리즘을 선택하고 구현했다는 점에 있다. 전기생산에 영향을 미치는 경제/금융관련 요소들 간의 복잡하고 긴밀하며 순환적인 연관성은, 근래 등장한 새로운 기계학습 알고리즘인 SSL을 적용하여 표현하였다. 또한, 전기수요에 영향을 미치는 기상관련 요소들과 전기가격과의 관계는, 입출력 변수 간의 비선형 함수관계 표현이 용이한 신경망으로 구현하였다. 넷째, 문제의 접근방식을 각 모듈의 특성에 맞게 분류문제와 예측문제로 각각 구성했다는 점에 있다. 크게 보면 경제/금융 관련 변수들이나 기상관련 변수들 모두 시계열 변수로 동일하게 여기고 다룰 수 있다. 그러나 전자는 오실레이션, 잡음, 불규칙성이 심하여 이로부터 전기가격의 값을 직접 찾는다는 것은 어려운 일이다. 따라서 문제를 단순화하여 등락에 대한 정보만을 얻어내는 합리적인 방법으로 접근하였다. 반면 후자는 전자에 비해 상대적으로 추세, 주기, 계절성 등의 규칙적인 패턴을 찾을 수 있으므로, 이로부터는 전기가격의 값을 얻어내는 방식을 취하였다. 다섯째, 제안하는 모델의 하이브리드 방법은 현실을 반영한다. 라는 점이다. 전기가격의 형성이 원유나 천연가스 가격, 환율 등에 크게 영향을 받는 국내 현실을 고려하여, 두 모듈간의 결합이 단순한 조합이 아닌 실효적인 예측치를 내도록 구조적 설계를 했다는 특성이 있다.

상기한 본 연구의 여러 특성들은 기존 연구들과 차별성을 갖는다. 본 연구에서 제안한 하이브리드 모델은 전기가격 예측뿐만 아니라 다양한 분야에도 적용이 가능하다. 가깝게는 전기 생산 및 수요 예측에도 직접 적용될 수 있고 금융분야의 주가예측이나 경제분야의 각종 경제지표 예측 등에도 적용 가능하다. 또 한편으로 본 연구는 예측 기간의 단위가 다른 연구에도 확장이 가능하다. 본 연구의 예측 단위는 전기가격에 대한 중기예측이었으나, 데이터가 주어진다면 년 단위의 장기예측 또는 일 단위나 시간 단위의 단기예측으로의 전환도 용이하다. 특히 추후 전력거래소가 활성화된다면 전력거래의 분 단위 거래에도 확장되어질 수 있다고 기대된다.

참고문헌

- Aggarwal, S. K. and Saini, L. M. *et al.* (2009), Electricity price forecasting in deregulated markets : A review and evaluation, *Electrical Power and Energy Systems*, **31**, 13-22.
- Amjady, N. and Daraeepour, A. (2009), Design of input vector for day-ahead price forecasting of electricity markets, *Expert Systems with Applications*, **36**, 12281-12294.
- Amjady, N. and Keynia, F. (2008), Day ahead price forecasting of electricity markets by a mixed data model and hybrid forecast method, *Electrical Power and Energy Systems*, **30**, 533-546.
- Amjady, N. and Keynia, F. (2009), Day-ahead price forecasting of electricity markets by a new feature selection algorithm and cascaded neural network technique, *Energy Conversion and Management*, **50** (2976-2982), 2976-2982.
- Ando, R. K. and Zhang, T. (2005), *A High-Performance Semi-Supervised Learning Method for Text Chunking*, Annual Meeting on Association for Computational Linguistics, Ann Arbor, Michigan.
- Azadeh, A. and Ghaderi, S. F. *et al.* (2008), Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors, *Energy Conversion and Management*, **49**, 2272-2278.
- Bair, E. and Tibshirani, R. (2004), Semi-Supervised Methods to Predict Patient Survival from Gene Expression Data, *PLoS Biology*, **2**, 511-522.
- Bishop and Christopher (1995), *Neural Networks for Pattern Recognition*, Oxford.
- Blum, A. and Chawla, S. (2001), *Learning from Labeled and Unlabeled Data using Graph Mincuts*, ICML 2001 Proceedings of the Eighteenth International Conference on Machine Learning San Francisco, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA.
- Blum, A. and Mitchell, T. (1998), *Combining labeled and unlabeled data with co-training*, COLT 1998 Proceedings of the eleventh annual conference on Computational learning theory New York, ACM New York, NY, USA.
- Catalao, J. P. S. and Mariano, S. J. P. S. *et al.* (2007), Short-term electricity prices forecasting in a competitive market : A neural network approach, *Electric Power Systems Research*, **77**, 1297-1304.
- Chan, K. F. and Gray, P. *et al.* (2008), A new approach to characterizing and forecasting electricity price volatility, *International Journal of Forecasting*, **24**, 728-743.
- Chapelle, O. and Scholkopf, B. *et al.* (2006), *Semi-Supervised Learning*, Cambridge, England, MIT Press.
- Deb, R. and Albert, R. *et al.* (2000), How to Incorporate Volatility and Risk in Electricity Price Forecasting, *The Electricity Journal*, **13**, 65-75.
- Ercan, P. and Soto, J. (2011), *A model for long term electricity price forecasting for France*, Master Thesis, KTH Royal Institute of Technology.
- Gareta, R. and Romeo, L. M. *et al.* (2006), Forecasting of electricity prices with neural networks, *Energy Conversion and Management*, **47**, 1770-1778.
- Ghiassi, M., Zimbra, D. K. *et al.* (2006), Medium term system load forecasting with a dynamic artificial neural network model, *Electric Power Systems Research*, **76**, 302-316.
- Gong, Y.-C. and Chen, C.-L. (2008), *Semi-supervised Method for Gene Expression Data Classification with Gaussian Fields and Harmonic Functions*, International Conference on Pattern Recognition, Tampa, FL.
- Joachims, T. (1999), *Transductive Inference for Text Classification using Support Vector Machines*, International Conference on Machine Learning, San Francisco.
- Kim, D.-Y. Lee, C.-J. *et al.* (2006), Development of System Marginal Price Forecasting Method Using ARIMA Model, *Trans. KIEE*, **55A**, 85-93.

- Lee, H. and Shin, H. (2011), Electricity Demand Forecasting based on Support Vector Regression, *IE Interfaces*, **24**(4), 351-361.
- Lee, J.-K. Park, J.-B. *et al.* (2005), A System Marginal Price Forecasting Method Based on an Artificial Neural Network Using Time and Day Information, *Trans. KIEE*, **54A**, 144-151.
- Liu, R. and Zhou, J. *et al.* (2006), *A Graph-based Semi-supervised Learning Algorithm for Web Page Classification*, International Conference on Intelligent Systems Design and Applications, China.
- Nigam, K. and McCallum, A. K. *et al.* (1999), Text Classification from Labeled and Unlabeled Documents using EM, *Machine Learning*, **39**, 1-34.
- Nogales, F. J., Javier Contreras, M. *IEEE et al.* (2002), Forecasting Next-Day Electricity Prices by Time Series Models, *IEEE TRANSACTIONS ON POWER SYSTEMS*, **17**, 342-348.
- Pao, H.-T. (2007), Forecasting electricity market pricing using artificial neural networks, *Energy Conversion and Management*, **48**, 907-912.
- Pino, R. and Parreno, J. *et al.* (2008), Forecasting next-day price of electricity in Spanish energy market using artificial neural networks, *Engineering Applications of Artificial Intelligence*, **21**, 53-62.
- Reich, Y. and Barai, S. V. (2000), A methodology for building neural networks models from empirical engineering data, *Engineering Applications of Artificial Intelligence*, **13**, 685-694.
- Shin, H. and Hill, N. J. *et al.* (2010), Graph sharpening, *Expert Systems with Applications*, **37**(12), 7870-7879.
- Shin, H. and Lisewski, A. M. *et al.* (2007), Graph sharpening plus graph integration : a synergy that improves protein functional classification, *Bioinformatics*, **23**, 3217-3224.
- Shin, H. and Tsuda, K. (2006), Prediction of Protein Function from Networks, *Semi-Supervised Learning*, O. Chapelle, B. Scholkopf and A. Zien, MIT press, 339-352.
- Subramanya, A. and Bilmes, J. (2008), *Soft-Supervised Learning for Text Classification*, EMNLP 2008 Proceedings of the Conference on Empirical Methods in Natural Language Processing Honolulu, Hawaii, Association for Computational Linguistics Stroudsburg, PA, USA.
- Tan, Z. and Zhang, J. *et al.* (2010), Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models, *Applied Energy*, **87**, 3606-3610.
- Torghabian, S. S. and Zareipour, H. *et al.* (2010), Medium-Term Electricity Market Price Forecasting : A Data-driven Approach, *North American Power Symposium (NAPS)*, Arlington, TX, 1-7.
- Vahidinasab, V. and Jadid, S. *et al.* (2008), Day-ahead price forecasting in restructured power systems using artificial neural networks, *Electric Power Systems Research*, **78**, 1332-1342.
- Weron, R. and Misiorek, A. (2008), Forecasting spot electricity prices : A comparison of parametric and semiparametric time series models, *International Journal of Forecasting*, **24**, 744-763.
- Yamin, H. Y. and Shahidehpour, S. M. *et al.* (2004), Adaptive short-term electricity price forecasting using artificial neural networks in the restructured power markets, *Electrical Power and Energy Systems*, **26**, 571-581.
- Yarowsky, D. (1995), *Unsupervised word sense disambiguation rivaling supervised methods*, ACL 1995 Proceedings of the 33rd annual meeting on Association for Computational Linguistics Stroudsburg, Association for Computational Linguistics Stroudsburg, PA, USA.
- Zhou, D. and Bousquet, O. *et al.* (2004), Learning with local and global consistency, *Advances in Neural Information Processing Systems*, **16**, 321-328.
- Zhu, X. (2005), *Semi-Supervised Learning with Graphs*, *Ph.D. dissertation*, Pittsburgh, PA 15213, Carnegie Mellon.
- Zhu, X. (2008), *Semi-Supervised Learning Literature Survey*, 1-60.
- Zhu, X. and Ghahramani, Z. *et al.* (2003), *Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions*, International Conference on Machine Learning(ICML 2003), Washington DC.

<부 록>

SSL의 레이블링 설정 시 서로 다른 이동평균을 사용했을 때의 차이 비교

금융/경제지표들의 관계를 그래프기반 SSL을 활용하여 전기판매단가의 등락을 예측할 때, 서로 다른 이동평균의 도입이 주는 민감도와 네트워크의 성능을 대비해서 판단하기 위해 이동평균값을 3개월에서 5개월로 확장시키고 그에 따른 SSL 레이블의 변화와 예측정확도를 비교해 보았다.

A1 : 이동평균 기간의 변화(MA3 vs. MA5)에 따른 SSL 레이블 비교

본 실험에서 한 동일 테스트 기간(2006년 8월부터 2008년 7월)에 대하여 SSL의 레이블 y_t 값 설정 시 3개월의 이동평균을 취한 경우(MA3)와 5개월의 이동평균을 취한 경우(MA5)를 비교하여 보았다. <Figure A1>에서 보듯이 24개의 테스트 기간에 대해 레이블 값이 거의 일치하였으나 3구간에서 차이를 보였다.

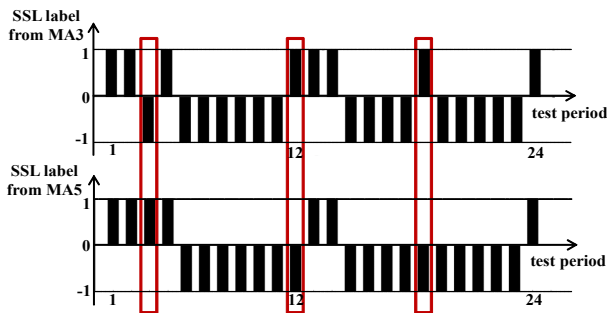


Figure A1. Comparison results(SSL label) : MA3 vs. MA5

식 (5)에서처럼 ANN의 예측과 SSL의 예측이 일치하지 않는 경우를 이동평균의 변화에 대해 찾아본 결과는 <Table A1>과 <Table A2>에 정리하였다.

<Table A1>은 테스트 기간에 대한 세 모델 즉, ANN, SSL(MA3), SSL(MA5)의 전기가격 등락 예측을 나타내며 <Table A2>는 ANN과 SSL의 등락예측이 SSL 레이블 설계 시 적용한 이동평균기간에 따라 어느 정도 일치하는지를 나타낸다. <Table A2>에서 빈 칸은 ANN과 SSL의 등락예측이 일치한 것을 나타내고 ■ 표시가 된 기간은 일치하지 않은 기간을 나타낸다. 3개월의 이동평균을 적용한 MA3는 총 24개월간 6개월의 차이를 보였으나 5개월의 이동평균을 도입한 MA5는 MA3보다 3개월 더 많은 9개월의 차이를 나타내었다. 즉, MA5의 레이블의

Table A1. Difference in signal prediction for the movement of electricity price

Test period	ANN signal	SSL signal(MA3)	SSL signal(MA5)
1	↑	↑	↑
2	↓	↑	↑
3	↓	↓	↑
4	↓	↑	↑
5	↓	↓	↓
6	↑	↓	↓
7	↓	↓	↓
8	↓	↓	↓
9	↓	↓	↓
10	↓	↓	↓
11	↑	↑	↓
12	↑	↑	↑
13	↑	↑	↑
14	↓	↓	↓
15	↓	↓	↓
16	↓	↓	↓
17	↑	↓	↓
18	↑	↑	↓
19	↑	↓	↓
20	↓	↓	↓
21	↓	↓	↓
22	↓	↓	↓
23	↑	↓	↓
24	↑	↑	↑

변화가 MA3보다 상대적으로 적은 것을 알 수 있다. 이는 본 논문은 전기가격의 중기예측을 위하여 월별 데이터를 사용하였고 금융/경제지표들은 주로 분기별로 계산하기 때문에 3개월 단위의 이동평균이 더 적절했다고 생각된다. 만일 단기 예측을 위한 일별데이터를 사용하는 경우라면 이때는 주 단위가 5일이기 때문에 MA5를 사용하는 방법이 더 적절할 수도 있을 것이다.

A2 : 이동평균기간의 변화(MA3 vs. MA5)에 따른 hybrid 모델의 예측정확도 비교

테스트 기간인 2006년 8월부터 2008년 7월까지의 기간에 대하여 제안하는 하이브리드 모델의 SSL레이블 설정 시 이동평

Table A2. Disagreement between SSL and ANN

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
MA3		■		■		■											■		■				■	
MA5		■	■	■		■					■						■	■	■				■	

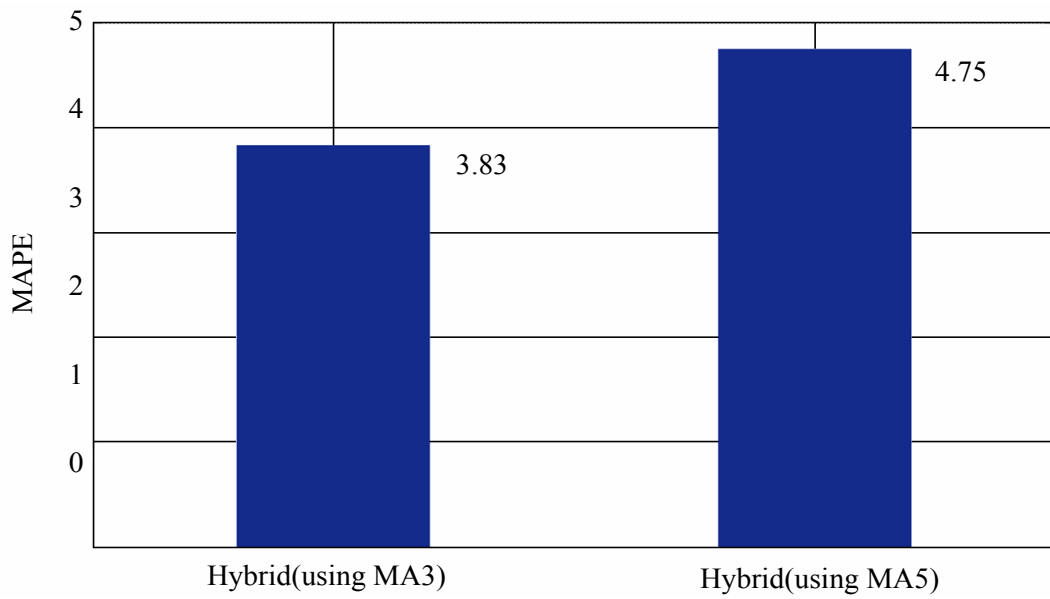


Figure A2. Comparison results(MAPE) : MA3 vs. MA5

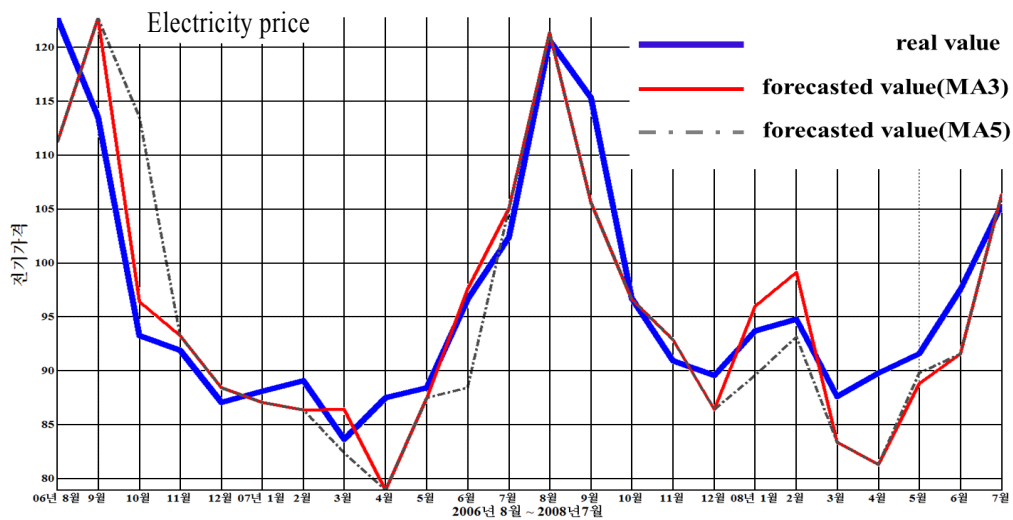


Figure A3. Graph fitting during the test period with the forecasted values by the hybrid model using MA3 and MA5

균기간을 3개월 및 5개월로 적용하여 실험한 결과 예측정확도를 MAPE로 나타내면 <Figure A2>와 같다. MAPE를 비교해 보면 MA3을 적용한 하이브리드 모델은 3.83, MA5를 적용한 하

이브리드 모델은 4.75로, MA3를 적용한 하이브리드 모델의 예측성능이 더 우수하였다. <Figure A3>은 두 경우 예측값들이 실제 전기판매단가를 얼마나 잘 적합했는지를 보여준다.