A Hybrid Model Development for Electricity Energy Price Prediction

Hangseok Kim[1], Kanghee Park[1], Hyunjung Shin[2,*]
[1] Department of Industrial Engineering, Ajou University
San 5, Wonchun-dong, Yeoungtong-gu, 443-749, Suwon, Korea
[e-mail:tdea, can17@ajou.ac.kr]
[2] Department of Industrial & Information Systems Engineering, Ajou University
San 5, Wonchun-dong, Yeoungtong-gu, 443-749, Suwon, Korea
[e-mail:shin@ajou.ac.kr]
*Corresponding author: Hyunjung Shin

Abstract

In this article, we propose a hybrid model for forecasting electricity price. The im pact of economic conditions on fluctuations of electricity price is designed using Semi-Supervised Learning. The real value of electricity price is predicted based on the climate indices using Artificial Neural Network. The results obtained by two al gorithms are combined through hybrid model.
.

Keywords: Artificial Neural Network, Semi-Supervised Learning (SSL), Hybrid Model.

Acknowledgment: The authors would like to gratefully acknowledge support from Post Brain Korea 21 and the research grant from National Research Foundation of Korea (2010-0007804)

A Hybrid Model Development for Electricity Energy Price Prediction

Hangseok Kim, Kanghee Park, Hyunjung Shin*

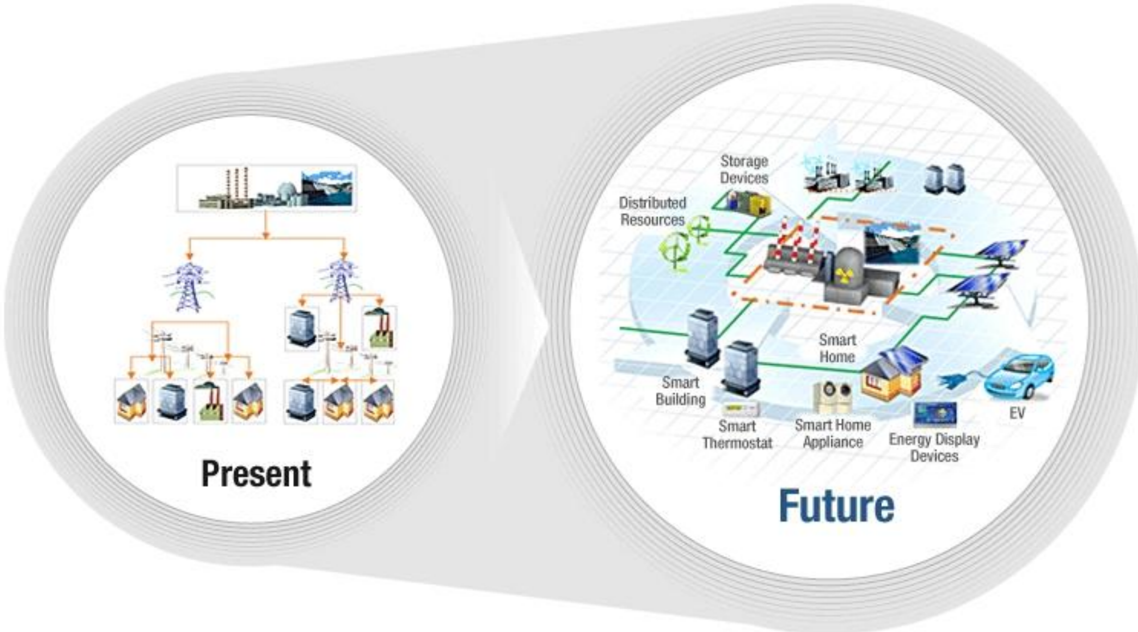
Data Mining Lab, Industrial Engineering Dept, Ajou University

*Corresponding Author: Hyunjung Shin, shin@ajou.ac.kr

Outline

Introduction

Introduction



1. Centralized Generation → Centralized & Distributed Generation
2. Large-scale Generation using Fossil Fuel → Greater Use of Renewables
3. One-way Flow of Power & Information → Two-way Flows of Power & Information (Real-time)
4. Supplier-oriented Facility Operation → Customer Participation in Facility Operation

Electricity Price

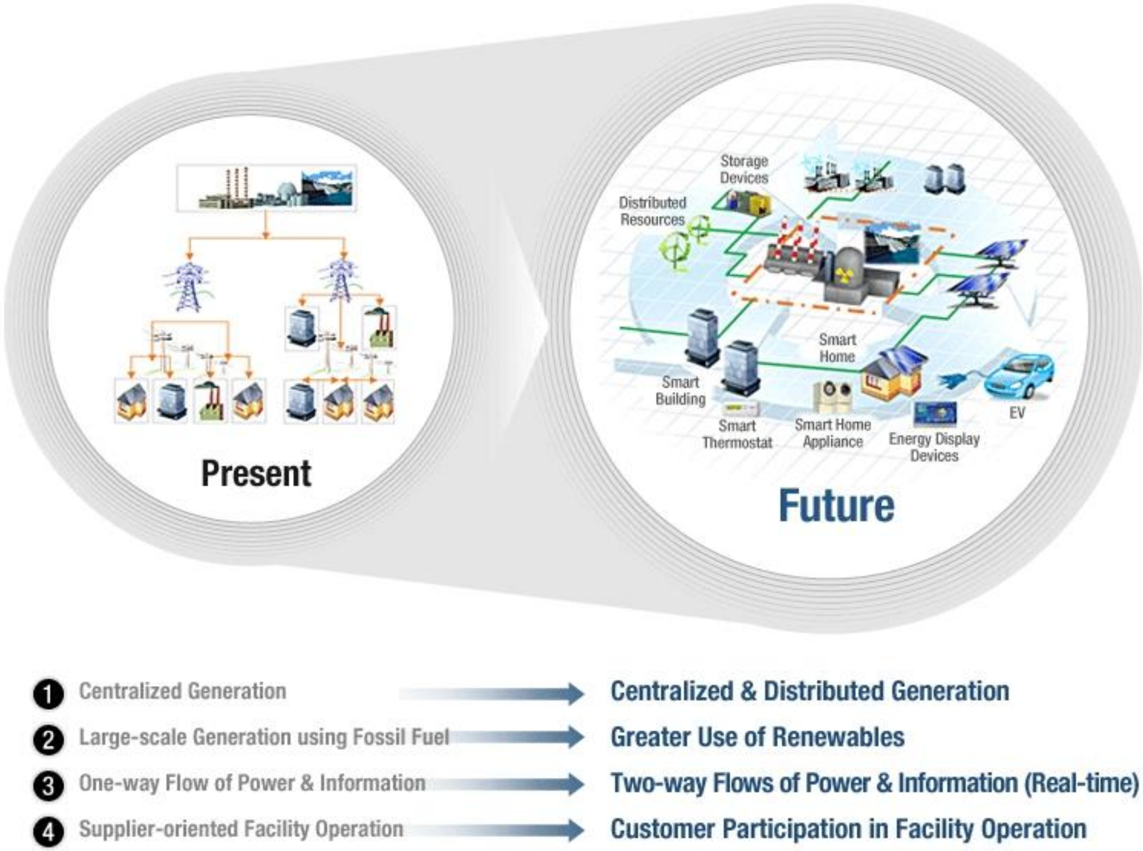
<Present>
Electricity Price
Producer -> Consumer
(simplex)

<Future>
Smart Grid
Producer <-> Consumer
(duplex)

The recently used method for determining electricity prices is a simplex way that dispatches the prices by a producer to consumers.

Reference: http://www.kepco.co.kr/sg/

Introduction



Electricity Price

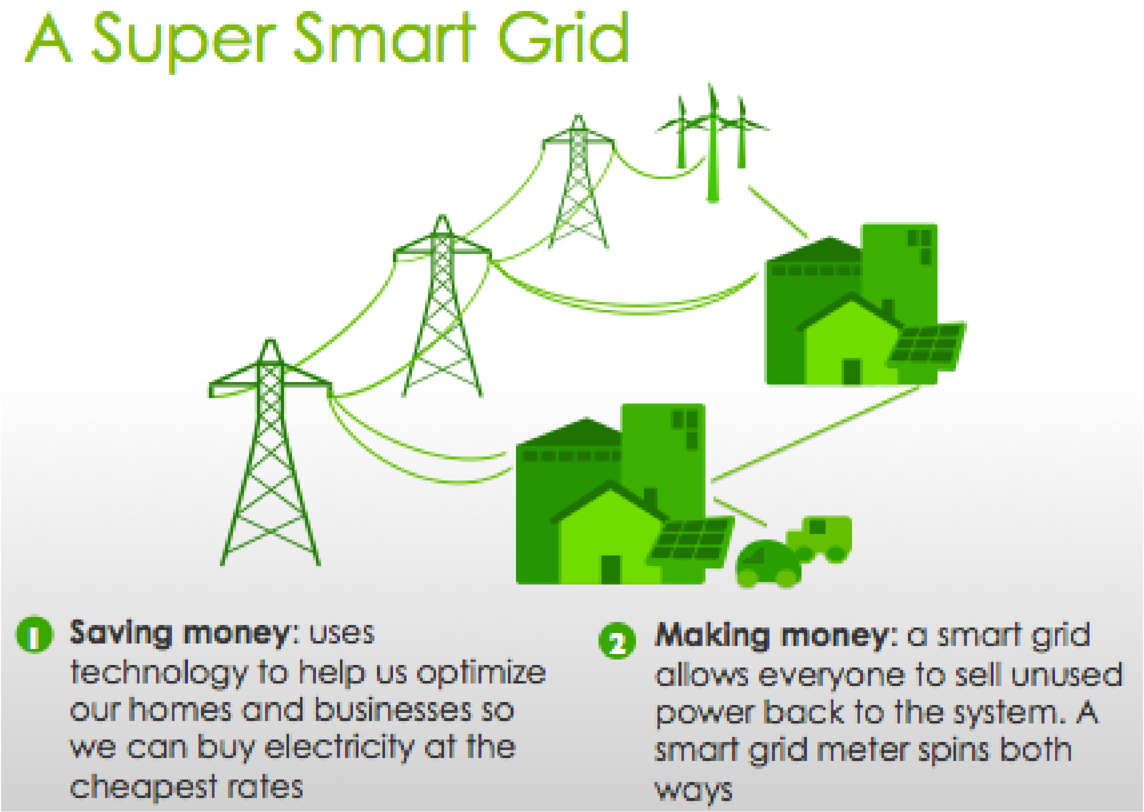
<Present>
Electricity Price
Producer -> Consumer
(simplex)

<Future>
Smart Grid
Producer <-> Consumer
(duplex)

However, in a future smart grid age, it will be determined as a duplex way between the producer and consumers.

Reference: http://www.kepco.co.kr/sg/

Introduction



Effects of Smart Grids

- It gives benefits to both the producer and consumers due to its effective uses in electricity.

Reference: http://www.kepco.co.kr/sg/

Introduction

The Need for Electricity Price Prediction

The premise of such smart grids requires a prediction in electricity prices.

Reference: http://www.kepco.co.kr/sg/

Introduction

The Need for Electricity Price Prediction

The prediction of the prices is also a common interest in both the producer and consumers. In addition, it surely affects reductions in energies.

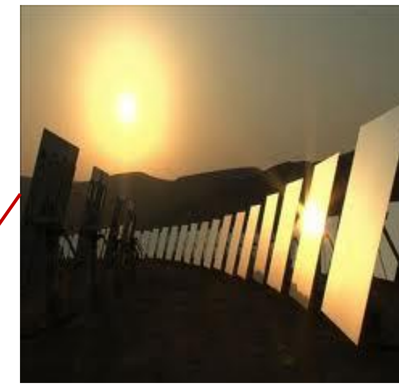
Reference: http://www.kepco.co.kr/sg/

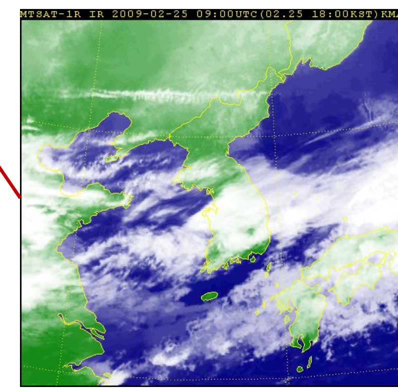
Introduction

Economical Factors

Climatic Factors

Electricity Price



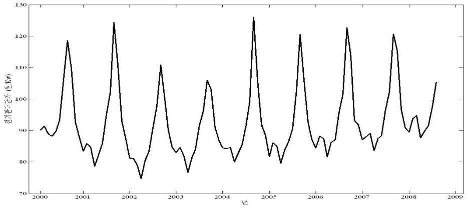


The electricity price includes some economical factors such as foreign exchange rates and oil prices, and climatic factors like solar radiation, temperature, and so on.

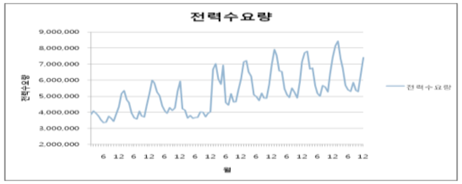
Introduction

Previous Research

Electricity Prices



Electricity Demands



Oil Prices

Climates

Regression

Artificial Neural Network

GARCH

ARIMA

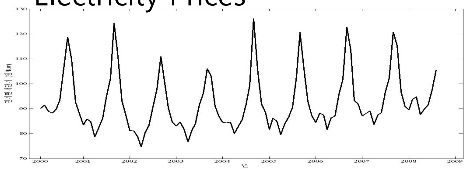
…

Electricity Price

Introduction

Previous Research

Electricity Prices



Electricity Demands

Oil Prices

Climates

Regression

Artificial Neural Network

GARCH

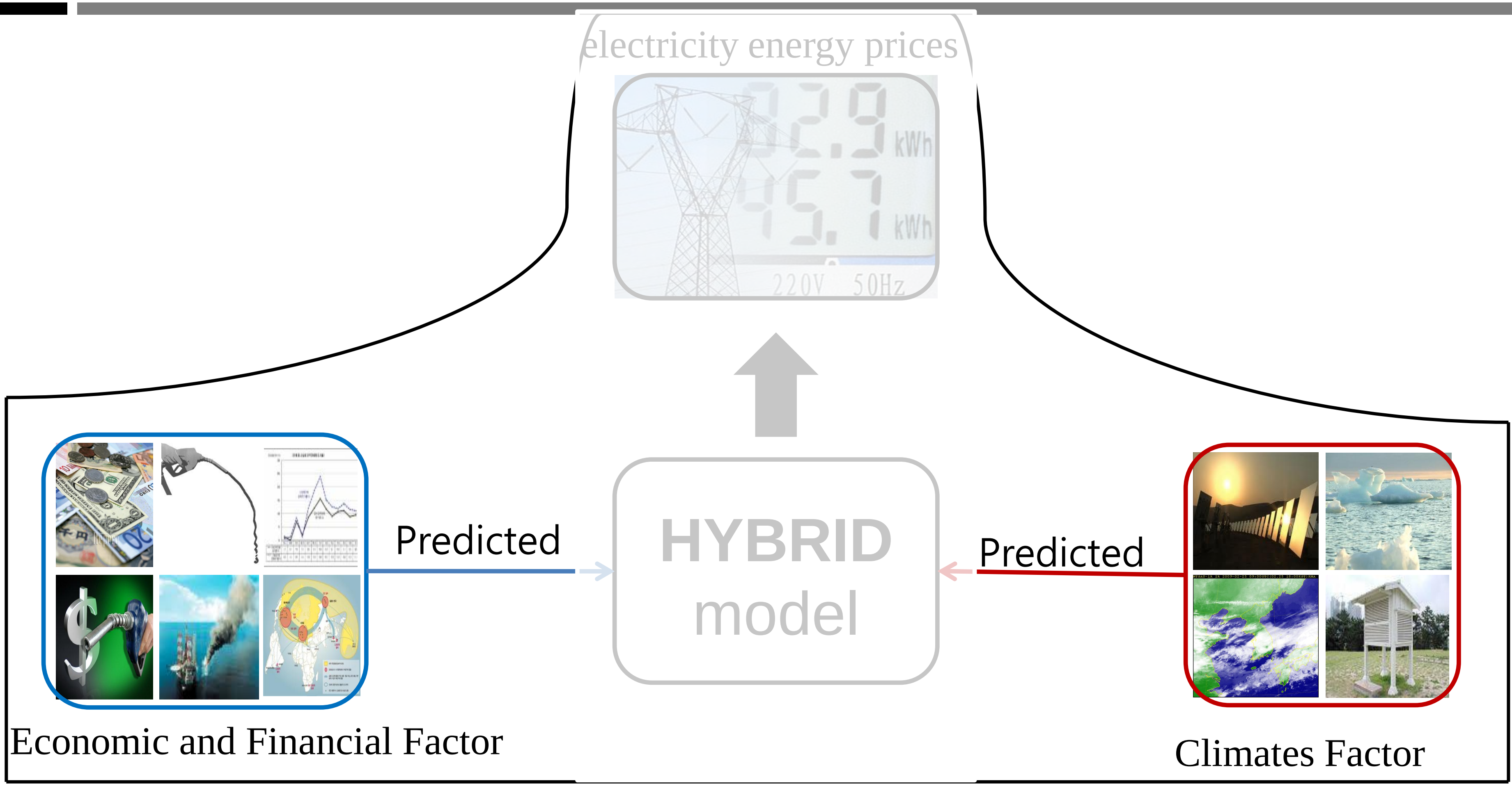
ARIMA

…

Electricity Price

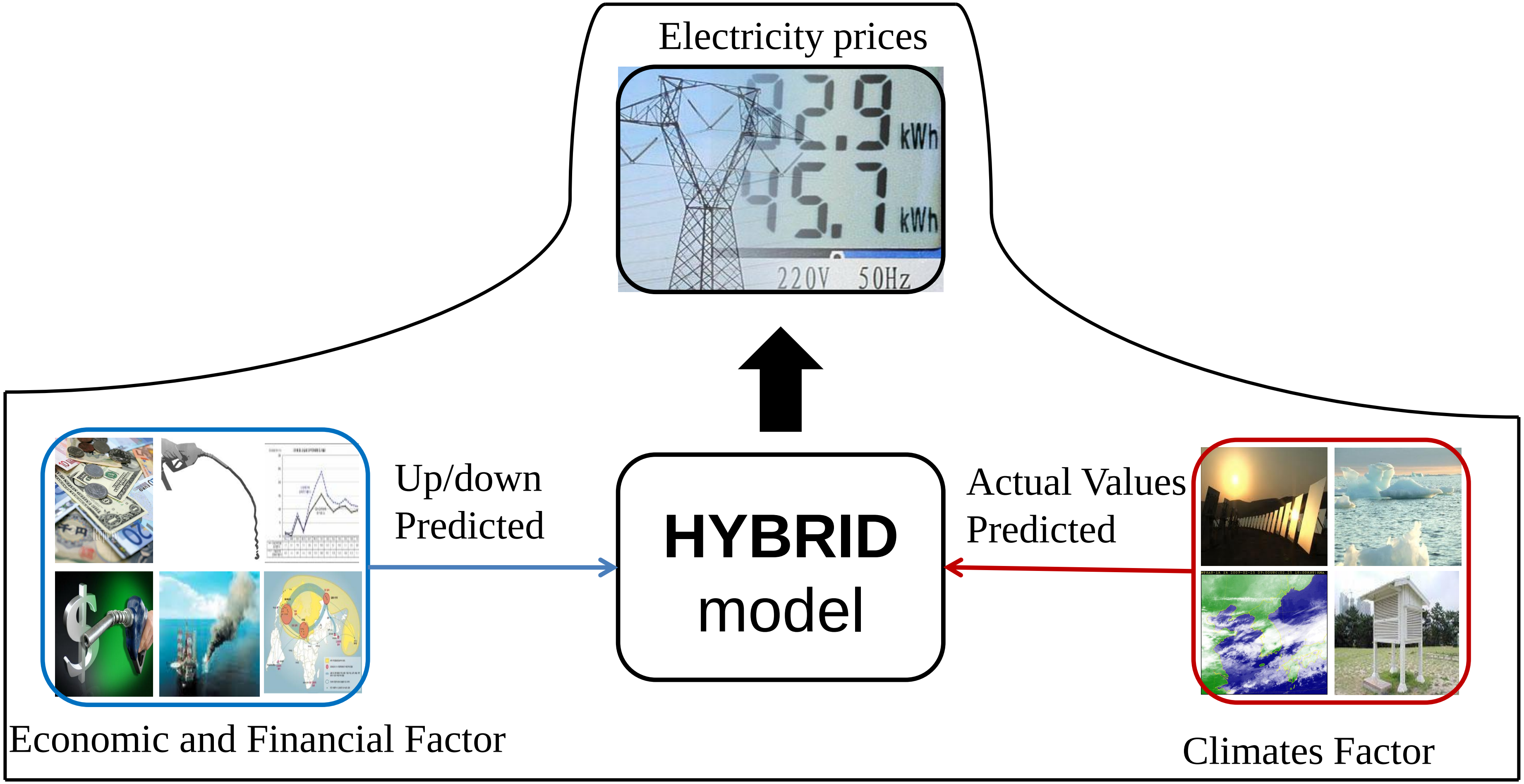
But these algorithms represent certain drawbacks in predicting such electricity prices because they do not consider the interrelation between these various factors for using such variables in the prediction.

Introduction



Thus, in this study, a model for predicting the up/down movement of some economically and financially related factors that consider their interrelations will be introduced. Also, a different model for considering climatically related factors will also be built.

Introduction



Then, a hybrid model based on these two models will be designed to predict the electricity prices.

Introduction

Features in this study

1. Variables consider both climatic and economic.
2. In the prices, it considers both predicted up/down and predicted actual values simultaneously.
3. Closely related variables are to be considered through grouping them.
4. A hybrid model is developed to perform an actual prediction.

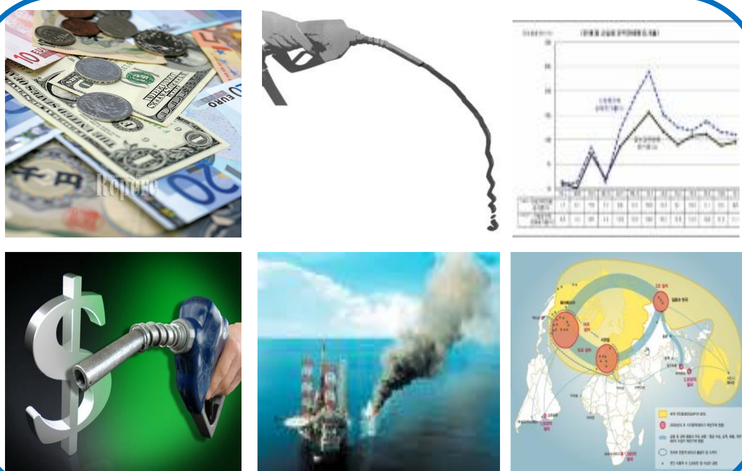
Motivation

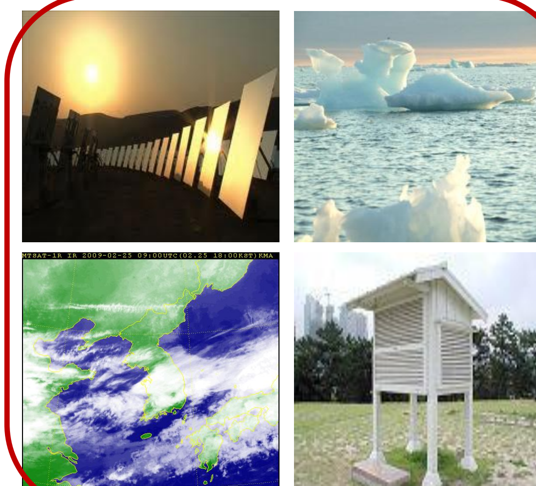
Motivation

UP/DOWN
Prediction

Real Value
Prediction

Climatically Factors : **Artificial Neural Network**





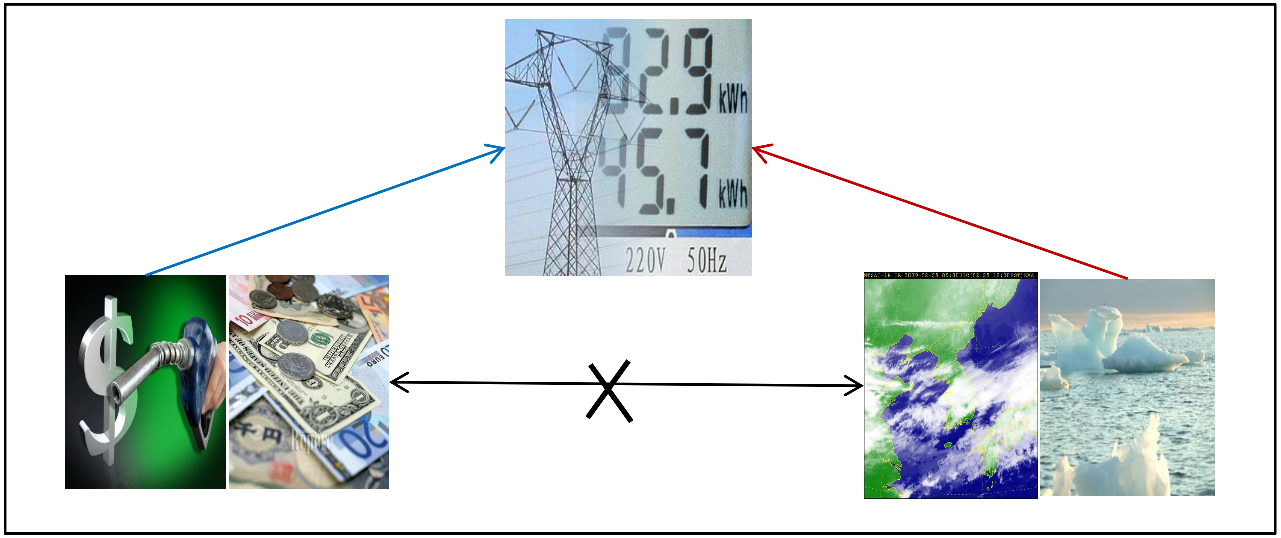
Economically and Financially Factors : **Semi-Supervised Learning**

Input variables are to be separated and considered using these two models.
The up/down prediction in economically and financially related factors are predicted using a SSL model. Climatically related factors are predicted using an ANN model.

Motivation

◆ The reasons of considering input variables into several groups

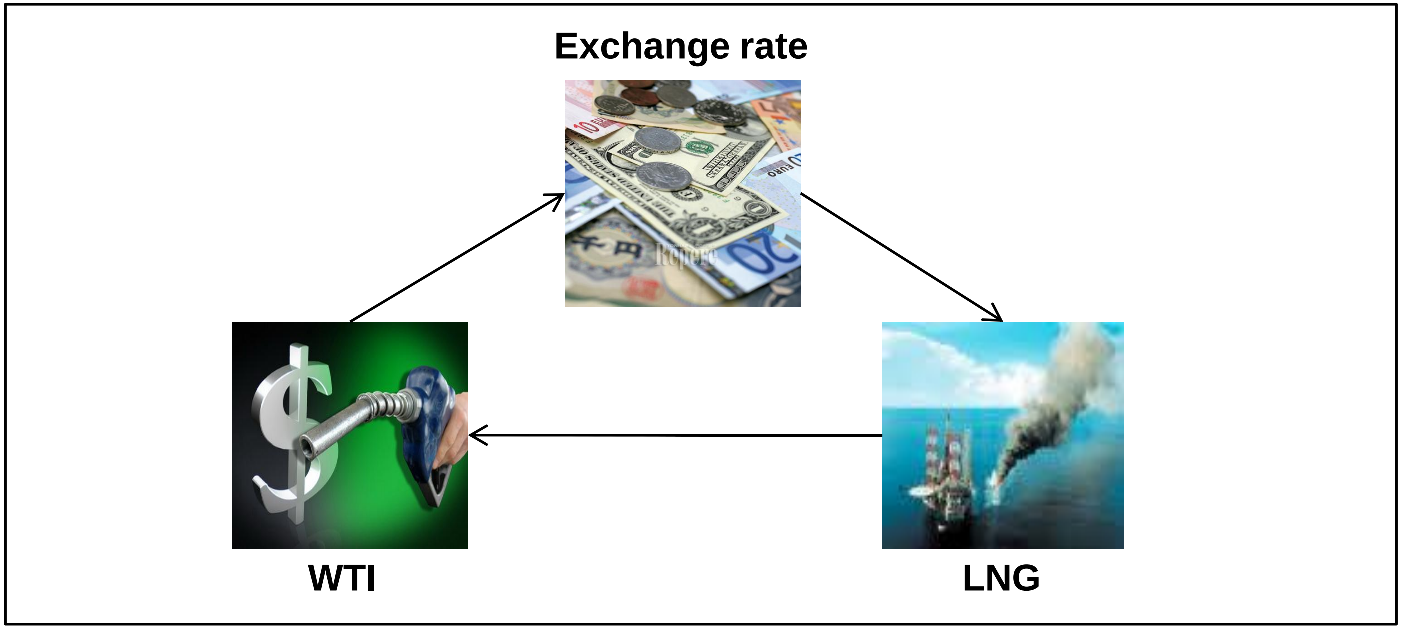
1. There is little probability that uses the economic/financial information together with the climatic information simultaneously.



Motivation

◆ The reasons of considering input variables into several groups

2. There is a strong correlation between economic and financial factors.



Motivation

◆ The reasons of considering input variables into several groups

3. The SSL algorithm represents a technical limitation that can be used to an issue of classification.

4. The influences of climatic factors on the demands/prices of electricity can be referenced by ANN.

Motivation

◆ The reasons of considering input variables into several groups

5. Finally, it is possible to overcome some technical difficulties in the applied algorithm and conceptual differences in different information on variables through developing a hybrid method.

Method

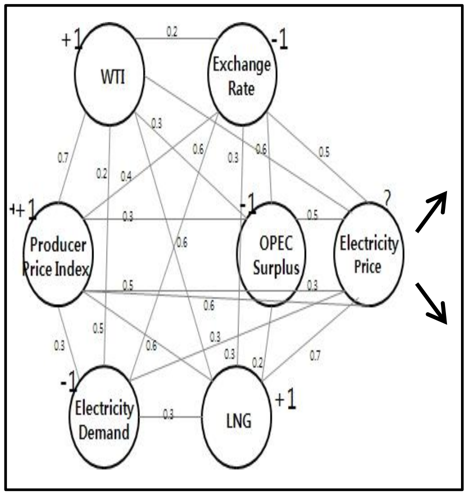
Modeling Process

Raw data

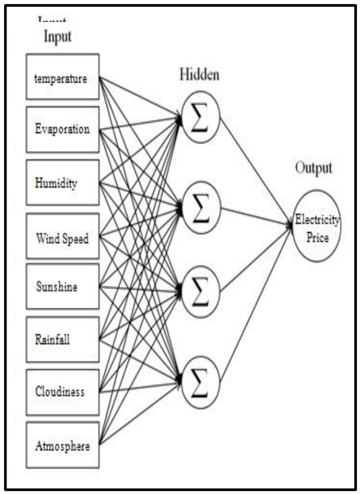
Data Preprocessing

Data

classification : SSL



regression : ANN



Financial variables

Climatic variables

Hybrid model

Method

Predicted value

Measurement

➢ MAPE

Measurement

Therefore, the modeling process for predicting the electricity price can be determined as follows.

Method

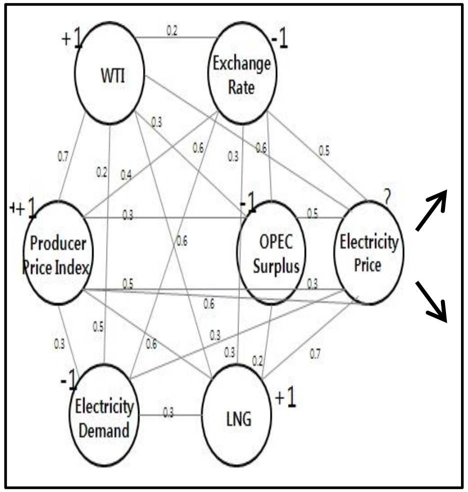
Method

Raw data

Data Preprocessing

Data

classification : SSL



+

Financial variables

regression : ANN

Climatic variables

Predicted value

Measurement

➢ MAPE

Measurement

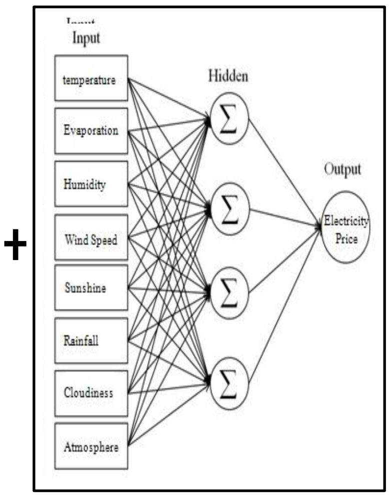
◆ First, The up/down prediction in electricity prices are determined: **Semi-Supervised Learning (SSL)** - Economic and financial factors

Method

- Second, the actual values in electricity prices are determined: **Artificial Neural Network (ANN)-** Climatic factors

regression : ANN

+



Climatic variables

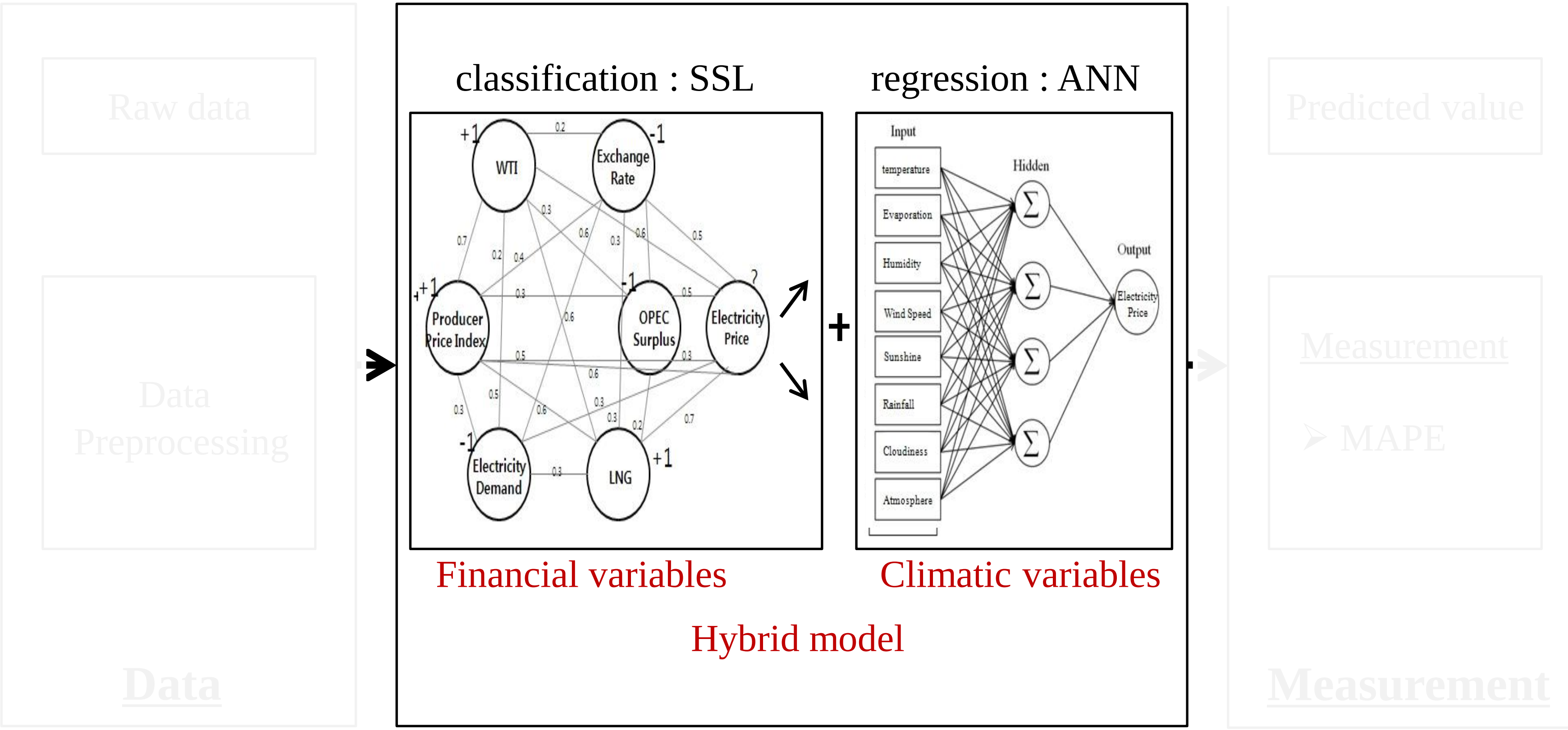
Predicted value

Measurement

- MAPE

Measurement

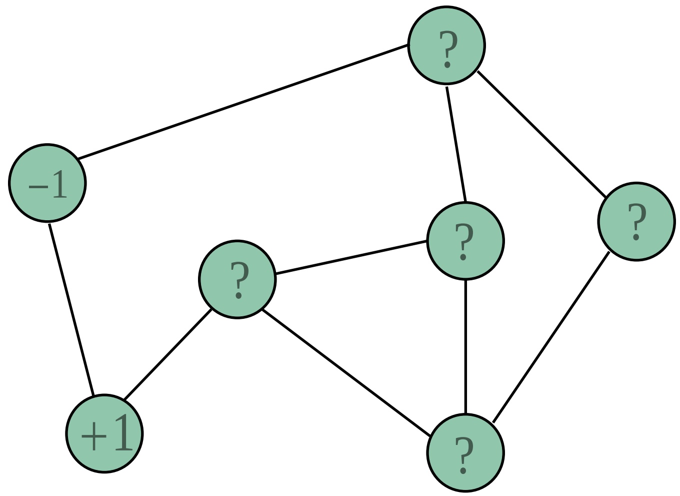
Method



◆ Finally, a hybrid model based on SSL and ANN is designed to determine the final electricity prices: **HYBRID Model (SSL+ANN)**

- Economic and financial factors + Climatic factors

Method: Semi-Supervised Learning



Objective Function

$$\min_{\boldsymbol{f}} \quad \mu\, \boldsymbol{f}^T L\, \boldsymbol{f} + (\boldsymbol{f} - \boldsymbol{y})^T (\boldsymbol{f} - \boldsymbol{y})$$

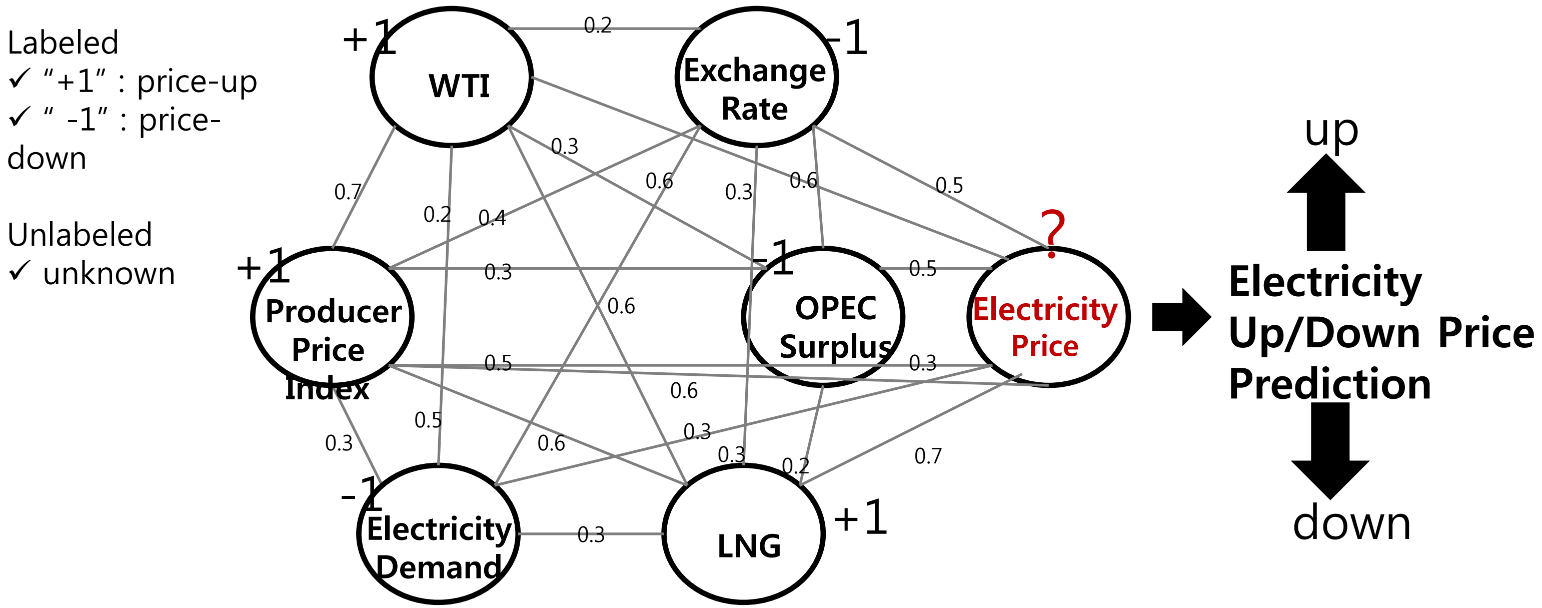
Solution

$$\boldsymbol{f} = \{\, \boldsymbol{I} + \mu L \,\}^{-1} \boldsymbol{y}$$

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

Method: SSL Model Composition

- The up/down movement of the economic, financial factors and electricity price are used input and output variables respectively.
- The up/down movement in these variables are presented by '+1' and '-1', which are marked at nodes.
- The label of the objective variable to be predicted is presented by '?'.

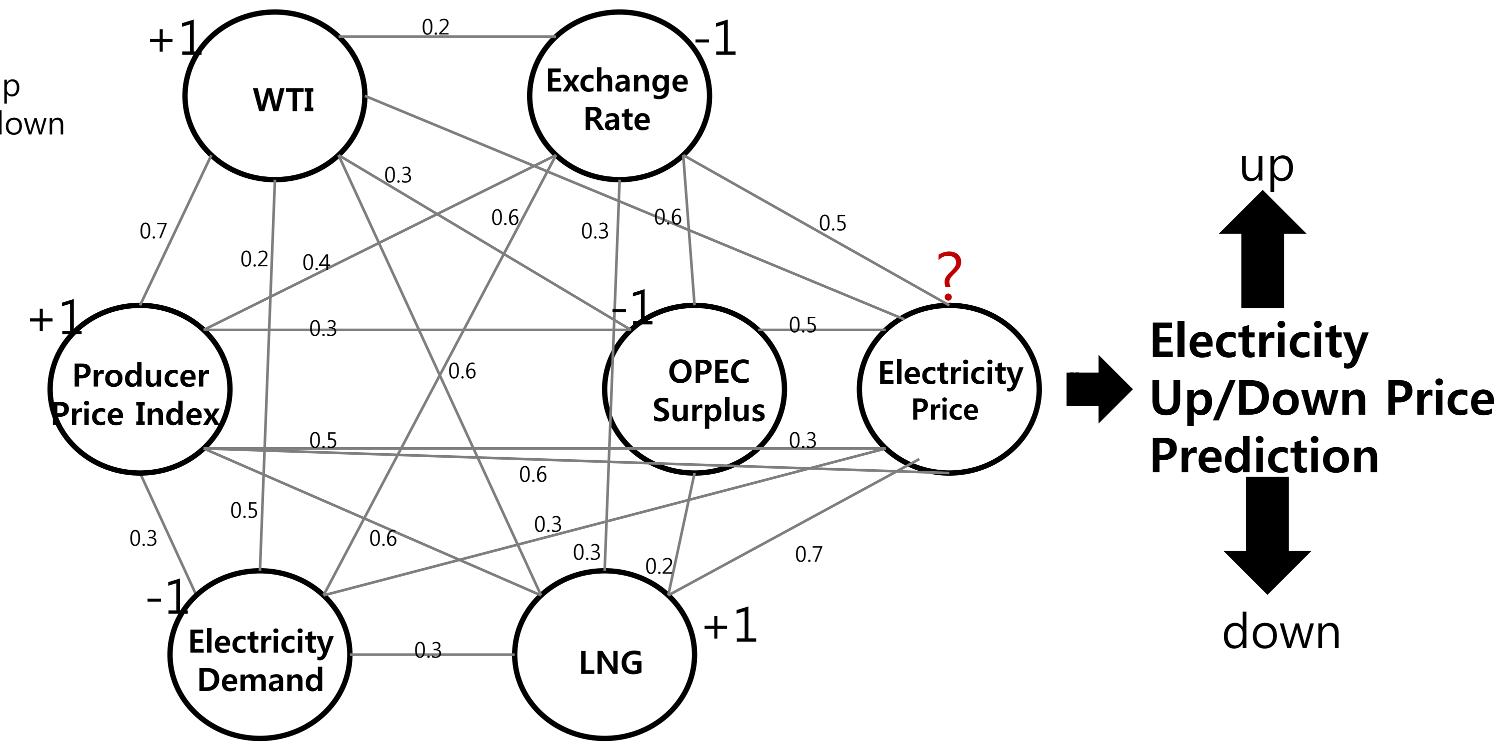


Method: Labeling, y

- The ‘up/down movement ’ of the variables presented by nodes are as follows.

Labeled
✓ “+1” : price-up
✓ “ -1” : price-down

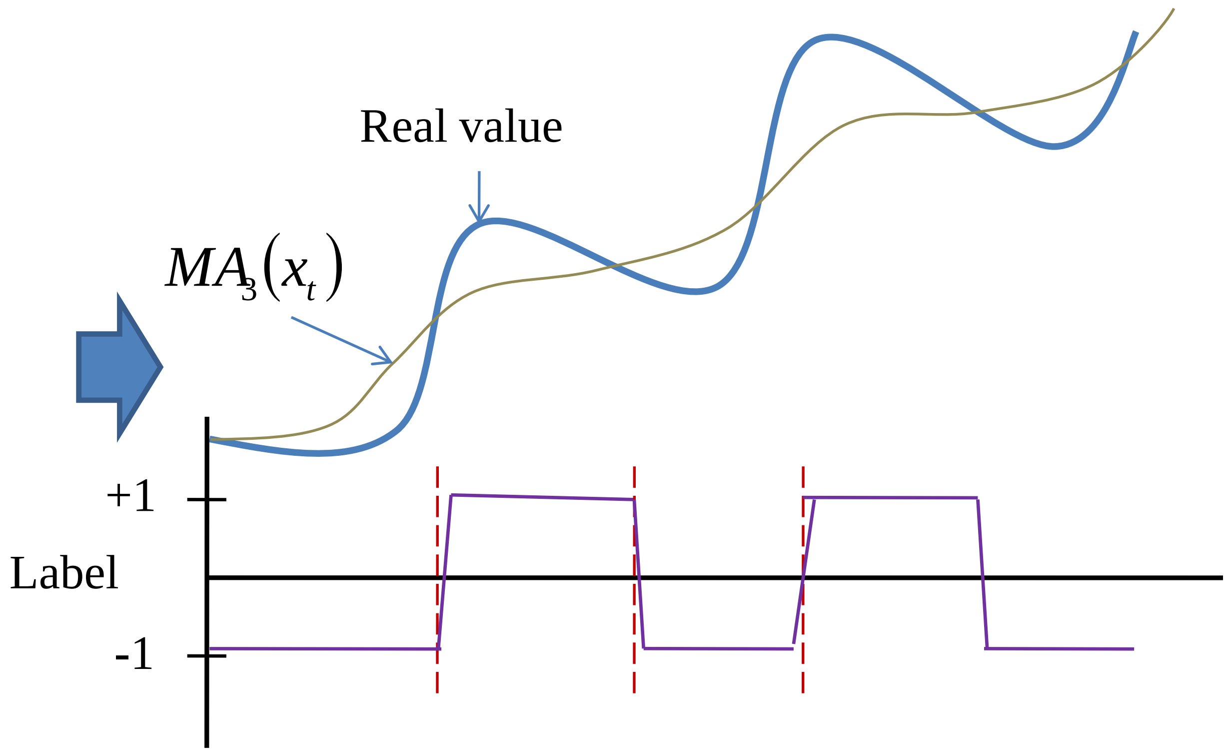
Unlabeled
✓ unknown



Method: Labeling, y

- The label corresponds to 'y' in the equation of SSL and the proposed label equation is as follows.

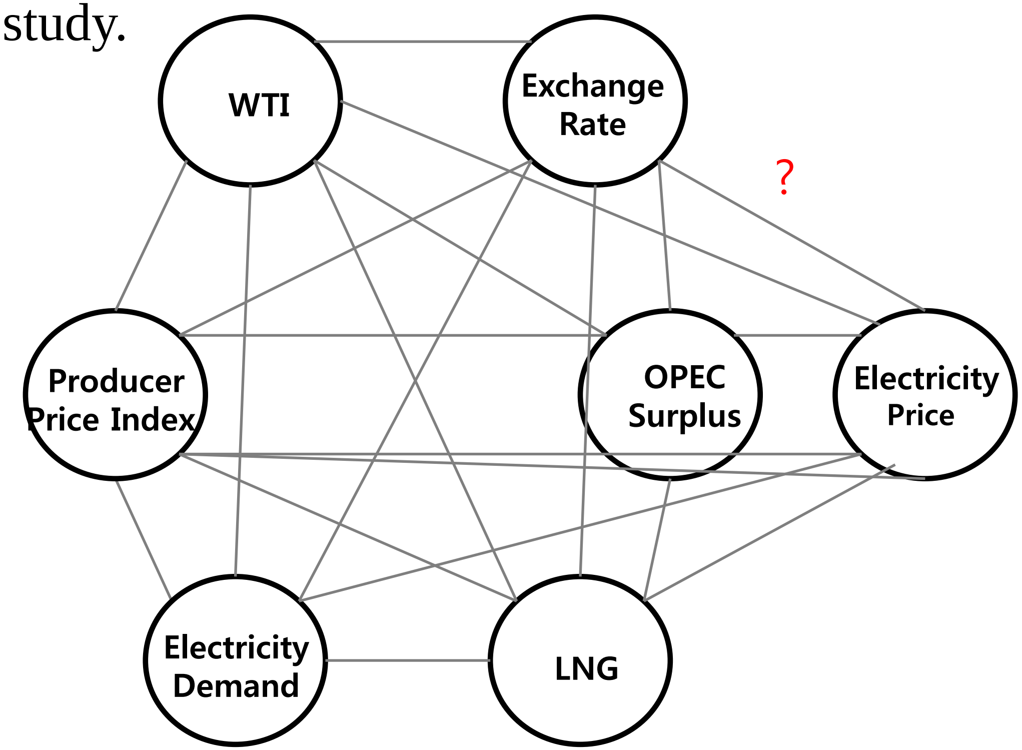
$$y = sign(x_t - MA_3(X_t))$$



Method: Similarity Matrix, W

Because the nodes in networks consist of time series data, it is not possible to apply a similarity measurement for the existing vectors.

Thus, a new method for measuring the similarity between time series data is proposed in this study.



Method: Similarity Matrix, W

- In the calculation method for the similarity, a labeling a for mentioned and a sign test are used.
- The sign test will be applied after marking the up/down(+1/-1) in each time series data for a specific period.
- Then, the absolute value of the kendall tau obtained from this test will be used as its similarity.

	1	2	3	4	5	6	7	8	9	10	11	12	**Kendal**
LNG	+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	

Method: Similarity Matrix, W

• For instance, if the MA3's up/down in the prices of LNG and electricity in the last year are given as follows, the similarity between them can be determined by the kendall tau of 0.7, which is obtained from the correlation test between two signs.

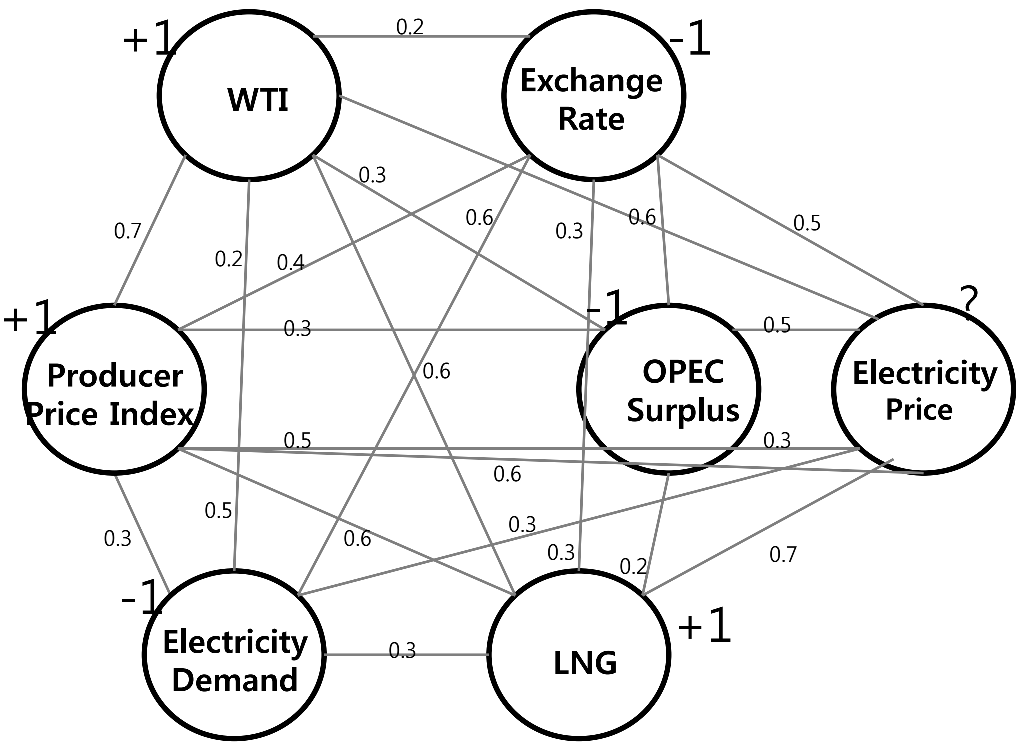
	1	2	3	4	5	6	7	8	9	10	11	12	**Kendal**
LNG	+1	-1	-1	+1	+1	+1	+1	-1	-1	-1	+1	-1	**0.7**
Electricity Price	+1	+1	-1	-1	+1	+1	+1	+1	-1	-1	+1	+1	

Method: Similarity Matrix, W

- Using this method, the similarities in 29 nodes for the given network can be measured.
- However, it is not necessary to update it every month.

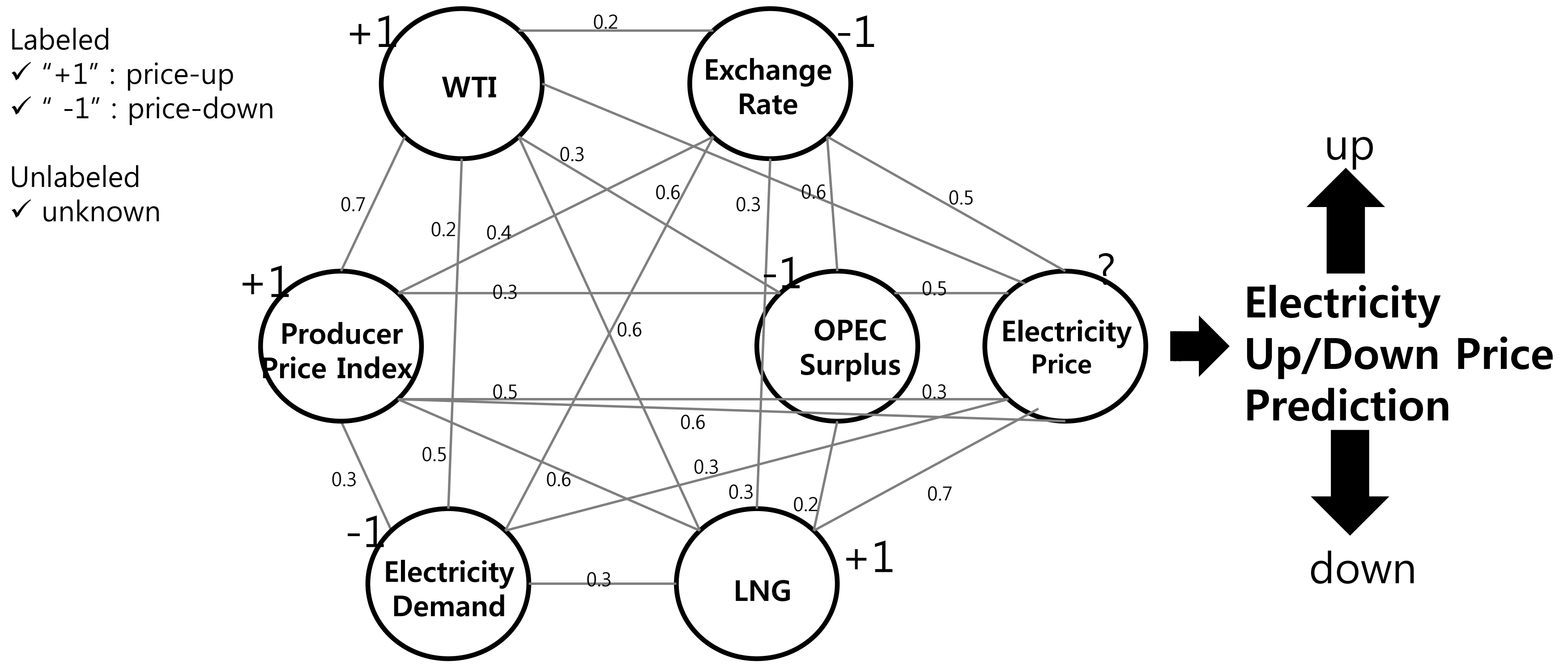
Labeled
✓ "+1" : price-up
✓ " -1" : price-down

Unlabeled
✓ unknown



Method: Similarity Matrix, W

As financial and economic networks are built, the prediction value of the electricity price can be obtained from these networks.

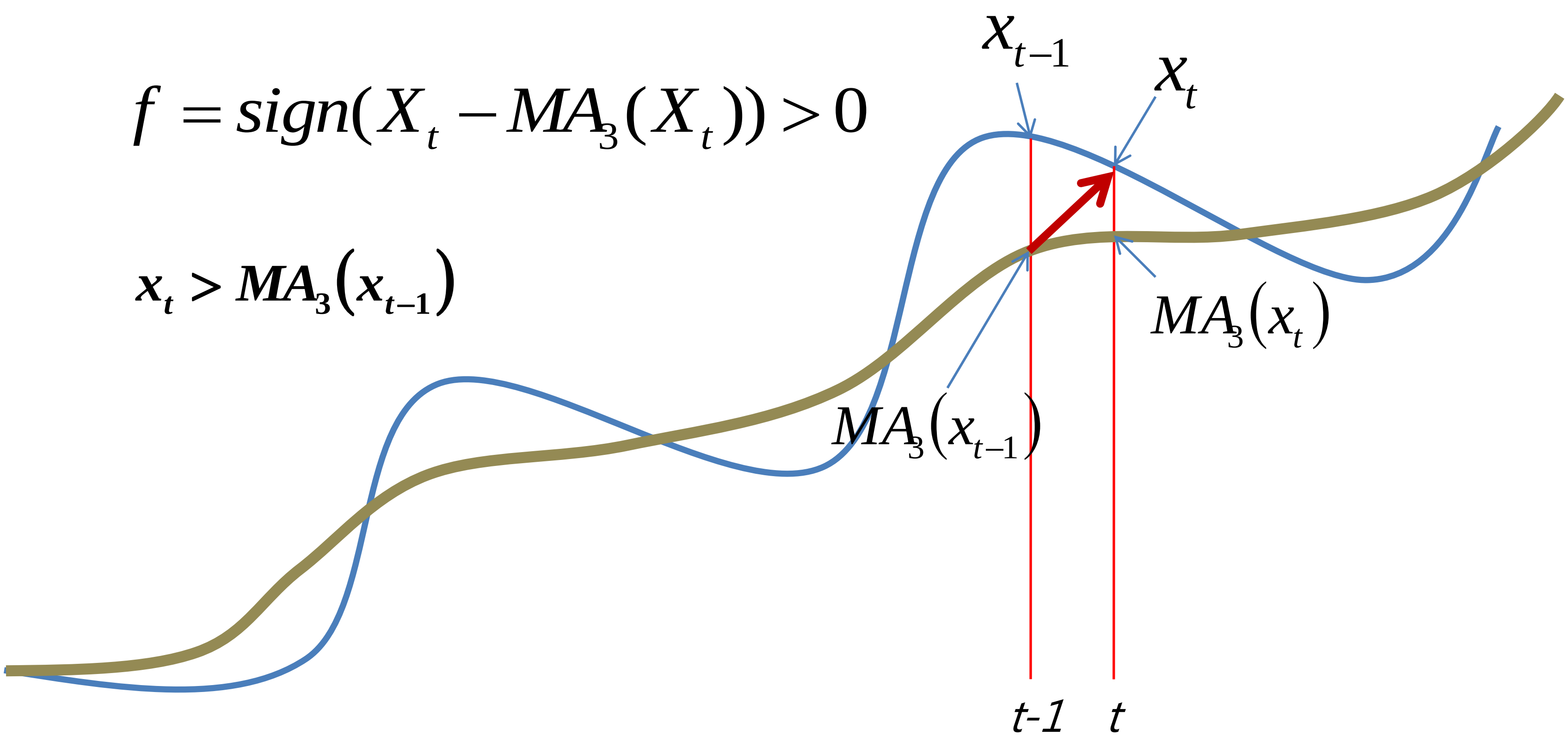


Method: Interpretation

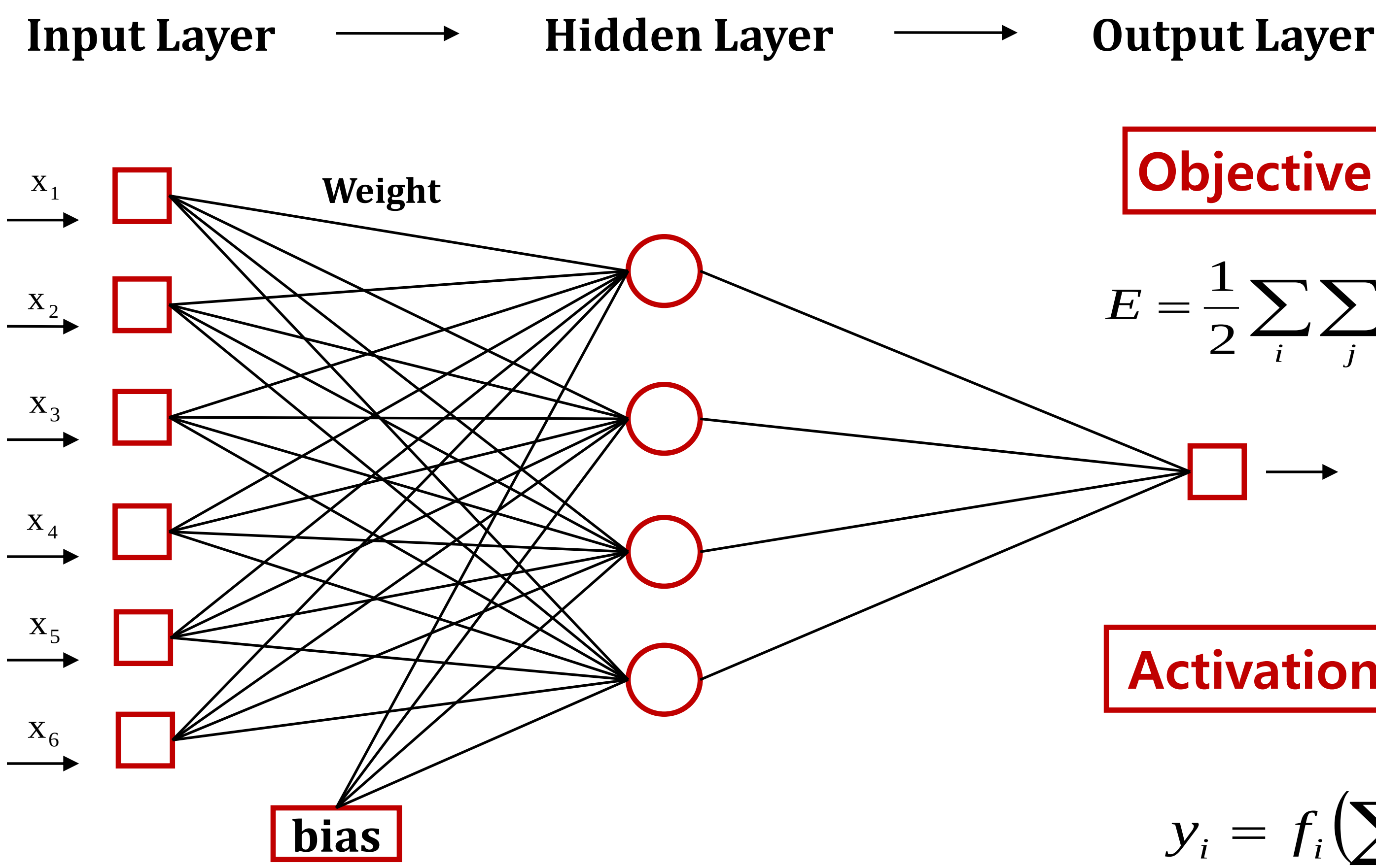
- The interpretation of the prediction value of the model is as follows.
  Label means, as it represents +1, it means that the next month price of electricity will be set by a higher price than the this month moving average.

$$f = sign(X_t - MA_3(X_t)) > 0$$

$$\boldsymbol{x_t > MA_3(x_{t-1})}$$



Method: Artificial Neural Network



Objective Function

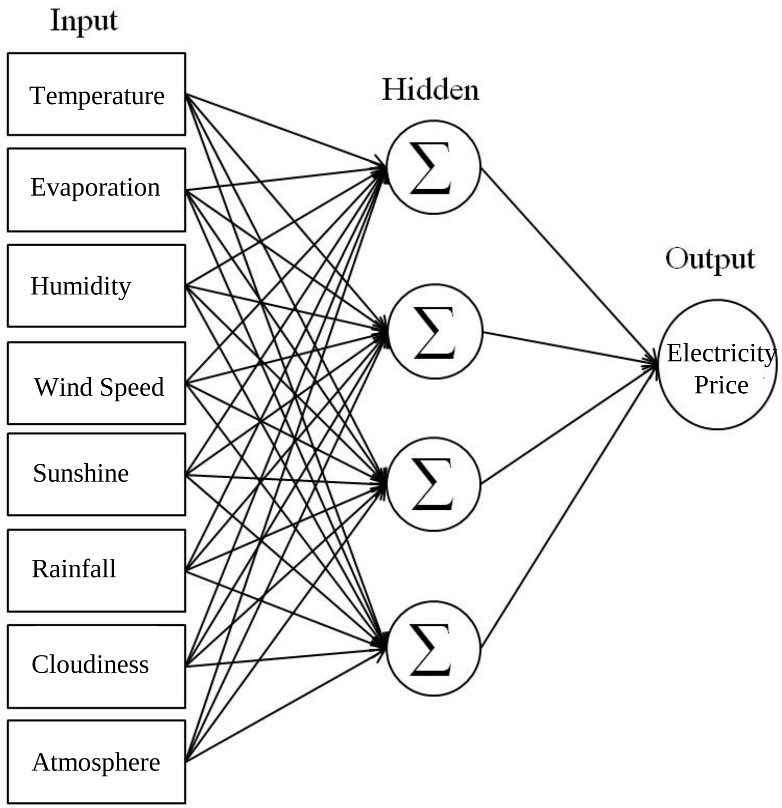
$$E = \frac{1}{2}\sum_{i}\sum_{j}\left(y_{ji} - Y_{ji}\right)^2$$

Activation Function

$$y_i = f_i\left(\sum_i \omega_{ij} X_i\right)$$

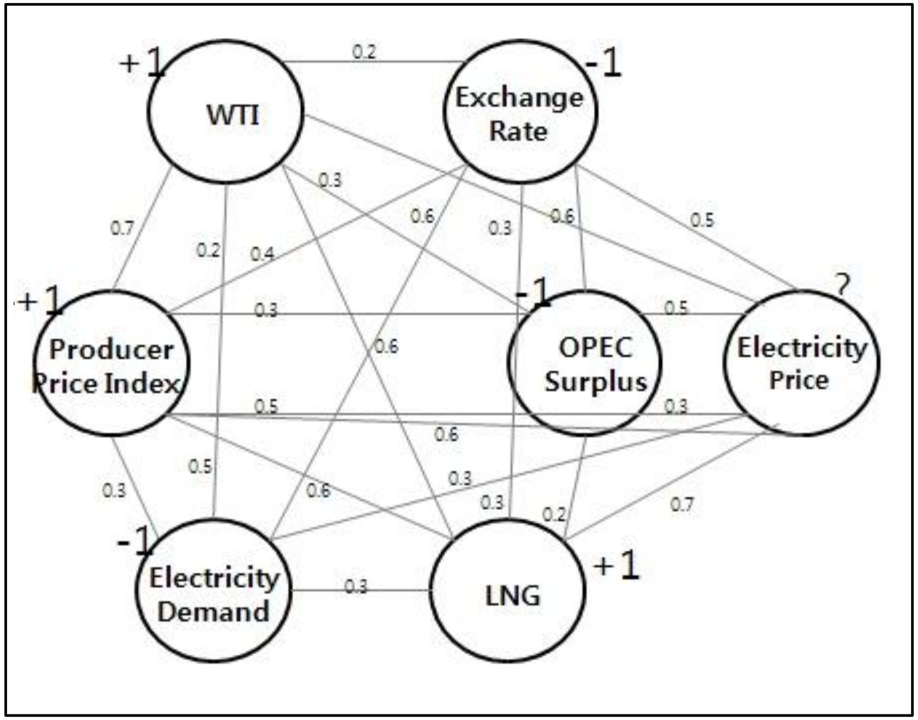
Method: ANN Model Composition

- Electricity price is used as an output variable for the ANN model, and the climatic variable is applied to an input variable.

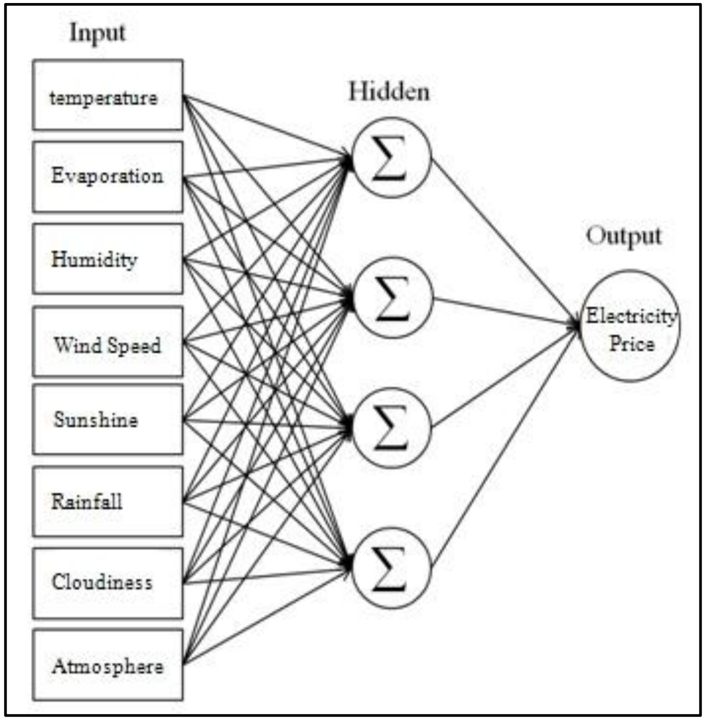


Method: SSL + ANN

A hybrid model(SSL+ANN) can be established by combining the prediction results of the up/down in the electricity price in SSL and the actual price of electricity predicted by using ANN.



+



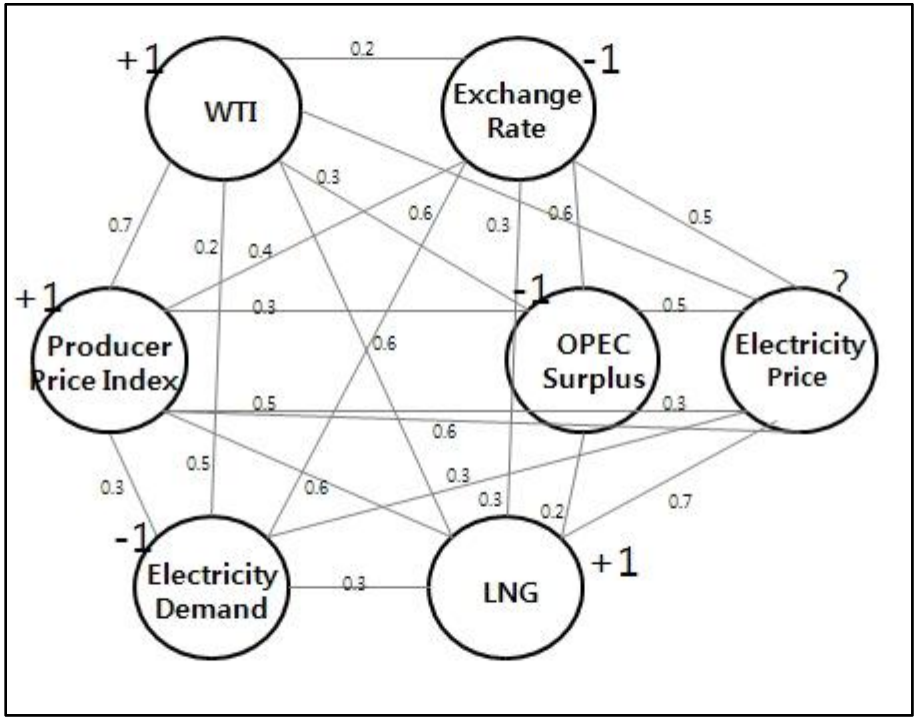
Method: SSL + ANN

- The definition of the output values for each model can be presented as follows:

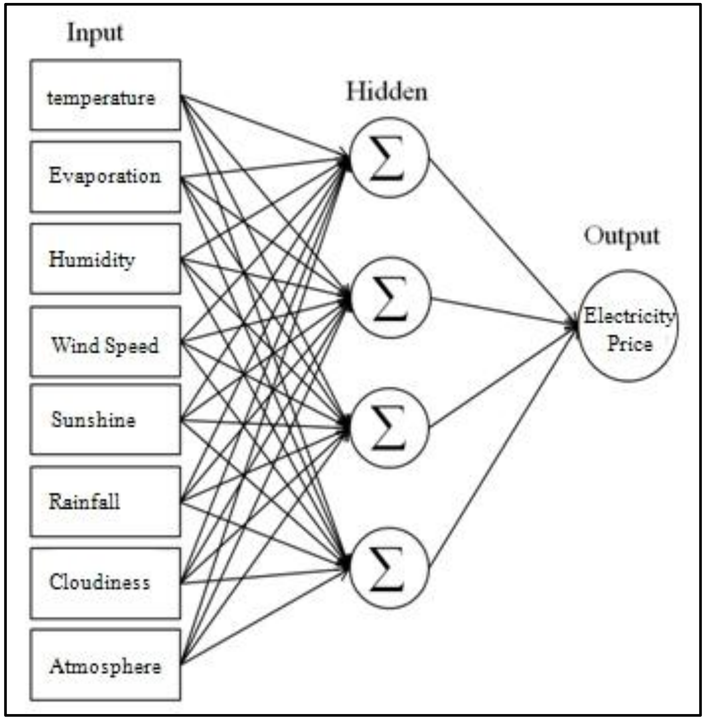
$y_t \in R$: y_t is the actual electricity price at that point of t

$f_{t+1}^{SSL} \in [-1,+1]$: It is the SSL prediction value at the point of t+1

$f_{t+1}^{NN} \in R$: It is the ANN prediction value at the point of t+1



+



Method: SSL + ANN

The final prediction value at the point of t+1 obtained using the hybrid model is as follows: $\hat{f}_{t+1}$

$$\hat{f}_{t+1} = \begin{cases} \boxed{f_{t+1}^{NN}} & if\ \ (f_{t+1}^{SSL})(f_{t+1}^{NN} - y_t) \geq 0 \\ y_t & otherwise \end{cases}$$

sign $(f_{t+1}^{SSL})(f_{t+1}^{NN} - y_t) \geq 0$: If there are some equality in the sign values between the SSL and ANN models, it will follow the previous case.

Method: SSL + ANN

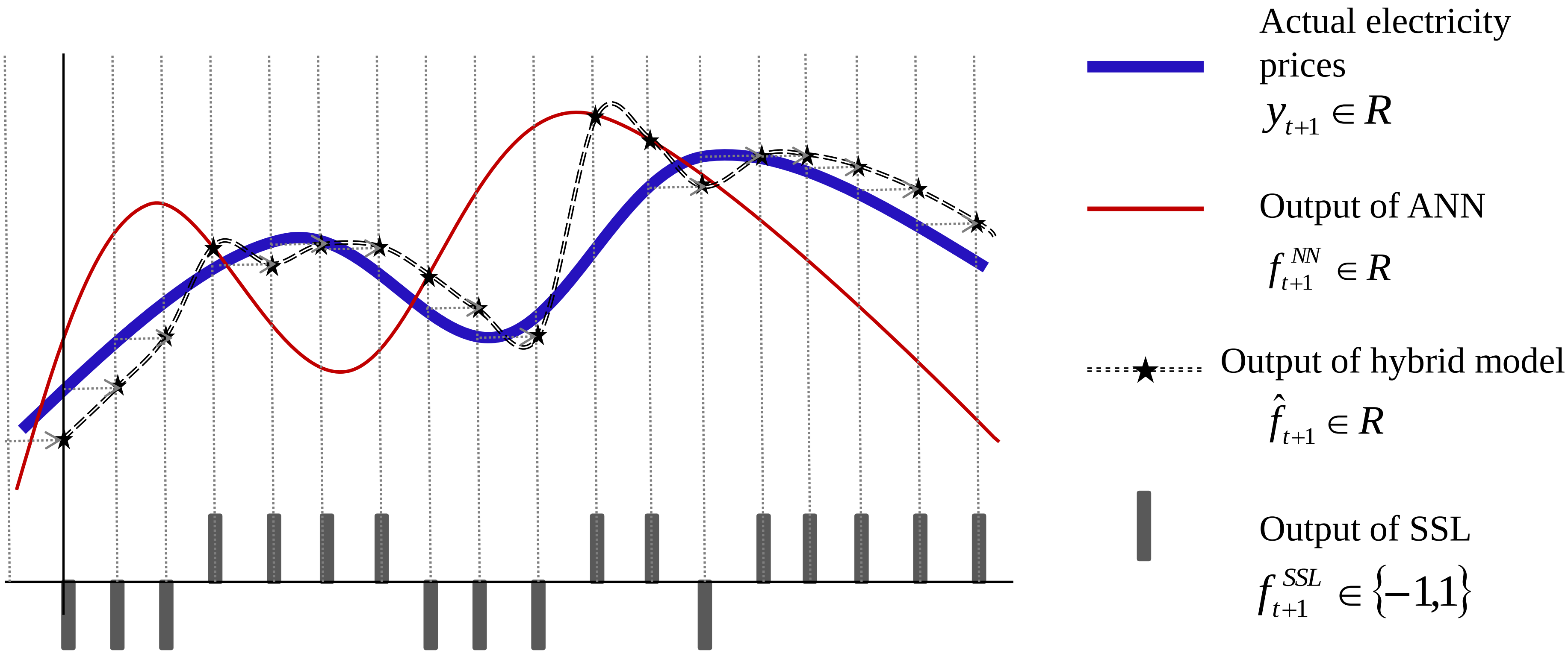
The final prediction value at the point of t+1 obtained using the hybrid model is as follows: $\hat{f}_{t+1}$

$$\hat{f}_{t+1} = \begin{cases} f_{t+1}^{NN} & if \ \ (f_{t+1}^{SSL})(f_{t+1}^{NN} - y_t) \geq 0 \\ \boxed{y_t} & otherwise \end{cases}$$

$\text{sign}\,(f_{t+1}^{SSL})(f_{t+1}^{NN} - y_t) < 0$: If there is complete different between two model, we will take current value.

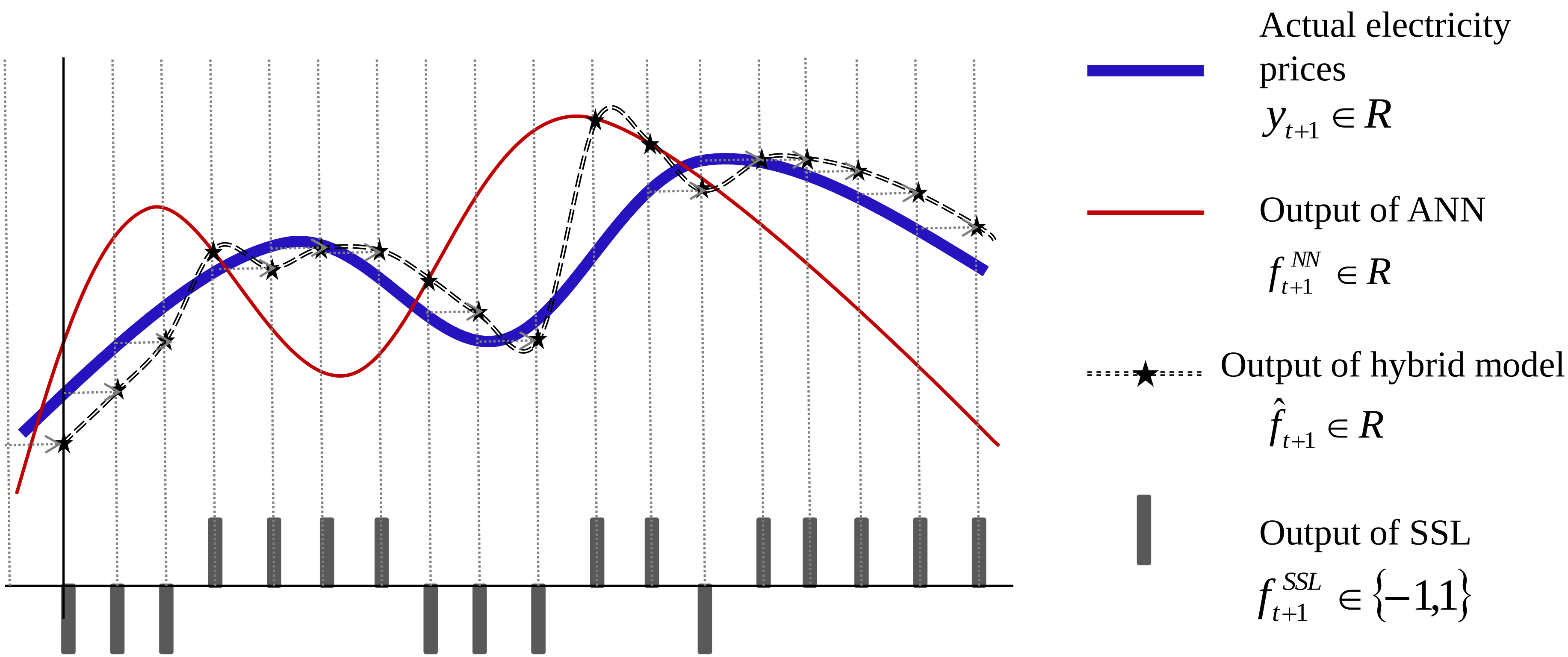
Method: SSL + ANN

• In this figure, the blue color shows the actual electricity prices and the red color shows the output of ANN.
• The gray bars show the output of SSL.
• The dotted line is the output of the hybrid model, which reflects the prediction value of ANN and the up/down prediction in SSL.



Method: SSL + ANN

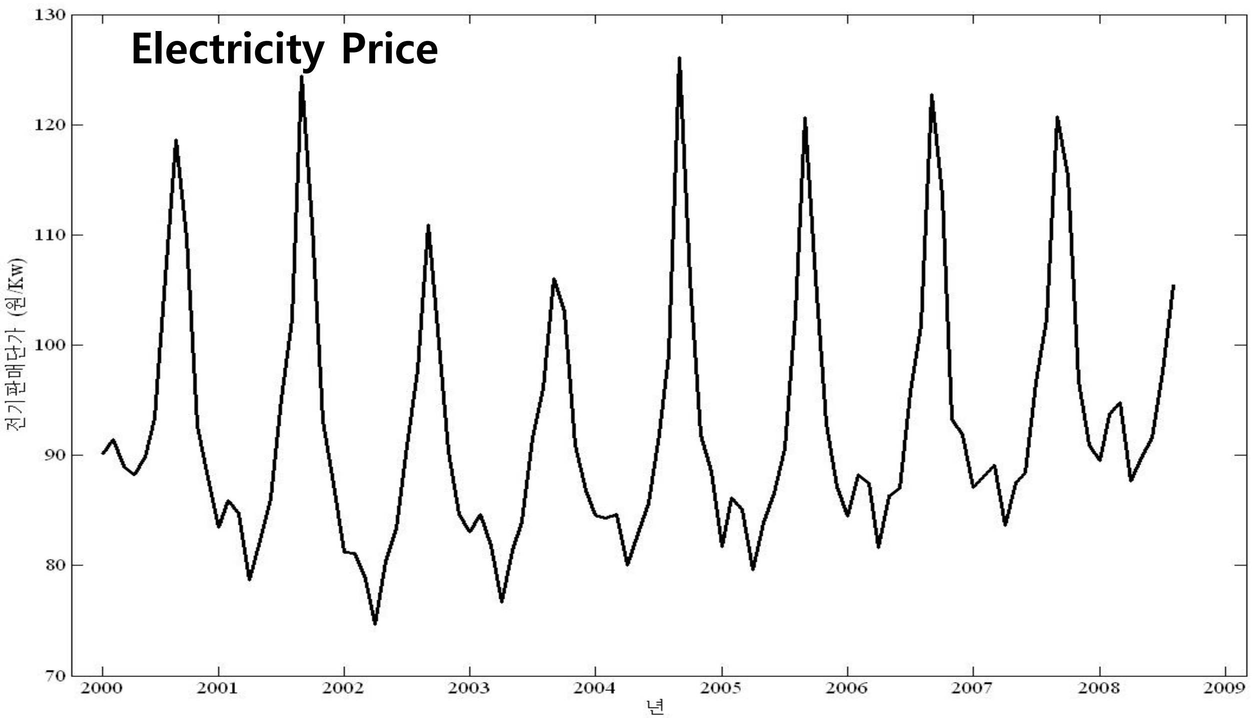
In the proposed combining method, it will change the final prediction value if the prediction value between these two models is different.
Also, it makes possible to perform more stable predictions.



Experiment & Result

Experiment: Data

- Period : 2000, 01 ~ 2008, 07
- Number of time point : Monthly data point
- Setting : Training Data – 2000, 01~ 2004, 07/ Validation Data – 2004, 08~ 2006, 07 / Test Data – 2006. 08~ 2008, 07



# of Data Point : 103

Experiment: Variable

✓ Input Variable

	Variable
Electricity related Factors	Electricity Demand
Financial/Economic Factors	West Texas Intermediate(WTI) Crude Oil Prices, Overall amount of world oil demand, amount of OECD demand, non-OECD demand, China demand, USA demand, OPEC production, Saudi production, Iran production, Iraq production, Kuwait production, non-OPEC production, USA production, Russia production, world production, producer price index, U.S. exchange rate, OECD commercial stockpiles, U.S. commercial stockpiles for crude oil, USA commercial stockpiles for oil, OPEC surplus production ability, NYMEX oil futures price, non-commercial real purchase (short), non commercial real purchase (long), commercial volume (short), commercial volume (long)
Climatic Factors	Temperature, Evaporation, Humidity, Wind Speed, Sunshine, Rainfall, Cloudiness, Atmosphere
Auto-Regression Factors (Lag period)	Electricity Price(t-1), Electricity Price (t-2), · · ·, Electricity Price (t-12),

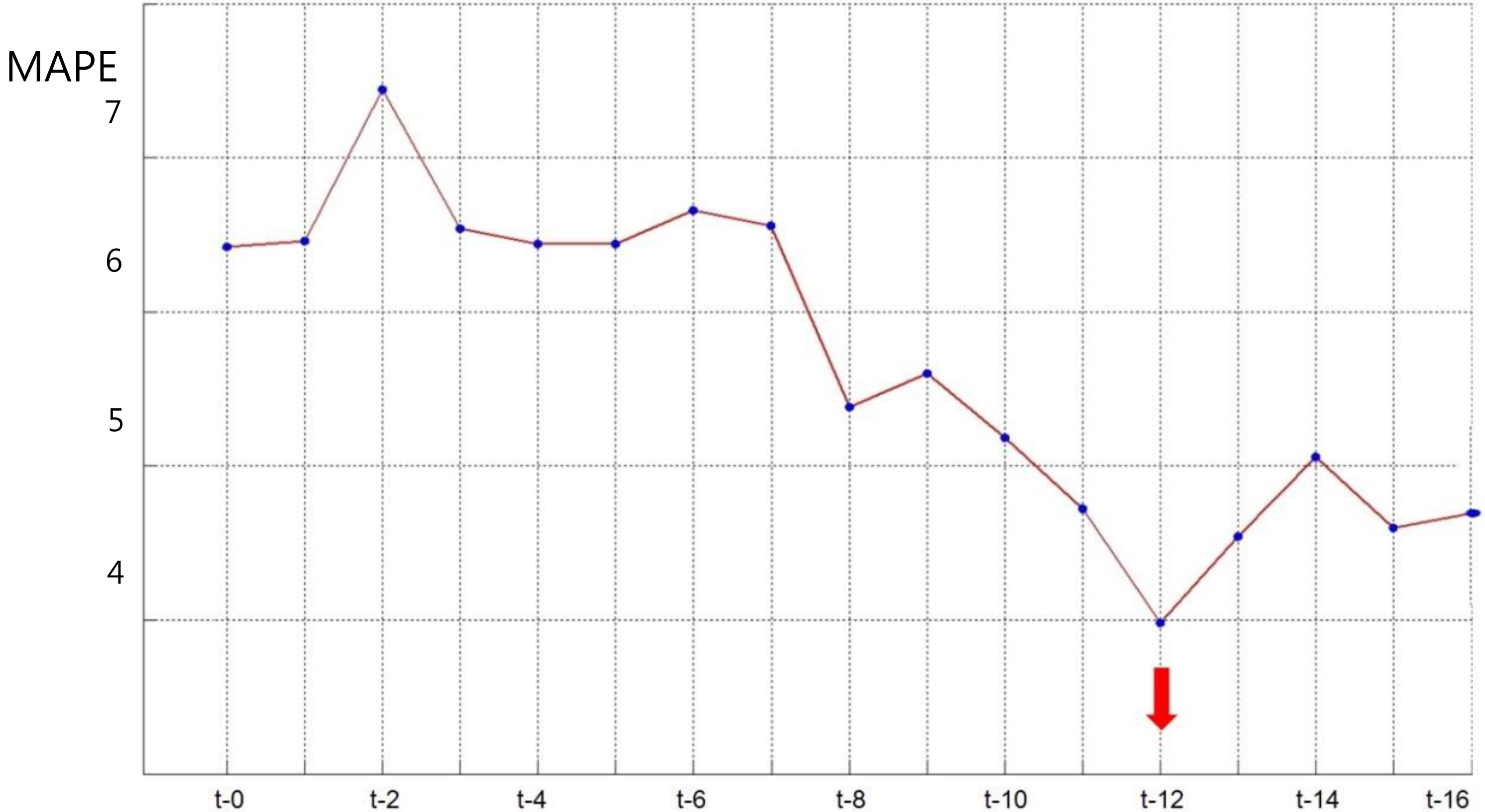
✓ Output Variable

	Variable
Power related Factors	Electricity Price

Experiment: Lag Period Selection

✓ Lag period of the auto-regression

- The lag period of the auto-regression represented the optimum values from the experiment that considers the point of t-18 including the point of t-12. Thus, the optimum order can be determined up to the point of t-12.



Experiment: Measurement

✓ MAPE(Mean Absolute Percentage Error)

$$MAPE = \frac{1}{n}\sum_{i}\left|\frac{y_i - f_i'}{y_i}\right| \times 100$$

Experiment: Hybrid Model Variable

✓ SSL Model

- The SSL model uses a total of 29 variables including 26 financial and economic variables, electricity prices, electricity demands, and LNG prices.

		Variable
Output variable		Electricity Price
Input variable	Electricity related Factors	Electricity Demand, LNG
	Financial/Economic Factors	West Texas Intermediate(WTI) Crude Oil Prices, Overall amount of world oil demand, amount of OECD demand, non-OECD demand, China demand, USA demand, OPEC production, Saudi production, Iran production, Iraq production, Kuwait production, non-OPEC production, USA production, Russia production, world production, producer price index, U.S. exchange rate, OECD commercial stockpiles, U.S. commercial stockpiles for crude oil, USA commercial stockpiles for oil, OPEC surplus production ability, NYMEX oil futures price, non-commercial real purchase (short), non commercial real purchase (long), commercial volume (short), commercial volume (long)

Experiment: Hybrid Model Variable

✓ ANN Model

- The ANN model uses a total of 8 climatic variables, auto-regression variables and the electricity prices.

		Variable
Output variable		Electricity Price
Input variable	Climatic variable factors	Temperature, Evaporation, Humidity, Wind Speed, Sunshine, Rainfall, Cloudiness, Atmosphere
	Auto-regression variable factors	Electricity Price(t-1), Electricity Price (t-2), · · ·, Electricity Price(t-12),

Experiment: Model Parameters Selection

- Semi-Supervised Learning(SSL) model parameter selection

K={3, 4, 5, 6, 7, 8, 9, 10, 15, 20} , μ(c1)={0.01, 0.05, 0.07, 0.1, 0.5, 0.75, 1, 10, 100}

- Best parameter combination K={15}, μ={0.01}

- Artificial Neural Network(ANN) model parameter selection

Hidden node= {1, 2, 3, 4, 5, 6, 7}

- Best parameter combination Hidden node= {3}

Experiment: Comparison Model

✓ For verifying the effectiveness of the hybrid model, the following three models were applied and compared.

- **ANN_A:** Electricity related Factors(2), Financial/Economic Factors(27), Climatic variables(8), Auto-regression variables(Lag period, 12)- total 49

- **ANN_B:** Climatic variables(8) and Auto-regression variables(12)- total 20.

- **SSL+ANN (Hybrid Model):** The hybrid model uses a total of 49 variables as the same as the ANN_A model and applies these variables to the SSL and ANN models separately.

Experiment: ANN_A Model vs. Hybrid Model

Input variable \ Comparison	ANN_A	ANN_B	HYBRID	
			SSL	ANN
Electricity related Factors (2)	●		●	
Financial/Economic Factors(27)	●		●	
Climatic variables (8)	●	●		●
Auto-regression variables(12)	●	●		●

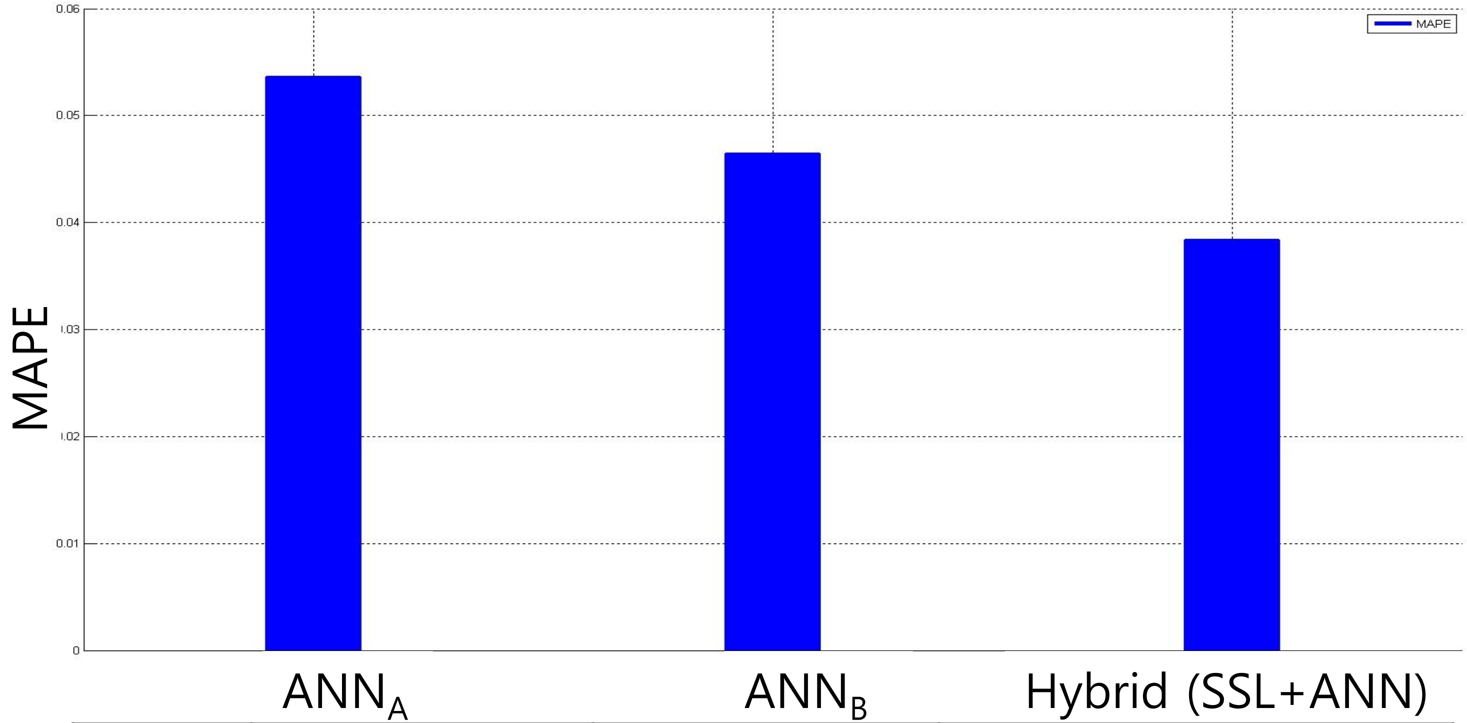
First, the comparison was applied to the ANN_A and Hybrid models. The same input variables were used in this comparison in which the effectiveness can be verified by separating these variables in the SSL and ANN models due to the characteristic of the hybrid model.

Experiment: ANN_A Model vs. Hybrid Model

Input variable \ Comparison	ANN_A	ANN_B	HYBRID	
			SSL	ANN
Electricity related Factors (2)	●		●	
Financial/Economic Factors(27)	●		●	
Climatic variables (8)	●	●		●
Auto-regression variables(12)	●	●		●

In the comparison of the ANN_B and Hybrid models, the effectiveness of economic and financial variables can be verified through comparing these two models.

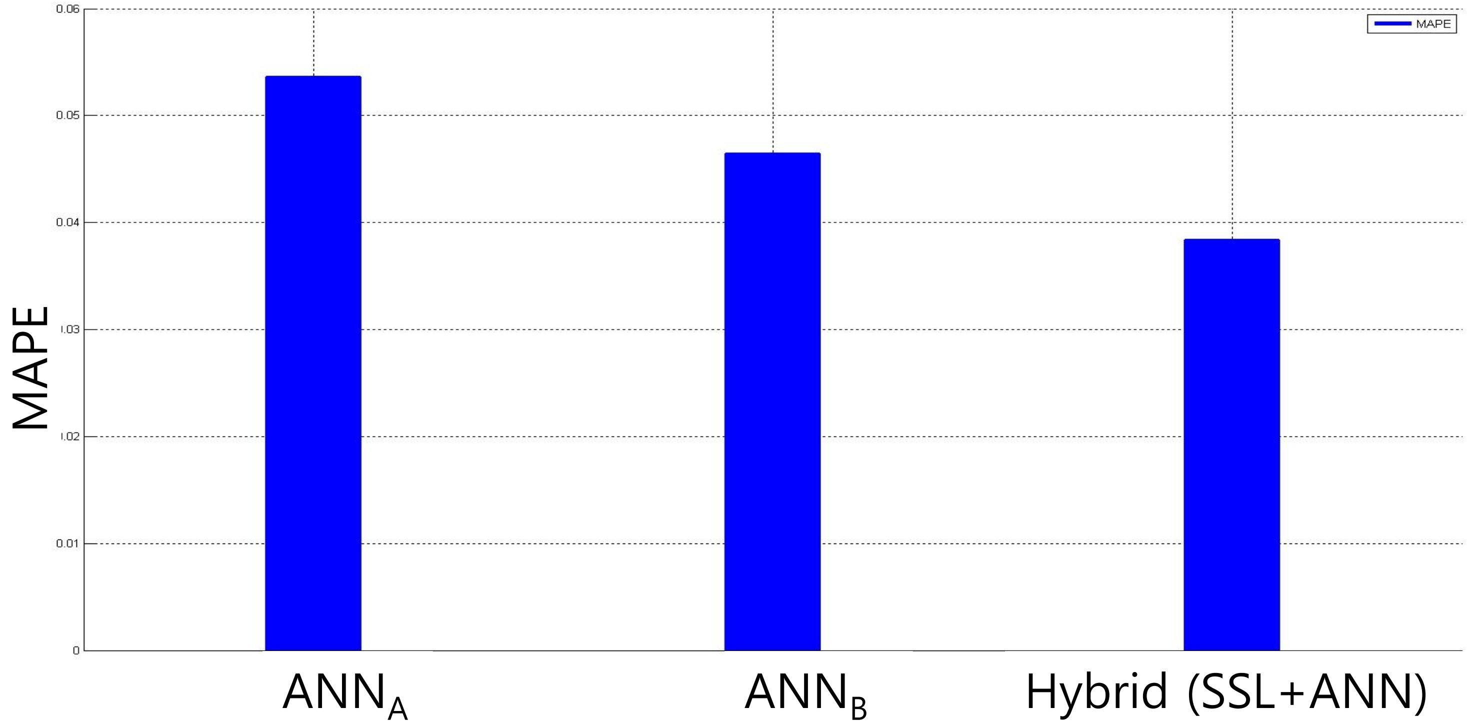
Result



Model	**ANN_A**		**ANN_B**		**Hybrid(SSL+ANN)**	
	validation	test	validation	test	validation	test
MAPE	6.45	5.35	6.02	4.64	**5.29**	**3.83**

In the results of the experiments using three models, MAPE can be determined as follows.

Result



Model	ANN_A		ANN_B		**Hybrid(SSL+ANN)**	
	validation	test	Validation	Test	Validation	Test
MAPE	6.45	5.35	6.02	4.64	**5.29**	**3.83**

The hybrid model among these three models showed the most excellent performance.

Result

Model	ANN_A		ANN_B		Hybrid(SSL+ANN)	
	validation	test	Validation	Test	Validation	Test
MAPE	6.45	5.35	6.02	4.64	**5.29**	**3.83**

In the comparison of the hybrid model and ANN_A, although these two models apply the same input variables, the hybrid model represented a higher accuracy than that of ANN_A. Because the hybrid model use these variables separately.

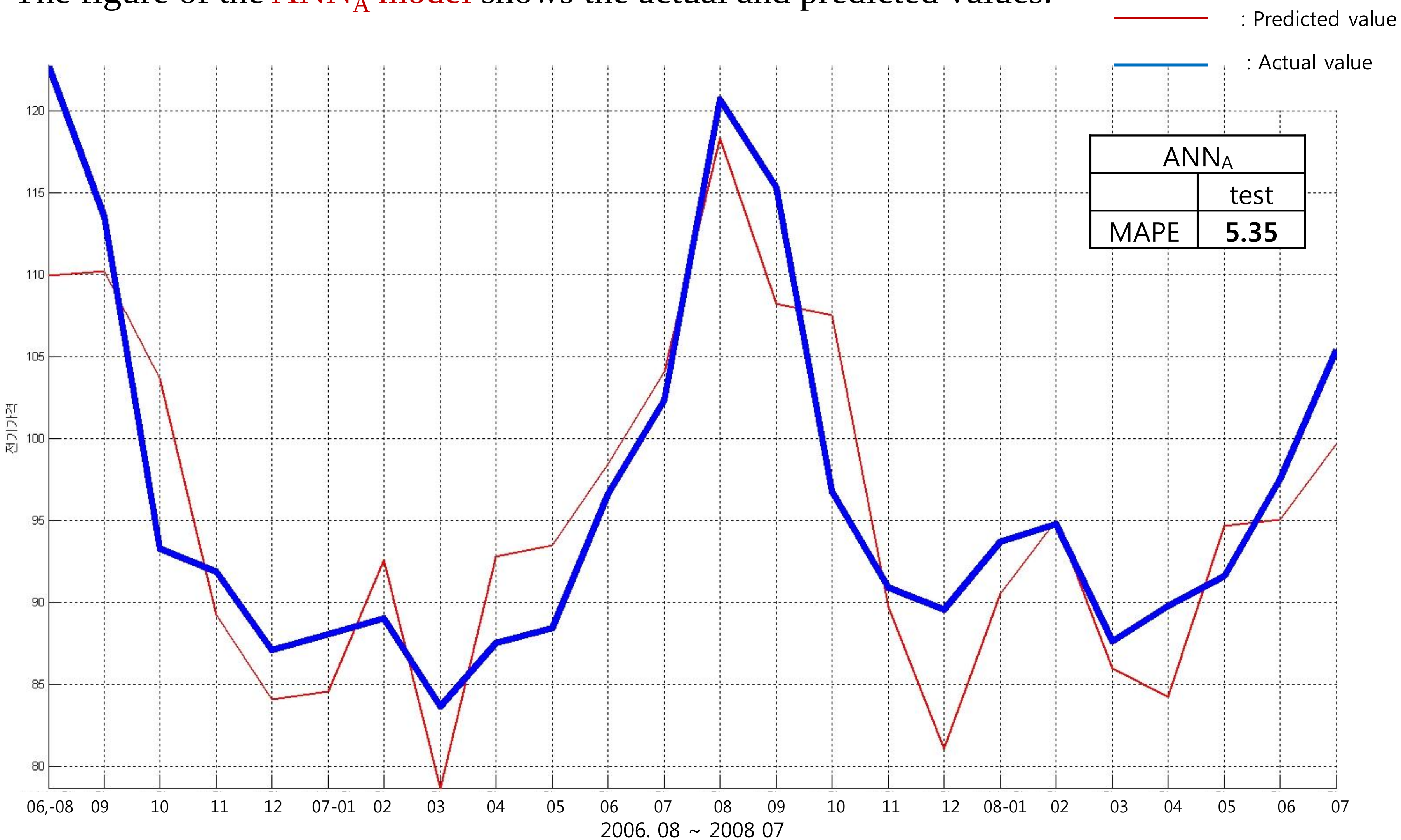
Result

Model	ANN_A		ANN_B		Hybrid(SSL+ANN)	
	validation	test	Validation	Test	Validation	Test
MAPE	6.45	5.35	6.02	4.64	**5.29**	**3.83**

In the comparison of the hybrid model and ANN_B, it revealed that the information obtained from financial/economic variables affect the prediction of the electricity price in the ANN.

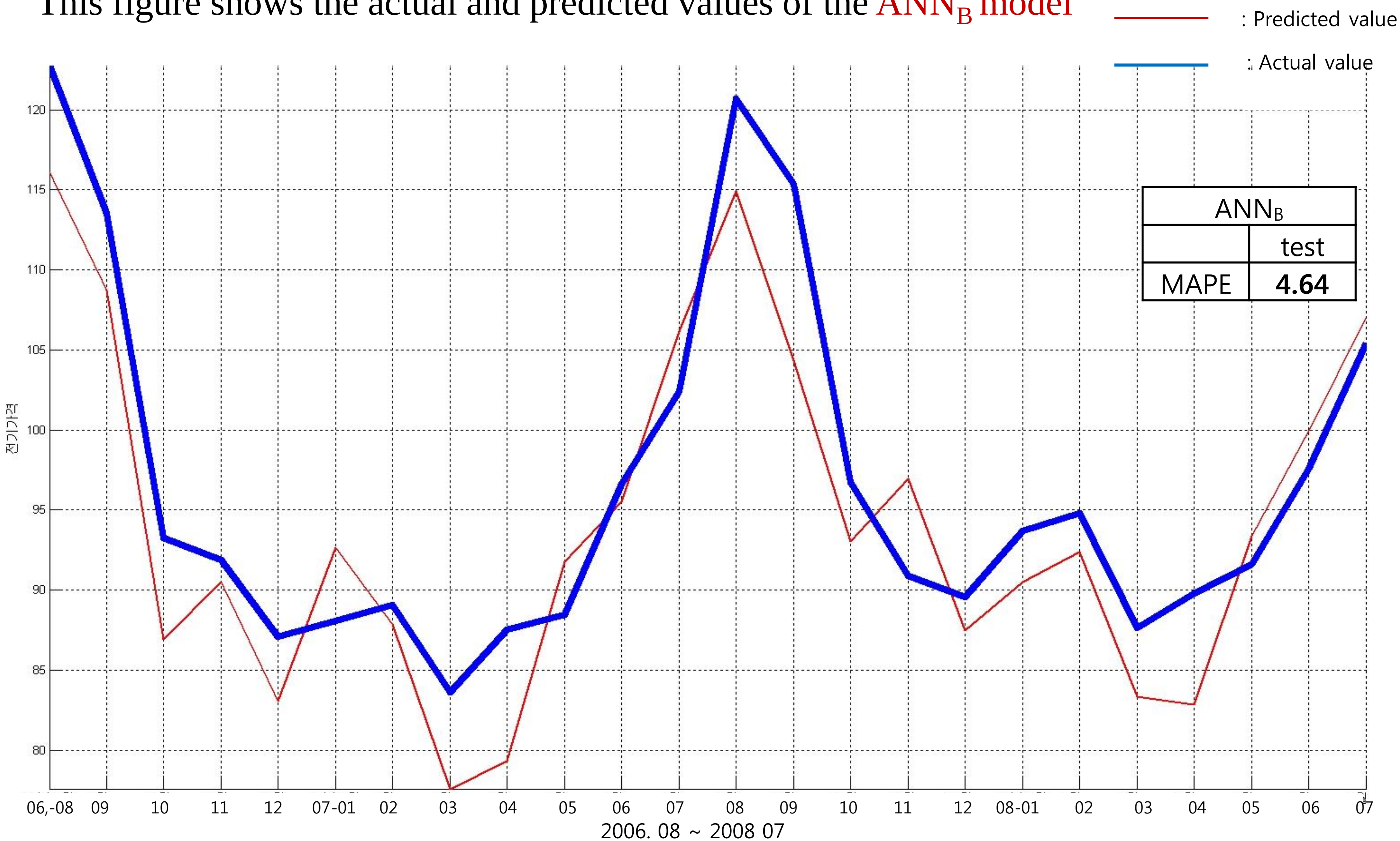
Result

The figure of the ANN$_A$ model shows the actual and predicted values.



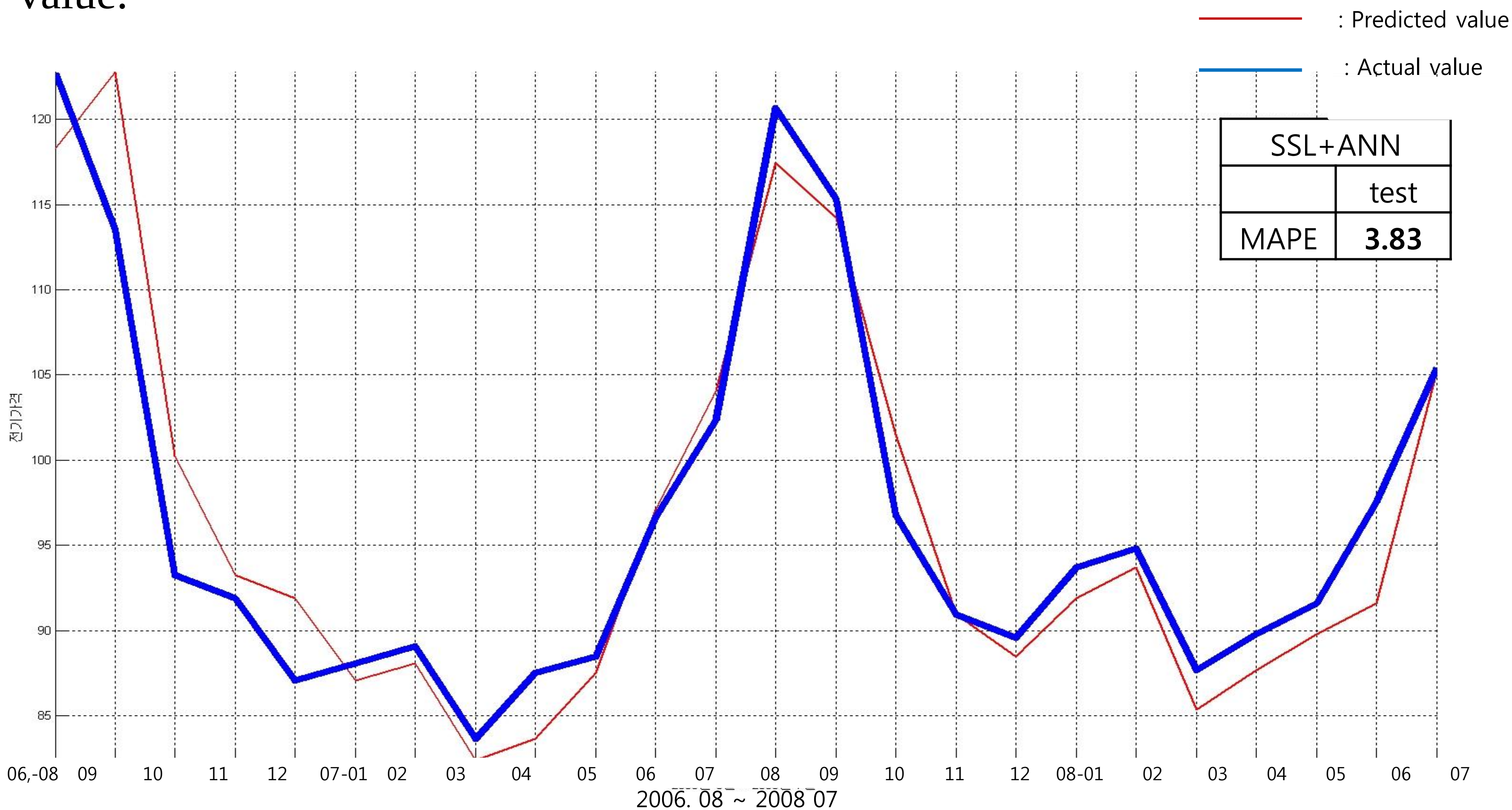
Result

This figure shows the actual and predicted values of the ANN$_B$ model



Result

Finally, this figure shows the prediction values of the hybrid model.
As shown in this figure, predicted value of hybrid model is most similar to actual value.



SSL+ANN	
	test
MAPE	**3.83**

Conclusion

Conclusion

- ✓ In this study, for predicting electricity prices, the most advanced algorithm, SSL (Semi-Supervised Learning), in the fields of data mining and machine learning, ANN (Artificial Neural Network) were used.

Conclusion

- ✓ As a result, the most excellent accuracy was obtained from the proposed hybrid model, which combines SSL and ANN, in which MAPE was 3.83%.
- ✓ Thus, it was verified that the proposed model represents an improvement in accuracies at least 0.8% to 3.2% compared to other models.
- ✓ Based on this performance, it revealed that the hybrid model applies input variables separately to SSL and ANN but obtains the information of up/down movement obtained from financial/economic variables from SSL and that shows an effectiveness in correcting prediction errors in the electricity prices.

Conclusion

- ✓ The proposed model shows excellent performances by reflecting the financial/economic information, electricity demands and related information, and climatic information to the prediction of the electricity prices.

- ✓ Also, it can be considered that the proposed model represents a possibility that agrees to the knowledge of domain experts in its methods and results.

Reference

Reference

- Dong-Seok Kong, Young-Hoon Kwak, Jung-Ho Huh, “Artificial Neural Network Based Energy Demand Prediction for the Urban District Energy Planning”, Journal of the Architectural Institute of Korea, Vol.26, No.2, pp.221~230, 2010.
- Bong-Woo Nam, Kyung-Bin Song, Kyu-Ho Kim, Jun-Min Cha, “The Spatial Electric Load Forecasting Algorithm using the Multiple Regression Analysis Method”, Journal of the Korean Institute of Illuminating and Electrical Installation Engineers, Vol.22, No.2, pp.63~70, 2008.
- Young-Min Wi, Guk-Hyun Moon, Jae-Hee Lee, Sung-Kwan Joo, Kyung-Bin Song, “Load Forecasting for the Holidays using a Polynomial Regression Incorporating Temperature Effect”, The Korean Institute of Electrical Engineers, pp.29~30, 2007
- Jin-Hee Kim, Je-Gon Lee, Jun-Min Cha, “Load forecasting and demand management considering with renewable energy”, The Korean Institute of Electrical Engineers, pp.2260~2261, 2009.
- Nak-Hoon Choi, Kwang-Myung Son, Tae-Gee Lee, “Daily peak load forecasting considering the load trend and temperature”, Journal of the Korean Institute of Illuminating and Electrical Installation Engineers, Vol.6, No.6, pp.35~42, 2001.
- Ja-Ho Shin, Yeon-Chan Hong, “GMDH Algorithm with Data Weighting Performance and Its Application to Power Demand Forecasting”, journal OF Control Automation and Systems Engineering Vol.12, No.7, 2006.
- Jin Yang, Hugues Rivard, Radu Zmeureanu, “On-line building energy prediction using adaptive artificial neural networks”, Energy and Buildings, Vol.37, Issue 12, pp.1250~1259, 2005.
- Ping-Feng Pai, Wei-Chiang Hong, “Support vector machines with simulated annealing algorithms in electricity load forecasting”, Energy Conversion and Management, Vol.46, Issue 17, pp.2669-2688, 2005.
- S.Mirasgedis, Y.Sarafidis, E.Georgopoulou, D.P. Lalas, M.Moschovits, F. Karagiannis, D. Papakonstantinou, “Models for mid-term electricity demand forecasting incorporating weather influences”, Energy, Vol.31, Issue 2-3, pp.208~227, 2006.
- E. Gonzalez-Romera, M.A. Jaramilo-Moran, D. Carmona-Fernandez, “Monthly electric energy demand forecasting with neural networks and Fourier series”, Energy Conversion and Management, Vol.49, Issue.11, pp.3135~3142, 2008.