

## 신경망, 시계열 분석 및 판단보정 기법을 이용한 교통량 예측 (Traffic-Flow Forecasting using ARIMA, Neural Network and Judgment Adjustment)

장석철, 석상문, 이주상, 이상욱, 안병하  
광주 북 오룡 1 광주과학기술원 기전공학과  
{stniron, soakbong, jusang, yashin96, bayhay@gist.ac.kr}

### Abstract

During the past few years, various traffic-flow forecasting models, i.e. an ARIMA, an ANN, and so on, have been developed to predict more accurate traffic flow. However, these models analyze historical data in an attempt to predict future value of a variable of interest. They make use of the following basic strategy. Past data are analyzed in order to identify a pattern that can be used to describe them. Then this pattern is extrapolated, or extended, into the future in order to make forecasts. This strategy rests on the assumption that the pattern that has been identified will continue into the future. So ARIMA or ANN models with its traditional architecture cannot be expected to give good predictions unless this assumption is valid; The statistical models in particular, the time series models are deficient in the sense that they merely extrapolate past patterns in the data without reflecting the expected irregular and infrequent future events Also forecasting power of a single model is limited to its accurate.

In this paper, we compared with an ANN model and ARIMA model and tried to combine an ARIMA model and ANN model for obtaining a better forecasting performance. In addition to combining two models, we also introduced judgmental adjustment technique. Our approach can improve the forecasting power in traffic flow. To validate our model, we have compared the performance with other models. Finally we prove that the proposed model, i.e. ARIMA + ANN + Judgmental Adjustment, is superior to the other model.

### 1. 서론

지난 수년간 보다 정확한 교통량을 예측하기 위하여 다양한 교통량 예측 모델들이 개발되어졌다. 특히 시간에 따라 순차적으로 발생하는 교통데이터들을 처리하기 위하여 많은 연구자들은 시계열 분석법에 의한 교통량 예측을 시도하였다. 최근에는 컴퓨터 성능의 고도화되면서 통계적 기반의 시계열분석(Autoregressive integrated moving average: ARIMA) 모델, 칼만필터링(Kalman filtering) 모델, 신경망(Artificial neural network: ANN) 모델들이 제시되고 있다.

특히 통계적기법 중에서 ARIMA 모델은 우수한 성능을 나타내는 예측모델로 단일변수 예측시 가장 작은 평균 제곱근오차(Mean square error: MSE)를 나타낸다. 그러나 이러한 우수성도 과거데이터가 선형적인 특성을 지니고 있는 경우에 해당되고, 비선형적 특성이 있는 경우에는 그 성능을 보장할 수 없다. 비선형적 특성을 나타내는 데이터의 경우 비선형회귀 기법이 존재하지만 많은 계산량을 요

구하기 때문에 실용적이 못한 단점이 있다.

1990년대 이후로 ANN 모델은 비선형 예측에서 우수한 성능을 입증하면서 각광을 받기 시작했다. 많은 전문가들은 ANN 모델을 추가예측, 신호예측 등 다양한 분야에 적용하여 전통적인 통계 모델보다 미래데이터를 예측하는 능력에 있어서 우수한 성능을 나타냄을 입증하였다. 또한 3계층(three-layer)의 전방결선(feed-forward)으로 이루어진 ANN으로 중간층(hidden layer)이 시그모이드(sigmoidal) 함수로 처리될 경우 실계수의 다중변수 함수에 대해서도 높은 예측성능을 나타낸다는 것이 수학적으로 증명되었고 [1][2], 비모수회귀 능력을 가지고 있음이 입증되었다[3].

그러나 대부분 이러한 예측 모델들은 과거 교통량을 분석하여, 그 패턴을 파악하고, 이를 외삽 및 확장하여 교통량을 예측하였다. 이러한 분석법은 과거의 교통 패턴이 미래에도 연장된다는 가정 하에 이루어지는 것이기 때문에 불규칙적인 교통 패턴은 예측의 신뢰성을 떨어뜨렸다. 특히, 시계열 분석법과 같은 통계적 기법은 단순히 과거 패턴을 외삽하는 것이므로 불규칙 변동에 적당하게 대처하지 못하였다[4]. 또한 ANN이나 ARIMA와 같은 단일 모델을 이용한 예측은 그 정확도 면에서 어느 정도의 한계를 내포하고 있는 실정이다.

우리는 위와 같은 불규칙적 패턴을 해결하기 위하여 판단보정기법(Judgmental adjustment: JA)을 도입하였다. 판단보정기법은 시계열분석에서 간과되거나 과거데이터의 외삽으로는 확인할 수 없는 요소들을 보정해주는 기법이다[5]. 예를 들어 데이터의 갑작스러운 변화가 생길 경우 예측담당자의 경험적 판단에 따라 모델의 예측결과를 보정하게 되는데 이것은 실생활에서 널리 사용되는 방법이다. 판단보정기법이 적용되는 요소는 다음과 같은 특징을 가지고 있다.

- 1) 데이터의 변화가 불규칙적이고 가끔 일어나는 경우로 이를 모델화하기에는 데이터의 양이 너무 적으나 결과에 미치는 효과는 매우 큼
- 2) 데이터의 변화가 일시적임
- 3) 데이터의 변화를 어느 정도 예측가능

본 논문에서는 가장 대표적인 예측 기법인 ANN 모델과 ARIMA 모델의 성능을 비교하고 나아가 정도 높은 예측 결과를 얻기 위하여, 두 기법의 통합을 시도 하였다. 또한 불규칙 변동에 대처하기 위하여 판단보정기법을 도입하고 각각의 모델들은 성능비교를 통하여 유효성을 검증하였다.

### 2. 실험방법

일반적으로 교통데이터라 함은 차량속도, 차량대기길이, 시간당 차량유입대수 등을 포함하지만 본 논문에서는

차량속도만을 다루기로 한다. 왜냐하면 차량속도를 통하여 다른 요소들은 어느 정도 예측가능하기 때문이다.

우리는 교통데이터를 광주 동문로에서 1999년 8월 15일 오전 9시 25분부터 8월 20일 새벽 0시 55분까지 5분 간격으로 총 1339개의 데이터를 얻어 실험을 하였다. 그림 1은 우리가 실험 데이터를 얻은 도로의 지도이다. 여기서 두 지점사이의 거리는 10km로 차량으로 운행시 약 5-10분 정도의 시간이 소요되며, 평균차량 속도는 약 40km/hour로 측정되었다.



그림 1 실험 데이터를 얻은 동문로 구간의 지도

그림2는 하루 동안 동문로를 지나는 차량의 평균속도를 나타낸다. 이 그림을 통해서 우리는 이 구간에서의 교통데이터의 특징을 파악할 수 있다. 즉 낮시간(07:00-19:00)대에는 이동차량의 증가로 인한 교통체증 때문에 평균 20km/hour이하의 차량속도를 나타내지만 이동 차량이 거의 없는 새벽시간(00:00-06:00)대에는 운전자에 운전 습관에 따라 50-140km/hour로 다양한 차량속도를 나타내고 있다. 그러나 새벽시간대의 속도편차는 예측모델의 효율을 떨어뜨리는 원인이 되기 때문에 본 논문에서는 70km/hour 이상의 차량속도는 70km/hour로 고정시켰다. 특히 차량이 드문 새벽에는 차량이 한 대도 지나가지 않는 경우가 발생하는데 이때 차량속도를 측정하는 루프검지기의 데이터가 0km/hour로 나오기 때문에 우리는 이것을 75km/hour로 변형시켰다. 이렇게 변형된 데이터 1339개 중에서 1039(8월 15일 09:25 - 8월 18일 23:55)개의 데이터는 각 예측 모델의 학습을 위하여 사용하였고 나머지 300(8월 19일 00:00 - 8월 20일 00:55)개의 데이터는 각 모델의 성능을 테스트하는데 사용되었다.

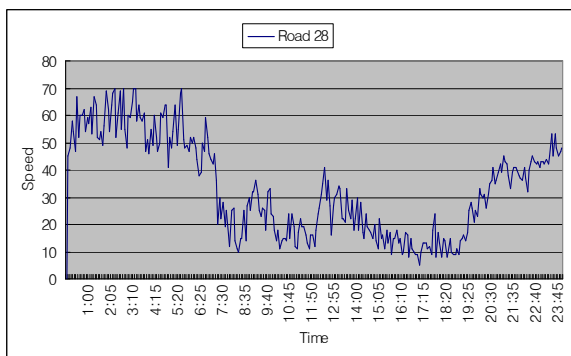


그림 2 동문로에서의 일일 차량속도

### 3. 예측모델

본 논문에서는 다음의 6개의 예측모델을 제안하고 각각의 모델은 road 28의 5분후의 속도를 예측하고 MSE와

평균절대값오차(Mean absolute error: MAE)를 이용하여 성능을 비교하였다. 이를 통하여 ANN 모델과 ARIMA 모델 중 어느 것의 우수한 성능을 보이는가와 판단보정기법의 적용이 교통량 예측능력향상에 도움이 되는가를 조사하였다.

- 1) A Simple ANN model
  - Back-propagation Training method 사용
  - Inputs: road 28에서 추출된 1039개의 데이터
- 2) A Simple ARIMA model
  - Inputs: road 28에서 추출된 1039개의 데이터
- 3) An ARIMA+ANN model
  - Back-propagation Training method 사용
  - Inputs
    - road 28에서 추출된 1039개의 데이터
    - ARIMA model의 예측데이터
- 4) An ANN+JA model
  - Back-propagation Training method 사용
  - Inputs
    - road 28에서 추출된 1039개의 데이터
    - road 28에 영향을 주는 road 4, 9, 12 데이터
- 5) An ARIMA+ANN model
  - 예측데이터 = ARIMA 예측데이터 + JA 예측데이터
- 6) An ARIMA+ANN+JA model
  - 예측데이터 = (ARIMA+ANN) 예측데이터 + JA 예측데이터

### 4. 실험결과

#### 4.1. A Simple ANN model

일반적으로 ANN의 성능은 입력 노드(node) 수와 은닉층(hidden layer)의 노드 수와 같은 네트워크 구조에 영향을 많이 받는다. 그러나 네트워크 구조를 결정하는 정형화된 방법은 아직 존재하지를 않기 때문에 휴리스틱한 방법에 의해서 최적화된 네트워크 구조를 결정한다. Hecht-Nielsen theorem에서 입력 노드 수가  $n$ 개 일때 은닉층의 노드 수는  $2n+1$ 개를 넘을 필요가 없다는 것을 보증하고 있기 때문에 이를 이용하여 네트워크 구조를 설계하였다[6].

입력변수는 3개의 변수( $t-1, t-2, d-1$ )를 기본변수로 하고 이후 시간변수( $t-3, t-4, t-5, t-6$ )를 하나씩 늘려나가며 각각의 예측 능력을 평가하였다. 은닉층의 구조는 Hecht-Nielsen theorem에 따라 설계하였고 노드 수는 짝수 개만을 실험하였다. 표1은 각 모델의 성능을 MSE와 MAE로 측정된 데이터이다.

- 1)  $t-p$  : 5\*p분 전의 차량속도
- $d-1$  : 하루 전의 차량속도

입력노드		은닉노드수	MSE	MAE
갯수	입력변수			
3	$t-1, t-2, d-1$	2	36.3646	4.7749
		4	33.6902	4.7096
		6	34.0149	4.7090
4	$t-1, t-2, t-3, d-1$	2	33.7409	4.7391
		4	33.1635	4.6945
		6	33.7135	4.9070
		8	33.9533	4.7003
5	$t-1, t-2, t-3, t-4, d-1$	2	34.6072	4.7294
		4	33.7879	4.6593
		6	35.7071	4.8102
		8	36.5722	4.8214
		10	35.8964	4.8404
6	$t-1, t-2, t-3, t-4, t-5, d-1$	2	34.7143	4.7333
		4	34.4038	4.6983
		6	36.6942	4.8393
		8	35.9956	4.7962
		10	39.7751	4.9747
		12	38.2846	4.9133
7	$t-1, t-2, t-3, t-4, t-5, t-6, d-1$	2	34.4528	4.6877
		4	35.4224	4.7416
		6	34.5496	4.6591
		8	37.4167	4.8431
		10	38.3701	4.9843
		12	40.1798	5.0963
		14	38.4336	4.9541

표 1 ANN의 예측성능

ANN은 1339개의 데이터중 1039개를 가지고 총 2000회 학습을 시켰으며 나머지 300개의 데이터로 성능을 평가하였다. 모든 실험은 지역해(local minimum)에 빠지는 것을 방지하기 위하여 8번 반복하여 실험하고 그 중 최저값을 취하였다. 위의 데이터에서 가장 좋은 성능을 나타낸 것은 4-4-12구조를 가진 것이었다.

ANN 모델의 특징은 다음과 같다.

- 1) 아침이나 낮시간과 같이 차량속도의 변화가 많지 않은 경우 정확한 예측능력을 보임
- 2) 예측 데이터는 예기치 않은 변화에 대하여 안정화하려는 경향을 보임

#### 4.2. A Simple ARIMA model

ARIMA 모델의 경우 Box-Jenkins 기법을 사용하여 설계하였으며 실험에 사용된 데이터는 ARIMA(0,1,2) 과정을 따르는 것이 확인되었다[7]. 이 모델의 평균성능은 MSE=30.698과 MAE=4.45661을 보였다. 모델식은 다음과 같다.

$$x_t = x_{t-1} + \varepsilon_t + 0.68661\varepsilon_{t-1} + 0.062871\varepsilon_{t-2}$$

ARIMA 모델의 특징은 다음과 같다.

- 1) 학습데이터가 많아질수록 우수한 성능을 보임
- 2) 충분한 학습데이터가 주어질 경우 ARIMA 모델이 ANN 모델보다 우수한 성능을 보임
  - 학습초기(8월 15일 09:25 - 8월 18일 23:55)에는

- 2) 입력노드수 - 은닉노드수 - 출력노드수

ARIMA 모델의 성능이 ANN 모델의 성능보다 열등하지만 시간이 지날수록(8월 19일 00:00 - 8월 20일 00:55) 성능이 우수해진다.

	MSE	MAE
ARIMA 모델의 초기 예측성능 (8월 15일 09:25-8월 18일 23:55)	39.941	4.9390
ARIMA 모델의 후기 예측성능 (8월 19일 00:00-8월 20일 00:55)	30.698	4.4566
ANN 모델의 예측성능 (8월 19일 00:00-8월 20일 00:55)	33.165	4.6945

표 2 ARIMA 모델과 ANN 모델의 성능비교

#### 4.3. An ARIMA+ANN model

이 모델은 ARIMA 모델의 결과값이 ANN 모델의 입력값으로 사용된다. 기존 연구결과는 ARIMA+ANN 모델이 ARIMA 모델이나 ANN 모델보다 우수한 예측 성능을 나타냈었다[8]. 그림 3은 이 모델의 대략적인 구조를 나타내고 있다.

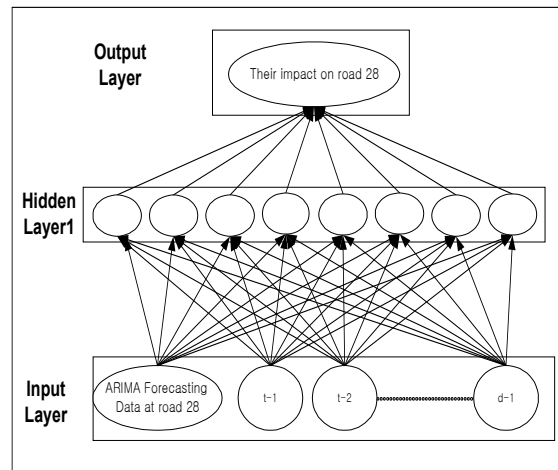


그림 3 ARIMA+ANN 모델의 구조

네트워크의 구조는 ANN 모델에서 좋은 성능을 나타내었던 3-4-1, 4-4-1, 5-4-1, 6-4-1, 7-2-1, 8-2-1만을 실험하였다. 표3은 네트워크 구조에 따른 성능을 나타내고 있다. 최고의 성능을 나타낸 구조는 5개의 입력값( $t-1, t-2, t-3, d-1, ARIMA\ output$ )을 가진 모델이었다. 이 실험으로 ARIMA+ANN 혼합모델이 ANN 이나 ARIMA 와 같은 단일모델보다 좋은 성능을 보임을 알 수가 있다.

Input factors	MSE	MAE
$t-1, t-2, d-1, ARIMA\ output$	28.7113	4.3369
$t-1, t-2, t-3, d-1, ARIMA\ output$	27.9322	4.1685
$t-1, t-2, t-3, t-4, d-1, ARIMA\ output$	28.3495	4.2628
$t-1, t-2, t-3, t-4, t-5, d-1, ARIMA\ output$	28.8632	4.3022
$t-1, t-2, t-3, t-4, t-5, t-6, d-1, ARIMA\ output$	27.9596	4.2197
$t-1, t-2, t-3, t-4, t-5, t-6, t-7, d-1, ARIMA\ output$	28.0852	4.1523

표 3 ARIMA+ANN의 성능비교

#### 4.4. 판단보정기법

판단보정기법은 예측모델이 고려하지 못하는 요소 때문에 발생하는 예측오차를 줄이기 위하여 도입되었다. 우리의 실험에서는 road 28의 속도를 예측하기 위하여 road 4, 9, 12의 5분전 및 10분전의 차량속도를 ANN의 입력변수에 첨가하여 판단보정을 시도하였다. 실제도로환경에서 road 4, 9, 12에서 road 28까지의 차량운행 소요시간은 5-10분이 소요된다. 그러므로 5-10분전에 road 4, 9, 12의 차량 속도는 road 28에 유입되는 차량의 속도에 영향을 미칠 것으로 가정하였다. 그림 4는 근처 도로의 차량속도와 그들이 road 28의 차량속도에 미치는 영향을 ANN을 사용하여 산출하는 과정을 표현하였다.

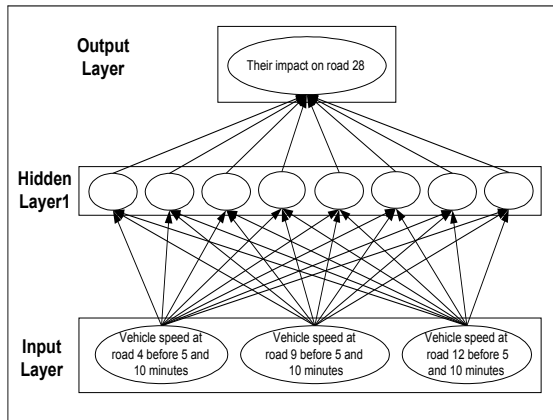


그림 4 판단보정기법 산출과정

ANN에서 road 28에 미치는 영향은 예측데이터와 실제데이터의 차이를 이용하여 산출하였다. 그림 5는 ARIMA+ANN+JA 모델이 예측한 교통데이터와 실제 교통데이터를 나타내고 있다.

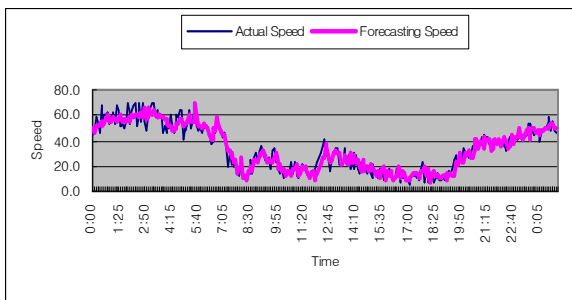


그림 5 ARIMA+ANN+JA 모델이 예측한 교통데이터와 실제 교통데이터의 차이

여기서 판단보정기법을 적용했다는 것은 단일한 예측모델(ARIMA 모델, ANN모델, ARIMA+ANN 모델)의 결과에 그림 4에서 산출된 값을 더하는 것을 의미한다. 표 4는 우리가 실험한 6가지 모델의 성능을 비교한 것이다. 여기서 판단보정기법의 적용이 성능향상에 도움이 된다는 것을 확인할 수 있다.

Input factors	MSE	MAE
ANN	33.1635	4.6945
ARIMA	30.6980	4.4566
ARIMA+ANN	27.9322	4.1685
ANN+JA	29.8105	4.3975
ARIMA+JA	28.5425	4.1217
ARIMA+ANN+JA	25.7167	3.9200

표 4 제안된 모델의 성능비교

## 5. 성능비교분석

지금부터 각각의 모델의 성능을 양측검정에 기반을 둔 Z-test를 이용하여 비교하고 다음사항을 조사할 것이다.

- 1) ARIMA 모델과 ANN 모델의 성능비교
- 2) ARIMA+ANN 모델과 단일모델(ARIMA 모델)의 성능비교
- 3) 판단보정기법의 유효성

### 5.1. ARIMA 모델과 ANN 모델의 성능비교

대부분의 선행연구에서는 ANN 모델이 전통적인 시계열분석 모델인 ARIMA 모델보다 성능이 우수하다고 보고되고 있다. 그러나 우리의 실험결과에서는 적은수의 훈련데이터가 있을 경우에는 ANN 모델이 우수하였으나 훈련데이터가 많은 경우 ARIMA 모델의 성능이 우수한 것으로 나타났다.

**가설1: ANN 모델의 예측능력은 ARIMA 모델의 예측능력과 차이가 없다.**

$$H_0: Error_{ANN} - Error_{ARIMA} = 0$$

$$H_1: Error_{ANN} - Error_{ARIMA} \neq 0$$

Z-test결과 ARIMA 모델이 ANN 모델보다 우수한 성능을 나타내고 있다. 이는 본 실험에서 쓰인 훈련데이터의 양이 다른 연구에서 쓰인 데이터의 양보다 상대적으로 많기 때문에 해석된다. p-value는 MSE와 MAE가 각각 0.2137과 0.0883으로 위의 가정이 매우 유의하다고 할 수는 없으나 어느 정도의 유의하다고 할 수는 있다.

### 5.2. ARIMA 모델과 ARIMA+ANN 모델의 성능비교

본 절에서는 ARIMA와 ANN 혼합모델과 ARIMA 단일 모델의 성능을 비교한다.

**가설2: ARIMA 모델의 예측능력은 ARIMA+ANN 모델의 예측능력과 차이가 없다.**

$$H_0: Error_{ARIMA} - Error_{ARIMA+ANN} = 0$$

$$H_1: Error_{ARIMA} - Error_{ARIMA+ANN} \neq 0$$

Z-test결과 ARIMA+ANN 모델이 ARIMA 모델보다 우수한 성능을 나타내고 있다. p-value는 MSE와 MAE가 각각 0.0244와 0.0033으로 위의 가정이 매우 유의하다고 할 수는 있다. 이는 단일 모델보다는 ARIMA+ANN 혼합모델의 예측 성능 우수하다는 것을 입증하고 있다.

### 5.3. ARIMA+ANN+JA 모델과 ARIMA+ANN 모델의 성능비교

본 절에서는 판단보정기법의 적용이 유효성 여부를 조사한다.

**가설3: ARIMA+ANN 모델의 예측능력은  
 ARIMA+ANN+JA 모델의 예측능력과 차이가  
 없다.**

$$H_0: Error_{ARIMA+ANN} - Error_{ARIMA+ANN+JA} = 0$$

$$H_1: Error_{ARIMA+ANN} - Error_{ARIMA+ANN+JA} \neq 0$$

Z-test결과 p-value는 MSE와 MAE가 각각 0.1302와 0.0469로 위의 가정이 유의하다고 할 수는 있다. 예측능력은 일반적으로 판단보정기법을 적용시 향상되는 것을 알 수가 있다. 또한 ARIMA+ANN+JA 모델이 가장 우수한 성능을 나타내고 있음을 입증하고 있다.

표 5은 Z-test를 통한 모델들의 성능비교 결과를 나타내고 있다.

Input factors	p-value	
	MSE	MAE
ARIMA vs. ANN	0.2173	0.0883
ARIMA vs. ARIMA+ANN	0.0244*	0.0033**
ANN vs. ANN+JA	0.1778	0.0074**
ARIMA vs. ARIMA+JA	0.0344*	0.0166*
ARIMA+ANN vs. ARIMA+ANN+JA	0.1302	0.0469*

표 5 Z-test를 통해 산출된 p-value  
 (\*: 유의수준 5%로 유의, \*\*: 유의수준 1%로 매우유의)

**6. 결론**

지능형교통시스템(ITS)에서 미래 교통량 예측은 효율적인 교통관리에 핵심적인 요소로 차내 교통정보제공시스템, 동적경로안내 시스템, 실시간 교통관리 시스템 및 유고감지 시스템 등에 이용되고 있다.

본 논문에서는 정확한 교통량 예측을 위해 6가지의 예측 모델을 제시하고 각각의 성능을 비교함으로써 다음과 같은 결과를 얻었다.

- 1) ARIMA 모델과 ANN 모델의 성능비교에서 기존의 연구처럼 언제나 ANN 모델이 ARIMA 모델보다 우수한 성능을 보이지는 않았다. 일반적으로 훈련데이터가 많은 경우에는 ARIMA 모델이 좀더 정확한 예측 데이터를 산출하였다.
- 2) ARIMA+ANN 혼합 모델은 단일모델(ARIMA 모델, ANN 모델)보다 우수한 성능을 나타내었다.
- 3) 판단보정기법은 예측능력향상에 도움이 되었다.
- 4) 6가지 모델중에서 ARIMA+ANN+JA 모델이 가장 우수한 성능을 나타내었다.

**Reference**

[1] K. Hornik, "Multilayer feedforward networks are universal approximators", Neural Networks, vol.2, pp.359-366, 1989.  
 [2] Funahashi, "On the approximate realization of continuous mapping by neural networks", Neural Networks, vol.2, pp.183-192, 1989.  
 [3] H. White, "Connectionist non-parametric regression:

multiplayer feedforward networks can learn arbitrary mapping", Neural Networks, vol.3, pp.535-549, 1990.  
 [4] Lee, Jae K., Yum, C.S. and Kim, W.J., "Neural Network Based Judgemental Adjustment for Time Series Forecasting", Proceedings of the International Conference EANN '95, pp.299-392, Aug., 1995.  
 [5] Yum Chang-Seon, "Judgemental Adjustment in Time Series Forecasting: Neural Network Approach", Theory in Graduate School of Management at KAIST, pp. 2-27, 1996.  
 [6] Fausett, L., "Fundamentals of Neural Networks", Prentice Hall, Inc., 1994.  
 [7] Vandaele, W., "Applied Time Series and Box-Jenkins Models", Academic Press, 1983  
 [8] Hong-Yong Oh, "Forecasting of the Short-Term Demand for the Natural Gas Using Time Series Analysis and Artificial Neural Networks", Unpublished Master Thesis, Graduate School of Management, KAIST, 1997.