

# A Study on Internet Traffic Forecasting by Combined Forecasts

Sahm Kim<sup>a,1</sup>

<sup>a</sup>Department of Applied Statistics, Chung-Ang University

(Received November 2, 2015; Revised November 16, 2015; Accepted November 23, 2015)

---

## Abstract

Increased data volume in the ICT area has increased the importance of forecasting accuracy for internet traffic. Forecasting results may have paper plans for traffic management and control. In this paper, we propose combined forecasts based on several time series models such as Seasonal ARIMA and Taylor's adjusted Holt-Winters and Fractional ARIMA(FARIMA). In combined forecasting methods, we use simple-combined method, MSE based method (Armstrong, 2001), Ordinary Least Squares (OLS) method and Equality Restricted Least Squares (ERLS) method. The results show that the Seasonal ARIMA model outperforms in 3 hours ahead forecasts and that combined forecasts outperform in longer periods.

Keywords: Fractional Seasonal ARIMA, adjusted Holt-Winters, internet traffic, combined forecasting

---

## 1. 서론

국내 시장에서 ICT(Information and Communications Technologies) 분야 제품들의 발달 및 대중화에 따라 많은 국내 시장에서의 인터넷 트래픽 사용량은 큰 관심 거리가 되고 있다. 특히 모바일, 웨어러블 기기 및 빅데이터가 ICT 분야에서 키워드로서 나타나는 한편, ICT 분야에서의 사물 인터넷에 사용으로 인하여 트래픽 사용량은 미래에 더욱 증가할 것으로 보여진다.

이러한 트래픽 사용량 증가에 따른 대책으로 트래픽 용량 증가는 가장 먼저 떠오를 수 있는 해결책이다. 그러나 이러한 방법은 많은 비용이 소모될 뿐 아니라 일시적인 해결 방안일 수 밖에 없다. 또 다른 해결 방안으로는 트래픽 사용량의 분배를 통해 트래픽의 수요예측 관리 방안으로서, 정확한 트래픽 예측을 통하여, 즉, 기기 및 가구 별 정확한 수요예측을 통하여, 트래픽을 분배하여 특정 부분의 트래픽의 과부하를 방지하는 한편, 트래픽의 관리 및 제어를 하는 방안이다. 정확한 트래픽 수요량의 관리와 제어를 위해, 트래픽 수요예측은 매우 중요한 요소이다.

ICT 분야에서 만이 아니라 인터넷 시장은 산업 분야에서 증추신경과 같은 존재로 기관들이 원할한 업무가 가능하게 할 뿐 아니라, 대중들의 정보 검색 및 SNS 사용, 소통에 있어서 큰 영향력을 미치고 있다. 최근 모바일 산업에서는 이전에 통화 기준 요금 정책에서 데이터 사용 기준 요금 정책으로 바뀐 것은 이

---

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2013R1A1A2008277).

<sup>1</sup>Department of Applied Statistics, Chung-Ang University, 221 Heukseok-Dong, Dongjack-Gu, Seoul 06974, Korea. E-mail: sahm@cau.ac.kr

리한 시대 변화를 반영된 것으로 보여진다. 현재 기기들의 발달에 따라 시간 및 장소에 상관없이 인터넷 사용이 가능함에 따라 트래픽에 대한 관리에 중요성이 증대되며, 특히 트래픽 예측의 실패는 통신망 마비 사태를 발생시킬 가능성이 있다.

국내외 최근 트래픽 분야에서의 최근 연구를 살펴보면, Liu 등 (1999)은 FARIMA(Fractional Autoregressive Integrated Moving Average) 모델을 이용하여 트래픽을 예측하였으며, Shu 등 (2003)은 계절형 ARIMA를 이용하여 무선 트래픽을 예측하였으며, Yu 등 (2010)은 중국 하이룽장 성에서 모바일 네트워크에서의 트래픽에 대하여 기기별 상관성에 따라 계절형 ARIMA 모델을 적용하여 예측 후 불확실한 이상 값들에 대하여 특성에 대한 상관 계수를 이용한 트래픽 이상 방안을 제시하였다. Zhou 등 (2005)는 네트워크 트래픽 분야에서 FARIMA와 ARIMA-GARCH 모델을 사용하였으며, Jiang과 Papavassiliou (2006)는 상관 모델을 이용하여 ARIMA 모델을 기반으로 이전 트래픽 수치들과의 상관관계를 고려하여 접근하였다. Tikunov와 Nishimura (2007)은 모바일 트래픽 예측에 있어서 Holt-Winters 방법을 사용하였으며, Au 등 (2011)는 모바일 트래픽 예측에서 시간별 자료를 이용하여 결합 예측 모델을 사용하였다. 이에 대하여 계절형 ARIMA 모델과 Holt-Winters 방법을 사용하여 SA(Simple Average), OLS(Ordinary Least Square), ERLS(Equality Restricted Least Squares)를 이용한 결합 예측 방법을 시도하였다.

트래픽 분야 외에서도 시계열 분석방법이 적용되어왔는데 대표적으로 전력 분야에서는 Taylor는 2003년에 오차항을 수정한 이중계절형 Holt-wintets 모델을 사용하였으며, 2010년도에는 삼중계절형 Holt-Wintets 모델을 사용하여 전력수요예측에서의 우수한 성능을 보였다.

이전 연구에서는 Kim (2007)은 Engle (1982)이 제안한 자기회귀이분산성(Autoregressive Conditional Heteroskedasticity; ARCH) 모델을 이용하여 자기회귀이분산성모형이 트래픽자료를 예측하는데 있어 AR 모형보다 성능이 우수함을 보였고, Ha와 Kim (2008)은 장기기억(long memory)특성을 설명하기 위하여 멱변환 GARCH(PGARCH) 모델을 소개하며 기존 GARCH 모형보다 우수성을 입증하였다. Kim (2011)은 Seasonal AR-GARCH 모델을 이용하여 Seasonal AR-GARCH이 Seasonal ARIMA와 AR-GARCH 모형보다 트래픽 자료를 예측하는데 성능이 우수함을 보였다. Sohn과 Kim (2013)은 장기기억 특성과 이분산성을 고려하여 AR-GARCH 모형, ARMA-GARCH 모형과 장기기억모형인 Fractional ARIMA와 장기기억과 이분산성을 고려한 Fractional ARMA-GARCH 모형의 사용으로 모형간 결합을 시도하였다.

본 연구에서는 트래픽 자료를 예측하는데 있어 지수평활법 중 Holt-Winters 방법과 Taylor에 의해 제안된 수정된 Holt-Winters 방법을 비교하며, 장기 기억특성을 고려한 수정된 Fractional Seasonal ARIMA 방법, 이중계절형 ARMA 모델을 이용하여 트래픽의 예측을 실시하며, 예측에 대하여 결합 예측 모형으로서 Aksu와 Bunter (1992)가 제안한 단순예측 결합방법과, OLS, ERLS 예측 결합 방법과 Armstrong (2001)이 제안한 MSE 기반 결합 예측 방법을 비교하고자 한다. 본 논문에서는 2장에서는 사용된 시계열 모형을 소개하며, 3장에서는 결합예측 방법을 소개한다. 4장에서는 분석 및 예측 성능을 비교하며, 4장에서는 이에 대한 결론을 나타낸다.

## 2. 시계열 모형 소개

### 2.1. Taylor의 수정된 이중계절형 Holt-Winters

본 논문에서는 지수평활법 중 가법 Holt-Winters에서 2개의 계절 주기를 가지는 이중계절형 Holt-Winters 방법에 대하여 예측에서 오차의 AR(1)을 고려한 Taylor의 수정된 이중계절형 Holt-Winters 모형 (Taylor, 2003)을 사용한다. 이에 따라 1일 주기인 288과 1주일 주기인 2016을 주기로 모형에

적용 하고자 한다.

이에 따라 이중 계절성을 적용시킨 수정된 이중계절성 Holt-Winters 모형은 (2.1)과 같이 정의된다.

$$\begin{aligned} L_t &= \alpha(Z_t - S_{t-s_1} - D_{t-s_2}) + (1 - \alpha)(L_{t-1} + T_{t-1}), \\ T_t &= \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}, \\ S_t &= \gamma(Z_t - L_t - D_{t-s_2}) + (1 - \gamma)S_{t-s_1}, \\ D_t &= \delta(Z_t - L_t - S_{t-s_1}) + (1 - \delta)D_{t-s_2}, \end{aligned} \quad (2.1)$$

여기서  $L$ 은 평활 수준,  $T$ 는 평활가법추세를 의미하며,  $S$ 는 첫 번째 주기에 대한 계절지표를 나타내며,  $D$ 는 두 번째 주기에 대한 계절지표를 나타낸다. 이에 따라  $\alpha$ ,  $\beta$ 는 수준과 추세에 대한 모수이고  $\gamma$ ,  $\delta$ 는 계절평활 모수이다. 이에 대하여 추정된 모수를 적용한  $h$  시점 후에 예측한 값은  $F_{t+h}$ 로서 이는 식 (2.2)와 같이 나타나며, 오차에서 예측에서 오차의 AR(1)을 고려한 예측값이다.

$$F_{t+h} = L_t + T_t \times h + S_{t+h-s_1} + D_{t+h-s_2} + \phi^h \{Z_t - L_{t-1} - T_{t-1} - S_{t-s_1} - D_{t-s_2}\}, \quad (2.2)$$

여기서  $\phi$ 는 오차의 AR(1)을 고려한 모수이며, 식 (2.1)에 초기값은 식 (2.3)를 통하여 구할 수 있다.

$$\begin{aligned} L_{s_1} &= \frac{\sum_{t=1}^{s_1} Z_t}{s_1}, \\ L_{s_2} &= \frac{\sum_{t=1}^{s_2} Z_t}{s_2}, \\ T_{s_1} &= \frac{1}{s_1} \left( \frac{\sum_{t=s_1+1}^{2s_1} Z_t}{s_1} - \frac{\sum_{t=1}^{s_1} Z_t}{s_1} \right), \\ T_{s_2} &= \frac{1}{s_2} \left( \frac{\sum_{t=s_2+1}^{2s_2} Z_t}{s_2} - \frac{\sum_{t=1}^{s_2} Z_t}{s_2} \right), \\ S_1 &= Z_1 - L_{s_1}, \dots, S_{s_1} = Z_{s_1} - L_{s_1}, \\ D_1 &= Z_1 - L_{s_2}, \dots, D_{s_2} = Z_{s_2} - L_{s_2}. \end{aligned} \quad (2.3)$$

이에 대하여 모수  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\phi$ 의 추정은 훈련자료에서의 최소제곱법을 통하여 추정된다. 즉, 1차 시점 예측으로 부터 SSE(Sum of Squared Error)를 최소화하는 모수의 추정치를 구하는 과정을 통해 계산된다. 이때 SSE는 식 (2.4)과 같다.

$$\text{SSE} = \sum_{t=2}^m (F_t - Z_t)^2. \quad (2.4)$$

## 2.2. 계절형 ARIMA

단변량 시계열 분석에서 가장 보편적인 방법으로 Box와 Jenkins (1994)의 계절형 자기회귀누적이동 평균(Autoregressive Integrated Moving Average; ARIMA) 모형이 있다. 이에 대하여 이중 계절형 ARIMA의 기본 형태는 식 (2.5)와 같다.

$$\phi_p(B)\Phi_{P_1}(B^{s_1})\Pi_{P_2}(B^{s_2})(1-B)^d(1-B^{s_1})^{D_1}(1-B^{s_2})^{D_2}(Z_t - c) = \theta_q(B)\Theta_{Q_1}(B^{s_1})\Psi_{Q_2}(B^{s_2})\epsilon_t, \quad (2.5)$$

여기서  $c$ 는 상수항을 말하고,  $B$ 는  $B^k Y_t = Y_{t-k}$ 에 연산을 맡는 후진연산자이고,  $\phi_p(B)$ 와  $\theta_q(B)$ 는  $p$ 차 AR과  $q$ 차 MA를 의미하며,  $\Phi_{P_1}(B^{s_1})$ ,  $\Pi_{P_2}(B^{s_2})$ ,  $\Theta_{Q_1}(B^{s_1})$ ,  $\Psi_{Q_2}(B^{s_2})$ 는  $s_1$ ,  $s_2$ 에 따르는  $P_1$ ,  $P_2$ ,  $Q_1$ ,

$Q_2$ 차 계절성 AR과 계절성 MA를 의미한다.  $d, D_1, D_2$ 는 차분의 차수를 말하며  $\epsilon_t$ 는 평균 0과 고정된 분산을 따르는 백색잡음과정이다. 다항식  $\phi_p(B) = 0, \theta_q(B) = 0, \Theta_{Q_1}(B^{s_1}) = 0, \Psi_{Q_2}(B^{s_2}) = 0, \Phi_{P_1}(B^{s_1}) = 0, \Pi_{P_2}(B^{s_2}) = 0$ 의 단위근은 단위원 밖에 위치한다.

Au 등 (2011)에서 이중 계절형 ARIMA 모형에 경우, 훈련 자료에 정상성을 조건을 충족시킨 다음, 최적의 모형 선택을 위하여 BIC(Basian Information Criteria)를 최소로 하는 모형을 선택한다. 본 논문에서도 이에 대하여 확장하여 적용을 시도한다. 5분 단위의 데이터를 기반으로 하기 때문에  $p, q$ 의 범위를 0에서 6사이로 확장하여 선택하며,  $P_1, P_2, Q_1, Q_2$ 의 범위를 0에서 2사이로 제한을 두어, 가장 작은 BIC를 가진 모형을 선택한다.

### 2.3. Fractional ARIMA(FARIMA)

FARIMA 모형은 Liu 등 (1999)이 자료의 장기기억 특성을 고려한 모형으로 제안되었다. 정상성을 만족하는 시계열 자료  $Z_t$ 의  $k$ 시차의 ACF( $\rho_k$ )의 절대값의 합이 유한할 경우  $Z_t$ 는 단기 기억(short memory)를 가지며,  $k$ 시차의 ACF( $\rho_k$ )의 절대값의 합이 유한하지 않을 경우 장기 기억(long memory)가진다.

Fractional ARIMA(FARIMA)는 장기기억 특성을 고려한 모형으로서 ARIMA( $p, d, q$ ) 모형에서  $d$ 가 실수를 가지는 모형으로 FARIMA( $p, d, q$ )에 대해 식 (2.6)과 같이 정의된다.

$$\phi_p(B)\Delta^d Z_t = \theta_q(B)\epsilon_t, \quad (2.6)$$

여기서  $\epsilon_t$ 는 백색잡음으로 모형의 정상성을 위해  $d \in (-0.5, 0.5)$ 이며,  $\Delta^d$ 는 분수차분연산자로 식 (2.7)과 같이 정의된다.

$$\Delta^d = (1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k. \quad (2.7)$$

### 3. 결합 예측 방법

Armstrong (2001)과 Aksu와 Bunter (1992)는 여러 결합 예측 방법으로 단순한 방법으로 예측의 정확도를 높일 수 있음을 보였다. Au 등 (2011)은 이에 대하여 2가지 모형을 기반으로 SA(Simple Average), OLS, ERLS 방법을 사용하여 비교를 실시하였다. 본 연구에서는 이러한 방법에, 추가적으로 MSE(Mean Squared Error)에 비율에 따라서 결합 예측에 대하여 비교를 실시한다.

본 연구에서는 Taylor의 이중계절형 Holt-Winters(HW), 이중계절형 ARIMA(A), FARIMA(FA) 모형의 예측치로서 각각  $F_t^{HW}, F_t^A, F_t^{FA}$ 로 표기한다.

먼저 SA(Simple Average) 방법으로  $F_t^{HW}, F_t^A, F_t^{FA}$ 의 동일 평균한 예측 값으로,  $F_t^{S.A.}$ 는 이를 단순 결합한 값이다. 이 방법은 가장 단순한 방법으로 이전 훈련 자료에서의 예측값의 적합치를 고려하지 않는다. 이는 식 (3.1)과 같이 표현된다.

$$F_t^{S.A.} = \frac{F_t^{HW} + F_t^A + F_t^{FA}}{3}. \quad (3.1)$$

OLS 결합 예측 방법은 훈련자료에서 적합지와 실제치와의 회귀분석을 통하여 산출된 추정치를 통하여 식 (3.2)에 수식과 같이 표현된다.

$$F_t^{OLS} = a + b \times F_t^{HW} + c \times F_t^A + d \times F_t^{FA}. \quad (3.2)$$

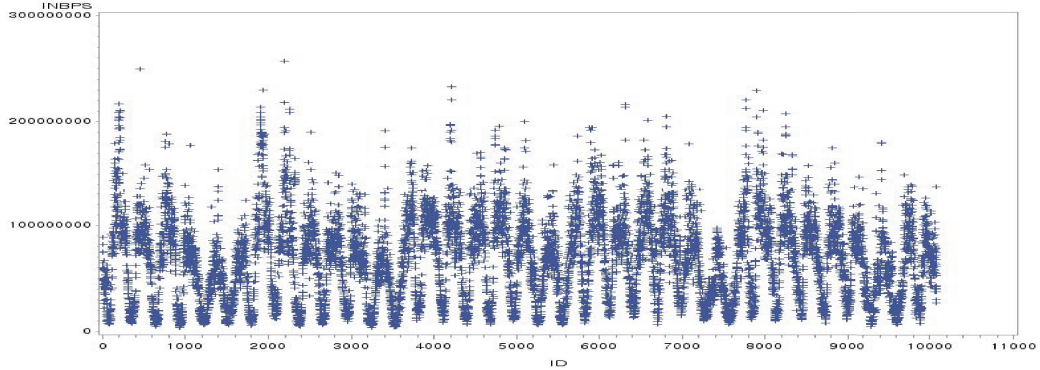


Figure 4.1. Original time series data.

ERLS(Equality Restricted Least Squares) 결합 예측 방법은 중 상수항이 없는 ERLS 결합 예측 방법을 선택하여 식 (3.2)에  $a = 0, b + c + d = 1$ 이라는 제한 조건하에서 식 (3.3)에 수식과 같이 표현된다.

$$F_t^{ERLS} = b \times F_t^{HW} + c \times F_t^A + (1 - b - c) \times F_t^{FA}. \tag{3.3}$$

Armstrong (2001)이 주장한 MSE(Mean Squared Error) 기반 결합 예측 방법은 MSE에 비율에 따라서 결합을 한다.  $m$ 개의 훈련자료에서의 MSE는 식 (3.4)와 같이 표현된다. 이에 따른 MSE기반의 예측값은 식 (3.5)와 같으며, 이는 훈련자료에서의 각 모형의 MSE 값을 고려하여 결합 예측하는 방법으로  $MSE^{HW}, MSE^A, MSE^{FA}$ 는 각 모형 별 훈련자료의 MSE로서,  $M$ 은  $M = MSE^{HW} + MSE^A + MSE^{FA}$ 로 정의된다.

$$MSE = \frac{1}{m} \sum_{t=1}^m (Z_t - F_t)^2, \tag{3.4}$$

$$F_t^{MSE} = \frac{(M - MSE^{HW}) F_t^{HW} + (M - MSE^A) F_t^A + (M - MSE^{FA}) F_t^{FA}}{2M}. \tag{3.5}$$

#### 4. 분석 및 예측 성능비교

본 연구에서 사용한 자료는 중앙대학교 교내 서버로 들어온 5분 단위 트래픽 자료로서 2015년 3월 31일부터 2015년 5월 4일 까지의 5주간의 자료를 사용하여 모형을 적합하였다. k-Ahead 예측을 실시하여 6시간(72-Ahead Forecasting)에서부터 12시간 후(144-Ahead Forecasting)에 대하여 예측을 실시하며 RMSE(Root Mean Square Error)를 기준으로 평가한다. RMSE 통계량은 3일간의 864개 자료를 통하여 산출되어 식 (4.1)과 같이 나타난다. 이 때의  $m$ 은 5주간의 자료로서 10080이다.

$$RMSE = \sqrt{\frac{1}{864} \sum_{t=m+1}^{m+864} (Y_t - F_t)^2}. \tag{4.1}$$

Figure 4.1는 훈련자료의 원자료 시계열도표로서 일별(288) 주기와 주별(2016) 주기를 가지는 패턴이 반복 되고 있음을 알 수 있다. 이에 따라서 본 논문에서는 이중 계절형을 고려한 모형을 작성하도록 한다.

**Table 4.1.** Parameter estimation of Holt-Winters

Parameter	Estimate
LEVEL( $\alpha$ )	0.11667
TREND( $\beta$ )	2.3e-09
SEASONAL1( $\gamma$ )	0.00156
SEASONAL2( $\delta$ )	0.35073
Error-term AR(1)( $\phi$ )	0.44598

**Table 4.2.** Parameter estimation of ARIMA

Parameter	Estimate	S.E.	Parameter	Estimate	S.E.
$\theta_1$	0.30462	0.01693	$\phi_1$	0.73916	0.01206
$\Theta_{1(288)}$	0.99800	0.00935	$\Phi_{1(288)}$	0.96392	0.01296

ARIMA = Fractional Autoregressive Integrated Moving Average.

**Table 4.3.** Parameter estimation of FARIMA

Parameter	Estimate	S.E.	Parameter	Estimate	S.E.
$\theta_1$	0.86902	0.01058	$\phi_1$	0.97610	0.00291
$\mu$	17.78570	1.64779	$\hat{d}$	0.45419	

FARIMA = Fractional Autoregressive Integrated Moving Average.

**Table 4.4.** Result of RMSE

	HW	ARIMA	FARIMA	SA	OLS	ERLS	Base on MSE
After 3hour	0.5808(4)	0.5176(1)	0.7598(7)	0.5557(2)	0.5832(5)	0.587(6)	0.5627(3)
After 6hour	0.8161(6)	0.6484(3)	0.8487(7)	0.6451(1)	0.6606(4)	0.6641(5)	0.6469(2)
After 9hour	0.9038(7)	0.7003(5)	0.8087(6)	0.6480(3)	0.6478(2)	0.6502(4)	0.6463(1)
After 12 hour	0.9643(7)	0.6392(5)	0.7195(6)	0.6164(4)	0.5951(1)	0.5973(2)	0.6058(3)

RMSE(Ranking)

RMSE = Result of Root Mean Square Error; HW = Holt-Winters;

ARIMA = Autoregressive Integrated Moving Average;

FARIMA = Fractional Autoregressive Integrated Moving Average; SA = Simple Average;

OLS = Ordinary Least Square; ERLS = Equality Restricted Least Squares;

MSE = Mean Squared Error.

본 연구에서는 Taylor가 수정한 이중계절형 Holt-Winters모형과 일반차분을 통하여 정상성을 만족시킨 자료에 이중계절형 ARIMA모형과 FARIMA 모형을 적용시킨다. 모형들의 차수 선택은 BIC를 기준으로 최적 모형을 선택한다.

이중계절형 ARIMA모형은 모형의 잔차 적합 검정에 만족하고 BIC(Basian information criterion)값을 기준으로 최소인 모형을 선택하였다. 예측된 모형은 이중계절형 모형보다 일별주기를 고려한 계절형 모형이 더 작은 BIC 수치로 나타남에 따라 ARIMA(1, 0, 1)×(1, 0, 1)<sub>288</sub>을 최종 계절형 ARIMA 모형으로 사용한다. FARIMA 모형도 BIC가 최소인 모형을 선택하여, 실수값으로 차분된 FARIMA(1,  $\hat{d}$ , 1)모형을 선택한다. 이 때의  $\hat{d}$ 의 추정값은 0.45419로 나타난다. Holt-Winters모형은 최적화된 가중치를 적용하여 모형을 결정한다. 이에 따라 최종 선택된 모형들의 모수 추정치는 Table 4.1부터 Table 4.3까지와 같다.

적용된 모형들에 대하여 본 논문에서는 결합 예측 방법으로서 SA(Simple Average), OLS, ERLS 방법 및 MSE(Mean Squared Error)에 비율에 따라서 결합 예측을 실시한다. Table 4.4의 결과를 보면 3시간 이후 예측에서는 계절형 ARIMA모형이 가장 낮은 RMSE로서 나타났으며, 그 다음으로는 단순결합

모형과 MSE기반의 모형이 좋은 예측력으로 보여졌다. 6시간 이후에 예측에서 단순결합 방법이 가장 작은 RMSE로서 사용된 방법이 단순하면서도 가장 좋은 예측력으로 보여졌으며, 다음으로 MSE기반에 결합 예측이 좋은 것으로 나타났다. 9시간 이후에 대한 예측에서는 MSE 기반에 결합 예측모형이 가장 높은 예측으로 보여졌으며, 다음으로 OLS 방법이며, 그 다음으로 단순결합 방법이 좋은 예측력을 보였다. 6시간에서 12시간 이후에 예측에서 결합 예측 방법들이 전반적으로 한 개에 모형을 통한 예측보다 작은 RMSE로서 좋은 예측 결과를 보여준다.

## 5. 결론

본 연구는 인터넷 트래픽 사용량에 대하여 시계열 모형 적용 및 분석을 통한 트래픽 변화 대응이 관건이다. 이에 대하여, 단변량 모형을 통하여 모형의 최적 예측을 사용할 것인지, 결합된 예측을 사용할 것인지에 대하여 생각해 보았으며,  $k$  시점 이후 예측에 따른 비교로 예측 기간에 따라 적합한 모형에 대하여 비교를 실시하였다.

이에 대하여 3시간 이후 예측에 대하여 최적 모형으로는 ARIMA 모형이 최적모형으로 나타났지만, 6시간 이후에 있어서는 오히려 결합 예측 모형이 최적모형으로서 적합하게 나타났다. 이는 향후 실제 분야에 적용 시 대응 시간에 따라서 최적 모형을 사용할 것인지, 결합예측 방법으로 사용할 것인지 결정하는데 도움이 될 것으로 사료된다. 특히 단순 결합 예측 방법은 6시간까지의 예측에서 결합 예측 방법 중에서는 가장 높은 예측 결과를 보여주어, 실제 분야에서 쉽게 사용할 수 있는 방법인 동시에 좋은 예측력을 보여준다.

향후 본 연구에 대한 확장 형태로 이분산성을 고려한 모형들의 결합 모형에 대한 연구 및 훈련자료의 길이에 따른 예측력에 영향이 있는지에 대한 연구가 필요하다고 생각되며, 실제 분야에 적용을 위한 자동화 방법들이 연구되어야 할 것으로 판단된다.

## References

- Aksu, C. and Bunter, S. (1992). An empirical analysis of the accuracy of SA, OLS, ERLS and NRLS combination forecasts, *International Journal of Forecasting*, **8**, 27–43.
- Armstrong, J. S. (2001). Combining forecasting, *International Series in Operations Research and Management Science*, 417–440.
- Au, S. T., Ma, G. Q. and Yeung, S. N. (2011). Automatic forecasting of double seasonal time series with applications on mobility network traffic prediction, In *2011 Joint Statistical Meetings*, July.
- Box G. E. P. and Jenkins, G. (1994). *Time Series Analysis: Forecasting and Control*, Prentice Hall.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U. K. Inflation, *Econometrica*, **50**, 987–1008.
- Ha, M. and Kim, S. (2008). Internet traffic forecasting using power transformation heteroscedastic time series models, *The Korea Journal of Applied Statistics*, **21**, 1037–1044.
- Jiang, J. and Papavassiliou, S. (2006). Enhancing network traffic prediction and anomaly detection via statistical network traffic separation and combination strategies, *Computer Communications*, **29**, 1627–1638.
- Kim, S. (2007). Time series models for performance evaluation of network traffic forecasting, *The Korea Journal of Applied Statistics*, **20**, 219–227.
- Kim, S. (2011). Forecasting internet traffic by using seasonal GARCH models, *Journal of Communications and Network*, **13**, 621–624.
- Liu, J., Shu, Y., Zhang, L. and Xue, F. (1999). Traffic modeling based on FARIMA models, In *IEEE Canadian Conference on Electrical and Computer Engineering*, 621–624.
- Shu, Y. T., Yu, M. F., Liu, J. K. and Yang, O. W. W. (2003). Wireless traffic modeling and prediction

- using seasonal ARIMA models, *Communications, 2003. ICC'03. IEEE International Conference on. IEEE*, **3**, 1675–1679.
- Sohn, H. and Kim, S. (2013). A study on the short term internet traffic forecasting models on long-memory and heteroscedasticity, *The Korea Journal of Applied Statistics*, **26**, 1053–1061.
- Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing, *Journal of the Operational Research Society*, **54**, 799–805.
- Tikunov, D. and Nishimura, T. (2007). Traffic prediction for mobile network using Holt-Winter's exponential smoothing, *Telecommunications and Computer Networks, 2007. SoftCOM 2007. 15th International Conference*, 1–5.
- Yu, Y. H., Wang, J., Song, M. N. and Song, J. D. (2010). Network traffic prediction and result analysis based on seasonal ARIMA and correlation coefficient, *2010 International Conference on Intelligent System Design and Engineering Application*, <http://dx.doi.org/10.1109/ISDEA.2010.335>.
- Zhou, B., He, D., Sun, Z. and Ng, W. H. (2005). Network traffic modeling and prediction with ARIMA/GARCH, In *Proceeding of HET-NETs Conference*, 1–10.



# 결합예측 방법을 이용한 인터넷 트래픽 수요 예측 연구

김삼용<sup>a,1</sup>

<sup>a</sup>중앙대학교 응용통계학과

(Received November 2, 2015; Revised November 16, 2015; Accepted November 23, 2015)

---

## Abstract

최근 들어 ICT 분야의 발달에 따라 데이터 사용량의 급격한 증가로 인터넷 트래픽 사용량 예측은 중요성은 강조되고 있다. 이러한 예측치를 적절한 트래픽 관리와 제어를 위한 계획 수립에 도움을 준다.

본 논문은, 5분 단위의 인터넷 트래픽 자료를 이용하여 결합 예측 모형을 제안하고자 한다. 이에 대하여 시계열의 대표적인 3개 모형인 Seasonal ARIMA, Fractional ARIMA(FARIMA), Taylor의 수정된 Holt-Winters 모형을 적용하였다. 모형 간 결합 예측 방법으로 예측치 간의 SA(Simple Average) 결합 예측 방법과 OLS(Ordinary Least Square)를 이용한 결합방법, ERLS(Equality Restricted Least Squares)를 이용한 결합 예측 방법, Armstrong (2001)이 제안한 MSE 기반 결합 예측 방법을 사용한다. 이에 따른 결과로서 3시간에서의 예측은 Seasonal ARIMA가 선택된 반면, 6시간 이후 예측에서는 결합 예측 방법이 좋은 예측 성능을 보여준다.

주요용어: Fractional Seasonal ARIMA, adjusted Holt-Winters, internet traffic, combined forecasting

---

---

이 논문은 2013년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (NRF-2013R1A1A2008277).

<sup>1</sup>(06974) 서울 동작구 흑석동 221, 중앙대학교 응용통계학과. E-mail: sahm@cau.ac.kr