

# 단기 측정 인터넷 트래픽 예측을 위한 모형 성능 비교 연구

하명호<sup>a</sup>, 손흥구<sup>a</sup>, 김삼용<sup>1,a</sup>

<sup>a</sup>중앙대학교 응용통계학과

---

## 요약

본 연구에서는 단기에 측정되는 트래픽 자료를 예측하기 위하여 Holt-Winters, Fractional Seasonal ARIMA, AR-GARCH, Seasonal AR-GARCH 모형을 사용하여 각 모형의 예측 성능을 비교하고자 한다. 예측에 이용된 시계열 모형에 대해 소개하고, 실제 트래픽 자료에 적용하여 트래픽 자료를 분석한 결과 Holt-Winters 방법이 예측력 측면에서 가장 우수하였다.

주요용어: 트래픽, AR-GARCH, Fractional Seasonal ARIMA, Seasonal AR-GARCH, Holt-Winters.

---

## 1. 서론

인터넷과 모바일 트래픽 산업의 급속한 성장으로 인하여 트래픽 예측에 대한 중요성은 통신사업자나 트래픽 관리자에게 매우 중요한 문제가 되고 있다. 이러한 요구에 부응하여 인터넷 트래픽, 모바일 트래픽 예측 연구에 많은 연구자들이 다양한 통계적 방법을 사용하여 연구를 수행하고 있다.

우선 Basu 등 (1996)은 자기회귀(Autoregressive) 모형을 도입하여 정상성을 만족하는 트래픽 자료를 예측하였고, Liu 등 (1999)은 FARIMA(Fractional Autoregressive Integrated Moving Average) 모형을 이용하여 트래픽을 예측하였다. Shu 등 (2005)은 계절형 ARIMA를 이용하여 무선 트래픽을 예측하였으며, Tikunov와 Nishimura (2007)은 모바일 트래픽 자료를 예측하는데 Holt-Winters의 방법을 사용하였다. Kim (2007)은 Engle (1982)이 제안한 자기회귀이분산성(ARCH; Autoregressive Conditional Heteroskedasticity) 모형을 이용하여 자기회귀이분산성 모형이 트래픽 자료를 예측하는데 있어 AR 모형보다 성능이 우수함을 보였고, Kim (2011)은 Seasonal AR-GARCH 모형을 이용하여 Seasonal AR-GARCH이 Seasonal ARIMA와 AR-GARCH 모형보다 트래픽 자료를 예측하는데 성능이 우수함을 보였다.

본 연구에서는 시계열 자료를 예측하는데 있어 많이 사용되는 Holt-Winters 방법과 장기 기억 특성을 설명할 수 있는 Fractional Seasonal ARIMA 방법, 분산의 이분산성을 설명할 수 있는 Seasonal AR-GARCH 모형을 이용하여 트래픽의 예측 성능을 비교하고자 한다.

## 2. 시계열 모형의 소개

### 2.1. Holt-Winters 모형

Holt-Winters (1960)의 승법적 계절지수평활(multiplicative seasonal exponential smoothing)은 Holt의 이중 모수 지수평활법을 확장시킨 방법으로 관측된 시계열이 선형추세성과 승법적 계절변동을 나타낼

---

이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2011-0003955).

<sup>1</sup> 교신저자: 서울시 동작구 흑석동 221, 중앙대학교 응용통계학과, 교수. E-mail: sahm@cau.ac.kr

때 적용하는 방법이다. Holt-Winters 승법 모형은 식 (2.1)과 같다.

$$\begin{aligned} L_t &= \alpha \left( \frac{Y_t}{S_{t-s}} \right) + (1 - \alpha)(L_{t-1} + T_{t-1}), \\ T_t &= \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}, \\ S_t &= \gamma \left( \frac{Y_t}{L_t} \right) + (1 - \gamma)S_{t-s}, \end{aligned} \quad (2.1)$$

여기서  $Y_t$ 는  $t$ 시점에서의 관측값,  $L$ 은 평활 수준을 나타내며,  $T$ 는 평활가법추세를 의미하며,  $S$ 는 계절 지표의 주기,  $s$ 는 계절 주기를 의미한다.  $\alpha, \beta$ 는 수준과 추세에 대한 모수이고  $\gamma$ 는 계절평활 모수로 가중치를 나타낸다. 평활 모수  $\alpha, \beta, \gamma$ 는 1차 예측단계에서 MSE를 최소화하는 최적값을 적용한다. 초기 값은 식 (2.2)의 수식에 따라 구해 질 수가 있다.

$$\begin{aligned} L_s &= \frac{1}{s} \sum_{t=1}^s Y_t \\ T_k &= \frac{1}{k} \left( \frac{Y_{s+1} - Y_1}{s} + \frac{Y_{s+2} - Y_2}{s} + \dots + \frac{Y_{s+k} - Y_k}{s} \right) \\ S_k &= \frac{Y_k}{L_s}, \quad k = 1, 2, \dots, s. \end{aligned} \quad (2.2)$$

식 (2.1)과 식 (2.2)를 사용하여 식 (2.3)과 같이 예측 할 수 있다.

$$F_{t+h} = (L_t + T_t \times h)S_{t+h-s} \quad (2.3)$$

이때  $F_{t+h}$ 는  $h$ 시점 후에 예측한 값을 의미한다.

## 2.2. Fractional Seasonal ARIMA 모형

Fractional Seasonal ARIMA(FSARIMA)는 표준 Seasonal ARIMA( $p, d, q$ ) 모형의 일종으로서  $d$ 가 자연수가 아닌 실수를 가지는 모형이다. FSARIMA( $p, d, q$ )에 대해 시계열  $X_t$ 는 다음과 같이 정의된다.

$$\Phi(B)\Phi'_s(B)\Delta^d X_t = \Theta(B)\Theta'_s(B)a_t, \quad (2.4)$$

여기서  $s$ 는 계절 성분의 주기이고,  $a_t$ 는 백색잡음이며 모형의 정상성을 위해  $d \in (-0.5, 0.5)$ 이다. 또한  $B$ 는 후진연산자(backward-shift operator)이고,

$$\begin{aligned} \Phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \Phi'_s(B) &= 1 - \phi_s B^s - \phi_{2s} B^{2s} - \dots - \phi_{ps} B^{ps} \\ \Theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \\ \Theta'_s(B) &= 1 - \theta_s B^s - \theta_{2s} B^{2s} - \dots - \theta_{qs} B^{qs} \end{aligned}$$

이며,  $\Delta^d$ 는 분수차분 연산자(fractional differencing operator)로 다음과 같이 정의된다.

$$\Delta^d = (1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k, \quad (2.5)$$

여기서  $\binom{d}{k} = d! / \{(d - k)!k!\}$ 이다. 만약  $d = 0$ 이면, FSARIMA 모형은 일반적인 SARMA 모형이 된다.

### 2.3. AR-GARCH 모형

AR-GARCH 모형은 AR 모형에 오차의 분산이 자기회귀적으로 변하는 조건부 이분산 자기회귀모형인 ARCH 모형의 일반화된 모형으로 본 연구에서 고려되는 AR-GARCH 모형은 다음과 같다.

$$\begin{aligned}\Phi(B)\phi_s(B)X_t &= \varepsilon_t \\ \varepsilon_t &= \sqrt{h_t} e_t \\ e_t &\sim iid(0, 1) \\ h_t &= \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}\end{aligned}\quad (2.6)$$

여기서  $\alpha_0 > 0, \alpha_i \geq 0 (i = 1, \dots, q), \beta_j \geq 0 (j = 1, \dots, p), p \geq 0, q > 0$ 이다.

만약  $p = 0$ 이면 GARCH( $p, q$ ) 모형은 ARCH( $q$ )로 축소된다. 또  $p = q = 0$ 이면 GARCH( $p, q$ ) 모형은  $\varepsilon_t \sim WN(0, \alpha_0)$ 가 되어 오차항의 분산이 상수로 일정한 즉 이분산성이 사라지게 된다.

### 2.4. Seasonal AR-GARCH 모형

Seasonal AR-GARCH 모형은 식 (2.7)과 같다.

$$\begin{aligned}\Phi(B)\phi_s(B)X_t &= a_t \\ a_t &= \varepsilon_t \sqrt{h_t} \\ h_t &= \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j} + \sum_{i=1}^{q'} \alpha_{is} a_{t-is}^2 + \sum_{j=1}^{p'} \beta_{js} h_{t-js}\end{aligned}\quad (2.7)$$

여기서  $s$ 는 계절 성분의 주기이며,  $\alpha_0 > 0, \alpha_i \geq 0 (i = 1, \dots, q), \beta_j \geq 0 (j = 1, \dots, p), \alpha_{is} \geq 0 (i = 1, \dots, q'), \beta_{js} \geq 0 (j = 1, \dots, p'), p \geq 0, q \geq 0, p' \geq 0, q' \geq 0$ 이다. 계절 성분을 고려하는 중요한 이유는 인터넷 트래픽 자료가 계절 패턴을 보이기 때문이다.

## 3. 모형의 성능 분석 결과

본 논문에서 이용한 자료는 Dingde Jiang와 Guangmin Hu (2009)이 사용한 것으로 6개월 동안 12개 도시에서 서로 주고 받은 5분 단위로 측정된 트래픽 양을 미국 애빌린주에서 모은 자료이다. 실제 분석에 사용한 자료는 23일치 6624개를 사용하였다. 모형 적합에 있어 계절 주기는 1일에 해당하는 288을 사용하였고, 전체 자료중에 21일치에 해당하는 6048개를 이용하여 모형을 적합하고, 2일치 자료 576개에 대해 적합 모형을 이용하여 예측하였다. 이 자료는 <http://www.cs.utexas.edu/users/szyzhang/research/AbileneTM/>에서 다운 받을 수 있다. 모형 적합과 정상성을 만족시키기 위하여 Fractional Seasonal ARIMA, AR-GARCH, Seasonal AR-GARCH에 대해 식 (3.1)과 같이 변수 변환하여 1차 차분과 계절 차분을 실시하여 모수를 추정 하였다.

$$Z_t = \log\left(\frac{Y_t}{Y_t - 1}\right), \quad X_t = Z_t - Z_{t-288}. \quad (3.1)$$

이와 같은 자료에 대한 시계열 그림은 그림 1과 그림 2와 같다. 그림 1은 6624개에 대한 원자료이고, 그림 2는 식 (3.1)과 같이 원 자료에 대해 로그 차분한 자료이다.

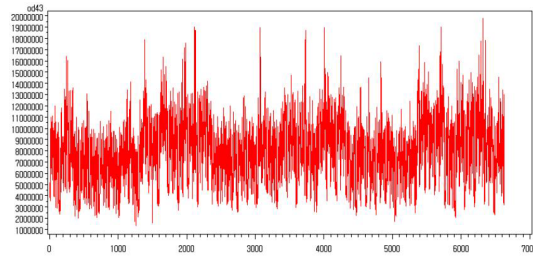


그림 1: 원 자료의 시계열 그림

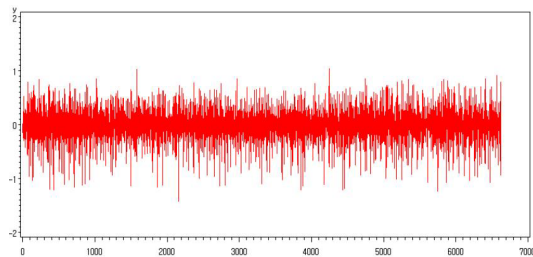


그림 2: 로그 차분한 시계열 그림

표 1: LM test 결과

Order	LM	Pr > LM
1	159.8028	< .0001
2	179.4223	< .0001
3	181.2740	< .0001
4	182.0671	< .0001
5	183.9434	< .0001
6	183.9779	< .0001
7	184.0611	< .0001
8	184.2419	< .0001
9	185.8451	< .0001
10	185.8527	< .0001
11	186.3341	< .0001
12	209.1624	< .0001

추정 방법에 있어 홀트-윈터스 방법은 원 자료를 이용하여 모수를 추정하였다. 추정 결과 ‘TREND’ 효과가 유의 하지 않고, AIC의 값이 ‘TREND’를 모형에서 제거하였을 경우 더 작기 때문에 예측값은 ‘TREND’ 효과가 없는 모형을 이용하였다. FARIMA 모형에서는 로그 변환한 자료에 대하여 를 추정하고, Seasonal ARIMA 모형을 이용하여 나머지 모수를 추정하였다.

이분산성 모형에서는 이분산성이 존재하는지에 대해 LM test를 실시하였다. 그 결과는 표 1과 같다. LM test 결과 이분산성이 존재하는 것으로 나타나 AR-GARCH, Seasonal AR-GARCH도 고려하였으며, 식 (3.1)과 같이 로그 차분한 자료를 이용하고, 모형 선택 기준은 AIC를 기준으로 선택하였다. Fractional Seasonal ARIMA, AR-GARCH, Seasonal AR-GARCH에 대한 AIC는 표 2에서 표 5와 같다.

그 결과 Holt-Winters 모형에서는 ‘TREND’가 없는 모형이 선택되었으며, Fractional Seasonal ARIMA 모형에서는  $ARIMA(3, -0.3884, 1)(3, 0, 1)_{288}$ , AR-GARCH 모형은  $AR(3)(3)_{288}-GARCH(1, 1)$ , Seasonal AR-GARCH 모형에서는  $AR(3)(2)_{288}-GARCH(1, 1)(1, 1)_{288}$  이 선택되었다. 선택된 모형에 대

표 2: Fractional Seasonal ARIMA 모형에 따른 AIC

모형	AIC
FARIMA(1, $d$ , 1)(1, 0, 1) <sub>288</sub>	-2998.51
FARIMA(2, $d$ , 1)(1, 0, 1) <sub>288</sub>	-2998.19
FARIMA(1, $d$ , 1)(2, 0, 1) <sub>288</sub>	-2998.49
FARIMA(2, $d$ , 1)(2, 0, 1) <sub>288</sub>	-3011.40
FARIMA(2, $d$ , 1)(3, 0, 1) <sub>288</sub>	-3019.57
FARIMA(3, $d$ , 1)(3, 0, 1) <sub>288</sub>	-3086.56

\*  $d$ 의 값을 추정된 결과  $\hat{d} = -0.3884$ 로 나타남

표 3: AR-GARCH 모형에 따른 AIC

모형	AIC
AR(1)(1) <sub>288</sub> -GARCH(1, 1)	-976.42
AR(2)(1) <sub>288</sub> -GARCH(1, 1)	-1223.55
AR(1)(2) <sub>288</sub> -GARCH(1, 1)	-1449.52
AR(3)(1) <sub>288</sub> -GARCH(1, 1)	-1222.71
AR(3)(2) <sub>288</sub> -GARCH(1, 1)	-1672.54
AR(3)(3) <sub>288</sub> -GARCH(1, 1)	-1837.29
AR(4)(1) <sub>288</sub> -GARCH(1, 1)	-1717.95
AR(2)(3) <sub>288</sub> -GARCH(1, 1)	-1836.39

표 4: Seasonal AR-GARCH 모형에 따른 AIC

모형	AIC
AR(1)(1) <sub>288</sub> -GARCH(1, 1)(1, 1) <sub>288</sub>	-2213.73
AR(2)(1) <sub>288</sub> -GARCH(1, 1)(1, 1) <sub>288</sub>	-2511.10
AR(1)(2) <sub>288</sub> -GARCH(1, 1)(1, 1) <sub>288</sub>	-2589.40
AR(3)(1) <sub>288</sub> -GARCH(1, 1)(1, 1) <sub>288</sub>	-2831.26
AR(3)(2) <sub>288</sub> -GARCH(1, 1)(1, 1) <sub>288</sub>	-2984.43
AR(3)(3) <sub>288</sub> -GARCH(1, 1)(1, 1) <sub>288</sub>	-2884.17

표 5: 홀트-윈터스 모수 추정 결과

Model	Variable	Estimate	Standard Error	$t$ Value	$p$ value	AIC
1	LEVEL	0.649314	0.008441	76.92162	< 0.001	190841
	TREND	0.001000	0.004293	0.23292	0.815834	
	SEASONAL	0.022781	0.002151	10.58905	< 0.001	
2	LEVEL	0.648905	0.008429	76.98337	< 0.001	190834
	SEASONAL	0.022739	0.002149	10.57944	< 0.001	

한 모수 추정값은 표 5에서 표 8과 같다.

추정된 각 모형의 모수를 이용하여 예측을 가장 잘하는 모형을 판단하기 위해 2일치 자료 576개의 자료에 대해 예측을 하였다. 예측에 대한 성능은 식 (3.2)와 같은 RMSE를 이용하였다.

$$RMSE = \sqrt{\sum_{t=n+1}^{n+n'} \frac{1}{n'} (Y_t - \hat{Y}_t)^2}, \tag{3.2}$$

여기서  $n$ 은 모형 적합을 위해 사용된 표본 수이고,  $n'$ 은 예측을 위해 사용된 표본 수이며,  $Y_t$ 는  $t$ 시점에서의 실제 값,  $\hat{Y}_t$ 는  $t$ 시점에 대한 예측값이다. 각 모형에 대해 식 (3.2)를 이용해 산출한 RMSE는 표 9와 같다. 표 9의 결과를 보면 홀트-윈터스 모형이 가장 잘 적합이 되는 것으로 나타났으며, 그 다음으

표 6: Fractional Seasonal ARIMA 모수 추정 결과

Variable	Estimate	Standard Error	t Value	p value
$d$	-0.3884			
$\theta_1$	-0.0796	0.0134	-5.93	< .0001
$\theta_2$	-0.0803	0.0131	-6.12	< .0001
$\theta_3$	0.8995	0.0130	69.18	< .0001
$\theta_{1(288)}$	0.7760	0.0110	70.05	< .0001
$\phi_1$	-0.1878	0.0185	-10.11	< .0001
$\phi_2$	-0.1858	0.0185	-10.04	< .0001
$\phi_3$	0.8081	0.0184	43.82	< .0001
$\phi_{1(288)}$	0.9932	0.0045	216.82	< .0001

표 7: AR-GARCH 모수 추정 결과

Variable	Estimate	Standard Error	t Value	p value
$\phi_1$	0.2101	0.0110	19.06	< .0001
$\phi_2$	0.1629	0.0109	14.97	< .0001
$\phi_3$	0.0179	0.0108	1.65	0.099
$\phi_{1(288)}$	-0.3081	0.0093	-33.22	< .0001
$\phi_{2(288)}$	-0.2152	0.0105	-20.54	< .0001
$\phi_{3(288)}$	-0.1716	0.0107	-16.11	< .0001
$\alpha_0$	0.0160	0.0010	15.32	< .0001
$\alpha_1$	0.2462	0.0160	15.42	< .0001
$\beta_1$	0.4236	0.0280	15.14	< .0001

표 8: Seasonal AR-GARCH 모수 추정 결과

Variable	Estimate	Standard Error	t Value	p value
$\phi_1$	0.2620	0.0116	22.61	< .0001
$\phi_2$	0.1772	0.0106	16.74	< .0001
$\phi_3$	0.0140	0.0095	1.47	0.1411
$\phi_{288}$	-0.3167	0.0131	-24.16	< .0001
$\phi_{576}$	-0.2499	0.0140	-17.89	< .0001
$\alpha_0$	0.0009	0.0002	3.40	0.0007
$\alpha_1$	0.0861	0.0078	10.99	< .0001
$\alpha_{288}$	0.1057	0.0101	10.51	< .0001
$\beta_1$	0.0224	0.0099	2.25	0.0244
$\beta_{288}$	0.7689	0.0174	44.15	< .0001

표 9: 각 모형에 따른 RMSE

	Holt-Winters	FSARIMA	AR-GARCH	Seasonal AR-GARCH
자료 1	0.3141	0.3656	0.4134	0.4004

로 트래픽의 성질이 장기 기억을 설명할 수 있는 FSARIMA의 모형의 적합도가 우수한 것으로 나타났다.

#### 4. 결론

일반적으로 트래픽 자료는 이분산성 뿐만 아니라 장기 기억 특성을 가지고 있는 것으로 알려져 있다. 본 연구에서는 지수 평활법을 확장시킨 홀트-윈터스(Holt-Winters) 방법, 장기 기억 특성을 설명할

수 있는 FARIMA 방법, 이분산성을 설명할 수 있는 AR-GARCH 모형, AR-GARCH 모형에 계절성분을 포함시킨 Seasonal AR-GARCH을 이용하여 실제 자료에 적합해 보았다. 그 결과 홀트-윈터스 모형의 적합성이 가장 우수 하였고, 그 다음으로 FARIMA 모형의 적합성이 우수 하였다. 향후에 좀 더 짧은 주기의 다양한 자료의 계절성과 특성을 고려한 모형의 성능평가를 시행하여 예측의 정확도와 신뢰도를 높이고자 한다.

### 참고 문헌

- Basu, A., Mukherjee, A. and Klivansky, S. (1996). Time series models for internet traffic, In *Proceedings IEEE INFOCOM 96, Fifteenth Annual Conference of the IEEE Computer Societies*, **4**, 24–28.
- Dingde, J. and Guangmin, H. (2009). GARCH model-based large-scale IP traffic matrix estimation, *Communications Letters, IEEE*, **13**, 52–54.
- Engle, R. F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of U. K. Inflation, *Econometrica*, **50**, 987–1008.
- 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**, .
- Liu, J., Shu, Y., Zhang, L. and Xue, F. (1999). Traffic modeling based on FARIMA models, *IEEE Conf. on Electrical and Computer Engineering*, 162–167.
- Shu, Y., Yu, M., Yang, O., Liu, J. and Feng, H. (2005). Wireless traffic modeling and prediction using seasonal ARIMA models, *IEICE-transactions on Communications*, **10**, 3992–3999.
- Tikunov, D. and Nishimura, T. (2007). *Traffic Prediction for Mobile Network using Holt-Winter's Exponential Smoothing*, Telecommunications and Computer Networks, 15th International Conference, 1–5.
- Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages, *Management Science*, **6**, 324–342.

## A Study on Performance Analysis of Short Term Internet Traffic Forecasting Models

M. H. Ha<sup>a</sup>, H. G. Son<sup>a</sup>, S. Kim<sup>1,a</sup>

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

---

### Abstract

In this paper, we first compare the performance of Holt-Winters, FSARIMA, AR-GARCH and Seasonal AR-GARCH models with in the short term based data. The results of the compared data show that the Holt-Winters model outperformed other models in terms of forecasting accuracy.

Keywords: Internet traffic, AR-GARCH, Fractional Seasonal ARIMA, Seasonal AR-GARCH, Holt-Winters.

---

---

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 (2010-0010237).

<sup>1</sup> Corresponding author: Professor, Department of Applied Statistics, Chung-Ang University, Dongack-Gu, Seoul 156-756, Korea. E-mail: sahm@cau.ac.kr