

# 인터넷 트래픽 예측 모형 성능 분석 연구

김삼용<sup>1</sup> · 하명호<sup>2</sup> · 정재윤<sup>3</sup>

<sup>1</sup>중앙대학교 통계학과, <sup>2</sup>중앙대학교 통계학과, <sup>3</sup>중앙대학교 통계학과

(2011년 1월 접수, 2011년 2월 채택)

## 요약

본 연구에서는 인터넷 트래픽 자료를 예측하는데 사용되는 Holt-Winters, FARIMA, AR-GARCH 모형을 트래픽 예측에 적용하여 각 모형을 성능을 비교하고자 한다. 각 시계열 모형에 대해 소개하고, 트래픽 자료의 특성인 장기 기억 특성을 설명하는데 적합한 모형을 알아보기 위해 실제 트래픽 자료에 적용하여 예측 성능을 비교하였다.

주요어: 트래픽, 장기기억, Holt-Winters, FARIMA, AR-GARCH.

## 1. 서론

현재 인터넷은 생활의 일부분으로 자리잡아 모든 분야에서 가장 기본이 되는 역할을 하고 있다. 이러한 발전에 따라 인터넷의 중요성이 커지고 있으며 이는 통신 사업자에게 있어 트래픽의 관리는 중요한 문제가 되고 있다. 이러한 요구에 상응하여 많은 연구자들이 인터넷 트래픽의 효율적인 분석과 예측을 위해 다양한 방법들을 연구하고 있고, 통계적 방법들도 도입되어 사용하고 있다. 우선 Basu 등 (1996)은 정상성을 만족하는 트래픽 자료를 예측하는데 자기회귀(Autoregressive) 모형을 도입하였고, Shu 등 (2005)은 계절형 ARIMA를 이용하여 무선 트래픽을 예측하였다. 또한, Tikunov와 Nishimura (2007)은 Holt-Winters의 방법을 이용하여 모바일 트래픽 자료를 예측하였으며, Kim (2007)은 Engle (1982)이 제안한 자기회귀이분산성(ARCH; Autoregressive Conditional Heteroskedasticity) 모형을 이용하여 트래픽 자료를 예측하는데 있어 AR 모형보다 성능이 우수함을 보였다. 또한 Liu 등 (1999)은 FARIMA(Fractional Autoregressive Integrated Moving Average) 모형을 이용하여 트래픽을 예측하였다. 본 연구에서는 시계열 자료를 예측하는데 있어 많이 사용되는 Holt-Winters 방법과 장기기억 특성을 설명할 수 있는 FARIMA 방법, 분산의 이분산성을 설명할 수 있는 GARCH 모형을 이용하여 트래픽의 예측 성능을 비교하고자 한다.

## 2. 시계열 모형의 소개

### 2.1. Short-memory와 Long-memory

정상 시계열 자료  $X_t$ 의  $k$ 시차의 ACF(AutoCorrelation Function)  $\rho_k$ 가 다음과 같이 합이 유한일 때 시계열 자료  $X_t$ 는 단기 기억(Short Memory)를 갖는다고 한다.

$$\sum_{k=0}^{\infty} |\rho_k| < \infty. \quad (2.1)$$

이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2010-0010237).

<sup>1</sup>교신저자: (156-756) 서울시 동작구 흑석동 221, 중앙대학교 통계학과, 교수. E-mail: sahm@cau.ac.kr

만약 시계열 자료  $X_t$ 의  $k$ 시차의 ACF  $\rho_k$ 가 다음과 같이 합이 유한하지 않으면 시계열 자료  $X_t$ 는 장기 기억(Long Memory)를 갖는다고 한다.

$$\sum_{k=0}^{\infty} |\rho_k| = \infty. \quad (2.2)$$

일반적으로 인터넷 트래픽 자료는 장기 기억의 특성을 가지는 것으로 알려져 있다.

## 2.2. 홀트-윈터스 지수 평활(Holt-Winters exponential smoothing) 모형

Holt-Winters (1960)의 승법적 계절지수평활(multiplicative seasonal exponential smoothing)은 Holt의 이중 모수 지수평활법을 확장시킨 방법으로 관측된 시계열이 선형추세성과 승법적 계절변동을 나타낼 때 적용하는 방법이다. 홀트-윈터스(Holt-Winters)의 승법적 계절지수평활법은 다음과 같이 시계열 패턴의 세가지 성분인 수평성, 추세성 그리고 계절성을 평활시키는 세 개의 방정식과 예측식으로 구성된다. 즉 현재 시점이  $n$ 인 경우에 1시점 후의 예측값  $F_{n+l}$ 은 다음과 같다.

$$F_{n+l} = (a_n + b_n l) S_{n+l-L}, \quad l = 1, 2, \dots, L. \quad (2.3)$$

식 (2.3)에서 각각의 요소는  $a_n$ 은 수평성분,  $b_n$ 은 추세성분,  $S_n$ 은 계절성분이고 다음과 같은 식으로 정의된다.

$$\begin{aligned} a_n &= \alpha \frac{Z_n}{S_{n-L}} + (1 - \alpha)(a_{n-1} + b_{n+1}), \\ b_n &= \beta (a_n - a_{n-1}) + (1 - \beta)b_{n-1}, \\ S_n &= \gamma \frac{Z_n}{a_n} + (1 - \gamma)S_{n-L}, \end{aligned}$$

여기서  $Z_n$ 은 원시계열 자료이고,  $L$ 은 계절성의 길이,  $\alpha, \beta, \gamma$ 는 평활상수를 나타낸다.

## 2.3. FARIMA 모형

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

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

여기서  $a_t$ 는 백색잡음이고 FARIMA 모형의 정상성을 위해  $d \in (-0.5, 0.5)$ 이며,

$$\begin{aligned} \Phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \Theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \end{aligned}$$

이다. 또한  $p$ 와  $q$ 는 음이 아닌 정수이며,  $B$ 는 후진연산자(backward-shift operator)이고,  $\Delta^d$ 는 분수차분 연산자(fractional differencing operator)로 다음과 같이 정의된다.

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

만약  $d = 0$ 이면, FARIMA( $p, d, q$ ) 모형은 일반적인 ARMA( $p, q$ ) 모형이 된다. 여기서 모수  $d$ 는 Hurst 모수를 추정함으로써 구할 수 있으며, 실제로  $H = d + 0.5$ 인 관계가 있다. 또한  $H$ 의 값이 1에 가까우면 강한 장기기억 특성을 가진다.

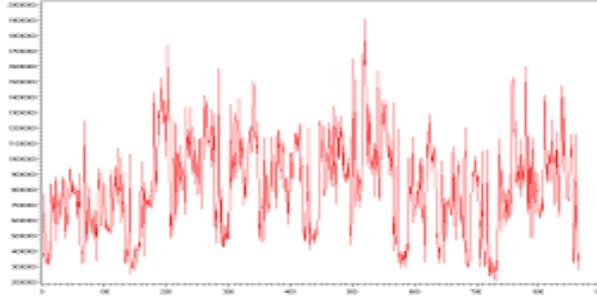


그림 3.1. 자료 1의 원자료 시계열 그림

### 2.4. GARCH 모형

GARCH 모형은 오차의 분산이 자기회귀적으로 변하는 조건부 이분산 자기회귀모형인 ARCH 모형의 일반화된 모형으로 GARCH( $p, q$ ) 모형은 다음과 같다.

$$\begin{aligned} \epsilon_t | \Psi_{t-1} &\sim N(0, h_t) \\ \text{Var}(\epsilon_t | \Psi_{t-1}) = h_t &= \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i} \\ &= \alpha_0 + A(L)\epsilon_t^2 + B(L)h_t, \end{aligned} \tag{2.6}$$

여기서  $A(L) = \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_q L^q$ ,  $B(L) = \beta_1 L + \beta_2 L^2 + \dots + \beta_p L^p$ 이고,  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$  ( $i = 1, 2, \dots, q$ ),  $\beta_i \geq 0$  ( $i = 1, 2, \dots, p$ ),  $p \geq 0$ ,  $q > 0$ 이다. GARCH( $p, q$ ) 모형은 조건부 분산이 지난 시점의 분산들의 선형결합과 지난 시점의 조건부 분산의 선형결합으로 표현됨을 알 수 있다.  $p = 0$ 이면 GARCH( $p, q$ ) 모형은 ARCH( $q$ )로 축소된다. 또  $p = q = 0$ 이면 GARCH( $p, q$ ) 모형은  $\epsilon \sim WN(0, \alpha_0)$ 가 되어 오차항의 분산이 상수로 일정한 즉 이분산성이 사라지게 된다.

GARCH( $p, q$ )는 ARCH( $\infty$ )로 표현할 수 있는데 ARCH( $\infty$ ) 모형은 차수가 너무 커서 추정해야 할 모수가 많아져 비효율적이다. 그래서 같은 의미를 가지는 GARCH( $p, q$ ) 모형을 이용하면 적은 개수의 모수를 이용하여 설명 가능하다.

### 3. 모형의 성능 분석 결과

본 논문에서 이용한 자료는 6개월 동안 미국 애빌린주에서 모은 자료로 12개 도시에서 서로 주고 받은 인터넷 트래픽 자료로써 Dingde와 Guangmin (2009)이 사용하였다. 자료는 5분 단위로 측정된 트래픽 양이며, 실제 분석에 사용한 자료는 3일치 자료 864개에 대해 3개의 자료를 사용하였다. 이 자료는 <http://www.cs.utexas.edu/usersyzhang/research/AbileneTM/>에서 다운 받을 수 있다. 이러한 방법으로 구한 자료에 대해 다음과 같은 방법으로 변수 변환하여 모수를 추정하였다.

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

그림 3.1, 3.3, 3.5는 원자료의 시계열 그림이고, 그림 3.2, 3.4, 3.6은 식 (3.1)과 같은 방법으로 로그 차분한  $Y_t$ 의 시계열 그림이다. 추정 방법에 있어 FARIMA 모형에서는 우선  $d$ 를 추정하고 난 후, ARMA모형을 이용하여  $p$ 와  $q$ 를 추정하였다. 그 결과 자료 1에서는 ARMA(1, 1)이 선택되었으며, 자료 2와

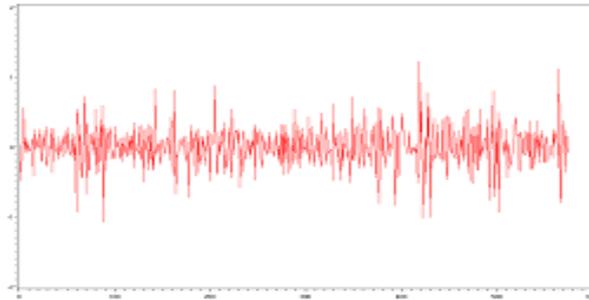


그림 3.2. 자료 1의 변수 변환 후 시계열 그림

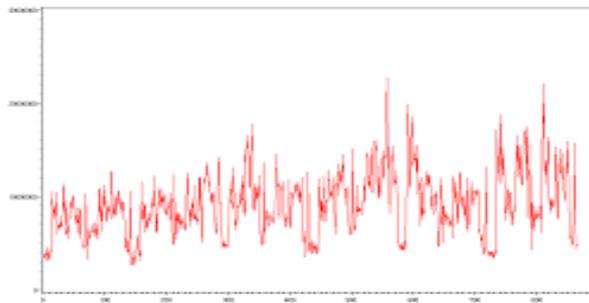


그림 3.3. 자료 2의 원자료 시계열 그림

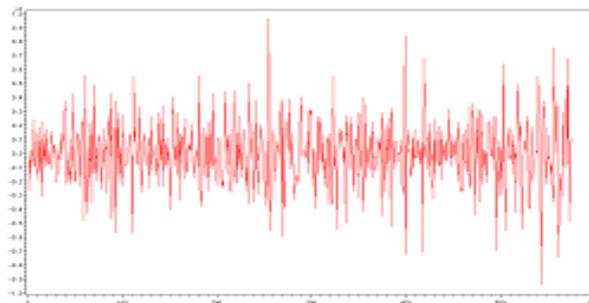


그림 3.4. 자료 2의 변수 변환 후 시계열 그림

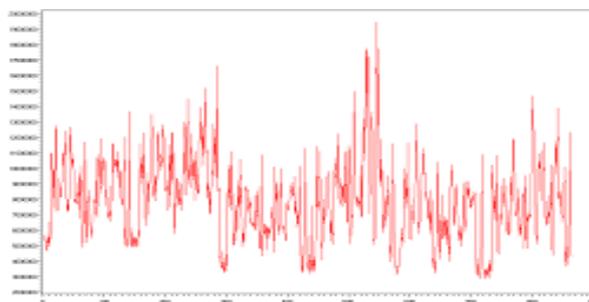


그림 3.5. 자료 3의 원자료 시계열 그림

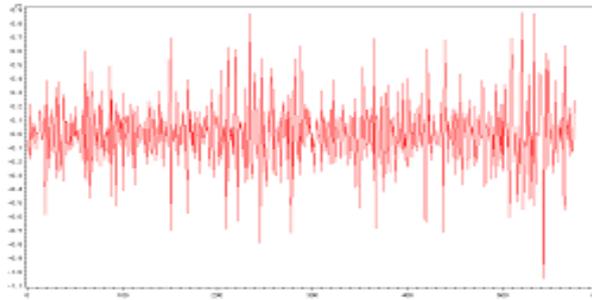


그림 3.6. 자료 3의 변수 변환 후 시계열 그림

표 3.1. 홀트-윈터스, FARIMA 모수 추정 결과

자료	홀트-윈터스		FARIMA					
	$\alpha$	$\beta$	$\phi_1$	$\phi_2$	$\phi_3$	$\phi_4$	$d$	$\theta_1$
자료 1	0.03	0.40	0.0344	-0.0405	-0.0215	-	0.4472	-
자료 2	0.03	0.23	0.3815	-	-	-	0.3188	0.0667
자료 3	0.02	0.28	0.5631	-0.1654	0.0982	-0.1668	0.4657	0.4670

표 3.2. AR-GARCH 모수 추정 결과

자료	AR-GARCH						
	$\phi_1$	$\phi_2$	$\phi_3$	$\phi_4$	$\alpha_0$	$\alpha_1$	$\beta_1$
자료 1	-0.5276	-0.1379	-0.1696	-	0.0151	0.1659	0.6056
자료 2	-0.5931	-0.0463	-0.0272	-0.0826	0.0214	0.2044	0.3210
자료 3	-0.4594	-0.0183	-0.1644	-	0.0189	0.3146	0.3046

표 3.3. FARIMA-GARCH 모수 추정 결과

자료	FARIMA-GARCH					
	$d$	$\phi_1$	$\phi_2$	$\alpha_0$	$\alpha_1$	$\beta_1$
자료 1	0.0344	-	-	0.0145	0.1374	0.6309
자료 2	0.3815	-0.2805	-	0.0353	0.2023	-
자료 3	0.5631	0.0266	0.1166	0.0189	0.3691	0.2684

표 3.4. 각 모형에 따른 RMSE

	홀트-윈터스	FARIMA	AR-GARCH	FARIMA-GARCH
자료 1	0.2977	0.2495	0.2567	0.2499
자료 2	0.2389	0.2096	0.2109	0.2098
자료 3	0.2547	0.2165	0.2194	0.2217

3에서는 ARMA(1,0)이 선택되었다. 홀트-윈터스(Holt-Winters) 방법에서는  $\alpha$ 와  $\beta$ 에 대해 0.01부터 0.99까지 0.01씩 증가시켜  $RMSE = (\sqrt{1/n \sum_{t=1}^n (X_t - \widehat{X}_t)^2})$ 가 가장 작은 값으로 추정하였으며, AR-GARCH에서는 비조건 최소 제곱추정량을 이용하였다. FARIMA-GARCH 모형에서는 FARIMA와 마찬가지로 우선  $d$ 를 추정한 후 ARIMA-GARCH 모형을 이용하여 모수를 추정하였다.

표 3.1에서 표 3.3은 FARIMA, 홀트-윈터스(Holt-Winters), AR-GARCH, FARIMA-GARCH 모형에 대한 모수 추정 결과이다. 각 추정치를 보면 모든 추정치 값이 각 모형의 제한 조건을 만족하고 있

음을 보여준다. 또한 FARIMA-GARCH 모형에서 자기회귀에 관련된 모수의 수가 FARIMA와 AR-GARCH 모형보다 줄어들고 있는 것으로 나타났다.

표 3.4는 각 모형의 추정 결과에 따른 RMSE(Root Mean Square Error)이다. 추정된 RMSE를 보면 FARIMA의 RMSE가 가장 작아 FARIMA 방법의 예측 오차가 가장 작게 나타나 트래픽의 특성을 설명하는데 있어 가장 적합한 것으로 나타났다. FARIMA-GARCH 모형에서는 자료 1과 자료 2에서는 RMSE의 값이 FARIMA와 큰 차이를 보이고 있지 않으나 자료 3에서는 차이를 보이고 있는 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 모형, 이분산성을 설명할 수 있는 AR-GARCH 모형, FARIMA와 GARCH 모형을 혼합한 FARIMA-GARCH 모형을 이용하여 실제 자료에 적합해 보았다. 그 결과 자료 1과 자료 2에서는 FARIMA 모형과 FARIMA-GARCH 모형의 적합성이 우수 하였으며, 자료 3에서는 FARIMA 모형이 우수하였다. 이러한 결과가 나타난 원인으로는 FARIMA 모형이 장기 기억 특성을 잘 설명했기 때문이고, 이분산성을 설명하는 AR-GARCH 모형보다 더 우수한 적합성을 보인 것으로 보아 트래픽은 이분산적인 특성보다 장기기억 특성이 주요한 특성이라고 사료된다.

#### 참고문헌

- 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.
- Liu, J., Shu, Y., Zhang, L. and Xue, F. (1999). Traffic modeling based on FARIMA models, *IEEE Canadian Conference 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.

# Performance Analysis of Internet Traffic Forecasting Model

S. Kim<sup>1</sup> · M. H. Ha<sup>2</sup> · J. Y. Jung<sup>3</sup>

<sup>1</sup>Department of Statistics, Chung-Ang University; <sup>2</sup>Department of Statistics, Chung-Ang University

<sup>3</sup>Department of Statistics, Chung-Ang University

(Received January 2011; accepted February 2011)

---

## Abstract

In this paper, we compare performance of three models. The Holt-Winters, FARIMA and ARGARCH models, are used in predicting internet traffic data for analysis of traffic characteristics. We first introduce the time series models and apply them to real traffic data to forecast. Finally, we examine which model is the most suitable for explaining the long memory, the characteristics of the traffic material, and compare the respective prediction performance of the models.

Keywords: Traffic, long memory, Holt-Winters, FARIMA, AR-GARCH.

---

---

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 Statistics, Chung-Ang University, Donggack-Gu, Seoul 156-756, Korea. E-mail: sahm@cau.ac.kr