

네이버 무선포털의 패킷량 분석에 관한 연구[†]

류귀열¹

¹서경대학교 컴퓨터학과

접수 2016년 4월 26일, 수정 2016년 5월 14일, 게재확정 2016년 5월 18일

요약

네이버 무선포털 패킷량을 분석하기 위해 2012년 9월 3일부터 2015년 10월 26일까지 조사하였으며, 한번 접속 시 6회 반복 측정하여 2,004개 자료를 수집하였다. 분석방법으로는 자기회귀오차모형을 사용하였으며, 종속변수는 패킷량이며 독립변수는 반복접속 횟수, 접속날짜, 접속시간, 접속요일, 접속월이다. 모형선택 기준은 AIC 기준과 R^2 기준으로 오차가 AR(36)을 따르는 모형이 선택되었다. 선택 모형으로부터 발견한 점들은 첫째로 날짜가 지남에 따라 평균 0.0752Kbyte 증가하고 있고, 둘째로 첫 번째 접속 시 다운로드되는 패킷량이 평균 156.965Kbyte로 재접속 시 다운로드되는 패킷량보다 평균 134.995Kbyte 많으며, 재접속 시 재사용률은 평균 82.76%라고 추정되었다. 셋째로, 시간대별 차이는 없었고, 넷째로 요일별 차이는 모두 유의하게 나타났다. 금요일이 가장 패킷량이 많았으며, 다음은 목요일이었으며, 수요일과 토요일은 비슷하였다. 다음으로 일요일이었으며 월요일이 가장 적었다. 다섯째로 월별 패턴에서는 5월과 8월이 각각 평균 13.98Kbyte, 12.48Kbyte적었으며 그 외 달은 유의한 차이를 보이지 않았다. R^2 에 의하면 우리의 모형은 실제 데이터 변동의 81.34%를 설명하고 있다. 연구의 한계는 패킷량에 영향을 많이 주는 데이터를 분석하지 못한 점이고 본 연구의 중요성으로 볼 때 다른 무선 포털을 분석 등 지속적인 연구가 요구된다.

주요용어: 터미변수, 무선포털, 자기회귀오차모형, 패킷량.

1. 서론

애플이 2007년 6월에 하나의 프로그램만 구동할 수 있는 아이폰을 출시하여 스마트폰의 시작을 알렸다. 2010년 6월에는 통신과 인터넷서비스 등 여러 가지 프로그램을 동시에 이용할 수 있는 멀티태스킹이 가능한 아이폰4를 출시하여 인터넷을 하면서 전화를 할 수 있게 됨으로써 본격적으로 스마트폰의 보급이 시작되었다. 이 후 신속한 개량을 실시하여 2011년 10월에 아이폰4s가 출시되었으며, 지속적인 개량을 통해 2015년 12월 현재 아이폰6s가 출시되고 있다. 우리나라에서는 KT가 2011년 11월11일부터 아이폰4s를 판매하기 시작하였다. 삼성전자는 2009년 10월에 윈도우모바일 운영체제를 탑재한 옴니아를 출시하였으나, 곧이어 2010년 6월에 안드로이드 운영체제를 탑재한 갤럭시를 출시하였다. 이 후 지속적인 개량을 통해 2015년 12월 현재 갤럭시6s까지 출시하였다. 아이폰과 안드로이드 폰의 경쟁으로 소프트웨어 안정성을 확보할 뿐 만 아니라 스마트폰 사용자들의 욕구를 만족시킴으로써 스마트폰은 폭발적으로 증가하였다. Ministry of Science, ICT and Future Planning (2015)에서 발표한 스마트폰 가입자 수와 이동전화 가입자 수가 Table 1.1에 나와 있다. Table 1.1에 보듯이 이동전화 가입자는 2011년 이후 서서히 증가하고 있으나 스마트폰 가입자는 2011년 출시 이후 급속도로 증가하여 2015년 말에는 2011에 비해 두배가 되었으며, 전체 가입자 중 74.09%가 되어 대중적인 단말기가 되었음을 알 수 있다.

[†] 본 연구는 2014학년도 서경대학교 교내연구비 지원에 의하여 이루어 졌음.

¹ (136-704) 서울시 성북구 정릉동 산16-1, 서경대학교 컴퓨터학과, 부교수. E-mail : gyryu@skuniv.ac.kr

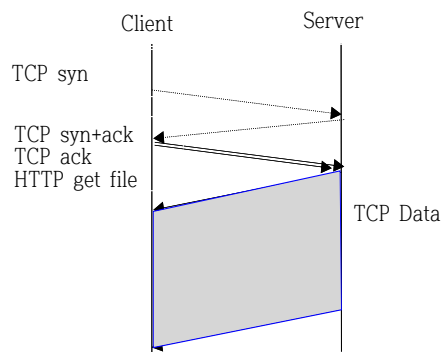
Table 1.1 Subscribers of smart phone and mobile services (unit: person)

Year	Smart phone	Mobile services	ratio
2011	22,578,408	52,506,793	43.00%
2012	32,041,079	53,624,427	59.75%
2013	37,516,572	54,680,840	68.61%
2014	40,698,151	57,290,356	71.04%
2015	43,667,799	58,935,081	74.09%

스마트 폰은 음성통화를 주로 하는 단말기가 아니라 인터넷서비스나 SNS 등 데이터 서비스를 동시에 많이 사용하는 단말기로서 스마트폰 가입자는 무선데이터를 많이 사용하게 된다. 무선데이터는 1기가바이트, 2기가바이트, 3기가바이트 등 이용 패킷량에 따라 요금을 책정하는 종량제 요금을 사용하기 때문에 사용패킷량에 민감하다. 그런데 무선데이터 이용자는 개별사이트의 사용량보다는 사용 총 패킷량만을 통신사로부터 알 수 있으므로, 개별 사이트의 패킷량은 알 수 없다. 무선인터넷 이용은 사용 편의성 등의 이유로 폭발적으로 증가하고 있다. 무선인터넷은 유선인터넷과 달리 무선 주파수를 사용하기 때문에 전송 매체의 한계로 인해 패킷량이 많으면 응답속도가 급속도로 느려진다. 따라서 무선포털 개별사이트의 패킷량에 관한 연구는 중요한 문제이다. 본 연구의 목적은 무선포털의 대표 사이트인 네이버 무선포털의 패킷량과 이들의 특징들을 분석하는 것이다. 주요 내용으로 인터넷 서비스는 다운로드 양을 줄이기 위해 첫 번째 접속할 때 보여주려는 콘텐츠를 모두 다운로드하고 재접속할 때는 다운로드된 데이터의 일부분을 재활용한다. 따라서 재접속 시 재활용되는 패킷량이 중요한데 이런 재활용 패킷량 비율을 추정할 것이다. 이와 더불어 접속일자, 접속시간, 접속요일, 접속월에 따르는 다운로드 되는 패킷량 변화를 추정할 것이다.

2. 응답패킷 측정과 연구모형

인터넷에서 데이터를 주고받는 방식은 HTTP (hypertext transfer protocol)로 FTP (file transfer protocol)에서의 파일전송의 비효율성을 줄이기 위해 제안된 것이다. HTTP에서 데이터를 주고받는 방식은 먼저 클라이언트와 서버의 채널을 열기 위해 RTT (round-trip time) 시간이 소요된다. 이는 클라이언트가 서버에 채널을 요청 (TCP syn)하는 데 소요되는 시간 0.5RTT와, 서버가 클라이언트에게 채널을 열어 (TCP syn+ack)주는데 소요되는 시간 0.5RTT이다. 채널이 열리면 클라이언트가 서버에 데이터를 요청하는데 소요되는 시간은 0.5RTT가 된다. 그 이후 요청한 데이터를 수신하게 된다. HTTP에서 데이터를 전송하는 프로토콜은 그림은 Figure 2.1에 나와 있다. 현재는 Oh 등 (2014)과 같이 인터넷 뿐 아니라 영상에서도 사용되고, Ryu 등 (2004)에서 보는 바와 같이 모바일에서도 표준으로 사용되고 있다.

**Figure 2.1** Protocol of data transmission in HTTP

무선포털은 유선인터넷 포털사이트를 무선인터넷 환경에 맞게 콘텐츠와 화면 등을 조정하여 구성한다. 사이트 주소는 www.naver.com에서 m.naver.com로 변경되며 스마트폰에서 네이버에 접속요청하면 자동으로 m.naver.com으로 접속한다. 따라서 스마트폰에서 naver포털에 접속하면 자동적으로 조정된 화면을 보게 된다. 네이버 무선포털의 응답패킷을 측정하기 위해 Ryu (2012)와 Ryu (2013a), Ryu (2013b)와 같이 네스케이프사에서 개발한 Firefox 브라우저를 사용하였으며, 응답패킷 측정은 Firebug에서 제공하는 네트워크 모니터링 도구를 사용하였다. Figure 2.2은 Firebug가 응답패킷을 측정하는 화면이다.

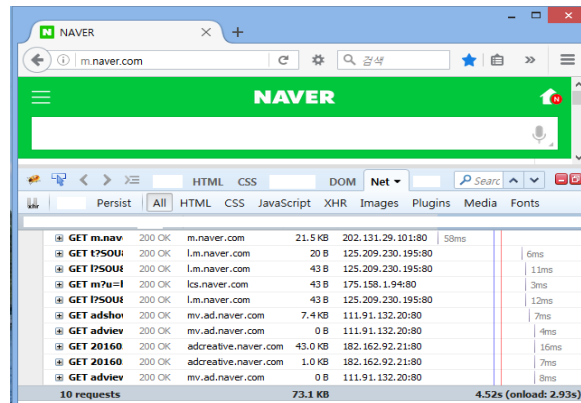


Figure 2.2 Response packets of naver's mobile portal using Firebug

분석하고자 하는 내용은 네이버 무선포털의 첫 번째 접속 시 다운로드되는 패킷량과 재접속 시 패킷량의 차이이다. 이는 두 번째 접속부터 이전에 다운로드받은 패킷을 재사용하는 량을 측정할 수 있으므로 매우 중요한 내용이다. 이와 더불어 접속일자, 접속시간, 접속요일, 접속월을 따르는 패킷량 변화를 추정할 것이다. 이를 위해 반복접속 횟수, 접속시간, 접속요일, 접속월을 독립변수로 하고 다운로드되는 패킷량을 종속변수로 선정하였다. 이들 변수는 시간에 영향을 받기 때문에 Ryu와 Kim (2013)와 Myoung 등 (2013), Lee (2012)가 설명한 바와 같이, 오차가 ARMA모형을 따르는 시계열회귀모형을 사용할 것이다. 시계열회귀모형은 종속변수가 설명변수에 영향을 받고 오차가 서로 상관관계에 있을 때 적합한 모형이다. 우리가 사용하는 시계열회귀모형은 식 (2.1)과 같다.

$$Y_t = \beta_0 + \beta_1 X_{1t} + \dots + \beta_k X_{kt} + a_t$$

$$a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_p a_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \quad (2.1)$$

여기서 $\epsilon_t \sim iid N(0, \sigma^2)$ 이다.

식 (2.1)에서 오차 a_t 는 ARMA(p, q)를 따른다. 자기회귀오차모형 선택은 두 단계로 수행된다. 먼저 회귀모형을 결정하기 위해 수정 결정계수 (Adjusted R^2) 기준과 AIC (Akaike Information Criteria) 기준으로 할 것이며, 두 번째로 잔차의 ACF와 PACF를 기준으로 ARMA(p, q)의 차수를 결정할 것이다. 모형진단은 Jo와 Son (2011)과 Jung (2009) 등이 설명한 바와 같이 모형선택 후 잔차의 자기상관함수 (autocorrelation; ACF)와 편자기상관함수 (partial autocorrelation; PACF)가 백색잡음 (white noise)과 같은 패턴인가로 진단할 것이다.

3. 연구 결과

네이버 무선포털 패킷량을 분석하기 위해 2012년 9월 3일부터 2015년 10월 26일까지 패킷량을 측정하였다. 측정 방법은 컴퓨터를 부팅 후 하루에 6회 반복하고 접속할 때 마다 브라우저를 다시 열어 측정하는 방식으로 총 2,004회 수신 패킷량을 측정하였다. 종속변수는 네이버 무선포털의 패킷량이며 독립변수는 반복접속 횟수, 접속날짜, 접속시간, 접속요일, 접속월이다. 본 논문의 연구목적에 의해 반복접속 횟수, 접속요일, 접속월의 개별속성 값의 효과분석이 필요하다. 이를 위해 이런 변수들을 더미변수로 분해하였다. 분해한 더미변수는 다음과 같다.

$$\begin{aligned}
 x_{11t} &= \begin{cases} 1 & \text{첫번째 접속} \\ 0 & \text{그외} \end{cases} & x_{21t} &= \begin{cases} 1 & \text{월요일} \\ 0 & \text{그외} \end{cases} & x_{31t} &= \begin{cases} 1 & \text{1월} \\ 0 & \text{그외} \end{cases} \\
 x_{12t} &= \begin{cases} 1 & \text{두번째 접속} \\ 0 & \text{그외} \end{cases} & x_{22t} &= \begin{cases} 1 & \text{화요일} \\ 0 & \text{그외} \end{cases} & x_{32t} &= \begin{cases} 1 & \text{2월} \\ 0 & \text{그외} \end{cases} \\
 x_{13t} &= \begin{cases} 1 & \text{세번째 접속} \\ 0 & \text{그외} \end{cases} & x_{23t} &= \begin{cases} 1 & \text{수요일} \\ 0 & \text{그외} \end{cases} & x_{33t} &= \begin{cases} 1 & \text{3월} \\ 0 & \text{그외} \end{cases} \\
 x_{14t} &= \begin{cases} 1 & \text{네번째 접속} \\ 0 & \text{그외} \end{cases} & x_{24t} &= \begin{cases} 1 & \text{목요일} \\ 0 & \text{그외} \end{cases} & x_{34t} &= \begin{cases} 1 & \text{4월} \\ 0 & \text{그외} \end{cases} \\
 x_{15t} &= \begin{cases} 1 & \text{다섯번째 접속} \\ 0 & \text{그외} \end{cases} & x_{25t} &= \begin{cases} 1 & \text{금요일} \\ 0 & \text{그외} \end{cases} & x_{35t} &= \begin{cases} 1 & \text{5월} \\ 0 & \text{그외} \end{cases} \\
 & & x_{26t} &= \begin{cases} 1 & \text{토요일} \\ 0 & \text{그외} \end{cases} & x_{36t} &= \begin{cases} 1 & \text{6월} \\ 0 & \text{그외} \end{cases} \\
 & & & & x_{37t} &= \begin{cases} 1 & \text{7월} \\ 0 & \text{그외} \end{cases} \\
 & & & & x_{38t} &= \begin{cases} 1 & \text{8월} \\ 0 & \text{그외} \end{cases} \\
 & & & & x_{39t} &= \begin{cases} 1 & \text{9월} \\ 0 & \text{그외} \end{cases} \\
 & & & & x_{310t} &= \begin{cases} 1 & \text{10월} \\ 0 & \text{그외} \end{cases} \\
 & & & & x_{311t} &= \begin{cases} 1 & \text{11월} \\ 0 & \text{그외} \end{cases}
 \end{aligned}$$

접속날짜는 엑셀의 날짜변환을 통해 숫자로 변환하였으며 2012년 9월 3일을 0으로 설정하였다. 네이버의 무선포털의 패킷량 그래프는 Figure 3.1에 나와 있다. 그래프는 Kbyte로 측정된 일별 평균 패킷이며, 시간이 지남에 따라 패킷량이 증가하는 추세를 보이고 있음을 알 수 있다.

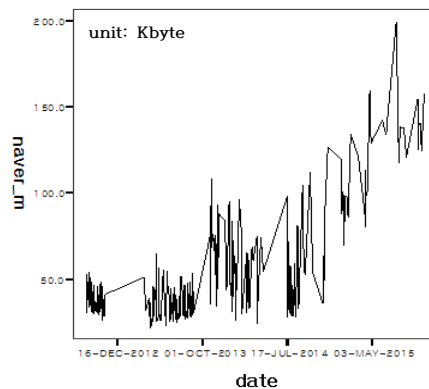


Figure 3.1 Packet data from Naver's mobile portal

독립변수는 반복접속 횟수 5개, 접속날짜, 접속시간, 접속요일 6개, 접속월 11개 총 24개이며, 오차가 ARMA를 따르므로 변수선택은 Park (2007)이 설명하고 있는 후방제거법 (Backward elimination)을 적용하였다. 그리고 오차가 AR(18), AR(24), AR(30), AR(36) 모형을 따를 때 잔차의 ACF와 PACF가 백색잡음의 형태를 보였다. 우리가 적용하는 모형은 식 (3.1)과 같다.

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_k X_{kt} + a_t$$

$$a_t + \phi_1 a_{t-1} + \phi_2 a_{t-2} + \cdots + \phi_p a_{t-p} = \epsilon_t \quad (3.1)$$

여기서 $\epsilon_t \sim iid N(0, \sigma^2)$ 이다.

오차가 AR(18)과 AR(24), AR(30), AR(36) 모형을 따를 때, 독립변수는 후방제거법으로 선택한 모형의 AIC와 R^2 가 Table 3.1에 나와 있다. AR의 차수가 높아짐에 따라 AIC가 낮아지고 R^2 은 높아지는 패턴을 보이고 있다. MA의 계수가 유의하지 않으므로 우리의 모형은 자기회귀오차 모형으로 적합되었다.

Table 3.1 AIC and R^2 of AR(18) and AR(24), AR(30), AR(36) model

Order	AIC	R^2
AR(18)	20431	0.7978
AR(24)	20378	0.8033
AR(30)	20301	0.8108
AR(36)	20268	0.8134

Box-Ljung 통계량으로 잔차가 백색잡음과 같은 패턴을 보이고 있는 것을 확인할 수 있는데, AR(18), AR(24), AR(30), AR(36)일 때 백색잡음의 형태를 보였으며, 잔차의 차수가 높아짐에 따라 백색잡음과 같은 형태가 더 뚜렷해졌다. 예시로 AR(18)과 AR(36) 모형의 Box-Ljung 통계량과 AR의 계수가 Table 3.2에 나와 있다. 결과를 보면 Box-Ljung 통계량은 백색잡음과 같은 패턴을 보이고 있음을 알 수 있다.

Table 3.2 Box-Ljung Statistic for residuals

Lag	AR(18)		AR(36)	
	Autocorrelation	Box-Ljung Statistic	Autocorrelation	Box-Ljung Statistic
1	0.025	0.270	0.035	0.116
2	0.008	0.514	0.018	0.212
3	0.000	0.721	0.012	0.333
4	-0.005	0.847	0.011	0.458
5	-0.017	0.853	-0.003	0.601
6	-0.034	0.629	-0.016	0.650
7	-0.024	0.597	-0.019	0.674
8	0.006	0.694	0.008	0.757
9	-0.007	0.771	-0.006	0.828
10	-0.006	0.835	-0.001	0.886
11	0.005	0.886	0.009	0.919
12	-0.079	0.104	-0.022	0.903
13	-0.012	0.133	-0.007	0.933
14	0.000	0.178	-0.003	0.957
15	0.011	0.218	0.012	0.967
16	0.007	0.268	0.010	0.977

그리고 AR(18) 모형의 잔차 ACF와 PACF는 Figure 3.2에 있다. Figure 3.2에서는 보는 것과 같이 시차가 6 또는 12에서의 피크가 한계선을 넘고 있는 것처럼 보이지만 Box-Ljung 통계량으로는 백색잡음의 형태를 보인다. AR(36) 모형의 잔차 ACF와 PACF는 Figure 3.3에 있다. Figure 3.3에서 보는 것

과 같이 AR(36) 모형에서의 잔차 ACF와 PACF 모두 한계선 안에 있음을 알 수 있다. 따라서 AR(36) 모형은 Table 3.1에서 보듯이 AIC 기준으로 볼 때 가장 낮고 R^2 기준으로 볼 때 가장 높아서 가장 적합한 모형으로 선택되었다. AR(36) 모형에서 독립변수를 후방제거법으로 선택된 모형의 결과가 Table 3.3에 나와 있다. 유의한 변수는 접속일, 첫번째 접속, 월요일, 화요일, 수요일, 목요일, 금요일, 토요일, 5월, 8월이며 오차에서 AR(6), AR(12), AR(18), AR(24), AR(30), AR(36)이었다.

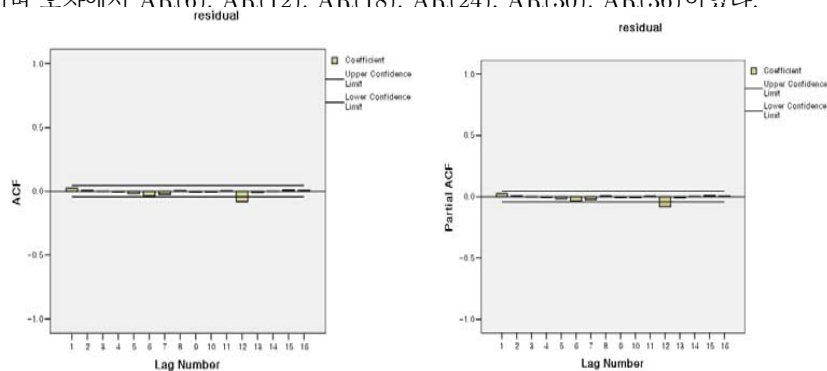


Figure 3.2 ACF and PACF of AR(18)

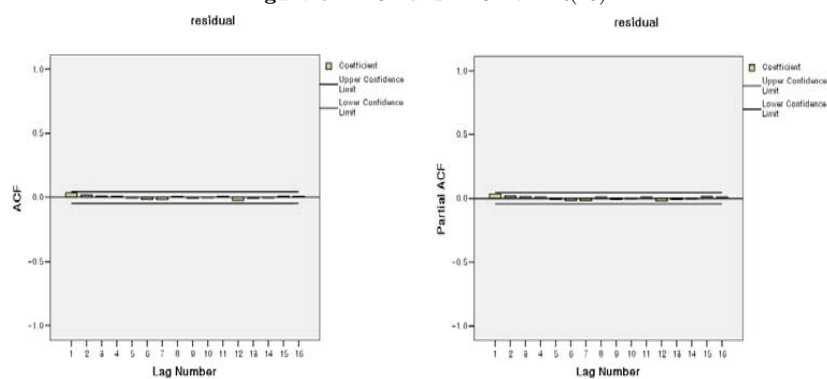


Figure 3.3 ACF and PACF of AR(36)

Table 3.3 Results of AR(36) model

Variable	df	Estimate (unit:Kbyte)	Standard Error	t -value	Pr > $ t $
date	1	0.0752	0.0098	7.58	<0.0001
rep1	1	134.995	11.8902	11.35	<0.0001
week1	1	-28.4133	9.0481	-3.14	0.0017
week2	1	5.5358	3.1772	1.74	0.0816
week3	1	6.1059	3.1544	1.94	0.0530
week4	1	7.5536	3.2306	2.34	0.0195
week5	1	12.6104	3.2343	3.90	<0.0001
week6	1	6.6421	3.9819	1.67	0.0955
month5	1	-13.9754	7.1351	-1.96	0.0503
month8	1	-12.4849	4.3524	-2.87	0.0042
AR(6)	1	-0.1367	0.022259	-6.14	<0.0001
AR(12)	1	-0.1559	0.022177	-7.03	<0.0001
AR(18)	1	-0.1514	0.022344	-6.77	<0.0001
AR(24)	1	-0.0979	0.022344	-4.39	<0.0001
AR(30)	1	-0.1608	0.022177	-7.25	<0.0001
AR(36)	1	-0.1227	0.022259	-5.51	<0.0001

모형의 특징을 보면 첫째로 날짜가 지남에 따라 평균 0.0752Kbyte 증가하고 있음을 알 수 있고, 이는 네이버 모바일 포털의 패킷량이 날짜가 지남에 따라 증가하는 추세를 반영하고 있다. 둘째로 첫 번째 접속 시 다운로드되는 패킷량이 재접속 시 다운로드되는 패킷량보다 평균 134.995Kbyte 많으며, 이는 재접속 시 평균적으로 134.995Kbyte 정도를 재사용하고 있다는 것을 의미한다. 반복접속 횟수별 평균 패킷량과 표준편차가 Table 3.4에 있다. 첫 번째 접속시 평균 156.965Kbyte 다운로드되며 그 이후 접속 시 다운로드 되는 패킷량은 26~30Kbyte 정도 수준이며 평균 28.129Kbyte로 나타났다. 두 번째 이후 접속 시 다운로드되는 패킷량은 평균 134.995Kbyte가 재사용되어 28.129Kbyte를 다운로드받고 있다고 말할 수 있다. 따라서 재접속 시 재사용률은 평균 82.76%라고 추정할 수 있다.

Table 3.4 Mean and standard deviation per repetition (unit: Kbyte)

Repetition	Mean	Standard deviation
1	156.965	123.195
2	30.173	19.498
3	27.340	17.497
4	27.916	19.062
5	26.915	19.517
6	28.302	21.133

셋째로 접속시간대별로 유의한 차이는 없었다. 넷째로 요일별 차이는 모두 유의하게 나타났다. 금요일이 가장 패킷량이 많았으며, 다음은 목요일이었으며, 그 다음은 수요일과 토요일로 비슷하였다. 마지막으로 월요일이 가장 적었으며 기준이 되는 일요일보다 평균 28.4133Kbyte만큼 적었다. 패킷량은 정보량과 밀접한 관계가 있기 때문에 요일별 제공하는 정보량도 다르다는 의미로도 해석될 수 있다. 다섯째로 월별 패턴에서는 5월과 8월이 각각 평균 13.98Kbyte, 12.48Kbyte 적었으며 그 외 달은 유의한 차이를 보이지 않았다. 오차에 대해서는 시간차가 6, 12, 18, 24, 30, 36일 때만 유의하고 그 외 시간차에서는 유의하지 않았다. 이는 사이트 접속 시 6회 반복 실험한 결과로 나타난 것으로 추정된다. 우리가 구한 모형은 식 (3.2)와 같다.

$$\begin{aligned}
 Y_t = & 0.0752 * Date_t + 134.995 * x_{11t} - 28.4133 * x_{21t} + 5.5358 * x_{22t} + 6.1059 * x_{23t} \\
 & + 7.5536 * x_{24t} + 12.6104 * x_{25t} + 6.6421 * x_{26t} - 13.9754 * x_{35t} - 12.4849 * x_{38t} + a_t \\
 a_t - & 0.1367a_{t-6} - 0.1559a_{t-12} - 0.1514a_{t-18} - 0.0979a_{t-24} - 0.1608a_{t-30} - 0.1227a_{t-36} = \epsilon_t
 \end{aligned} \quad (3.2)$$

여기서 $\epsilon_t \sim iid N(0, \sigma^2)$ 이다.

실제 데이터와 예측치의 정확성을 직관적으로도 검토해 보기 위해 Figure 3.4과 같이 그래프를 이용하였다. 2,004개의 모든 데이터에 대해 그래프를 한번에 보이기 어려워 그 중 일부분인 최근의 자료에 대한 그래프만 보였다. Figure 3.3을 보면 실제 데이터에서 첫 번째 접속 시 많은 양의 패킷을 다운받고 두 번째 접속부터는 적은 양의 패킷을 다운받고 있는 패턴을 확인할 수 있으며, AR(36) 모형도 첫 번째 접속 시 예측값과 이후 접속 시 예측값이 확연히 차이를 보이며 실제 데이터 패턴을 따라가고 있는 것을 알 수 있으며, 우리의 모형이 잘 작동하고 있음을 알 수 있다. 우리의 모형의 예측치들은 실제 데이터 변동의 81.34%를 설명하고 있다. 그런데 Figure 3.4에서 가장 큰 차이를 보인 데이터는 2015년 7월 22일 811.11Kbyte이었으며 예측값은 425.78Kbyte으로 차이가 컸었는데 이는 수신 데이터가 이례적으로 많아서 발생한 문제이다. 비교 대상인 AR(18) 모형의 R^2 는 79.78%이었다.

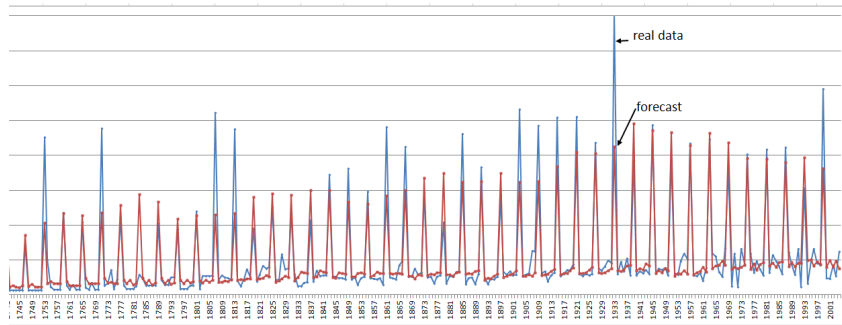


Figure 3.4 Real data and forecasts of AR(36) model

4. 결론

네이버 무선포털 패킷량을 분석하기 위해 2012년 9월 3일부터 2015년 10월 26일까지 접속 시 6회 반복 측정하여 총 2,004회 조사하였다. 종속변수는 네이버 무선포털의 다운로드되는 패킷량이며 독립변수는 반복접속 횟수, 접속날짜, 접속시간, 접속요일, 접속월이다. 본 논문의 연구목적은 달성하기 위해 반복접속 횟수, 접속요일, 접속월의 개별속성 값의 효과분석이 필요하여 이 변수들을 더미변수로 분해하였다. 분석방법은 자기회귀오차분석 모형을 사용하였으며 독립변수 선택은 후방제거법을 사용하였다. 모형선택 기준은 AIC와 R^2 을 사용하였으며, 모형의 적합성은 잔차 ACF와 PACF로 판단하였다.

오차가 AR(36)을 따르는 모형이 AIC 기준으로 볼 때 가장 낮고 R^2 기준으로 볼 때 가장 높아서 가장 적합한 모형으로 선택되었다. AR(36) 모형에서 유의한 독립변수는 접속일, 첫번째 접속, 월요일, 화요일, 수요일, 목요일, 금요일, 토요일, 5월, 8월이며 오차에서는 AR(6), AR(12), AR(18), AR(24), AR(30), AR(36)이었다. 모형 적합성에서는 잔차 ACF와 PACF도 백색잡음과 같은 패턴을 보여 적합한 모형으로 판단되었다.

네이버 모바일 포털 패킷량의 특징으로 첫째로 날짜가 지남에 따라 평균 0.0752Kbyte 증가하고 있고, 둘째로 첫 번째 접속 시 다운로드되는 패킷량이 재접속 시 다운로드되는 패킷량보다 평균 134.995Kbyte 많으며, 이는 두 번째 이후 접속 시 다운로드되는 패킷량은 평균 134.995Kbyte가 재사용되어 28.129Kbyte를 다운로드받고 있다고 말할 수 있다. 따라서 재접속 시 재사용률은 평균 82.76%라고 추정되었다. 세 번째로, 접속시간대별 유의한 차이는 없었고, 네 번째로 요일별 차이는 모두 유의하게 나타났다. 금요일이 가장 패킷량이 많았으며, 다음은 목요일이었으며, 수요일과 토요일은 비슷하였다. 마지막으로 월요일이 가장 적었으며 기준이 되는 일요일보다 평균 28.4133Kbyte만큼 적었다. 다섯 번째로 월별 패턴에서는 5월과 8월이 각각 평균 13.98Kbyte, 12.48Kbyte 적었으며 그 외 달은 유의한 차이를 보이지 않았다. 오차에 대해서는 시간차가 6, 12, 18, 24, 30, 36일 때만 유의하고 그 외 시간차에서는 유의하지 않았다. 실제 데이터와 예측치의 정확성을 검토해 보기 위해 그래프를 이용하였다. 그래프에 의하면 예측값이 실제 데이터 패턴을 따라가고 있으며, R^2 에 의하면 예측치들은 실제 데이터 변동의 81.34%를 설명하고 있다는 것을 알 수 있다.

무선인터넷은 주로 종량제 서비스이므로 이용자들의 만족을 높이려면 패킷량을 줄이려는 노력이 동반되어야 한다. 그러므로 본 연구의 주제인 네이버 무선 포털의 패킷량에 관한 연구는 패킷량의 크기와 특징을 파악하게 함으로써 패킷량을 줄일 수 있는 동기와 방법을 제공하기 때문에 패킷을 줄이는 기반연구라고 할 수 있다. 본 연구를 기초로 효율적으로 포털을 구축하기 위한 전략을 수립할 수 있고 이를 실현한다면 비용을 줄일 뿐만 아니라 무선통신의 한계를 극복하며 빠른 서비스도 가능하여 이용자들의 만족도를 높일 수 있다. 본 연구의 중요성으로 볼 때, 다른 무선포털 분석 등 지속적인 연구가 요구된다.

References

- Jo, S. and Son, Y. (2011). *Time series analysis using SAS / ETS*, Yulgok press, Seoul.
- Jung, D. (2009). *Demand forecasting of time series I*, Hannarae Academy, Seoul.
- Lee, H. (2012). Analysis of time series models for consumer price index. *Journal of the Korean Data & Information Science Society*, **23**, 535-542.
- Ministry of Science, ICT and Future Planning(2015). Statistical data, Retrieved from <http://www.msip.go.kr>.
- Myoung, S., Kim, D., Lee, D., Kim, H. and Jo, J. (2013). An analysis of time series models for toilet and laundry water-uses. *Journal of the Korean Data & Information Science Society*, **24**, 1141-1148.
- Oh, J., Yun, D. and Jung, J. (2014). A Router buffer-based congestion control scheme for improving QoS of UHD streaming services. *Journal of KIISE*, **41**, 974-981.
- Park, S. (2007). *Regression analysis*, Minyoungsa, Seoul.
- Ryu, G. (2012). A study on response time of WiBro depending on signal intensity. *Journal of the Korean Data & Information Science Society*, **23**, 1119-1128.
- Ryu, G. (2013a). A study on comparing response times between Wibro and wired internet using portals. *Journal of the Korean Data & Information Science Society*, **24**, 23-32.
- Ryu, G. (2013b). A study on distribution comparison of response packets for major portal sites. *Journal of the Korean Data & Information Science Society*, **24**, 473-444.
- Ryu, S. and Kim, J. (2013). Time series regression model for forecasting the number of elementary school teachers. *Journal of the Korean Data & Information Science Society*, **24**, 321-332.
- Ryu, S., Park, S., Lee, I., Jang, H. and Baek, D. (2004). SNMP information based hierarchical routing mechanism for fast handoff in mobile IP. *Journal of KIISE*, **31**, 131-145.

A study on analysis of packet amount of Naver's mobile portal[†]

Gui-Yeol Ryu¹

¹Department of Computer Science, SeoKyeong University

Received 26 April 2016, revised 14 May 2016, accepted 18 May 2016

Abstract

The purpose of this paper is to build a model of packet amount of Naver mobile portal. We collected 2004 cases by measuring the sixth per access from September, 2012 to October, 2015. We use regression model with autoregressive errors, in which predictors incorporated into the model were replication, date, time, week, month. It has been found the model which errors follow AR(36), based on AIC and adjusted R^2 . We found some characteristics from our model as follows. In addition to model building, we also have discussed some meaningful features yielded from the selected model in this paper. Considering the importance of this topic, continuous researches are needed.

Keywords: Autoregressive error model, dummy variable, mobile portal, packet amount.

[†] This Research was supported by Seokyeong University in 2014.

¹ Associate professor, Department of Computer Science, SeoKyeong University, Seoul 136-704, Korea.
E-mail: gyryu@skuniv.ac.kr