

음성 트래픽과 인터넷 트래픽 추정에 관한 연구

황정연 · 강병용 · 전경표

한국전자통신연구원 교환 · 전송기술연구소 통신시스템 연구부 시스템 분석팀

A Study on the Voice Traffic and Internet Traffic Estimation

Jung-Yeon Hwang · Byung-Ryong Kang · Kyung-Pyo Jun

On this study we selected some variable which affect on the estimated of the voice traffic, and estimated daily average traffic by years according to the variables. We applied nonlinear growth curve model to future traffic forecast with estimated historical traffic data. As a result of the forecasting, this study investigates the year in which the internet traffic goes far than the voice traffic.

1. 서론

정보통신서비스가 음성서비스 위주에서 인터넷 및 멀티미디어 정보통신서비스로 급속히 이행됨에 따라 기존의 음성 서비스의 정보통신 네트워크는 인터넷 서비스를 효율적으로 수용하여 수익성을 개선할 수 있도록 망구조 및 기능의 고도화를 요구받고 있다. 그리하여 각 나라별 통신사업자는 정보통신 네트워크의 고도화 및 관련 기술의 개발을 위한 연구개발 투자에 국가적인 역량을 집중하고 있다. 특히 21세기에 가장 촉망받는 산업분야로서 정보통신 분야가 갖는 중요성은 이 분야에 대한 관심을 증대시키는 기폭제 역할을 하고 있다. 현재 우리 사회는 산업화 사회를 거쳐 정보화 사회로 진입하고 있다. 근래에 들어 기업의 사무실, 학교, 기업 등에서 업무용 전산기기가 급속하게 보급되면서 이들 컴퓨터 상호간 또는 중앙의 호스트와 개인용 컴퓨터간에 정보와 자원을 공유할 필요성이 증대되어 빌딩, 학교, 기업 등과 같은 한정된 지역 내에 독자적인 LAN을 구축하여 사용해 왔다. 그래서 상호 네트워크를 연결하는 인터넷에 대한 수요가 급격히 늘어 인터넷 데이터 트래픽이 증가하고 있는 추세이다. 인터넷 서비스 등 데이터 통신 위주의 통신망을 구축하기 위해서는 네트워크 자체에 대한 투자의 증대뿐만 아니라, 관련기술의 효율적인 연구개발 역시 중요하다. 이때 통신사업자에 있어서 망에 대한 투자의 적절한 시기를 판단하기 위한 미래의 트래픽 예견은 올바른 의사결정을 내리는데 있어서 매우 중요하다.

본 논문은 기존 음성 트래픽과 인터넷 트래픽에 대한 추정을

하기 위하여 영향을 미치는 변수를 조사하여 연도별 하루 평균 트래픽을 추정한다. 이 추정된 연도별 과거 트래픽 자료를 가지고서 미래를 예측하기 위하여 비선형 성장곡선 모형중 최적모형을 선택하여 적용한다. 그리고 예측된 결과를 가지고서 비교하여 향후 몇 년 후에 인터넷 트래픽이 음성 트래픽을 앞서는지에 대한 추세를 예측하여 앞으로 광중망 설계에 있어서 참고자료가 될 수 있도록 한다.

2. 예측모형

예측은 미래의 알려지지 않은 상황을 사전에 예견하는 것으로 이에 대한 분석방법은 주어진 자료의 특성 및 성격에 따라 시계열분석(Time series Analysis), 회귀분석(Regression Analysis), 로짓모형(로지스틱 회귀모형), 비선형 성장곡선 모형 등 주어진 자료의 형태에 따라 여러 가지 방법이 있다. 자료의 형태가 시간별 PDF(Probability Density Function) 형태인 경우에는 흔히 잘 알려진 시계열 분석으로 Box & Jenkins의 ARIMA 모형을 주로 이용한다. 그리고 회귀분석은 종속변수(반응변수) Y와 독립변수(설명변수) Xs 들간의 선형결합 형태의 모형으로 Y에 대한 Xs의 설명력이 높은 변수들을 선택하여 최소자승법(OLS: ordinary least squares method)에 의하여 추정하는 것으로 주로 단기 예측에 적용하며, 자료의 형태는 연속형 자료이다. 로짓모형은 주로 설문지 형태의 범주형 자료에 대하여 분석하는 모형으로 서비스 이용자 특성에 따른 서비스 이용자 성향 분석을 할 때 적용한다. 마지막으로 성장곡선모형은 연속적인 시

간(t)의 흐름에 따라 얻어지는 누적 수요자료를 예측하기 위하여 적용하는 모형으로 크게 선형모형과 비선형모형으로 분류된다. 이때 성장곡선 모형의 선형모형은 맨필드-블랙맨 모형(Manfiled-Blackman model), 선형 고펜르츠 모형(Linear Gompertz Model), 와이블 모형(Weibull model), 베스 모형(Bass model), NSRL 모형(Nonsymmetric Responding Logistic model), 하베이 모형(Harvey model) 등으로 Young (1993)의 논문에서 잘 나타나 있다. 이 선형모형은 대부분 상한값 K를 상정하고서 분석하는데, 상한값을 미리 상정하고 행하는 분석의 경우, 장기 예측시점에서는 잔차가 크게 발생하게 된다. 다음 비선형 성장곡선 모형은 상한값 K를 모수로 추정하여 구할 수 있는 장점이 있으며, 모형으로는 로지스틱 모형(Logistic model), 고펜르츠 모형(Gompertz model) 등이 널리 사용되며, 그 외 최근에 인터넷 자료를 통하여 분석하면서 발견된 모형으로 음의지수 모형(Negative Exponential model)과 지수 모형(Exponential model)이 있다. 인터넷 자료의 특성은 급격히 상승하는 형태의 자료로서 일반적인 형태의 완만한 증가를 나타내지 않은 점을 특성으로 들 수 있다.

대표적인 비선형 성장곡선 모형의 함수 식을 살펴보면 아래와 같다.

(1) 로지스틱 모형(Logistic Model)

$$Y_t = \frac{K}{1 + \exp(\alpha - \beta t)} + \epsilon_t \quad (1)$$

(1)의 모형은 펄 성장곡선 모형(Pearl Growth Curve Model)으로 불리기도 한다. 그리고 주어진 자료가 완만한 형태의 곡선으로 이루어진 경우에 가장 잘 적합한 모형으로 알려져 있다 (Young, 1993).

(2) 프로빗 모형(Probit Model)

$$Y_t = K + \Phi(\alpha + \beta t) + \epsilon_t \quad (2)$$

(3) 고펜르츠 모형(Gompertz Model)

$$Y_t = K \exp(-\alpha)^{\exp(-\beta t)} + \epsilon_t \quad (3)$$

(4) 음의지수 성장곡선 모형(Negative Exponential Growth Curve Model)

$$Y_t = \alpha(1 - e^{-\beta t}) + \epsilon_t \quad (4)$$

(5) 지수 성장곡선 모형(Exponential Growth Curve Model)

$$Y_t = \alpha(e^{-\beta t}) + \epsilon_t \quad (5)$$

(4), (5)의 모형은 자료의 형태가 급격히 폭발적으로 증가하는 인터넷 자료의 형태에 대해서 가장 적합한 모형이다.

위 비선형 성장곡선 모형에서 Y_t 는 시간에 대한 누적 자료이고, K 는 최대 상한값(upper limit), α, β 는 추정하여야 할 모수(parameters)를 나타내며, $\epsilon_t \sim i.i.d. N(0, \sigma^2)$ 을 가정한다.

위 모형을 이용한 자료분석은 SAS/STAT NLIN 명령어를 이용하였다. 이때, NLIN 명령어를 이용하면 모형에서 얻어진 예측값과 신뢰대를 OUTPUT에 저장하여 바로 SAS/GRAPH를 이용한 그래프 표현이 가능하다.

일반적인 비선형 모형을 위한 방정식을 나타내면 다음과 같다.

$$Y = F(\beta_0^*, \beta_1^*, \dots, \beta_n^*, Z_1, Z_2, \dots, Z_n) + \epsilon = F(\beta^*) + \epsilon \quad (6)$$

이때 Z 는 독립변수에 의한 행렬, β^* 는 모수에 관한 벡터, ϵ 은 오차항의 벡터이다. 그리고 F 는 독립변수와 모수에 관한 함수이다.

비선형 정규 방정식의 해는 아래와 같다.

$$X'F(\beta) = X'e \quad (7)$$

이때 X 는 아래와 같다.

$$X = \partial F / \partial \beta \quad (8)$$

비선형의 경우 X 와 $F(\beta)$ 는 β 에 대한 함수이며, 일반적으로 해는 존재하지 않는다. 따라서 NLIN 명령어에 의한 해는 수치 해석을 이용한다. 이 경우, β 에 대한 초기값(starting value)은 오차의 제곱합 $e'e (L(\beta))$ 가 최소값을 갖도록 연속적으로 개선되어 구해진다. 그리고 비선형 방정식의 수치해석 계산은 β 와 $e = Y - F(\beta)$ 에 의한 최근값(current value)으로 계산된다. 이때 오차는 β 에 의해서 얻어지며, 행렬 X 는 선형회귀에 의한 유사한 계열을 적용한다. 수치해석에 의한 반복 계산은 어느 한 점 β 에서 시작한다. 이때 X 와 Y 의 계산은 아래의 Δ 에 의해서 반복 계산되어진다.

$$L(\beta_0 + \alpha \Delta) < L(\beta_0) \quad (9)$$

위 식에서 어떤 방법의 Δ 를 적용하느냐에 따라 모수에 대한 벡터가 변화되어 계산된다. 이때 Δ 를 적용하는 수치해석 계산 방법은 아래의 방법에 따라 다르지만 같은 해를 얻게 된다.

(1) Steepest descent 방법

$$\Delta = -X'e \quad (10)$$

(2) Gauss-Newton 방법

$$\Delta = (X'X)^{-1}X'e \quad (11)$$

(3) Newton 방법

$$\Delta = G^{-1}X'e \quad (12)$$

(4) Marquardt 방법

$$\Delta = (X'X + \lambda \text{diag}(X'X))^{-1}X'e \quad (13)$$

위 관련 각 방법의 알고리즘에 대한 자세한 내용은 수치해석책이나, SAS/STAT NLIN 명령어를 참고한다.

3. 음성트래픽 추정 및 예측

음성 트래픽은 실제 트래픽 자료를 구하기가 쉽지 않아서 음성 트래픽에 영향을 미치는 변수를 조사하여 고려하였다. 일반전화 가입자수 및 이동전화 가입자수에 대한 연도별 자료 [출처 : 한국정보통신진흥협회(연도별 일반전화가입자수, 이동전화가입자수, <http://203.254.87/DB/>)]를 가지고서 음성 이용 평균 대역폭, 평균 점유시간, 평균 통화수 및 음성 가입자당 트래픽 볼륨을 상정하여 추정하였다. 그리고 이동전화가입자의 경우는 월 평균 사용시간을 적용하여 이동 가입자당 평균 트래픽 볼륨에 의하여 분석하였다.

- 통화당 음성 평균대역폭 : 64Kbps=8Kbyte

표 1. 음성전화가입자수와 이동전화가입자 수에 의한 음성 트래픽 추정

연도	음성전화 가입자수	이동전화 가입자수	일반전화가입자 트래픽(Tera/Day)	이동가입자 트래픽(Tera/Day)	음성 전체 트래픽 (Terabyte/Day)
1990	13,276,449	80,005	113.91	0.04	113.96
1991	14,572,585	166,198	125.03	0.11	125.14
1992	15,593,454	271,868	133.79	0.20	133.99
1993	16,632,593	471,784	142.71	0.40	143.11
1994	17,646,614	960,258	151.41	0.98	152.38
1995	18,600,203	1,641,293	159.59	2.04	161.63
1996	20,543,691	3,180,989	176.26	4.90	181.16
1997	20,982,382	6,910,496	180.03	13.13	193.16
1998	21,373,020	13,458,837	183.83	30.81	214.19
1999.9	21,687,344	22,087,900	186.08	50.57	236.65

- 통화당 평균점유시간: 99.71초

(출처: 서울 양천전화국에서 '98.1.~'99.3.까지 67,000여 음성가입자를 대상으로 월평균 통화당 점유시간을 측정하여 얻은 결과)

- 가입자당 하루 평균이용건수:

$$\text{가입자당 하루 평균이용건수/day} = 10.76 \text{calls/day}$$

(출처: 한국정보통신진흥협회 '97년도 국내 전체 총 전화 이용건수/총 전화가입자수) (/12개월/30일)

- 일반전화 가입자당 평균 트래픽 볼륨(traffic/day) = 8Kbyte * 99.71초 * 10.76 / 1,000 = 8.58Mbyte/Day

- 이동전화 가입자 월 평균 사용시간: '98년 기준 143분 /month, 286sec/day

(출처: <http://www.Strategisgroup.com/press/pubs/celltr98.html>)

- 이동전화 가입자당 평균 트래픽 볼륨

$$\begin{aligned} (\text{traffic/day}) &= 8\text{Kbyte} * 290\text{초} = 2,320\text{Kbyte}/1,000 \\ &= 2.32\text{Mbyte/Day} \end{aligned}$$

위 <표 1>은 한국정보통신진흥협회에서 제공한 자료(연도별 일반전화 가입자수, 이동전화가입자수)를 가지고서 트래픽 볼륨에 의하여 음성 전체 트래픽을 추정하여 얻은 결과이고, 그 결과에 대한 그림은 다음과 같다.

음성 트래픽 예측은 위 <표 1>의 음성 총 트래픽 추정 결과를 가지고서 2상의 다섯 가지 비선형 성장곡선 모형을 적용하여 비교 분석하였다. 그리고 분석을 위해서는 SAS(Statistical

Analysis System) 패키지를 활용하여, SAS/STAT NLIN와 ETS MODEL 명령어를 이용하였다. 이때 다섯 가지 비선형 방정식의 해(모수)는 식 (13)의 Marquardt방법을 적용 반복측정(iteration method)에 의하여 해를 구했으며, 그 결과는 <표 2>와 <그림 2>로 나타났다.

표 2. 음성 트래픽에 대한 다섯 가지 성장곡선 모형의 모수 추정 결과

모형	모수		
	K Prob> T	α Prob> T	β Prob> T
로지스틱	4471.78 0.0051***	3.662839 0.0021***	0.083721 0.0001***
프로빗	5113.32 0.8892	-2.019568 0.5069	0.035987 0.4388
곰페르츠	6029.88 0.8598	4.001268 0.4856	0.022361 0.5327
음의 지수	-	1013.21 0.8050	0.032536 0.8258
지수	-	112.11693 0.0001***	-0.080587 0.0001***

위 <표 2>는 다섯 가지 성장곡선모형의 모수 추정 결과에 대한 개별 검정 결과를 제시한 것이다. P-value 값이 유의수준 $\alpha=0.05$ 보다 작아 귀무가설 ($H_0: K, \alpha, \beta=0$)이 모두 유의한 모수로 추정된 모형은 로지스틱 성장곡선모형과 지수성장곡선 모형만이 인정되어 추정된 모수를 선택할 수 있다. 그러나 지수모형에서는 상한값(K)을 추정할 수는 없다. 다음 아래 <표 3>은 비선형 성장곡선모형에서 최적모형 선택을 위한 적합성 검정 통계량들로 음성 트래픽에 대해서 다섯 가지 성장곡선 모형의 모형선호기준 통계량의 결과들을 비교한 것이다.

<표 3>의 결과는 모형선호 기준 통계량을 제시한 것으로 정확한 실제 결정계수와 수정결정계수를 제외한 나머지 통계량은 모형간에 서로 비교하여 가장 작을수록 좋은 모형으로 판정된다. 위 결과에서는 MSE를 제외한 모든 통계량에서 로지스틱 성장곡선모형이 모형선호 기준통계량에서 가장 좋은 모

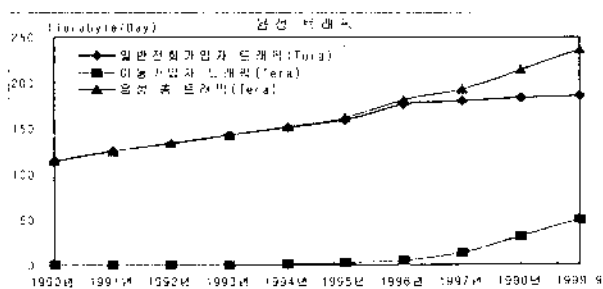


그림 1. 음성 전체 트래픽 추정.

표 3. 음성 트래픽에 대한 다섯 가지 성장곡선 모형의 모형선택기준 통계량 결과

통계량	로지스틱	프로빗	곰페르츠	음의지수	지수
ME	94.4258	317.7669	109.2463	167.5348	856.6916
MPE	68.8865	213.5720	58.6892	101.1028	528.4896
MAE	97.7430	317.7669	112.8757	167.5348	856.6916
MAPE	70.28825	213.5719	61.8796	101.1027	528.4896
MSE	16.47668	21.09555	25.38407	384.1621	13.37058
RMSE	113.9499	328.5336	133.4328	172.0416	864.3355
RMSPE	87.2309	234.6840	68.87396	101.1041	530.6740
R_c^2	0.9920	0.9898	0.9877	0.3327	0.9911
$\overline{R_c^2}$	0.9897	0.9868	0.9841	0.3214	9.9889
UI	0.6709	1.9344	0.7857	1.0113	5.0893
U	0.2643	0.5014	0.5584	0.9993	0.7183

형으로 판명되었다. 그리고 비선형 성장곡선모형의 결정계수는 선형회귀모형에서 구하는 결정계수와는 식이 다르며, 결정계수 값이 높을수록 좋은 모형으로 설명한다. 다음 비선형 성장곡선 모형의 모형 판정 비교를 위한 모형선택기준통계량에 대한 식은 아래와 같다.

(1) 평균오차(Mean Error)

$$ME = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i) \tag{14}$$

(2) 평균백분율오차(Mean Percentage Error)

$$MPE = \frac{100}{N} \sum_{i=1}^N (\hat{y}_i - y_i) / y_i \tag{15}$$

(3) 평균절대오차(Mean Absolute Error)

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \tag{16}$$

(4) 평균절대백분율오차(Mean Absolute Percentage Error)

$$MAPE = \frac{100}{N} \sum_{i=1}^N |(\hat{y}_i - y_i) / y_i| \tag{17}$$

(5) 평균제곱오차(Mean Squared Error)

$$MSE = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 \tag{18}$$

(6) 평방근 평균제곱오차(Root Mean Square Error)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \tag{19}$$

(7) 평방근평균제곱백분율오차(Root Mean Square Percentage Error)

$$RMSPE = 100 \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i / y_i)^2} \tag{20}$$

이상의 통계량들은 그 값이 작을수록 해당 모형을 선택한다.

(8) 정확한 실제 결정계수(Corrected Actual R-square)

$$R_c^2 = 1 - \frac{SSE}{CSSA} \tag{21}$$

$$SSE = \sum_{i=1}^N (\hat{y}_i - y_i)^2$$

$$SSA = \sum_{i=1}^N y_i^2$$

$$CSSA = \sum_{i=1}^N y_i^2 - \left(\sum_{i=1}^N y_i \right)^2 / N = \sum_{i=1}^N (y_i - \bar{y})^2$$

y_i = 실제값(actual value),

\hat{y}_i = 예측값(predicted value)

SSA는 실제(actual) 자료의 제곱합(Sum of the Squares)을 의미하고, CSSA에 있어서 \bar{y} 는 실제 자료의 평균(actual mean)에 의해서 구해진다.

(9) 정확한 실제 수정결정계수(Corrected Actual Adjust R-square)

$$\overline{R_c^2} = \frac{(n-1) R_c^2 - k}{n - k - 1} \tag{22}$$

$k = p - 1$,

n = 표본 수,

p = 추정된 모수의 수

결정계수의 범위는 0과 1사이의 값을 가지며 1에 근접한 값을 가질수록 적용한 모형은 좋은 설명력을 갖는 모형이다.

(10) 예측의 정확성 통계량(a statistic measuring the accuracy of a forecast) UI

$$UI = \frac{MSE}{\sqrt{\frac{1}{N} \sum_{i=1}^N y_i^2}} \tag{23}$$

(11) Theil's 부등식 계수(inequality coefficient) U

$$U = \frac{MSE}{\sqrt{\frac{1}{N} \sum_{i=1}^N y_i^2 + \frac{1}{N} \sum_{i=1}^N \hat{y}_i^2}} \tag{24}$$

Theil's의 부등식 계수 UI과 U의 계수 값은 0에 근접한 때 좋은 모형으로 결정된다. 그리고 계수 값이 0.55 이하이면, 추정된 모형에서 예측된 결과가 정확한 것으로 판정한다(Young, 1990). 다음 아래 <그림 2>는 최적모형으로 선택된 로지스틱 성장곡선모형에 의한 음성 트래픽의 예측 결과를 나타낸 것이다.

4. 인터넷 트래픽 추정 및 예측

인터넷 트래픽은 실제 트래픽 자료를 구하기가 쉽지 않아서 인터넷 트래픽에 영향을 미치는 변수를 조사하여 고려하였다. ISP전용회선 가입기관, ISP PPP/Shell 가입자, PC통신 가입자에 대한 1995년부터 1998.11까지의 월별 자료(출처: 한국전산원 인터넷 통계자료, <http://www.nic.or.kr/net/>)를 가지고서 각각에 로지스틱 성장곡선 모형을 적용하여 2002년까지를 예측하였다. 그리고 2002년까지 예측된 결과를 가지고서 ISP전용회선 가입기관, PPP/Shell 가입자, PC통신가입자에 대하여 한국전산

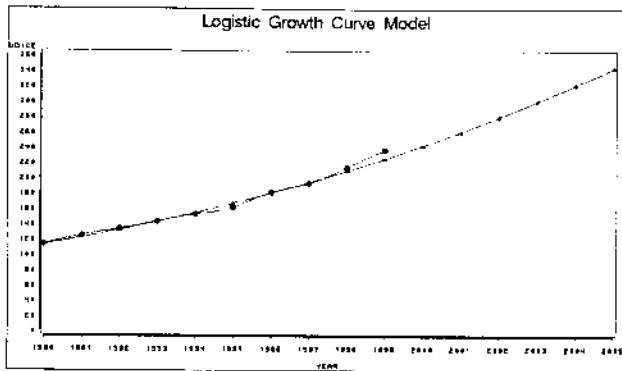


그림 2. 로지스틱 성장곡선 모형에 의한 음성 총 트래픽예측.

원에서 발표한 “한국인터넷정보센터 운영월보”에 제시한 인터넷 이용자 산출방법을 적용하여 국내 인터넷 이용자수를 추정하였다.

인터넷 트래픽은 2002년까지 예측된 연도별 인터넷 이용자수(기관: ISP 전용회선가입자, 개인: ISP PPP/Shell 가입자, PC통신 가입자)에 대하여 인터넷 트래픽의 평균 대역폭, 평균 사용시간에 대하여 인터넷 이용자당 평균 트래픽 볼륨을 적용하여 추정하였다.

4.1 시나리오 1

시나리오 1의 트래픽 볼륨은 이용자별 주어진 가입자 정합장치(개인: 모뎀(33Kbps, 56Kbps), 기관: LAN)별 인터넷상에서 자료를 다운로드 받을 경우 나타나는 실제 대역폭을 측정하였다. 이때 측정방법으로 가정에서는 모뎀으로 PC통신에 접속하여 인터넷에 들어가 자료를 다운로드 하였고, 기관의 경우는 PC LAN환경에서 인터넷에 접속하여 다른 인터넷 사이트의 자료를 다운로드시 나타나는 실제 대역폭을 측정하는 방법으로 각각의 자료를 랜덤하게 시간대별로 가정에는 50명의 학생이 10회 이상 측정하였고, 기관의 경우는 학교와 연구소에서 총 100번을 측정하여 얻어진 자료에서 인터넷 트래픽의 평균대역폭을 얻었다.

· 인터넷 이용자수 산출 방법

$$\text{인터넷 이용자수} = (\text{전용선이용기관} \times A) + \text{PPP/Shell 이용자} + (\text{PC통신 이용자} \times B)$$

· A, B의 보정계수 의미

(출처 : 한국인터넷정보센터 설문조사에 의한 A, B 보정계수 설정)

· A는 전용선 이용기관을 6개 기관으로 세분화(ac, co, go, or, re, nc, 지역)하여 200여 기관을 표본추출하여 전화문의 설문조사 결과 얻어진 한 기관당 평균 이용자수를 보정계수로 적용하여 계산

· A는 전용선을 이용하는 인터넷 이용기관으로 기관의 규모에 따라 기관당 인터넷 이용자수의 편차가 수십 명에서 수백 명으로 분산이 크게 나타날 것으로 예상된다.

표 4. 전용선을 이용하는 기관별 평균 이용자수

(1998. 12. 기준)

도메인 구분	비율	추정된 기관수	한 기관당 평균 이용자수	기관 이용자수 (기관수×기관당 평균 이용자수)
CO	92%	12,736	56	713,191
OR	4%	763	763	422,488
AC	1%	138	454	62,847
GO	1%	138	163	22,564
RE	1%	138	23	3,184
NE	1%	138	32	4,430

· B는 PC통신을 통한 실제 인터넷이용자 비율 (20% 적용)을 보정계수로 적용

· 위 표에서 한 기관당 평균 이용자 수는 6개 기관으로 구분하여 200여 개의 기관을 랜덤하게 표본을 추출하여 전화응답에서 평균 이용자수를 구한 결과임.

- PC통신가입자 중에서 인터넷 이용자는 '98년 기준 평균 20% 정도가 이용하며, '99년 부터 PC통신가입자 중 인터넷서비스는 기본서비스로 이용하게 되어, PC통신이용자중 인터넷 이용자수는 크게 증가될 전망이다.

· 개인 가입자의 경우 기존 모뎀을 이용 최대 56Kbps의 대역폭을 이용하였으나, 현재는 ISDN, ADSL, CATV 모뎀으로 대체되어 보다 높은 대역폭에서 인터넷을 이용하는 사용자 수가 크게 증가될 전망이다.

- 인터넷 이용자당 평균 대역폭은 공중백본망 및 가입자망 업그레이드로 인하여 '98년 이후부터는 매년 1.4 ~ 1.5배 정도로 증가될 전망이다.

· 인터넷 이용자중 ISP전용회선이용 기관의 LAN 사용 평균 대역폭 : 전용회선을 이용하는 연구소와 학교에서 '99년 4~5월에 임위의 시간대별 총 100회를 실시 배 인터넷을 통한 자료를 다운로드시 나타나는 대역폭의 자료를 수집하여 얻은 평균 대역폭임.

· 인터넷 이용자 중 개인 PC 모뎀 사용 평균 대역폭 : 가정에서 50명의 학생이 랜덤하게 시간대별로 10회 이상 PC통신을 이용 인터넷에 접속하여 자료를 다운로드시 나타나는 대역폭의 자료를 수집하여 얻은 평균 대역폭임.

- 인터넷 평균사용시간 : 1hours/day(인터넷 사용빈도 72% : 1~2시간)

(출처 : 인터넷 이용형태 조사(1차) 결과, 한국인터넷정보센터, 1997.12.)

아래 <표 5>는 한국전산원 인터넷 통계자료에서 제공한 자료(월별 ISP전용회선 가입기관, ISP PPP/Shell 가입자)와 한국정보통신진흥협회에서 제공한 자료(연도별 PC통신 가입자)를 가지고서 이용자별 주어진 가입자 정합장치(개인: 모뎀, 기관: LAN)별 인터넷 이용자 트래픽 평균 대역폭 기준에 의하여 인터넷 총 트래픽을 그 결과는 아래 그림과 같다.

표 5. ISP 전용회선 가입기관, ISP PPP/Shell 가입자, PC통신 가입자에 대한 예측결과 및 인터넷 이용자수, 인터넷 트래픽 추정 결과

연도	ISP전용회선 가입기관	ISP PPP/Shell 가입자	PC통신 가입자	인터넷 이용자	인터넷 총 트래픽 (Terabyte/Day)
1995	613	21,774	766,192	301,439	0.31
1996	2,212	85,764	1,711,013	803,158	0.94
1997	6,186	319,344	3,105,321	1,869,540	2.52
1998	13,843	754,680	4,828,115	3,668,870	6.44
1999	18,750	826,109	5,561,625	5,113,259	11.96
2000	22,162	855,399	6,190,534	6,189,769	22.40
2001	23,389	858,574	6,498,554	6,914,531	43.29
2002	23,769	858,977	6,636,972	7,480,437	86.14

인터넷 트래픽 예측은 위 <표 5>의 인터넷 트래픽 추정 결과를 가지고서 2상의 다섯 가지 성장곡선 모형을 적용하여 분석하였으며, 이에 대한 모수 추정 결과는 아래와 같다.

위 <표 6>은 다섯 가지 성장곡선모형의 모수 추정 결과에 대한 개별 검정 결과이다. 위 결과에서 P-value 값이 유의수준

(α)를 추정할 수 없다. 다음 아래 <표 7>은 비선형 성장곡선 모형에서 최적모형 선택을 위한 모형선호기준 통계량이다. 이것은 시나리오 1에 대한 인터넷 트래픽의 다섯 가지 성장곡선 모형의 모형선호기준 통계량들에 대한 결과를 비교한 것이다.

표 7. 시나리오 1에 의한 인터넷 트래픽의 다섯가지 성장곡선 모형의 모형선호기준 통계량 결과

통계량 모형	로지스틱	프로빗	곰페르츠	음의지수	지수
	ME	17.4935	477.6767	20.9293	21.7292
MPE	147.7693	33939	68.3521	99.7842	641.5870
MAE	19.5489	477.6767	21.9227	21.7292	14.0835
MAPE	238.6046	33939	208.6929	99.7841	656.8769
MSE	0.38711	1.01772	2.46914	0.41502	0.29580
RMSE	32.7086	481.6496	35.3249	35.3108	17.3396
RMSPE	429.9570	66727	371.9409	99.7844	1203
R_c^2	0.9997	0.9992	0.9980	0.9996	0.9997
\overline{R}_c^2	0.9996	0.9989	0.9972	0.9995	0.9997
U_1	0.9257	13.6315	0.9998	0.9994	0.4907
U	0.8260	0.8985	0.9613	0.9986	0.2785

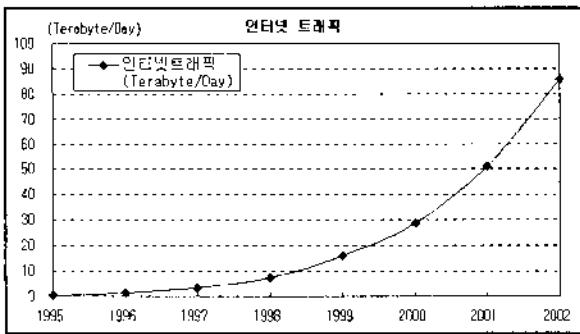


그림 3. 시나리오 1에 의한 인터넷 트래픽추정.

$\alpha=0.05$ 보다 작아 귀무가설 ($H_0: K, \alpha, \beta=0$)이 모두 유의한 모수로 추정된 모형은 음의지수 성장곡선모형과 지수성장곡선 모형만이 인정되어 추정된 모수를 선택할 수 있다. 그러나 음의지수 성장곡선모형과 지수성장곡선모형에서는 상한값

표 6. 시나리오 1에 의한 인터넷 트래픽의 다섯 가지 성장곡선 모형의 모수 추정 결과

모수 모형	K	α	β
	Prob> T	Prob> T	Prob> T
로지스틱	4531.16 0.8289	8.711071 0.0963	0.680705 0.0001***
프로빗	52436.73 0.8387	-4.310711 0.0059**	0.195564 0.0427*
곰페르츠	30414.33 0.8036	11.959374 0.0085**	0.101451 0.1446
음의 지수	-	-0.887727 0.0001***	-0.654504 0.0001***
지수	-	0.768306 0.0001***	-0.673948 0.0001***

<표 7>의 모형선호 기준통계량을 모형간에 서로 비교한 결과 지수성장곡선모형이 가장 좋은 모형인 듯 보인다. 하지만, 아래 <표 8>에 의한 정규성 가정 검토를 실시하면 다음과 같다.

<표 8>의 Shapiro-Wilk 정규성 검정통계량은 P-value값이 $P < \alpha$ 작은 확률 값을 구한 것으로, 귀무가설 (H_0 : 정규성 분포를 따른다)에 의한 정규성 검정결과 지수성장곡선모형만이 유의하게 판정된다. 따라서 지수성장곡선모형을 제외한 나머지 모형은 모두 정규성 가정을 충족하는 것으로 나타난다. 결국 시나리오 1에서 최적모형으로는 음의지수 성장곡선모형이 선택되었음을 알 수 있다.

다음 <그림 4>는 최적모형으로 선택된 음의지수 성장곡선

표 8. 시나리오 1에 의한 다섯 가지 성장곡선 모형의 정규성 검정 결과

Shapiro-Wilk	로지스틱	프로빗	곰페르츠	음의 지수	지수
$\hat{W} =$	0.885	0.937	0.931	0.949	0.754
$P_r < W =$	0.215	0.584	0.533	0.704	0.009

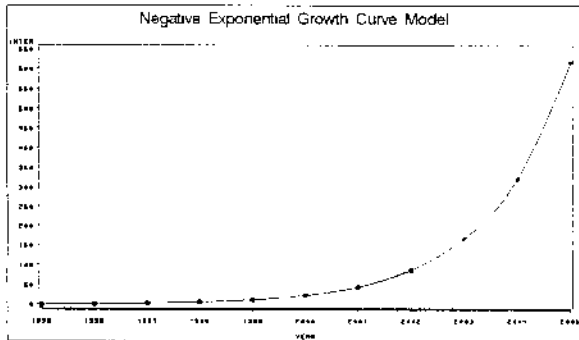


그림 4. 시나리오 1의 음의지수 성장곡선모형에 의한 인터넷 총 트래픽 예측.

모형에 의한 시나리오 1의 인터넷 트래픽 예측 결과를 나타낸 것이다.

4.2 시나리오 2

시나리오 2는 인터넷 국내 이용자에 대하여 통신환경(백본망, 가입자망)이 약간 앞서는 아시아 국가 중 호주의 연도별 평균 증가 대역폭(Ovem, 1998, "Internet Market Forecast," P. 162) 자료를 적용하여 인터넷 트래픽을 산출한다. 호주의 연

도별 증가 대역폭을 적용하면, 실제 국내 인터넷 대역폭 보다 약간 앞설 것으로 생각되며, 앞4.1절의 시나리오 1에 의한 국내 인터넷 트래픽의 결과와 비교하여 실제로 시나리오 2의 인터넷 트래픽 결과가 약간 높은 형태로 나타나는지를 살펴본다. 시나리오 1의 인터넷 예측치가 국내 인터넷 트래픽을 정확히 설명하는지를 살펴본다.

시나리오 2의 인터넷 트래픽 예측은 시나리오 1의 <표 5>에서 ISP 전용회선 가입기관, ISP PPP/Shell 가입자, PC통신 가입자에 대한 예측결과 및 인터넷 이용자수를 가지고서 기관과 개인에 대하여 호주의 연도별 인터넷 트래픽 대역폭을 적용하여 인터넷 트래픽을 추정 및 예측한다. 즉, 국내의 연도별 인터넷 대역폭 자료가 제시되어 있다면, 이것을 적용하여 시나리오 1과 시나리오 2의 예측 결과가 동일한지 살펴본다. 같은 경우, 시나리오 1의 예측결과가 맞은지를 알 수 있으나, 현재 국내에 대한 트래픽 연도별 평균 증가 대역폭 자료를 얻을 수 없어서

표 9. 아시아 각국의 인터넷 이용자 연도별 평균 증가 대역폭 (단위: Kbyte/User)

구분	1998	1999	2000	2001	2002	2003	2004	2005
아시아기관	0.23	0.32	0.44	0.90	1.06	1.24	1.45	1.26
개인	0.14	0.23	0.36	0.57	0.87	1.27	1.82	2.56
호주 기관	0.24	0.33	0.45	0.93	1.11	1.30	1.51	1.33
개인	0.10	0.17	0.29	0.46	0.72	1.08	1.59	2.29
홍콩 기관	0.23	0.32	0.43	0.87	1.04	1.23	1.42	1.27
개인	0.12	0.22	0.37	0.60	0.96	1.47	2.21	3.25
일본 기관	0.25	0.34	0.47	0.93	1.11	1.30	1.51	1.35
개인	0.16	0.25	0.39	0.62	0.97	1.47	2.16	3.11
싱가포르기관	0.23	0.31	0.42	0.85	1.02	1.21	1.41	1.27
개인	0.12	0.22	0.37	0.60	0.96	1.47	2.20	3.23

<출처: Ovum, 1998, Internet Market Forecasts Indicators for Global Internet Growth 1998-2005>

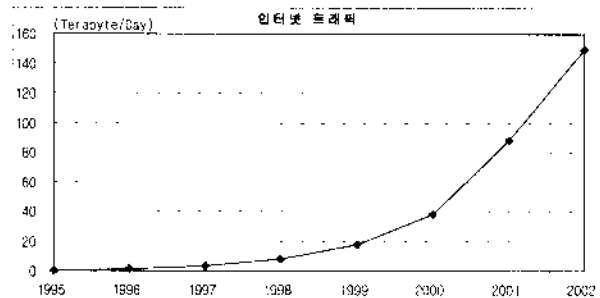


그림 5. 시나리오 2에 의한 인터넷 트래픽 추정.

상대적 비교를 위해서 호주의 연도별 평균 증가 대역폭을 적용하게 되었다. 그리고 호주의 인터넷을 이용하는 통신환경은 국내보다 앞서서 백본망 및 가입자 망을 고속화하여 망을 설치 운영하고 있다. 한편, 국내는 2000년까지 6대 광역시 및 신도시에 ATM백본망을 설치하고, 2002년까지는 시내 통화권역 144개지역을 확대 설치하여 1.5 M~2Mbps급 서비스를 전국 아파트를 대상으로 전국도시의 45% 정도 서비스를 실시할 예정이다. 다음 <표 9>는 아시아 각국의 인터넷 이용자 연도별 평균 증가 대역폭을 나타내고 있다. 인터넷 이용자 평균 대역폭은 아시아 지역에 있어서 개인 이용자의 경우 평균 대역폭은 기관 이용자에 비하여 현재는 대역폭이 낮은 상태이지만, 호주는 2004년, 홍콩은 2003년, 일본은 2003년, 싱가포르 경우도 2003년 정도가 되면 가정에서 인터넷을 이용하는 평균 대역폭이 기관에서 이용하는 업무용 보다 가입자망의 고도화에 의하여 가정에서 이용하는 인터넷 평균 대역폭이 높아질 것으로 전망하고 있다. 다음 아래 <표 10>은 시나리오 2에 대한 인터넷 트래픽 추정 결과를 나타낸다.

위 <표 10>은 한국전산원 인터넷 통계자료에서 제공한 자료(월별 ISP전용회선 가입기관, ISP PPP/Shell 가입자)와 한국정보통신진흥협회에서 제공한 연도별 PC통신 가입자)자료를 가지고서 이용자별 주어진 가입자 정합장치(개인: 모뎀, 기관:

표 10. ISP 전용회선 가입기관, ISP PPP/Shell, PC통신 가입자에 대한 예측결과 및 인터넷 이용자수, 인터넷 트래픽 추정 결과

연도	ISP전용회선 가입기관	ISP PPP/Shell 가입자	PC통신 가입자	인터넷 이용자	인터넷 트래픽 (Terabyte/Day)
1995	613	21,774	766,192	301,439	0.53
1996	2,212	85,764	1,711,013	803,158	1.65
1997	6,186	319,344	3,105,321	1,869,540	3.49
1998	13,843	754,680	4,828,115	3,668,870	6.86
1999	18,750	826,109	5,561,625	5,113,259	12.27
2000	22,162	855,399	6,190,534	6,189,769	27.37
2001	23,389	858,574	6,498,554	6,914,531	60.04
2002	23,769	858,977	6,636,972	7,480,437	107.03

LAN)별 인터넷 이용자에 대하여 호주의 평균 대역폭을 적용하여 트래픽량을 산출하여, 그 결과를 가지고서 인터넷 총 트래픽을 추정하였다. 이 추정된 결과를 그림으로 나타내면 아래와 같다.

시나리오 2에 대한 인터넷 트래픽 예측은 <표 10>의 인터넷 트래픽 추정 결과를 가지고서 2장의 다섯 가지 비선형 성장곡선 모형을 적용하여 분석하였다. 그리고 이때 산출된 모수 추정 결과를 표로 나타내면 아래와 같다.

표 11. 시나리오 2에 의한 인터넷 트래픽의 다섯가지 성장곡선 모형의 모수 추정 결과

모수	모형		
	K Prob> T	α Prob> T	β Prob> T
로지스틱	253.68964 0.0813	6.445016 0.0001***	0.876076 0.0001***
프로빗	10637.25 0.7846	-3.925442 0.0024***	0.229197 0.0464*
곰페르츠	10733.80 0.6348	11.027110 0.0002***	0.124776 0.0441*
음의 지수	-	-1.22013 0.0011***	-0.642261 0.0001***
지수	-	1.036365 0.0010***	-0.664322 0.0001***

위 <표 11>은 다섯 가지 성장곡선모형의 모수 추정 결과에 대한 ($H_0: K, \alpha, \beta$ 개별 검정 결과이다. 위 결과에서 P-value 값이 유의수준 $\alpha=0.05$ 보다 작아 귀무가설 $\beta=0$)이 모두 유의한 모수로 추정된 모형은 음의지수 성장곡선모형과 지수성장곡선모형만이 유의하므로 추정된 모수를 선택할 수 있다. 그러나 음의지수 성장곡선모형과 지수성장곡선모형에서는 상한값(K)을 추정할 수 없다. 다음 아래 <표 12>는 비선형 성장곡선모형에서 최적모형 선택을 위한 적합성 검정 통계량이다.

이것은 시나리오 2에 대한 인터넷 트래픽의 다섯 가지 성장곡선 모형의 모형선호기준 통계량들에 대한 결과를 비교한 것이다.

앞 <표 11>에서 비선형 성장곡선모형간의 모수 추정에 대

표 12. 시나리오 2에 의한 인터넷 트래픽의 다섯 가지 성장곡선 모형의 모형선호 기준통계량 결과

모형 \ 통계량	로지스틱	프로빗	곰페르츠	음의지수	지수
ME	-23.5154	472.0630	-27.2693	-27.3429	-26.7585
MPE	-58.0471	20747	-87.1786	-99.8381	-79.9220
MAE	25.1609	472.063	27.2693	27.3429	26.7585
MAPE	154.7081	20747	87.17856	99.83812	79.9220
MSE	1.06645	2.92145	2.39873	4.15464	4.95997
RMSE	42.6535	477.0777	44.7435	44.7233	44.3278
RMSPE	257.2045	39119	91.8684	99.8382	85.2941
R_c^2	0.9995	0.9985	0.9988	0.9975	0.9970
\overline{R}_c^2	0.9993	0.9980	0.9983	0.9971	0.9965
U	0.9532	10.6619	0.9999	0.9995	0.9907
\overline{U}	0.8766	0.8746	0.9959	0.9989	0.9773

한 개별검정에서 음의지수 성장곡선모형과 지수성장곡선모형만이 선택되어, 두 성장곡선모형에서 적합성 검정을 위한 모형선호기준통계량 <표 12>를 가지고서 서로 비교하여 보면, 음의지수 성장곡선모형과 지수성장곡선모형에서 적합하게 판정되었다. 다음 아래 표는 성장곡선모형에 대한 정규성 가정에 대한 검정 결과를 나타낸다.

<표 13>의 Shapiro-Wilk 정규성 검정통계량은 P-value값이 $P < \alpha$ 작을 확률 값을 구한 것으로, 귀무가설(H_0 : 정규성 분포를 따른다)에 의한 정규성 검정결과 지수성장곡선모형만이 유의하게 판정된다. 따라서 지수성장곡선모형을 제외한 나머지 모형은 모두 정규성 가정을 충족하는 것으로 나타난다

결국 시나리오 2에서 최적모형으로는 음의지수 성장곡선모형이 선택되었음을 알 수 있다. 다음 <그림 6>은 최적모형으로 선택된 음의지수 성장곡선모형에 의한 시나리오 2의 인터넷 트래픽 예측 결과를 나타낸 것이다.

표 13. 시나리오 2에 의한 다섯 가지 성장곡선 모형의 정규성 검정 결과

Shapiro-Wilk	로지스틱	프로빗	곰페르츠	음의 지수	지수
$\hat{W} =$	0.938	0.921	0.931	0.916	0.745
$P_r < W =$	0.597	0.784	0.527	0.403	0.008

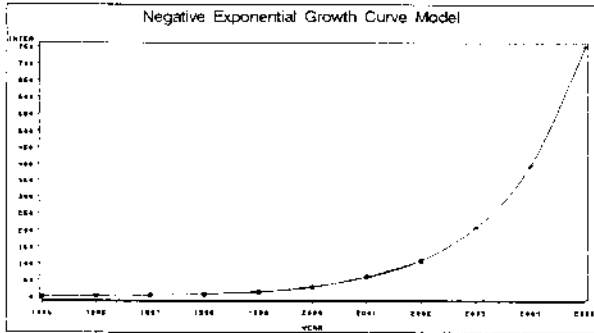


그림 6. 시나리오 2의 음의 지수 성장곡선모형에 의한 인터넷 총 트래픽 예측 결과.

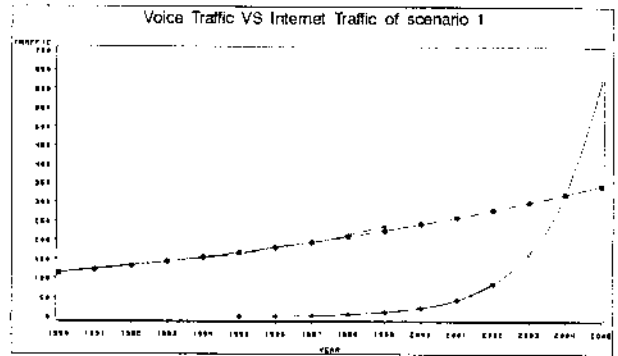


그림 7. 음성 트래픽과 시나리오 1을 적용한 인터넷 트래픽 간의 예측 비교.

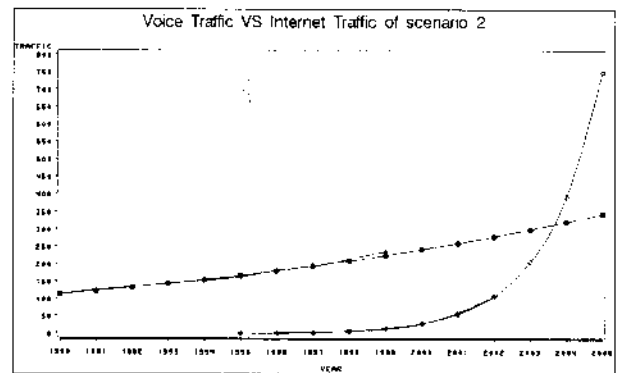


그림 8. 음성 트래픽과 시나리오 2를 적용한 인터넷 트래픽 간의 예측 비교.

5. 결론

음성 트래픽과 인터넷 트래픽간의 예측 결과를 비교하는데 있어서 앞 4장에서 음성 트래픽에 대해서는 최적모형으로 선택된 로지스틱 성장곡선모형을 적용하여 음성 트래픽을 예측하였다(황정연, 1997). 그리고 인터넷 트래픽에 있어서 시나리오 1의 경우는 기관 및 개인 이용자 형태별 대역폭을 구분하여 추정된 자료를 가지고서 최적모형으로 선택된 음의지수 성장곡선모형을 적용하여 인터넷 트래픽을 예측하였다. 또한, 시나리오 2는 인터넷 국내 이용자에 대하여 통신환경(백분망, 가입자망)이 앞서 망을 고도화한 아시아 국가 중 호주의 연도별 평균 증가 대역폭을 Ovrm, 1998년 ("Internet Market Forecast," p. 162)에 발간된 자료를 적용하여 시나리오 2의 인터넷 트래픽을 산출한 것이다. 국내의 연도별 평균 증가 대역폭에 대한 자료가 존재하여 이것을 시나리오 2에 적용하여 트래픽을 예측, 시나리오 1과 비교하여 동일한지 살펴보면 가장 쉽게 알 수 있는 것이다. 그러나 현재 국내에 대한 연도별 평균 증가 대역폭의 자료를 구할 수 없어서 아시아 국가 중 국내보다 먼저 망 고도화를 실시한 호주의 연도별 증가 대역폭을 적용하였다. 그리고 실제 국내 인터넷 대역폭 보다 약간 앞설 것으로 생각한다.

이에 따라서 4.1절의 시나리오 1에 의한 국내 인터넷 트래픽의 결과치를 시나리오 2의 인터넷 트래픽과 비교하여 보면 예측 결과가 약간 낮은 형태로 나타나는지를 살펴보면 된다. 그래서 시나리오 1의 국내 인터넷 트래픽이 정확히 예측되었는지를 살펴볼 수 있을 것이다. 실제 예측된 결과를 비교하기 위

하여 아래 <그림 7, 8>에서 살펴보면, <그림 7>의 시나리오 1의 경우 음성 트래픽에 대해서는 로지스틱 성장곡선 모형에 의하여 예측한 결과이고, 인터넷 트래픽은 음의지수 성장곡선모형을 적용하여 예측한 결과이다. 이때, 인터넷 트래픽이 2004년을 지나면서부터 2005년에는 인터넷 트래픽이 음성 트래픽을 앞서는 것으로 예측되었다.

다음 아래 <그림 8>은 시나리오 2에 의하여 예측된 결과로 음성 트래픽은 시나리오 1과 마찬가지로 동일한 결과이고, 인터넷 트래픽은 최적모형으로 선택된 음의 지수성장곡선모형을 적용하여 예측한 결과에 대한 그림이다. <그림 8>에서는 2003년을 지나 2004년에는 인터넷 트래픽이 음성 트래픽을 앞서는 것으로 예측되었다. 즉, 시나리오 1의 적용에 의한 인터넷 예측값은 시나리오 2의 예측값보다 1년 정도 낮게 예측되는 것으로 보아 시나리오 1의 적용은 타당함을 알 수 있다. 즉, 아래 <그림 8>은 시나리오 2에 해당하는 것으로 호주의 인터넷 연도별 평균 대역폭을 적용하여 인터넷 트래픽을 예측한 결과로 인터넷 트래픽은 음성 트래픽에 2003년 후반기를 지나면서부터 2004년에 이르러 크게 인터넷 트래픽이 음성 트래픽을 앞서는 형태로 <그림 7>의 결과보다 1년정도 앞서는 것으로 예측되었다. 국내 인터넷 트래픽은 약 2004년을 지나면서부터 2005년에는 인터넷 트래픽이 음성 트래픽을 앞서게 될 것으로 나타났다. 그리고 참고적으로 AT&T에서 분석한 "ATM Year '98"

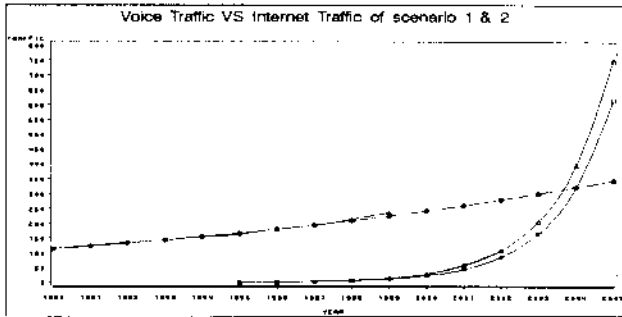


그림 9. 음성 트래픽과 시나리오 1, 2를 적용한 인터넷 트래픽 간의 예측 비교.

보고서에 의하면 미국의 경우 2000년을 지나 2001년부터 인터넷 트래픽이 음성 트래픽을 앞서는 것으로 나타났다.

참고문헌

황정연(1997). 비선형 성장곡선 모형의 분석 절차에 관한 연구, *대한품질*

질경영학회지, 25(1).

황정연, 강병용(1998). 최적모형으로 선택된 로지스틱 모형과 음의자수 성장곡선모형에 의한 예측과 신뢰구간 추정, NCS'98 (The Second Conference on Next Generation Communication Software), *한국통신학회 차세대 통신소프트웨어 학술대회*.

홍종선(1992). 통계자료분석, 탐진.

Bewley, R., and Fiebig, D. G.(1988), Flexible logistic growth model with application in telecommunications, *International Journal of Forecasting*, 4, 177-192.

Lindberg, B.C.(1982), International comparison of growth in demand for a new durable consumer product, *Journal of Marketing Research*, 19, 364-371.

Daniel, B.(1998), Internet market forecasts indicators for global internet growth 1998-2005, *Ovum Ltd*, 158-162.

SAS/Procedure and ETS Guide for PC, Ver. 6Ed.,(1990), SAS Institute Inc.

SAS/Procedure and STAT Guide for PC, Ver. 6Ed.,(1990), SAS Institute Inc.

Young, P., and Ord, J.K.,(1990), "Model Selection and Estimation for Technological Growth Curves", *International Journal of Forecasting* 5, 501-514.

Young, P.(1993), "Technological Growth Curves: A Competition of Forecasting Models", *Technological Forecasting and Social Change* 44, 375-389.



황정연

1993년 중앙대 통계학 석사
1997년 충남대 통계학과 박사과정
현재: 한국전자통신연구원
시스템분석팀 선임 연구원
관심 분야: 경제성공학, 시스템분석,
신경망이론, 데이터마이닝



전경표

1979년 KAIST 산업공학 석사
1988년 Univ. of North Carolina State 산업공학 박사
현재: 한국전자통신연구원 통신시스템
연구부장 책임 연구원
관심 분야: 큐잉이론, 시스템분석,
신경망이론, 트래픽 분석



강병용

1983년 경북대 경제학 석사
현재: 한국전자통신연구원 시스템
분석팀 장책임연구원 재직
관심 분야: 경제성공학, 시스템분석,
벤치마킹, 수요예측