Communications of the Association for Information Systems

Volume 17 Article 5

January 2006

Global Diffusion of the Internet IX: Predicting Global Diffusion of the Internet: An Alternative to Diffusion Models

Somnath Mukhopadhyay
University of Texas at El Paso, smukhopadhyay@utep.edu

Communications of the Association for Information Systems

GLOBAL DIFFUSION OF THE INTERNET IX: PREDICTING GLOBAL DIFFUSION OF THE INTERNET: AN ALTERNATIVE TO DIFFUSION MODELS

Somnath Mukhopadhyay
Information and Decision Sciences Department
The University of Texas at El Paso,
smukhopadhyay@utep.edu

ABSTRACT

This research builds Internet growth forecasting models based on existing knowledge of diffusion and connectionist theories. It shows that a simple connectionist multi-layered perceptron artificial neural network (MLP) model can create a flexible response function to forecast Internet growth for the near future. This paper identifies the most suitable diffusion models that generate predictions for the Internet diffusion with low errors. However, the MLP model is superior to the best diffusion model on both the calibration and the validation samples of Internet growth data. This research also investigates the process of combining diffusion and connectionist models. The findings will encourage researchers to use connectionist models to predict diffusion of other innovation processes also.

Keywords: Internet Growth, Diffusion Models, Neural Networks, Forecasting

I. INTRODUCTION

Studying the diffusion of the Internet is important for both government policy makers and business investors [Wolcott and Goodman, 2003; Press, 1997]. Inaccurate predictions of Internet growth can lead to inadequate capacity planning. Models that explain and predict the Internet growth are useful for policy makers, e-market planners, hardware and software companies, training enterprises, and e-commerce related companies. These companies may adjust their strategic plans to account for Internet growth in the potential markets. Multinational enterprises involved in electronic commerce can use global Internet growth predictions as an attribute in selecting the International market of their choice for entry. E-commerce and other business planners can benefit by orienting their strategic plans to exploit Internet diffusion [Samaddar et al., 2002]. Measuring Internet growth with precision is difficult [Press, 1997]. This research compares alternative models for predicting Internet growth.

This study makes contributions to information systems (IS) research in several ways. In the last 20 years, numerous studies on diffusion models sought to explain the diffusion of an innovation process [Gurbaxani, 1990; Mahajan et al., 1990; Venkatraman et al., 1994; Mahajan et al., 1998; Rai et al., 1998]. However, few studies use these models to forecast growth. To our knowledge this paper reports on the first use of connectionist models in conjunction with diffusion models to forecast Internet growth. This research is also the first to compare diffusion models with connectionist models on Internet growth data. The findings from this research will encourage IS researchers and practitioners to use connectionist models in addition to diffusion models to predict diffusion of the Internet and other innovation processes.

Section II discusses the choice of the diffusion models for prediction of the Internet growth. Section III explains the research method and data for this study. Section IV reports the results with analysis. Finally, section V summarizes the results and analysis with conclusion and future research direction.

II. MODELING ALTERNATIVES OF THE INTERNET GROWTH

Understanding Internet growth patterns involves assessing alternative models for Internet diffusion [Rai et al., 1998]. There were significant researches on diffusion models using historical data to explain the adoption of an innovation process [Gurbaxani, 1990; Mahajan et al., 1990; Mahajan et al., 1998, Rai et al., 1998]. One assumption behind most diffusion modeling is that there are a fixed number of potential adopters of new technologies [Rogers, 1983]. Therefore, the adoption process targets an ever smaller number of adopters as time goes by. The diffusion process follows a *simple logistic curve* (s-shaped) over time through imitation [Mansfield, 1961]. Two main factors responsible for the growth process are *imitation* and *innovation* [Bass 1969]. These factors were later called *internal* and *external influences* [Mahajan and Muller, 1979]. Internal influence is the influence from early adopters on potential late adopters. Late adopters imitate early adopters if early adoption is successful. External influence is the impact of factors other than imitation on the growth process. For example, over time new and similar innovations (external influence) may cause the growth of the original innovation to decline. Favorable government policies may cause a sudden acceleration of growth, recognized by a one-time jump in the cumulative growth curve. Diffusion models are, therefore, of three basic types: internal-influence, external-influence, and models with both internal and external influences [Venkatraman et al., 1994]. Diffusion models are a logical first choice in modeling the Internet growth process since many studies frequently used diffusion models in predicting technological growth.

The literature on new technology diffusion is really about S-curves. S-curves are roughly consistent with the facts because s-curves do not consider failure of an innovation process [Geroski, 2000]. Other studies suggest that one should look for alternate approaches [Rai, et al., 1998; Dekimpe et al., 1998]. This paper offers a new approach which utilizes the power of artificial intelligence (AI) modeling to forecast the growth of the Internet. This study uses one of the most popular modern modeling techniques: *artificial neural network* or simply *Neural Network (NN)* also called *connectionist models of computations*. Neural network models are based on a theory of connectionist learning network developed out of a motivation to study the neuro-physiological functions of a human brain [Rumelhart *et al.,* 1988]. The reason for choosing neural networks is simple. Their successful application to difficult problems were well documented in the 1980s [Elman and Zipser, 1987; Sejnowski and Rosenberg, 1987]. Neural network models research in the 1990s improved generalization for forecasting [Sarle, 1995]. Flexibility and generalization are viewed as the two most powerful aspects of neural network modeling [Wieland and Leighton, 1988]. Neural network can become a causal forecasting model for Internet growth with additional meaningful input attributes other than time.

In spite of their promise, neural network models do not always generalize for many applications when used for prediction in extrapolation [Roy and Mukhopadhyay, 1997]. Connectionist search techniques may find a local minimum instead of the global one without proper network structure [Lippmann, 1987]. NN models must achieve at least the same degree of accuracy as the diffusion models to be an alternative. The challenge for this research is to show that the connectionist approach *is* competitive when modeling the growth of the Internet.

CHOICE OF DIFFUSION MODELS

In choosing a set of diffusion models for this research, we looked at similar previous studies [Young, 1993; Rai et al., 1998]. One extensive study applied nine different growth curve models to various time-series data sets to determine which models achieved the best forecasts for differing types of growth data [Young, 1993]. The study showed that the Harvey model works the best with the longer data sets (more than 15 observations). We chose the Harvey model since we have more than 15 observations in our data sets. In addition, many similar studies used two growth models, *Logistic* and *Gompertz* [Young, 1993; Rai et al., 1998; Meade and Islam, 1998; Samaddar et al., 2002]. Exponential models do not usually work on diffusion data [Samaddar et al., 2002]. Exponential models are preferable to Logistic and Gompertz models during the early stage of the Internet growth [Rai et al., 1998]. However, our preliminary analysis shows that contrary to the findings, exponential model does not perform well on the Internet growth data. Our initial analysis is also in line with a recent similar study [Samaddar et al., 2002]. We did not consider exponential models since we wanted to select the best diffusion models to compare with the neural network model. Research in over 200 studies demonstrates that combining forecasts produces consistent but modest gains in accuracy [Armstrong, 1989; Meade and Islam, 1998]. We chose to combine two competing models which have relatively good performances and different forecast directions, high and low, in calibration samples.

We give below the equation forms of five models (three diffusion models, one combined and one neural network) used in this research. For all the models below, Y_t is the cumulative number of existing adopters of a given innovation at a time period *t* = T.

Gompertz Model

In Gompertz models the rate of diffusion is a function of existing adopters and the difference between the logarithms of the number of adopters at the saturation level and the existing number of adopters. This relation leads to the following integral form [Gurbaxani, 1990]:

$$Y_t = KA^M \qquad (1)$$

where, M is equal to B^t. For $0<A<1$ and $0<B<1$, Y_t is an increasing S-curve which reaches the saturation point of K (total number of potential adopters of the innovation) as time *t* approaches infinity. Diffusion growth rate is the highest at inflection point after which the growth rate starts to decrease. Inflection point is at $Y_t = K/e$ where e is Euler's constant (approximately 2.7027). Y_t reaches 37% of its saturation level at the inflection point. We estimated parameters K, A, and B from calibration sample using non-linear least squares.

Logistic Model

Logistic models do not use the logarithmic form of the number of adopters in determining the rate of diffusion [Gurbaxani, 1990]:

$$Y_t = 1/(K + A^*M) \qquad (2)$$

For $A>0$ and $0<B<1$, Y_t is an increasing S-curve which reaches the upper bound or the saturation point of 1/K as time *t* approaches infinity. Inflection point occurs when Y_t reaches 50% of its saturation level at $Y_t = K/2$. We estimated parameters K, A, and B from calibration sample using non-linear least squares

Harvey Model

The Harvey model is a rate-of-change ($y_t = dY_t/dt$) model which allows time *t* as an independent variable. The functional form is as follows [Harvey, 1984; Young, 1993]:

$$\log y_t = b_o + b_1{*}t + b_2{*}\log Y_{t-1} + \epsilon_t \qquad (3)$$

The predictive estimate of Y_t is:

$$\hat{Y}_t = Y_{t-1} + \exp(\log \hat{y}_t) \qquad (4)$$

MLP Model

MLP offers two major advantages over diffusion models. First, MLP is flexible in looking for nonlinear patterns in data. Second, MLP does not require *a priori* knowledge of relationships and distributional assumptions about the data. A previous study used MLP forecasting models for time-series data [Heravi, Osborn and Birchenhall, 2004].

MLP network consists of a layer of input nodes, one or more layers of hidden nodes, and a layer of output nodes. First hidden layer nodes connect with input layer nodes. Second hidden layer nodes connect with the first hidden layer. Output layer nodes connect with the last hidden layer nodes. Connection strengths, called *weights*, are connection values. The output of each node in an MLP, called *activation value*, is a function of its inputs from previous layer and the corresponding weights. Activation value of an input layer node is the value of the input variable. Activation value of the output layer unit is the estimated value of the dependent variable (target). A training algorithm learns the mathematical relationship between input variables and the target by assigning proper weights to all network connections.

BP Training Algorithm

We used an MLP model trained by back-propagation (BP) algorithm [Rumelhart *et al.,* 1988]. BP training algorithm estimates a target value from input variable values of the first sample point by assigning initially a set of arbitrary weights to all network connections. The method compares actual target value with the estimated value. Error signal is the difference between the actual value and the estimated value. The training process changes all weights in proportion to the error signal. Learning rate is the constant of proportionality. The method produces no error signal if there is no difference between the actual and the estimated value. The training method starts changing weights from the top layer connections. The process of updating weights propagates back through the network from top layer to the first layer connections. The larger the learning rate the larger is the weight change. The process of updating weights repeats over all sample points to complete a full iteration. After an iteration, the method computes *summed squared error* value over all sample points. Training stops when the *summed squared error* value is less than a low predefined value.

The nonlinear regression equation form of one hidden layered MLP is as follows:

$$\log \hat{y}_{t+h} = \hat{\beta}_{\phi,h} + \sum_{j=1}^{n} \hat{\beta}_{j,h} \, f(\mathbf{I}_t, \hat{w}_{h,j}) \tag{5}$$

where *h* is forecast horizon. $\mathbf{I}_t$ is input vector of current time period value and logarithm of lagged value of $\hat{Y}_{t+h}$. $\hat{W}_{h,j}$ is the network weight vector corresponding to forecast horizon *h* and *jth* hidden node. We used the logistic form of activation function *f* at each node:

$$f(\mathbf{I}_t, \hat{w}_{h,j}) = (1+e^{-z})^{-1} \tag{6}$$

where

$$z = \hat{w}_{h,j,\phi} + \hat{w}_{h,j} * t + \sum_{i=1}^{l} (\hat{w}_{h,j,i} * \log y_{t+h-i}) \tag{7}$$

and *n* is the number of hidden nodes. Logistic activation functions (equations 6 and 7) introduce nonlinearity in the model. The number of lagged time periods of Y_t is *l*. We used *l = 1* for all *n* logistic functions. Activation functions have to be differentiable for BP training algorithm. We used differentiable sigmoid function (equations 6 and 7) to compute activation values of hidden and output layer nodes.

MLP Network Architecture and Parameter Values

We followed the guidelines proposed by a recent study on architecture selection of MLP [Xiang et al., 2005]. The study suggests that one should first try with a three-layered MLP. The number of hidden units should match the minimum number of line segments (hyper planes in high dimensional cases) required to approximate the target function (similar to an *S-curve* in this case) for a minimal architecture. Functions learned by a minimal net over calibration sample points work well on new samples. We used three layers of network: one input layer for input variables (time *t* and logarithm of Y_{t-1} or *loglag*), one hidden unit layer, and one output layer of one unit (logarithm of Y_t value or *loghost*). We chose three hidden units (n = 3) as it is the minimum number required to approximate an *S-curve*. The network connects all hidden nodes with all input nodes. The output node connects to all hidden nodes. Learning is rapid with high values of learning rate. However, the learning process can jump back and forth in the error surface if the learning rate is too high. This phenomenon is called *oscillation*. One way to increase the learning rate without leading to oscillation is to include a *momentum* factor in the weight change formula. We used 0.1 for learning rate and 0.9 for momentum factor as recommended by a previous research [Rumelhart, et al., 1988].

Combined Forecast

We combined forecasts from two methods by minimizing error variance of the combined forecast [Granger, 1980; Stock and Watson, 2004]. The weight on each method (W_i) is as follows:

$$W_i = (1/ MSE_i) / \sum_{\forall i} (1/ MSE_i) \tag{8}$$

where MSE_i is the calibration *mean-squared-error* of forecasts from method *i*.

III. RESEARCH METHOD

DATA

We used host counts as a measure of the Internet size consistent with a previous study [Rai et al., 1998]. A similar study used number of Bitnet nodes to model growth pattern of computing networks [Gurbaxani, 1990]. We collected Internet usage data from Internet Engineering Task Force (IETF) reports (ftp://ftp.nw.com/pub/zone/). Table 1 below shows the Internet host count data.

METHOD

Information systems (IS) researchers often used the diffusion models to explain growths of various innovative processes. However, forecasting studies using diffusion models are rare. In this research we studied Internet growth forecasts from diffusion models. This study is the first to calibrate and validate MLP models to forecast Internet growth.

Table 1. Internet Host Count Data. Source: IETF Reports (1981-2004)

Quarter	Time Period	Number of hosts	Quarter	Time Period	Number of hosts	Quarter	Time Period	Number of hosts
Jan-82	1	225	Jul-88	27	33,000	Jan-95	53	4,852,000
Apr-82	2	233	Oct-88	28	56,000	Apr-95	54	5,747,000
Jul-82	3	279	Jan-89	29	80,000	Jul-95	55	6,642,000
Oct-82	4	344	Apr-89	30	105,000	Oct-95	56	8,057,000
Jan-83	5	409	Jul-89	31	130,000	Jan-96	57	9,472,000
Apr-83	6	475	Oct-89	32	159,000	Apr-96	58	11,176,500
Jul-83	7	540	Jan-90	33	197,500	Jul-96	59	12,881,000
Oct-83	8	628	Apr-90	34	236,000	Oct-96	60	14,513,500
Jan-84	9	727	Jul-90	35	274,500	Jan-97	61	16,146,000
Apr-84	10	826	Oct-90	36	313,000	Apr-97	62	17,843,000
Jul-84	11	925	Jan-91	37	376,000	Jul-97	63	19,540,000
Oct-84	12	1,024	Apr-91	38	455,500	Jan-98	65	29,670,000
Jan-85	13	1,258	Jul-91	39	535,000	Jul-98	67	36,739,000
Apr-85	14	1,493	Oct-91	40	617,000	Jan-99	69	43,230,000
Jul-85	15	1,727	Jan-92	41	727,000	Jul-99	71	56,218,000
Oct-85	16	1,961	Apr-92	42	890,000	Jan-00	73	72,398,092
Jan-86	17	2,221	Jul-92	43	992,000	Jul-00	75	93,047,785
Apr-86	18	2,926	Oct-92	44	1,136,000	Jan-01	77	109,574,429
Jul-86	19	3,853	Jan-93	45	1,313,000	Jul-01	79	125,888,197
Oct-86	20	4,780	Apr-93	46	1,486,000	Jan-02	81	147,344,723
Jan-87	21	8,641	Jul-93	47	1,776,000	Jul-02	83	162,128,493
Apr-87	22	13,968	Oct-93	48	2,056,000	Jan-03	85	171,638,297
Jul-87	23	19,295	Jan-94	49	2,217,000	Jan-04	89	233,101,481
Oct-87	24	24,622	Apr-94	50	2,757,948	Jul-04	91	285,139,107
Jan-88	25	28,863	Jul-94	51	3,212,000	Jan-05	93	317,646,084
Apr-88	26	30,932	Oct-94	52	3,864,000			

Model Calibration and Validation

We performed this research in three steps. In step 1 we calibrated two diffusion models and MLP on the same data used by a similar study [Rai et al., 1998]. In step 2, we created 36 new calibration samples from the same data to do a more robust analysis with rolling forecasts. Finally, in step 3 we used more Internet growth data to test whether the best models from steps 1 and 2 can learn a sudden jump (like an external influence) in Internet host counts We performed, therefore, three different analyses by breaking down the data (January 1982 – January 2005) into several pieces. In step 1 we assumed that external factors did not have much influence on Internet growth. We compared MLP and diffusion model forecasts generated at a point in time from one calibration sample. In step 2 we performed a robust rolling forecast accuracy analysis on 36 calibration samples to compare MLP and diffusion models. In step 3 we studied model responses to a sudden jump in host counts in calibration sample. We treated the jump in host counts as an external factor.

Step 1: Forecasts at a Point in Time

We calibrated two diffusion models, *Logistic* and *Gompertz*, and MLP model on January 1982 through January 1994 data. We combined two methods, Logistic and MLP, by assigning complementary weights (equation 8) to each method because the two methods had different forecast biases (high and low forecasts). We, therefore, generated forecasts for all test sample points at one fixed point in time (January 1994).

Step 2: Rolling Forecasts

We chose three years of host count data (January 1994 through October 1996) in 3 months (1 quarter) interval as our test sample. We generated 1-*quarter forecast* for each test sample point from four models: Logistic, Gompertz, Harvey, and MLP. We, therefore, created 12 calibration samples for 12 test data points for 1-quarter rolling forecasts. For example, models calibrated on January 1982 through October 1993 data produced one step ahead 1-quarter forecasts for January 1994 actual host count. Similarly, 1-quarter forecasts for April 1994 actual host count came from models built on January 1982 through January 1994 data. We repeated the process for 1-year and 3-year forecasts. We generated rolling forecasts at different points (successive) in time to make a robust comparison of methods.

Step 3: Internet Growth Data with Pseudo-external Influence

We used January 1982 through July 1999 data for calibration and January 2000 through January 2005 data for test. IETF modified the process for estimating the host counts during 1998. As a result, there was a one time upward shift in estimated Internet usage numbers. We treated the jump in estimated host counts as a *pseudo external influence*. For example, a global policy change favorable to Internet adoption will cause a similar upward shift in host counts. We chose the two best methods from step 2 analysis. Step 3 analysis answers the research question: *Which of the two models respond to a sudden jump in Internet host counts better*?

IS researchers need to validate their research instruments thoroughly [Straub, 1989]. However, the validation process is different for forecasting instrument developed from historical data than for instruments calibrated from primary data. Models calibrated on historical data must perform well on new samples before implementation. However, a model with good fit statistics does not always perform well on new data. Models may remember each sample point location to minimize the calibration error during training. However, the location specific memory fails when the locations of sample points change in new samples. Memorization occurs when MLP networks remember the locations of calibration sample points. An over-sized MLP network over-fits data causing memorization. We chose a simple network and a robust validation method of rolling forecast analysis to avoid reporting results from memorization.

Forecast Error Measures

A previous study [Armstrong and Collopy, 1992] evaluated measures for making comparisons of errors across 90 annual and 101 quarterly time-series data. The study recommended median absolute percent error (MdAPE) statistic to select the most accurate methods when many time-series data are available. Researchers should not choose mean absolute percent error (MAPE) when they expect large errors because low forecasts usually produce lower MAPEs. The study also concluded that root mean square error (RMSE) is not reliable. However, most practitioners prefer RMSE to all other error measures since it describes the magnitude of the errors in terms useful to decision makers [Carbone and Armstrong, 1982]. We report both RMSE and MdAPE for all test samples. We considered MdAPE as the criterion for choosing the best forecasting model. The error statistics are as follows:

$$\text{MdAPE} = \text{Median value of } (\text{ABS}((F - A) / A) \qquad (9)$$

$$\text{RMSE} = ((\sum_{\forall i} (F - A)^2) / N)^{0.5} \qquad (10)$$

where F and A are the forecast and the actual for observation *i* respectively.

IV. RESULTS AND ANALYSIS

STEP 1 RESULTS

Table 2 shows model estimates and performance measures of Logistic, Gompertz and MLP models on calibration sample (January 1982 – January 1994). All models have R^2 value greater than 0.99. All estimates from diffusion models are significant at $p < 0.01$. Gompertz model has the smallest average error (-2736). MLP is the best model (MdAPE =3.96). The best diffusion model is Gompertz. Logistic is the only model to forecast high. We, therefore, combined MLP forecasts with Logistic forecasts to generate a set of combined forecasts for test sample points. The weights were 40% and 60% on Logistic and MLP model respectively. MLP network used two inputs: *t* and *loglag*. T_hu1 in table 2 is the weight from input variable node *t* to the hidden unit 1. Loglag_hu1 is the weight from input variable node *loglag* to hidden unit 1. Hu1_loghost is the weight from hidden unit 1 to output node. MLP model has a minimal network since all weights from input layer nodes to the hidden layer nodes are significantly large.

Table 2. Estimates and Performance of Models on Calibration Sample

Model	Parameter	Estimate	Mean Error	MdAPE
Logistic $R^2 > 0.99$	K	0.0000001654	3279	63.83
	A	0.002202671		
	B	0.832704965		
Gompertz $R^2 > 0.99$	K	257740000	-2736	10.78
	A	0.0000000269		
	B	0.9738		
Neural Pseudo $R^2 > 0.99$	T_hu1	0.376	-2856	3.96
	Loglag_hu1	-2.054		
	T_hu2	1.149		
	Loglag_hu2	0.913		
	T_hu3	1.428		
	Loglag_hu3	-0.242		
	Bias_hu1	-2.305		
	Bias_hu2	-0.557		
	Bias_hu3	-2.339		
	Hu1_loghost	-6.503		
	Hu2_loghost	4.088		
	Hu3_loghost	4.747		
	Bias_loghost	8.818		

Table 3 reports the model performances on test sample (April 1994 through July 1997). MLP model RMSE (4,677,350) and MdAPE (35.81) are the lowest. MLP model is, therefore, a promising alternative to diffusion models for Internet growth prediction.

Table 3. Forecasts and Performance of Models on Test Samples

Quarter	Actual	MLP Forecast	Gompertz Forecast	Logistic Forecast	Combined Forecast
4/1/1994	2,757,948	2,554,766	2,535,557	2,509,346	2,536,846
7/1/1994	3,212,000	2,905,192	2,861,934	2,781,510	2,856,394
10/1/1994	3,864,000	3,296,163	3,220,091	3,057,665	3,202,065
1/1/1995	4,852,000	3,727,983	3,611,894	3,333,232	3,572,237
4/1/1995	5,747,000	4,198,146	4,039,199	3,603,675	3,963,602
7/1/1995	6,642,000	4,722,421	4,503,844	3,864,787	4,384,048
10/1/1995	8,057,000	5,303,088	5,007,632	4,112,942	4,833,525
1/1/1996	9,472,000	5,925,025	5,552,326	4,345,273	5,301,746
4/1/1996	11,176,496	6,609,193	6,139,637	4,559,753	5,800,602
7/1/1996	12,881,000	7,347,679	6,771,210	4,755,201	6,324,836
10/1/1996	14,513,496	8,152,155	7,448,620	4,931,209	6,881,355
1/1/1997	16,146,000	9,025,220	8,173,356	5,088,030	7,471,831
4/1/1997	17,842,992	9,963,361	8,946,815	5,226,434	8,094,442
7/1/1997	19,540,000	10,962,256	9,770,296	5,347,562	8,747,021
RMSE		4,677,350	5,244,312	7,279,429	5,702,186
MdAPE		35.81%	39.61%	51.54%	42.02%

Figure 1 shows plot of actual host counts versus model predictions.

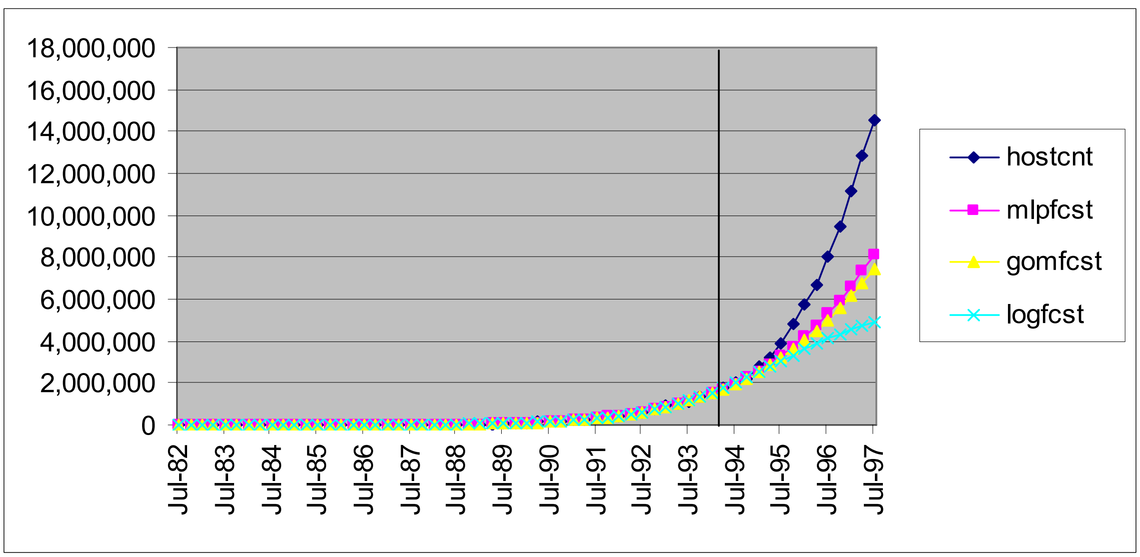

Figure 1. Actual versus Model Predictions for Dataset 1

The vertical line on the plot separates calibration and test samples. All models performed well on calibration sample. However, MLP is the best on test samples. Logistic forecasts dip on test sample after a while indicating that maximum penetration occurred around 1996. MLP and Gompertz models followed the trend in test data to some extent.

STEP 2 RESULTS

We give below results of 1-quarter, 1-year and 3-year intervals from four models: Logistic, Gompertz, Harvey and MLP.

1-Quarter Rolling Forecasts

Table 4 shows 1-quarter rolling forecasts and performance on 12 test sample points. MLP performance is the best (MdAPE and RMSE are 3.6% and 362,213 respectively). Logistic model is close second. Performances of Gompertz and Harvey models are poor relative to the top two models.

Table 4. 1-Quarter Rolling Forecasts

Quarter	Actual	Logistic Forecast	Gompertz Forecast	MLP Forecast	Harvey Forecast
1/1/1994	2,217,000	2,300,970	575,217	2,287,088	4,636,524
4/1/1994	2,757,948	2,509,348	643,421	2,505,494	4,968,228
7/1/1994	3,212,000	3,042,040	727,891	2,958,388	6,180,200
10/1/1994	3,864,000	3,637,498	823,623	3,410,282	7,177,160
1/1/1995	4,852,000	4,455,532	936,705	4,069,468	8,623,536
4/1/1995	5,747,000	5,791,040	1,077,326	5,529,016	10,835,984
7/1/1995	6,642,000	7,086,384	1,239,621	6,419,752	12,815,456
10/1/1995	8,057,000	8,190,660	1,421,521	7,489,016	14,776,544
1/1/1996	9,472,000	9,845,176	1,638,019	9,380,304	17,917,168
4/1/1996	11,176,496	11,552,336	1,885,997	10,796,328	21,032,112
7/1/1996	12,881,000	13,510,656	2,171,512	12,977,296	24,784,736
10/1/1996	14,513,496	15,393,808	2,491,336	14,309,840	28,503,792
RMSE		405,734	6,707,223	362,213	7,415,042
MdAPE		5.09%	81.30%	3.60%	88.85%

1-Year Rolling Forecasts

Table 5 shows 1-year rolling forecasts and error statistics on the same 12 test sample points. MLP is the best model (MdAPE and RMSE are 18.3% and 1,614,826 respectively). Logistic is close second. Gompertz and Harvey models did not perform well.

Table 5. 1-year Rolling Forecasts

Quarter	Actual	Logistic Forecast	Gompertz Forecast	MLP Forecast	Harvey Forecast
1/1/1994	2,217,000	1,971,692	402,336	2,159,898	5,334,860
4/1/1994	2,757,948	2,183,070	453,153	2,270,008	6,032,988
7/1/1994	3,212,000	2,686,982	512,636	2,641,000	7,049,632
10/1/1994	3,864,000	3,239,458	579,342	3,339,678	8,056,496
1/1/1995	4,852,000	3,333,238	647,573	3,345,234	8,689,472
4/1/1995	5,747,000	4,335,748	732,117	3,973,472	10,388,216
7/1/1995	6,642,000	5,456,448	827,925	4,530,784	11,866,992
10/1/1995	8,057,000	7,276,328	941,107	6,171,744	13,936,912
1/1/1996	9,472,000	11,625,160	1,081,882	7,689,776	17,134,464
4/1/1996	11,176,496	15,048,872	1,244,338	8,537,688	20,131,104
7/1/1996	12,881,000	15,795,736	1,426,388	11,312,216	23,187,744
10/1/1996	14,513,496	19,170,368	1,643,066	12,410,696	27,790,624
RMSE		2,182,948	7,187,050	1,614,826	6,905,926
MdAPE		21.74%	87.40%	18.30%	80.83%

3-Year Rolling Forecasts

Model *MdAPEs* are 56.01, 66.31, 95.15 and 340.10 for MLP, Logistic, Gompertz and Harvey respectively. RMSE numbers are very high (8,137,181 approximately) for all models. All methods start to lose accuracy fast if forecasts are extrapolated too far in future like 3 years. MLP model performed better than the other methods for 3-year forecasts also. Logistic is close second. We, therefore, chose MLP and Logistic for step 3.

Figures 2 and 3 show rolling forecast plots of MLP model and Logistic model respectively. MLP forecasts are smoother than Logistic forecasts across all the forecast horizons. Both models consistently forecast reasonably well for the near future.

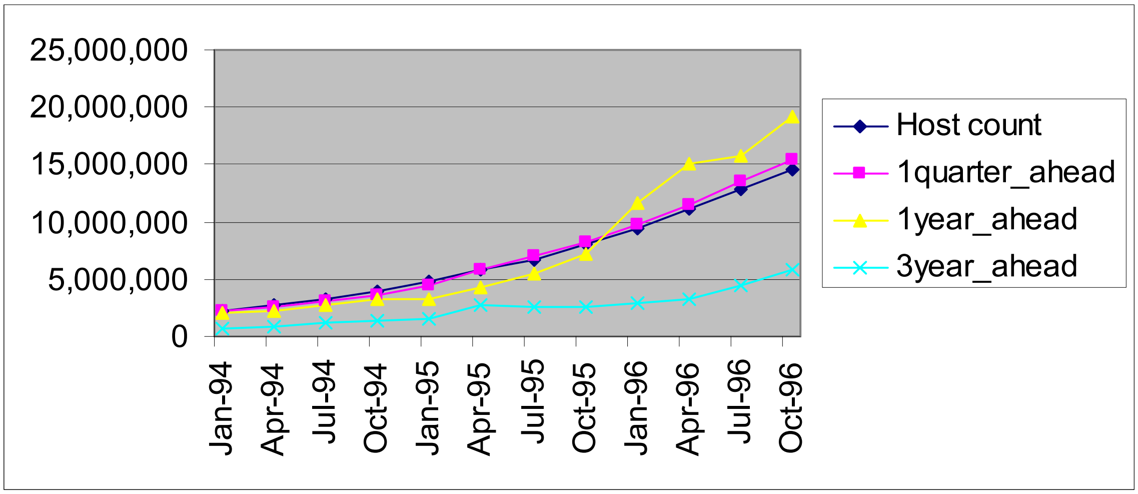

Figure 2. Actual versus Rolling Forecasts of Logistic Model

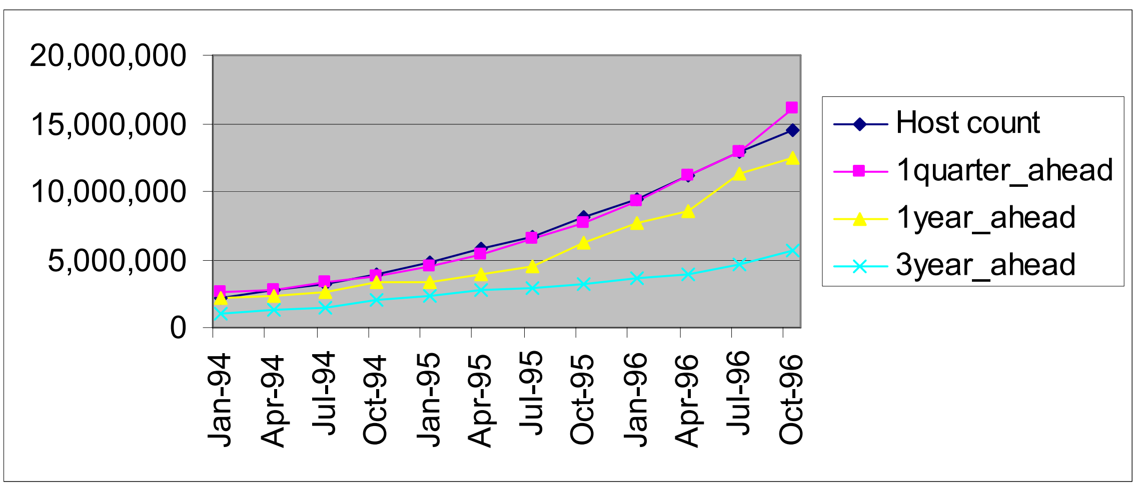

Figure 3. Actual versus Rolling Forecasts of MLP Model

STEP 3 RESULTS

Table 6 reports model estimates and error statistics on calibration sample (January 1982 – July 1999).

Table 6. Estimates and Performance of Models on New Calibration Samples

Model	Parameter	Estimate	Mean Error	MdAPE
Logistic R2 > 0.99	K	0.000000007	-22,890	11.24
	A	0.000950012		
	B	0.852096591		
	B	0.2535		
Neural Pseudo R2 > 0.99	T_hu1	2.472	46,943	4.55
	Loglag_hu1	0.511		
	T_hu2	-1.087		
	Loglag_hu2	3.540		
	T_hu3	-.907		
	Loglag_hu3	-0.823		
	Bias_hu1	1.289		
	Bias_hu2	3.881		
	Bias_hu3	1.938		
	Hu1_loghost	4.015		
	Hu2_loghost	5.098		
	Hu3_loghost	-7.657		
	Bias_loghost	10.694		

MLP model performance again is the best (*MdAPE* = 4.55%). Table 7 reports corresponding forecasts and error statistics on test samples (January 2000 – January 2005).

Table 7. Model Forecasts and Error Statistics on Test Sample

Quarter	Actual	MLP Forecast	Logistic Forecast
1/1/2000	72,398,080	67,597,383	66,698,228
7/1/2000	93,047,744	82,031,168	78,126,113
1/1/2001	109,574,400	97,904,271	89,226,061
7/1/2001	125,888,192	112,730,115	99,489,150
1/1/2002	147,344,640	127,631,473	108,555,121
7/1/2002	162,128,384	143,549,087	116,246,348
1/1/2003	171,638,272	158,030,793	122,550,673
1/1/2004	233,101,440	183,330,365	131,487,415
7/1/2004	285,138,944	203,556,570	134,482,623
1/1/2005	317,646,080	220,351,043	136,744,295
RMSE		44,689,897	85,518,232
MdAPE		11.65%	27.31%

MLP model adapted to the sudden jump in calibration sample better than Logistic model. Figure 4 shows the plot of model predictions versus actual host count on test sample.

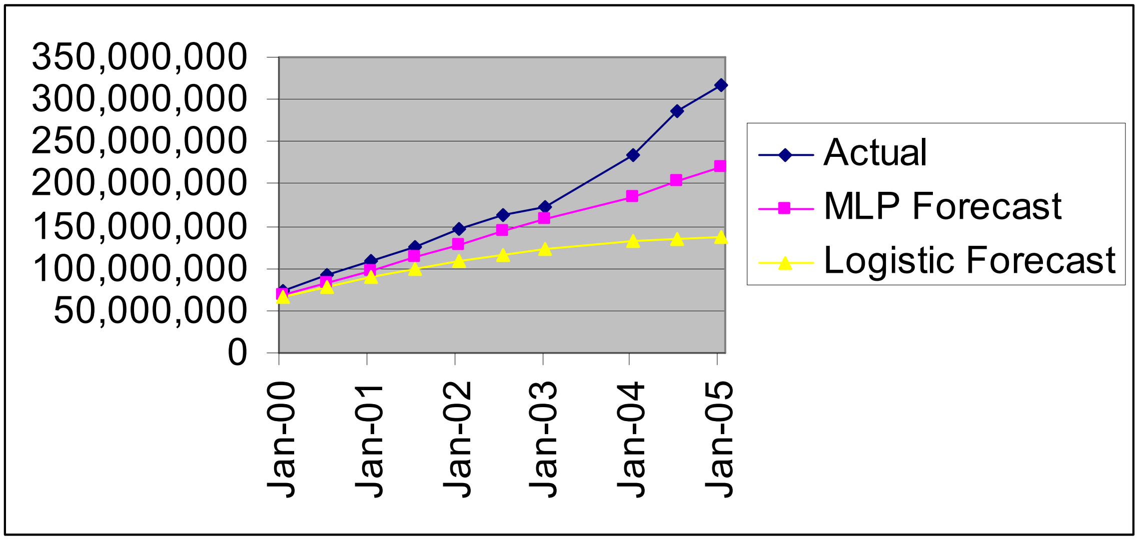

Figure 4. Actual versus Model Predictions for Step 3 Test Samples

Logistic model predicts full market penetration around 2004. Results are consistent with the theory that diffusion models sometimes underestimate actual growth process [Van den Bulte and Lilien, 1997].

V. DISCUSSION AND CONCLUSION

MLP, the proposed new alternative approach, consistently outperformed diffusion models on all test samples with one network structure. This study shows that MLP is a better choice than diffusion models in forecasting Internet growth. Diffusion models have several limitations as a forecasting tool: instability with limited available data, environmental differences, and systematic underreporting of estimated time to attain total number of first purchase sales [Heeler and Hustad, 1980]. Estimation of unknown ceilings of total number of adopters is often closer to the number of adopters in the last observation period than it is to reality [Van den Bulte and Lilien, 1997]. Flawed estimates are problematic to the users of diffusion models, including market forecasters and strategic market planners. Diffusion models are inflexible because the models attempt to fit a fixed s-shaped function by adjusting the values of the shape parameters. Our results show that MLP models do not have the above limitations when forecasting Internet growth.

Figure 5 shows the prediction surface of the MLP model for step 1 data. Prediction surface shows the geometry of complex nonlinear mapping from input variables to target. Mapping function estimates values of target variable *loghost* for each sample point. Diffusion models unlike MLP cannot account for factors other than time, which might influence Internet growth. For example, for time period $t = 50$ in figure 5, diffusion models forecast only one value from individual functions (equations 1 and 2). However, MLP forecasts a range of values (between 17 through 36 in figure 5) by accounting for additional influence on target from second factor *loglag*. MLP, therefore, will often find mapping functions closer to optimal mapping functions than diffusion models.

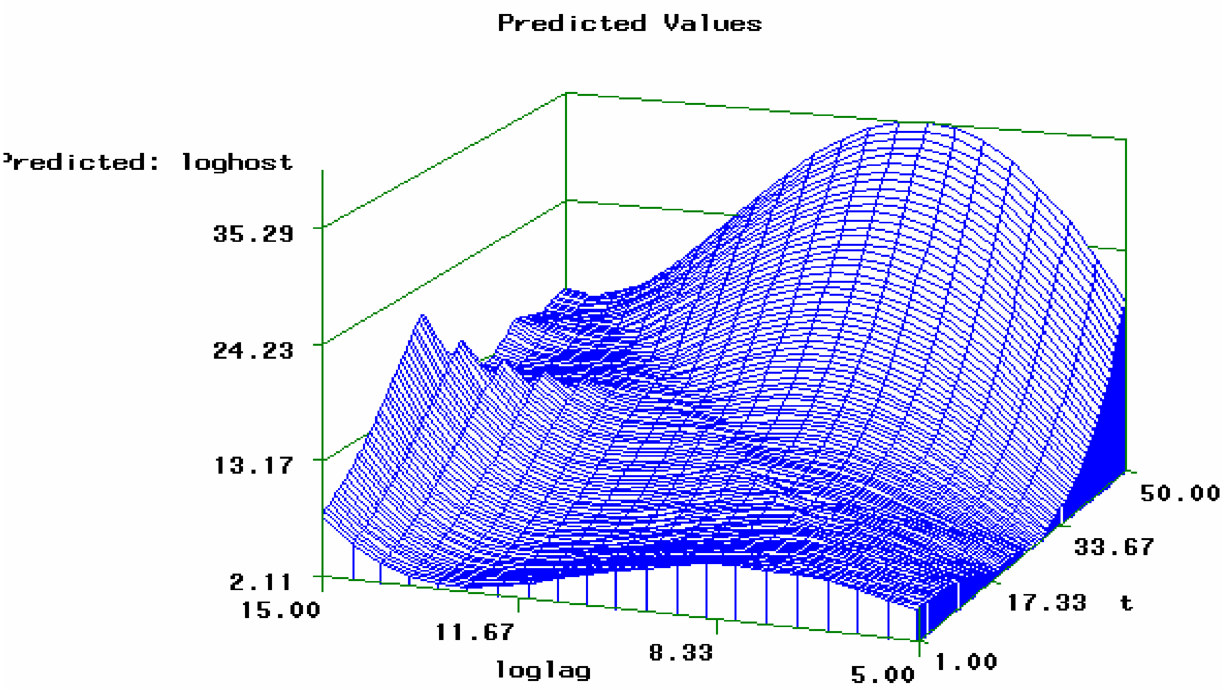

Figure 5. Estimated Prediction Surface of Neural Network Model

Corresponding contour plot (Figure 6), a two dimensional flat projection of prediction surface, shows the rich decision boundaries from MLP model.

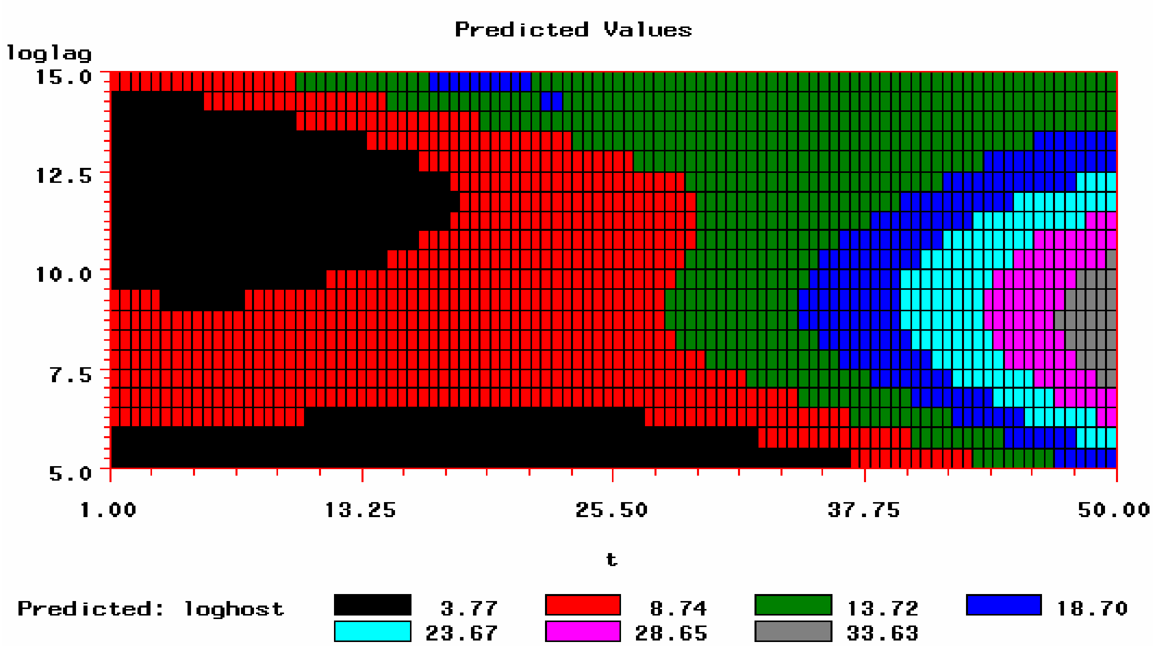

Figure 6. Contour Plot of Neural Network Model Prediction

Rich contours and surfaces of MLP models indicate that the models can find the best relationship between the input and the output [Weiland and Leighton, 1988]. However, oversized MLP networks can over-fit data. Robust validation results of our research confirm that we did not over-fit the models. MLP models are adaptable to changes in environment [DeLurgio and Bhame, 1997]. Step 3 results show that MLP adapted well to a sudden jump in host counts. This research however, confirms the findings from previous studies that diffusion models perform well on calibration data [Gurbaxani, 1990; Rai et al., 1998; Mahajan et al., 1998].

Managers and policy makers like to see a model which forecasts reasonably well at least in the near future. The findings of this research will be useful to them. The results of this research will encourage IS researchers and practitioners to investigate connectionist models to predict diffusion of other innovation processes. IS researchers can combine other artificial intelligence tools with MLP to build a more powerful hybrid forecasting system.

Like many innovations Internet is a global phenomenon. How a group of users adopt Internet depends heavily on local, technological, economic, political, and social conditions [Wolcott and Goodman, 2003]. Future research can include attributes related to the above factors in growth models. MLP is the most convenient choice to accommodate additional attributes because of its flexible architecture.

Editor's Note: This article was received on May 15, 2005. It was with the author for two revisions for a total of three months and was published on January 25, 2006.

EDITOR'S NOTE: The following reference list contains the address of World Wide Web pages. Readers, who have the ability to access the Web directly from their computer or are reading the paper on the Web, can gain direct access to these references. Readers are warned, however, that

1. these links existed as of the date of publication but are not guaranteed to be working thereafter.

2. the contents of Web pages may change over time. Where version information is provided in the References, different versions may not contain the information or the conclusions referenced.

3. the authors of the Web pages, not CAIS, are responsible for the accuracy of their content.

4. the author of this article, not CAIS, is responsible for the accuracy of the URL and version information.

REFERENCES

IETF Report (1981-2004): Retrieved last on December, 2004 from ftp://ftp.nw.com/pub/zone

Armstrong, J. S. (1989) "Combining Forecasts: The End of the Beginning and the Beginning of the End?", *International Journal of Forecasting*, 5, pp. 585-588.

Armstrong, J. S. and Collopy, F. (1992) "Error Measures for Generalizing about Forecasting Methods: Empirical Comparisons", *International Journal of Forecasting*, 8, pp. 69-80.

Bass, F. M. (1969) "A new Product Growth Model for Consumer Durables", *Management Science,* (15)5, pp. 215-227.

Carbone, R. and Armstrong, J. S. (1982) "Evaluation of Extrapolative Forecasting Methods: Results of a Survey of Academicians and Practitioners", *Journal of Forecasting*, 1, pp. 215-217.

Dekimpe, M.G., Parker, P.M. and Sarvary, M. (1998) "Staged estimation of international diffusion models - An application to global cellular telephone adoption," *Technological Forecasting and Social Change*, (57, 1-2), pp.105-32.

DeLurgio, S. A. and Bhame, C. D. (1997) "Neural Network and AI Forecasting Successes", *International Conference Proceedings, APICS*, pp. 496-500.

Elman, J.L. and Zipser, D. (1987) "Learning the Hidden Structure of Speech", *UC San Diego Institute of Cognitive Science Report 8701.*

Fourt, L. A. and Woodlock, J. W. (1960) "Early Prediction of Market Success for New Grocery Products", *Journal of Marketing (*25), pp. 31-38.

Geroski, P.A. (2000) "Models of technology diffusion," *Research Policy*, (29, 4-5), pp.603-25.

Granger, C. W. J. (1980) "Forecasting in Business and Economics", *Academic Press*, *New York*.

Gurbaxani, V. (1990) "Diffusion of Computing Networks: The Case of Bitnet", *Communications of the ACM, (33)*12, pp. 65-75.

Harvey, A. C. (1984) "Time Series Forecasting Based on the Logistic Curve", *Journal of the Operational Research Society* (35), pp. 641-644.

Heeler, R. M. and Hustad, T. P. (1980) "Problems in Predicting New Product Growth for Consumer Durables", *Management Science*, (26)10, pp. 1007-1020.

Heravi, S, Osborn, D. R. and Birchenhall, C. R. (2004) "Linear versus Neural Network Forecasts for European Industrial Production Series", *International Journal of Forecasting* (20), pp. 435-446.

Lippmann, R. P. (1987) "An Introduction to Computing with Neural Nets", *IEEE ASSP Magazine*, April 1987, pp. 4-17.

Mahajan, V. and Muller, E. (1979) "Innovation Diffusion and New Product Growth Models in Marketing", *Journal of Marketing*, (43)4, pp. 55-68.

Mahajan, V., Muller, E. and Bass, F. (1990) "New product Diffusion Models in Marketing: A Review and Directions for Research", *Journal of Marketing, (54)*, pp. 1-26.

Mahajan, V., Sharma, S. and Bettis, R. (1998) "The Adoption of M-Form Organizational Structure: A Test of Imitation Hypothesis", *Management Science, 34*, pp. 1188-1201.

Mansfield, E. (1961) "Technical Change and the Rate of Imitation", *Econometrica*, (29)4, pp. 741-766.

Meade, N. and Islam, T. (1998) "Technological forecasting - Model selection, model stability, and combining models," *Management Science*, (44) 8, pp.1115-1130.

Press, L. (1997) "Tracking the Global Diffusion on the Internet", *Communications of the ACM, 40*(11), pp. 11-17.

Rai, A., Ravichandran, T. and Samaddar, S. (1998) "How to Anticipate the Internet's Global Diffusion", *Communications of the ACM, 41*(10), pp. 97-106.

Rogers E. M. (1983) "Diffusions of Innovations", *The Free Press, New York*.

Roy, A. and Mukhopadhyay, S. (1997). "Iterative Generation of Higher-order Nets in Polynomial Time Using Linear Programming", *IEEE Transactions on Neural Networks*, 8(2), pp. 402-412.

Rumelhart, D. E., Hinton G. E. and Williams, R. J. (1988) "Learning Internal Representations by Error Propagation" in *Parallel Distributed Processing Explorations in the Microstructure of Cognition*, V 1, MIT Press, Cambridge, pp. 328-330.

Samaddar, S., Nargundkar, S. and Chatterjee, S. (2002) "E-market Infrastructure Planning and Internet Growth", *8th American Conference on Information Systems*, pp. 721-729.

Sarle, W. S. (1995) “Stopped Training and other Remedies for Overfitting”, *Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics*.

Sejnowski, T. J. and Rosenberg, C. R. (1987) “Parallel Networks that Learn to Pronounce English Text”, *Complex Systems* (1), pp. 145-168.

Stock, J. and Watson, M. W. (2004) “Combination Forecasts of Output Growth in a Seven-Country Data Set”, *Journal of Forecasting*, 23, pp. 405-430.

Straub, D. W. (1989) “Validating Instruments in MIS Research”, *MIS Quarterly*, June pp. 147-169.

Van den Bulte, C. and Lilien, G. L. (1997) “Bias and Level Diffusion Models”, *Marketing Science*, (16)4, pp. 338-353.

Venkatraman, N., Loh, L. and Koh, J. (1994) ”The Adoption of Corporate Governance Mechanisms: A Test of Competing Diffusion Models”, *Management Science, (40)*, pp. 496-507.

Weiland A. and Leighton R. (1988) “Geometric Analysis of Neural Network Capabilities”, *Technical Report, Arpanet III,* pp*.* 385-392.

Wolcott P. and Goodman S. E. (2003) “Introducing Global Diffusion of the Internet Series”, *Communications of the Association for Information Systems* (11), pp. 555-559.

Xiang, C., Ding, S. Q. and Lee, T. H. (2005) “Geometrical Interpretation and Architecture Selection of MLP”, *IEEE Transactions on Neural Networks*, (16)1, pp. 84-96.

Young, P. (1993) "Technological Growth-Curves - a Competition of Forecasting Models”, *Technological Forecasting and Social Change*, (44, 4), pp.375-89.

ABOUT THE AUTHOR

Somnath Mukhopadhyay is Assistant Professor in Information and Decision Sciences Department at The University of Texas at El Paso. His research interest is to solve various business problems using tools like Linear Programming and Neural Networks. He published with others in journals like *IEEE Transactions on Neural Networks, Informs Journal on Computing*, *Neural Networks,* and *Neural Computations*. He has consulting experience over 15 years in quantitative modeling in various industries such as airlines, hotel, thermal engineering, and telecommunications. He was visiting research assistant in Stanford University parallel distributed processing (PDP) group.

Communications of the Association for Information Systems

ISSN: 1529-3181