A Forecasting Method for Increasing Trend of Internet Nodes

SUN Xin-Bo[1], *FAN Zhi-Ping*[2]

[1, 2]School of Business Administration, Northeastern University, Shenyang 110004, P.R.China

Summary
This paper proposes a forecasting method for increasing trend of internet nodes based on dynamic logistic equation. According to the node increasing data provided by the Cooperative Association for Internet Data Analysis (CAIDA) and China Internet information center, the parameters in the logistic equation are fitted. By using logistic forecasting method, parameters are fitted with known nodes data and the problem of parameter changing with time is solved. An experiment is conducted to illustrate that the proposed method can forecast the regularities of the increasing trend of internet nodes.

Keywords:
Internet, Logistic equation, Nodes, Parameter estimation

1. 1. Introduction

Internet has been developing very fast with the speed of twice more than that of the previous year and that is beyond the prediction of Internet engineers. This kind of increasing speed of Internet is called by the way of exponential increasing. Recently, to describe and forecast the increasing regularities of Internet nodes has received more and more attention. To study the increasing regularities of Internet, increasing of Internet nodes should be synthesized in order to forecast the increasing of the nodes and to collaborate the development of all of the factors of the system. This paper analyzes the data provided by the Skitter's project of CAIDA (The Cooperative Association for Internet Data Analysis) and CNNIC (China Internet Network Information Center), and establishes a Logistic nodes increasing model. Then the increasing trend of internet nodes is forecasted. The result of experiment shows that the proposed method can forecast the increasing trend of Internet nodes well.

2. 2. The Forecasting Model for Nodes

2.1 The Logistic Equation for Forecasting Internet Nodes

Logistic equation was proposed by Verhulst in 1837 to study the population development model under the condition of limited resources. The general form is as follows [1].

$$\frac{dN(t)}{dt} = r(1 - \frac{N(t)}{k})N(t)\,, \qquad (1)$$

where $N(0) = N_0$ and $t \geq 0$.

Because the development of internet nodes is similar with the development process of population, this paper uses the improved logistic method to forecast the increasing trend of internet nodes. Here, $N(t)$ denotes the quantity of network nodes at time t, K is the peak load of limited resources, and r is the increasing rate of internal network nodes that can be affected by the band, IP address, economy and politics etc.

Based on the data [1,2] from October 1997 to July 2002 provided by CAIDA and Chinese Internet nodes development, the nodes increasing situation is forecasted with Logistic equation by month.

By Logistic equation (1), when t =0 (October 1997), $K = 146$, $r = \ln(325/252)$, part of errors of the forecasting are as follows: 0.02% (July 1999), 0.04% (July 2000), 0.28% (July 2001), 0.22% (July 2002). It can be found that the forecasting errors begin to increase since 2000. Through analyzing, the parameters (K, r) in Logistic equation will change with the time. Then it has become the critical problem that how to forecast the increasing situation of Internet nodes more precisely, that means how to fit several groups of (K, r) better.

2.2 Integral Fitting of Parameters

To fit the parameters of Logistic equation is actually a reverse problem of differential coefficient equation that has already a lot of method such as visual measurement method, three points method, quasi linear method ,integral fitting method ,nelder meade algorithm method, recursive analysis and fitting method etc. By analysis and comparison, the integral fitting method can provide more perfect parameters fitting results for the increasing forecasting model of nodes [3, 4]. As a result, this paper adopts integral fitting method [5, 6] to forecast the increasing curve of Internet network nodes [7].

The fitting steps are as follows.

Step 1. Construct the super-coupled equation set, the process is as follows.

We transform Eq. (1) into the following form

$$\frac{dN}{dt} = x_1 N - x_2 N^2\,, \qquad (2)$$

$$x_1 = r\,, \qquad (3)$$

$$x_2 = \frac{r}{k}\,. \qquad (4)$$

We integrate Eq. (2) from t_0 to t_i, i.e.,

$$N(t_i) - N_0 = x_1 \int_{t_0}^{t_i} N dt - x_2 \int_{t_0}^{t_i} N^2 dt \cdot$$

Let $b_i = N(t_i) - N_0$, $a_{i1} = \int_{t_0}^{t_i} N dt$ and. $a_{i2} = \int_{t_0}^{t_i} N dt$, where $a_{01} = 0$ and $a_{02} = 0$. Then we can get the super-coupled equation group as follows

$$b_i = a_{i1} x_1 - a_{i2} x_2, \quad i = 0,1,2,..., m-1,$$

i.e.,

$$AX=B. \qquad (5)$$

Step 2. Calculate least-square solution of Eq. (5), i.e.,

$$X = (X_1 - X_2)^T = (A^T A)^{-1} A^T B. \qquad (6)$$

Step 3. Calculate the value of K and r by Eqs. (3), (4) and (6), i.e.,

$$r = x_1, \; K = x_1 / x_2 . \qquad (7)$$

Step 4. Calculate the fitted N_i and its forecasting value, i.e.,

$$N = \frac{K}{(1 + (K/N_0 - 1)e^{-rt})}, \qquad (8)$$

where $N(t_0) = N(0) = N_0$.

3. Experiments

CAIDA is an international cooperation organization with the study focus on the global Internet structure and data. CAIDA is located in San Diego Supercomputer Center of Californian University At present, 30 research institutions have joined into the SDSC, including the initiators CISCO SYSTEM, DARPA, ENDACE MEASUREMENT SYSTEM, SUN Micro System and other research institutions and most of the study institutes in North America, Europe and Japan. The authors of this paper are the members who study the nodes in China.

In order to demonstrate the validity of the Logistic function, this paper takes Internet data from CAIDA center and CNNIC as research object and analyzes the increasing regularities to prove the validity of the method in this paper. So the Internet data from July 1997 to July 2000 was used to fit by Logistic curve, the increasing situation of nodes was forecasted from July 2001 to July 2002 by using the fitted Logistic curve.

Firstly , construct the super-coupled equation set according to step1 :

$$\begin{bmatrix} 56.31 & -1263.84 \\ 178.10 & -2437.49 \\ 764.64 & -4379.11 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 10.03 \\ 68.11 \\ 136.56 \end{bmatrix},$$

And the least-square solution of super-coupled equation set are $x_1 = 0.076535$, $x_2 = 0.000324$. Then, by Eq. (7) we get r=0.076535 and k=236.2191358.

In the end, the improved fitted Logistic curve is as follows:

N=236.2191358/ (1+ (236.2191358/29.9-1)e$^{-0.076535t}$)

The result of the nodes increasing situation from 2001 to 2002 by the improved formula fitted is like Table 1.

Table 1：Increasing situation forecasted by improved formula

Period	Practical Nodes (10,000)	Forecasted Nodes (10,000)	Error
July 1997	29.9	29.9	0.000
August 1997	30.1	30.1	0.000
September 1997	34.8	34.8	0.000
…	…	…	…
July 1998	54.2	54.2	0.000
…	…	…	…
September 1999	146	149	0.001
…	…	…	…
July 2000	650	677	0.017
…	…	…	…
September 2000	837	970	0.036
…	…	…	…
July 2001	1002	1270	0.098
…	…	…	…
July 2002	1254	1530	0.1723

From Table 1, we can see that the fitting error of improved model is at least 5% less than the previous model, and the increasing speed of the errors of the improved model is slower. It can be shown that it is more suitable to fit the increasing of Internet nodes by the improved Logistic equation. What is more, the fluctuation of band, economy and politics can greatly affect the value of K. The comparison result is as Fig. 1.

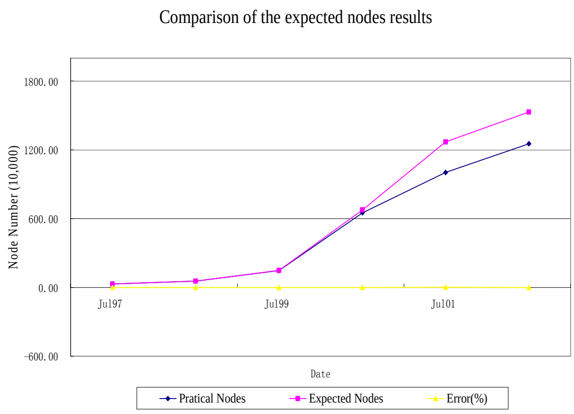

Fig. 1 Comparison of the expected node results

Fig. 1 demonstrates the fitting error made by the improved Logistic equation for the increasing of Internet nodes is much less than that made by previous Logistic equation, and this means the improved Logistic equation is a feasible way to expect the increasing situation of Internet nodes.

3. 4. Conclusions

By analyzing the data of the increasing Internet nodes from CAIDA and CNNIC, this paper constructs dynamic Logistic integral fitting model, it can not only solve the problem of precise and continuous change of the parameters fitting in Logistic node model, but also can forecast the increasing situation of Internet nodes. So the model has great value for the forecasting and control the Internet development trend.

Acknowledgements

This work was partly supported by the Teaching and Research Award Program for Outstanding Young Teachers (TRAPOYT) in Higher Education Institutions and Research Fund for the Doctoral Program of Higher Education (Project No. 20040145018), Ministry of Education, China.

The author would like to express his sincere thanks to referees for their constructive comments and suggestions.

3.1 References

[1] G.F. Gause, The struggle for existence, Balitimore, Williams and Wilkins. pp. 90-103,145, 1987.
[2] E. Simaxisheng, Peacl Curvl Parameters. Forecasting, no. 1, pp. 51-52, 1994.
[3] S.H. Zhou. Estimating the parameters of logistic equation. Application of Statistics and Management, vol. 11 no.5, pp.32-35, 1992.
[4] S.E. Yangexun. Zoology Model Principle. Shanghai: Shanghai Translation Press, 1988.
[5] Z. J. Yang, Z.L. Xu. Forecast of the population growth in the country of Heilongjiang by the forecast method of dynamic logistic. Journal of Heilongjiang Bayinongken University, vol.9, no.2, pp. 23-28, 1997.
[6] L. Jiang. Theory of System Science. Beijing: Huaxia Press, 1990.
[7] X.Y. Wu. A numerical value method for logistic curve. Journal of Biology & Mathematics, vol.5 no.1, pp.26-32, 1990.

SUN Xin-Bo received the B.S. degrees in industrial automation from Shenyang Gold College in 1996, received the M.S. degrees in Management Science and Engineering from Northeastern University (NEU) in 2003, respectively. He is currently lecturer in the Department of Management Science and Engineering, School of Business Administration, NEU. He is currently interested in information management and knowledge management.

FAN Zhi-Ping received the B.E. degree in industrial automation from Northeastern University (NEU), Shenyang, China, in 1982, the M.S. degree in industrial automation from NEU in 1986, and the Ph.D. degree in control theory and applications from NEU in 1996. He is currently Professor in the Department of Management Science and Engineering, School of Business Administration, NEU. He was Research Fellow at City University of Hong Kong in 2001, 2003, 2004 and 2005, and he has published more than 20 papers in international journals. He is currently interested in decision analysis, operations research, information management and knowledg