

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

High-Speed Railway Passenger Flow Volume Prediction Model Based on Factor-Response Simulation Method

Du Xuedong, Shang Rui and Gao Ziyu
Engineering Center, College of Information Science and Engineering,
Shandong University of Science and Technology, Qingdao 266510, China

Abstract: A new high-speed railway Passenger Flow Volume (PFV) prediction model was proposed according to the factor-response simulation method, and the percentage of the five factors' influence. The five factors were confirmed from many factors that affect PFV. It can be concluded that the proposed prediction model was precise and practical through an experimental test.

Key words: Factor-response simulation, mean absolute percentage error, consumption function, regional economic conditions

INTRODUCTION

Now-a-days, people's trip generation rate has grown rapidly along with the improvement of people's income and values (Shi *et al.*, 2007). A good prediction model of PFV is helpful for the National Railway Administration to draw up an effective railway operation plan to achieve the greatest economic benefit.

However, PFV is affected by a lot of nonlinear factors in real life which increase the complexity of PFV predictive analysis tremendously (Dearien and Plum, 1993; Wang *et al.*, 2002). Considering the importance of PFV prediction, many people try to create or use all kinds of nonlinear factors to upgrade the prediction of PFV. Traditional prediction techniques includes multiple linear regression, stochastic time series, general exponential smoothing, state space and Kalman filter (Davis and Nihan, 1991; Moghram and Rahman, 1989; Okutani and Stephanedes, 1984). Traditional prediction techniques make it hard for the researchers to test. Recently, someone try to find the similarities between some popular networks models and transportation, they come up with prediction models like Neural Tree model (NT), Artificial Neural Networks (ANN), Back-Propagation Neural Network (BPN) (The effect of tannic acid on *in vitro* gas production and rumen fermentation of sunflower meal (Dougherty, 1995; Zurada, 1995; Lan *et al.*, 2008; Lan, 2009; Zafar *et al.*, 2006) and they have acquired remarkable achievement. Especially for the popular BPN

which has been successfully used in many transportation fields. Nevertheless, these researchers all try to match dynamic and changeable PFV with fixed existing models which make it passive and hard to predict PFV precisely.

Five factors and their influence percentage on PFV are discussed in this paper and a new PFV prediction model (Averill and Kelton, 2000) is proposed, at the same time some correction is made on the proposed PFV prediction model. Through the test of the prediction model on Qingdao's PFV data, the Mean Absolute Percentage Error (MAPE) of the model is under 20% totally. Therefore, the PFV prediction model is persuasive and practical (Tsung-Hsien and Chi-Kang, 2003; Lewis, 1982; Yu *et al.*, 2008; Steven and Keh, 2011).

PFV INFLUENCE FACTOR-RESPONSE SIMULATION

Factor-response simulation method is a technique that spread out as many interrelated factors as possible and makes use of the factorial design method to get the influence percentage of every factor. Due to limited space, Table 1 only shows the selected factors and their properties that have a greater impact to PFV.

In Table 1, the factors are classified into three levels A, B and C by the deciding of whether the factor is qualitative or quantitative and controllable. The influence A, B and C reduces by turns. It is not hard to

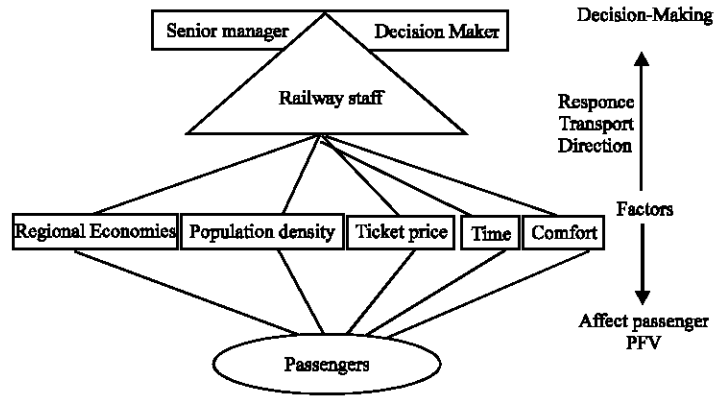


Fig. 1: Relationship between factors and PFV

Table 1: Factors that may affect the PFV

System factor	Possible factors	Quantitative?	Quantitative?	Controllable?	Uncontrollable?	Influence degree	Possible response
External factor	National railway policies		✓	✓		B	Speed raised
	Economic development level		✓		✓	B	Stabilization
	Income level	✓		✓?	✓?	B	Increase steadily
	Consumption rate	✓		✓?	✓?	A	Increase steadily
	Population density	✓			✓	A	Increase rapidly
Internal factor	Traveling fever		✓	✓?	✓?	B	Fluctuate
	Tricket price	✓		✓		A	Appropriately increase
	Time (speed)	✓		✓?		A	Speed increase
	Tricket-selling service		✓	✓		C	Quality improved
	Safety		✓	✓?	✓?	A	Higher safety factor
	Comfort		✓	✓?	✓?	B	More comfortable
	Vehicle maintenance		✓	✓		C	In time
	Traffic density	✓	✓			C	Increase

Table 2: Factors design matrix

Factor combination	F_{eco}	F_{pop}	F_{tp}	F_{tm}	F_{saf}	Response
1	-	-	-	-	-	R_1
2	+	-	-	-	-	R_2
3	-	+	-	-	-	R_3
4	+	+	-	-	-	R_4
5	-	-	+	-	-	R_5
6	+	-	+	-	-	R_6
7	-	+	+	-	-	R_7
8	+	+	+	-	-	R_8
9	-	-	-	+	-	R_9
10	+	-	-	+	-	R_{10}
11	-	+	-	+	-	R_{11}
12	+	+	-	+	-	R_{12}
13	-	-	+	+	-	R_{13}
14	+	-	+	+	-	R_{14}
15	-	+	+	+	-	R_{15}
16	+	+	+	+	-	R_{16}
17	-	-	-	-	+	R_{17}
18	+	-	-	-	+	R_{18}
19	-	+	-	-	+	R_{19}
20	+	+	-	-	+	R_{20}
21	-	-	+	-	+	R_{21}
22	+	-	+	-	+	R_{22}
23	-	+	+	-	+	R_{23}
24	+	+	+	-	+	R_{24}
25	-	-	-	+	+	R_{25}
26	+	-	-	+	+	R_{26}
27	-	+	-	+	+	R_{27}
28	+	+	-	+	+	R_{28}
29	-	-	+	+	+	R_{29}
30	+	-	+	+	+	R_{30}
31	-	+	+	+	+	R_{31}
32	-	-	-	-	+	R_{32}

see from the Table 1 that there are six main factors: Economic development level, geographic locations, population density, ticket price, time and safety. The economic development level and different geographic locations are replaced as one factor that is regional economic conditions. Therefore, the factors become regional economic conditions (F_{eco}), population density (F_{pop}), ticket price (F_{tp}), time (F_{tm}) and safety (F_{saf}). Factor-design matrix based on 2^k (k is the number of factors, here it equals to 5) will be given below to analyze the influence from factors to PFV while R_i indicates the average PFV of every month (Table 2), here + means the relevant factor has a positive affect on the PFV and-means the relevant factor has a negative affect on the PFV.

The relationships between factors and PFV can be easily seen in Fig. 1 and the influence percentages of every factor can be calculated as follows:

$$R_i = R_{i+12} = R_{i+24} = \frac{\sum_{j=1}^5 \sum_{k=1}^{12} (m_{jk})}{5} \quad (i=1,2,...,12) \quad (1)$$

e_{F_i} indicates the influence of factor F_i to average PFV of every month, their expressions are:

Table 3: Parameters table

Factor	-	+
F_i	α	β

Table 4: Coding chart for factors

C_i	Total consumption of t period
α	Autonomous consumption parameter
β	Marginal propensity to consume parameter ($\beta \in [0,1]$)
λ, μ_t	Constant
t	Period
Y_t	Total income of t period

$$\left\{ \begin{array}{l} e_{F_{eo}} = \frac{\sum_{i=2,4,\dots,32} \sum_{j=1,3,\dots,31} (R_i - R_j)}{16} \\ e_{F_{qv}} = \frac{\sum_{i=4}^{32} (R_i + R_{i-1} - R_{i-2} - R_{i-3})}{16} \\ e_{F_{vp}} = \frac{\sum_{i=5}^8 (\sum_{j=0}^3 (R_{8i+j-35} - R_{8i+j-39}))}{16} \\ e_{F_{im}} = \frac{\sum_{i=9}^{16} (R_i - R_{i-8}) + \sum_{j=25}^{32} (R_j - R_{j-8})}{16} \\ e_{F_{af}} = \frac{\sum_{i=17}^{32} (R_i - R_{i-16})}{16} \end{array} \right. \quad (2)$$

Table 3 shows the meaning of +, -, it means the change of PFV when the value of factor f changes from α to β . It will be easy to predict the PFV when we figure out the change law of f (i). Analysis of 5 factors and their expressions will be given in the next section.

FACTORS ANALYSIS

PFV prediction is affected by many complex nonlinear factors. Qingdao's regional economic conditions, population density, ticket price, time and safety five factors are chosen here for comprehensive analysis and obtain Qingdao's PFV prediction model.

REGIONAL ECONOMIC CONDITIONS

Regional economic is mainly measured by consumption level of the specific region and consumption level can be described by the consumption function. Consumption function is a function that determines consuming behavior and the key indicator that reflect regional economic conditions. Consumption function was first proposed by English economist J.M. Keynes (Keynes, 1936; Davis, 1952).

Hall (1978) introduce rational expectation consumption into consumption function in 1978 and proposed adaptive expectations consumption function model:

$$C_t = \lambda a + (1 - \lambda) C_{t-1} + \lambda \beta Y_t + \mu_t \quad (t = 1, 2, 3, \dots, T) \quad (3)$$

Table 4 shows the meanings of every parameters in the Eq. 3:

Using generating function method to recursively solve the Eq. 3. Assume $H(x) = C_0 + C_1x + C_2x^2 + \dots$, multiply $(1-\lambda)x$ to, $H(x)$, it becomes $(1-\lambda)xH(x) = (1-\lambda)x C_0 + (1-\lambda) C_1x^2 + (1-\lambda) C_2x^3$ subtract these two expressions above, extract common factor and divide $(1-\lambda)x$ both sides of the expression, it becomes:

$$H(x) = \sum_{k=0}^{\infty} (\lambda a + \lambda \beta Y_t + \mu_t) \frac{1 - (1-\lambda)^k}{\lambda} x^k,$$

since $H(x) = \sum_{i=0}^{\infty} C_i x^i$, then:

$$C_t = C(t) = (\lambda a + \lambda \beta Y_t + \mu_t) \frac{1 - (1-\lambda)^t}{\lambda} \quad (t = 1, 2, 3, \dots, T) \quad (4)$$

POPULATION DENSITY

Population density will affect travel number of people. This study use Logistic Curve Model to realize the prediction of Qingdao permanent population:

$$Y_t = \frac{\kappa}{1 + \alpha \beta^t} \quad (t = 1, 2, 3, \dots, T) \quad (5)$$

where, κ , α and β are unknown parameters in the expression, t indicates time. Using modified exponential curve parameters method (Pangjin and Wu, 2008), because:

$$S_1 = \sum_{t=0}^{m-1} Y_t^{-1}, \quad S_2 = \sum_{t=m}^{2m-1} Y_t^{-1} \text{ and } S_3 = \sum_{t=2m}^{3m-1} Y_t^{-1}$$

Substitute S_1, S_2, S_3 into Eq. 5, therefore:

$$\left\{ \begin{array}{l} \beta = \left(\frac{S_3 - S_2}{S_2 - S_1} \right)^{\frac{1}{m}} \\ \alpha = \frac{m(\beta - 1)}{S_1(S_2 - S_1)(\beta^m - 1)^2} \\ \kappa = \frac{m(\beta^m - 1)}{S_1\beta^m - S_2} \end{array} \right. \quad (6)$$

Choose $m = 6$ and select as one month's permanent population of Qingdao randomly, they are 9091.99, 8731.22, 8397.81. Substitute values into Eq. 6 we get:

$$\left\{ \begin{array}{l} \beta = \frac{8397.81 - 8731.22}{8731.22 - 9091.99}^{\frac{1}{6}} = 0.986 \\ \alpha = \frac{6 \times (0.986 - 1)}{9091.99 \times (-360.77) \times (0.924 - 1)^2} = 4.1321 \times 10^{-6} \\ \kappa = \frac{6 \times (0.924 - 1)}{9091.99 \times 0.924 - 8731.22} = 1.38 \times 10^{-3} \end{array} \right. \quad (7)$$

Table 5: Population table

Year	City permanent population/10,000	t	Year	City permanent population/10,000	t
1988	651.69	1	1999	702.97	12
1989	657.15	2	2000	706.64	13
1990	666.64	3	2001	710.48	14
1991	670.92	4	2002	715.65	15
1992	673.10	5	2003	720.68	16
1993	675.34	6	2004	731.12	17
1994	678.52	7	2005	740.90	18
1995	684.63	8	2006	749.38	19
1996	690.26	9	2007	757.99	20
1997	695.43	10	2008	761.56	21
1998	699.56	11	2009	762.91	22

Get:

$$Y_t = \frac{1.38 \times 10^{-3}}{1 + 4.1321 \times 10^{-6} \times 0.986^t} \quad (8)$$

Qingdao's permanent population of every year and its prediction (Table 5) and population growth is given below.

TIME, TICKET PRICE AND SAFETY

Travel time (T_t) and waiting time (W_t) are the important factors that affect the selection of means of transportation from the view of passengers (Ruhe and Yan, 2005). Passenger's time-consuming in one travel can be expressed as (μ , η are parameters):

$$T_{sum} = \mu T_t + \eta W_t \quad (9)$$

Passengers' total consuming time T_{sum} in one transportation has a passive influence on passengers' choose of this means of transportation. There are three types of trains EMU, express and local train in Qingdao railway station. T_t is approximately equal to the Average Travel Time (ATT) of every train. There are altogether 19 EMUs and ATT equals to 4.16 h, 2 express trains with ATT equals to 18.98 h and 22 local trains with ATT equals to 22.54 h. ATT of all the trains T_t equals to 14.25 h. W_t is hard to investigate, we assume W_t to be 0.5 h for the convenience of simulation.

Ticket price is also an important factor that affects PFV. Ticket price varies from travel distance to travel distance. Assuming that there are χ stations from origin stop to last stop, the distance between the i station and the $i+1$ station is P_i ($1 \leq i \leq \chi$), the ticket price between the i station and the $i+1$ station is P_i , so the average price from station to station is shown as follows:

$$\bar{P}_{avg} = \frac{\sum_{i=1}^{\chi-1} (\chi-1) P_i}{\sum_{i=1}^{\chi-1} (\chi-1) P_i} \mathfrak{R}(t) \quad (10)$$

$\mathfrak{R}(t)$ is a power function on t .

Safety is one of the important factors that affect the passengers' choice of train. It is mostly affected by drivers' work time and the train's speed (Kecklund and Ingre, 2001). \bar{S} means safety coefficient, W_h is work time, μ and ρ are variable parameters, T_{sp} is the speed of the train. Safety can be described as:

$$\bar{S} = \mu \operatorname{sgn}(W_h - 8) W_h + \frac{\rho}{T_{sp}} \quad (11)$$

PFV PREDICTION MODEL AND ERROR CORRECTION

We can view nonlinear factors into some simple unities, therefore, factors affect PFV can be expressed as factors linear superposition. Take Qingdao's PFV as experimental data from 2005 to 2009 (Table 6).

First, e_{Feco} assuming that is the percentage of the influence from regional economic conditions to PFV, therefore:

$$e'_{Feco} = \frac{e_{Feco}}{e_{Feco} + e_{Fpop} + e_{Ftp} + e_{Ftim} + e_{Fsa}} \times 100\%$$

other factor's influence percentages can be figured out in the same way. PFV therefore, can be expressed as:

$$P_{PFV} = e'_{Feco} C_t + e'_{Fpop} Y_t + e'_{Ftp} \bar{P}_{avg} + e'_{Ftim} T_{sum} + e'_{Fsa} \bar{S} \quad (12)$$

Assume that means years and months to be predicted, for example 2020.04 and 2030.12 means April of 2020 and December of 2030. t means month, $t = (T - [T]) \times 100$. Prediction model is a combining form of power function and exponential function on t . Substitute every into PFV prediction model, we get PFV of every month of every year, compared them to experimental data, we can see the comparison in Fig. 2 and 3. Figure 2 illustrates that we use the previous prediction model to predict the PFV, then we compared it with the actual PFV, we can see that there are obvious different between these

Table 6: PFV of Qingdao city

Month	Year (PFV×10,000)				
	2005	2006	2007	2008	2009
1	128.0	61.2	73.4	110.4	115.1
2	116.7	42.3	72.6	97.0	101.2
3	116.0	47.5	69.4	98.4	101.5
4	42.0	63.5	70.4	105.0	118.4
5	47.6	72.2	81.4	110.0	125.6
6	40.0	64.3	98.4	105.7	109.4
7	59.0	93.3	130.1	123.7	148.9
8	61.0	92.4	129.2	120.6	158.9
9	53.0	75.7	98.9	114.5	119.4
10	35.5	81.3	100.3	110.8	126.5
11	48.7	56.2	86.9	91.2	99.5
12	44.1	55.5	90.2	90.8	100.7

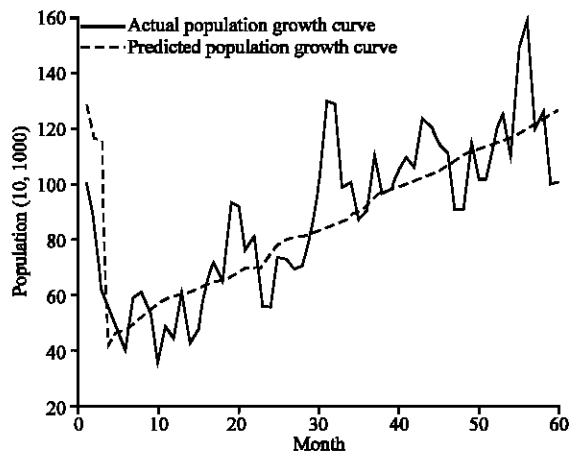


Fig. 2: Preliminary PFV prediction

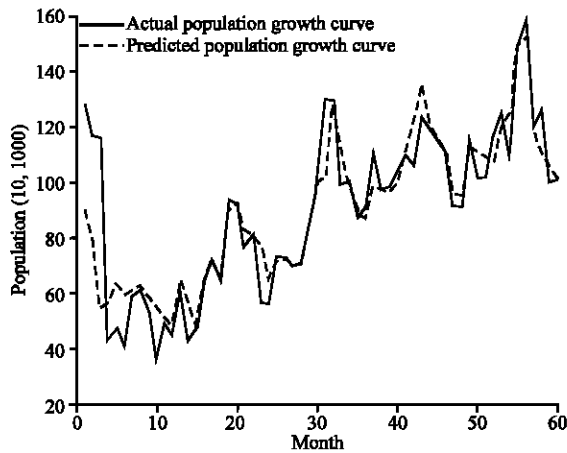


Fig. 3: Final PFV prediction

two groups of data and in Fig. 3, we use prediction model expressed in expression 17 and we find the prediction data is very similar with the actual data, it means our latter prediction model is much better in prediction.

Table 7: MAPE value

Month	MAPE (100%)				
	2005	2006	2007	2008	2009
1	29.68	7.84	3.26	10.32	0.60
2	31.79	27.60	0.55	0.00	8.79
3	28.57	0.42	1.87	7.63	7.19
4	33.30	0.15	1.42	4.76	1.09
5	34.40	2.07	4.17	0.00	4.37
6	47.50	1.55	1.52	0.86	14.35
7	2.20	3.64	77.70	9.53	0.067
8	2.62	0.21	0.69	0.49	4.21
9	10.94	9.37	1.21	0.43	0.58
10	56.30	1.72	1.09	0.00	12.96
11	3.28	3.62	14.15	21.50	6.43
12	9.07	16.03	3.99	4.51	0.19

This simulation is generally in accordance with actual trend but below expectation. When analyze the monthly PFV line chart from 2005 to 2009, we will find PFV tending to ascending trend year by year totally (The x-coordinate represent 60 months of the 5 years). Table 6 in specific, PFV grow rapidly during vacation and drop rapidly after vacation, then come to gradual advance trend. This trend can be used to describe for other year. It means we should deal with the prediction of PFV in vacation and in peacetime separately, including fluctuation. The modified prediction model is shown as follows:

$$P_{PFV} = (e'_{T_{eco}} C_1 + e'_{T_{top}} Y_1 + e'_{T_{sum}} T_{sum} + e'_{T_{avg}} \zeta_{avg} + e'_{T_{ref}} \bar{S}) f(t) + \sigma(t) \quad (13)$$

$f(t)$ is a power function on t . According to curve trend of experimental data, we assume that $\sigma(t)$ equals to:

$$\sigma(t) = \begin{cases} 1.25 \frac{T-2010}{\alpha} t \times 10^3 & (t=4,7,12) \\ 4.0 \frac{T-2010}{\beta} t \times 10^5 & (t=1,5,8,9,10) \\ 1.0 \frac{T-2010}{\gamma} t \times 10^5 & (t=2,3,6,11) \end{cases} \quad (14)$$

α , β and γ are constant which are used to assure $\sigma(t)$ increase regularly year by year. According to (Lewis, 1982), if MAPE is lower than 20%, the model can produce a fine prediction:

$$MAPE = \frac{1}{M} \sum_{i=1}^M \left| \frac{av(i) - pv(i)}{av(i)} \right| \% \quad (15)$$

av (i) and pv (i) means actual value and predictive value, M means month, that is how many months' experimental data it will use. The over striking values in Table 7 are numbers that fail to make MAPE under 20%. Since all kinds of emergency may affect PFV greatly and values which fail to meet the requirement only take a little part of total values will be deemed to be within tolerance. In summary, the PFV prediction model proposed in this study is effective.

CONCLUSION

In this study, a PFV prediction model is proposed by using the method of factor-response simulation and the factorial design method on the factors which affect PFV mostly. We make a test on prediction model by using the experimental data of Qingdao railway station and we find the model's MAPE is under 20% totally. The precise and practical prediction model can help the National Railway Administration to draw up an effective railway operation plan to achieve the greatest economic benefit.

ACKNOWLEDGMENTS

Project supported by the State Key Laboratory of Rail Traffic Control and Safety (Contract No. RCS2008K001), Beijing Jiaotong University and National High-Tech Research and Development Program of China (863 Program) grants 2009AA062703, China.

REFERENCES

- Averill, M.L. and W.D. Kelton, 2000. Simulation Modeling and Analysis. 3rd Edn., Tata Mcgraw-Hill Publishing Company Limited, Tsinghua University Press, pp: 622-666.
- Davis, T.E., 1952. The consumption function as a tool for prediction. *Rev. Econ. Statis.*, 34: 270-277.
- Davis, G.A. and N.L. Nihan, 1991. Nonparametric regression and short-term freeway traffic forecasting. *J. Trans. Eng.*, 117: 178-188.
- Dearien, J.A. and M.M. Plum, 1993. The Capital, energy and time economics of an automated, transportation system. *IEEE Aerospace Elec. Sys. Mag.*, 8: 28-32.
- Dougherty, M., 1995. A review of neural networks applied to transport. *Trans. Res. Part C Emerging Tech.*, 3: 247-260.
- Hall, R.E., 1978. Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence. *J. Econometrics*, 86: 971-987.
- Kecklund, L. and M. Ingre, 2001. Railway safety and the train driver information environment and work station. *Signalling safety 2001*, London, pp: 1-12
- Keynes, J.M., 1936. *The General Theory of Employment Interest and Money*. Macmillan, London, UK., ISBN-10: 1573921394, pp: 403.
- Lan, T.Y., C.Y. Lo, M.Y. Wang and Y.A. Yu, 2008. Multi-quality prediction model of CNC turning using back-propagation network. *Inform. Technol. J.*, 7: 911-917.
- Lan, T.S., 2009. Neural-based GA optimization on multi-objective CNC turning. *Inform. Technol. J.*, 8: 839-846.
- Lewis, C.D., 1982. *Industrial and Business Forecasting Method: A Practical Guide to Exponential Smoothing and Curve Fitting*. Butterworth Scientific, London pp: 143.
- Mogham, I. and S. Rahman, 1989. Analysis and evaluation of five short-term load forecasting techniques. *IEEE Trans. Power Syst.*, 4: 1484-1491.
- Okutani, I. and Y.J. Stephanedes, 1984. Dynamic prediction of the traffic volume through Kalman filtering theory. *Trans. Res. Part B. Methodol.*, 18: 1-11.
- Pangjin, B. and B. Wu, 2008. The forecast of the permanent residents of Qingdao based on the logistic model. *J. Shandong Univ. Sci. Tech.*, 27: 102-108.
- Ruhe, X. and Z. Yan, 2005. Comprehensive appraisal oil intercity railway passenger transportation. *J. Guangzhou Univ.*, 4: 58-64.
- Shi, F., L. Deng and L. Huo, 2007. Boarding choice behavior and its utility of railway passenger. *China Railway Sci.*, 6: 117-121.
- Steven, K.C.L. and H.C. Keh, 2011. Embedding a multi-agents collaboration mechanism into the hybrid middleware of an intelligent transportation system. *Inform. Technol. J.*, 6: 1113-1125.
- Tsung-Hsien, T. and L. Chi-Kang, 2003. An artificial neural networks approach to forecast short-term railway passenger demand. *J. Eastern Asia Soc. Trans. Stud.*, 5: 221-235.
- Wang, B., Z.H. Dong and Z.J. Cao, 2002. Study on induced-increasing traffic theory. *J. Hebei Univ. Tech.*, 31: 102-106.
- Yu, X., H. Gao and Z. Deng, 2008. The research on control of lunar rover with rocker bogie based on bus network driving. *Inform. Technol. J.*, 7: 1001-1008.
- Zafar, M.F., D. Mohamad and R.M. Othman, 2006. Writer independent online handwritten character recognition using a simple approach. *Inform. Technol. J.*, 5: 476-484.
- Zurada, J.M., 1995. *Introduction to Artificial Neural Systems*. PWS Publishing, Boston, MA, USA.