A Study of Speculating the Number of Tourists by Web Information - A Case Study of Sansia Old Street

Chia-Hui Yu[1] Meng-Long Shih[2*]

Abstract

A Taiwanese proverb: “The crowd would cause money to influx”. The amount of tourists is the element of the tourism market boom, and it is the priority to accurately estimate the amount of tourist for the manager of spring resort. The amount of tourist is the very important basic information for managing the spring resort, it is helpful to grasp tourism trend and understand the tourist needs for tourism managers. The forecast of amount could provide the spring resort to estimate the amount of tourist for the future, and thus the managers could forecast the tourist needs of recreation facilities, travel information, and service demand to become the basis of planning future management strategy development.

In this study, blog data were collected from a total of 1,457 articles from 2009 to 2013 about Sanxia Old Street’s tourism from the Yahoo website. Based on literature exploration, the emotional dimensions of blog posts were analyzed and converted to emotional scores. Furthermore, the Internet search popularity of Google Trends and five tourism-related economic indicators were taken as independent variables. The back-propagation network (BPN) of artificial neural network was used to construct the prediction models for combinations of different types of variables for verification and forecast accuracy evaluation. The empirical results show: 1. Taking advantage of electronic word-of-mouth , Internet search popularity on Google Trends combined with economic indicators can really be used in forecasting the tourism market acceptances. 2. The forecast model that combines economic indicators and electronic word-of mouth with Internet search popularity on Google Trends is superior to artificial neural network models that solely use economic indicators. 3. The empirical results show that the best forecasting ability is when emotional scores of Internet reputation are moved forward for four months, indicating

Keywords：Search heat of Google trends, Emotion indicator

[1] Professor and Corresponding Author, Department of Culture Resources and Liesure Industries, Tatung University.
[2*] Master, Department of Cultural Resources and Leisure Industries,National Taitung University, Corresponding Author E-mail：mlshih@nttu.edu.tw

I. Introduction

The implementation of the policy of two-day weekends in Taiwan has increased domestic travel time and spending. The supporting measures of government agencies and the diverse tourism products developed by the tourism industry have also indirectly promoted the development of local community. In the reflection of the wave of globalization and localization, due to the rising awareness of cultural asset protection of the local community and the governmental promotion of local industries, image district plan, community building and other issues, the declining old streets regain the attention of the mass tourism market. The old streets with great tourism market potentials have indirectly become the popular sites of attraction for two-day weekends. Sansia Old Street is a good example. With the promotion of media and the Internet, combined with the local cultural belief of Sansia Tzushr Temple as well as the tour packages of neighboring attractions, Sansia Old Street has gradually become a tourist attraction of foreign and domestic tourists.

Market temperature is a figurative term, like the temperature of the socio-economic prosperity. In the time of economic growth, people expect a promising future of the economy, and it is often referred to as the economic boom. In contrast, for the tourism market, “crowd means money flows”. The number of tourists represents the booming degree of the tourism market. The accurate analysis of tourism market fluctuations can help decision makers to make the right judgments, in particular, the analysis of the number of tourists. For policy makers, the number of tourists is a very important basic data of tourist attractions management, which can be used to grasp the trend of tourists, and understand the needs of tourists. Direct information can be provided for the reference of the tourist attraction managers in attraction planning, as well as the development of management strategies. In addition, information about the number of tourists can be used to evaluate the demand and supply of tourist attractions resources, in order to ensure sustainable use of resources and enhance the quality of tourism. Therefore, if the tourism business decision-makers can accurately predict the number of tourists, they can control costs, reduce risk and create profits.

With the advancement of information technology and popularity of social network and the booming of the tourism industry, the Internet has been widely used in the tourism industry. The Internet provides a large amount of useful information, which for the industry is good news. For example, consumers can share their travel experiences and opinions on the Internet forums, discussion boards and message boards. The industry can take use the valuable information (e.g., consumer’ feedbacks) as a guide to improve the direction of management decisions.

Web information (including movie reviews, news or WOM (Word of Mouth)) has been widely used in various fields of research. In terms of industries, some studies have used movie reviews or WOM as prediction indicators (Brewer et al., 2009; Joshi et al., 2010). In terms of the social sphere, Yang, Tsai, Huang and Peng, (2011) used the Internet search quantity data in the study of suicide. Shih(2014) used the web information to speculate the number of hot spring resort tourists. Smith, (2012) used the Internet search amount data to predict fluctuations in foreign exchange. To sum up, there is a certain association between web information and the real social behaviors. Web information has a guiding role for the customer decision-making, and the travel behaviors of tourists have precursory effects.

Current studies have explored the relationship between economic indicators and tourism, using economic indicators to forecast and predict tourism demand (Law and Au,1999; Gounopoulos,Petmezas and Santamaria 2012). Economic indicators used in these studies include the real GDP, the consumer price index, exchange rate, the unemployment rate, etc. Song, Gao, and Lin (2013) used web information and economic data to forecast the number of tourists in China, Taiwan, Japan, Australia, the U.K., and the U.S. They found that the model combining web information and economic data is superior to the model of economic data. Therefore, this study aims to use the web search information and economic indicators to establish the forecasting model for the number of tourists, and analyze the advantages and disadvantages of the forecasting model. The findings can provide a reference for the tourism industry in the future planning, design, and business management strategy development. Moreover, it can provide a direction for researchers in forecasting the number of tourists by forecasting variable indicators.

II. Literature Review

2.1 Definition of the number of tourists

The number of tourists is a direct indicator for the assessment of Sansia Old Street leisure industry's performance. More tourists mean more revenue for the Sansia Old Street leisure industry. The number of tourists in this study refers to the actual number of tourists arriving in the Sansia Old Street.

2.2 Estimation, prediction and statistical technology development regarding the number of tourists

Accurate forecasting of the number of tourists is the basis of developing business plans, as well as designing hotel facilities and services, and providing high-quality services to enhance tourist satisfaction.

Styness (1983) summarized the methods for forecasting the number of tourists, which are Delphi technique, time series, structural models, and system or simulation models (Lin and Lin, 2010).

Witt and Witt (1995) reviewed published articles on empirical research of tourism demand, and found that from 1966 to 1992, there were 19 articles using the dependent variable of the number of tourists. The empirical methods included linear model, log- linear model, logit model and the probit method. Those research methods have advantages and disadvantages.

Law and Au (1999) used the average price of hotel, service price, foreign exchange rates, population marketing costs, and gross domestic production to forecast the number of Japanese tourists visiting Hong Kong. The results confirmed that neural network is better than exponential smoothing method, multiple regression method and mobile average method.

Cho (2003) used exponential smoothing, ARIMA and neural network in six countries (the US, the UK, Singapore, Japan, Taiwan, and Korea) to predict the number of tourists visiting Hong Kong. The results indicated that except for the UK, the neural network is better than exponential smoothing and ARIMA.

Medeiros et al. (2008) used neural network to predict the international tourism demand in Balearic Islands and Spain. The method could provide the number of tourists over the time.

Song and Li (2008) reviewed 121 articles on tourism demand published in SSIC (Social Science Citation Index) from 2000 to 2007, and found 45 research methods use in those studies, including SVR (Support vector regression) and ANN (Artificial neural network). However, no single method showed optimal predictive results in all cases.

Gounopoulos, Petmezas, and Santamaria (2012) used exponential smoothing and ARIMA to forecast the number tourists in Greece, and explored the impact on the overall economy. The results showed that ARIMA is better than exponential smoothing.

Song, Gao, and Lin (2013) used the tourism information prediction system to forecast the number of tourists in China, Taiwan, Japan, Australia, the UK, and the US. The results showed that the model combining network information and economic data is better than the economic data model.

To sum up, there is no single method that produces optimal predictive results in all cases. Therefore, this study plans to use ANN as the method to forecast the number of tourists.

2.3 Google Insight for search-related applications

Google is currently the most popular Internet search engine in the world with more than 65% market share. Six billion searches per month globally can bring Google 25 billion USD (Smith, 2012) in revenue each year. Google Insight for Search is derived from Google Trends. Users can enter keywords to find the most popular terms among web users. If the same word becomes increasingly popular over a period of time, it means the amount of searches increases. The high attention and concern of the web users means that the product or thing represented by the word may have a certain changing (increase or decrease) trend (Lin and Chiu, 2012). Many people use Google Insight for Search for a wide range of purposes, including financial activities, social and medical behaviors. Ginsberg et al. (2009) used it to predict the flu epidemic, and found that the results can be used to accurately predict the flu epidemic around regions in the U.S. Vosen and Schmidt (2011) used Google Trends with three macroeconomic indicators to conduct a comparative study with the traditional indicators in the forecasting of private consumption. The results showed that Google indicators have better forecasting power in the training and testing samples. Shih (2014) used the Google Trends with three macroeconomic indicators to forecast the number of tourists in spa resorts, and found the forecasting power to be excellent. Smith(2012) forecasted the foreign exchange market volatility based on the search amount of specific keywords, including economic crisis, financial crisis and economic recession, in order to enhance the predictive power of GARCH (1,1). The results indicated that the information of search amount on Google is useful to the financial market. Frijters, Johnston, Lordan, and Shields (2013) used the Internet search information to study the relationship between macroeconomic conditions and problematic drinking.

2.4 Select the macroeconomic indicators

The macroeconomic indicators used in relevant studies may vary due to research subjects. For example, Song, Gao, and Lin 2013) used GDP, CPI, substitution price and other economic indicators to forecast the number of tourists. Smith (2012) studied the foreign exchange market volatility forecasting regarding specific keyword search amount by using five economic indicators, including the unemployment rate, stock trading volume, private consumer spending, housing prices and the volatility of stock returns. Vosen and Schmidt (2011) forecasted private consumption used three macroeconomic indicators, namely three-month national treasury bond interest rate and stock price index (Lin and Chiu, 2012). Gounopoulos, Petmezas and Santamaria (2012) used the real GDP, CPI, foreign exchange rate, and the unemployment rate to forecast the number of tourists in Greece, and explored the impact on macroeconomics. Shih (2014) used CPI, NTD-USD conversion rate, stock price index, the unemployment rate, and GDP to forecast the number of tourists in spa hotels. Based on the above studies, and considering that most tourists in Sansia Old Street are local tourists, this study excludes the foreign exchange rate as the indicator, and adopts five macroeconomic indicators, which are money supply, unemployment rate, GDP, CPI, and stock price index.

III. Research and Method

3.1 Research Structure

Among current studies, the forecasting of the number of tourists commonly uses macroeconomic indicators. This study combines Google Trends, emotional indicators, and macroeconomic indicators to forecast the future number of tourists of Sansia Old Street. The research framework is shown in Figure 1. This study consists of three parts: 1) collect related articles and indicators; 2) data preprocessing; 3) construct the forecasting model. The proposed forecasting model is then validated for the feasibility.

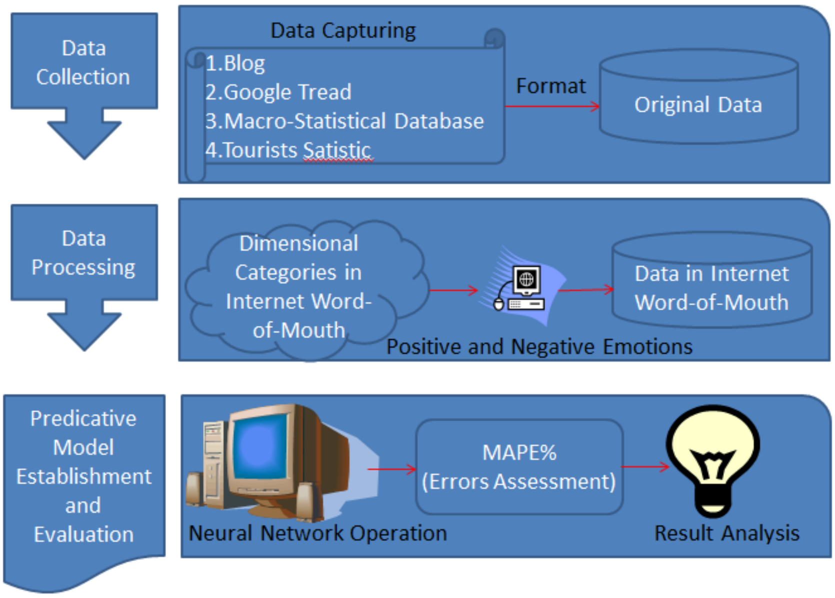

Figure 1. Research Framework

3.2 Data Sources

3.2.1 Number of tourists in Sansia Old Street

This study collected the data of "number of tourists in major tourist attractions" on the Executive Information System of the Ministry of Transportation and Communication, and selects the statistics of Sansia Old Street. From January 2009 to December 2013, there are 60 pieces of monthly data.

3.2.2 Google Trends

On Google Trends, the keyword of "Sansia Old Street" is inputted to obtain the monthly search trend, as shown in Figure 2. From January 2009 to December 2013, there are a total of 60 samples of monthly data.

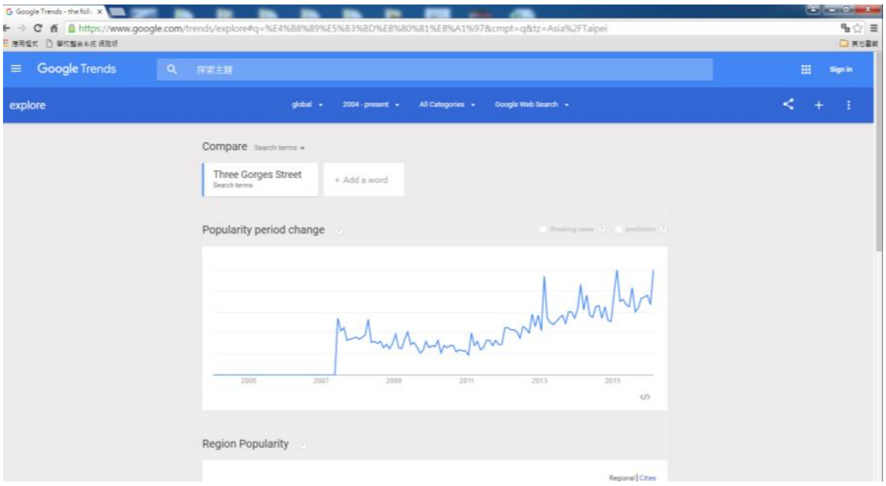

Figure 2 Sansia Old Street Google Trends

Source: Google Trends

3.2.3 Extraction of emotional indicators

Chiu, Hsu, Lin, and Peng (2013) explored the correlation among automobile sales volume, economic indicators, and search term popularity. The web emotional data were obtained by experts to label the articles with positive or negative scores before computing the emotional scores by month. Finally, they obtained seven types of emotional indicators. This study refers to the computational method of the emotional score, as described below:

(1) The number of positive articles (P_m_count)

X_i is the i-th positive article, d_i is the i-th article, and the number of positive articles in the m-th month is computed.

$$P_m\_count = \sum_i X_i \ , \ X_i = \begin{cases} 1 & if \ \ Score(d_i) > 0 \\ 0 & others \end{cases} \qquad (1)$$

(2) The number of negative articles (N_m_count)

Y_i is the i-th negative article, d_i is the i-th article, and the number of negative articles in the m-th month is computed.

$$N_m\_count = \sum_i Y_i \ , Y_i = \begin{cases} 1 & if \ \ Score(d_i) > 0 \\ 0 & others \end{cases} \qquad (2)$$

(3) Emotional score (PN_m)

The subtraction of the number of positive articles and the number of negative articles every month shows the emotional score in the m-th month.

$$PN_m\_Score = P_m\_count - N_m\_count \qquad (3)$$

This study shifts the input variables to learn how many months from the future can affect the number of tourists to achieve the purpose of using previous indicators to forecast the future number of tourists. The example of shifting months is shown in Figure 3. Shift1 refers to the shifting of all input variables of the original month by one month; Shift 2 refers to the shifting of all input variables of the original month by two months, Shift 3 refers to the shifting of all input variables of the original month by three months, Shift1~3 refers to the combination of all input variables of shifting by 1, 2 or 3 months.

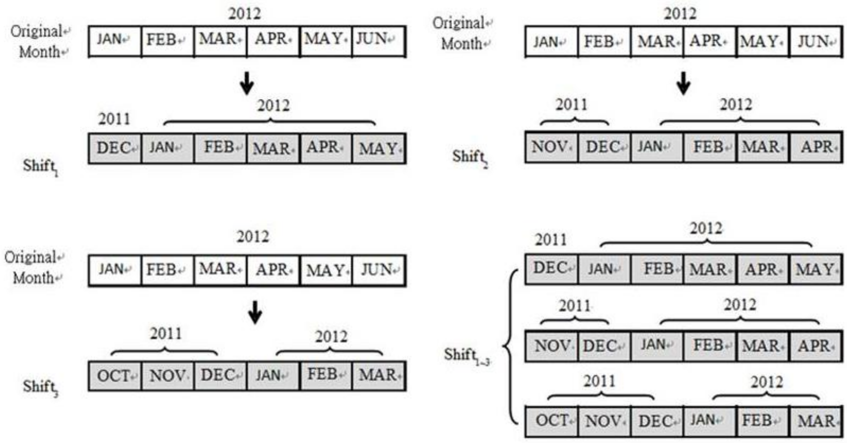

Figure 3 example of shifting months

Source: Chiu, Hsu, Lin, and Peng (2013)

3.3 Forecasting model establishment and evaluation

3.3.1 Forecasting model establishment

The emotional scores include cultural resources, price, food, environment, service, transportation, overall satisfaction; the economic indicators include M1B money supply, unemployment rate, GDP, CPI, stock price index, and Google Trends as independent variables. PCNeuron4.0 is used to input variable values and added with weighted value and biased values to obtain the dependent variables. The amount of the biased values is computed by the division of the addition of the independent variables and the dependent variables by 2. The proposed forecasting model is as shown in Figure 4.

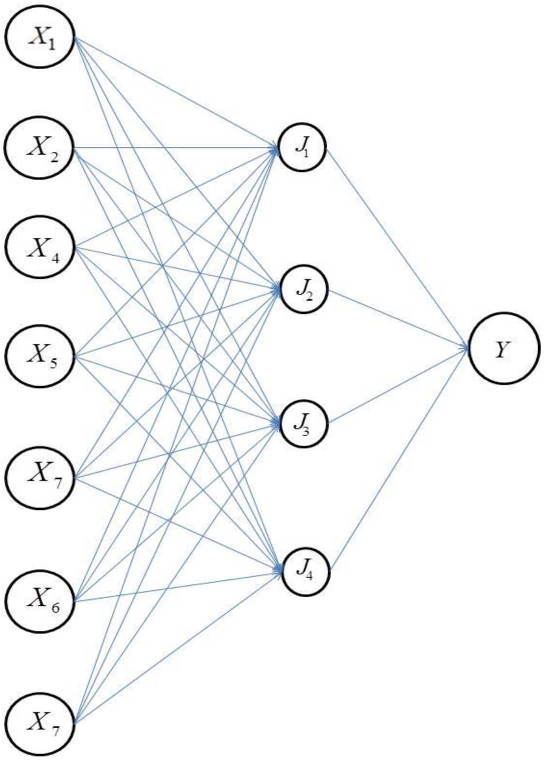

Figure 4 The proposed forecasting model

Source: Compiled by this study

Variables as shown in Figure 4 are illustrated as below:

X_1 =emotional score, X_2 =money supply, X_3 =the unemployment rate, X_4 =GDP, X_5 =CPI, X_6 =stock price index, X_7 = Google Trends, J_n ,n=1~4: weighted value, Y=dependent variable (number of tourists). This study uses the ANN web software of PCNeuron 4.0 to forecast the number of tourists visiting Sansia Old Street. The computation is divided into three parts: 1) training and testing sample; 2) model validation; 3) forecasting. This study uses the emotional score and 5 macroeconomic indicators, and Google Trends search amounts, as the input variables for training and testing of samples by different variable combinations. The impact of various variables on the accuracy of forecasting is verified.

3.3.2 Training, testing and validation samples

To avoid excessive learning, the BPNN requires a group of new samples to validate the weight modification accuracy known as the “Validation Set”. Therefore, after selecting the training and testing samples, this study validates the samples to confirm the accuracy of the forecasting model. The error of the above web model can be validated by RMS (Root of Mean Square). A smaller RMS indicates better error convergence and the forecasting model is closer to the real value. The definition equation of RMS is as follows:

$$\text{RMS}=\sqrt{\frac{\sum_{p}^{M}\sum_{j}^{M}\left(T_{jp}-Y_{jp}\right)^{2}}{M\times N}}$$

where T_{jp} =the target output value of the j-th output unit of the p-th sample

Y_{jp} =the inferred output value of the j-th output unit of the p-th sample

M=number of samples

N=number of processing units in the input layer

3.3.3 Web parameter settings

This study input various variables into PC Neuron4.0 for forecasting computation, and uses samples in 54 months from January in 2009 to June in 2013 for training and testing. The forecasting model is the constructed. Referring to Yeh (2009), the settings are the optimal parameter settings after being tested by thousands of neural network models. The number of learning cycles is 1000 times, and the number of testing periods is 10 without using the batch learning and the learning network-related weighted value. The range of the weighted value is 0.3, the random number seed is 0.456, the initial value of learning speed is 1, the learning speed deduction coefficient is 0.95, the learning speed lower limit is 0.1, the initial value of the inertia factor is 0.5, the inertial factor reduction coefficient is 0.95, and the inertial factor lower limit is 0.1.

3.4 Forecasting capability verification

The forecasting value obtained by using forecasting model is compared with the statistical data of tourists to Sansia Old Street by the Tourism Bureau for verifying the accuracy. This study uses indicators proposed by Chen, Lai, and Yeh (2012) including MAPE (mean absolute percentage error), RMSE (root mean square error), and MAD (mean absolute difference) to measure accuracy. The above three indicators are generally used to measure whether the time series statistics and target value fitness is correct. If the value is smaller, it means that the forecasting value is closer to the real value, and the forecasting capability is better. The computation equation of three evaluation indicators is as shown below:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t-F_t}{A_t}\right|\times 100\%$$

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mathrm{A_t}-\mathrm{F_t})^2}$$

$$MAD=\frac{1}{n}\sum_{i=1}^{n}\left|A_i-F_i\right|$$

A_t : target value

F_t : estimated value

As the denominator of MAPE is the real value represented by percentage, there is no problem of inconsistent comparison basis due to value difference. According to Lewis (1982), MAPE is divided into four grades. When MAPE<10%, it means that the error between the real condition and the forecasting value is smaller. If the value is closer to 0, it means that the forecasting capability is better. The principles of MAPE evaluation forecasting accuracy are as shown below.

Table 1 The principles of MAPE evaluation forecasting accuracy

MAPE	Forecasting capability
<10%	Excellent forecasting capability
10%~20%	Good forecasting capability
20%~50%	Reasonable forecasting capability
>50%	Incorrect forecasting capability

Source: Yeh (2011), Chen (2012)

IV. Analysis of Empirical Results

4.1 Data description

Using the above research method, this study collected the sample data of various variables from 2009 to 2013 to establish the sample original database for forecasting analysis. The independent variables include emotional indicators, search term popularity and economic indicators. The variables of economic indicators include money supply, unemployment rate, GDP, CPI, and stock price index. The statistical descriptions of various variables are as shown in Table 2.

Table 2 The statistical descriptions of various variables

Variable type	Variable description	Minimum value	Maximum value	Average value	Standard deviation
Dependent variable	Number of tourists (people)	94300.00	335284.00	173376.41	47887.47
The emotional indicators	Emotional score (point)	6.00	238.00	84.05	52.56
Search term popularity	Google Trends (samples)	28.00	100.00	49.90	16.18
Economic indicators	Money supply (NTD)	8327910.00	13274088.00	11293000.00	1218730.00
	Unemployment rate (%)	4.06	6.13	4.77	0.69
	GDP (NTD)	2986363	3812911	3419200	206141
	CPI	96.13	104.06	100.18	2.11
	Stock price index	4475.14	8970.76	7614.13	936.90

Source: Compiled by this study

The data of economic indicators are sourced from the PC-AXIS Overall Statistics Database of the Directorate-General of Budget, Accounting and Statistics, including the monthly indicators and seasonal indicators. The monthly indicators include CPI, NTD-USD conversion rate, stock price index, and unemployment rate; the seasonable indicator is GDP. As the forecast tourist number refers to the monthly data, the seasonal indicators are converted into monthly values to obtain consistency in data presentation.

4.2 Analysis of empirical results

This study used different variable combinations in BPN forecasting to construct the forecasting models for the tourism market temperature of Sansia Old Street. The variables include EM (economic index model), EGM (Economic, Google Trends index model, EMM (Economic, Mood index model), and EGMM (Economic, Google Trends, Mood index model). The empirical results of the following forecasting models are discussed below.

Using 54 months' data from January 2009 to June 2013 as samples, this study input five variables of economic indicators for the training and testing of forecasting model. The EM analysis and evaluation results are as shown in Table 3.

Table 3 The EM analysis and evaluation results

Forecasting model	Item		Value
Economic index model (EM)	Analysis item	(Training, testing) samples	(40,14)
		The Number of input layers	5
		The number of hidden layers	2
		The number of output layers	1
		Testing model RMS	0.10673
		Model validation RMS	0.10991
	Evaluation item	MAPE	21.89%
		RMSE	38572.69
		MAD	27665

Source: Compiled by this study

As shown in Table 5, the selected EM uses 40 samples in training and 14 samples in testing, and implements with two hidden layers. The forecasting model testing RMS is 0.10673, indicating that the forecasting model error convergence is good. The error between the forecasting value and the real testing value is very insignificant. In addition, 10 samples are additionally used in validation, showing the model validation RMS is 0.10991. This means that the model's training and learning effect is good without the problem of excessive learning. The training model RMS and model validation RMS are insignificantly different, suggesting that the training model has generality and similar forecasting capability for samples out of the training samples.

Ten samples in forecasting and the model evaluation mentioned in the above section are used to compare the forecasting value and the real value. For the three evaluation indicators, MAPE is 21.89%, RMSE is 38572.69, and MAD is 27665. Compared to the MAPE forecasting capability scale table, EM has forecasting accuracy of a reasonable range.

4.2.1 EGM results analysis

By following the forecasting model analysis of economic indicators, search term popularity is added into economic indicators for forecasting, and the training and testing sample combinations are used to select the optimal forecasting model. The forecasting model analysis and evaluation of economic indicators combined with search term popularity is as shown in Table 4.

Table 4 The forecasting model analysis and evaluation of economic indicators combined with search term popularity

Forecasting model	Item		Value
EGM	Analysis item	(Training, testing) samples	(40,14)
		The Number of input layers	5
		The number of hidden layers	2
		The number of output layers	1
		Testing model RMS	0.10673
		Model validation RMS	0.10991
	Evaluation item	MAPE	21.51%
		RMSE	37217.20
		MAD	27474

Source: Compiled by this study

EGM uses 40 samples in training and 14 samples in testing as well as one hidden layer. The forecasting model testing RMS is 0.10673, suggesting that the error between the expected and actual forecasting values is very slight. In other words, the forecasting model error convergence is good. In addition, ten 10 samples are input for validation and the model validation RMS is 0.10991, suggesting the learning effects of the model training is good without the problem of excessive learning. The training model RMS and model validation RMS are low, indicating that the training model has generality and has close forecasting capability for samples out of the training samples.

This study forecasts 10 samples by using the proposed model. The forecasting values obtained by using the model evaluation method proposed in the previous section are compared with the real values. It is found that three evaluation indicators are: MAPE is 21.51%, RMSE is 37217.20, and MAD is 27474. Compared to the scale of MAPE forecasting capability, EGM has the forecasting accuracy in the reasonable range.

4.2.2 EMM results analysis

EMM is also tested with combinations of training and testing sample combinations to select the optimal forecasting model. The EMS analysis and evaluation results are as shown in Table 5.

Table 5 The EMS analysis and evaluation results

<table>
<tr><td>Forecasting model</td><td colspan="2">Item</td><td>Value</td></tr>
<tr><td rowspan="9">EMM</td><td rowspan="6">Analysis item</td><td>(Training, testing) samples</td><td>(40,14)</td></tr>
<tr><td>The Number of input layers</td><td>6</td></tr>
<tr><td>The number of hidden layers</td><td>2</td></tr>
<tr><td>The number of output layers</td><td>1</td></tr>
<tr><td>Testing model RMS</td><td>0.10221</td></tr>
<tr><td>Model validation RMS</td><td>0.10901</td></tr>
<tr><td rowspan="3">Evaluation item</td><td>MAPE</td><td>21.34%</td></tr>
<tr><td>RMSE</td><td>36511.07</td></tr>
<tr><td>MAD</td><td>27442</td></tr>
</table>

Source: Compiled by this study

The proposed EMM is a forecasting model with 40 samples in training, 14 samples in testing in addition to two hidden layers. RMS is 0.10221, suggesting that the error between the expected and actual forecasting values of the proposed forecasting model is very slight. It also means that the forecasting model error convergence effect is good. In addition, 10 samples are input for validation and the model validation RMS is 0.10901, suggesting that model training learning effect is good without excessive learning problem. The training model RMS and model validation RMS values are very low and the close, indicating that the training model has generality and has the similar forecasting capability for samples out of the training samples.

Next, 10 samples are input into the proposed model for forecasting, and the obtained forecasting values and actual values are compared by using the model evaluation method. The three valuation indicators are: MAPE is 21.34%, RMSE is 36511.07, and MAD is 27442. Compared to the scale of MAPE forecasting capability, EGM has the forecasting accuracy in the reasonable range.

This study validates the time points of the impact of web emotions on the forecasting, and shifts the emotional scores by 1 to 6 months using the lagging method and inputs them in the forecasting model. The obtained forecasting values are compared with the real values and evaluated. The relevant research methods have been described in the previous section of research design. The empirical results are as shown in Table 6.

Table 6 the result of forecasting for shifts the emotional scores by 1 to 6 months

Emotional score's shifting months	MAPE
0 month	21.34%
1 month	21.60%
2 months	21.42%
3 months	20.49%
4 months	18.46%
5 months	20.51%
6 months	21.41%

Source: Compiled by this study

As shown above, regarding the impact of the previous emotional scores on the forecasting of future Sansia Old Street tourism market, the forecasting accuracy of shifting by 4 months is the highest. This suggests that the web WOM emotions have an impact on the forecasting of tourism market temperature of Sansia Old Street.

4.2.3 EGMM results analysis

Finally, this study combines the economic indicators, search term popularity and web emotions to construct EGMM. The search term popularity and web emotional variables are shifted backward by 4 months to test the model forecasting capability. The results of EGMM analysis and evaluation are as shown in Table 7.

Table 7 The results of EGMM analysis and evaluation

Forecasting model	Item		Value
EGMM (search term popularity and web shifted backward by 4 months)	Analysis item	(Training, testing) samples	(40,14)
		The Number of input layers	7
		The number of hidden layers	2
		The number of output layers	1
		Testing model RMS	0.11500
		Model validation RMS	0.10030
	Evaluation item	MAPE	15.42%
		RMSE	28770.46
		MAD	19165

Source: Compiled by this study

The selected EGMM is a forecasting model using 40 samples in training, 14 samples in testing, and two hidden layers in implementation. The forecasting model testing RMS is 0.11500, suggesting that the error between the forecasting expected values and the actual values of the forecasting model is very slight and the error convergence of the forecasting model is good. Beside using 10 samples, the model validation RMS is 0.10030, suggesting that the learning effect of the proposed model training is good without the problem of excessively learning. The training model RMS and model validation RMS indicate that the training odel has generality, and the forecasting explanatory power is relatively stable with similar forecasting capability for samples out of the training samples.

Next, by using the proposed model, 10 samples are inputted for forecasting. With the model evaluation method as descried in the previous section, the forecasting values and actual values are compared. The three evaluation indicators include: MAPE is 15.42%, RMSE is 28770.46, and MAD is 19165. Compared to MAPE forecasting capability scale, EGMM has good forecasting accuracy.

4.3 Comparative analysis of the economic indicators, search term popularity and emotional model performance

According to the above empirical results, this section summarizes the results as shown in Table 10 to illustrate the difference and performance of the forecasting models established on the basis of four different types of variables.

Table 8 the forecasting models established on the basis of four different types of variables.

Model description	EM	EGM	EMM (shift 4 month)	EGMM (shift 4 month)
(Training, testing) samples	(40,14)	(40,14)	(40,14)	(40,14)
The Number of input layers	5	5	6	7
The number of hidden layers	2	2	2	2
The number of output layers	1	1	1	1
Testing model RMS	0.10673	0.10673	0.10221	0.11500
Model validation RMS	0.10991	0.10991	0.10901	0.10030
MAPE	21.89%	21.51%	18.46%	15.42%
RMSE	38572.69	37217.20	32297.11	28770.46
MAD	27665	27474	23631	19165

Source: Compiled by this study

Table 8 illustrates the forecasting models of the four combinations of variables with the same number of training and testing samples as well as hidden layers as the comparison basis of models of different variable combinations.

Most previous studies on number of tourists forecasting analyze economic indicators. However, the empirical results of this study suggest that, compared to the forecasting method of using economic indicators, the proposed method uses economic indicators, search term popularity, and web emotions as the variables and shift the search term popularity and web emotional variables backward by four months. The forecasting capability is improved from 21.89% of EM to 15.42% of EGMM, indicating that the forecasting accuracy has been improved from the reasonable to the good level.

The empirical results suggest that the forecasting model using economic indicators, search term popularity, and web emotions as variables while shifting the search term popularity and web emotional variables backward by four months has the best forecasting performance among the four models of different variable combinations.

Regarding the other two evaluation indicators of RMSE and MAD, EGMM has the lowest values among the four forecasting models. This can better validate that EGMM can enhance the forecasting performance of the economic indicators model. Figure 5 illustrates the comparison of the forecasting values and the real values by a broken line graph.

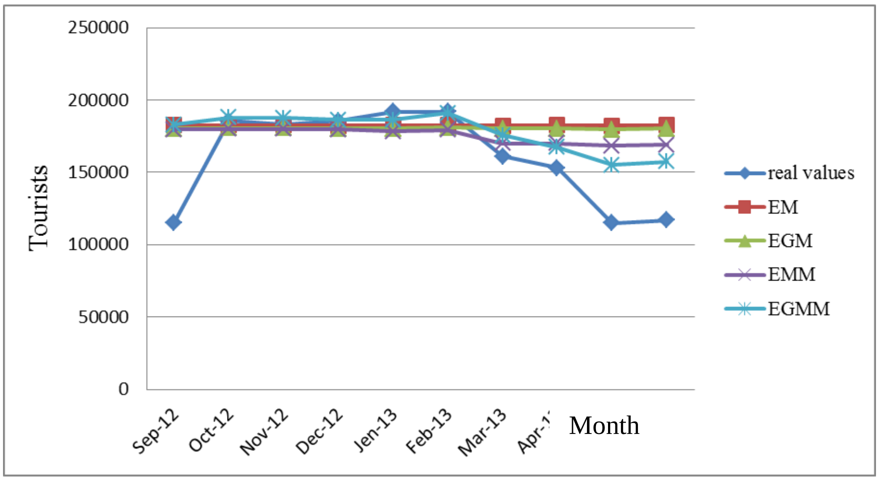

Figure 5 Comparison of the forecasting values and the real values by a broken line graph

Source: Compiled by this study

As shown in Figure 5, the forecasting value trends and the real values of EGMM are close, indicating that the forecasting model integrating economic indicators, search term popularity, and web emotion is better than the other three models.

4.4 Forecasting of future tourism market temperature

Based on the empirical results, by using the EGMM model and shifting the variables of search term popularity and web WOM backward by 4 months, this study forecasts the tourism market temperature of Sansia Old Street in the following six months of the forecasting model samples from July to December, 2013. The forecasting values and real values are compared. The resulting forecasting capability MAPE is 11.89%, indicating that EGMM is considerably accurate in forecasting the future tourism market temperature of Sansia Old Street.

Figure 6 illustrates the comparison of the forecasting values and the real values of Sansia Old Street in the following 6 months.

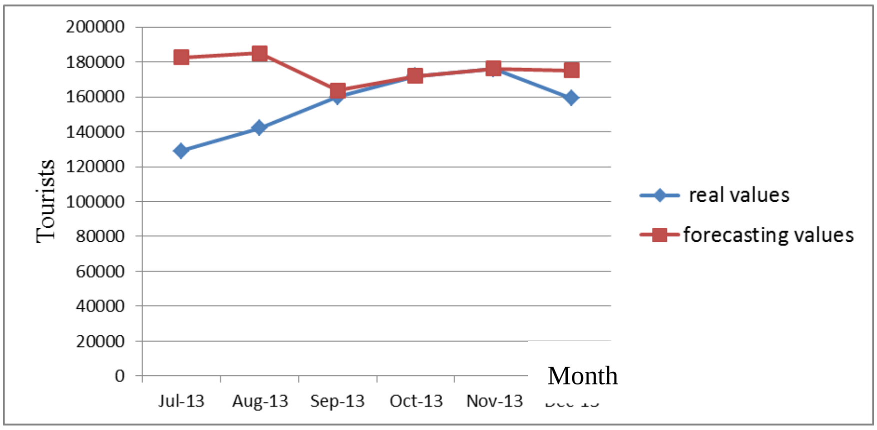

Figure 6 Comparison forecasting values and real values of tourism temperature of Sansia Old Street from July to December 2013

Source: Compiled by this study

As shown above, forecasting values tend to get close to the real values, indicating that the proposed model has good forecasting capability to forecast the tourism market temperature of Sansia Old Street.

To sum up, this study has empirically found that the forecasting model integrating economic indicators, search term popularity, and web emotions has the best forecasting performance. Moreover, previous web emotions have an impact on the forecasting of the future tourism market temperature of Sansia Old Street in 4 months ago. Meanwhile, the proposed forecasting model has good forecasting capability to forecast the future tourism market temperature of Sansia Old Street.

V. Conclusion and Suggestions

5.1 Conclusion

This study collected 1457 samples of Sansia Old Street-related blog data on Taiwan Yahoo from January 2009 to December 2013 as the basis for literature review. The emotional dimensions of the articles were analyzed, and the results were converted into emotional scores. Using Google Trends and 5 economic indicators as independent variables, as well as the number of tourists in Sansia Old Street as the dependent variable, this study constructed neural network forecasting models. The results showed the forecasting model integrating web WOM, search term popularity, and economic indicators can accurately forecast the tourism market temperature of Sansia Old Street. The empirical results are shown below：

This study used economic indicators, search term popularity and web WOM emotional scores to construct an EGMM. The model has 7 input layers, 2 hidden layers, and 1 hidden layer. About 75% of the total samples are used as for the training of forecasting model, and 25% of the total samples for the testing of the model. The EGMM parameter settings are: learning cycle times is 1000, testing cycle times is 10, weighted value range is 0.3, the random number seed 0.456, the initial value of learning speed 1, the learning speed deduction coefficient 0.95, the learning speed lower limit 0.1, the initial value of the inertia factor 0.5, the inertial factor reduction coefficient 0.95, the inertial factor lower limit is 0.1.

The forecasting model integrating economic indicators, search term popularity and web WOM is better than the neural network of economic indicators. The research results suggested that among four models proposed in this study, if judged by forecasting evaluation indicators including MAPE, RMSE and MAD, the MAPE value of EGMM is the lowest, indicating that the model forecasting value is closest to the real value and its forecasting capability is the best.

According to the empirical results, the forecasting capability of the forecasting model with web WOM emotional scores being shifted backward by 4 months is the best. The data mining also suggested that the number of tourists in Sansia Old Street and web WOM are correlated. In EGM, variables including emotional score are shifted backward by 1 to 6 months to forecast the tourism market temperature of Sansia Old Street. The results indicated that the forecasting capability of the model of shifting backward by 4 months is the highest and the value of MAP is 18.46%. Meanwhile, the time point of previous web emotion affecting the forecasting of tourism market temperature of Sansia Old Street is 4 months.

5.2 Suggestions

For the tourism industry and operators, the proposed model is an objective, innovative and convenient forecasting tool. Based on the results, the suggestions for local business operators or administrators of Sansia Old Street are proposed as follows:

5.2.1 Suggestions to the industry:

The findings suggested that web WOM emotions, search term popularity and economic indicators can affect forecasting tourism market temperature. Thus, the industry should refer to the search term popularity of “Sansia Old Street” by web users and focus on web marketing to create interesting topics to increase and maintain search term popularity.

As web emotion has a more influential impact on the forecasting of tourism market temperature, the industry should pay more attention to web WOM of Sansia Old Street. Web WOM and blogs can be used to introduce Sansia Old Street, and the contents should be constantly updated to create the positive web emotion. Meanwhile, the communication with tourists should be strengthened to raise tourism service quality and reduce negative web emotions.

The industry should refer to the trends of economic indicators including domestic money supply, unemployment rate, GDP, CPI and stock price index, in order to accurately forecast number of tourists, regulate tourism resources, and maintain the stability of service quality.

References

1. Chiu, C.C., C.H. Hsu, Y.C. Lin, and G.Y. Peng, 2013, A Study on the Correlation between Automobile Sales Volume and Economic Indicators and Search term popularity–a Case Study of Mazda Automobile, a research report of the Information Management Department, Yuan Ze University.
2. Chiu, C.C. and Y.T. Lin, 2012, Forecasting Taipei’s Movie Box Office by Using Web WOM, a research report of the Information Management Department, Yuan Ze University.

3. Lin, B.S. and Y.J. Lin, 2010, "Estimation and Forecasting of Island-type Recreational Area's Tourist Numbers," Journal of National Park, 20(1):1-14.
4. Shih, M.L., 2014, "A Study of Using Web Information to Deduce Spa Hotel Tourist Numbers – a Case Study of A Resort in the Rural Area," Rural Development Perspectives, 16:1-17.
5. Belsley, D.A., E. Kuh, and R.E. Welsch, 1965, Regression Diagnostis:Identifying Influential Data and Sources of Collinearity, Ithaca: Deportment of Agricultural Economics, Cornell University ,A.E.179.
6. Brewer, S. M., J. M. Kelly and J. J. Jozefowicz, 2009, "A Blueprint for Success in the US Film Industry," Applied Economics, 41(5):589-606.
7. Frijters, Paul D.W., G.L. Johnston, and M.A. Shields, 2013, "Exploring the relationship between macroeconomic conditions and problem drinking as captured by Google searches in the US," Social Science & Medicine , 84:61-68.
8. Chen, C.F., M.C. Lai, and C.C. Yeh, 2012, "Forecasting tourism demand based on empirical mode decomposition and neural network," Knowledge-Based Systems ,26:281-287.
9. Ginsberg, J., M.H. Mohebbi, R.S. Patel, L. Brammer, M.S. Smolinski, and L. Brilliant, 2009, "Detecting influenza epidemics using search engine query data," Nature, 457:1012–1014.
10. Gounopoulos, D., D. Petmezas, and D. Santamaria, 2012, "Forecasting Tourist Arrivals in Greece and the Impact of Macroeconomic Shocks from the Countries of Tourists," Origin, Annals of Tourism Research, 39(2):641-666.
11. Joshi, M., D. Das, K. Gimpel, and N. A. Smith, 2010, "Movie Reviews and Revenues: An Experiment in Text Regression," Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 293-296.
12. Law, R. and N. Au, 1999, "A neural network model to forecast Japanese demand for travel to Hong Kong ," Tourism Management , 20(1):89-97.
13. Cho, V., 2003, "A comparison of three different approaches to tourist arrival forecasting," Tourism Management , 24(3):323-33.
14. Lewis, C.D., 1982, International and Business Forecasting Methods, London: Butterworths.
15. Medeiros, M.C., M. McAleer, D. Slottje, V. Ramos, and J. Rey-Maquieira, 2008, "An alternative approach to estimating demand-Neural network regression with conditional volatility for high frequency air passenger arrivals," Journal of Econometrics ,147(2):372-383.
16. Smith, G.P., 2012, "Google Internet search activity and volatility prediction in the market for foreign currency," Finance Research Letters, 9(2):103-110.
17. Song, H. and Gang Li, 2008, "Tourism demand modeling and forecasting—A review of recent research," Tourism Management , 29(2):203-220.
18. Song, H., B.Z. Gao, and V.S. Lin, 2013, "Combining statistical and judgmental forecasts via a web-based tourism demand forecasting system," International Journal of Forecasting , 29, (2):295-310
19. Vosen, S. and T. Schmidt, 2011, "Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends," Journal of Forecasting, 30:565-578.

20. Wang, P.W., Y.J. Su, M.L. Shih, and S.D. Lou, 2010, “Analysis of online Word-of Mouth in Online Forums Regarding Notebook Computers,” Journal of Convergence Information Technology, 5(5):118-124.
21. Witt, S.F. and C.A. Witt, 1995, “Forecasting tourism demand: A review of empirical research,” International Journal of Forecasting ,11(3):447-475.
22. Yang, A.C., S.J. Tsai, N.E. Huang, and C.K. Peng, 2011, “Association of Internet search trends with suicide death in Taipei City, Taiwan, 2004-2009,” Journal of Affective Disorders, 132(1-2):179-184.

利用網路訊息推測遊客量之研究–三峽老街為例

余家蕙[1]施孟隆[2*]

摘要

台灣有句諺語：“人潮就是錢潮”。遊客數量是旅遊市場熱潮的要素，準確估計遊客數量是度假村經理人員的重要事情。遊客數量是管理度假勝地的重要基礎資料，有助於掌握旅遊趨勢，了解旅遊管理者的旅遊需求。預測量可以為未來遊客量預測提供一定的依據，因此管理者可以預測遊憩設施的旅遊需求，旅遊信息和服務需求，從而成為規劃未來管理策略發展的基礎。

本研究從 2009 年到 2013 年共收集了 1457 篇關於三峽老街旅遊的文章，來自雅虎網站。在文獻研究的基礎上，對遊客文章的情感維度進行分析，並轉化為情緒分數。此外，谷歌趨勢的互聯網搜尋熱度和五個旅遊相關的經濟指標被視為自變量。採用人工神經網絡的 BPN（Back Propagation Network）模型構建不同類型變量組合的預測模型，進行驗證和預測精度評估。實證結果表明：1.利用電子口碑的優勢，互聯網搜索谷歌趨勢與經濟指標相結合，可真正用於預測遊客量。 2.經濟指標和電子口碑相結合的預測模型與谷歌趨勢的互聯網搜尋熱度相比，優於單純使用經濟指標的人工神經網絡模型。3.實證結果表明，情緒分數向前推進四個月時，其模型預測力最佳。

關鍵字：搜尋熱度、情緒指標

[1]國立台東大學文化資源與休閒產業學系碩士
[2*]國立台東大學文化資源與休閒產業學系教授與通信作者 Email:mlshih@nttu.edu.tw