Analyzing Google Trends with Travel Keyword Rankings to Predict Tourists into a Group

Jui-Hung Chang, Chien-Yuan Tseng

[1] Computer and Network Center, Department of Computer Science and Information Engineering, National Cheng Kung University, Taiwan
[2] Department of Computer Science and Information Engineering, National Cheng Kung University, Taiwan
changrh@mail.ncku.edu.tw, P76031488@mail.ncku.edu.tw

Abstract

This study explored the correlation between tourism-related popular words in search engines and the number of tourists, which is a topic worth discussing for the tourism industry. When individuals decide to sign up for a tour planned by travel agencies, they often have no idea whether the minimum number of participants will be reached until it is confirmed a few days before the departure, which is of great inconvenience when arranging itineraries. Hence, predicting whether the minimum number of participants in a tour can be reached is closely related to both the tourism industry and the tourists. In this regard, the number of Taiwanese traveling to Japan was predicted based on the popularity of keywords concerning travel in Japan searched on Google Trends. The scores of popular words concerning travel in Japan on the Google search engine and words concerning travel in Japan mentioned in tourism articles in e-news networks were also summarized. The experimental results indicated that the popularity of tourism keywords on Google was highly correlated to the number of Taiwanese tourists traveling to Japan. After the number of Taiwanese traveling to Japan within the following month was classified by the ANN model, the mean square error reached 0.13. Furthermore, by using the data of travel agencies in Taiwan to match the Google Trends data, the research predicted whether tours to Hokkaido would reach the minimum requirement for participants. The prediction accuracy of the ANN model was 68%.

Keywords: Tourism-related popular words, ANN model, Google trends

1 Introduction

The Google search engine records a wide range of information. Before traveling abroad, many people will search for information concerning food, clothing, housing, and transportation, and Google will reflect the popularity of tourist destinations. It is necessary for the tourism industry to accurately predict future popular tourist destinations and the number of tourists so as to promptly adjust locations and prices, hoping to launch the most popular itinerary and increase profits. In recent years, studies pertaining to Taiwan's tourism have drawn considerable attention from both the Taiwan government and Taiwanese individuals.

This research found that studies concerning tourism are related to tourism recommendation systems [1-2], with a few studies sourcing from community websites [3-4] or heterogeneous platforms (such as Wikipedia, Wikivoyage, Wikitravel, Panoramio, Flick, etc.) [5]. However, technologies used for forecasting the number of tourists include a variety of statistical methods. A large number of studies have relied on historical data to predict the most popular tourist destinations in the future, while the most recent studies have taken advantage of a vast amount of reviews and community data on the internet, such as Trip Advisor or Flickr, to predict future popular tourist sites [6-7]. Also, some studies have employed the popularity of tourism-related words on search engines for prediction. For instance, some studies have use Baidu and Google to predict the popularity of Hainan Island [8]. Based on a period of lag model, the correlation coefficient calculation is conducted between the popularity scores of words concerning traveling on Hainan Island and the future number of tourists on Hainan Island to identify the words that can predict the degree of popularity. In this aspect, the information obtained from the auto-regression moving average (ARMA) models could improve the accuracy of the model.

Currently, there is neither a study that uses the popularity scores of tourism-related words on Google Trends to set up a neural network model to predict the number of tourists traveling in Japan, nor a study that employs the popularity scores of tourism-related words on Google Trends or the internal data of travel agencies to establish a neural network that can predict whether the minimum number of participants in a tour can be reached, which could enhance the evaluation of the tourism industry and the convenience for

*Corresponding Author: Jui-Hung Chang; E-mail: changrh@mail.ncku.edu.tw
DOI: 10.3966/160792642019012001023

individuals to make relevant arrangements. Thus, there were three topics in this study. The first area used the scores of popular words concerning travel in Japan on Google the and tourism information of e-news networks to improve the accuracy of predictions. After conducting word and sentence segmentation of the travel articles, tourism related words were summed up to analyze the correlation between the above two kinds of information and the number of Taiwanese tourists traveling to Japan. The second area used Google Tens or Flow to set up an ANN model and took data from the Google search engine as the input to predict the number of Taiwanese individuals traveling to Japan in the next n months as a reference for governments and the tourism industry. The third area used information concerning the popularity scores of tourism-related words on Google and the internal data from travel agencies, including travel year, travel month, name of tour, tour fee, and whether these tours reached the minimum requirement for participants to predict whether tours to Hokkaido would reach the minimum number of participants through the ANN model as a reference for governments and the tourism industry.

This study hoped to provide the following contributions:

(1) Experiments using the Pearson correlation coefficients revealed that Taiwanese individuals prefer to search information concerning locations and air tickets for relevant tours in Japan through Google three and four months prior to traveling. Words related to traveling in Japan that appeared in the tourism section of e-news sources had a low correlation with the number of Taiwanese tourists traveling to Japan (the correlation coefficient was less than 0.5), indicating that information from e-news sources had no impact on travel destination.

(2) In the experiment that predicted the number of Taiwanese individuals traveling to Japan, the popularity scores of tourism-related words on Google Trends were input into the ANN model to determine the number of Taiwanese tourists traveling to Japan in n months. The experimental results indicated that when n was equal to 1, the mean squared error of the model classification of 0.13 was the best, indicating that inputting multiple popularity scores of tourism-related words into the ANN model was the best method to predict the number of tourists traveling to Japan after one month.

(3) In the experiment that predicted whether a tour would reach the minimum number of participants, the popularity scores of words related to tourism in Hokkaido on Google Trends were input into the ANN model to determine whether the tours would reach the minimum number of participants. The experimental results revealed that the accuracy rate of the test data was 68%, which indicated that it was possible to efficiently predict whether the tour would reach the minimum requirement for participants through the ANN model.

As to the rest of this paper, Section 2 describes the relevant technical background used in this study, Section 3 describes the methodology, and Section 4 describes the experiment. Finally, the conclusions of this paper are summarized in Section 5.

2 Related Work

As one of the popular topics in domestic and international tourism-related studies, in recent years, a large number of studies have confirmed the positive correlation between the popularity scores of tourism-related words on Google Trends and number of tourists. This section will introduce the related literature of this topics, the ANN architecture, and the algorithms used in this study, including Google TensorFlow, a tool used to establish the ANN model.

Tourists often search for information concerning tourist spots on Google and book air tickets and hotels over the internet. In 2012, a number of studies used data from five Google search engines to predict the number of hotel reservations through the autoregressive-moving-average model [9]. In 2015, a study predicted the popularity of tourism on Hainan Island through the search engines of Baidu and Google [8]. The Pearson correlation coefficient was adopted to sum up words related to traveling on Hainan Island. Then, the lag time algorithm and relevant experiments were conducted to calculate the correlation between words concerning traveling on Hainan Island and the future number of tourists on Hainan Island. The highly relevant words were listed. Later, experiments using auto-regression moving average (ARMA) models showed that data from search engines could help reduce experimental errors. A number of studies also predicted the number of tourists through multiple data from Google [10] to improve the accuracy. In 2016, some studies proposed a new model and compared the accuracy of the new model with that of the original ARMA model [11].

2.1 Travel Related Researches

Multiple travel-related papers have been based on recommendation systems for studying tourism. In recent years, studies on recommendation systems have gradually recognized the importance of using multiple selection criteria for improving recommendation output. [12] Proposed a new multi-criteria recommendation method based on the summarization of “preference lattices” (partial orders) of the standard preferences of users. The study assumed that some of the selection criteria for a project (product or service) would dominate the overall rankings, and that these dominant criteria would be different for different users. Following this assumption, users would be grouped according to their standard preferences to create preference lattices. Users' recommendation output was based on the ratings of other users from the same or similar groups. After introducing the general method of aggregation, the study developed three alternative recommendation methods to substantiate the steps: (a) the standard aggregate function, (b) the overall project ratings, and (c) combining aggregation with collaborative filtering. Then, a set of experiments on a data set of service rankings was used to assess the accuracy of these three methods and compare them with the traditional collaborative filtering method to cover multiple standard layers. The results showed that the combination of aggregation and extended collaborative filtering was of the highest accuracy.

[13] described the development of a travel recommendation system based on user-enhanced profiles consisting of basic user information, users' correlation with a set of prototypes, and user functional levels. The work aimed to evaluate whether users' physical and psychological functional level in the creation of the users' profiles would have a significant impact on the recommendation result. In addition, the purpose of describing the work in this study was to categorize points-of-interest (POI) in different ways in order to consider whether they were capable of receiving tourists with a certain degree of physical and psychological problems.

Recommendation systems and adaptive systems have been introduced into tourism applications to support travelers' decision-making processes. These systems should be able to address unexpected changes during travel. Therefore, the first step is to sense the volume, needs and preferences of tourists before, during, and after the trip. In addition, information concerning accommodation, flights, cities, events and destinations are collected from different sources to provide personalized messages. The current tourism systems fail to collect all the information about tourists and travel products from different sources or provide a series of recommendations based on preferences and travel stages of tourists (i.e. main destination, expected budget, duration of travel, accommodation, transportation, activities and restaurants). These systems also do not support any customization. [14] proposed and implemented an adaptive tourism recommendation system backed by a suitable tourist recommendation framework, program, architecture and system. The tremendous growth of social networks on the internet has created new platforms and tools for social interaction and exchange among people. [15] incorporated messages into social networks to improve traditional recommendation systems, including user preferences and influence of social friends. User interest ontology was summarized, and these messages were used to develop personalized advice. The study proposed a social recommendation system that centers on user interest ontology. The system can improve the recommendation quality of tourism in Tunisia and can be used on Tunisian tourism websites to facilitate users who are interested in visiting Tunisia.

Improving accuracy has always been one of the most prominent issues in studies concerning recommendation systems. In recent studies, multiple criteria recommendation systems adopting multiple standards to evaluate overall scores have received considerable attention. [16] proposed a neural network model to improve the accuracy of multi-standard recommendation systems. The neural network was trained through simulation algorithms and integrated with two single-level recommendation system samples. The study introduced the experimental results of two single scoring techniques and the corresponding neural network model. To analyze the performance of methods, this study conducted a comparative analysis of the performance of each rating-based technology with the proposed multi-standard model. The experimental results showed that the proposed model was far superior to the existing technology. [17] looks to improve the accuracy of wind turbine error prediction, uses the K-means of unsupervised learning to classify similar characteristics, and utilizes the classified cluster in the BP neural network for error prediction, thus reducing the maintenance cost. The artificial neural network of back propagation (BP) can expand the accuracy by 3.5% in comparison to the traditional artificial neural network. The principle component method and function approximation are used to discuss what kind of news that users prefer. [18] abstracts several important characteristics of networks and selects the important part from multiple influential factors. Considering the high dimension property of the grading model, the artificial neural network (ANN) is employed to predict the popularity of news, with an accuracy of 95%.[19] aims at the recognition of human activity in combination with geometric parameter properties and the artificial neural network of Radial Basis Function (RBF). The image shape calculation takes vectors including region and shape to calculate the geometric parameter properties, which are then put into the RBF neural network for recognition. The improved RBF function increases the learning efficiency of ANN. The experimental results prove it attains better recognition accuracy more efficiently. [20] uses the BP neural network in Microblog, motivated by the openness and rapid transmission of Microblog. At present, users can spread their own opinions everywhere rapidly through blogs, and so the author wants to create a system for predicting the trend of public opinion (common people's opinion) in Microblog. This system is able to collect and process data automatically and predict the trend of public opinion.

[3] explored recommendation systems based on data from social networks using multiple recommendation algorithms, system functions, different interfaces, filtering techniques, and artificial intelligence

technologies. A location-oriented recommendation system based on the social pertinent trust walker (SPTW) was proposed and compared with the existing baseline random walk models. It successfully enhanced the recommendation SPTW model of group users. [4] Introduced a recommendation system that can be used by travel agencies to assist tourists with their travel itineraries, especially for tourists who have no specific direction or purpose, during which the dialogues in the message box between the agency and consumers were described. A text mining technique was adopted to study one of the blocks. The system then used the database to obtain tourism options such as attractions and cities to provide tour packages tailored to the needs of consumers. Travel agencies would make an itinerary through test approaches using the Rocchio classification algorithm. Then, the ontological specifications (OS) were compared with and without ontological specifications (WOS). The correlation of ontological specifications presented positive results. The system proposed by [5] was a smart space-based recommendation system (e.g. Wikipedia, Wikivoyage, Wikitravel, Panoramio, Flickr, etc.) that may collect tourist attractions on different internet platforms. This system generates rating scores based on the degree of preference and the current status of the attractions. Tourists can rate the attraction based on their own experience, and then the system will divide tourists with the same interests into the same group and use the measurement mechanism that may be applied to this group [1]. The POST-VIA 360 is a platform dedicated to supporting the life-cycle of tourism loyalty generated after the first tour. This system expands to collect information with different methods after the first tour. When analyzing the information, POST-VIA 360 will generate accurate post-travel information, which will then be used to provide relevant recommendations based on the locations and the bionic recommendation systems. To validate this system, the study tested the POST-VIA 360 with specialty items. The results indicated that the accuracy of the system took the lead in the field. The recommendation system in [2] has been frequently used to provide tourism information so as to reduce the amount of explosive information available to consumers. A growing number of recommendation systems intend to provide more abundant tourism experiences, more pluralistic content, context-aware services, browsing, and ratings to users on mobile devices. In mobile computing, wireless networking, web technology, and social networking, there are plenty of opportunities to provide highly accurate and effective travel recommendations, which respect the context parameters of personal preferences, access usability, personality, sociality, and environmentalism all the more. This paper used a systematic approach to review this advanced field and the mobile travel recommendation systems. Also, it respected and emphasized the challenges and prospects faced by studies concerning recommendation systems in the tourism industry.

2.2 Pearson Correlation Coefficient

The Pearson correlation coefficient is often used to test whether two variables are correlated, and it is often used to show the correlation between continuous variables. Hence, Pearson's correlation coefficient formula was used in this study, such as Eq. (1). When r = 1, it indicates that it is of complete correlation; when r is between 0.7 and 0.99, it means a high degree of correlation; between 0.4 and 0.69 indicates a moderate correlation; between 0.1 and 0.39 indicates a low degree of correlation, and between 0.01 and 0.09 means there is no correlation.

$$r = \frac{\Sigma_i (x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\Sigma_i (x_i - \bar{x}) \Sigma_i (y_i - \bar{y})}} \qquad \textbf{(1)}$$

2.3 In-depth Learning Model (Artificial Neural Network)

In-depth learning has become one of the most popular machine learning fields in recent years, with common models including artificial neural networks (ANN) and convolutional neuron networks (CNN). Deep neural networks have better classification and regression ability than shallow networks. In recent years, there have been major breakthroughs in the field of in-depth learning, such as handwriting recognition, image recognition, and speech recognition, and practical applications have also been gradually expanded. For example, the autonomous driving of unmanned vehicles and the AlphaGo AI Go program are the most recent cases of in-depth learning. The principle of in-depth learning is to use neural networks with multi-hidden layers and back propagation to adjust the parameters and fine-tune the model. The process is described in detail as follows. The formula of one neuron is shown in Eq. (2), in which σ represents the activation function, w_i represents weight, and b represents bias:

$$f_{W,b}(x) = P_{W,b}(C_1 \mid x) = \sigma(\Sigma_i w_i x_i + b) \qquad \textbf{(2)}$$

To verify the effectiveness of the model, a loss function should be defined for the model's performance. Then, the loss function is gradually adjusted to a weight, such as Eq. (3), by means of back propagation, where η represents the learning rate, which enables the adjusted weight to reach the lowest point when be substituted into the loss function, indicating that the parameter weight is the local optimum solution:

$$w_i \leftarrow w_i - \eta \Sigma_n - (y^n - f_{W,b}(x^n)) x_i^n \qquad \textbf{(3)}$$

2.4 TensorFlow

TensorFlow [22] is an open source software library used for the machine learning of various perception and language comprehension tasks. In recent years, more and more researchers have used TensorFlow to construct in-depth learning models.

3 Research Method

3.1 System Architecture

The research steps of the computing platform of this study are shown in Figure 1, which is divided into four units: the data set, data preprocessing unit, analysis unit, and database. Details of the steps are described in subsequent sections.

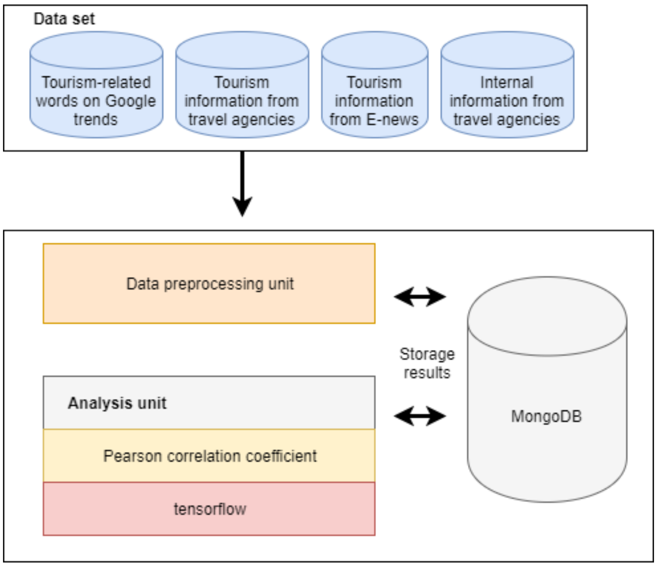

Figure 1. System Architecture

(1) Data set

The main source of information is divided into three types. The first one uses a crawler to collect tourism-related information, popularity scores of tourism-related words on Google Trends, and articles on e-news platforms. The second type is the information published by the government, including the monthly number of Taiwanese individuals traveling to Japan, gender and age, etc. provided by the Taiwan Tourism Bureau website. Detailed information is presented in Table 1. The information collected by the crawler is in disorder, which entails pre-processing. Further explanation to relevant processing programs is given in the data preprocessing section. The third type is the information concerning whether Hokkaido tours are confirmed, as well as the Google Trends information related to tourism key words about Hokkaido and internal details from travel agencies, which are shown in Table 2.

Table 1. Research Data Sources

Tourism-related words on Google Trends	In terms of the popularity scores of tourism-related words on Google Trends, this study adopted a crawler to collect information concerning the popularity scores of tourism-related words on Google Trends during the period from January 2012 to April 2017
Tourism information from E-news sources	Using tourism information from E-news sources, this study employed a crawler to collect information during the period from September 2015 to April 2017, including dates, titles and news content.
Number of Taiwanese individuals traveling to Japan, provided by the Taiwan Tourism Bureau website	The number of Taiwanese individuals traveling to Japan is summed up on a monthly basis. This study collected information during the period from January 2012 to April 2017.

Table 2. Information about Hokkaido on Google Trends and Internal Information from Travel Agencies

Tourism-related words on Google Trends	This study adopted a crawler to collect information concerning the popularity scores of tourism-related words about Hokkaido during the period from January 2012 to April 2016
Tour information from travel agencies	Information during the period from January 2016 to December 2016 included 8,000 items concerning years of departure, tour fees, and whether the tours reached the minimum number of participants.

(2) Data preprocessing (Preprocessor)

There are four main projects in the pre-processing part:

(a) Information concerning tourism-related words on Google Trends

Google Trends [23] is a webpage used for searching the popularity scores of tourism-related words, which are summed up on a weekly basis as shown in Figure 2. The weekly scores are converted to monthly scores via Eq. 4, where S_m represents the average popularity scores of tourism-related words, S_W represents the weekly popularity scores of tourism-related words, and n represents the number of data items.

$$S_m = \frac{\Sigma s_W}{n} \quad \textbf{(4)}$$

In a study that predicted the number of Taiwanese tourists traveling to Japan, “Japan” was taken as the keyword by the open source software pytrends [24] to collect information. Then, pytrends was used to sum up the popular words concerning travel in Japan from

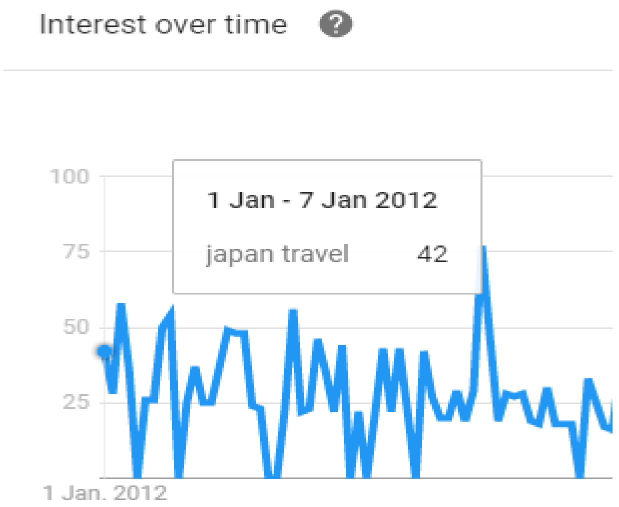

Figure 2. Popularity Scores of Tourism-related Words on Google Trends

2012 to 2016, which are listed in Table 3. In the study that predicted whether tours to Hokkaido could reach the minimum requirement for participants, the keyword was set as "Taiwan" and then pytrends was used to sum up popular words related to travel in Hokkaido during the period from 2014 to 2016, which are listed in Table 4.

Table 3. Tourism-related Key Words for Prediction of the Number of Taiwanese Individuals Traveling to Japan

Key Words	Related Key Words
Japan	"Japan Must-Buy", "Japan Tokyo", "Japan Travel", "Japan Osaka", "Japan Airlines", "Japan Weather", "Japan Map", "Japan Food", "Japan Free Trip", "Japan Kyoto", "Japan Hokkaido", "Japan Meteorology", "Japan Exchange Rate", "Japan Okinawa", "Japan Kyushu"
Japan Travel agencies	"Japan Travel Agency Travel", "Japan Travel Agency Recommendation"

Table 4. Key Words for Prediction of whether Hokkaido Tours Will Reach Minimum Requirement for Participants

Key Words	Related Key Words
Hokkaido	"Hokkaido Must-Buy", "Japan Hokkaido", "Hokkaido Travel", "Hokkaido Weather", "Hokkaido Map", "Hokkaido Food", "Hokkaido Free Trip"
Hokkaido Travel agencies	"Hokkaido Travel Agency", "Hokkaido Travel Agency Recommendation"

(b) Collect information from e-news webpages

The main steps for collecting information from e-news webpages are as follows:

1. Segment words and sentences in the titles and content of e-news sources through the taiba word and sentence segmentation system, and store the results in the MongoDB Database.

2. Write a pre-processing program through Python and sum up the number of Japan-related words in each month. Japan-related words include counties and cities in Japan. The source of information concerning counties and cities in Japan is the Wikitravel webpage collected by the crawler.

(c) Number of Taiwanese individuals traveling to Japan provided by Taiwan Tourism Bureau

The number of Taiwanese individuals traveling to Japan, gender proportion, age proportion, and so on collected by Taiwan Tourism Bureau are saved in the MongoDB for further analysis in this study.

(d) Internal tour information provided by travel agencies

The internal tour information provided by travel agencies in Taiwan are stored in MongoDB, and then Python is used to write a program to collect information concerning Hokkaido tours, using the names of tours as digital vectors. The suite gensim and jieba in Word2vec are used to convert the words into vectors when the model is completely trained [21].

(3) Pearson correlation coefficient calculation unit

Based on previous studies [8], the Pearson correlation coefficient model adopted the lag model to compare the popularity scores of tourism-related keywords on Google to the number of tourists after n months, and then calculate the correlation coefficient between these two.

This study calculated the correlation between the popularity scores of tourism-related keywords about Japan on Google and the number of tourists traveling to Japan after n months. x_i in (1) represents the popularity scores of tourism-related keywords of Japan, $\bar{x}$ represents the average popularity scores of tourism-related keywords of Japan, y_i represents the number of Taiwanese tourists traveling to Japan, and $\bar{y}$ represents the average number of Taiwanese individuals traveling to Japan. This model was also used to verify whether the information from e-news sources was related to the number of tourists.

(4) ANN unit

(a) Study on the number of Taiwanese individuals traveling to Japan

The model architecture that predicts the number of Taiwanese individuals traveling to Japan is shown in Figure 3. The data input and output formats are shown in Table 6, with the model input being the monthly popularity scores of tourism-related words on Google Trends, and the model output being number of tourists traveling to Japan n months later. k represents the number of Taiwanese tourists traveling to Japan and is classified based on the following criteria:

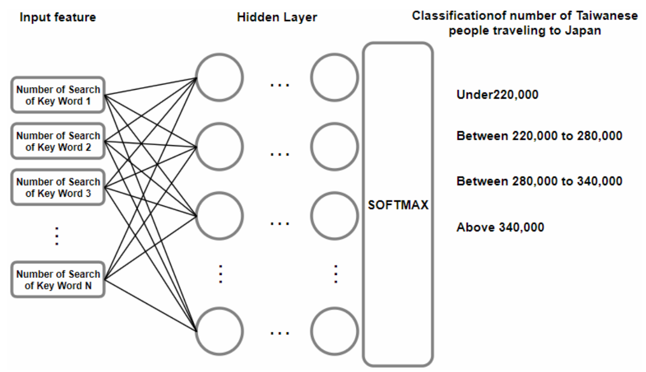

Figure 3. ANN Model for Predicting the Number of Taiwanese People Traveling to Japan

k<=220000	Classification 1
220000=k<=280000	Classification 2
280000<=k<340000	Classification 3
k>=340000	Classification 4

The ANN model and activation function was set up through Google TensorFlow. In the setting of ReLU and TensorFlow, the loss function is set as cross entropy. Then, the Adam function is used to speed up the convergence rate of the model. There are two hidden layers, with 12 neurons in each layer.

(b) Study on whether Hokkaido tours will reach the minimum requirement for participants

The ANN model architecture for predicting whether Hokkaido tours will reach minimum requirement for participants is shown in Figure 4, with the model input being the tour information from travel agencies and the popularity scores of tourism-related words of Hokkaido on Google Trends, and the model output being whether the tours will reach the minimum requirement for participants. Its neural network is set as Table 5.

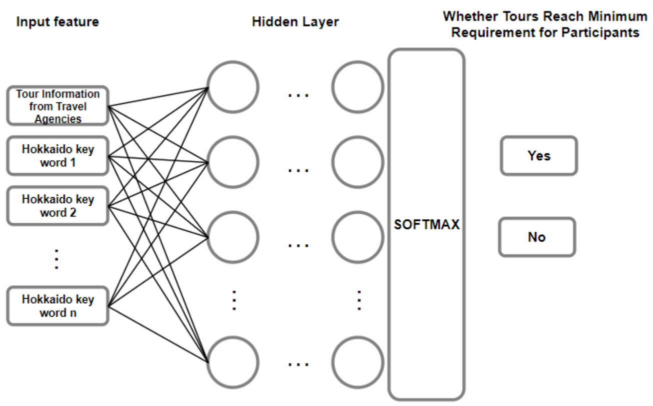

Figure 4. ANN Model Architecture for Predicting whether Hokkaido Tours Will Reach the Minimum Requirement for Participants

4 Experiment

The experimental environment was an Intel Core i5-6500 with a memory of 8GB.

Table 5. Relevant settings of the neural network

Units of Neural Network	Names
Activation function	ReLU
Loss function	Cross Entropy
Convergence algorithm	Adam
Hidden layers	3
Number of neurons hidden in the first layer	100
Number of neurons hidden in the second layer	50
Number of neurons hidden in the third layer	10

4.1 Pearson Correlation Coefficient Experiment

This study adopted the popularity scores of tourism-related words on Google Trends and information from e-news platforms during the period from September 2015 to April 2017. The threshold of the correlation coefficient was set at 0.5. The correlation coefficient result of each tourism keyword is listed in Table 7. The results indicated that these tourism keywords were related to the number of Taiwanese individuals traveling to Japan. For example, in the one-month lag experiment, the Japan exchange rate showed a positive correlation with the number of Taiwanese tourists traveling to Japan, which suggested that if the number of searches on the Japan exchange rate in n months increased, the number of Taiwanese individuals traveling to Japan in n +1 month would also increase.

Information from e-news sources after word and sentence segmentation was used to calculate the correlation between the popular words related to Japan during each month and the number of Taiwanese tourists traveling to Japan in the future n months, with the threshold of the correlation coefficient r set as 0.5. The experimental result revealed that the correlation coefficient between the popular words related to Japan in each month and the number of Taiwanese individuals traveling to Japan was less than 0.5, which indicated that the correlation is low.

4.2 Experiment preDicting Number of Taiwanese Tourists Traveling to Japan

The data were sourced from the popularity scores of tourism-related words on Google Trends in each month from 2012 to 2016, with the ANN output being the number of Taiwanese individuals traveling to Japan n months later. Of the data, 80% were training data, and the other 20% were test data. The accuracy of the training data was over 95%. The experimental results of the test data are shown in Table 8. The results indicated that the classification of the number of Taiwanese tourists traveling to Japan in the following month was conducted based on the popularity scores of tourism-related words on Google Trends, and the optimum result of the mean square error was 0.13.

Table 6. Data of ANN Model

Date	Popularity Scores of Number of Search of "Japan"	Popularity Scores of Number of Search of "Japan Travel"	Popularity Scores of Number of Search of "Japan Exchange Rate"	Classification of Number of Tourists Traveling to Japan
2012/1	51.6	26.8	15.2	1
2012/2	49.5	29.25	16.75	1
2012/3	54.25	31.25	26.25	1
2012/4	51.8	30	19	1

Table 7. Correlation between Key Words on Google Trends and Number of Tourists Traveling to Japan

Tourism-related Key Words on Google Trends	Pearson Correlation Coefficient	Lag Month
"Japan Food"	0.5	0
"Japan Exchange Rate"	0.6	1
"Japan Tokyo"	0.54	3
"Japan Travel"	0.58	3
"Japan Airlines"	0.72	3
"Japan Air Tickets"	0.65	3
"Japan Air Tickets"	0.57	3
"Japan Airlines"	0.72	4
"Japan Osaka"	0.52	4
"Japan Exchange Rate"	0.61	4
"Japan Air Tickets"	0.68	4
"Japan Okinawa"	0.61	5
"Okinawa Air Tickets"	0.67	6

Table 8. MSE of ANN Model

Future n Months	MSE
n=0	0.28
n=1	0.13
N=2	0.14

4.3 Experiment Predicting Whether Hokkaido Tours Would Reach the Minimum Requirement for Participants

The data were sourced from the internal information of travel agencies from 2014 to 2016, with the ANN input being the popularity scores of tourism-related words about Hokkaido on Google Trends. The internal information of travel agencies included years of department, months of departure, names of tours, and tour fees. The ANN output was whether the tours reached the minimum number of participants. Of the data, 80% were training data, and the other 20% were test data. The accuracy of the training data was 70% and the experimental results of the test data were 68%, which indicated that the model could effectively predict whether the tours would reach the minimum requirement for participants.

5 Conclusion

The first topic of this study was to predict the number of Taiwanese individuals traveling to Japan using the popularity of tourism-related words about Japan on Google Trends. This study found that the popularity scores of tourism-related words about Japan on Google could accurately predict the number of Taiwanese tourists traveling to Japan in the next three or four months. On the contrary, information from e-news sources showed a low degree of correlation, indicating that the number of tourists would not increase due to the related reports. Hence, the ANN model was used to predict and analyze the multiple popularity scores of tourism-related words about Japan on Google. The results showed that classification of the number of Taiwanese individuals traveling to Japan had the highest accuracy in the following month. The second topic was to use the internal information from Taiwan travel agencies as well as information on Google Trends to predict the number of Taiwanese individuals traveling to Japan. The results indicated that the accuracy of the test data for Hokkaido tours reached 68% through the ANN model and could predict whether the tours would reach the minimum requirement for participants. The experimental results could be used to enhance the arrangements of the tourism industry and facilitate people's travel itineraries.

This study found that tourism-related data on Google may help to predict the number of tourists going to relevant travel destinations in the future, and the results could be provided as a reference to the tourism industry for planning itineraries. Future studies may combine Google Trends with internal data from travel agencies to conduct in-depth analysis on tourists, gender proportion, age, area and preferences for the destinations, which may allow travel agencies to make more precise judgments for tours.

Reference

[1] R. Colomo-Palacios, F. J. García-Peñalvo, V. Stantchev, S. Misra, Towards a Social and Context-aware Mobile Recommendation System for Tourism, *Pervasive and Mobile Computing*, Vol. 38, pp. 505-515, July, 2017.

[2] D. Gavalas, C. Konstantopoulos, K. Mastakas, G. Pantziou, Mobile Recommender Systems in Tourism, *Journal of Network and Computer Applications*, Vol. 39, pp. 319-333, March, 2014.

[3] L. Ravi, S Vairavasundaram, A Collaborative Location Based Travel Recommendation System through Enhanced Rating Prediction for the Group of Users, *Computational Intelligence and Neuroscience*, Vol. 2016, No. 7, March, 2016.

[4] C. Devasanthiya, S. Vigneshwari, J. Vivek, An Enhanced Tourism Recommendation System with Relevancy Feedback Mechanism and Ontological Specifications, *Advances in Intelligent Systems and Computing*, Vol. 398, December, 2015.

[5] A. Smirnov, A. Kashevnik, A. Ponomarev, N. Teslya, M. Shchekotov, S. I. Balandin, Smart Space-Based Tourist Recommendation System, *Lecture Notes in Computer Science*, Vol. 8638, 2014.

[6] H. Q. Vu, G. Li, R. Law, B. Haobin Ye, Exploring the Travel Behaviors of Inbound Tourists to Hong Kong Using Geotagged Photos, *Tourism Management*, Vol. 46, pp. 222-232, February, 2015.

[7] J. Shen, J. Shen, T. Mei, X. Gao, Landmark Reranking for Smart Travel Guide Systems by Combining and Analyzing Diverse Media, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 46, No. 11, pp. 1492-1504, 2016.

[8] X. Yang, B. Pan, J. A. Evans, B.Lv, Forecasting Chinese Tourist Volume with Search Engine Data, *Tourism Management*, Vol. 46, pp. 386-397, February, 2015.

[9] B. Pan, D. C. Wu, H. Song, Forecasting Hotel Room Demand Using Search Engine Data, *Journal of Hospitality and Tourism Technology*, Vol. 3, No. 3, pp. 196-210, 2012.

[10] P. F. Bangwayo-Skeete, R. W. Skeete, Can Google Data Improve the Forecasting Performance of Tourist Arrivals? Mixed-data Sampling Approach, *Tourism Management*, Vol. 46, pp. 454-464, February, 2015.

[11] X. Li, B. Pan, R. Law, X. Huang, Forecasting Tourism Demand with Composite Search Index, *Tourism Management*, Vol. 59, pp. 57-66, April, 2017.

[12] L. Liu, N. Mehandjiev, D.-L. Xu, Multi-criteria Service Recommendation Based on User Criteria Preferences, *Proceedings of the fifth ACM conference on recommender systems*, Chicago, IL, 2011, pp. 77-84.

[13] F. Santos, A. Almeida, C. Martins, P. Oliveira, R. Gonçalves, Tourism Recommendation System Based in User's Profile and Functionality Levels, *Proceedings of the Ninth International C* Conference on Computer Science and Software Engineering (C3S2E 2016)*, New York, NY, 2016, pp. 93-97.

[14] L. Etaati, D. Sundaram, Adaptive Tourist Recommendation System: Conceptual Frameworks and Implementation, *Vietnam J. Computer. Sci.*, No. 2, pp. 95-107, 2015.

[15] M. Frikha, M. Mhiri, F. Gargouri, Designing a User Interest Ontology-driven Social Recommender System: Application for Tunisian Tourism, in: J. Bajo et al. (Eds.), *Trends in Prac. Appl. of Agents, Multi-Agent Sys. and Sustainability, Advances in Intelligent Systems and Computing 372*, Springer, 2015, pp. 159-166, DOI:10.1007/978-3-319-19629-9_18.

[16] M. Hassan, M. Hamada, A Neural Networks Approach for Improving the Accuracy of Multi-Criteria Recommender Systems, *Appl. Sci.*, Vol. 7, No. 9, p. 868, August, 2017.

[17] X. Liu, M. Li, S. Qin, X. Ma, W. Wang, *A* Predictive Fault Diagnose Method *of* Wind Turbine Based *on* K-Means Clustering *and* Neural Networks, *JIT,* Vol. 17, No. 7, pp. 1521-1528, December, 2016.

[18] W. Wu, W.-C. Du, H.-Z. Xu, H. Zhou, M.-X. Huang, Popularity Prediction of Online News Based on Radial Basis Function Neural Networks with Factor Methodology, *JIT*, Vol. 17, No. 5, pp. 915-927, September, 2016.

[19] H.-T. Li, Human Action Recognition Using Geometric Parameter Features and a Radial Basis Function Neural Network, *JIT*, Vol. 17, No. 2, pp. 255-261, March, 2016.

[20] N.-Y. Chen, Y. Liu, Z.-J. Zhang, A Forecasting System of Micro-Blog Public Opinion Based on Artificial Neural Network, *2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing,* Kitakyushu, Japan, 2014.

[21] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Efficient Estimation of Word Representations in Vector Space, *Advances in Neural Information Processing Systems*, arXiv. 1301.3781, 2013.

[22] https://www.tensorflow.org/.

[23] https://trends.google.com/trends/.

[24] https://github.com/GeneralMills/pytrends.

Biographies

Jui-Hung Chang is currently a full-time Professor with the Computer and Network Center, Department of Computer Science Information Engineering, National Chen Kung University. Her current research interests include data base management, digital learning system, system analysis and object oriented design, program design, data mining, cloud calculation application, and project management and planning. Furthermore, she is also interested in active participation in related travel information prediction and recommendation, geographical system, agricultural statistical big data, biological big data mining with industry academy cooperation plan and research projects.

Chien-Yuan Tseng received the degree in computer science and information engineering from National Cheng Kung University. His research areas include the populous neural network, Spark, Hadoop distributed file system, and the programming of websites. He is immersed in calculating and analyzing big data and has published several conference papers.