

# 利用網路訊息推測遊客量之研究

## - 墾丁國家公園為例

蔡柏偉<sup>\*</sup>、施孟隆<sup>\*\*</sup>

### 摘要

網路平台提供了消費者一個表達商品心得與交流的地方。愈來愈多的消費者在從事消費行為前，會先上網搜尋的使用心得作為參考，因而網路訊息的影響力也與日俱增。網路口碑具匿名性且龐大資料來源的特質，經常被用來對消費市場做分析預測。而旅遊市場常會受到網路訊息、氣候環境，乃至經濟發展等多種因素的影響。因此網路資訊相當適合運用來探討觀光旅遊業中的遊客的真實感受。

本研究利用資料探勘方法，蒐集 2009 年 1 月至 2013 年 12 月止所有公開討論墾丁國家公園之網誌文章，共 11,745 篇，經過情緒特徵詞選取，萃取網站討論者之情緒，量化轉換成情緒分數，藉以獲取墾丁國家公園之網路情緒指標。並進一步加入谷歌搜尋熱門度和經濟指標等因素作為投入變數，透過類神經網路 (Neural Network) 之倒傳遞網路 (BNP)，建構不同變數類型組合之預測模型，進行驗證與預測準確度評估。

研究結果發現：1. 網路情緒指標、谷歌搜尋熱門度及經濟指標確實可利用於預測遊客量；2. 以經濟指標結合網路情緒指標或谷歌搜尋熱門度所建構之預測模型均能達到良好預測能力；3. 將網路情緒指標和谷歌熱門度平移 3 個月的預測能力最佳，MAPE 值達 14.36%，說明網路情緒對墾丁國家公園旅遊遊客量的影響時間為 3 個月。本文以客觀、創新的預測方法，運用即時的網路資訊，實務上可以提供給相關單位作為資源分配與計畫的參考。

關鍵字：谷歌搜尋熱度；情緒指標；倒傳遞網路

關鍵字：搜尋熱度；情緒指標；倒傳遞網路

---

\* 國立台東大學碩士。

\*\* 國立台東大學文化資源與休閒產業學系教授，通訊作者。

## 壹、緒論

網路平台提供了消費者一個表達商品心得與交流的地方。愈來愈多的消費者在從事消費行為前，會先上網搜尋的使用心得作為參考，因而網路訊息的影響力也與日俱增。網路資訊具匿名性且龐大資料來源的特質，經常被用來對消費市場做分析預測。而旅遊市場常會受到網路資訊、氣候環境，乃至經濟發展等多種因素的影響。因此網路資訊相當適合運用來探討觀光旅遊業中的遊客的真實感受。

市場溫度乃一比喻性的名詞，如同社會經濟景氣溫度，通常當社會呈現經濟成長、民眾對於未來經濟一片看好之時，通常稱之為經濟景氣熱絡。相對地，對於旅遊市場而言，「人潮即錢潮」，造訪的遊客量多寡通常也顯示該旅遊市場熱絡的程度。而準確的分析旅遊市場波動的要害，將有助於決策者做出正確的判斷，特別是遊客量的分析。對決策者而言，遊客量更是管理觀光旅遊景點相當重要的基礎資料，可用於掌握遊客動向、瞭解遊客需求、遊客旅遊情形等，並提供直接資訊，供觀光旅遊景點管理者對於未來景點規劃設計、經營管理策略發展上之參考。此外，亦可以遊客量資訊評量遊客使用觀光旅遊景點資源之旅遊供需狀況，以確保資源永續利用及旅遊品質提升，故觀光旅遊景點業者或決策者若能準確預測遊客量，則可達成控制成本、降低風險及創造利潤之目的。

隨著資訊科技進步、社群網路普級，以及旅遊業的蓬勃發展，網際網路在旅遊行業中得到廣泛應用，網際網路充斥著大量有用的資訊，對於業者是一項好消息，諸如消費者藉由網路論壇、討論區和留言板等平台分享自己旅遊後的經驗與意見，業者可利用有價值資訊(如消費者之網路評價)，作為改善經營管理決策之指導方向。

網路訊息(包括電影評論、新聞或者是口碑(Word of Mouth; WOM))已廣泛運用在各領域研究中，在產業領域方面，有些學者將電影評論或口碑加入為預測的指標(Brewer et al., 2009; Joshi et al., 2010);在社會領域方面，Yang et al (2011)利用互聯網搜索量數據研究自殺關聯；施孟隆(2014)利用網路訊息推測旅遊溫泉度假遊客量之研究。另外Smith(2012)利用互聯網搜索量數據預測外匯之波動。根據前述研究顯示，網路訊息與現實的社會行為存在有一定的關聯，網路訊息對顧客決策具有導引作用，遊客的旅遊行為具有前兆效應。

國內外學者對於經濟指標與旅遊關研究很多，它們僅採用經濟指標對旅遊需求進行推估與預測(Law and Au, 1999; Gounopoulos, Petmezas and Santamaria 2012)，這些研究採用的經濟指標包括實質國內生產毛額、消費者物價指數、匯率、失業率..等，另外，Song, Gao, 與 Lin(2013)採用

網路資訊與經濟數據對中國大陸、台灣、日本、澳大利亞、英國與美國，進行遊客量之預測，研究結果表明，結合網路資訊與經濟數據之模式優於經濟數據之模式。因此，本文擬分別採用網路搜索資訊與經濟指標，建立遊客量預測模型，分析遊客量預測模型之良窳，歸納結果作為旅遊業之經營者，未來規劃設計與經營管理策略發展之參考；並提供研究者進行遊客量之預測時，採用預測變數指標的另類思考方向。

## 貳、文獻回顧

### 一、墾丁國家公園選定與遊客量定義

#### (一) 墾丁國家公園選定與遊客量概況

目前台灣有九座國家公園，分別是墾丁國家公園、玉山國家公園、陽明山國家公園、太魯閣國家公園、雪霸國家公園、金門國家公園、海洋國家公園，台江國家公園與壽山國家公園等。表1說明民國98年至102年期間，各國家公園遊客人數統計：

表1 國家公園遊客人數統計表

年度	98	99	100	101	102	比例
墾丁	4,938,421	6,531,302	6,341,160	7,008,107	7,050,085	32.2%
太魯閣	6,501,603	3,702,083	3,687,050	4,819,139	4,776,482	24.4%
陽明山	4,205,795	3,779,349	3,367,445	3,625,756	4,087,216	19.8%
金門	806,657	1,020,861	1,319,263	1,479,779	1,389,216	6.3%
玉山	1,428,731	718,879	973,821	1,002,090	1,121,303	5.5%
雪霸	794,766	829,981	916,095	826,703	1,343,443	4.9%
壽山					3,870,000	4.0%
台江		247,979	867,680	808481	823,972	2.9%
海洋	12,190	9,526	9,657	7,325	14,294	0.1%

資料來源：參考內政部營建署國家公園網站<http://np.cpami.gov.tw>整理

觀光局統計中發現，從2011年到2012年，平均有93.8%的國人曾在國內旅遊，

每人一年平均旅遊7.5次，其中國家公園兩年的旅遊人次共計34,417,781人，佔全臺主要遊憩區旅遊人數的7%，可見國家公園旅遊是國人選擇休閒遊憩的重要景點之一。如表1所示，在各國家公園遊客人數統計中，以墾丁國家公園居首位，太魯閣國家公園次之、陽明山國家公園排名第三，因此本研究選定墾丁國家公園作為研究對象。

墾丁國家公園管理處於1982年9月公告成立，是我國第一座成立的國家公園，為我國少數同時涵蓋陸域與海域的國家公園之一。園區三面臨海，海域面積15206.09公頃、陸域面積18083.50公頃，合計共33,289.59公頃，行政轄區包括恆春鎮、滿洲鄉之大部分及車城鄉部分等三鄉鎮。依據國家公園法及生態資源特性，將區內土地劃設為生態保護區、特別景觀區、史蹟保存區、遊憩區及一般管制區等五種分區。

位於台灣最南端的墾丁國家公園屬熱帶性氣候，氣溫終年暖和，擁有熱帶原始林、稀有植物、野生動物群集珍貴，並為候鳥遷移之路線。由於百萬年來地殼運動不斷的作用，陸地與海洋彼此交蝕影響，造就本區高位珊瑚礁及海蝕地形景觀，實為優異之熱帶生態資源之景觀地區。然而近年來，每年約有五百萬人次的遊客進入，觀光遊憩及商業等行為已經對當地生態環境造成衝擊，並產生種種負面影響，也成為該國家公園經營管理上之最大挑戰。園區內共設有鵝鑾鼻公園、貓鼻頭公園、後壁湖遊艇港、南仁山等四個管理站以及龍坑管制站、砂島、瓊麻展示館、龍鸞潭賞鳥館、南灣遊憩區、關山觀日落等六個據點共有十個以上景觀點，實為一多元化且為我國唯一擁有海域珊瑚礁天然景觀的國家公園（墾丁國家公園管理處，2013）。

## 二、遊客量之推估、預測與統計技術的發展

基於遊客量的掌握的重要性，準確遊客量的預測是研擬經營計畫之基礎，有助於經營管理者對於遊憩設施、旅遊相關設施與服務需要之配置，提供高品質的服務，讓遊客擁有具深刻的旅遊經驗。

Styness(1983)歸納遊客量預測方法包括德爾菲法(Delphi Technique)、時間數列法(Time series)、結構性模型(structural models)與系統或模擬模型(system or simulation models)（林寶秀、林晏州，2010）。

Witt and witt(1995)於旅遊需求實證研究回顧文獻表明，從1966~1992研究期間，有19篇研究依變數採用遊客量，其採用實證方法包括線性迴歸(linear model)、非線性迴歸(log-linear model)、羅吉斯法(logit model)和probit方法等，各種研究方法擁有優點與缺點。

Law 與 Au(1999) 利用服務價格平均酒店價格外匯匯率人口行銷費用及國內生產毛額等預測日本遊客往訪香港的遊客數，研究證實類神經網路優於指數平滑法、複迴歸法與移動平均法。

Cho(2003)利用指數平滑法、ARIMA 與類神經網路，運用於從六個國家(美國、英國、新加坡、日本、台灣和韓國)，預測來訪香港遊客數，研究結果發現，除英國外，類神經網路優於指數平滑法、ARIMA。

Medeiros et al (2008)利用類神經網路預測巴利阿里群島，西班牙的國際旅遊的需求，研究結果可隨時間變化的差異提供遊客數。

Song 與 Li(2008)回顧 2000~2007 年間刊登於社會科學期刊文獻資料庫(Social Science Citation Index ;SSIC)有關旅遊需求之文章共 121 篇與包括支持向量迴歸法(Support vector regression ;SVR )、類神經網路(Artificial neural network ; ANN)等 45 種研究方法。研究結果表明沒有任何一種方法，在所有情況下擁有最佳的預測效果。

Gounopoulos, Petmezas 與 Santamaria(2012) 利用指數平滑法、ARIMA 預測希臘的遊客量，並探討對總體經濟的衝擊。研究結果表明 ARIMA 優於指數平滑法。

Song; Gao 與 Lin(2013)採用旅遊資訊預測系統對中國大陸、台灣、日本、澳大利亞、英國與美國，進行遊客量之預測，研究結果表明，結合網路資訊與經濟數據之模式優於經濟數據之模式。

綜合前述，研究結果表明沒有任何一種方法，在所有情況下擁有最佳的預測效果。因此，本研究擬倒傳類神經網路作為預測遊客量之方法。

### 三、谷歌趨勢搜尋(Google Insight for Search)之相關運用

谷歌(google)是目前最流行的世界互聯網搜索引擎，有超過 65% 的市場份額。在全球各地每月六十億搜索查詢，每年為谷歌賺進 250 億美元 (Smith, 2012)。谷歌搜尋透視，由搜尋趨勢(Google Trends)演變而來，研究人員可以輸入相關字詞，找出網路使用者最感興趣的詞彙，如果同一詞彙在一段時間內的熱門度提高，代表搜尋量提高，網路使用者目前對於此詞彙高度的關注且感興趣，意味著該字詞所代表的產品或事物未來可能有某種變化(上升或下降)的趨勢(林彥佐，邱昭彰，2012)。Google 搜尋透視使用人很多，涉及範圍廣泛，包括財經活動，社會、醫療行為等，Ginsberg et al. (2009)利用 Google 搜尋資訊，進行流感疫情的預測，研究結果表明可以正確預測美國各地區每週的流感疫情。Vosen 與 Schmidt(2011)利用搜尋趨勢搜索熱門度，結合三個總體經濟指標，與傳統調指標進行預測私人消費的比較研究，結果表明谷歌趨勢搜索熱門度的指標在訓練及測試的樣本其預測力優於傳統的指標；施孟隆(2014) 利用谷歌趨勢搜索熱門度，結合三個總體經濟指標情緒指標推測旅遊溫泉度假遊客量，結果表明

其預測力優於傳統的經濟指標。Smith(2012)針對特定關鍵字谷歌互聯網搜索數量變化進行外匯市場波動性預測，利用經濟危機、金融危機和經濟衰退等關鍵詞，增加 GARCH (1, 1) 之預測能力，研究結果也表明谷歌搜索量資訊之訊息是金融市場的寶庫。Frijters, Johnston, Lordan 與 Shields(2013)利用互聯網的搜索資訊研究宏觀經濟條件和問題飲酒的關係。

#### 四、總體經濟指標之選取

學者在相關研究採用的總體經濟指標，由於研究的對象不同，其採用的總體經濟指標亦不同，如 Song, Gao, 與 Lin(2013) 針對遊客量之預測時，採用國內生產毛額、消費者物價指數與替代價格等經濟指標。Smith(2012)針對特定關鍵字谷歌互聯網搜索數量變化進行外匯市場波動性預測研究。採用失業率，股票交易量，私人消費支出，住房價格和股票收益的波動性等五個經濟指標。Vosen 與 Schmidt(2011) 於預測私人消費研究採用實際個人所得、三個月的國庫券利率以及股價指數等三個總體經濟指標(林彥佐，邱昭彰，2012)。Gounopoulos, Petmezas 與 Santamaria(2012) 採用實質國內生產毛額、消費者物價指數、匯率、失業率經濟指標，預測希臘的遊客量，並探討對總體經濟的衝擊。施孟隆(2014)採用消費者物價指數、台幣兌美元、股價指數、失業率、國內生產毛額等總體經濟指標推測旅遊溫泉度假遊客量。根據前述研究為基礎，但考慮本研究中所探討之墾丁國家公園遊客仍以國內遊客為主，故不把匯率指標納入討論。綜上所述，本研究擬選擇國內生產毛額、消費者物價指數、貨幣供給額、失業率及股價指數等五項指標，作為經濟指標變數。

### 參、研究設計與方法

#### 一、研究架構

在傳統的研究中，在遊客量預測上，通常是以總體經濟指標進行分析預測，本研究是結合搜尋熱度、情緒指標與總體經濟指標，預測未來墾丁國家公園的遊客量。研究架構，如圖 1，主要分成三個部分，第一、進行文章及指標資料蒐集，第二、資料前處理，第三、建立預測模型，進行預測模型的實驗並做模型的評估，以確認預測模型的可行性。

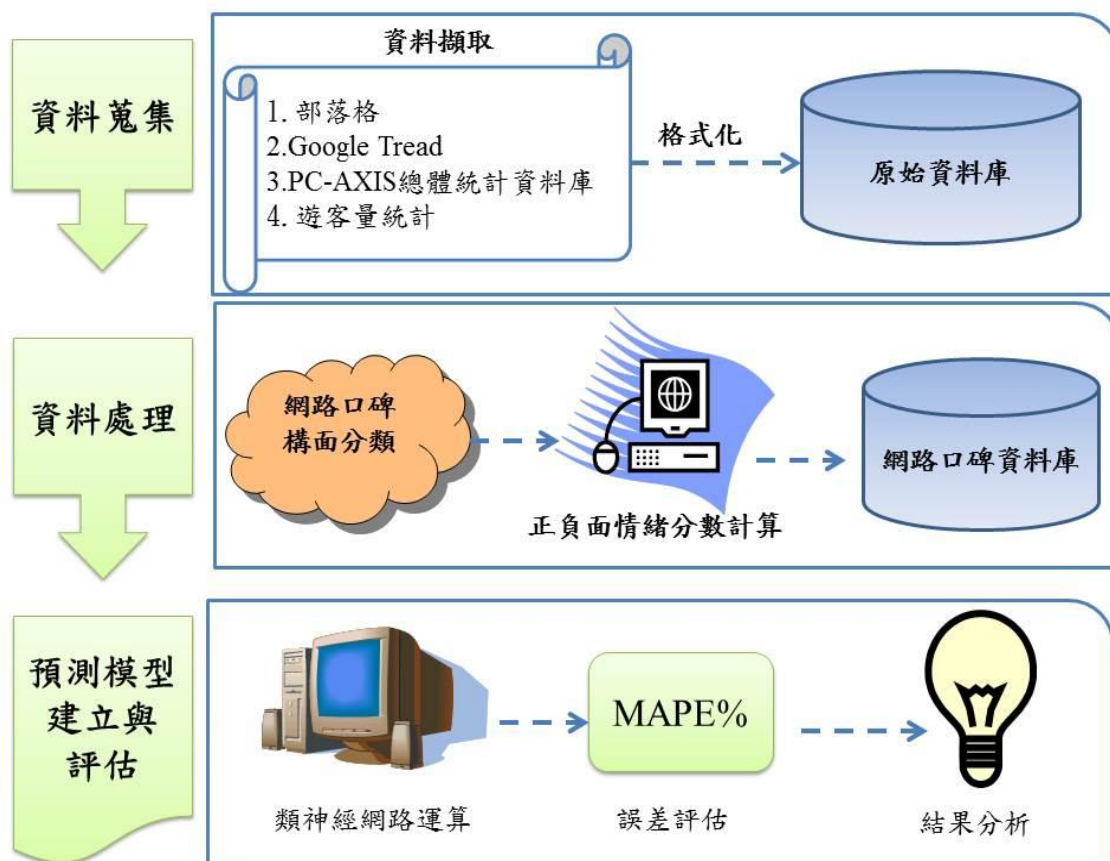


圖 1 研究架構圖

## 二、資料蒐集與處理

為收集網路口碑情緒資料，本研究利用谷歌搜尋輸入關鍵字蒐集與墾丁相關的網誌，並在谷歌搜尋時間設定選項中，選擇以年為單位呈現搜尋結果，以便資料分類。透過Crawler工具抓取特定網頁之討論內容，收集2009年1月至2013年12月止所有公開討論文章，共11,745篇。由於並非所有的網頁內容皆為所需資訊(如：廣告、圖片、...)。因此還須針對該討論谷歌的HTML網頁格式撰寫程式，以將所需之內容從網頁原始碼中分離出來。將所得之純文字資訊存入資料庫以做下一階段的特徵詞處理。

### (一) 谷歌搜尋熱門度

根據創市際公司在2010年的調查，谷歌搜尋的頁數已超越所有其他的搜尋網站，同時也成為台灣網友停留時間最長的搜尋引擎。本研究透過谷歌搜尋趨勢取得"墾丁"在2009年至2013年的關鍵字搜尋熱度，詳見圖2。

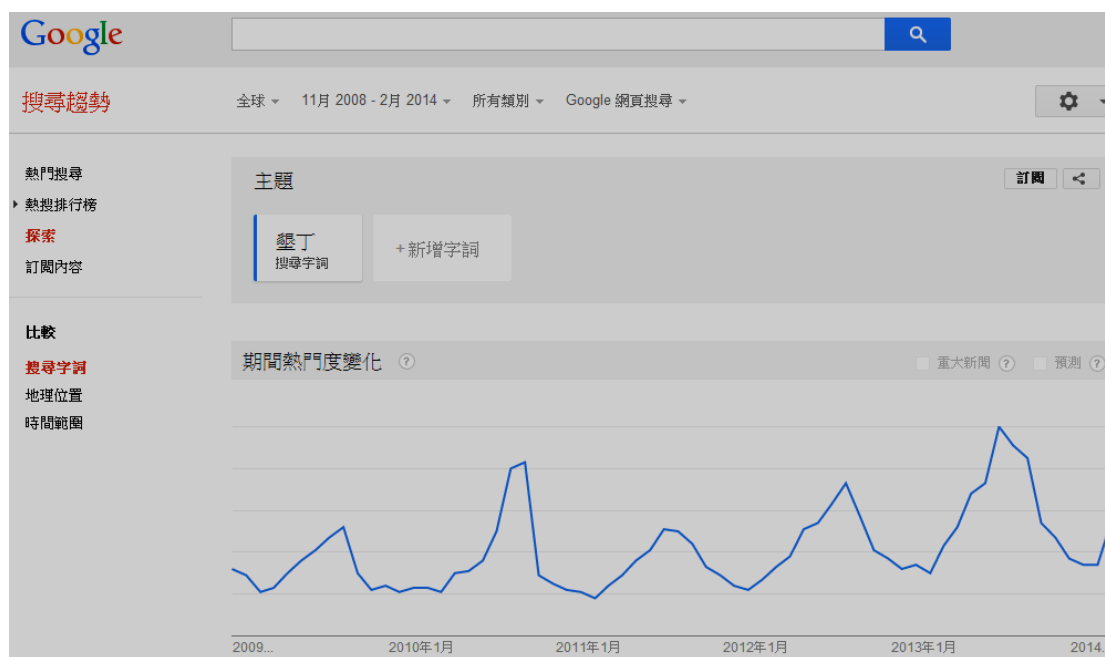


圖2 谷歌搜尋趨勢關鍵字熱門度

資料來源：谷歌搜尋趨勢網站擷取

由於從谷歌搜尋趨勢取得的是關鍵字每週熱門度，因此須將週資料轉換成月資料，統計出以"墾丁"為關鍵字的每月搜尋熱門度次數，以利後續資料處理分析。下表2為墾丁國家公園，2013年1月到12月的搜尋熱門度，

表 2 谷歌搜尋趨勢網路搜尋熱門度次數

日期	Jan-13	Feb-13	Mar-13	Apr-13	May-13	Jun-13
搜尋次數	131	139	205	274	277	377
日期	Jul-13	Aug-13	Sep-13	Oct-13	Nov-13	Dec-13
搜尋次數	350	288	202	162	172	202

資料來源：本研究整理



## (二)網路情緒分數轉換處理

### 1. 特徵詞

此步驟是透過對墾丁國家公園遊客旅遊經驗相關之情緒特徵詞的篩選，作為每份文件共同的特徵屬性。前一階段保留的字詞經過詞頻的計算後，保留詞頻值較高的特徵詞，彙整如表 3；

表 3 高詞頻之特徵詞

---

經選取之特徵詞：

---

美味，讚，棒，滿意，有趣，好玩，快樂，幸福，貼心，舒服，開心，加分，爽，乾淨，遺憾，溫暖，倒楣，大推，扣分，難吃，爛，糟，失望，無聊，不爽，髒，不推，難過，扣分，冷淡，誇張。

---

資料來源：本研究整理

接著研究者對特徵詞進行挑選，本研究將篩選後之特徵詞，請教墾丁地區旅遊相關業者及生態專家對特徵詞進行挑選。被選取的特徵詞，經過在討論區中之文件搜尋比對，確定該特徵詞於絕大部份的文件，均能識別為對當次旅遊經驗表達想法，則將該詞保留。蒐集到的資料會透過中文斷詞處理等方式作為特徵詞群組，並經由過濾篩選後成為具有代表性的情緒特徵詞。透過此特徵詞篩選的步驟，保留對墾丁旅遊景氣能有效表達的特徵詞，希望藉由此篩選動作，我們可以將精選出來的特徵詞，使每份文本可以有共通的特徵屬性，進而確保研究結果的一致性。相關蒐尋過濾的例子如下：

- 沙灘旁剛好有位遊客在嘔吐. 整條路上有點髒亂。
- 船帆石是我自己最喜歡的一個地方。它的地形獨特，風景相當特別，大推!
- 這次住的民宿我覺得既貼心又浪漫，還有美味的 BBQ 食材。
- 墾丁兩天一夜都嫌無聊，你們去三天兩夜實在令人佩服。
- 墾丁必訪的墾丁大街，大碗滿意的青蛙下蛋一定要嘗試一下。

將選取後特徵詞，整理為表 4。

表 4 選取後特徵詞

---

選取後特徵詞：

---

美味，讚，棒，滿意，好玩，貼心，舒服，開心，加分，爽，乾淨，大推，扣分，難吃，爛，糟，失望，無聊，不爽，髒，不推

---

資料來源：研究者整理

## 2. 情緒分數

Oelke et al. (2009)在一篇討論客戶回饋意見分析的文章中，定義所謂意見信息字(Opinion signal word，包括像"great"、"need"、"like"等字)，該字詞依與產品特定之屬性字(Attribute)的距離，分別給予不同的分數，並依意見字的正、負屬性，分別乘上 +1, -1 的極性值(Polarity)。而單篇文章的意見分數(Opinion score)即為上述所有正、負意見分數之總合。基於這樣的方式，並根據旅遊滿意度相關文獻探討中，本研究歸納出墾丁國家公園旅遊網路口碑情緒之構面，包括「價格」、「餐飲」、「人文資源」、「環境設施」、「服務」、「交通」、「整體滿意度」等七大構面。本研究設計了一套單篇文件的情緒分數計算方法。自部落格取得之某份文件，將所表現之情緒區分為"正面經驗"、"負面經驗"，並以情緒分數 +1、-1 表示。以上述七大構面為衡量依據，搜尋網路部落格中之墾丁國家公園相關文章，判斷其文章傳達正面或負面之看法。若為正面之看法即給予+1 分數；負面之看法則給予-1 分數，依此類推，計算每段期間各構面之總分，用以表示情緒分數，並將其格式化為網路口碑資料庫，見表5。

表5 網路口碑情緒指標資料庫

期間	人文資源		價格		餐飲		環境		服務		交通		滿意度	
	正 面	負 面	正 面	負 面	正 面	負 面	正 面	負 面	正 面	負 面	正 面	負 面	正 面	負 面
Jan-13	50	2	48	1	63	3	72	2	66	1	45	5	73	2
Feb-13	63	1	58	2	73	3	70	1	68	2	72	3	46	4
Mar-13	152	2	120	1	133	2	147	2	137	1	89	1	159	1
Apr-13	130	2	115	4	159	6	135	2	139	1	87	3	161	2
May-13	125	1	120	2	102	2	119	2	105	0	87	4	131	2
Jun-13	159	3	182	4	212	5	168	5	176	1	127	1	160	4
合計	679	11	643	14	732	21	711	14	691	6	507	17	730	15
百分比	14%	11%	14%	14%	16%	21%	15%	14%	15%	%	10%	6%	16%	15%

資料來源：研究者整理

接著將篩選得出的特徵詞儲存至 crawler 爬蟲程式，透過 crawler 程式利用 Google 搜尋截取部落格上的資訊，程式能針對設定時間 2009 年至 2013 年內，網路部落格文字內容進行比對計分，並計算出發文者對墾丁國家公園的情緒分數。以 2013 年 1 月至 12 月資料為例，如下表 6 所示：

表 6 網路情緒分數

日期	Jan-13	Feb-13	Mar-13	Apr-13	May-13	Jun-13
情緒分數	401	434	927	893	763	1178
日期	Jul-13	Aug-13	Sep-13	Oct-13	Nov-13	Dec-13
情緒分數	1051	1366	1048	1140	966	816

資料來源：本研究整理

為了達到用過去的變數預測未來遊客量之目的，本研究將輸入變數進行平移 (Shift)，並嘗試平移一至三個月份，以找出距離多久的指標對遊客量的影響程度較大。如圖3 所呈現的平移月份之範例，Shift<sub>1</sub> 是將原始月份的輸入變數平移1個月份，Shift<sub>2</sub> 是將原始月份的輸入變數平移2個月份，Shift<sub>3</sub> 是將原始月份的輸入變數平移3個月份，Shift<sub>1-3</sub> 即為結合平移1個月、2月份、3月份的輸入變數。

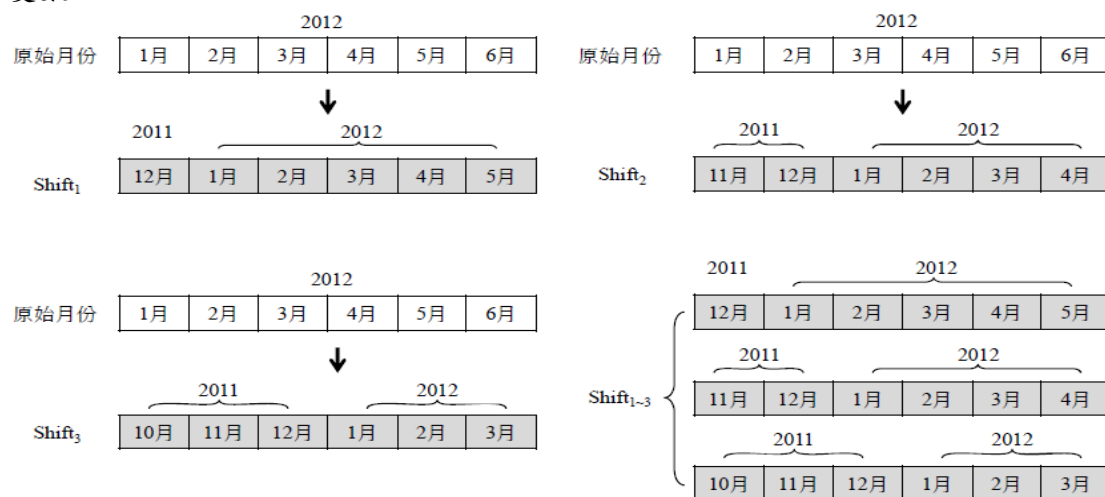


圖 3 平移月份之範例

資料來源：徐嘉鴻、林榆青、彭國彥(2013)

### 三、經濟指標

本研究從行政院主計處的PC-AXIS總體資料庫取得研究所需之經濟指標數據，包括：MIB貨幣供給額、失業率、國內生產毛額、消費者物價指數、股價指數等五項經濟指標以進行預測。其中部分資料之統計時間係以季均值顯示，因此須將季資料轉換成月資料，以利後續資料處理分析。以2013年1月至12月資料為例，詳見表7所示。

表7 遊客量、網路搜尋熱門度與經濟指標資料庫

期間	MIB 貨幣供給額	失業率(%)	國內生產 毛額	消費者物價 指數(%)	股價指數
Jan-13	12,407,945	4.16	3,531,731	102.41	7759.25
Feb-13	12,634,244	4.24	3,531,731	102.90	7932.68
Mar-13	12,595,646	4.17	3,531,731	101.42	7906.99
Apr-13	12,556,675	4.07	3,540,563	102.05	7893.42
May-13	12,681,028	4.06	3,540,563	102.24	8272.34
Jun-13	12,709,184	4.14	3,540,563	102.75	7988.80
Jul-13	12,822,892	4.25	3,679,037	102.59	8101.91
Aug-13	12,988,059	4.33	3,679,037	102.72	7932.02
Sep-13	13,036,629	4.24	3,679,037	104.06	8193.46
Oct-13	13,014,770	4.24	3,812,911	103.78	8366.18
Nov-13	13,068,044	4.16	3,812,911	103.03	8236.52
Dec-13	13,274,088	4.08	3,812,911	102.94	8431.40

資料來源：本研究整理

#### 四、遊客量統計

利用墾丁國家公園管理處的政府資訊公開之統計資料，蒐集自2009年1月至2013年12月止，共計60筆月資料，見表8所示。

表8 2009年1月至2013年12月墾丁國家公園之遊客量

期間	遊客量	遊客量	遊客量	遊客量	遊客量
	2009	2010	2011	2012	2013
Jan	375160	310442	299142	482917	381746
Feb	347170	530143	517067	381637	612735
Mar	327741	432723	400086	540826	533821
Apr	475449	615814	581957	627934	730726
Mar	492353	618874	516355	678355	591087
Jan	429888	551822	555997	532353	582764
Jul	558700	742192	644632	788354	591862
Aug	326550	656974	567621	675778	730422
Sep	339490	455693	509497	482099	534808
Oct	385722	529116	571468	652801	656458
Nov	422880	566635	651051	567387	551473
Dec	457318	520874	526287	597666	552183

為了避免所有輸入變數彼此間的值域差距過大，而間接影響遊客量預測效果，所以將所有輸入變數進行資料正規化 (Normalized)，使得各個輸入變數的值域一致，本研究採取的正規化為極值正規化，其計算公式如下：

$$\text{Normalized\_value} = \frac{\text{Original\_value} - \text{Min\_value}}{\text{Max\_value} - \text{Min\_value}}$$

## 五、預測模型建立與評估

### (一) 預測模型建立

本研究將MIB貨幣供給額、失業率、國內生產毛額、消費者物價指數、股價指數、谷歌趨勢搜尋熱門度及網路情緒分數等七項指標作為自變數，並使用PC Neuron4.0類神經網路軟體，以其中的倒傳遞網路模式進行墾丁之遊客人數預測。預測模型如下圖4所示。

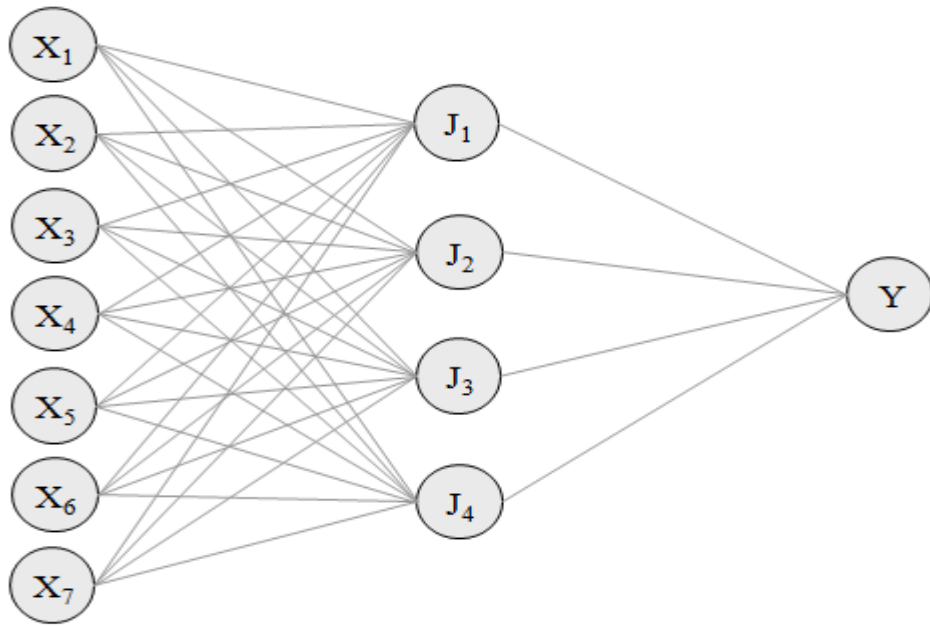


圖4 本研究預測模型圖

資料來源：本研究繪製

$X_1$ =情緒分數

$X_2$ = 谷歌趨勢熱門度

$X_3$ =貨幣供給額

$X_4$ =失業率

$X_5$ = 國內生產毛額

$X_6$ =消費者物價指數

$X_7$ =股價指數

$J_n, n=1\sim 4$ : 加權值

$Y$ =遊客量

在實證操作上可分為三個部份：1. 訓練及測試樣本；2. 驗證模型；3. 預測。研究以5項總體經濟指標為基礎樣本，以搭配谷歌趨勢熱門度和情緒分數等不同之變數組合來訓練、測試樣本，進而檢驗模型之預測準確度。

## (二)預測模型評估

### 1. 樣本訓練、測試與驗證

本研究保留 60 筆原始資料中的最後 6 筆資料作為模型預測能力檢測，運用前面 54 筆資料來建構模型，為了避免過度學習問題，本研究利用一組樣本資料來驗證預測模型的正確性。預測模型結果將以誤差均方根(Root of Mean Square，RMS)進行檢驗，所得誤差均方根越小表示預測模式越接近實際值，誤差均方根定義公式如下：

$$RMS = \sqrt{\frac{\sum_p^M \sum_j^M (T_{jp} - Y_{jp})^2}{M \times N}}$$

$T_{jp}$ =第p個樣本的第j個輸出單元之目標輸出值

$Y_{jp}$ =第p個樣本的第j個輸出單元之推論輸出值

M=樣本數

N=輸出層處理單元數

### 2. 類神經網路參數設定

本研究將2009年1月至2013年6月共54筆月資料投入PC Neuron4.0軟體進行運算，其模型參數設定如下表9所示。



表9 倒傳遞類神經網路參數設定值

項目	設定值	項目	設定值
1. 輸入單元數	6、7	10. 是否使用以學習之 網路連結加權值	否
2. 第一隱藏層單元數	3、4	11. 加權值值域	0.3
3. 第二隱藏層單元數	未使用	12. 亂數種子	0.456
4. 輸出處理單元數	1	13. 學習速率初始值	1
5. 訓練範例數	40	14. 學習速率折減係數	0.95
6. 測試範例數	14	15. 學習速率下限值	0.1
7. 學習循環數目	1000	16. 慣性因子初始值	0.5
8. 測試週期	10	17. 慣性因子折減係數	0.95
9. 是否使用批次學習	否	18. 慣性因子下限值	0.1

資料來源：研究者整理

上表各項參數設定依不同變數組合，設定會有所調整。上表中第1至3項處理單元數依投入指標數量不同而有6、7個輸入單元數，連帶影響第一層隱藏層單元數有4、5個兩種設定，第7至18項設定則參考葉怡成(2009)之建議設定。上述設定值為本研究之基本設定，若RMS值偏大，無法獲得正確學習模型，始調整參數設定以求最適當模型。

### 3. 預測能力評估

預測準確性係指衡量模型所得預測值與實際值間之誤差大小，藉以評估模型配適的好壞。學者Makridakis(1993)指出，平均絕對誤差百分比具有整合不同衡量指標的最佳特質，建議評估不同模型預測能力應以MAPE為準則，能提供決策者作為判斷依據，陳少棠(2012)曾利用MAPE評估房地產交易溫度之模型誤差，徐嘉鴻、林榆青、彭國彥(2013)也利用MAPE評估汽車銷售量之模型成效，余家蕙(2014)

同時利用MAPE評估三峽旅遊市場溫度之模型準確度，故本研究將採用MAPE來檢測模型預測值的準確性。MAPE常被用於測量時間序列統計值與目標值的適配精確度，通常以百分比表示，其公式如下。

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

$A_t$  代表實際值

$F_t$  代表預測值

n 為期數

一般而言，MAPE 因為其分母為實際值，並以百分比表示，因此不會因數值之大小而產生比較基礎不穩固之問題。根據學者(Lewis 1982)將MAPE 分為四種等級，當MAPE 值<10%時，即表示其實際與預測的誤差越小，並有較佳的預測能力，MAPE 評估預測準確度之準則如表10所示。

表10 MAPE預測能力尺度

MAPE	預測能力
<10%	預測能力極佳（愈接近0愈好）
10%~20%	預測能力優良
20%~50%	預測能力合理
>50%	預測能力不正確

資料來源：Lewis(1982)

## 肆、實證結果分析

### 一、資料描述

藉由前章所述之研究方法，本研究蒐集 2009 年至 2013 年間各項變數之樣本資料，建立樣本原始資料庫後以進行預測分析。其中自變數類型共可分為情緒指標、搜尋熱門度及經濟指標等三項。經濟指標內之變數則有「貨幣供給額」、「失業率」、「國內生產毛額」、「消費者物價指數」、「股價指數」等 5 個變數，各項變數之描述性統計分析如下表 11 所示。

表 11 變數描述統計表

變數類型	變數名稱	最小值	最大值	平均數	標準差
網路資訊	情緒分數 ( $X_1$ )	275	1366	753.13	266.85
谷歌趨勢 搜尋熱門 度	谷歌趨勢搜尋熱 門度 ( $X_2$ )	79	426	225.15	91.69
經濟指標	貨幣供給額 ( $X_3$ )	8327910	13274088	11292688	1218733
	失業率 ( $X_4$ )	4.06	6.13	4.77	0.69
	國內生產毛額 ( $X_5$ )	2986363	3812911	3419180	206141
	消費者物價指數 ( $X_6$ )	96.13	104.06	100.19	2.11
	股價指數 ( $X_7$ )	4475.14	8970.76	7614.13	936.9
產出變數	遊客量(Y)	299142	788354	531151.3	114936.1

資料來源：本研究整理

經濟指標的資料來源為「行政院主計處 PC-AXIS 總體統計資料庫」，其包含月指標及季指標二種，月指標包含消費者物價指數、台幣兌美元、股價指數、失業率；季指標包含國內生產毛額。因預測遊客量為每個月，故將季指標以均等轉換的方式，轉換成月值，讓資料呈現一致性。

## 二、實證結果分析

本研究為達成研究目的，分別以不同變數類型組合投入倒傳遞網路進行預測，建立墾丁國家公園旅遊市場溫度之預測模型。其分別為：網路情緒結合經濟指標預測模型(Mood index & Economic Model; MEM)、搜尋熱門度結合經濟指標預測模型(Google Trends index & Economic Model; GEM)、經濟指標結合網路情緒及搜尋熱門度預測模型(Economic、Mood index & Google Trends

Model；EMGM)、利用 EMGM 模型結合資料平移(EMGMS)等 4 個預測模型。以下分別分析各預測模型之實證結果。

(一)網路情緒結合經濟指標預測模型(MEM)

本研究依前述研究設計，以 2009 年 1 月至 2013 年 6 月共計 54 筆月資料為樣本，並輸入網路情緒分數及 5 項經濟指標進行預測模型之訓練測試，MEM 預測模型之設定與評估結果如表 12。

表 12 MEM 模型分析表

MEM 預測模型	類別	數值
	(訓練、測試)樣本	(40, 14)
分析項目	輸入層單元數	6
	隱藏層	1
	輸出層單元數	1
	測試模型之 RMS	0.00598
	驗證模型之 RMS	0.01177
評估項目	MAPE	19.21%

資料來源：本研究研究整理

如表 12 所示，MEM 模型以 40 個樣本訓練，14 個樣本測試，並以 1 層隱藏層執行運算，本模型測試之誤差均方根(RMS)為 0.00598，由此可得知模型誤差收斂良好，預測值與實際值誤差小；另外再投入 10 個樣本進行驗證，驗證模型之 RMS 為 0.01177，說明本模型訓練學習效果良好，且驗證模型之 RMS 值並無過小問題可避免過度學習，對訓練樣本以外的樣本有接近之預測能力。最後以前述探討採用 MAPE 對 MEM 模型進行評估，其數值為 19.21%，對照 MAPE 預測力尺度表，MEM 模型具有良好之預測能力。

## (二) 搜尋熱門度結合經濟指標預測模型(GEM)

為比較不同預測組合找出最佳預測模型，本研究將關鍵字搜尋熱門度結合 5 項經濟指標投入模型運算，GEM 預測模型之設定與評估結果如表 13。

表 13 GEM 模型分析表

GEM 預測模型	類別	數值
	(訓練、測試)樣本	(40, 14)
分析項目	輸入層單元數	6
	隱藏層	1
	輸出層單元數	1
	測試模型之 RMS	0.00535
	驗證模型之 RMS	0.01479
評估項目	MAPE	18.12%

資料來源：本研究研究整理

如上表 13 所示，GEM 模型同樣以 40 個樣本訓練，14 個樣本測試，並以 1 層隱藏層執行運算，本模型測試之誤差均方根(RMS)為 0.00535，顯示模型誤差收斂相當好；另外再次投入 10 個樣本進行驗證，驗證模型之 RMS 為 0.01479，說明本模型訓練學習效果良好，且訓練模型與驗證模型之 RMS 值差距小，對訓練樣本以為的樣本有接近之預測能力。最後以前述探討採用 MAPE 對 GEM 模型進行評估，其數值為 18.12%，對照 MAPE 預測力尺度表，GEM 模型具有良好之預測能力。

## (三)網路情緒結合經濟指標及搜尋熱門度預測模型(EMGM)

本研究將5項經濟指標和網路情緒分數及關鍵字搜尋熱門度共7項指標投入模型運算，嘗試瞭解在結合所有變數後，對預測模型的準確度的提升效能，EMGM預測模型之設定與評估結果如下表14。

表 14 EMGM 模型分析表

EMGM 預測模型	類別	數值
	(訓練、測試)樣本	(40, 14)
分析項目	輸入層單元數	7
	隱藏層	1
	輸出層單元數	1
	測試模型之 RMS	0.00503
	驗證模型之 RMS	0.01868
評估項目	MAPE	19.18%

資料來源：本研究研究整理

如上表 14 所示，EMGM 模型同樣以 40 個樣本訓練，14 個樣本測試，並以 1 層隱藏層執行運算，本模型測試之誤差均方根(RMS)為 0.00503，顯示模型之預測期望值與實際測試值誤差很少；另外再次投入 10 個樣本進行驗證，驗證模型之 RMS 為 0.01868，說明本模型訓練學習效果良好，且訓練模型與驗證模型之 RMS 值相較偏低，說明本模型夠有較佳的普遍性和解釋力。最後以前述探討採用 MAPE 對 EMGM 模型進行評估，其數值為 19.18%，對照 MAPE 預測力尺度表，EMGM 模型具有良好之預測能力。

#### (四) EMGM 模型結合資料平移(EMGMS)

最後，本研究採遲落法將 EMGM 模型運算當中的網路情緒分數、搜尋熱門度及經濟指標往前平移 1~3 個月，嘗試檢驗此模型的預測能力，EMGM 預測模型之設定與評估結果如下表 15。

表 15 EMGMS 模型分析表

EMGMS 預測模型	類別	數值
	(訓練、測試)樣本	(40, 14)
分析項目	輸入層單元數	7
	隱藏層	1
	輸出層單元數	1
平移 1 個月 EMGMS <sub>1</sub>	MAPE	15.81%
平移 2 個月 EMGMS <sub>2</sub>	MAPE	17.28%
平移 3 個月 EMGMS <sub>3</sub>	MAPE	14.98%

資料來源：本研究整理

如上表 15 所示，EMGMS 模型同樣以 40 個樣本訓練，14 個樣本測試，並以 1 層隱藏層執行運算，當平移 1 個月時(EMGMS<sub>1</sub>)，模型 MAPE 值為 15.81%；當平移 2 個月時(EMGMS<sub>2</sub>)，模型 MAPE 值為 17.28%；當平移 3 個月時(EMGMS<sub>3</sub>)，模型 MAPE 值為 14.98%，對照 MAPE 預測力尺度表，以平移三個月的 EMGMS<sub>1</sub> 模型預測能力最高，且三個模型皆具有良好的預測能力，結果顯示網路口碑情緒、Google Trends 及經濟指標對於墾丁未來的旅遊市場溫度之具有優良預測力。

#### (五) MEM、GEM、EMGM、EMGMS<sub>3</sub> 模型效能分析比較

根據研究結果，將四種預測模型以表 16 進行差異性比較。

表 16 墾丁國家公園旅遊市場溫度模型分析表

模型名稱	MEM	GEM	EMGM	EMGMS <sub>3</sub>
(訓練,測試)	(40,14)	(40,14)	(40,14)	(40,14)
輸入層單元數	6	6	7	7
隱藏層	1	1	1	1
輸出層單元數	1	1	1	1
測試模型 RMS	0.00598	0.00535	0.00503	0.00532
驗證模型 RMS	0.01177	0.01479	0.01868	0.00786
MAPE	19.21%	18.12%	19.18%	14.98%

資料來源：本研究研究整理

由表 16 顯示，本研究運用相同之訓練及測試樣本數設定，對四種不同變數類型組合之預測模型進行演算，並用 MAPE 評估模型預測能力，結果顯示所有模型均有良好的預測能力。

以往的遊客量的預測模型大多以經濟指標或自然環境等因素做為變數進行推估，但透過上表之研究結果顯示，加入網路情緒分數及 Google Trends 網路搜尋熱門度確實能增強模型預測能力；此外，若將網路情緒分數及 Google Trends 網路搜尋熱門度及經濟指標加以平移，其預測力能更為提升，其中又平移三個月所建構的預測模型，為所有模型組合中預測效能為最佳。各模型預測值與實際值關係如下圖 5 所示



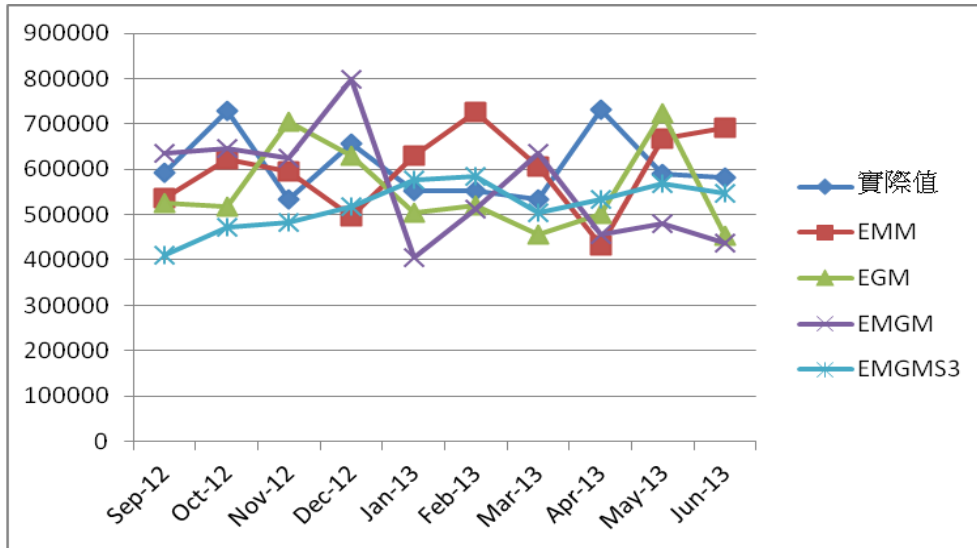


圖 5 各模型預測值與實際值之關係折線圖

資料來源：研究者整理

由上圖可以看出 EMGMS<sub>3</sub> 模型的預測值趨勢和實際值較為接近，更加驗證本研究之假設，利用網路情緒結合經濟指標及搜尋熱門度並加以平移之模型，其預測力優於其它三種模型。

### 三、旅遊市場溫度預測

#### (一) 旅遊市場預測

在前述研究結果的基礎下，本節選擇以四種預測模型當中，MAPE 值最佳的 EMGMS<sub>3</sub> 模型，輸入各項變數，並將 2013 年 4 月至 2014 年 3 月，網路情緒分數、搜尋熱門度及經濟指標等資料平移三個月，對 2013 年 7 月至 2014 年 6 月遊客量進行未來旅遊市場溫度預測。最後將推估的預測值與實際值進行比對換算，得到的 MAPE 值為 14.36%，由此可說明本研究所建構的 EMGMS<sub>3</sub> 模型對於墾丁國家公園未來的旅遊市場溫度有良好的預測力。

表 17 即為本研究利用 EMGMS<sub>3</sub> 模型，對墾丁國家公園 2013 年 7 月至 2014 年 6 月進行旅遊溫度預測值與實際值比較；圖 6 為墾丁國家公園 2013 年 7 月至 2014 年 6 月旅遊溫度預測對照折線圖，從圖表中可以發現，EMGMS<sub>3</sub> 模型的預測值能隨著實際值的走勢起伏變動，即透過輸入即時的網路情緒分數、搜尋熱門度及經濟指標，便能有效推測出墾丁國家公園遊客量。

表 17 旅遊溫度預測值與實際值比較

EMGMS <sub>3</sub> 模型	實際值	預測值	誤差值	MAPE
Jul-13	591862	587136	4726	14.36%
Aug-13	730422	678066	52356	
Sep-13	534808	605333	-70525	
Oct-13	656458	595644	60814	
Nov-13	551473	505562	45911	
Dec-13	552183	643685	-91502	
Jan-14	440869	563985	-123116	
Feb-14	654104	423766	230338	
Mar-14	644630	756999	-112369	
Apr-14	848301	746957	101344	
May-14	703073	619355	83718	
Jun-14	708256	797765	-89509	

資料來源：研究者整理

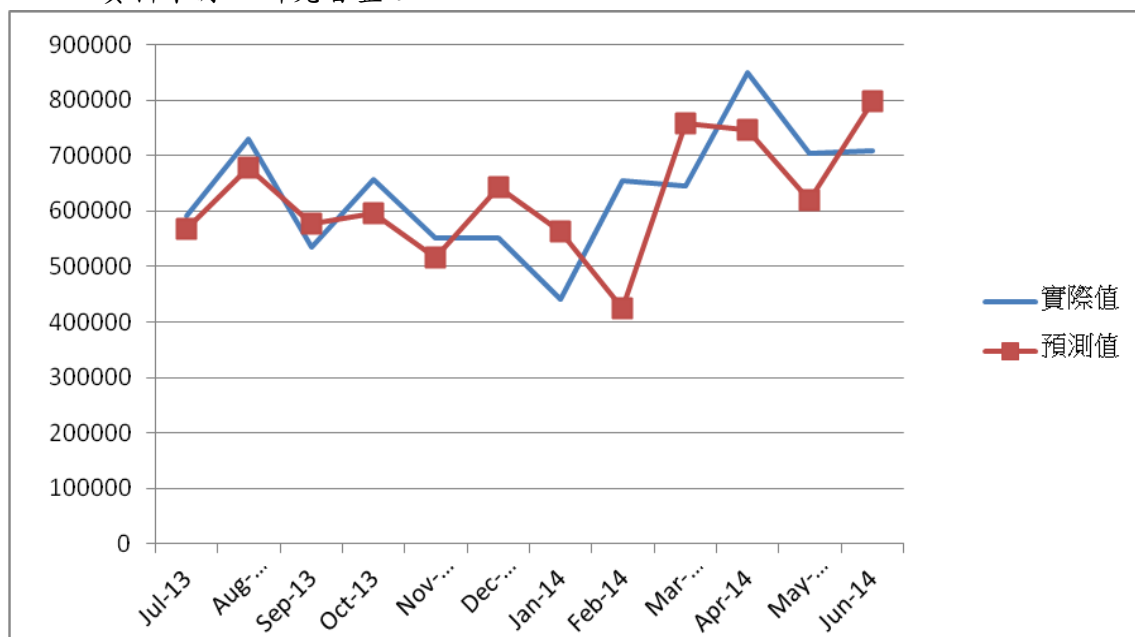


圖 6 墾丁國家公園 2013 年 7 月至 2014 年 6 月旅遊溫度預測對照折線圖

資料來源：研究者整理

綜合本研究結果，發現利用經濟指標結合平移後的網路情緒分數、搜尋熱門度及經濟指標，所建構出的預測模型有良好的預測能力，當平移時間為三個月時，預測效果最佳。同時，本研究建構之預測模型對墾丁國家公園未來旅遊市場溫度具有良好的預效能力。

## 二、預測效果較差之月份

進一步觀察，部分的月份中，有幾個月份的 MAPE 值相當的高。從表 5 可以得知，在 2014 年的 1 月至 4 月，這幾個月份預測人數誤差大於 10 萬人次。根據預測模型，我們將時間點往前平移三個月，推測其中原因，可能是 2013 年 9 月-12 月黃色小鴨活動帶來的排擠效應。黃色小鴨由高雄市政府在 2013 年 9 月 19 日引進，並相續在桃園、基隆等地巡迴展出，引發媒體及網路爭相報導搜尋。根據高雄市政府的統計，黃色小鴨共吸引超過三百萬的人潮前來觀看，陸續在桃園、基隆等地的展覽活動，都吸引上百萬人的目光。影響所及不僅是遊客的造訪，網路口碑、搜尋熱門度也隨之改變，像這類型的變數是無法得知的，也就會造成預測效果不佳的情況發生。

## 伍、結論與建議

### 一、結論

本研究利用資料探勘方法，蒐集收集 2009 年 1 月至 2013 年 12 月止所有公開討論墾丁國家公園之網誌文章，共 11,745 篇，經過情緒特徵詞選取，萃取網站討論者之情緒，量化轉換成情緒分數，藉以獲取墾丁國家公園之網路口碑。並進一步加入 Google Trends 搜尋熱門度和經濟指標等七項指標，作為自變數，透過類神經網路(Neural Network)之倒傳遞網路(BNP)，建立不同組合模型，並以墾丁國家公園遊客數量，作為旅遊市場溫度之依變數，建構出預測模型。研究發現，利用經濟指標，結合平移後的網路情緒分數及搜尋熱門度，所建構出的預測模型，對墾丁國家公園旅遊市場溫度有良好的預測能力，其研究結果分述如下：

(一)本研究將網路情緒分數及 Google Trends 搜尋熱門度平移後，結合經濟指標利用類神經網路，建構之最佳預測模型 EMGMS<sub>3</sub>，MAPE 值達 14.98%，其各項數值設定如下：輸入層單元數為 7 個，隱藏層為 1 層，輸出層單元數為 1 個，並將總樣本 54 個的 3/4 樣本數 40 個作為訓練模型，1/4 樣本數 14 個作為驗證模型。各項參數設定為：學習循環數目 1000 次、測試週期 10 回、加權值域 0.3、亂數種子 0.456、學習速率初始值 1、學習速率折減係數 0.95、學習速率下限值 0.1，慣性因子初始值 0.5、慣性因子折減係數 0.95、慣性因子下限值 0.1。

(二)本研究將 MEM、GEM、EMGM 及 EMGMS<sub>3</sub> 等模型，進行效能分析比較，結果顯示所有模型均有良好的預測能力，其中以 EMGMS<sub>3</sub> 模型預測能力較佳，MAPE14.98%。驗證本研究之推論，利用網路情緒結合經濟指標及搜尋熱門度並加以平移之模型，

其預測力優於其它三種模型。

(三)研究利用 EMGMS<sub>3</sub> 模型，將 2013 年 4 月至 2014 年 3 月，網路情緒分數、谷歌趨勢搜尋熱門度及經濟指標等資料平移，並推測三個月後，即 2013 年 7 月到 2014 年 6 月墾丁國家公園之旅遊市場溫度，所得 MAPE 值達 14.36%，即誤差值小，有良好的預測能力，說明過去網路指標及經濟指標，對墾丁國家公園旅遊市場溫度影響時間點為 3 個月。本研究突破傳統遊客量預測限制，克服資料落遲性問題，運用即時的線上資料，包括網路情緒分數、Google Trends 搜尋熱門度及經濟指標，建構出墾丁國家公園旅遊市場溫度之預測模型，可作為相關單位資源分配之依據。

## 二、建議

對於業者及管理者而言，本研究提供具有客觀、創新及便利性之預測方式，實務上可運用於經營者編制未來營運計畫時預測之參考依據。基於此，綜合以上研究結論，針對墾丁國家公園業者或管理者提出以下建議：

### (一)對業者之建議

1. 研究結果顯示墾丁國家公園的網路指標與旅遊市場溫度，起伏變化相當大，有明顯淡旺季差異。建議業者於每年淡季，11 月至 2 月前三個月，即 9 月至 11 月應創造話題，引發網路討論效應，吸引遊客前往，並結合網路口碑行銷，讓遊客進行旅遊體驗後，於部落格上發表正面口碑之相關文章，藉以提升網路情緒分數，並帶動旅遊市場溫度的增溫。
2. 本研究發現，網路口碑情緒、谷歌搜尋熱門度與經濟指標，對於預測墾丁國家公園旅遊市場溫度具有影響力。因此，業者應注重經營網路行銷，對於遊客於網路發表之意見與評斷，能即時回應。實務作為方面，可經營部落格或網路平台，加強與遊客溝通，並能即時回應遊客疑問或需求，以提升旅遊服務品質，創造網路正面情緒分數。
3. 從研究結果得知，各縣市重大活動，特別經過網路、媒體大量曝光，對網路搜尋度、遊客量會有影響，業者需掌握國內各縣市，重大旅遊活動，確實調配旅遊資源，擬定價格與行銷方案，以維持服務品質的穩定。

### (二)對研究者之建議：

1. 本研究僅針對墾丁國家公園，國內遊客人數進行預測，未來研究可進一步納入非台灣之外籍遊客量進行探討。
2. 網路趨勢變化迅速，近年來 FB、LINE 等通訊軟體使用率大增，受限於帳戶隱私設定，本研究只針對網誌文章做擷取，當作情緒分數指標，未來可進一步加入 FB 粉絲專頁中的評論，做為本研究建立模型的輸入變數，藉以提升預測效力。
3. 本研究針對台灣國家公園進行預測模式之探討，未來研究可將其類似預測

模型，針對不同性質、不同領域之產業進行驗證比較。

4. 本研究結合網路情緒分數與類神經網路建構出旅遊市場景氣溫度，未來研究者可根據本研究之預測模式，結合不同預測方法進行預測驗證，或以不同之組合方式建立預測模型，以期獲得更佳之預測績效。

## 參考文獻

1. 邱昭彰、徐嘉鴻、林榆青、彭國彥，2013，「汽車銷售量與經濟指標及網路搜尋熱門度關聯性之研究 - 以 Mazda 汽車為例」，元智大學資訊管理系，打字複印。
2. 邱昭彰、林彥佐，2012，「利用網路口碑預測台北市電影票房交易量，元智大學資訊管理系」，元智大學資訊管理系，打字複印。
3. 林寶秀、林晏州，2010，「島嶼型遊憩區遊客量推估與預測」，國家公園學報，20(1):1-14。
4. 施孟隆，2014，「利用網路訊息推測旅遊溫泉度假遊客量之研究-某鄉村地區 A 渡假村為例」，鄉村發展，16:1-17。
5. 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.
6. Chen, C.F.; M.C. Lai, C. and C., Yeh, 2012, "Forecasting tourism demand based on empirical mode decomposition and neural network," *Knowledge-Based Systems*, 26 : 281-287.
7. Cho, V., 2003, "A comparison of three different approaches to tourist arrival forecasting," *Tourism Management*, 24(3):323-33.
8. Frijters, Paul, D. W. Johnston, G. L, and Michael 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.
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. 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.
12. Lewis C.D., 1982, *International and Business Forecasting Methods*, London: Butterworths.

13. 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.
14. Smith, G. P. , 2012, "Google Internet search activity and volatility prediction in the market for foreign currency," *Finance Research Letters*, 9( 2): 103-110.
15. Song, H. and Gang Li, 2008, " Tourism demand modeling and forecasting—A review of recent research," *Tourism Management*, 29(2):203-220.
16. 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.
17. Vosen, S. and T. Schmidt, 2011, "Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends," *Journal of Forecasting*, 30:565-578.
18. 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.
19. Witt, Stephen F. and Christine A. Witt, 1995, "Forecasting tourism demand: A review of empirical research," *International Journal of Forecasting*, 11(3):447-475.
20. 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.

# **The research of predicting tourist amount by internet information – Using Kenting National Park as an Example**

**Po-Wei Tsai\***      **Meng-Long Shih\*\***

## **Abstract**

Internet platform provides consumers a place to express their ideas and experience. More and more consumers are used to searching information from internet before they by shop, thus the influence of electronic word of mouth (e-WOM) is also increasing day by day. Travel market might be effected by e-WOM, climate, environment, even economic variation, Therefore, e-WOM is a appropriate method to discuss the tourists true feeling in tourism industry.

In this study, blog data were collected from a total of 11,745 articles from 2009 to 2013 about Kenting National Park tourism from the Google 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 internet emotion indicators, 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 internet emotion indicators or Internet search popularity on Google Trends with economic indicators can achieve in a good value of MAPE(14.36%). 3. The empirical results show that the best forecasting ability is when emotional indicators ,Google trends of Internet reputation are moved forward for 3 months, indicating the influential time point for internet emotion indicators to predict the future volume of tourism in Kenting National Park travel market is 3 months. This study provides an objective, innovation and convenient way of predicting. It can be used practically as a predicting reference for operators when preparing plans of future operations.

**Keyword: Google trends; emotion indicator; Back-propagation Network**

---

\*master.and \*\*professor and corresponding author , Department of Cultural Resources and Leisure Industries, National Taitung University