

黃芝璇、馬麗菁 (2021),『以圖形方式摘要化顧客評論』,  
資訊管理學報, 第二十八卷, 第二期, 頁 125-154。

## 以圖形方式摘要化顧客評論

黃芝璇

國立聯合大學資訊管理學系

馬麗菁\*

國立聯合大學資訊管理學系

### 摘要

隨著網際網路和社群網絡的盛行，消費者經常在採購或預訂產品或服務之前，瀏覽其他顧客的評論，以做為選擇產品或服務的參考。但是，針對所需的產品或服務，常常會有非常多的線上評論，並且評論的內容大多是以文字形式呈現的非結構化資料，因此消費者往往需要花費大量時間，來逐一瀏覽其他顧客的評論，以找到適合自己的產品或服務。另一方面，企業管理者也常需要費時一一瀏覽大量評論，來了解顧客對所提供產品或服務的反饋意見，做為其改善或調整經營決策的參考。本研究基於文字探勘、情感分析和社交網絡圖的概念，提出了一個圖形方法來摘要化顧客評論，並呈現在二維圖形上，且以台灣三大著名觀光飯店為例，取得其在 Booking.com 的線上顧客評論資料，進行分析比較。本研究先利用文字探勘在顧客評論中找出高頻名詞，並使用相似性分析和多維尺度分析法，計算每個高頻名詞的相對座標位置，接著根據社交網絡圖的概念呈現這些名詞的相關性。此外，本研究進一步使用情感分析，來顯示顧客評論中正面和負面情緒的趨勢和程度。藉由觀察圖形取代傳統閱讀複雜且冗長的文字內容，消費者可以在更短的時間內，依據自己的喜好做出較合適的選擇。企業管理者或經理人亦可從圖中觀察顧客的反饋意見，這些意見或建議可以做為改善或調整營運策略的參考，以提昇企業的競爭優勢。

**關鍵詞：**文字探勘、情感分析、視覺化、顧客評論、社交網絡圖

---

\* 本文通訊作者。電子郵件信箱：lcma@nuu.edu.tw  
2020/12/23 投稿；2021/01/29 修訂；2021/02/02 接受

Huang, Z.X. and Ma, L.C. (2021), 'Summarizing customer reviews with visual representations', *Journal of Information Management*, Vol. 28, No. 2, pp. 125-154.

# Summarizing Customer Reviews with Visual Representations

Zhi-Xuan Huang

Department of Information Management, National United University

Li-Ching Ma\*

Department of Information Management, National United University

## Abstract

**Purpose**— With the popularization of the Internet and social networks, consumers often browse other customers' reviews before making reservations or consumption, as a reference for making choices. Consumers often need a lot of time to browse online reviews one by one. On the other hand, business managers often need to spend a lot of time browsing a large number of reviews page by page to understand customer feedback on the products or services provided. This study aims to propose a graphical approach to summarize online reviews.

**Design/methodology/approach**— This study proposes a graphical approach to summarize online customer reviews on two-dimensional graphs based on the concept of text mining, sentiment analysis and social network diagram. This research first finds out high-frequency nouns in the customer reviews, uses similarity analysis and the multidimensional scaling to calculate the coordinates of each high-frequency noun, and then displays these nouns based on the concept of social network diagram. In addition, this study further employs sentiment analysis to show the tendency and degree of positive and negative emotions in customer reviews.

**Findings**— Online customer reviewers of three famous tourist hotels in Taiwan, retrieved from Booking.com, are taken as an example to illustrate the proposed approach. From the perspective of consumers and hotel managers, the features and improvement

---

\* Corresponding author. Email: lcma@nuu.edu.tw  
2020/12/23 received; 2021/01/29 revised; 2021/02/02 accepted

suggestions for these three hotels are provided respectively.

**Research limitations/implications** — There are many special cases in Chinese semantics. Because the sentence structure of special case words is more complicated and there are many exceptions, this research did not deal with these special case words separately. In addition, graphical display methods often have errors, and this research did not verify that the relationships in the drawn graphics are completely correct.

**Practical implications** — This paper provides several implications for consumers and business managers. With the help of graphs instead of complex and lengthy texts, consumers can make more appropriate choices according to their preferences in a shorter time. Managers can also observe guest feedback or opinions from the graphs, which can be used as a reference for improving or adjusting business strategies to enhance their competitive advantages.

**Originality/value** — This study proposes a graphical approach to summarize online customer reviews. Consumers and business managers can directly observe the results of summary comments in a more efficient way. Greatly shorten the time spent browsing online reviews page by page. This research framework can also be widely used in other industries.

**Keywords:** Text mining, sentiment analysis, visualization, customer review, social network diagram

## 壹、緒論

隨著網際網路和社群網絡的盛行，消費者經常在採購或預訂產品或服務之前，會瀏覽其他顧客的評論，以做為選擇產品或服務的參考。但針對所需的產品或服務，常常會有成千上萬筆的線上評論，並且評論的內容大多是以文字形式呈現的非結構化資料，因此消費者往往需要花費大量時間，來逐一瀏覽其他顧客的評論，以找到適合自己的產品或服務。另一方面，企業管理者也常需要費時一一瀏覽大量評論，來了解顧客對所提供產品或服務的反饋意見，做為其改善或調整經營決策的參考。因此，如何協助消費者或企業管理者在較短時間內總結顧客評論，是一項重要議題。

本研究提出了一個圖形方法來摘要化顧客評論，並呈現在二維圖形上，且以台灣三大著名觀光飯店為例進行分析比較。現代人因工作繁忙壓力大，外出旅遊成為民眾放鬆紓壓常見的方式之一，依交通部觀光局的統計資料顯示，近十年來國內旅遊的人次大幅成長，如圖 1 所示。隨著國內旅遊人次的增加與國民所得的提升，民眾開始注重旅遊的品質，對於旅遊住宿的需求也相對提高，國內飯店、旅館、民宿等旅宿業家數也隨之快速成長。隨著觀光旅館家數的增加，旅宿業之間的競爭也越發激烈，如何協助業者了解顧客需求，進而在眾多的旅宿業中找到競爭優勢，以及幫助顧客在面對各式各樣的選擇時，能夠快速找到理想中的旅遊住宿，是一個值得關注的議題，故本研究以台灣三大著名觀光飯店為例進行探討。



圖 1：97 年至 108 年國人國內旅遊總旅人次統計圖（來源：交通部觀光局）

由於網際網路的普及，顧客常會在訂房之前瀏覽其他顧客的住宿評論，依照前人的經驗去判斷該旅店的好壞，做為選訂旅店的參考（李智明等 2018）。評論或評價常常能反應出顧客的想法，許多研究都以網路評論作為分析資料，例如，莊維荃（2017）利用問卷調查的方法，探討使用者在瀏覽過產品的網路評論後，對其來源及信任程度，是否會影響到使用者的購買決定。但由於評論資料多為文字資料且為非結構化架構，傳統問卷調查法不易處理文字評論，因此，逐漸有學者以文字探勘技術分析線上評論資料，例如，Cao、Duan 與 Gan（2011）針對線上顧客評論的有用性進行探討分析，透過文字探勘技術從評論內容中提取語意特徵，分析那些特徵會影響評論的有用性。然而，傳統文字探勘分析出來的結果多使用文字敘述，往往無法讓讀者一目了然，若能結合圖形化的方式呈現出資料分析結果，能使讀者更快速的了解資料代表的意涵，因此有學者開始發展結合文字探勘與圖形化呈現的方法，例如，李佳儒（2017）將文字探勘應用於網路的電影評論，建立一個 R-Shiny 視覺化系統，提供使用者透過系統查詢電影的影評，以及該部電影的詞頻分析、關聯規則分析和評價分類等，並以文字雲的方式呈現出影評的高頻詞彙。

本研究以臺灣三家知名觀光飯店為例，進行圖形化分析與探討，希望透過文字探勘技術，找出評論內容中的高頻名詞，並以相似性分析及多維尺度分析法（Multidimensional Scaling; MDS）找出各個高頻名詞的座標位置，再以社交網路圖概念，呈現該家飯店高頻名詞間的關係；此外，本研究並進一步利用情感分析找出評論中描述高頻名詞的正負情感傾向，並以圖形化方式呈現。本研究資料來源以 Booking.com 的旅宿評論為主，因 Booking.com 的評論不僅包含與其他訂房網站相同的基本資料，如分數、類別評分、日期及標題等，更包括了顧客正向及負向的評論，且評論筆數充份較易於分析比較。

本研究結果一方面可協助消費者，藉由觀察圖形取代傳統一一閱讀複雜而冗長的文字評論，能更有效率地找出較適合的選擇。另一方面，企業管理者或經理人亦可從圖中觀察客人的反饋意見，做為改善或調整營運策略的參考，以提昇企業的競爭優勢。

## 貳、文獻探討

隨著網際網路的快速發展，人們花在網路上的時間日趨增長，尤其在 Web 2.0 的興起後，人們開始可以在網路上互動，進而分享經驗與心情。網路評論是一種以文字形式，描述體驗感受的一種方式，有使用經驗的民眾在網路上留下他們的感受，而尚未體驗的民眾，則常會參考先前的評論做相關的決策。許多研究根據這些評論資料進行分析探討，找出其中有趣的資訊，例如：林漪寒（2016）透過

顧客的評論內容，分類出兩間公司之間是否為競爭者。蔡孟穎（2018）以美妝網站的評論為資料集，預測分析對評論人氣有顯著影響的重要字詞。劉雨鑫（2018）以網易雲音樂平台為研究對象，找出歌詞、評論與標籤之間的商業意義和相關性。楊宗晟（2019）透過網路爬蟲收集各電子商務網站品牌清淨機的線上顧客評語，並利用文字探勘和關聯規則分析呈現出視覺化文字雲，提供企業改進及審視決策參考。本研究亦以線上訂房網站 Booking.com 為例，抓取訂房網站上的評論資料，做為分析的資料來源，進而以文字探勘、情感分析、圖形化分析等方法進行研究。

在文字探勘方面，這個資訊爆炸的時代，我們每天都接觸到大量的資訊，包括各種文字、圖片、影像、數據等，傳統的資料探勘主要是針對結構化的資料進行處理，但是以半結構化或非結構化格式的資料也隱藏著許多重要的資訊（鍾任明等 2007）。在大量的資料中，常包含許多以非結構化形式呈現的文字資料，為了分析這些文字資料，並從中找出有意義的規則或樣式，許多研究採用文字探勘技術進行分析。文字探勘是從文本中獲取有用訊息的過程，即從大量文字資料中找出有趣的樣式，可運用在不同的領域上，主要過程步驟包括數據收集、預處理、數據清理、數據挖掘、建模以及模型評估（Zuo 2018）。文字探勘常被應用在線上評論上，例如，葉珊妤（2018）利用文字探勘與機器學習法預測部落格遊記的人氣數與留言數，找出遊記文章中各景點、餐廳及住宿之間的關聯，預測結果能提供給遊客做為規畫行程之參考。Ma 與 Huang（2020）結合文字探勘與基因演算法，分析線上飯店評論，並找出關鍵名詞間的關係。本研究的資料為線上訂房網站的評論內容，由大量的非結構化文字所組成，因此亦運用文字探勘技術，將評論內容進行高頻名詞的萃取，並進一步結合情感分析，以圖形化的方式呈現。

在情感分析方面，情感分析也稱為意見探勘（Opinion Mining）（Peng & Lee 2008），是指利用自然語言處理、文字探勘以及計算機語言學等方法來進行處理判斷，分析一篇文章或是一段文字的情感分類為正向或是負向情感。常用的分類方法有兩種，一為基於機器學習的方法，機器學習法是在文本中進行特徵選取，藉由訓練資料集以及測試資料集，利用 K-近鄰演算法（K-Nearest Neighborhood；KNN）、單純貝氏分類器（Naïve Bayes Classifier）或支持向量機（Support Vector Machine；SVM）等方法進行情感分類，最後將分類結果用於訓練文本以做出正確的情感標註。另一方法為採用情感字典法，字典法主要是透過對文本進行拆句及分析，根據文本中所包含的正向情感詞以及負向情感值進行情感分類，並利用情感字典計算情感值，最終以情感值作為判斷情感傾向的依據（謝麗星等 2012）。

近年來情感分析方法被廣泛地應用在顧客評論及社交網絡上。例如，Amrani、Lazaar 與 Kadiri（2018）提出一種基於隨機森林和支援向量機的情感分析混合方法來識別亞馬遜的產品評論，該研究比較了隨機森林、支援向量機以及

混合分類法，結果顯示使用混合分類法能獲得較好的研究結果。而 Chen、Chang 與 Yeh (2017) 提取 YouTube 影片的標題、標籤、說明及評論，透過文字探勘進行分析，並將影片分為六種情感類別，利用無監督式及監督式學習法的集成模型，將 YouTube 影片根據情緒進行分類。Zhang 等 (2018) 提出一種基於情感詞典的中文微博文本情感分析方法，透過程度副詞詞典、網絡單詞詞典、否定單詞詞典以及其他相關詞典的提取和建構，來擴展情感詞典，並通過計算權重獲得微博文本的情感價值，進一步將某個主題的微博文本分為正面、負面和中性。本研究亦利用字典法，將描述高頻名詞的形容詞進行情感分析方法，分類為正向情緒或負向情緒，並在最後將形容詞及高頻名詞的關係，以圖形化方式呈現。

隨著社交網絡的盛行，顧客對產品和服務的大量評論，常導致資訊超載，如何將大量數據以更有效率與效能的方法與技術來探勘與呈現，已成為一個具有挑戰性的問題。近年來文本摘要技術被廣泛應用在分析線上顧客評論上。文本摘要技術是指從初始文本中萃取出最重要的資訊，並產生簡化版本的文本。Gambhir 與 Gupta (2017) 指出文本摘要技術可以分為兩類：萃取性和抽象性摘要，萃取性摘要主要是選擇原始部分文檔，而抽象性摘要的目的是找出能代表原始文檔的具凝聚力和簡潔性概要。文本摘要又可分為單文本摘要和多文本摘要，其中單文本摘要僅處理單個文檔，而多文本摘要涉及多個文檔較為複雜 (Gupta & Lehal, 2010)。Lloret 等 (2015) 發展了一個概念層級的多文本摘要法，透過將句法簡化、句子再生和內部概念呈現，產生超簡潔意見摘要。Moussa、Mohamed 與 Haggag (2018) 進行了一項針對社交媒體的意見摘要技術的調查，該調查顯示了目前意見摘要技術的挑戰，並涵蓋了意見摘要中常使用的各種方法，包括視覺化 (Visualization)、抽象 (Abstractive)、基於方面 (Aspect based)、以聚焦查詢 (Query-focused)、即時 (Real Time) 和更新 (Update) 總結等。

有不少學者利用神經網絡的方法於文本摘要研究上 (Rush et al. 2015)。例如，Nallapati 等 (2016) 使用循環神經網路 (Recurrent Neural Networks; RNN) 開發了抽象文本摘要模型，研究結果顯示它們在兩個不同的語料庫上有良好的成效。Pecar (2018) 在顧客的旅行評論中，使用了神經網絡來形成摘要，該摘要納入了隨著時間變化的觀點。亦有不少學者利用文字探勘方法或結合情感分析進行文本摘要研究。例如，Zhuang、Jing 與 Zhu (2006) 提出了一種基於多知識的多文本摘要方法用於電影評論，該方法整合了文字探勘、統計分析和電影知識集。Hu、Chen 與 Chou (2017) 提出了一種多文本摘要方法，利用 *k*-medoids 群集演算法從線上飯店評論中找出最重要的 *k* 個句子，同時考量了內容與情感相似度，來決定兩個句子間的相似度。本研究亦結合了文字探勘與情感分析法，於多文本中產生圖形化抽象性摘要。

在圖形化呈現方面，隨著顧客對產品和服務的評論大量增加，如何讓使用者

快速、方便的看出資料所代表的意涵，視覺化技術有很大的助益。語義分析方法和視覺化技術提供了文件內容的簡明視圖，幫助使用者更容易地從大量的原始文件中，發現和了解關鍵資訊。近年來有許多研究，開始運用各種的視覺化技術呈現文件摘要的結果。有些研究使用傳統的統計圖表顯示文件摘要結果，他們多採用長條圖和圓餅圖，以圖形結構顯示文件摘要（Moussa et al. 2018）。例如，Kamal 與 Islamia（2015）利用機器學習和自然語言處理法設計了一個特徵層級統一意見探勘和情感分析框架，利用各種統計圖表（例如長條圖和圓餅圖）來呈現特徵摘要。標籤雲（Tag Cloud）亦是常見的視覺化呈現方式，以字體或圖形的大小代表文字出現的頻率，例如，Rinaldi（2019）開發了一種基於結合本體論和標籤雲的語義標籤雲方法，用於瀏覽和摘要 Web 文檔；Rinaldi 與 Russo（2021）提出了一種基於統計和語義分析的文件摘要與視覺化技術，利用視覺語義標籤雲（Visual Semantic Tag Cloud），呈現多媒體文件的類別和摘要。Carmeli（2020）提出了一個名為 EXPLAINIT 的評論摘要系統，該系統利用監督和非監督式學習的組合來挖掘評論中的觀點，並將其組織在觀點因果關係圖（Opinion Causality Graph; OCG）中，以總結因果關係。Hu 與 Li（2017）提到基於網絡的圖形（Network-based graph）（例如 NodeXL）可以幫助用戶分析和視覺化社交網絡數據，並發展基於網絡的導航圖，來協助使用者快速地找到所需的 TED Talks 影片。

Liu 等（2018）提到圖形匯總的類型可以根據輸入類型分為同質圖和異質圖，其中同質圖較為普及，具有一種節點類型和一種鏈接類型；異質圖對各種類型的節點及其之間的關係的處理方式不同，能呈現多元的資訊。前面所述的視覺化方式多為同質圖的呈現方式，本研究是採用異質圖的方式，並結合標籤雲及基於網絡的圖形概念呈現，分別以不同節點圖形代表名詞、正向形容詞、負向形容詞，並以節點大小、節點間的距離、與鏈接連線寬度分別代表不同的資訊。

## 參、研究方法

本研究的研究方法分為五部份說明，分別是研究架構、資料蒐集與前置處理、文字探勘流程、圖形化分析、與情感分析。

### 一、研究架構

本研究的研究架構圖如圖 2 所示，分成三個部份，包括資料蒐集、資料前處理與文字探勘、情感分析與圖形化呈現。本研究的資料來源為交通部觀光局的旅宿業相關公開資料，以及訂房網站 Booking.com 中的顧客評論內容，並篩選所需資料進行後續分析。第二部份是將蒐集到的評論內容利用文字探勘及語料庫進行

斷詞等處理，找出詞性為名詞的高頻詞彙，建置評論內容與高頻名詞的矩陣。最後是情感分析與圖形化呈現，將評論內容與高頻名詞矩陣計算相似性，得到高頻名詞之間的距離矩陣，接著以視覺化方式呈現，以點座標的大小表示高頻名詞出現的頻率；並以點座標之間的距離代表相似性。另外將詞性為形容詞的詞彙，結合情感字典進行情感分析，計算形容詞的正負向情感值，並結合否定詞及程度副詞加權計算情感程度，最後以圖形化方式呈現出顧客對該飯店評價的分析結果。

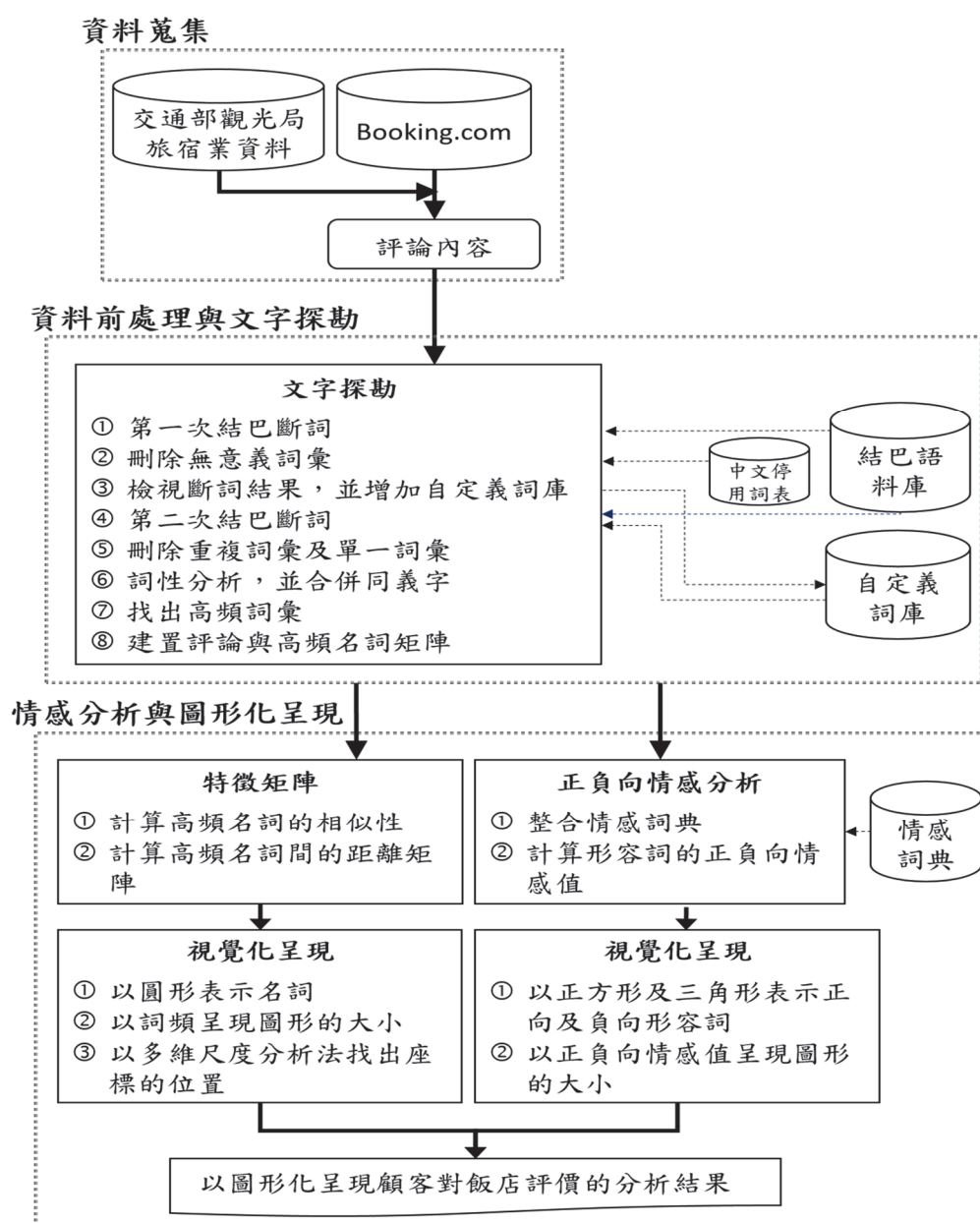


圖 2：本研究架構圖

## 二、資料蒐集與前置處理

本研究資料來源為交通部觀光局的旅宿業相關資料及訂房網站 Booking.com 中的顧客評論資料。首先由交通部觀光局的公開資料「臺灣地區觀光旅館家數及房間數」整理出房間數前十名的飯店，再搜尋這十家飯店於 Booking.com 上的評論數，挑選資料筆數較多的飯店做為研究對象，各家飯店評論數相關資料比較如表 1 所示。

表 1：各家飯店評論數比較表

| 地區  | 旅館名稱          | 2019 /1 公告房間數<br>(交通部觀光局) | 2018/05/01-2019/04/30<br>篩選後評論數 |
|-----|---------------|---------------------------|---------------------------------|
| 台北市 | 台北君悅酒店        | 866                       | < 200                           |
| 台北市 | 台北寒舍喜來登大飯店    | 692                       | < 100                           |
| 高雄市 | 義大皇家酒店        | 656                       | 468                             |
| 高雄市 | 君鴻國際酒店        | 592                       | 207                             |
| 台北市 | 晶華酒店          | 569                       | < 100                           |
| 高雄市 | 漢來大飯店         | 540                       | 203                             |
| 桃園市 | 台北諾富特華航桃園機場飯店 | 519                       | < 200                           |
| 台北市 | 台北凱撒大飯店       | 478                       | 337                             |
| 屏東縣 | 墾丁福華渡假飯店      | 458                       | 798                             |
| 高雄市 | 高雄國賓大飯店       | 457                       | 294                             |

本研究資料蒐集期間為 2018 年 5 月 1 日至 2019 年 4 月 30 日一年內資料，並只篩選出中文評論，因 Booking.com 為旅遊訂房的著名網站，本研究以 Booking.com 的評論為例，透過爬抓程式至 Booking.com 取得飯店評論內容，並選擇評論數超過 300 則的飯店，分別是「義大皇家酒店」、「台北凱撒大飯店」以及「墾丁福華渡假飯店」，做為研究對象進行分析。三家飯店各抓取的資料為評論填寫日期、入住月份、分數、標題、正面評論以及負面評論。將蒐集好的資料，進行資料前處理及文字探勘。

## 三、文字探勘流程

經過篩選後取得三家飯店的評論資料後，「義大皇家酒店」有 468 則評論、「台北凱撒大飯店」有 337 則評論、「墾丁福華渡假飯店」有 798 則評論，總計 1,603 則有效評論。因評論資料為文字型態，因此本研究利用文字探勘的方式進行

評論資料的處理。

目前最常被使用的中文斷詞系統有兩種，分別是中研院所開發的 CKIP Tagger，以及為使用 Python 語言所開發的 Open Source，結巴 (Jieba) 中文斷詞系統 (<https://github.com/APCCLab/jieba-tw>) (Day & Teng 2017)。本研究採用結巴斷詞系統進行文字探勘的資料處理，雖然結巴與 CKIP 的斷詞一樣都有詞性標註，但是結巴斷詞可以依照自身的研究資料自訂詞彙，並建立自建詞典，因本研究的資料為住宿的評論資料，有不少口語化的用詞，常需使用到自建詞彙，所以採用結巴斷詞系統。

首先採用結巴中文斷詞系統將三家飯店的所有評論進行第一次的斷詞處理，共拆解出 43,924 個詞彙。接下來進行詞彙的前置處理，處理過程如下：

1. 刪除無意義的字彙，利用中文停用詞表，將標點符號、表情符號等特殊詞彙刪除。如：「&」、「.....」等，共刪除 20,183 筆詞彙，剩下 23,741 筆詞彙。
2. 第二次斷詞，在刪除無意義字彙後，檢視第一次的斷詞結果，並增加旅遊住宿常用詞彙至自定義詞庫內，例如：「摩天輪」、「接駁車」、「客訴」等，再結合結巴語料庫及自定義詞庫進行第二次的斷詞。共縮減 262 筆詞彙，剩下 23,479 筆詞彙。
3. 整合重複詞彙，從第二次斷詞後剩餘的 23,479 筆詞彙裡，刪除停用字後，依詞性及詞頻整合重複出現的詞彙，整合後剩 4,766 筆詞彙。並刪除外語（如 XDD）、數詞（如 8501）以及無法判別的特殊字（如㗎），等共 474 筆，剩餘 4,292 筆詞彙。
4. 刪除單一詞彙，將文字長度為一且不具代表性的詞彙視為單一詞彙；如：「雙」、「找」、「蓋」等，共刪除 272 筆詞彙，剩下 4,020 筆詞彙。
5. 詞性分析，從 4,020 筆詞彙中篩選出名詞共 1,803 筆，進行同義字合併，如：「風景」、「景觀」、「景色」視為同義詞彙，合併後剩下 1,763 筆詞彙。
6. 找出高頻名詞，從剩餘的 1,763 筆詞彙中，計算出詞頻平均數為 5.21，故本研究取詞頻數為 6 以上的詞彙做為高頻名詞，共 202 筆詞彙。

經過上述前置處理過後，共有 202 個詞彙為三家飯店 Booking.com 評論內容的高頻名詞，並將高頻名詞分別對三家飯店的評論內容進行比對與計數，建置評論與高頻名詞矩陣。本研究將評論及高頻名詞矩陣轉為二元矩陣處理，當高頻名詞出現於該則評論當中時以 1 表示，未出現則以 0 來表示。

#### 四、圖形化分析

本研究利用相似性分析計算高頻名詞之間的距離矩陣，再運用多維尺度分析

法找出高頻名詞的座標位置，並利用詞性為形容詞的關鍵詞整合情感字典，計算出形容詞的正負向情感值，最後再以社會網絡分析圖概念呈現各家飯店的分析結果。

先利用相似性分析計算高頻名詞之間的距離矩陣，再找出該高頻名詞相對應的座標位置。相似性的部分是使用 Jaccard similarity (Jaccard 1901) 來計算，Jaccard similarity 可以計算出兩高頻名詞間的相似度，計算出的結果介於 0 到 1 之間，而 Jaccard distance 為相異性，與 Jaccard similarity 相反，利用 Jarrard distance 相異性計算出高頻名詞間的距離，並轉換成距離矩陣。設高頻名詞  $i$  與高頻名詞  $j$  皆出現的文章篇數以  $t$  表示；高頻名詞  $i$  有出現且高頻名詞  $j$  沒有出現的文章篇數以  $r$  表示；高頻名詞  $i$  沒有出現且高頻名詞  $j$  有出現的文章篇數以  $s$  表示。

令  $S_{i,j}$  為高頻名詞  $i$  與高頻名詞  $j$  間的 Jaccard similarity (相似值)，令  $D_{i,j}$  為 Jaccard dissimilarity (相異值)，相異值越大，呈現距離越遠；反之，相異值越小，則呈現距離越近，相似值及相異值的計算方式如公式(1)與(2)。本研究利用 R 語言計算相似性與距離矩陣。

$$S_{i,j} = \frac{t}{r+s+t} \quad (1)$$

$$D_{i,j} = 1 - \frac{t}{r+s+t} \quad (2)$$

接著，本研究運用多維尺度分析法 (Borg & Groenen 1997; Cox & Cox 2000; Ma 2012)，找出高頻名詞的座標位置。多維尺度分析法屬於統計方法中多變量分析的方法之一，把高維度的樣本映射到一個低維度的空間上，同時又保留樣本間的原始關係，是一種維度縮減方法，同時也是一種視覺化方法，被廣泛應用在社會學、行銷學等統計實證分析中。多維尺度分析法主要先計算兩兩樣本間的相異性，以建構距離矩陣，接著轉換成內積矩陣，然後做奇異值分解 (singular value decomposition)，以找出對應的投影方向。這樣的統計降維法，可以用來以圖形化的方式呈現樣本間的關聯性，兩樣本間的距離越短，代表相似性越高。多維尺度分析法分為公制 (metric) 與非公制 (nonmetric) 兩類型，前者採用歐幾里得距離，後者則泛指任何非歐幾里得距離。本研究採用較常用的公制式 MDS，亦稱為經典多維尺度分析法 (Classic MDS)，分析工具為 R 語言。

在高頻名詞視覺化呈現部份，本研究採用社交網絡圖的概念，來呈現各家飯店高頻名詞間的關係，利用圓形圖示表示高頻名詞，而圓形圖示的大小來表示該高頻名詞於評論裡出現的頻率，高頻名詞之間的距離代表相似性，兩兩高頻名詞越接近代表相似性越高。本研究所使用的圖形化工具為社會網絡分析常用的 NodeXL。

## 五、情感分析

本研究經由文字探勘分析後，將所得出的結果以圖形化方式呈現出來，每間飯店於評論當中，常被討論的詞彙會依照頻率高低以圓形大小顯示，但是所呈現的結果僅能代表常被提及的高頻名詞而已，例如，「房間」為高頻名詞，但無法知道評論中，對於該飯店房間的評論是好或是不好的，也不知房間是那方面好或是不好。為了更進一步了解該詞彙出現在評論中的描述是正向還是負向的，本研究利用情感分析方法找出描述各個高頻名詞的正、負向形容詞，並且依否定詞及程度副詞算出加權情感值，最終將形容詞連結名詞，以圖形化的方式呈現。

本研究利用常用的字典法進行情感分析，為了計算所代表的正負向情感程度，將情感詞典細分為四個部分：正向形容詞、負向形容詞、否定詞以及程度副詞情感詞典。本研究之情感分析架構圖如圖 3 所示，先將文字探勘中的詞性分析結果，篩選出形容詞表，進行合併同義字後，利用國立台灣大學情感詞典（National Taiwan University Semantic Dictionary; NTUSD）比對形容詞表，NTUSD 是由臺灣大學整理並建立的情感詞典（<http://nlg.csie.ntu.edu.tw/download.php>），經過比對後找出正向形容詞及負向形容詞，再將形容詞表中未列在 NTUSD 詞典中的詞彙，以自建詞典的方式加入。自建詞典的方面，本研究使用多數決判斷的方式，請三位相關研究者，判斷該詞彙為正向、負向或者無法判斷的詞彙，判斷過程中若有一位或一位以上學者認為該形容詞無法判斷，則該形容詞設定為無法判斷將不列入詞典；若有兩位以上的學者認為是正向，則該形容詞設定為正向形容詞，否則反之，最後整合出正、負向的形容詞詞典。

為計算正負向情感值，首先將評論內容拆解成句子，並找出包含高頻名詞的句子，將句子拆解成斷詞，並利用正、負向形容詞詞典進行比對，表 2 列出本研究正向形容詞及負向形容詞的範例及其形容詞值，正向形容詞包括方便、舒適、乾淨等，負向形容詞包括老舊、髒、昏暗等；若比對為結果為正向形容詞，形容詞值設為 1，若為負向形容詞則形容詞值設為-1。再以否定詞典進行比對，本研究採用 Yong 等（2014）建立的否定詞典，範例如表 3 所示，否定詞包含不、沒、無等詞彙，若形容詞前出現否定詞，否定詞值設為-1，若未出現否定詞，則否定詞值設為內定值 1。最後以程度副詞詞典進行比對，本研究利用著名的中文情感字典「How Net 知網」（[http://www.keenage.com/html/c\\_index.html](http://www.keenage.com/html/c_index.html)）的程度級別詞語做為基礎，知網將程度副詞分為六個等級，共 219 個中文程度副詞，經由本研究整理後，將較雷同的級別整合成三個等級，本研究的程度副詞級別分成一到四等級，如表 4 所示，分別是整合知網的三個級別以及未出現程度副詞的內定值級別，程度副詞值對應級別的範圍從 0.75 到 1.5，若形容詞前出現程度副詞，便依照程度級別設定程度副詞值。

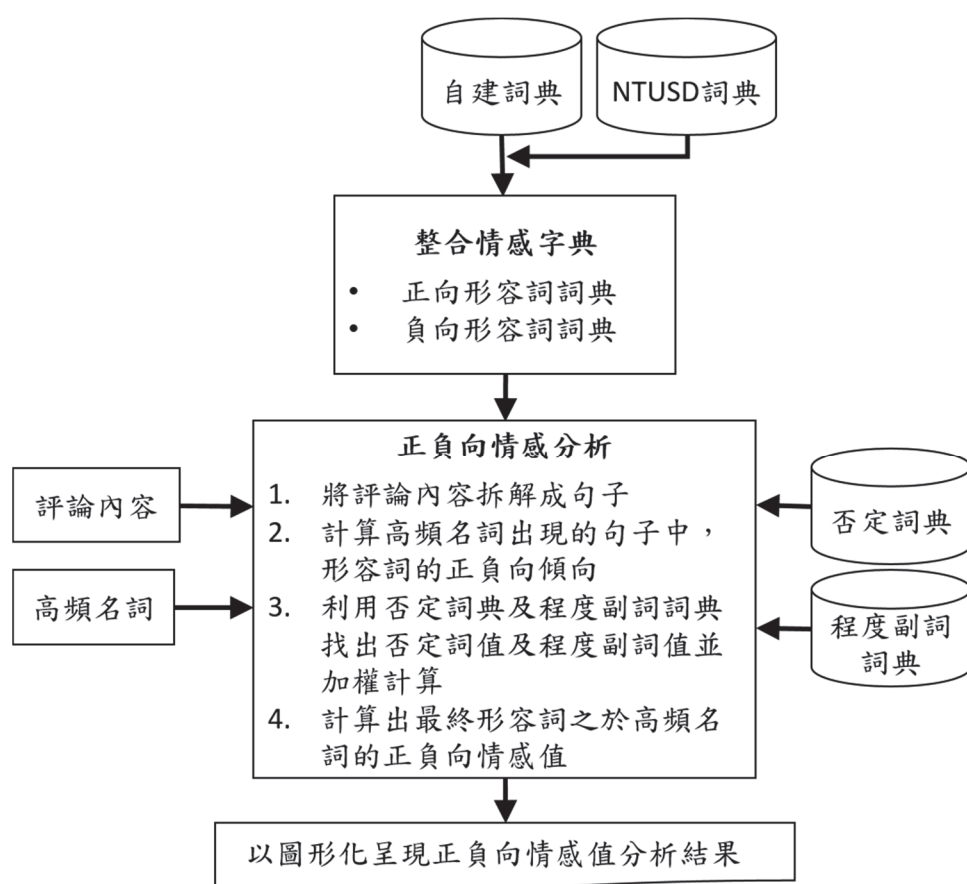


圖 3：本研究情感分析架構圖

本研究將形容詞的正負向情感值，定義為以正負向形容詞值乘以否定詞值，再乘以程度副詞值，計算方式如公式(3)，代表該形容詞之於某一高頻名詞的正負向情感，各個形容詞的正向情感值的範圍會落在 0.75 到 1.5 之間，而負向情感值的範圍會落在 -1.5 到 -0.75 之間。

表 2：本研究正、負向形容詞範例及形容詞值列表

| 類別    | 範例                    | 形容詞值 |
|-------|-----------------------|------|
| 正向形容詞 | 方便、舒適、乾淨、親切、豐盛、寬敞、整潔等 | 1    |
| 負向形容詞 | 老舊、髒、昏暗、無聊、糟糕、潮濕、難吃等  | -1   |

表 3：本研究否定詞範例及否定詞值列表

| 類別   | 範例                     | 否定詞值  |
|------|------------------------|-------|
| 否定詞  | 不、沒、無、非、莫、勿、未、別、難道、沒有等 | -1    |
| 非否定詞 | 未出現否定詞                 | 內定值 1 |

表 4：本研究程度副詞範例及程度副詞值列表

| 等級 | 範例               | 程度副詞值 |
|----|------------------|-------|
| 4  | 超級、非常、極、備感、完完全全等 | 1.5   |
| 3  | 很、可、多、太、特、何等、格外等 | 1.25  |
| 2  | 未出現程度副詞          | 內定值 1 |
| 1  | 略、略為、稍、稍為、有點、有些等 | 0.75  |

$$\text{正負向情感值} = (\text{程度副詞值}) \times (\text{否定詞值}) \times (\text{正負向形容詞值}) \quad (3)$$

正負向情感值的計算方式，舉例說明如表 5 所示，表中列出三句評論內容，及其正負向情感值的計算方法。例句 1「這間房間乾淨」中，「房間」為名詞，例句中未出現程度副詞及否定詞，只出現正向形容詞「乾淨」，故形容詞值為(1)，否定詞值及程度副詞值均用內定值(1)，故正負向情感值為 $(1) \times (1) \times (1) = (1)$ ，表示用「乾淨」形容「房間」為正向情感，情感值為 1。例句 2「餐廳非常不乾淨」中，「餐廳」為名詞，形容詞「乾淨」為正向形容詞，在形容詞前，出現否定詞「不」及程度副詞「非常」，於是將形容詞值(1)乘以否定詞值(-1)後，再乘以程度副詞對應到的程度副詞值 (1.5)，得出正負向情感值為(-1.5)，即用「乾淨」形容「餐廳」的情感值為 -1.5。例句 3「停車位有點少」中，「停車位」為名詞，形容詞「少」為負向形容詞，並且形容詞前有程度副詞「有點」，將正負向形容詞值(-1)乘以否定詞內定值(1)，再乘以所對應的程度副詞值(0.75)，可以得出最終的正負向情感值(-0.75)，代表用「少」形容「停車位」的情感值為-0.75。

表 5：評論例句及正負向情感值計算

| No. | 例句                                    | 計算方法(公式 3-4)                   | 正負向情感值 | 形容詞之於名詞 |
|-----|---------------------------------------|--------------------------------|--------|---------|
| 1   | 這間 (房間)(乾淨)<br>名詞+形容詞                 | $(1) \times (1) \times (1)$    | +1     | 乾淨→房間   |
| 2   | (餐廳) (非常) (不) (乾淨)<br>名詞+程度副詞+否定詞+形容詞 | $(1.5) \times (-1) \times (1)$ | -1.5   | 乾淨→餐廳   |

|   |                              |                                 |       |       |
|---|------------------------------|---------------------------------|-------|-------|
| 3 | (停車位)(有點) (少)<br>名詞+程度副詞+形容詞 | $(0.75) \times (1) \times (-1)$ | -0.75 | 少→停車位 |
|---|------------------------------|---------------------------------|-------|-------|

設某飯店評論中共有  $K$  個句子包含高頻名詞  $N_j$  及形容詞  $A_i$ ，設在第  $k$  個包含  $N_j$  與  $A_i$  的句子中，形容詞  $A_i$  之於名詞  $N_j$  的正負向情感值為  $E_{i,j}^k$ ，設形容詞  $A_i$  之於名詞  $N_j$  的平均正負向情感值表示為  $E_{i,j}$ ， $E_{i,j}$  可依公式(4)計算，為  $K$  個  $E_{i,j}^k$  的總和除以總數  $K$ 。

$$E_{i,j} = \frac{\sum_{k=1}^K E_{i,j}^k}{K} \quad (4)$$

計算出正負向情感值之後，再以圖形化方式將結果呈現出來，在原本第一階段圖形化呈現的結果中，於該高頻名詞周圍畫出相對應的形容詞，以正方形表示正向、三角形表示負向，並以圖形的大小來表示情感值的高低，正方形越大表示正向情感值越高，若三角形越大則表示負向情感值越高。

值得注意的是本研究針對高頻名詞是先分開計算處理的，故若一個句子中出現了兩個以上的高頻名詞都是個別計算，例如，一個句子中出現「房間很大但是廁所很暗」，名詞「房間」與「廁所」是會先分開計算，但是形容詞部份僅會計算最靠近該名詞的形容詞情感值，以上述例句來說，會視為「很大」用來形容「房間」，「很暗」用來形容「廁所」。對於同一句子中若有兩個以上形容詞用來形容同一名詞，則會忽略距該名詞較遠的形容詞。

## 肆、研究結果與分析

本研究結果分為摘要統計分析、第一階段高頻名詞的圖形化分析結果以及第二階段圖形化情感分析結果，分述如下。

### 一、摘要統計分析

本研究蒐集三家飯店一年內之評論數共 1,603 則評論，分析所有評論後得到 1,763 個關鍵詞彙，名詞詞頻平均數為 5.21，根據詞頻平均數選定 202 個出現頻率超過平均數的高頻名詞，表 6 顯示排名前 10 名高頻名詞的詞頻次數，其中「房間」出現頻率為 613 次最高，其次是「早餐」出現頻率為 463 次，「員工」則出現 353 次，可見顧客對飯店的整體評論中，最重視早餐、房間及員工，被提及的次數最多。接著分別於三家飯店找出個別的高頻名詞，各家飯店前 10 名的高頻名詞統計表如表 7，由表中可以看出，「早餐」、「房間」、「員工」、「設施」以及「服務」皆共同出現在三間飯店的排行前十名內；台北凱撒大飯店的「位置」、「交通」、「車站」是其餘兩家飯店沒有出現的詞彙，可能因該飯店位於台北車站附

近，交通便利位置佳；而墾丁福華渡假飯店比較特別的是「沙灘」，可能因其臨近沙灘，深受喜愛。

表 6：前十名高頻名詞詞頻次數統計表

| 排名 | 詞彙                      | 出現頻率 |
|----|-------------------------|------|
| 1  | 房間                      | 613  |
| 2  | 早餐                      | 463  |
| 3  | 員工/人員/工作人員/服務人員/服務員/服務生 | 353  |
| 4  | 設施/酒店設施/公共設施/設備         | 321  |
| 5  | 位置/地理位置/地點/地方           | 300  |
| 6  | 床                       | 202  |
| 7  | 小孩/小朋友/小孩子/兒童/孩童/孩子     | 169  |
| 8  | 游泳池/泳池                  | 168  |
| 9  | 服務                      | 168  |
| 10 | 浴室/淋浴間/衛生間/淋浴/衛浴        | 148  |

本研究在經由結巴斷詞後，所得到共 282 個形容詞，形容詞詞頻平均數為 8.58，根據詞頻平均數，選取詞頻為 8 及以上的形容詞定為高頻形容詞，共有 38 個形容詞，其中正向形容詞有 25 個，負向形容詞有 7 個，6 個形容詞無法判斷其正負傾向，故最終篩選出的高頻形容詞共有 32 個。高頻形容詞列表如表 8。

表 7：三家飯店前十名高頻名詞

| 名次 | 義大皇家酒店 | 台北凱撒大飯店 | 墾丁福華渡假飯店 |
|----|--------|---------|----------|
| 1  | 早餐     | 位置      | 房間       |
| 2  | 房間     | 房間      | 早餐       |
| 3  | 員工     | 員工      | 設施       |
| 4  | 設施     | 交通      | 員工       |
| 5  | 床      | 設施      | 沙灘       |
| 6  | 小孩     | 台北      | 位置       |
| 7  | 服務     | 早餐      | 游泳池      |
| 8  | 游泳池    | 浴室      | 床        |
| 9  | 浴室     | 服務      | 服務       |
| 10 | 態度     | 車站      | 小孩       |

表 8：高頻形容詞表

| 形容詞 | 形容詞值 | 形容詞 | 形容詞值 | 形容詞 | 形容詞值 | 形容詞 | 形容詞值 |
|-----|------|-----|------|-----|------|-----|------|
| 方便  | 1    | 佳   | 1    | 清楚  | 1    | 豐盛  | 1    |
| 舒適  | 1    | 豐富  | 1    | 精緻  | 1    | 老舊  | -1   |
| 不錯  | 1    | 整潔  | 1    | 優越  | 1    | 不足  | -1   |
| 乾淨  | 1    | 友善  | 1    | 寬敞  | 1    | 髒   | -1   |
| 大   | 1    | 安全  | 1    | 便利  | 1    | 貴   | -1   |
| 棒   | 1    | 衛生  | 1    | 容易  | 1    | 發黴  | -1   |
| 親切  | 1    | 耐心  | 1    | 良好  | 1    | 昏暗  | -1   |
| 舒服  | 1    | 好   | 1    | 明顯  | 1    | 不滿  | -1   |

## 二、第一階段高頻名詞的圖形化分析結果

本研究先利用 NodeXL 圖形化呈現三家飯店於評論中常出現的高頻名詞及其之間的關係。本研究將三家飯店常出現的高頻名詞，以前 20 名高頻詞彙為例進行圖形化呈現，圖中點的大小代表此高頻名詞於該飯店評論中出現的筆數，點越大表示此高頻名詞出現的評論則數越多，點越小則反之。圖中兩點之間的距離，代表兩兩高頻名詞之間的相似度，兩點間的距離越近，表示兩者間的相似性就越高，即高頻名詞同時出現在評論中的篇數愈多。

三家飯店的圖形化分析結果分述如下。義大皇家酒店的圖形化分析結果，如圖 4 所示，由圖形中點的大小可以看出「早餐」、「房間」、「員工」的點較大，分別為義大皇家酒店出現的高頻名詞前三名，即高頻名詞於評論中出現的頻率以「早餐」最高，故圖形呈現出該點為最大。高頻名詞兩兩之間的距離代表其相似性，距離越近表示相似性越高，即表示兩詞會同時出現在評論中的頻率越高，由圖形中可以看出「小孩」與「游泳池」的距離相對「小孩」與「景觀」的距離較近，在相似性分析中的距離矩陣顯示「小孩」與「游泳池」的相異值為 0.16，而「小孩」與「景觀」的相異值為 0.74。

在義大皇家酒店的分析圖中，高頻名詞「小孩」、「設施」、「游泳池」的點很相近，代表帶小孩的顧客可能較在意飯店的「設施」與「游泳池」；高頻名詞「購物中心」及「遊樂」的點很相近，可能因該飯店附設有購物中心和遊樂園，故常一起被提及。對消費者來說，可以由此圖看出這家飯店有小孩喜愛的游泳池與設施，亦有遊樂園及購物中心，適合親子旅遊。對企業管理者來說，可以針對親子旅遊設計相關的活動或促銷方案，例如，在購物中心消費滿一定金額送遊樂園門票等。

台北凱撒大飯店的圖形化分析結果，如圖 5 所示，台北凱撒大飯店出現的高頻名詞前三名為「位置」、「房間」、「員工」，由圖中可以看出這三個點較大。而從圖形中可以看出，「電梯」、「捷運」及「出口」於評論中常被顧客一起討論，可能因為該飯店近捷運出口且有電梯可直達飯店；此外，「台北」及「車站」的位置相近，可能因該飯店位於台北車站附近地點便利，故顧客常在評論中一同提及。對消費者來說，由此圖可以看出這家飯店靠近台北車站，近捷運交通便利，適合不開車自助遊或商務旅遊。對企業管理者來說，可以針對自由行或商務客設計相關的活動或促銷方案，例如，住宿送捷運一日券或提供短程旅遊規劃服務等。

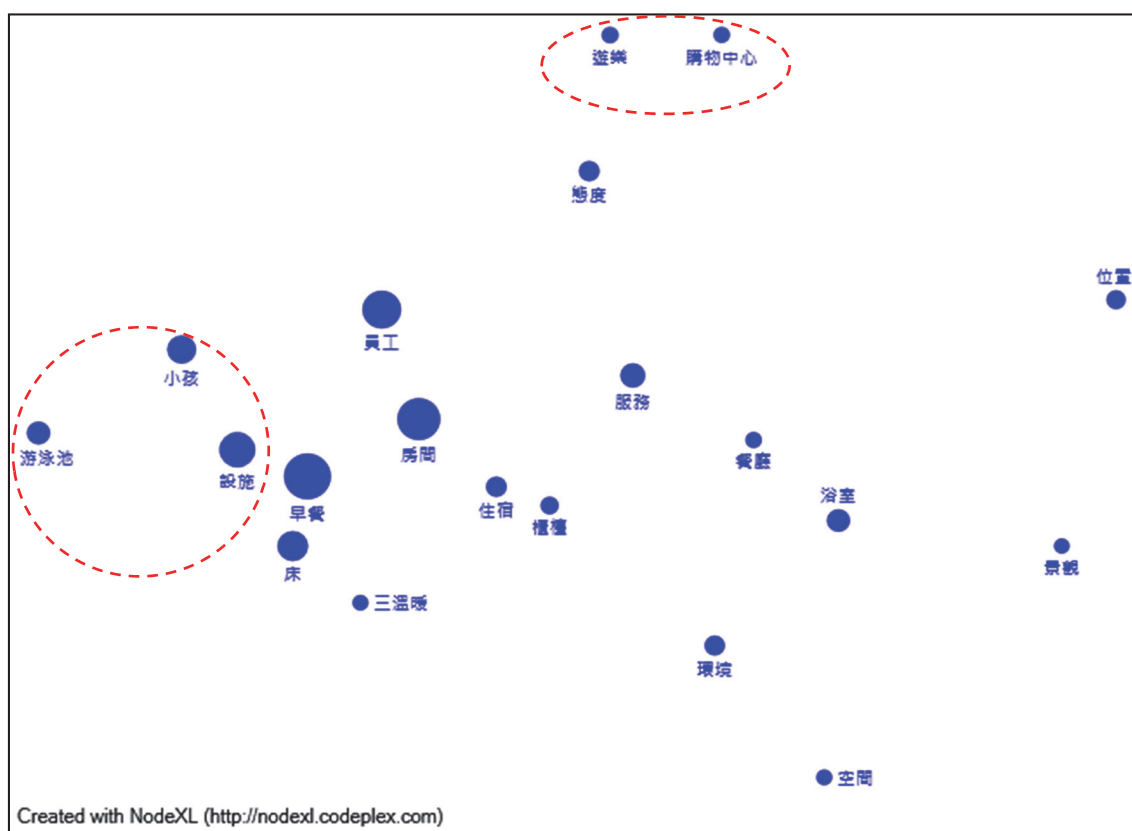


圖 4：義大皇家酒店之第一階段圖形化分析結果

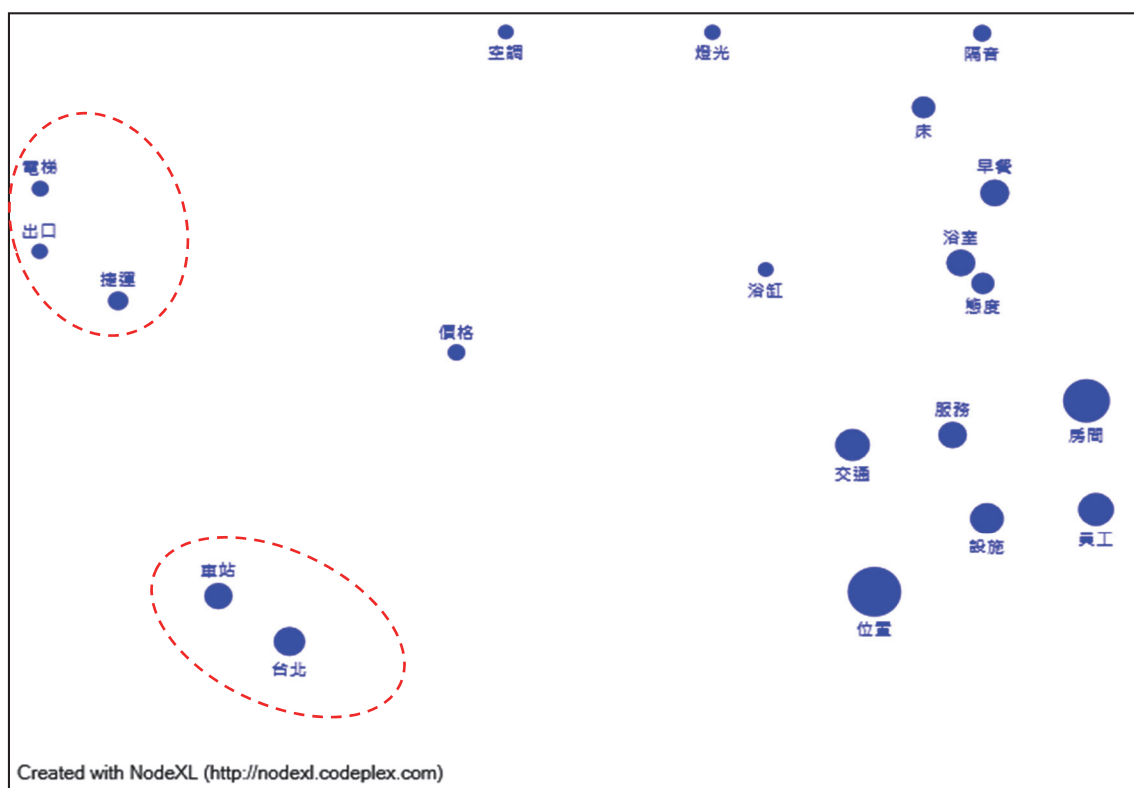


圖 5：台北凱撒大飯店之第一階段圖形化分析結果

墾丁福華渡假飯店的圖形化分析結果，如圖 6 所示，墾丁福華渡假飯店的高頻名詞前三名為「房間」、「早餐」、「設施」。從圖中可以看出與其他兩間飯店不同的是「位置」、「沙灘」、「小灣」、「墾丁大街」距離很近，可能因這間飯店的位置很靠近小灣沙灘與墾丁大街；「房間」與「空間」常同時被提及，可能因為該飯店房間空間很大；「小孩」與「設施」亦常一起被討論，可能因為該飯店有不少受小孩喜愛的設施。對消費者來說，由此圖可以看出這家飯店靠近小灣、沙灘與墾丁大街，地點很適合喜歡戲水或逛墾丁大街的遊客，房間大且有小孩遊戲設施，亦適合親子旅遊。對企業管理者來說，可以針對喜愛戲水、逛墾丁大街或親子旅遊的顧客，設計相關的活動或促銷方案，例如，提供水上活動服務或住宿送墾丁大街消費抵用券等。

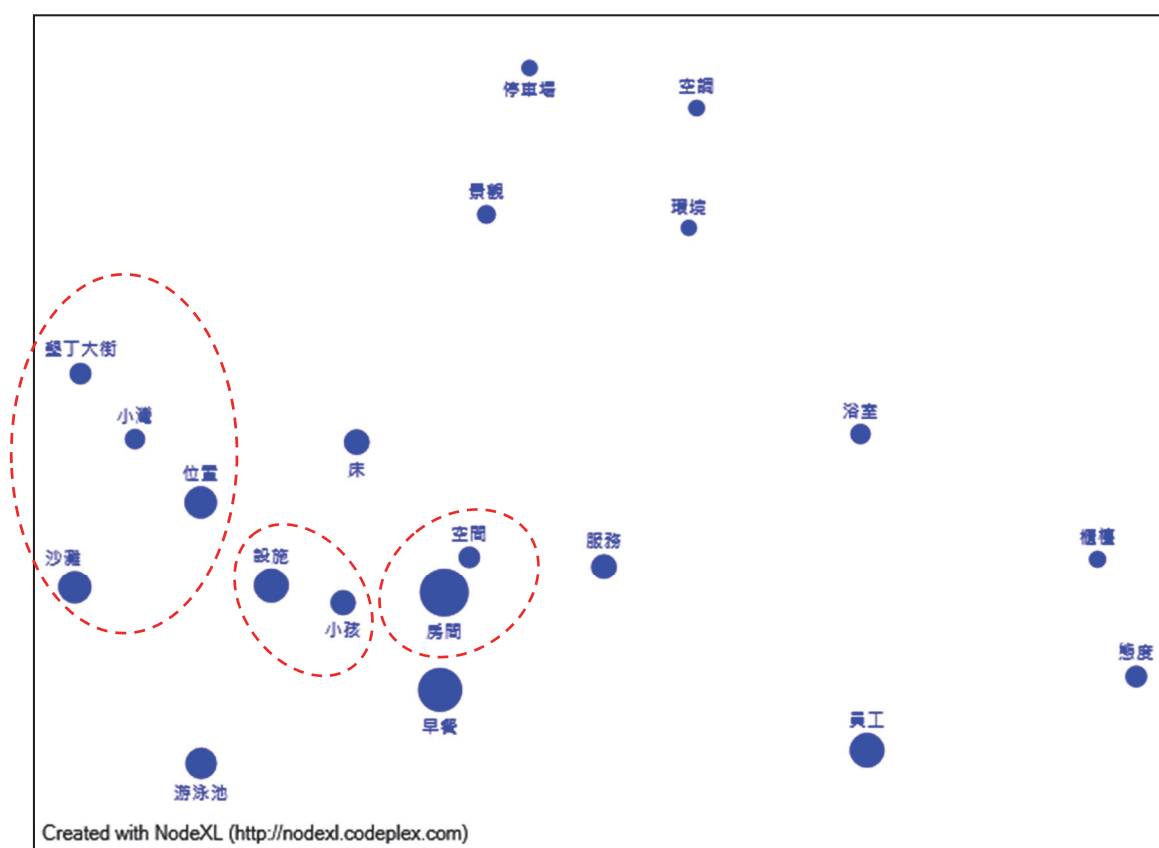


圖 6：墾丁福華渡假飯店之第一階段圖形化分析結果

### 三、第二階段圖形化情感分析結果

完成第一階段高頻名詞的圖形化分析後，第二階段納入形容詞、副詞與否定詞進行情感分析。先計算各形容詞的正負向情感值，為簡化圖形本研究只取出現篇數為 2 篇以上的高頻形容詞進行圖形化呈現。在圖形化呈現中，圓形圖示代表名詞的部分，點的大小代表該高頻名詞於該飯店評論中出現的篇數；正方形代表正向形容詞，三角形代表負向形容詞，圖形的大小代表情感值的高低，圖形越大表示該高頻形容詞的情感值越高、情感程度越強，圖形越小則反之，並在圖形旁標示情感值；線條的寬度，代表該形容詞出現於評論中的篇數，線越粗表示篇數越多，線越細則反之。

義大皇家酒店的圖形化情感分析結果，如圖 7 所示，圖形中可以看出形容「房間」的形容詞有「大」、「舒適」、「乾淨」以及「好」，其中形容詞「大」為正向形容詞，該形容詞於評論中的篇數為 19 篇，情感值為 1.21，相較於形容「房間」的其他形容詞，其情感值最高，故形容詞「大」的圖形較其他形容詞的圖形

大，表示「大」的情感程度較強。另外形容「房間」的形容詞「好」，從圖形中可以看見有正方形以及三角形，代表在評論中同時有正向以及負向的形容詞在形容「房間」，形容詞「好」形容「房間」的正負向情感值為正向的 1.15，以及負向的 -0.96，表示雖然有負向評論存在，但負向的程度不及正向的強度，本研究將正、負向形容詞都呈現在圖上，並未將其合併相消滅，是為了將結果完整呈現出來，讓使用者可以清楚的了解正、負向評論的情感強度。而形容「設施」的形容詞有「好」、「不錯」、「老舊」以及「不足」，表示顧客的評論中認為義大皇家酒店的設施很好、還不錯，但是有些老舊與不足。圖中可以很清楚的看出在「床」的部分，有 41 篇形容「床」很「舒服」的評論，且其情感值為 1.22，表示顧客對義大皇家酒店的床舒適程度有很高的評價。

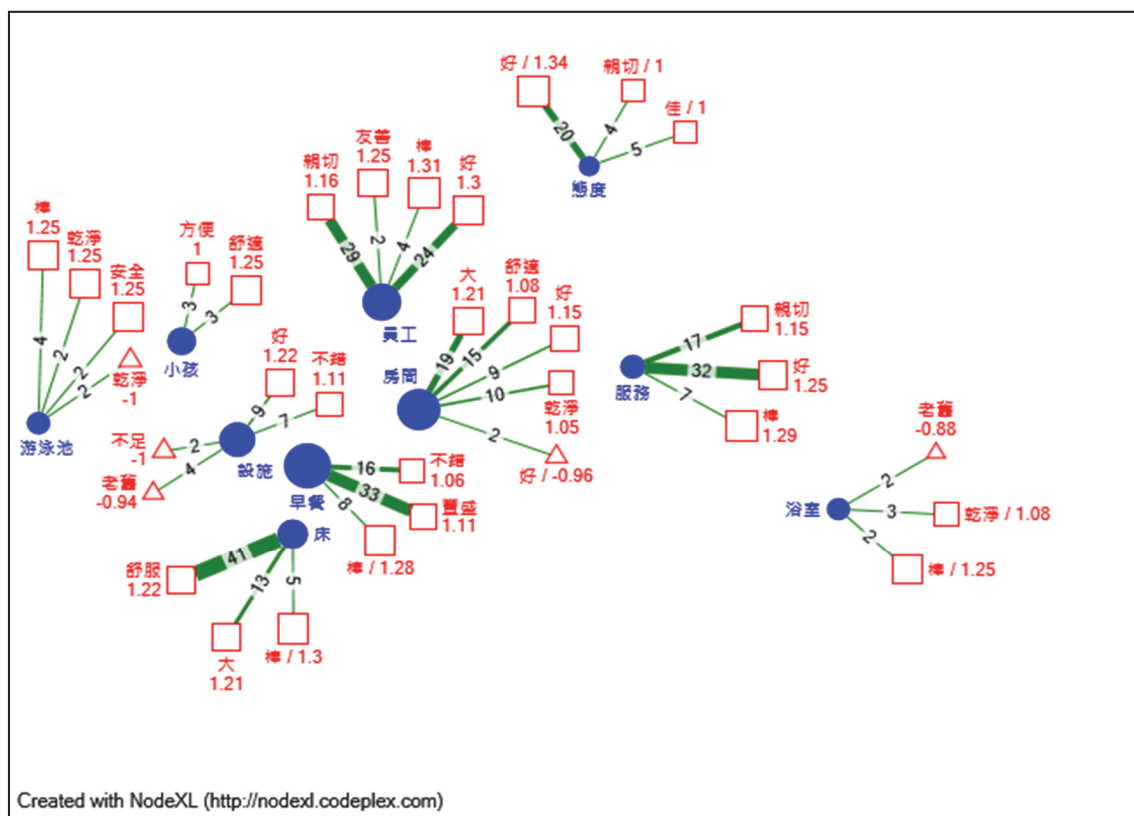


圖 7：義大皇家酒店之第二階段圖形化情感分析結果

對消費者來說，由此圖可以看出很多人評論這家飯店的床很舒服、早餐豐盛、服務很好、員工親切，適合較重視睡眠品質、早餐與服務的消費者。對飯店管理者來說，可以看出部份顧客對設施老舊、浴室老舊與游泳池不乾淨不滿，建議更新設施與浴室，並加強游泳池清潔管理。

台北凱撒大飯店第二階段的圖形化情感分析結果，如圖 8 所示，由圖形中可以看出形容「房間」的形容詞有「大」、「乾淨」、「方便」以及「老舊」，其中形容詞「方便」的情感值為 1.2，相較於形容「房間」的其他正向形容詞，其情感值最大，故形容詞「方便」的圖形較其他形容詞的圖形大，表示「方便」的情感程度較強，但該形容詞於評論中的篇數只有 5 篇，連線較細表示評論數不多；另外形容「房間」的形容詞「老舊」為負向形容詞，從圖形中可以看出該形容詞的情感值為 -1.34，其情感值很大，代表顧客對於房間老舊的程度已經對此感到很不滿，且有 6 篇評論認為房間很老舊。形容「設施」的形容詞有「不錯」、「棒」、「好」以及「老舊」，雖然「老舊」的評論篇數有 13 篇，但是其情感值為 -0.98，情感程度不算高，表示在顧客的評論中認為台北凱撒大飯店的設施很棒、還不錯，但是有點老舊。而圖中可以很清楚的看出在「位置」的部分，形容詞有「好」、「方便」、「不錯」、「棒」以及「優越」，有 56 篇形容「位置」很「好」的評論，且其情感值為 1.23，而以「棒」形容「位置」的情感值最為高，情感值為 1.35，表示關於台北凱撒大飯店，顧客對於其飯店位置給予很高的評價。

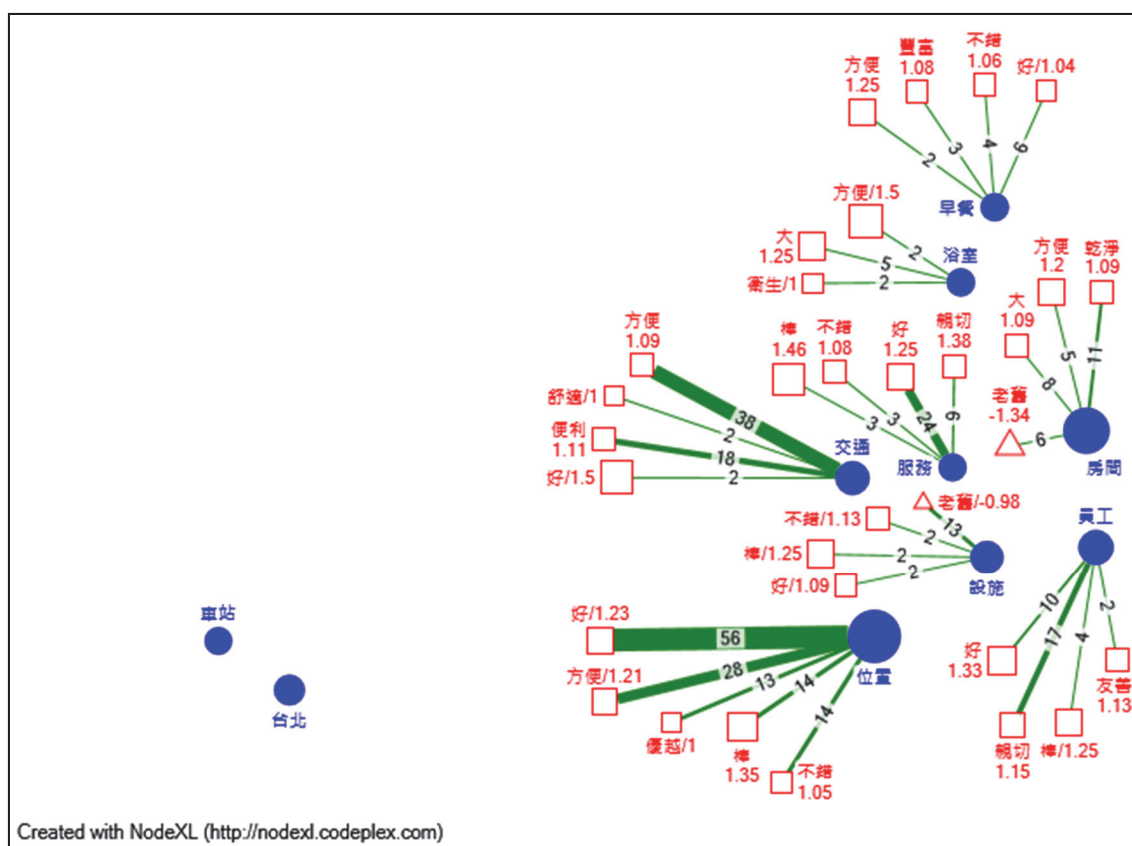


圖 8：台北凱撒大飯店之第二階段圖形化情感分析結果

對消費者來說，由此圖可以看出很多評論提及這家飯店的位置很好、交通方便、服務很好、員工親切，適合自助旅行、商務出差及重視服務品質的消費者。對飯店管理者來說，可以看出部份顧客對設施老舊與房間很老舊不滿，建議應編列經費翻修房間與更新設施。

墾丁福華渡假飯店第二階段的圖形化情感分析結果，如圖 9 所示，從圖形中可以很明顯地看出形容「房間」的形容詞有「大」、「舒適」、「寬敞」、「乾淨」以及「老舊」，其中形容詞「大」的評論篇數高達 90 篇，雖然其情感值為 1.16，正向的情感強度不算很強，但相較於形容「房間」的其他形容詞，其情感值最大；另外形容「房間」的形容詞「老舊」為負向形容詞，從圖形中可以看出該形容詞的情感值為-0.98，雖然情感強度不算太強烈，但是也有 10 篇評論在講述房間老舊，表示關於墾丁福華渡假飯店的評論，多數顧客認為房間很大，且舒適、寬敞、乾淨，但是有點老舊。而形容「服務」的形容詞有「好」、「親切」、「棒」以及「不滿」，對於「服務」感到「不滿」的評論篇數雖然只有 2 篇，但是其情感值

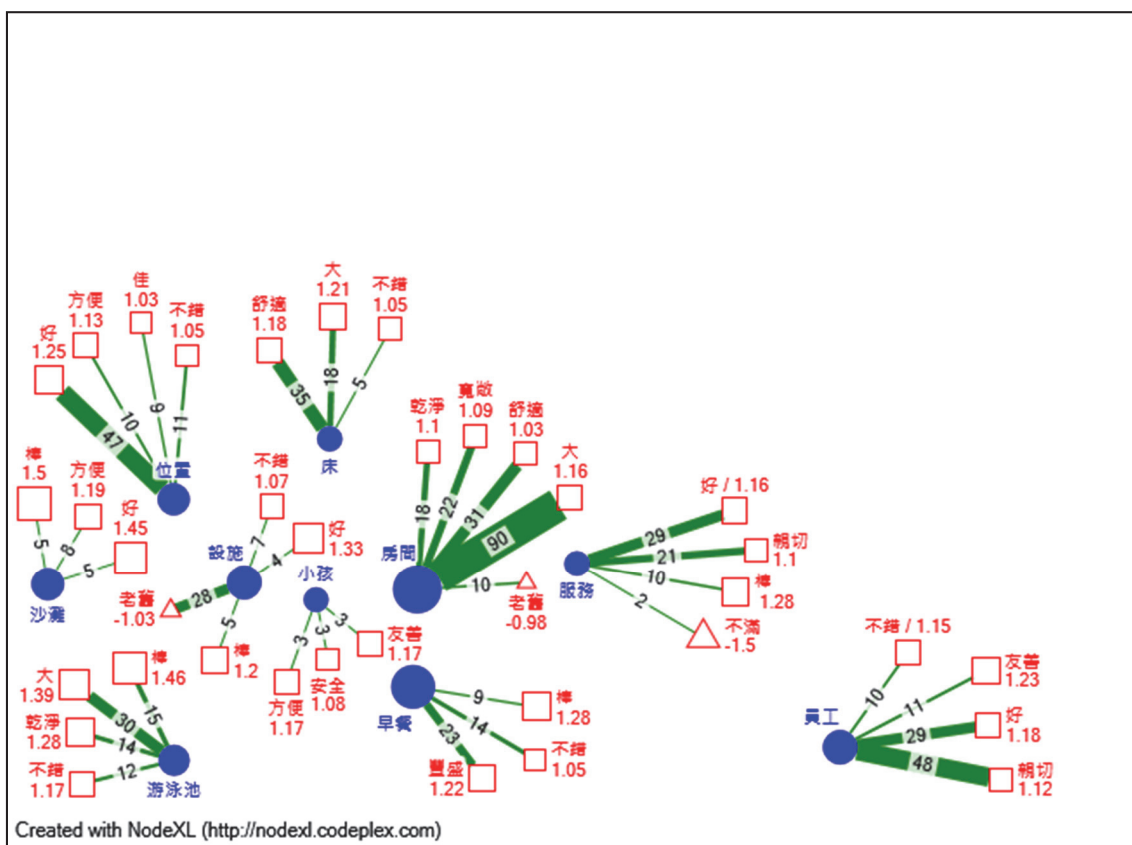


圖 9：墾丁福華渡假飯店之第二階段圖形化情感分析結果

為最高程度的-1.5，表示在墾丁福華渡假飯店的評論中，有少數顧客對於該飯店的服務感到非常不滿，但是大部分的顧客認為服務很好、親切、很棒。而該飯店相較於其他飯店沒有的名詞為「沙灘」，圖中可以看出形容「沙灘」的形容詞有「方便」、「棒」以及「好」，其中形容詞「棒」跟「好」的情感值各為 1.5 以及 1.45，表示關於墾丁福華渡假飯店，顧客對於該飯店的沙灘有相當高的評價。

對消費者來說，由此圖可以看出很多評論顯示這家飯店的房間大、員工親切、位置很好、床舒適、游泳池很大很乾淨，很適合親子旅遊、喜歡戲水或重視睡眠品質的消費者。對飯店管理者來說，可以看出不少顧客覺得設施老舊，應儘速加強設施更新或維護。

綜合上述分析，在三家飯店相同處部份，三家飯店的早餐均不錯，很豐盛；員工都很親切，友善，人也很好；服務好、親切、很棒，唯有少數評論對墾丁福華渡假飯店的服務感到不滿；三家飯店的房間均很大且乾淨，但是有一些評論認為台北凱撒大飯店及墾丁福華渡假飯店的房間老舊；而在設施部份，三家飯店的評論均認為設施很好、還不錯，但是也均認為設施較老舊。顯示對於這三家飯店而言，早餐、員工、服務都相當不錯，是共同的優點，但是設施老舊都是需要去改善的部份。

在相異處部份，由分析結果可以看出，在義大皇家酒店部份，顧客評論員工態度很好受到肯定，但是浴室的部份有點老舊，可考慮進行浴室的翻新工程，進一步的改善設備環境；而在台北凱撒大飯店部份，由分析結果可以看出，有 56 篇形容位置很好的評論，代表該飯店的位置相當好、很方便，且交通也很便利，受到青睞；最後關於墾丁福華渡假飯店的部份，由分析結果可以看出，顧客評論認為沙灘非常棒、很方便，且游泳池很大、很棒，可能因墾丁福華渡假飯店臨近沙灘，飯店本身也設有游泳池設備，故顧客對於沙灘及游泳池有很高的評價，是該飯店獨具的特色。

## 伍、結論

本研究提出了一個圖形方法來摘要化顧客評論並呈現在二維圖形上，並以台灣三大著名觀光飯店為例，分析臺灣三家知名觀光飯店的線上顧客評論，透過文字探勘的技術，以相似性分析及多維尺度分析法分析出高頻名詞及其座標位置，進行第一階段圖形化分析，呈現出各個高頻名詞之間的關係，再透過納入情感分析方法，依情感程度加權計算，第二階段再以圖形化方式呈現出各形容詞之於高頻名詞的正負向情感傾向。

第一階段圖形化分析結果顯示，在「義大皇家酒店」方面，常出現在該飯店評論的前三名高頻名詞為「早餐」、「房間」、「員工」，而「小孩」、「設施」及

「游泳池」常一起被討論，代表帶小孩的顧客可能較在意飯店的「設施」與「游泳池」；「購物中心」及「遊樂」亦常一起被提及，可能因該飯店附有購物中心和遊樂園。在「台北凱撒大飯店」方面，出現的高頻名詞前三名為「位置」、「房間」、「員工」，而「台北」及「車站」常在評論中一同被提及，「電梯」、「捷運」及「出口」亦常同時出現，可能因該飯店的地理位置於台北車站及捷運出口附近，位置便利。在「墾丁福華渡假飯店」方面，高頻名詞前三名為「房間」、「早餐」、「設施」，而「位置」、「沙灘」、「小灣」、「墾丁大街」常一起被提及，可能因這間飯店的位置很靠近小灣沙灘與墾丁大街；「房間」與「空間」常一起出現，可能因為該飯店房間空間很大；「小孩」與「設施」亦常一起被討論，可能因為該飯店有不少受小孩喜愛的設施。

第二階段圖形化情感分析結果顯示，在「義大皇家酒店」方面，顧客認為義大皇家酒店的床很舒服、早餐豐盛、服務很好、員工親切，但亦有部份顧客對設施老舊、浴室老舊與游泳池清潔不滿，建議更新設施與浴室，並加強游泳池清潔管理。在「台北凱撒大飯店」方面，顧客認為該飯店的位置很好、交通方便、服務很好、員工親切，但部份顧客對設施與房間老舊不滿，建議應編列經費翻修房間與更新設施。關於「墾丁福華渡假飯店」的評論，顧客認為該飯店的房間大、員工親切、位置很好、床舒適、游泳池很大且乾淨，但有不少顧客覺得設施老舊，建議應儘速更新或維護設施。

本研究結果可提供給飯店業者參考，飯店業者可透過呈現出來的圖形，瞭解住客的意見回饋以及評論重點，可以做為改進及調整營運策略的參考，以提昇企業的競爭優勢。此外，消費者可直接由圖形中看出該飯店的概觀評論結果，可快速地了解該飯店的評論重點，而不需要費時地一篇一篇瀏覽所有評論，並可依照自己的偏好做最適合的選擇。本研究方法可廣泛應用於其他產業上，研究架構亦可做為相關研究之參考。

在研究限制的部份，由於每則評論的內容字數不一致，有些評論內容甚至只有一個單詞，還有少數評論內容出現錯字的部份，無法透過電腦進行辨別，因此可能會造成些許影響；此外，由於中文字博大精深，結巴的斷詞系統所斷出來的詞性、詞頻結果常不如預期，故無法完全依靠斷詞系統產生的結果進行分析，常須配合人工檢查修正；另外，中文語意上有許多特例，例如：好不容易、好不快樂等詞彙，因特例詞彙的文句架構較為複雜且例外多，本研究並未將這些特例詞彙另行處理。再者，圖形化顯示方式常會有誤差，本研究目前並無法證實所繪出圖形當中的關係是完全正確的。

在未來研究方面，可朝下列幾個方向努力：(1)加入「入住月份」、「評分」等欄位資料，並結合機器學習方法，透過深度學習結合情感分析進行預測，分析當某些關鍵詞出現時，顧客給予的總體評論分數是否會受到影響；(2)本研究對於同

一句子中若有兩個以上形容詞用來形容同一名詞，則會忽略距該名詞較遠的形容詞，後續研究也許可朝均納入但採不同權重方式處理；(3)本研究尚未實證此方法是否實際能提高使用者閱讀的感受，後續研究亦可進一步分析探討；(4)可製作三家飯店的比較圖，作為飯店經營者策略規劃的參考。

## 誌謝

本研究受科技部計畫 MOST 108-2410-H-239-011-MY3 補助。

## 參考文獻

- 李佳儒 (2017),『文字探勘結合網路電影評論建立 R-Shiny 視覺化系統』,未出版碩士論文,輔仁大學統計資訊學系應用統計研究所,新北市。
- 李智明、玄燁、萬文隆 (2018),『兩岸消費者選擇訂房網站關鍵因素研究』,東吳經濟商學學報,第九十七期,頁 63-102。
- 林漪寒 (2016),『從使用者評論探勘競爭者』,未出版碩士論文,國立中山大學資訊管理學系研究所,高雄市。
- 莊維荃,2017。『網路評論影響你的購買決定?』,未出版碩士論文,國立中央大學資訊管理學系研究所,桃園市。
- 楊宗晟 (2019),『文字探勘在線上顧客評論分析之應用』,未出版碩士論文,明新科技大學資訊管理系,新竹縣。
- 葉珊妤 (2018),『以文字探勘與機器學習法於部落格遊記之研究』,未出版碩士論文,國立聯合大學資訊管理學系,苗栗縣。
- 蔡孟穎 (2018),『預測線上評論的幫助性:誰是關鍵評論者?』,未出版碩士論文,國立臺灣科技大學管理學院 MBA 研究所,台北市。
- 劉雨鑫 (2018),『廣告音樂歌詞與評論的商業性研究-Python 的應用』,未出版碩士論文,國立高雄大學國際高階經營管理碩士在職專班,高雄市。
- 鍾任明、李維平、吳澤民 (2007),『運用文字探勘於日內股價漲跌趨勢預測之研究』,中華管理評論國際學報,第十卷,第一期,頁 1-31。
- 謝麗星、周明、孫茂松 (2012),『基於層次結構的多策略中文微博情感分析和特徵抽取』,中文信息學報,第二十六卷,第一期,頁 73-83。
- Amrani, Y.A., Lazaar, M. and Kadiri, K.E.E. (2018), 'Random forest and support vector machine based hybrid approach to sentiment analysis', *Procedia Computer Science*, Vol. 127, pp. 511-520.
- Borg, I. and Groenen, P. (1997), *Modern Multidimensional Scaling*, Springer, New York.
- Cao, Q., Duan, W. and Gan, Q. (2011), 'Exploring determinants of voting for the

- “helpfulness” of online user reviews: A text mining approach’, *Decision Support Systems*, Vol. 50, pp. 511-521.
- Carmeli, N., Wang, X., Suhara, Y., Angelidis, S., Li, Y., Li, J. and Tan, W.C. (2020), ‘EXPLAINIT: Explainable review summarization with opinion causality graphs’, arXiv:2006.00119v1 [cs.CL].
- Chen, Y.L., Chang, C.L. and Yeh, C.S. (2017), ‘Emotion classification of YouTube videos’, *Decision Support Systems*, Vol. 101, pp. 40-50.
- Cox T.F. and Cox M.A.A. (2000), *Multidimensional Scaling*, Chapman & Hall, London.
- Day, M.Y. and Teng, H.C. (2017). ‘A study of deep learning to sentiment analysis on word of mouth of smart bracelet’, *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Sydney, Australia, July 31-August 3, pp. 763-770.
- Gambhir, M. and Gupta, V. (2017), ‘Recent automatic text summarization techniques: A survey’, *Artificial Intelligence Review*, Vol. 47, pp. 1-66.
- Gupta, V. and Lehal, G.S. (2010), ‘A survey of text summarization extractive techniques’, *Journal of Emerging Technologies in Web Intelligence*, Vol. 2, No.3, pp. 258-268.
- Hu, C.C. and Li, H.L. (2017), ‘Developing navigation graphs for TED talk’, *Computers in Human Behavior*, Vol. 66, pp. 26-41.
- Hu, Y.H., Chen, Y.L. and Chou, H.L. (2017), ‘Opinion mining from online hotel reviews - A text summarization approach’, *Information Processing and Management*, Vol. 53, pp. 436-449.
- Jaccard, P. (1901), ‘Étude comparative de la distribution florale dans une portion des alpes et des jura’, *Bulletin de la Société Vaudoise des Sciences Naturelles*, Vol. 37, pp.547-579.
- Kamal, A. and Islamia, J.M. (2015), ‘Review mining for feature based opinion summarization and visualization’, *International Journal of Computer Applications*, Vol. 119, No. 17, pp. 6-13.
- Liu, Y., Safavi, T., Dighe, A. and Koutra, D. (2018), ‘Graph summarization methods and applications: A survey’, *ACM Computing Surveys*, Vol. 51, No. 3, Article 62.
- Lloret, E., Boldrini, E., Vodolazova, T., Martinez-Barco, P., Munoz, R. and Palomar, M. (2015), ‘A novel concept-level approach for ultra-concise opinion summarization’, *Expert Systems with Applications*, Vol. 42, No. 20, pp. 7148-7156.
- Ma, L.C. (2012), ‘Screening alternatives graphically by an extended case-based distance approach’, *OMEGA*, Vol.40, pp. 96-103.
- Ma, L.C. and Huang, Z.X. (2020), ‘A graphical approach to analyze online hotel reviews

- based on text mining', *International Journal of Innovation in Management*, Vol. 8, No. 1, pp. 31-36.
- Moussa, M.E., Mohamed, E.H. and Haggag, M.H. (2018), 'A survey on opinion summarization techniques for social media', *Future Computing and Informatics Journal*, Vol. 3, pp. 82-109.
- Nallapati, R., Zhou, B., Santos, C., Gulcehre, C. and Xiang, B. (2016), 'Abstractive text summarization using sequence-to-sequence RNNs and beyond', *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning. Association for Computational Linguistics*, pp. 280-290.
- Pecar, S. (2018), 'Towards opinion summarization of customer reviews', *Proceedings of ACL 2018, Student Research Workshop*, pp. 1-8.
- Peng, B. and Lee, L. (2008), 'Opinion mining and sentiment analysis', *Foundations and Trends in Information Retrieval*, Vol. 2, pp. 1-135.
- Rinaldi, A.M. (2019), 'Web summarization and browsing through semantic tag clouds', *International Journal of Intelligent Information Technologies*, Vol. 15, No. 3, pp. 1-23.
- Rinaldi, A.M. and Russo, C. (2021), 'Using a multimedia semantic graph for web document visualization and summarization', *Multimedia Tools and Applications*, Vol. 80, pp. 3885-3925.
- Rush, A.M., Chopra, S. and Weston, J. (2015), 'A neural attention model for abstractive sentence summarization', *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 379-389.
- Yong, W., Xueqiang, L., Lianchun, J. and Shibin, X. (2014), 'Sentiment classification for Chinese microblogging based on polarity lexicons', *Computer Applications and Software*, Vol. 31, No. 1, pp. 35-37.
- Zhang, S., Wei, Z., Wang, Y. and Liao, T. (2018), 'Sentiment analysis of Chinese microblog text based on extended sentiment dictionary', *Future Generation Computer Systems*, Vol. 81, pp. 395-403.
- Zuo, C. (2018), 'Defense of computer network viruses based on data mining technology', *International Journal of Network Security*, Vol. 20, No. 4, pp. 805-810.
- Zhuang, L., Jing, F. and Zhu, X. Y. (2006), 'Movie review mining and summarization,' *In Proceedings of the 15<sup>th</sup> ACM International Conference on Information and Knowledge Management*, pp. 43-50.

