

# 以數位口碑為基礎之流行性商品銷售預測模式

范有寧\*

國立臺灣大學資訊管理學系

黃心惟

國立臺灣大學資訊管理學系

陳靜枝

國立臺灣大學資訊管理學系

## 摘要

市場瞬息萬變，企業必需不斷調整腳步，才有機會取得先機，而企業的採購與生產計劃源頭即是銷售預測，足見銷售預測的重要性。即使如此，企業在銷售起伏較大的流行性商品上的銷售預測仍不是非常準確。從文獻探討中可知，口碑是影響消費者購買的動機之一，消費者可由身邊的親朋好友口中獲得口碑，亦可利用網際網路取得數位口碑。但過去只對口碑形成的因素與影響層面做研究，而未實際量化口碑對銷售數字的影響，故本研究欲將數位口碑實際應用在銷售預測上。

本研究提出一將數位口碑運用於銷售預測之演算法，分成兩個階段進行流行性商品之銷售預測，首先自動擷取網路中的商品相關討論文章，利用簡單貝氏分類器判斷文章的評價，並將其轉換成量化評價以用作預測。除了建立多個目標函式以釐清口碑與銷售量的關係外，同時亦嘗試各種時期的口碑作為自變數，以決定口碑與銷售量的期間差距為何。最後找出所有組合中預測誤差最小，且具效度的模型用作預測，以提升流行性商品銷售預測的準確度。本研究以台灣知名連鎖藥妝店的十大熱銷商品為樣本，發現本模型適用具話題性、能在網路上引起足夠的討論量商品。且其預測誤差皆低於常見的移動平均法、指數平滑法、趨勢指數平滑法等。可知將本模型應用於流行性商品銷售預測時，可有效地增進預測準確度。

**關鍵詞：**流行性商品、數位口碑、銷售預測、文本挖掘、需求管理

---

\* 本文通訊作者。電子郵件信箱：d96725003@ntu.edu.tw  
2011/03/15 投稿；2011/05/16 修訂；2011/05/18 接受

# A Solution for Sales Forecasts of Fashion Products Based on Electronic Word-of-Mouth

Yu-Neng Fan\*

Department of Information Management, National Taiwan University

Hsin-Wei Huang

Department of Information Management, National Taiwan University

Ching-Chin Chern

Department of Information Management, National Taiwan University

## Abstract

Forecasts are essential to the business's decision making and planning processes. Better forecasting can contribute to better price structuring and better inventory management. However, it is a challenging problem owing to the volatility of demand which depends on many factors. The situation is especially prominent in fashion product due to its sales versatility. Past research shows that disseminating information through word-of-mouth communication is one of the most effective mediums for relaying important product and company information. It not only plays an important role in the evaluation of products but also plays an important role in society as well.

In this study, an automatic mining approach is proposed to resolve the aforementioned issues. A text mining technique and Naive Bayes classifier will be used to determine the rating of each product-related article extracted from the Internet. Based on regression model, some target functions have been designed to clarify the relationship between the rating of word-of-mouth and the sales. Performances of our model are evaluated by the real sales data of a large cosmetic chain store in Taiwan. The proposed method is proved to outperform several traditional sales forecasting methods. Therefore, we believe that this model can effectively enhance the prediction accuracy when applied to fashion products.

**Keywords:** Fashion Product, Electronic Word-of-Mouth, Sales Forecasts, Text-Mining, Demand Management.

---

\* Corresponding author. Email: d96725003@ntu.edu.tw  
2011/03/15 received; 2011/05/16 revised; 2011/05/18 accepted

## 壹、緒論

為了在不斷變化的環境裡提升企業的競爭優勢，組織的管理階層必須依現有的資訊即時做出正確的決策以因應瞬息萬變的市場趨勢。然而企業一切的採購、生產計劃的源頭即為銷售預測 (Kuo & Xue 1999)，唯有依據正確的銷售預測才能定出恰當的採購數量以取得足夠的生產原料進行生產，而後生產時所需人力物力資源亦需事先規劃以避免額外支出 (Sheikh 2003)。一旦銷售預測不準確，那麼所影響的並非只是終端零售業者，而是包括了從製造商、批發商等整條供應鏈上下游的各個生產或銷售計劃，牽一髮而動全身 (Chopra & Meindl 2007)。

銷售預測對企業利潤的影響是如此重大，因此許多企業都定期執行銷售預測，但預測出的結果卻往往不盡理想 (Dalrymple 1987)。檢視其原因，發現銷售預測時考量的幾個重點通常為過去的銷售數據、商品本身特性、促銷事件、節慶、季節等企業內部已知因素。然而許多時候在加入此類屬性後做出的預測仍有一段落差，出現許多不規則之處，顯示出影響商品銷售成績的變因不只上述已知資訊，可能還包含了外部因素。由於影響銷售數字的最直接因素為消費者行為，因此在外部因素方面可由消費者行為著手。消費者在制定決策前會大量搜尋口碑 (Word of Mouth) 資訊 (Olshavsky & Granbois 1979)，口碑為非正式且人與人之間基於非商業的理由，藉由語言來傳遞包括品牌、商品、服務等的資訊 (Arndt 1967)。口碑多寡與好壞皆對消費者決策具有重大影響 (Gremier et al. 2001)。但口碑不一定需由身邊的親朋好友口中獲得，利用網際網路，消費者亦可從相關商品使用者所發表的文章來取得數位口碑 (Electronic Word of Mouth) (Ratchford et al. 2001)。

近年來網際網路使用者自行發表的文章越來越多，包括部落格 (blog) 與討論區的文章如雨後春筍般冒出。隨著部落格作者知名度、讀者數量、討論區參與者的增加，此類文章的影響力也隨之擴大。舉例而言，在購買美容保養商品時，網友口碑甚至比專家名人推薦更有影響力。而欲了解某商品的口碑如何，可請消費者表示相關意見，亦可主動收集已存在的意見。許多廠商會以問卷模式進行，請消費者表示使用特定商品或服務的相關感受，但問卷填答的成本高昂，且樣本數有限，無法輕易的擴大規模至更多樣商品，故本研究不採問卷填答方式收集資料，而意欲尋求系統化、自動化口碑收集方式。口碑的傳播不限於口耳相傳，亦可藉文字呈現 (Hennig-Thurau et al. 2004)。考慮到近年網路社群蓬勃發展，網路上各式使用者自行發表的意見更不斷增加，其中有關消費性商品的使用經驗文章亦不在少數，因此可視為一項重要的口碑來源 (Ratchford et al. 2001)。

然而該類文章數量過多，若人工閱讀分析必然十分耗時。問題在於，網路輿論對消費者行為的影響究竟有多大？而網路文章中隱涵的資訊應如何被有系統的

抽取而出？且挖掘出的資訊又該如何應用並產生價值？故本研究欲發展出一套自動探勘模型，系統化的分析處理相關資訊，套用至銷售預測模型後與真實銷售數字做驗證，嘗試更精準的預測銷售成績。故本研究將針對預測之商品，自動化的搜集網路相關討論文章，使用文本挖掘（Text Mining）技術將文章分依語意分類成正、負與中立評價等三類以成為量化指標，再將該量化指標與銷售數據一起放入迴歸模型中，並嘗試多個目標函式以釐清口碑與銷售量的關係。最後再以平均絕對百分比差（MAPE）為標準，找出預測誤差最小，且符合模型假設的迴歸曲線作為預測模型，以發展出更精準的預測模型。本研究針對供應鏈末端的零售商，一家台灣地區知名連鎖藥妝店的銷售資料為實際案例，進行研究方法之驗證，並假設目標商品均具有充分的歷史銷售資料。

## 貳、文獻探討

本研究為流行性商品之銷售預測，因此文獻探討將由銷售預測著手，發覺前人之銷售預測中有不能解釋的異常銷售量後，再探討消費者行為與口碑，欲從消費者的角度發覺銷量異常的原因，最後認定為口碑（Word of Mouth）因素。口碑的採集若以問卷型式進行因成本高昂而無法輕易擴大規模，有鑑於網際網路上的討論區與部落格等消費者自行發表的商品意見增加，因此本研究欲從討論文章中找出量化的數位口碑，採用文本挖掘（Text Mining）的方式，套用於銷售預測模型。因此文獻探討，將以：銷售預測，消費者行為與口碑及文本挖掘三部分，切入討論。

### 一、銷售預測

#### （一）預測方法

流行性商品與一般長銷型商品不同（Fisher & Rajaram 2000; Golder & Tellis 1997），其銷售預測的模式也會大不相同，因此無法使用與一般長銷型商品相同的銷售預測方法進行分析。銷售量是一種時間序列（time-series）的資料。統計學家將時間序列分成四個部分：(1)長期趨勢（trend）、(2)周期性（cyclical effect）、(3)季節性（seasonality）、(4)其它無法預測的變動（irregularity）（Black 2004; Taylor 2004）。針對一項產品，想要了解銷售趨勢時，會先將季節性、促銷等活動所造成的效果移除，再引用適合的統計模型來了解趨勢的走向，最後再把這樣外在的因素加入，得到較為合理的預測值。

然而流行性商品的干擾因素又較一般長銷型商品更為複雜，其中的干擾因素如促銷與特殊因素等，又來得更多（Geurts & Whitlark 2000; Kahn 1998; Schrieber 2005），為使未來的銷售預測更正確，可改進輸入資料、降低時間序列資料的差異

與供應鏈中資訊共享等 (Geurts & Whitlark 2000; Schrieber 2005)，或如流行性商品的銷售資料時間序列分析中，應先以產品生命週期的概念 (Klepper 1996; Polli & Cook 1969; Ting 2007)，分析流行性商品的流行週期，以便釐清流行性商品的干擾因素，再分為流行週期與非流行週期的兩大部份銷售資料時間序列再依不同方法進行銷售預測。但過去有關流行性商品的銷售預測演算法中，都沒有加入口碑對銷售數字的影響，甚至直接將流行期間去除後再做預測 (Ting 2007)，因此本研究將針對此處做改善。

## (二) 預測準確性

一般常見的預測準確度評估方式有平均方差 (Mean squared error, MSE)、平均絕對差 (Median absolute deviation, MAD) 與平均絕對百分比差 (Mean absolute percentage error, MAPE) 等三種比較標準 (Black 2004; Sheikh 2003; Taylor 2004)。

其中只有 MAPE 具備去除單位與絕對數值之差異的能力，且最常被一般企業所使用 (Kahn 1998)，其中有半數企業使用預測值做為 MAPE 的分母，而另外半數使用歷史實際值做為分母。由於較大的分母會使得誤差比率減少，故檢視這兩種方式，可發現前者較易發生高估銷售值的問題；而後者則較易出現低估銷售值的情形，另外，學者 Jain (1998) 在其研究提出使用累積銷售比率進行銷售預測的方法，相對於預測準確度的計算，也應考慮採用累積銷售值的 MAPE 計算方式，即除了評估單期的 MAPE，亦可評估多期累積銷售的 MAPE。總結以上，本研究將採用 MAPE 來驗證本模型的誤差值，其目的在去除單位與絕對數值之差異，並採用銷售預測值為分母的公式，以減少低估流行性商品銷售的情形，且考慮使用多期累積銷售的 MAPE。

本研究將採用多期累積銷售的 MAPE 進行績效評估，其中使用銷售預測值為分母，公式如下：

$$MAPE_i = \frac{1}{SN_i} \sum_{t=0}^{SN_i-1} \left| \frac{y_{it} - \hat{y}_{it}}{y_{it}} \right|$$

$y_{it}$  為第  $i$  個商品的第  $t$  期歷史銷售量， $\hat{y}_{it}$  為第  $i$  個商品的第  $t$  期預測銷售量。本研究首先提出了幾種可能的線性與非線性模型，並使用最小方差來求出該模型中的最佳解，接著使用 MAPE 來評估使用該模型預測時的準確度。若該模型 MAPE 與一般常用預估方法 MAPE 相比較高 (不準確)，則重新提出別種模型運算，直至找出良好適用的模型為止。

## 二、消費者行為與口碑

影響消費者行為的因素有許多，其中口碑多寡與好壞皆對消費者決策具有重

大影響 (Bansal & Voyer 2000; Buttle 1998; Gelb & Johnson 1995; Gremler et al. 2001)。口碑為非正式且人與人之間基於非商業的理由，藉由語言來傳遞包括品牌、商品、服務等的資訊 (Arndt 1967)。但口碑不一定需由身邊的親朋好友口中獲得，利用網際網路，消費者亦可從有相關商品使用者所發表的文章來取得數位口碑 (Ratchford et al. 2001)。數位口碑為消費者藉由提供消費者發表意見的網站平台，對廣大的其他消費者所分享他們對商品與服務的意見或經驗 (Hennig-Thurau et al. 2004)。

雖然已知此事實，但在收集口碑時，廠商們通常是以某些商品或服務為目標進行問卷調查，此為被動地請消費者提供意見之調查，不僅費時費力，且只能針對單一商品做研究，欠缺可擴充性。而本研究雖同為探討口碑與消費者行為的關係，但採用的方法為主動至網路上搜尋文章，電腦化作業後不僅縮短時效，同時更可輕易地將研究目標擴大至許多商品。

### 三、文本挖掘 (Text Mining)

欲量化文本中隱含的口碑資訊，需使用文本挖掘方法。而文件資料通常為非結構性 (Unstructured) 或半結構性 (Semi-structured) 的資料，因而難以直接抽取並處理，然而隨著資訊爆炸的時代來臨，文本挖掘的技術日新月異，以下就各種目的分別討論文本挖掘之相關研究。

#### (一) 資訊抽取 (Information Extraction)

資訊抽取為將片斷的訊息從文件中抽取出以滿足使用者的資訊需求 (Yang & Lee 2008)。文本挖掘中的資訊抽取通常需先取出關鍵詞 (Keyword)。關鍵詞的選擇一般而言分為三種 (Lo 2008)：字典方法 (Dictionary Approach)、語言學方法 (Linguistic Approach) 與統計方法 (Statistical Approach)。字典方法利用事先定義的語彙字典來取出關鍵字 (Sakurai & Ueno 2004)。語言學方法為利用文法分析，找出文章中的所有名詞、動詞、代名詞後，再將無意義的詞彙剔除 (Kroha et al. 2006)。統計方法為計算單詞在該篇文章與所有文章中出現的頻率後，將較重要的詞彙挑出。其中較為廣泛使用的統計方法為詞頻與逆向文章頻率：tf-idf (term frequency-inverse document frequency) (Salton & McGill 1983)。

#### (二) 分類 (Classification)

若已知既定之標籤分類，則將文本分至各標籤的行為稱為文本分類 (Sebastiani 2002)。Kroha、Baeza-Yates 與 Krellner (2006) 先行計算正反關鍵詞數量，再利用簡單貝氏分類器 (Naive Bayes Classifier) 將財經新聞分為上昇與下降兩種市場趨勢，最後利用分類的結果用來預測股市走向。

本研究為量化商品評價，即為將文章分為好、壞與中立評價，關鍵詞中以形容詞占多數，故與 Izumi、Matsui 與 Matsuo (2007) 與 Kroha 等 (2006) 較相近。其中 Izumi 等 (2007) 使用決策樹將市場好壞評價分為 7 種程度，Kroha 等 (2006) 則使用簡單貝氏分類將新聞分為市場上升與下降兩種趨勢，由於決策樹與簡單貝氏的預測準確度相近 (Huang et al. 2003)，故本研究決定採用較簡單 Kroha 等 (2006) 的簡單貝氏分類將文章分為好、壞與中立三種評價。

總結本章，從銷售預測、消費者行為與口碑、文本挖掘等三節的文獻探討中可知，前人的銷售預測方法常不能準確的預測流行性商品的銷售量，因為流行性商品皆有持續一段時間的異常銷售量出現。而消費者行為相關的文獻提到口碑的質量對消費者決策具有重大影響，故銷售量的流行可能起因於口碑的流行。經過文本挖掘文獻探討後，本研究決定採用簡單貝氏分類將討論區與部落格的文章分為好、壞與中立三種評價，以作為數位口碑的轉換依據。

## 參、問題描述

本研究欲利用數位口碑之探查，來增進流行性商品在流行期間之銷售預測的準確度。首先欲找出特定商品的評價，但商品評價隱含在各討論區或部落格之中，無法直接取得量化資訊，故需使用文本挖掘 (Text Mining) 技術來抽取出特定商品的量化評價，最後再找出使用這些量化評價經變數轉換而得數位口碑來預估銷售流行期間與期間銷售量的方式。本章中將介紹本研究如何量化商品評價、銷售量流行與口碑流行之呼應及本研究之假設條件。

### 一、量化商品評價

本研究之目的雖為利用數位口碑來改善銷售預測模型，但數位口碑的取得實為一文本挖掘之結果，故本研究的商品需在網路上存在數位口碑，如討論區或部落格等。首先針對不同商品，將在該商品的重要討論區中找尋相關文章，如睫毛膏的重要討論區之一為 Fashion Guide。文本挖掘方式分類商品。由於目前的文本挖掘技術常常著重在自動化地找出名詞性關鍵字以辨別文章的主題，但本研究卻是評斷文章為好評價或壞評價等此類語意性質，其中關鍵字以形容詞居多。且目前的中文文本挖掘研究較難正確地切割形容詞，或自動找出影響文章語意的關鍵形容詞，所以此處需請對相關商品有深入了解的專家定義該商品的重要關鍵字字典，其中包含與商品相關的名詞與形容詞，以做為分類之用。如有關睫毛膏的正面評價可能為「好卸」、「好用」、「濃」、「持久」，負面評價可能為「暈」、「掉屑」、「不推」。

討論區文章中的一部分作為簡單貝氏分類器 (Naïve Bayes Classifier) 的學習

樣本，其餘作為分類對象，使用經專家預先定義的關鍵詞作為分類依據，將文章分為三種，正評價、負評價、中立評價。最後計算同一商品在某段期間內每個時距  $t$  的正評價數、負評價數、中立評價數，分別定義為  $P_t$ 、 $N_t$ 、 $M_t$ 。其中同個商品的  $P_t$ 、 $N_t$ 、 $M_t$  可能呈負相關（正評價越多，負評價越少），或其他各種可能的關係。而由於不確定這三種評價可能各自對未來的銷售量產生何種影響，故此處需將  $P_t$ 、 $N_t$ 、 $M_t$  三者經變數轉換為綜合的口碑  $W_t$ ，以與未來的銷售量做呼應。而如何找出最恰當的變數轉換方法與係數也是本模型欲求的重點。如圖 1 為其中一種可能的變數轉換方式： $W_t = P_t - N_t$ 。

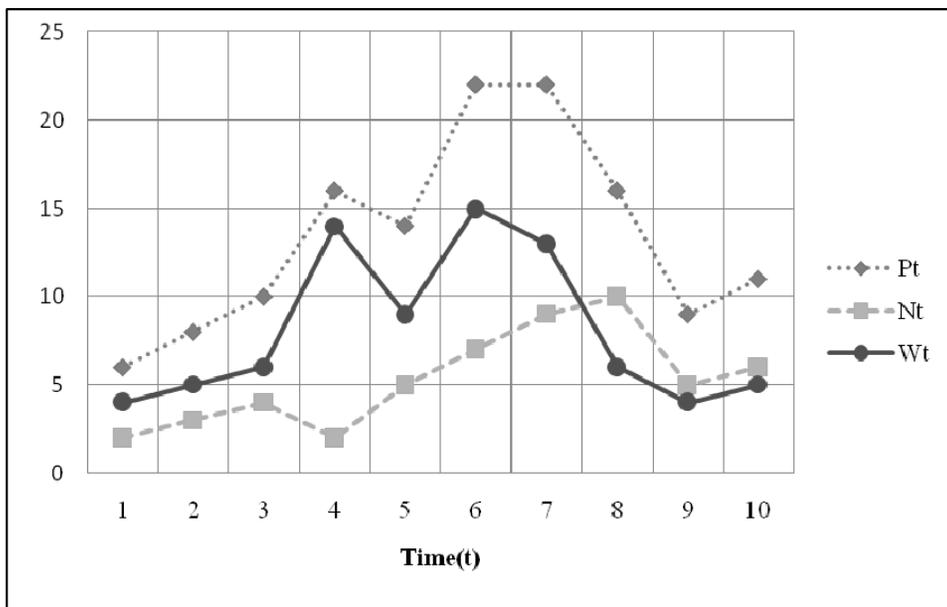


圖 1：口碑變數轉換範例  $W_t = P_t - N_t$

## 二、銷售量流行與口碑流行之呼應

根據消費者心理學可知，消費者在購買商品時，受口碑影響效果很大。以因果關係來看，若口碑為因，則銷售量為果，故會出現口碑流行在前，而銷售量流行在後，且兩者流行趨勢一致的呼應現象。但本研究的重點在於，口碑流行開始與銷售流行開始的期數差  $L$  為多少，即口碑比銷售早幾期。由於不同性質的評價對不同時期的銷售可能產生不同影響，所以本研究將分別嘗試將各種時期的  $P_{(t-L_2)}$ 、 $N_{(t-L_3)}$ 、 $M_{(t-L_4)}$  轉換為第  $t$  期的口碑  $W_t$ ，用作預測第  $t$  期銷售量  $S_t$  的參考。其中  $L_2$ 、 $L_3$ 、 $L_4$  不一定相等，舉例來說，有可能三期前的正評價、一期前的負評價與二期前的中立評價，共同影響了當期的銷售量，則此時  $L_2 = 3$ 、 $L_3 = 1$ 、 $L_4 = 2$ 。

### 三、假設條件

#### (一) 商品具歷史銷售數據

本研究為將過去的銷售數據與數位口碑之數據做適當之連結，以發展出整合模型做為未來銷售預測之用。故研究之商品必需有一段夠長的歷史銷售數據，包括廠商提供直接的消費點的銷售資料，或是間接的出貨資料，以做為模型之輸入資料。

#### (二) 商品為重要商品

商品必需是銷售量大或價值高，且為廠商重要商品。因為此類商品通常有具代表性的銷售量：量大且持續一段時間。如彩妝商品中的睫毛膏即是此類商品。

#### (三) 商品具流行性與話題性

本研究為利用數位口碑以提高流行性商品之銷售預測準確度，故研究之商品需為具流行期間之流行性商品、且具備夠突出的產品特色，如數位相機或筆記型電腦，才能在網路上引起足夠的討論量，即足夠的數位口碑，以利本研究從中挖掘出商品口碑評價並用於模型中。

#### (四) 商品具數位口碑

本研究是以數位口碑為基礎的商品預測，故需有數位口碑的存在以套用至本研究的模型。而數位口碑為使用者在網路上發表的意見，故網路中需有評論該商品的討論區或部落格等，如 3C 商品討論區，以供本研究採集數位口碑。

#### (五) 商品具地區性

即使在現今全球化的網際網路環境中，由於受限於語言、商品特性、使用者習慣等緣故，消費性商品的討論區仍是具地區性的。畢竟討論區是由一群使用者共同貢獻自己的意見，而大多數的消費者通常較習慣閱讀母語或使用母語發言；且即使是全球化的商品，仍會有地區性的發售時間或商品型號與名稱，故商品討論區通常侷限於特定地區的商品，如台灣討論區的商品仍以有在台灣販售的商品居多。而本研究的商品亦具地區性，並有相對應的地區性討論區。

#### (六) 口碑流行與銷售流行時間間隔有限

商品口碑的流行將引起銷售量之流行，即使銷售量無法立即反應出口碑流行的程度，但口碑與銷售之間隔時間長度仍是有限的。其中因商品特性的不同，口碑流行與銷售流行之間隔時間也有所不同，意即間隔長度將依商品特性的不同而有不同的限制，在此定義口碑最早會早於銷售  $LL$  期。

## 肆、以口碑為基礎的銷售預測演算法

### 一、演算法概述

本研究欲利用數位口碑的收集與分類，來增進流行性商品銷售預測的準確度。本研究將自動擷取網站上的討論文章，將其轉換成量化評價以用作預測。除了建立多個目標函式以釐清口碑與銷售量的關係外，同時亦嘗試各種時期的口碑作為自變數，以決定口碑與銷售量的期間差距為何。最後找出所有組合中預測誤差最小，且具效度的模型用作預測，以提升流行性商品銷售預測的準確度。本演算法的主要流程說明如下：

#### (一) 量化商品評價

針對商品特性，於重要網站中搜集相關討論文章後，使用文本挖掘技術將討論文章分類至正、負與中立評論文章。統計各時距內的正、負與中立評論文章數，作各時期的商品評價指標。

#### (二) 以最小方差法求得最適迴歸模型

設計多種目標函式，並嘗試過去各種時期的評價與銷售量作為最小方差法的參數，以找出預測誤差最低的曲線用來預測未來的銷售量  $S$ 。

#### (三) 迴歸曲線檢定

檢查迴歸模型求出的最佳曲線是否符合模型的假設與標準。若不符合，則應從迴歸曲線中選擇次好且符合假設與標準的曲線用作預測；若符合則輸出結果。

#### (四) 輸出結果

將符合模型假設與標準的最佳迴歸曲線，相關之變數與係數輸出，作為未來預測的公式。

### 二、量化商品評價

本研究欲利用商品的口碑來增進銷售預測準確度。所以首先應收集商品評價並量化評價，流程包含，先以網路為基礎的評論收集，針對商品特性，於重要網站中搜集相關討論與評論文章；接著，關鍵字計算，請專家觀察收集到的文章，建立關鍵字字典，並計算文章與關鍵字字典相符合的關鍵字數量；並建立特徵向量，分析關鍵字頻率後可得文章的  $tf-idf$  向量，再加上相對應的產品名稱，即為文章的特徵向量；最後進行簡單貝氏分類，使用一部分文章的特徵向量以用於訓練簡單貝氏分類器，訓練的結果可用於分類其餘文章。以下就針對每一步驟詳細

敘述其流程。

### (一) 以網路為基礎的評論收集

首先針對商品特性，找到該商品的相關討論區或部落格，接著設定商品名稱的關鍵字，如品牌、外形、暱稱等，剖析該網站架構，針對商品的關鍵字進行自動搜尋，擷取討論串標題或內文中包含該關鍵字討論串。同時亦針對各個商品剔除最後一筆銷售資料之後的討論。本研究的探討對象為台灣某彩妝通路的十大暢銷商品，皆為美容保養相關產品。檢視 2008 年台灣百大網站排名 (Bnext 2008)，可發現 Fashion Guide 為唯一入選的美容網站專門討論區，故可選擇此網站討論區作為商品相關討論文章收集對象。

### (二) 關鍵字計算

討論文章收集後，交由專家觀察文章，並建立所有商品的關鍵字字典，同義詞視為同個關鍵字。接著依同義詞長度排序，從長度最長的同義詞開始，搜尋文章中符合同義詞的部分，找到後清除符合的部分，重複至最短的同義詞為止，如此可保證關鍵字數量計算的正確性。

### (三) 建立特徵向量 (Feature Vector)

本研究使用產品名稱與  $tf-idf$  建立每篇文章的特徵向量以利後續分類。其中  $tf-idf_{ij}$  衡量在一群文章中，某關鍵字  $i$  對該某文章  $j$  的重要性。公式如： $tf-idf_{ij} = tf_{ij} * \ln \frac{N}{N_i}$ 。其中

- $tf_{ij} = \frac{\text{第 } j \text{ 篇文章中，第 } i \text{ 個關鍵字出現的次數 } N}{\text{第 } j \text{ 篇文章中，所有關鍵字出現的次數 } N_i + 0.01}$
- $idf_i = \ln \frac{\text{所有文章數 } N}{\text{包含第 } i \text{ 個關鍵字的所有文章數 } N_i}$
- $N = \text{所有文章數}$
- $N_i = \text{包含第 } i \text{ 個關鍵字的所有文章數}$

由於各個關鍵字可能對不同商品造成不同影響，因此此處的特徵向量需加上一名目性 (Nominal) 的屬性：產品名稱。令全部的關鍵字數量為  $k$ ，那麼對於第  $j$  篇文章，有一個向量  $V_j = (\text{Product}, tf-idf_{1j}, tf-idf_{2j}, \dots, tf-idf_{kj})$ 。其中若  $N_i$  為 0，則代表關鍵字  $i$  不曾任一篇文章中出現，不具鑑別力，故不列入考慮，且計算向量  $V_j$  時直接省略該維度。

本研究有關口碑的原始資料為一般使用者所撰寫的文章，為非結構性資料，故需使用資訊抽取方式先行處理文章。由於商品評價的關鍵字通常為形容詞而非專有名詞，且針對商品的不同，相對應之正負形容詞又不相同，且可能同義多詞。故本研究採用字典方法來統計關鍵字，並計算文章的特徵向量以利接下來的分

析。由於關鍵字可能對不同商品造成不同影響，因此此處的特徵向量需加上一名目性 (Nominal) 的屬性：產品名稱。令全部的關鍵字數量為  $k$ ，那麼對於第  $j$  篇文章，有一個向量  $V_j = (\text{Product}, tf-idf_{1j}, tf-idf_{2j}, \dots, tf-idf_{kj})$ 。其中若  $N_i$  為 0，則代表關鍵字  $i$  不曾在任一篇文章中出現，不具鑑別力，故不列入考慮，且計算向量  $V_j$  時直接省略該維度。而討論區中每篇評論的字數通常很少，而可能出現第  $j$  篇文章不包含任何關鍵字的狀況，故在  $tf_{ij}$  的分母加上少量的 0.01 以避免除以零的情況。

#### (四) 簡單貝氏分類

在上一點建立特徵向量後，將所有文章的特徵向量放入資料探勘軟體 Weka (Hall et al. 2009) 中的簡單貝氏分類，將文章分為正評價、負評價或中立評價 3 種分類。其中一部分文章用於訓練簡單貝氏分類器，其餘文章則用於測試分類器的分類效果。

由於簡單貝氏分類在遇到數值性 (Numeric) 資料時，會假設資料為常態分配，並依常態分配計算機率。但本研究中產生特徵向量的文章為討論區中使用者的意見，通常非常簡短，一篇文章中包含的關鍵字可能極少，而使得特徵向量中包含大量的 0，並使得各個屬性的分配不為常態分配。

因此在進行簡單貝氏分類前，必需先將所有的  $tf-idf$  屬性離散化，此處使用常見的離散化方法中較佳的 MDL (Minimum Description Length) 法則來離散化  $tf-idf$  屬性 (Witten & Frank 2005)。接著用貝氏定理 (Bayes Theorem) 計算某筆資料  $V = (v_1, v_2, \dots, v_n)$  在每個分類  $C_i$  的事後機率 (Posteriori Probability)，並將該筆資料  $P(C_i|V)$  機率最大的分類  $C_i$ 。最後總合用來訓練的文章與分類後的文章，統計各時距內三種分類 (正評價、負評價與中立評價) 的文章數量，並計算出所有商品、所有時距內的正評價數  $P_{it}$ 、負評價數  $N_{it}$ 、中立評價數  $M_{it}$ ，作為迴歸模型的參數。

### 三、以最小方差法求得最適迴歸模型

由前一節得到量化商品評價後，統計各時距內的各種評價文章數量，並作為上一章所述迴歸模型的參數。針對第  $i$  個商品，依據該商品的正評價數  $P_{it}$ 、負評價數  $N_{it}$ 、中立評價數  $M_{it}$  的兩兩相關係數  $r$ ，選擇迴歸模型 Case 1、Case 2、... 或 Case 5 操作。令選擇的多元迴歸模型包含常數有  $BN_i$  個係數，則有  $BN_i - 1$  個自變數，依據  $BN_i$  的不同，可將所有的目標函式分類如下：

If  $BN_i = 6$

Case 1, 有 5 個自變數： $S_{it}$ 、 $P_{it}$ 、 $N_{it}$ 、 $M_{it}$ 、 $t$

If  $BN_i=5$

Case 2, 有 4 個自變數:  $S_{it}$ 、 $P_{it}$ 、 $M_{it}$ 、 $t$

Case 3, 有 4 個自變數:  $S_{it}$ 、 $P_{it}$ 、 $N_{it}$ 、 $t$

Case 4, 有 4 個自變數:  $S_{it}$ 、 $N_{it}$ 、 $M_{it}$ 、 $t$

If  $BN_i=4$

Case 5, 有 3 個自變數:  $S_{it}$ 、 $P_{it}$ 、 $t$

所有  $BN_i-1$  個自變數中, 除了  $t$  之外, 均需嘗試  $k=1\sim 3$ 、 $L_{i1}=1\sim LL_i$ 、 $L_{i2}=1\sim LL_i$ 、 $L_{i(BN_i-2)}=1\sim LL_i$  的迴歸函式。以虛擬程式碼 (pseudo code) 表示如下, 其中針對每個迴歸函式均需做步驟 1~5。當所有的迴歸函式的所有步驟都執行完後, 則找出其中具最小 MAPE 的迴歸曲線。

For  $k$  from 1 to 3

For  $L_{i1}$  from 1 to  $LL_i$

For  $L_{i2}$  from 1 to  $LL_i$

⋮

For  $L_{i(BN_i-2)}$  from 1 to  $LL_i$

1. Use the least squares method to find the best curve

2. For each node of the curve

If  $|\text{StandardDeviation}| > 2$  then

Remove the node

3. If amount\_of\_nodes  $< (BN_i-1) * 6$  then

Give up the curve

4. Use the least squares method to find the best curve

5. Compute the MAPE of the curve

Find the curve with the smallest MAPE of all the curves

### (一) 單次最小方差法實作

跳脫出前述設定, 專注於每個組合都要做一次最小方差法。若多元迴歸模型中共有  $n$  筆資料與  $k$  種自變數, 其中  $y_i$  為第  $i$  筆資料中欲預測的應變數, 而  $x_{ij}$  為第  $i$  筆資料中, 第  $j$  個自變數。詳列如下:

- $(y_1, x_{11}, x_{12}, \dots, x_{1k})$
- $(y_2, x_{21}, x_{22}, \dots, x_{2k})$
- $(\dots\dots\dots)$
- $(y_n, x_{n1}, x_{n2}, \dots, x_{nk})$

若多元迴歸模型為  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$ , 則可用  $\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k$  來估計  $y$  值。即

- $\hat{y}_1 = b_0 + b_1x_{11} + b_2x_{12} + \dots + b_{1k}x_{1k}$
- $\hat{y}_2 = b_0 + b_1x_{21} + b_2x_{22} + \dots + b_{2k}x_{2k}$
- .....
- $\hat{y}_n = b_0 + b_1x_{n1} + b_2x_{n2} + \dots + b_{nk}x_{nk}$

上述可用矩陣運算來表示：

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \hat{Y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}, B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}, \hat{Y} = XB$$

由於  $Y=XB$  通常無解，意即  $Y$  通常不在  $X$  的 column space 上，因此利用最小方差法找出  $\hat{Y}=XB$ ，使  $Y$  與  $\hat{Y}$  的距離最小，即最小化  $\|Y - \hat{Y}\| = SSE = Error\ Sum\ of\ Square = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ 。方法為令  $X^T(XB - Y) = 0$ ，即  $X^T X B = X^T Y$ ，可得  $B = (X^T X)^{-1} X^T Y$  為迴歸之係數矩陣。

#### 四、迴歸曲線檢定

前述所得之最佳迴歸曲線，需符合迴歸模型  $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k + \varepsilon$  的假設。包括殘差  $\varepsilon$  為平均值為 0 的隨機變數、殘差  $\varepsilon$  的標準差不隨自變數而更改，殘差  $\varepsilon$  為獨立，殘差  $\varepsilon$  為常態隨機變數。

由於迴歸模型所求之殘差平均值必為 0，因此符合第 1 個假設。而其後假設之符合則由下列方式檢定：

- 標準化殘差需通過卡方常態分配檢定，沒有足夠證據推論其分配不為常態分配。
- 標準化殘差需通過 Durbin-Watson 檢定，沒有足夠證據推論有一階自我相關 (First-Order Auto-Correlation)
- 銷售量與標準化殘差散佈圖中，標準化殘差無明顯隨銷售量增加、減少或週期性波動的趨勢。

最後，選定的曲線必需表現出正評價  $P_{it}$  對銷售量有正面影響，負評價  $N_{it}$  對銷售量有負面影響。意即函式中  $P_{it}$  的係數必需為正， $N_{it}$  的係數必需為負。以 Case 4 的  $k=1$  為例，可得  $Z_{i1} = \sum_{t=0}^{SN_i-1} (S_{it} - (b_{i0} + b_{i1}S_{i(t-L_{i1})} + b_{i2}N_{i(t-L_{i2})} + b_{i3}M_{i(t-L_{i3})} + b_{i4}t))^2$ ，其中求得的  $b_{i2}$  必需為負值。若求得的曲線通過上述所有檢定與符合係數的正負標準，則可視為一有效的迴歸曲線，否則應找尋該 Case 中  $MAPE_{ik}$  次大的曲線，直至曲線通過上述所有檢定與符合係數標準為止。

## 五、演算法複雜度

本演算法一共包含了三個部分：量化商品評價、以最小方差法求得最適迴歸模型與迴歸曲線檢定。其中量化商品評價在每次執行本演算法時只會做到一次，相較其餘二個部分可能重複數千甚至數萬次以上，量化商品評價對整體演算法的效率影響較小。綜觀分析整個迴歸模型的複雜度為  $O(LL_i^4 * SN_i^2)$ ，迴歸曲線檢定的複雜度  $O(LL_i^4 * SN_i)$ ，可得整個演算法的複雜度為  $O(LL_i^4 * SN_i^2)$ ，不為 NP 演算法。

## 伍、實務案例分析

本研究以台灣知名連鎖藥妝店的十大熱銷日系與美系彩妝與保養商品為例（表 1 依銷售金額排名），檢驗「以口碑為基礎的銷售預測演算法」的正確性，訂定所有商品的最大延遲時間皆為半年（ $LL = 26$ ），並從彩妝與保養商品的重要討論區 Fashion Guide 中，搜尋與此十大熱銷彩妝與保養商品相關討論文章。全部十大熱銷商品的文章有 1554 篇，各項商品的討論文章（Number of DWOM）如表 1：

表 1：十大商品與口碑

Product ID	Brand	Product	Number of DWOM	Number of Period	DWOM /Period
1	M 牌（日系）	睫毛膏（mascara）	269	85	3.16
2	T 牌（日系）	卸妝油（makeup remover）	87	92	0.95
3	M 牌（日系）	睫毛膏（mascara）	126	54	2.33
4	U 牌（日系）	洗面乳（facial cleanser）	1	92	0.01
5	T 牌（日系）	護唇膏（lip balm）	233	78	2.99
6	M 牌（日系）	睫毛膏（mascara）	126	67	1.88
7	C 牌（美系）	洗面乳（facial cleanser）	23	82	0.28
8	T 牌（日系）	護唇膏（lip balm）	233	78	2.99
9	M 牌（日系）	卸眼液（eye makeup remover）	44	92	0.48
10	I 牌（韓系）	去光水（polish Remover）	2	92	0.02

由表 1 可發現，口碑量除以期數的比率較高的商品為日系 M 牌睫毛膏、日系 T 牌護唇膏，而比率較低的商品為日系 T 牌卸妝油、日系 U 牌洗面乳、美系 C 牌洗面乳、日系 M 牌卸眼液、韓系 I 牌去光水。使用 Weka 軟體的簡單貝氏分類器訓練所有產品 86% 的文章（1336 篇）來學習文章分類，並用學習的結果分類餘下文章，可得 76.1468% 的準確度（在 218 篇文章中正確分類 166 篇）。在此，我們將

額外討論分類的結果。在分類不正確的情況下，把正面評價的文章分成負面或是負面評價文章分成正面的狀況只占了總分類錯誤比率的 32.7%。大多數 67.3% 的分類錯誤狀況，是將正面評價分成中性、或是負面評價分成中性、亦或是中性評價分成正面或是負面這三種狀況。從此可知，本研究所提出的分類方法，具有一定的效度，因為只有少數的狀況，會產生極端錯誤的分類結果，如：將負面評價的文章判斷為正面評價的文章。而大多數情況是將，負面或正面評價文章判斷為中性，因此對於整體的預測準確度來說並不會造成太嚴重的影響。

將數位口碑法預測 (DWOM) 結果與目前一般預測時常用的 3 期移動平均法 (Moving average)，指數平滑法 ( $\alpha=0.8$ ) (Exponential Smoothing)，趨勢指數平滑法 ( $\alpha=0.8$ 、 $\beta=0.2$ 、起始值=1) (Exponential Smoothing with Trend) 比較。其中由於移動平均法無前 3 期預測資料、指數平滑法無第 1 期預測資料，故統一取第 4 期起至最後一期之 MAPE 為比較標準。最後可得 MAPE (%) 如表 2：

表 2：十大商品預測結果

Product ID	MAPE for				DWOM /Period	Number of Positive comments	Number of Negative comments	Number of Neutral comments
	DWOM	Moving Average	Exponential Smoothing	Exponential Smoothing with Trend				
1	<b>*23.11%</b>	34.38%	43.66%	28.12%	3.16	88	39	142
2	<b>*66.19%</b>	75.84%	269.33%	89.72%	0.95	21	42	24
3	<b>*26.94%</b>	49.93%	66.74%	36.40%	2.33	16	32	78
4	N/A	13.97%	<b>*13.01%</b>	14.04%	0.01	1	0	0
5	<b>*34.86%</b>	70.26%	70.26%	35.99%	2.99	28	55	150
6	<b>*21.84%</b>	37.19%	52.39%	30.07%	1.88	16	32	78
7	30.92%	31.83%	55.37%	<b>*28.80%</b>	0.28	8	8	7
8	<b>*24.22%</b>	47.36%	68.32%	68.32%	2.99	25	3	205
9	<b>*20.78%</b>	33.34%	37.29%	26.13%	0.48	13	13	18
10	22.79%	25.25%	33.71%	<b>*22.09%</b>	0.02	2	0	0

以下針對表 2 中十個商品做分析，分為不適用數位口碑法的商品、適用數位口碑法且預測表現佳的商品與適用數位口碑法但表現不佳的商品。探討後可知當口碑與期數的比率在  $1/3 \approx 0.33$  以上時，適用數位口碑法，相較於其他預測方式可望得到較佳的預測準確度。表 2 中“\*”者標示 MAPE 較小的預測方法，由此結果可知，商品 4 (日系 U 牌洗面乳) 因本身具有數位口碑數量較少，導致不適用數位口碑預測法；商品 7 (美系 C 牌洗面乳) 則因口碑與期數的比率只有 0.28 因此也不適用數位口碑預測法；最後，商品 10 (韓系 I 牌去光水) 亦因口碑與期數的比率只有 0.02 因此也不適用數位口碑預測法。

## 一、不適用數位口碑法的商品

由研究假設條件可知，適用本研究的商品需具話題性，才能在網路上引起足夠的討論量，以利本研究從中挖掘出商品口碑評價並用於模型中。由表 2 可知數位口碑表現不佳或誤差高於其他方法的商品有商品 4（日系 U 牌洗面乳）、商品 7（美系 C 牌洗面乳）與商品 10（韓系 I 牌去光水）。並可知數位口碑表現不佳或誤差高於其他方法的商品，有其共通點為：口碑數量不足。若與銷售資料期數相比較，可發現這三個商品的口碑數量除以期數的比率分別為 0.01、0.28、0.02，未滿  $1/3 \approx 0.33$ ，而其餘七個商品的比率皆在 0.33 以上。故以  $1/3$  為分野，可發現適用本模型的商品為具有足夠數量口碑的商品，即口碑量與期數的比率在  $1/3$  以上。若商品的口碑與期數的比率未滿  $1/3$ ，本研究則建議使用移動平均法、指數平滑法、趨勢指數平滑法等其他方法，可能產生較高的預測準確度。以下圖 2、圖 3、圖 4 展示不適用數位口碑法的商品的銷售實際值與預測值折線圖。

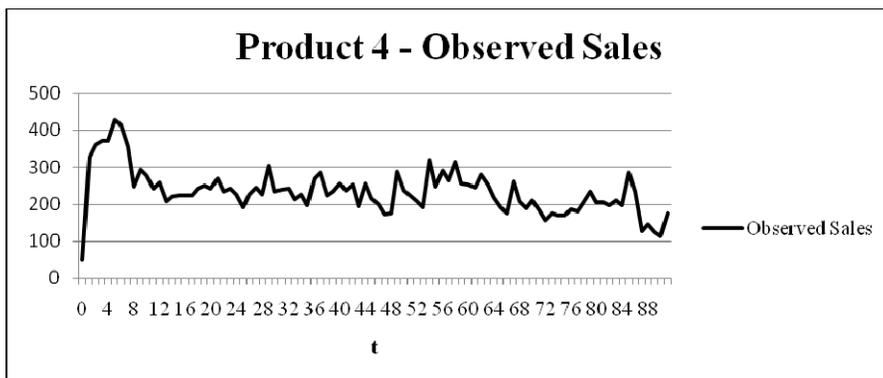


圖 2：產品 4 的銷售實際值折線圖

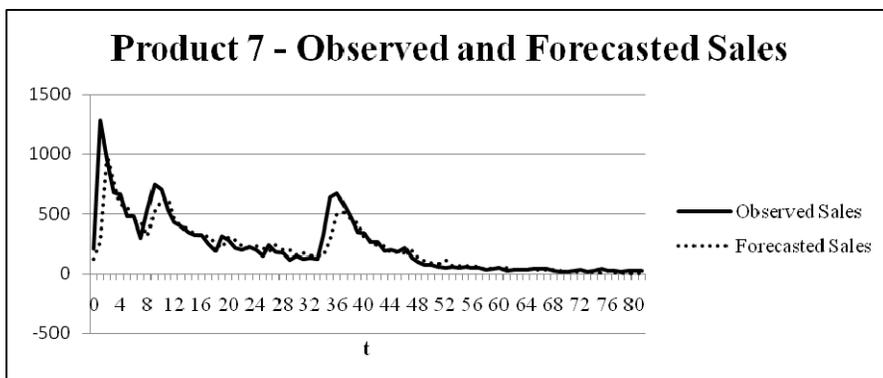


圖 3：產品 7 的銷售實際值與預測值折線圖

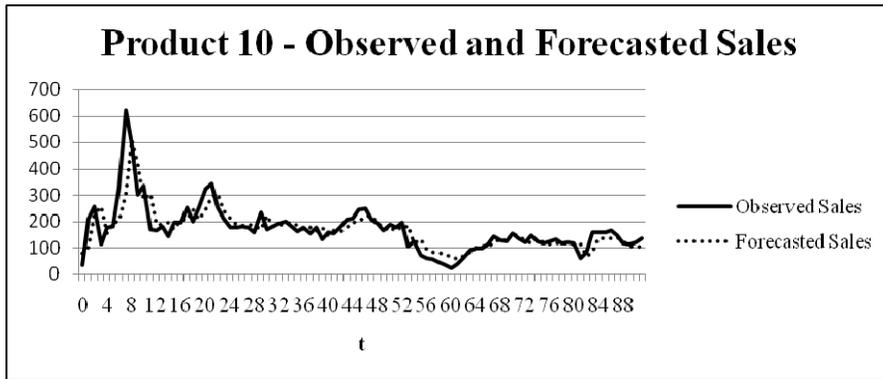


圖 4：產品 10 的銷售實際值與預測值折線圖

## 二、適用數位口碑法且預測表現佳的商品

商品 1 (日系 M 牌睫毛膏)、商品 3 (日系 M 牌睫毛膏)、商品 5 (日系 T 牌護唇膏)、商品 6 (日系 M 牌睫毛膏)、商品 8 (日系 T 牌護唇膏)、商品 9 (日系 M 牌卸眼液) 的所有預測方法中，以數位口碑法 (DWOM) 的預測誤差最小，且平均 MAPE 只有 25.29%。從表 1 故可知這些商品口碑與期數的比率在 1/3 以上，除了商品 9 (日系 M 牌卸眼液) 的數位口碑數與期數的比率較小外，其他商品的數位口碑數與期數的比率均超過 1、甚至 2，表示這些商品推出之後短期間就引起網路討論區的熱烈討論，進一步影響了消費者的購買行為，因此如果流行性商品的商品口碑與期數的比率在 1/3 以上，則可考慮使用本研究提出的數位口碑預測方法，以得到較低的預測誤差與較高的預測準確度。以圖 2、圖 3、圖 4 展示不適用數位口碑法的商品的銷售實際值與預測值折線圖。以圖 5、圖 6、圖 7、圖 8、圖 9、圖 10 展示適用數位口碑法且預測表現佳的商品的銷售實際值與預測值折線圖。

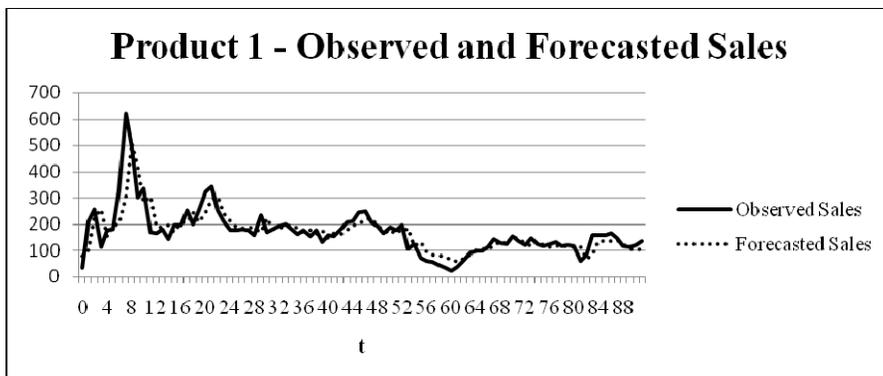


圖 5：產品 1 的銷售實際值與預測值折線圖

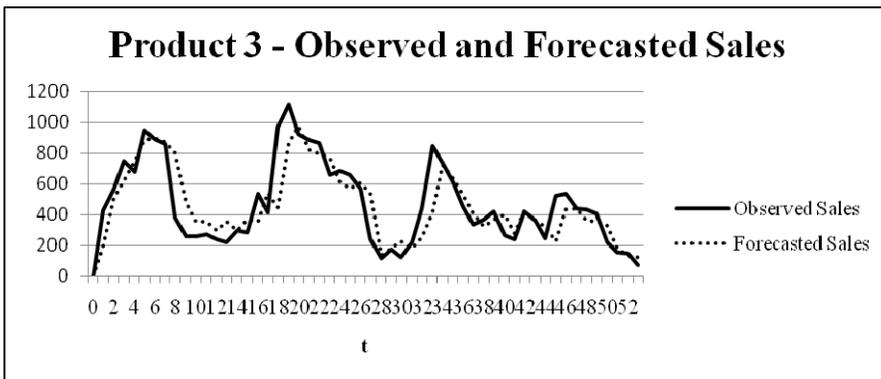


圖 6：產品 3 的銷售實際值與預測值折線圖

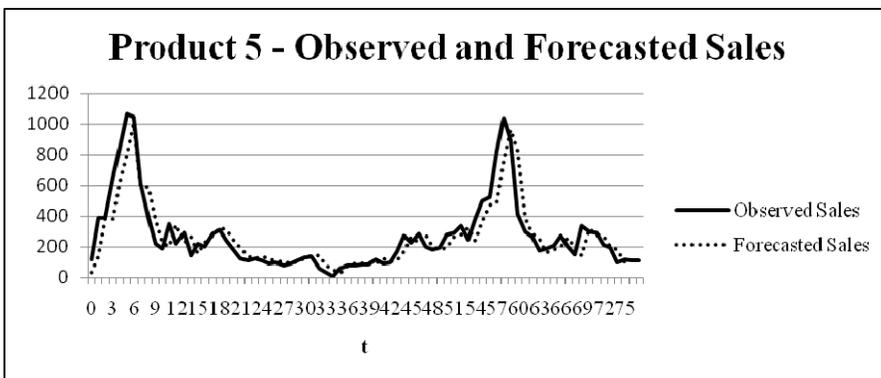


圖 7：產品 5 的銷售實際值與預測值折線圖

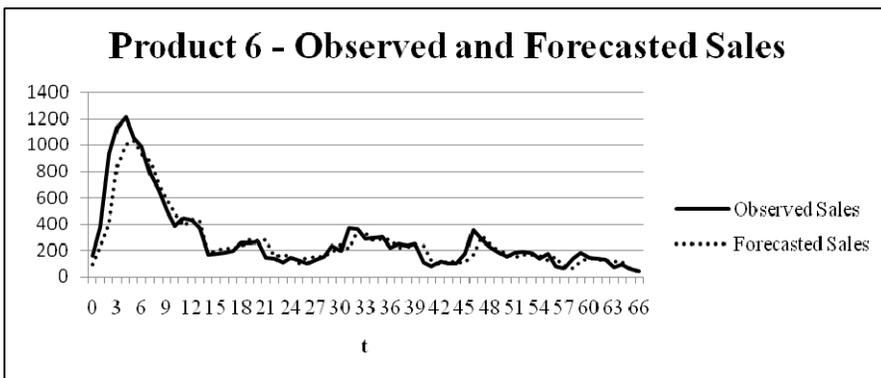


圖 8：產品 6 的銷售實際值與預測值折線圖

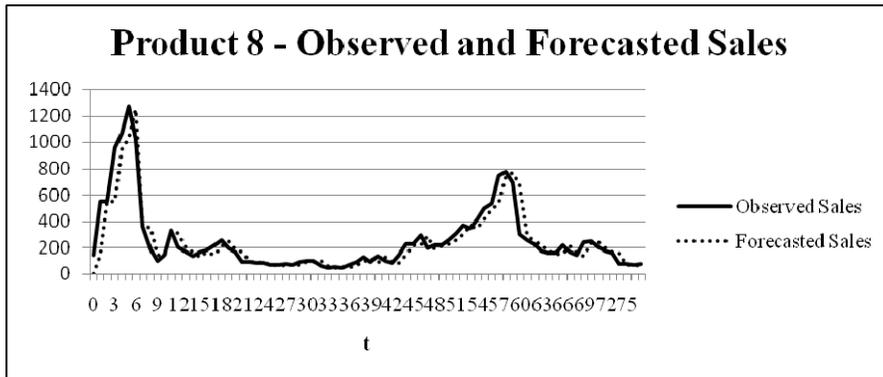


圖 9：產品 8 的銷售實際值與預測值折線圖

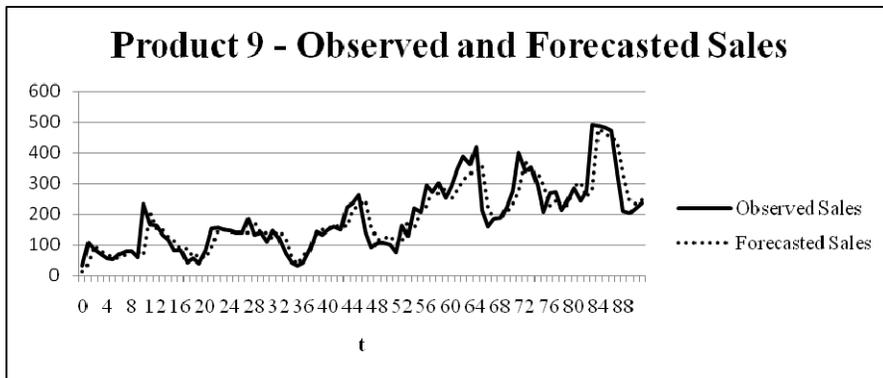


圖 10：產品 9 的銷售實際值與預測值折線圖

### 三、適用數位口碑法但預測表現不佳的商品

雖然商品 2 (日系 T 牌卸妝油) 在所有的預測方法中，以數位口碑法的預測誤差最小，但 MAPE 仍高達 66.19%。檢查該商品的銷售數據，可發現此商品的銷售起伏非常大，有時幾乎沒有銷售量 (0)，亦有時高達 1000，起伏動輒數百。就算用其他預測方法，MAPE 也都在 75% 以上，其中使用移動平均法預測的 MAPE 甚至高達 269.33%。故可知此商品的起伏過大，無論用何種預測方法皆很難準確地預測，但其中仍以本研究的數位口碑法表現最佳。以下圖 11 展示適用數位口碑法但預測表現不佳的商品的銷售實際值與預測值折線圖。

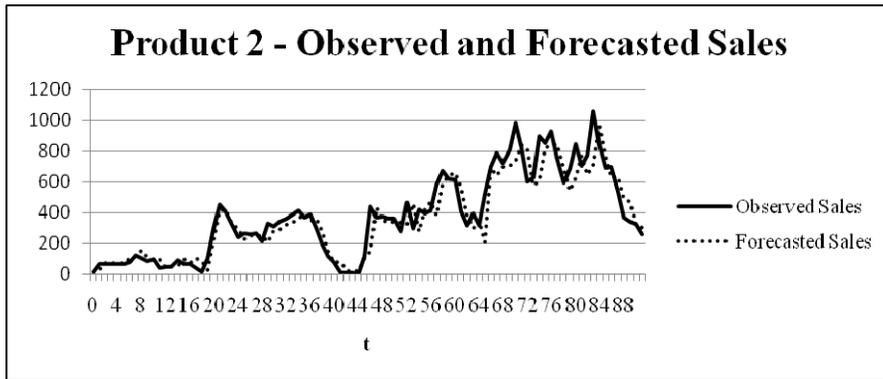


圖 11：產品 2 的銷售實際值與預測值折線圖

#### 四、演算法應用與優缺點

綜觀上述分析數據可知，本演算法適用具有足夠數位口碑的商品，其中以口碑對期數比率在 1/3 以上視為適用。在具有足夠數位口碑的商品上，使用本演算法可得到良好的預測準確度，且準確度皆高於一般常用的移動平均法、指數平滑法與趨勢指數平滑法。但在口碑較少的產品上，使用本演算法則不一定能得到較佳的預測結果，此時應考慮使用其他預測方法。

#### 陸、結論

流行性商品的銷售量大，且起伏也大，使用傳統預測方法往往很難準確預測。因此本研究從口碑著手，想辦法利用口碑來增進銷售預測準確度。但若用問卷調查的方式來收集口碑，不僅成本高昂，且很難輕易地規模化。故本研究設計自動化的收集商品討論文章方式，使用文本挖掘方法分類文章以量化口碑，並用作預測。

由於目前的文本挖掘技術常常著重在自動化地找出名詞性關鍵字以辨別文章的主題，但本研究卻是評斷文章為好評價或壞評價等此類語意性質，其中關鍵字以形容詞居多。且目前的中文文本挖掘研究較難正確地切割形容詞，或自動找出影響文章語意的關鍵形容詞，所以本研究建立了關鍵字字典，其中包含與商品相關的名詞與形容詞，以做為分類之用。

分類後量化後的商品評價可用於以最小方差法求得最適迴歸模型中作為參數，本研究設計多種目標函式，並嘗試過去各種時期的評價與銷售量，以找出預測誤差最小，且符合模型假設與檢定的曲線，作為預測方法。最後以台灣知名連鎖藥妝店的十大熱銷商品為樣本，進行實務案例分析。發現若商品需具話題性、能在網路上引起足夠的討論量（以口碑與期數比率達 1/3 以上為標準），則適用本

模型。且找出的迴歸曲線，其預測誤差 MAPE 皆低於常見的移動平均法、指數平滑法、趨勢指數平滑法等。

未來可望將本模型應用於流行性商品銷售預測，以增進預測準確度。且使用本模型預測某一商品而得之預測方法，亦可套用在廠商後來推出的同類型或改版商品。另外，觀察求得的迴歸曲線可發現，口碑通常會延後發酵，即商品討論串出現時，不會直接反映在銷售上，而會延後一段時間才於銷售數字顯現出效果。且同為熱銷商品，口碑除以期數的比率卻以睫毛膏、護唇膏較高，而卸妝油、洗面乳、卸眼液與去光水較低。代表比起清潔性質的產品，消費者更傾向於討論化妝或保養性質的商品。故廠商在行銷化妝與保養性質的商品時，若要創造銷售量，除了使用傳統的廣告與促銷活動等方法外，亦可針對數位口碑，嘗試在網路中引起討論、帶動商品話題，以提昇商品銷售量。最後，由本研究得知，數位口碑變數將有利於強化或擴增過去預測模式而非取代過去模型，期待未來研究能加入此一議題並深入討論。

## 參考文獻

- Arndt, J. (1967), 'Role of product-related conversations in the diffusion of a new product', *Journal of Marketing Research*, Vol. 4, No. 3, pp. 291-295.
- Bansal, H.S. and Voyer, P.A. (2000), 'Word-of-Mouth processes within a services purchase decision context', *Journal of Service Research*, Vol. 3, No. 2, pp. 166-177.
- Black, K. (2004), *Business Statistics: For Contemporary Decision Making* (4 ed.), John Wiley & Sons, Inc., Hoboken, USA.
- Bnext (2008), Business Next Publishing Corp, Taipei.
- Buttle, F.A. (1998), 'Word of mouth: understanding and managing referral marketing', *Journal of Strategic Marketing*, Vol. 6, No. 3, pp. 241-254.
- Chopra, S. and Meindl, P. (2007), 'Supply chain management: strategy, planning, and operation', *Business and Economics*, pp.265-275.
- Dalrymple, D.J. (1987), 'Sales forecasting practices: results from a united states survey', *International Journal of Forecasting*, Vol. 3, No. 3-4, pp. 379-391.
- Fisher, M. and Rajaram, K. (2000), 'Accurate retail testing of fashion merchandise: methodology and application', *Marketing Science*, Vol. 19, No. 3, pp. 266-278.
- Geurts, M.D. and Whitlark, D.B. (2000), 'Six ways to make sales forecasts more accurate', *The Journal of Business Forecasting Methods & Systems*, Vol. 18, No. 4, pp. 21-30.
- Gelb, B. and Johnson, M. (1995), 'Word-of-Mouth communication: causes and

- consequences', *Journal of Health Care Marketing*, Vol. 15, No. 3, pp. 54-58.
- Golder, P.N. and Tellis, G.J. (1997), 'Will it ever fly? Modeling the takeoff of really new consumer durables', *MARKETING SCIENCE*, Vol. 16, No. 3, pp. 256-270.
- Gremler, D.D., Gwinner, K.P. and Brown, S.W. (2001), 'Generating positive Word-of-Mouth communication through customer-employee Relationships', *International Journal of Service Industry Management*, Vol. 12, No. 1, pp. 44-59.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I.H. (2009), 'The WEKA data mining software: an update', *SIGKDD Explorations*, Vol. 11, No. 1, pp. 10-18.
- Hennig-Thurau, T., Gwinner, K.P., Walsh, G. and Gremler, D.D. (2004), 'Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the Internet? ', *Journal of Interactive Marketing*, Vol. 18, No. 1, pp. 38-52.
- Huang, J., Lu, J. and Ling, C.X. (2003), 'Comparing naive Bayes, decision trees, and SVM with AUC and accuracy', *Third IEEE International Conference on Data Mining*, pp. 553-556.
- Izumi, K., Matsui, H. and Matsuo, Y. (2007), 'Integration of artificial market simulation and text mining for market analysis', *Advances in Hybrid Information Technology*, Vol. 4413, pp. 404-413.
- Jain, C.L. (1998), 'Quick and easy ways to monitor forecasts', *Journal of Business Forecasting Methods and Systems*, Vol. 17, No. 2, pp. 2-30.
- Kahn, K.B. (1998), 'Benchmarking sales forecasting performance measures', *Journal of Business Forecasting*, Vol. 17, No. 4, pp. 19-23.
- Klepper, S. (1996), 'Entry, exit, growth, and innovation over the product life cycle', *American Economic Review*, Vol. 86, No. 3, pp. 562-583.
- Kroha, P., Baeza-Yates, R. and Krellner, B. (2006), 'Text mining of business news for forecasting', *Proceedings of the 17th International Conference on Database and Expert Systems Applications*, IEEE Computer Society, Krakow, Poland, pp. 171-175.
- Kuo, R.J. and Xue, K.C. (1999), 'Fuzzy neural networks with application to sales forecasting', *Fuzzy Sets and Systems*, Vol. 108, No. 2, pp. 123-143.
- Lo, S. (2008), 'Web service quality control based on text mining using support vector machine', *Expert Systems with Applications*, Vol. 34, No. 1, pp. 603-610.
- Olshavsky, R.W. and Granbois, D.H. (1979), 'Consumer decision making-fact or fiction? ', *Journal of Consumer Research*, Vol. 6, No. 2, pp. 93.

- Polli, R. and Cook, V. (1969), 'Validity of the product life cycle', *The Journal of Business*, Vol. 42, No. 4, pp. 385-400.
- Ratchford, B.T., Talukdar, D. and Lee, M.-S. (2001), 'A model of consumer choice of the Internet as an information source', *International Journal of Electronic Commerce*, Vol. 5, No. 3, pp. 7-21.
- Sakurai, S. and Ueno, K. (2004), 'Analysis of daily business reports based on sequential text mining method', *Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics*, Hague, Netherlands.
- Salton, G. and McGill, M.J. (1986), *Introduction to Modern Information Retrieval*, McGraw-Hill, New York, USA.
- Schrieber, J. (2005), 'Demand visibility improves demand forecasts', *Journal of Business Forecasting*, Vol. 24, No. 3, pp. 32-37.
- Sebastiani, F. (2002), 'Machine learning in automated text categorization', *ACM Comput. Surv.*, Vol. 34, No. 1, pp. 1-47.
- Sheikh, K. (2003), *Manufacturing Resource Planning (MRP II) with an Introduction to ERP, SCM, and CRM*, McGraw Hill Companies, Inc..
- Taylor, B.W. (2004), *Introduction to Management Science*(8 ed.), Pearson Education, Inc., Upper Saddle River, USA.
- Ting, T.-W. (2007), 'A collaborative planning forecasting and replenishment solution for retail industry', *Department of Information Management*, National Taiwan University, Taipei, Taiwan.
- Witten, I.H. and Frank, E. (2005), *Data Mining, Practical Machine Learning Tools and Techniques* (2 ed.), Morgan Kaufmann Publishers, San Francisco, USA.
- Yang, H.-C. and Lee, C.-H. (2008), 'Image semantics discovery from web pages for semantic-based image retrieval using self-organizing maps', *Expert Systems with Applications*, Vol. 34, No. 1, pp. 266-279.