

網路服務品質探勘與管制

Web-Service Data Quality Mining and Control

羅淑娟 Shu-Chuan Lo

國立台北科技大學工業工程與管理研究所

Department of Industrial Engineering and Management, National Taipei University of Technology

柯秀奎 Shiou-Kuei Ke

國立台北科技大學工業工程與管理研究所

Department of Industrial Engineering and Management, National Taipei University of Technology

林晶璟 Ching-Ching Lin

國立台北科技大學通識教育中心數學組

Math Group of General Education Center, National Taipei University of Technology

摘要：一個受歡迎的網站可能會從網路會員處收到千百封的留言，最重要的訊息是屬於顧客關於技術需求及不滿意的怨言。本研究提出一個自動機制(WebQC)基於文件探勘和支援向量機(SVM)技術來分類顧客的留言，此機制可以自動過濾抱怨的留言並且正確地增加客服部的生產力以及顧客滿意度。本研究在抱怨率上使用了不良率管制圖(p-control chart)來檢查服務品質是否低於網站運行的期望水準，並以一社群網站案例作為實驗案例，根據實驗結果顯示，其分類正確率或檢定力(如果留言為抱怨，SVM可以辨識為抱怨)的正確率超過83%(平均值為89%)，而不良率管制圖可適時地反應出非隨機的狀況。

關鍵字：文件探勘；支援向量機；文件分類；補償不良率管制圖；網路服務品質。

Abstract: A popular Web site may receive hundreds of thousands of messages each day from site users. The most useful customer feedback consists of technical requests and complaints. In this research, we propose the WebQC software to classify a user's feedback messages by using text-mining techniques and a support vector machine (SVM). Our software can filter the messages as

complaints or other kinds of messages automatically and thus effectively increase the productivity of the customer service department and as well as customer satisfaction. In this research, we employ the p -control chart on the complaint scale to check the quality of the Web site services if this quality is lower than expected. We used a community Web site as our test case. The empirical results show that our ability to classify a message accurately (i.e., if a message is a complaint, the SVM recognizes it as a complaint) is over 83% (on average 89%). Also, the p -control chart has the ability to reflect a normal situation in real-time.

Keywords: Text mining; Support vector machine (SVM); Classification; compensating p -control chart; Web service quality

1. 緒論

社群網站通常具備了多種功能，包含很多有趣的內容、社群和電子商務，希望吸引更多的網路使用者加入，企圖增加交易次數和交易金額，以達成獲利的目的。由於有太多類似的社群網站分食了有限的市場，社群網站面對了嚴重的挑戰。為了增加會員的數量以及收入，社群網站的執行者常以持續不斷地更新網站功能，作為在嚴格的競爭環境下的生存的策略。

大部分的網站都提供了留言板的功能，收集顧客的怨言和需求。一個受歡迎的社群網站每天會湧入上千百封的留言，造成客服部“資料爆炸”的現象，目前雖然有些自動答詢機制在發展中(Berger et al., 2000; Hirschman and Gaizauskas, 2001)，由於語意解析還不甚精準，網路公司若要維持現有服務水準，還須仰賴人工辨識，現有文獻中，多著重於自動答詢技術的發展，而忽略了技術以外的管理連結，因此本研究僅以自動機制來過濾留言板中具有服務需求意義的留言，並將這些需求予以分類，可作為預先處理分派的基礎，本研究更將此機制連結品質管制機制，當抱怨留言件數高過以往水準時，便提出警訊，此研究可作為網站服務品質管制的具體建置參考，對於經常更新經營策略的網站來說，可利用此本研究所提之方案，監控新的經營策略，檢視客戶的報怨反應是否較以往嚴重，若是將可及時獲得警訊，以有效地進行挽救。

本研究提出的自動機制可以辨識抱怨留言，並可在怨言量超過平常的水準時發佈警告訊號，將此機制稱為中文網站怨言品管機置(Web Quality Control, WebQC)。我們先使用文件探勘技術對關鍵字作TFIDF(Term-Frequency Inverse-Document-Frequency)的權重計算(Salton and McGill, 1983)，並使用 SVM(Support Vector Machine)將留言分類為四大類(Cortes and Vapnik, 1995)；分別為非中文留言或亂碼、技術問題、不滿意怨言及其他類別，本研究以技術問題及不滿意的留言認定為怨言，我們使用不良率管制圖來管制抱怨率，如果抱怨率高於管制上限，則 WebQC 將給予警告訊息顯示服務品質降低了。

本文其餘章節內容如下：第二節我們回顧一些文件探勘的相關技術；第三節描述 WebQC 的自動機置結構及在本研究所使用的研究方法，包括了TFIDF、SVM 及不良率管制圖；第四節提出 WebQC 運用在一個社群網站上的實驗結果；最後在第五節作出結論。

2.文件探勘

2.1 文件分類

文件分類是將自然語言依事前定義的集合作主題歸類 (Sebastiani, 2002)。文件分類始於 60 年代，但是直到 90 年代早期才變成主要的子領域。80 年代晚期，大部分受歡迎的文字件分類分法是文件分類在給定分類的條件下，根據專家知識規則下進行辨識編碼的知識工程。然而 90 年代在文件分類上一般方法是建構在機器學習的自動文件分類上，機器學習是藉由從預先分類好的文件集合來進行學習。

文件分類是對於每一組 $(d_i, c_j) \in D \times C$ 指派一個布林數值， D 為整個文件集合， C 為預先定義分類的集合。在 (d_i, c_j) 中的布林值為 T 時，表示文件 d_i 隸屬於 c_j 類別，當布林值為 F 則表示 d_i 不屬於 c_j 類別。文件分類可被表示為 $\Phi: D \times C \rightarrow \{T, F\}$ ， Φ 為一未知函式。分類器 ϕ 為 Φ 的估計值，本研究在分類上不重疊，也就是說文件 $d_i \in D$ ， $c_j \in C$ ，如果分類器 $\phi(d_i, c_j) = T$ ，則其它的 $\phi(d_i, c_k) = F$ ，其中 $k=1, \dots, |C|$ 且 $k \neq j$ ， $C = \{c_1, \dots, c_{|C|}\}$ 具有 $|C|$ 個獨立類別。

2.2 中文斷詞

斷詞將文件區分成語言學上的基本單元，成為一些共同的字，對於文件分類來說是首要的條件。我們藉由關鍵字對抱怨作資訊(語意)擷取，並為每個抱怨類別建立關鍵字規則。目前有三種主要的中文斷詞與文字擷取方法，分別是詞庫式斷詞(Case and Zeng, 1991)、統計式斷詞(Fan and Tsai, 1988; Sproat and Shih, 1990)以及混合式斷詞(Nie et al., 1996)。

詞庫式斷詞是基於字/片語字典或是詞典，利用最大配對(或最長配對)演算法選擇一串字使其包含最長的詞；統計式方法是以訓練資料依手動斷詞產生的詞集合，其中字詞的出現頻率，作為該詞的統計資訊。簡單的統計方法就是計算一個詞在字串 S 中出現的機率 $P(S) = \frac{CW(S)}{C(S)}$ ， $CW(S)$ 是訓練資料中字串 S 的某一斷詞出現數量， $C(S)$ 是訓練資料中字串 S 出現數量。當給定一個輸入字串作斷詞時，其最佳解是一連續組合詞 S_i 使得 $\prod_i P(S_i)$ 最高。

混合式方法結合了詞庫斷詞與統計斷詞方法。使用詞庫字典來分割所有種類的字詞，然後使用統計機率來決定最佳的分割。

2.3 資訊擷取

資訊擷取或稱文件擷取，是從文件中萃取出某些有意義的內容和方向。基本上每個文件都是由許多關鍵字所組成的。如果我們使用向量來表示文件，那麼文件就是由所有關鍵字投影長度所組成的向量。關鍵字擷取主要用來辨識有意義且文件中代表性的詞彙或片語。

文件切割後產生了多種的詞彙包括了名詞、動詞、代名詞、冠詞、連接詞以及介系詞等，這些結果對於資訊擷取而言並非全部必要。如果我們刪除了代名詞、冠詞、連接詞及介系詞的話，文件的意義仍然不會被扭曲。然而，關鍵字擷取在文件斷詞後才會作處理。

一般對於關鍵字擷取有三種方法，分別是詞庫比對法、文件剖析法及統計分析法。詞庫比對法使用預先定義的詞組字典擷取關鍵字；文件剖析法是藉由語言學與一些規則過濾無意義的片語後，擷取文件中所有的名詞、動詞、代名詞或介系詞等；統計分析法藉由一個文件或文件集合項目中的權重比例擷取關鍵字。Salton 與 McGill (1983) 提出如果我們想要決定文件中某項目的重要性與代表性，我們可以計算此文件中該項目頻率(TF)及此項目在其他文

件中的出現頻率，稱為反轉文件頻率(IDF)。當一個項目的 TF 值愈高時，就意味著此項目在此文件中愈重要；而 IDF 的目的則是找出文件中可用來區別的項目。當一個項目的 IDF 愈低時，就意味著這個項目愈可以拿來區隔其他的文件。統計資訊擷取結合了這兩種索引方式形成 TFIDF，其定義如下：

$$\begin{aligned} TFIDF(W_i, M) &= TF(W_i, M) \times IDF(W_i) \\ &= TF(W_i, M) \times \log\left(\frac{|D|}{DF(W_i)}\right) \end{aligned}$$

$TFIDF(W_i, M)$ 是在文件 M 中關鍵字 W_i 的權重； $|D|$ 為文件中所有留言的數目； $DF(W_i)$ 是在留言集或文件集之中關鍵詞 W_i 出現的頻率。

本研究我們同時使用詞庫比對法與統計分析法進行關鍵字的擷取，先藉由訓練資料建立詞庫並依詞庫比對法擷取候選關鍵字，然後再使用統計分析法計算 TFIDF 值，並建立一門檻值過濾候選值低於門檻值的關鍵字。留言可以表示為一個關鍵字向量，所形成的矩陣 M 定義如下：

$$M = \begin{bmatrix} w_{11} & w_{12} & \dots & \dots & w_{1p} \\ w_{21} & w_{22} & \dots & \dots & w_{2p} \\ \vdots & \vdots & \dots & \dots & \vdots \\ \vdots & \vdots & \dots & \dots & \vdots \\ w_{m1} & w_{m2} & \dots & \dots & w_{mp} \end{bmatrix},$$

其中， w_{ik} 為文件 $i(i=1, \dots, m)$ 中關鍵字 k 的權重(TFIDF 之值)， $m=|D|$ ； $k=1, \dots, p$ ， p 為我們在所有留言中所擷取的關鍵字數量。

2.4 自動答詢機制

隨者網路資料庫的聯通與分享，使用者可以藉由網路取得大量的資料與資訊，篩選龐大的網路資訊而進一步產生有用且符合需求的知識，僅憑藉人工或傳統的搜尋引擎來完成此工作，將是一沉重的負擔，且執行者也需要相當的背景知識與訓練，才有能力執行此工作，目前針對此問題的解決方案較具體的目前有常見問題查詢技術(Frequently Asked Question, FAQ)、自動詢答技術(Automatic Question Answering System)和自動問答探索技術(Automatic Answer-Finding System, AAF System)，此三種技術內容與差異，將

簡述如下。

常見問題查詢技術是搜集一些工作環境或生活中常見的問題，將這些問題以一問一答(Q&A)的方式呈現，使用者以自然語言問句輸入系統，問句可視為一些索引關鍵詞的組合，一般配合「分類目錄式架構」或計算問句與常見問題集的相似度，作為可能答案的搜尋，再由使用者確認答案。目前較為知名的常見問題查詢有 AskJeeves (<http://www.ask.com>)(吳志鴻，民 90；賴育昇，民 91)、FAQFinder (<http://www1.ics.uci.edu/~burke/faqfinder>)(Hammond et al., 1995；Burke et al., 1997)和 FAQ.org (<http://www.faq.org>)。

自動詢答技術是指可接受使用者所提之自然語言問句，並由一堆文件中尋找此問題的解答，回覆使用者的答案所包含的前後文，足以讓使用者確認其正確性(Hirschman and Gaizauskas, 2001)。答詢系統與 FAQ 的不同在於 FAQ 視問句為關鍵詞組合，利用搜尋引擎至預存的問題集中搜尋可能答案，但答詢系統需具備有解析問句語意能力，而其也不需預存文件集，而是由可接觸的任何資料庫中找尋答案。

自動問答探索技術(Berger et al., 2000)是將知識工作者曾經回答過的問題及答案視為文件集(即常見問題集)，此種文件有別於一般精檢索的文件集，對於某種知識領域，自動問答探索系統利用常見問題集回答使用者所提問題，比透過傳統檢索技術所得文件集或透過自動詢答系統所得關鍵詞答案，有較佳的解答效果與完整概念，自動問答探索系統可說是結合常見問題查詢技術與自動詢答技術優點發展而成的系統。

本研究的文件處理技術較接近 FAQ，但本研究的應用目的並非在文件探勘技術，而是應用探勘技術提升網路公司客服處理效率與輔助監控服務品質，此概念的提出，就我們所知並未見之於文獻中。

3. 研究方法

3.1 WebQC 架構

圖 1 為 WebQC 的架構，在此架構中我們抽取訓練用留言樣本並且先手動將它們辨識分類，接著使用中文文件斷詞工具將留言切割成詞彙或項目，本研究我們使用由中央研究院詞庫小組(CKIP, 2006)研發的混合式斷詞系統作為我們中文斷詞工具，稱為 AutoTag，此中文詞庫約有 10 萬個詞。使用 Autotag 斷詞後，我們任意選取一個門檻值，經由 TDIDF 規則擷取關鍵字。

如果一個詞彙或項目的 TFIDF 值大於門檻值，那麼就認定此詞彙/項目為關鍵字，並將這些關鍵字儲存於留言特徵資料庫中。我們將訓練用的留言利用 TFIDF 值轉換成關鍵字向量並送入 SVM 分類器分類，所得到的結果與先前手動分類的正確結果作一比較，利用門檻值的改變試著找出最大的正確率。右邊的流程建立了留言特徵資料庫與 SVM 的分類規則，並估計了 SVM 的分類誤差。在結束訓練步驟後便開始進行左邊的流程，建立所有留言的 TFIDF 留言矩陣 M 。在分類後我們可以從不良的類別中計算不良率，並嘗試建立初期的補償 p 管制圖，在經過一些調整後建構連續的管制標準。這些標準可作為對於網站服務品質的一測量方法。如果不良率超出標準，那麼管制圖會發布一個警訊給網站管理者以達到品質控制的目的。

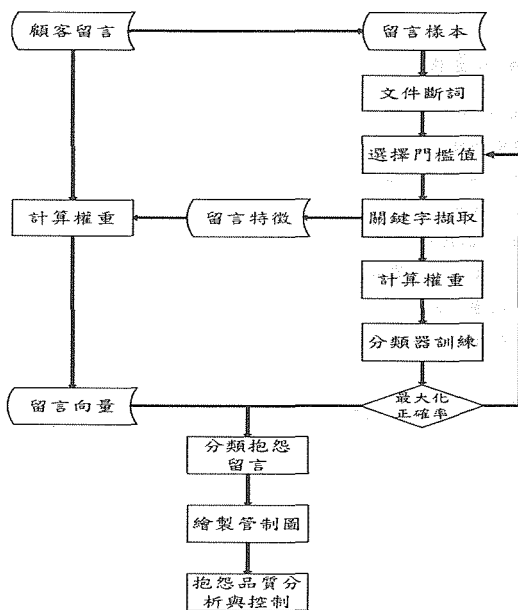


圖 1 留言分類架構

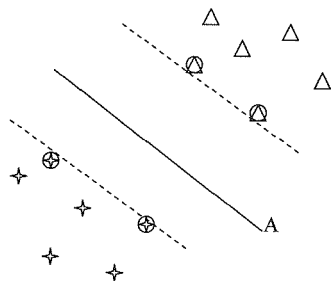


圖 2 支援向量機(SVM)

3.2 支援向量機

支援向量機(SVM)是由 Cortes 與 Vapnik (1995)所提出的,其目的為解決在分類辨識問題中兩個群體的分類問題。SVM 找出一個最佳的分割面(決策面)將所有的資料區分為兩類。決策面是一個能將兩個類別的向量分開最大邊的超平面,如圖 2。

此超平面藉由訓練資料中的部份資料所決定,在圖 2 中圓圈起來的訓練資料稱作支援向量,實線 A 為超平面並滿足條件 $w \cdot x + b = 0$, 其中 w 為超平面的標準向量, x 為支援向量, b 為一數值。當兩個不同類別間訓練向量的投影量距離最大時決定了超平面 $w/|w|$ 的方向。此距離 $\Phi(w)$ 如下所示:

$$\Phi(w) = \min_{\{x: y=1\}} \frac{x \cdot w}{|w|} - \max_{\{x: y=-1\}} \frac{x \cdot w}{|w|} \quad (1)$$

我們將訓練資料的集合標記為 (x_i, y_i) , $i=1, \dots, l$, $x \in R^n$, n 為輸入維度;

$y \in \{+1, -1\}$, 如果存在一向量 w 與一數值 b 則此線性分隔可表示為下列不等式:

$$\begin{aligned} w \cdot x_i + b &\geq 1 && \text{當 } y_i = +1, \\ w \cdot x_i + b &\leq -1 && \text{當 } y_i = -1, \end{aligned} \quad (2)$$

上式適用於訓練資料集的所有元素。公式 2 可以被改寫成下式：

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0 \quad , \quad i = 1, \dots, l \quad (3)$$

最佳的超平面 (\mathbf{w}_0, b_0) 就是公式(1)最大化的距離的解。最佳的超平面也可寫成下式：

$$\min \Phi(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2, \quad (4)$$

需滿足條件(3)且其解唯一。在建構一個最佳的超平面為一二次方程式問題，這可以藉由 Lagrange 乘數法來解決。從(3)與(4)中我們可得：

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^l \alpha_i \{y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1\} \quad , \quad \alpha_i \geq 0, i = 1, \dots, l \quad (5)$$

藉由 Fletcher (1987)所提出的 Karush-Kuhn-Tucker (KKT)理論，我們得到：

$$\alpha_i (y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1) = 0 \quad \forall i$$

從上列等式中 α_i 不等於 0 時，那麼會導出：

$$y_i(\mathbf{x}_i \cdot \mathbf{w}_0 + b_0) = 1 \quad (6)$$

換句話說如果 $\alpha_i \neq 0$ ，則向量 \mathbf{x}_i 將會稱為支援向量。在公式(2)中此分類函式可以被改寫為：

$$f(\mathbf{x}) = \mathbf{x} \cdot \mathbf{w}_0 + b_0 = \sum_{i=1}^l y_i \alpha_i (\mathbf{x} \cdot \mathbf{x}_i) + b_0 \quad (7)$$

如果 $f(\mathbf{x}) > 1$ ，則可表示為此資料向量 \mathbf{x} 屬於在標記 $y = +1$ 中的分類，反之向量 \mathbf{x} 屬於在標記 $y = -1$ 中的分類。

關於文字分類器的選擇，許多學者指出 SVM 的正確性較其它的演算法優越(Lee and Yang, 2005; Joachims, 1998)，雖然在 SVM 中訓練資料的樣本大小較小，卻在分類上仍然可以得到不錯的結果。所以本研究使用 SVM 做為本研究的分類器且選擇 Statistica V6.0 版本中 SVM 資料探勘模組作為我們分析工具。

3.3 怨言品質管制

當一則留言經過轉換為 TFIDF 關鍵字向量後我們稱為留言向量，並將留言向量輸入到 SVM 分類器中以決定屬於哪一個分類。本研究我們建立四個分類，分別為非中文或亂碼類別、技術問題類別、抱怨類別以及其他類別。我們認為留言屬於技術問題類別或抱怨類別是屬於不良的服務，所以我們記錄了每週的不良率，不良率管制圖(p control chart)在此被使用來檢查線上服務的不良率是否保持穩定且達到服務的品質管制目標。不良率管制圖在製造業的品質管制上是一個通用的工具，我們確信管制圖在服務品質管制上同樣是一個有效的工具，在 3.3.1 節將作不良率管制圖基本理論的描述。當分類器檢查有誤差時，需使用補償不良率管制圖，其內容於 3.3.2 節中解釋。

3.3.1 不良率管制圖

不良率管制圖是一個統計上的時間序列圖形，主要目的是識別隨機差異與非隨機差異，其基本理論是從統計學上得到。管制圖在一些統計的樣本分配上建立一個管制界限。藉由中央極限定理得知當樣本數夠大時，樣本平均數的樣本分配近似於常態分配，樣本平均數分配的期望值為母體平均數。型 I 誤差是指母體平均數未改變但樣本平均數卻落在管制界限外的一種錯誤判斷，管制圖基本原理即藉由型 I 誤差的規劃來建立管制界線，如果我們設定型 I 誤差機率為 0.027 時，則管制界限為樣本分配的平均數 ± 3 倍樣本分配的標準差。在二項式的實驗中對於不良率 p 的點估計的統計量為 $\hat{p} = x/n$ ， x 為 n 次實驗中的成功次數。在每一次二項式實驗中不良品被指定為 1，良品則被指定為 0，而成功的次數 x (不良品的個數)可解釋為 n 個 0 與 1 的總合，而樣本不良率 \hat{p} 則是 n 個值的樣本平均數。因此藉由中央極限定理，當 n 值夠大時， \hat{p} 進似於常態分配，其期望值為：

$$\mu_{\hat{p}} = E(\hat{p}) = E\left(\frac{x}{n}\right) = \frac{np}{n} = p$$

變異數為：

$$\sigma_{\hat{p}}^2 = \sigma_{x/n}^2 = \frac{\sigma_x^2}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

所以我們可以確定不良率管制圖的中心線(CL)=樣本分配 \hat{p} 的平均數， $p = \hat{p}$ 。

$$\text{管制上限(UCL)} = \hat{p} + 3 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n_i}}$$

$$\text{管制下限(LCL)} = \hat{p} - 3 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n_i}} \quad (8)$$

\hat{p} 可藉由 $\bar{p} = \frac{\sum n_i \hat{p}_i}{\sum n_i}$ 估計， \hat{p}_i 為第 i 個樣本的不良率， n_i 為第 i 個樣本的樣本大小。

3.3.2 不良率管制圖於檢驗誤差的修正

在理想情況下，分類器能正確的判斷顧客的留言而沒有錯誤的分類，但是在現實的分類器中在分類判斷上是會有一些錯誤的。SVM 在留言分類的判斷上會有兩種錯誤，一種是非抱怨留言卻被分類成抱怨留言(此錯誤機率表示為 $e1$)；另一種是抱怨留言被分類成非抱怨留言(此錯誤機率表示為 $e2$)。因為這兩種錯誤在管制界限的決定上可能造成嚴重的誤差。所以當不良率管制圖中檢驗錯誤存在時我們需要修正(8)式。不良率圖檢驗誤差的修正式如下(Case, 1980)：

$$CL_{pe} \text{ (補償管制中心線)} = (1 - \bar{e1} - \bar{e2}) \times CL + \bar{e1},$$

$$UCL_{pe} \text{ (補償管制上限)} = (1 - \bar{e1} - \bar{e2}) \times UCL + \bar{e1},$$

$$LCL_{pe} \text{ (補償管制下限)} = (1 - \bar{e1} - \bar{e2}) \times LCL + \bar{e1}, \quad (9)$$

在此 $\bar{e1} = \frac{\sum e1}{k}$ ， $\bar{e2} = \frac{\sum e2}{k}$ ， k 為樣本組數。舉例來說，一周內有共 75 篇留言，其中 43 篇為抱怨留言；假設分類器辨識出其中 38 篇為抱怨留言，檢驗誤差 $e1$ 和 $e2$ 分別為 0.13 和 0.18，則根據(8)式和(9)式我們得到：

$$CL_p = \bar{p} = \frac{43}{75} = 0.58,$$

$$UCL_p = 0.58 + 3 \times \sqrt{\frac{0.58 \times (1 - 0.58)}{75}} = 0.75,$$

$$LCL_p = 0.58 - 3 \times \sqrt{\frac{0.58 \times (1 - 0.58)}{75}} = 0.409,$$

$$CL_{pe} = (1 - 0.13 - 0.18) \times 0.58 + 0.13 = 0.53,$$

$$UCL_{pe} = (1 - 0.13 - 0.18) \times 0.75 + 0.13 = 0.65,$$

$$LCL_{pe} = (1 - 0.13 - 0.18) \times 0.409 + 0.13 = 0.412.$$

如果在檢驗誤差 $e_1=13\%$ 和 $e_2=18\%$ ，真實的不良率為 58% ，則 $C_{Lp}=0.58$ 應被修正為 $CL_{pc}=0.53$ ，原 $UCL_p=0.75$ 、 $LCL_p=0.409$ 應分別被修正為 $UCL_{pc}=0.65$ 與 $LCL_{pc}=0.412$ 。

4. 實驗與討論

4.1 資料背景

本實驗資料來自台灣一社群網站，網站的會員人數大約有七萬人，有 90% 的人的年齡小於 30 歲。會員在網站上購買服裝及配件來裝飾自己圖像模型來表現自我，並在網站上利用日記和生活照圖片的分享來交朋友。本研究在 2004 年三月到十二月間收集了 7143 篇有效的會員留言，並且希望監控服務水準的變化來了解網站的服務品質是否在管制下。經過和網站管理者的討論，我們使用週作為時間週期單位，因為大部分的管理者認為時間單位週是最經濟的管理週期。我們計算每週的不良率，週不良率即為我們不良率管制圖的管制物件。

為了建立具代表性的留言字詞資料庫，本研究隨機抽取三到六月中 1000 則留言樣本並使用中央研究院詞庫小組 CKIP 中文自動斷詞系統 AutoTag 軟體進行最長字詞斷詞，部份加入一些人為知識來建立主要的留言特徵資料庫。藉由 TFIDF 統計分析從留言特徵資料庫中擷取的關鍵字。在擷取的過程中，我們需要設立一個門檻值，如果字詞 TFIDF 的權重小於此門檻值，那麼此字詞將不被包括在關鍵字集合中。本研究建立非中文或亂碼、技術問題、抱怨留言及其他四個分類，並人工預先標記 1000 篇留言的分類，再將所有的標記留言都送入 SVM 分類器進行分類。我們調整門檻值使分類器訓練留言的正確性最大化。根據 WebQC 的架構，在 SVM 分類正確率最大化目標下， 1000 篇留言中關鍵字為 224 個，平均正確率可以達到 89% ，如表 1 所示：

表 1 SVM 分類正確率

| 類 別 | 訓 練 數 | 正 確 分 類 數 | 正 確 率 |
|--------|-------|-----------|-------|
| 非中文或亂碼 | 21 | 18 | 86% |
| 技術問題 | 531 | 494 | 93% |
| 抱怨留言 | 12 | 10 | 83% |
| 其 它 | 436 | 411 | 94% |

4.2 研究結果

由於在三月時網站提供許多新的服務，為了避免新服務造成不穩定地結果，所以我們正式對顧客留言進行分類選擇從五月開始。留言屬於問題類以及抱怨類的我們認為在服務上屬於是不良的留言。我們嘗試藉由五月和六月的留言資料，估計這兩種分類誤判風險，並於六月留言中估計週不良率，結果列於表 2。

表 2 第二行為總留言數，第三行為人為判斷的不良率，第四行為 SVM 判斷的不良率，而 $e1$ 與 $e2$ 則分別列於第 5 行與第 6 行，最後一行則是 SVM 分類的正確率。從表 2 可以得到 $\bar{p} = \frac{\sum ni\hat{p}_i}{\sum ni} = 0.57$, $\bar{e1} = \frac{\sum e1}{k} = 0.13$, $\bar{e2} = \frac{\sum e2}{k} = 0.18$ 。藉由公式(9)六月份的修正後的管制線與管制界限列於表 3。

從表 3 可以看出不良率在六月 14 到 20 日期間不良率超過管制上限，我們將這筆不良率資料剔除並重新計算試行的管制中心，得到 $\bar{p} = 0.55$ ，修正後的中心線與管制界限列於表 4。不良率有上升是服務水準惡化的信號，網站管理者就需要關心注意，反之不良率平均下降表示服務品質改善，因此本研究將管制上限表示為實線，代表週不良率超過管制上限為一種品質警訊，而管制下限表示為虛線代表週不良率超過管制下限為一種可能的品質改良。圖 3 則是六月份剔除第二週(六月 14 到 20 日期間)的修正後不良率管制圖，所有的不良率都在管制界限中。在 $e1 = 0.13$ 和 $e2 = 0.18$ ， $\bar{p} = 0.55$ 與 3 週平均樣本大小 $n = 159$ 下，我們得到 $CL_{pe} = 0.51$ ，補償上下管制界線分別為 $UCL_{pe} = 0.59$ 和 $LCL_{pe} = 0.43$ 。

我們使用這些試用標準在七月的不良率管制圖上，來檢測試用標準是否穩定以及能否被使用在下個月。圖 4 為七月的不良率管制圖。所有的資料都在管制界限中，因此我們使用六月的 3 筆資料及七月的 4 筆資料重新計算 $\bar{p} = 0.61$ 及七週平均樣本大小 $n = 178$ ， $UCL = 0.72$ 和 $LCL = 0.5$ ，在 $\bar{e1} = 0.12$ 和 $\bar{e2} = 0.2$ 下，從公式(9)得到修正後補償管制標準分別為 $CL_{pe} = 0.53$ 、 $UCL_{pe} = 0.61$ 及 $LCL_{pe} = 0.46$ 。

我們使用這些修正後的標準在八月到十二月的不良率管制圖中，如圖 5。從圖 5 可以觀測出有 5 個點在管制上限外，其中特別從 10 月 15 日到 11 月 15 日有 3 個點是連續的，有很高的可能性顯示非隨機的誤差發生在這個時期。我們聯繫線上公司，公司回報說他們在這個時期更新了一個新的郵件認證政策，因此一些會員在這個新的機制下沒有有效的密碼，這是造成抱怨增加的主要原因，而我們的管制圖在此時也反映出問題，其它的 2 個點則可能是不正確的警報。

九月份後除了 3 個點超過管制上限外，大部分的不良率都低於管制中心線，我們認為可能有兩個原因。第一網站在此時已經改進了一些問題，因此如果沒有新的服務發佈的話，不良率也隨之下降。第二我們沒有更新關鍵字資料庫，這也可能造成一些抱怨或是問題不能被分類器辨識出來。從我們的實驗資料指出，從六月到九月的不良率有很多在管制圖上呈現對稱，我們假設關鍵字資料庫在網站系統運作機制微幅變動狀況下，每四個月需要更新一次；但是當運作機制有做大幅度的修正時，則需要利用 WebQC 架構重新建構新的管制圖。

表 2 不良率管制圖訓練資料

| 週次 | 留言數量 | 人工判定 不良(率) | SVM 判定 不良(率) | e1 | e2 | 檢定力 |
|-----------|------|---------------|-----------------|------|------|-----|
| 5.3~5.9 | 192 | 64 (0.33) | 73 (0.38) | 0.14 | 0.25 | 84% |
| 5.10~5.16 | 209 | 86 (0.41) | 91 (0.44) | 0.13 | 0.19 | 85% |
| 5.17~5.23 | 155 | 86 (0.55) | 80 (0.52) | 0.14 | 0.13 | 82% |
| 5.24~5.30 | 190 | 79 (0.42) | 74 (0.39) | 0.12 | 0.19 | 81% |
| 5.31~6.6 | 121 | 57 (0.47) | 54 (0.45) | 0.1 | 0.13 | 83% |
| 6.7~6.13 | 75 | 45 (0.60) | 37 (0.49) | 0.06 | 0.23 | 82% |
| 6.14~6.21 | 151 | 94 (0.62) | 95 (0.63) | 0.22 | 0.13 | 78% |
| 6.21~6.27 | 184 | 121 (0.66) | 107 (0.58) | 0.11 | 0.18 | 83% |
| 6.28~7.4 | 219 | 98 (0.45) | 95 (0.43) | 0.14 | 0.22 | 81% |

表 3 訓練資料管制圖中心線與管制界限

| 週次 | 管制 上限 | 中心線 | 管制 下限 | 補償管制 上限 | 補償管制 中心線 | 補償管制 下限 |
|-----------|----------|------|----------|------------|-------------|------------|
| 6.7~6.13 | 0.74 | 0.57 | 0.40 | 0.64 | 0.52 | 0.40 |
| 6.14~6.20 | 0.69 | 0.57 | 0.45 | 0.60 | 0.52 | 0.44 |
| 6.21~6.27 | 0.68 | 0.57 | 0.46 | 0.60 | 0.52 | 0.45 |
| 6.28~7.4 | 0.67 | 0.57 | 0.47 | 0.59 | 0.52 | 0.45 |

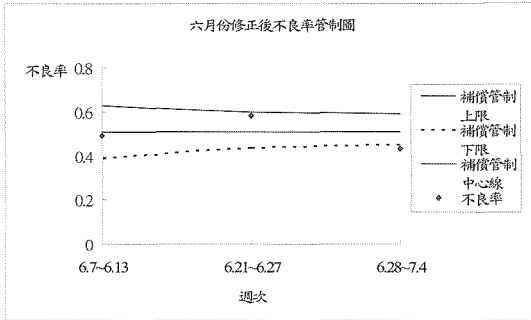


圖 3 六月份修正後不良率管制圖

表 4 六月份管制圖修正界限

| 週次 | SVM 分類後之不良率 | 補償管制上限 | 補償管制中心線 | 補償管制下限 |
|-----------|-------------|--------|---------|--------|
| 6.7~6.13 | 0.49 | 0.63 | 0.51 | 0.39 |
| 6.21~6.27 | 0.58 | 0.58 | 0.51 | 0.43 |
| 6.28~7.4 | 0.43 | 0.58 | 0.51 | 0.44 |

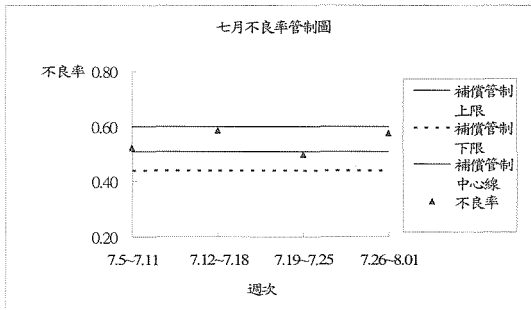


圖 4 七月不良率管制圖

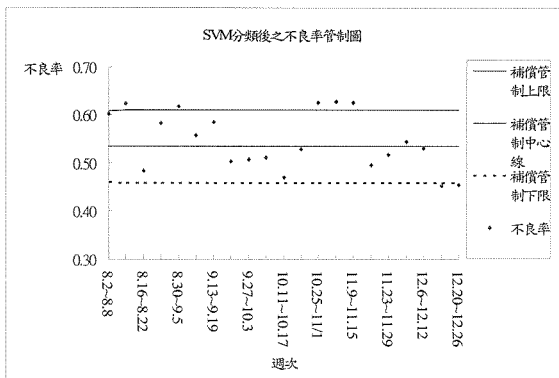


圖 5 SVM 分類後之不良率管制圖

5. 結論

現在網站管理是需要相當的資訊技術，目前網站管理者缺乏工具來了解網站的執行情況，管理者若以人工方式來處理客服留言，排除因疲勞產生的誤差，將會比目前 SVM 分類機制有更高的準確度，但此為一理想狀態，因為人有疲勞因素，在加上網站的客服量非常大，若不以自動機制作為解決方案，其人工成本將會是經營上極大的負擔。本研究試著解決以上的問題，我們應用文件探勘的技術來自動分類顧客的留言，在實驗資料方面，我們建立一個包含 224 個關鍵詞的字庫，使用 SVM 分類的正確率超過 80%，分類管制機制可減少客服部門的人力，且可以更快速的回應而提升服務品質。經過留言分類後，我們使用不良率管制圖來控制服務品質水準，經實證研究，不良率管制圖可以明確呈現出品質資訊且及時發覺異常的狀況。沒有這個管制圖機制，管理者很難去了解網站上的品質是什麼情況，全部的資訊都在技術人員之下零散地散佈著。本篇研究主要的研究目標和貢獻是提供一簡易判讀的自動化機制 WebQC，讓線上公司的管理者能即時了解網站服務品質資訊，使得當網站服務品質下降時，可以及時進行修正與挽救。

6. 參考文獻

- 吳志鴻 (民 90), 應用關鍵頁搜尋及知識分類技術於 Q&A 系統之研究與設計, 元智大學資訊管理研究所碩士論文。
- 賴育昇 (民 91), 自然語言處理於網際網路常用問答集檢索之研究, 國立成功大學資訊工程學系博士論文。
- Berger, A., Caruana, R., Cohn, D., Freitag, D. and Mittal, V. (2000), "Bridging the lexical chasm: statistical approaches to answer-finding," in *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 192-199.
- Burke, R., Hammond, K. and Kulyukin, V. (1997), "Natural language processing in the FAQ finder system: Result and prospects," *1997 AAAI Spring Symposium Technical Report*, 10, 97.
- Case, K. E. (1980), "The p Control Chart under Inspection Error," *Journal of Quality Technology*, 12(1), 1-9.
- Case, Y. and Zeng, M. (1991), "Development of an Automated Indexing System Based on Chinese Words Segmentation (CWSAIS) and its Application," *Journal of Information Science*, 10, 352-367 (in Chinese).
- CKIP, Autotag (2006), <http://rocling.iis.sinica.edu.tw/CKIP/>, Academia Sinica, Taiwan.
- Cortes, C. and Vapnik V. (1995), "Support-Vector Networks," *Machine Learning*, 20(3), 273 – 297.
- Fan, C. K. and Tsai, W. H. (1988), "Automatic word identification in Chinese sentences by the relaxation technique," *Computer Processing of Chinese and Oriental Languages*, 4(1), 35-56 (in Chinese).
- Fletcher, R. (1987), *Practical Methods of Optimization* (2nd ed.), New Jersey: John Wiley & Sons.
- Hammond, K., Burke, R., Martin, O. and Lytinen, S. (1995), "FAQ finder: A case-based approach to knowledge navigation," in *Proceedings of the 11th Conference on Artificial Intelligence for Application*, 80-86.

- Hirschman, L. and Gaizauskas, R. (2001), "Natural Language Question Answering: the view from here," *Natural Language Engineering*, 7(4), 275-300, Cambridge University Press, United Kingdom
- Joachims, T. (1998), "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," in *Proceedings 10th European Conference on Machine Learning (ECML)*, 1398, 137-142.
- Lee, C. and Yang, H. (2005), "A Classifier-based Text Mining Approach for Evaluating Semantic Relatedness Using Support Vector Machines," in *Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05)*, Las Vegas, Nevada, USA.
- Nie, J., Briscois, M. and Ren, X. (1996), "On Chinese Text Retrieval," in *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, Zurich, Switzerland.
- Salton, G. and McGill, M. (1983), *Introduction to Modern Information Retrieval*, New York, NY: McGraw-Hill.
- Sebastiani, F. (2002), "Machine Learning in Automated Text Categorization," *ACM Computing Surveys*, 34(1), 1-47.
- Sproat, R. and Shih C. (1990), "A Statistical Method for Finding Word Boundaries in Chinese Text," *Computer Processing of Chinese and Oriental Languages*, 4, 336-351.