

商業智慧系統之建置

吳昌憲^{1*} 羅元皓²

¹國立虎尾科技大學資訊管理所助理教授

國立虎尾科技大學資訊管理所

²國立虎尾科技大學資訊管理所研究生

國立虎尾科技大學資訊管理所

*cswu@nfu.edu.tw

摘要

資料量暴增的時代中如何有效的利用現有的資訊是成功的關鍵因素，利用資料探勘在雜亂無章的資訊中去取得有效的資訊是重點，而操作介面的部分是不是能夠讓一般的使用者所接受，都能夠快速的了解並上手是本研究的主要重點。本研究以 MATLAB 做為開發的平台，利用 GUIDE 工具設計簡單的操作介面，再結合資料探勘的分析功能，最後再將分析出的數據與現有的線上系統所分析出的數據做比較，看是否能夠做到簡易操作並且達到數據的準確性，與其他大型分析系統媲美。

關鍵字：資料探勘、MATLAB、商業智慧

一、緒論

1.1 研究背景與動機

時代在改變的同時技術也在跟著改變，為了應付現在資訊化的時代，處事情的方式也在轉變，身處在現今資訊化的時代中，如何去善加處理雜亂無章的資訊是一大困難點，往往上司主管或是一般工作者面對龐大的資訊量，常常是心有餘而力不足想用卻不知道該怎麼去使用，而更不用說可以幫助到他們解決問題，或是拿到什麼可以擬定策略的資訊。而企業在建置 ERP 系統後，又競相導入電子商務的情況下，這些交易資料及在網站上所收集到的交易與客戶資料，如果沒有系統化的處理保存，那就只是一堆堆雜亂無章的資料，但是如果能夠充分整合自身的資料，以獲得所需要的資訊，那對公司來說將會是一份很重要的資產。

商業智慧也就是順著這個資訊大爆炸的環境下而衍生出來的，但是商業智慧這個名詞也是近年來才慢慢出現並且被大家所接受，它是隨著大家的需求有所不同而改變，經過一連串的演進過程而來的，造就了現在商業智慧的技術，現在不僅僅可以協助企業統計、挖掘與分析隱含在數據資料背後的知識，將相關數據資料轉化為有助於企業決策的有用知識，以往需要許多的人力去分析和計算，但是透過商業智慧的系統化整理，只需要花費些許的時間就可以拿到所需要的數據，而決策者就可以有多餘的時間可以依據得到的數據去思考下一步的決策制定。

1.2 研究目的

商業智慧近幾年來在各企業界的使用愈趨流行，雖然每個產業對於商業智慧的應用需求各不相同，但是絕對都是抱著一樣的目的和想法來使用，讓操作的使用者能夠適時的獲得適當的資訊，對公司做出最為有利的決策；這份研究主要是挑出資料探勘演算法的部分以及設計操作介面來跟現有的系統做比較，比較出之間的差異性。其主要目的有二：

1. 利用 MATLAB 設計一個能讓使用者容易理解上手的操作介面。
2. 資料探勘的準確率。

這麼做的目的無非就是為了要讓一般使用者都能夠輕易使用，並且能夠正確的拿到所需要的分析數據，不必擔心到自己是否學過程式碼或是會不會輸入指令這方面的問題。

二、文獻探討

2.1 商業智慧的定義

美國Gartner公司的分析師 Howard Dresner 在1989 提出了”商業智慧”(Business Intelligence, BI)這個名詞。根據1996年Greene為商業智慧所下的定義：「商業智慧是將與企業運作有關的重要性資訊蒐集整理，用來管理目前和將來的商業運作環境的一個過程」。也就是說以使用實際資料為基礎的系統，來增進企業決策制定的概念和方法。而BI的定義隨著時間的演化日趨完整，許多學者對商業智慧也給予不同的定義，根據文獻加以整理彙整如表一

表1 商業智慧定義整理表

學者	定義
陳界銘(1998)	把資料轉換成知識的一種過程。
White C.J(1999)	以一組技術及產品來提供使用者解決商業問題所需的資訊，以支援戰略性和策略性之商業決策。
Michael.J. Berry and Gordon S. Linoff (2000)	結合先進資訊與企業專業知識，以提高企業競爭力為最終目標。
Intelix Inc.(2001)	一種可以將資料轉換為具有意義的資訊的能力，換言之，就是提供使用者更多與企業相關的深度資訊、做為判斷未來走向的能力
Jonathan Wu(2002)	利用線上分析處理(Online Analytical Processing, OLAP)等工具，讓使用者能回答商業問題、並看出資訊所顯現出的重大趨勢或模式的系統。
童啟晟(2002)	提供業務經理做成重大決策的一種情報依據(提供具體可見的報表來提醒決策者)，同時讓業務策略及流程可做適時地修正或調整，來增加企業的競爭優勢並改善業務的營運與企業利潤的獲得。
蔡正仁(2002)	將企業所累積的龐大資料加以整理分析，使決策者能夠快速、容易利用資料，進而獲得有效、即時的決策參考。
吳姝蓀(2002)	利用資訊科技，將現今分散存在於企業內、外部各種結構化資料

	彙整，並依據某些特定的需求進行分析與運算，再以最適的方式，將這些結果呈現給決策者、管理者或是知識工作者等，以協助這些組織角色在管理組織績效或是決策判斷時的重要參考
樂斌(2002)	商業智慧乃是將企業內各種的資料轉換為有意義的資訊，用以提供企業了解現況或是未來展望，更能讓企業快速掌握關鍵商機，將不同平台的異質性資料，透過智慧型的轉換分析，產出結構化知識的整合互動式分析工具，以利企業內部決策、判斷、分析的依據基礎，使企業改善決策制訂的方法與過程。
楊志強(2003)	商業智慧是對資料倉儲(Data Warehouse)的龐大資料進行資料採礦(Data Mining)，藉由存取產生報表分析，最後提供企業決策模式，也因此BI 先要有龐大資料的基礎，需要ERP、CRM 等商業應用系統的成熟。
David Loshin(2003)	流程、技術和工具需要將資料轉換成資訊，資訊轉換成知識，再將知識轉換成可獲得公司效益的行動計畫。商業智慧包含資料倉儲、企業分析工具、內容/知識管理。
王茁(2005)	利用現代資訊技術收集、管理和分析結構化和非結構化的商務資料和資訊，創造和累積商務知識和見解，改善商務決策品質，採取有效的商務行動，完善各種商務流程，提升各方面商務績效，增強綜合競爭力的智慧和能力。
林東清(2005)	企業利用快速、即時、整合的資訊科技，來蒐集、分析、企業外部環境的競爭資訊與內部經營的重要關鍵指標，來提供即時、多維度的資訊，以支援決策者的判斷，提昇企業競爭能力的一種流程與資訊系統。
Chaudhuri(2006)	將BI 視為一種策略，其流程、技術和工具需要將資料轉換成可獲得公司效益的行動計畫。
Anonymous(2006)	當正確地設計與建置BI 系統時，它可以為不同層級的人及跨部門的人提供必要的分析並製定決策。BI 系統並非一項技術工具，而是可以幫忙非技術人員做明智的決定。

總結以上各位專家學者對商業智慧的定義整理可知，商業智慧就是結合有效的資訊與企業專業知識，將資料庫裡的資料做整合，把不同來源、不同型態的資料經過萃取、轉換、載入，把資料轉換成知識的一種過程，使決策者能夠快速、容易的利用資料，進而獲得有效、即時的決策參考。

2.2 商業智慧的演進

商業智慧的出現是一個漸進的、複雜的演變過程，它經歷事務處理系統(TPS)、高級管理人員資訊系統(EIS)、管理資訊系統(MIS)和決策支援系統(DSS)等階段，最終演變至今天的商業智慧。它是一個可包含企業所有知識的系統，服務管理決策層或部門經理，幫助其進行分析和決策。商業智慧（Business Intelligence，BI）的概念出現之後，很多人經常會搞不清楚商業智慧的範疇究竟包含哪些部

分，其實商業智慧是統整了「管理」、「決策」及「資訊科技」等三項要素的有效分析機制（商業經理人手冊，2002）。

因為現今資訊科技與網際網路的興起，使用到商業智慧的機會越來越多，範圍也越來越廣，不管是業界常常使用到的顧客關係管理，還是供應鏈管理甚至企業資源規劃，都是商業智慧在實務上的應用。為了要使得企業中的決策人員能即時地取得正確及其所需要的資料，商業智慧的作業性層面工具可以說是商業智慧最重要的核心，這些工具包含了資料倉儲(Data Warehouse)，資料超市(Data Mart)、線上即時分析(OLAP)、線上交易系統(OLTP)、資料探勘(Data Mining)等，商業智慧完整架構如圖1，而本研究主要是要針對資料探勘的部分提出應用和討論。

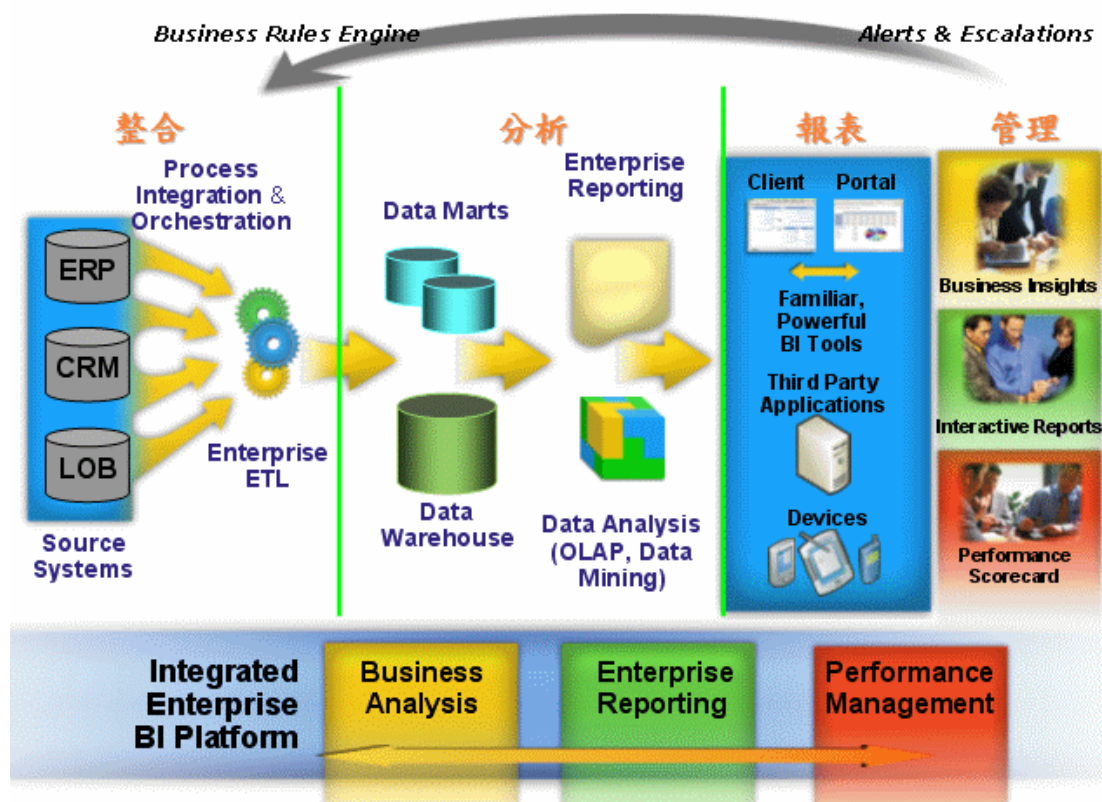


圖1 商慧智慧架構圖

資料來源：李玉秀、莫介中，2006

2.3 資料探勘

2.3.1 資料探勘的定義

資料探勘最早是由 Usama Fayyad 所提出。面對資訊爆炸時代的來臨，電子化資料也愈來愈多，累積愈來愈多的資料就造成了龐大的資料庫。隨著時間的累積，過多的資訊使的企業決策單位無法有效的利用，甚至產生混淆與誤用。所以從資料庫中利用新的技術及工具，結合智慧化及自動化處理資料，轉換成有用的資訊及知識，愈顯重要(Chen et al.1996)。從大量資料中，藉由各種不同的資料分析方式來尋找事前未知、有趣且可以付諸行動的規則或知識(Burton and Kleer,1982 ; Berry and Linoff,1997 ; Fayyad and

Smyth,1996；Fayyad and Uthurusamy,1996)。經過資料探勘後，所挖掘到的知識，其主要目的為提供豐富及前所未知之資訊，做為使用者知識發現及決策支援之用(Olaru and Wehenkel，1999；Fu，1997)。

2.3.2 資料探勘的功能

Michael J.A.Berry&Gordon S.Linoff 認為，Data mining 的功能包含了六大功能:分類、推估、預測、關聯分組或同質分組、群集化、描述及視覺化。分述如下：

1. 分類：
依分析對象的屬性分門別類並加以定義，建立類組。如將保單貸款者的風險屬性，區分為高、中、低度風險申請者。使用的技巧有決策樹，記憶基礎推理等。
2. 推估：
分類的結果仍是不連續的，而推估所得的結果則是連續性的數值。如按照信用申請者知教育程度、行為別來推估其信用卡消費量。使用的技巧包括統計方法上之相關分析、迴歸分析以及類神經網路方法。
3. 預測：
根據對象屬性之過去觀察值來推估該屬性之未來值。例如由顧客過去之刷卡消費量預測其未來之刷卡消費量。使用技巧包括迴歸分析、時間數列分析與類神經網路方法。
4. 關聯分組：
從所有物件決定哪些相關物件應該放在一起。例如超市中相關之盥洗用具(牙刷、牙膏、牙線)，放在同一間貨架上。在顧客行銷系統上，用來確認交叉銷售的機會以設計出吸引人的產品群組
5. 群集化：
將異質母體中區隔為較具同質性之群組，相當於行銷術語中的區隔化。群集化事先並未對於區隔加以定義，使資料自然產生區隔。使用的技巧包括 K-means 法及 agglomeration 法。
6. 描述及視覺化：
資料探勘之目的有時只是單純的想利用更容易瞭解的方式，來描述一個複雜的資料庫中所存在的狀況。對一種尚未被認識的狀態而言，一個正確的描述足以啟發許多對該狀態的認識與解釋，而資料視覺化即為一種強而有力的描述方式，透過人類所擅長讀取的圖象表示方法來呈現，比一千條列表的規則更有價值。

2.3.3 資料探勘的方式

引述 Y.Fu 於 1997 年將資料探勘的技術依不同研究領域區分說明：

1. 統計方法：
資料探勘使用許多統計工具，包括 Bayesian 網路、迴歸分析、相關分析及分群分析，通常統計模式是經由訓練資料集來建立，然後從模式中尋求規則及特徵。
2. 機器學習方法：
就像統計方法一樣，機器學習方法尋求一個最佳模式來符合測試資料。不同之處在於大多數機器學習方法利用資料自動化學習過程，自動歸納出分類規則及建立模組。資料探勘最常使用的機器學習方法包括類神經網路以及決策樹歸納。
3. 資料庫導向方法：

資料庫導向方法不像前兩個技術在尋找最佳化模式，轉而著重在處理現有的資料是一種屬性導向歸納，反覆從大量的資料中找出共同的規則及模式。

4. 視覺探索：

將多維度的資料轉換成視覺化物件，如點、線及區域，讓使用者可以動態檢視及探索有興趣的部分，進而分析出資料的模式。

5. 其他方法：

如類神經網路運用於資料的分類及預測，約略集合運用於分類及分群這些技術可以互相整合來處理複雜的問題及提供替代方案，如統計方法可以結合視覺探索技術，用以輔助系統處理資料模式及趨勢分析。

表 2 常見的資料探勘方法整理

方法	說明	技術
歸納(分類)法	將相同屬性的資料歸納並以一個類別標籤代表該族群，將這些族群透過訓練資料來導出一些規則，這些規則可以被用來找出目標屬性應有的值。	線性迴歸、決策樹、基因演算法、K-鄰近法、類神經網路、模糊邏輯
族群推算(集群)	歸納法是已知要找的資料屬性為何，族群推算是將類似的資料放在同一個族群，一個族群內的資料很相似，不同族群之間的資料相似度就很低	相似度演算法、K-Means、模糊邏輯、基因演算法
關聯法	找出哪些屬性是互相關聯，通常以「if A, then B」的邏輯方式展現，已就是要找出某一事件中會同時出現的東西	案例庫推理、集合理論、統計方法、Priori 法、GRI 法
順序法	與關聯法很像，但不同的是會以「時間」來區隔	線性迴歸、多重迴歸、類神經網路、決策樹、鄰近法
迴歸分析	在歷史趨勢和統計基礎上，用已知的資料預測未來或事件的趨勢的方法	簡單線性迴歸分析、複迴歸
時間序列	只依賴時間的資料值進行預測	類神經網路、統計方法

資料來源：梁定澎, 2006 & 李卓翰, 2006 & 王茁, 2005

2.4 MATLAB

MATLAB 這個名詞是由 MATrix(矩陣)和 LABora-tory(實驗室)這兩個單字的前三個字母所組合而成的，它是基於矩陣運算的軟體。雖然它是以一種科學軟體的面貌出現，但它更像一種語言，透過工程人員比較容易理解和學習的方式，藉助積木般的構建和解決問題的方式，透過工具箱(Toolbox)化簡與解決目前工程和科學界重要的問題。MATLAB 有自己專屬的程式語言，而且本身提供了一個支援豐富的預設函式庫，使得程式設計的工作變得更為簡單，而且更有效率。相較於科學計算所用的傳統語言，MATLAB 擁有許多優勢，包括：

1. 容易使用：

MATLAB 是一種直譯式語言，可以像書寫便條紙般，在命令列輸入一段敘述式來計算結果，也可以執行事先寫好的大型程式。可以容易的書寫並修改程式，並使用偵錯器進行偵錯。MATLAB 也提供

許多程式開發工具，像整合性的文字編輯器/偵錯器，線上說明文件及使用手冊，工作區瀏覽器以及大量的程式使用範例等。

2. 平台獨立性：

能支援多種作業系統，在任何平台撰寫的 **MATLAB** 程式都可以在別的平台執行，連輸出的檔案也不例外可以直接讀取，所以如想要移到新的平台上開發是不被影響的。

3. 預設函式：

本身具備大量的函式庫，足以提供解決一些基本的問題。另一個特點是還提供了有特別功能的工具箱(**ToolBoxes**)，來幫助使用者解決在特定領域中較為複雜的問題。

4. 與裝置無關的繪圖：

MATLAB 擁有許多完整繪圖及影像處理的指令，這些繪圖及影像可以直接顯示在任何執行 **MATLAB** 電腦的圖形輸出裝置上，這個功能讓 **MATLAB** 成為一個非常突出的視覺化資料處理工具。

5. 使用者圖形介面：

有了這項功能，程式設計者可以建立互動式的使用者圖型介面，提供給一些比較沒有程式寫作經驗的人使用。本研究在這個部分特別針對那些沒有經驗的使用者作討論以及設計。

6. **MATLAB** 編譯器：

一個獨立的 **MATLAB** 編譯器，可以將 **MATLAB** 程式轉換成可以在沒有安裝 **MATLAB** 的電腦上執行的程式，這樣也就可以普及到任何一台電腦上執行了。

2.4.1 GUI

圖型使用者介面(**Graphic User Interface, GUI**)是程式語言必備工具。主要目的是在提供使用者一個互動的平台，藉著簡易的操作便能執行該軟體所提供的所有功能。有別於在命令視窗進行指令式的操作，使用門檻因此降低不少，一般稱之為友善的使用者介面(**Friendly User Interface**)。給使用者方便，其實是給創作者一些額外的工作，在原來的程式之外，必須考慮使用者的習性與背景、操作動線等，與程式執行不直接相關的細節，再透過另一組指令完成這件工作。

使用 **GUIDE** 工具的優點如下：

1. 可以產生每一個控制物件的標準反應函式，使用者只要填入所需的指令或敘述。
2. 可以快速地「拉」出介面，具有「所見即所得」的性質。

使用 **GUIDE** 工具的缺點如下：

1. 比較不適用於複雜的 GUI 介面。
2. 無法將 GUI 應用程式集中於一個檔案。
3. 執行效率可能比較差一些。

三、研究方法

一開始主要是針對相關文獻以及未來要測試的資料數據作查找與收集的動作，再來是針對收集的的文獻作深入的探討像是演算法的類型和分類，挑選出適合的演算法來作研究的基礎，以及構想什麼樣的介面比較適合一般使用者的思維，當全都決定好的時候最後再將兩者做結合，最後再對分析出來的數據跟線上現有的大型系統做比較，做最後的檢討與討論。其流程彙整如圖 2

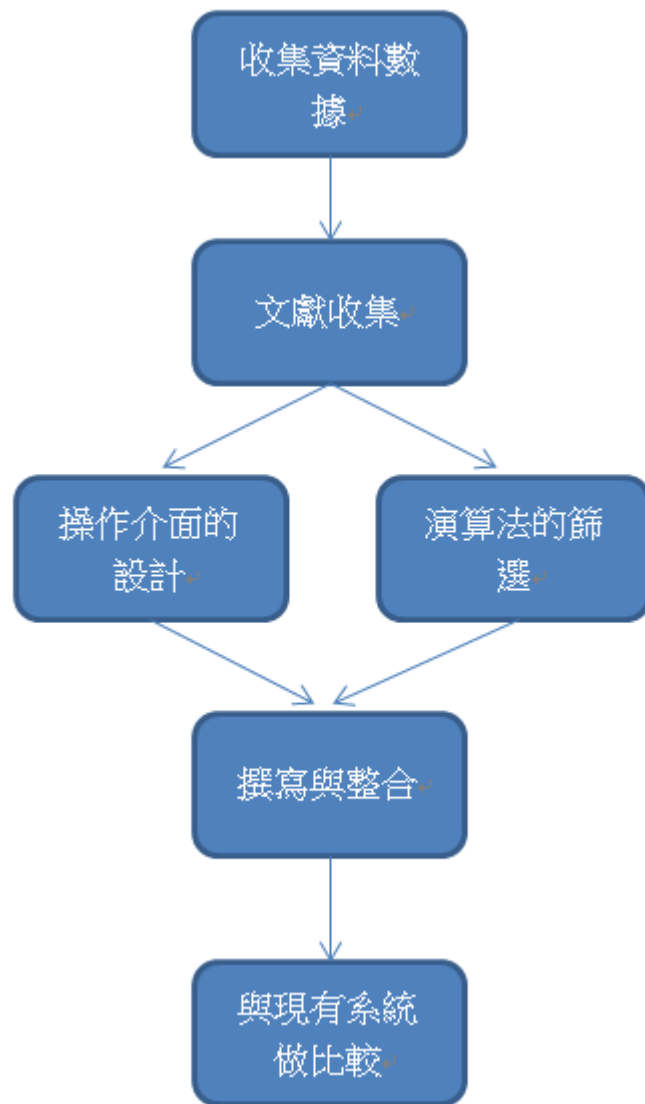


圖 2 研究流程圖

3.1 演算法的篩選

Data mining 對資料的探討和分析，擷取出有用的資訊這點，對於一般企業來說可以改善他們的行銷、銷售等等，這些迷人的吸引力使得 Data mining 在電子商務的環境中成為大家所喜愛分析技術，而 Microsoft 當年則藉由 SQL Server 2000 Analysis Services 把 Data mining 結合其中，而當時唯一導入的只有兩個演算法，分別為決策樹(decision tree) 群集(clustering)，雖然不敢完全說是因為導入了 Data mining 而遠遠的超越了競爭對手，但是我想這個部分對於當時的市場情況來說，一定佔了不少的比重。這兩個演算法剛好也屬於不同的兩個類型分別為分類和群集，所以本研究決定就以決策樹以及群集這兩個演算法來實作。

3.2 操作介面的設計

期望最後能利用 Matlab 裡的 GUIDE 工具設計出以點、選的方式來操作，免去了輸入指令的步驟，

既可以選擇想要的資料檔案，又可以選擇想要使用的演算法，畢竟不是所有的使用者一開始就都能理解指令的操作，如果想要貼近一般使用者的想法，一目了然的介面是最符合一般大眾的了，如果操作順利也能真正達到方便又省時的效果。另外分析出來的數據準確率，希望能夠拿來跟現有比較大型的分析系統來做比較後，跟比較完善的系統比起來不會相差太多，甚至如果準確率比他們更高會更好。

以下圖 3 為操作介面的構想圖，圖 3 中的開啟檔案功能按鈕為的就希望像我們平時打開檔案時一樣用點選的方式開啟如圖 4 所示，而圖 3 中的圖表類型按鈕功能為，依使用者的需要選取想要數據分析出來後以什麼樣的圖案來呈現如圖 5，圖 3 中的開始分析按鈕當使用者點選完自己想用的選項時，按下此按鈕就直接開始分析載入的數據，而畫面就會呈現在中間視窗中，圖 3 中有幾個 Check Box 則是設定為一些修改的小功能，例如修改圖表中的一些背景格線方面的問題，讓圖表更容易清楚劃分出界線。

最後還是希望所呈現出來的操作界面和功能，可以達到研究的主旨「簡單操作、快速上手、數據準確」，順利的解決一般使用者在操作上所遇到的困難，而在分析數據方面，希望在數字上的準確性可以和其他系統比起來毫不遜色。

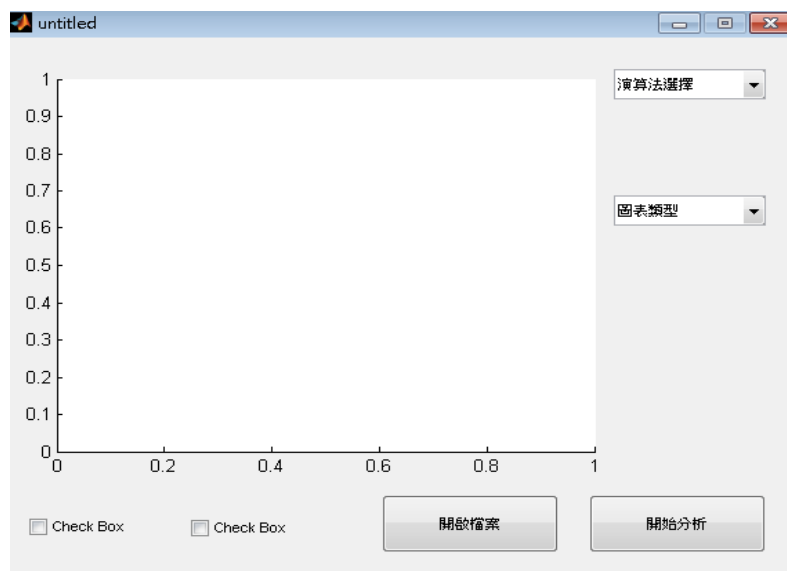


圖 3 操作介面構想圖

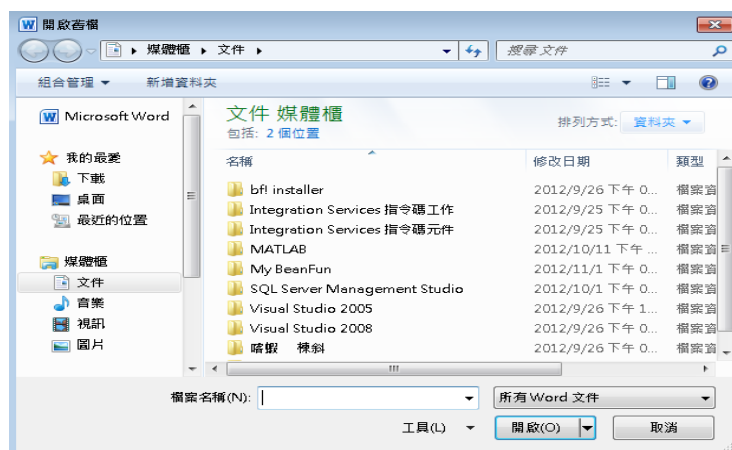


圖 4 開啟檔案畫面

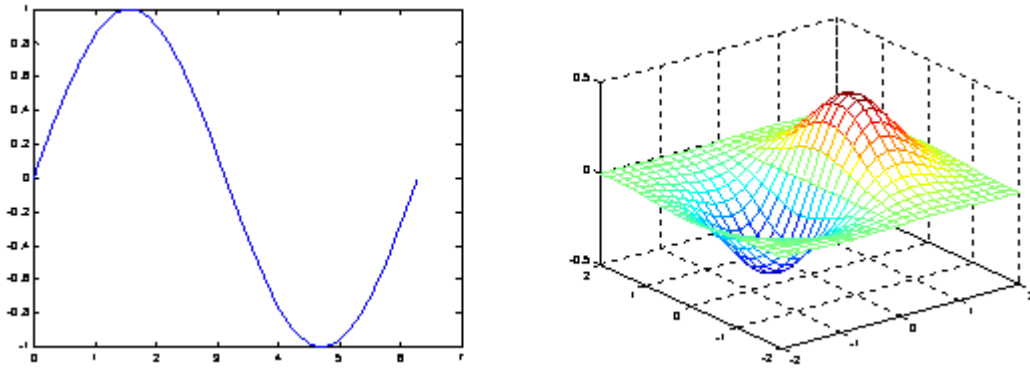


圖 5 圖表類型

四、預期成果

將建置的 MATLAB 系統來對商業數據進行探勘分析與圖表展現，期望和現有線上 SQL Server 系統取得一致性的決策結論。

參考文獻

1. 王茁：商業智慧。初版。台北：博碩文化股份有限公司，2005；1-9。
2. 王偉如：淺談商業智慧。資策會教育訓練處，2010.01.31
自<http://www.iiiedu.org.tw/professional/920171.htm>
3. 吳姝蓓，“商業智慧為企業所建立的五項新價值”，eBusiness Executive Report 電子化企業經理人報告，2002 年四月，第32 期。
4. 郭進聖，商業智慧系統建置-以製造業某上市公司為例，碩士論文，國立台灣大學管理學院會計與管理決策組，2006。
5. 郭姿君，MATLAB程式設計與應用，滄海書局，98年11月出版印刷。
6. 童啟晟，2001，商業智慧的定義、型態、工具、軟體與發展趨勢，電子化企業經理人報告，十一月，38~42。
7. 楊適綺，顧客知識管理應用於零售產業之探索性研究，碩士論文，國立成功大學資訊管理研究所。
8. 劉德泰：淺談商業智慧。資策會數位教育研究所。
http://www.iiiedu.org.tw/knowledge/knowledge20021130_2.htm
9. 劉宜妝，資料採礦之應用研究－台灣地區漁市場行情資料庫之關聯法則分析，碩士論文初稿，國立中興大學行銷學系，2002。
10. 蔡淑賢，商業智慧(BI)讓企業變聰明了，經濟部商業現代化雙月刊，2002 年 2 月，第 50 期。
11. 蕭乃綺，應用資料採礦技術於我國壽險業顧客價值分析之研究，碩士論文，東海大學工業工程與經營資訊研究所，2006。
12. 電子化企業經理人報告，商業智慧經理人手冊，ARC 遠擎管理顧問公司，2002年四月，第32期。

13. Berry, M.J.A., & G.S. Linoff, Data Mining Techniques: for Marketing, Sales, and Customer Support., New York: Wiley Computer, 1997.
14. Harding W. BI crucial to making the right decision. Financial Executive 2003 ; 19(2) ; 49-50.
15. Fu, Y. "Data mining task, technique and applications," IEEE POTENTIALS, October/November 1997.
16. Stephen J. Chapman (2008), MATLAB Programming for Engineers, 4e