

<<從資料中挖金礦>>

图书基本信息

书名：<<從資料中挖金礦>>

13位ISBN编号：9789867889942

10位ISBN编号：9867889940

出版时间：2010-3-5

出版时间：經濟新潮社

作者：[日] 岡嶋裕史

译者：李弘元

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

<<從資料中挖金礦>>

前言

第二章商業上使用的資料採礦 資料倉儲的建構 凡是能夠成為話題、普及的技術，一定都有需求存在。

但如果工具或分析技巧只是技術優越、想法有趣，沒有人想去使用，熱度則會消退。

資料採礦蓬勃發展至今已經超過十年的歲月，一定存在著相當的期待和需求，它究竟被使用在什麼用途上呢？

但資料採礦主要的需求來自以CRM（客戶關係管理，也就是企業應用資訊系統，架構與客戶之間長期關係的方法）為主的業務活動、販賣促銷活動。

資料採礦並不是被侷限於特定領域的分析系統，稍後將介紹它被應用到其他領域的狀況。

一般聽到資料採礦，首先應該會想到POS系統（店員在銷售商品的同時，蒐集顧客的性別、年齡、商品名稱、數量、金額等資料，分析運用到經營管理層面的系統）這樣的東西吧。

日本7-11最先採用的POS系統，劃時代以資料採礦為導向。

從一開始導入，就不把POS只當成是高功能的端末，而將其視為統合型的介面（資料往來的媒介機器）。

這個策略的目的是，藉由嚴格區分暢銷商品和滯銷商品，設法消除不良庫存和缺貨狀況；而為了讓不同立地條件和客層的店舖，都能夠執行有效的分析，因此建構了蒐集項目極多的「資料倉儲」

（data warehouse）。

藉由網路連結總部、各店舖、資料倉儲，使得商品的管理可以細分到單品的程度。

POS系統執行的工作當中，金額的計算只佔了很小的一部分。

如同圖2-1所示，店員將正在購物的客戶資料：大約幾歲？

性別？

有無同伴？

等資料輸入到POS系統，系統則藉由網路持續地將資料傳送、儲存到資料倉儲。

建構資料倉儲是資料採礦的第一步。

執行資料採礦的大前提是必須有可供分析的資料，但並不是只要有資料就可以。

資料必須經過完善的整理，並且維持在隨時可以取用的狀態。

符合這個條件的資料儲存系統就是資料倉儲。

以前有專門負責資料儲存的機構。

事實上儲存本身並沒有意義。

是否處於能被馬上利用的待機狀態，才是影響分析性高低的主因。

缺乏策略、雜亂儲存的資訊，無法成為有用的資料，大部分的情形形同積壓的垃圾，假設要被使用也必須重新投入相當的成本。

當今隨著Web 2.0的普及，雜亂繁多的資訊可以很容易地以低成本再度編輯；才不久前如果要實現這樣的事情，就必須在某個時間點事先建構好資料倉儲。

另外，資料倉儲的一大特徵是，可以按照時間順序儲存／展開資料。

之後展開資訊進行分析時，有無時間軸來進行評斷，會對分析的精準度產生莫大的影響。

這是因為觀測連續、持續性資料的傾向，會比只評斷某個時間點的資料，得到更多的訊息。

這在日常生活上也是相同的道理。

調查商圈 說明至此，讀者可能還難以想像資料採礦如何具體被應用。

接下來看看常被提及的範例。

例如，圖2-2是包圍自己商店的商圈預測圖。

按照一般的常識，就是以自己的店舖為圓心，畫一個圓來代表商圈吧。

如果要做宣傳活動，就是在這個預設的商圈裏發宣傳單、強力曝光。

但是經由實際的問卷調查卻發現，大部分的顧客是住在預設的商圈之外（圖2-3）。

這樣的結果讓店家感到困惑。

結果顯示，無論在預設的商圈裏如何地發傳單，效果依然很有限。

<<從資料中挖金礦>>

當然可以依照問卷調查的結果重新設定商圈，改變發送傳單的範圍吧。

如果「知道了實際的商圈」還放任不管，實在說不過去。

為什麼商圈不是集中在一般所想像的範圍內呢？

如果這時擁有地圖、其他店鋪資料，或許就可以執行資料採礦了。

例如，配合這樣的資料一起看（圖2-4）。

首先是競爭對手店鋪的資料。

競爭店鋪會從商圈裏搶走自己的顧客。

這個資料中顯示，原先預設的商圈裏確實有一家人氣商店。

而且這家店接近車站，比自家店擁有更好的立地條件。

這就能解釋為什麼勢力範圍內的顧客會被搶奪了。

但是，光是這個資料還無法說明，預設範圍以外的顧客為什麼願意特地光臨本店？

再仔細看看地圖，就會發現自己的店鄰接著主要幹道。

而經由這條主要幹道，自己的店比競爭店鋪擁有更好的立地位置。

主要幹道再下去有一些新興社區，設有完善的停車場。

如果距離車站遠、擁有車子，那麼開車來購物就變得理所當然。

消費者從社區開車前來購物，則自家店的優勢會大過競爭店鋪。

由於開車的緣故，商圈會比一般所預測的商圈更遠些，然而距離卻不會造成顧客過多的負擔。

這就可以解釋商圈偏移的問題了。

這個假設正確與否需要另外的驗證，但卻提示了一個思考方向。

那就是，如果這個分析是正確的，就必須更改發傳單的地點、擴建停車場以追求來客的穩定。

了解顧客的傾向 再看另外一個珠寶店的例子。

A先生從事高級珠寶的販賣，準備打廣告來促進銷售。

可供廣告的媒體有好幾種選擇，但若考慮到成本與效果，就不能無限制地打廣告。

錢要花在刀口上，如何分配最適當呢？

?電視廣告 ?報紙廣告 ?廣播廣告 ?網路廣告 ?傳單廣告 首先，如果顧客看不到的話就沒有用，所以先進行問卷調查，看看住在商圈裏的人會接觸到怎樣的廣告（圖2-5）。

看電視的人不在少數這是可以預期的。

網際網路的影響力不斷增加讓人印象深刻。

相反地，從報紙、廣播得到訊息的人顯得很少，令人有些遺憾。

最多的是傳單。

的確，塞進信箱裏的東西很難不看一眼。

根據這個結果，A先生的珠寶店是不是採用傳單廣告就可以了呢？

如果只是經由這個資料來判斷，那麼這個決定是正確的。

但是常被看到的不同於常被利用的。

就好像住在東京常常看到東京鐵塔，但不代表常常上去東京鐵塔。

接著假設有這樣的資料又如何呢？

就是詢問消費者相信哪個媒體的廣告（圖2-6）。

結果和先前詢問所接觸廣告的問卷調查，順序上有極大的變化。

A先生的珠寶店，銷售的都是高級珠寶。

消費者對於高價物品的買賣，相對上會採取比較謹慎的態度。

這時候資訊源的可信度將會是重要的判斷依據。

所以無論傳單可以替珠寶店提高多少知名度，如果傳單的可信度很低，那麼對於行銷業務上反而可能造成不好的印象。

這個資料同時也顯示，網路廣告有同樣的問題。

相對於此，傳統的電視、報紙廣告的可信度比較高。

和先前常接觸到廣告的資料一起考慮的話，那麼對於珠寶店而言，電視廣告應該是最有效的吧。

但是如果圖2-7這樣的資料出現了呢？

<<從資料中挖金礦>>

以年齡來分析先前常接觸到的廣告的資料。

這類資料常常容易被忽略。

如果A先生珠寶店的商圈中，住著大量五十歲以上的人口呢？

那麼先前電視廣告最能吸引顧客的調查結果，將產生急遽的動搖。

即便整體廣告收視戶中，電視廣告是最常接觸的、可信度最高的媒體，但如果住在自己店舖商圈中的人們不常看電視廣告的話，效果就會大打折扣。

「所以，報紙比較好嗎？

」、「會不會還有導致其他結果的資料出現呢？

保險起見是不是也花些經費投入網路廣告」等，看來A先生嘗試錯誤的日子還要持續下去。

捨棄奧客 八 : 二 是常被提及的經驗法則。

也就是說一家企業八 %的利潤，是來自僅占二 %的優良顧客。

如果這是正確的，對於企業而言，重要的是如何把握住二 %的優良顧客。

只要些許的投資，常客就會持續成為常客。

比起獲取新顧客，這些常客對企業的貢獻度更高。

另外在顧客投資上，要小心不要過度投入到奧客上。

要從奧客身上獲得一千日圓的利潤，可能需要一萬日圓以上的投資！

看到這裏，企業的確會有把全體顧客劃分為優良顧客、一般顧客、奧客的想法（圖2-8），但是卻無法直接詢問客人：「您是優良顧客嗎？

」 所以需要加以定義，滿足怎樣的條件是優良顧客、怎樣又是奧客。

【優良顧客】 每年消費一千萬圓以上 每年消費十二次以上，而且每次的消費額在五十萬圓以上 【奧客】 每年消費一萬圓以下 抱怨電話每個月超過三十通 每個月在食品賣場的試吃次數超過六十次 滿足以上任一條件，則可定義為優良顧客或奧客。

如此一來，所有的顧客都可以被分為優良顧客、奧客和一般顧客等三類。

根據這個分類，可以對資源進行最適當的分配，提供無微不至的服務給優良顧客，而對於奧客則僅提供最小限度的服務。

只不過這只能算是事後的分析，只是對優良顧客貼上「優良」的標籤。

當然光是這樣也具有意義，不過若能在更早的階段，例如對可能成為優良顧客的客人事先進行行銷活動、誘導可能成為奧客的客人成為優良顧客或是捨棄他們，則可以更有效率地提升利潤。

在此嘗試使用資料採礦追溯優良顧客或奧客的動向。

圖2-9是現有的顧客資料。

或許讀者會想：這家公司究竟在調查顧客什麼啊？

就姑且相信有這樣一份資料吧。

從這份資料可以預測出優良顧客和奧客嗎？

如果漠視的話，就很難掌握到隱藏於資料中的趨勢。

先前定義了優良顧客和奧客，因此如圖2-10一般，就依照「每年消費一千萬圓以上」及「每年消費十二次以上，而且每次消費額在五十萬圓以上」的屬性來作分類。

以此類推，奧客的歸類也是依照之前的定義。

「每年消費一萬圓以下」、「抱怨電話每個月超過三十通」、「每個月在食品賣場的試吃次數超過六十次」。

如此一來就可以相當明確地抓出顧客的特性。

為了更容易理解，可以如圖2-11將其視覺化。

圖表化之後，可以看出明顯的趨勢。

光頭族成為優良顧客的機率很高，而要求沒有金額和日期的發票、或使用假名填寫資料者成為奧客的機率很大。

如果一開始就知道這個趨勢，就可以分別對顧客群採取適當的對應，這對於追求企業利潤有很大的幫助。

當然，現實社會裏的資料採礦，是由更複雜的要素所組成，且適用的領域也不侷限於CRM的範圍

<<從資料中挖金礦>>

。下一章將詳細討論使用到商業領域以外的資料採礦，以及資料採礦的程序。

<<從資料中挖金礦>>

內容概要

為什麼你的網站不賺錢？

！

要了解Google、Amazon網站成功的祕密，就在這裡！

本書以最簡單的方式介紹Data Mining（資料採礦、資料探勘）在資料分析、商業行銷方面的強大威力，也揭露Google、Amazon.com賴以成功的關鍵。

例如搜尋引擎，其中就使用了資料採礦的技巧。

資料採礦的目的，就是在龐大的資料中擷取有用的資訊。

由於儲存資訊的成本已經降到接近0，資料採礦的可行性也大幅增加，除了可用在商業行銷，也可運用於社會，甚至預測未來。

本書用許多日常生活的例子，來解釋資料採礦的觀念和工具，例如：?以喝牛奶與身高的關係解釋「迴歸分析」；?用青少年約會的例子說明「決策樹」；?以機動戰士、天文的分類來說明「群集分析」；?「自我組織映射圖」則是舉了戰機和連鎖咖啡店為例；?「關聯法則」中則有購物籃、將棋、Amazon網站的例子；?最後的「類神經網路」以花粉症為例。

而且，資料採礦與我們的個人資料、隱私權也息息相關，例如：電信業者來電促銷寬頻產品、銀行來電說有優惠利率或是投資訊息……這些都可能是資料採礦延伸出來的「問題」！

如果不好好管理個人資料，將使自己曝露在沒有隱私、高風險的環境中！

甚至，企業會比你自已更了解你！

正如作者在書末所說：「……個人對資訊機器或網路的理解程度，會影響個人的社會地位。

不精通資訊的人，將在毫不知情的情況之下被監視、被榨取資料、被管理。

相對與此，精通資訊的人，可以利用積存的資訊，使自己處於管理他人的優勢地位，甚至知道如何處身於監視之外……」在高度數位化的世界，如何活用、善用資訊，掌握自己的未來，本書將是最好的參考。

<<從資料中挖金礦>>

作者簡介

岡嶋裕史 1972年生於日本東京。
中央大學總合政策研究所博士。
曾任職於富士總合研究所，現擔任關東學院大學經濟學部經營學科情報部門、經濟學研究所副教授、中央大學Web Service Initiative技術部會副部會長。
著作有《一星期弄懂 - - 資安管理員的集中專題研究》（日本經濟新聞社）、《個人識別碼為什麼是四位數字？》（光文社新書）、《從郵政與有線電話了解網路的機制》（集英社新書）等。

譯者簡介 李弘元 日本明治大學管理研究所畢業。
曾任職日本丸紅株式會社、台灣丸紅的電子產業部，及台灣三菱電機系統工程師。
現任職台灣電子產業。
譯作有《你的手機我的夢》、《IT韓潮》、《中國製造》（合譯）、《經略中國，布局大亞洲》（合譯）、《人口減少經濟時代》等。

<<從資料中挖金礦>>

書籍目錄

序言 第一章 運用資料採礦，發現隱藏的定律 資料採礦不是萬靈丹 / 資料採礦與傳統統計分析的不同 / 儲存資料的成本變低了 / 從龐大的資料中淘金 / 預測未來 / 新手也能上手嗎？ / 發現的定律未必有用 / 理解基本原理和構造的必要性 / 定義因人而異第二章 商業上使用的資料採礦 資料倉儲的建構 / 調查商圈 / 了解顧客的傾向 / 捨棄奧客第三章 資料採礦的步驟 先決定想知道什麼 / 資料的蒐集方法 / 某些資料在使用上的困難 / 捨棄不需要的資訊 / 線索在哪裏？ / 先從單一資料下手 / 失敗經驗有助於往後的分析 / 下一個屬性 / 異常值的處理 / 根據單一屬性設定假說的極限 / 發現橫跨好幾個屬性的隱藏定律第四章 尋找妥協點—迴歸分析 更貼近正確的分析方法 / 建立假說 / 最小平方法 / 自變數增加會導致複雜化 / 迴歸分析無法順利處理的資料 / 善用不同的分析方法第五章 快速判斷—決策樹 定性的資料 / 過度培育決策樹會造成「過度學習」第六章 分類就會清楚—群集分析 老師教給我們？ 或是自己思考？ / 自己思考沒有教師的分類 / 機動戰士的分類第七章 複雜的分類—自我組織映射圖 模糊程度 / 壓縮多次元資料成為二次元 / 自我組織映射圖第八章 購物籃裏面的東西是？ —關聯法則 再談「紙尿布和啤酒」 / 找得到有用的規則嗎？ / 關聯性不被認同的規則、已知的規則 / 試作關聯法則 / 諺語中具有很多關聯法則 / 人力資料採礦 / 長尾效應 / 資料是競爭力的泉源第九章 模仿神經細胞—類神經網路 強大的學習能力與並列性 / 模仿神經迴路 / 基本架構非常簡單 / 學習的方法 / 神經細胞之間的結合第十章 資料採礦與資料管理 變質的資料採礦 / 資料採礦的黎明期 / 經由開放化，增加資料流入 / 他人比自己還了解自己 / 如何控管自己的資料 / 無法控管的網路 / 個人資料保護法能夠保障消費者嗎？ / 管理本身資料的必要性第十一章 監視社會與資料採礦 監視應運而生 / 曖昧的界線 / 監視的目的改變了 / 人透過資料被認識 / 監視資料的共有與分析 / 方便性的圈套 / 資訊強者壟斷資源

<<從資料中挖金礦>>

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>