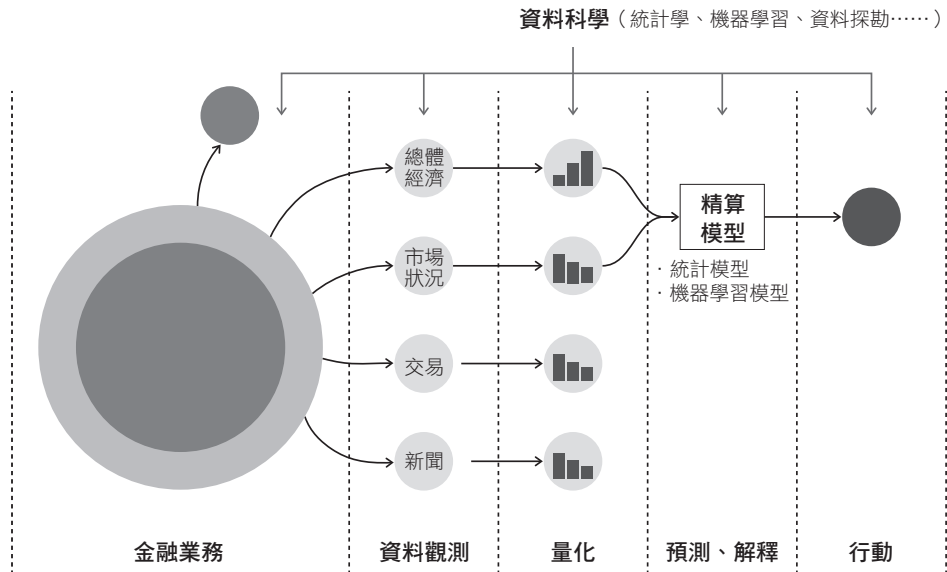


資料科學對金融機構的重要性

金融業又稱作資料生產產業，因此一直以來，資料化的程度皆領先其他產業，以健全的金融體系與縝密的金融理論為背景，將業務資料系統化，並使用電子化後的資料。

可是近年來如同在 FinTech 提及，金融業務本身的結構將產生巨大變化。業務結構變化意味著，與業務有關的資料產生機構及各種資料的意義出現轉變，因此必須以過去不同的手法處理資料，或運用未曾處理過的資料，建構掌握金融業務結構變化的新服務。

在這種背景下，提高了**資料科學**在金融機構的重要性。資料科學是與資料處理手法有關的學問，由統計學、機器學習、資料探勘、資料庫等學術領域組成。必須使用資料科學的各種手法，掌握金融業務的現況，並根據實際的情形，加強金融服務。本章將說明金融機構今後要持續成長，不可或缺的「資料科學」，其中會特別著重說明「機器學習」的手法。



◆ 金融業務與資料科學

金融機構的資料是什麼？

一般認為資料是觀測或調查的記錄結果，現在這些都以電子方式記錄在電腦上。本章將根據資料性質，分成「結構化資料」與「非結構化資料」，說明符合各種資料特性的資料科學手法。

本章描述的結構化資料是指以電子方式記錄時，具有列（記錄）與行（屬性）的表格式資料，如關聯式資料庫。在所有資料科學的入門書中提到「資料」時，幾乎都是指這種表格式資料。一般表格式資料是由含有多種屬性的記錄集合構成。此外，這種屬性集合稱作「特徵量」或「特徵向量」，代表資料的性質。

然而非結構化資料是指以電子化方式記錄資料時，不採用表格顯示的資料，例如文字、影像、聲音等。如果要分析這種資料，就得把記錄調整成具有特徵量的表格式資料，或類似表格式資料的狀態。近年來為了處理這種非結構化資料，設計了各種手法（請參考 3-5 ~ 3-7）。

何謂機器學習？

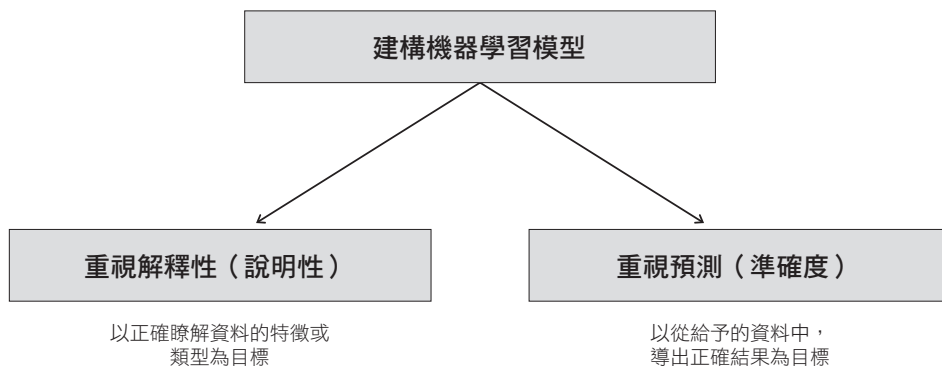
今後將參與機器學習專案者的入門知識

何謂機器學習？

機器學習是用機器找出潛藏在給予資料中的規則（類型）。過去的程式設計是由人類定義，機器執行動作，但是機器學習是把資料與機器學習手法提供給機器，讓機器學習資料的特徵，找出隱藏在其中的原則。

一般說明機器學習時，會根據業務條件，把重點放在「解釋性（說明性）」或「預測（準確度）」上。「解釋性」是指給予的資料具有何種性質，理解導致該結果的原因。然而，「預測」就像從過去的經驗中討論未來，學習過去的資料，料中將來的資料能帶來何種結果。

例如，證券零售業務的「解釋性」是指，從購買某種金融商品的結果，找出其中存在著何種客戶特性。「預測」是指篩選出比較可能購買該金融商品的客戶。



◆機器學習重視的條件

銀行的功能

我們平常使用銀行的服務時，幾乎不會注意到銀行的功能，但是屬於金融機構的銀行，具備了以下三種功能。第一種是「**信用創造功能**」。根據存款提供貸款，貸出的資金就像變成貸款者的存款，由存款衍生出新存款，進而增加存款貨幣。第二是「**支付功能**」。利用存款帳戶轉帳或付款，無須使用現金。第三是「**金融仲介功能**」。從資金寬裕的個人或企業收集存款，融資給需要資金的個人或企業。

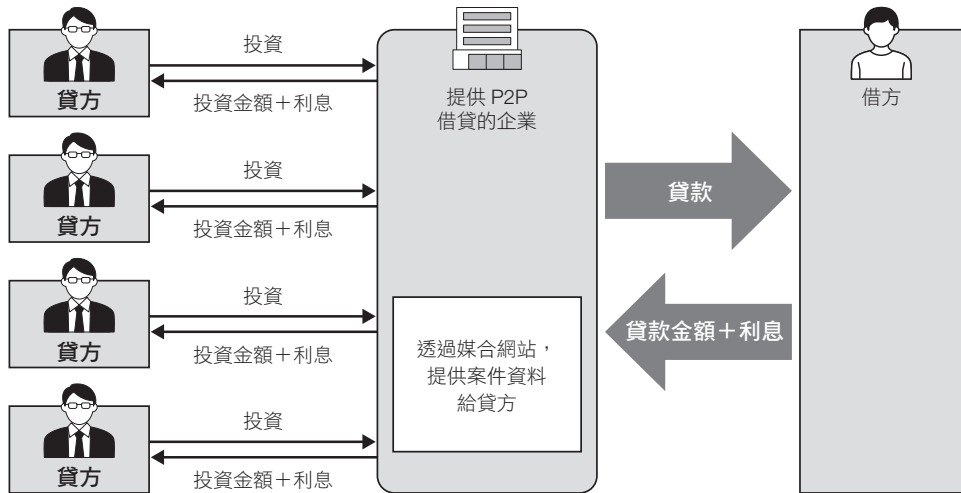
其中最受到重視的是「**金融仲介功能**」。融資給企業獲得的利息成為了銀行的主要收益來源，銀行為了發揮這個功能，必須盡量避免融資後，貸款者因經營狀態惡化而無法回收資金，出現倒帳風險（**縮減信用成本**），以及消除貸方與借方之間產生與借方信用狀況有關的資料落差（**資料生產功能**）。因此，為了提高資料生產功能，必須提升信用評估。

信用評估的分析種類

銀行的信用評估主要組合了兩種分析。一種是**定量分析**，這是使用記載在營業額或經常利益等財報上的數值，以及從這些數值計算出來的財務比率（例如經常利益 ÷ 營業額 × 100% 計算出來的營業利益率）等定量資料進行分析。另一種是**定性分析**，這是使用訪問企業及實地調查，收集業界動向及技術力、客戶、經營者資質等定性資料進行分析。

這些分析是針對每位貸款者進行的。如果貸款對象主要是中小企業，也會包含高倒帳風險的企業。由於每件貸款金額小，從中獲得的利息也少。

在日本，受到貸款業法的限制，貸方無法直接向借方融資，所以提供 P2P 借貸的企業不多。未來法規若能逐漸完善，參與的企業應會陸續增加。



◆ 日本 P2P 借貸的結構

• 交易型貸款

交易型貸款是指在網路商店或網路銀行等提供特定服務的平台，根據往來記錄進行信用評估再融資的服務。用來評估信用的資料包括信用卡的付款記錄、雲端會計系統每天輸入的交易傳票、存款帳戶的存提款資料等。融資對象必須是平台的使用者，根據平台上每天往來的高準確度電子化資料進行信用評估，因此審核時間比銀行短，通過審核之後，就能獲得融資，但是利率會比較高。

在日本，提供電子商務的企業愈來愈多，部分金融機構也開始進入這塊市場。

為什麼資料科學會受到保險業的關注

保險需要分析各式各樣與企業或個人風險有關的資料，一直以來都是使用稱作**保險數理**的統計手法來分析資料。保險可以說是以統計學為基礎的業務，但是近年來，運用 IoT 及大數據，開發新型保險商品的需求愈來愈強烈，因此機器學習及 AI 等資料科學開始受到重視。

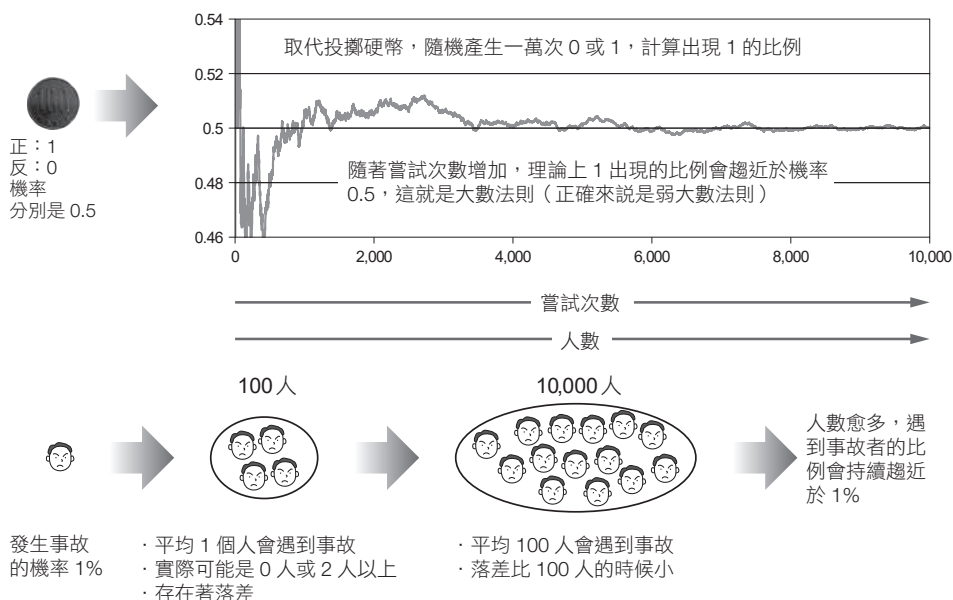
除了商品開發之外，在客戶接觸點或後台，也運用了資料科學來提高業務的效能與效率。保險的基本業務是根據資料分析風險，但是這一點卻可能因為資料科學的運用而產生極大的變化。

使用資料科學細分風險

保險是什麼時候出現的呢？這點眾說紛紜，據說現在我們使用的近代保險誕生於工業革命之後。工業革命後的機器化發展，大幅擴大了人類的行動範圍，因此對保險的需求也變得多元化。在背後支撐這一切的是十八世紀發展出來的機率論，以及該理論為基礎的保險計算技術，因而累積了資料。

保險包含了稱作保險原理的原則，支撐這個部分的是機率論的基本定理——**大數法則**。大數法則是指，機率 p 產生的現象，經過 n 次嘗試之後，該現象產生了 r 次。此時，嘗試次數 n 愈大， r/n 愈趨近 p 。

1 出現的比例



◆大數法則

根據大數法則，收集大量具有同質風險的個人，以平均事故率來設定保險費率，保險費的總額與保險金的總額會達到平衡。這代表著在每個人的保險交易中，理賠的保險費與實際獲得的保險金期待值可能相等。收到的保險費與理賠的保險金相等，稱作「收支相等原則」。投保人根據實際風險支付保費，所有人公平負擔保險費稱作「給付、反對給付均等原則」。這兩種原則稱作保險原理，有了大數法則，這些保險原理才會成立。

保險對象有各種風險，細分之後，可以增加保險的類別，滿足使用者的多元化需求。可是有時因為細分方法不同，可能出現成為保險對象的事件不符合大數法則，或沒有累積足夠資料的情況。此時，保險原理不成立，即使有風險，也很難成為保險的對象。

保險業在風險細分與保險原理彼此對立的情況下，花費了各種心思，開發出多元化商品，例如與吸煙者相比，非吸煙者的壽險保費比較便宜，而資料科學可以針對這點發揮效果。例如用資料科學分析醫療收據及健檢資料等大數

數位行銷與一對一行銷

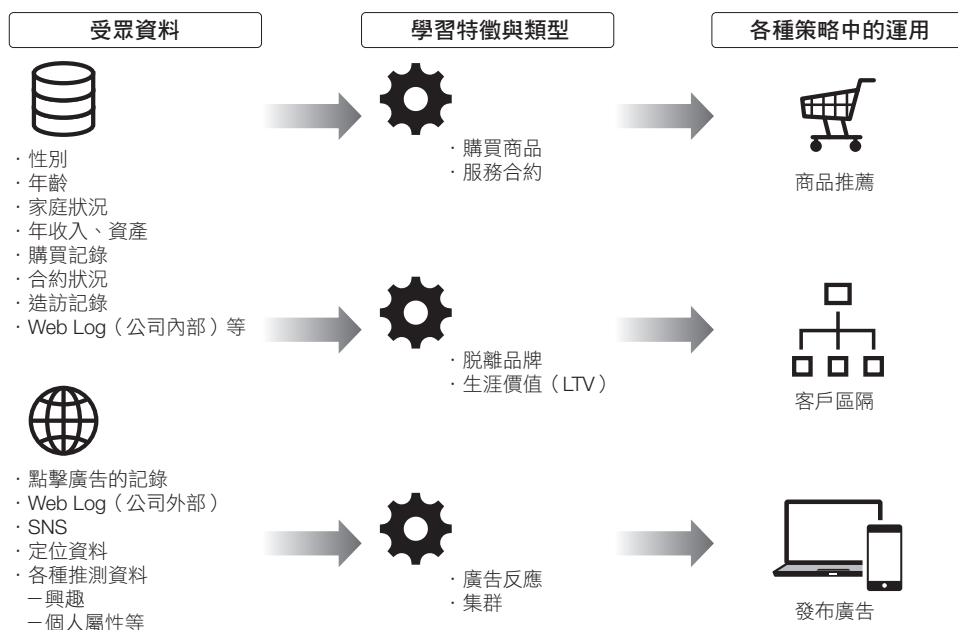
在企業從事的經濟活動中，行銷的處理範圍廣泛，包括市場調查、促銷、流通、與客戶建立關係等功用。行銷活動的內容包羅萬象，主要的功能是有效率地針對需求，提供商品與服務。

就功能面而言，**數位行銷**與傳統的行銷沒有什麼兩樣。數位行銷的特色是，除了過去使用的客戶資料，也把網站上的數位資料運用在行銷上。收集、累積的網站資料中，包含了網站及 App 的行為 Log 等。

如下一頁的圖示，2017 年網路廣告的市場規模超過雜誌與報紙等媒體，接近市占率最大的電視媒體。隨著網際網路在客戶接觸點的機會與商品促銷通路的比例增加，網路上的行銷活動變得更重要，同時也提高了網頁資料的價值。

透過網頁取得資料，不僅拓展了處理資料的範圍，也能運用高即時性的資料，例如登入會員網站、啟動 App、來店或通過特定地點等。有效運用每個人的客戶屬性及行為 Log 等資料，即時掌握使用者的需求，達到商品、訴求方法、時機等符合個人最佳化的行銷（**一對一行銷**）。

如果資料種類及分量不多，或想掌握的現象與資料關係明確，就算是人類，也能輕易區隔或評分。例如按照年齡與性別分別顯示一般廣告、橫幅廣告，或從過去的購買記錄中，大致分類現有客戶的重要度。可是，使用的資料若超過數百種，人類很難從未知的關係中，預測出結果或推論原因，不過機器學習與統計學卻很擅長這一點。由於處理大量資料以及重視高預測準確度的案例愈來愈多，因此除了以傳統統計學為基礎的手法之外，也開始運用機器學習等資料科學。



◆行銷運用的資料科學

利用外部資料擴大受眾資料

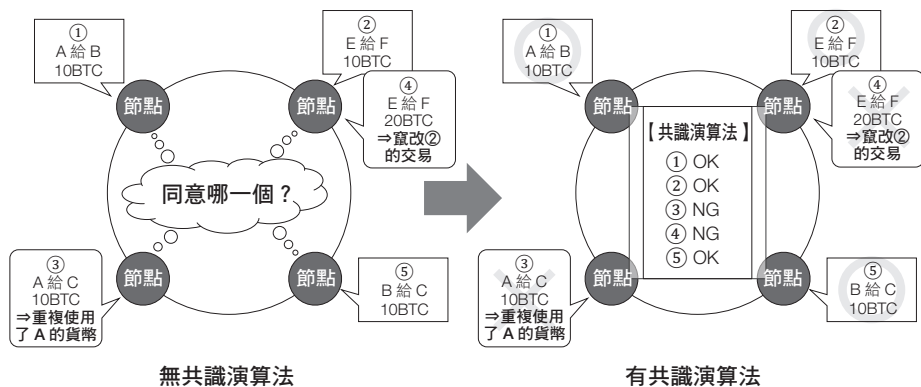
個人的客戶屬性、行為 Log 稱作受眾資料。在行銷常用的客戶屬性中，包括年齡及性別等人口統計資料，以及興趣嗜好、商品偏好等心理資料。另一方面，行為 Log 之中，還包含了商品的購買記錄，以及網站、實體門市、廣告等各種通路的接觸時間、接觸次數等。客戶屬性資料可以透過加

何謂共識演算法？

過去的主從式系統是由伺服器統一管理資料。然而，5-2 說明的 P2P 型區塊鏈在各個節點都有管理者，一邊共享資料，一邊維持彼此資料的整合性。即使分散的多個節點有某個部分故障了，仍能維持整合性的結構運作。

為了實現這一點，區塊鏈使用了演算法，根據特定原則，在各個節點之間取得共識 (= Consensus)，這種演算法稱作**共識演算法**。

共識演算法是區塊鏈技術的「關鍵」，也象徵了世上各種區塊鏈的重要性質。下圖簡單顯示了匯款時，共識演算法的作用。沒有共識演算法（圖左），就沒有驗證交易的方法，可能因此同意了非法交易。然而，導入了共識演算法的系統（圖右）可以整理、驗證網路上的交易，排除交易③與交易④等非法交易。



◆ 共識演算法的作用

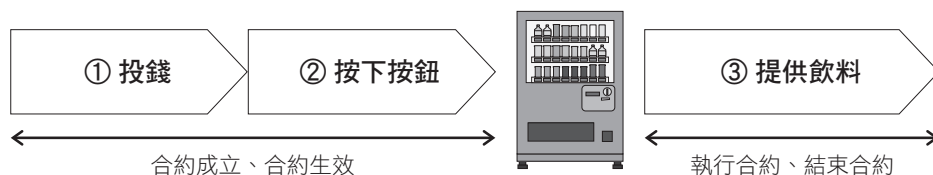
何謂智慧合約？

自動化處理合約減少信任成本

掌握區塊鏈普及的關鍵功能

智慧合約（Smart Contract）是部分區塊鏈擁有的功能，可以自動化處理合約，一般認為智慧合約掌握了區塊鏈普及的關鍵。

智慧合約的概念在區塊鏈出現以前就已經存在，1990 年由美國研究學者 Nick Szabo 提倡。Smart Contract 直譯的意思是聰明（Smart）的合約（Contract），其實是指自動執行的合約。Szabo 以自動販賣機為例，說明初期的智慧合約。把錢投入自動販賣機，按下飲料的按鈕，就會自動提供飲料。換句話說，自動販賣機是由管理者準備、會自動執行的合約。



◆初期的智慧合約實例（自動販賣機）

智慧合約的優點是，**降低與簽約者之間的信任成本**。以上述的自動販賣機為例，半數以上的消費者完全不擔心投錢之後是否真的會掉下飲料。

信任成本在 P2P（請參考 5-2）的交易中，明顯地表現出來。我們以網路拍賣交易為例，在這種交易方式中，賣家會擔心買方是否確實支付費用，而買家會擔心賣家是否確實寄送商品。一般是由經營拍賣的網站仲介負責預防這種問題並執行應變方法，使用者要負擔信任成本，當作仲介的手續費。使用智慧合約有機會能降低這種成本。

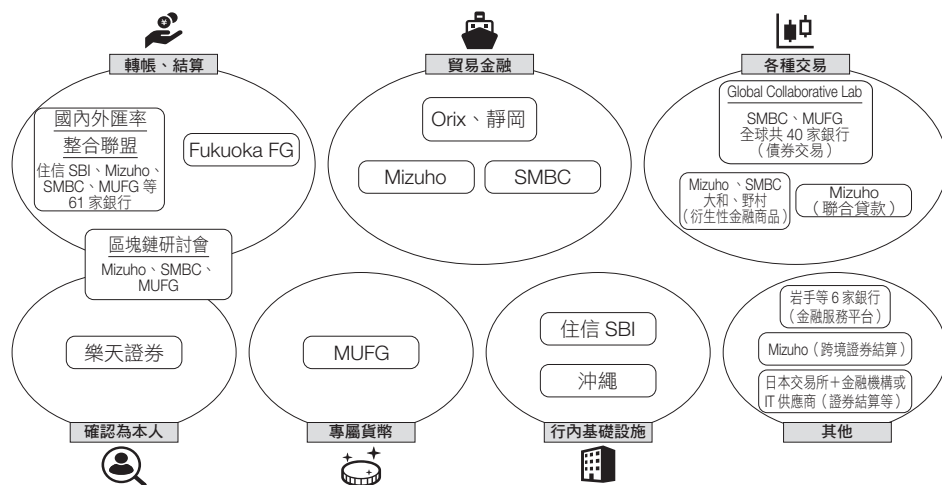
業務實例 (2) 金融機構的運用方法

正在評估運用在轉帳、結算、貿易金融、證券交易等廣泛領域

金融機構的區塊鏈運用方法

就連現行的金融機構也開始評估擴大運用區塊鏈的範圍。下圖是日本金融機構的運用案例。由此可以得知在轉帳、結算、貿易金融、各種交易、確認為本人、專屬貨幣、行內基礎設施等廣大領域皆開始加以運用。

以轉帳及結算為例。日本國內外匯率整合聯盟發表了運用區塊鏈技術，以智慧型手機為主的轉帳 App 「Money Tap」。利用新的轉帳基礎設施，可以進行安全且 24 小時即時的轉帳步驟。此外，在貿易金融領域，SMBC 等組織正在開發能統一管理與貿易有關的合約狀況及掌握貨物位置的平台。



(注) 加上底線的部分是由多家金融機構參加的聯盟名稱

◆日本國內金融機構的區塊鏈運用結構