

增訂版

大數據的 關鍵思考

行動×多螢×碎片化時代的商業智慧

阿里巴巴前副總裁／數據委員會會長
紅杉資本中國專家合夥人

車品覺——著

數據力決定競爭力！
進入數據驅動的商業世界，
「數據思考」是每個工作人的必備能力！

| 增訂3萬字全新內容，揭示大數據近未來的6大趨勢與5大觀戰重點 |

何英圻 91APP董事長 | 沈南鵬 紅杉資本全球執行合夥人 | 林之晨 台灣大哥大總經理、AppWorks董事長暨合夥人
林奕彰 eBay大中華區CEO | 許景泰 SmartM世紀智庫執行長 | 程九如 AppWorks合夥人 | 張建鋒 阿里巴巴技術
委員會主席 | 詹宏志 網路家庭董事長 | 鄭緯筌 「寫作力」與「內容駭客」網站創辦人——**熱烈推薦**

天下雜誌

增訂版

大數據的 關鍵思考

行動×多螢×碎片化時代的商業智慧

阿里巴巴前副總裁／數據委員會會長

紅杉資本中國專家合夥人

車品覺——著

數據力決定競爭力！
進入數據驅動的商業世界，
「數據思考」是每個工作人的必備能力！

| 增訂3萬字全新內容，揭示大數據近未來的6大趨勢與5大觀戰重點 |

何英圻 91APP董事長 | 沈南鵬 紅杉資本全球執行合夥人 | 林之晨 台灣大哥大總經理、AppWorks董事長暨合夥人
林奕彰 eBay大中華區CEO | 許景泰 SmartM世紀智庫執行長 | 程九如 AppWorks合夥人 | 張建鋒 阿里巴巴技術
委員會主席 | 詹宏志 網路家庭董事長 | 鄭緯筌 「寫作力」與「內容駭客」網站創辦人——**熱烈推薦**

各界好評

新零售時代，掌握數據為品牌核心競爭力。消費環境已虛實融合（OMO），但尚有不少零售場景老舊，線上線下資料無法打通融合，又怎麼驅動數據創造商機？如本書所提，要打造數據化的營運模式，才能把數據拿來營運。品牌數位變革，首要建立OMO循環外，更要有數據思維來加速決策，方能成為數據驅動的零售科技公司。

——何英圻 | 91APP董事長

品覺的《大數據的關鍵思考（增訂版）》一書淺顯易懂地解釋了「大數據」這個今天科技界的熱門詞。從用數據到養數據，娓娓道來大數據應用的實戰故事，是和數據打交道的各行各業人士值得一讀的好書。

——沈南鵬 | 紅杉資本全球執行合夥人

我跟品覺相識多年，在大數據還沒流行以前，他就開始關注這個趨勢，研究數據應用的意義，以及怎樣真正讀懂數據和利用數據。這本書深入淺出，把深奧的大數據講得生動有趣，非常好讀；書中匯集了品覺個人的很多真知灼見，非常值得一讀。閱讀此書是一種享受，正如我享受與品覺的友誼一樣。

——林奕彰 | eBay大中華區CEO

本書的出版絕對是「大數據思維」商業運用的寶典！不論你在什麼產業，我強力推薦每一位企業領導人或經理人，都應積極地讓大數據思維

融入企業體系。因為企業未來的競爭力，絕對跟活用大數據有著緊密相依的關係！

——許景泰 | SmartM世紀智庫執行長

品覺在大數據應用領域有著豐富的經驗，書中提出的觀點對大數據的決策者很有意義。大數據使企業經營者擁有了望遠鏡和顯微鏡，除了能對已知事物觀察得更清楚之外，更關鍵的是，使我們能看到之前從來沒有想到、沒有觀察到的事物。品覺在這方面提供了有益的建議。

——張建鋒 | 阿里巴巴技術委員會主席

推薦序

培養數據思維，找到全新的藍海

文 / 程九如 | AppWorks合夥人

蝙蝠可以在昏暗的叢林中快速飛行及獵食，靠的是對情報的理解與運用，而不是靠有限的視野或過往的記憶。具備數據思維與能力的組織，可以透過對數據的收集、分析、探索與運用，在數據的維度中找到全新的藍海，有效掌握產業趨勢，並打造優勢的立基。

組織想要具備數據思維與能力，整體必須經歷一番蛻變的歷程，才有可能在整體的認知、行為、能力、思維，甚至文化上，獲得徹底的翻新。

作者以親身的經歷清楚地點出了組織蛻變的關鍵步驟與心法，並分享了如何循序漸進地在潛移默化中引導並推動組織整體的數據化轉型與升級，且最終獲得了豐碩的成果。

對事的書我看的是技術，對人的書我看的是藝術，這本著作在每個章節裡都交融著技術與藝術，好不精彩，值得所有重視數位經濟的企管夥伴們細細品味。

推薦序

以人為本，探究用戶的真實需求

文 / 鄭緯筌 | 「寫作力」與「內容駭客」網站創辦人

自從離開媒體與網路產業之後，我便以企業顧問、職業講師和專欄作者的身分行走江湖，儘管工作忙碌，卻不以為忤。因為工作的緣故，每天接觸的資訊龐雜而眾多，要看的報告更動輒超過二位數，但我卻安之如飴，也已經習慣了與數據為伍的生活。

曾有寫作班的學生說我是「念理工科的寫作大神」，其實，自己雖然一開始在專科時代讀的是電子工程，在碩士階段讀的是工業工程，但我也喜歡嘗試跨領域學習，所以包括傳播、文創和商管，都在涉獵的範疇之中。這些歷練不但讓我增廣見聞，也培養了宏觀的視野，在面對大數據的時候能夠更有系統和章法。

讓我們把焦點回到這本書《大數據的關鍵思考（增訂版）》。還記得不久前，大數據還是一個熱門的關鍵字，如今卻已經百花齊放。大家都知道大數據可能很重要，卻很少有人弄得清楚要如何運用數據的力量，來解決人們在工作與生活面向的各種問題。

作者車品覺告訴我們，大數據雖然聽起來酷炫，但並非萬能。想知道問題能否憑藉數據的解讀來解決，首先必須思考五件事以定義需要解決的問題，包括：這是什麼問題？誰的問題？這個問題是由我們來解決嗎？需要現在解決嗎？資料能夠解決嗎？唯有當這五個問題的答案都是肯定的時候，或許就是當下可以靠資料解決的問題了。

如果未能事先定義問題的範疇，就不知道各種騰空出世的應用該如何衡量其成效。所以，想讓大數據發揮效用的話，首先必須清楚地定義問題，換句話說，就是追尋從知道到不知道的過程。

他認為，企業界能夠運用數據解決的問題，基本上都是圍繞著商業公式與商業價值打轉。好比新近在成長駭客領域相當受到大眾關注的議題，像是：如何用更低廉的成本攬客？如何留住客戶？如何擬定價格策略？看起來雖然千頭萬緒，但其實都能夠透過大數據的演繹來找到答案。

一如他在書中所提到的重點，大數據的本質，首重還原用戶的真實需求。在這個數位科技日新月異的年代，從事數據分析仍需以人為本，講求使用者體驗。大數據的應用開始變得多元，我們更應該審時度勢，用更開闊的胸襟與視野去迎接大數據的年代。話說回來，既不能讓當前的技術限制未來產品的發展，也不可只看重數據的力量，而忽略了人的溫度與本質。

我永遠記得第一次拜讀車老師的大作時，除了佩服他的技術本色，也為他的浪漫情懷所折服。這回，很高興有機會再跟各位讀者朋友們推薦車老師這本大作的增訂版。如果您對大數據有興趣的話，請不要錯過了這本書——要知道，好書往往經得起時代的淘洗，車老師的這本代表作就是最好的例證之一。

新版序

我在阿里巴巴的六年

回顧我在阿里巴巴的六年，從經歷來說，是先負責支付寶，繼而負責淘寶，最終負責整個阿里巴巴集團的數據工作。實際上，這個過程同樣也是我對數據應用的一個歷程：第一階段用數據，第二階段養數據，第三階段從看數據到用數據。

加入支付寶

在加入支付寶之前，業界都認可我是一個比較懂得用數據去設計、優化、疊代產品的產品長（CPO）。我在支付寶大膽地結合產品與數據分析部門，累積了不少經驗。

說起加入支付寶，其實有些偶然。當時，支付寶邀請我去分享如何用數據做產品。分享會之後，我去了華星大廈的一個會議室，裡面坐著Lucy（彭蕾，時任支付寶CEO）、Eric（井賢棟，時任螞蟻金融COO），他們問了我幾個如何能用好數據的問題。當時，我反問了Lucy一個問題：「你每天大概會看多少數據？」

「大概幾十個吧。」

「不對，你不應該一天看幾十個數據那麼多，我覺得你應該只看幾個數據就夠了。」

或許他們已經不記得我提出的這個問題了，但是我卻記得非常清楚。因為在支付寶的兩年裡，這是我做數據產品時一直遵循的原則。

堅持數據產品化的理念

正式加入支付寶後，我面對的第一件事情，就是著手組建數據分析師團隊。2010年的支付寶，公司大部分業務團隊對數據團隊都不太滿意，而且高層對數據能產生的價值也沒有太多認識。而數據負責人的調職，也導致了支付寶的數據分析和數據技術部門被分開管理。

面對數據分析和數據技術團隊的目標不一致，加上業務部門對於為什麼要用數據，以及如何用好數據充滿疑問，甚至有人認為數據分析僅僅是門面工夫，所以數據技術團隊有時候會跟我的團隊「對著幹」，因為兩個團隊的老闆不是同一人，他們也沒有義務一定要支持我，有時候甚至會撇開我們，直接跟業務部門對接。

但是最後，我堅持的數據分析產品化路線終於得到了雙方的認同，數據技術團隊甚至還主動過來問我：「我們怎麼合作比較好？」所以那時，雙方的關係就從「對著幹」變成了「我們會全力支持你」。

終於有一天，Lucy走進我的辦公室跟我說：「你知道嗎？你們做的產品讓我覺得很爽，我很喜歡這個產品。」而當我在支付寶的第二個產品「地動儀」出來時，她對我說：「這就是我夢寐以求的數據產品。」

同事和老闆對結果的支持，讓我肯定自己在支付寶走這條路是對的。如果讓我總結這其中的精髓，我的祕訣就是——數據分析也要講求使用者體驗。

簡單來說就是，別說是數據產品，哪怕一份小小的數據報告，我們都必須要知道以下五個問題：

- 這是什麼問題？
- 這是誰的問題？
- 這個問題現在必須馬上解決嗎？
- 這個問題我們能解決嗎？

- 我們有足夠的數據來解決嗎？

把報告當成產品來做

當我「苛刻」地提出把月度經營報告當成產品來做，目的是讓這份報告像iPhone一樣好用的要求時，整個團隊都傻了，報告怎麼能像產品一樣！那麼，為什麼數據報告不能做到行雲流水，讓人看得很爽快呢？

所以，我是這麼要求的：首先，報告的前三頁一定要吸引人，讓人有欲望繼續看下去；其次，當別人看到一個數據，心裡在猜測數據背後的各種原因時，我的下一頁報告就要解答他心中的疑問。

如果要用產品管理的理念來運行分析團隊的話，必須要問「問題是什麼」，以此作為切入點。接著，解決使用者的問題，得到他們的信任，然後就可以做更多的好產品。這是一套做產品的理念，因為產品需要不斷疊代，並非一勞永逸。

那麼，我怎麼評估報告是否成功呢？簡單來說，如果一份月度經營報告僅僅五分鐘就被pass掉了，那它肯定是不成功的。

我在支付寶時，幾乎每次月度報告都會讓管理層討論三、四個小時，一份數據報告居然會成為管理層的討論焦點，每次都有很多高層管理者關注我在講什麼，這時，我就知道自己肯定成功了。

當然，我同時也會提醒分析師，千萬別以為靠一份報告就可以解決所有事情，一定要讓業務部門認為這份業務報告是大家一起探索的，這是我管理團隊的一個基本理念。

管理團隊、推動理念的祕訣

總之，數據產品化的理念最終得到了Lucy和其他高層，以及團隊和同事們的認可。那麼，針對這個理念，我是如何在團隊內一步步推動的呢？

首先，我讓團隊做好基礎服務——「白米飯」。為什麼叫白米飯？因為對人來說，要是飯不夠就會出大問題。對公司來說，如果數據基礎服務滿足不了使用者，他們就會跳起來，那時候再高大上的報告也沒用。所以，我要求分析團隊一定要跟業務部門達成共識，知道什麼東西如果我們不提供給他們，結果將會很嚴重。

但是，長期做「白米飯」的分析部門是沒有前途的，所以其次，我抽出團隊一部分人成立了「突擊隊」。每當我聽到管理層對數據的需求時，就把資訊帶到團隊，讓團隊成立兩、三支「突擊隊」，讓管理層用到有用的、合適的，以及帶有探索性的內容。把自己最精銳的部隊放在突擊隊裡，這個決定非常重要。

值得注意的是，成立突擊隊有一個前提，那就是必須要有人能接觸到最高管理層的資訊通道，否則無法得到高層的資訊。而這個角色就是我，我接觸最高管理團隊，並且把資訊輸送給突擊隊。

以「白米飯」和「突擊隊」為基礎，最後一步，我們開始沉澱數據分析的框架。為什麼數據分析框架如此重要？因為它是數據產品的養分。我用建立數據產品為目標的理念來經營數據團隊，而數據分析框架是給數據產品養分的關鍵點。

在支付寶時，基於Lucy和Eric對我的支持，數據產品都是自上而下推行。我先做了「觀星台」，這是給高階管理者的數據儀表板工具，然後做了「地動儀」，一個給中階管理者的數據產品，接著還有「客服360」等。

所有這些數據產品都是從數據分析框中提煉出來的。把分析的理念和框架變成數據產品，本質上是數據泛化的過程。這個過程非常重要，因為數據報告的需求會越來越多，如果不這麼做，分析團隊將永遠被冗雜和重複的工作所困。

不讓當前的技術，限制未來的產品

在這期間，我們也做過一個改變。因為「觀星台」和「地動儀」還是傳統的分析工具，只不過讓你有更好的「眼睛」，看到更好的數據。用我一直強調的理念來講，這屬於描述性的報告，用於描述一家公司今天做得怎麼樣。但由於我們有了分析框架，可能稍微會帶一點診斷性，所以，這種產品主要還是描述性產品，偶爾會做一些診斷，但還未達到探索預測的狀態。

在這個看法的基礎上，我有了個新想法，想法的靈感源於Facebook。當時Facebook積極發展探索性的數據產品，視覺化及交互功能已經非常靈活。探索和靜態報告的區別在於，使用者可以選擇自己的角度靈活地看數據，真正做到「逛」數據。

因此，我們創造了一個產品叫「黃金策」，可以讓使用者很輕鬆地在三秒鐘內，就看到在不同標籤下用戶購物行為的異同。比如，30~35歲的上海女性，擁有一張信用卡和多張信用卡的兩個不同群體，看她們的購買行為有什麼不一樣；或者，看廣東和上海的女性在購物表現上有什麼不一樣。使用者可以非常輕鬆地比較不同用戶群的購物表現，我認為，好的數據產品應該是可以激發使用者靈感的。

在做「黃金策」這個產品的過程中，讓我深有感觸的是，數據產業的人，很容易讓技術制約產品的發展。當我提出做「黃金策」時，大部分技術工程師都斷言，三秒鐘就要有結果是不可能的。因為以前類似的靈活配置的報告，一般都不能做到即時查詢，如果要快速產出的話，只能找分析師。即便如此，也可能因為資源排期等原因等待很久。

當時只有一個人站出來說「我可以幫你搞定三秒鐘」，這個人就是蔣傑（現任騰訊數據中心總經理）。於是很自然地，我就跟蔣傑一起合作了這個專案，並且成了好朋友。後來，蔣傑很出色地幫我完成了這個專案。所以我覺得，做技術的人，不應該讓當前的技術和思維，限制未來的產品。

就使用者體驗來說，使用者之所以不使用我們的產品，其中一個很

重要的原因是，沒有人願意耗時太久等一份報告。因為當報告出來之後，他的問題可能早已經解決了。試想，在做決定時，拍一下腦袋只需兩分鐘，但等一份報告要兩個星期，而且報告的結果還不一定能幫上他，你說，他會選拍腦袋還是等報告？所以，這也是我們產品人的理念：到底怎樣才能解決使用者的問題？而這也是產品人和數據人不一樣的地方。

讓使用者喜歡數據，團隊就會喜歡自己

除了以上這些有關數據的工作外，那時我還做了另外一件事，我啟動了一個專案叫「西湖品學」，每年請公司外的數據達人來分享經驗，這是一場數據人的盛會，當時很受歡迎。我做西湖品學有兩個原因，首先，我認為人們不使用數據，並不是他真的不想使用；其次，沒有人推廣使用數據的好處。

那時，我常在支付寶說一個比喻：「我右手用得好好的，你突然讓我用左手，那你就要告訴我用左手的好處，不然我無法改變我的習慣。」

用數據的習慣也是一樣，當他知道數據好用時，就肯定會用。所以，每一年的西湖品學，我都請外面很多人來講數據到底有什麼好處，讓公司內部的數據人知道數據應用的商業場景是怎樣的，從而激起更多人對數據的渴望和期待。西湖品學就是這樣一個全新的嘗試。

在數據分析人員中，我們大部分分析員都被戲稱為「表哥」、「表弟」，因為他們每天都要做很多表。但是，每天僅僅做表的人會有幸福感嗎？每天做這麼多表，也不知道別人怎麼用。沒有結果，肯定就沒有幸福感可言。而西湖品學讓他們知道，他們的責任在哪裡、怎樣跟業務對接，他們不僅僅是「表哥」、「表弟」。所以我的目的就是，不讓他們抱怨自己的才華被埋沒了，要讓他們知道自己努力的價值所在。

當我的分析師看到管理層很激烈地討論他們做的報告時，我能看見他們臉上的笑容，「不枉我們兩個星期不眠不休地把這份報告做出來。」其實，大多數數據人都不怕辛苦，怕的就是做出來的東西沒人看。

總之，我對分析師的要求有些不同，我要求他們對商業有一定理解，這就是我在這本書中提到的「混、通、曬」——混在商業情境中，打通數據與商業場景，將數據產品當成載體，把這個東西曬出來，讓更多人使用。

當你看到你的數據使用者不能沒有你，當你看到你的使用者滿意的時候，你自然會感到興奮，整個循環就能激勵人心。

總結起來，數據產品化的理念，是我在支付寶對整個數據分析部門的頂層設計，我一直用這個理念在運行部門，決定資源如何配置。怎樣讓我的數據使用者喜歡用我的產品？取悅使用者的辦法，就是幫他解決問題。

然後就是讓數據更泛化。因此，我們必須宣傳數據的好處，教育我們的使用者如何使用數據。我們不能假定使用者最終會明白數據的好處，而是要一邊宣傳理念，一邊教授技巧，讓使用者知道怎麼用數據、怎麼用好數據，了解數據的價值。

最後一點則是，讓我們部門的員工喜歡自己，喜歡自己的工作。

我在支付寶的成功，也離不開我的堅持。首先，我益於自己堅持的頂層設計的理念，我的老闆Lucy和Eric也十分相挺，願意給我機會，給我很大的支持和包容。最後就是，我有很棒的下屬，能夠實踐我的理念。

記得當我晉升到副總裁，成了當時阿里巴巴數據領域職位最高的人時，Lucy曾經說：「在中國，品覺是數一數二的分析師。」這句話當然是對我個人的偏愛，但如果她指的是電商平台數一數二最好的分析師，我覺得能勉強接受。

每每回想起我在支付寶的兩年，甚至離開支付寶，到了淘寶之後很久，公司內外依然有很多人對我說支付寶的數據分析團隊是最棒的。每當聽到這樣的讚美，我都覺得十分欣慰，那段經歷讓我感到非常幸福。

淘寶的挑戰

有一天，Lucy把我叫過去說：「現在淘寶很需要你，淘寶的數據體系需要你去整理。」那是2012年底，也是我在支付寶的黃金時代，就這樣，我來到了淘寶，開始了我在阿里巴巴後四年的時期。

淘寶前兩年的經歷非常豐富，回想起來，就好像過去了十年一般，我的白髮基本上也都是在這段時間長出來的。

從支付寶來到淘寶，意味著我要重新打一仗。由於我已經在支付寶打了一場勝仗，可以照搬很多頂層設計到淘寶使用，但我也不能完全照搬，因為淘寶使用的頂層設計有些跟支付寶一樣，有些卻完全不一樣。

淘寶跟支付寶相比其實要複雜得多。淘寶本身是個自下而上的組織，是個極其複雜的生態圈。有趣的是，當我來到淘寶時，面對的情況恍如回到剛入職支付寶時的情形：團隊的人差不多都走光了，我感覺就像踏上了一塊杳無人煙的新大陸——分析師只剩下幾個人，而且我被告知這幾個人還會陸續離開。

面對這一切，我做了一個今天回想起來仍然覺得很勇敢的決定：我說我不會帶支付寶的一兵一卒過來。所以說，我在支付寶組建了一個完整的團隊，然後在淘寶又重新組建了一支全新的團隊，這段經歷其實讓我很痛苦，但也是極為珍貴的人生經歷。

和支付寶時代面對的團隊不一樣的是，在淘寶，我需要同時管理分析及技術團隊，所以不用經歷支付寶前期那段分析團隊和技術團隊目標不一致的痛苦。

當時，淘寶的數據技術團隊很亂，而且是累積多年的亂，所以，要整理淘寶的數據體系，只用頂層設計和數據產品的方法去做，肯定是行不通的。

但還好，此時我已經有很多經驗，加上當時的總裁三豐跟我感情非常好，比起我在支付寶時，其實辦事更容易。比如，在支付寶時我還不能直接問Lucy到底你要什麼，但是在淘寶，我可以隨時到三豐的辦公室問他要什麼。在支付寶我只是一步步嘗試滿足Lucy的需求，而在淘寶，我有管理層的絕對支持。所以，我遇到的問題都是來自內部，沒有外在問題。

在淘寶的分析團隊中，我同樣分別推出了「白米飯」和「突擊隊」，以及分析框架。不一樣的是，我們很難給淘寶定KPI，說達到某個程度就是做得好。如果你說這樣就好了，你會發現另一個指標可能就不好了。

也正因為如此，當我們建數據產品時，我犯了一個錯誤，我想用自上而下的方法去做，最終發現這在淘寶行不通。儘管得益於我在eBay的經驗，對淘寶這樣的雙邊市場比較熟悉，但後來我依然覺得淘寶的生態圈非常複雜，其分析框架和支付寶非常不一樣。

當我發現自上而下的方法不行之後，我就逐步去做更多能發現問題的數據產品。總體來說，還是堅持數據泛化的方向。在淘寶這樣自下而上的組織，這相對來說比較容易，因為權力已經下放給每個小二（客服人員），在需要數據幫他們做決策時，讓他們每個人都用數據產品，其實比在支付寶時期更容易。反而因為數據非常零散，要做好一個數據產品，知道如何化繁為簡更重要。因為雖然有資源，但使用資源的人也很多。

關於在淘寶所做的數據產品我這裡不再一一贅述，值得一提的是，我們當時做了一個非常重要的決定，就是行動終端。那時候，很多淘寶的消費者都開始使用智慧型手機，我就決定，我們一定要做行動版的阿

里數據App。「淘數據」這個電腦版產品在我來淘寶之前就已經存在，只是一直不溫不火，不是沒人使用，而是產品的使用者體驗不好。因此，當我們用行動版取代電腦版之後，發現很多高階管理者及中階管理者一下子變成了午夜常客，我就知道自己的決定是對的。

此外還有一件事，那就是數據服務，當時需求也已經顯現出來。因為我們這個部門承擔的責任不僅是數據分析，還有數據底層的搭建。雖然我的業務分析團隊當時只服務淘寶，但是數據技術團隊要服務天貓、淘寶、聚划算等多個業務部門。這需要我們在數據服務上分幾個層面去做：一是除了淘寶之外，天貓、聚划算事業部的數據產品，我們也幫他們做了；二是我只建數據中間層，別的部門可以用我們的數據中間層來建他們的數據產品；三是提供數據服務，我用早期的數據服務方式把數據輸出給其他部門使用。

這就是我在淘寶前期的情況，上半身（分析師）是淘寶，下半身（數據底層）是集團。

啟動阿里巴巴大數據

進入阿里集團的這幾年裡，大部分業務部門的分析師都歸我管理，同時，數據平台和技術底層也都在我這邊，所以我的業務就演變成了集團數據技術及產品部，內部叫DT（Data Technology）。

回想淘寶時期，我一直覺得，如果我能早點整合數據底層團隊，收穫會更大。整合數據要花的力氣往往跟數據的離散、不規範、各自為政很有關係。有時，白紙一張重新開始，比起消耗大量時間在處理老問題要簡單得多。

而對於我們整個DT部門，這時候有兩個新的使命：一是做對外的數據產品，即生意參謀，同時也逐步把集團中其他對外數據產品，例如

「數據魔方」及「量子報告」等，綜合到生意參謀上；二是建立數據公共層，這是結合集團各業務群的共用資源而成的數據服務，希望可以成為內部的DaaS（Data as a Service，數據即服務）模型。

數據公共層的起緣，是因為當時集團各個部門建了很多自己的數據「煙囪」，重複建設嚴重，導致數據的存儲、加工等資源嚴重浪費。怎麼把各個煙囪中的數據分為專有和公有，並且把公有的數據放在數據公共層，這是我在這段時間內耗費精力最多的事情。在我眼裡，這個數據公共層也是我認為在大數據中最重要的東西，我也在其中學到了很多經驗。

前面提到的那些應用，我們都可以稱之為「數據」，而這時才是「大數據」。大數據的意義就在連結，數據公共層就是跨業務群的精華，讓所有人都能利用其他人的數據冗餘，這時，大數據才算做了起來。

我這麼說，不是說我們以前沒有數據共用，而是從前都是零散和沒有計畫地分享及同步數據，有點小打小鬧。到了正式建立數據公共層的時候，才有人去整合全集團的所有數據，把它當成一個數據資產共建來看待。所以，數據公共層後來就被CEO定義為公司唯一的數據資產收集與加工部門。

打通數據閉環

在淘寶的幾年生涯裡，有個非常重要的專案讓我永遠難忘。

有一年，馬雲要我們把聚划算自動化，在我眼裡，其實這就等同於商業版的無人駕駛技術，因為，要把這個部門的品類規畫、選品、定價等流程自動化，讓機器決定什麼產品哪天上架。它要求整個過程中的決策都要自動化，而沒有人參與其中，這對於我們在數據的收集、加工、提煉等整個鏈條的能力都提出了非常大的考驗。

而且，這個專案也是真正實驗如何在業務營運中，使數據技術成為主幹的一步。數據能力、演算法，要完全融入業務流程當中，每個環節中的數據回流等，都將成為循環的一部分。

好在，這時我們部門已經是一個具備全方位能力的部門了，能夠覆蓋數據的準備、加工、收集、分析、整合到使用的整個鏈條，而且我的數據科學團隊也開始參與集團中很多不同業務的專案，甚至連交通這樣的政府專案都有涵蓋。

對於這段時間的經歷，我想說，我很榮幸，能夠用我的白髮換來親身經歷一場人機大戰的故事，讓我可以在過去三年中接觸到這麼多數據，甚至整個數據鏈，最終我實現了數據應用的閉環。要做好數據，我們首先要在數據使用的閉環中，讓數據越來越容易使用，同時在數據準備的閉環中，讓製作數據的成本越來越低（見圖1）。

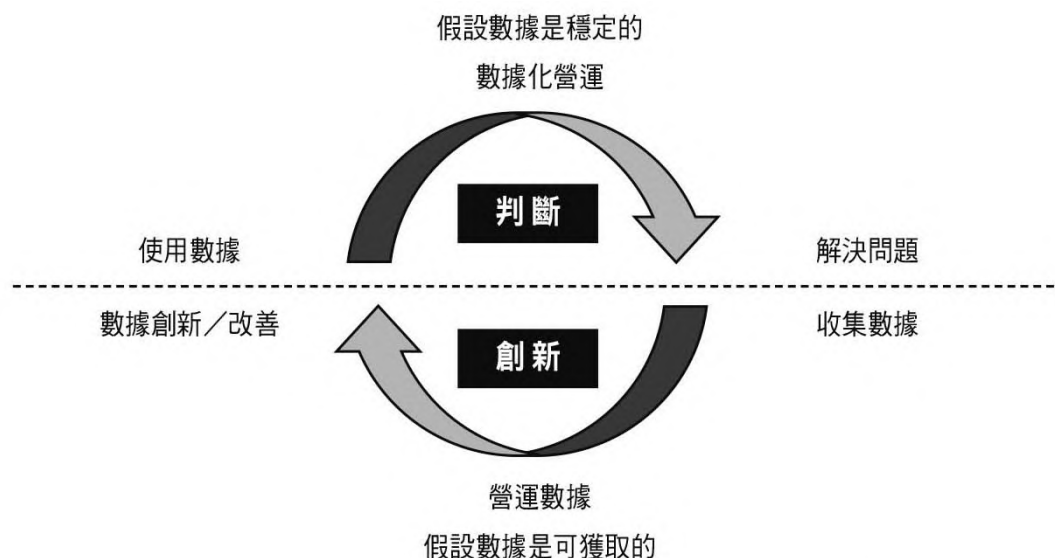


圖1 數據化營運和營運數據的閉環系統

在這個循環系統中，虛線以上，使用者能很簡單地使用數據，虛線以下，是讓數據製作變得更快、更自動化。一個是使數據產生價值，如

何讓多種、多樣、多變的數據原料，可以快速靈活地重組，並且相對穩定，這是做數據基礎建設天天都要去想的問題。另一個是如何讓使用數據變得更簡單，這是兩個閉環。

這就是「存、管、用」和「混、通、曬」。「存、管、用」就是用技術而不是人工方法，去做資產或原料管理。怎樣評估做得好不好呢？得看數據有沒有、細不細、全不全、穩不穩、快不快。所以，一旦把數據管理好了，我們就會很快、很靈活地做出數據，並且在使用時產生很大的價值，到底我們提煉到的會是垃圾還是黃金，就取決於此。

以往，用數據要「混、通、曬」，養數據要「存、管、用」，在經歷了支付寶和阿里集團的六年後，我覺得以上每一個環節都已經出現了很多創新，比如，感知技術的創新，讓數據收集層面多姿多彩；存儲的格式更適合數據靈活調動；人工智慧除了直接服務於產品之外，還可以增加數據的識別度；大家都已經意識到，數據的互聯互通及品質保證，是數據科技的基礎。

數據團隊要學會取捨

在淘寶的幾年，可以說是我在支付寶成功之後的收成期。得到高層的信任，拿到很多重要的專案，很多人都想加入我的團隊，人才也不再是一個問題。

從我進來淘寶到現在，數據團隊一直是急速奔跑的團隊。我進淘寶時，團隊僅僅七人，如今已經發展到七百多人，規模擴大了一百多倍。與此同時，數據部門承載管理層的期望也越來越高，也越來越在聚光燈下。

團隊發展時，我一直比較糾結員工配合的問題，工程師和分析師的合作，是數據部門的一個「死門」，因為分析師總是想把髒活給技術的人做，而技術的人覺得這是個苦差事，總是想做一些比較偏業務的事

情，兩者常常都不滿意各自所在的位置，想跳出自己的圈圈。此外，分析師也有一些配合的問題，比如決策分析和業務分析，兩種分析師到底該怎麼區分，怎麼讓大家一起合作，一直是個難題。

這樣的配合，我認為是數據和業務團隊的整合。中央部門的數據團隊一定要面對一個很殘酷的問題：什麼是我們該做的？什麼又是當時機成熟之後，要放手到業務部門的？

簡單來說就是，誰做得好，就應該由誰來做。2014年時，我也曾經想把商業分析團隊放回到各個業務部門中，結果業務老大都不同意，認為由我管更好，誰做得好就歸誰。

在很多專案中，我發現，只要業務部門一旦看到數據能幫上他的忙，而且是非常重要的東西時，他們就會把應用的東西拿回去。所以中央部門不用強求，只能做小專案。在大專案中，大部分業務團隊都有自己比較好的應用場景，能更好地進行試驗。所以，中央部門慢慢要形成好的數據中間層，這是關鍵。

中央部門很難做好應用，分析師也一樣，應用和分析師都需要對業務場景非常理解才能做得好。而底層數據怎樣更快、更靈活、更容易提供給使用者使用，肯定是中央部門需要做的。

但劃分不是絕對的，而是你中有我，我中有你。決策分析最好放在中央，因為決策分析是服務整家公司，幫助CEO看大盤，肯定不能用業務部門的視角。

數據則分公有數據和專有數據，專有數據要回到離應用比較近的地方，公有數據是公司最有用的資產，由中央部門集中管理，讓它成為很重要的原料，以供別人使用。那麼，邊界怎麼分？我只能說：只要大數據越泛化，中央部門要承擔的責任就越重，也越集中、越具體。

這些年來，我覺得現在阿里巴巴的數據依然沒達到我理想中的藍圖，主要有兩個原因：一是整個大環境還未形成，二是阿里巴巴的數據環境才剛開始，演變是循序漸進的過程，這個過程是急不來的。

我看數據科學

在2010年底，我就意識到大數據的出現，所以當時就跟Lucy提出，支付寶要有一些數據科學的人才儲備。但可惜的是，我雖然感覺到了新時代即將來臨，卻未能正確地指出機器學習的趨勢，所以沒有大力引進。

到了淘寶後，我是第一個在管理層力推大量招聘數據科學家的人，當時還有其他高層質疑我的做法。但在當時淘寶總裁的大力支持下，我終於如願。記得我招到第一位數據科學家後，我們每週都會在我的辦公室會談，用非技術語言聊數據科學家是做什麼的，很快地加深我對數據科學的理解。

傳統的數據科學家常常是博士或博士後，研究的方向是機器學習。數據科學家的出現，其實是個演變的過程，幾年前，企業的數據應用都是看得多、用得少，不夠直接，也很少涉及機器學習、演算法等領域。

從數據科學家加入的那天開始，我就認為他們是不可或缺的，他們的工作不是今天的數據分析師和數據挖掘工程師做得了的。雖然數據分析團隊中也有數據挖掘的人，但數據科學和數據挖掘是有區別的。

我認為，合格的數據科學家應該具備產品意識，為了解決問題，會更偏重於找到問題的原因，然後找出最好的方法來解決它。所以可以說，區別就在「科學」這兩個字。

數據科學是科學，不是工程。工程追求效率，科學追求真理。如果數據科學家做得太過實作，就會變成數據挖掘工程師；如果他「飄」在空中，將無法在企業中生存。企業想要找到既能上天、也能落地的數據科學家很難。在我眼裡，一個企業如果有這樣的人，就像三國時期劉備有了關羽。但可惜的是，這世上關羽少之又少。

在這種情況下，企業應該把事情分為長、中、短線來看。頂層設計

要清楚，讓數據科學家做一些比較有難度的事情，讓能落地打仗的工程師去做落地的事情。這兩者如果沒有好好區分，他們之間就會「打架」，這樣企業耗不起。此外，從企業環境來說，一家公司到底有沒有科學精神，也會影響數據科學人才的落地。

經歷過對數據科學的理解後，回想我在支付寶做的最後一個產品「黃金策」，如果今天讓我重新做這個產品，我不會用以前的方法，因為今天的技術和過去的技術已經不一樣了。

「黃金策」給人的感覺是可以發現很多東西，但同時，能發現的東西也是有局限的。雖然有六百個或八百個用戶變量讓你自己組合，但數量還是只有這樣，能不能有六千個或八千個呢？從今天的技術來講，用深度學習的方法，有些環節機器會做得比人更好。

六年來，我所做的工作都是從人的經驗提出一個假設，然後用數據來證明哪個假設更有可能。到今天為止，由數據來告訴我們新的發現，還是少數情況。但數據帶來的這些少數新發現，是人類以前的邏輯中所沒有的。換句話說，數據重組了這個世界的「可能」。

這是人類為什麼要花很多資源去做大數據的原因所在，因為人類的思考是有局限的，但運用集體智慧，用深度學習的方法來找出人類提不出來的假設，然後透過人機協作，找出未來的新世界、新智慧，這就是數據科學要走的路。

數據科學如果不是每天都在進步，每天都不滿於今天的狀態，去PK以前的理論，就不是數據科學，而是死工程。

數據科學每天都在突破以前的理論，因為以前的理論只能解決某個點，以前的頂層設計只適合以前的情況，未來的頂層設計怎麼樣，需要我們不斷去探索，而科學，就是探索。

前言

忘掉大數據

我們正處在歷史的轉捩點上，數據技術快速變革，大數據成了人們競相議論的熱門詞彙，但鮮少有人提及，在這場巨大變革中，人們需要具備的能力。無數的企業及個人望「數」興嘆：「大數據和我有什麼關係？」

未來是大數據的時代，未來的競爭，就是數據的競爭。也許，我們早該忘掉那些華而不實的喧囂，讓大數據真正從「看」到「用」，真正「活」起來。

大數據的力量，來自觸類旁通的關聯。過去，我們總是用數據來證明或企圖說服工作上的盲點，如今數據不再是一加一的依據，而是具備了預測和開創新機的能力。

用數據找機會

我第一次見證大數據的魅力是在十七年前，只是當時我並不知道那就是「大數據」。

1990年代初期，我結識了一些以博奕為生的朋友，這些人組成了一個團隊，每年透過賽馬就能獲利數億港幣。我非常驚訝，要知道，很多人在賭馬場上可是血本無歸，而他們卻能把這種機率遊戲，變成穩定的獲利工具。

原來，他們的祕密就是使用了一套「養數據」策略——將每一場比賽的過程都錄下來。當時我覺得這個做法很奇怪，電視上已經在轉播比

賽了啊，還另外錄下來做什麼？後來我才得知，他們居然在每一場比賽中，錄下賽馬不同角度的影像。

透過這些錄影，他們分析騎師、馬匹有哪些失誤動作，這些動作會帶來怎樣的後果，然後再把這些數據「清洗」出一個更準確的數據（smart data）。賽馬過程中有許多意外，他們用數據來還原——在沒有意外發生的情況下，馬匹會因為不同場地與不同騎師，而有不同的速度。就這樣，他們可以更準確地判斷每匹馬的實力和獲勝機會，透過無聲無息的數據收集，每年數億港幣入袋。

令我最為震驚的是，他們竟然不看表面數據，而是從無限的數據中尋找核心數據。

這正體現了大數據與過往數據之間最大的不同。以前，我們是「有問題找數據」，在大數據時代，其最核心的特質是「用數據找機會」。

做大數據，必須預先判斷，哪些數據是你必須提煉出來以解決盲點的。賽馬的結果其實充滿意外，但新的數據角度幫助我們一窺真實結果，那就是「用數據還原真實」。

數據，決勝未來的商業利器

現在，大數據的概念紛繁複雜，媒體上充斥著各種關於大數據的報導，但其中不乏牽強附會、濫竽充數的言論，有些媒體甚至把簡單的統計也冠上「大數據」的頭銜。

作為跟數據打交道十幾年的人，我深深知道，從「看」到「用」，再從「用」到「養」的營運數據，本身就是一個複雜的過程，而我們目前最應該做的，是暫時忘記大數據的概念。

行勝於言，只有具備實效性的數據才是正道。我希望從實用角度撥開大數據的迷霧，告訴每個人大數據的具體運作應該是怎麼樣的，而

且，數據量不是最重要的問題，我們要的不是數據的量，而是有「質」的量，這正是我寫本書的目的。

在這個風雲變幻的數據時代，只有讓數據成為商業利器，才能決勝千里。

首先，我們需要一套具商業敏感度的數據決策框架，可以使企業看得更準，並且能夠判斷近期所做的事情哪些是對的、哪些是錯的。如此一來，快速的數據反饋，可以讓每個決策誤差得到及時修正。其次，讓數據真正從「看」到「用」，讓用數據成為建構企業生產力的重要部分。再者，讓數據技術策略深入企業的每個角落，使數據從生產、收集、使用、分享到反饋，變得簡單易用。最後，讓數據技術策略落地，還要特別注意數據的穩定、準確、時效和有效落實。

在阿里巴巴，我學習到一個很重要的經驗——人和事是分不開的。想成為一家數據化的企業，文化的培養必不可少。「混、通、曬」及「存、管、用」，這兩套內功是讓企業的血液（數據）流動的關鍵。數據流動得宜，則神清氣爽；相反，數據若出現停滯或品質問題，企業將性命不保。

鍛鍊數據力和思考力

大數據時代的到來，讓我們第一次碰上了數據化帶來的機遇和滾滾紅利，也讓我們的思維方式出現了重大改變。很多時候，我們欠缺的不是解決問題的方式，而是定位問題的能力。我們應該思考，有沒有數據可以改善我們的盲點，學會以「假設數據是可獲取的」來重新思考周圍的一切。當數據化思維成為你的直覺式思考時，就能夠融合數據的力量和思考的力量，產生新的、無與倫比的合力。到那時你會發現，周圍的一切都將因此煥發出新的生機。

最後，衷心感謝幫助我完成這個看似不可能完成的任務，寫完這本看似不可能完成的書，最終付諸實踐的老師及朋友們。感恩給予我寫這本書勇氣的雪謙寺擁珠扎西仁波切；感謝一直幫助我推動這本書前期工作的林君君；感謝用業餘時間協助我完成寫作及案例工作的趙軍、李江、吳慧敏、歐吉良、張大紅；感謝湛廬文化專業的策畫和編輯團隊；更感激家人一直容忍我的忽略。

正如數據是一種信仰，萬法唯心，每一刻的進步都來自真誠的對待。

PART 1

從「數據化營運」到「營運數據」

數據十誡

- ▶ 一切從定義問題開始，問題問好了，答案就在裡面。
- ▶ 在萬物皆數據的時代，要以「假設數據都能獲取」為前提去思考問題。
- ▶ 數據助力企業四部曲：描述現況、深入診斷、預測趨勢、指揮行動。
- ▶ 「快＋準」的數據，讓我們可以從已知規律中尋找價值。
- ▶ 「廣＋亂」的數據，給我們從發現中獲取顛覆過去規律的能力。
- ▶ 大數據不是獨奏，而是不斷連接無處不在的數據。
- ▶ 數據技術就是加速和積累（數據、分析、服務）的能力。
- ▶ 大數據生態的连接，需要建立標準與規範。
- ▶ 大數據是許多小數據的組合。
- ▶ 數據是一種信仰。

第1章

大數據，很多人只會談、不會做

有人說，大數據就像國王的新衣，每個人都在國王面前說著動聽的話，國王信以為真，其實他並不知道自己在裸奔。

的確，網路上有很多人在談大數據，但是他們只會談，不會做，因為他們根本就沒有做過，包括那些所謂的大數據專家，他們真的做過嗎？沒有。事實上，這些人對大數據內在的問題一點都不了解，更別說知道大數據的水有多深了。

目前，在大數據方面，無法深入應用的原因在於，從收集到使用的大數據價值鏈出現了問題。從理論上來說，從收集到使用的螺旋式循環是一個巨大渦輪，只有先數據化營運，才能營運數據。然而現在的情況是，用數據的人不知道大數據從哪裡來，做數據的人不知道大數據如何使用。

想用的人不敢用，因為擔心大數據的真實性；做的人不知道怎麼用，因為大數據的複雜性。這個問題造成的結果就是，數據量變得越來越大，而且越來越無法有效使用。

大數據從來不是免費的午餐

我先問一個數據管理上最現實的問題：大數據如何備份？毫不誇張地說，大數據已經這麼龐大了，如果再備份一次，成本起碼會增加一倍。

做大數據，基本上都要從大量收集數據開始，因為這些數據在未來

會大有用處，但是，你不可能無止境地收集下去。在這裡，你已經看到了一個再清晰不過的偽命題——大數據的確可以備份，但成本會增加兩、三倍。然而，「以前重要的數據一定都需要備份啊，」你自然會問，「如果不能備份，我該怎麼辦？」這就是大數據管理中必然會遇到的問題。

我再問一個問題：如果你在數據使用方面一直得心應手，整個商業鏈和數據緊密相關、相輔相成，但是，現在數據鏈忽然斷了，或是不再有效了，你怎麼辦？

需要注意的是，這個問題並不是說你有沒有好好使用數據，而是說，曾經你可以獲得的數據，現在無從取得了；或是無法在線上收集，只能從線下取得，成本也相應提高了；抑或是，這些數據儲存在其他地方，你不能使用了。在諸如此類的情況下，你怎麼辦？

當然，如果你對於數據的使用本來就很生疏，也覺得無關緊要，那麼數據鏈消失了、失效了、斷了，你也不會有太大的煩惱。但假如，在你使用數據正得心應手的時候，出現這樣的問題，你也只能束手無策，眼睜睜看著機會溜走。

大數據從來不是免費的午餐。伴隨大數據熱潮的到來，關於大數據的新問題層出不窮——大數據會夾雜虛假資訊；大數據的數據量很大，但有用的資訊不一定多，甚至還會破壞核心資訊；大數據的來源是多管道的，偏倚、隨機的誤差總是存在。

但是，我們也需要客觀認識到，大數據現在面臨的這些問題，其實就是把小數據中的一些問題放大了。小數據中難道就沒有噪音會破壞我們的核心資訊嗎？當然也有，只不過當大數據把數據量放大和變多的時候，噪音的破壞性也相應增大。小數據中難道就沒有管道偏倚和隨機的問題嗎？當然也有，但是在數據的背景，問題被更明顯地放大了。

斷層才是最大問題

說了這麼多，事實上，還沒有觸碰到如今大數據面臨的最大問題——人。

很多人都會問：大數據能帶來什麼價值？如何衡量大數據創造的價值？事實上，最直接的衡量標準就是，在經營上，它為你賺了多少錢，帶來了多少實際的利潤提升。

對於這個問題，現在很多人最傾向使用的解決方法是計算在用了大數據之後，點擊率提高多少、轉換率提高了多少。但是要知道，轉換率和點擊率提高的數據，可能根本不是投資人或公司最高管理者對大數據的期望。對業務人員來說，轉換率能提升五個百分點就已經非常好了，如果能將轉換率從2%提高到3%，簡直就是奇蹟。但是對公司最高管理者來說，這並不是他想要的大數據。

你需要認識到，斷層才是大數據所面臨最嚴重的問題。收集數據的人並不清楚未來使用數據的人要做什麼，這是目前大數據運用的一大關鍵問題。

在使用大數據時，通常是先收集數據，因為我們知道在未來的某一刻，這些數據對我們可能有用。不過，「未來可能有用」就注定會引發一個問題——收集數據的人不知道未來使用數據的人要做什麼。這時候，如果你再問收集數據的人：「如何才能更好地收集數據？」數據的使用就會陷入死循環。

事實上，不僅是收集數據的人，就連設計數據模型的人也同樣不清楚數據是如何獲得的。設計數據模型是使用數據的關鍵環節，也就是從以往經驗中找到一些潛在的規則，然後把這些規則結合起來以解決問題。

舉個例子來說，我現在身上只有10元人民幣，我可以用它買一本雜誌，也可以用它搭幾次公車，或者買泡麵充饑。這三種選擇在一般情況

下都是可行，但是，如果再加上「我沒有吃早餐」的情境，在以上三種可能性中，我選擇吃泡麵的機率自然比較大。這其實就是一個簡單的模型——輸入「擁有10元人民幣」和「沒有吃早餐，肚子餓」這兩個情境，輸出「買泡麵」這個結論，幫助我們快速選擇一條解決之路。

選擇「吃泡麵」的情境，雖然看似簡單，但同樣是經驗的總結，這個經驗就是「肚子餓了要吃飯」。把設計數據模型的過程說得更複雜或更專業一點，就是基於很多以往的經驗，進行總結，或是發現舊有數據使用規則，將經驗和數據相結合，最後輸出一個可以被套用的業務規則。

不過，此時負責設計數據模型的人不一定完全清楚數據是如何產生的，這是客觀存在的難題。就大數據研究而言，很多時候，我們使用的是他人的數據，既然是他人的數據，又怎麼能夠保證自己完全清楚呢？

關於大數據的運用，最近流行一個例子：假設我在搜尋引擎中發現某個地方搜尋「感冒藥」的頻率非常高，於是，我就斷定這個地方可能出現了流行性感冒。但這個數據是從何而來？我雖然知道很多人搜尋了「感冒藥」，卻完全不知道搜尋這個詞語的人到底是誰。提供數據的人既沒有責任告訴我數據是怎麼來的，也不必告訴我數據的品質如何，更不必告訴我數據會不會有偏差。這種情況推及到大數據領域就是，研究數據的人完全不清楚數據是怎麼來的。

數據從何而來？

當使用模型的人不知道模型數據從何而來時，模型的使用角度就要改變，具體來說，就是視情境而定。比如說，今天你要到杭州的淘寶城，有人告訴你：「今天你應該從文一西路過來。」你問原因，他回答說：「因為模型是這樣提供路線的。」他不會跟你解釋，就說模型是這

樣的，因為下午2點到4點，走文一西路比較快。

這就是使用模型的一個情境，選擇文一路則是模型的輸出。那麼，你要不要走其他路？走文二路或文三路？可是，這種選擇的結果就未知了。

對很多人來說，模型是一個黑盒子，充滿神祕性和未知性。所以，即便模型設計者將模型公開，也會因為設計者和使用者專業能力和知識背景的不同，使兩者之間出現資訊不對稱的情況，使用者自然也就不知道為什麼要選用這個解決方法。

也就是說，設計模型的人可能不知道模型的效果好不好，而使用模型的人也不知道該如何反饋使用的結果。如此一來，這種資訊不對稱會越來越嚴重。

這個問題恰恰體現了大數據實踐中非常嚴重的斷層——收集數據的人不知道將來的人怎麼使用數據，設計模型的人不知道自己採用的數據在未來是否穩定，使用模型的人也不清楚數據的來源或加工過程，這些都是普遍存在且現實的問題。

從公司管理層的角度來看，投資人了解數據的意義是什麼嗎？高階管理者對數據的期望和中階管理者又有什麼不同？他們知道數據能幫助企業做什麼嗎？這些問題的答案完全會因立場不同而異。

中階管理者大多不知道數據能幫助他們做什麼，他們沒有高階管理者的視野，相較之下，你只需要告訴他們數據能解決什麼問題即可。但數據分析師可能會更加困惑不解：「我做了這麼多東西，為什麼你們不用？」

每個層級和每個部門都是一個斷層，而且對數據價值的內在衡量都不一樣，所以，當我們講到數據價值時，沒有人能給出一個合理的定位，原因就在於有幾個關鍵問題沒有區分清楚：一是要明確這是誰心中的數據價值，投資人、高階管理者、中階管理者、數據分析師，各自心中對數據的價值都不同；二是要明確數據的分類，不同類型數據所產生

的價值各不相同。

為什麼我們在大數據應用方面存在障礙？一個重要原因便在於，應用人員對於數據價值和數據分類沒有明顯的界定。關於這個主題的探討，我會在後續章節重點講述。

更主動的管理和更多的創新

在目前的大數據環境裡，數據其實與商業模式密不可分，每個人都認識到大數據巨大的經濟價值，但今日的大數據發展趨勢之快，對很多公司而言，變得更加虛無縹緲，難以掌控，讓每個人抓狂、手足無措。

在這方面，阿里巴巴已經對數據化營運做了不少有益的嘗試。淘寶一直致力於一件事情，即用數據幫助企業營運和解決問題。在不斷使用數據的同時，也發現了數據本身的問題——大數據需要更主動的管理，也需要更多的創新。

數據化營運，就是用數據解決問題，但是如果我們想把數據做得更好，解決更多新的問題，就需要做一件從未做過的新事情——營運數據。對阿里巴巴來說，這件事情是從2011年才開始有計畫地進行，主動收集數據，並且以此創造更優質的新數據，讓新數據更好地服務企業的營運。

這是從「用數據」到「養數據」的過程，也是從「數據化營運」到「營運數據」的過程，更是從「看」到「真用」的過程。

從數據化營運到營運數據是一個循環，今天，電商企業正走到了其中一個節點上。在經歷了一開始的喧囂之後，大家終於感受到，要使大數據產生真正的商業價值，需要關注的內容並非4V^[1]那麼簡單，應該將焦點放在如何真正讓數據落地，也就是從數據化營運到商業管理能力的提升。

關鍵思考 | 問題就是答案

大數據改變了人類學習、工作和生活的方式，更改變了人類的思維方式，而我所談的數據化思維，正是從大數據的特色中產生的。西方式的思考著重系統性分析，因此對一件事情的理解過程，大多是利用數據進行細分 / 歸類、對比、溯源，以及從三者的動態趨勢中，找出變化的規律。我們可以從以下這幾點著手：

- 目的（**Purpose**）——假設我們的目的是如何讓自己更「好運」。
- 定義（**Identify**）——我們要定義什麼是「好運」，例如，可以細分為財運、健康或人緣。
- 量化（**Measure**）——要如何量化「好運」？參照前面的定義，尋找可量化的目標，例如，工資增長率、體檢中某項指標的高低等。量化之後，我們可以從趨勢中預判做什麼可以更「好運」。
- 評估（**Assess**）——執行之後，評估什麼做對了、什麼做錯了。

可見，西方式思考中的溯源，就是從事物中找出因果關係，尋找能夠有效獲得成功的驅動力。而東方式思考中的溯源，則是從本質找出根源，例如，「好運」的本質，可能在經過分析後，顯示為「快樂」，但量化出來的結果卻是「笑」。

然而，不管東方或西方，要溯源，一切答案都得從問題開始。因此，良好的提問技巧，能使我們在尋找答案的道路上事半功倍。好問題（**deep question**），可以使我們離答案更近一

步。

曾經有位知名CEO很得意地問了我的老師一個問題：「你覺得到底是eBay的商業模式好，還是Amazon的商業模式好？」他的言外之意是，今天Amazon做得比eBay好，是不是也就意味著它的商業模式比較好。

我的老師並沒有直接回答這位CEO的問題，而是反問了一個問題：「你的問題也許跟商業模式無關，產生這樣的結果，會不會只是因為Amazon的CEO做得比eBay的CEO好？」無須多言，問題就是答案。所以，對一件事情的認知，要看你是否具備了「用問題解決問題」的能力。

2011年，阿里巴巴開始從數據化營運轉向營運數據，當時，我負責支付寶的營運數據工作，我十分苦惱，毫無頭緒，於是，我打電話向老師求助，諮詢他對於這個問題我應該如何思考。

接下來發生的事情讓我非常難忘，老師並沒有直接回答我，而是問了我三個問題：

- 你所在的公司目前面對的三大問題是什麼？
- 未來三個月中，公司要解決的問題是什麼？
- 在過去的一個月裡，你做對了什麼，做錯了什麼？

「我現在給你五分鐘時間，如果你說不出公司目前面臨的三大問題，而你作為商業智慧部門的負責人，基本上，就不該繼續坐在這個位置上了。」他這麼說。

當然，我明白這三個問題裡已經包含了答案。當你要做數據化營運或營運數據時，你必須問自己這些問題。如果你沒有問自己這些問題，就去想如何收集數據的話，那肯定會出問

題，這是很關鍵的一點。

所以，思維方法很重要，在有了數據、有了對商業理解的前提下，思維方法是不可或缺的關鍵。思考的方法是無窮盡的，就像兩個武林高手過招，假設一方面對的是功夫高手李小龍，作為局外人的我，自然無法給出得勝的建議，這需要他自己動態地尋找答案。

在尋找答案的過程中，有一點很重要。很多時候，我們會受一些固有問題的影響，如果不懂得變通、不懂得提煉核心的話，就會產生思考盲點。當你的思考出現盲點，就注定問不出關鍵問題，就好比，對方明明是男性，還問他是不是男性，那就沒有價值了。

有時，有些東西並非如你所想的那樣，才會連問問題都懶得思考，就問出像前面那位CEO「eBay和Amazon誰的商業模式比較好」那樣奇怪的問題。

話說回來，在知道自己的努力有盲點的情況下，即便問題問得不好也並不丟人。但如果自視甚高，什麼問題都不問，盲點注定會相伴一生。我曾經見識過一位超級CEO，他的厲害之處就在於，他凡事都說：「我不懂，麻煩你再解釋一下。」

最重要的就是思考，儘管你懂，但如果在問問題、看問題的角度上沒有拓寬視野，想解決問題依舊很難。就好像詠春拳裡的「問手」，它不能用於過招，而是尋找答案的一種方式。你一碰，它就有答案；你不動，它就不動，而且會避開，你只能尋找破綻再出擊。「問題」是用數據來取得數據，而「問手」就是用一個問題引出另一個問題。

在大數據的商業環境裡，要既懂數據，又懂商業，還要擁有一套好的思維方法，而數據化思維正是這樣的嶄新事物。

當我們想知道問題的答案時，首先必須學會如何問對問

題，因為一個好問題會引發出很多答案，這是一門藝術。在我看來，人的思考方式在大數據充斥的商業時代尤為重要。後面的「關鍵思考」中，我將談一些有關思考的例子，供大家參考。

答案不重要，思考的角度才重要。要習得一套巧妙的數據化思考方式，三分靠想法，七分靠實踐。所以，切勿空談。

-
1. 4V是大數據的經典定義，包括海量的數據規模（Volume）、快速的數據流轉和動態的數據體系（Velocity）、多樣的數據類型（Variety）和巨大的數據價值（Value）。🔗

第2章

還原用戶的真實需求

當身邊的人都興奮地討論大數據時，你是否心存不解——雖然我們每天都在收集數據，但我們為什麼要關注數據？這些數據又如何為我們所用？

的確，當回歸到「數據能幫助你產生什麼價值」這個問題時，如果你自己都沒弄清楚，那麼就算給你再多數據，你也只會手足無措。我們身邊的大數據現象越來越普遍，數據的廣度和深度都呈現幾何級、甚至指數級增長的趨勢。在思考數據的價值時，我認為，可以從三個面向來考慮：

- 你能否清楚識別（Identify）用戶的身分？
- 你是否清楚收集的數據對你的價值（Value）是什麼？
- 你是否知道收集數據時的情境（Situation）是什麼？

識別身分，串聯散落的行為數據

很多人都同時擁有手機、個人電腦、平板電腦等多個數位裝置。比如，某人擁有兩支手機、一部iPad、三張信用卡，這個人每天都登入你的網站。在這種情況下，你能否知道這是同一個用戶？你有沒有這種識別能力呢？

銀行雖然可以透過信用卡識別靜態個人的身分，但缺少當事人使用電腦、手機或平板電腦等多元上網裝置時的身分識別能力。

這時，有誰能夠識別這三個裝置和三張信用卡背後的使用者是同一

個人？目前，企業有多大的能力識別「碎片化了的個人」，將是巨大的考驗，也會是巨大的機會。

在識別過程中，分辨使用者其實並非難事，比如，企業可以查看用戶電腦裡的cookie，獲取用戶的e-mail地址、網購時的收貨地址，甚至私人手機號碼、親密連絡人、信用卡與身分證資訊等等，這些都可以作為鑒別「碎片化了的個人」的依據。

需要注意的是，這些屬性有些是唯一的，有些卻不是，但它們依然有很大的參考價值。舉例來說，你會經常將手機、電腦借給別人使用嗎？當然不會。那麼信用卡呢？想必更不會外借了吧。

了解識別身分的方式之後，我們就可以把散落在網站內外、似是而非的行為數據串聯起來。不同的網站，收集到的數據不同，數據的價值也千差萬別。如果某個網站連接了你的三個裝置和三張信用卡，那麼，這個網站收集到的內外數據，就很容易將你識別出來。所以，是否知道用戶是誰，決定了企業數據收集行為的意義大小。

企業價值VS.客戶價值

數據收集的價值包含兩個面向。一方面看你是否能衡量這個數據對企業產生的價值——你不必將用戶的所有行為都記錄下來，而是記錄那些對企業有幫助的數據，即企業價值。另一方面則是看你能不能衡量數據對顧客的價值——這個數據將如何幫助企業，為客戶提供更好的服務，即客戶價值。

從企業價值來看，數據收集實現的是企業資源的合理分配，例如，把推薦系統做得更好，讓更多用戶可以有更高的機率找到他們想要的商品，如此就能提高商品的購買率，為企業創造更大的經濟效益。

從客戶價值來看，數據收集實現的是顧客體驗的提升，例如，便捷

的搜尋引擎，可以使顧客更容易找到他們需要的商品，創造更優質的購物體驗。

數據提供的價值，從不同面向看，會有不同的結果。比如，每天下午2點，銀行門口有很多客戶大排長龍。如果以客戶價值為核心，那麼思考的角度就應該是，每個在下午2點去銀行辦理業務的人，平均要等多少分鐘才會得到他需要的服務。但是，銀行的做法是研究大量的數據，目的是想出如何減少在下午2點時排隊的人數，這就是以企業價值為核心的考量。

以企業價值為核心去審視就是，如何幫助企業在下午2點時分流排隊的人數；以客戶價值為核心去審視則是，如何讓客戶減少等候的時間，並得到他需要的服務。

同樣一件事情，兩者的提煉不一樣。一是思考能否縮短在這家銀行排隊的時間，讓客戶提前兩分鐘享受服務；二是思考能否使用手機解決排隊問題，透過手機銀行，減少實體銀行裡的排隊人數。

我的觀點是，數據價值必須區分是從誰的角度出發。所以，當我們在談論大數據的價值時，第一點要注意的就是，角色不一樣，對數據價值的看法也就不同，因此，在衡量價值時，要考慮到接受者和給予者兩個對立面的不同看法。在這個銀行的例子中，給予者銀行看到的是，2點時有二十多位客戶在排隊；而接受者哪會對排隊人數有敏感度，他們只會覺得，若十分鐘內沒有得到服務，那就是等很久了。

從電子商務的角度看，推薦系統的做法跟銀行類似。現在，所有電商公司在考核推薦系統時，都是以點擊率作為KPI^[1]。

從企業考核推薦系統的角度來說，點擊率是主要指標，想的是如何讓用戶點擊更多自己的產品，但對於消費者來說，卻並非如此。用戶想的是：「今天我有一個很想買的東西，我能不能買到呢？」

所以，在使用數據改善用戶的購物體驗時，企業應該是僅僅改善推薦系統，還是要了解用戶整個購買決策過程的結構？

可能的答案是，企業需要知道用戶在登入網站之前的情況，而不是登入之後的情況，這就需要企業全面客觀地審視自家網站是不是有能力還原用戶的需求了。

你知道當時所有的情境嗎？

在準確識別和衡量數據價值後，我們再來看數據收集時的情境。現在，大數據的價值，吸引了很多企業和組織的興趣，因此他們非常重視數據的收集與儲存，但是，在現實操作的過程中，問題仍然層出不窮。大量的碎片化數據是噪音，讓事實串聯變得非常困難。

而值得思考的另一方面是：當我們把這些枯燥的數據串聯起來時，就一定能代表事實嗎？

例如，早上，你在上班的路上看見有個男生穿著一件非常好看的T恤，你心動了。於是，到了公司，你坐在位子上的第一件事，就是在購物網站上搜尋T恤。結果呢？出現了十萬個相關商品。

當你正在猶豫怎麼挑選時，老闆突然召集大家到會議室開會。你坐在會議室裡，發現開會好無聊，於是打開手機App，繼續想該怎麼買到那款T恤，你篩選了一些比較中意的品牌，點進商品介紹頁，但依然沒有找到你在路上碰到的那款讓你心動的T恤。

最後，手機螢幕上彈出一則手機促銷活動的廣告，你發現一款你非常喜歡的手機，價格很便宜，雖然你已經有手機了，但是你依然毫不猶豫地買下它。

那麼，在向你推薦產品時，一家公司到底要有多大的能力，才能還原你當時所處的情境呢？

搜尋引擎暴露出你目前住在上海的事實——你是最近搬過來的，網路上的收貨地址換到了上海，所以，它推薦了許多上海的賣家給你。

你在上班途中才看到那件T恤，覺得很好看，所以決定搜尋T恤。但是，在你還沒有登入購物網站之前，T恤這個因素只有你知道，電商不知道。因此，當你在搜尋引擎描述一件T恤時，你跟這家電商的第一個接觸點就出現了，所以，電商還原的情境是「早上9點半，你搜尋了T恤，搜出十萬個結果，但你一個都沒有點擊進入」。電商完全不知道，你不點擊的原因僅僅是老闆突然召集開會。

會議開始後不久，你打開手機App再次搜尋T恤，這是你和這家電商的第二次接觸。後來，在無數的品牌裡，你忘記了早上讓你心動的那款T恤，而選擇了一部促銷中的手機，有誰知道你為什麼這麼選擇？最後，你買了手機，選擇了上海的賣家。但事實上，你只是到上海出差幾個月，而現在你在杭州。

試問，一家公司有多大的能力猜到這些複雜的情境，以及每天每時每刻，每一個進入網站的用戶背後發生的特別故事？

所以，每個數據分析師都閉著眼睛說：「我不管用戶之前看過什麼T恤，反正他搜尋之後什麼都沒有點擊，可能是推薦的演算法不夠好。」

而在手機端，又有另一名數據分析師說：「這是這個用戶第一次登入行動終端，也是第一次點擊了T恤搜尋選項。」

正如你看到的，本來應該有那麼多相關性的數據，但在購買T恤的案例中，卻沒有任何相關性，那麼，企業要如何還原你的購物情境呢？每個人都在猜想可能的原因。值得一提的是，這個案例僅僅涉及用戶在一個網站上的購買行為，而沒有做過比價，在真實的購物情境中哪會這麼簡單？即使是這麼簡單的一個情境，企業又該如何還原呢？

每天都有大量的碎片化數據產生，每天我們都在做假設，那麼，用這種狀態分析數據，能有多可靠？

所以，企業在收集數據時，一定要明確知道，自己是否有能力識別用戶？是否有能力收集用戶在網站中發生的所有行為？是否有能力區分

行動端和電腦端？另外，企業又擁有多大能力看到具體的情境是怎樣的，比如，今天南方有沒有下雨？東北有沒有下雪？

歸根究柢，我想說的是，作為企業，你到底有多大的能力還原用戶的真實需求？

數據的價值，來自情境的還原

當我們在還原用戶的情境時，必須認清數據收集的領域是什麼。在不同領域收集到的數據，可以找到與所在領域不同的東西，比如，在搜索引擎和社群網站（SNS）所得到的數據就不一樣。

企業首先要做的，是確認用戶是否為同一個人，比如，在社群網站裡，涉及的資訊主要都是聊天內容，如果我是做社群網站的，我就會多去尋找這個人和其他人的關係，他今天跟張三聊了三分鐘，明天跟李四聊了五分鐘，這項數據在社群網路領域裡可以獲得。但是，當我們要真實地還原這個人的行為時，最好要有不同領域作為互補，這會讓你掌握更多、更全面的資訊。

你有多大的能力知道哪些數據是同一個用戶的，這是企業首先必須解決的問題。然後再去關注，收集到的數據量這麼大，廣度這麼寬，價值在哪裡？如果企業不清楚收集到的數據是否屬於同一個用戶的，這個數據又有什麼用？所以，在大數據裡，最重要的還是收集人的數據。

數據的價值，如同我們前面分析的，必須來自情境。對於消費者數據的收集，其中一個瞄準器就是，你能否還原用戶購買行為的情境。

科技不斷進步，如果有一天，Google眼鏡成為每個人的標準配備，或是有一天，我們買的每一部電腦都綁定手機，那麼，這兩種交叉數據就很容易獲取。

為什麼情境如此重要？情境是否被準確地表達？情境是否會成為一

件事情的背景，用來還原整件事情？

有一年的黃金週，我們發現很多用戶使用iPad購物，為什麼？你或許不會想到，因為那年第一次實行黃金週期間高速公路不收費的政策，很多人堵在路上，沒有其他事情做，所以只能使用iPad購物。如果企業在分析數據時，沒有考慮到當時整個中國高速公路都出現壅堵的情境，就無法還原整個情境，也就無法解釋這個現象。

我還發現了一個新情境。有一天，我們研究了一些行動數據，用QR Code讓用戶造訪我們預想讓他造訪的頁面。我們看到，iPhone手機在掃完QR Code後就到達了頁面，但是Android手機卻沒有。中國很多Android系統的手機在掃完QR Code後，都無法自動連結頁面，這時，Android手機就成為用戶登入網站購物的重要情境。

不管是iPhone、三星，還是其他品牌的手機，手機的大小和系統的相容性，都會對情境產生巨大的影響。如果我們盲目地觀察數據本身，就自以為是地認為用戶沒有點擊，那就大錯特錯了，事實上，是他點不了。

可見，很多看似無關緊要的東西都在情境裡，而在行動終端的世界裡，這個情境又增加了很多其他的東西，這都需要我們仔細判別。

當我和數據分析師聊天時，我總是會特別提醒，在研究行動數據時，要特別注意用戶在每天移動的時間點和非移動的時間點做了什麼。從起床到睡覺，有幾個時間點基本上是固定的，就可以判斷每個人一天的行為變化。

我曾經在一次電商大會的圓桌論壇上提出一個觀點，當時與會者都很認同。我建議把一些以前用來觀察用戶忠誠度的框架，比如RFM模型^[2]，拿來作為收集數據的瞄準器。有什麼數據能夠讓我更清楚地看到用戶近期的購買行為、消費頻率和消費金額？

RFM模型是數據收集的一個面向，而電腦、手機、平板電腦，則是另一個終端情境的面向，電腦更容易收集近期購買行為，手機更容易收

集消費金額，這樣就可以透過情境的不斷變換來收集更多數據。

現在，一些終端的確可以收集到以前收集不到的數據。過去，我們不知道一些數據的收集背景是不是移動的，但是現在可以。用戶做一件事情的時候是不是正在移動？他是不是在某個百貨商場裡？

面對不同的情境，我們的框架也要相應改變。所以，現在做數據分析報告，最後的一個問題變成了：「裝置變了，這個報告的結果還是一樣嗎？你的報告是不是也應該變一下？」

為未來的數據大戰做好準備

我認為，數據的本質就是還原，這是收集元數據（meta-data）的關鍵方法。如果沒有這個概念，你就不知道未來需要什麼數據，就更不懂得什麼是重要數據，最後只會產生越來越多無從辨別的數據。一旦數據多到公司沒有辦法處理時，那麼，其他會處理的人、公司和國家，就會毫不留情地把你擠出市場。

未來一定是國家和國家之間的數據大戰，公司與公司之間的數據大戰！

資訊時代催生了海量數據的出現，這個世界上，每時每刻都在產生大量的數據。大數據已經不再是一個單純的概念，而是像空氣一樣圍繞在每個人身邊，每個人都是數據的製造者。

也正是因為每個人都在透過不同的裝置生產數據，使得數據在「量」這個維度上不斷地膨脹，但是，「量」的單純膨脹，卻對企業真正了解一個用戶的需求產生極大的挑戰。所以，如何識別各個裝置的使用者是否為同一個人，如何更好地理解用戶在不同情境下表現出來的不同需求，以及數據在融合後產生的價值，將是未來商業中，每一個企業都必須考慮的問題。

在不久的將來，隨著O2O（線上到線下）的深入和穿戴裝置的興起，企業和企業之間必須進行更多的數據融合與交換，進行更多的跨產業數據交流，這樣才能更精準地還原用戶真正的需求，讓用戶在任何情境中，都能獲得數據帶來的便利。

總而言之，更深化的數據連接，使海量數據經過提煉，更真實地還原了事實，也使我們運用數據科技解碼未來的需求成為可能。

關鍵思考 | CEO關心哪三個數據？

我在面試數據分析師時，一定會問他們一個問題：「假如我是一家知名電商的CEO，現在是星期一早上9點，請提供我三個數據指標，向我證明，在過去的一週裡，企業營運一切正常，好讓我心裡踏實。你認為，會是哪三個指標呢？」

絕大多數應徵者對這個問題的回答比較一致：第一是流量，第二是交易量，第三是其他，包括轉換率、交易額等。

當他們這樣回答完後，我會反問他們：「剛剛我問的問題，你真的聽清楚了嗎？」

這時候，有人會回答說，聽清楚了，答案就是這三個數據。

往往這時候，我會再提醒應徵者：「請注意，我要的數據是給CEO看的，而且還是超級電商的CEO，時間軸是一週的敏感性數據。」

面試進行到這個環節，我就發現大部分面試者根本聽不懂「CEO」的含義。事實上，既然是CEO，就意味著他是公司的最高領導層，那麼，給他看的東西顯然要與其他人不同。

在這個例子中，我們會發現，絕大多數的應徵者很少會换位思考。也就是說，他們都是從自己的角度思考，而不是以數據分析師，一個要給CEO彙報三個數據指標的分析師身分來思考問題。

那麼，什麼是以數據分析師的身分思考問題呢？通常，在我提出問題時，作為數據分析師的你，首先要想的是，CEO會關注什麼數據，是長期的，還是短期的？是風險最大的，還是風險一般的？或是最近發生了什麼事情？以及提供給CEO的數據有哪些注意事項等等。

所以，我再問問應徵者：「當你坐在面試桌對面給我答案的時候，有沒有想過，在星期一早上，這家知名電商的CEO真正想看的是什麼？」

再想想這個問題，你到底有沒有聽清楚「CEO」、「知名電商」、「一週數據」這些關鍵字？CEO要的是「踏實」——他聽完就可以安心吃早餐了。

在面試時，如果應徵者不針對這幾個問題提出詢問就貿然回答，滿分是10分，我只會給5分。因為這個問題裡面就有很多問題，比如，什麼是踏實？踏實是一個概念，你不問清楚踏實的含義，就給我三個指標，無論如何都是錯的。

在正常情況下，首先，不要急於回答我提出的問題，而是先問清楚什麼是踏實，切勿自己先做假設。以下，我們可以假設一個相對理想的面試情境。

你反問：「什麼是踏實的狀況？」

我回答：「最近這家電商和另一家電商在打價格戰，而它最近又新推出了圖書項目，那麼，CEO最關注的自然這些圖書業務做得好不好。」

你再問：「什麼是好？是基於每天新增用戶和原有用戶購

書的數量？CEO是希望用書來吸引新用戶，還是想透過圖書業務的推廣，讓現有用戶交叉購買？」

在這些思考結束之前，你絕對不能給出指標。因為，在沒有解決一個問題的內涵之前，任意給出的指標，必錯無疑。所以，我才會問應徵者到底有沒有聽清楚問題。

在我做面試官的經歷中，很多看似有經驗的數據分析師，往往在我提出的問題還沒有解釋清楚時，就搶著作答。絕大多數人在思考不到一秒鐘的時間就給出答案，而這一秒鐘的答案，我可以確定他們根本沒聽清楚我的問題。

通常這個時候，我會再給他們一次機會，問他們：「剛才你給我的這個答案，如果我給你滿分10分，你會給自己打幾分？」此時，大部分的人只會打6~7分。

當應徵者給自己打7分時，我會反問：「另外3分丟掉的原因是什麼？」他開始反思，說剛才給的指標可能不是CEO想要的，因為他對這家電商的近況並不是很了解……

接下來，我再反問：「剛才我的問題是，假如我是知名電商的CEO，現在是星期一早上9點，請你給我三個數據指標，證明在過去的一週裡，企業營運一切正常。你聽清楚了嗎？如果你確定自己清楚了，能請你再給我一次答案嗎？」

這時候，聰明的人不會再用一秒鐘就給我答案了，而是重新思考，開始問問題，再給出答案。這時候的答案，當然會比第一個答案好得多。



最後，當我再問他：「現在，滿分10分的話，你給自己打幾分？」此時，他們給自己的分數通常都會高一些。至此，我的面試也結束了。

事實上，關於這個問題，我根本就不關心打分數的結果。當然，如果評價是10分，那就不用面試了，因為沒有仔細考慮

過答案，就自信滿滿地回答，這種人必然無法承擔數據分析師的責任。雖然自信是對的，但是思考更重要。

作為數據分析師，如果你不把自己的分析與當下結合，是無法進步的。

有趣的是，當我把這個問題貼在網路上時，還是有很多人追問我答案是什麼，CEO關心的到底是哪三個數據。這時候，我真的很想說，答案不是結果，方法才是。

-
1. KPI（Key Performance Indicators），關鍵績效指標，是衡量管理工作成效的一個重要指標。 
 2. RFM模型，指透過客戶近期購買行為（Recency）、消費頻率（Frequency），以及消費金額（Monetary）這三項指標，來描述一個客戶的價值。 

第3章

「活」的數據，才是大數據

無論是企業還是個人，無論我們承認與否，在大數據時代，我們周圍到處充斥著碎片化的數據，為此，我們很容易迷失在海量的數據之中。我們盲目地投資大數據，收集越來越多的數據，然而，令人沮喪的是，這些數據卻是「死」數據。

那麼，什麼是「死」數據？在我看來，死數據就是單純儲存在數據庫中，無法被分析和使用，並且無法產生價值的數據。

要如何啟動這些死數據，讓整個大數據活起來，成為企業營運的牽引力？答案就是：把數據用起來！

大數據的真正價值，是將數據用於形成主動收集數據的良性循環中，以帶動更多的數據進入這個自循環中，並應用於各個產業。

舉個最簡單的例子，現在很多網站都有推薦功能，很多推薦的東西，不論是音樂、影音，還是商品，都可以讓用戶選擇喜歡或不喜歡（當然，不點擊未必代表不喜歡）。如此一來，企業就可以透過用戶的選擇，基於電腦後台的演算法，為用戶重新推薦。這就形成了一個循環——從已有的數據進行「分析—推薦—反饋—再推薦」的過程。

當然，自循環遠不止這一種形式，多樣化的自循環方式，打開了大數據之門，而進入這個循環的關鍵就是，從解決問題出發。

目前，企業對於數據應用仍抱持半信半疑的態度。多年來，企業因無法建立數據收集的循環，導致營運數據大多建立在直覺的判斷和分析基礎上。當面對周圍海量的消費者數據時，充滿了危機的大數據，更難為企業的運作提供清晰的思路。

無從下手，成為企業面臨大數據時代的核心弱點。這時，如果沒有

找出相應的關鍵解決方法，企業就會在由海量數據構成的新興市場中，錯失發展良機。

在數據的自循環中，有兩個核心關鍵：一是「活」做數據收集，二是「活」看數據指標。

所謂「活」做數據收集，就是指企業不要局限於只收集自己用戶產生的數據，還要把別人的數據收集過來，進行綜合分析。

「活」看數據指標，則是指企業不要局限於已有的數據框架，應該結合用戶需求的不同情境，靈活應用收集到的活數據。

抓住相關性，靈活收集數據

關於數據收集，中國研究大數據的著名學者、電子科技大學的周濤教授，曾經分享一個非常重要的觀點。他說，做大數據收集不外乎兩方面，一方面是自己用，即用外面的數據來增加自己手上數據的精準度；另一方面是給別人用，即把我的數據貢獻給外面需要的人，提高他的數據的精準度。

Amazon的數據收集有一個很經典的例子，恰恰顯示了這一點。在很多年前，Amazon就主動收集用戶的IP位址，然後從IP位址破譯出在用戶所在位置附近多少公里內是否有書店。工作人員從收集到的數據中了解到，一個人是否選擇在網路上買書，很重要的原因是他附近有沒有書店。

Amazon是主動收集數據，即透過收集外部數據，來幫助自己判斷線下是否存在潛在競爭對手。

企業在做數據收集時，並非總是能直接收集到需要的關鍵數據，這時候就需要變通一下了。我曾經在一個外貿B2B（企業對企業）平台工作，當時很多人不知道我是如何找出新品項的趨勢，我也沒有對任何人

說過我的招數，有的人甚至以為我背後有高手相助。

事實上，我之所以發現新品項，是透過追蹤美國電商廣告的關鍵字變化。

中國製造商跟美國消費者距離如此遙遠，我們如何才能收集到美國消費者的數據？一般來說是不可能的。有的廠商是從中國海關購買數據，但海關的數據是落後的，無法讓企業及時獲得產業的發展趨勢。不過，這種趨勢就在美國購買關鍵字的那些經理手上，這個數據可以從國外網站搜尋到，於是答案就這樣到了我手上。

在美國，這些搜尋引擎行銷（SEM）專家會為公司設計購買關鍵字的方案，我主要就關注這些人的數據，看他們換了什麼關鍵字，以了解美國的商品發展趨勢。這樣，我就可以很快地告訴品項經理，美國現在流行什麼樣的東西，這就是利用外部數據來幫助自己做決策的方法。

以鞋類為例，具體來說，我的做法是先觀察在美國做得最好的幾家鞋類網站買了什麼關鍵字、變換了什麼關鍵字、哪些關鍵字是常態、哪些關鍵字是在季節更迭時才購買。觀察到這些關鍵字之後，我放到Google去觀察關鍵字的增長趨勢，到eBay看看這一款鞋子有沒有交易及價格的變化。在知道eBay有交易之後，我再放到中國電商平台中搜尋這個產品，如果沒有出現相應的搜尋結果，就意味著這可能是一個良機。

所以，做大數據收集，有時候需要更多的靈活變通。從上面的例子可以看出，Amazon的案例的確經典，因為它找到了消費者購買決策過程中的一個關鍵點。B2B平台尋找關鍵字的重點在於，如何觀察第三者數據。每個人都知道，在收集消費者數據時，最好是觀察直接用戶，但如果沒有這個數據，你觀察什麼數據？答案就是，去觀察業界對這個數據最敏感的那些人。

生活中其實也有這樣的例子，我曾經和我的朋友分享了一個簡單且有趣的例子。如果你想知道香港某家酒樓生意好不好，你問問門口賣報紙的人就知道了，因為香港人喜歡去喝茶時買一份報紙。

其實，這個規律不是我觀察到的，而是香港稅務局發現的。香港稅務局如果擔心酒樓虛報營業額，就可以去查賣報紙的商家賣了多少份報紙來判斷，這是一個非常有趣卻相當實際的靈活收集使用者數據的案例。

「活」做數據收集，就是要跳出既定思維的框架，從相關的產業和業務中，收集能夠為現在所用的數據，找到能夠更好地佐證企業現有業務決策和發展的數據。「活」做數據收集的一大好處就是，能夠規避現有數據框架的弊端，更真實地反映用戶的實際需求和企業的實際情況。

媒體常用的「克強指數」是活做數據收集非常好的案例，透過耗電量、鐵路運貨量和銀行貸款發放量這三大數據的結合，來觀察經濟的發展，有助於剔除GDP統計數據中的差距。

結合情境，動態使用數據

我們不僅要靈活地收集數據，還要注意到，收集數據只是第一步，如果不讓數據活起來，僅僅把收集的數據堆砌在一起，是沒有意義的。

阿里巴巴的十八個性別標籤

消費者在註冊網站時，常常需要填寫自己的性別，我填寫的答案是男性，但如果你分析我的購買行為，你會發現，很多時候我的購買行為告訴網站，這些商品的目標用戶並不是我自己，因為我也會為我的妻子和父母買東西。

當收集到的數據不能為企業所用時，企業就永遠不知道關於我的這個數據原來是不準確的。這些數據看似準確地描述了我的性別，但是卻不能準確地描述我的搜尋和購物行為。

在梳理阿里巴巴的數據時，我們驚訝於為什麼阿里巴巴會有十八個

性別標籤，乍聽之下很不可思議。你肯定會想，阿里巴巴是不是在閉門造車，自己憑空捏造出這麼多性別？

事實上，每個性別表現並非看上去的那麼簡單，而是基於用戶在不同情境中的不同表現而做的分類。這就揭示了一個問題，我們每個人都不會只呈現出簡單的一面，比如在獨處時和在人前時，我們就會表現出不一樣的自我。不同的性別標籤其實就是應用了這一點——同樣的人在搜尋商品時可能會表現出不一樣的行為特點，這些不一樣的行為就是我所說的情境。結合情境以應用數據，就是「活」用數據。

Amazon的動態模型

其實，有多少個性別標籤並不重要，重要的是如何讓用戶在不同的情境中，獲得更好的服務。

Amazon一直在自己的商業活動中應用這個理論。多年以來，Amazon就是使用動態模型——相對於「歷史的你」，它的推薦更相信「現在的你」。所以，它相信今天登入網站的你有什麼需要和興趣，比起歷史的你更重要。

我們不妨擴大一下思路，假設你在網站上買了一份保險，又買了一件衣服。這兩件事情，哪一個更能證明你是「本我」，而哪一個更能證明你是「被別人影響的我」？

推薦系統的運作原理，是透過推薦一件商品去影響你的購買決策。一般來說，當你要為自己做一個長遠的購買決定時，會綜合考慮一些因素，這時的購買決策會跟你自己的性格產生緊密相關。比如，你的性格是不愛炫耀，不愛刺激的活動，那麼你就不會買一輛跑車，這是你深思熟慮後的決定。

對於用戶購買行為的情境，我們大多無法判定它的產生，到底是因為長期積累的性格多一點，還是受今天心情的影響多一點。而且前提

是，沒有其他人用你的帳號購物。如果今天是你的女友借了你的帳號購物，這個系統可能就無法做出正確的推薦了。

讓數據活起來，從靜態數據變成動態數據，必須用情境來驗證，靜態數據是沒有用的。理所當然，企業很容易一開始就假定「你告訴我身分證上你是男的，肯定就是最準確的」。從表面上看，企業根據用戶自己填的數據來做決策會更精準，但結果卻相反。在做了很多測試後我們發現，動態地觀察數據，更能準確反映出使用者的資訊和需求。

但是，當用戶在做一個比較重要的購買決定時，真實性別、真實職業，以及一些描述個人的靜態數據，對企業就更有具參考性了。對消費者來說，購物就是決策，凡是決策，都是經對比、細分、溯源、看趨勢。所謂的衝動消費是什麼都不管，直接買下，那就要分析購物當天的情境，或許那天是雙11購物節？

別再死盯著投資報酬率

從公司價值的角度來看，我在B2B平台得到一個重要經驗——如果企業不活用數據，只是死盯著一個數據的話，企業的營運肯定不會理想。

當時，我任職的公司在購買廣告關鍵字時，一直以來用的重要考核數據，是投資報酬率（ROI），即每個廣告關鍵字能為公司貢獻多少交易量。那麼，如果我們靈活使用數據，會怎麼樣？

我當年就對公司的管理者提出一個建議：把一些用得很好（排行前30%）的廣告關鍵字拿出來，調查根據這些關鍵字造訪網站購物的消費者，三個月後是否還會來買東西。

事實證明，這些用得很好的廣告關鍵字，有一半只吸引了消費者一次性的購買行為。因為，這些關鍵字都是「最便宜的XX」，對某些消

費者來說，這些關鍵字有著很大的吸引力，但並不一定會產生巨大的效用。

也因為這樣，企業吸引的都是一般消費者，而不是真正在做批發的人。企業管理者完全不知道，當時掌控投資報酬率的部門，其實是在做賠本生意。

所以，機械式地使用數據的結果就是，企業會因為只看狹義投資報酬率而死得很慘。投資報酬率只能體現廣告關鍵字在當天使用的效果，並不能反映出這個關鍵字及吸引而來的用戶對網站的長期價值。

當時，公司把所謂最好的廣告關鍵字刪掉之後，管理者發現，三個月內，企業的業績並沒有受到任何影響。

活用數據就是，你能否看出這個數據的局限是什麼。一方面要看，我們的數據為使用者體驗改善了什麼；另一方面則是要看，企業在使用數據時，解決了什麼問題或開創了什麼機會。要牢牢記住，活用數據很重要。

活的數據，是活用數據的精髓所在。《孫子兵法》曰：「兵無常勢，水無常形，能因敵變化而取勝者，謂之神。」這句話在數據應用方面同樣適用。如果企業能夠基於情境和相關的活數據，將數據應用發揮出最大的價值，那麼，開創新的商業模式，也會在不遠的將來成為可能。

關鍵思考 | 別再做「碰巧遊戲」

在開始之前，我先講一個有趣的故事。有一天，全球最大的專業人士社群網站忽然發現，從雷曼兄弟公司登入的來訪者變多了，但是並沒有深究其原因。第二天，雷曼兄弟宣布倒閉。

原來，前一天雷曼兄弟的員工都到專業人士社群網站找工作了。

Google宣布退出中國的前一個月，我在LinkedIn發現一些平時很少露面的Google產品經理上線，這也是相同的道理。

試想，如果LinkedIn能夠針對某家上市公司，選擇性地分析數據，會不會挖出其中潛在的商業價值呢？我相信，很多網站現在還停留在「註冊用戶數」這種簡單的數據收集，而不知道還要收集上述有價值的數據資訊，使企業獲得發展。

好的分析師，相當於市場解碼器

我常常說，今日的數據分析師，需要有點軍師的味道——從枯燥的數據中，看到解開市場發展密碼。簡單來說就是，一個具商業意識的數據分析師，如果監測到網站上有6歲以下孩子的用戶群增加，那麼，他基本上就可以預測奶粉的銷量也可能呈上升趨勢。

從更深一層來看，和傳統賣場一樣，網站中，每項商品的作用大不相同，有的產品為了賺錢，有的產品為了促銷，有的則是用來吸引流量，這也就解釋了不同商品在網站上擺放位置不同的現象。同理，只有具備商業敏感度的數據分析師，才會懂得用什麼數據來驅動公司實現經營目標，絕對不會單憑交易量就決定商品策略。

評估新電商平台，你會觀察什麼數據？

要評估一個新電商平台的競爭力，你會觀察什麼數據？交易量？不對，因為平台是雙邊市場，不同於單邊的B2C（企業對個人）賣場，除了買家流量之外，對於一個需要活力的電商

平台來說，每天有多少新的優質賣家加入也至關重要。比如，他們賣了多少東西和賣了什麼東西，這對新電商平台的競爭力來說，都有著不可或缺的作用。因為，此階段的商業競爭，最核心的影響因素是人氣，而非實質交易量。如果新加入的賣家賣不出東西，只有老賣家的交易量在增長，即使最後的交易量每天都呈增長趨勢，也無法大幅提升競爭力。

換個場景看，一家剛踏入市場的B2C公司和一家已經占領大部分市場的B2C公司，兩者的關注點會一樣嗎？當然不可能，因為前者看的是流量，賺人氣；但流量對後者的意義就沒那麼大了，因為它更關注成交率、轉換率和回頭率。

組裝你的數據儀表板

我把數據資訊的梳理過程，理解為數據世界裡的「儀表板」。比如，在開車時，如果水溫過高，儀表板就會亮燈提示，同樣，網站的關鍵資訊也可以監控，在電商交易中，行為數據和商業模式之間的邏輯關係組成了儀表板——好的儀表板，對於商業活動中的好現象和壞現象，都會敏銳地反應。

為了方便理解，我想出兩個名詞：前端行為數據和後端商業數據。前者指的是訪問量、瀏覽量、點擊率及站內搜尋等，是反映用戶行為的數據；後者更則側重於商業數據，比如交易量、投資報酬率、顧客終身價值（LTV）。

分析數據「走火入魔」的人會明白，數據就像散布在夜幕裡的繁星，彼此之間布滿了關係網，只要輕按其中一個亮點，就會驅動另一個亮點產生變化，而這個動態的網，就是儀表板。

當我們將行為數據與商業數據對接起來時，大家可能會比

較關心網站的用戶群，以下就是一個相關例子。

某天，某網站發現前端註冊量增加不少，訪問量也提高，但交易量卻未見大幅度的提升。原因是什麼？想必數據分析師們每天都在思考這個問題。那麼，我們就來分解這個問題，尋找相應的答案。

處在網站前端的人只知道點擊量等數據，比如，誰一直重複購買？誰影響了5%~15%核心用戶群的購買行為？誰在傳播網站的正面或負面評價？然而，很少有人關注後端的商業數據。

如果操作網站後端交易環節的工作人員只知道賣東西，卻不知道提出這幾個問題：一個客戶登入網站，平均停留時間是十五分鐘，還是三十分鐘，對將來重複購買的影響大嗎？一個客戶登入網站論壇和沒有登入論壇，對交易量有影響嗎？此時，很少有人關心前端數據。

你發現這中間的問題了嗎？行為數據與商業數據沒有對接起來。

於是，前後端數據形成分裂，沒有人知道它們的內在關係。作為網站的決策者，不知道網站核心用戶群的行為特徵，也不知道怎麼去刺激核心用戶增加，更不知道從一個用戶登入網站之後到他離開，中間哪些環節需要疏通，不死也是萬幸。

當然，這個例子也只是管中窺豹。一個平台營運商，反映使用者行為的前端數據與後端的商業數據是千千萬萬，賣家和買家也是千千萬萬，但是，在了解前端哪一個數據，對整個網站後端的交易量有最大的影響後，就只須針對這個前端數據下工夫，以刺激後端數據的增加。

反過來，後端哪一個交易數據比較高，弄清楚是從哪一個管道收集的、主要貢獻用戶是誰，網站的產品設計就要向他們

傾斜，如此才會提升管道前端的轉換率等關鍵數據。

如果一個網站的核心用戶群，每月以10%的速度增長，不爆紅也是怪事。

遺憾的是，今天許多電商公司，每天都在做「碰巧」遊戲，今天推薦A家的產品，明天又撤下A家的產品；今天做低價促銷，明天又做線下活動。這些決策的改變，沒有儀表板的指示或良好的監控，都是閉著眼睛在碰巧。

世上哪有這麼多碰巧，還是快睜開眼睛，認真看路吧！

第4章

行動數據，大數據的顛覆者

2013年被稱為行動元年，智慧手機的出貨量超過了三億部、4G牌照的發放、微信等應用的快速崛起、阿里巴巴全力進軍行動平台的發布，這些現象和事件無不意味著行動時代的到來。行動數據不斷改變人們的生活，也不斷改變數據資訊的全貌。現今的大數據環境，與之前相比，變得更加複雜，也為數據的分析帶來了諸多挑戰。

讓我們再來回想一下第二章描述過的、很多人都有可能碰到的場景：

早上你在路上發現一個路人身上穿的一件T恤很好看，於是，在進到辦公室的第一件事情，就是打開電腦，登入網站搜尋這件衣服。還在挑選的時候，突然被老闆叫去開會，開會時覺得有些無聊，拿出手機，打開App繼續搜尋，後來被App上的手機促銷活動吸引，最終買下了一部手機。

這個看似簡單的情境，從數據分析的角度來說卻非常不簡單。如果按照過去的做法，我們用電腦的視角分析這一數據，會發現這個用戶搜尋了T恤，瀏覽了T恤的幾個頁面，但在幾個小時後卻買了一部手機。這不僅僅會讓數據分析師感到困惑，更讓整個基於瀏覽和購買的推薦系統產生疑問——此種行為會將T恤和手機這兩個內容產生一種相關，但事實上這兩類東西顯然是不相關的。

這時候，我們需要的是一種「多螢」思維——需要考慮到現在使用者使用網路的多情境問題。現在，「多螢」可能會包括多部電腦、手機和平板，在可預見的不遠將來，可能會有更多的穿戴裝置出現，例如，用智慧眼鏡搜尋一件衣服，用智慧手錶錄下一段語音，搜尋某一種好吃

的零食。當多螢變成一種常態時，不管是數據分析師的分析方法，還是推薦系統的推薦演算法都必須多螢化。因為當多螢時代到來後，一切數據都在發生變化.....

行動數據帶來的「噪音」

在探討行動數據讓數據產生「噪音」的問題前，有必要稍微談一下關於數據收集的方法。個人電腦時代的網路，數據收集是基於瀏覽器。從1990年代開始蓬勃發展，在這二十年左右的時間，網路已經有比較成熟的數據收集方案——透過cookie識別用戶身分。

如果是一個擁有用戶體系的網站，就能夠以用戶ID與cookie的關聯性，識別用戶的行為。在電腦時代，用戶在使用瀏覽器時，cookie通常都是長期存在，而且對於大部分用戶而言，很少會在意它的存在，因為cookie對用戶的意義不大，對網站的意義才大，這是由於網站可以透過它，長期收集使用者的行為數據。一個電子商務網站能夠知道使用者看了什麼、什麼時候看、看了多長時間、結果買了什麼.....這一系列的情境都能夠瞭若指掌。這種收集數據的方法，讓網站能夠識別用戶，而這也是研究用戶行為的第一步。

可是，行動時代到來後，一切都發生了變化。原本可以收集使用者完整行為數據的網站變得不再可靠，而數據還原使用者行為的方式，也發生了改變。為什麼行動的出現會讓這樣的情況發生呢？

在這裡，簡單講解行動數據收集的原理。行動基本上分為兩種形式，一種是WAP（無線應用協議），另一種是App（手機應用程式）。對WAP來說，整體的工作原理和電腦差不多。在行動網路早期，用戶可以透過手機的瀏覽器，造訪專門為手機設計的頁面。

現在行動網路的主流應該是App，是基於手機端的應用程式。App

的發展速度非常快，尤其在最近幾年。隨著iOS和Android的高速發展，整體的App用戶也在快速增長，但是，數據收集的技術卻沒有像電腦那樣成熟。App數據收集的方式包括兩種，一是收集用戶連網時請求伺服器的紀錄，二是將使用者的行為數據記錄下來後，適時傳給網站。但不論是什麼樣的收集方式，行動數據最終的表現是在沒有帳戶體系的情況下，和電腦的用戶行為完全沒有辦法連結，這也就意味著使用者的數據出現了斷層。

如果只收集行動裝置的使用者數據，是不是就不會出現斷層問題呢？答案也是否定的。在個人電腦時代，雖然瀏覽器各有不同，但是基於cookie的整套數據收集體系很成熟，不同的瀏覽器不會對數據的收集造成很大影響，但是行動App時代就完全不一樣。App對用戶的識別是基於手機的個別代碼，但是這一識別標準在不同的作業系統上也存在差異。這種差異主要體現為以下兩點：一、操作系統本身的差異；二、同一作業系統上的版本分化。

作業系統的差異比較好理解，現在主流的操作系統iOS、Android、Windows Phone差異性就很大，安全性原則也非常不同，比如，iOS的封閉式系統嚴格控制外部收集者的介入，限制獲取使用者的資訊。而版本的分化，對數據收集形成更大的挑戰，例如，iOS6的作業系統允許App獲取手機的MAC位址（媒體存取控制位址），到了iOS7就無法獲取了，只能獲取廣告識別碼，但用戶又可以限制獲取，這就導致基於MAC地址識別使用者的方法，在使用者升級作業系統之後完全失效。

Android在這方面的表現更加誇張。Android是一個開放的系統，不同的廠商，甚至其他組織都可以修改它，這就使得在一部分手機上可以使用的方法，在其他手機上就行不通。典型的例子就是透過手機掃QR Code、打開應用，用小米手機掃描就能打開，而三星手機也許就不行。

以App而言，有很多App會內嵌網頁，這時往往會收集到App本身的數據和網站瀏覽的數據，使得數據變成兩份。如果沒有順利傳遞App內

的識別使用者標記，即使在手機端，數據也會變得不完整。

再退一步講，即便上述兩點都沒有對數據收集產生影響，App的版本更新也會對數據收集造成障礙，例如，一個App開發的時間長了之後，會更新多個版本，而現實中也總有不願意更新版本的使用者，這使得數據收集的結果會有很多份，而不像電腦時代，只產生一份數據。比如，淘寶行動版的最新版本可能是4.0，但手機預裝的版本可能還是3.4，在這種常見的情況下，數據分析師可以從更廣的視角來看待數據收集的問題。

電腦數據與行動數據的關係

在兩份數據同時存在的情況下，就需要衡量電腦和行動的關係。我的看法是，保證電腦和行動兩份數據的完整，透過使用者體系將兩份數據連結起來，就可以在分析時，用彼此的融合，還原使用者行為。

首先，從數據來看，電腦數據和行動數據，由於技術上的差異，兩者本身就是分離的，這就意味著資料倉儲（Data Warehouse）需要用額外的空間來保存行動數據，而不能單純將行動數據混入電腦數據裡。而且，行動數據中，還存有App性能的數據，例如後台啟動、程式當機等，所以分開儲存是有必要的。對於不需要還原使用者行為的數據，比如，整體頁面的點擊率、用戶造訪時間長短等，基於應用程式本身的數據，則可以透過行動數據分析，效率更高。

其次，對於數據收集和管理策略而言，必須要有一套帳號體系，能夠使原本沒有關係的數據發生關係。由於技術不同，電腦數據和行動數據天生就不會發生關係，要讓關係發生的唯一辦法就是人，或是用戶的帳號。這對於有帳號系統的平台，比如騰訊和阿里巴巴，不會是多大的問題，但是，對於不依賴帳號系統的平台，想讓數據之間發生關係十分

困難。為了讓用戶的行為能夠串聯，現在很多App都要求用戶登入，這樣即使用戶換了手機，用戶的行為仍然能延續，而企業也會清楚此用戶是老用戶。否則，在換手機如換衣服的今天，可能所有用戶手機的生命週期都不到兩年，那麼，企業就永遠都不會知道誰是自己的老用戶了。

多螢思維下的電子商務

面對行動數據，要解決的主要問題有：

- 1 做到高效準確的收集
- 2 培養數據分析師的多螢思維

所謂做到高效準確的收集，就是需要建立一套多面向的作業系統、多版本、自動、可靈活配置的使用者行為獲取系統。透過這個系統，開發工程師不再需要採用人工寫代碼的方式收集數據，也就保證了數據的純淨無噪音。但是，這套系統還需要應對一個問題，那就是因操作系統的變化，導致使用戶識別失效的情形。因此，企業需要建立自己的使用者識別方式，獨立於裝置號以外的識別代碼。同時，這套系統還需要做到多系統相容，要保證同一個版本在Android和iOS上的數據，具有一致性。

培養數據分析師的多螢思維，則是一個長期的過程。我的方法是，先派出精鍊的數據分析師小團隊，去探索行動數據的情境，因為即便像阿里巴巴這樣的大公司，行動數據的分析也是一個全新領域，所以，需要有分析師去探索。經過探索後，我們就能發現行動數據的各種問題，為清晰的數據分析打下基礎，從而更有效地還原使用者行為，以便分析在多螢環境下的電子商務。

最後，在數據儲存領域，也需要更合適的模型框架來應對海量的行動數據。在行動出現之前，資料倉儲都是基於個人電腦的思路，按照主體所建立，比如，使用者屬性、瀏覽行為、訂單中心等；在行動出現之後，很多數據倉庫工程師，將行動也作為一個並列於使用者屬性或瀏覽行為的主體保存。

其實，這種做法的誤解是，沒有真正理解行動的含義，行動應當是作為一種橫向的基礎模型，穿插於各個垂直的主題模型之中，如果將行動作為一個獨立的模型，在還原使用者行為和分析用戶時，會變得很困難。

古往今來，每當顛覆者出現時，機會也會隨之出現。行動網路是當今商業發展不能避開的鴻溝，此時，企業應該做的是，在顛覆者行動之前，提早做好準備和規畫。在行動時代和多螢時代，企業需要設計更好的技術方案，以更容易識別使用者，用更好的產品情境，讓使用者享受到多螢的便利，必須及早培養數據使用人員的多螢思維，及早規畫面對行動情境的數據底層設計方案。

從數據化營運角度來看，雲是計算，端是應用。「雲+端」的模式，使數據科技的價值得到極大的發揮，就像「INTEL inside」（英特爾處理器品牌認知計畫）一樣。從營運數據的角度來看，端是感測器，雲是「儲存+管理」，而「多螢+行動」就是未來數據創新的孵化器。

在未來的商業大戰中，積極備戰行動領域的公司，將會在行動網路的情境下獨占、超越其他公司的優勢，並享受更為豐厚的收益。

關鍵思考 | 留意樣本的偏見

假設，你想買一款新手機，於是去詢問了三位朋友的建議

——到底是買iPhone，還是買三星？如果我們把這種經歷當作一次抽樣調查，我確定，它絕對不可能告訴你真實的結果。

我們在做抽樣調查時，樣本往往從一開始就存在最常見的偏見——誰願意回饋我的問卷，誰的建議就被反映在調查結果中；不願意回饋問卷的人，他的意見就不會反映在調查結果裡。有趣的是，那些被我們忽視、沒有回饋的樣本，很多時候卻是我們最需要知道的結果。

比如，今天淘寶做了一項調查，詢問：「你愛淘寶嗎？」收到一萬份回覆為「愛」，於是，我們就得出「大家都愛淘寶」的結論。但事實上，我們總共發出十萬個樣本，這結果說明，有九萬個樣本對這項調查保持沉默。此外，我們在做調查時，會贈送禮物給被調查者，有些人是因為收到了禮物，才覺得應該說點好話。

在了解這樣的情況後，你還會輕易得出「大家都愛淘寶」的結論嗎？所以，當我們分析一份數據時，一定要問自己：樣本的背後有什麼是我們不知道的？

我們經常會看到易觀、艾瑞、Comscore等機構公布某一事物的調查數據，在看到這樣的數據時，我常常會問，他們的樣本是怎麼挑選的？據我所知，他們有一個第三方監控軟體，以每月數百元的費用「購買」樣本，把監控軟體置於後台。試想，如果你想做一個高端用戶樣本的調查，那麼，這種做法可行性很小，就算有可能，結果也不會準確。因為他們即使每個月出1500元人民幣的資金，也無法獲取準確的數據。原因顯而易見，要月薪5萬元人民幣以上的用戶把軟體放在後台，監測自己每天做了什麼，他們當然不會答應。所以，在這種取樣結果中，注定有一部分數據缺失，最終導致結果不正確。

在現實生活中，我們往往不會去想採樣的問題，造成我們

對很多調查結果深信不疑，我們的決策就會有對號入座的嫌疑。這種嫌疑極為常見，比如，某人告訴你，他身邊很多朋友都是天秤座，這是因為他心裡能想到的朋友很多都是天秤座，事實上，其他他沒想到的人可能是別的星座。

再舉兩個現實中的例子來解釋樣本的偏見。某人說，他開車兩千次，從沒出過意外。今天，他喝了酒，但那兩千次裡並沒有酒後駕駛的樣本，所以，兩千次這個數據在今天是沒有參考意義的，因為無法證明他喝完酒之後還能安全駕駛。但如果這兩千次中有兩百次喝了超過三瓶白酒，而且喝完之後還開了五十公里的路回家，那麼，這兩千個樣本就有參考性了。

樣本數量是否足夠和是否平衡的問題，是另一個常見的偏見。

某雜誌對用戶偏愛的手機作業系統做了一項調查，結論是，現實中使用Android系統的人比較多。但是，這本雜誌的讀者主要是IT業界人士——這個取樣一開始就存在偏見，而抽樣又來自讀者，再加上IT業界人士也比較傾向接受Android系統，所以，樣本的不平衡，就造成了結論的失真。

另外一個影響樣本的因素是時間。阿里巴巴有一個360度調查，用於反映員工對部門的評價。如果我想拿高分，我會在調查前請大家吃飯，幫大家培訓……做好調查中列出的每一個項目。要相信，短期記憶是十分強大的。這就是典型的時間樣本，這樣的抽樣是典型的似是而非。那麼，如果想消除時間因素對結果的干擾，該怎麼辦？其實很簡單，你只要隨便抽一天詢問員工：「你的部門好嗎？」這往往才會得出有價值的答案。

此外，偏見還會因為我們抽樣時的詢問方式而出現。有一個很好的例子，Amazon的調查就讓人信服，因為它的問卷第一

句話是：「你的親朋好友裡，有沒有Amazon的員工？」如果你回答「是」，那麼很可能你的意見就不會出現在這份樣本的結果裡。Amazon的員工、親友，和一般消費者的觀念，很可能存在差異。Amazon的問法，使得調查得到的數據更精確，也讓存在偏見的樣本有了意義，因為它能區分兩者的差異到什麼程度。

在調查過程中，如果這個樣本在你調查前就是有目的的，那麼，這份存在偏見的數據也是有價值的。比如，你可以透過交叉資訊來驗證結果是否準確。這在我們在生活中也很常見，舉例來說，你問三個朋友，到底是買iPhone，還是買三星。但首先你得問：「你是用什麼手機？」如果他是用iPhone，卻推薦三星，那你就要追問為什麼了。

同樣，我現在詢問淘寶小二：「你選擇在哪裡購物？」他們熟悉淘寶，必定會受到影響。如果針對淘寶小二進行調查，結果顯示更多人願意去京東、當當購買商品，這就證明京東和當當確實不錯——起碼就某些品項來說是不錯的。此時，存在偏見的樣本對企業仍然有價值。

值得注意的是，樣本跟大數據不同。大數據相信全量數據，而非樣本；是分析得出，而不是抽樣獲得。

第5章

數據分類與數據價值——找出你的核心數據

在大數據時代，企業首先要做的是收集大量數據，但收集數據，並非僅是把收集到的數據放到硬碟裡面那麼簡單，更重要的是對數據進行分類、存放及管理。不然就如同一個雜亂的儲藏室——東西放進去時很輕鬆，等到要把東西找出來時，就會浪費大量的時間，甚至可能再也找不到。

如今，數據在企業中已經充當了一種生產原料的角色。既然是原料，就要被估值。

從數據角度來說，估值就是透過不同面向去思考數據的價值。只有基於對數據分類和數據價值的認識，才能對數據做篩選。如果連有多少種數據都不知道，我們很難描述數據收集起來後，到底要如何進行篩選，也不知道哪一個數據更有價值，更不知道哪一個數據需要更安全的處理，甚至不知道在它的生命週期中，備份的策略是什麼。

對數據的認知，完全取決於我們是否能夠認知自己所擁有的數據，是否能夠篩選出「到底什麼才是我的核心數據」、「到底什麼數據會被我們頻繁地使用」。

數據分類為什麼重要？

在大數據時代，每做一件事情，我們都要知道切確的目的是什麼。從個人的角度考慮，當我們為數據分類後，真的就更容易理解大數據嗎？

「數據」本來就是一個既清晰又模糊的概念。之所以說清晰，是因為每個人對數據都有概念，而且每個人每天都會接觸到各式各樣的數據；而說模糊，則是指數據包含了各種不同的類型，能夠產生完全不一樣的價值。

至於數據到底該怎麼分類，則需要視不同的情境而定。就如同「人類」，如果拿人類和動植物做比較，人類必然有不同於其他物種的特性，這時，人類就是一個非常清晰的概念。但是，如果要對人類做更深層次的分析，顯然需要對不同類型的人進行比較和分析，在很多情況下，可能還需要對不同種族和不同年齡層的人進行分類和分析。

數據作為一種資產，不同的數據，含金量必然不同，自然會產生不同的價值。而且，就同一組數據而言，在不同的環境下，甚至會呈現出不同的價值。就好比衡量黃金和鐵這兩種金屬的經濟價值，同樣重量的黃金，其價值必然遠勝於鐵。但如果是要用來做刀劍，柔軟的黃金顯然就不是合適的材料。黃金和鐵的合金可能又在另一個場合有其他的應用。

對數據而言，不同的情境，也會產生不同的價值。有些數據可能會變成一個營運指標，能夠讓管理者更容易做業務決策；有些數據可以用來預測未來；不同的數據有機結合之後，又能夠更精準地為用戶推薦商品。一個好的工匠，必然對各種材料的性質及使用方法瞭若指掌。而一個好的數據從業者，必須對各個數據的價值和穩定性洞若觀火。

遺憾的是，並不是每個數據從業者都能夠妥善地對數據進行分類和估值。很多時候，我們可能只看到數據的一些應用情境，卻無法更加深入地進行數據的管理和應用。

當我們將焦點放在一組數據能夠產生什麼價值時，往往會忽略其生產過程，因為很多數據是經過加工才產生價值的，但往往加工鏈上的中間數據卻未被區分出來。就好像我們在吃一盤番茄炒蛋，好吃是結果，但如果問番茄為「好吃」貢獻了什麼，雞蛋又貢獻了什麼，你可能就回

答不出來了。

古語有云：「治大國若烹小鮮。」其實做數據也要有這樣的感覺，尤其在大數據時代，更要做到精細化。

還是拿番茄炒蛋這個例子來解釋。我們這次端出這樣一盤番茄炒蛋，那麼，下次是不是還能保證口味完全一樣呢？如果要做到口味完全一樣，我們應該怎麼做？當然，你可以說這個過程已成為一種習慣，是一種感覺。但是，一旦需要做到標準化，或者所有人都要做到一樣，與習慣相比，就不如找出當中的規律了。

這個過程是一個定量分類和管理的過程，也是一個標準化的過程——需要多少番茄和多少雞蛋。如果再精細一點，還要看番茄的產地、雞蛋的養殖場、番茄的熟度怎麼樣、每顆雞蛋要多少克等等，所有這些內容都要有精細的規定。

在運用數據上，這樣一個精密的過程就更不可少了。比如，這次可能運用了某個模型，效果不錯，採用的這組數據效果也不錯，那麼，下次是不是還能保證有類似的效果呢？或者說，換了情境後，還能保證類似的效果嗎？這些都是需要認真思考的問題。

數據分類的四大面向

今天，我們多數人把自己關注的焦點放在結果上，而忽略了過程。事實上，這些過程其實也蘊含了數據管理和數據分類的過程。有些關鍵數據必須做好保護，如果這幾個數據變了，或者被污染了，前面的價值也就無法保證。

有人肯定會問，管理數據是一個過程，能不能分解？能不能區分哪些元素比較重要？哪些有代替品？哪些無可取代？哪些是不可或缺的核心？其實，這些問題就包含數據戰略。

所以，我們不能含糊應對，必須更妥善地為數據做歸類和整理，從而更清晰地識別數據的價值。我們可以從四種面向，為數據進行分類。

面向1——以是否可以再生為標準

按照是否可以再生的標準，數據可以分為不可再生數據和可再生數據。

不可再生數據通常是最原始的數據，比如，用戶在造訪網站時，瀏覽紀錄會追蹤用戶的行為，如果當時沒有記錄下來，就沒有其他數據來還原用戶的行為了。這有點像拿相機拍閃電，抓拍很重要，一旦錯過，就不可能再重複剛才那瞬間的光影。因此，針對用戶日誌等不可再生數據，必須要有完善的保護措施和嚴格的權限設置。

現在，很多系統都有備份多份數據的功能，理想情況應該是，因為磁片損壞而造成數據遺失的案例應該越來越少，然而，因為系統升級失敗和操作失誤等，造成數據遺失，在各家公司都屢見不鮮，見怪不怪了。

可再生數據，則是透過其他數據可以生成的數據。原則上，指標類數據的衍生數據都是可再生的——只要原始的不可再生數據還在，就可以透過重新運算而獲得。

不過，千萬不能因為聽到「可再生」就掉以輕心，有些可再生數據是經過很長時間的累積，不斷加工而成，是長時間從海量數據中計算出來的，比如，某個用戶在數個月內的連續購買行為所產生的規律。如果未做保護，雖然可再生，但是再生的時間會給企業帶來問題。因為即便是具備頂尖計算能力的公司，都可能要花費數日，甚至是數週、數月，而這個時間過程可能就會對公司的某一項核心業務造成毀滅性的打擊。

以不可再生數據來說，已有的數據要嚴格保護；想要、但是還沒有的數據，就要及早收集。例如，很多電商網站並不關注客戶在商品介紹

頁有沒有滾動螢幕。如果這類數據沒有記錄下來，企業就無從得知商品介紹頁的有效性。當頁面要進行改版時，需要對比這類數據進行參考，就無法獲得相應的數據支持，最後，只能等數據收集到了之後，再做決策，這就造成了決策的延誤。

對可再生數據而言，要及早做好業務的預判和數據處理的規畫，這樣一來，在需要數據的時候，就能快速應用，我們把這一數據叫作數據中間層。

面向2——以數據的儲存層為標準

按照數據的儲存層來看，數據可分為基礎層、中間層和應用層。

從數據的儲存角度來說，數據有很多層次。基礎層通常與原始數據一致，也就是僅僅儲存最基本的數據，不做匯總，以盡量避免失真，從而作為其他數據研究的基礎。中間層是基於基礎層加工的數據，通常也被認為是數據倉庫層，這些數據會根據不同的業務需求，按照不同的主體存放。應用層則是針對具體數據問題的應用，比如，作為解決具體問題的數據分析和數據挖掘的應用層數據。

在儲存層這個層面上，最大的問題就是數據的冗雜和管理的混亂。尤其是對一些擁有海量數據的大公司而言，數據的冗雜問題尤為嚴重，造成了大量的浪費。在大公司中，進行數據分析、開發、挖掘的人，可能有數十、甚至數百人，這些人可能歸屬於不同的業務團隊，為了滿足不同的業務，各自分析數據應用。如此一來，不同的人可能都從頭建立起一套包含基礎層、中間層和應用層的數據，彼此之間又沒有合適的交流方式，也就造成了工作的浪費。

那麼，是不是應該把所有的數據做更好的歸納或管理呢？我認為這也不是絕對的。任何管理方法，無論是集中式管理，還是分散式管理，都各有利弊。而且人和業務多了之後，企業也很難進行集中管理。

我的建議是，基礎層必須統一，因為這是最基本的數據，而且是原始數據。除了備份的需求以外，沒有必要在各個場合保留多份數據。只要保證這個數據有良好的元數據管理方式，就能大大降低成本。至於中間層和應用層，則要視具體情況而定，如果公司的業務相對單一且成本壓力比較大，建議集中式管理；如果公司的業務量非常大，則可由多個數據團隊進行分散式管理和應用，以確保基礎層單位有最高的靈活性。

面向3——以業務歸屬為標準

按照數據的業務歸屬來區分，數據可以分為多個數據主體。按照業務歸屬分類的意思就是，將數據按照不同的業務主體，分門別類進行歸納。就好像倉庫一樣，將不同的物料分類存放，可以提高使用和管理的效率。

按照業務歸屬分類的數據，在不同的公司可能體現出不同的內容，在平台型電商，可以分為交易類數據、會員類數據、日誌類數據等。交易類數據是指平台型電商的訂單編號，其中包含買家、賣家在什麼時間成交了什麼商品；會員類數據記錄了買家、賣家的身分資訊，例如註冊時間、身分證字號、信用等級等資訊；日誌類數據則是指用戶的行為，即哪個用戶在什麼時段造訪了平台的哪個頁面，點擊什麼按鈕等。

數據的分類主體，要根據業務的特點進行歸納，並沒有一個特別的硬性規定。總體原則就是讓數據的儲存空間更少，分析及挖掘的過程更簡單、更快速。

面向4——以是否為隱私為標準

按照是否為隱私來區分，數據可以分為隱私數據和非隱私數據。

顧名思義，隱私數據就是需要嚴格的保密措施來保護的數據，否則會對用戶的隱私造成威脅。例如，用戶的交易紀錄就屬於隱私數據，對

一家擁有良好數據管理機制的公司而言，一般的管理方法是對數據的隱私級別進行分層。

從安全的角度，數據可以分為兩種類型、四個層次。兩種類型就是企業級別和用戶級別。企業級別的數據，包括交易額、利潤、某大型活動的成交額等；用戶級別的數據，就像是身分證字號、密碼、用戶名、手機號碼等。四個層次則分別是公開數據、內部數據、保密數據、機密數據。

當然，也有隱私數據保護做得不好的企業，之前很多隱私外洩的案例，都對用戶造成了很大的損傷。比如，某些網站幾十萬筆的訂房資訊外洩、數百萬個帳戶密碼外洩等，都是類似的故事。

隨著擁有大量數據的網站和公司越來越多，數據安全成為核心，需要投入專門的人和專門的團隊來管理數據安全。推動數據安全，初期往往會受到一線員工的反對，因為任何一個安全系統，都意味著已有的許可權被收回，也會因為改變工作方法而降低效率。

擁有大數據的企業高層主管必須關注數據安全，否則，數據越大，對惡人的吸引力就越大，最終，用戶和公司的損失也越大。

數據的五大價值

如果說，數據分類的目的，是為了對數據有更好的認知和管理，那麼，對不同數據價值的認知，就是讓數據被更好地應用的前提。如果僅是收集數據而不運用，那麼，除了占用儲存空間、增加儲存成本以外，沒有任何價值。但是在實際運用中，也需要認清數據到底能夠產生什麼價值。有時候，同一組數據可能會在不同場合產生完全不一樣的價值；有時候，單一的數據沒有什麼特別的價值，需要組合起來才能產生價值。

價值1——識別與串聯價值

顧名思義，識別的價值，肯定是唯一能夠鎖定目標的數據。最有價值的數據，比如身分證、信用卡、e-mail、手機號碼等，這些都是識別和串聯價值很高的數據。

前面多次提到的搜尋T恤的例子中，電商網站識別用戶的方法，就是登入帳號。千萬不要小看這個帳號，如果沒有這個帳號，網站就只能知道有一些商品被瀏覽了，但是卻無法知道是被哪個用戶瀏覽，更不可能還原用戶的購買行為。

當然，識別用戶的方法不止登入帳號一種，識別用戶的傳統方法還包括cookie。cookie是在你瀏覽器裡面的一串字元，對網路公司來說，這就是識別用戶身分的一個標記（當然也可以選擇不被追蹤）。所以你會發現，你在搜尋引擎上搜尋過一個詞語，之後在很多網站都會看到相關資訊或商品的推薦，這就是透過cookie實現的。很多網路公司都非常依賴cookie，所以會採用各種cookie來記錄不同的使用者類別，單一的cookie沒有價值，將使用者造訪不同頁面的行為串聯起來，才產生了核心價值，這就是串聯價值。

然而，現代社會已經不再是單純的「電腦＋瀏覽器」的時代。2013年被很多人認為是行動網路元年，之後幾年，人們上網的方式快速地轉移至「電腦＋手機＋平板」的形式，網路已經進入了多螢時代。在資訊多元化時代，單純運用cookie識別使用者行為的方式已經過時了。為了解決這個問題，企業開始使用一定的帳號系統對用戶進行設限，在各種場合提醒用戶使用手機號碼註冊，或用手機號碼換取更多的權益，本質上就是希望能夠在多螢時代「認」出用戶。

網路是一個自由開放的社會，很多用戶可能因為不同目的而註冊多個帳號。如果單純從帳戶的角度看，會被認為是多個用戶；但是，從現實的角度看，他又是實實在在的同一個人。在一般的情況下，擁有多個

帳號可能不會是什麼特別的問題，但是，一旦涉及反欺詐等需要識別身分的情況時，就必須識別出來。在這種情況下，識別用戶，就需要靈活處理串聯數據，比如一些隱密的識別和串聯數據——密碼。密碼並不能讓企業認定兩個帳號是同一個人，但是，如果兩個帳號密碼很類似、甚至完全一樣，就極有可能是同一個人。為了方便，很少人會給自己不同的帳號設不同的密碼。

當然，身為用戶的你大可放心，因為在大公司中，你的密碼都是被加密過的，沒有人知道你真正的祕密是什麼，對比密碼的方式也僅僅是對加密後的一串字元做比較。

在美國，另一項非常重要的數據是車牌號碼，也對識別身分具有重大意義。據說，美國FBI對人和車的跟蹤是分開的，他們認為車牌號碼很重要，和人有非常高的相關性，能串聯起個人的許多資訊。

如果你想知道日常生活中有哪些很有價值的識別和串聯數據，可以回想一下，你遺失信用卡時，打電話到信用卡中心，對方會問你的問題。一般來說，對方會問你的出生年月日、住家地址、電話號碼等問題，這一系列問題就是把你的個人數據做識別和串聯。因為銀行必須確認是不是本人，這時候，生日、電話號碼是有權重的。可能在有了兩、三個這樣的數據後，即使你沒有密碼，銀行還是會相信你，為你重新辦卡。

我在營運數據時，經常觀察這樣的數據，有時候能夠直接識別，有時候需要間接識別。生日、電話號碼，這些都是非常重要的數據，因此，我認為有些人把自己的生日放在社群網站上，是非常不妥的做法。

所以，千萬不要小看識別數據的價值，我的經驗告訴我，能夠識別關係和身分的數據是最重要的。這些數據應該是有多少存多少，永遠不要放棄。在大數據時代，越能還原用戶真實身分和真實行為的數據，就越能夠讓企業在大數據競爭中，保持戰略優勢。

價值2——描述價值

還是用T恤的例子來說明。你搜尋一件T恤，搜尋的是什麼品牌、材質、尺碼等，這些都是描述數據，可以用來刻畫研究對象。研究對象可以是商品，可以是企業，可以用戶。反過來說，描述數據也可以幫助我們更理解研究對象。

在網路上，我們經常會聽到很多關於好男友的標準，比如，身高170～180公分、體重65～75公斤、月收入5000～10000元人民幣、不抽菸等，這其實就是將好男友這樣一個感性的指標數據化，而這裡用到的數據，就充當了描述研究對象的作用。

在一般情況下，描述數據是以一種標籤的形式存在，它們是經過初步加工的一些數據，這也是數據從業者平時最基礎的工作。一家公司一年的營業收入、利潤、淨資產等數據，都是描述性的數據。

在電商平台這類企業的日常經營中，描述業務的數據包括成交額、成交用戶數、網站的流量、商品介紹頁的流量、成交的賣家數等，我們就可以透過這些描述業務的數據，觀察交易活動是否正常。

但是，對企業來說，數據的描述價值與業務目標的實現並非成正比關係，也就是說，描述數據不是越多越好，而是應該收集和業務緊密相關的數據。例如，一家兼有電腦平台和行動平台業務的電商公司，在電腦平台上，可能更關注成交額；而在行動平台上，應該更關注活躍用戶數。

對業務人員來說，描述數據能夠讓他更了解業務發展的狀況，讓他們對日常業務有更清楚的認知；而對管理層來說，經常關注業務數據，也能夠讓他更了解企業的發展，做出明智的決策。

描述數據最好的一種方式，就是分析數據的框架。在複雜的數據中抽出核心，讓使用者能夠在極短的時間裡看到經營狀況，同樣，又能夠讓使用者看到更多他想看的數據細節。分析數據的框架，是對數據分析

師的基本要求——基於對數據的理解、對數據進行分類和有邏輯的展示。一般來說，優秀的數據分析師都具備非常好的數據框架分析能力。

價值3——時間價值

如果你不是第一次在電商網站上買東西，你的歷史購買行為就會呈現出時間價值。這些數據已經不僅僅是在描述T恤了，還展示出在這一時間軸上你曾經買過什麼，以便讓網站對你將要買什麼做出最佳預測。

在考慮了時間面向之後，數據會產生更大的價值。針對時間的分析，在數據分析中是非常重要的、但往往也比較有難度的部分。我們可以透過以下兩個案例做更進一步的說明。

第一個案例是電商A首頁的廣告，是一個關於茶葉的廣告，但這個廣告並不是對每一個人都顯示，對我顯示的原因是我之前在電商A買過茶葉。這個數據的價值就是，廣告系統透過對我的歷史數據的判斷，將我識別成一個可能會買茶葉的人，並透過廣告系統的內在競價機制，推薦我這一款茶葉。

第二個案例是關於一個品項的演進路線，其中存在一些很有意思的現象。比如，烘焙用品會有一個明顯的演進路線，奶油、模具、起司、包裝盒等用品，會隨著烘焙周邊產品的演進，不斷出現新的需求。這就是在大數據的基礎上，基於用戶的時間，來對產品的演進進行分析，讓消費者更快找到需要的商品。

大數據其中一個非常重要的作用就是，它能夠基於大量歷史數據進行分析，而時間絕對是代表歷史的一個面向。數據的時間價值，是大數據運用最直接的體現，透過對時間的分析，能夠歸納出用戶對於一種情境的偏好。知道了用戶的偏好後，企業就能更精準地為用戶推薦商品。

時間價值，除了體現歷史的數據之外，還有另一種價值是「即時」。網路廣告領域的即時競價（RTB），便是基於即時的一種運用。

即時競價就是當使用者進入某一個情境後，各家需求方平台（DSP）就會開始競價，對用戶現實情境進行數據推送。比如，用戶正在瀏覽和一個化妝品有關的頁面，或者正在商場逛街，在這個情境中，就會出現和化妝品有關的資訊。這個化妝品廣告不是預先設置好的，而是在這個具體的情境中，透過即時競價才出現。

價值4——預測價值

數據的預測價值分成兩種。第一種是對某一項單品進行預測。在電商領域中，凡是能夠產生數據、能夠用於推薦的，都會產生預測價值。比如，推薦系統推薦了一款T恤，它有多大的可能性被點擊，這就是預測價值。

預測價值本身沒有什麼價值，它只是在估算這項商品是有價值的，可以讓你對未來可能出現的情況做好準備。比如，推薦系統估算今天會有十個用戶買這件T恤，這就是預測。再追加一個問題：你有多大的信心今天能賣出十件T恤？你說有98%的可能性，這就是對未來的預判及準確度的預估。

網路上很多情境都會運用對單品的預測。在電商A中，哪些商品能主推，哪些商品不能主推，都可以透過數據的預測價值來呈現。比如，電商A的團購活動，每天有大量的賣家和商品想報名，那到底該如何選擇呢？在這個過程中，就要透過這個賣家和商品的歷史數據，預測某項商品能不能賣到一定的交易額，如果能賣就能上，如果無法達到一定的交易額，那就不能上。

第二種預測價值，是對經營狀況的預測，即對公司的整體經營進行預測，並且能夠用預測的結論指導公司的經營策略。在今日的電商中，行動端是一個重要的部門，對於新的行動業務來說，核心指標之一就是每天的活躍用戶數，而且這個指標也是考核行動團隊的重要依據。

作為行動團隊的負責人，到底如何判斷現在的經營狀況和目標之間存在多大的差距呢？這就需要對數據進行預測。透過預測，將活躍用戶分成新增和留存兩個指標，進一步分析兩者對目標的貢獻度分別是多少，並分別對兩個指標制定相應的產品策略，然後分解目標，進行日常監控。這類型的數據能夠對公司整體的經營策略產生非常大的影響。

價值5——產出數據的價值

從數據的價值來說，很多數據本身並沒有特別的含義，但是在組合幾個數據，或是整合部分數據之後，就會產生了新的價值。

在電商中，這樣的情境很常見，比如，在電商開始初期，很多人都會關注誠信問題，那麼，要如何評價誠信呢？於是就產生了兩個衍生指標：一個是好評率，一個是累積好評數。這兩個指標，就是目前在電商平台的頁面上經常看到的賣家好評率和星鑽級別，用戶能夠藉此了解這個賣家的歷史經營狀況和誠信。

但是，僅以這兩個指標來評價賣家，又顯得略微單薄，因為它們無法很精確地衡量賣家的服務水準，於是，又衍生出更多的指標，像是與描述相符、物流速度等，這些指標最終變成了一個新的指標，叫作店鋪評分系統（DSR），可以用來綜合評價這個賣家的服務水準。

當然，單一商品在電商網站上可能會出現幾千條評價，而評價又是用戶站在自己的立場描述，但是推及到某個用戶上，如果每次買一樣東西都要閱讀幾千條評價，顯然是不太可能，因此，就需要把這些評價重新定位，以產生出新的、能夠幫助用戶做出明智購買決策的數據，這些數據就是關鍵概念的抽取。我們看圖2，就知道如何讓評價的數據產生新的數據。



圖 2 電商網站中會影響用戶購買行為的幾大關鍵概念

在認識數據的分類和明確數據的價值後，我們就更能識別出哪些是我們想要的核心數據，從而讓數據發揮更大的作用。精細的數據分類，嚴格的數據生產、加工過程，將讓我們在使用數據時受益匪淺。

關鍵思考 | 用傻瓜的視角去觀察

有一次，我在廣東看舞獅，突然有人跑來問我：「老師，中國有獅子嗎？」是啊，中國並沒有獅子。中國有老虎，沒有獅子。在日常生活中，我們對舞獅或石獅已經司空見慣，所以沒有人會去思考這樣的問題。那麼，中國是什麼時候開始舞獅，什麼時候開始在門口擺放石獅的？於是，我到處去查證。

事實上，換一種視角去觀察，我們可以避免理所當然的盲點。這讓我想起微信創始人張小龍，他曾經說過，在做微信時，他有一個祕密武器——他可以一下子把自己變成傻瓜，從傻瓜的角度去研究產品的設計方法。

我們每一個人，即使是成天在嘴上說自己不聰明的人，多少都有些自戀，也會認為自己是聰明的。其實，許多慣性思維會讓我們變成真正的傻瓜，因為你永遠不會去思考中國有沒有獅子這個問題，也可能會設計出一個只有你自己才懂得如何使用的產品。

比如，我給你一張白紙，命題的關鍵字是「貝殼」，你會

很快地把與命題相關的東西填滿。因為從小老師就教我們背書，背得越快越好，誰快誰就聰明，在這裡，快速填滿相關的東西就是聰明。

如果我們從思考這個角度出發，這種做法雖然看似聰明，但其實很笨。因為從反向的角度來看，你填得越快、越熟練，或許會越快成為一個只會複製的機器，最終失去自主思考的能力。

那麼，當我們用傻瓜的視角觀察問題時，我們又應該用何種視角做判斷呢？我的建議是，不要以短期利益為重，眼光要放長遠，要利及眾生。

在做決策時，不應該只以短期利益作為考量標準。試想，一個為了當前利益而斤斤計較的人，長此以往，怎麼會有真正的朋友？

像傻瓜一樣觀察，用利他的眼光做判斷，這用一句話總結來說就是一大智若愚。先是若愚，後是大智。

我覺得「若愚」的意思就是：我知道我不知道。我舉一個例子，記得我的前老闆有一個習慣，他在提問前總會說：「我很笨，剛剛我沒聽懂，你能再重複一遍嗎？」

把自己放在一個很低的位置去提問，這是很聰明的做法，但在現實中，對很多人來說卻很難做到。特別是在擁有了一定的地位、能力後，人們就更不願意承認「我不知道」。但你要知道，只有當你知道自己不知道時，才能理所當然地說：「對不起，我不懂，你能再重複一遍嗎？」

在我所在的公司高層裡，還有一位非常聰明的人。有一次開會，某個同事在講他的規畫，說到一半，CEO就打斷他說：「等等，你說的是什麼？是說我們的伺服器不夠用，還是說有很多伺服器不能用？」總之，還有一堆後續問題等你解答，這

讓很多人覺得CEO是個令人討厭的人。但他堅持的理念是，讓問題在清晰的思路下執行，這是依託在非常仔細的觀察和解答完所有疑問後才做出決定的「傻瓜習慣」。

另外，當我們看得長遠以後，很多事情就顯得公平許多。同樣道理，在人生裡，擁抱大愛，學會幫助別人，你會發現，整個世界的時間彷彿都被拉長了，很多東西都變得公平了，而選擇也變得容易了。

第6章

從「用」數據到「養」數據

過去，有一些問題一直困擾著我：「現在企業獲取數據如此容易，數據的增長速度如此之快，那麼，對於企業來說，到底要收集什麼樣的數據？要收集多少數據？收集數據的邊界在哪裡？」

數據收集要有目的性

後來，我在美國遇到一位高人，他認為，過去收集數據很難，而現在獲取數據資源變得越來越容易，但是，如果收集數據的出發點不是為了解決問題，那麼，收集再多的數據也沒有意義。

同時，許多企業還有一個疑問：「現在收集數據不難，成本也不高，為什麼不先收集了再說？等以後需要數據來解決問題時，再拿出來用，不是也可以嗎？」

這位高人同樣也給出了答案，他對此持否定觀點，並指出用這樣的理念來設計數據應用，注定會失敗。

數據收集沒有邊界，我為此也痛苦了好一段日子。比如，收集一個人的生日，雖然可以精確到幾分幾秒，但這麼精確的數據又能用在什麼領域，產生什麼價值呢？

事實上，數據是有生命週期的。舉例來說，一家電商的婦幼網站會主動收集用戶的寶寶資訊，包含寶寶的生日、性別、小名、身高、體重，這裡，我們就要清楚這幾個數據的用途和生命週期。

例如，寶寶的身高和體重，媽媽填寫的是當下的數據，但我們知

道，寶寶的身高、體重變化非常快，也就是說，這兩個數據的生命週期很短，在當時的推薦中可能有用，但過了一段時間後，這兩個數據就失效了。而寶寶的生日和性別這兩個數據的生命週期就很長，可以從生日的年分推算寶寶的年齡，而性別基本上是終生穩定的。

保存數據及數據收集時的背景，也是一件不容易的事情。還是以收集寶寶的年齡為例來闡述，我們可以透過用戶購買特定的商品集合，例如適合不同年齡層的奶粉和尿布等，知道寶寶所處的年齡層，以推算出寶寶目前的年齡。但如果僅僅是保存寶寶的年齡這個數據，也會很快失效，因為人的年齡是動態變化的，因此，你還需要同時保存獲得這個數據的時間，即在哪年哪月計算得知寶寶的年齡，這個資訊就是背景數據。

另一種更加聰明的做法，即透過寶寶目前的年齡，反過來推算寶寶是什麼時候出生的。

所以說，在收集數據時，我們必須知道這些數據在未來可以用來做什麼，如果今天想像不出來的話，日後就更不可能了。

如今，很多電商高階主管會詢問數據分析師，商品的重複購買率是多少，於是，工作人員收集數據，計算重複購買率，卻很少想到，高階主管需要知道重複購買率，是要做什麼決定，這就好比刻舟求劍的故事。事實上，在世事多變的大數據時代，我們不能只是機械式地套用方法或指標。

重複購買率有不同的定義，而做不同的決策，就需要考慮不同定義的重複購買率。如果是一家投資公司想收購A公司，就會從重複購買率來看A公司的營運優劣或用戶品質等。如果是從A公司本身營運的角度，則應該更關注日、週級別的重複購買率變化趨勢，或者當月新增客戶中，有多少人在三個月後重複購買，從而衡量出每個月新增及既有客戶的忠誠度和品質，找出改善的空間。

在知道了以上的背景之後，再去選擇使用什麼數據，不是更有目的

性嗎？

數據應用講求小而美

2011年底，作為數據分析師，我開始思考怎麼從「用數據」，轉變為「養數據」，即從數據化營運，轉變為營運數據。這段時間，我為應該收集什麼樣的數據而特別煩惱，我也曾試圖做出一個特別大且適合多數人使用的數據應用，但是後來發現，這在數據應用的起步階段幾乎是不可能的，一是因為要找到可以解決大部分人需求的數據應用並不容易；再者，當時公司的數據非常豐富，需要考慮的因素很多，因素之間的關聯性又很複雜。

在開發數據應用時，數據就等於原料，當原料一直在變化的情況下，做出來的產品就很容易出問題。體會到數據和應用的關係後，我最後決定從小角度切入，先把小應用做出來，這就是很好的瞄準器。

這裡所說的「小」，不是指數據量，而是指應用的目標很具體。

很多人在沒有獲取足夠數據，並且對數據缺乏理解的情況下做出決策，其實是在「享受」自己的無知。舉例來說，對於一款數據應用，如果我的目的是分辨兩種決策哪個比較好，以及差異在哪裡，這就是一個很具體的問題；但如果我的目標是想知道如何讓公司獲利，這就是一個空泛的目標。

經過這樣的一番周折後，按照小角度切入的想法設計數據應用，就可以做得具體而快速，也可以避免因原料的變化而導致數據缺失的問題。

把數據放進框架中

在大數據的背景下，必須考慮數據之間的關聯性。一個單獨的數據是沒有意義的，要把數據放在一個「數據框架」中，才能看出存在的問題。為了把這個問題說得更明白，這裡我舉一家電商公司找我討論的問題為例。

A公司問我，要不要撤去導購網站的廣告，因為他們懷疑，許多老客戶是從導購網站造訪官網，而不是直接造訪官網。把這個說得再直白點，就是他們想弄清楚在導購網站投放廣告與公司業務之間的關係。那麼，接下來就要觀察，用什麼「數據框架」，有助於做出決策。

1 公司目前的投入產出比率

- 導購網站引進的新舊用戶占比為何？
- 導購網站引進的新舊用戶的投入產出比率和轉換率為何？
- 若撤去導購網站，對流失新舊用戶的影響為何？

2 與競爭對手的博弈

有一個問題可能會被忽略，那就是，如果你不做導購網站廣告，你的對手就會馬上投放廣告。在做數據框架時，要特別注意，框架不是靜態的，而是存在博弈的，需要把競爭對手的因素考慮進來。

3 時間因素

- 用現在、過去和未來的眼光來審視導購網站，看看導購的品質是不是越來越好。
- 需要注意，時間有延遲性，引進的流量會有一些延遲，兩、三個月後才能知道新用戶的價值。

總之，數據框架是商業分析師的靈魂所在，可以從框架中找尋問題

的關鍵因素及答案。不同的問題有不同的框架，無法完全在此闡述。

如何用框架做決策？

對此，我總結四個步驟：

- 1 確定問題，從解決問題的角度出發去收集數據。**
- 2 將收集到的數據整理好，放入一個數據框架中。**這個框架是用來幫助決策者做決定的，讓決策者更清楚地看到數據與決策之間的關係，比如，A公司在框架內要知道競爭情況、新舊客戶的比例等因素，以及多種因素之間的關係。
- 3 看框架與決策的關係。**比如，A公司與導購網站有三種選擇——完全不合作、部分合作和全面合作。數據分析師可以根據數據框架告知公司該怎麼決策。如果發現數據框架與決策不能匹配，就必須返回第二步驟。
- 4 根據決策行動，然後檢查行動是否達到目的。**如果行動後，發現根本沒有達到目的，就要檢討整個過程，尋找問題出在哪裡。是數據有問題嗎？還是因為框架不對？或是決策錯誤？是否還有數據沒考慮進去？

所以，想要解決的問題越複雜，框架也就越複雜。對於現在多數還沒開始做數據應用的電商公司來說，一開始框架千萬不要過於複雜，一定是鎖定某個需要解決的具體問題，開始搭建框架，使框架與決策緊密

相關。

所以，在搭建框架前問問自己：「目前的問題是什麼？決策是什麼？框架又該怎麼搭建？」雖然你的回答可能多種多樣，但是，最重要的前提是從小角度切入，從「小」做起。

養數據，重要的數據策略

「養數據」還有一個重要含義，就是決定要收集哪些數據。這個決定不容忽視，因為這可能是公司數據戰略中非常重要的一個環節。很多公司的做法，是有什麼數據就收集什麼，完全沒有主動性，但這樣會遇到一個核心問題，那就是——公司開始喪失自主思考的能力，對該收集什麼樣的數據並不了解。

在我看來，被動收集數據的行為是「收集」，而主動收集數據的行為則是「養」數據。再拿前面講到的電商婦幼網站為例，如果收集主體是醫院，那麼，寶寶的姓名、性別、出生年月和媽媽的姓名等數據的收集，就是收集數據，因為在住院檔案和出生證明上，這些都是不可少的。但如果是婦幼用品，要針對家庭進行行銷，那麼這些資訊顯然不足，寶寶奶粉的品牌、家庭收入、職業等資訊，需要企業主動收集，這些數據就是「養」出來的。

養數據通常有兩類，一類是網站自己沒有的數據，需要用戶主動提供；另一類是公司擁有、但沒有進行收集的數據。

如果要收集公司沒有的數據，在養數據時，通常需要花費更多的精力和技巧。舉例來說，現在有很多個性化的手機購物App，會根據使用者的偏好推薦商品，讓每個人有完全不一樣的購物體驗。具體的運作原理是，當用戶第一次使用這些App時，介面中會彈出一些問題，以互動的方式來收集用戶的資訊數據，比如，讓用戶在幾款衣服中挑選一款喜

歡的，這樣，在幾個問題之後，App就知道用戶的喜好了。

這類App通常還會有另一個特點，那就是讓用戶按「讚」，這個按讚的功能就是讓用戶明確地告訴App自己喜歡什麼。如此一來，用戶就可以在過程中不斷透露自己的偏好，App就可以「養」出使用者的核心數據。

至於公司擁有、但沒有收集數據的狀況，更多時候是因為內部資源協調和公司決策的問題。很多提供餐飲和團購資訊服務的網站會有一個功能，那就是把資訊透過簡訊發送到用戶手機，這樣用戶就不需要特別記錄，不過，有些網站會記錄用戶的手機號碼，有些網站不會。那麼，企業需不需要保存這些手機號碼呢？這就是一個與業務決策和資源協調相關的問題。如果決定記錄，則需要開發人員在後台進行功能的優化；如果不記錄，就會遺失這部分資訊。

依我的觀點，當然是記錄這些數據，也就是養數據，因為，這不僅能夠檢驗用戶的手機號碼是否有效，還能夠慢慢建立起企業與用戶之間的社交關係。

總之，「用數據」大致上一種方法論，而「養數據」，則是一種數據策略，是基於深入業務理解的更高層次商業決策。養數據的時間越早，累積的數據也就越多。養數據同樣也是一種管理和商業藝術，在養之前，可能誰也不確定最終會出現什麼結果，但是，一旦養成，則會產生非常大的商業價值。

關鍵思考 | 遠離**Yes / No**選擇

我們做的很多決策都是Yes或No，比如看電影，看或不看，只有一個選擇。

在遇到壓力時，我們非常容易陷入Yes / No選擇，會一直盯著看或不看電影這個選擇，而不會考慮「可不可以去喝茶」，這也是選擇，但在我們心中卻是餘地非常小的選擇。因此，當我們面對這樣兩難的問題時，一定要從壓力中迅速跳出來思考：可不可以不是「或」，而是「也」。要不要去看電影？或者也去.....

比如，有一天，你走進一家古董店，看到一件很喜歡的東西，就問：「老闆，這個多少錢？」老闆看到你進來時眼睛一直盯著這件古董，於是聰明地對你說：「您是今天第一位客人，很有緣，您隨便開個價吧。」於是，你開價500元人民幣，老闆馬上說：「不行啊，這個價值1500元。」你說：「不，我就出500元。」老闆說：「好吧，今天第一筆生意，就當作送你了。」事實上，500元人民幣成交的這件商品，或許只值100元人民幣。

在這個過程中，其實只要你一張口開價，便意味著老闆已經成功把你引入Yes / No選擇中。老闆心裡笑了：「吃定你了。」開價，就等於給你一個Yes / No選擇：500元人民幣，要或不要？沒有其他。

但是，如果你夠聰明，當看到老闆想把你引入Yes / No選擇時，就應該立即跳出來。此時，你無須開價，而是問：「多少錢？」老闆說：「1000元人民幣。」你說：「為什麼值1000元人民幣？」老闆會說：「這是明朝的東西。」你再問：「怎麼證明是明朝的？」之後，你會覺得豁然開朗許多。

別把自己逼到Yes / No選擇的境地，這在邏輯思考中非常致命。你必須跳出來，問「為什麼」。

討價還價時，停一停，「為什麼是1000元人民幣？」

討論今晚是否看電影時，等一等，「為什麼要看電影？」

別人說數據增長20%時，想一想，「為什麼是20%？」

想當一名數據分析師，你的腦海裡就要長期裝著這個想法：「這個數據是怎麼來的？」你要打破局限，不要把自己封閉在狹隘的思想中。在未來商業中，誰被逼到Yes / No選擇的境地，往往誰就吃虧。

Yes / No選擇在電商的活動中也很常見。「促銷最後一天」就是給你壓力，讓你不知不覺走進買或不買的選擇中。那麼，這時你該怎麼做呢？你應該快速退一步思考：「為什麼是最後一天？」其實，這是一個假資訊，早在前兩天就已經是所謂的最後一天了。就像在街頭我們常見的那種店——最後三天大促銷，但三天又三天過去了，半年過去了，還是最後三天大促銷。這些店就是在騙那些剛好今天路過的新顧客，把他們推到買或不買的境地。其實，對這些新顧客來說，如果不走到這裡，就會有很多的選擇。

如果你夠細心，你會發現，小孩大多是用Yes / No選擇。當小孩問：「現在能不能看電影？」如果你否決了，他們不會建議去打高爾夫球。在還沒接受教育之前，小孩不會有這樣的想法，我們都是在經過教育之後，才慢慢開始懂得怎麼做出選擇。但是長大後，在外界的壓力下，很多人往往又會變回小孩，逼得自己沒有選擇的餘地。

把這樣的思考應用在我們的成長中，很多人就是受到Yes / No的概念影響，將自己不成功的原因歸結為某件事情沒有做好，他們不會想到，失敗的原因，可能是因為自己沒做另一件事情。

我在澳洲曾經勤奮地練了兩年擊劍，擊敗了很多。但我發現，當我升上新的層級後，遇上高手總是輸。教練對我說，如果想要擺脫現狀，就要加強防守。教練說：「過去你贏對

手，是因為進攻很厲害，同級中比你差的選手，會因為你的快速進攻而敗陣。但現在，你想快速進攻，就會發現，勝利的可能性已經很小了。事實上，不是你進攻不夠快，而是防守不夠好。當你防守做得很好時，就可以做到假進攻、真防守，或是假防守、真進攻。這樣當你攻擊時，對手就不會太早防禦你。你發現沒有，別人輸給你，是因為怕防守，而不是怕進攻。」

同樣的道理，在面對Yes / No選擇的時候，我們很容易被慣性牽絆。很多時候，你需要遠離Yes / No的概念，才不會被困住而渾然不覺。

第7章

數據的盲點與負面數據的力量

如果數據存在盲點，核心數據就無法輕易顯現出來。而分析有盲點的數據，就像在清水中煉油，耗時耗力，仍有可能一無所獲。那麼，為什麼有人能在相同的數據中找出關鍵點，有人卻一無所獲？

大數據時代，在思考的過程中，有一個問題必須得到解決，那就是假設數據都是可獲取的情況下，什麼樣的數據才是關鍵？

數據的物理盲點

既然談到盲點，就必須對盲點下一個定義。一般來說，盲點是指眼睛沒看到的現實存在體。數據上的定義也類似，數據的盲點，就是指被我們無意中忽略的重要數據或角度在數據中，盲點可以分為兩類：一類是物理盲點，另一類是邏輯盲點。

所謂的物理盲點，是指在數據庫中不存在這樣的數據，即企業沒有收集到應該收集的數據。這類數據問題的產生，通常是數據收集策略出了問題。

讓我們回顧一下商品介紹頁的案例，過去，我們在評價一個商品是不是被用戶有效地瀏覽時，使用的數據指標是用戶在頁面上的停留時間，但這個指標存在天生的缺陷。

首先，停留時間並不完全代表用戶對商品的喜好。在企業收集到的數據情境中，有很大一部分可能是用戶的無意識停留行為，比如，剛打開網頁，就被主管叫去開會或談話等，導致用戶在某些頁面停留的時間

過長，這種數據就沒有收集的必要。

其次，通常停留時間不容易計算清楚。在傳統的日誌模式下，計算停留時間，都是以開啟下一個頁面的時間，作為上一個頁面結束的時間。因此，當一個人在多個視窗同時停留時，就會陷入停留時間過短的誤解。

在評價使用者是否對頁面感興趣時，這個數據就出現了物理盲點——企業沒有精確的數據來描述使用者是否對這項商品感興趣。

面對這種情況，就需要企業在商品介紹頁上用技術手段做標記，以記錄使用者在頁面上是否有動作、是否滾動螢幕看更多的內容。在明確了用戶具體看到哪些內容後，企業就收集到了相對有效的數據，能夠妥善地對商品介紹頁進行後續分析。

如果物理盲點是出現在電腦網頁上，算是萬幸；如果是出現在行動用戶端，那就是比較大的問題了。為什麼這麼說呢？因為網站可以隨時調整電腦網頁的模式——上一秒做完改版，加上代碼，發布後，下一秒用戶就能看到全新的內容，企業需要的數據也會在下一秒開始收集。但是，一旦物理盲點出現在App上，就會為數據收集帶來很大的難度。

造成這種現象的主要原因有兩個，第一是行動端的數據收集技術還不是很成熟，即便已經想到要收集手機的哪些數據，也會因為開發工程師的疏忽等非主觀原因而造成數據的忽略，當然，這在電腦端也可能會發生。

第二個原因則會讓這個問題加重，如果在電腦端出現因為疏忽而造成數據遺漏的情況，企業可以透過快速修復網頁來彌補，一般只會造成幾分鐘或幾小時數據的遺漏。但在行動端就沒有那麼容易了，因為行動端要求使用者經常更新版本，但在實際生活中，並不是所有用戶都會安裝新版本。另外，即便所有用戶都願意安裝新版本，新版本的推出也會受到系統方的制約，比如，以iOS系統而言，申請發布新版本可能需要等待數日，這時候，整個數據收集活動就會受到很大的影響。

所以，在數據收集的規畫期，一定要盡可能想清楚需要什麼數據來衡量業務，盡量避免可能出現的盲點。有時候，為了避免盲點，企業需要盡可能多收集一些數據，以減少物理盲點出現的可能性。

數據的邏輯盲點

相對於物理盲點的數據遺漏情況，邏輯盲點就是有數據、但是沒有好好地發掘出來。數據邏輯盲點的出現，很多時候與數據分析師或數據使用者的經驗和敏感度都有關係。在分析數據時，相關從業者必須對數據抱持敬畏之心，不可輕易放過任何能產生問題的可能性。

通常，企業在分析數據時，會將焦點放在宏觀的數據上，比如進入商品頁瀏覽的總人數，很少有人會去關注個別商品頁的瀏覽人數，即使有些商品的瀏覽人數明顯激增，分析師往往也不會投入精力去詳細研究。由於日常工作繁忙，這類數據常常被當作一個異常數據而忽略了，這就形成一個盲點。邏輯盲點如果未受到重視，很可能對企業的業務造成極大的損害。

曾經，有個電商網站就出現過這樣的案例，某項商品的瀏覽人數非常多，但商品的交易量卻很有限，這與常理並不相符。通常大量的瀏覽一定會帶來大量的成交，那麼，為什麼會出現如此違背邏輯的現象呢？

後來，該電商網站的一名員工發現，他雖然沒有瀏覽過這項商品，卻在自己的瀏覽紀錄裡發現了這個商品，疑惑之餘，他馬上就上報系統故障問題。直到這時，大家才投入精力優先解決這個問題。經過分析後發現，有人用了一種非常「聰明」的方法在刷數據。該電商網站的用戶ID是一串有序增長的數字，這項商品的賣家就模擬其他用戶的ID，造訪自家的商品頁，並且成功騙過日誌系統的防作弊機制，使得沒瀏覽過這項商品的用戶，在自己的瀏覽紀錄裡發現了這項商品。那麼，賣家為什

麼要做這樣的事情呢？因為網站會在後台記錄這項產品，並且誤認為它符合用戶的喜好，並向用戶推薦相關產品。

在這個案例中，從問題出現到問題解決，經過了一段時間，雖然沒有造成重大損失，但也對部分的使用者體驗造成影響。其實，這樣的問題如果能夠在出現之初就得到注意，並且認真地研究異常數據、尋找根源，就能夠避免帶來的影響。

還有一些數據也容易被忽略。舉例來說，對一些平台型電商或自營電商來說，成交額總是能獲得更多的注意，而對退貨情況關注得較少。各家電商通常都會對成交額設定指標，每個營運項目的團隊或員工都有成交額的KPI。大部分時候，考核成交額並沒有太大的問題，但就是有些人會去鑽「重成交、輕退款」的漏洞，這種情況在各家公司中屢見不鮮。一些經營平台的員工會讓賣家「想方設法」達成成交額，並允諾將支援活動資源以作為回報。所以，有些賣家就採取了「自己買進，然後退款」的手段，以提升成交額。這樣一來，雖然員工得到了業績，賣家得到了資源，但是公司卻沒有獲得任何好處。其實，在這個過程中，只要注意品項或商品的退款率，就能輕易地發現這些數據，盲點自然就會消失。

在邏輯盲點中，最大的盲點是將電腦端的數據和行動端的數據混著看。很多網站都會統計轉換率，即購買用戶除以瀏覽用戶所得出的數據。在單純電腦端的情況下，這個轉換率大致可用，還是可以描述宏觀的經營狀況。但是，在行動終端進入後，這個數據就有問題了。行動端成交的數據記錄在成交中，但是瀏覽的數據卻沒有計算至瀏覽中，就會導致轉換率越來越高。如果忽略這一點，就有可能遺漏重要商機。

人為製造的盲點

數據除了物理盲點和邏輯盲點，甚至還有一些是人為製造的盲點。人為製造的盲點就是故意掩蓋數據，或者人為調整數據的標準。在一些網站，定義UV（獨立訪客），主要是考慮自主UV，也就是說，只有當用戶主動造訪頁面時才算數。但是，有些人為了自己的績效，可能會動一些手腳，比如，對外投放彈出式廣告，也算用戶主動造訪；或是在頁面加上嵌入模組，當用戶造訪一個頁面時就會有兩條數據，藉以增加UV的數據。這些情況通常比較隱晦，甚至能欺騙管理層。

所以，管理層在聽取數據彙報時，不僅要看數據的變化和趨勢，也需要對數據的標準有一定的理解，最重要的是，要問清楚定義是什麼。如果涉及轉換率，就要問清楚計算的分子、分母是什麼，才不至於掉入數據的盲點。

幸好，現在商業的整體環境對數據都比較重視，我接觸到的高層主管也越來越關心數據，越來越常使用數據做決策，也越來越重視數據的分析和挖掘對業務產生的價值。如此一來，數據的盲點也會越來越少，數據也會變得越來越有價值。

善用小偷思維

數據的盲點有沒有價值？讓我們舉一個生活中的例子。

在中國，很多人都會在黃金週時外出旅遊，你的房子有七天都沒開燈，這項資訊對你而言沒有任何用途，但是對小偷來說就是很好的機會。對小偷來說，一間屋子三天不亮燈，並不代表沒人在家。小偷會一直觀察環境有沒有突然變化，會對零散的資訊，如噪音特別敏感，他將所有感官能力都提升到了警戒線。小偷的成功，一個很重要的原因在於，他們會衡量風險，也就是會使用「負面思考」看待問題。

人類是有惰性的，我們喜歡盡量把自己想要思考的東西變得簡單。

「壞人」卻很願意承認不可預見性的存在。就像我們常常判斷一件事情是好或不好，而不是「未知」。舉例來說，我參與一場賭博，賭注為200元人民幣，在這個事件中，就我看來，贏就是好，輸就是不好。但是對壞人來說，他們一直在衡量一些未知的情況，比如，這家賭場是不是黑店、警察會不會突然闖入沒收賭資，或者會不會一開始就輸了200元。

壞人之所以會對未知的情況如此關心，是因為當壞人的風險實在太高了。對正常人來說，家裡丟了一件東西，也許無關緊要，但是對小偷來說，失敗是要坐牢的，這就注定了他得細心觀察正常人看不見的數據。

想成為小偷裡的「成功人士」，首先想的肯定不是如何偷東西，而是偷完以後怎麼不被發現。小偷必須懂得利用蛛絲馬跡來降低風險。

當失敗的代價越大、風險越高時，人類天生有著動物般的敏銳度。也許你不知道自己已經在風口浪尖，但是當你身處其中時，你一定會提高對風險的評估能力，不是嗎？動物世界就有這樣的現象，能力越小的動物，越具有優秀的感知風險能力，猛獸則沒有。

也許你會問，對於我們這些不準備當小偷的人來說，負面思考又有多大的意義？

我們都會受到經歷和經驗的影響，當我們退居劣勢時，就會像小動物一樣能夠感知風險。高風險能讓一個人對不可預見性非常敏感，到那時候，負面思考和避免風險的能力就會自然而然地出現了。

所以，如果我們平時也懂得運用這樣的小偷思維，就能避免很多錯誤。這時候，如果我們再站在壞人的角度做負面思考，覺得別人出錯就是你的機會，很多人出錯就是你的大機會，這甚至能創造出一條很有價值的產業鏈。

在美國和澳洲，有種燈可以設置偶爾亮、定時亮和晚間亮，研發者可說是一個典型具備負面思考能力的人。可見，我們身邊的這些負面思

考是可以產生巨大價值的。

再看看我們身邊的商業環境，這樣的負面思考也很常見。網路公司奇虎360的周鴻禕就是具備負面思考能力的人，看他的軌跡，從3721到360，他注意到了很多人的錯誤，觀察到了很多人需要免費防毒軟體的需求。

當你希望能達成某個目標時，你必須用心觀察別人觀察的事物，這樣才會提升你的觀察力。我們都意識到培養正向思考的重要性，也都意識到要用心觀察成功人士是如何成功的。但同時也要注意，應該培養自己運用負面思考的眼睛，去觀察別人是如何失敗的。面對數據的盲點，正面思考告訴你怎麼做可以到達成功的終點，負面思考則告訴你怎麼做才不會失敗。

當對手的數據就擺在那裡時，不看的人才是傻瓜，這並不涉及道德問題。偷看不是數據世界的本質，但觀察別人怎麼失敗，卻可以實實在在地降低自己的成本。值得注意的是，在正常情況下，獲取數據情報很不容易，這個時候，你得想著不能破壞公司的名譽，還要想著不能讓別人偷看自己的數據，更要把別人來「偷」數據的後門關上。

在上述過程中，我只是在模仿小偷。總之，壞人觀察到的數據，正常人很難看得懂，但如果你學會用負面思考的方式去觀察，必定對你有很大的提升。一般人想要訓練到這樣的程度是很難的事情，畢竟你的心態是正常的，你不會去注意誰的包包拉鍊沒拉上，不是嗎？

無意識地培養負面思考，那叫亂想。但如果你是為了特定目的而做規範訓練，你就會成為很厲害的「小偷」；在日常生活中，那叫作很厲害的「觀察者」；而在數據世界裡，你就是很厲害的「數據分析師」。

「小偷思維」讓我們可以換個角度觀察數據，在現實中，我們可以用這種思考方式來衡量數據盲點的價值。這兩個內容看似分離，但核心是一樣的——有沒有看到應該看到的數據，有沒有錯失不應該錯失的數據。

負面數據的價值

先提一個假設性問題：兩百公尺外忽然發生爆炸，你如何透過大數據鑑定是否有人蓄意而為？如果是，如何快速鎖定嫌犯？

一般情況下，我們主要分析兩個數據：在爆炸之前，現場是否有一些形跡可疑的人出沒；在爆炸發生後，人群中是否有行為怪異的人。例如，在爆炸發生後，大部分人都是往事故發生的方向走，以察看誰受傷了，或者單純湊熱鬧。如果這時，有兩個人是往反方向快速離開現場，就算是行為怪異了。

也許有人會問，這和大數據有關嗎？當然有，而且還能引發我們思考另一個問題：數據的盲點在哪裡？

平時大家接觸到的都是資訊，其中，可以量化的資訊稱為數據。所謂大數據時代，就是指可以記錄和調用的數據變得越來越多。網路和行動網路就像兩部巨大的超級感測器，把人、物、時間、空間等數據即時記錄下來，讓我們在時間和空間裡都留下線索。數據從太少到太多，從簡單到複雜，還衍生出無限的關聯數據。面對海量數據，我們常常疑惑，這些數據究竟是不是足夠了？是多了，還是少了？

回到前面提到的案例，人們常常可以預見正面的資訊（現場圍觀群眾的行為），但很容易漏掉負面的數據（把數據範圍放大，你可以注意到急速離開現場的人）。這裡面的盲點指的是，我們在觀察數據時，很容易只將焦點放在正面數據，而忽略負面數據。

這種案例不勝枚舉，以我的日常工作為例，當團隊在對高階主管做報告時，我通常不看報告，主要記錄下主管們「什麼時間點頭」、「什麼時間針對性地做了筆記」和「什麼時間沒怎麼在聽，在看手機」。不要小看這三個數據，與當時報告的內容對應（關聯）起來，我就能輕易

知道，剛才的報告中，哪部分不吸引人，哪部分需要改進。尤其是主管們看手機的時間，完全可以算是理論上的負面數據，但是對報告的改進，其實非常有價值。

再比如說，公司每個月有10億元人民幣的交易量，雖然額度已經很大了，但你是否分析過，其實還有2億人民幣的潛在交易量沒有做成。為什麼？因為有人看了不買。為什麼不買？透過分析這些負面數據，其實很可能發掘出非常有價值的平台交易增長關鍵點。

當心經驗也會成為盲點

還有一類數據盲點，常常發生在我們賴以依存的經驗上。我的導師說過：「『假設』這件事情，常常是誤判的源頭。『假設』會讓人盲目。」

很多時候，人們在做決策時會依靠經驗，但經驗有可能是錯的，這會導致「假設」錯誤，繼而影響執行的結果。比如，今天你看到一張桌子，上頭擺放著很好吃的食物，但是你把手伸過去時卻被燙到了，之後，你再看到這張桌子，恐怕都會因為之前的經驗而不敢靠近。顯然，錯誤的經驗會讓你不斷犯錯。

經驗會成為盲點，還有一個重要原因是，你沒有注意到環境變化的趨勢。有時候，經驗反而會讓人經歷慘敗，因為這件事情已經出現趨勢上的變動，但大家卻未察覺。之前的經驗連結到當下的決策，會讓你認定事情必然會發生，甚至是規律。但是這個規律會騙人，股票市場中的教訓已經太多了。

數據分析師經常會受一些企業邀請去做趨勢判斷，比如，針對兩家公司的員工增減趨勢做分析。表面上看，一家公司的員工比另一家少，但如果再看產業的背景，一家是新創公司，一家是有十年歷史的公司，

此時，再結合產業的發展趨勢，你就會做出不同的判斷。想要做出正確判斷，必須先抓住發展的趨勢。

古語有云：「無知最窮。」在大數據時代，那些自詡對大數據什麼都懂的人，其實有無數盲點而不自知。他們淺薄地認為，數據越多就越能解決問題，實在是無知的表現。

在海量的大數據面前，我們最大的盲點就是：衡量應該收集多少數據才算足夠，這是「存」；如何管理從未見過、如此多而廣的數據，這是「管」；如何使用它們以實現更大的價值，這是「用」。

關鍵思考 | 為什麼數據會騙人？教你區分謊言

有一部講述魔術的美國電影《頂尖對決》，片頭一開始就對觀眾發出一句考問：「你真的看見了嗎？」很多人以為自己看見了，但因為缺乏真正的觀察，其實還是離真相很遠。

什麼是「觀察」？觀察就是在事物內部找到相對性，而且是從共通性裡找出相對性。

所以，當我們想要判斷一個人講述的事實是不是謊言時，就要基於以下兩種情況進行判斷：一種是，這個對你說話的人，是你人生當中第一次見面，或是見面頻率很低；另一種情況則是，在日常生活中，你經常和這個人見面。

這對於判斷一個人是否說謊具有不一樣的意義。因為對於熟悉的人，要判斷他是否說謊其實很容易；而對於一個只見過一次面的人，要判斷他是否說謊就要難一些，因為你沒有經驗做對比，只能透過短時間的觀察，很難找出相對性。

在此，我們首先要了解兩個詞語：「常態」和「時態」。

在說明常態之前，我們先來了解「情態」的概念，蘊藏在體內為情，表露在體外為態。情有七種，《禮記·禮運篇》中指出，七情為喜、怒、哀、懼、愛、惡、欲。歡喜、憤怒、哀傷、恐懼、愛慕、厭惡、欲望，都是人們內心的情態。這七種情態在人體內不斷地擾攘，被大腦的邊緣系統激發後，就會衍生出「時態」。有些人賭博贏了，高興到旁若無人，其實是邊緣系統出了問題。當他冷靜下來之後，也許會驚訝於自己剛才的表現：「唉呀，剛才失態了，真丟臉。」

我們在觀察一個人的時候，最容易觀察到他的時態。但如果要觀察出一個人時態的差異，就必須了解這個人的「常態」。常態是什麼呢？常態分為以下六種：弱、狂、嘩、周旋、慵懶、媚。所謂弱態，是指輕聲細語、動作溫柔、包容性強，像小鳥依人一般。所謂狂態，是指不修邊幅、倔強好勝、言談舉止旁若無人。對比自己身邊的朋友，就可以輕易判斷出他的常態。

狂態的人在被問及意見時，通常願意主動表達觀點，並堅信自己是對的。弱態的人則不會表露太多意見，會給出模稜兩可的意見。嘩態的人通常遇到事情會大手一揮：「你們都別說，讓我來說。」周旋態的人在被問及選擇時，說到A會說A很好，說到B會說B其實也不錯。慵懶態比較漫不經心，什麼都不是很在乎。媚態則意指諂媚。

在人們的日常狀態中，每個人並非單獨屬於一種態，而是經常混在一起。但有一個交友原則：遠離弱且媚的人。如果是狂且媚的人，其實是非常厲害的人。

我們日常在做判斷時，一定要先知道這個人的常態，才有判斷的依據。當你向一個常態是慵懶態的人彙報工作時，他忽然呈現正襟危坐的時態，你就可以判斷這次的工作內容也許特

別重要。

想要區分謊言，就要先區分常態和時態。如何知道一個人的常態呢？要靠觀察。所以我才說，對於熟悉的人，要判斷他是否說謊很容易，因為了解其常態；而對於剛見面的人，則比較難做出相應的判斷。

在我們的日常用語中，還有變態一詞，其實就是由常態轉變為時態而衍生的異常肢體動作或不理智行為。如果洞悉常態和時態，其實就很容易觀察出一個人變態的起點和變化點。

另外，還有很重要的一點要觀察，當一個人的常態發生改變時，你要思考，促使他改變的是什麼環境和事件？比如，當一個人被抓到警局時，會恐懼、暴躁不安、肢體動作粗魯，但這並不能推論他一定是壞人，環境的壓力也應該要考慮進去。

如果是爆炸案，在爆炸發生前後，是否有不符合常態的人出現在現場？案發前，現場是否有不正常逗留的人？案發後，當有人受傷倒地，圍觀人群忙著搶救時，有兩個人行色匆匆，拿著手機，往反方向離開現場，這兩個人的犯罪嫌疑就很大，因為這就屬於常態中的變態。

以人推「數」，當我們在觀察數據時，要考慮以下幾個問題：數據的常態和時態是什麼？為什麼數據會騙人？透過直覺感受和經過仔細觀察所得出的結論，也許會大相徑庭。

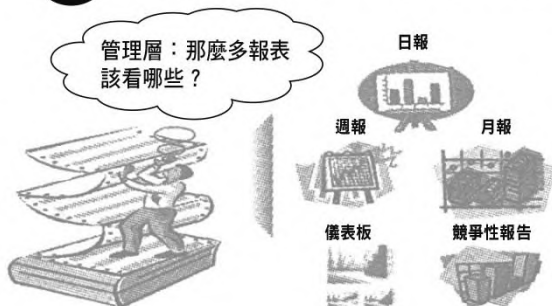
PART 2

阿里巴巴的大數據祕密

數據化營運實踐過程中的問題

堵

日常報表資訊量大，難以捕捉有效資訊



獨

資訊分散在不同部門，缺少有效整合



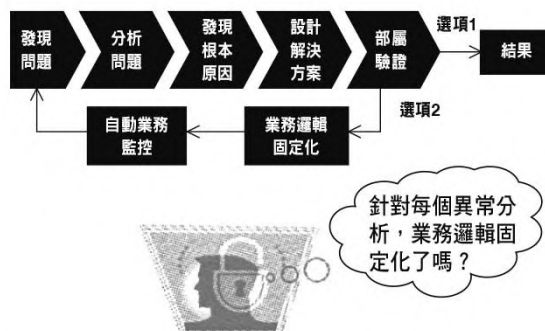
慢

業務異常的處理，往往是自上而下推動



漏

關鍵分析取得成效，但未實現沉澱



第8章

阿里巴巴的大數據實踐

按照一般理解，一個人在註冊網站時，性別只會登記為男性或女性，但不知道你是否能想到，阿里巴巴竟然有十八個性別標籤。

在經過仔細調查後，我們發現了一些可以識別用戶性別的方法，比如，某個用戶早上的瀏覽行為更男性一些，晚上就會變得更女性化（也許是妻子在用）。真實的性別只有0與1的關係，而現實卻是0～1的關係，比如70%是男性，30%是女性。

在每一個情境中，性別的變化都是其來有自，例如搜尋、社交和廣告等。靜態的真實性別在A/B測試^[1]中的表現，不如動態的虛擬性別有效，這恰恰證明了數據需要營運，絕對不可以閉門造車，這樣的例子在阿里巴巴不勝枚舉。

落實大數據的方法

談到數據化營運，在我們公司內部，流程是這樣的，舉例來說，我們要為童裝商品做行銷推廣，招攬新客戶，我們會先找到目標客群，利用標籤，把對這個品項感興趣的消費者找出來，再透過發送郵件或簡訊等方式，吸引他們的關注，這其實就是數據化營運，也就是用數據幫助企業解決問題。最簡單的解決辦法是找出所有曾經看過、但沒有購買的客群，或者從相關品項去擴大目標客群。

數據從「用」升級到「養」，即營運數據，我們就會嘗試在整個淘寶用戶中查找，以12歲以下兒童為目標的商品為例，此時我們的做法就

是在用戶中搜尋，有多少用戶家中有12歲以下的孩子、卻未曾購買過這類商品。

你會發現，以前只能針對有過購買行為的幾百萬用戶，現在，我們的營運範圍竟然可以達到幾千萬可能有12歲以下孩子的用戶。從幾百萬到幾千萬，這就是營運數據，這是一個需要從主動收集數據到營運數據，再到產生新數據的過程。

在我的理解中，從數據化營運到營運數據，是不斷運行的循環。在這個循環中，會容納許多新的、不同面向的數據，這些數據經過在整個循環中的適應過程，然後再運用到數據化營運，並且改變原有的營運方式，這個過程就是我認為落實大數據的方法。

假設數據是穩定的

數據化營運的前提是，假設數據是穩定的，並且以此改變企業的營運。以前的數據都是結構化數據，比如說，早期大部分企業都用財務數據來幫助企業營運，因為財務數據經過多年的沉澱，相對來說較為穩定。但自從非結構性數據或半結構性數據產生之後，數據化營運的難度就變大了。比如，音頻是一種典型的非結構化數據，你很難把音頻變成結構化數據。有了這樣的數據之後，用數據指導企業營運，面臨的技術難度遠遠大於以往。

我們做數據化營運，前提是假設數據是穩定的；而營運數據，則是假設數據都是可獲取的，而且是不穩定的（見圖3）。

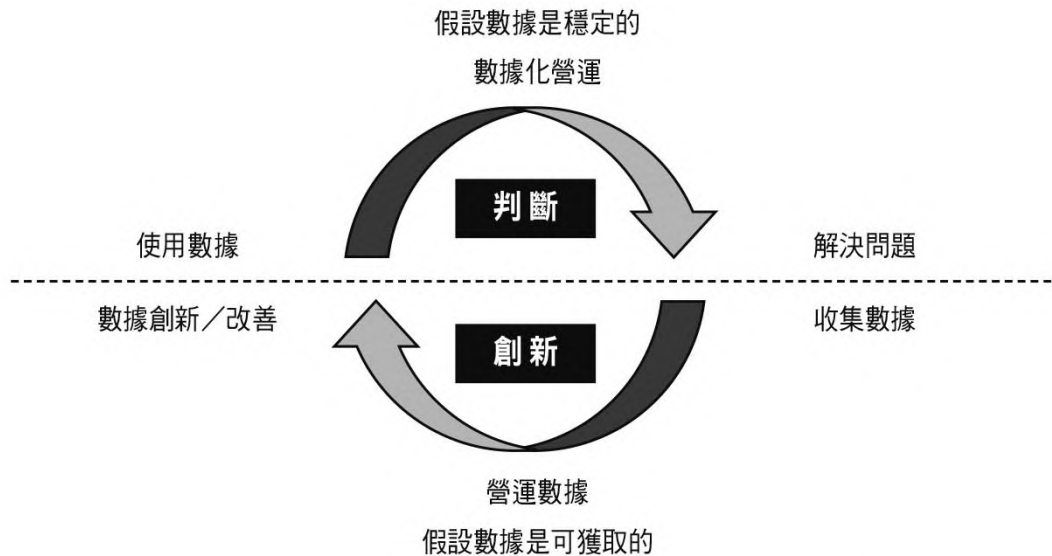


圖3 數據化營運和營運數據的閉環系統

這又如何理解呢？所謂穩定，例如關鍵字的投資報酬率，這項數據聽起來很穩定，但反映的卻是短期利益。數據化營運需要和商業結合得非常緊密，所以數據也是混合在商業裡，以假設穩定的方法，做業務上的對比和細分，以及趨勢預測。

這並不是說穩定是對的，而是因為企業對數據營運的分析需要假設很多穩定因素。

我們以點擊率為例，數據營運人員可能會覺得，這項數據我們一直在使用，有什麼特別的呢？但在我眼裡，點擊率是變化的。我的問題是，當用戶點進商品介紹頁時，有多少用戶完全沒有滾動螢幕往下看？如果這個比例很高，那就意味著點擊率本身有誤差。

所以，在數據化營運裡，我們往往假設數據穩定不變，因為在我們的認知中，我只是數據的使用者，只要假設數據是穩定的就可以了，這是普遍心態。那麼，為什麼我要去注意數據穩不穩定？而且，假設數據穩定，也意味著我們習慣不去尋找一些新數據。

當鑰匙不見了，我們一般的做法是去可能丟掉東西的地方找。同

理，在運用數據時，如果出了問題，我們也是一樣的心理，不會去思考慣性思維以外的其他可能性。而且，當這些地方無法解決問題時，我們就會不解，這是數據化營運中很常見的一條死路。

假設數據是可獲取的

於是，到了養數據，也就是營運數據的階段，我們的做法就必須改變。此時，在思考靜態的性別時，就要問自己：「會不會不好用？是不是需要改善？」

從企業的角度，數據化營運一般是自上而下，管理者強調KPI指標，員工按照目標進行分解。所以，數據化營運是計畫性的，而凡計畫性的，一定都是穩定且結構化的。

營運數據卻是自下而上，經過無數的錯誤和實驗，慢慢地向上追溯。就像阿里巴巴有十八個性別這件事情，這十八個性別並非管理層憑空想像出來，而是將大家發現的性別，經過交叉比對，分析到底在什麼情境下更適合。可見，這樣的結果是試出來的。

在我們的日常工作中，你很少會聽到負責業務的管理者說：「我要收集這項數據，它很有用。」這樣的場景一般不會出現。我有過這樣的經歷，是因為我是數據部門的管理者，我在跟其他公司的人討論時產生了這樣的想法，回到公司後便提出：「我們也來試試這個數據。」比如，前面提到Amazon根據IP位址觀察用戶附近十公里內有沒有書店的例子，同樣的營運數據的方法我就在淘寶嘗試過。

我想說的是，如果沒有營運數據這個部分，收集來的數據絕對就不是大數據。

從假設數據是可獲取的角度來思考問題，是數據化營運裡很重要的方法，這與「如何利用數據拿到更多數據」的問題並不一樣。

今天，我已經知道了你的一些數據了，但是並不多，此時我就要想辦法，用我知道的東西，拿到你更多的數據，這在數據化思考中，稱為「有意識地用數據拿數據」。

現在，最常見的一種用數據拿數據的方法就是A/B測試。原理是，我本來就知道你對某項東西很感興趣，今天，我做了一個測試，本來是A頁面，我給你B頁面，看看你有什麼樣的反應（有點像功夫裡的假動作）。其實，B頁面拿到的數據，一定是基於A頁面拿到的數據。

比如，你經常走某條路，今天，在你走過這條路時，我安排了一位美女從這裡經過，測試一下你的路徑有沒有變化，以及你有沒有停下來。以前收集到的數據告訴我，你在路過時是不會停留的，而今天你停了下來，那我就得到了「美女的出現可以让你停下來」的數據。

這時，美女就是我拋出來的問題，即用問題來獲得更多數據，這就是用一部分數據拿到更多數據最常採用的方法。

建立大數據創新循環

打通數據化營運和營運數據，就有可能創造出一些意想不到的創新。

我們在大數據環境裡營運數據時，有可能走不通的原因是，現在的數據已經龐大到需要管理才能達到創新的數據化營運。這個循環能否成型，往往需要長時間的努力。

互聯網金融就是近幾年比較新穎的一個例子，現在也有很多走在前端的公司在做類似的事情。這麼做的原因是，公司發現，用戶的資產鏈是評估其經濟狀況一個有用的資訊。將收集到的這些數據應用在信用評估上，可以更動態地決定用戶的信用額度，降低借出方的風險。

比如，某金融網站擁有用戶公司的財務數據，大概知道用戶的資金

流向——原來還有30%的訂單在處理，還有不少訂單在路上，有些則是押在一些擔保交易上，所以這個用戶不是沒錢，而是錢押在供應鏈的各個環節上。萬一用戶不還錢，網站也有能力卡住用戶的資金流，這讓網站有能力降低自己遭受呆賬的風險。

其實，阿里巴巴在數據化營運到營運數據的循環系統中，能做到現在的成績，跟管理層更相信數據、更願意付出，以及更願意投資有很大的關係。這個循環系統的關鍵在於，管理層如何應對數據價值的斷層，這對於整個循環能否走下去影響非常大，而阿里巴巴應對這個循環的內功，我認為答案在於「混、通、曬」。

先不要一頭霧水，「混、通、曬」，簡單來說，是我們數據分析師的三訣竅，我會在下一章詳細介紹。

關鍵思考 | 先開槍，後瞄準

前幾年，有一個關於淘寶的說法甚囂塵上——淘寶正在經歷野蠻成長。

什麼是野蠻成長？這個概念現在常常被提及，事實上，野蠻成長應該是在資源有限的情況下才成立，在資源豐富的時候，或許就不存在野蠻成長的概念。現今的淘寶，已經是一個有錢、有資源的「富二代」，已經不存在野蠻生長的前提了。

在我看來，馬雲是非常懂得利用野蠻成長的領導者。在一次會議中，有位同事說明一項提案，但是在同事說明的同時，馬雲卻一直搖頭。會後，馬雲說：「我給你三個目標，你用一年的時間去執行這件事情。」這三個目標都是在資源有限的情況下進行，這才叫野蠻成長。

那麼，為什麼馬雲一直搖頭，他不相信這件事會成功，卻依然讓這位同事去嘗試呢？後來，馬雲提起，當初他想做阿里巴巴時，也被很多人否定了。所以，即使這位同事的提案看起來成功的可能性很低，還是要讓他去嘗試，因為不試怎麼知道結果呢？

對啊，不試怎麼知道呢？但「試」也是有方法的。

我經常聽到很多淘寶賣家說：「先開槍，後瞄準。」我常常思考，這句話在邏輯上怎麼講得通？正常的邏輯應該是「先瞄準，後開槍」才對啊？後來我知道了，其實這句話是有前提的，那就是，當你沒有數據的時候，如何在探索之中找到有參考價值的數據，最後發現目標，做出決策。所謂開槍，就是在一堆海量數據裡盲選；而瞄準，則是在海量數據裡盲選出有價值的數據。這樣，「先開槍，後瞄準」才是有價值的。

將這個觀點延伸，有關開槍和瞄準的順序問題，就像現實中，很多人容易把目標和目的混為一談，目的和目標區別在於，「目的」是要到達的終點，而「目標」是衡量到達終點的價值標準。

如果用一句話來總結野蠻成長，那就是——是，但不是；不是，但是。

這句話怎麼理解呢？我們用「是，但不是」舉個例子，我們常常這麼評價一個人：「他是好人，也是壞人。」他是好人，是因為他在外面對朋友很好；他是壞人，是因為他對自己家人很壞。


再說「不是，但是」，比如，某個人創業失敗了，但是不是完全失敗了呢？這其中還是有很多數據價值的。

比如，我很欣賞化妝品購物網站NALA的劉勇明，他原本是學海洋學的科學家，懂得如何探索和觀察，後來當了商人，

對用數據解決問題這個領域並不是很得心應手。他很重視客服，最擅長的做法是透過客服來觀測數據，透過最前端來決定後端，這就是探索。

我很讚賞他說的：「發現交易量急速上漲時，就要停掉廣告。」這句話體現了他作為科學家的思維邏輯——當交易量急速上漲時，訂單的處理能力就會下降，所以，減少廣告，可以更妥善地控制好局勢。

「先開槍，後瞄準」，對新創公司來說是正確的做法。對新創公司而言，管理者不需要把眼光放得過於長遠。從短期目標出發，在不斷的觀察中發現問題，並且解決問題，可能會更容易獲得成功，事實上，很多人的成功就是這樣造就的。

-
1. A/B Test，數據分析中的一種對比分析方法，可以更明確地展示實驗組和控制組的實驗結果。 

第9章

混、通、曬——阿里巴巴的數據化營運三訣竅

談到阿里巴巴的數據化營運，我首先想到的就是「人」。我們大多數時間都花在討論應該要做什麼，卻很少反過來想，要落實數據化營運，首先要從人做起。

因此，我想跟大家分享的祕密是，數據化營運的內功是什麼？簡單來說，就是好好利用「混、通、曬」這三訣竅。

「混」出數據

現在很多數據分析師，在面對專業領域「怎麼算回歸」、「怎麼畫函數」的問題時，常常是遊刃有餘，但在實際工作中，卻往往缺乏商業意識。如果數據分析師缺乏商業意識，那麼，公司就成了「盲人」——分析師不知道該以什麼樣的邏輯分析數據，公司的決策層也得不到任何有價值的參考意見。

現在絕大多數CEO都在抱怨，每天要看一大堆零零散散的數據。造成這種局面的原因就是，數據分析師只是單純地把數據傳遞給管理者，卻沒有向管理者解釋，這些呈現使用者行為的數據，和能夠在商業上產生價值的數據之間的內在關係。

CEO沒有多餘的精力解讀PV（頁面瀏覽量）和UV（獨立訪客）等數據，他們只需要知道數據是否有問題、反映了什麼問題、最近有什麼新的發現，以及需要我們做什麼樣的改變。

簡單來說，具有商業意識的數據分析師，在監測到網站上嬰兒車銷

量增加的情況時，基本上，就可以預測奶粉的銷量也會隨之上升。而且，也只有具備商業敏感度的數據分析師，才懂得用什麼數據驅動公司實現經營目標。

那麼，數據分析師如何才能擁有商業敏感度？我覺得就要靠「混」。

什麼是「混」？舉個例子，我以前公司的CEO就跟我說：「你每週要跟業務部的負責人吃兩頓飯，最起碼兩次，這就是你的KPI。」我認為，這就是「混」。作為數據分析師，如果不和業務部混在一起，又怎麼會知道業務部都在做些什麼。

商業敏感度是靠「混」出來的，它並不會憑空出現在你面前。從更深層次來說，數據部的人要和業務部的人經常混在一起，比如，參加業務部的週會、業務規畫討論、腦力激盪等，或者定期做跨部門的工作輪調、互通週報郵件，一起喝茶、吃飯等。現在在公司裡，我最不希望看到數據分析師長時間坐在自己的位子埋頭苦幹，我更願意看到他們和業務人員混成一片。

我要求數據分析師在給我的週報裡，一定要講到業務方的動態。而且，我給他們的考評標準是，千萬不要讓我看見業務方發過來的週報裡有的內容，你的週報裡沒有。我認為，要實現這個要求，最基礎的出發點是，數據分析師一定要跟業務方溝通，才有可能服務他們。

打「通」混出來的數據

當你與業務人員混得夠熟時，在看到某些數據後，你自然就會明白：「喔，這個數據跟商業決策絕對有著莫大關係。」

今天，各家電商公司在評估公司的經營狀況時，越來越依賴數據，但是，很少有電商敢完全肯定地說，自己掌握了體現公司狀況的完整數

據。對公司主管而言，一是因為，很多電商在開始收集數據時，會發現數據非常散亂，分布在不同的數據收集管道和營運人員（公司的核心員工）手裡，使得數據流非常「堵」。另一個問題是，絕大多數電商缺乏大數據營運的經驗，只是收集了很「散」的數據，不知道要如何利用，也不知道該讓哪些數據建立關聯。

從客觀角度來看，數據營運的各個方面都可能存在影響數據精準度的「噪音」。數據本身是客觀的，但是它很容易受到產品和營運人員的影響——產品的目的導向會影響營運人員的想法，而營運人員的想法會影響樣本獲取的精準度，這就造成數據在不同人眼中出現不同結果的情形。

以轉換率為例，市場部門和營運部門對轉換率的想法是不同的，如果公司內部的數據標準沒有打通，取得一致，那麼，公司在做決策時，被數據迷惑和誤導的可能性就會增大。

因此你會發現，問題最後還是要歸結到人和公司上。如果不能打「通」到商業環境裡，即使數據再多，也沒有任何價值。堅持帶著業務問題觀察數據，並帶著數據觀察業務，同時擁有這兩種敏感度，就是做到了「通」。有些人可以在很短的時間裡判斷出數據是否有價值，就是因為「混通」了。

值得注意的是，「通」有兩種情形。第一種情形，比如，現在有一個商業情境和一堆數據，當這兩者產生關係時，就是商業模式和數據之間的「通」，我們可以稱它為「數據中間層」，簡單來說，就是能夠敏感地反映商業變化的數據群（smart data set）。

另外一種更深入的「通」，存在於公司組織中的數據，比如，有一個商業問題，需要數據A和B、C、D互通，才能解決。舉例來說，今天退款非常多，我會做的第一件事情，可能是去察看行銷部門的數據，如果我看到今天UV爆增，所以退款才特別多，這是合理的。但是，如果我從客服數據中，發現一些退款是來自少數賣家，那就另當別論了，因

為這說明這個商業情境可能存在作弊，此時，就需要綜合行銷、客服和風險控管等多組數據以進行解釋。

所以，把數據打通，一種是商業理解和數據之間的打通，另一種就是部門數據和部門數據的交叉。

現實中的「不通」，也有兩種，一種是從人的角度來說，比如人在思考上的「不通」；另一種則是從公司的角度來說，比如架構上的「不通」。這兩者角度不同，造成的損失也有差別。大部分的公司都會出現不通的情況，比如，部門與部門之間的隔離，打亂了整個流程的連貫性，財務部門對產品、市場和營運等的不敏感，可能會把數據打散，這樣一來，商業活動就根本沒有辦法繼續進行。

想實現數據的積累和沉澱，想打通數據，建立合理的系統是不二之選。首先，做好數據安全工作，確保公司內部不同職位的員工可以察看不同的數據；其次，統一不同部門的數據標準和定義，使公司內部數據有統一的介面，避免混亂；最後，連結不同部門的數據，創造機會讓數據的營運可以擴散至數據部門之外。

「通」是「混、通、曬」裡最關鍵的連接點，以前，數據量沒這麼大的時候，公司「混」完，就「曬」了，完全憑藉商業敏感度在營運數據。而現在，海量數據成為主流，「通」也就成為營運數據不可或缺的一部分。

將經過混和通的數據「曬」出去

「曬」是在「混」和「通」的基礎上產生的最終數據表現，是基於人、商業和數據結合後的一種看數據和用數據的方法論。

在曬數據這個層面上，通常是透過數據回答這幾個問題：

- 業務好不好？

- 數據如何改變，可以讓業務更好？
- 如何利用數據幫助業務發現機會？
- 如何利用數據創造新的商業價值？

這幾個問題看起來是遞進關係，其實不然，因為具體應該用數據解決什麼問題，要根據業務的情境來決定。每一個問題都有不同的回答方法，接下來，就具體討論這些方法。

用數據解答業務的好壞

數據如何證明業務好不好？這個問題用更專業的語言來說，就是如何搭建數據框架，並透過數據框架來解答業務的好壞。所謂的框架就是，對一項業務進行指標化的分解，並透過有限的多個指標來客觀描述業務的狀況。指標化的分解，是為了能夠快速定位問題。

舉例來說，每個人一生中都可能因為感冒去醫院看病，醫生一般會先讓病人驗血，再從化驗單判斷是病毒感染，還是細菌感染，或是其他感染因素，根據化驗單的指標對症下藥。數據框架的作用正是如此。

比如，某個電商網站今天的成交額提升了30%，那麼，這是不是一個值得高興的數據呢？我們無法直接給出答案，此時，就應該讓數據告訴我們，假設是因為投放外部廣告，才使成交額提升，這就是一個客觀的描述，我們要做的，就不是看成交額的高低，而是評價廣告效果的好壞。

在電商領域，評價業務水準通常有兩套指標，一套是常用的計算成交額的方法，即「流量×轉換率×客單價＝成交額」，它能夠評價一個類別、一個商品的健康程度。

另外一套指標則多用在大型促銷活動的時候，即「活動預熱期加入

購物車的商品數×商品單價×經驗轉換率×經驗成交額占比＝活動成交額」，因為經驗轉換率和經驗成交額占比通常是兩個固定的數據，所以只要看購物車有多少商品及單價，基本上就能算出大概會有多少成交額了。這就是數據框架的作用，使我們清楚業務的狀況。

但是，業務是好是壞，需要不斷地進行比較，單獨一個數字是沒有意義的，比如，180這個數據，從表面上看，沒有任何意義，但是如果有比較，這個數據就有了意義。如果是身高180公分的男性，那就算高個子，因為中國人的平均身高可能只有170公分；如果是汽車時速180公里，那就是危險駕駛，因為高速公路的限速是時速120公里，已經超速50%了。

推及至業務上，業務需要進行比較才能判斷好壞，而比較的前提，就是要不斷地尋找比較的對象，以電商的大型促銷活動來說，至少需要三個參照組的比較，才能區分業務的好壞，分別是：這次活動對比平時的漲幅、這次活動對比上次同等規模活動的漲幅、這次活動的同期漲幅對比上次活動的同期漲幅。

經過這些比較，才能客觀地評論大型促銷活動的效果到底好或不好，如果沒有選對參照組，就很可能得出錯誤的結論。要選對參照組，就需要分析師具備很強的商業敏感度，而這種商業敏感度，就來自「混」和「通」。

除了搭建數據框架之外，好的展現形式也必不可少，最典型的就是圖和表。但是，現實中很多人不知道圖和表到底該用在什麼地方，現在就介紹幾個小訣竅。如果業務情境是要查找精確的數據，這時候用圖就錯了；如果業務情境是要觀察變化趨勢，卻用了表，那也錯了。一個好的分析師，不論是設計數據產品，還是製作圖表，都需要很好的「曬」的意識，讓「曬」出去的數據能夠吸引到更多人的注意。

用數據提升業務

數據如何改變，可以讓業務更好？這個過程就是透過數據發現問題，並解決問題。就如同前面提到的化驗單的例子，透過數據的分析找出問題的癥結，並進行針對性地優化，這樣的例子在阿里巴巴幾乎是天天都在發生，以下就是一個相關的例子。

淘寶有一個「二手閒置」業務，與淘寶主站的筆記型電腦類別有一個以舊換新的合作專案。用戶成功在二手閒置網站「甩舊」（賣出二手商品）後，就可以獲得在主站的筆記型電腦類別以優惠價「買新」的資格，完成「甩舊＋買新」這兩步驟後，還可以獲得數百點的集分寶。從專案描述來看，整個交易原理並不複雜。

那麼，在專案上線後，整個市場反應如何，用戶買單嗎？根據淘寶後台的數據顯示，整個活動的流量漏斗效應非常嚴重。大量參與活動的用戶中，最後成功「以舊換新」的寥寥無幾，整體轉換率非常低。那麼，問題出在哪裡呢？事實上，還是要從數據切入，查找原因：

- 「估價」到「售舊」環節，用戶流失率最高。
- 在流量來源監控中，其中一個管道的轉換率幾乎為零。
- 手機及平板的回收轉換率明顯高於筆記型電腦。
- 在成功回收的商品中，宏碁、聯想、三星等品牌占比最高。
- 在成功購新的商品中，價位以3000～4000元人民幣為主。
- 來訪用戶中，以男性居多。

透過以上數據反映出的問題，我們可以從四個方面來改進：

- 整體步驟過多，漏斗效應明顯——首先針對「估價」到「售舊」環節進行優化，簡化流程。

- 儘管活動帶來不少流量，但部分流量幾乎無後續轉化——流量來源需要調整。
- 用戶對品牌、價格有偏好——針對櫥窗展示商品的品牌和價格進行優化。
- 來訪用戶以男性居多，細化分析後發現，21～25歲和26～30歲的年輕男性為主要群體——廣告定向投放。

基於以上的數據和發現，後續的數據落實行動也就清晰起來，具體做法有以下三個方面：

- 流程：以簡化為原則，針對用戶流失嚴重的環節進行較大的調整。
- 流量：第二期流量投放上，選擇以年輕男性為主的流量來源，放棄有流量而無轉化的管道，提升精準度。
- 商品：將聯想、宏碁等品牌放在較明顯的位置；下調推薦商品的價位，由原來的7000元人民幣以上，下調至3000～4000元或4000元人民幣以上。

經過這樣的調整後，淘寶「以舊換新」第二期活動成效明顯，整體轉換率提升了六倍左右。

用數據發現機會

利用數據幫助業務發現機會，涉及具體數據在業務場景中的應用，這裡最核心的價值就是，使數據變成人人都要用，而且是人人都必須用。

我曾經提出一個理念，即人人都是數據分析師，也就是讓公司的業務人員都能夠透過數據做決策，讓數據驅動業務，實現真正的數據化營

運。

有個團購網站就做得非常好，營運高層和一線營運主管都有非常好的數據意識，他們和數據分析師「混」和「通」了之後，非常清楚知道數據能夠解決什麼問題，他們和數據分析師一起針對網站的主要流程進行剖析，最終得出以下五個觀點：

1. 網站流量和頁面有限，必須個性化。
2. 商品是公司成敗的關鍵，必須選對商品。
3. 銷售過程中可能出現斷貨的情況，必須預測銷量。
4. 價格是競爭的核心，必須關注競爭對手。
5. 服務是重點，必須重視發貨速度和用戶投訴。

定位了這些問題之後，就開始著手解決，針對個性化、選品和預測銷量的問題，數據分析師迅速跟進，建立數據模型。針對價格和服務的問題，他們建立了專屬的預警系統進行觀測。如此一來，數據營運得非常順利，對業務的提升也有很大的幫助。

用數據創造新的商業價值

我舉阿里金融的一個例子來說明。阿里巴巴有一句話是「讓信用變成財富」，阿里金融正是這樣。阿里金融的核心就是利用數據計算客戶的信用水準，並根據客戶的信用進行授信，讓客戶能夠方便快捷地獲得需要的資金。基於這種信用模式，有越來越多的相關產品被開發出來。當中的核心就是結合業務和數據，將數據「曬」出來，形成一種競爭力。

「曬」是產生力量的過程，以上是數據化營運中的「曬」，同樣適

用於企業內部。對數據分析師來說，說服別人才是「曬」，以產品情境來說，需要說服產品經理把某個東西放在產品裡，這就是數據工具體驗上的一大難題。

公司架構上沒問題了，但要是數據工具不方便使用，使用者體驗自然不會滿意。所以，數據能不能做到在獲取、使用、分享、協同、連接、組合上，變得超級簡單和便捷，這是數據化營運非常重要的一點，也正是「曬」的內容。

數據分析師一定要讓數據變得簡單，只有這樣才不會出現如前面所說的，收集數據的人不知道怎麼用，用數據的人不知道數據是怎麼來的。正因為數據很複雜，倘若數據分析師無法將它變得簡單，管理和營運的工作人員自然會產生困惑，也就不可能真正做到「通」和「曬」。

阿里巴巴的「混、通、曬」三訣竅，其實是配合了數據方法論與人的修練，能做到借事修人，讓用數據的人在數據中成長，循序漸進地讓每個人都成為數據分析師。

關鍵思考 | 思考，要學會關視窗

我一直覺得，每個人每天思考能力和池子的水一樣，是有限的。每個人情況不一，有些人思考能力強一些，就可以做更多的思考；有些人思考能力弱一些，只能做更少的思考。但不論思考能力強弱，每思考一次，都會讓自己一天的思考池子的水減少，降低自己的思考和決策能力。

有這種想法的，不僅僅我一人。我曾經在《哈佛商業評論》網站看過一篇文章，很有意思，標題是〈Boring is Productivity〉，中文直譯過來就是「無聊是一種生產力」。文

章裡面提到一個案例，有人問美國總統歐巴馬：「為什麼你每天都穿一樣的衣服？」歐巴馬的回答是：「因為這樣可以減少我思考要穿什麼衣服的時間，我就可以將更多的時間留下來做更有價值的決策。」這樣的例子在文章中可以找到不少，其核心意思就是，我們每天思考能力是有限的，如果要做更有意義的思考，就不要把有限的思考資源浪費在不能產生價值的思考上。

這個觀點，套用社會上一個比較時髦的詞語來說，就是思考也要注意「節能環保」，我們要保護的是自己有限的思考資源。至於如何做到節能環保，我自己是使用「關視窗」的方法。每個使用電腦的人都清楚，如果開啟過多視窗的話，會有兩個可能的影響：一是增加電腦的負擔，電腦運行的速度會變慢；另一個是，即便電腦的性能夠強大，不至於讓視窗影響運行的速度，但我們的大腦也無法同時處理多個資訊。我們能做的，就是關掉多餘的「視窗」，認真考慮一件我們認為非常重要的事情。

關視窗，在電腦世界中很容易執行，但在現實生活中卻很有難度，尤其是在動態的工作過程中。為什麼我會覺得困難？因為我覺得，在我們思考的過程中，永遠不可能只有一件事情可以做，對管理者來說尤其如此，不斷被其他人影響，不斷被即時通訊軟體的訊息通知打斷，不斷被電子郵件提醒去做事……這一切都會讓我們的思考中斷。所以，才會有不少時間管理方面的學者建議，每天要處理很多事情的大忙人，可以在一個固定的時段處理電子郵件，以提高辦事效率。

但是，在動態的過程中，典型的狀態就是一場談話或一場會議，那就不像不收電子郵件和不開即時通訊軟體那麼簡單了。談話的過程是動態的，而且這個過程可能不會嚴格按照預

定的會議主題來進行，你會被各種細微末節的小問題和各種情緒干擾，最終導致浪費了不少時間，卻沒有得到應有的效果。

對於這樣的動態談話過程，我自己的做法是，永遠只開一個視窗。意思是，在談話過程中，我的大腦中始終有幾個問題會放在首位，這幾個問題不會隨著談話的過程而隨意改變。這幾個問題包括：我這次要解決的是什麼問題？對方（包括客戶、同事等）要的是什麼？現在提出的方案解決這個問題了嗎？在只開一個視窗的前提下，如果會議中出現不符合主題的內容，我一般不會過多參與，或是有意識地結束這些與會議無關的話題。

以前，我參加一些電商的會議時，總是會被一些細節干擾，即便最終提出建議和解決方案，也可能不是真正的解決方案。比如，在會議之前，我可能想透過解決產品的問題來提高轉換率，不過，這些問題一旦放到會議中討論，就可能會受到各種因素影響——管理者可能會說是產品的問題；而產品部門總是有數據可以說明他們的產品沒問題，實際上是營運能力有問題；而營運部門又會拿出種種數據來解釋營運是有效的，是吸引到的用戶有問題。結果往往是會議離主題越來越遠，如果再碰到一個不是特別善於演講和辯論的市場部負責人的話，問題的矛頭最終可能將指向市場部，市場部就會背起這個黑鍋，至於到底是什麼問題引發轉換率低，其實誰也不知道。

這種時候，對我們干擾最大的就是數據，每一份數據一定都是精心準備的，看起來在邏輯上非常嚴密，對於那些快速思考的人來說，非常容易獲得認同。所以，此時我就一定不要被各種其他因素干擾，如果在每一個細節上過度爭論，會減少我對最終核心問題的判斷精力。

我會在心裡告訴自己，我的精力是有限的，不要把時間浪

費在不值得的地方。這又是一句說起來很容易，但是做起來非常難的事情。比如，開會的時候，並不是每個人都是決策者，他們一定會從自己的角度出發，一定會提出各種觀點，此時我會做的事情就是閉上眼睛，好好想一想，到底我們要解決的是什麼問題，我應該用什麼邏輯來解決，在拋開一切雜念，獲得驗證的思路之後，再來看細節問題，決策就會顯得非常簡單。其實，以邏輯來思考的問題或發現的問題，就是一個本質問題。

第10章

存、管、用——阿里巴巴的營運數據三絕招

2011年，阿里巴巴開始從數據化營運向營運數據發展，並形成一個良性循環——從數據化營運的三訣竅「混、通、曬」，走到了營運數據的三絕招「存、管、用」。

存：數據收集的開始

「存、管、用」，從「存」開始，營運數據的第一件事情就是把數據收回來、集中起來、儲存起來。

為什麼要收集數據？因為過去的數據化營運產生了一些問題，雖然我們解決了一些，但還是有一些問題並沒有得到解決。當談到要如何更妥善地解決這些問題時，我們想到了運用數據可能是有效的，在這樣的情境中，我們開始收集數據、管理數據和使用數據。

以電商的數據收集為例，比如，在婦幼類別，我們可以透過填寫寶寶的年齡或購買的特定品項（不同年齡層的奶粉、尿布和衣服）等，推算寶寶目前的年齡階段；在汽車類別，我們可以透過客戶購買的機油、空氣濾清器型號等，推算客戶是否有汽車，以及車型。甚至衣服也可以用來作為數據收集的情境，比如，從一個用戶購買衣服的歷史尺寸，觀察用戶是否有身材上的變化。關於數據收集，前文已經有較多篇幅介紹，這裡就不再贅述。

就數據的收集而言，最重要的不是看我們收集了什麼數據，而是要思考，這些數據如何使用，以及收集這些數據能夠產生什麼作用。用一

句話來說就是，收集數據不是目的，收集起來的數據如何產生價值，才是最終的目標。不過，如何收集在未來具有價值的數據，的確是一個難題，當中就需要一些經驗判斷了。

管：保護好儲存的數據

數據儲存之後，數量和廣度都很大，需要有完善的管理。數據管理的內容包含很多方面，比如，數據的來源、如何讓數據不遺失、如何保護數據的安全、如何讓數據準確和穩定，以及如何更妥善地運用數據，這些都是數據營運中的「管」。

然而，「管」並沒有一個標準可循。我在和一些投資人及國外網路公司的數據專家交流時，他們都認同我的一個觀點：大數據管理到底要怎麼做？答案就是「沒有確切答案」。

其實，關於數據的管理，整個大數據產業和其他產業一樣，都經歷過很多起起落落。比如，IT技術剛興起時，各家公司對IT和OA（辦公自動化）系統的追隨十分盲目，最後卻沒有產生期望內的回報。之後，資訊化管理開始興起，MIS（執行資訊系統）被很多企業推崇，不管企業規模大小，都開始使用ERP（企業資源規畫）和MRP（物料需求規畫）系統，結果，仍然沒有達到預期。就數據而言，2004年左右，美國的一些數據管理經驗在中國造成很大的轟動，許多網路公司紛紛建立BI（商業智慧）團隊，但是到了2009年左右，各家公司又開始不完全認同BI數據部門，這也是事實。

但也正是在那個時候，中國頂尖網路公司的數據化營運開始啟動。阿里巴巴的數據管理，最不一樣的地方在於，阿里巴巴非常依賴數據產品，希望用數據產品來解決獲取及使用數據的問題。

當時，我們犯了一個很久之後才發覺的錯誤，那就是，我們的做法

跟美國一些公司的想法一致，即「不管怎麼樣，我們先收集數據，將來肯定有用」。我認為，在我職業生涯的這些年，最值得反思的就是這句話。因為沒有一家公司可以讓你無止境地收集數據，之後再使用，這根本是不現實的。

這就是不做決定的代價。因為，在這個世界上，有一些決定我們一定得做。從營運數據的角度來說，如果我們只收集數據而不做分析和應用的話，代價就是很沉重的儲存成本。

無須多言，這種儲存成本的代價是巨大的，即便是一家富有的公司，即便它機器比較多，也只能短時間承擔這種損失。

即使是阿里巴巴這樣的公司，到最後也會發現，這樣做是不行的。因為不管你有多少機器，這些數據都在呈指數式成長，當提到怎麼備份時，問題就出來了。

你一定會問，在這種情況下要如何備份？此時，你就要決定，什麼東西需要先備份，什麼東西可以先放在「冷藏庫」裡。「冷藏庫」的意思是一些成本比較低的伺服器，但是放在裡頭的數據不能隨時使用，需要調出來才可以使用。

針對這種情況，有人說，我們把三年前的數據都放進去，夠嗎？答案是：還是太多了。有人說，那我們可以把一年半前的數據都放進去吧？我說不行，因為利用數據觀察業務發展趨勢的分析師，一般都要看三年的數據，所以這種做法也不現實。

在面對「決定什麼數據要放進冷藏庫」和「決定什麼數據在緊急情況下一定要保護」的問題時，你就會發現以前我們所講的觀點——數據先收集起來，將來再使用，完全是一個偽命題。以前，從來沒有人對這個偽命題表示過異議，無論是銀行，還是其他金融機構，甚至是網路公司。而當大數據出現後，這個觀點就成了一個藉口、一個偽命題。

是的，這是一個很難下的決定，但這就是你必須做的決定。如果你在將來的某天，發現你需要的數據沒有提前保存的話，就只能錯失這次

的發展機會。事實上，這就是企業的博弈。

你可能會問，一家企業並不會從事所有的商業，那為什麼所有的數據都要收集？事實就是如此，這是數據人在管理上的不負責任，但平心而論，這個責任也非常難以承擔。

大公司正在數據管理這條路上學習，當前，我們面臨很多以往不曾預見的問題。比如，關於數據管理，我們應該在各個部門運作，還是應該集中管理？我們應該在數據安全的前提下更開放，讓更多人找到數據的價值，還是應該更封閉，讓洩露數據的可能性更小？另外，個人隱私怎麼去保護？我們如何才能成為一家負責任的數據管理公司？這些都是代表性的難題。

現在，大型的網路公司通常同時擁有成千上百個正在開發的專案，它們都直接或間接地改變著數據，在這種情況下，如何保障數據安全？事實上，數據的源頭已經「髒」了，而下游使用數據的人還不知道。同時，源頭的數據使用者也沒有責任告訴下游的人這些數據已經「髒」了。

所以，我的觀點是，如果你本來數據就用得不好，這對你的發展也不會有太大的影響。但是，如果你數據用得好，並且將它作為公司的核心競爭力，那麼你的麻煩就大了。因為你的數據來自各個地方，而每一個來源都沒有責任要告訴你，今天的數據是正常的、可靠的。特別是在大數據出現後，數據的精準與否更加重要。因為，大數據在很多情況下，是利用外部數據來協助調整內部數據，如果你的內部數據難以保證「乾淨」的話，那麼外部數據同樣無法保證「乾淨」。

數據管理，是大數據產業的髒活、苦活和累活，是最悲催、最難解的事情。如果沒有這些背景支持，我對很多公司在做所謂的大數據營運，就持有懷疑態度了。所以，我覺得很多所謂的「大數據專家」都是作家，而非真正的數據人。

用：從收集數據到管理數據

到底該如何應用數據，才能解決當今公司存在的一些問題？在「存、管、用」的過程中，如果產生問題，我們會詢問是決策的問題，還是執行的問題。如果是決策問題，我會責問相關人員是否明確了收集回來的數據的作用，以及有沒有保管好這些數據，倘若不知道，那就是一個偏差。如果不是數據結合過程中出錯，就是數據來源出了問題，那就要針對數據的源頭——「存」數據，更加嚴格。這時候你會發現，「存、管、用」也是一個反覆驗證的過程。

在「存」數據上，我看到的問題是，很多人往往只看到數據的表象，比如，根據有過購買行為的用戶來做數據收集，大部分在「冰山」之下的數據就放任不管。收集數據的面向是多種多樣的，假設什麼數據都可以獲取，如果中國每家醫院的數據庫都是開放、可以合作的，那麼這個數據就是既龐大又準確。

在「用」數據的問題上，數據的分裂和重組，都能做到顛覆性創新。用戶的生理性別是決策的重要依據，而現在可以在購物屬性上變成「早男晚女」，這種改變並不違背營運數據的原則，因為這是購物屬性。這個創意來自「本來不可分裂的東西分裂了，本來不應重組的東西產生了新的力量」的理念。

舉例來說，你只知道用戶的地址，再根據房地產租售服務平台安居客，得知房子的租金，加上一些外部資訊的補充，比如，十公里內有沒有銀行和電影院，有沒有大型購物中心，你就可以在數據世界中重現他的財富圈和生活圈。當這個地址是辦公室時，就可以得出他的辦公生活區和日常生活區這兩個生活圈。

從使用數據的角度來說，電商產業有很多值得其他產業借鏡的地方，可以讓數據真正地被使用，並且產生實際的商業價值。先來看看現在電商的背景，不論是以阿里巴巴為代表的平台型電商，還是以京東為

代表的自營型電商，或者以1號店為代表的垂直類電商，它們的共同特點就是商品種類非常豐富，商品數量動輒百萬、千萬，平台型電商的商品數量可能更多。

對消費者來說，當他進入電商網站的首頁，並不需要看到那麼多商品，如果消費者有明確的購物訴求，那麼，他可能會直接進入電商網站的搜尋引擎，開始尋找商品；如果沒有明確的購物訴求，則可能在電商網站提供的分類和促銷活動等區域隨意尋找。這時問題就來了，頁面的內容有限，消費者的時間有限，而且需求有偏好，但是商品數量非常大，電商的目標又是為了讓消費者能夠透過閒逛產生成交額，那麼，如何找到合適的商品放在首頁，就成了問題的關鍵。

面對這樣的問題，我的解決方案是透過一套數據中間層，生成使用者在特定市場的個性化標籤。電商企業不同類別的營運員工，透過演算法或人工選品來實現使用者標籤和商品的匹配，從而實現用戶「逛」的效率最佳化，進而提高用戶的購買轉換率。

如何建立用戶標籤？

建立標籤，簡單來說，就是透過數據的分析，對用戶的偏好進行描述。建立標籤通常有以下三種方法：透過業務規則結合數據分析、透過模型，以及透過模型的組合。

方法1——透過業務規則結合數據分析來建立標籤

這類標籤和業務人員的經驗緊密結合，這裡可以舉幾個例子，以對這類標籤的設置有更加具體的認識。

比如，業務人員可以從用戶購買的汽車，判斷他可能擁有這款車，此時，就可以透過數據進行分類，把用戶分為福斯Golf車主和福特

Focus車主等，這個時候，當用戶進入汽車配件類別時，就可以直接為用戶推薦相應的汽車配件，直到用戶出現搜尋別的汽車用品的明確行為，再進行數據調整。

再舉一個例子，有些用戶平時很少網購，但是一到大型節日前就會大量購買商品，這一類用戶通常都是企業的採購人員，這時候就可以在禮品等類別做相關的企業禮品推薦，甚至直接推薦自家網站的儲值卡。

還有，對於中老年人的識別，可以透過用戶經常使用的位址和包裹的寄送位址來區別，比如，經常寄送包裹到另一個固定地方，再結合購買的商品，就可以判斷包裹的接收人可能是年長者，給用戶貼上「家中有年長者」的標籤，在衣服類別中，就會出現中老年服飾。

方法2——透過模型來建立標籤

比如，用戶在婚慶類別的特定行為，當然，特定行為是透過數據模型識別出來的，此時，我們就可以判斷這個用戶即將結婚，這樣就可以結合時間，為用戶貼上婚慶的標籤，也可以持續觀察這一類用戶，在未來可能會貼上居家裝潢和婦幼的標籤等。

結合用戶的手機儲值和收貨地址等行為，就可以用模型計算出該用戶是否是為自己購買，還是作為網購的中心者，為他人購買。如果能判斷他經常為他人購買，則可以貼上類似「網購影響力中心」這樣的標籤，在不同類別的場景中運用。

方法3——透過模型的組合來生成新的標籤

任何一個模型都是有生命週期的，或者說，企業內部不同的模型設計人員，可能會對同一個用戶做出不同的判斷，所以，我們需要不斷地整合模型。一般情況下，可以採用模型投票的方法，從多個模型中找出合適的標籤。比如，在三個模型中，兩個模型認為寶寶是3~6個月，一

個認為寶寶是12個月以上，透過模型的整合，應該可以判斷寶寶為3～6個月。

如何有效應用標籤？

應用標籤，是指在電商網站的首頁或類別頁，進行標籤的使用。標籤的使用，最核心的關鍵就是數據中間層和前台業務層的對接，以及能夠讓營運人員非常方便進行商品的設置。

這裡涉及兩個核心點，一是中間層和業務層的銜接，二是中間層的易用性，下面分別就這兩個內容做探討。

中間層和業務層的對接

目前，對接在網路廣告中是非常熱門的概念，典型的應用之一就是DMP（數據管理平台）。在這個系統中，用戶以標籤化的形式存在，也就是說，之前建立的用戶標籤有了一個管理的平台，終端使用者可以在這個系統中選擇用戶，選擇完成後就會產生一個投放計畫。DMP還會和前台業務平台打通，簡單來說就是，用戶登入首頁之後，系統就會認出用戶身上的標籤，便可以根據DMP設置的計畫，產出不一樣的內容。

中間層的易用性

對終端使用者來說，選擇標籤需要夠簡單，並且能夠非常清楚地知道這個標籤具體代表的含義是什麼。對數據從業者而言，讓數據變得超級簡單，是非常重要的使命，所以，介面設計和後台管理等內容都十分重要，否則可能會失去標籤系統的價值。

對大數據來說，「用」是讓數據發揮價值的最大一步，在這裡，我也只舉了一個數據應用的簡單例子——標籤系統。這是數據和營運數據

緊密結合的一個案例，也是數據營運或數據驅動的一個典型案例。

只有先結合大數據的技術，將數據化營運做好，才能讓數據從成本轉化成利潤，真正發揮出大數據的價值。

關鍵思考 | 用化骨綿掌解決本質問題

大多數人可能都習慣在現象層面思考問題，因為不需要花太多的力氣思考。事實上，我們的邏輯思考是一個慵懶的系統，如何找到本質，才是解決問題的關鍵。

一方面，要找到本質問題，需要我們在思考時排除過多的東西，使我們能夠更集中精力去解決重要的問題；另一方面，我們要有方法發現本質問題。能夠透過現象看本質，是一個非常重要的方法，對這個方法最簡單的描述就是——不斷用邏輯方法分解進行問題，直到不能分解為止，然後從根本處去解決這個問題。

我們還是以電商產業為例，說明如何從現象出發，找到本質。大部分的電商，其實傳統產業也是一樣，關注的都是成交額，因為利潤就來自於此。但如果僅僅問一個問題：如何提高成交額？那就非常難回答了。所以，我們必須把這個問題分解了來回答。

通常，在電商產業，我們會把成交額拆分成幾個指標：日均UV、瀏覽轉換率（UV到達商品介紹頁的轉換率）、購買轉換率（進到商品介紹頁的用戶到達購買的頁面）、筆單價（平均每一筆的金額）、人均筆數，這五個指標。

不過，值得注意的是，這些指標都是平均值，從診斷整體

健康程度的角度來說，這些指標是可以使用的，但是，如果要用來分析更為細節的業務或應用於別的公司，就要注意數據的陷阱，盲目套用，很有可能會造成巨大的損失。

在指標分解完成後，我們就可以用這些指標來分析公司的業務。我們把這些數字圖形化，見圖4，這樣是不是更直接指出了問題所在？最近一週成交額上漲，是因為筆單價提升。

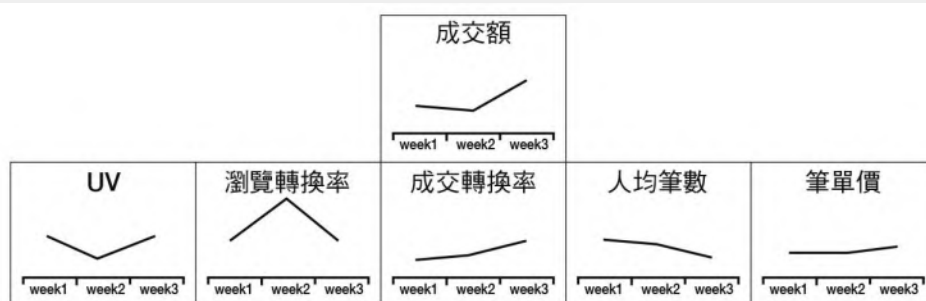


圖 4 五大指標對電商 A 第 1 ~ 3 週業務的影響

我們再來看看圖5，這會讓問題更直接，能夠使我們更清晰地看到問題所在。從總體交易額的角度看，圖5和圖4趨勢差不多，但是，當我們更深入分析原因時，發現前兩個指標出現了截然不同的走勢。

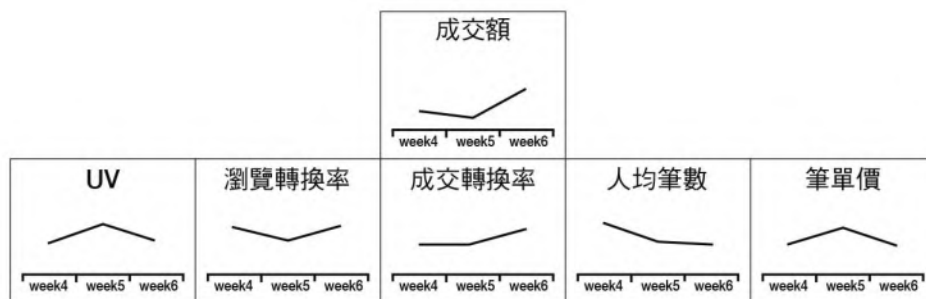


圖 5 五大指標對電商 A 第 4 ~ 6 週業務的影響

如果我們不深入分析，只是不斷猜測到底是什麼因素導致

銷售額下降，也就只能做到猜測的層面，對企業的發展毫無實際價值。

在分析問題的本質時，有一個很好的方法可以用，叫作 MECE（Mutually Exclusive and Collectively Exhaustive），意思是彼此獨立、完全窮盡，這是來自麥肯錫的方法。這個方法很像圖4和圖5這兩個分析銷售額的分解圖。在解決問題的時候，只有做到彼此獨立且完全窮盡，才能夠找到問題最深層的本質原因，才能夠就這個原因找到合適的解決方案。

關於銷售額的分解圖，其實，在不同的業務場景下，還可以進一步細分，比如，UV還能夠細分成廣告和非廣告，或者首次和非首次等。

總之，當我們遇到問題時，一定要問自己，這到底是現象還是本質。如果是現象，那就一定要找到本質。被表面現象迷惑，是很多人都會犯的錯，而所謂的成功，就是少犯錯，也就是早一點從本質出發來思考問題。

第11章

大數據，未來商業的利器

我們正處於決策成本巨變的爆發點，過去那些想盡辦法都無法獲取的數據，在今天唾手可得；當一些表面上看起來完全不相關的產業數據連結起來時，居然產生了新的商業價值。

更重要的是，過去，我們常常是帶著問題去尋找能夠驗證自己觀點的數據，而今天，我們卻可以使用數據去預測可能出現的問題。海量數據可以使人的智慧得到更大的發揮，並且變得更加規模化。

大數據的本質是人，數據研究的極點就是難以捉摸的人性。一旦我們掌控了數據後的數據，就擁有制勝未來商業的無敵利器。

假設數據是髒的

美國有一家新創公司，專門做與地理位置相關的數據收集、整理和查詢服務，它將地理位置的相關指標，按照飯店和旅館等屬性，劃分為不同類別，對外提供基於位置資訊的即時查詢，為包括美國最大評論網站Yelp在內的多個知名應用提供底層數據服務。

這家公司最令人印象深刻的是，對於收集來的數據，它會提供數據品質評分，以反映數據的可信度和品質水準。它會對這些數據的源頭，以及處理數據階段所使用的演算法進行評分，也就是說，這家公司在提煉數據的每一個階段都進行了數據化管理。

這家公司的做法讓我們看到一個趨勢，也是非常重要的一個趨勢。它先接受了數據來源肯定是髒的，以及數據來源一定會被污染的事實，

所以，當它在處理數據的時候，會像汙水處理廠一樣，每一步都問自己要如何處理這些汙水。到底是因為數據來源髒了，還是因為數據提煉過程做得不好？這個過程我們一定要區分，而且這樣的區分是可取的。

這家公司做管理數據，是假設數據是髒的，而不是假設數據是穩定的。以假設數據是髒的來處理數據，在大數據時代將是一個非常重要的趨勢。

事實上，我們今天在處理的大數據，依然只是冰山一角，而更大的數據隱藏在我們的語言中，比如，我們說的話和寫的字。所以，將來我們要準確地從互動中抓取數據，也一定得依賴對自然語言的處理。現在，美國很多數據研究人員都在瞄準非結構性數據，即語言處理這一領域。

學會盤點數據

想確定數據的優先順位，就要先解決以下幾個問題。一是數據的標準化，在大數據時代，我們需要一個標準化的東西供我們進行交流。二是數據的穩定性。我們到底如何對接和交換數據，如何在交換的時候保持數據的穩定性。比如自然語言，比如在行動裝置和電腦不同情境下受到的影響，這些情況都會衍生出許多新問題

第三個重要問題是數據的儲存，這將涉及數據的時效性問題。有人曾經提出一個很有價值的觀點，即現實中，網站最大的情境變化就是網站改版，因為重新設計網站就會影響數據，例如公司的介紹頁和首頁，任何改變都會影響數據。如果等到一年至三年後，你才說得出數據的改變是因為促銷、用戶行為，或是改版所引起，這個數據就已經沒有任何價值了，這就是數據的時效性。

所以，美國出現一個概念，叫作數據淡化（data decay），意思很

明顯，數據會慢慢淡化。我們要更清楚地認識到，數據是有優先順位的，在數據中有些是特別核心的，有些即使遺失了也沒有多大問題。我們要學會坐下來，好好盤點那些對公司最有價值、對用戶最有價值的數據，這是一個非常重要的趨勢。

數據的標籤化管理

數據的屬性標籤是人類經驗判斷的數據，是數據後的數據。例如，當你要為一件物品貼上標籤時，其實就已經動用了你的經驗數據分析，並進行歸納總結，結合當下的環境做出判斷。這種經驗加上直覺的判斷，是不穩定且難以解釋的，但是從數據收集的角度來看，數據的屬性標籤又是一個潛力極大的數據。

在數據屬性的管理上，對用戶來說，每個人身上都貼著各式各樣的標籤，但是對企業而言，如何統整這些標籤，如何用一個點去串聯標籤，又如何把這些點連起來去描述這個用戶，這才是核心問題。

比如，你要去應徵一家公司，A公司對你的評價是「很可靠」，B公司對你的評價是「不可靠」，而C公司準備雇用你，卻看到了A公司和B公司給你貼的兩個標籤：可靠和不可靠，就會感到困惑。

這樣的問題常常發生，那麼，我們應該怎麼管理標籤呢？

首先，我們要明確的是，「可靠」這個概念是沒有標準化意義的，除非在定義標籤之前，我們就清楚界定了「可靠」的標準是什麼。否則，「可靠」的標準是準時還錢、說話算數，還是他向來都很守時，我們將無從得知。如果這些標準就是「可靠」，這就給了我們得以還原數據的能力。所以，在屬性管理中，假設屬性是「X」，那麼，我們一定要定義清楚什麼是「X」，在沒有定義清楚的情況下，這個數據的屬性毫無價值，而且將來你依然不知道怎麼使用這個數據。

在經過觀察後，將標籤加進平台，和直接加進去是不一樣的。在電商平台中，就有一些標籤是在觀察後加進去的，如果由賣家自由地加標籤，必定會造成混亂。所以，標籤的屬性管理，在營運數據中非常重要。

標籤化管理，是一個非常重要的趨勢。電商企業今天面對的一些問題，在美國的電商企業中同樣存在，可見，我們發現的問題，別人也在面對，不過這些問題並不需要現在去解決。

屬性管理的層級化十分必要，但是在使用數據前，必須要了解數據的情境，數據是如何放進去的，以及數據的情境是什麼。在這一切未知之前，就說數據如何好用，那是不可能的。所以，現在企業營運數據的趨勢是，我們應該找出一些屬性進行歸類，再慢慢考慮如何提煉，這對於未來非常重要。

數據和數據之間的關係才是重點

Google做了一件非常驚人的事情，它甚至能在不明白某個網頁語言的情況下，就知道網頁的內容是什麼。

試想一下，如果你懂俄語，要看懂一個俄語網頁在講什麼當然很簡單。但是，如果你僅僅透過看字詞的排列和網站的分類，就知道網頁的內容，這是不是很令人驚嘆嗎？

這就是知識圖譜（Knowledge Graph），它是一個無窮盡的世界。事實上，知識圖譜並不是數據，而是數據和數據之間的關係。但這裡有個大弊端，就是數據的儲藏量非常大、儲存的方法也很複雜，且稍微改變一點點關係的定義，整體就會產生巨大的變化。

比如，有個知識圖譜在講電商平台用戶之間的關係，那麼數據資訊就非常龐大了。試想，今天電商平台裡有多少個用戶跟你有關係？假設

有二十五個人，這二十五個人的關係就演變成了 25×25 條關係。這時候，我再問你，什麼是關係？見過面就算關係，還是一起買過東西才叫關係？這問題就具備了一定的難度。

建立關係的維度是無限的，而且只要定義稍微改變，整個儲存和整個數據庫都會發生變化，所以，控制知識圖譜是有難度的。

舉個貼近我們生活的例子，比如，銀行讓你辦信用卡，促使銀行做出這個決策的，不是你的個人關係，而是總關係。銀行決定是否貸款給你，還要看你的配偶是什麼職業、家人的經濟情況如何，當這種種關係連起來時，就會產生一個極為重要的知識圖譜。

以往我們談大數據時的本錢，莫過於「我有這種數據，你沒有」。在未來，數據和數據之間的關係才是重點中的重點，而不是單純的數據本身。

數據的即時化處理

以上談到的很多內容，都是關於數據收集和數據管理層面，而在數據的處理上，我在LinkedIn上看到一個很有趣且很有價值的做法。LinkedIn在處理數據時，會把公司的數據服務分為幾層，一方面是緊急性，另一方面是重要性，比如，它會把數據分為「快數據不重要」、「慢數據重要」等級別。

我覺得，對數據即時性分層的做法是合理的。有些人覺得數據的即時化處理是趨勢，但是我持有一定的懷疑態度。real time是「即時」，right time是「恰當的時機」。在我看來，數據處理不一定要即時，因為在其他數據沒有到位的情況下，數據即時化的價值也不大。比如，我們常見的情況是，每家公司都有財務的相關數據，這類數據的處理通常都是「T+1」，意思是，你想要的數據在第二天才可以拿到。

但是，換個情境來看，如果銀行要判斷用戶的信用卡是否被盜用，那就一定需要進行即時化處理。

我們從商業的角度看數據的即時化，值得注意的是，在具備了即時化的數據處理能力後，很多以前不能解決的情境，開始變得能夠解決。在未來的某一天，編碼的工程師能夠在編碼時直接寫上「如果一個三天前只瀏覽、未購買的客戶回來了，就給他一個2元人民幣的紅包」。這個程式是完全編好的，用戶登入就成了一個即時標籤，快速的運算會讓每個網站都具備最強的時效性。

我們再換一個角度思考，如今，手機、電視、遊戲機和電腦等多螢運作的時代下，一個網站有多大的能力，能夠以非常快的速度，在極小的時間點裡抓住消費者，賣出產品，這種即時的能力會在未來的商業中變得越來越重要。

一個網站必須要讓自己的即時能力更高，甚至用戶接下來的每一步你都應該猜到。但是，我們千萬不要把所有的能力都用在處理即時化上，因為依然有大量的數據需要在「恰當的時機」處理，有的數據是重要但不緊急。

未來是人機的結合體

很多人會問，大數據目前發展到什麼階段了？我的答案是——落差還很大的階段。

毋庸置疑，大數據已經大大地影響著我們的社會，但還遠遠沒有達到它的爆發點。因為有了大數據，人的經驗開始和數據結合，兩者相互刺激、活化，讓人的智慧規模化地放大，這也使得整個社會開始伴隨著數據的發展，產生巨大的改變。

但是，在大數據的使用上，未來的發展空間注定會比現在取得的成

績更加宏大。「數據化營運」和「營運數據」這個數據與人之間的閉環系統也會越來越完善，人機結合仍然有巨大的空間等待我們去挖掘。

未來，數據的種類將超出我們想像。以前，數據大多集中在外部行為的監控上，不論是網路購買行為，還是網路社交行為，都是用戶在遠距離提供數據，即便這樣，我們仍然還沒有好好運用這些數據。

隨著穿戴裝置的出現，數據和人將真正融為一體，類似Google眼鏡這樣的裝置，將讓我們看到的東西即時數據化；類似智慧健康手環的裝置和可以深度收集腦電波數據的裝置，將隨時使人體的活動轉化成數據。目前，記錄睡眠狀況只是初步的應用，在不遠的未來，用數據記錄我們每一秒鐘的生活也將成為可能。

當萬物皆數據的時刻到來時，商業發展的更多新機會將會出現，數據將會幫助我們做出更好的判斷，比如，什麼時候最適合吃飯、什麼時候身體疲憊應該睡覺、什麼時候記憶力最好等，這些都能透過數據來進行預判。甚至，當記錄了人類足夠的數據之後，數據就可能告訴我們此時此刻應該做些什麼及最佳策略。也許在那時，決定一個人聰明與否的指標，已經不是IQ，而是能否擁有足夠優質的數據。

人和機器的結合，或者人和數據的結合，將是未來的一種進步模式，人類將透過數據變得更加聰明。

最後，我想用兩句話來總結我對數據的期待：

當下，我們要學會人機分工，讓人做人最擅長的事情，讓機器做機器最擅長的事情。

未來，我們要相信人機結合，人機的界線已經模糊，無人駕駛汽車已經變成可能。未來人類身上流動著的是血液，以及數據。

關鍵思考 | 忽略了趨勢，過去的價值一文不值

我們通常認為，經驗應該是越多越好。一個有經驗的人，總是更能夠在很多事情上做出正確的判斷，因為積累使然。但事實上，我們發現，有時候經驗越多，似乎越容易讓人犯一些低級錯誤。這就好像，你會發覺，天天研究彩券和股票的人，總是發不了大財，往往是一個菜鳥莫名其妙就中了大獎和買了一支牛股。

為什麼經驗有時會使人犯低級錯誤呢？我們應該分成兩種情況來看待，一種是資訊不對稱，另一種則是邏輯錯誤。

在討論資訊不對稱時，首先要講的就是經驗。經驗是對過去的度量，但不是所有經驗資訊的品質都很好，在經驗的數據庫裡，一定有一些資訊是正確的，有一些是錯誤的。當經驗中混雜著很多噪音干擾時，我們會跟隨錯誤的經驗做判斷，此時，我們就會發現自己變笨了。

比如，使用未經評估的經驗，你可能用A方案成功了，用B方案卻失敗了。而且，評估也不是單純的0或1，而是用0~1的範圍來度量。說到這個，就要提到數據分析師了。一般而言，數據分析師對資訊是很有潔癖的，也就是在經驗的累積上，他們對品質的掌控十分嚴格，如對數據資訊的排序、分析可靠的資訊來源進行多次使用、了解資訊的出處和資訊的提供者等。根據這些，數據分析師才做出決定。這意味著，你所有的資訊來源都需要有正確的途徑和管道，不然，這個決策也會出現偏差。

除此之外，還有一個因素導致經驗使人變笨，那就是在分析時忽略了趨勢，這個道理可以用賽馬來舉證。

在香港，你經常可以看到一些人拿著一大堆的資料在分析馬匹的數據，比如，有人會根據一匹馬跑一千兩百公尺的時間，計算牠未來跑完全程可能需要的時間。但我們發現，還是

有很多人算錯了，導致賭馬失敗。為什麼？因為歷史數據和我們今天面對的情況之間，出現了一些假象。

香港大多數賭馬的人最終收集到的數據都是受到影響的數據，而不是經過清洗的數據，當然不準確。每一個賭馬的人都在看過去的數據——馬會會提供前三場賽馬的數據給每一個賭馬者，大家只會關注這個結果，而不會去關注當天發生了什麼。如果是我，我會回去看錄影，就有可能發現其他的情況。比如，一匹馬本來想衝刺，但前面的馬擋住了牠，才會扣了兩秒鐘；或者騎師揚鞭，鞭子掉了，扣了五秒鐘；又或者馬匹發脾氣，偏離跑道，也會扣除秒數。排除所有意外算出的時間，就是乾淨的、沒有影響因素的真正經驗了，這時候，我們算出的第一至三名的數據，和最終比賽結束公布的第一至三名的數據就會不一樣。

除此之外，還要觀察馬匹成長的趨勢問題。事實上，每一場賽馬都會出現很多意外，如果這些意外都不出現，隨著馬匹成長，也會直接導致結果不同。因為每匹馬幼年 and 成年狀態完全不同，因此數據價值的評估也是不一樣的。3~4歲的馬，等同於人類18~25歲，體能上會出現非常大的變化。又或者，年幼的馬對騎士的體重非常敏感，多一、兩公斤就會有非常大的影響，但是長到5歲後，負重對馬的影響會慢慢降低，這就是硬趨勢。

即便有很多經驗，但是當過往經驗的數據不能完全匹配你今天要做的事情時，意味著，經驗數據需要清洗——把當天的影響因素都找出來，並且還原，這時候得到的數據才是正確的經驗。

結論就是，經驗使人變笨的原因在於，之前的經驗本身就存在誤差，即數據來源本身就存在問題，而這種誤差一般人看

不出來。因此，在過去的經驗積累就有問題的情況下，根據經驗得出的結論，自然會使人犯低級錯誤。

永遠不要假設這個世界是真空的，你還需要多多觀察頻繁出現的新數據。當有新的數據出現時，過往的經驗就需要重新評估了。

PART 3

沒有數據，就沒有未來

大數據的近未來

- ▶ 應用行動化：讓終端裝置與數據收集的作業可以更為彈性而有效率。
- ▶ 資訊數據化：讓資訊的流通、交換、加工、運用，更趨標準化及結構化。數據的應用變得更即時、直接。
- ▶ 交易無紙化：徹底改變了我們的交易行為與資金流，並賦予未來微經濟商業模式更多創新思考的可能性。
- ▶ 人類智慧化：人的思維與新科技將會遇上前所未有的碰撞，人機協作是個新的機遇。
- ▶ 決策即時化：透過大數據即時收集及加工，改變了決策與資訊的關係。
- ▶ 線下線上化：未來仍將呈現線下更廣泛運用線上數據的趨勢。線上與線下將接在一起，不能分割。

第12章

未來趨勢，蘊於數據之中

「凡是預先設定的，都缺乏彈性亦不夠全面；面對一場決鬥，必須要根據實際變化去決定下一步的行動。」這是李小龍在他的武術筆記中說過的一段話。在我看來，對待數據也是如此。

數據積累永遠滯後，孤立的局部觀察並不真實，所以，一知半解所造成的錯誤決定，有時更危險。

從大量數據的收集到找出有用的規律，預測結果，必須經得起時間的考驗，非常不容易。如果不能將預測付諸行動，或者與現實偏差太大，便前功盡棄。因此，數據業界有個說法：「花六個月做的數據模型，使用兩個月就不能用了。」這不是笑話，而是常見的現象。

隨著技術進步，數據應用經歷了幾個重要階段：從最早期的描述性分析（發生了什麼），到診斷性分析（為什麼發生），直至今天每個產業和垂直領域都在努力發展的預測性分析（即根據已有的歷史數據預測未來）。

實際上，最重要的是，真正改變生活的數據應用正悄然興起，商業分析正朝這個方向走，這就是指示性分析（Prescriptive Analytics），預測不同行動產生的結果，並根據預測，提供具體的行動建議。這是商業分析的突破，預示我們已走入商業核心，創造新的商業模式指日可待。

用大數據預判趨勢，創造商機

美國新創公司FLYR Labs把預測性分析和指示性分析用於預訂機

票，他們根據大量數據和對航空公司定價的理解，預測未來的機票價格。其實，預測價格並不新奇，但是利用預測有效地鎖定低價機票，這種產品很有意思，情況就好像將火藥變成了手槍。

FLYR Labs開發了一個新產品叫Foresight，並運用在預訂機票網站Getflyr，從而進入到指示性分析階段。當你猶豫是要今天買300美元的機票，還是過幾天再買時，Getflyr或許可以幫你解決煩惱。你只需要花3~30美元，就可以鎖定未來七天內的機票價格，而且不必擔心機票加價，因為Getflyr會為漲幅埋單，如果票價下跌，你就可以享受低價的機票。這一切都是基於大量的即時計算和模型學習調整，最終呈現給用戶明確的解決方案。

上述例子就是一個機率遊戲，如果這個「為確定性而買單」（Paid for Certainty）的例子普及到日常生活中，可以衍生出無數的可能性。在機票的例子中，動態定價能力是這個產品的關鍵，即時票價的趨勢預判和參與人數，讓風險更容易控制。

再舉個例子，有天你把車停在路邊，擔心不知道會不會被交警開罰單。如果我跟你說，給我5元人民幣，有罰單我買單，你會接受嗎？你認為什麼才是這個例子的關鍵數據？這個數據容易獲取嗎？我想，我首先會求助於Google地圖。

大數據的重要性已經不言而喻，而大數據技術在商業活動、社會生活中的作用也越來越明顯、越來越重要。

物聯網，催生創新金融產品

最近，香港提出未來要做金融科技，想做好金融科技，就離不開大數據。在美國，銀行、保險、投資，甚至房地產都有不少例子，其中借助大數據獲取新客戶最為常見，也較容易實現。預測客戶偏好的產品及

管道，也可以令雙方關係更為緊密。

在金融產品中，風險管理跟定價相關，可以參考以下的創新例子。

依你的駕駛行為定保費

假設你是20～25歲的男性，駕駛一部跑車，如果你是安全駕駛者，你會發現，你為別人付出了更高的保險費用。因為在你這個年齡層群體中，有一部分人發生意外的機率很高，所以保險公司會向你收取較高的保費。大部分保險公司是根據一個群體發生意外的機率，而不是你的駕駛行為來收費，他們不知道你每天的駕駛情況，比如駕駛習慣及路線等，也無法評估你真實的出意外機率。

我們發現，在美國，有兩家保險公司已經收集了逾百萬用戶的駕駛數據，他們讓駕駛者安裝一個名為Telematics的感測器，即時收集駕駛行為數據，從中得知駕駛者有沒有超速、粗暴駕車等情況。整合了大量數據後，保險公司可以知道一般人的駕駛習慣是怎樣的，在不同的路段上，每個人與一般的駕駛習慣相比是怎樣的，如此一來，就更容易確定如何定價，相應的產品被稱為「Pay How You Drive」，即依你的駕駛行為付費。

另一家公司則採用更進一步的定價方法——「Pay As You Drive」，即在你駕駛時才付費，對不使用的車只收取很少的保費。這種方法更划算，尤其是駕駛里程低的用戶，可以節省數百美元的保費。簡單來說，大多數保險是在你不使用時也要付費，而在這個例子中，按使用付費其實更為合理。

前一個例子說明了在細分用戶群及提供服務的不同情境中，可能存在較大的成本差異，物聯網提供了更多的大數據做精準風險評估，因此定價可以不一樣。後一個例子則說明，我們可以準確知道客戶何時，甚至如何使用這項服務，因此定價可以更靈活。

這兩個例子都可以在金融業或服務業裡舉一反三地應用。

根據社交網站動態定保費

不出所料，美國一家公司把健康保險定價也列入上述討論範疇中，只要使用者願意分享自己的部落格、社交網站等數據，保險公司便可據此給不同的人定出不同的價格。但這樣可能會弄出一些笑話，例如，你在Facebook上貼了一張你假裝抽菸的圖片，可能因此令你的保費被提高。當然，這些比較創新的服務還有非常多的因素，需要更多源頭數據進行精準定義，而非根據偶然的行為便敏感地提高價格。

物聯網的出現，可以讓我們知道過去很多不知道的事情，例如，一部汽車是怎麼被使用的；冷氣、冰箱等電器的使用情況如何；我們每天到底是怎麼生活的……這些數據或許會成為我們為每種產品或服務付費的新依據。在不久的將來，物聯網將與我們的生活息息相關，也會在生活中催生更多創新。物聯網離我們不再遙遠。

以精算模型預測人類行為

你算過命嗎？算命先生是否說過你將來會大富大貴？但多年後，你的生活還是沒有任何大富大貴的跡象，你會為此感到苦惱嗎？又或者，像我這樣，被誤診為癌症，後來幸運地發現只是虛驚一場？

機率遊戲就在我們身邊，常常在不知不覺間從四面八方向你突襲，你甚至會疑惑，這一切到底是一場夢，還是真有其事。虛擬實境將改變我們的世界，下面的例子或許可以讓你有所啟發。

哈佛大學醫學院的一份研究報告提到，他們在研究參加過戰爭的退伍軍人的戰後創傷和精神治療時，開發出一個精算模型，這個模型基於近一百萬名退伍軍人的數據，能夠預測退伍軍人的犯罪傾向。美國軍方

也曾在2004年至2009年試圖建立退伍軍人數據庫，以研究退伍後軍人的行為風險，這個數據庫表明，其中近六千名退伍軍人曾發生暴力犯罪，從謀殺到縱火等等。

哈佛大學醫學院的研究者將這些數據和軍隊數據、人口數據，以及犯罪和醫療數據相結合，開發出精算模型，以預測參戰退伍軍人未來的犯罪率。最終結果證明，超過50%的犯罪都發生在模型判定的高風險人群中。

未來，或許潛在嫌犯會被提前監控。儘管在傳統意義上，警方的使命是破案而不是防止犯罪，但現在的大數據技術為預測犯罪提供了可能，預防並阻止惡性槍擊事件的發生，將在一定程度上為民眾的生活帶來安全保障。

這是好事嗎？未必。這類措施也可能會出現誤差。假設你連續三天在超市買了三個壓力鍋，這是你出於正常、安全的目的而做的購買行為，但結果可能引來飛虎隊半夜破門而入！為什麼？因為警方的大數據顯示，你這種購買行為可能是有風險的。這個例子不是玩笑，這樣的事情在美國發生過，結果警方賠償了幾千美元才了事。

當然，凡事都可以用一個角度去看，如果能在機率遊戲中看到玄機——掌握大數據與人性之間的關係，那麼我們都能從中獲利。透過這個技術，或許有一天，你也能當算命先生，字字珠璣，妙語生花。

以大數據因應「不測風雲」

2015年，巴黎氣候大會緊鑼密鼓地進行一場科學界的「華山論劍」，各國政府和研究機構都在利用大數據技術，來預測氣候變遷並尋求解決措施，大數據技術已經成為人們從各個角度研究氣候變遷的重要方法。

美國航太總署（NASA）的氣候模擬中心，透過數千部電腦進行平行計算，分析在全球收集到的各種類型的氣候數據，例如風如何影響洋流、雲含有的水量、南北極的冰川融化速度等，並建立物理模型，預測天氣變化的趨勢。

此外，美國EMC公司和美國國家公園的科學家展開合作，透過群眾外包方式取得關於鳥類、森林的數據，讓國家公園管理員了解氣候變遷對野生動物的影響。同時，透過衛星圖像、群眾外包數據和公共數據，分析各國森林砍伐的狀況及對氣候變遷的影響。

值得一提的是，隨著網路、物聯網、社交網站的發展，美國不少民間機構也為氣候數據的收集做出了貢獻。科學家使用大數據模擬地球上的生物種群，研究生物與氣候、污染、人類活動，以及全球碳循環之間的關係。

大數據對美國的戰略地位可見一斑。當前，美國的數據技術一枝獨秀，其實是有賴於企業的開放心態及美國政府的大力支持所帶來的成果。這些投資不僅培養了很多創新企業，還加速了民生的改善，例如減少交通堵塞、降低犯罪率、降低天氣變遷造成的損失、提升城市服務效率，以及廣義的經濟增長等。現在，美國有超過二十座城市透過和大學及技術公司合作，加入智慧城市專案中。

數據科學的差距，意味著國家的未來、公司的生存、個人的競爭力，大數據絕對不是感性的運動，而是理性的落實。中國人一直用「人心難測」、「天有不測風雲」這樣的語言來解釋不確定性，那麼，面對未知，能否借古鑒今固然是聰明人的基礎，但在不確定性中借助數據科學做判斷並採取行動，才是大智慧。

行動智慧的關鍵在於數據

我們每天都會收到天氣預報，這對我們來說只不過是個訊息，它能

幫助我們決定如何適應今天的天氣，比如出門是否帶傘、是否需要多穿衣服等，但美國有家公司卻把天氣數據運用在商業中。

這家公司叫Planalytics，是一家領先全球、專門做天氣分析的公司，服務來自美國、歐洲及南美各國的兩百多家企業客戶。這家公司在天氣分析上積累了大量經驗，特別是在細節上，在美國多家大型零售商和製造商都有成功案例。

到目前為止，Planalytics已經分析了上萬億的交易數據及八千五百個類別，根據這些分析，再將天氣數據轉化為有助企業決策的指標（可理解為數據中間層），進而優化業務結果。

數據的重要價值在於，它能让人們把決策提前。透過數據，哪怕只是比別人早一秒對事情做出準確的預判，我們就比競爭對手更有優勢。準確的預判，加上快速的行動，就是現在我們所講的「行動智慧」

（Actionable Intelligence）。想達到這個目標，我們需要足夠的資源，最關鍵的原料就是數據。

然而，一些重要數據之間的關係往往不那麼直接，人們經常會忽略這些數據之間的關聯性。會發生這種情況有兩種可能：一是這些數據之間的關係比較隱性，不容易直接被人發現；二是過去不容易拿到這個數據，又或者能拿到，但數據太粗糙或不夠準確等。

連結數據，就能創造價值

春天的到來，對零售商來說意味著春裝的採購、戶外用品銷售的回升等，然而，今年的春天是否會和以往相同？有數據表明，重複去年天氣的機率僅有15%，但商家的規畫往往是基於往年的銷售，並假設同樣的天氣會再現。

天氣對商業的影響，遠大於我們意識到的，商家對天氣的反應也往往是滯後而臨時的。在美國，天氣對服飾業和百貨業的影響都高達30億

美元，對園藝、五金工具的影響則達到40億美元。

這些對天氣敏感的產業，利用天氣數據，將業務天氣化（weatherized）帶來的效益是非常可觀的。天氣數據看似簡單，卻價值連城，但其價值並非來自數據本身，而是藉由與具體商業問題的結合，發掘出對商業的影響，從而創造價值。

30℃的天氣對業務意味著什麼？這取決於所在的地區、所處的產業、所在的季節、該地區的歷史天氣，以及歷史銷售紀錄等。因此，天氣看似是小數據，但是要利用天氣數據驅動業務，卻需要融入多個來源的大數據。

一直以來，我也在觀察天氣與服裝銷售之間的關係，我發現一個很有意思的現象，原來，每年冬天剛結束，就有人開始尋找夏裝，當廣州的天氣首度轉熱時，夏裝的需求就爆發了。反過來，在冬天，北方城市首次降溫，也會帶來搜尋冬裝的熱潮，而最早開始搜尋的這些人，往往都是服裝達人。

能像這樣了解天氣與服裝銷售之間關係的人，可能就發小財了，這就是數據連結的價值，是大數據的奇妙之處。別人的冗餘，收集起來廢物利用，變成了我的寶貝，點石成金。這種經濟效益絕對奇妙無窮，也是我從事大數據的動力。如果我們能把更多的環境數據與人們衣食住行的習慣連結起來，將有可能挖掘出非常有價值的商業機會。

以大數據解讀人心

「相形不如論心。」這句話出自《荀子·非相》。所謂論心，即是從收集一個人的行為資訊來判斷他的意志、情緒、涵養及心計的表現。

為什麼我認為這是發展行為大數據的機會？我想用楚莊王參謀的一句話來總結：「臣非能相人也，能觀人之友也。」大數據時代的思考方

法就是要全面了解一個人的行為，不是單純地看而已，要有計畫地多角度觀察，甚至要觸類旁通地觀察。

分析情緒，確保廣告效用

我們的言語就像臉部表情一樣，會透露出我們的感受。最近，一家名為BehaviorMatrix的公司發布了新一代的數據產品，能夠幫助理解交流和決策背後的情緒。這項產品提供的情緒信號探測功能，透過分析上下文及一系列不同數據來源的交流對話，追蹤情緒的反應鏈，並透過識別參與交流的人群模糊網路，發現誰是真正有影響力的人。我們不僅能知道人們在討論什麼，還能知道他們為什麼會有這樣的感受，是什麼在影響他們，以及驅動他們意見和決策的情緒是什麼。

同時，這項產品在分析出情緒後，還會預測這樣的情緒下可能產生怎樣的行為。據稱，這個情緒信號處理系統是該領域第一個，平台收集來自三百萬不同來源的數據，包括社交媒體、新聞、部落格、論壇、調查，甚至廣播電視字幕、印刷刊物。應用領域包括客戶服務品質的提升、品牌意識、新產品開發、廣告效果分析等。

BehaviorMatrix曾經分享一個案例，公司分析製藥公司針對家庭客戶的一則電視廣告，當增加一個新角色時，會引起觀眾情緒如何變化。結果發現，這個新角色對品牌的作用是負面的，角色本身和出現的頻率都引起了觀眾的反感，對品牌造成破壞。製藥公司希望透過這樣的情緒分析，探索在電視廣告中，怎樣的情緒將觸動觀眾渴望購買這項商品。

這個情緒分析產品，在技術上超越了目前常用的語義分析，用一種類似人腦的方式進行自然語言處理，能夠理解文字背後細微的情緒和情感。

還原，是深入洞察的根本。我深信，當大數據時代衍生的多媒體數據越豐富，科技識別的能力越精細，情境的定位越明確，數據從收集到

使用的規範化，都有助於未來的分析更具追溯性和關聯性。

在前例中，如果語音、筆跡、臉部表情皆成為情緒計算的一部分，我相信，今天我們見到的，都只是冰山一角。

思維的開拓，是利用數據的關鍵

顧客購買的數據重要嗎？當然重要，但有時候，顧客拒絕購買的數據可能更重要。

我曾在一些會議上分享一家珠寶公司的例子，他們不僅收集顧客買了什麼首飾的數據，還收集顧客看了什麼首飾、卻沒有買的數據。這個案例很有意思，正面資訊和負面資訊疊加起來，便可更清晰地判斷顧客需求，從而幫助公司的生產和銷售。

最近我們發現，美國數據公司Granify跟這家珠寶公司的做法有異曲同工之妙，他們透過觀察顧客的「數位肢體語言」，來了解顧客為什麼不買，同樣是從負面資訊去尋找機會。

解讀肢體語言，負面資訊也有價值

Granify的口號是：「Do you know which shoppers are not going to buy? We do. And we can change their minds.」（你知道哪些顧客不會消費嗎？我們知道，而且懂得如何改變他們的想法。）

對於導入Granify平台的電商網站，Granify收集並觀測用戶整個流覽旅程的每一個細微行為，透過每秒觀測超過四百個變量，以理解使用者的「數位肢體語言」，並預測他們下一步的行為。一旦決策引擎判斷某種行為可能會影響客戶的購買行為，系統將自動彈出資訊或優惠來打消他們的顧慮。

這家公司的可貴之處在於，它不但能預測客戶行為，還能採取行

動，直接影響客戶的決定。公司成立Influence Lab，專門研究客戶行為影響，透過市場專家、使用者體驗設計師和數據科學家三方的共同努力，採用消費心理學的技術及分析，發現用戶的想法，並系統化地影響用戶的決定。例如，當發現用戶對運費有顧慮時，就會彈出滿額免運的資訊。

Granify目前與十九個國家的大型零售商都有業務合作，一些成功案例顯示，用戶轉換率可提升高達35%。

這讓我想起，早在十年前，我的老師就幫Amazon做過這樣的數據實驗，從一個人敲鍵盤的頻率及時間間隔，來猜測這個人到一個網站是為了買東西，還是只是閒逛，以及購買意願有多強烈等。

今天，我們的科技已高度發展，大家都在談論「人臉辨識」，但我聽說，市場上已經有這樣的技術，憑一個人穿的衣服，就能追蹤他在百貨公司裡的行為，從而收集到顧客所有的正面和負面數據。

因此，未來的零售店關鍵是，能否真正從一個人進入這家店後的肢體語言，例如，他看了什麼、沒買什麼、最後買走什麼，來判斷顧客真正需要的是什麼，才能更好地滿足客戶需求。所以，很多時候我們不是沒有數據，而是沒想到要收集哪些數據。

因此，要想利用大數據，不要只關乎科技的發展，還要注重思維的開拓。只有先進的科技，加上開闊的數據思維，大數據的價值才能發揮得淋漓盡致。

未來，人人都是數據分析師

當你遇上難題時，你曾經說一聲「天知道」就草率下決定嗎？一般人平均每天要做上百個決定，小到早餐吃什麼，大到要不要跳槽，都需要我們大傷腦筋。有些小的選擇，選錯了無傷大雅，但一些關係到你人

生的重大選擇就要注意了。

當你面對既重要又緊急的問題，卻沒有任何經驗時，該怎麼辦？人人都希望自己做出英明的決策，但如何保證？如果有一套清晰的決策框架幫助你，是否好過迷迷糊糊地做出選擇呢？其實，只要你願意，人人都可以成為自己的分析師，在面對重大的決策時，更能深思熟慮。只需謹記：好的問題裡面潛藏了答案。

明確你的問題

在決策前，首先要明確你的問題是什麼。先把問題裡的一些關鍵字定義清楚，然後反覆用不同的角度重組，直到問題越來越清晰。例如，我要找一份更好的工作，但什麼是更好？薪資高就是好嗎？

當問題定義好後，決策就好辦了。因為釐清了問題之後，就比較容易找到相關的數據和資訊。在還沒弄清楚問題之前，便沒辦法進行下一步的思考和行動。如果有人問我，他換什麼工作好，我會不斷問他：你想解決的問題或目的是什麼？找到下一份工作之後，如何知道是否達到目的了？如何比較現在與未來？在我們釐清問題的過程中，很容易就會知道衡量各種選擇的邏輯。

搞清楚已知和未知

在已知的資訊中，要判斷哪些與你的判斷邏輯相關，是否支持你的選擇，同時評估資訊對選擇的重要性和準確性，當中如果有決定性的因素，更要細分及追本溯源。在數據化時代，量化的資訊（包括半結構化數據）尤其重要。

當然，世事哪有這麼完美，你不知道的往往才是決策的陷阱。不知道的資訊分為兩種，一種是你花點時間就能找到，另一種是「你不知道你不知道」，比如，你想換工作，除了薪資、職位、發展前景等條件

外，可能還需要了解公司的穩定性、新老闆對待員工如何等，這些是你花點時間就能得知的資訊，但是你可能萬萬沒想到，第一天上班，部門就改組了。

在不確定中做判斷

除了已知和未知，可能還要看你知道的資訊中，有多少是可靠的。分析師在看到數據的第一個反應就是，這個數據是從哪裡來的？準確嗎？他們需要不斷地校驗數據的準確性。

當釐清問題，並且收集到「足夠」的資訊後，就可以對選擇進行對比，包括前後對比、內外對比、自他對比等。分析師行內有「六字真言」：對比、細分、溯源，可謂字字珠璣。其中，「溯源」是高手的必爭地，無他，基礎也。

總結來說就是，分解問題，找到準確資訊，把資訊融合到判斷邏輯中，在有限的資源下做出有利的決策。

兵無常勢，水無常形，判斷的能力不僅僅是一種科學，更是經驗的積累，事後覆盤是學習的唯一途徑，用結果倒推來審視：問題是否夠清晰？判斷邏輯對不對？資訊是否匹配？從決策的過程中得到回饋，不斷優化。

好的分析師會把自己也作為分析物件。在如此循環往復的分析判斷中，最關鍵的是，長期保持「新手」的心態。

關鍵思考 | 大數據生態正走向平民化

我曾經跟一位美國零售集團的高階主管交流後得知，數年前，當傳統零售業意識到電子商務帶來的改變時，除了招聘大

量數據科學家外，第一件事就是去收集線上對手大量的即時商品數據。

幾年後，如今美國的大型零售公司都已具備選擇商品、動態定價及多終端多管道的管理能力。為了支持每一秒鐘與對手的競爭，可以想像，收集更廣、更即時數據的重要性，即使只是簡單的天氣變化，都有可能令你當前商品的競爭力落後對手。

《孫子兵法》有云：「知己知彼，百戰不殆。」對於企業應該如何使用數據，這句話最為貼切。當很多公司來問我應該如何使用數據時，我一定會告訴對方，第一步是用數據了解自己（知己），第二步就是用數據來了解競爭對手（知彼），並根據對手的動態變化做出攻守判斷。大數據的首要關鍵，就在於如何量化數據大小與決策好壞的關係。

你可能會問：這麼高難度的事情，小企業如何跟得上呢？好消息是，在這幾年間，美國的大數據生態已逐步走向平民化。

最近，美國一家專門做數據收集爬取的公司Import.io，讓我眼睛為之一亮。網路是最大的開放數據來源，透過Import.io的工具，用戶可以瞬間將眾多不同網站轉化為數據表格或API（應用程式介面）。Import.io每天可以從網路上搜集千萬條紀錄，至今已從三十萬個網站上收集了上千億條數據紀錄。Import.io在同類型公司中技術領先，曾多次被評選為最佳新創公司。目前這項工具對個人用戶免費開放，對企業用戶則源源不斷地提供大規模、訂製化等極具品質保證的數據。

讓我感到興奮的是，他們把收集半結構化數據這項本來只有技術工程師才能做的事情，變成了人人都可以使用的平民化服務。而健康的大數據生態，應該可以讓數據從收集、加工到

應用的各個環節變得越來越精細，整個數據處理的過程形成一個互惠互利的產業鏈，大家都在大數據的海洋中分工協作。

不得不讚嘆美國大數據產業鏈的發達，有人做數據中間層，比如，結合天氣與零售的數據，提供給需要的人使用；有人把機器學習的門檻降低，變成MLaaS（Machine Learning as a Service，機器學習即服務），讓不懂數據的人也具備數據學習的能力；也有人把數據視覺化變成簡單易用的工具。Import.io 這樣的公司，就是把數據收集，變成如水、電、煤一樣，每個人都能使用的服務。

大數據從原始數據再到加工、使用，是很長的鏈條，如果有好的數據生態和產業鏈，每個人專注於其中的一個環節，各司其職、分工協作，就能把數據的價值發揮到最大。希望在不遠的將來，我們也能跟美國一樣，早日建立豐富、完整的數據產業鏈。

第13章

未來洞察力，才是核心競爭力

大數據公司最大的痛苦是什麼？不是沒有數據，而是有太多數據。各種各樣，包括結構性和非結構性的數據從四面八方而來，但哪些有用？如何挑選？這才是最難的。

大數據變革在於數據本身

有一次，我在美國巧遇Google的前數據部門主管，他跟我說一件他覺得很痛苦的事：提出數據需求的人，往往70%的情況都沒有使用他們所提供的數據。他為此非常生氣，便質問業務方：「為什麼你們老跟我要數據，結果卻不用？」對方回答：「因為我用了之後才知道這些數據沒有用。」

對於一個每天都要準備數據的人來說，這好比你是一家餐廳的老闆，廚師每天準備了很食材，你卻只用一小部分。所以，我從事大數據工作一直有個夢想：如果有一天可以把數據處理自動化，該多好！

以創新方法實現數據處理自動化

說起數據處理的自動化，不得不提一家在數據領域有突破性創新的公司Paxata，它是美國少有的接地氣的自助型數據開發平台。這家公司的專業領域令人興奮，因為目前市場上幾乎所有的數據分析，包括商業智慧軟體和創新，都聚焦在解決數據分析的演算法、視覺化、模型等的自動化，卻忽略了一個事實：現實工作中，數據科學家和分析師首先需

要花費80%（甚至更多）的時間在數據清洗和準備上。而Paxata採用機器學習的創新方法解決數據開發，其功能如下：

- 讀取數據：讀取任何形式的數據，自動解析和識別數據的類型及數據的意義。
- 數據探索：通過人機交互的形式，幫助用戶主動發現數據品質問題，比如，提供全文搜尋、數據品質熱力圖、發現規律，來發現數據中的錯誤、重複、稀疏、缺失等問題。
- 數據清洗和修改：Paxata平台的核心是語義融合和機器學習引擎，基於演算法智慧識別並進行數據變換，以改善數據。

除了上述功能，Paxata還透過人機交互的形式，完成數據合併、發布、改進、分享、安全控制等數據準備的所有環節。其中一個有趣的功​​能是資料豐富化（data enrichment），它能基於第三方數據庫，建議用戶增加更多的變量以提升挖掘價值，並自動完成數據的轉化。他山之石，可以攻玉，同類公司關注的變量及挖掘路徑可以互相借鑒。

今天，大數據之所以說起來容易、做起來難，主要有兩個原因，其一是數據的易用性，其二是數據處理的成本，因此，底層數據處理的自動化才非常重要。我們需要將使用數據的經驗數據化，並把這些經驗轉化為其他公司對自己數據的理解和使用能力。

或許在不久的將來，我們就會發現，只要我們對某一領域使用數據的經驗有足夠的沉澱，哪怕給我一堆毫無關係的數據，我也能根據數據經驗庫，告訴你我從中看到了什麼。世界日新月異，變化之快令人咋舌，數據人只有不斷顛覆自己使用數據的能力，才能找出大數據未來的方向。

用數據治理數據

2015年，我參加「中國優秀首席資訊官評選」頒獎大會，現場很多人都在談大數據，其中有傳統產業，有銀行，也有醫院。我注意到，大部分演講嘉賓都集中在討論如何應用大數據，或是如何利用數據進行創新。跟以往的很多會議一樣，我很少聽到數據作為原料，應該怎麼管理。

你可能會說，銀行業、通訊業等不是早就在做數據管理了嗎？的確，數據管理並非新鮮事，二十年前就有人在做，但是，在數據呈指數級增長之下，我們現在講的數據跟以往已經不一樣了。不僅是數據的大小，包括數據的內容、來源、結構等都很不一樣。以往我們可能根本不敢想像，Facebook的日均新增數據量已經達到600太位元組（TB，1TB＝1024GB）！與此同時，無數以往不可能出現的演算法和應用也隨之呈爆炸式增長。

數據在創新，演算法在創新，應用也在創新，我們不禁會問，以往的數據管理思路，能適應新形勢的要求嗎？

自上而下的管理費時費事

在一些傳統銀行，如果你要看某個數據，就必須拿著申請單找一堆人簽字，最後，從一個類似醫院取藥的窗口，將按照流程簽滿字的紙遞給數據管理員。這個管理員會告訴你，你要等幾個小時才能拿到數據。因為他要從一大堆備份中找出你要的數據，載入完成後才能給你。

這就是傳統的數據管理方式，嚴格按照自上而下的流程進行，甚至要做到滴水不漏。之後，銀行發生了一件對他們來說具有劃時代意義的改變——那個幫你找數據的管理員變成了機器人。雖然機器人的效率大大提升了，但數據管理的本質卻沒有變化。

在大數據時代，數據將會促成新技術的發展，以及更多新數據的產

生。開放的網路環境、頻繁的數據更新、豐富的數據種類，加上數據保護的需求激增、數據生態圈的物種增多、流動性加快……這些複雜多變又未知的環境，對中央式管理來說是極大的挑戰。試想，若還是按照以往的方式，會出現怎樣的情形？也許，數據的審批人會比數據的使用者還多；也許，數據的創新會停滯不前；也許，數據的應用價值會比現在晚幾年才被發現。

數據治理需要人人參與

因此，我們必須意識到，數據治理不等同於數據管理，絕非依靠自上而下的貫徹執行便可解決。相反，數據治理需要每個人的參與和協同，要求每個人都有意識去治理好數據。而且數據治理本身也可以變成應用和創新，成為更多應用和創新的「水、電、煤」。

在數據產生價值之前，不管看數據還是直接用數據，首先都要注意：數據有沒有？數據準不準？粒度細不細？是否穩定？如果行政總裁每天看的數據都是錯誤的，或者不是穩定產出，那麼企業依賴數據的決策將付出極高代價。所以，企業若不做數據治理，或者還是按照以往的思維來管理數據，那麼，日後對數據的依賴越深，便可能越快出問題。

數據治理的新思路，不僅是指組織結構上要從自上而下，變成全體協同，而且要在技術上創新，提供更智慧的治理工具，幫助大家提高數據品質、保護數據安全及有效控制數據成本。

大數據時代的數據治理，一定是將有形的管理策略，化成無形的智慧產品；從一紙命令，變成根植在每個人腦中的信念和下意識的習慣。我們要用大數據的思維方式，用數據來治理數據。

在大數據風暴中，取得發展的指南針

《大數據時代的致勝決策》（*Data Crush*）^[1]一書中提及「指數思維」，這個概念來自美國知名發明家、未來學家雷·庫茲威爾（Ray Kurzweil）提出的加速回報定律（Law of Accelerating Returns），是指當技術呈現指數級增長時所需的洞察力。這種指數級增長產生的影響力，是我們用過去的思維方式無法理解的，因此我們需要用新的思維方式——「指數思維」來看待明天的變化和機遇。

短短的時間，我們從個人電腦到網路，再到智慧型手機的普及，激發各種行動網路的應用，各種感測器的發明，引發物聯網激增，甚至最近有人提出以人為中心的「人聯網」（Internet of Me）。大數據技術創新，加速了很多應用創新，應用創新又反哺出大數據的增長。凡此種種，都在我們眨眼之間產生了大量人與人、物與物、人與物之間的數據。這些數據經過先進的演算法處理，人工智慧也將變得無處不在。

行動網路、物聯網，甚至人聯網，在這場指數級變化中，擔當了加速變革的重要角色，人類此刻正初次嘗到「人機合一」的貼身享受。就像伊隆·馬斯克（Elon Musk）所說：「基於當前科技發展的速度，人類正在變得擁有上帝般的能力。」

以前，我們需要拿遙控器控制房間裡的各式電器，現在，我們只要拿著手機就能感知周圍的物理環境，並做出更好的決策。例如，當我不知道從A地到B地怎麼走時，手機導航可以馬上告訴我路線。結合行動網路、物聯網和各種穿戴裝置的三種數據，我們和物理環境有了更深度的連結。只要有手機，便可跟全世界溝通，亦可跟現實環境溝通。若說網路世界或虛擬世界有時給人的感覺並不是那麼真實，那麼，有了物聯網之後，虛擬世界和現實世界將變成一體。

舉例來說，我們透過社交網站與朋友保持聯繫，某天中午，有一個小時的空檔要打發，我看了一下我的社交網站，想找朋友喝咖啡、聊天。這時，我只需按下某個按鍵「我想找朋友消遣一個小時」，手機就會告訴我現在離我最近的朋友是誰，以及附近最適合我們見面的餐館在

哪裡。手機還會告訴我這一個小時的陽光很好，建議我們曬曬太陽，以及今天周圍發生了什麼新鮮事可以跟這位朋友聊。

指數思維讓我想起了一個新詞「datafication」，意思是數據化。這個世界的人、物和事，正以前所未有的速度被數據化，很多過去我們以為不可能有的數據，如今都唾手可得，基於海量數據的深度挖掘及人工智慧都在爆發中，數據創新、演算法創新、應用創新的成果，將以指數級增長，擴散到商業的每一個角落，穿戴技術和物聯網的威力也將超出我們的想像。

不管你是否喜歡，大數據這場風暴已經離我們越來越近。在這場風暴中，只有能做出深度回應的企業，才可獲得發展的指南針和救生衣。

關鍵思考 | 讓數據透過產品「說話」

有一次，我打算搭晚上7點從杭州到北京的航班，結果7點整，一個提供航班資訊的App告訴我航班延誤了。一個小時後，它再次告訴我航班還在延誤，一直到晚上11點。這個App確實一直在提供資訊，但這些資訊並不能幫助我做出更多的決策判斷。

如果，當時這個App能夠結合更多資訊，它是否可以告訴我另一家航空公司有班8點的航班可以立即改簽飛走？即使做到這一點，這還不算是數據產品，只是把資訊善用在產品上，就立刻使產品有了新價值。但是，當時我並沒有收到這些服務，一直從晚上7點等到半夜11點，App不斷提供的延誤資訊只會讓我越來越苦惱。

資訊化和數據化的區別在於，資訊化為我們提供了參考，

而數據化讓我們可以直接行動。從參考到直接行動，這之間有著巨大的差距。

為什麼數據的價值無法實現？為什麼目前大量企業的數據案例難以凸顯價值？其中很重要的一點是，產品經理沒有真正懂數據。很多產品經理還停留在過去靠感覺來做產品的階段，不知道如何運用數據來改善產品，更沒有意識到數據已經成為產品的核心原料。

過去的 IT 時代，我們只是簡單地使用數據，很少為了解決問題而提煉數據。為什麼我要強調提煉數據？因為如果要讓數據產生價值，讓更貼身的數據分析框架來解決用戶的實際問題，就需要將數據嵌入產品或生產流程中，在提煉數據的最後一哩路，讓數據在產品中「說話」。

如何讓數據「說話」？在過去的資訊時代，我們最擅長的方法，是根據歷史統計數據的規律來指引行動，比如，我們會統計過去一個月，週五下午6點的計程車小費，計算出平均數，然後告訴用戶「建議給5元人民幣小費」。

那個時候，我們使用的大多是單一角度且靜態的數據，而現在，我們更希望得到全景且動態的數據。比如，我們可以獲取不同街道的堵塞程度，從而計算司機對小費的敏感度；我們可以結合附近的天氣情況、演唱會散場的時間等數據，來預測某個時段、某個地段可能成交的小費金額。這樣的演算法就是利用更全面的大數據，透過更多的環境動態數據，而非歷史統計數據來提供服務。

產品化×數據化×商業眼光


未來，產品經理要懂得如何利用數據來增值，這其中有三

個關鍵點——產品化、數據化和商業眼光。目前，很多產品經理更關注的是產品化，而忽略了數據化。

那麼，如何利用數據來增值呢？答案是，讓數據前置。

假設我要為女兒選擇一所學校，如果等到三個月後，考試成績都出來了，才知道這所學校不適合我女兒，那豈不是為時已晚？如果我能夠事先根據數據計算，衡量這所學校適不適合我女兒，不是更好嗎？數據價值的關鍵，就是數據前置，即讓更多數據嵌入產品之中，將數據分析框架應用到公司業務中，而不是停留在用統計數據做決策參考。

也許不用很久，產品經理就會發現數據化對於產品的重要意義。數據必須和產品結合，不然數據的價值難以實現。讓數據變成產品的核心原料，是產品經理最大的挑戰，也是產品經理最大的機遇和想像力。

-
1. 2014年getAbstract國際圖書獎獲得者、商業數據系統與業務模式的設計者克里斯多夫·蘇達克（Christopher Surdak）的顛覆力作，書中描繪了一幅大數據時代企業創新業務模式的路線圖。 

第14章

大數據的近未來

對於大數據產業來說，過去這幾年冒出了很多新名詞，但在我看來，真正的大數據應用和市場才剛剛開始萌芽。我希望大家先認清一個關鍵，那就是，所有的數據都是基於應用而產生，而數據經過收集、整合後，又再落實到自身或其他應用情境中。大數據的創新價值可以來自新連結的數據、演算法或產品本身。

過去，大數據的成長和智慧型手機的普及有著緊密的關係，智慧型手機發揮了媒體、通訊、社交及感測器於一體的功能。同時，物聯網的浪潮正在醞釀中，線上與線下的接合，帶來了更深度的數據連結，觸碰到消費者的全通路行為收集。

在一波接著一波的趨勢浪潮下，很多人問：「未來，大數據的趨勢會怎麼發展？」一般來說，這種問題我可以有很多不同的猜想與觀點，但是認真想想，與其給出我的答案，不如和讀者分享一下我對發展趨勢的思考邏輯，希望能幫助讀者在自己的產業挖掘真正有價值的發展趨勢。這部分內容不是手把手教讀者「捕魚」，而是傳授一點有關「魚竿」和「釣魚」的知識，讓讀者有能力在趨勢大海中，獵捕屬於自己的大數據觀點。

變，是唯一的不變

說到趨勢，人們往往習慣從改變的現象做觀察。如果我們把視野格局放大，在大數據的長期發展趨勢中，找出變化的本質，反倒讓我們更

容易看清楚動向。所以，從這個角度出發，我至少可以看到，未來的兩、三年內，有幾項已經存在的情況正在過渡中：

- 應用行動化
- 資訊數據化
- 交易無紙化
- 人類智慧化
- 決策即時化
- 線下線上化

上述這些趨勢不必贅述，其實已經是我們現實生活的一部分了。

應用行動化，提供了更大的便利性與行動性，讓終端裝置與數據收集的作業可以更為彈性而有效率。

資訊數據化，讓資訊的流通、交換、加工、運用，更趨標準化及結構化。數據化時代，數據的應用變得更即時、直接。

交易無紙化，徹底改變了我們的交易行為與資金流，並賦予未來微經濟商業模式更多創新思考的可能性。

人類智慧化，描繪大數據所產生的創新價值如何與人類交互並深入生活之中，人的思維與新科技將會遇上前所未有的碰撞，人機協作是個新的機遇。

決策即時化，透過大數據即時收集及加工，改變了決策與資訊的關係。過去，我們認為數據不能以低成本獲取，難以達到決策的即時性和精確性。

線下線上化，也就是大家熱議的全通路議題，未來仍將呈現線下更廣泛運用線上數據的趨勢。未來，線上與線下將接在一起，不能分割。

這些本質上的轉變，會持續好幾年，上述這六大趨勢會在各自的體

系內深化發展與創新，而大數據的發展趨勢集中在兩個方向：

- 首先，大數據的價值會落實於各個產業當中，數據技術會成為各行各業的優化工具或催生顛覆性創新。
- 其次，大數據的發展也會被自我顛覆，數據的收集、更新、識別、關聯將會變得越來越自動化，落後者將會是被超越與淘汰的輸家。

跳脫惰性的乘法思維

每一個理論的產生，都是為了讓我們更容易去解釋各種現象。趨勢的變化，可能會受到很多商業與政策背景的影響，而有不同的呈現形式，但是背後的原則，或者說脈絡，要能夠經得起重複的檢驗。所以，我試著用所謂的大數據生態循環鏈來描述這個趨勢。

圖6的兩個維度恰恰是我一再提到的觀念：大數據是被需求驅動的，而需求來自於對現有已知或未知問題的解決。所以，在縱軸上，我將問題區分為「明確問題」和「模糊或複雜問題」，這說明，在大數據落實應用之前，我們要弄清楚需要解決的問題是什麼。橫軸是定義大數據在原料端的呈現形式，一端是條理清楚的「集中化數據」，另一端則是結構模糊的「碎片化數據」。在了解基本定義後，我對大數據的發展趨勢，基本上可以用圖6來闡述。

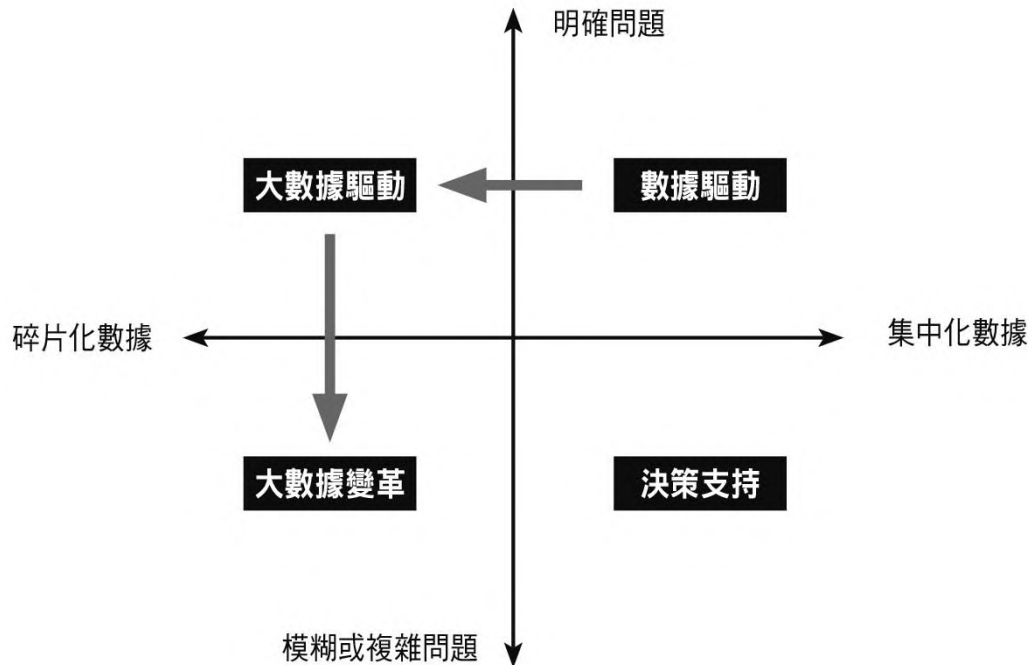


圖 6 大數據創新循環鏈

首先，所有的數據應用都是從第一象限「數據驅動」開始，這部分的數據集中化，而且要解決的問題很明確，已經可以開始運用數據來優化我們的決策。在這個階段可以發現人的惰性，導致無法以數據作為他們的核心競爭力，畢竟這些人手邊不是沒有數據，而是不知道，或是沒有很積極地用數據來解決問題，所以透過數據驅動，讓這些人把手邊的數據用起來，嘗試去解決一些老問題。

隨著數據不斷地快速增加，很多碎片化數據或其他冗餘的數據開始加入，這個時期我們可以稱為「大數據驅動」。能否應用來自第三方的冗餘數據來解決自身的問題，是辨別進入大數據應用領域與否的重要判定基準。然而，這個階段面臨的挑戰是來自決策者過去的習慣，以及數據人的惰性。

我們要了解，在第一階段和第二階段，處理數據所需要的能力是很不一樣的。在第二階段，因為數據零散，我們需要一套新的方法來收

集、加工不同的數據，所以我們在這個階段常常只聽人家講，而沒什麼產出。因為處在這個階段，雙方都不習慣，商業的人說我不習慣用數據做決策，數據的人說我們也不習慣用零散的數據，試圖穩定地解決問題。

從第一階段到第二階段，我認為都還是加法類型的演進，可以稱為連續性創新，也就是讓既有的東西做得更好，而且是基於以往的經驗疊加上去的。

到了第三階段的「大數據變革」，我們將面對模糊的問題、碎片化的數據。過去我們不太習慣在還不清楚問題的時候，就使用數據找出問題所在，這時，我們往往會陷入一種矛盾，數據零散、混亂、沒標準、不規範，問題也不清不楚，究竟該怎麼辦呢？畢竟兩者都超出了我們過去的習慣與經驗法則，然而，這恰恰是人類未來要變革的地方。

不同於前兩個階段的加法邏輯，第三階段開啟的是乘法思維。透過模糊問題與碎片化數據的碰撞，我們將會看見新的問題與新的機會，我認為，這部分才是大數據真正的價值所在。

最後，我想強調的一點是，這幾個階段不是線性的延展，而是一個滾動的循環。因為第三階段碰撞出來的問題與機會，經過系統化的驗證與標準化後，我們需要的數據來源，以及能夠解決的問題，也會逐漸穩定下來，形成新的服務或方案，接著再次歸回到數據驅動的解決層次。也正因为這種特性，我將其定義為大數據的創新循環鏈。

抓住趨勢中的觀戰重點

前面講述了趨勢的架構和思考方法，最後，我還是忍不住要跟讀者分享幾個獨門觀點，讓各位在各自產業中觀察大數據趨勢的時候能夠很快地抓到重點，可以讓各位不光跟著看熱鬧，還能從中看出創新商機的

門道。

重點1——數據安全問題

數據安全事件頻繁發生，隨著全球各個國家開始採用新的數據安全技術和新的數據保護法律法規，對數據安全的監督要求會變得更為嚴格。大家對個人隱私的保護比商業機密的洩露更為關注，已經到了沒有一個政府或企業不關注的時候。

但是，數據安全背後代表的是數據開放的風險與疑慮，當數據風險沒有辦法有效地管控，並建立個人對數據的信任感時，這對正在發展中的大數據產業來說，將會形成一種阻礙。

個人隱私、公司機密，乃至於國家和國家之間的數據保護，都將會是快速成形的趨勢。當數據成為商業重要且關鍵的資產時，隨之衍生的可能會是像「首席數據隱私官」這樣的職業。或許很多人會覺得匪夷所思，但是我相信，未來，許多擁有大量數據的公司，都將把首席數據隱私官視為一個重要而關鍵的角色。

重點2——數據分析的簡化與外包

講數據分析工作的外包，其實是一個概念上的舉例，其實我要談的是大數據背後將會形成的產業鏈分工，這是一個值得大家關注的發展趨勢。

隨著大數據落實應用，少有企業可以獨立完成從原始資料收集、加工、分析，乃至於落實應用的完整程式。這意味著，未來在不同的數據處理階段，都將有機會發展出專門的技術公司，以協助企業完成大數據實務應用前的準備工作。

回想我們前面提到的大數據創新循環的概念，每一次大數據變革階段激盪出新的問題與機會後，當這些新的問題開始聚焦，對應的數據來

源也趨於集中時，就代表一個新的產業鏈機會隨之產生，這些中間層的服務與創新，對大數據產業的發展將扮演至關重要的角色，同時其中也蘊藏著可觀的商機。

重點3——政府對數據的態度

從整個數據地圖來看，政府其實是擁有最多數據的「財主」。因為政府鎖定了很多公共服務領域的關鍵數據來源，這是公共數據開放的大資源，也是一把大數據驅動的金鑰匙。政府的數據涵蓋能源、金融、交通、治安、醫療、環境、食品等不同產業，這些數據都是相對集中且非常重要的。

所以，政府數據的開放，是促成產業創新的催化劑，這意味著，政府的大數據政策，對整體數據產業的發展有多麼關鍵。我們可以期待政府對公共數據開放的態度，而各個產業也可以跟隨政府數據政策的腳步開始嘗試練習，進入大數據驅動，乃至於大數據變革的第三階段。

重點4——多螢時代

過去幾年，我們看到電腦被手機顛覆了，那麼，手機會被顛覆嗎？雖然短期內還不知道，但我可以預見有兩個新的「螢幕」出現：一是智慧電視，二是物聯網汽車。

智慧電視是你家裡的螢幕，收集你看節目的數據，以及推薦你節目的數據，形成了天然的數據閉環。物聯網汽車則是第二個非常關鍵的螢幕，將來，所有汽車內部都會像特斯拉一樣，由一個大螢幕控制汽車的每個部分，並記錄汽車行駛中的各種數據，因此產生資訊的流動。

最近愛立信公司（Ericsson）針對全球四十個國家、十萬名消費者進行一項未來載具的調查，研究結果顯示，超過一半的受訪者都認為，智慧型手機將在五年後被淘汰，取而代之的是具備人工智慧的新裝置。

如果你問我會不會有第三個、第四個螢幕出現，我的觀察很簡單，即從兩個層次的分配來思考這個問題：時間分配和載具分配。人在不同的時間區段，會因為當時的環境狀態不同，而對不同的裝置有不同程度的依賴。在家的時候，對智慧電視的依賴會比手機高一些。當我們離開家，往下一個目的地移動的時候，在公共交通工具上，我們需要的是手機，如果是自己開車，汽車導航或行車電腦的螢幕就會成為主要的關注焦點。

所以，說實話，我不太在意五年後我們拿在手上的那塊螢幕是否仍是手機，我更在意的是，人如何與那塊螢幕互動，以及在互動的過程中，我們如何收集到有價值的數據，並進一步對使用者的日常生活做出優化的回饋。

重點5——數據趨於產業化

所有大數據的實踐都發生在產業裡，過去我們看到，網路影響比較大的產業，必然容易數據化，已經冒出頭來的有金融、醫療、電商等產業。下一步的大數據，應該是不同的領域各自發展，不會有一個全盤通吃的方案轉移成每一個領域的解決方案。

在中國，不管是零售、醫療、教育、金融等產業，都會受到「互聯網+」的帶動而發展，這是對等的，因為很多小公司剛起步，產生了很多小數據，這是一個從0到1，然後整合碎片化數據，最後再到積累大量數據的三步進程。這三步進程的時間點，加上不同的應用，鑄就了產業大數據。

另一方面，我們可以觀察到，大數據未來將會從過去的淺層連結（weak link），轉變為深層連結（deep link）。在大數據由淺而深的演變中，我們可以發現，過往大家在看待網路的時候都習慣從產業出發，思考網路（數據）可以幫自己做些什麼；等到互聯網和大數據的時代，

是時候做些顛覆性的改變了，要嘗試以網路（數據）為出發點切入思考，再把產業的思維放進來碰撞，看看可以激蕩出什麼樣的創新思維。就好比Uber、Airbnb都顛覆了傳統產業運用網路的概念。這種思考與創新的方式才能將跨領域的東西提煉出來。

提到產業，我簡單提幾個比較顯著的發展，來看看未來產業怎麼看見中間層的機會，從中看到新的問題與創新的機會。

金融業與保險業的機會

如果要我用一個字道盡金融在大數據時代的機會，那就是「微」（micro）。過去很多的創新都被技術和數據的能力所局限，而未來，數據的收集、加工和應用，都將實踐個人化的價值，這將會激發出很多新的金融商業模式。其動態的意義，一方面展現在金融保險體系中對客戶的掌握，因為過去個人信用評比往往無法有效反映最新的個人信用風險，導致銀行或保險公司無法提供最符合顧客需求和利益的服務。另一層意義在於，解決服務合理性的問題，也就是當顧客使用服務的時候才收取費用。

過去，我們對汽車保險的保費設定，是基於客戶過去的駕駛肇事紀錄來調整費率，其背後所代表的意義是，這個汽車保險的遊戲規則，是維繫在投保車主的駕駛安全行為上。但是認真想想，肇事紀錄已經是相對落後的指標，在大數據時代，難道沒有更動態的數據可以預測危險駕駛的風險嗎？當然有，而且來源還不止一個。

未來的汽車都會像特斯拉一樣，車上的感測器會記錄駕駛人怎麼踩油門（比如，習慣性的緊急煞車，就反映出一種危險駕駛的信號）、變換車道時是否會打方向燈、是否會頻繁地按喇叭等，這些資訊都從側面反映了駕駛人的駕駛習慣是否安全。如果把這個駕駛人的行車路線數

據，比對政府公布的危險肇事路段數據，這樣就可以知道這輛車每天上下班的路線屬於怎樣的安全等級。綜合以上這兩類數據，即便沒有肇事紀錄，保險公司仍能根據這些數據，調整對這部車的風險評比，隨時機動地調整保費費率，也就是採取動態費率。

同樣，如果車險是為了確保客戶在駕駛期間的風險，那麼，透過車輛感測器，就可以清楚知道這輛車有多少時間是停在車庫、又有多少時間是行駛在路上，保險公司就可以根據車輛實際承受風險的時間來計算保費，並向客戶收費，這就實現了前面提到的動態計價。

醫療產業的機會

醫療產業目前面臨的最大問題是數據不整合，例如，我在A醫院看病，卻拿不到之前在B醫院的就醫病歷。另一個大問題是，高齡化導致了嚴重的醫療負擔，只要一生大病就沒錢買藥或就醫，只有想辦法降低醫療成本或藥物成本，減少濫用資源等情況，才能從根本上減輕政府負擔，把資源配置到真正需要的人手中。

美國福特公司的三十萬員工，每年享有高達30億美元的醫療保險預算，但這筆錢過去只有一個人在管理。1997年，福特公司採用他們第一個數據應用分析之後發現，竟然有人150歲了還在領醫療保險，有人一年領兩次懷孕補助等等。這些都是資源的浪費，如果不透過數據分析，福特公司可能永遠也不會發現這樣的謬誤。

醫療是一種連續性的行為，一個人從健康、亞健康到生病的階段，都不是突發的，背後都有遺傳或生活習慣、飲食習慣的脈絡可循。醫療數據的互通，可以讓醫生了解一種疾病和另一種疾病之間的關係，很多疾病的危險信號常常因為資訊不通而被忽略，導致無法在第一時間察覺，最終延誤治療。美國就曾經透過傳染病傳播數據預估，提前預測要

生產多少疫苗，以及各區的疫苗使用狀況，從而對疾病的防治帶來重要幫助。

零售業的機會

對零售業來說，最重要的就是如何利用數據，讓供應與零散的需求能夠配合。大數據讓賣家知道消費者心裡面要什麼，讓賣家以最快的途徑找到消費者想要的東西，同時給予消費者最好的價格，並使用對消費者來說最方便的方式付款，最後，在恰當的時間送達，這樣的賣家才是贏家。

從供應方來看，如何用最小的庫存、最快的方法、最合理的利潤率，來滿足消費者的需求，從而使供應鏈的處理變得更好，成本浪費減少到最低。

以數據驅動為基礎的線上零售已經發展十幾年了，那麼，線下未來會出現什麼情況？當興趣點（Point of Interest）技術逐漸成熟，人人手上一都有一支手機，處處都能發揮大數據的連接能力，時刻都是機會。但是，人們不會特地打開某個App來購物，即便線上購物很發達，它在某些時候仍然不是最方便的購物管道，因此，改善消費者體驗型態的最好方法，就是覆蓋全通路。對零售業來說，線上、線下都能覆蓋到是最好的，只有線上App是不夠的，最好連線下通路都能拿到，不然很容易被對手「彎道超車」。因此，全通路的打通和合作策略，是零售業在未來必須關注的第一個重點。

我還想談談推薦功能。現在，我們可以在同一時間享受多種服務，流覽多種產品資訊，當幾十萬個跟你有關的商品展現在你眼前時，你要怎麼選擇？主觀上，大家都希望可以自己選擇，但是有這麼多選擇的時候，選擇反而變成了一種負擔。因此，未來的推薦功能，應該是游離

於「優選」與「逛」之間。

大數據讓手機變成個人消費助理，不斷跟著消費者走，也不斷領著消費者走，手機成為商家和消費者之間的一個媒介。

我最後要談的是作為零售業根本的生產問題。數據是不是可以成為產品創新和改良的依據？我們可以觀察並不斷優化從設計到生產、包裝、銷售、售後等過程，最終生產出符合顧客需求的商品。

雖然是老生常談，但仍要強調，大數據時代對零售業來說不是單純的轉型問題，而是一個攸關生死存亡的問題。

關鍵思考 | 走出大數據和小數據的迷思

我們談了許多有關大數據的趨勢和觀點，最後我想用一個思考性的問題來對這些趨勢做個總結。每次講到大數據，大家總是喜歡在數據大小這件事上爭論不休。那麼，數據的大或小，究竟對我們意味著什麼？

簡單來說，企業對數據收集的認識發生了從量到質的轉變。越來越多公司意識到他們收集的大部分數據除了占據存儲空間外，並沒有發揮太多作用。慢慢地，企業在決定要收集哪些數據的時候，不再像以前那樣粗放，不再覺得數據可能有用，就先收集了再說。

過去，人們把太多的注意力放在大數據本身，所以忽略了一個根本性的問題，當有穩定的數據提供給你時，你會用它來解決問題嗎？當集中的小數據還未被利用，可能是你對問題、對商業的理解還不夠透徹。因此，很多大企業在理解了這一點之後，不再迷信大數據，而是更踏實地收集一些對解決企業當

前問題有用的數據。不少企業對大數據的態度開始變得更謹慎，並且走向專業領域。

不用再斤斤計較於大數據的詞義，將關注焦點集中在怎麼樣看待問題、看懂問題，然後試著透過數據思考，最終落地實踐，這才是企業應該認真思考的下一步。

請記住，如果每個人都不去實踐，而是一味地談論趨勢，那麼，這些趨勢最終也不過是路邊算命先生的信口開河而已。

結語

開啟你個人的大數據管理

過去幾年，數據分析師問我最多的問題是，如何才能擁有大局觀，洞悉一切。一般人問我最多的問題則是，如何才能保持幸福快樂的生理和心理狀態。

這一切的答案，要從命運開始說起。

我一直認為，命由天注定，而運是可以改變的。人生絕非宿命，公司和個人並無二致。

人一生下來，體質、DNA和天賦往往都是注定的。而運呢？卻是由不斷的選擇和決策所組成。

命就像一輛離開生產線，出了工廠的汽車，它的品質和耐用程度，從離開生產線時就注定了。運卻是這輛車後天開闢出來的道路，大路與小路不同，每一條路的選擇是對是錯，對人生的影響或大或小。有時，小路也能使人生出現極大的轉折。

在我看來，成功路上，有四種東西可以累積，分別是財富、人際關係、知識及思維方式。

前兩者大多數人都能夠直接明白，比如，沒有錢很多事情都辦不好，多個朋友就多條路。知識則是指人從小學、中學、大學，以及進入職場獲得的各種知識和經驗，在現在的職場上，每個人的競爭能力及賺錢能力，大多和知識有直接關係。

不被絕大多數人知覺的是第四點——思維方式。思維方式與你所做的選擇和決策密切相關，冥冥之中決定了運的走向。

我們生活在大數據時代，各種數據充斥在生活環境裡，每天都能收集所見的文字、影像、聲音，以及各方面的資訊。在這樣一個資訊氾濫

的環境裡，很多人完全沒有關注過如何累積、分類，以及使用當前越來越容易獲得的大數據，更別說讓大數據變得能為個人所用。

數據的累積，從收集到還原

有的人每天都會在同一個十字路口迷路，導致每天都會遲到十分鐘。為什麼會出現這樣的錯誤？這是因為，他根本沒有重視數據的累積，也沒有想過使用這些數據去產生價值。

在我們之中，絕大多數的人對事物的態度也只是觀察，卻不會思考。我的做法是，我要累積數據，累積我所在的產業中出現的資訊。

這是一個有限時間和無限大數據博弈的時代。如果沒有累積數據的意識，沒有培養自己數據化思考的意識，改變命運，就真的成為一種投機行為。

古語有云：「人無志不立。」如果你有志向，你就會知道自己應該累積什麼樣的知識、財富和關係。如果你進行大數據管理，到了某個階段，你又擁有很好的數據化思考方式，你會發現，財富、關係和知識的積累完全能夠為你所用，到那時，你的人生就會有很大的不同。

讀到這裡，你應該已經感覺到了，本書所講的大數據內容，其實完全可以為個人所用。我可以用我的經歷為例，我相信，如果你現在開始用大數據的方式來管理個人知識，開啟個人大數據的累積，不超過六個月，你就會發現自己開始變得不一樣了。

我的個人大數據管理方式

我現在每天早上4點半起床，熟悉我的人都知道，我每天早上必做的一件事情是「品覺晨讀」。做這件事情大概有六年了，它是我個人大

數據管理中的數據收集工具，就好比營運數據時的養數據概念。

今天，如果你開始關注某一個領域，想累積某一種知識，在養數據的過程中，你會如何開始？這就是一個棘手的問題。沒有人能夠在第一次營運數據時，就知道這個數據是好的、是對自己是有用的，或者知道什麼數據不好、對自己沒有用。就像開始做企業大數據時，我根本不知道哪些數據一定有利，或者在未來可能對我產生重要的價值。

所以，我一開始會把每天關注的那些數據收集回來，放到我的Evernote裡，為什麼我選擇用Evernote呢？因為方便我做數據管理，我可以用各種裝置，包括用手機來收集數據，把每一份收集回來的數據進行分類、加上標籤（結構化），方便我隨時隨地調用。

時至今日，在我累積的數據中，已經有超過上萬篇相關的簡報及早上收集的電商新聞，如果沒有一個很好的分類和標籤管理系統，我如何能做到這一點？所以，要完成大數據積累的循環，放進去的數據，一定要能夠非常快速、即時地調用。我一直在訓練自己，如何使用兩、三個關鍵字，十秒內就調出我累積的有效數據。當我想要使用數據幫我做決策時，有效的數據就在手邊，這個時候，營運數據，即養數據，就已經產生了作用。

現在，我依然每天繼續收集數據，我可以透露自己收集數據的一個「狠招」。在我的微博關注裡，我按照兩個面向來分類我關注的人：一是他們懂什麼東西，即有什麼特長；二是根據重要性，分為五～二十分鐘，或者四十五分鐘的關注度。這樣劃分的意思是，如果今天早上我只有五分鐘，我就只看重要朋友發布的資訊，如果我有十多分鐘，我會把重要性為五～二十分鐘的朋友發布的資訊都看完。

從收集數據的角度來看，首先，我在選擇有效的數據來源；其次，我有重要性的選擇，所以我能很好地控制我的時間。每天我做完這些數據收集的時間大約是一小時，時間久了，我現在對每個數據來源，甚至提供數據的記者能力都瞭若指掌。某位學者的觀點是怎麼樣的，騰訊的

記者能力如何；哪個網站週六、週日不用關注；鳳凰網比較勤奮，週六是有人上班的；騰訊網最努力，每天都有新聞。甚至，各家媒體參加了哪些會議，這些我都知道。

我現在已經對數據積累與決策品質之間的關係具備了一種直覺，所以我開始對電商有「覺」了。這個過程就如同前面所說的三訣竅——「混、通、曬」，我每天混在電商的環境中，不斷吸收和尋找規律，這是「混」；打通每天收集來的數據和每天所處的電商情境，則是「通」；至於「曬」，就是這些數據如何影響我在工作中的決策。

2013年年底時，我聽到一些數據分析師在討論年報，他們不斷在講2013年阿里巴巴的年報。我問他們有沒有看新聞，「我們的競爭對手最近有什麼新動作？京東的CEO劉強東還想做什麼？新聞是真是假，你們關注了嗎？」所以，單獨討論這份報告是沒有意義的，因為他們沒有將情境融入其中，只講阿里巴巴本身，報告的意義也就無從談起。

當然，在營運數據和數據化營運的閉環裡，我們一定要假設一切都在變化，一定要去關注還有沒有新的數據收集源頭，現有的數據來源會不會出現問題、我還欠缺哪些新的數據源頭，比如，近年興起像虎嗅網這樣的新媒體、199IT，以及特別喜歡發表大數據文章的人。我作為所謂的「大數據專家」，怎麼可以不知道這個領域裡誰有料、誰沒料？誰是人云亦云呢？

這就像和業內的高手過招，我們在積累的過程，是在動態中不斷迂迴地走向目標，不斷衡量這個產業的變化，就像李小龍一樣，在和高手博弈。

個人大數據管理的挑戰

在養數據的過程中，關鍵字標籤是數據管理最重要的工具，透過標

籤，我們可以快速定義一篇報告的內容，而每次調用數據的順暢情況，則可以用來衡量這個標籤是否有效。

在個人大數據管理中，關於標籤最常見的矛盾是，你為某篇文章貼上一個標籤，但這個標籤是動態變化的，雖然在你設置標籤的當下是正確的，但是過了一段時間，你會發現這個標籤失效了。舉個例子，你給某個朋友貼上的標籤是「年輕女性」，但是過了十年，這個標籤就失效了，必須進行升級。我們把這個現象叫作標籤的生命週期。

標籤失效後，你需要更新索引，這個過程所花費的力氣非常大。多年來，我不斷尋找對應電商的標籤，以下是我總結的三點經驗，可以分享給大家：

- 越有效的標籤，越能讓我快速調取數據。標籤的用途多樣，比如，知識分類、識別來源、描述情景或人物，以及表明時序等等。
- 要注意知識範疇的培養。例如，在電商中，我運用各種分類標籤，有產業標籤，如B2B、B2C、C2C、外貿B2C、外貿B2B，最新的還有O2O；以及技術方面的標籤，有大數據、行銷方法、使用者體驗、個體行銷等等。
- 情境的標籤，我分為公司與人物，再加上時間。

可見，由時間鏈、共識面（公司、人物、來源）、知識體系（標籤）和格式（內容）所構成的知識圖譜，是解碼決策分析的依據。表1是我做數據分析時使用的知識圖譜部分內容，以供參考。

收集時間	公司	人物	來源	電子商務	格式
過去兩週	淘寶	馬雲	新浪科技	B2B	產業報告
3 個月內	天貓	馬化騰	TechWeb	B2C	數據分析
6 個月內	京東	劉強東	騰訊科技	C2C	新聞
6 個月外	1 號店	陳年	艾瑞	O2O	
	唯品會		易觀	物流	
	凡客			行動	
				社群化	
				大數據	

表 1 各大電商平台的知識圖譜

假設你現在想參考馬雲一篇關於物流的文章中的數據，你回憶這好像是上週的新聞，所以，調用這篇文章的面向就是「過去兩週＋馬雲＋物流」。

如果想多一個面向去看的話，想增加共識面，就可以用「過去兩週＋物流」，看看在這兩週內還有誰提到了物流。

你會發覺，在這個知識圖譜中，有些標籤比較歷久彌新，即生命週期比較長，比如，人的名字，不論這個數據從哪裡來、在哪裡、誰說的，都不會對數據標籤產生影響。但有些標籤是對趨勢的界定，就需要經常總結或細分。

收集到的數據，我都會像這樣貼上標籤，當需要調取數據時，我會在心裡問一下數據的來源是哪家媒體、哪個人、什麼時候獲取的，如此一來，時間線和來源就成為我鎖定所需要數據的重要線索。

從另一個角度來說，當我想知道Amazon在物流方面的新動向時，我會先鎖定物流產業和公司標籤，再配合時間線做決策。如果我想到了「大數據＋周濤＋三個月內」，我根本不用想文章的細節，就可以將文

章調取出來。這就是大數據帶來的改變——我們只需要記住幾個標籤作為線索，就可以輕鬆調用一萬篇文章中的數據。

但千萬不要覺得你可以一勞永逸了，因為這些標籤會隨著時間變化而失效，數據管理就成為重要的一環。

數據分析師都有著近乎變態的數據潔癖，他們對數據的來源極為看重。所以，我將人和公司區別開來，人是特別的，可以離開某家媒體和公司，是流動的。如果我將數據管理定位在人之上，就表示我認為人更重要；如果定位在公司和媒體之上，那就表示公司或媒體更重要。

這就是變化，在大數據的世界中，沒有人要求你懂細節，也沒人要求你成為數據分析方面的專家，但是要求你，在你需要拿到一個專家的數據時，能夠快速地調用。能夠做到這一點，你就突破了人類短期記憶的弱點。

這也正是過去六年裡我做品覺晨讀的原因，一方面是方便我做個人大數據的知識管理，另一個目的就是提醒我，這個數據化營運和營運數據的循環系統不能出現中斷，提醒我要堅持不懈地維持這個循環。

品覺晨讀也不斷帶給我回饋，比如，我發完微博後，有多少人轉發、多少人評論、多少同行專家幫助我，一起優化我的大數據內容和標籤。這使得我在做數據收集的時候，也了解到數據收集源頭的重要性。借助別人的力量，即在社會上過濾，會讓你的數據收集變得更有效。

當然，品覺晨讀的文章只是我收集文章的一部分，在我經歷多次數據化營運和營運數據的循環之後，我可以快速知道一篇文章是不是對我有價值，這加快了我收集文章的速度，但是，快思（加快判斷）的同時，很容易出現盲點。為了減少盲點，每當我懷疑自己的判斷時，就會列出一部分可能出現盲點的文章，放在微博，其他人自然會告訴我哪些是有用的，然後我再觀察這些數據，甚至還可能從中找出新的數據分類或標籤。

人人都能成為數據分析師

個人的大數據管理，我已經做了六年，過程中困難重重，但就像大數據管理一樣，不能因噎廢食。我保證，只要你開始用這個方法，三個月後，最晚也是六個月之後，你會變成另一個人。

《思考》（*Thinking*）中有一個例子，和我的思考，以及我在這六年裡所做的事情極為相似。書中說，如果你把一個人的左眼蒙起來，三天後，從醫學的角度來看，他的大腦和左眼相關的部分會退化。意思就是，眼睛與大腦是互相影響的，你看到的將影響你的思考，反過來，你的思考方式又影響著你看到的東西。

所以，如果你六個月內不斷地累積知識，你會發現，你的思考結構會改變了——不僅僅是知識量的改變，更是思維方式顛覆性的改變。這是因為你收集數據的方法已經不一樣了，你累積的東西也不一樣了，你變成了另一個人，你會覺得你「混」得特別好。

也許你已經躍躍欲試，但是我從這幾年的經歷提煉出一些建議，希望值得你關注。

首先，你選定的題目一定要是你感興趣的，或者跟你的職業相關。比如，你喜歡寵物，想成為養貓的專家，你就每天收集養貓的數據，當別人問到怎麼養貓時，你能很快地知道某個專家說過相關的內容。

其次，做個人大數據管理和做大數據一樣，一定要從小處著眼。比如，你的目標是要成為電商專家，這個目標在一開始就顯得過大了。以我來說，一開始，我只收集出口的相關資料和數據，在這個小領域裡，形成一個小的大數據循環，熟悉了這個循環後，再涉獵手機支付產業方面，然後是C2C和B2C產業等。

在經歷過小循環之後，我才允許自己去觀察別的領域。現在，在別人眼裡，我是一個數據營運的專家，其實我也是這樣「混」出來的。即

便是現在，我也一直在訓練這個循環。

再者，盡量讓數據的收集變得簡單。當你發現有用的知識後，一定要跟自己說，不要多做思考，先收集。現在就收集，不是明天，更不是後天。別人在簡報時，看到有價值的內容立刻拍下來，而不是待會再跟他索取檔案。

知識的收集永遠是「現在式」，而且這個操作越方便越好，越快越好。我的收集就很簡單，設置一個快捷鍵，一下子就能將數據抓取過來。

最後，你一定要把這個「營運數據—數據化營運—營運數據」的循環打通，也要利用今天社會化的優勢，學會借助別人的力量。我認為這才是真正的大數據管理——形成一個有效的循環，形成一個可以不斷獲取數據和反饋的動態循環。

個人大數據管理能夠為個人產生力量，讓知識累積，成為你在人生道路上，從「命」到「運」最重要的分水嶺。

現在，除了我自己這樣做知識和數據管理，我還想讓我的助理也加入這個循環系統中，只要這個方式的方向正確，做好協同，我甚至可以找八個助理一起用同樣的方式轉動這個循環。而我只需要確認一個步驟：當我問他問題時，我該用什麼關鍵字。

這就相當於建立一個新的知識體系，建立一個新的循環，我和我的助理就是這個圈子大數據知識的協同者。

無論如何，我們的人生就是要想盡一切辦法累積知識，首先要明確累積的目的，之後才是累積的方式，最後是如何調用累積的東西，以及明確它是否影響了我的決策。只有對這個數據有感覺的時候，數據才能對自己的職業產生影響。

這樣下來，我堅信，兩年後，你就可以成為某個領域的專家，而且這套方法是可以被複製的。如果我喜歡貓，從今天開始，我要成為一個精通養貓的人，只要按照這套方法做，我想，兩年後，多數人將無法在

相關知識領域跟我相比，因為我有一套累積知識的系統性方法。

如果這個例子得以複製，將會有一大群人跟我一樣，那我就建立了一個聯盟。第一天，收集、決策、行動、管理；第二天，再繼續收集、決策、行動、管理，每天都這樣實施，開啟你個人的大數據管理，改變你的知識積累和思維方式，成為能夠掌握自己未來的人。

後記

像李小龍格鬥般思考

每個男人小時候心裡都住著一個武打明星，或多或少都會作一些武俠夢，幻想自己能夠天下無敵。在我小時候，崇拜的偶像就是功夫巨星李小龍，他不僅是電影明星，更是一位武術大家，他用他的功夫開創了一片武俠新天地。

李小龍曾經這麼描述功夫：

「功夫是一種特殊的技巧，是一種精巧的藝術，而不是一種體力活動。這是一種必須使智力同技巧相配合的精妙藝術。功夫的原理不是可以學得到的，它好像科學一樣，需要尋求實證，而由實證中得到結論。

「在經過四年嚴格的功夫訓練後，我開始了解，也感覺到柔能克剛的道理，也就是如何消除對手的力道，減少自己力道浪費的手法，而這一切，都必須先求得氣定神閑。話聽起來很簡單，但實際做起來卻很困難，一旦和對方交上手後，我的思想就再也難以保持清明而不受擾亂的狀態。尤其是在對過幾招之後，我就忘了柔的理論，唯一想到的只是，不管怎麼樣我也得打贏他。」

「以無法為有法，以無限為有限」，是李小龍創立格鬥術截拳道的綱領和要義。他認為，這種方法不固定的格鬥術的精髓就在於——在動態中制勝。

他認為，穩定應該是在動態中的穩定。之所以要保持動態中的穩定，則是因為任何的格鬥都充滿了不確定性，如果只能做到特定環境下的靜態穩定，那麼，在瞬息萬變的格鬥環境中，一定會出現逆轉機會，而這個逆轉通常就是失敗或成功的關鍵。對動態中穩定的關注，目的就是希望能夠把環境中不確定的東西，變得相對確定而有規律。

如果我們把格鬥一般化，就可以推廣到商業的博弈中；如果把李小龍的格鬥術一般化，就會變成一個非常重要的商業方法論。

忘掉必殺技

很多人在剛開始學習一個全新的領域時，總是會很崇拜老師高超的本領，常常迫不及待地問他們：「有沒有什麼絕招能讓我快速達到像您一樣的水準呢？」答案是否定的。

從經濟學的角度分析這件事情，你也會發現，這個世界上其實不存在輕而易舉就能獲取的絕招。試想，如果輕而易舉就能獲得絕招，結果是每個人都可以獲得所謂的絕招，當人人都有時候，絕招就再也不是絕招了。

在我和一些電商討論營運數據的過程中，也經常會被問到類似的問題，比如：「我的首頁要如何改版？」「我這個按鈕應該放在左邊還是右邊？」「我應該觀察什麼數據？」等等。其實，這些問題都是追求必殺技的一種問法，而且這樣的問題都是沒有固定答案的。所以，從現在開始，忘掉必殺技吧，因為世界上永遠沒有必殺技。

如果沒有必殺技，在與高手過招時，應該用什麼招術呢？招術就是針對對手出的招，反制他的招，這就是有用的制勝技。

所以，當我被問到首頁要如何改版時，我沒有絕對的答案，我可能會反問：「你們公司首頁的主要功用是什麼？」

當我被問到按鈕是放在左邊好還是右邊好時，我會接著問：「放左邊和放右邊的區別是什麼？」

當我被問到企業應該觀察什麼數據時，我的問題就會變成：「你看數據的目的是什麼？」

只要問我的人給我的答案不同，那我給的建議一定也是不一樣的。

如果我得到的答案還不足以讓我獲得足夠的資訊，我就會針對答案不斷提出新的問題。

為什麼我要不斷地問問題？原因很簡單，就是我想知道問題到底出在哪裡，找出每一個問題背後不一樣的原因。這個過程是我在思考問題時常會用到的方法，我把它稱為「迂迴」。

在迂迴中尋找實踐機會

迂迴，不是直奔一個目標，而是透過不斷地尋找實踐點，慢慢進行。可能某個目標看起來很簡單，似乎馬上就能實現，實則是困難重重。在格鬥中，每個人都知道擊中對手要害的重要性，可以KO獲勝，但如果直接朝對方頭部出重拳，那一定會被對手防死，反而使自己露出很大的破綻，讓對手有機可乘。所以，很少有比賽能夠瞬間擊倒對手（除非強弱分明），高手都是在不斷嘗試的過程中發現機會。

有一次，我在和一家電商公司的管理者交流，談到購買流程，他問了我應該凸顯「立即購買」，還是凸顯「加入購物車」的問題。

他之所以在思考這樣的問題，是因為他認為購買流程對電商至關重要，而之所以會有這樣的困惑，則是因為他發現了兩種不同的購買流程。在商品介紹頁上，淘寶凸顯的是「立即購買」，而京東和Amazon凸顯的則是「加入購物車」。這讓他十分不解，淘寶是最大的電商，之所以這麼做肯定有它的道理，但是淘寶和該電商公司的性質不同，京東和Amazon的模式與其更相像，然而，這兩家公司又沒有淘寶做的規模大。

從這個問題來看，這名管理者思考問題的角度已經非常深入了，對這個問題的各個方面都已經考慮過。當時，我並沒有給他我的解決方案，因為我也不知道該怎麼解決，我只說了一句話：「先切出一部分流

量來試試。」

這麼做的目的是，我希望在開局之前，先能夠有一些變化呈現出來，再利用這個變化來破局。破局之後，一定會有數據收集回來，有了數據之後，就有一個可以參考的原料。就像在格鬥中，幾下試探之後，高手通常就能找到對手的攻防路數，再針對對手的路數做有效攻擊。

然而，破局後，問題也會隨之而來：收集了數據後又該怎麼做？怎麼樣算好，怎麼樣算不好？這又是一個大問題。要解決這個問題，就要去衡量破局之後的每一個數據變化。

衡量每一個變化

在變化中，找到正確道路的方式是，要能夠判斷什麼是對的、什麼是錯的，然後，保持對的，屏除錯的。如果無法有效衡量這些變化，這樣的變化就是無意義的。

我們可以再看一下前面提到的，應該凸顯「立即購買」還是「加入購物車」的案例。有了數據之後，如果最後測試的結論是，凸顯「立即購買」的轉換率，高於凸顯「加入購物車」的轉換率，這時，是不是就應該下結論——這個電商網站應該進行網頁改版，做成像淘寶那樣凸顯「立即購買」，或至少也應該在「立即購買」和「加入購物車」上設置同樣的比重。問題如果如此簡單的話，那思考也太容易了。

在討論這個指標之前，我們還有一個更重要的目標是——要讓這個電商網站有更高的銷售額。一般來說，電商有一個非常簡單的公式：銷售額＝訪問人數×轉換率×客單價。所以，衡量這次的改版，除了看轉換率有沒有下降之外，還需要關注另外兩個指標。在這個案例中，會不會因為直接購買更方便，而導致購買的筆數減少、單價降低？這兩個指標的降低，減弱了轉換率提高的效果，甚至會對銷售額的增長有反向作

用。從人的角度來看，到底哪些人對這樣的改變比較敏感，哪些人不敏感？對於透過廣告投放吸引進來的人來說，有沒有造成什麼改觀？

所以，衡量就是在目標確定的基礎上，能夠對一些變化做出有效判斷。就如同在格鬥中，我需要判斷，對方的變化對我的獲勝有什麼影響，我的出招對獲勝有什麼影響。如果無法衡量一次變動所帶來的各個方面影響，這樣的變動就只能算是一次沒有積累的變動，就有一點自娛自樂的嫌疑了。

對於電商企業，衡量變動有兩大前提：一是要保證數據能夠被有效記錄，二是數據能夠被有效解讀。

不少產品經理在設計自己的產品時，可能會花費很多精力在產品本身的設計上，卻沒有花費精力去思考如何衡量產品成功與否。對於成功與否這個話題，通常只會在他們的產品檔案中寫上一句類似「使用者體驗有所提升」的空話。這樣的產品自然不會設計數據收集的方案，自然也就沒有衡量的行為可言。

還有一類產品經理，設計了數據收集的方案，但是這些數據並沒有被實際運用，或只是被有限地使用。那麼，這樣的衡量也是缺失的，不利於進一步改進。

當然，和格鬥一樣，一招定勝負是不可能的，通常一次戰鬥會持續好幾回合，所以，我們需要對每一次的變動進行衡量，同時也要對每一次的變動有所反饋，用這次數據作為下次變動的基礎。

在整個格鬥過程中，反饋就像鏈條，只有透過反饋，把變動串聯在一起，才能讓整個變動過程朝勝利的方向發展。

持續的反饋

在思考過程中，反饋是一個非常重要的環節。如果說，格鬥中的穩

定來自於動態，那麼，思考中的穩定，就來自於變化環境下的不斷反饋。

每一次的反饋都是迂迴中的一個新起點，也是一個新的衡量開始。在持續的反饋過程中，整個思考過程將會變成一個有效的迴路。有效的迴路，就如同腳踏車上的鏈條，能夠將動力傳導到車輪，帶動車子向前行駛。

大數據環境瞬息萬變，基本上每秒都會產生不同的規則。十年前，誰也不會預測到網路零售能夠超過兆元人民幣的規模。同樣，十年後，我們也無法預知社會到底會變成什麼樣子。在面對飛速變化的環境時，對環境變化的適應尤為重要。就像在格鬥中，面對對手不斷變化的招式，自己也需要不斷地調整招式來應對。否則會像程咬金的三道板斧一樣，雖然看起來威力無窮，但是一旦三招用完，就再也沒有新東西了，只好催著他的寶馬逃跑了。

我曾經和一些電商交流，在談到營運數據的問題時，我指出，很多時候，我們往往只是簡單地衡量數據，並沒有持續地反饋。

所謂持續的反饋，就是在衡量的基礎上，評價一個變動到底起了多大的作用，並確定數據的變化確實是因為這次的變動而出現。現在，很多電商吸引用戶的方式都是採用線上廣告，這就是一個需要不斷反饋的過程，比如，今天我投放的媒體到底怎麼樣。可能僅一天的數據無法解釋，需要透過很多天的數據才能做決定。需要觀察這個媒體的投放效果好不好，對哪些商品好，對哪些商品不好；效果好的商品是不是可以加大投入，效果不好的商品是不是需要更換媒體；對哪些用戶是好的，對哪些用戶又是不好的等等，這都是需要解決的問題。對於一些新成立不久，還沒有建立良好口碑的企業而言，透過不斷的反饋，將會在很短的時間內建立起屬於自己的經驗體系。

尋找爆發點

像李小龍一樣做商業博弈，其中涉及很多方法，這些方法的最終目的，當然是要獲取戰鬥的勝利。所有的迂迴、衡量和反饋，最終目的就是為了尋找機會的出現。

我在上海認識一位研究國學的高人，我曾經向他請教如何讀《孫子兵法》，他當時對《孫子兵法》做了非常精要的提綱——戰機。要讀懂《孫子兵法》，就必須從中讀懂如何尋找機會。

《孫子兵法》中有很多章節，與我們寫到的迂迴、衡量和反饋的方式非常接近，書中有一句話叫「多算勝，少算不勝，而況於無算乎」。用現在的話來講，就是選擇迂迴前進，在營運數據中衡量每一個變動，並透過變動的結果，調整自己的下一個迂迴動作。

我想用一個簡單的例子來說明，有兩輛坦克，一輛射程是三千公尺，另一輛是三百公尺（我不是軍事專家，可能射程的說法有些不妥），如果射程三百公尺的坦克想獲勝，它應該如何尋找機會？

首先，這輛射程三百公尺的坦克必須要知道對方的射程和己方的射程，如果處在對方的射程之內、自己的射程之外，那只有當靶子的份。其次，要找到一個合適的方案，是迂迴前進，還是做好偽裝，等待對方靠近。最後，當對手出現在射程內，要用什麼方法擊中對手，這就是尋找機會的過程。當然，如果對手沒有按照自己的假設行動，則可以根據對手的情況，不斷地改進方案，最後找出正確的方案。

類似的情況在企業中也是屢見不鮮，我認為，未來的數據角色，對企業而言就像決鬥，如果兩個人程度差不多，多「算」的人就會勝利，而數據就是「算」的核心。但是，在尋找機會和「算」的過程中，還有一點需要非常注意，即分析方法一定要客觀和理性，在經驗不是特別豐富的情況下，千萬不能盲目地用數據去觀察問題，要保持相當清醒的頭腦，不能讓潛意識迷惑自己。

總之，機會就是基於不斷的迂迴、衡量和反饋，透過合理、有效的邏輯思考方法，找到問題的關鍵，即問題的主幹，因為抓住主幹才能獲得成功。

前面提到了很多都是在動態中尋找機會的方法，在動態中找到的機會，如果能夠抽取出來，就是一件事情的主幹。主幹就是一件事物最原始、核心的規律。舉個簡單的例子，我們會說「打蛇打七寸」，「七寸」就是打蛇的主幹，這個主幹一定是根據無數的打蛇經驗得出的。碰到蛇，如果沒有打到七寸，我們可能會被蛇攻擊，甚至可能喪命；如果打到七寸，就可以輕鬆獲得成功。

我把這種抓住主幹，透過主幹解決一系列問題的情況，稱作「爆發」。但是有一點一定要明確，除非我們可以肯定自己的環境與別人完全相同，否則別人的爆發點不一定就是自己的爆發點。換句話說，別人的主幹問題不是你的主幹問題，所以，簡單套用解決方案的做法，可能會害死自己。以前，傳統企業一窩蜂地導入ERP系統，這些年IT公司又一窩蜂地搞雲端運算，就是簡單地套用別人的方案來解決問題的失敗案例。

使用數據的過程中，其實也存在爆發點。我以前寫過一篇文章，提到數據扮演的三種角色，分別是：

- 第一個階段，有多少數據就收集多少數據，數據只從點上指導營運，並不指導戰略和方向。
- 第二個階段，數據要精準化，數據之間開始有了關係，擔當的角色從點到線，或者面，就可以作為戰略參考。
- 第三個階段，做數據模型，分析外圍數據開始變得非常重要，可以為公司的下一步戰略找到出路。

在這三個階段中，數據真正能夠充當爆發角色的，一定是第三階

段。但是，如果沒有前兩個階段收集和累積問題作為基礎，第三階段的爆發點就不一定會出現，即便出現了，也需要考量其真實性。

爆發，是做每一件事情和解決每一個企業問題的人都需要考慮的，尤其是做數據的人，一定要時時刻刻想到數據能夠為企業產生什麼價值，能夠用數據解決什麼問題，是不是能夠用數據發現企業中的爆發點。如果做數據的人時刻都有這樣的意識，那麼，企業將會因為數據獲得非常大的收益。

所以，我始終有一個看法，尋找爆發點的過程，其實就是自己和自己戰鬥的過程。在這之中，需要不斷忍受痛苦和折磨，不斷嘗試和精進，直到找到自我突破點。

請記住，很多時候，決鬥對象不一定是他人，而是自己。

附錄

回應台灣讀者的問題

未來，沒有一家公司不是大數據公司。2014年11月11日，阿里巴巴集團透過長達十六公尺的大螢幕，即時呈現雙11購物節的交易額（GMV）、各熱門品類最受歡迎的品牌、各市場的交易額排名等反映出消費者行為的豐富數據，開賣不到三分鐘，成交額破10億元人民幣（約新台幣50億元），全天的商品成交總額達到571億元人民幣（約新台幣2855億元）。這種活用大數據的驚人本領，人人都想學。

但在中小企業占九成的台灣，大數據似近似遠，阿里巴巴集團前副總裁車品覺，特別根據自身研究大數據的經驗，針對台灣讀者最需要了解的七大問題提出看法。

大數據，跟誰都有關係

問：對很多人來說，大數據只是個流行詞，誰會需要大數據呢？

答：任何可以成為流行詞的東西，肯定是社會對它充滿巨大的期望和想像。大數據也不例外。建議企業在考慮使用大數據之前要從當前的問題著手，然後思考：

1. 理解業務中有沒有一些決策做得不好，使用數據可以解決嗎？
2. 在營運的工作流程中，有沒有錯誤可以利用數據快速糾正呢？
3. 數據是否有可能幫助提升產品或服務的品質？甚至直接成為創新的元素？

以上幾個問題，是希望大家明白，大數據的運用，和你對業務是否有深入理解相關。從這個角度去看，數據化營運無孔不入，跟誰都有關係。

然而，在使用數據之前，企業還要關注一些基本功，數據平台的軟體支援、如何做好數據的加工準備（包括新舊數據整合）、提煉數據到解決問題的能力。如此，才有資格談大數據。

所以，中小企業都應該用數據去認知自己哪些做得好或不好（數據化營運），數據作為一種工具，可以幫助中小企業了解自己，也可以優化業務。但是，經營的本質還是取決於創始人的方向與管理，大家不能本末倒置，一味地期待透過大數據就能解決企業的所有挑戰。

問：台灣의 中小企業可以如何應用大數據？

答：大數據的力量來自分享、整合和產品化。例如，Google地圖之所以能告訴你前面的路塞車，其實是有賴於每個使用Google地圖的位置分享。所以我認為政府的推動，可以讓小企業減少得到數據的門檻，增加業界的數據共用，這樣就更有利於讓小企業也享受到大數據的科技。從產業鏈來看，小公司聯盟，把數據統一，用數據來解決一些業內彼此都不能解決的問題。

問：台灣大多是中小企業，相較於大公司，中小企業做大數據的方式、思維有何不同？

答：不管是中小企業還是大企業，在運用任何一種新科技的時候必然要量力而為。而中小企業和大企業的區別，在於中小企業的資源肯定沒有大企業那麼多，所以，中小企業不容易像大公司一樣有龐大的數據團隊。因此，中小企業在運用數據的時候一定要有更穩妥的方法，注重使用數據的效益，可以嘗試從小專案著手，再逐步拓展。

問：您曾提到大數據的應用講求跨界和創新，在實踐大數據的過程中，最困難的地方為何？

答：大數據應用講求跨界和創新，更準確地說，大數據的價值來自可以從多角度來看同一件事，全景觀察可以減少誤差及創造新的機會。但並不是要求你能夠認知到全部外面的世界，而是能讓其他人的數據為你所用。大數據實踐中最困難的地方在於你對自身的理解，再加上，隔行如

隔山，外部整合回來的數據可能很有價值，但同時也有很多噪音，大家並不完全清楚數據的來源和定義。

問：在資訊氾濫的環境，數據越來越容易收集，但也代表「噪音」越多。企業該如何找到核心數據並成功應用？

答：根據過去的經驗，我認為初期不要貿然就開始一個非常大的大數據專案，應該是要從小處開始。數據比較適合以小、具體、容易評估效果作為起點的專案，以此鍛鍊自己收集、加工、使用數據來做決策，以及衡量這個數據價值的能力，即以小知大。從小的情境開始，用數據在商業場景中不斷優化。

問：成為一名傑出的數據部門主管應具備哪些關鍵能力？

答：根據我的觀察，目前非常缺乏一種數據管理人才：對業務有足夠的理解、明白數據能為業務起什麼作用、了解技術更新對價值產生的關係、數據收集到加工、新數據與歷史的整合、使用數據的便利性等等。其中，對業務和商業的理解，絕對是成為數據主管所需要的基本條件，但若是想達到傑出的程度，肯定要懂得如何在人才匱乏的大數據產業中，吸引和保留住人才的眼光和能力了。

問：如何培養對大數據的敏銳度？

答：當在公司遇到業務問題時，問問自己：現在擁有的數據能幫我解決

問題嗎？假設所有數據都可以獲取，我需要什麼數據來解決問題？要怎麼做才能更容易獲取需要的數據呢？

舉例來說，我過去看到路上的交通狀況時曾經想過，大城市裡的計程車服務會不會有可能改善？我那時想著，如果計程車上有個燈能顯示過去客戶對他的評價，那麼司機為了保持好評價，應該會提供更好的服務，這就是數據可能解決的一個簡單例子，下一步才是如何設計一個容易的方法讓顧客去評價，而現在的叫車軟體就是一個很好的實現案例。

這是訓練數據敏感度的好方法，也是過去十年我個人一直在用的方法——透過周遭事物訓練數據敏感度，讓數字「說話」。

致謝

在修行路上，不要單打獨鬥

我相信，一切善良的力量都會有善果，生命中那些有聲和無聲的人，都會以他們的方式存在、生長，在你和他人修行的道路上，看見彼此、發現彼此和祝福彼此。

結緣甘孜

九年前，我隻身來到上海。有一天，忙累之餘，我學佛的上師忽然問我：「品覺，你要不要去甘孜看一下？」我當時沒有想太多，就回答說：「好啊，去就去吧。」是的，正是這一去，完全改變了我後來的生活。

我的上師是雪謙寺的活佛，在我與他結緣時，甚至連甘孜的「孜」怎麼唸都不知道。上師很少和我說一些大道理，第一次去甘孜，他只是帶著我，也不讓我問行程。到了甘孜後，上師讓我去看看當地的孩子。說實話，他們的生活狀態帶給我極大的衝擊。五、六十個孩子沒有鞋子穿，也吃不飽，和老師們一起擠在破帳篷裡。吃、住和教學都在十二坪的破帳篷裡，映入眼簾的，是他們破舊的穿著和臉上的笑容，我當時什麼都沒有說，身為一個飽經滄桑的老男人，在那一刻，我竟然哭了。

我不騙你，現在回想起來，我都不知道自己為什麼會流淚，也正是從那個時候起，我就下定決心，我要為這些孩子們做些什麼，而且不能再等了。有些事情可能是冥冥之中注定的，從2009年起，我開始拉身邊的人捐款，不夠的就自己貼，以確保每個月給一百個學生每人100元人

民幣的生活費。

你可能會問：「為什麼那時候你如此信任那些幫你們管理捐款的人？」因為那時候的管理者是五明佛學院的大湛布，他們負責幫學校設計課程，管理孩子的衣食住行，對於學佛的人而言，這一點非常可靠，也讓我們這些捐助者沒有後顧之憂。

自此，在工作之餘，我的生命裡多了一股無法妥協的力量——如果月底不按時匯錢過去，孩子就會沒飯吃。這種原動力，好像一條幸福的鞭子，一直在身後激勵著我。

信任，電商人的溫暖

甘孜受助的孩子中，有的是孤兒，有的是家境困難。2011年，我發現自己貼的錢越來越多，因為孩子們從一百人增加到兩百人，而且還在不斷增加。當時我日思夜想的就是，我必須要想辦法讓增加的孩子也能吃飽飯。

經常有人問我：「為什麼要自己掏錢，你也只是一個普通的打工者而已，這是一個無底洞。」我說不清楚，也答不上來，再說我普通話也不好，也就無從為自己辯護。可是每一次去甘孜，看到孩子們身體越來越好，老師們也很認真在教學，再看到孩子考試時自信的樣子，我就知道，我必須堅持下去。這種力量，是親眼所見帶給我的，是外人無法體會的。

後來，我覺得不能太依靠身邊人的捐助，我必須建立一個長期滾動的體系。因此，2011年，在家人及朋友的支持下，我開了淘寶店「桑珠貝瑪」，主要是銷售一些佛珠和茶葉，很多朋友也會將商品放到上面義賣，所得款項全部捐給甘孜的孩子們。正是因為這個淘寶店，我開始養成早上4點半起床做「品覺晨讀」的習慣，主要是為店鋪引流量，偶爾

也做做客服，並利用空餘時間採購，當時我幾乎是一個人把所有事情都搞定了。所以，從另一面來看，我深深了解到當一個淘寶賣家有多難。就這樣不斷堅持，之後店裡每個月都有7000、8000到1萬元人民幣的收入，也算大大緩解了當時的危機。

但是我明白，桑珠貝瑪淘寶店能有如此的收入，大部分並不是因為我，而是歸功於身邊的好心人，而且我估算，有一半的訂單都是阿里巴巴的同事，以及電商圈的朋友下訂的。本來，在阿里巴巴，我作為高階主管，本人以及親人都是不能開淘寶店的。在開店時，我對阿里巴巴廉政部說，如果我開店會給其他高階主管藉口，我寧可不開店。可是廉政部的同事對我說：「你繼續做吧，只要每個月公布收支，後面的事情我們來處理。」聽完他們的話，我覺得自己沒有來錯阿里巴巴。

就在前不久，忽然有人捐2萬元人民幣到我的支付寶，這2萬元讓我驚喜不已。後來我才知道，原來是我在支付寶時的老闆和同事捐的。他們現在常常開玩笑說：「以後同事之間打賭，輸錢的一方都給車品覺助學用。」雖然只是一句玩笑話，但同事卻真的這麼做了，讓我很感動。

還有一些電商圈的朋友，看都沒看過甘孜的孩子，而且一次都沒去過甘孜，卻仍慷慨捐助，例如，唯品會的沈亞、億邦動力網的鄭敏、廖斌等好心人一直在暗中幫助我。有一次，唯品會的財務在改動薪酬系統，到了月底，捐助的錢還沒匯過來，沈亞對財務說，先把孩子的錢支出來，其他再說。他們對我沒有任何懷疑，就這樣相信我，這種信任是支撐我在助學路上一直堅持走下去很重要的力量。

因為有一幫這樣善良的人在幫助我，我更要讓每一分錢都對得起捐助的人。有時候，帳目錯了，少了點錢，我就自己補進去。佛教裡面說：「別人捐的錢是公德，一毛錢都不能進自己的口袋。」這也就是我們在講善的用心。

做公益，最重要的其實不是錢

在甘孜助學的過程中，我和我的朋友開始明白，做公益，錢很重要，但真的不是最重要的，比錢更重要的還有兩樣東西。

一是要專業。記得2013年，一個朋友捐了上百個羽絨睡袋，那時正逢天氣忽然從十幾度驟降到零下。說實話，物資從那麼遠的地方運到甘孜，其實很不容易。當我們發睡袋時，我的心溫暖極了。從收到睡袋到發放，我們都有條不紊，大家配合得極好。

還有一次，我們捐衣服給孩子，但是發到最後，少了一件，有一個孩子沒拿到，他哭得很厲害。當時，我們把自己身上的衣服脫給他，但他就是不肯要。可能有人會責怪孩子不懂感恩，但事實並非如此，孩子是很單純的，他只是想追求和大家一樣的東西。所以，做公益真的不是一件容易的事情，其中需要計畫得很細緻，還要考慮到各個方面。因此，專業很重要，否則非常容易在過程中無意傷到孩子。

另一樣重要的東西就是愛和用心。在甘孜助學的路上，這麼多年，我有幸可以碰到許多有心的人，他們甚至從來不問結果，只是選擇無條件的信任。很多人捐錢、捐東西，從來不會尋求所謂的回饋和回報，在這些人心裡，只有一個信念：「你做了，我相信你，我相信只要對孩子們好，我們用心就好了。」這讓我很感動。

2010年，我們去甘孜的每個學校看孩子，很多學校連路都沒有，中途迷路了，凌晨兩點才到了一個小鎮，安頓下來後，我做了番茄炒蛋給同行的志工吃，他們說這是他們吃過最好吃的東西。這盤菜真的炒得那麼好嗎？不過是在那種情境下，大家都有一種使命感，覺得必須找到那所學校，而圍在一起吃一盤菜的溫馨畫面，確實讓人動容。

有很多人都在默默地做很多事情，有的愛很直接，有的愛默默無聞，例如，我曾經和銀泰網的Robin和今夜特價酒店的Mars說，我們自己是靠專業技能吃飯的，我的願望就是，不僅要讓無法讀書的孩子讀

書，還要讓他們也能靠自己的專業技能謀生，這才是真正的好事。隨後，他們兩個人都開始把出去講課賺的錢捐了出來。

同行的很多人，總是令我覺得自己做得還不夠。今年，我們再次一起去甘孜，去之前我就在籌畫寫一本數據方面的書，籌畫這本書的初衷是，一方面，利用自己這麼多年做數據的經驗和專業知識，把一些方法集合起來，分享給大家；另一方面就是，想籌集更多的錢來幫助更多的孩子吃飽飯，接受教育。而在這本書的籌備中，很多人都在出力。

在去甘孜的路上，上師對我說：「不忘初心，才會收穫善果。」這句話深深植入我心，因為任何一個人善意的初心，一定會換來善果。

我也改變了自己

我的一生並不像別人那樣按部就班地學習、畢業、工作，一路走來我都是跌跌撞撞。我想，如果運氣不好，現在的我很可能在路邊幫人修車，在廚房為人炒菜，甚至加入黑社會。我大學畢業時，連我老爸都哭了，他覺得這個孩子是被救回來的。我曾經有機會好好讀書，卻沒有把握，之後才浪子回頭，我深深知道讀書對於人有多重要，那是進入社會的門票。而現在，我看到這些孩子連讀書的機會都沒有，頓時覺得世界的不公，這大概就是與這群孩子結緣的原因吧。

有人問我，是不是看到一個孩子在自己的幫助下成長，會很有成就感。實際上，我們並沒有特別關注哪個孩子，因為語言不通，與孩子之間交流比較少，只是每年會去看望他們一次。我最大的快樂就是，這些孩子能吃飽，在吃飽的同時還能接受教育。

因為結了這樣的緣，忙與累也變成了一種快樂，比如，這些年我每天堅持4點半起床，把每天重要的電商新聞摘出來發微博，因為我知道，每天只發一條淘寶店義賣的微博，是不會有人看的，只有累積，大

家才會看到你的用心。除此之外，我還養成了長期堅持早上閱讀和打坐的習慣，這也讓我對產業的了解更深，對於我的工作和生活來說都是好事。

正因為如此，上師對我說：「你知道嗎？2006年的品覺是一個很討厭的人，現在卻變化很大。」上師說的沒錯，在這個過程中，我也能感覺自己心態的變化，我開始更少考慮自己，像上師說的，我開始學會把自己放在更低的位置，在這場修行中，我學會了放鬆，而世界似乎也更廣闊了。

這是一個網路的世界，每個人都不應該單打獨鬥，而我也還在修行的路上，一切還沒結束.....

天下財經 359

大數據的關鍵思考（增訂版）

行動×多螢×碎片化時代的商業智慧

作者 / 車品覺

封面設計 / FE設計

責任編輯 / 白詩瑜、劉珈盈

EPUB3檔案製作 / Cultivar播蒔數位出版有限公司

發行人 / 殷允芃

出版部總編輯 / 吳韻儀

出版者 / 天下雜誌股份有限公司

地址 / 台北市104南京東路二段139號11樓

讀者服務 / (02) 2662-0332 傳真 / (02) 2662-6048

email / bill@service.cw.com.tw

天下雜誌GROUP 網址 / <http://www.cw.com.tw>

法律顧問 / 台英國際商務法律事務所．羅明通律師

出版日期 / 2020年3月4日第一版第一次印行

電子版出版日期 / 2020年3月4日

本書由Cheerspublishing 湛庐文化正式授權

Complex Chinese Translation copyright © 2020

by Commonwealth Magazine Co., Ltd.

Published by arrangement through Bardon-Chinese Media Agency

博達著作權代理有限公司 ALL RIGHTS RESERVED

書號: BCCF0395P

eISBN: 978-986-398-530-3

天下網路書店 <http://www.cwbook.com.tw>

天下雜誌我讀網 <http://books.cw.com.tw/>

天下讀者俱樂部 Facebook <http://www.facebook.com/cwbookclub>

天下讀者俱樂部 Instagram <https://www.instagram.com/bookcw1/>