

今天的演講由南山管理資訊分析中心的賴昌作協理以「客戶數據分析與數位應用」為題進行演講。賴協理首先以 NBA 為例切入，例如有了球員表現的相關數據，我們可以建模預測他下次的得分，或是透過球員 Twitter 的情緒（文字等非結構化資料），就可以透過 NLP 去剖析並推測他明天會不會打架罵人。「數據分析跟 AI 取代傳統經驗」，賴協理認為這是為什麼各行各業的總經理都年輕化，因為「經驗就在數據裡」，甚至連 Marketing 也開始用 Data。

然而我認為這樣講其實不夠妥善，數據確實會說話，然而問話者的技巧也相當重要。賴協理提到「觀察思考並運用數據解決業務問題」，可是「有數據」的下一步其實無法直接就到解決問題。即使近年來 Deep Learning 興起，Data 的 Preprocessing 依然是每個專案成敗與否最重要的部分。因此除了數據本身，解構數據者，Domain Knowledge 的供給者通通都同樣重要，我認為不該獨尊技術，而忽略了每一個能將模型 tune 的更完美的領域專家。

接著賴協理與我們分享南山現有的數據專案，目前南山除了客人基本資料這些「量化」資料以外，也有導入網頁點擊等非結構化資料。我認為點擊和客人看到試算結果後的反應確實相當具代表性，正如同我們說的「嘴巴說不要，身體很誠實」，會不會客人表面上對保險興趣缺缺，私底下卻不斷查看和試算？

在聽南山「客戶分群」、「業務和被保人的網路圖」以及「家戶金錢支配」等專案的時候，我心裡有一個疑問，那就是或許是因為南山本質上還是一間金融公司，其似乎很偏好「可解釋性」高的模型，諸如決策樹或是隨機森林，且協理對於技術的掌握度也不盡理想。

在「客戶分群」的專案中南山將客戶分為「貴客、稀客、常客、普客」等等，然而這個概念其實是 Classify 而非 Clustering。數據很重要，然而如我前述所說，運用數據的人同樣甚至更為重要，Cluster 分出的是電腦「客觀」上認為最相近的觀察值，Classify 則受資料標記者的「主觀」影響。這當然是個很難的問題，到底是可解釋性重要，還是準確性重要？以我自己的觀點來說，我認為專業人員的存在，就是成為專業技術和一般人員之間的橋樑，而不是偏向一方，太過追求或排斥理性上的完美。

而賴協理提到的人物關係網路圖、家戶金錢支配圖，我認為這兩者則只停留在「數據探索」，如計算、統計的階段，其中並沒有任何「機器學習」的概念存在。賴協理甚至提到說南山現在利用「Deep Learning」的技術將使用者在不同裝置上的資料進行歸戶，但顯然資料比對或是 Cookie 都是更為合理的作法。資

管所的盧信銘老師曾在課堂上提到機器學習是「No Free Lunch」的，意即沒有一個方法能永遠最好的解決問題。不只南山，我認為每一間金融公司在接納數位科技的時候，最重要的都是要理解到「因地制宜」的概念，而不只是盲目的追求最新最炫的名詞，終究只有真金才通過的了火煉，而且很重要的一點不是公司有沒有 XGBoost，有沒有深度學習，而是數據科學「解決」了什麼問題。

因此在聽完今天的演講後，我認為最重要的不是賴協理提到什麼可以解決什麼，反而是最一開始的那個問題：「今天我面對了什麼問題？我要怎麼透過數據或是其他科技來解決？」，這是我們必須持續修行，持續挑戰和嘗試的功課。還有一點就是：「除了數據，哪些專業人員的參與可以讓這件事變得更好？」

最後演講中特別讓我有印象的是賴協理有提到我上次心得中討論的「保險差別定價」，協理說數據確實可以協助保險公司對客戶進行差別定價，但抵處了保險集眾人之力的精神。在執行很多個數據專案的經驗中，我也知道 ROC CURVE 有的時候就是會卡在 0.75, 0.8 無論怎麼修改就是上不去，想要一次就估算出要怎樣才能把比較不健康的人無差別待遇的納入保障之中是很難的。從信用卡到電子支付的時代，金融業也是一次又一次的突破關卡，因此對於像我一樣健康弱勢的人來說，第一個問題應該是：「我可不可以有選擇把自己納入保護網的權利？」這也又回到了今天演講後我思考的最深的一個問題：「我們的目標是什麼？我們又要如何達到？」是公平重要，還是保障重要呢？