2018-12-17

**陳如月**

**機器學習：商業規劃的下一個新領域**

如果你是一名藥商主管，很可能你已經投入了大量的時間和精力探索機器學習的可能性。無論你是支持者還是懷疑論者，圍繞機器學習的大膽主張和不停的嗡嗡聲都會吸引你的注意力並挑戰你的想法，要求你評估其潛力並對其適合你的業務採取立場。

機器學習代表數據進階分析的巨大飛躍。整合現實世界數據來源並建立如此深入、細微和精確性的市場、病人和醫療服務提供者的動態圖像的能力，開啟未開發的模式和見解的巨大金礦，可以提供信息(並有可能改變)戰略，並推動業務成長。通過創建一個動態的、多方位的景觀，機器學習可以讓數據自己說話、揭示真實的故事，而沒有人為的偏見。

**機器學習如何改變商業規劃**

簡單地說，機器學習使團隊能夠同時查看所有數據，展開一種演算法來揭示哪些是最重要的數據模式。它超越了線性分析，提供市場的動態理解，有助於一些規劃活動，如市場規模、病人區隔、目標、醫療服務提供者區隔、支付方區隔和訊息傳遞，以及健康經濟與效果研究（Health Economics and Outcomes Research，簡稱HEOR）活動。

因為病人是以一種整體的、多方面的方式被描述的，所以它使我們能夠瞭解他們有哪些其他診斷，以及他們正在使用哪些醫療服務、他們正在接受哪些其他治療，以及數據來源中的任何其他特徵。透過建立已知病人的檔案，我們可以將這些資料應用到一般人群中，以找到未確診的病人。這種方法在未確診的人群中尤其有效。

一個例子是非酒精性脂肪性肝炎(non-alcoholic steatohepatitis，簡稱NASH)市場，其中包括大部分未意識到的和無症狀的病人群體。在最近的一個專案中，Decision Resources Group專家部署了機器學習技術，目的是將NASH人群擴大到被確診的病人之外，並建立有意義的子市場區隔。這需要分析演變成NASH的情況，以確定潛在病人的關鍵特徵，匹配電子健康記錄(Electronic health record，簡稱EHR)和求償記錄，以建立已知病人的概況，並進行機器學習建模，以根據關鍵特徵標記未確診病人（如，肝指數異常功能檢查升高、共病）。

分析發現，至少有25%的非酒精性脂肪肝疾病(nonalcoholic fatty liver disease，簡稱NAFLD) 病人已演變成NASH，突顯出該類別內診斷不足的真實程度。機器學習使客戶端能夠建立四個臨床上不同的NASH區隔市場以及三個混合區隔市場。它還發現了一個特別有趣的洞察力，那就是，多達10%的NASH病人沒有表現出典型的疾病特徵。在這種情況下，在使用的案例適當的情況下，機器學習可以成為深入理解和支援病人群體的寶貴工具。

但它並不一定是機器學習和傳統方法之間的較量。在某些情況下，運行舊式的分析是非常好的，因為商業問題可能不需要先進的方法。事實上，在某些情況下，機器學習可能會使分析過於複雜。與經驗豐富的數據科學夥伴合作，一起確定案例的適當性非常重要。

**機器學習可以回答的具體商務問題的例子**

|  |  |  |
| --- | --- | --- |
| **活動** | **說明商業問題** | **機器學習如何幫助提供** |
| 估算市場規模 | 我的未確診病人人數有多少？ | –從我們已知的病人建立模型–驗證並建立優化的模型–對病人進行整體評分，以識別出患有該疾病的高機率病人 |
| 病人區隔 | 如何在飽和的市場中建立流行疾病的病人市場區隔？ | –利用EHR/求償資料建立一種演算法，該演算法可識別更廣泛的病人群體中自然發生的病人特徵–確定不同類型的臨床鑑別指標–應用於較大的疾病人群，以估計更廣泛人群中的病人群體規模 |
| 定標 | 在確診之前，我如何定標那些很有可能患有患有特定罕見疾病的病人呢？ | –從我們已知的病人建立模型–驗證並建立優化的模型–對病人進行整體評分，以識別出患有該疾病的高機率病人–重新對病人進行整體評分，以提醒這一群體的診斷前 "觸發因素"–識別並通知病人的對應的醫療服務提供者 |
| 健康經濟與效果研究（HEOR） | 如何利用EHR數據中病人報告的結果構建成本效果的高效的HEOR預測模型？ | –整合求償/ EHR數據，建立一個探索性模型使我們能夠識別區分治療成功或成本高的病人的臨床特徵–使用折半樣本驗證模型–應用這些無偏見的特性來優化預測模型的性能–總結結果，公佈給更廣泛的醫療社群 |
| 醫療服務提供者的區隔/教育 | 如何在不依靠醫師完成的調查偏差的情況下，確定哪些信息對於吸引醫療專業人員和驅動處方行為最有效？ | –綜合EHR、求償和溝通資料，建立一個探索性模型，可以識別與醫師的經驗相關的特徵模式和針對不同類型的品牌資訊採取的行動過程–確定醫療服務提供者之間的差異指標–使用不同類型的訊息傳遞驗證和建立優化模型–識別品牌資訊的差距和機會，並根據每個市場區隔的既定特徵對其進行優化–繼續標竿瞄準模型性能，並優化醫療服務提供者的設定檔和區隔 |
| 產品銷售分析 | 如何找出上市產品表現不佳的原因？ | –建立一個集醫療和藥品求償資料於一體的模型–確定影響銷售的因素 (例如，醫療服務提供者是否有處方、保險公司是否核准、病人是否領取處方的藥品、或者他們只嘗試過一次？)–應用這些特徵識別消除銷售障礙的機會，以及定標醫療服務提供者提供品牌訊息的機會 |

* **怎樣才能展開機器學習？什麼團隊需要？**
1. **明確的目標、使用案例和產出要求**

機器學習需要人的指導。知道你想要實現的目標、想解决什麼樣的難題，以及你是如何定義病人的，這一點至關重要。因此，你需要確定你的商業問題、了解如何處理這個專案、定義什麼將構成成功，並思考你將如何實行你的發現。

1. **願意採納這種方法**

機器學習為商業規劃活動提供了一種新的思維方式，並與藥業的傳統智慧背道而馳。它需要你以開放的心態接近它。你還必須願意克服人的偏見 - 從分析中排除猜測並不能降低你的角色價值，但它確實會增加你洞察力的價值。

機器學習將有益於各種使用情況，只要以適當的方式應用適當的演算法。但並不是每一種情況都有意義：例如，你不可能把它應用於發現未確診的常見疾病病人，比如糖尿病。機器學習不是萬靈丹。它仍然需要人類的專業知識來產生見解、應用分析和制定實施戰略。成功的專案可以利用傳統分析和機器學習的組合。

1. **數據和數據科學的強大夥伴**

機器學習要求你能夠獲得出色的數據和出色的數據專業知識。如果你自己購買數據，那麼你就有可能最終得到不適合你想吸引的病人類型的模式集合。而如果你無法獲得完整、整體性的數據，你就無法找到未確診的病人。

一個好的數據科學夥伴將引導你完成專案的每一步，並根據你的目標和案例使用情況提出建議。他們應該能夠顯示出對整個醫療保健部門的深刻和整體了解，在你的特定治療類別和市場方面有很強的專業知識，以及使用最新的數據科學技術進行成功專案的豐富經驗。

他們還必須具有職業誠信，根據你的具體需求而不是他們的需求提出建議。有時候，正確的建議是你的專案根本不需要機器學習。不幸的是，一些藥商過去曾因與不符合這些標準的夥伴合作而受到傷害。

* **採用機器學習的各個階段**
1. **開發**

在初始階段，數據科學合作夥伴將與你的團隊合作，確定他們將設計和構建初始模型的規格。

每個病人群體的數據看起來都不一樣，因此合作夥伴將進行探索性分析，以深入理解其與目標人群相關的特徵。其中一些人群將與醫療保健系統進行大量互動，即使在罕見疾病中也是如此，但其他的則並非如此。合作夥伴將提出建議並幫助你確定具體的業務問題，應用治療領域的專業知識，以建立識別病人群的最佳方法。然後，合作夥伴將建立初始機器學習模型。

1. **驗證**

 在此階段，將使用真實世界的數據對模型進行測試和改進。從本質上說，它將被測試是否有能力識別前一年已卻診的病人。模型是對那些已知病人的選擇過高還是不足？然後應用業務邏輯和治療專業知識幫助確定哪些業務規則和篩選器將改進你的模型並優化其性能。

1. **實作**

 一旦驗證階段完成，並且演算法已被優化，你的模型就可以上線。根據你向合作夥伴請求的輸出週期 (例如，每週、每月等) ，你將開始收到一份報告或儀表板。你可以開始使用這些見解來制定商業策略，然後透過現場銷售活動和訊息傳遞實作。

1. **重新校準**

 必須不斷地重新調整機器學習模型，將變動納入醫療保健領域。比如說，你正在一個罕見疾病的市場進行一項研究，這個市場最初有1000名已知的病人。然而，在第一年，你已經確定目前正在治療另外的250名病人。可以將這些額外病人的經驗和特徵結合到模型中，以進一步優化其性能。本質上，隨着市場的發展，你會繼續改變這個市場的樣子。如果你不繼續校準模型，你就會錯失機會。

※　　　※　　　※

產品的生命週期越來越短。因此，找到好的數據科學合作夥伴，正確使用機器學習，把握契機，務必使產品ㄧ上市就成功。

（取材自PharmExec.com）

–End–