

講題

小資料科學家的生存之道

摘要

雖然現在是大數據，人工智慧跟深度學習的時代，但身為資料科學家，實務上還是有很多機會要處理小資料。有時候客戶以為的大數據，在我們眼中是小資料。常常客戶抱著大數據來，卻不知道他的數據有什麼用？能解決他什麼痛點？針對這些大數據，我們會先了解客戶的痛點，找出問題，訂定評估方法的方式(evaluation)。接著我們可能需要說服客戶標記跟問題相關的標準答案，標記需要金錢，人力，時間，在現實的考量下，我們拿到手的也會是小資料。小資料有時候更適合跑簡單的機器學習模型，啟發式規則或正則表示法。有時候客戶也會要求解決方案的可解釋性，那可能就要跑決策樹之類的模型。在我的工作經驗裡，溝通能力也是很重要的，介紹自己的工作內容，什麼可以做，什麼有困難。跟客戶，老闆，合作團隊溝通。最後也想跟大家**分享自然語言處理最簡單又最困難的點。**

講者

Cicilia Lee

哈囉大家好，我是資料科學家 Cicilia Lee，台大資工所畢業，主修自然語言處理(Natural Language Processing)跟文字探勘(Text Mining)。畢業後先當了軟體工程師，工作上一直有接觸機器學習跟資料探勘，後來自然而然就成為了資料科學家。待過大公司也待過新創公司，做過的專案有輿情探勘(Opinion Mining)跟智慧型語音助理(Intelligent Agent)。畢業前最紅的機器學習演算法是支援向量機(Support Vector Machine)，畢業後才遇到深度學習跟BERT的浪潮，被沖得東倒西歪之餘，發現其他的機器學習演算法跟自然語言處理知識應用在小資料也有不錯的表現。