



食物辨識 預測你的食譜

洪儷娟 文

「欸！這個好好吃噢！但是看起來也不會太難做？」很多人可能都聽過類似的對話。在日益講求健康的社會，我們會想知道攤販、夜市和餐廳等等食物的營養成分，甚至對於眼前的食物也有意願自己動手完成。台灣人喜歡在飯前留下食物影像的習慣，多數是為了發文紀錄。但是現在，藉由新的技術讓我們在拍照的同時，獲得更多的食物資訊，增添另一層健康意義，告訴我們吃了什麼以及推薦我們如何製作，這值得我們去了解最新的食物辨識系統如何發展。

深度學習淺知識

在進入食物辨識的討論前，先簡單了解一下什麼是人工智慧和深度學習。根劇〈科技新報〉解析，機器學習屬於人工智慧的一部份，機器學習是將人在腦中判斷的過程運用到機器身上，讓機器具備人的學習能力，其中包含訓練和預測，而機器以四個步驟進行判斷：獲取數據、分析數據、建立模型、預測未來。

深度學習 (Deep Learning) 則為一種機器學習方式，是利用深度神經網路讓電腦可以自動分析資料找出「特徵值」，最重要的地方在神經網路可以自我學習，就像人類的神經一樣可以重複來回的判斷資訊。人類複雜的神經系統使我們在判別資訊時的錯誤率可以大為下降，所以機器的深度學習技術也需要透過多層神經網路來識別資訊。但是神經網路層數過多，機器也會犯暈，導致效果不佳，因此，編碼器的使用成為神經網路判讀的重要輔助角色，讓神經網路傳遞不打結。

食物辨識首例 食物圖像變食譜

麻省理工學院 (MIT) 卡塔爾計算研究所 (QCRI) 所發表的《[Learning Cross-modal Embeddings for Cooking Recipes and Food Images](#)》論文中，研發的人工智慧系統Pic2Recipe是目前最新的食物辨識系統。Pic2Recipe就字面上意思來看就是Picture to recipe，透過分析食物圖像來預測你的食物成分，進而推薦相似的食譜。研究人員建置了一個網羅二十多個主流食譜網站的資料庫——Recipe1M，整合超過100萬個食譜以及80萬張食物圖像。Pic2Recipe系統以此做為依據，訓練神經網路來尋找食物圖像和食譜的關聯，利用這些圖像和文字的比對，培訓神經網路自動理解食物。

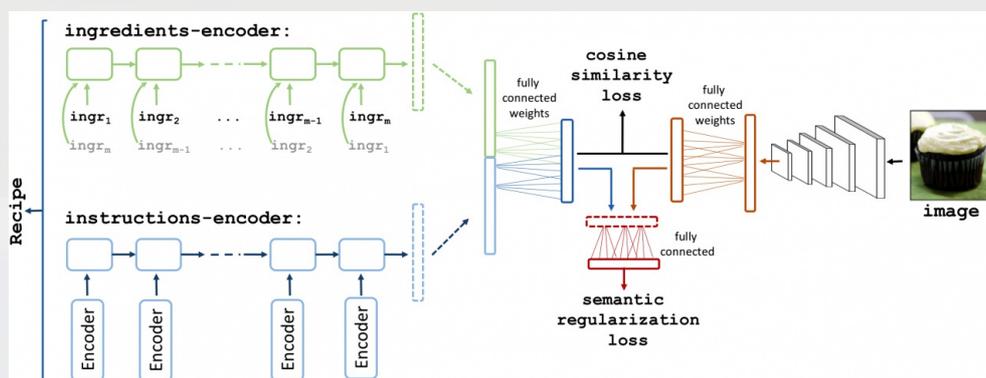
Pic2Recipe系統辨識食物的介紹影片 (影片來源 / [MITCSAIL](#))

Pic2Recipe的最新技術建構

食譜數據的收集主要是將線上食譜轉成原始碼 (HTML檔) 以獲得相關的文字，不符合研究人員需要的空白鍵和其他字符會一併屏除，之後再下載內文的圖像，所以數據結構內容分為兩層，第一層是基本訊息，包括標題、成分列表和步驟說明，第二層是建立在第一層之下的圖像。

因為食譜的主要內容包含材料和步驟說明，Pic2Recipe系統給予個別材料一個代表編碼，而這個代表編碼是為了讓神經網路學習並作為判斷對應材料的基準。為了這樣做，實際材料名稱是從每個食譜的材料中提取文字。例如「2湯匙橄欖油」中，「橄欖油」被提取作為成分名稱，並作為一個單位處理。換句話說，就是給這項材料一個明確且讓神經網路看得懂的名字。步驟說明的部分因為敘述較長，所以需要比較複雜的編碼過程去把逐項步驟表示給神經網路看。如此一來，神經網路就會建構材料和步驟的文字訊息識別，創造一個語義規則化的空間。

至於圖像的部分，主要是採用卷積神經網路來識別食物。人想要判別一張圖像裡呈現的3D物體是容易的，但是對機器來說它只看得懂2D平面圖像。最新的殘差神經網路已經能將3D物體轉換成2D平面圖像給機器判讀，而在此使用的卷積神經網路則能預測3D圖像的聚焦對象，也就能使Pic2Recipe系統辨識使用者拍攝的食物主體。



整個系統是由三種形式 (材料、步驟說明、圖像) 的編碼器組成的聯合嵌入器。(圖片來源 / [CSAIL-QCRI論文網站](#))

工研院也有食物辨識技術

MIT開發多模式神經網絡來共同學習嵌入圖像和食譜的模型為最新食物辨識的指標，工研院也不遑多讓，在判讀食物影像辨識上也有相當的成就。「享食添糖」是由工研院研發來紀錄日常飲食的APP，同樣是利用深度學習的方法自動分析食物影像的技術，目標是找出食物中的成分，並透過食材營養資料庫評估其營養值，提供使用者食物的營養含量。享食添糖和Pic2Recipe最大的差異在於，前者的資料庫為食物營養值，後者的資料庫為食譜。

資料庫的選擇解釋了為何Pic2Recipe能成為食物辨識的最新指標。因為食物營養值的資料很容易標示和區別，在機器的深度學習判斷上不太會有太大的誤差，工研院也顯示其判別率可達80%。然而MIT選擇了一個不好控制的資料庫，因為食譜範圍過大，數據難以管控，判別的精確性會降低。尤其要針對各式食譜的文字進行編碼和操作，形成非常龐大的工程，這就是為什麼以往較少看到將深度學習應用在語義的領域上。因此，Pic2Recipe能初步整合大範圍的語義資料實為深度學習在食物辨識上的先驅。

目前技術的限制及未來推進方向

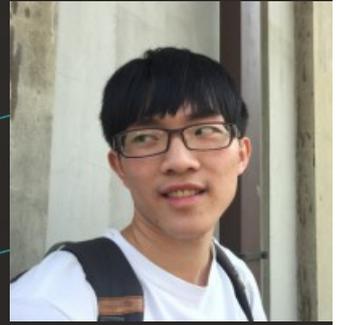
Pic2Recipe雖然在食物辨識技術中有新的突破，資料庫數據量也有一定的規模，但仍然受限於種類不完整，導致在圖像配對上準確度不足。使用者上傳圖像的品質不一也是一大問題，會嚴重造成機器判斷的困難。再者，同樣食材以相似做法完成的食譜也會讓機器無法正確判讀。此外，Recipe1M資料庫內多為烘焙類食譜，這同時是Pic2Recipe的食物研究主軸，所以進行烘焙食品的圖像判別會較為準確，其餘食品的錯誤率較高，而目前整體的識別率為65%。

現在的系統可以分析你手中食物圖像的內容物及推薦相應的食譜，試著反過來想，以後我們的冰箱內有哪些食材，就推薦我們使用煎、炒、煮、炸各種方法製作的食譜，將可能改變我們的飲食習慣。食物辨識系統從記錄的角色漸漸變為引導的角色，期待日後食物辨識能持續進步，輕易地就能為每個人找到適合的食譜，使食譜系統成為管理我們健康的助手，這不管對老人、小孩、病患，或是想控制體重的人，都會是一大福音。





記者 洪儷娟



編輯 鄭仰珉