

一袋水果怎麼排隊

本課講 2015 年的一篇文章。Vinyals、Bengio 和 Kudlur 三位科學家發現：同一組東西，換個順序交給電腦，結果就會不一樣。文章的名字就叫《順序很重要》。

適合 10 歲左右

約 40 分鐘

需要：鉛筆、12 張小卡片、2 位同伴

1 熱身：三個水果排幾種隊

書包裡有 3 個水果：蘋果、香蕉、橙子。把它們從左到右排成一排，一共有幾種排法？前兩種已經寫好了，請寫出其餘幾種。

① 蘋果、香蕉、橙子

② 蘋果、橙子、香蕉

③ _____

④ _____

⑤ _____

⑥ _____

3 個水果有 6 種排法。再加 1 個梨，就有 24 種。10 個水果呢？超過 360 萬種！

可是不管怎麼排，書包裡的東西都一樣。「一袋水果」本來就沒有先後。

2 有的東西有順序，有的沒有

「我愛媽媽」和「媽媽愛我」，字一樣，意思不同。像句子這樣，先後一換意思就變的一串東西，叫**序列**。一袋水果、抽獎箱裡的號碼球，先後換了也沒關係。像這樣只管「有哪些」、不管「誰先誰後」的一組東西，叫**集合**。

東西	序列	集合	東西	序列	集合
一首歌的歌詞			今天來了哪些同學		
做蛋糕的步驟			跳棋盤上剩哪些棋子		

請在表裡打 ✓。麻煩在這裡：第 1 課講過帶「小本子」的電腦，它**一個一個地讀**，讀一個記一筆。要把一袋水果交給它，就得先排成一隊。這種一個一個讀、一個一個寫的程序，叫「序列到序列」模型。

3 換個順序，結果就變了

按道理說，水果怎麼排都一樣。科學家發現，**電腦其實很在乎**。

一個真實的例子：倒著讀，翻譯得更好

2014 年，有科學家教電腦把英語翻譯成法語。他們把英語句子**倒過來**交給電腦讀，電腦反而翻譯得明顯更好。句子沒變，只是換了讀的順序。這篇文章收集了這類發現，問了 2 個問題：順序到底重不重要？如果重要，該怎麼辦？

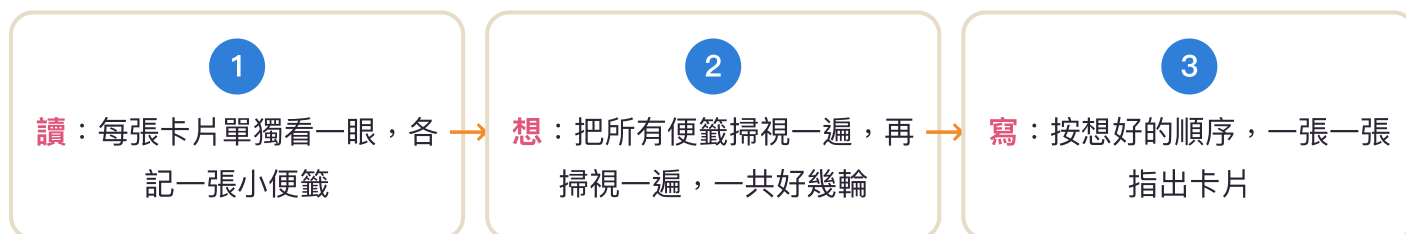
道理上一樣，做起來不一樣

第 2 課說過，小本子只有一頁，早讀的東西會慢慢變淡。所以先讀誰、後讀誰，本子上的筆記就不一樣。

就像背購物清單，按超市貨架的順序背最好記。每種順序在道理上都一樣對，**可是有的順序電腦學得快，有的學得慢**。

4 辦法一：把卡片攤開，多看幾眼

排 6 張數字卡片時，你會**全部攤在桌上，來回看幾遍**，再動手。科學家給電腦的辦法也分 3 步：



第 2 步最關鍵。每一輪掃視，電腦都會決定哪幾張多看兩眼。這種「把眼光多放在有用的地方」的本領，叫**注意力**，第 13 課和第 15 課講過它。



示意：每輪都看全部卡片，橙框是多看兩眼的

為什麼這樣就不怕順序了？像榨果汁

每一輪，電腦把看到的東西**按輕重混在一起**，變成一個想法。這就像榨果汁。先放蘋果還是先放香蕉，榨出來的果汁一樣。卡片怎麼擺，混出來的想法也一樣。掃視幾輪可以自己定，輪數越多，電腦動手前「想」得越久。

5 例子：給打亂的數字排隊

科學家用排序來考電腦。電腦拿到一組打亂的小數，都在 0 和 1 之間。它要從小到大，一個一個**指出**它們在哪兒。比如上圖 4 張卡片依次是 0.73、0.21、0.95、0.40（示意），電腦就回答：「第 2 張、第 4 張、第 1 張、第 3 張。」這種「指一指」的寫法叫**指針網絡**，第 18 課會專門講。

科學家試了 5 個數、10 個數和 15 個數，發現兩件事：

1. **掃視的輪數越多，排得越準**。先多想幾輪再動手，比看一眼就動手強。
2. **數越多，越難排準**。這和你自己排卡片的感覺一樣。

6 辦法二：寫的時候，也挑個好順序

順序的麻煩不只在「讀」，也在「寫」。有時候，答案本身就是一個集合。

老師的答案順序不一樣，電腦就犯糊塗（示意）

題目：「袋子裡有哪些水果？」答案是蘋果和香蕉。第 1 次，老師寫「蘋果、香蕉」。第 2 次，另一位老師寫「香蕉、蘋果」。兩個答案都對。可是電腦一個字一個字地學寫，它會想：「第一個該寫蘋果，還是香蕉？」它搖來擺去，學得很慢。

文章裡還有一個真實的例子：把一句話拆成大塊和小塊，寫成「結構圖」。同一張圖按不同的先後寫，電腦學得好不好也不同。

他們的辦法是：**練習的時候，讓電腦自己挑一個最好寫的順序**。每道題都先看幾種寫法，電腦現在覺得哪種最順手，就按哪種練。剛開始練習時，要故意多試幾種順序，免得太早認定一種壞寫法。作者在一個人造的練習題上試過，這個辦法確實能幫電腦找到好寫的順序。

7 這和今天有什麼關係

第 13 課講的 Transformer，是今天聊天 AI 的大腦結構。不告訴它字的位置，它就像看一桌攤開的卡片，分不出先後。第 21 課的程序就用了本文「多看幾輪」的辦法。

今天的聊天 AI 也會被順序影響。2021 年有研究發現，只換幾道例題的先後，AI 答對的比例就可能變化很大。還有研究發現，AI 記住了「小明的媽媽是王老師」，卻不一定答得出「王老師的孩子是誰」。讓 AI 「先寫想法、再寫答案」常常更準，這也是在挑一個更好寫的順序。

8 動手：排隊聽 vs 攤開看

玩法（2 個人，約 15 分鐘，需要 12 張小卡片）

A 出題，B 當「電腦」。每輪 B 都要把 6 個數從小到大寫出來。3 組數只給 A 看：

第 1 組 47、12、85、30、66、9 第 2 組 58、23、91、14、70、36 第 3 組 81、5、44、63、27、99

1. **排隊聽**：A 每秒念 1 個第 1 組的數，只念一遍。唸完 B 才動筆。
2. **攤開看 1 眼**：第 2 組寫成卡片，打亂攤開。B 看 5 秒，翻過去再動筆。
3. **攤開看 3 眼**：用第 3 組數做卡片。B 看 5 秒，卡片翻過去 5 秒，重複 3 次後再動筆。
4. 數一數 B 排對了幾個位置，填進表裡。交換角色時，新出題人自己編 3 組數。

第幾輪	排隊聽	攤開看 1 眼	攤開看 3 眼
B 排對幾個（共 6）			
交換後排對幾個			

你會發現：排隊聽時，前面的數容易忘。卡片攤開時，可以來回比較。多看幾眼，通常排得更準，就像文章裡「多想幾輪」的辦法。

9 動手：哪個順序最好記

玩法（3 個人，全班更好；約 15 分鐘）

1. 3 個人各拿一行，詞一樣，先後不同。**不要偷看別人那一行。**
2. 默讀 30 秒後蓋住，再用 1 分鐘寫下記得的詞，順序不限。
3. 數一數記住幾個，填進表裡。全班玩時分 3 組，算每組的平均數。

順序	12 個詞	記住幾個
甲：打亂	跳繩、餅乾、尺子、游泳、糖果、書包、足球、薯片、橡皮、跑步、酸奶、鉛筆	
乙：按類排	鉛筆、橡皮、尺子、書包 餅乾、薯片、酸奶、糖果 足球、跳繩、游泳、跑步	
丙：編故事	書包、鉛筆、足球、餅乾、橡皮、跳繩、酸奶、尺子、游泳、薯片、跑步、糖果（背上書包，帶好鉛筆，去踢足球……）	

你會發現：按類排或編成故事，通常比打亂的好記。電腦練習時挑「最好寫的順序」，就像你挑最好記的順序。

10 詞語小卡片

集合

只管「有哪些」的一組東西，比如一袋水果。

序列

先後一換意思就變的一串東西，比如一句話。

讀—想—寫

單獨看每張卡片，掃視幾輪，再指出答案。

挑順序

練習時讓電腦自己挑最好寫的答案順序。

11 想一想

1. 「全班同學的名單」是集合還是序列？「按身高排的隊」呢？為什麼？

2. 生活中什麼事先後換了結果不變？什麼事一換就不行？各舉一個例子。

3. 背單詞、背課文時，你會特意排一個順序嗎？什麼樣的順序對你最好記？

一句話記住：東西沒順序，電腦卻在乎；攤開多看幾眼，挑好寫的順序。

給家長和老師：本課講的是 Oriol Vinyals、Samy Bengio、Manjunath Kudlur 2015 年的論文 [Order Matters: Sequence to sequence for sets](#) (ICLR 2016)。論文指出 seq2seq 對輸入和輸出的排列都敏感，證據包括翻譯中倒序輸入更好 (Sutskever 等 2014) 和句法樹線性化方式的影響。針對集合輸入，作者提出讀—處理—寫結構：逐元素編碼，用無外部輸入的 LSTM 做 T 步內容注意力（後稱 Set2Set），再用指針網絡輸出；在 5、10、15 個數的排序任務上，處理步數越多準確率越高。針對集合輸出，作者提出訓練中在候選順序裡選當前概率最高者計算損失，並在前期採樣探索，在人造的圖模型聯合分佈估計任務上做了驗證。

「小本子」「小便籤」「掃視」「榨果汁」分別對應 RNN 狀態、逐元素編碼、多步注意力讀取和加權求和的排列不變性。0.73 等 4 個數、三輪掃視圖、水果答案、購物清單、活動的數字和詞表都是示意；「小明的媽媽」是對「反轉詛咒」研究的通俗改寫。講義未引用論文的具體準確率。想深入可看本站的 [第 16 天中文導讀](#)，代碼見開源項目 [30u30 第 16 天目錄](#)（英文）。