

第 2 課

記憶傳送帶和三扇門

本課講 2015 年的一篇文章。Christopher Olah 用一組圖，講清了一種會「挑著記」的電腦程序。它叫 LSTM。

適合 10 歲左右

約 40 分鐘

需要：鉛筆、便利貼、幾位同伴

1 熱身：誰還記得開頭？

先填一個簡單的：天上飄著白 _____。再讀一個小故事，填最後的空。讀的時候不要往回看。

奶奶家養了一隻白貓，叫雪球。暑假裡，樂樂在奶奶家住了一個多月。他每天早上跑步，下午寫作業，晚上和表哥下象棋。週末，姑姑帶他們去海邊撿貝殼。回家那天，樂樂抱了抱奶奶，又抱了抱 _____。

第一題只看前面幾個字就夠了。第二題的線索藏在**最開頭**，中間隔了好幾句話。這種「線索離得很遠」的題，對上節課的循環神經網絡來說特別難。

2 小本子太小了

上節課說過，循環神經網絡帶著一個「小本子」，每讀一個字就記一筆。問題是，這個小本子只有一頁。每讀一個字，它都要**整頁重寫**一遍。新內容寫進去，舊內容就被擠掉一點。讀了幾十個字，開頭的線索就快看不清了。

小本子裡的「雪球」，越讀越淡（示意）：

讀完第 1 句

雪球

讀完第 3 句

雪球

讀完第 5 句

雪球

讀到最後

雪球

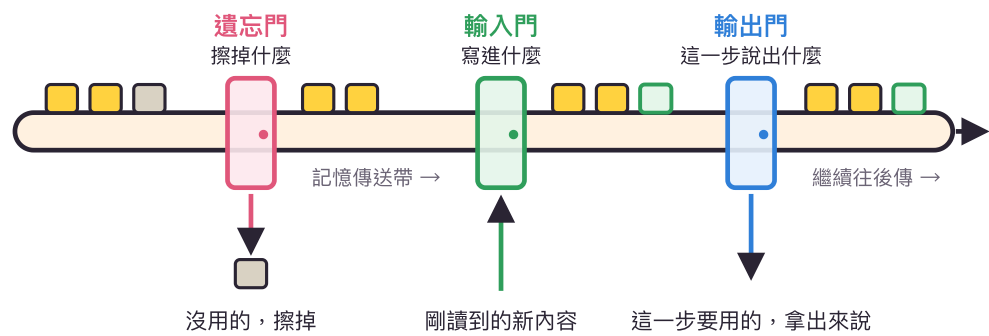
還有一個原因藏在「訓練」裡。「你猜錯了」這個提醒，要從後往前一步一步傳回去。每往前傳一步，提醒就變弱一點。傳了幾十步，就幾乎沒有了。所以，按道理說它能記住遠處的事，實際上卻很難學會。科學家在 1990 年代就發現了這個毛病。

3 LSTM：多了一條記憶傳送帶

1997 年，兩位科學家 Hochreiter 和 Schmidhuber 發明了一種升級版的循環神經網絡。它叫**長短期記憶網絡**，英文簡稱 **LSTM**。2015 年，Olah 畫了一組圖，一步一步把它講清楚。很多人學 LSTM，都從

這篇文章開始。

LSTM 也有上節課的小本子，但它還多了一條**記憶傳送帶**。科學家管它叫「細胞狀態」，意思就是長期記憶。



傳送帶從故事開頭通到結尾。上面的東西基本**原樣往後傳**。每讀一個字，它只做兩件小事：擦掉一點，加上一點。不用整頁重寫，開頭的線索就能傳得很遠。訓練時，「你猜錯了」的提醒也能順著傳送帶往回走，不容易變弱。

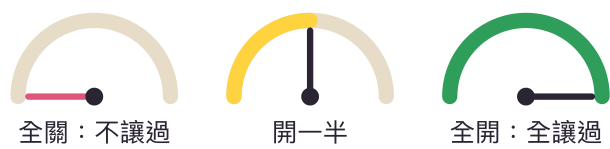
4 三扇門：擦掉、寫進、說出

傳送帶旁邊站著三扇**門**。門就是一個開關，管著有多少東西能通過。這就像開學整理書包：

哪扇門	它決定什麼	整理書包時
遺忘門	舊記憶哪些擦掉、哪些留下	新學期，拿出上學期的舊作業
輸入門	新內容哪些寫上傳送帶	放進新發的課本
輸出門	這一步說出哪些記憶	數學課只拿數學書，別的書留在書包裡

5 門可以只開一半

這三扇門不是隻有「開」和「關」兩種樣子。它們更像水龍頭，可以擰到任何大小。**門開多大，就讓多少東西通過。**



比如遺忘門開一半，舊記憶就留下一半。它沒有消失，只是變淡了。就像練習本還剩一半空頁，你撕掉寫過的頁，留著空頁繼續用。

誰來擰水龍頭？沒有人幫 LSTM 擰。每讀一個字，它都根據剛讀到的字和上一步的想法，決定三扇門各開多大。怎麼決定，是它在訓練中**自己學會**的。

6 看 LSTM 一句一句讀故事

LSTM 在玩「猜下一個字」，要猜空裡填「他」還是「她」。看看三扇門每一句做了什麼（示意）：

讀到的句子	遺忘門擦掉	輸入門寫進	輸出門說出
小明是二年級的男孩。	沒什麼要擦	主角：小明，男孩	不用說
____ 最喜歡畫恐龍。	沒什麼要擦	主角愛畫恐龍	空裡填「他」
新同學小美是個女孩。	舊主角小明	主角：小美，女孩	不用說
____ 帶來了一盒新蠟筆。	沒什麼要擦	主角帶了蠟筆	空裡填「她」

7 這和今天有什麼關係

2013 到 2015 年前後，聽寫語音、翻譯的程序很多都用 LSTM，上節課 Karpathy 的程序也用它。後來 LSTM 被換下來了，因為它只能一個字一個字按順序讀，練起來慢。今天的 ChatGPT、Claude 這些聊天 AI，用的是 2017 年出現的新方法 Transformer。

可是 LSTM 的好點子還在。今天的大 AI 裡也有一條「直通的路」，每一層只加一點、改一點。這和記憶傳送帶是同一個想法。Olah 在文章結尾還猜：下一個大方向是「注意力」，就是每一步挑出最該看的信息。2 年後 Transformer 正是靠它。

8 動手：傳話鏈 vs 記事本

玩法（5 到 8 個人站成一排，約 10 分鐘）

第 1 輪 傳話鏈：第 1 人看紙條：「週五下午 3 點，到 3 樓美術室，帶紅色水彩筆。」悄悄傳給下一人。每人傳話時，末尾再加一樣要帶的東西，比如「還要帶蘋果」。最後一人大聲說出來。

第 2 輪 記事本：換一句：「週三上午 10 點，到操場東門，穿藍色運動鞋。」寫在紙上往後傳。每人不用重說，只在紙上添一樣東西。最後一人大聲讀出來。

輪次	原話的 4 個要點	還剩幾個	最後說的關鍵詞
傳話鏈	週五、3 點、美術室、水彩筆		
記事本	週三、10 點、東門、運動鞋		

你會發現：傳話鏈每過一個人，整句話都要重說一遍，開頭很容易丟。這像普通循環網絡的小本子。記事本上，原話一直待在紙上，每人只加一點。這就像 LSTM 的記憶傳送帶。

9 動手：三扇門便利貼遊戲

玩法（2 到 4 個人，約 15 分鐘）

1. 一個人當「讀故事的人」，其他人當 LSTM。LSTM 面前放 3 張空便利貼，這就是記憶傳送帶。只能用這 3 張。
2. 讀故事的人每次只讀一句。LSTM 不能看下面的故事，也不能回頭問。
3. 每聽完一句，LSTM 做三件事：**擦掉**沒用的便利貼（拿不準的，可以只把字擦淡）；把要記的事**寫進**空便利貼，一張只寫幾個字；被提問時，只看便利貼**說出**答案。
4. 每一句做完，在記錄表裡寫一筆。

① 週一早上，樂樂帶著一把藍色雨傘去上學。② 路上下起了大雨，同學小林沒有帶傘。③ 樂樂把藍傘借給了小林。④ 中午出了太陽，大家去操場踢足球。⑤ 小林踢進了 2 個球，全班都在歡呼。⑥ 放學時，天又下雨了。

提問：樂樂的藍傘現在在誰手裡？小林踢進了幾個球？

第幾句	擦掉了什麼	寫進了什麼	說出了什麼
①			
②			
③			
④			
⑤			
⑥ 和提問			

你會發現：只有 3 張便利貼，你必須挑著記。讀到第 ③ 句，你大概會擦掉「傘在樂樂那」，寫上「傘在小林那」。LSTM 每讀一個字，也在做這三個選擇。不同的是，沒人教它規則，它在練習中自己學會該擦什麼、記什麼。

10 詞語小卡片

長短期記憶網絡 (LSTM)

升級版的循環神經網絡，1997 年發明。它會挑著記，能記住很遠以前的線索。

記憶傳送帶 (細胞狀態)

LSTM 的長期記憶。從頭通到尾，上面的東西基本原樣往後傳。

門

管著有多少東西能通過的開關。可以全關、全開，也可以只開一半。

三扇門

遺忘門決定擦掉什麼，輸入門決定寫進什麼，輸出門決定這一步說出什麼。

11 想一想

1. 如果 LSTM 的遺忘門壞了，什麼都不擦，一直往傳送帶上加東西。讀一個很長的故事時，會出什麼問題？

2. 說一件你自己「只開一半門」的事。比如整理房間時，有什麼東西你沒扔，只是先收起來？

3. 你是怎麼記住一年前的事的？更像普通循環網絡的小本子，還是更像 LSTM 的傳送帶？為什麼？

一句話記住：LSTM 有一條記憶傳送帶和三扇門，會挑著擦掉、寫進、說出，所以記得更遠。

給家長和老師：本課講的是 Christopher Olah 2015 年的博文 [Understanding LSTM Networks](#)。LSTM 由 Hochreiter 和 Schmidhuber 1997 年提出，遺忘門由 Gers 等人 2000 年加入。博文用圖解釋細胞狀態 (cell state) 和遺忘門、輸入門、輸出門，並說明普通 RNN 難學會長程依賴的原因 (梯度消失)。

講義做了簡化：門的取值在 0 到 1 之間，講義說成「門開多大，就讓多少東西通過」；遺忘門的值表示舊記憶保留多少。講義沒有講公式。「雪球」故事、「越讀越淡」圖、小明和小美的讀故事表、書包比喻、雨傘故事和傳話遊戲都是我們編的示意，不是原文內容。真實 LSTM 的記憶是一串數字，不是文字。

想深入可看本站的 [第 2 天中文導讀](#)，代碼見開源項目 [30u30 第 2 天目錄](#) (英文)。