

電腦零件輪流請假

本課講 2014 年的一篇文章。Zaremba、Sutskever 和 Vinyals 三位科學家發現：讓電腦的零件輪流「請假」，電腦就不再死記硬背。但是請假要請對地方。

適合 10 歲左右

約 40 分鐘

需要：鉛筆、彩筆、1 個骰子、3 位同伴

1 熱身：誰真的學會了？

小明和小紅要考乘法。老師先發了一張練習卷，上面有 5 道題和答案。

小明把 5 個答案全背了下來：「第 1 題是 12，第 2 題是 35……」

小紅沒有背答案。她弄懂了每道題是怎麼算出來的。

練習卷再做一遍，兩人都全對。考試那天，老師換了 5 道**新題**。

誰考得好？請圈出來： 小明 / 小紅 / 一樣好 為什麼？ _____

小明背的是舊題的答案，新題他一個也沒背過。這節課我們會看到：電腦也會犯小明的毛病。

2 電腦也會死記硬背

前面的課講過，循環神經網絡靠練習「猜下一個詞」來學本領。它練習用的文章，就是它的**練習卷**。電腦的「大腦」裡有很多小零件。零件越多，大腦越大，記性越好。可是大腦太大、練習卷太少時，電腦會像小明一樣，**把練習卷直接背下來**。

	練習捲上的句子	沒見過的新句子
背答案的電腦	猜得幾乎全對	猜得很差
找到規律的電腦	猜得不錯	也猜得不錯

練習時全對，遇到新題就錯，這種毛病叫**過擬合**。這篇文章裡，電腦的練習卷是一套英文新聞，大約 93 萬個英文單詞。對大腦很大的電腦來說，這不算多。所以在這篇文章以前，大家只敢用小的循環網絡。小腦袋不容易背答案，可是本領也有限。

3 治死記硬背的辦法：輪流請假

2012 年，Hinton 帶領的團隊想出一個辦法，叫 **Dropout**，我們叫它「請假法」。

想像小組 4 個人做手工。如果總是同一個人剪紙，他一請假，小組就完蛋了。現在換個規矩：每次隨機選幾個人請假，剩下的人必須把活幹完。這樣，每個人都要學會好幾樣本領，誰也不能只靠別人。電腦練習時也這樣做：



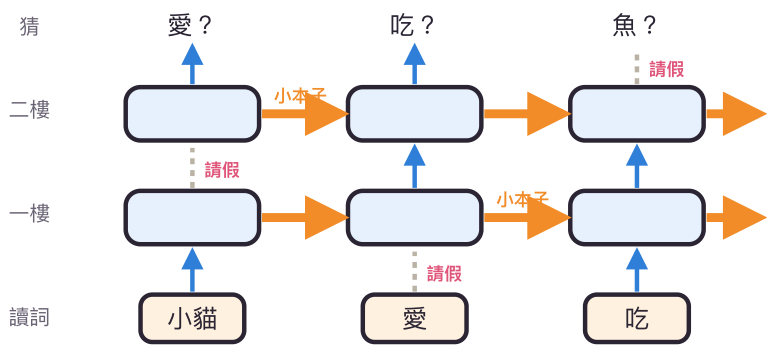
練習時一直這樣輪流請假

請假只在練習時發生。真正使用電腦時，所有零件**全員到齊**。每個零件都練過獨當一面，合在一起更可靠。

請假法在很多種電腦大腦上都很好用。可是用在循環網絡上，效果卻不好。為什麼呢？

4 循環網絡裡有兩條路

這篇文章用的循環網絡，像一棟兩層的小樓。每一層叫一個「**層**」。每讀一個詞，消息要走兩條路：



示意圖：藍色箭頭往上傳，灰色虛線是這一次請假的連接。橙色粗箭頭往右傳小本子。

哪條路	它在做什麼	要走多少步
往上傳（層與層之間）	一樓想一想剛讀的詞，交給二樓。二樓再交給「猜」。	只有 3 步
往右傳（小本子線）	每層帶著自己的小本子，讀完一個詞就交給下一個詞。	句子多長走多少步，常常幾十步

5 關鍵發現：紙條不能撕

以前，有人把請假法用在**所有**箭頭上，包括往右傳的小本子線。結果小本子每傳一次，就丟掉一些內容。想像全班傳一張紙條，每傳一次，就撕掉一小塊：

寫好的紙條	下午三點在操場門口見
傳了 5 次	下午□點在□場門□見
傳了 10 次	□□□點□□場□□見

傳到最後，誰也看不懂了。句子越長，小本子傳的次數越多，丟的也越多。電腦就記不住前面讀過什麼。

三位科學家的辦法：只在樓梯上請假

他們只讓**往上傳**的連接請假，**往右傳**的小本子線一點也不碰。就像傳紙條時，可以隨機換人來傳，但是**紙條本身不能撕**。

這樣，不管句子多長，一條消息從讀進來到猜出去，最多隻遇到 3 次請假。這個改動很小，卻讓大家終於敢把循環網絡做大了。

6 結果：猜下一個詞更準了

科學家用「**猶豫分**」來打分。它的意思是：猜下一個詞時，電腦好像在多少個詞中間猶豫不決。猶豫分越小，說明猜得越準。

電腦的大腦	練習時請假嗎	猶豫分
小（每層 200 個零件）	不請假	約 114
中（每層 650 個零件）	請假	約 83
大（每層 1500 個零件）	請假	約 78

第一行是不請假時最好的成績。不請假的大腦再做大，反而會死記硬背。用了新的請假法，大腦可以放心做大，猶豫分明顯下降。英譯法翻譯、聽懂冰島語說話、看圖寫一句話，這 3 個任務的成績也都提高了一些。

7 這和今天有什麼關係

這個辦法後來成了訓練循環網絡的常用做法。今天的聊天 AI 換了一種新結構，但還在用同一個道理：**請假可以，但不要打斷負責傳消息的主線**。

也有變化。最大的聊天 AI 讀的書太多，多到根本背不下來。所以它們最開始大量讀書時，常常乾脆不請假。用少量材料學新本領時，請假法又常常派上用場。同一個好辦法，放對地方是幫忙，放錯地方是搗亂。

8 動手：請假遊戲

玩法（4 個人，約 15 分鐘，需要 1 個骰子、白紙、彩筆）

1. 小組合畫一座房子。1 號畫屋頂，2 號畫牆，3 號畫門和窗，4 號畫樹和太陽。
2. 第 1 輪：沒人請假。大家在同一張紙上各畫各的，限時 1 分鐘。
3. 第 2—4 輪：先擲骰子，擲到幾號誰就請假（5、6 重擲）。剩下的人替他畫完。
4. 考試輪：全員回來，每人獨自畫一整座房子，限時 2 分鐘。畫齊 4 個部分的有 _____ 人。

第幾輪	誰請假	誰替他畫了	4 個部分畫齊了嗎
第 2 輪			
第 3 輪			
第 4 輪			

你會發現：玩了幾輪，每個人都會畫好幾個部分了。小組不再依賴某一個人，電腦的零件也是這樣。

9 動手：背答案，還是找規律？

玩法（2 個人，約 10 分鐘）

神秘機器按一個秘密規則工作。它的「練習卷」：放進 1 吐出 3，放進 2 吐出 5，放進 3 吐出 7，放進 6 吐出 13。

1. 一人當「背書蟲」：只許看練習卷。練習捲上沒有的數，只能從捲上挑一個數來猜。
2. 一人當「偵探」：先找出秘密規則，再用規則算。
3. 兩人分別答題，再一起找出正確答案，打 ✓ 或 ✗。然後換角色再玩。

題目	背書蟲的答案	偵探的答案	正確答案
練習題：放進 6			
新題：放進 4			
新題：放進 10			

你會發現：新題只有偵探能做對。背書蟲就是「過擬合」的電腦。

10 詞語小卡片

過擬合

練習時全對，遇到新題就錯。就像只背答案、不懂道理。

Dropout (請假法)

練習時隨機讓一部分零件請假，逼每個零件都學會獨當一面。真正使用時全員到齊。

層與層之間

消息往上傳的那條路，從樓下傳到樓上。這篇文章只在這裡請假。

小本子線 (記憶線)

消息往右傳的那條路，把記下的線索交給下一個詞。這裡不能請假。

11 想一想

1. 為什麼真正使用電腦時，要讓所有零件回來，不再請假？

2. 傳紙條時，每傳一次就撕掉一個字。傳 20 次後，紙條會怎樣？這和循環網絡有什麼關係？

3. 你自己學習時，什麼時候是在「背答案」，什麼時候是在「找規律」？各舉一個例子。

一句話記住：請假要請對地方。層與層之間可以請假，小本子那條記憶線不能斷。

給家長和老師：本課講的是 Wojciech Zaremba、Ilya Sutskever、Oriol Vinyals 2014 年的論文 [Recurrent Neural Network Regularization](#)。論文只在多層 LSTM 的非循環連接（層與層之間）上加 dropout，時間方向的循環連接不加。在 Penn Treebank 詞級語言模型上，測試困惑度從最佳無正則模型的 114.5 降到 82.7（中型）和 78.4（大型）。

講義中的「猶豫分」就是困惑度，「在多少個詞中間猶豫」是簡化的理解方式。小明和小紅的故事、兩層小樓示意圖、撕紙條的例子、請假遊戲和神秘機器都是示意，不是原文內容。神秘機器的規則是「乘 2 再加 1」。想深入可看本站的 [第 3 天中文導讀](#)，代碼見開源項目 [30u30 第 3 天目錄](#)（英文）。