

电脑零件轮流请假

本课讲 2014 年的一篇文章。Zaremba、Sutskever 和 Vinyals 三位科学家发现：让电脑的零件轮流“请假”，电脑就不再死记硬背。但是请假要请对地方。

适合 10 岁左右

约 40 分钟

需要：铅笔、彩笔、1 个骰子、3 位同伴

1 热身：谁真的学会了？

小明和小红要考乘法。老师先发了一张练习卷，上面有 5 道题和答案。

小明把 5 个答案全背了下来：“第 1 题是 12，第 2 题是 35……”

小红没有背答案。她弄懂了每道题是怎么算出来的。

练习卷再做一遍，两人都全对。考试那天，老师换了 5 道**新题**。

谁考得好？请圈出来： 小明 / 小红 / 一样好 为什么？ _____

小明背的是旧题的答案，新题他一个也没背过。这节课我们会看到：电脑也会犯小明的毛病。

2 电脑也会死记硬背

前面的课讲过，循环神经网络靠练习“猜下一个词”来学本领。它练习用的文章，就是它的**练习卷**。电脑的“大脑”里有很多小零件。零件越多，大脑越大，记性越好。可是大脑太大、练习卷太少时，电脑会像小明一样，**把练习卷直接背下来**。

	练习卷上的句子	没见过的新句子
背答案的电脑	猜得几乎全对	猜得很差
找到规律的电脑	猜得不错	也猜得不错

练习时全对，遇到新题就错，这种毛病叫**过拟合**。这篇文章里，电脑的练习卷是一套英文新闻，大约 93 万个英文单词。对大脑很大的电脑来说，这不算多。所以在这篇文章以前，大家只敢用小的循环网络。小脑袋不容易背答案，可是本领也有限。

3 治死记硬背的办法：轮流请假

2012 年，Hinton 带领的团队想出一个办法，叫 **Dropout**，我们叫它“请假法”。

想象小组 4 个人做手工。如果总是同一个人剪纸，他一请假，小组就完蛋了。现在换个规矩：每次随机选几个人请假，剩下的人必须把活干完。这样，每个人都要学会好几样本领，谁也不能只靠别人。电脑练习时也这样做：



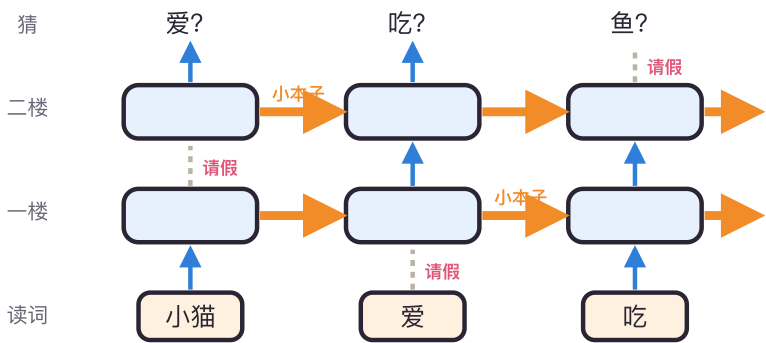
练习时一直这样轮流请假

请假只在练习时发生。真正使用电脑时，所有零件**全员到齐**。每个零件都练过独当一面，合在一起更可靠。

请假法在很多种电脑大脑上都很好用。可是用在循环网络上，效果却不好。为什么呢？

4 循环网络里有两条路

这篇文章用的循环网络，像一栋两层的小楼。每一层叫一个“层”。每读一个词，消息要走两条路：



示意图：蓝色箭头往上传，灰色虚线是这一次请假的连接。橙色粗箭头往右传小本子。

哪条路	它在做什么	要走多少步
往上传（层与层之间）	一楼想一想刚读的词，交给二楼。二楼再交给“猜”。	只有 3 步
往右传（小本子线）	每层带着自己的小本子，读完一个词就交给下一个词。	句子多长走多少步，常常几十步

5 关键发现：纸条不能撕

以前，有人把请假法用在**所有**箭头上，包括往右传的小本子线。结果小本子每传一次，就丢掉一些内容。想象全班传一张纸条，每传一次，就撕掉一小块：

写好的纸条 下午三点在操场门口见
传了 5 次 下午□点在□场门□见
传了 10 次 □□点□场□□见

传到最后，谁也看不懂了。句子越长，小本子传的次数越多，丢的也越多。电脑就记不住前面读过什么。

三位科学家的办法：只在楼梯上请假

他们只让**往上传**的连接请假，**往右传**的小本子线一点也不碰。就像传纸条时，可以随机换人来传，但是**纸条本身不能撕**。

这样，不管句子多长，一条消息从读进来到猜出去，最多只遇到 3 次请假。这个改动很小，却让大家终于敢把循环网络做大了。

6 结果：猜下一个词更准了

科学家用“**犹豫分**”来打分。它的意思是：猜下一个词时，电脑好像在多少个词中间犹豫不决。犹豫分越小，说明猜得越准。

电脑的大脑	练习时请假吗	犹豫分
小（每层 200 个零件）	不请假	约 114
中（每层 650 个零件）	请假	约 83
大（每层 1500 个零件）	请假	约 78

第一行是不请假时最好的成绩。不请假的大脑再做大，反而会死记硬背。用了新的请假法，大脑可以放心做大，犹豫分明显下降。英译法翻译、听懂冰岛语说话、看图写一句话，这 3 个任务的成绩也都提高了一些。

7 这和今天有什么关系

这个办法后来成了训练循环网络的常用做法。今天的聊天 AI 换了一种新结构，但还在用同一个道理：**请假可以，但不要打断负责传消息的主线**。

也有变化。最大的聊天 AI 读的书太多，多到根本背不下来。所以它们最开始大量读书时，常常干脆不请假。用少量材料学新本领时，请假法又常常派上用场。同一个好办法，放对地方是帮忙，放错地方是捣乱。

8 动手：请假游戏

玩法（4 个人，约 15 分钟，需要 1 个骰子、白纸、彩笔）

- 1. 小组画一座房子。1 号画屋顶，2 号画墙，3 号画门和窗，4 号画树和太阳。
- 2. 第 1 轮：没人请假。大家在同一张纸上各画各的，限时 1 分钟。
- 3. 第 2—4 轮：先掷骰子，掷到几号谁就请假（5、6 重掷）。剩下的人替他画完。
- 4. 考试轮：全员回来，每人独自画一整座房子，限时 2 分钟。画齐 4 个部分的有 _____ 人。

第几轮	谁请假	谁替他画了	4 个部分画齐了吗
第 2 轮			
第 3 轮			
第 4 轮			

你会发现：玩了几轮，每个人都会画好几个部分了。小组不再依赖某一个人，电脑的零件也是这样。

9 动手：背答案，还是找规律？

玩法（2 个人，约 10 分钟）

神秘机器按一个秘密规则工作。它的“练习卷”：放进 1 吐出 3，放进 2 吐出 5，放进 3 吐出 7，放进 6 吐出 13。

- 1. 一人当“背书虫”：只许看练习卷。练习卷上没有的数，只能从卷上挑一个数来猜。
- 2. 一人当“侦探”：先找出秘密规则，再用规则算。
- 3. 两人分别答题，再一起找出正确答案，打 ✓ 或 ✕。然后换角色再玩。

题目	背书虫的答案	侦探的答案	正确答案
练习题：放进 6			
新题：放进 4			
新题：放进 10			

你会发现：新题只有侦探能做对。背书虫就是“过拟合”的电脑。

10 词语小卡片

过拟合

练习时全对，遇到新题就错。就像只背答案、不懂道理。

Dropout（请假法）

练习时随机让一部分零件请假，逼每个零件都会独当一面。真正使用时全员到齐。

层与层之间

消息往上传的那条路，从楼下传到楼上。这篇文章只在这里请假。

小本子线（记忆线）

消息往右传的那条路，把记下的线索交给下一个词。这里不能请假。

11 想一想

1. 为什么真正使用电脑时，要让所有零件回来，不再请假？

2. 传纸条时，每传一次就撕掉一个字。传 20 次后，纸条会怎样？这和循环网络有什么关系？

3. 你自己学习时，什么时候是在“背答案”，什么时候是在“找规律”？各举一个例子。

一句话记住：请假要请对地方。层与层之间可以请假，小本子那条记忆线不能断。

给家长和老师：本课讲的是 Wojciech Zaremba、Ilya Sutskever、Oriol Vinyals 2014 年的论文 [Recurrent Neural Network Regularization](#)。论文只在多层 LSTM 的非循环连接（层与层之间）上加 dropout，时间方向的循环连接不加。在 Penn Treebank 词级语言模型上，测试困惑度从最佳无正则模型的 114.5 降到 82.7（中型）和 78.4（大型）。

讲义中的“犹豫分”就是困惑度，“在多少个词中间犹豫”是简化的理解方式。小明和小红的故事、两层小楼示意图、撕纸条的例子、请假游戏和神秘机器都是示意，不是原文内容。神秘机器的规则是“乘 2 再加 1”。想深入可看本站的 [第 3 天中文导读](#)，代码见开源项目 [30u30 第 3 天目录](#)（英文）。