

第 2 课

记忆传送带和三扇门

本课讲 2015 年的一篇文章。Christopher Olah 用一组图，讲清了一种会“挑着记”的电脑程序。它叫 LSTM。

适合 10 岁左右

约 40 分钟

需要：铅笔、便利贴、几位同伴

1 热身：谁还记得开头？

先填一个简单的：天上飘着白 ____。再读一个小故事，填最后的空。读的时候不要往回看。

奶奶家养了一只白猫，叫雪球。暑假里，乐乐在奶奶家住了一个月。他每天早上跑步，下午写作业，晚上和表哥下象棋。周末，姑姑带他们去海边捡贝壳。回家那天，乐乐抱了抱奶奶，又抱了抱 ____。

第一题只看前面几个字就够了。第二题的线索藏在**最开头**，中间隔了好几句话。这种“线索离得很远”的题，对上节课的循环神经网络来说特别难。

2 小本子太小了

上节课说过，循环神经网络带着一个“小本子”，每读一个字就记一笔。问题是，这个小本子只有一页。每读一个字，它都要**整页重写**一遍。新内容写进去，旧内容就被挤掉一点。读了几十个字，开头的线索就快看不清了。

小本子里的“雪球”，越读越淡（示意）：

读完第 1 句

雪球

读完第 3 句

雪球

读完第 5 句

雪球

读到最后

雪球

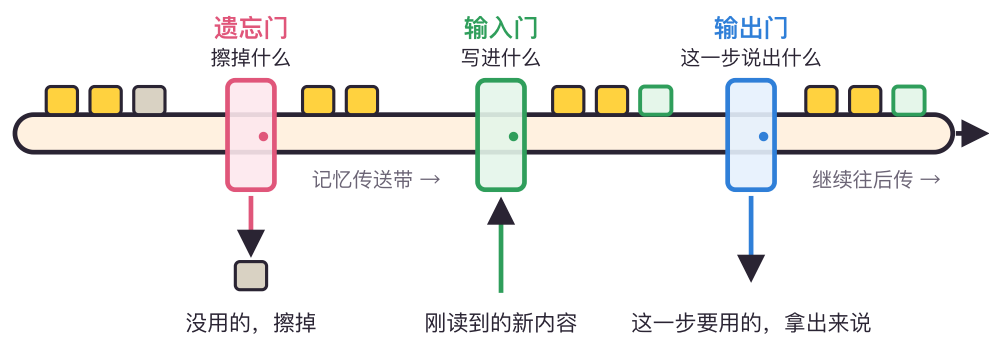
还有一个原因藏在“训练”里。“你猜错了”这个提醒，要从后往前一步一步传回去。每往前传一步，提醒就变弱一点。传了几十步，就几乎没有了。所以，按道理说它能记住远处的事，实际上却很难学会。科学家在 1990 年代就发现了这个毛病。

3 LSTM：多了一条记忆传送带

1997 年，两位科学家 Hochreiter 和 Schmidhuber 发明了一种升级版的循环神经网络。它叫**长短期记忆网络**，英文简称 **LSTM**。2015 年，Olah 画了一组图，一步一步把它讲清楚。很多人学 LSTM，都从

这篇文章开始。

LSTM 也有上节课的小本子，但它还多了一条**记忆传送带**。科学家管它叫“细胞状态”，意思就是长期记忆。



传送带从故事开头通到结尾。上面的东西基本**原样往后传**。每读一个字，它只做两件小事：擦掉一点，加上一点。不用整页重写，开头的线索就能传得很远。训练时，“你猜错了”的提醒也能顺着传送带往回走，不容易变弱。

4 三扇门：擦掉、写进、说出

传送带旁边站着三扇门。门就是一个开关，管着有多少东西能通过。这就像开学整理书包：

哪扇门	它决定什么	整理书包时
遗忘门	旧记忆哪些擦掉、哪些留下	新学期，拿出上学期的旧作业
输入门	新内容哪些写上传送带	放进新发的课本
输出门	这一步说出哪些记忆	数学课只拿数学书，别的书留在书包里

5 门可以只开一半

这三扇门不是只有“开”和“关”两种样子。它们更像水龙头，可以拧到任何大小。**门开多大，就让多少东西通过。**



比如遗忘门开一半，旧记忆就留下一半。它没有消失，只是变淡了。就像练习本还剩一半空页，你撕掉写过的页，留着空页继续用。

谁来拧水龙头？ 没有人帮 LSTM 拧。每读一个字，它都根据刚读到的字和上一步的想法，决定三扇门各开多大。怎么决定，是它在训练中**自己学会**的。

6 看 LSTM 一句一句读故事

LSTM 在玩“猜下一个字”，要猜空里填“他”还是“她”。看看三扇门每一句做了什么（示意）：

读到的句子	遗忘门擦掉	输入门写进	输出门说出
小明是二年级的男孩。	没什么要擦	主角：小明，男孩	不用说
____ 最喜欢画恐龙。	没什么要擦	主角爱画恐龙	空里填“他”
新同学小美是个女孩。	旧主角小明	主角：小美，女孩	不用说
____ 带来了一盒新蜡笔。	没什么要擦	主角带了蜡笔	空里填“她”

7 这和今天有什么关系

2013 到 2015 年前后，听写语音、翻译的程序很多都用 LSTM，上节课 Karpathy 的程序也用它。后来 LSTM 被换下来了，因为它只能一个字一个字按顺序读，练起来慢。今天的 ChatGPT、Claude 这些聊天 AI，用的是 2017 年出现的新方法 Transformer。

可是 LSTM 的好点子还在。今天的大 AI 里也有一条“直通的路”，每一层只加一点、改一点。这和记忆传送带是同一个想法。Olah 在文章结尾还猜：下一个大方向是“注意力”，就是每一步挑出最该看的信息。2 年后的 Transformer 正是靠它。

8 动手：传话链 vs 记事本

玩法（5 到 8 个人站成一排，约 10 分钟）

第 1 轮 传话链：第 1 人看纸条：“周五下午 3 点，到 3 楼美术室，带红色水彩笔。”悄悄传给下一人。每人传话时，末尾再加一样要带的东西，比如“还要带苹果”。最后一人大声说出来。

第 2 轮 记事本：换一句：“周三上午 10 点，到操场东门，穿蓝色运动鞋。”写在纸上往后传。每人不用重说，只在纸上添一样东西。最后一人大声读出来。

轮次	原话的 4 个要点	还剩几个	最后说的关键词
传话链	周五、3 点、美术室、水彩笔		
记事本	周三、10 点、东门、运动鞋		

你会发现：传话链每过一个人，整句话都要重说一遍，开头很容易丢。这像普通循环网络的小本子。记事本上，原话一直待在纸上，每人只加一点。这就像 LSTM 的记忆传送带。

9 动手：三扇门便利贴游戏

玩法（2 到 4 个人，约 15 分钟）

- 1. 一个人当“读故事的人”，其他人当 LSTM。LSTM 面前放 3 张空便利贴，这就是记忆传送带。只能用这 3 张。
- 2. 读故事的人每次只读一句。LSTM 不能看下面的故事，也不能回头问。
- 3. 每听完一句，LSTM 做三件事：**擦掉**没用的便利贴（拿不准的，可以只把字擦淡）；把要记的事**写进**空便利贴，一张只写几个字；被提问时，只看便利贴**说出**答案。
- 4. 每一句做完，在记录表里写一笔。

① 周一早上，乐乐带着一把蓝色雨伞去上学。② 路上下起了大雨，同学小林没有带伞。③ 乐乐把蓝伞借给了小林。④ 中午出了太阳，大家去操场踢足球。⑤ 小林踢进了 2 个球，全班都在欢呼。⑥ 放学时，天又下雨了。

提问：乐乐的蓝伞现在在谁手里？小林踢进了几个球？

第几句	擦掉了什么	写进了什么	说出了什么
①			
②			
③			
④			
⑤			
⑥ 和提问			

你会发现：只有 3 张便利贴，你必须挑着记。读到第 ③ 句，你大概会擦掉“伞在乐乐那”，写上“伞在小林那”。LSTM 每读一个字，也在做这三个选择。不同的是，没人教它规则，它在练习中自己学会该擦什么、记什么。

10 词语小卡片

长短期记忆网络 (LSTM)

升级版的循环神经网络，1997 年发明。它会挑着记，能记住很远以前的线索。

记忆传送带 (细胞状态)

LSTM 的长期记忆。从头通到尾，上面的东西基本原样往后传。

门

管着有多少东西能通过的开关。可以全关、全开，也可以只开一半。

三扇门

遗忘门决定擦掉什么，输入门决定写进什么，输出门决定这一步说出什么。

11 想一想

1. 如果 LSTM 的遗忘门坏了，什么都不擦，一直往传送带上加东西。读一个很长的故事时，会出什么问题？

2. 说一件你自己“只开一半门”的事。比如整理房间时，有什么东西你没扔，只是先收起来？

3. 你是怎么记住一年前的事的？更像普通循环网络的小本子，还是更像 LSTM 的传送带？为什么？

一句话记住：LSTM 有一条记忆传送带和三扇门，会挑着擦掉、写进、说出，所以记得更远。

给家长和老师：本课讲的是 Christopher Olah 2015 年的博文 [Understanding LSTM Networks](#)。LSTM 由 Hochreiter 和 Schmidhuber 1997 年提出，遗忘门由 Gers 等人 2000 年加入。博文用图解释细胞状态 (cell state) 和遗忘门、输入门、输出门，并说明普通 RNN 难学会长程依赖的原因 (梯度消失)。

讲义做了简化：门的取值在 0 到 1 之间，讲义说成“门开多大，就让多少东西通过”；遗忘门的值表示旧记忆保留多少。讲义没有讲公式。“雪球”故事、“越读越淡”图、小明和小美的读故事表、书包比喻、雨伞故事和传话游戏都是我们编的示意，不是原文内容。真实 LSTM 的记忆是一串数字，不是文字。

想深入可看本站的 [第 2 天中文导读](#)，代码见开源项目 [30u30 第 2 天目录](#) (英文)。