

给电脑大脑配块记忆板

本课讲 2014 年的一篇文章。DeepMind 公司的 Graves、Wayne 和 Danihelka 给神经网络接上了一块“记忆格子板”。他们管它叫神经图灵机。

适合 10 岁左右

约 40 分钟

需要：方格纸、便利贴、12 张卡片、2 位同伴

1 热身：记在脑子里，还是记在纸上？

请同伴读一遍第 1 串词，只读一遍。读完后，你把记住的词按顺序写进格子。

第 1 串（只能记在脑子里）：苹果、火车、月亮、小猫、雨伞、铅笔、西瓜、足球

第 2 串（可以边听边写）：香蕉、飞机、太阳、小狗、帽子、橡皮、葡萄、篮球

第 1 串								
第 2 串								

第 1 串记对了 _____ 个，第 2 串记对了 _____ 个。第 2 串容易得多，因为你有了一块**脑子外面的记忆板**。

2 普通神经网络：只能记在脑子里

神经网络就是电脑里的“大脑”，由很多小零件连成。前面几课的神经网络，都把要记的东西**揉在零件里**。第 1 课的小本子、第 2 课的传送带，都长在网络里面。零件既要算，又要记。

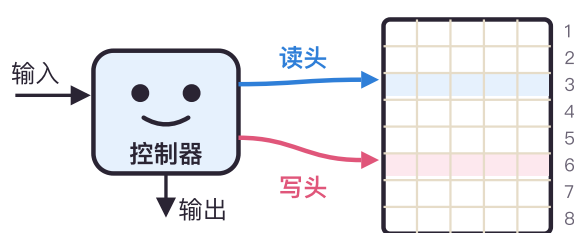
麻烦是：小本子的大小是固定的，东西一多就挤不下。想多记一点，只能把整个大脑做大。这就像为了记住更长的购物清单，先换一个更大的脑袋。能不能给神经网络一张“草稿纸”？

3 电脑的办法：一块单独的内存

你用的电脑，早就把“算”和“记”分开了。处理器（CPU）只管算。**内存**只管存，像走廊里一排储物柜。每个柜子都有号码，号码叫**地址**。电脑说“打开 37 号柜子”，就能直接读或存。这叫按地址读写。

1936 年，数学家图灵设想了一台机器：一条长纸带分成很多格子，一个“头”能读、写、往左右挪一格。它叫**图灵机**。神经图灵机的名字就从这里来。

4 神经图灵机 = 大脑 + 记忆格子板



控制器：一个神经网络，就是大脑。它决定读哪里、写什么、往哪挪。

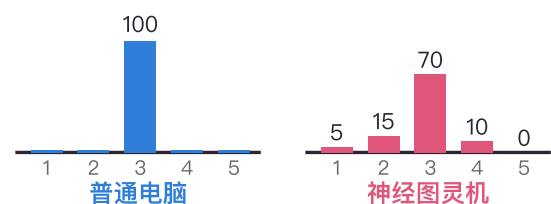
读头和写头：大脑伸出去的两只“手”。读头负责读。写头负责写，分两步：先擦掉一点旧的，再写上新的。

记忆格子板：大脑外面的一块板子，像一张方格纸。论文实验里，板子有 128 行，每行 20 个小格子，每格放一个数。科学家管它叫**记忆矩阵**。想多记一点，换一块更大的板子就行，不用把大脑做大。

5 难点：读写头必须是“模糊”的

普通电脑一次只读一个格子：要么读第 3 格，要么读第 4 格，没有中间状态。可是神经网络学本事，靠的是第 1 课说的“猜错了就悄悄调整自己”，也就是把很多小旋钮各拧一点点（第 4 课讲过旋钮）。

只读一个格子的读头，像只有“开”和“关”的电灯开关。旋钮拧一点点，读到的还是同一格，电脑得不到任何提示。拧过了头，又一下子跳到别的格子。所以它没法一点点学。



神经图灵机的读写头像**调光灯**。它把 100 份注意力分给所有格子。大部分集中在一两个格子，其他格子只分到一点点。

读出来的东西，是各格内容按份数混在一起。写的时候，份数多的格子改得多。这样，旋钮拧一点点，读到的也变一点点，电脑就知道往哪边调。这种“分份数”的办法，就是第 13 课和第 15 课讲过的**注意力**。那两课的注意力只能读，这里的格子还能写。

6 找格子的两种办法

找法	怎么找	生活里的例子
按内容找	说出“我要找长这样的”，跟每一行比一比。越像的行，分到的注意力越多。	在操场上找“戴红帽子的人”
按位置找	从上一步的位置出发，往前或往后挪一格，或者不动。	读完这一行，就读下一行

两种找法可以搭配：先按内容找到一个东西，再按位置往后读。每一步多用哪种，是大脑自己学会的。挪格子时注意力会散开一点，像手电筒的光变淡，所以它挪完还会“收一收”。

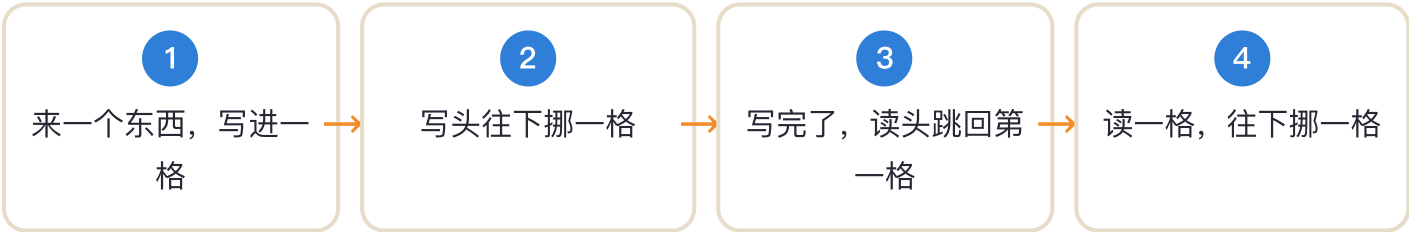
7 实验：它自己学会了小程序

科学家没给它写程序，只给它看很多题目和正确答案。论文测了 5 个任务，这里讲 4 个：

- **复制**：给它一串东西，让它原样说一遍。每个东西是一排 8 个小灯，亮灭随便排。
- **重复复制**：给它一串东西和一个次数，让它连着复制几遍。
- **联想回忆**：给它好几组东西，再拿出一组问：“这组后面是哪一组？”
- **排序**：每个东西带一个分数，让它把分数高的那些从高到低排出来。

复制任务：练短的，考长的		
练习时，每串只有 1 到 20 个东西。考试时，串加长到 120 个。		
串有多长	神经图灵机	LSTM（第 2 课）
1 到 20 个（练过）	学得更快，错得更少	也能复制，错得多一些
超过 20 个（没练过）	到 120 个还大体正确	很快就出错

科学家把读写头每一步的注意力画出来，发现它学会的做法很像一段程序：



只要格子板还有空格，这段程序不管串多长都管用。LSTM 把东西揉在零件里，串一长就挤不下了。

8 这和今天有什么关系

说实话，今天几乎没人直接用神经图灵机。它练起来不稳定，速度慢，很难做大。可是它的想法活了下来：

- **按内容找**：ChatGPT、Claude 这些聊天 AI 的核心零件里，用的就是同样的找法，叫注意力。
- **外接记忆板**：有的 AI 先去资料库里查东西，再回答。有的 AI 干长活时，把中间结果写进文件，再读回来。
- **练短的，考长的**：今天的 AI 还不擅长。比如学过短的加法题，换成位数很多的加法，常常算错。

9 动手：记忆格子板复制游戏

玩法（3 个人，约 15 分钟）

准备：方格纸上画一行 10 个格子，第 1 格画 ★。粉色便利贴当写头，黄色当读头，都贴在第 1 格旁。A 出题，B 当控制器，C 检查。B 只照程序卡做，**不许用脑子记**。

程序卡

写：听到一个符号，画进写头指的格子，写头往下挪一格。

听到“完”：读头找到画着 ★ 的格子（按内容找）。

读：说出读头指的格子里的符号，读头往下挪一格。遇到空格就停。

1. A 慢慢说 4 个符号，比如“圆、三角、星、方”，再说“完”。B 照卡写、照卡读，C 检查。
2. 擦干净，A 换 8 个符号再来一遍。程序卡不改。
3. 最后一轮不用格子板：A 说 8 个符号，B 用脑子记，再说出来。

记一记：B 说错了几个？第 1 轮 _____ 个，第 2 轮 _____ 个，第 3 轮 _____ 个。

你会发现：同一张程序卡，4 个和 8 个符号都能用。只靠脑子记，8 个就容易错。

10 动手：按内容找 vs 按位置找

玩法（2 到 4 个人，约 10 分钟）

准备：12 张卡片，每张画一种动物，比如猫、狗、兔子、大象、鱼。正面朝上排成一排，**不写号码**。一人念题，其他人抢答，一人计时。比完再自己出 3 道新题。

题目	用的找法（圈一个）	用了几秒
从左数第 9 张是什么动物？	内容 / 位置 / 两种	
有长鼻子的动物在哪张？用手指出来。	内容 / 位置 / 两种	
找到猫，再往右挪 2 张，是什么动物？	内容 / 位置 / 两种	

你会发现：找“长鼻子”一眼就看到，这是按内容找。找“第 9 张”得一张张数，这是按位置找。有的题要两种合起来。神经图灵机每一步也要选用哪种。

11 词语小卡片

控制器

神经图灵机的大脑，决定读、写和挪动。

记忆矩阵

大脑外面的记忆格子板。

读头和写头

大脑的两只“手”，一只读，一只写。

模糊读写（注意力）

注意力按份数分给很多格子。

12 想一想

1. 你的书包、铅笔盒、记事本，哪个最像神经图灵机的格子板？为什么？

2. 读头如果只能“要么这格要么那格”，电脑为什么很难练会？用电灯开关和调光灯说一说。

3. 在班里找同学时，什么时候你按内容找？什么时候按位置找？各举一个例子。

一句话记住：神经图灵机给电脑大脑接上一块记忆格子板，用模糊的读写头一点点学会读写。

给家长和老师：本课讲的是 Alex Graves、Greg Wayne、Ivo Danihelka 2014 年的论文 [Neural Turing Machines](#)。论文给控制器（前馈网络或 LSTM）接上外部记忆矩阵（实验用 128×20 ），读写都用对所有行的连续权重分布完成，整个系统可用梯度下降端到端训练。寻址分 4 步：内容寻址、插值、循环卷积移位、锐化。5 个任务是复制、重复复制、联想回忆、动态 N 元语法和优先级排序。复制任务训练长度 1 到 20，测到 120 时神经图灵机仍大体正确，LSTM 超过 20 后很快出错。

讲义做了简化：权重之和为 1，讲义说成“100 份注意力”；擦除和添加说成“先擦掉一点旧的，再写上新的”；锐化说成“收一收”；排序任务是输出优先级最高的若干个向量。讲义没有讲公式。热身词表、柱状图里的份数（5、15、70、10、0）、8 行格子板图、电灯开关和储物柜比喻、两个动手游戏都是我们编的示意，不是原文内容。真实的格子板里存的是数，不是文字或符号。想深入可看本站的 [第 17 天中文导读](#)，代码见开源项目 [30u30](#) [第 17 天目录](#)（英文）。