

# 一袋水果怎么排队

本课讲 2015 年的一篇文章。Vinyals、Bengio 和 Kudlur 三位科学家发现：同一组东西，换个顺序交给电脑，结果就会不一样。文章的名字就叫《顺序很重要》。

适合 10 岁左右

约 40 分钟

需要：铅笔、12 张小卡片、2 位同伴

## 1 热身：三个水果排几种队

书包里有 3 个水果：苹果、香蕉、橙子。把它们从左到右排成一排，一共有几种排法？前两种已经写好了，请写出其余几种。

① 苹果、香蕉、橙子

② 苹果、橙子、香蕉

③ \_\_\_\_\_

④ \_\_\_\_\_

⑤ \_\_\_\_\_

⑥ \_\_\_\_\_

3 个水果有 6 种排法。再加 1 个梨，就有 24 种。10 个水果呢？超过 360 万种！

可是不管怎么排，书包里的东西都一样。“一袋水果”本来就没有先后。

## 2 有的东西有顺序，有的没有

“我爱妈妈”和“妈妈爱我”，字一样，意思不同。像句子这样，先后一换意思就变的一串东西，叫**序列**。一袋水果、抽奖箱里的号码球，先后换了也没关系。像这样只管“有哪些”、不管“谁先谁后”的一组东西，叫**集合**。

| 东西     | 序列 | 集合 | 东西        | 序列 | 集合 |
|--------|----|----|-----------|----|----|
| 一首歌的歌词 |    |    | 今天来了哪些同学  |    |    |
| 做蛋糕的步骤 |    |    | 跳棋盘上剩哪些棋子 |    |    |

请在表里打 ✓。麻烦在这里：第 1 课讲过带“小本子”的电脑，它**一个一个地读**，读一个记一笔。要把一袋水果交给它，就得先排成一队。这种一个一个读、一个一个写的程序，叫“序列到序列”模型。

### 3 换个顺序，结果就变了

按道理说，水果怎么排都一样。科学家发现，**电脑其实很在乎**。

#### 一个真实的例子：倒着读，翻译得更好

2014 年，有科学家教电脑把英语翻译成法语。他们把英语句子**倒过来**交给电脑读，电脑反而翻译得明显更好。句子没变，只是换了读的顺序。这篇文章收集了这类发现，问了 2 个问题：顺序到底重不重要？如果重要，该怎么办？

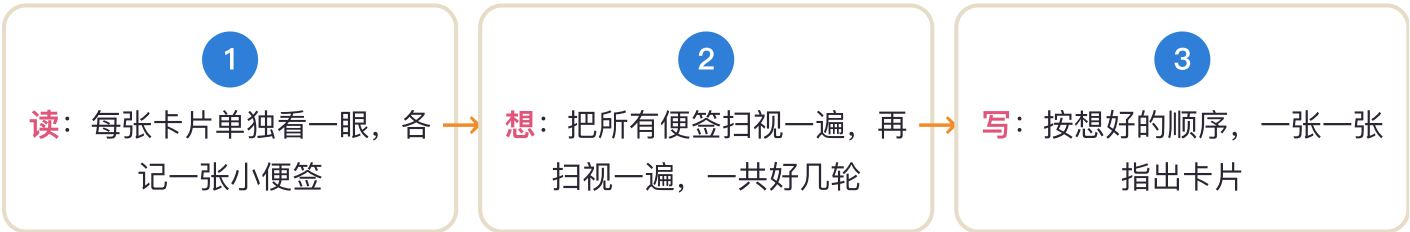
#### 道理上一样，做起来不一样

第 2 课说过，小本子只有一页，早读的东西会慢慢变淡。所以先读谁、后读谁，本子上的笔记就不一样。

就像背购物清单，按超市货架的顺序背最好记。每种顺序在道理上都一样对，**可是有的顺序电脑学得快，有的学得慢**。

### 4 办法一：把卡片摊开，多看几眼

排 6 张数字卡片时，你会**全部摊在桌上，来回看几遍**，再动手。科学家给电脑的办法也分 3 步：



第 2 步最关键。每一轮扫视，电脑都会决定哪几张多看两眼。这种“把眼光多放在有用的地方”的本领，叫**注意力**，第 13 课和第 15 课讲过它。



#### 为什么这样就不怕顺序了？像榨果汁

每一轮，电脑把看到的東西**按轻重混在一起**，变成一个想法。这就像榨果汁。先放苹果还是先放香蕉，榨出来的果汁一样。卡片怎么摆，混出来的想法也一样。扫视几轮可以自己定，轮数越多，电脑动手前“想”得越久。

## 5 例子：给打乱的数字排队

科学家用排序来考电脑。电脑拿到一组打乱的小数，都在 0 和 1 之间。它要从小到大，一个一个**指出**它们在哪儿。比如上图 4 张卡片依次是 0.73、0.21、0.95、0.40（示意），电脑就回答：“第 2 张、第 4 张、第 1 张、第 3 张。”这种“指一指”的写法叫**指针网络**，第 18 课会专门讲。

科学家试了 5 个数、10 个数和 15 个数，发现两件事：

1. **扫视的轮数越多，排得越准**。先多想几轮再动手，比看一眼就动手强。
2. **数越多，越难排准**。这和你自己排卡片的感觉一样。

## 6 办法二：写的时候，也挑个好顺序

顺序的麻烦不只在“读”，也在“写”。有时候，答案本身就是一个集合。

**老师的答案顺序不一样，电脑就犯糊涂（示意）**

题目：“袋子里有哪些水果？”答案是苹果和香蕉。第 1 次，老师写“苹果、香蕉”。第 2 次，另一位老师写“香蕉、苹果”。两个答案都对。可是电脑一个字一个字地学写，它会想：“第一个该写苹果，还是香蕉？”它摇来摆去，学得很慢。

文章里还有一个真实的例子：把一句话拆成大块和小块，写成“结构图”。同一张图按不同的先后写，电脑学得好不好也不同。

他们的办法是：**练习的时候，让电脑自己挑一个最好写的顺序**。每道题都先看几种写法，电脑现在觉得哪种最顺手，就按哪种练。刚开始练习时，要故意多试几种顺序，免得太早认定一种坏写法。作者在一个人造的练习题上试过，这个办法确实能帮电脑找到好写的顺序。

## 7 这和今天有什么关系

第 13 课讲的 Transformer，是今天聊天 AI 的大脑结构。不告诉它字的位置，它就像看一桌摊开的卡片，分不出先后。第 21 课的程序就用了本文“多看几轮”的办法。

今天的聊天 AI 也会被顺序影响。2021 年有研究发现，只换几道例题的先后，AI 答对的比例就可能变化很大。还有研究发现，AI 记住了“小明的妈妈是王老师”，却不一定答得出“王老师的孩子是谁”。让 AI “先写想法、再写答案”常常更准，这也是在挑一个更好写的顺序。

## 8 动手：排队听 vs 摊开看

玩法（2 个人，约 15 分钟，需要 12 张小卡片）

- A 出题，B 当“电脑”。每轮 B 都要把 6 个数从小到大写出来。3 组数只给 A 看：
- 第 1 组 47、12、85、30、66、9    第 2 组 58、23、91、14、70、36    第 3 组 81、5、44、63、27、99
1. **排队听**：A 每秒念 1 个第 1 组的数，只念一遍。念完 B 才动笔。
  2. **摊开看 1 眼**：第 2 组写成卡片，打乱摊开。B 看 5 秒，翻过去再动笔。
  3. **摊开看 3 眼**：用第 3 组数做卡片。B 看 5 秒，卡片翻过去 5 秒，重复 3 次后再动笔。
  4. 数一数 B 排对了几个位置，填进表里。交换角色时，新出题人自己编 3 组数。

| 第几轮         | 排队听 | 摊开看 1 眼 | 摊开看 3 眼 |
|-------------|-----|---------|---------|
| B 排对几个（共 6） |     |         |         |
| 交换后排对几个     |     |         |         |

**你会发现**：排队听时，前面的数容易忘。卡片摊开时，可以来回比较。多看几眼，通常排得更准，就像文章里“多想几轮”的办法。

## 9 动手：哪个顺序最好记

玩法（3 个人，全班更好；约 15 分钟）

1. 3 个人各拿一行，词一样，先后不同。**不要偷看别人那一行。**
2. 默读 30 秒后盖住，再用 1 分钟写下记得的词，顺序不限。
3. 数一数记住几个，填进表里。全班玩时分 3 组，算每组的平均数。

| 顺序    | 12 个词                                                 | 记住几个 |
|-------|-------------------------------------------------------|------|
| 甲：打乱  | 跳绳、饼干、尺子、游泳、糖果、书包、足球、薯片、橡皮、跑步、酸奶、铅笔                   |      |
| 乙：按类排 | 铅笔、橡皮、尺子、书包   饼干、薯片、酸奶、糖果   足球、跳绳、游泳、跑步               |      |
| 丙：编故事 | 书包、铅笔、足球、饼干、橡皮、跳绳、酸奶、尺子、游泳、薯片、跑步、糖果（背上书包，带好铅笔，去踢足球……） |      |

**你会发现**：按类排或编成故事，通常比打乱的好记。电脑练习时挑“最好写的顺序”，就像你挑最好记的顺序。

## 10 词语小卡片

### 集合

只管“有哪些”的一组东西，比如一袋水果。

### 序列

先后一换意思就变的一串东西，比如一句话。

### 读—想—写

单独看每张卡片，扫视几轮，再指出答案。

### 挑顺序

练习时让电脑自己挑最好写的答案顺序。

## 11 想一想

1. “全班同学的名单”是集合还是序列？“按身高排的队”呢？为什么？

---

---

2. 生活中什么事先后换了结果不变？什么事一换就不行？各举一个例子。

---

---

3. 背单词、背课文时，你会特意排一个顺序吗？什么样的顺序对你最好记？

---

---

**一句话记住：东西没顺序，电脑却在乎；摊开多看几眼，挑好写的顺序。**

给家长和老师：本课讲的是 Oriol Vinyals、Samy Bengio、Manjunath Kudlur 2015 年的论文 [Order Matters: Sequence to sequence for sets](#) (ICLR 2016)。论文指出 seq2seq 对输入和输出的排列都敏感，证据包括翻译中倒序输入更好 (Sutskever 等 2014) 和句法树线性化方式的影响。针对集合输入，作者提出读-处理-写结构：逐元素编码，用无外部输入的 LSTM 做 T 步内容注意力（后称 Set2Set），再用指针网络输出；在 5、10、15 个数的排序任务上，处理步数越多准确率越高。针对集合输出，作者提出训练中在候选顺序里选当前概率最高者计算损失，并在前期采样探索，在人造的图模型联合分布估计任务上做了验证。

“小本子”“小便签”“扫视”“榨果汁”分别对应 RNN 状态、逐元素编码、多步注意力读取和加权求和的排列不变性。0.73 等 4 个数、三轮扫视图、水果答案、购物清单、活动的数字和词表都是示意；“小明的妈妈”是对“反转诅咒”研究的通俗改写。讲义未引用论文的具体准确率。想深入可看本站的 [第 16 天中文导读](#)，代码见开源项目 [30u30 第 16 天目录](#)（英文）。