

两两握手找关系

本课讲 2017 年的一篇文章。DeepMind 的 Santoro 和同事给电脑加了一个小零件。它让电脑学会比较东西之间的关系，回答“谁离谁最远”这类问题。

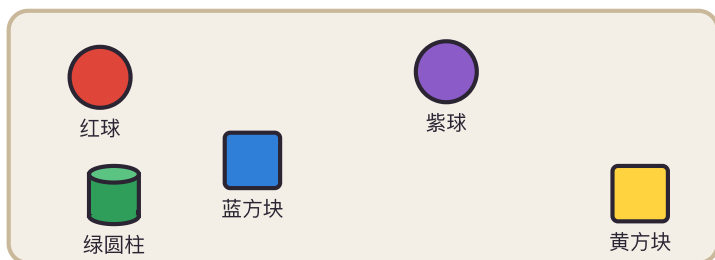
适合 10 岁左右

约 40 分钟

需要：铅笔、尺子、5 个小物件、一位同伴

1 热身：两种看图题

看下面这张图。图里有 5 个东西。请回答下面两组题。



A 组：看一眼就行	B 组：得比一比
图里有绿色的东西吗？ _____	离红球最远的东西是什么颜色？ _____
紫色的东西是什么形状？ _____	和蓝方块形状一样的是什么颜色？ _____

A 组题，盯着**一个东西**就能答。B 组题，眼睛得在好几个东西之间**来回跑**。对电脑来说，这两种题难度很不一样。

2 “是什么”和“谁和谁怎么样”

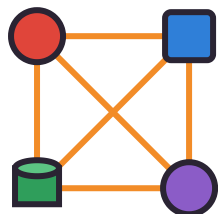
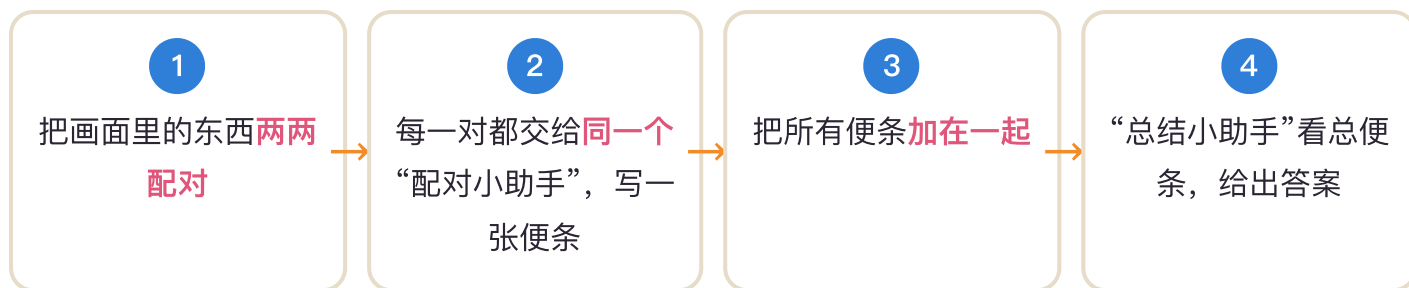
第 8 课讲过，电脑用**小窗口**一格一格扫图片，很会认出“这是一个红球”。可是 B 组题问的不是“是什么”，而是“**谁和谁怎么样**”：谁离谁远，谁在谁左边，谁和谁形状一样。

东西和东西之间的这种联系，叫**关系**。靠比较关系来回答问题，叫**关系推理**。

小窗口每次只看一小块。红球和黄方块隔得远，不会出现在同一个小窗口里。普通程序也能慢慢学会比较，可是要看非常多的例题。科学家想：能不能直接告诉它“**请你一对一对地比**”？

3 办法：全班两两握手

想象全班同学两两握手。每握一次手，这两个人就一起说一句话：“我们俩之间是什么关系？和老师的问题有关吗？”说完写在一张便条上，投进箱子。最后，老师把箱子里的便条合在一起看，得出结论。这篇文章教电脑做的，就是这件事：



4 个东西配成 6 对，就是图上 6 条线。每一对都用**同一个**配对小助手，它学会的本领每一对都能用。

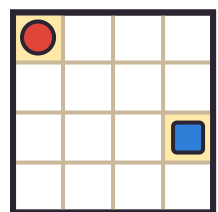
两个小助手都是会自己学习的小神经网络，合起来叫**关系网络**。它像一块积木，可以插在第 8 课的小窗口后面：小窗口认东西，关系网络做比较。

算一算：几个人握几次手

新同学要和前面每个人各握一次。第 4 个人来了，再握 3 次： $3 + 3 = 6$ 。请接着填：

几个人	2	3	4	5	6	10
握几次手	1	3	6			

4 东西从哪来？问题怎么听？



没人告诉电脑“这是一个球”。小窗口扫完后，图变成一格一格的结果。**每一格就当作一个“东西”**。

每一格还贴上**位置标签**，比如“第 1 行第 1 列”。没有标签，小助手就不知道两个东西隔多远。

每一对握手前都先听题目。问远近，小助手就留意远近。

电脑更“死板”：A—B、B—A 各算一次，自己和自己也算。比如 64 格，就要算 4096 对。

5 为什么先后顺序不要紧

第 16 课讲过：一袋水果没有顺序，叫**集合**。画面里的东西也是集合。关系网络不怕顺序，秘密在“**加在一起**”：每张便条是一串数字，电脑把它们对应相加，得到总便条。

加法不管先后： $3 + 5 + 2 = 10$ $2 + 3 + 5 = 10$

便条先投哪张、后投哪张，总便条都一样，就像第 16 课的“榨果汁”。

6 考试成绩：比人还准

科学家用一个专门考关系的看图问答题库来考它，叫 **CLEVR**。画面里有方块、球和圆柱，颜色、大小不同，有的是金属，有的是橡胶。题目像“和大金属球颜色一样的东西有几个？”（示意）。电脑只能看图片本身。



横条是答对的比例。只加了这一个小零件，答对的比例就从论文对比的此前最好成绩 68.5% 升到 95.5%，超过了人类测试者。科学家还考了另外 3 种题：

考什么	怎么考、考得怎样
简单小图	作者自己画的平面小图，“认东西”和“比关系”分开考。普通程序做关系题明显落后，加上关系网络后两种题都超过 90%。
短故事问答	读几句小故事再答题，比如“小明拿起苹果，走到花园。苹果在哪？”（示意）。每句话当一个“东西”。20 种题做对了 18 种。
会动的小球	有些小球之间连着弹簧。电脑只看小球怎么动，就能猜出哪些连在一起。

7 这和今天有什么关系

第 13 课讲过 Transformer，它是今天聊天 AI 的大脑结构。在 Transformer 里，每个字都要“**抬头看**”其他所有字。换句话说，每一对字都要比一比，打个分，再按分数轻重加在一起。

这和关系网络很像，只是 Transformer 会给重要的便条更大的分量。从这个角度看，聊天 AI 读一段话，也在做大量的关系推理。

两两配对的麻烦，今天还在。想想第 3 节的握手表：1000 个字两两配对，就有将近 50 万对。所以聊天 AI 读长文章又慢又费电，科学家还在想办法少算一些对。也有被换掉的地方：“每一格当一个东西”太粗糙，后来的程序学会了找出完整的东西。CLEVR 这套题现在也难不倒电脑了。

8 动手：两两握手游戏

玩法（2 个人，约 15 分钟，需要尺子）

- 1. 桌上摆 5 个颜色、形状不同的小物件，比如橡皮、瓶盖、硬币、积木、铅笔帽，起名 A 到 E。
- 2. 两人轮流当“配对小助手”：每一对量出相隔几厘米，再写一句关系，比如“A 比 B 大”。每对写一张小便条，也抄进下表。

配对	隔几厘米	写一句关系	配对	隔几厘米	写一句关系
A—B			B—D		
A—C			B—E		
A—D			C—D		
A—E			C—E		
B—C			D—E		

- 3. 当“总结小助手”，只看表回答：离 A 最远的是 _____；挨得最近的是 _____和 _____。
- 4. 把 10 张便条洗乱，按新顺序读一遍再答。答案变了吗？ _____

你会发现：答“离 A 最远”要看 4 对，答“哪两个最近”要看 10 对。洗乱便条，答案不变。

9 动手：我出题你来答

玩法（2 个人，约 10 分钟）

- 1. A 用 3 到 6 个小物件摆一个画面。
- 2. A 出 1 道“看一眼”题和 1 道“比一比”题，像热身那样。
- 3. B 回答，说出自己比了几对，A 记下。然后换人，摆的个数要变。

谁出题	摆了几个	“看一眼”题比了几对	“比一比”题比了几对
A 出题，B 答			
B 出题，A 答			

你会发现：“看一眼”题不用比；“比一比”题，东西越多，要比的越多。

10 词语小卡片

关系推理

比较东西之间的联系来答题，比如谁离谁最远。

两两配对

一组东西每两个配成一对，一对也不漏。

关系网络

配对小助手给每对写便条，总结小助手给答案。

加在一起

把便条相加。加法不管先后，东西怎么排都一样。

11 想一想

1. 如果撕掉每一格的位置标签，电脑还能答出“离红球最远的是什么”吗？为什么？

2. 生活里哪些问题要用关系推理？举 2 个例子，比如“班里谁最高”。

3. 如果总结小助手不把便条加在一起，只看第 1 张便条，会出什么问题？

一句话记住：认东西靠看，比关系靠两两配对，再加在一起。

给家长和老师：本课讲的是 Adam Santoro、David Raposo、David G. T. Barrett 等人 2017 年的论文 [A simple neural network module for relational reasoning](#) (NeurIPS 2017)。论文提出关系网络 (RN)：共享的 MLP g_θ 处理每一对物体，求和后交给 MLP f_ϕ 输出答案，求和使结果与物体顺序无关。物体取自 CNN 特征图的格子并拼接坐标，问题的 LSTM 编码拼进每一对。CLEVR 像素输入上 RN 达到 95.5%，人类 92.6%，论文对比的此前最好模型 (CNN+LSTM+空间注意力) 68.5%。另有 Sort-of-CLEVR (两类题都超过 90%)、bAbI (联合训练解决 20 个任务中的 18 个) 和由运动轨迹推断弹簧连接的物理任务。

两个“小助手”对应 g_θ 和 f_ϕ ，“便条”对应关系向量，“位置标签”对应坐标，“先听题目”对应问题条件化。热身图、配对图、CLEVR 例题、短故事、活动物件都是示意；“1000 个字约 50 万对”是我们按两两配对算的。想深入可看本站的 [第 19 天中文导读](#)，代码见开源项目 [30u30 第 19 天目录](#) (英文)。