

## H. ARC 人類表現估計 *H-ARC*

time limit 1s

memory limit 256MB

### Statement

張詩大最近在研究「抽象與推理語料庫基準中人類表現的穩健估計」(A Robust Estimate of Human Performance on the Abstraction and Reasoning Corpus Benchmark)，簡稱為 H-ARC，以下是他的論文摘要：

抽象與推理語料庫 (ARC) 是一個視覺程式合成的基準測試，旨在測試人類和機器在具有挑戰性的分佈外泛化能力。自 2019 年以來，現有的人工智慧方法在此挑戰上進展有限。比較人類和機器的表現對於基準測試的有效性很重要。雖然先前的工作探索了人類解決 ARC 基準測試任務的能力，但他們要麼只使用原始數據集中的一部分任務，要麼使用 ARC 的變體，因此只提供了人類表現的初步估計。在這項工作中，我們透過評估 1729 名人類在原始 ARC 問題集中全部 400 個訓練任務和 400 個評估任務上的表現，獲得了對人類表現更穩健的估計。我們估計人類在訓練集上的平均正確率介於 73.3% 到 77.2% 之間，報告的實證平均為 76.2%；在公開評估集上的平均正確率介於 55.9% 到 68.9% 之間，報告的實證平均為 64.2%。然而，我們也發現 800 個任務中有 790 個任務至少有一個人能在三次嘗試內解決，這表明絕大多數公開可用的 ARC 任務原則上都能被典型的網路眾包工作者解決。值得注意的是，雖然這些數字略低於早期的估計，但人類的表現仍然大大超過了目前解決 ARC 的最先進方法。為了促進 ARC 的研究，我們公開發布了我們的數據集，稱為 H-ARC (human-ARC)，其中包含所有人類參與者的提交內容和操作軌跡。

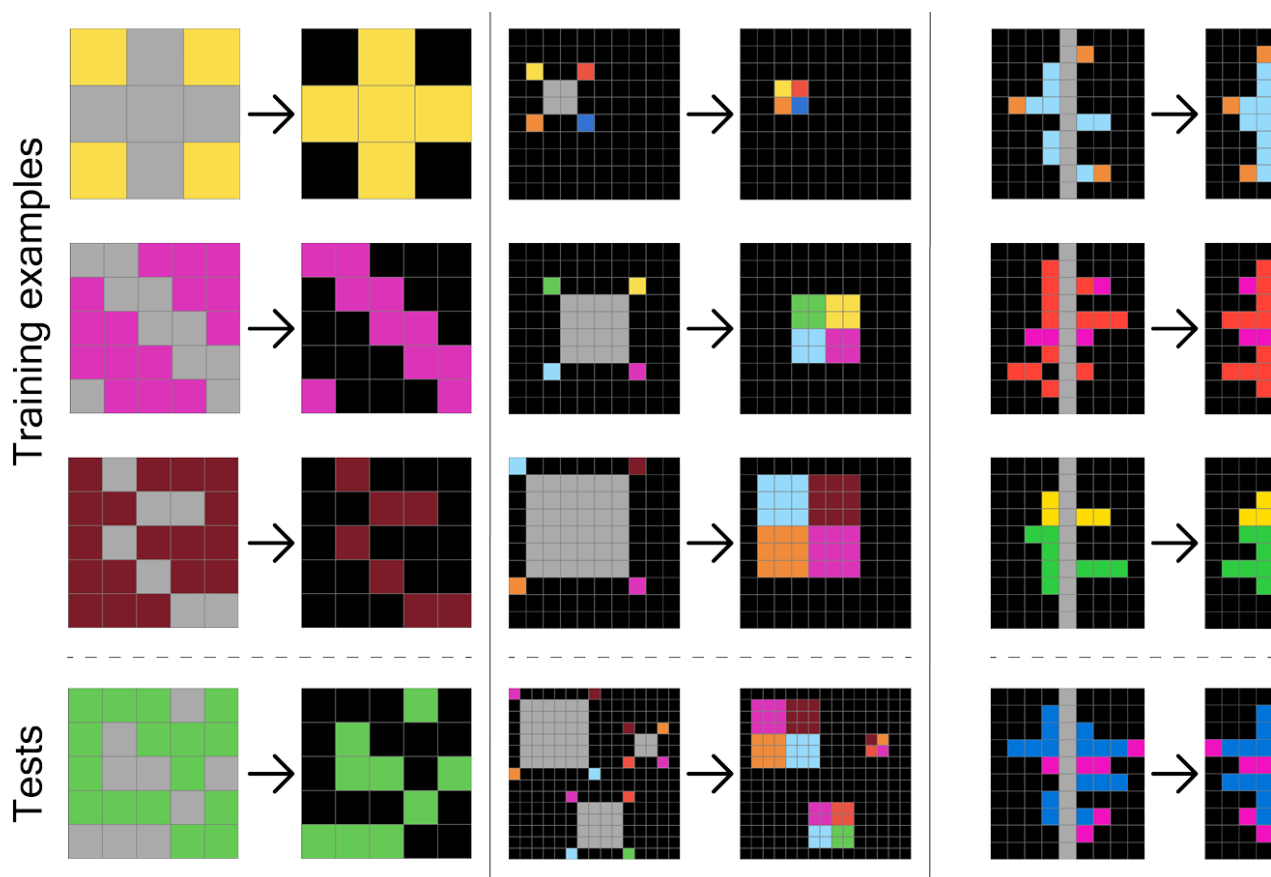
這篇論文對學術界有以下重要貢獻：

- 提供更穩健的人類表現估計：過去對 ARC (Abstraction and Reasoning Corpus) 基準測試中人類表現的評估不夠全面。這篇論文透過大規模評估，來提供了更穩健且全面的數據，證明了人類在 ARC 任務上的表現遠超目前的頂尖人工智慧模型，這對於基準測試的有效性至關重要。
- 揭示人機智能差距：研究結果清楚地突顯了人類智能在原則上能夠執行解決幾乎所有 ARC 任務所需的計算，這與目前任何已報告的 AI 模型都不同。這為 AI 系統設定了一個具有挑戰性的目標，並指出當前 AI 方法與人類在抽象推理和泛化能力上的顯著差距。
- 分析人類錯誤模式：論文對人類在 ARC 任務中錯誤的類型和模式進行了初步分析，並與機器錯誤進行了比較。結果表明人機犯錯的本質不同，這進一步強調了現有 AI 方法尚未捕捉到人類解決 ARC 問題的核心方式，為未來 AI 模型改進提供了方向。

不過他目前還缺少了部分人類表現的資料，於是他邀請了彰中的學生當受試者，並且實驗內容如下：

- 給定 3 張不同的圖片作為輸入-輸出範例
- 根據這 3 個範例，推斷出其中一種轉換方式
- 給你第 4 張測試圖片，試問其輸出結果如何

目前已知共有三種轉換方式，如下圖所示：



(arXiv: 2409.01374 [cs.AI])

不過張詩大為了照顧色盲，於是讓顏色只有 5 種，且每種顏色用數字 0、1、2、3、4 表示。並且考量不同大小的圖片，整理資料用起來很麻煩，所以將圖片大小固定為  $4 \times 4$ 。  
故上面三種轉換方式，變成如下述表示：

第一種：

給定  $4 \times 4$  的矩陣，每個元素只有 0 或 1

```
1001      0110
0110      1001
0110  -> 1001
1001      0110
```

(規則：將所有 0 替換為 1，所有 1 替換為 0，即顏色反轉)

第二種：

給定  $4 \times 4$  的矩陣，每個元素為 0 ~ 4

四個角落非為 0，其他必為 0

```
1004      0000
0000      0140
0000  -> 0230
2003      0000
```

(規則：將四個角落的非零元素移動到中心  $2 \times 2$  區域的對應位置，其他元素變為 0)

第三種：

給定  $4 \times 4$  的矩陣，每個元素為 0 ~ 4

```
0004      4000
1210      1100
0320  -> 0200
2013      3100
```

(規則：從中間對折 (折線為垂直方向)，用右邊的顏色 (除了 0) 把左邊的顏色蓋掉，之後右半邊全變成 0。)

**Mingyee** 很榮幸的受邀當他的受試者，但實際上他只想蹭獎勵金，不過亂猜會被發現而領不到錢，所以他寫一個程式來快速且精準地通過實驗。**Mingyee** 必須根據給定的三個輸入-輸出範例，推斷出其中一種轉換方式，然後將其應用於第四張測試圖片以得出其輸出結果。

## ***Input***

$c_{1,1}$   $c_{1,2}$   $c_{1,3}$   $c_{1,4}$

$c_{2,1}$   $c_{2,2}$   $c_{2,3}$   $c_{2,4}$

$c_{3,1}$   $c_{3,2}$   $c_{3,3}$   $c_{3,4}$

$c_{4,1}$   $c_{4,2}$   $c_{4,3}$   $c_{4,4}$

此為一張圖片的輸入格式。

共給七張圖片，每張圖片隔一行，前面每兩張代表一組輸入輸出結果，最後一張為測試圖片。

## ***Output***

*Ans*

輸出測試圖片其轉換後的結果。

### Sample Input

```
1101
1011
0001
0010
```

```
0010
0100
1110
1101
```

```
0000
1001
0010
0100
```

```
1111
0110
1101
1011
```

```
1010
1110
1011
1111
```

```
0101
0001
0100
0000
```

```
0110
1111
1010
1010
```

### Sample Output

```
1001
0000
0101
0101
```

### **Note**

- $c_{i,j} \in \{0, 1, 2, 3, 4\}$

### **Task**

- **task:** 100% **As statement**