

2023 年 MCM
问题 C: 预测 Wordle 的结果



背景介绍

Wordle 是**纽约时报**目前每天提供的一个流行谜题。玩家通过在六次或更短时间内猜出一个五个字母的单词来**解决**这个谜题，每次猜测都会得到反馈。在这个版本中，每个猜测必须是一个实际的英文单词。不允许猜测不被比赛认可的单词。Wordle 继续受到欢迎，该游戏现在有 60 多种语言的版本。

纽约时报网站关于 Wordle 的说明指出，在你提交你的单词后，瓷砖的颜色会改变。黄色的瓷砖表示该瓷砖中的字母在单词中，但它的位置不对。绿色瓷砖表示该瓷砖中的字母在单词中，并且在正确的位置。灰色的牌表示该牌中的字母根本不在这个词中（见附件 2）^[2]。图 1 是一个解决方案的例子，其中正确的结果是在三次尝试中找到的。

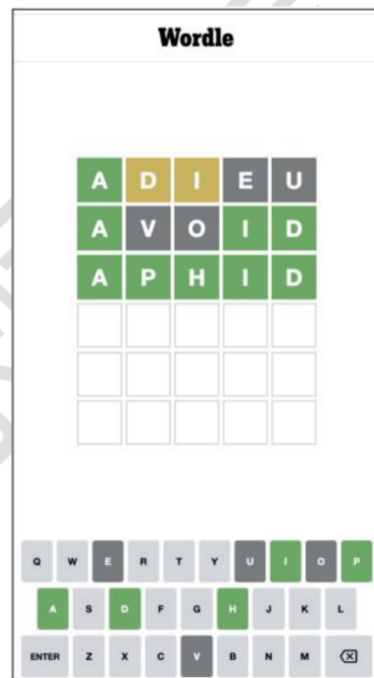


图 1: 2022 年 7 月 21 日的 Wordle 谜题的解决实例^[3]

玩家可以在普通模式或“困难模式”中进行游戏。Wordle 的“困难模式”使游戏更加困难，它要求玩家一旦在一个词中找到一个正确的字母（瓷砖是黄色或绿色的），这些字母必须在随后的猜测中使用。图 1 中的例子是在困难模式下进行的。

部分用户在 **Twitter** 上报告他们的分数。针对这个问题，MCM 生成了一个 2022 年 1 月 7 日至 12 月 31 日的每日结果文件（见附件 1）。这个文件包括日期、竞赛编号、当天的单词、当天报告分数的人数、使用困难模式的玩家人数，以及在一次、两次、三次、四次、五次、六次中猜中该单词或无法解开谜题的百分比（用 X 表示）。例如，在图 2 中，2022

年 7 月 20 日的单词是 "TRITE", 其结果是通过挖掘 Twitter 得到的。尽管图 2 中的百分比之和为 100%, 但在某些情况下, 由于四舍五入, 这可能不是真的。

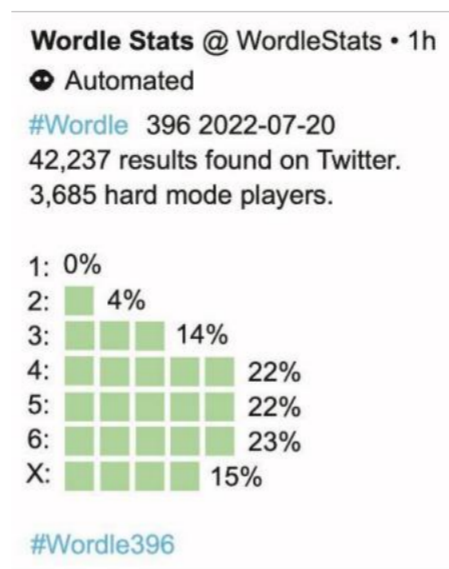


图 2: 2022 年 7 月 20 日对 Twitter 的报告结果分布^[4]

要求

纽约时报要求你对这份文件中的结果做一个分析, 回答几个问题:

- 报告结果的数量每天都在变化。建立一个模型来解释这种变化, 并使用你的模型为 2023 年 3 月 1 日的报告结果数量建立一个预测区间。词语的任何属性是否影响到报告的在困难模式下进行的分数的百分比? 如果是, 如何影响? 如果没有, 为什么没有?
- 对于一个给定的未来日期的解字, 建立一个模型, 使你能预测报告结果的分布。换句话说, 预测未来某一天的 (1、2、3、4、5、6、X) 的相关百分比。你的模型和预测有哪些不确定因素? 请举一个具体的例子, 说明你对 2023 年 3 月 1 日 EERIE 这个词的预测。你对你的模型的预测有多大信心?
- 开发并总结一个模型, 按难度对解题词汇进行分类。找出与每个分类相关的给定单词的属性。使用你的模型, EERIE 这个词的难度如何? 讨论一下你的分类模型的准确性。
- 列出并描述这个数据集的一些其他有趣的特征。

最后, 在写给《纽约时报》谜题编辑的一到两页的信中总结你的成果。

你的 PDF 解决方案的总页数不超过 25 页, 应包括:

- 一页的总结表
- 目录
- 您的完整解决方案
- 一到两页的信
- 参考文献列表

注意: MCM 比赛有 25 页的限制。你提交的所有内容都计入 25 页的限制 (摘要表、目录、报告、参考文献列表和任何附录)。你必须为你的想法、图像和报告中使用的任何其他材料注明来源。

附件

1. 数据文件 [Problem C Data Wordle.xlsx](#)

所附的数据文件包含了你应该用于此的唯一数据

问题：这个问题需要的所有信息都在问题陈述和数据文件中给出。你不需要访问纽约时报网站或 Twitter 网站。在这些网站上没有其他信息可供查找。

数据文件输入说明

日期：指的是某个 Wordle 谜题的日期，格式为 mm-dd-yyyy（月-日-年）。

比赛编号：Wordle 谜题的索引，从 2022 年 1 月 7 日的 202 开始。

字：玩家在相关的日期和比赛号码上试图猜出的解词。

报告结果的数量：当天在 Twitter 上记录的分数总数。

硬模式下的数量：当天在 Twitter 上记录的硬模式的得分数量。

1 **尝试：**玩家在一次猜测中解开谜题的百分比。

2 **尝试：**玩家在两次猜测中解开谜题的百分比。

3 **尝试：**玩家在三次猜测中解开谜题的百分比。

4 **尝试：**玩家在四次猜测中解开谜题的百分比。

5 **尝试：**玩家在五次猜测中解开谜题的百分比。

6 **尝试：**玩家在六次猜测中解开谜题的百分比。

7 **或更多的尝试 (X)：**在六次或更少的尝试中无法解开谜题的玩家的百分比。注意：由于四舍五入的原因，百分比的总和可能不一定是 100%。

2. 发表在《纽约时报》网站上的 Wordle 的方向^[2]。

How To Play

Guess the Wordle in 6 tries.

- Each guess must be a valid 5-letter word.
- The color of the tiles will change to show how close your guess was to the word.

Examples

W E A R Y

W is in the word and in the correct spot.

P I L L S

I is in the word but in the wrong spot.

V A G U E

U is not in the word in any spot.

词汇表

纽约时报：位于美国纽约市的一份日报，以印刷和网络形式出版。

推特：一个社交网站，允许用户广播不超过 280 个字符的短文（从最初的 140 个字符增加）。

解决（Wordle 谜题）：按照正确的顺序输入正确的字母，形成当天的 Wordle 单词。

参考文献

注意：我们提供以下引文来支持问题陈述，我们从这些资源中提取了重要的观点。解决这个 MCM 问题不需要这些网站上的额外信息。解决这个问题不需要访问《纽约时报》或

Twitter 网站。

- [1] Wordle logo from The New York Times website. Accessed on December 13, 2022 at <https://nytco-assets.nytimes.com/2022/08/cropped-Screen-Shot-2022-08-24-at-8.49.39-AM.png>.
- [2] "Wordle-The New York Times." The New York Times, 2022. Accessed December 13, 2022 at <https://www.nytimes.com/games/wordle/index.html>.
- [3] "Wordle-The New York Times." The New York Times, July 21, 2022.
- [4] "Wordle Stats." Twitter, July 20, 2022.

数模乐园微信公众号