Uncover the Hiddnd Secret in the Wordle Results

Summary

Since Wordle has become a popular puzzle game, it has accumulated a large amount of-data. In this paper, we define a series of metrics and build several models to explore thehidden information in Wordle results.

First, after **preprocessing** the given data and analyzing the time series diagram of thenumber of reported results, we found that the changes can be divided into 3 stages. Toforecast the number of reported results, we developed a weighted optimization modelbased on ARlMA and BP neural network. The prediction interval is then given usingthe Bootstrap method. We packaged this process as **ARIMA-BP Interval PredictionModel Based On Bootstrap**. Thus, we finally predicted the interval prediction valueobtained on March 1. 2023 at 95% confidence level to be about (19504.74, 20383.26)

Then, we defined 3 qualitative and 4 quantitative attributes of words and used them tobuild a **Multiple Linear Regression Model** with the percentage ofhard mode's playersWe found that the proportion will decrease by an average of0.618 when the initial letterchanges from a vowel to a consonant while it will increase by an average of0.017 foreach one-unit increase in word internal distance.

After that, we made the percentage distribution prediction of the reported results basedon **LSTM Model**. To ensure the percentage is around 100%, we first processed thecomponent data using a **spherical coordinate transformation**. Then we use them asoutput variables, the 7 word attributes and number ofresults as input to train our LSTMmodel. The prediction of EERlE based on this are [2%,11%, 25%,24%,19%,14%.5%. We changed the model's parameters and added noise to do **sensitivity analysis**. Meanwhile, we introduced COV to measure the uncertainty of the model predictionand found that it is around 0.4. For **error analysis**, we use MSE, RMSE and R tomeasure the prediction accuracy, and their values are shown in Table 7.

We extracted 6 indicators: RDC, TE, SK,NFC, NON, and HL to measure the dificultyof words. We built a **GMM Clustering Model** based on these indicators and thusclassifying 5 difficulty levels. We classified the word EERlE as dificulty level lll

In addition, by counting the frequency of each letter in five positions, we found S as theinitial letter has the most frequency and more specific statistical results are shown inTable 9.We also used the **Association Rule Model** based on **Apriori algorithm** tomine the word combination pattern in Wordle, ldeally, we found that the letters A.S Eand F,TL usually appear together in Wordle.

Finally, we evaluated and refined the model and reported the findings in a letter to thethe Puzzle Editor of the New York Times.
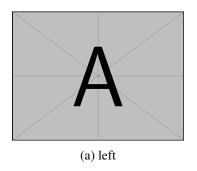
Keywords: ARIMA-BP,LSTM, GMM, Apriori Algorithm, Word Attributes

Contents

1 Introduction

1.1 Background

Wordle is an online word puzzle game invented by Josh Wardle during the epidemicThe New York Times newspaper which is well known for the games it publishes hasbought Wordle on February 2021[1]. Wordle only allows one game to be played perday, and every player in the world plays to guess the same five-letter word in six triesor less each day[2]. Players can play in regular mode or hard mode. They can sharetheir scores via Twitter, thus attracting more people to play and share
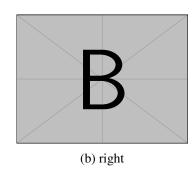
1x2,1x3,2x2 的几种组图：

(a) left   (b) right

Figure 1: Two images

Figure **??** gives an example of subfigures. Figure **??** is on the left, and Figure **??** is on the right.
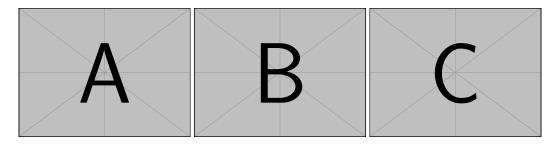

Figure 2: Three images

1.2 Literature Review

Two major problems are discussed in this paper, which are:

- Doing the first thing.
- Doing the second thing.

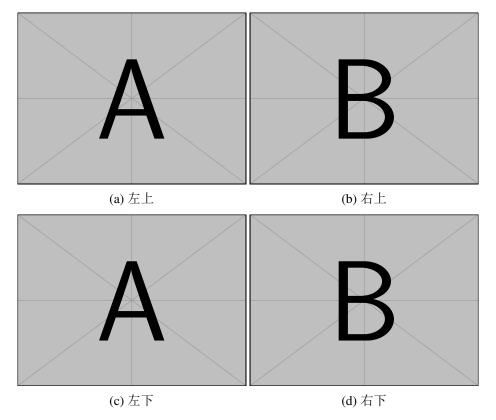A literatrue[?] says something about this problem ...

(a) 左上

(b) 右上

(c) 左下

(d) 右下

Figure 3: 组图组图变变变

1.3 Our work

We do such things ... 这部分直接上图

1. We do ...

2. We do ...

3. We do ...

2 Preparation of the Models

2.1 Assumptions and Explanations

To simplify the problem, we made the following assumptions, each of which has a corresponding reasonable explanation.

- *Assumption 1:* 假设
  ↪ *Explanation:* 理由
- *Assumption 2:* 假设
  ↪ *Explanation:* 理由

- *Assumption 3:* 假设
  ↪ *Explanation:* 理由
- *Assumption 4:* 假设
  ↪ *Explanation:* 理由

Additional assumptions are made to simplify analysis for individual sections. These assumptions will be discussed at the appropriate locations.

2.2 Notations

Table **??** lists some important mathematical notations used in this paper.

Table 1: Notations used in this paper

Symbol	Description
x_i	Longitude within the i-th Wildfire Grid
y_i	Latitude within the i-th Wildfire Grid
Ω_i	The area of the i-th grid
d_{ki}	the distance d_{ki}
SC_k	Score for evaluating the k-th wildfire grid
$x_{ki}^{(\alpha)}$	the SSA_α drone sent by the k-th EOC to the i-th wild-fire grid
$x_{ki}^{(\beta)}$	the RR_β drone sent by the k-th EOC to the i-th wildfire grid
t_{fly}^{δ}	The flight time of drones

*Some variables are not listed here and will be discussed in detail in each section.

2.3 Data

2.3.1 Data Collection

Websites, where we collect data, are listed in Table **??**.

Table 2: Notations used in this paper

Database Names	Database Websites
Google Scholar	https://scholar.google.com
Wikipedia	https://www.wikipedia.org
wolframalpha	https://www.wolframalpha.com

2.3.2 Data Processing

3 Model 1

3.1 Details about Model 1

The detail can be described by equation (**??**):

$$\frac{\partial u}{\partial t} - a^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) = f(x,y,z,t) \tag{1}$$

4 Model 2

4.1 Conclusion of Model 2

The results are shown in Figure **??**, where t denotes the time in seconds, and c refers to the concentration of water in the boiler.
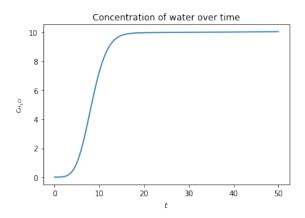

Figure 4: The result of Model 2

再来一个伪代码，默认样式为隐藏行号的三线表形式的伪代码

可在`ldmcm.sty`中修改样式，更详细的用法请参考 algorithm2e 宏包文档

Algorithm 1: How to write algorithms

Input: 输入
Output: 输出
initialization;
while *not at end of this document* **do**
    read current;
    repeat
        do these things;
    until *this end condition*;
    if *understand* **then**
        go to next section;
        current section becomes this one;
    else
        go back to the beginning of current section;
    do
        do these things;
    while *this end condition*;

4.2　Commetary on Model 2

The instance of long and wide tables are shown in Table **??**.

Table 3: Basic Information about Three Main Continents (scratched from Wikipedia)

Continent	Description	Information
Africa	Africa Continent is surrounded by the Mediterranean Sea to the north, the Isthmus of Suez and the Red Sea to the northeast, the Indian Ocean to the southeast and the Atlantic Ocean to the west.	At about 30.3 million km^2 including adjacent islands, it covers 6% of Earth's total surface area and 20% of its land area. With 1.3 billion people as of 2018, it accounts for about 16% of the world's human population.
Asia	Asia is Earth's largest and most populous continent which located primarily in the Eastern and Northern Hemispheres. It shares the continental landmass of Eurasia with the continent of Europe and the continental landmass of Afro-Eurasia with both Europe and Africa.	Asia covers an area of 44,579,000 square kilometres, about 30% of Earth's total land area and 8.7% of the Earth's total surface area. Its 4.5 billion people (as of June 2019) constitute roughly 60% of the world's population.
Europe	Europe is a continent located entirely in the Northern Hemisphere and mostly in the Eastern Hemisphere. It comprises the westernmost part of Eurasia and is bordered by the Arctic Ocean to the north, the Atlantic Ocean to the west, the Mediterranean Sea to the south, and Asia to the east.	Europe covers about 10,180,000 km^2, or 2% of the Earth's surface (6.8% of land area), making it the second-smallest continent. Europe had a total population of about 741 million (about 11% of the world population) as of 2018.

5　Model 3

6　Test the Model

6.1　Sensitivity Analysis

6.2　Robustness Analysis

这部分很重要，不能缺！

7   Conclusion

7.1   Summary of Results

7.2   Strengths

- The sensitivity analysis of the model demonstrates the effectiveness of the model under different parameter combinations and prove the robustness of the mod
- Second one ...

7.3   Weaknesses and Improvements

- The analysis of fish migration can be more accurate if we have more complete data;
- Some approximate analysis methods are applied to model the management of fishing companies, which may lead to a situation contrary to the actual one in extreme cases.

Memorandum

To:
From: Team #1887415157 of 2023 MCM
Date: February 31, 2023
Subject:

A randomly generated piece of English that has no real meaning. A randomly generated piece of English that has no real meaning. A randomly generated piece of English that has no real meaning. A randomly generated piece of English that has no real meaning.

A randomly generated piece of English that has no real meaning. A randomly generated piece of English that has no real meaning. A randomly generated piece of English that has no real meaning. A randomly generated piece of English that has no real meaning. A randomly generated piece of English that has no real meaning. A randomly generated piece of English that has no real meaning. A randomly generated piece of English that has no real meaning. A randomly generated piece of English that has no real meaning. A randomly generated piece of English that has no real meaning. A randomly generated piece of English that has no real meaning. A randomly generated piece of English that has no real meaning. A randomly generated piece of English that has no real meaning. A randomly generated piece of English that has no real meaning. A randomly generated piece of English that has no real meaning.

> ## Note
>
> 1. Text Text Text Text
> 2. Text Text Text Text Text Text Text Text Text
> 3. Text Text Text Text Text Text Text Text Text Text Text Text
> 4. Text Text Text Text

This part must be a bit of an unordered list or an ordered list or something like that, otherwise it's really all text and will look super ugly.

- Solution 1. **Build more shopping centers**. Explain solution1 explain solution1 explain solution1 explain solution1 explain solution1 explain solution1.
- Solution 2. **Build more shopping centers**. Explain solution2 explain solution2 explain solution2 explain solution2.
- Solution 3. **Build more shopping centers**. Explain solution3 explain solution3 explain solution3in solution1 explain solution1 explain solution1.
- Solution 4. **Build more shopping centers**. Explain solution4 explain solution4 explain solution4 explain solution4 explain solution4 explain solution4 explain solution4.

A part summarizes[?, ?] the words of the table above. Ura Ura Ura Ura Ura Ura Ura Ura. Aba Aba Aba Aba Aba Aba Aba Aba Aba Aba Aba Aba Aba Aba Aba Aba Aba. Blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah. Blah blah blah blah blah blah blah blah blah blah blah blah.

Conclusion. This part of writing conclusive things, one or two sentences to summarize it, do not write too much.

References

[1] Hao H, Wang Y, Xia Y, Zhao J, Shen F. Temporal convolutional attention-based network for sequence modeling. arXiv preprint arXiv:2002.12530. 2020Feb 28.

[2] Harvey AC. 1990 Forecasting, structural time series models and the kalman filter. Cambridge, UK: Cambridge University Press.

[3] Krizhevsky A, Sutskever I, Hinton GE. 2012 ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems 25 (NIPS) (eds F Pereira, CJC Burges, L Bottou, KQ Weinberger), pp. 1097–1105.