

Research on Predicting Wordle Results

Qiuping Zhong*, Mingfei He

Liberal Arts College, Guangdong Lingnan Institute of Technology, Guangzhou, China, 510663

*Corresponding author: qiu609@126.com

Abstract. Wordle is a very popular puzzle game, but the number of players is decreasing. In order to reduce this phenomenon, this paper intends to solve it through relevant methods, such as ARIMA model, Credibility, Multicollinearity model; Ridge regression algorithm; Fuzzy synthetic evaluation method. First, the data in the game is cleaned up, mainly including the modification of the number of letters, the number and percentage of special letters in the word competition result report, etc. Secondly, the use of the establishment of the relevant model, and to be analyzed related factors and circumstances, to obtain clear results. It is worth noting that the model itself has certain advantages and disadvantages. Finally, based on the above steps, put forward four feasible suggestions, hope that the game can be adopted, the game complement.

Keywords: Wordle, ARIMA model, Multicollinearity model, Fuzzy synthetic evaluation method.

1. Introduction

1.1. background

Wordle is a popular puzzle currently offered daily by the New York Times. Wordle continues to grow in popularity and versions of the game are now available in over 60 languages.

The rules of the wordle game are as follows:

Players try to solve the puzzle by guessing a five-letter word in six tries or less, receiving feedback with every guess.

Each guess must be an actual English word.

The color of the tiles will change after you submit your word. A yellow tile indicates the letter in that tile is in the word, but it is in the wrong location. A green tile indicates that the letter in that tile is in the word and is in the correct location. A gray tile indicates that the letter in that tile is not included in the word at all.[1]

If the sixth tries could not solve the puzzle means the game fails (indicated by X).

Players can play in regular mode or "Hard Mode". Wordle's Hard Mode makes the game more difficult by requiring that once a player has found a correct letter in a word (the tile is yellow or green), those letters must be used in subsequent guesses.[2]

1.2. Assumptions and Justifications

In order to facilitate solve problems, we make the following assumptions and simplifications according to the actual situation:

Assumption 1: Wordle players can use dictionary.

Justifications: The wordle game time is unlimited and open. There are only six opportunities to challenge every day. So dictionary data, including word or letter frequency, is allowed.

Assumption 2: The sum of percentages allows an error of plus or minus 2.

Justifications: It due to rounding.

Assumption 3: The percentage is calculated using the number of reported results.

Justifications: Wordle is divided into regular mode or "Hard Mode." So number of reported results is the sum to number in regular mode and number in hard mode

2. Data cleaning

Data cleaning 1: Modifying two four-letter words to five letters. We found the word listed are "tash" and "clen". They're all four letters, so these two words are wrong. The way we deal with it is to assume that the word contains these four letters, and then pick a five-letter word in the dictionary that contains those four letters. Finally, choose the right word according to the principle that the order of the four letters remains the same or changes the least. Based on the above method. There are 12 five-letter words with "clen" in them, we change "clen" to "clean". There are 52 five-letter words with "tash" in them, we change "tash" to "stash".

Data cleaning 2: Modifying special letters in words. Because the words in wordle are made up of 26 letters, but we found that the special word had the special letter " ï " in it. Modify " ï " to " i ", and the word naive is a correct word. So, we modify "naïve" to "naive".

Data cleaning 3: Modifying the number of reported results. We found the number of reported results is 2569, this is obviously wrong. Since interpolation method is simple and the prediction accuracy is high. So we use spline interpolation method to predict the number of reported results. Let Contest number and Number of reported results as two variables, using Mat-lab to calculate the predicted value, the result is 21935.

Data cleaning 4: Modifying the Percent of contest number 281. We sum the 7 data (percent) for each day. The result of contest number 281 is 126. This is obviously wrong. Therefore, we use a proportional method to modify the data.

3. Build the ARIMA model and Analysis the prediction result

3.1. Build the ARIMA model

The steps for modeling with ARIMA[3] are as follows (see Figure 1):

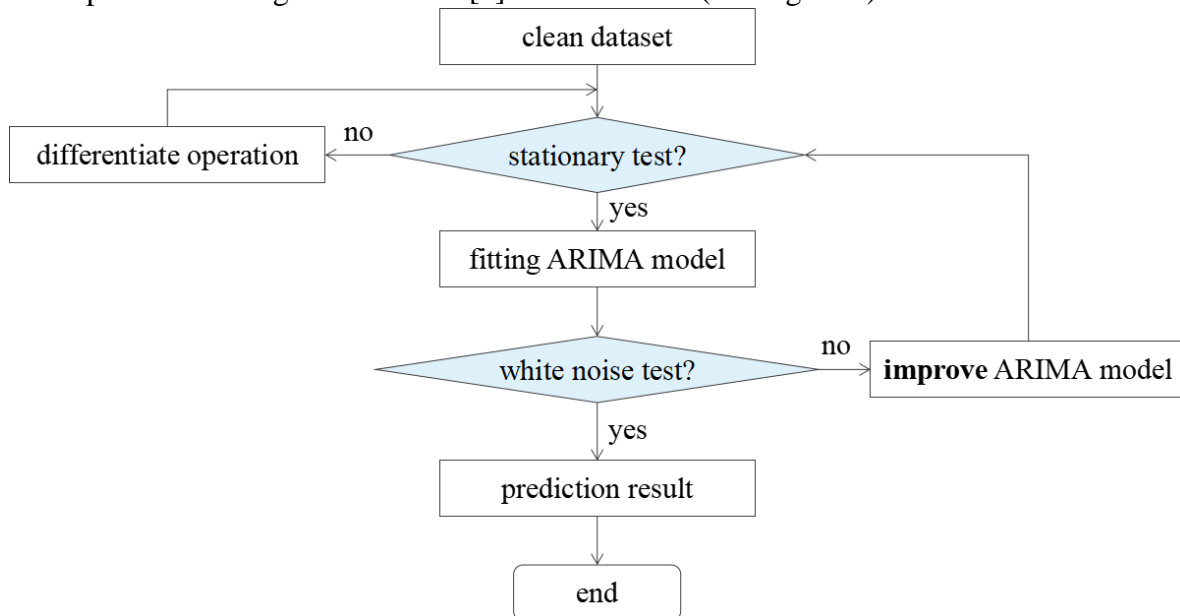


Figure 1: The steps for modeling with ARIMA

3.2. Analysis the prediction result

The model parameters were calculated using SPSS (see Table 1).

Table 1: Model parameter

	coefficient	std	t	P> t	0.025	0.975
constant	-168.296	463.997	-0.363	0.717	-1077.713	741.121
ar.L1.D.Number	-0.362	0.049	-7.33	0	-0.459	-0.265

Remarks: ***, ** and * respectively represent the significance level of 1%, 5% and 10%. Therefore, the model of ARIMA (1,1,0) is

$$\hat{y}_t = 168.296 - 0.362y_{t-1} \tag{1}$$

Goodness of fit R²=0.982, and the fitting figure 2, both indicate that the fitting effect is very good..

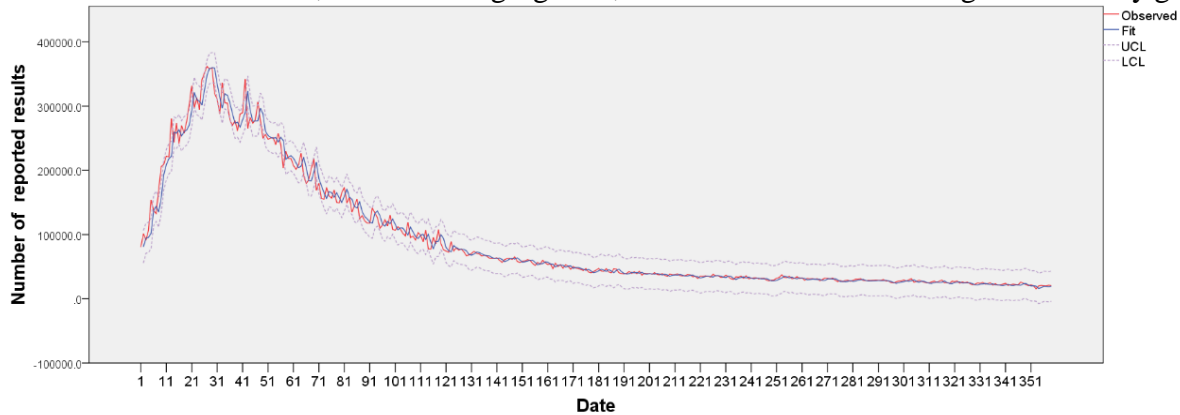


Figure 2: Fit graph of ARIMA

Because the Contest number on March 1, 2023 is 620. Substituting t=620 into the model (Equation 1), the calculated predicted value is 10456.

But the predicted value is subject to error, to improve the prediction accuracy, formula (1) can be used to calculate the prediction interval.

$$CI = \hat{y}_t \mp Z \cdot SE \tag{2}$$

Where, CI is the confidence interval, Z is the quantile of the standard normal distribution and SE is the standard error.

Put Z=1.96, SE=463.997 and $\hat{y}_t = 10456$ into the calculation and the prediction interval is [9546.6, 11365.4].

In addition, we also predicted the number of reported results for the contest No. 561-619 (see Figure 3).

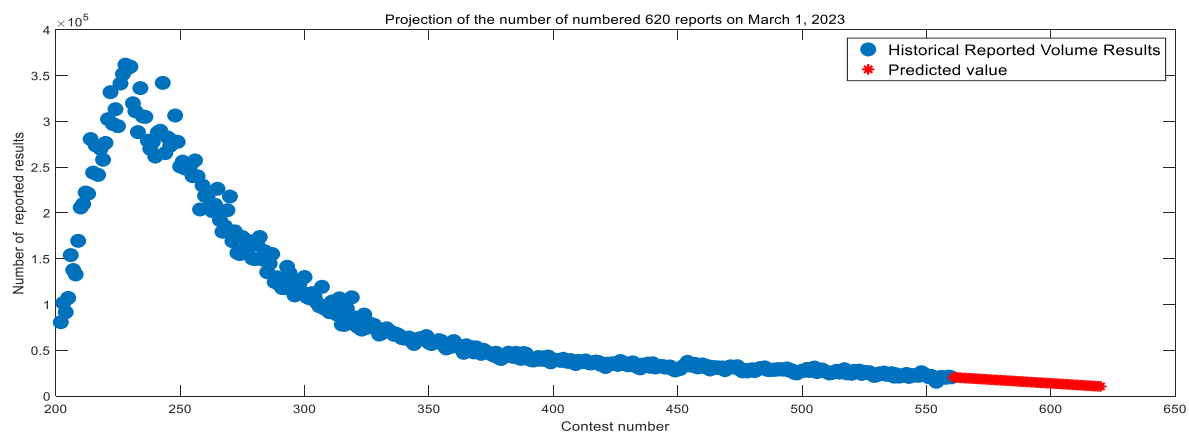


Figure 3: Predicted value for contest No. 561-620

4. Multicollinearity model predict percentage

4.1. Build Multicollinearity model

We need to predict the percentage, so let percentage be the dependent variable p, where one try, two tries, three tries, four tries, five tries, six tries and could not solve the puzzle is denoted as pi (i=1,...,7). [4]Because there is a precise correlation between percentages as follow:

$$\sum_{i=1}^7 p_i = 1 \quad (3)$$

Sometimes the sum to 100% is not be true due to rounding.

According to the data, the contest number and number of reported results are assumed as dependent variables x_1 and x_2 .

The above is in compliance with Multicollinearity. Therefore, we can set up Multicollinearity model.

$$\begin{cases} \hat{Y} = \beta X + \varepsilon \\ \sum_{i=1}^7 p_i = 1 \end{cases} \quad (4)$$

Where, \hat{Y} is the seven prediction target variables, X is the matrix of the arguments x_1 and x_2 . β is the coefficient vector, and ε is a random error vector.

4.2. Ridge regression algorithm solve the model

Ridge regression algorithm is a linear regression algorithm that prevents overfitting by limiting the size of the regression coefficient. It makes the regression coefficient tend to be stable and smooth in the training process, and reduces the correlation between the features. Ridge regression algorithm performs well on features with high correlation and small sample data sets.[5]

4.3. Explore interesting features

We found some interesting features that we hope will help players and game developers to better experience the game.

Firstly, we divided historical wordle words into letters and counted the frequency of 26 letters. The most frequent 26 letters are e,a,o,r and t. So, the 1 try we play the wordle, we try to guess the word with those letters in it. This increases the success rate and reduces the number of games played.

Secondly, the proportion of players in hard mode increased. This leads to a decrease in the success rate of the game.

Finally, based on the previous analysis, we found that wordle had a high success rate, up to 97.1% for the most 6 times. Use elimination to finish the game, can improve the game pass rate, even if your luck is very bad.

First guess five words with different letters, assuming you don't guess a single letter the first four times, 20 letters can be eliminated, leaving only six letters to choose. Then classify and discuss words according to different letter numbers. Let's take four different letters, the " $\sqrt{\quad}$ " in the figure means the letters are correct, including out of order. According to the elimination method, you can guess at least three letters correctly on the fifth time. Take the remaining letter and guess the correct three letters, combine them with the position order, and you will find the correct word, because the word that meets these conditions is almost always unique. That's why the accuracy is so high.

4.4. Fuzzy synthetic evaluation method to assess the difficulty of words

According to the comprehensive evaluation index and evaluation elements of words, Take word features, for example, This is a two-level fuzzy comprehensive evaluation and analysis model, Determine the evaluation factor set.[6]

When we evaluate the attributes of words, Divide it into different levels. Because the complexity of a word affects the rate at which people play a game, Therefore, from the perspective of cognitive level, we classify words into three levels: easy, medium, and difficult.[7]

The characteristics of each word were extracted and manual screening was adopted. Put them into each evaluation result and count the frequency obtained by each evaluation result. According to each

evaluation index, it is divided into the second-level evaluation elements, and the membership degree of each evaluation result is obtained by using the formula.[8]

First level indicators are the three indicators of the complexity of word construction. The weight of the second-level index is based on the frequency of letter occurrence, the degree of letter repetition and the percentage of passing through the game.[9]

5. Conclusions and Suggestions

5.1. Conclusions

When building the ARIMA model to predict the number of reported results, we found that the number of reported results decreased slowly with time.

When we analyze the impact of word attributes on players, we find that word attributes have little effect on success rate, only on the speed of success.

The Multicollinearity model is established to predict the percentage with high reliability, and the passing rate of the game is found to be 97%, indicating that the game is not difficult.

A fuzzy comprehensive evaluation model is established based on word attributes to distinguish the difficulty level of words. We found that there were not many difficult words, and most of them succeeded on the fifth or sixth time, which was a high success rate, indicating that the game was not difficult.[10]

5.2. Strengths and Weaknesses

(1)Strengths. Firstly, the prediction accuracy of time series model is high. Secondly, Multicollinearity model predict percentage has high credibility. The calculation speed of Ridge regression algorithm is fast. The reliability of predicting EERIE is 46.4%. Thirdly, the comprehensive evaluation model considers comprehensive factors and determines three difficulty levels for the game. The difficulty of identifying "EERIE" is medium, which is consistent with the predicted value in Table 10. It shows that the classification model is highly accurate. Finally, the analysis of word attributes comprehensively found three interesting features. And based on our findings, provide useful advice to gamers and developers.

(2)Weaknesses. On the one hand, the model accuracy is not high enough, and it does not go deep enough when studying credibility. On the other hand, due to time constraints, no ensitivity Analysis was performed, We will continue to study this topic after the game.

5.3. Suggestions

Firstly, random words every day, make the puzzle different for each player , thus reduce plagiarism, improve the loyalty of the game.

Secondly, increase the difficulty of the game, attract and retain more players, and keep wordle popular. such as 6 letter words, Limit the length of the game to avoid spending a lot of time looking up the dictionary, which will affect work and study.

Thirdly, make the game more interesting. such as add educational learning elements, after finishing the game, there are learning knowledge points or story related to the words, making the game more practical and valuable.

Finally, add a challenge mode to keep older players enthusiastic.

References

- [1] "Wordle-The New York Times." The New York Times, 2022. Accessed December 13, 2022 at <https://www.nytimes.com/games/wordle/index.html>.
- [2] SPSSPRO. You can follow the ARIMA model tutorial without the knowledge of statistics [D], https://zhuanlan.zhihu.com/p/440372527?utm_id=0.

- [3] Yang Haimin, Pan Zhisong, Bai Wei. A survey of time-series pre-test method [J]. Computer Science, 2019, 46(01): 21-28.
- [4] Yao Jinhai, Zou Jiajun. SVM-ARIMA model construction and numerical simulation of CPI prediction [J]. statistics and decision, 2022, 38(21): 48-52.
- [5] Gao Guanyin, Cao Jingwei, Yu Siyang, et al. Research on quantitative evaluation method of network security risk based on fuzzy comprehensive evaluation theory [J]. Post and Telecommunications Design Technology, 2023(08): 71-74.
- [6] Bai Ping, Zhang Li. Study on the evaluation of standards implementation effect based on fuzzy comprehensive evaluation method-local standards in the field of safety production (chemical industry) [J]. Safety, 2023, 44(07): 60-66.
- [7] Wang Zhiling, Li Yanan. Evaluation and selection of plastic suppliers based on fuzzy comprehensive evaluation method [J]. Modern Commerce, 2023(13): 24-27.
- [8] Xie Donghui, Rui Huang, Liang Bingrong, et al. Study on the quality evaluation of prefabricated buildings based on fuzzy comprehensive evaluation [J]. Urban Architecture, 2023, 20(10): 204-207.
- [9] Zhu Lianbo, Wang Tingting, Shen Lu, et al. Research on performance evaluation of government investment project construction management-based on combination right and fuzzy comprehensive evaluation method [J]. Engineering Construction, 2023, 55(04): 59-66.
- [10] Ji Mian, chenglong. Application of Ridge Regression Algorithm in Ozone Concentration Ensemble Forecast [J]. Journal of Anhui University (Natural Science Edition), 2018,42(04): 93-102.