Prediction and Analysis of the Future Development Trend of Wordle

Yixiao Wang[1*], Wenhao He[1a], Yuxuan Wang[2b]

[*] Corresponding author: wangyixiaoleo@qq.com
[a] email: henwenhao0814@163.com, [b] email:1764635058@qq.com

[1] College of Engineering, China Agricultural University, Beijing, China 100083;
[2] College of Information and Electrical Engineering, China Agricultural University, Beijing, China 100083

Abstract: Wordle is a popular puzzle game provided by New York Times. Exploring the factors that affect games makes sense for the future development of Wordle. We mainly study the future trend of the number of reported results and how to predict the percentage of scores. First, we set up the propagation model of Wordle with the reference of the model of epidemic diseases. Then we use particle swarm optimization algorithm to optimize the unknown parameters, we use the adjusted iterative model to simulate the changes of the number of Wordle reports. Moreover, in order to probe whether any attribute of the word will affect the percentage of difficult mode records, we selected several important attributes including parts of speech and frequency to label data and conducted correlation analysis on them. Finally we analyze the impact of the word attributes on the percentage in difficult mode. Due to the need to predict the percentage scores for a future date, we added the change of time into the prediction model and build the regression model with time variables containing contest number, the day of the week, if the day is weekend and word attributes variables containing part of speech and frequency as independent variables. Finally, we get the coefficient relationship between different tries and these variables, and regression prediction equations were constructed. Finally, We predicted the score percentages of word 'EERIE' at 2023/3/1 and we also judge the accuracy of the model by testing it with MAE and MSE.

Keywords: Wordle, Prediction, Model of epidemic diseases, Correlation analysis, PLS regression model

1. INTRODUCTION

Wordle is a popular puzzle game provided by the ***New York Times***. At present, the game version has more than 60 languages. For the rules of the game, players have to guess a five-letter English word six times or less to solve the puzzle, and will be prompted after each attempt. The color of the tiles will change after submitting the word. Yellow indicates that the letter exists in the correct answer, but the position is wrong; Green indicates that the letter exists in the correct answer and the position is correct; Gray indicates that the letter does not exist in the correct word. Of course, every guess we make must be a real English word. At the same time, players can choose regular mode or "Hard Mode". In the "Hard Mode", the letters in the green or yellow tiles must be used in the following attempts.

Based on the information we searched, we obtained the daily report results and the percentage of different daily scores from 2022.1.1 to 2022.12.31. Using these data, we could predict the future results and different score percentage, and try to find out whether various factors such as the attributes of the words

MSEA 2023, May 26-28, Nanjing, People's Republic of China
Copyright © 2023 EAI
DOI 10.4108/eai.26-5-2023.2334422

and the future results and different score percentage, which have guiding significance for the development and improvement of future games.

2. METHODOLOGY AND DATA

Combined with the actual situation, in order to simplify the model, we made the following assumptions. And the whole paper is based on the following assumptions.

- Assume that there are 10 million people involved in this system, who may become players. Define the amount of all users as $N_{\text{all users}} = 10000000$.
- Assume that 15% of people will share their results on Twitter after finishing the game of the day, which form score records. Define the probability of sharing as $P_{\text{twitter}} = 0.15$.
- Assume that a twitter can affect 5 people. Define the spreading rate as $R_{\text{spread}} = 5$.
- Assume that the probability of players losing interest in the game is 10%. Define the probability of leaving the game of $P_{\text{leave}} = 0.1$.
- Ignore all influencing factors except for time and word attributes.
- Set *adjectives and determiners*, *nouns* and *verbs* as three categories of words, and assume that parts of speech only contain these three categories.

2.1 Infectious disease model

This model is set up with the reference of epidemic disease model (**Fig. 1**)[1,2]. All users are divided into new users, players and old users. When the Wordle game has not yet appeared on Twitter, every user is a new user (Corresponding to the infectious disease model, all people have been infected with the virus but have not yet become ill).

We can gain the formulas below from Figure 1 easily:

$$N_{\text{players}}' = N_{\text{players}} + N_{\text{new players}} - N_{\text{old players}} \tag{1}$$

$$N_{\text{new users}}' = N_{\text{new users}} - N_{\text{new players}} + N_{\text{refreshed users}} \tag{2}$$

$$N_{\text{old users}}' = N_{\text{old users}} + N_{\text{old players}} - N_{\text{refreshed users}} \tag{3}$$

In our model, people will constantly change their types under the influence of Twitter. Twitter may make a user who has not played the game start to play, or make a user who has been playing the game regain interest. Players who are playing games will post tweets, increasing the number of game related tweets, resulting in an expansion of the influence of the game in turn.

Twitter corresponds to the virus in the infectious disease model, and its strength is determined by the number of players. Assuming that the probability of everyone sharing to Twitter after finishing the game is P_{twitter}, the formula for calculating the number of tweets N_{tweets} is:

$$N_{\text{tweets}} = N_{\text{players}} \cdot P_{\text{twitter}} \tag{4}$$

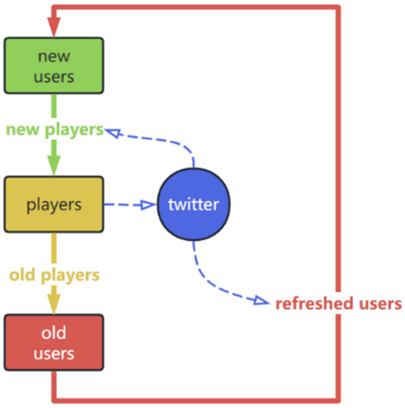


Figure 1. Schematic of the model of epidemic diseases

Assuming that the number of people affected by one single Twitter is R_{spread}, the formula for calculating the number of people affected by Twitter is:

$$N_{affected\ users} = \min\{N_{tweets} \cdot R_{spread}\ ,\ N_{all\ users}\} \tag{5}$$

In each iteration:

Some new users are affected by tweets of the game, becoming new players and join players (corresponding to the symptoms' appearance of infected people). Assuming that the probability of a new user becoming a new player is $P_{new\ player}$, the calculation formula of the number of new players is:

$$N_{new\ players} = N_{affected\ users} \cdot \frac{N_{new\ users}}{N_{all\ users}} \cdot P_{new\ player} \tag{6}$$

Some players may get bored with the game, becoming old players and join old users (corresponding to the rehabilitation of patients). Assuming that the probability of players losing interest is P_{leave}, the calculation formula of the number of old players is:

$$N_{old\ players} = N_{players} \cdot P_{leave} \tag{7}$$

Affected by tweets, some old users become interested in the game again. They become refreshed users then join new users (corresponding to the second-time virus infection after recovery). Assuming that the probability of old user becoming new user is $P_{old\ player}$, the formula for refreshed user number is:

$$N_{refreshed\ users} = N_{affected\ users} \cdot \frac{N_{old\ users}}{N_{all\ users}} \cdot P_{old\ player} \tag{8}$$

To sum up, the model can be represented by the following function with 7 parameters:

$$\text{Number of reported results}_{prediction} = F(N_{all\ users}, P_{twitter}, R_{spread}, P_{new\ player}, P_{old\ player}, P_{leave}, \text{Contest number}) \tag{9}$$

Use the earliest data to initialize the model. The number of reported results on 2022/1/7 is 80630. According to the model assumption, the number of players $N_{\text{all users}}$ on this day is:

$$N_{\text{players}} = \frac{N_{\text{tweets}}}{P_{\text{twitter}}} = 80630 \div 15\% \approx 537533 \tag{10}$$

Adjust parameters $P_{\text{new player}}$ and $P_{\text{old player}}$ in the model to fit the actual data. After simple manual adjustment, we determined the range of the remaining two parameters:

$$P_{\text{new player}} \in [0.1, 0.4] \tag{11}$$

$$P_{\text{old player}} \in [0.1, 0.4] \tag{12}$$

Use the particle swarm optimization algorithm to optimize parameters. Take the error between the predicted value and the actual value as the objective function. Let N be a function with Number of reported results as the dependent variable and Contest number as the independent variable

$$O(P_{\text{new player,}} P_{\text{old player}}) = \text{w} = \sum_{n=202}^{560} \frac{|F(P_{\text{new player,}} P_{\text{old player,n}})\text{-}N(n)|}{N(n)} / 359 \tag{13}$$

Set the number of particle attributes to 2, the number of particles to 3, the range of particle position to [10, 40], the maximum speed to 150, the social cognition-coefficient to 0.1, and the self-cognition coefficient to 0.8.

The formula for calculating $P_{\text{new player}}$ and $P_{\text{old player}}$ from particle position is:

$$P_{\text{new player}} = \text{Particle Position.x}_1 \div 100 \tag{14}$$

$$P_{\text{old player}} = \text{Particle Position.x}_2 \div 100 \tag{15}$$

After 100 iterations, the optimal solution is as follows:

$$P_{\text{new player}} = 0.2853188482 \tag{16}$$

$$P_{\text{old player}} = 0.2419533705 \tag{17}$$

$$O = \omega = 0.25199217129919005 \tag{18}$$

The comparison between the prediction model and the real data is plotted in **Fig.2a**, and the change of the objective function value with the number of iterations during the iteration process is plotted in **Fig 2b**.

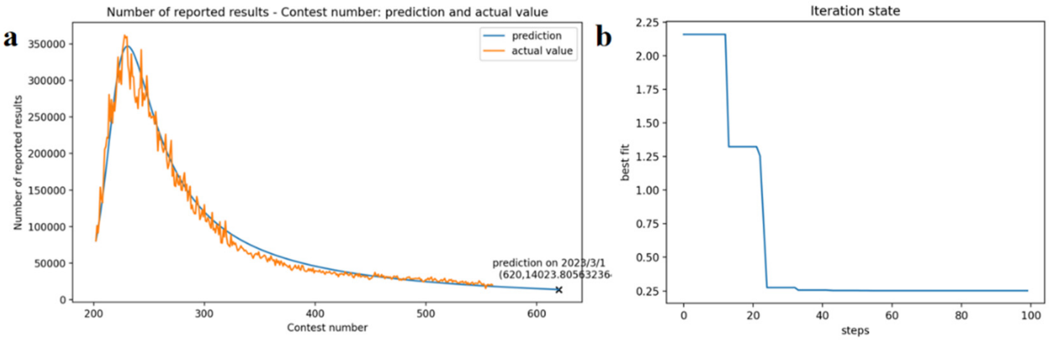


Figure 2. a)The comparison between the prediction model and the real data, b)The change of the objective function value with the number of iterations

2.2 Error analysis

Since the model is used to predict the data of 2023/3/1, the fitting accuracy of the data with a later date is given priority. The error between the model and the data in the last two months is more suitable when evaluating the prediction error. The above error calculation method is still used, and the relative error is:

$$\Omega = 0.1019328447627061 \tag{19}$$

Using the relative error above to predict interval of the number of reported results on 2023/03/01, the interval is:

$$\left[\frac{F}{w+1}, \frac{F}{1-w} \right] \tag{20}$$

In other words, out prediction of *number of reported results* on 2023/3/1 is:

$$14023.805632364674 \pm 2287.178244009907 \tag{21}$$

2.3 Explanation for changing

We can roughly divide the change of number of reported results over time into three stages: rising period, decline period and stable period (**Fig. 3a**).

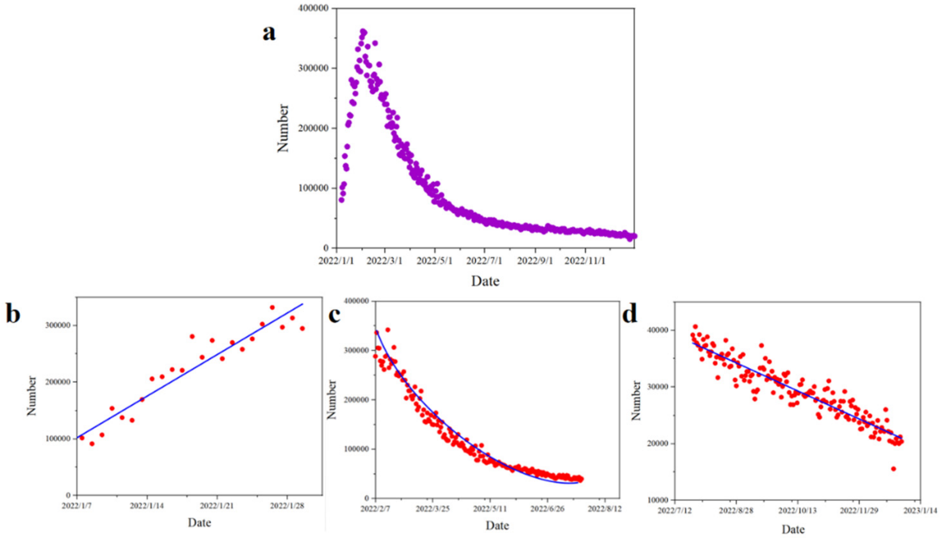


Figure 3. a)The relationship diagram of the number of reported results and date, b,c,d) Curve fitting

At first, new users accounted for the majority of the total users. Under the condition of the same spread rate of R_{spread}, tweets of the game can affect more people than in the later period. On the whole, the number of reported results is growing rapidly in this period (**Fig. 3b**).

As time goes by, the number of players $N_{players}$ increases, and the number of tweets N_{tweets} also increases, which plays a positive role in the rising speed; At the same time, the number of new users is getting smaller and smaller, which has a negative effect on the rising speed. Under the joint action of the two above, the growth of number of reported results is approximately linear.When the number of reports reaches the maximum, the number of new users is too small to maintain the growth trend, and the number of reports begins to decline(**Fig. 3c**). In the process of decline period, $N_{players}$ will decrease, and N_{tweets} will also decrease, resulting in a decrease in the influence(viral infectivity) of the game. Therefore, at the beginning of the decline period, the number of reports decreased faster and faster. Over time, some old users became new users again and began to re-participate in the game. The trend of decrease in the number of reports eased gradually. The data change during the reduction period is approximately exponential.

When the number of users in several groups gradually stabilized, the data changes began to become flat. Under the dynamic balance, the data change in the stable period is also approximately linear(**Fig. 3d**).

The number of reported results in the stable period is linearly fitted, and the analytical formula is:

$$\widehat{Y}=-106.49x+80591 \tag{22}$$

Use this function and its error between the real data and itself to predict the data of 2023/3/1, and get:

$$14567.2\pm1334.61 \tag{23}$$

It can be seen that there is not a big difference between this result and the established infectious disease model above (**Fig. 4**), which verifies the correctness of our model.

The disturbance of the data around the average value is affected in many ways, such as whether there is any major news to divert people's attention on the day, the amount of advertising push, weather factors, whether it is a holiday, and so on.

For example, the figure below shows that the number of reported results suddenly increased slightly when the number of Contest was around 450, which may be due to the promotion of the game made unintentionally by a piece of news visible to the public.

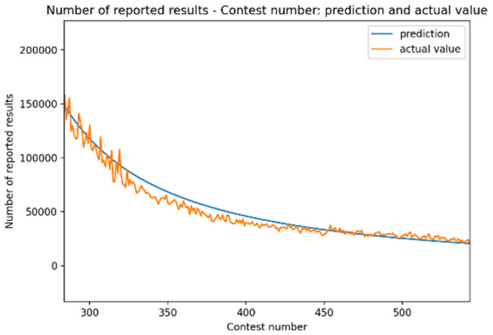


Figure 4. The diagram of prediction and actual value

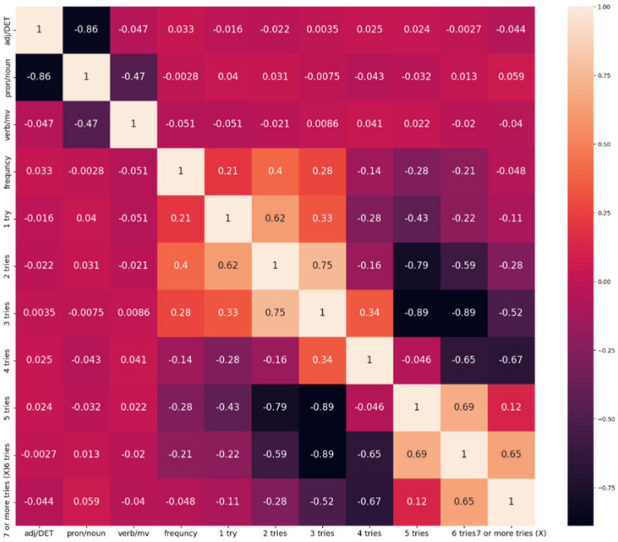
2.4 Correlation Analysis

Different words have different attributes, so there are different frequency and the use of scenes in daily life, which will cause people to guess the word with subconsciously using high frequency of vocabulary or common nouns and verbs. However, some rare prepositions or adverbs and words with unusual letters will not be given priority. This phenomenon can also make people score differently every day when playing the Wordle in hard mode. That is to say, different words will have different degrees of difficulty.

Therefore, we use different attributes to annotate the words given in the Appendix, in which we only focus on the two important attributes, part of speech and frequency. The part of speech includes adjectives/adverbs/determiners, nouns/pronouns and verbs/modal verbs. For frequency, we counted the frequency of each letter in the daily word. The sum of the frequency represents how frequent the word is. The correlation analysis of times of attempts, speech and frequency can generate heat map[3] as **Fig. 5**

This heat map reflects the correlation of three types of speech and frequency with different times of attempts. It can be known that there is a large correlation gap between the two attributes.

Since the correlation coefficient between part of speech and almost all times of attempts is extremely small, we can infer that part of speech has little effect on the proportion of different times of attempts. However, there is some correlation between frequency and times of attempts due to the slightly large correlation coefficient.



	adj/DET	pron/noun	verb/mv	frequncy	1 try	2 tries	3 tries	4 tries	5 tries	6 tries	7 or more tries (X)
adj/DET	1	-0.86	-0.047	0.033	-0.016	-0.022	0.0035	0.025	0.024	-0.0027	-0.044
pron/noun	-0.86	1	-0.47	-0.0028	0.04	0.031	-0.0075	-0.043	-0.032	0.013	0.059
verb/mv	-0.047	-0.47	1	-0.051	-0.051	-0.021	0.0086	0.041	0.022	-0.02	-0.04
frequncy	0.033	-0.0028	-0.051	1	0.21	0.4	0.28	-0.14	-0.28	-0.21	-0.048
1 try	-0.016	0.04	-0.051	0.21	1	0.62	0.33	-0.28	-0.43	-0.22	-0.11
2 tries	-0.022	0.031	-0.021	0.4	0.62	1	0.75	-0.16	-0.79	-0.59	-0.28
3 tries	0.0035	-0.0075	0.0086	0.28	0.33	0.75	1	0.34	-0.89	-0.89	-0.52
4 tries	0.025	-0.043	0.041	-0.14	-0.28	-0.16	0.34	1	-0.046	-0.65	-0.67
5 tries	0.024	-0.032	0.022	-0.28	-0.43	-0.79	-0.89	-0.046	1	0.69	0.12
6 tries	-0.0027	0.013	-0.02	-0.21	-0.22	-0.59	-0.89	-0.65	0.69	1	0.65
7 or more tries (X)	-0.044	0.059	-0.04	-0.048	-0.11	-0.28	-0.52	-0.67	0.12	0.65	1

Figure 5. The heat map

When times of attempts are less than 4, the frequency and times of attempts are positively and strongly correlated, even reaching 0.4 at 2 times of attempts; When times of attempts are greater than or equal to 4, the frequency and times of attempts are negatively correlated, and the correlation is slightly less than the low times of attempts.Compared with high times of attempts, the frequency mainly affects the proportion of low times of attempts. To minimize this effect, we can use the model to evaluate the daily word, and make the word attribute weighted scores equal as much as possible.

2.5 PLS regression model

To predict the distribution of the reported results on future dates with a solution word given, we need to build a model that takes both time and word attributes into account. Based on the word attribute quantification in the first question, we built a PLS regression model[4,5,6,7,8] for prediction.

In this model, we choose three time variables containing contest number, the day of the week and if the day is weekend; Word attributes variables contains three parts of speech and frequency. Since we have to consider both time and word attributes, we can build a sequence of seven independent variables:

$$\mathbf{X}=(x_1 ,\ x_2 ,\ x_3 ,\ x_4 ,\ x_5 , x_6 , x_7) \tag{24}$$

Constitute the original independent variable data matrix:

$$\mathbf{X}=[x(1), x(2),\ldots\ldots, x(n)]^{T} \qquad n=1, 2, \ldots\ldots, 359 \tag{25}$$

Similarly, using times of attempts to build sequences of seven dependent variables:

$$\mathrm{Y}=\left(y_1 ,\ y_2 ,\ y_3 ,\ y_4 ,\ y_5 , y_6 , y_7\right) \tag{26}$$

Constitute the original dependent variable data matrix:

$$\mathrm{Y}=[y(1), y(2),\ldots\ldots,y(n)]^{T} \qquad n=1, 2, \ldots\ldots, 359 \tag{27}$$

The first pair of linear combination components *t* (1) and *u* (1) are extracted, where w and v are both weight vectors

$$\mathrm{T}(1)=\mathrm{w}_1^{\mathrm{T}}\mathrm{x}(1)^{\mathrm{T}} \quad \mathrm{w}_1=(\mathrm{w}_{11},\cdots,\mathrm{w}_{15}) \tag{28}$$

$$\mathrm{U}(1)=\mathrm{v}_1^{\mathrm{T}}\,\mathrm{y}(1)^{\mathrm{T}} \quad \mathrm{v}_1=(\mathrm{v}_{11},\cdots,\mathrm{v}_{17}) \tag{29}$$

Both t (1) and u (1) should extract the difference information of the variable matrix as much as possible, and the correlation should reach the maximum. Using the *X* and *Y* matrices, we can calculate the first pair of component score vectors:

$$\widehat{\mathrm{T}}(1)=\mathbf{X}\mathrm{w}_1 \tag{30}$$

$$\widehat{\mathrm{U}}(1)=\mathbf{Y}\mathrm{v}_1 \tag{31}$$

Establish the regression of *y* to *t* (1) and *x* to *t* (1) respectively:

$$\begin{cases}\mathbf{X}=\hat{\mathrm{t}}(1)\alpha_1^{\mathrm{T}}+\mathbf{X_1}\\ \mathbf{Y}=\hat{\mathrm{u}}(1)\beta_1^{\mathrm{T}}+\mathbf{Y_1}\end{cases} \tag{32}$$

Where *α* and *β* are both parameter vectors with least squares estimates as follows, and *X1* and *Y1* are residual matrices respectively.

$$\begin{cases}\mathrm{A}_1=\mathbf{X}^{\mathbf{T}}\dfrac{\hat{\mathrm{t}}(1)}{\|\hat{\mathrm{t}}(1)\|2}\\ \beta_1=\mathbf{Y}^{\mathbf{T}}\dfrac{\hat{\mathrm{t}}(1)}{\|\hat{\mathrm{t}}(1)\|2}\end{cases} \tag{33}$$

Later, the step is repeated with the residual matrix instead of the original matrix. When the absolute value of the element in the residual *Y* is approximately 0, the regression accuracy established by the first component is considered to be sufficient.

The components and their corresponding MAE and MSE are as follow:

$$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i-y_I\right| \tag{34}$$

$$\mathrm{MSE}=\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i-y_i\right)^2 \tag{35}$$

Table 1. MAE and MSE of different components

COMPONENTS	MAE	MSE
1	3.335510198	23.01720318
2	3.284213912	22.49427433
3	3.278357031	22.45168815
4	3.279905112	22.39389226
5	3.2780227	22.37652189
6	3.275652855	22.35282811
7	3.275359531	22.32235436

According to the **Table 1**, we can know that MAE and MSE begin to converge when the number of selected components is 4.

If the *X* is *n* * *m*, set rank for $r \le \min(n-1,m)$, and the *t* (1) ... *t* (r) must be existed to make:

$$\begin{cases}\mathbf{X}=\hat{t}(1)\alpha_1^T+\cdots+\hat{t}(r)\alpha_r^T+\mathbf{X_r}\\ \mathbf{Y}=\hat{u}(1)\beta_1^T+\cdots+\hat{u}(r)\beta_r^T+\mathbf{Y_r}\end{cases}\tag{36}$$

Substitute

$$t_k=w_{k1}x_1+\ldots+w_{k7}x_7\ (k=1,2,\cdots,r)\tag{37}$$

into

$$Y=t_1\beta_1+\ldots+t_r\beta_r\tag{38}$$

and obtain PLS regression equation for the dependent variable:

$$y_j=a_{j1}x_1+\ldots+a_{j7}x_7,\ (j=1,2,\ldots,7)\tag{39}$$

Finally, we use Python to solve the seven coefficients to obtain the relationship of times of attempts and each attribute as **Table 2**.

Table 2. Coefficients of different variables

TIMES OF ATTEMPTS	Contest number	week	is_weekend	adj/ DET	pron/ noun	verb/mv	frequency
1	-0.093	0.011	0.021	-0.009	0.014	-0.012	0.092
2	0.044	0.012	-0.010	-0.051	0.082	-0.072	0.861
3	0.440	0.067	-0.100	-0.078	-0.001	0.137	0.285
4	0.599	0.046	-0.197	-0.086	-0.041	0.227	-1.596
5	-0.247	-0.133	0.039	0.100	-0.109	0.041	-0.626
6	-0.618	-0.028	0.182	0.104	-0.024	-0.131	0.282
7	-0.125	0.026	0.066	0.020	0.080	-0.189	0.702

Regression prediction equations were constructed from the coefficients given in the table. We predicted the score percentages of EERIE at 2023/3/1 and get the following matrix:

[0.04987342 5.41395334 21.5305623 32.47839538 25.09226636 11.68217328 3.75277592]

2.6Model Accuracy

Since only time and word attributes are considered in the modeling process, there are more variables to influence the final score proportion, such as whether the participants are native English speakers and so on.

In our model, the MAE and MSE are small in 10-fold cross-validation, which means the accuracy of the model is very ideal. So we are confident in the prediction (**Table 3**).

Table 3. MAE and MSE of different folds

FOLD	MAE	MSE
0	3.394	22.116
1	2.878	15.813
2	3.172	22.237
3	3.404	28.976

4	3.665	25.300
5	3.122	19.921
6	3.476	24.858
7	3.561	28.310
8	3.648	24.835
9	3.197	19.892

3. CONCLUSIONS

In this paper, we use the model of epidemic diseases, PSO, correlation analysis, PLS regression model to predict the trend of the number of results in the future and the time and word properties on the percentage of scores respectively.

From the point of view of dynamics and principles, we set up the propagation model of Wordle with the reference of the model of epidemic diseases. Then we use particle swarm optimization algorithm to optimize and adjust the unknown parameters according to the data, we use the adjusted iterative model to simulate the changes of the number of Wordle reports; Then we analyze the error performance before predicting the result interval on 2022/3/1 and get the prediction interval [11151.8,16311.0]. In the future, the number of reported results will also continue to decline slowly. Moreover, in order to probe whether any attribute of the word will affect the percentage of difficult mode records, we selected several important attributes including three parts of speech and frequency to label the data and conducted a correlation analysis on them. Finally we obtain the heat map and analyze the impact of the word attributes on the report percentage in difficult mode. We find that part of speech has little effect on the percentage of scores but frequency has more.

Due to the need to predict the percentage scores for a future date, we added the change of time into the prediction model. We choose PLS regression model to build the regression model with time variables containing contest number, the day of the week, if the day is weekend and word attributes variables containing part of speech and frequency as independent variables. Finally, we get the coefficient relationship between different tries and these variables, and regression prediction equations were constructed from the coefficients. Finally, We predicted the score percentages of EERIE at 2023/3/1 and get the matrix of its percentage scores [0.04987342 5.41395334 21.5305623 32.47839538 25.09226636 11.68217328 3.75277592]. We also judge the accuracy of the model by testing it with MAE and MSE.

However, there are factors we haven't discussed about, such as the influence of minority people or cultural environment. As a result, there is still a deviation between the final prediction and the actual value. In future work, we can continue model the remaining factors to achieve more precise results, which is also very helpful to the future development of Wordle games.

REFERENCES

[1] Chen Tengfei, Longhua, Shao Yubin, Du Qingzhi, Zhang Yanan, Song Xiaoxiao. Research and analysis of infectious disease models based on individual behavior [J]. Disease Surveillance, 2022,37 (06): 813-820.

[2] CAI Lijiao. Kinetic analysis of a class of infectious disease model [D]. Tianjin Polytechnic University, 2020.DOI:10.27357/d.cnki.gtgyu. 2020.001017.

[3] Pan Mengfei, Pan Lifang, Ruan Linzhi, Gao Chong. Take part of the settlements in Jiubao Street in Hangzhou city as an example [J / OL]. Journal of Zhejiang Sci-Tech University (Social Science edition): 1-9 [2023-02-20]. http://kns.cnki.net/kcms/detail/33.1338.TS.20221201.1143.003.html

[4] Zhao Yang, Sun Mingyue, Gao Hongyang, Zhang Wantong, Pu Xinyi, Yang Xiaochen, Yi Danhui, Gao Rui. Example study of comprehensive efficacy evaluation of new Chinese drugs based on partial least squares-second order factor model [J]. Chinese Journal of New Drugs, 2022,31 (18): 1774-1778.

[5] Wang Gang, Zhang Fuyin, Li Minghui, Wang Jinlong, Wang Yibo, Wu Chuanwei. Research on the air quality monitoring system based on the partial least squares regression algorithm [J]. Sensors and Microsystems, 2022,41(01):37-40+49.DOI:10.13873/J.1000-9787 (2022) 01-0037-04.

[6] Fu Shuai, Qian Donghai, Sun Jiajun, Li Hao. Research on QR code Visual Positioning Algorithm based on nonlinear least squares method [J]. Automation and instrumentation,2023(01):1-5.DOI:10.14016/j.cnki.1001-9227.2023.01.001.

[7] Chen Ziyun, Huang Xiaoxia, Yao Wanqing, Peng Mengjia. Rapid detection of citral content in the essential oil by NIR combined with partial least squares [J]. Anhui Chemical Industry, 2022,48 (01): 121-124.

[8] Huang Yiting, Shen Jianxin, Wang Yunchong, Chen Yidong. Slide mode control of variable parameters of permanent magnet synchronous servo motor based on recursive least squares observer [J]. Chinese Journal of Electrical Engineering, 2022,42(18):6835-6846.DOI:10.13334/j.0258-8013.pcsee.211725.