Research on Wordle Guessing Game Result Prediction Based on SIRS Model and Simulation

Bojin Liu [1,*], Rui Zhang [2], Xuqing Wang [1]

[1] School of Mathematics and Science, Hebei GEO University, Shijiazhuang, China, 050031

[2] School of Management, Hebei GEO University, Shijiazhuang, China, 050031

* Corresponding author: 15200943695@163.com

Abstract. Wordle is a daily word guessing game where players aim to guess a five-letter word in six or fewer attempts. To predict user participation, this study establishes a SIRS model to explain and predict the quantity of reported results, based on an analysis of the text propagation mechanism. To predict user guessing results, the player group is divided into entertainment players and strategic players. An Analog Simulation Model is established to simulate the word-filling process of these two types of players, thereby obtaining the result distribution for these two types of players. Subsequently, the Bayesian formula is used to calculate the proportions of entertainment players and strategic players, which are found to be 75.3% and 24.7%, respectively. However, given that each player's intuition, proficiency in different words, and desire to share vary, it is posited that individual player differences impact the predictions.

Keywords: SIRS Model, Analog Simulation, Wordle Guessing Game.

1. Introduction

The New York Times offers a daily puzzle known as Wordle, which has gained significant popularity. In this game, players are tasked with identifying a five-letter word within a maximum of six attempts. Upon each submission of a proposed word, feedback is provided to the player through the use of color-coded tiles. Players have the option to engage in either the regular mode or the more challenging Hard Mode. In the latter, the difficulty of the game is notably increased, and the number of guesses made by the player is represented as a percentage.

The number of reported guessing results in previous years varies daily. This paper uses a model to explain this variation and creates a prediction interval for the number of results reported on March 1, 2023, using the model. It then analyzes whether any attributes of the words affect the percentage of scores reported in the difficult mode. Subsequently, the paper predicts the distribution of reported results for given future solution words on future dates.

The application of the SIRS (Susceptible, Infected, Recovered, Susceptible) model and Analog Simulation has been prevalent in various fields, particularly in predicting outcomes based on historical data and current conditions. The research "An Analog Technique to Improve Storm Wind Speed Prediction Using a Dual NWP Model Approach" used the analog ensemble method to improve the prediction of wind speed for storms. This shows the application of analog simulation in weather prediction. The paper "Homogenous mixing and network approximations in discrete-time formulation of a SIRS model." proposed a discrete-time deterministic epidemic model using the SIRS model to understand the contagious dynamics and the behavior observed in the incidence of real infectious diseases. This illustrates the use of the SIRS model in epidemiology. [1-6]

The use of infectious disease models and simulations in the context of a word guessing game like Wordle is an innovative approach. The SIRS model, traditionally used to model the spread of diseases, can be adapted to model the spread of information or guesses in the game. The analog simulation can be used to model the behavior of different types of players and predict their guessing results. This approach takes advantage of the similarities between the spread of diseases and the spread of information in a game, providing a novel perspective on game analysis. Figure 1 shows that. [7]

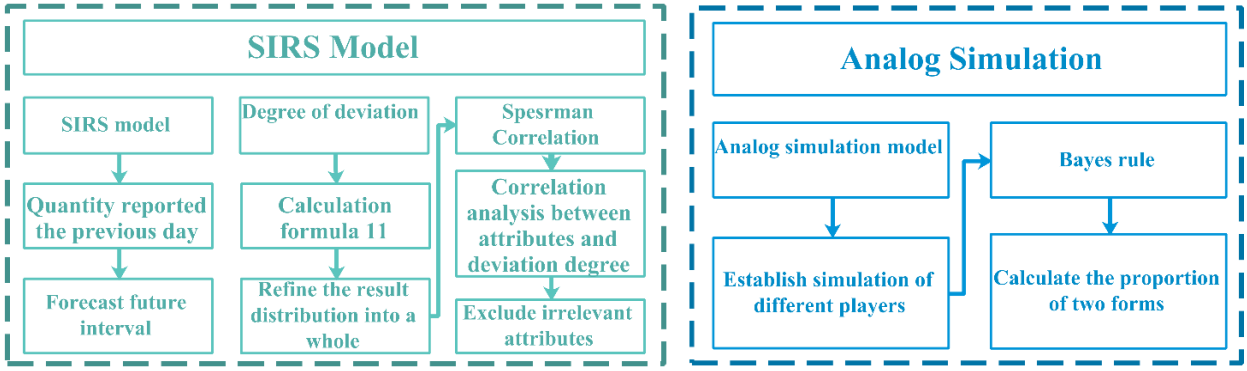


Figure 1. Structural Model

2. SIRS Model and Correlation Analysis

The SIRS model is constructed to elucidate and forecast the quantity of reported outcomes, drawing upon an analysis of the propagation mechanism inherent to Wordle. Subsequently, to enable correlation analysis with word attributes, the result distribution is refined into a measure of deviation, encapsulating the outcome as a whole. Ultimately, an analysis is conducted to ascertain the correlation between word attributes and the degree of deviation, thereby determining its potential influence on the percentage of reported results in Hard Mode.

2.1. Inference of changes in the number of reported results

Despite the seemingly simple rules of Wordle, the game possesses numerous attributes that facilitate its propagation. Primarily, the provision of only one puzzle per day creates an artificial scarcity, thereby enhancing the challenge and anticipation for players. Additionally, the ingenious design of Wordle's sharing function, which uses color-block emojis to represent game results, is not only highly recognizable and easily disseminated, but also prevents spoiler exposure.

Moreover, the traditional popularity of Wordle, coupled with its straightforward, user-friendly game design, has promoted its continuous propagation during the initial stages. However, as the spread of Wordle continues, the number of reported results gradually reaches saturation. As the novelty diminishes, the number of reported results begins to exhibit a declining trend, mirroring the spread of infectious diseases. For this reason, a modified infectious disease model is established for simulation.

2.2. SIRS model interpretation and prediction change

The implications of S\I\R in the SIRS model for Wordle disseminate as shown in Table.1.

Table 1. The connotation of SIRS model

Potential player (Susceptible)	Share results player (Infected)	Former player (Recovered)
A group of players who don't familiar with the Wordle and are easily recommended by I-type players.	The current players will share the game with S-type players, which will drive them to become I-type players.	The players who no longer play Wordle at present, but have the possibility of becoming an S-type player.

The choice of the SIRS model over the SIS model is motivated by the consideration that individual's reporting of results may be intermittent, implying that they do not acquire complete "immunity", and may report results again in the future. Subsequently, the formulation is established. To address this issue, the differential is converted into a difference for analysis from a discrete perspective.[8] Let S1 denote the initial quantity of potential players. The changes in type S players are then represented as follows:

$$S_{t+1} = S_t - \lambda S_t I_t + \alpha R_t \tag{1}$$

Where $\lambda S_t I_t$ represents the number of people recommended by I-type players to play Wordle, λ indicates the probability of playing Wordle after being recommended. The number of former players who may use Wordle again is reflected by αR_t. Besides, α stands for the probability that R-type will become S-type players.

The changes of type I players:

$$I_{t+1} = I_t + \lambda S_t I_t + \mu I_t \tag{2}$$

Where μI_t means the number of people who have never played Wordle. μ is the probability that a I-type player changes into a R-type player, which is inversely proportional to people's freshness.

The changes of type R players:

$$R_{t+1} = R_t + \mu I_t - \alpha R_t \tag{3}$$

We set the initial parameters according to the situation on January 7, 2022: $I_1 = 80630, S_1 = 600000, \mu = 0.05, \lambda = 0.5, \alpha = 0.01$.

With the above given parameters, the numerical simulations are as shown in Figure 2.

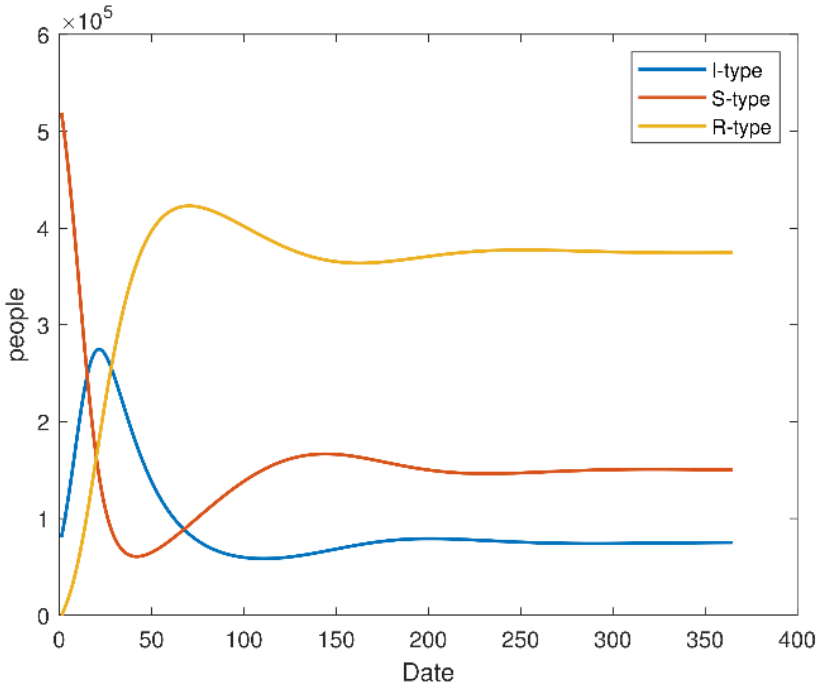


Figure 2. Numerical Simulation

As discerned from the figure, the overall trend of the type I value aligns closely with the actual scenario, suggesting that the SIRS model provides a reasonable interpretation of data fluctuations. However, under the given parameters, the number of type I players is not precisely captured. At this juncture, it is observed that the novelty for individuals varies over time, initially increasing and subsequently decreasing, implying that λ should first decrease and then increase. As a result, the actual value is continuously obtained through the model, and the parameters are dynamically adjusted. The average error between the calculated actual value and the measured value (the last 30 data points in the table) is approximately 277, which is used to determine the prediction interval range. The results are as shown in Figure 3 and Figure 4:

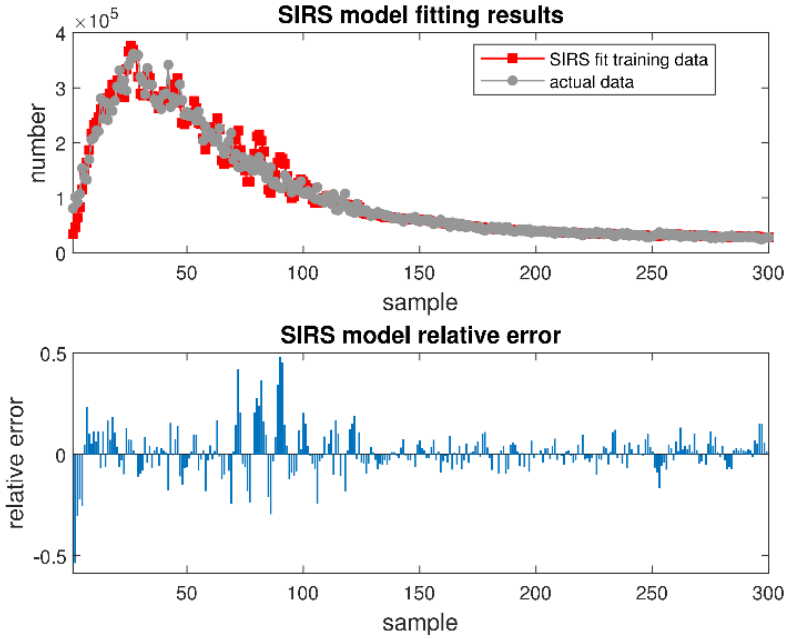


Figure 3. SIRS model fitting

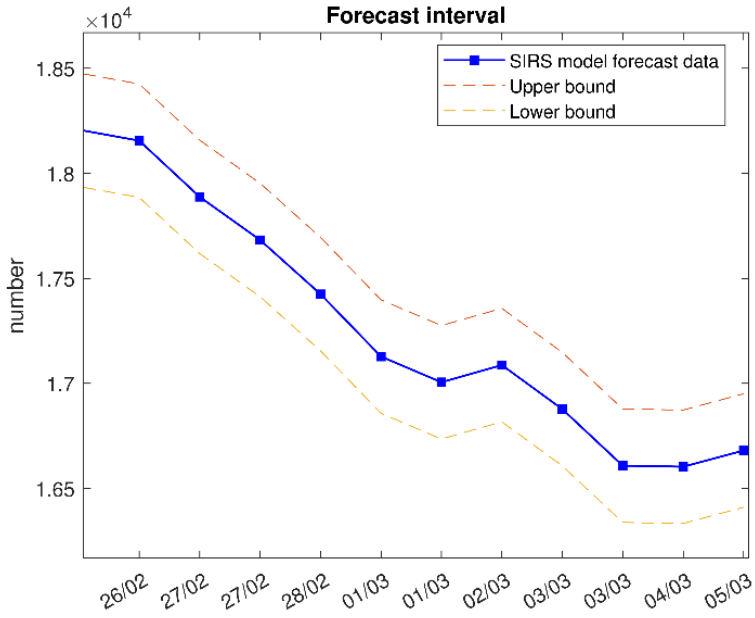
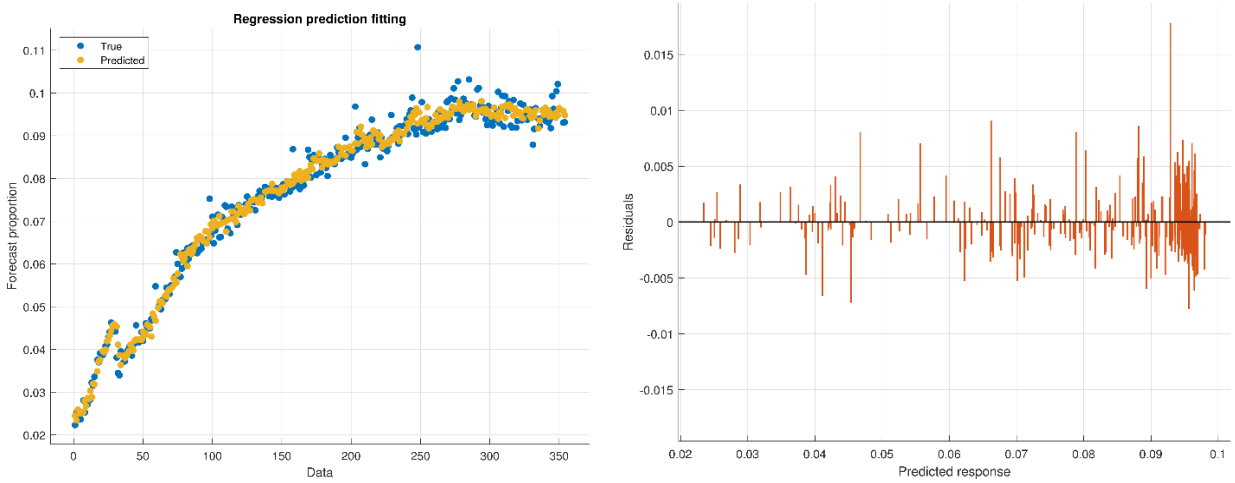


Figure 4. Forecast interval

As can be seen, the forecast interval of the number of results reported on March 1, 2023 is [16901,17355].

Furthermore, to facilitate subsequent analyses, a prediction of the proportion of future players is made.[9] Given the relatively minor degree of volatility, a regression prediction method is employed in Figure 5:



(a) Regression prediction fitting (b) Residuals

Figure 5. Percentage of future difficult mode players

And RMSE is about 0.0026, which is still a satisfactory result.

2.3. Daily guess word deviation degree

The daily reported results comprise a dataset that forms a total of seven aggregates, which is not conducive to subsequent correlation analyses with other aggregates. Consequently, a numerical value is considered to reflect the degree of guessing deviation for the day. The deviation degree 'e' for players with varying numbers of attempts is represented in the following Table 2.

Table 2. Player deviation degree

Attempts	1 try	2 tries	3 tries	4 tries	5 tries	6 tries	more tries(X)
e	0	2	4	6	8	10	12

And then calculate the degree of guessing deviation E on the day:

$$E_i = \sum_{j=1}^{7} res_{ij} \,.\, e_j = \begin{pmatrix} 1\% & 3\% & \cdots & 9\% & 1\% \\ & & \vdots & & \\ & & \vdots & & \\ & & \vdots & & \\ 0\% & 2\% & \cdots & 12\% & 2\% \end{pmatrix}_{359\times7} \begin{pmatrix} 0 \\ 2 \\ \vdots \\ 10 \\ 12 \end{pmatrix}_{7\times1} = \begin{pmatrix} 6.26 \\ 6.44 \\ \vdots \\ 6.3 \\ 6.7 \end{pmatrix}_{359\times1} \tag{4}$$

Where res_{ij} is the result distribution.

The greater the deviation, the more difficult to guess the solution word of the day.

2.4. Correlation analysis of word attributes and deviation degree

In this section, 'E' denotes the degree of deviation. The terms Part of Speech, Letter Frequency, Letter Order, Letter Layout and Word Frequency correspond respectively to the word's Part of Speech, Letter Frequency, Letter Order, Letter Layout, and Word Frequency. Initially, the data is standardized, and an equidistant constant is added to each column of data to facilitate the observation of the trend of each curve in the Figure 6:

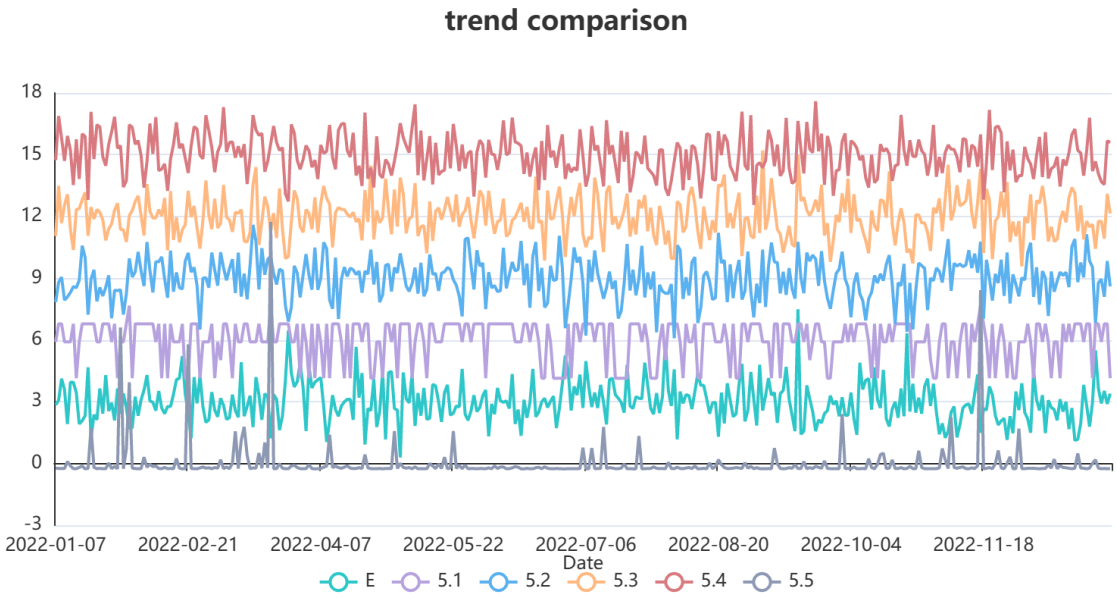


Figure 6. Data standardization

An initial inference drawn from the figure suggests that the attribute of Letter Order exhibits the highest correlation with the degree of deviation.

Subsequently, a total of six aggregates, comprising word attributes and deviation degrees, are subjected to normality testing. Given that the volume of data is less than 5000, the Shapiro-Wilk test is employed to assess its significance. The test is conducted as follows in Table.3.

Table 3. Shapiro-Wilk test

Variable	sample size	median	average	standard deviation	skewness	kurtosis	S-W test
E	359	6.34	6.385	0.801	0.586	1.17	0.98(0.000***)
Part of Speech	359	3	3.109	1.137	-0.971	-0.443	0.742(0.000***)
Letter Frequency	359	1.133	1.13	0.131	-0.256	-0.011	0.993(0.079*)
Letter Order	359	0.022	0.023	0.008	0.176	-0.072	0.996(0.434)
Letter Layout	359	0.075	0.074	0.015	0.065	-0.509	0.994(0.160)
Word Frequency	359	2146	25675.164	97655.552	7.888	74.296	0.225(0.000***)

Note: ***, **, * represent the significance levels of 1% 5% and 10% respectively.

If the result does not exhibit significance (P > 0.05), it suggests conformity to the normal distribution, and vice versa. However, in practical scenarios, if the absolute value of the sample kurtosis is less than 10 and the absolute value of the skewness is less than three, it can be considered as essentially conforming to the normal distribution when combined with the normal distribution of the histogram, the P-P chart, or the Q-Q diagram.[10] Consequently, it can be inferred that Letter Frequency, Letter Order and Letter Layout fully conform to the normal distribution. 'E' is essentially in line with the normal distribution, while Part of Speech and Word Frequency do not conform to the normal distribution.

Given that some data do not satisfy the normal distribution, it is posited that the analysis of the Spearman correlation would yield more accurate results.

In practical applications, the relationship between variables is not relevant. Subsequently, ρ can be calculated through simple steps, with the calculation formula as follows:

$$e = 1 - \frac{6\sum grade_i^2}{n(n^2-1)} \tag{5}$$

Where $grade_i$ is the difference of the grades of two observed variables. n represents sample size. The Figure 7 below is visualization of the calculation results:

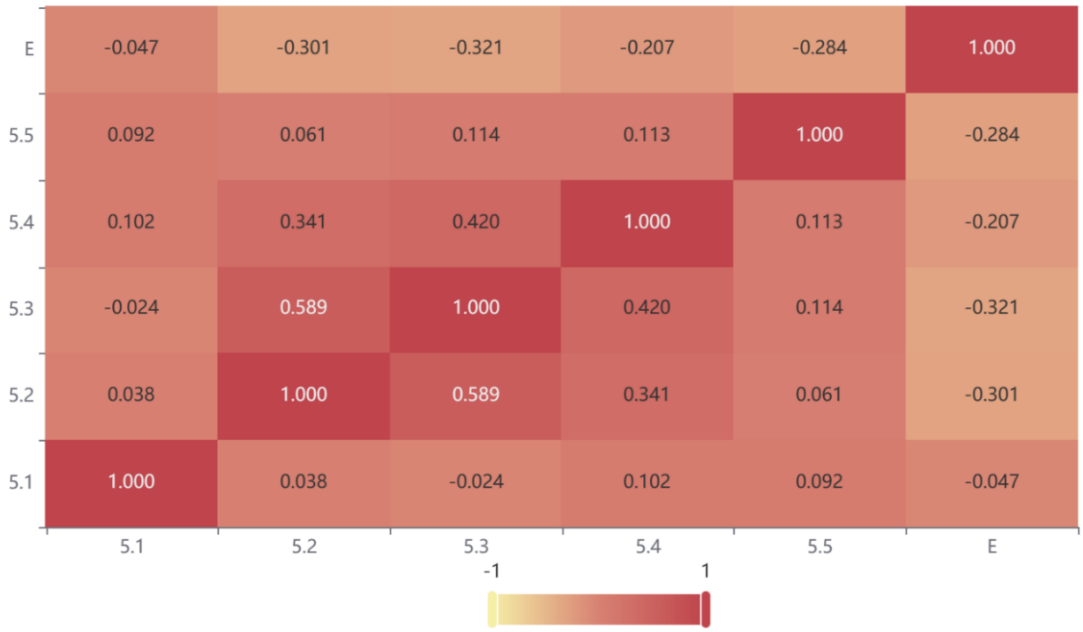


Figure 7. Correlation heat map

As observed from the figure, the word attributes exhibit a negative correlation with the degree of deviation, which aligns with expectations. For instance, words with high letter frequency are easier to guess correctly, resulting in a lower degree of deviation, thus indicating a negative correlation. It can be inferred that Letter Frequency has the highest correlation with the degree of deviation, while Part of Speech has the lowest correlation. Generally, in practical scenarios, if the absolute value of the correlation coefficient exceeds 0.2, a correlation is assumed to exist. Therefore, for word parts of speech with a correlation coefficient of less than 0.2 (Word Frequency), it is considered that they will not impact the degree of deviation and are thus eliminated from subsequent analyses.

In conclusion, the letter frequency, letter order, letter layout, and word frequency attributes of words will influence the percentage of reports in Hard Mode. However, the part of speech of a word does not exert an effect on this.

3. Based on the Distribution of Prediction Results of the Analog Simulation Model

Particular interest is directed towards players who can guess correctly in 1 or 2 tries. According to research, the theoretical limit of grades cannot be lower than 3. Therefore, it is posited that players participating in Wordle should be classified into two categories: entertainment players and strategic players. Most of the players who can guess correctly in 1 or 2 tries are likely to be entertainment players, as shown in Figure 8.

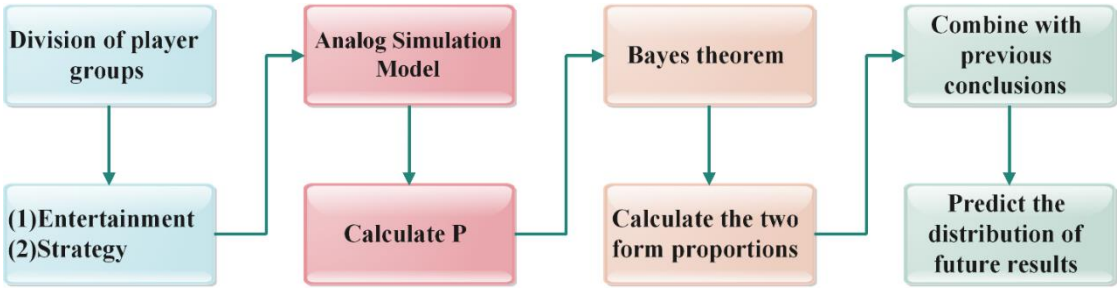


Figure 8. Flow chart of second question

Among them, the improvement of the Hard Mode is to reduce the opportunity to test which letters are to test, because the letters in green or yellow from the last input word cannot be lacking. In other

words, 10 different letters cannot only use two times to detect whether the answer exists. This affects the number of guesses of strategic players.

3.1. Divide the player groups

Entertainment Player: These players typically guess based on commonly used words in their daily life. The higher the word frequency, the higher the likelihood of the word being used for guessing. Simultaneously, the field in which the players operate also influences their guesses, and they base their guesses on the feedback received from previous attempts. However, they do not employ complex strategies. In essence, they rely on their intuition to guess the correct results, which introduces a strong element of randomness.

Strategic Player: These players aim to identify the correct result with the fewest possible guesses each day. They do not simply fill in words at each attempt, but choose words that provide the most feedback information. They then meticulously analyze the received feedback information to inform their next decision, taking into consideration the word attributes related to the degree of deviation. In brief, their objective is to narrow down the range of optional words as quickly as possible. As mentioned earlier, the theoretical limit of strategic players' performance cannot be lower than three. Based on this, it is assumed that only 10% of the players whose results are 1 or 2 tries are strategic players, as the majority of cases require three or more attempts to find the correct answer.

3.2. The establishment of the Analog Simulation Model

Consideration is given to the uncertainty of individual player differences affecting the predictions. This is due to the fact that each player's intuition, proficiency with different words, and desire to share vary. An attempt is made to reflect this difference using the rand formula in the model, but it does not completely align with reality. It is assumed that the choice for recreational players is simple, that they are primarily concerned with word frequency, and that there are no limitations.

To calculate the proportion of entertainment players and strategic players in the group in Section 3.3, a solution word with moderate deviation is taken as a starting point. After simulating 100,000 entertainment players, it is concluded that the proportion of players who guess correctly in 1 or 2 tries among entertainment players is approximately 7.55%.

3.2.1. Analog Simulation of strategic players

In terms of simulating strategic players, it is necessary to formulate their specific strategies, shown in Figure 9:

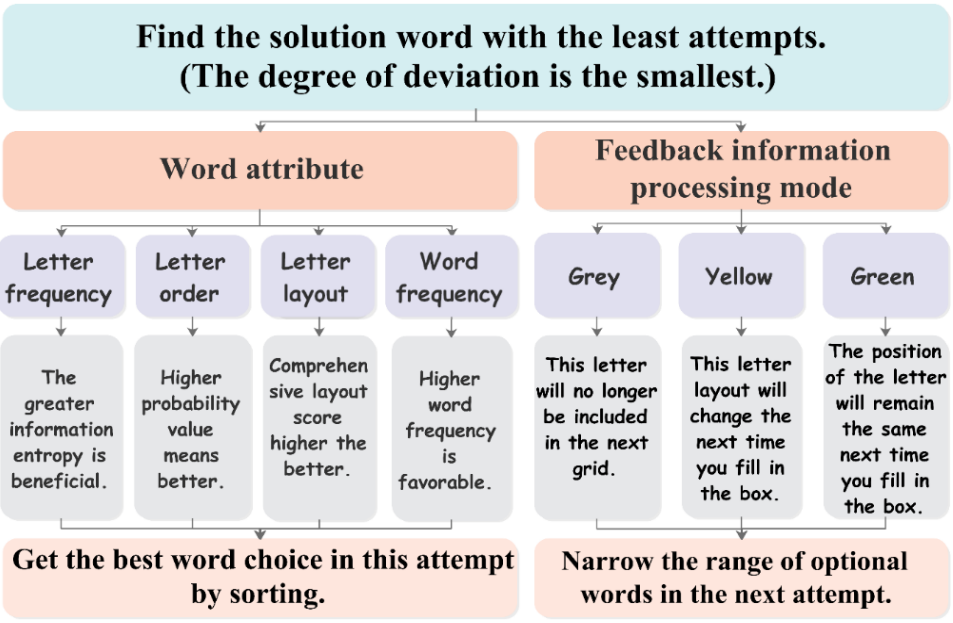


Figure 9. Related strategies of strategic players

Note: Where ω indicates the weight corresponding to the four words attributes, and the higher the correlation with the degree of deviation, the greater the weight ω. In accordance with the results of the section 2.4, the weight is assigned by the $w_i = \frac{e_i}{\sum_{i=1}^{n} e_i}$. The value are as follows: $w_1 = 0.270, w_2 = 0.289, w_3 = 0.186, w_4 = 0.255$,and the corresponding values of the word attributes are standardized in advance. In order to obtain more information, players generally do not follow the feedback information processing method until the third attempt.

On the foundation of this strategy, the simulation of strategic players is conducted:

Algorithm Analog Simulation Strategy Player

Input: The solution word vector $S_{word}\{d_1, d_2, \cdots d_5\}$ of the day and the number of strategy players G.

Output: Distribution of results for all strategy players $t_1, t_2, t_3, t_4, t_5, t_6, X$.

for people = 1toG **do**
  An initial thesaurus W_0 is formed, which contains the attribute values corresponding to each word.
  for $try = 1$ to 2 **do**
    The purpose of the first two attempts is to get the maximum information. Using roulette to choose the best words in the thesaurus.
    if $try_{word} = S_{word}$ than **then**
      $t_{try} = t_{try} + 1$ Count how many times each player needs until gets the word.
      Continue
    else
      In the normal mode, the initial thesaurus w only removes this t.
      When it is a difficult mode, the initial lexicon w is reduced to w.
      All retain feedback information.
    end
  end
  According to the feedback information of the first two times, narrow the range of words that can be selected in the third attempt. Eventually form a thesaurus H.
  for $try = 3$ to 7 **do**
    Selecting the best word in thesaurus W_0 by roulette
    if $try_{word} = S_{word}$ than **then**
      $t_{try} = t_{try} + 1$ Count how many times each player needs until gets the word.
      continue
    else
      Based on the processing method of feedback information in the strategy, the thesaurus is reduced to W'_{try}.
    end
  end
end
return Distribution of results for all strategy players

However, not all players are as thoughtful and consistent as algorithms. To address this issue, a roulette-style simulation is employed for players to choose words. The more effective the word in the algorithmic strategy, the greater the probability of it being selected. This introduces a degree of decision-making disturbance, thereby reflecting the individual differences among players.

3.3. Bayesian Theorem calculates the proportion of two types of players

As the Analog Simulation is conducted by categorizing into different types of players, the result obtained represents the distribution of outcomes based on a specific type of player, which is a

conditional probability. The relationship between two conditional probabilities can be described using the Bayesian formula.

In this context, the results distributed in 1 or 2 tries are assumed as event A, while the rest are event B. Entertainment players are denoted as ent, and strategic players as str. The average percentage distribution, can be obtained as follows:

$$P(A) = 0.0632 \quad P(B) = 0.9368 \tag{6}$$

Further, based on the analysis of section 2.1 and section 2.2:

$$P(ent|A) = 0.9 \quad P(A|ent) = 0.0755 \tag{7}$$

Finally, combined with the Bayesian formula:

$$P(ent|A) = \frac{P(ent)P(A|ent)}{P(A)} \tag{8}$$

It can be deduced from Equation 8:

$$P(ent) = \frac{P(A)P(ent|A)}{P(A|ent)} \approx 0.753 \tag{9}$$

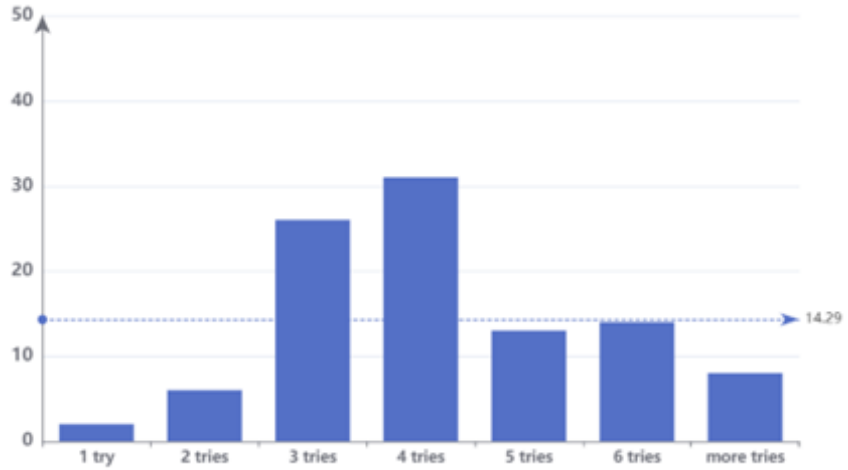
$$P(str) = 1 - P(ent) = 0.247 \tag{10}$$

In conclusion, it has been determined that the proportion of entertainment players in the reported results is 75.3%, while strategic players account for 24.7%.

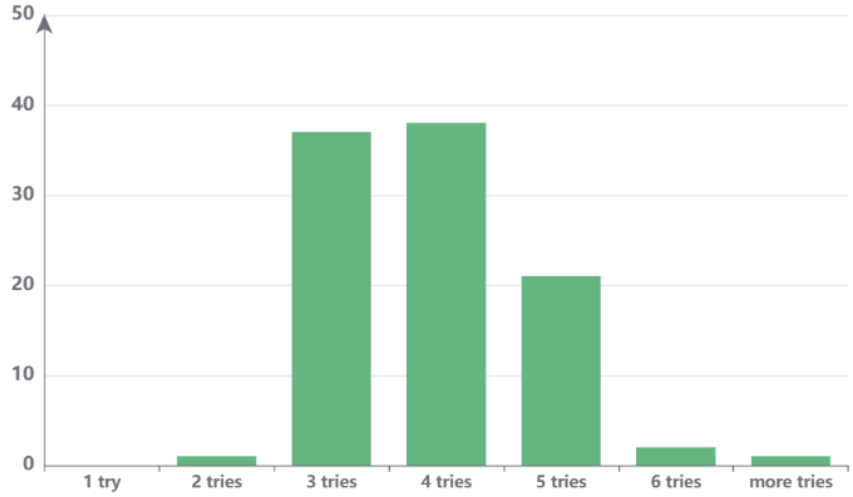
3.4. Predict the future results distribution

For the distribution of future forecast results, the SIRS model is initially combined with the regression prediction to estimate the number of reported results and the proportion of players in Hard Mode, thereby obtaining the number of Hard Mode players. Subsequently, based on the proportion of the two types of players in the population, the number of entertainment players and strategic players are determined, and these are separately substituted into the corresponding Analog Simulation Algorithm. Finally, after setting the solution word of the day, the result distribution for the day can be predicted.

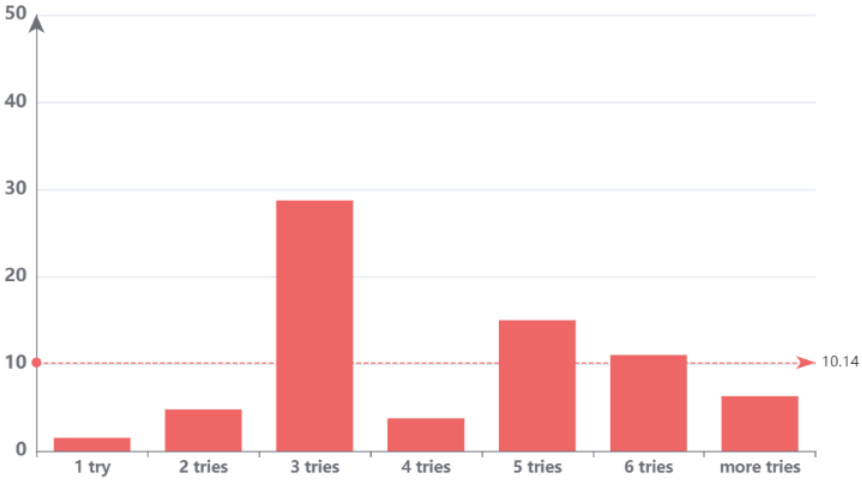
The next step involves predicting the distribution of results when the solution word is 'EERIE' on March 1, 2023. Initially, the numbers of entertainment players and strategic players in Hard Mode are calculated to be 1759 and 577, respectively, and these are substituted into the corresponding Analog Simulation Algorithm to obtain in Figure 10:



(a) Total



(b) Strategy



(c) Entertainment

Figure 10. Word attribute result visualization

An examination of the various attribute values of 'EERIE' is then conducted, as shown in Table.4.

Table 4. Various attribute values of EERIE

Attribute	Letter Frequency	Letter Order	Letter Layout	Word Frequency
EERIE	1.5789	0.0357	0.0792	2272
Average	1.1316	0.0229	0.0742	10099

From the table, it can be observed that the various attribute values of 'EERIE' are generally greater than the average value of the word attributes in the data table. Given that the attribute value is negatively correlated with the degree of deviation, the degree of deviation is less than the average value, indicating that players require fewer attempts to find the solution word. This essentially aligns with the result distribution of the model prediction, providing sufficient confidence in the model's prediction.

4. Conclusion

In conclusion, this study provides an in-depth analysis of the propagation mechanism of Wordle, suggesting that the fluctuation in the number of reported results is influenced by the sharing of Wordle and the novelty it offers to individuals. The SIRS model, established in this context, effectively explains and predicts the number of reported results, with a prediction range of [16901,17355] for the reported results on March 1, 2023.

The study further refines the distribution of results into the degree of deviation to facilitate correlation analysis with word attributes. The Spearman correlation analysis reveals that word attributes, namely Letter Frequency, Letter Order, Letter Layout, and Word Frequency, significantly influence the percentage of reports in Hard Mode. However, the part of speech of a word does not exhibit a similar effect.

The player group is categorized into entertainment players and strategic players, and an Analog Simulation Model is developed to simulate the word-filling process of these two types of players. This leads to the derivation of the result distribution for these two types of players. Subsequent application of the Bayesian formula calculates the proportions of entertainment players and strategic players at 75.3% and 24.7%, respectively.

The model's prediction of the solution word as 'EERIE' on March 1, 2023 aligns with the distribution shown in Figure 10. However, the study acknowledges the impact of individual player differences, such as variations in intuition, proficiency in different words, and desire to share, on the predictions.

The findings of this study highlight the feasibility of using the SIRS model and Analog Simulation Model for predicting game outcomes in Wordle. Future research could focus on refining these models to account for individual player differences more accurately, thereby enhancing the precision of predictions.

References

[1] Alessandrini, S., L. Delle Monache, S. Sperati, and J. N. Nissen, 2015a: A novel application of an analog ensemble for short-term wind power forecasting. Renewable Energy, 76, 768 – 781.

[2] T. Appel et al., "Cross-Task and Cross-Participant Classification of Cognitive Load in an Emergency Simulation Game," in IEEE Transactions on Affective Computing, vol. 14, no. 2, pp. 1558 - 1571, 1 April-June 2023.

[3] Liying Ma, Bo Lv, Yongping Hou, Xiangmin Pan, "Prediction Model of the Mechanical Behavior of a Fuel Cell Stack under Strengthened Road Vibrating Conditions", Modelling and Simulation in Engineering, vol. 2021, Article ID 6671547, 11 pages, 2021.

[4] Wenting Chen, Bopeng Jin, "Modeling and Simulation of Cultural Communication Based on Evolutionary Game Theory", Complexity, vol. 2021, Article ID 6115606, 12 pages, 2021.

[5] A. Ścibior, V. Lioutas, D. Reda, P. Bateni and F. Wood, "Imagining the Road Ahead: Multi-Agent Trajectory Prediction via Differentiable Simulation," 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), Indianapolis, IN, USA, 2021, pp. 720 - 725.

[6] Schwägele, S., Zürn, B., Lukosch, H. K., & Freese, M. (2021). Design of an Impulse-Debriefing-Spiral for Simulation Game Facilitation. Simulation & Gaming, 52 (3), 364 – 365.

[7] Ilaria Renna (2021) Homogenous mixing and network approximations in discrete-time formulation of a SIR'S model, Journal of Biological Dynamics, 15: 1, 635 - 651.

[8] Zhou X. and Wang M. (2022) “Dynamic analysis of a fractional-order SIRS model with time delay”, Nonlinear Analysis: Modelling and Control, 27 (2), pp. 368 - 384.

[9] Al-Akashi, Falah Hassan Ali. "Stock Market Index Prediction Using Artificial Neural Network." JITR vol.15, no.1 2022: pp.1 - 16.

[10] Bao Chen. Visualization of temporal text based on dynamic word cloud. 2021. Shandong University, MA thesis.