Research on Outcome Prediction Model Based on LSTM Neural Network Model: An Example of Wordle Game

Yuan Feng *, Xinjie Song

Central South University of Forestry and Technology, Changsha, Hunan, 410004

* Corresponding Author Email: a122595798@gmail.com

Abstract. Wordle is popular worldwide, and predicting the results of the game has attracted many people's attention. Based on this, this article uses the LSTM neural network model, information entropy theory to predict and analyze the number of gamers and word difficulty of Wordle's daily words. The results indicate that: (1) The number of players in the Wordle game shows a daily decreasing trend; (2) The information entropy of words with two or more duplicate letters is lower than words without duplicate letters; (3) There is no significant correlation between the information entropy and the pass rate in the Wordle difficult mode.

Keywords: Wordle, LSTM neural network model, information entropy theory.

1. Introduction

Wordle is a popular alphabet puzzle game provided by The New York Times and popular among users from all over the world. After inputting five letters, players determine whether the letters appear within the word and whether their positions are correct based on the color changes of the tiles (green, yellow, gray). During the game, many factors may affect the number of attempts players make to guess words, including the player's vocabulary, frequency of word usage, and whether the words have duplicate letters.

To accurately predict the results of the game and provide players with a better gaming experience, this article utilizes the LSTM neural network model, information entropy theory, and Spearman analysis method.

The information entropy theory originated from Shannon and has been widely used in information analysis. Wang [1] sets with conditional information entro, and analyzed and discussRough sets' information theory and algebraic perspectives ts to find similarities and differences. Based on this, they obtained the decision table reduction of conditional information entropy; Chen [2] combined information entropy Theory with fractal theory pioneered the extension of urban land use structure information into fractal dimension; Zhou [3] analyze the characteristics of information entropy, applied information entropy to comprehensive evaluation to standardize the objective weight of evaluation indicators; Zhang [4] introduced information entropy into the evaluation of urban ecosystem evolution based on the wide application of information entropy theory in the urban context; Peng [5] believed that privacy is a type of information, applied information entropy theory to the field of communication privacy protection, and creatively Developed a privacy definition protection mechanism and an evaluation analysis of the attack range. In foreign research, Denis et al. [6] applied information entropy in the field of chemistry and described that discrete information entropy can be used to quantify molecules into sets, while continuous information entropy can quantify the electron density of molecules and atoms; Tsai et al. [7] stipulate that the measurement of information entropy should be used to evaluate image quality, propose a simple and direct TI parameter, and apply a single parameter to comprehensively evaluate the mathematical ray irradiation method to evaluate the quality of the image.

However, information entropy has fewer applications in the field of text research. Information entropy can measure the entropy value contained in information. By calculating the entropy value, the amount of information contained in text can be judged. In this paper, the size of the information contained in the five-letter words that appear in the daily word guessing game in Wordle is used to predict and analyze the results based on the LSTM neural network model. The research results are

useful for the application of information entropy theory in English word formation. The research has a certain guiding significance.

2. Model building

The node output of a recurrent neural network (RNN) is determined only by weights, biases, and activation functions. It is a chained structure where the same parameters are used for each time slice. However, RNN has a long-term dependence problem. During the processing of long sequences, the error gradient disappears exponentially with the increase in the time between events, resulting in gradient decline and increased computational difficulty, which may cause gradient explosion. In any case, the layer where the RNN obtains a small gradient will no longer learn. So RNN has only short-term memory.

A long short-term memory network (LSTM) [8-9] is one of the variants of a recurrent neural network (RNN). Its core lies in the cell state and the structure of the " gate". Compared with RNN, the three "gate" structures of LSTM can determine whether the input is important enough to be remembered and whether it can be output, which gives them the ability of long-term memory.

To minimize the training error, gradient descent methods such as applying the temporal backward pass algorithm can be used to modify the weights each time based on the error. When the LSTM block is set up, the error is also calculated through the backward passback, affecting each gate from the output to the input stage until the value is filtered out. Therefore, normal backward pass neural training of LSTM block memory for long-term values is effective.

LSTM can solve the long-term dependency problem of RNN because LSTM introduces a gate mechanism to control the circulation and loss of features. LSTM consists of a series of cell units, and its chain structure is shown in Figure 1.

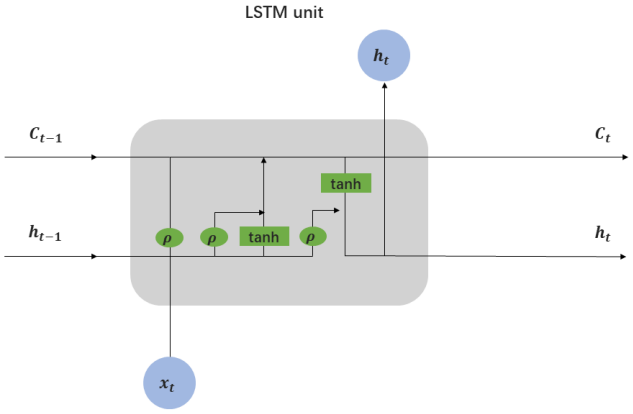

Figure 1. Schematic diagram of LSTM structure

The concept of information entropy evolved from physical entropy. It is often used as a measure of the degree of ordering of a system. It can be understood as the probability of a certain type of information appearing at discrete random times and can be further used as a target or parameter for the optimization of system equations [10-11]. Therefore, information entropy can be used as a certain attribute of words, and the size of information entropy corresponds to the amount of information carried by words. Because the information entropy is cumulative, the information entropy of a word is regarded as the sum of the information entropy of each letter. According to the definition of entropy, we can get the formula of information entropy as:

$$H(\alpha) = -\Sigma_{K=1}^{N} p(\alpha_i) \log p(\alpha_i) \tag{1}$$

Among them, $p(a_i)$ random events are represented a as a_i probability, and α is a set for information entropy calculation.

3. Result analysis

3.1. Result analysis of the LSTM neural network model

Considering that the given data is affected by rounding, we preprocessed the data. We deleted some data whose sum of percentages exceeded 100% and imported the remaining data into the LSTM algorithm for training.

During the training process of LSTM, set the parameters as follows: Max Epochs is set to 100 times, Initial Learn Rate is set to 0.1%, and Gradient Threshold is set to 0.01. Using the data set from January 7, 2022, to December 31, 2022, as training set for training.

This paper uses the predicted model to predict the data for 2022/03/01. As can be seen from Figure 2 and Figure 3, the simulated data fits the real data well.

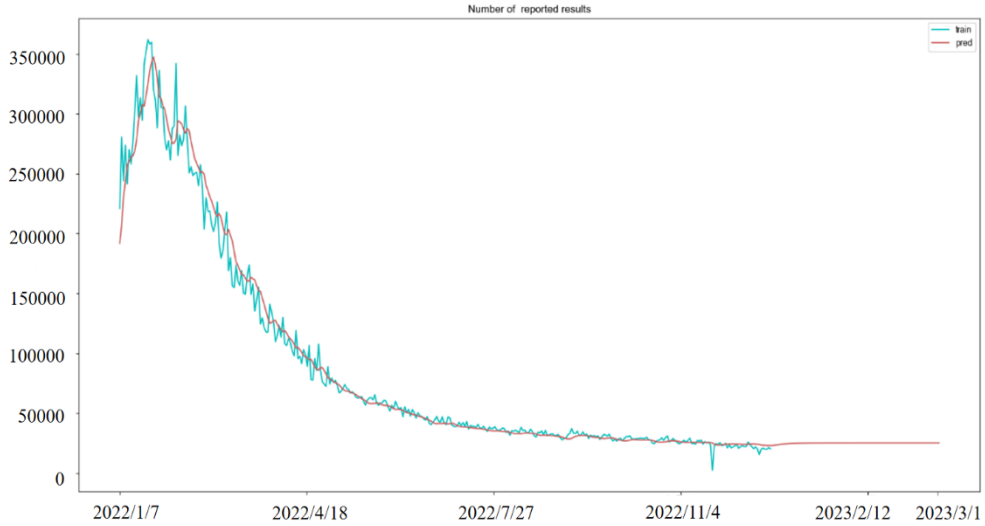

Figure 2. Number of report results

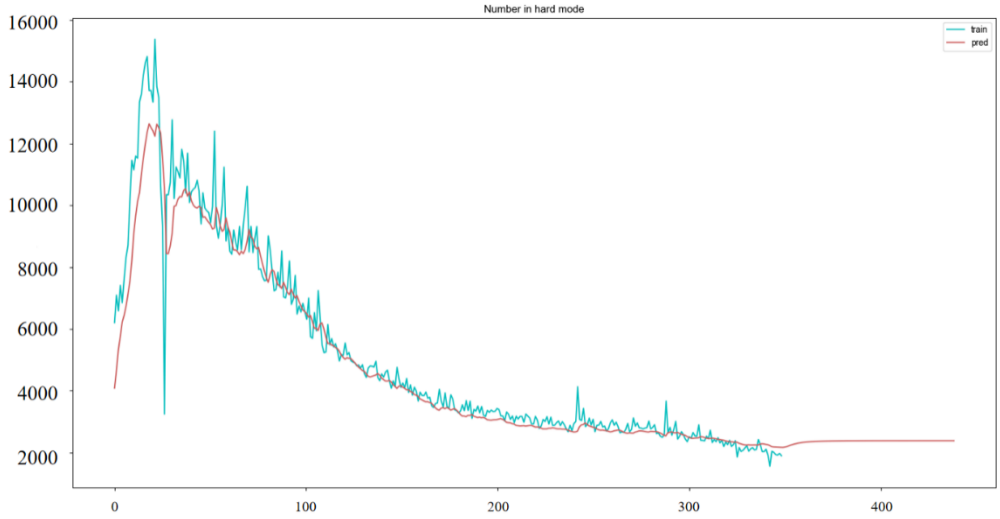

Figure 3. Quantities in Hard Mode

Using two common indicators MSE and SMAPE in time series forecasting to test the performance of the model. MSE (mean squared error) is the average of the sum of squared errors between the observed value and the error value, and is the second moment of the error, including the variance of the estimate, The machine bias, is a measure of the quality of the estimate.

$$MSE(\mathrm{y},\hat{\mathrm{y}}) = \frac{1}{n}\Sigma_{i=1}^{n}|\mathrm{y}_i - \hat{\mathrm{y}}_i| \tag{2}$$

SMAPE (Symmetric Mean Absolute Percent Error) is an indicator of the prediction accuracy of the prediction method. Their formulas are as follows:

$$SMAPE(y,\hat{y}) = \frac{1}{n}\Sigma_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{\frac{|y_i|+|\hat{y}_i|}{2}}\right| \times 100\% \quad (3)$$

According to the provided formula, the MSE and SMAPE between the simulated data and the predicted data in normal mode are 0.003 and 0.077%, respectively, while the MSE and SMAPE in hard mode are 0.001 and 0.173%.

Since there will be errors between the model's predictions and the true values, use a 95% confidence interval to determine the interval. Within the 95% confidence interval, the number of people playing normal mode and hard mode on 2022/03/01 is shown in Table 1:

Table 1. Predicted result values

Number of reported results	Number in hard mode
20936	2143

The 95% confidence intervals of the reported number and the predict number are [19448, 22922] and [1826, 2152], respectively, and the prediction results of two are 20936 and 2143, resin, both within the 95% confidence interval, and the prediction of the model result is valid.

3.2. Information entropy result prediction

Using all five-letter words to calculate the frequency of occurrence of 26 letters and then substituted them into the formula to find the information entropy of the letter AZ. The information entropy of 26 letters in AZ is shown in Figure 4:

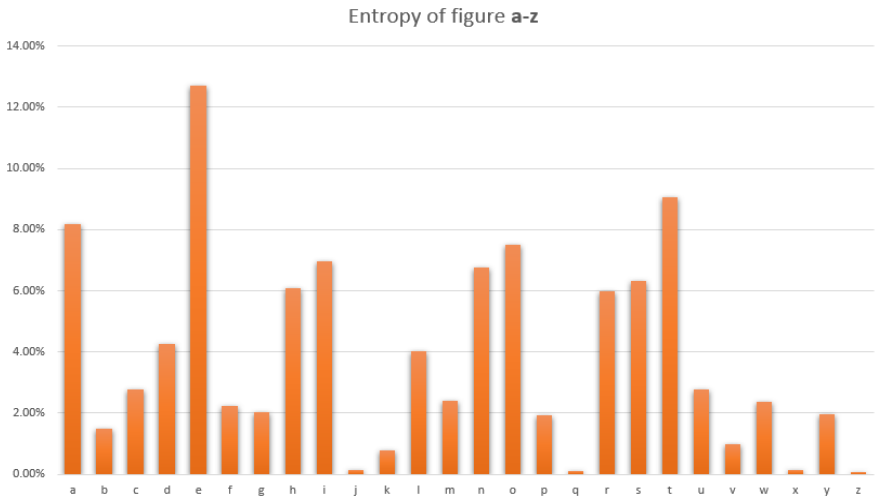

Figure 4. Information entropy of letters AZ

According to the additive nature of information entropy, can get the information entropy of a word composed of five letters. The Spearman test was then used to determine the correlation between information entropy and the percentage of scores passing the hard mode (Table 2).

Table 2. Predicted result values

	Percentage of hard mode in all	Entropy
Percentage of hard mode in all	1(0.000***)	-0.048(0.362)
Entropy	-0.048(0.362)	1(0.000***)

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

The table above shows the parameter results of the model test, including the correlation coefficient (-0.048) and the significance p-value (0.362). It is found that the p-value (0.362) is significantly

greater than 0.05, and there is no correlation between information entropy and difficult mode passing percentage.

The running result is shown in Figure 5:

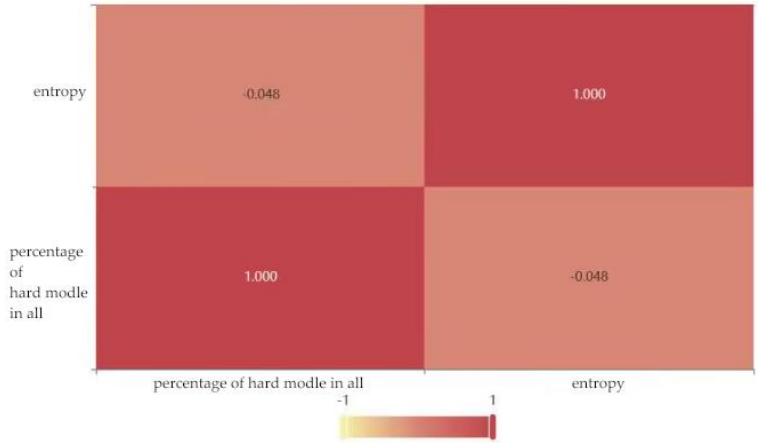

Figure 5. Heatmap of Spearman analysis

4. Conclusion

To discover the word rules of the Wordle guessing game, this paper uses the information entropy theory and LSTM neural network algorithm to judge the word rules based on the daily data of Wordle games. The research results are as follows:

1) There is no obvious correlation between the passing percentage in Wordle difficulty mode and the information entropy of the word.

2) In a word consisting of 5 letters, the more repeated letters, the smaller the information entropy.

3) This paper establishes a simple model based on the definition of information entropy, without adding the rules of word formation related to words. Further research is needed on the changes of specific word formation rules to information entropy

4) This paper proposes to apply information entropy to word composition, which is worthy of further study.

References

[1] Wang Guoyin, Yu Hong, Yang Dachun. Decision table reduction based on conditional information entropy [J]. Journal of Computer Science, 2002 (07): 759-766.

[2] Chen Yanguang, Liu Jisheng. Quantitative description of urban land use structure and morphology: from information entropy to fractal dimension [J]. Geographic Research, 2001 (02): 146-152.

[3] Zhou Wei, Li Xiaojing. Comprehensive evaluation method based on information entropy theory [J]. Science, Technology and Engineering, 2010, 10 (23): 5839-5843.

[4] Zhang Yan, Yang Zhifeng, He Mengchang, et al. Analysis of urban ecosystem evolution based on information entropy [J]. Journal of Environmental Science, 2005 (08): 1127-1134.

[5] Peng Changgen, Ding Hongfa, Zhu Yijie, et al. Information entropy model and measurement method for privacy protection [J]. Journal of Software, 2016, 27 (08): 1891-1903.

[6] Sabirov, DS; Shepelevich, IS Information Entropy in Chemistry: An Overview. Entropy 2021, 23, 1240.

[7] Tsai, DY., Lee, Y. & Matsuyama, E. Information Entropy Measure for Evaluation of Image Quality. J Digit Imaging 21, 338–347 (2008).

[8] Yang Li, Wu Yuqian, Wang Junli, Liu Yili. A Review of Recurrent Neural Network Research [J]. Computer Applications, 2018, 38 (S2): 1-6+26.

[9] Liu Haoran, Zhao Cuixiang, Li Xuan, Wang Yanxia, Guo Changjiang. Research on a Neural Network Optimization Algorithm Based on Improved Genetic Algorithm [J]. Journal of Instrumentation, 2016, 37 (07): 1573-1580.

[10] Xu Tianyun, Zhou Jun. Fault Analysis of RV Reducer Planetary Gear Based on Information Entropy and VMD [J]. Noise and Vibration Control, 2023, 43 (01): 122-127.

[11] Application of information entropy principle in structural optimization [J]. Chinese Journal of Applied Mechanics, 1992, (02): 66-75.