Predicting Lottery Numbers through Network Analysis

Jeonghui Lee [1]

Abstract

The lottery system is generally perceived as a game of chance, but sometimes statistically significant patterns emerge that are difficult to explain by mere randomness. This study approaches the relationships among lottery numbers from a network analysis perspective to develop a predictive model. Each lottery number is considered a node in the network, and the Expectation Reflection algorithm is used to analyze the interactions between numbers. This approach explores the potential for predicting future lottery numbers based on previous draws. The model's prediction results indicate that the model has a certain level of predictive capability, considering the approximate 30% probability of matching at least two numbers by random chance. This study presents a novel approach to predicting lottery numbers and suggests that more sophisticated predictions could be achieved in future research with more extensive data and advanced modeling techniques.

1. Introduction

The lottery system is generally perceived as a game of chance, but sometimes unusual events occur that challenge this notion. For example, there are cases where the same person wins the lottery consecutively, or the distribution of winning numbers shows statistically significant patterns. These phenomena suggest that a new approach is needed to analyze and predict the relationships between lottery numbers, as they are difficult to explain by mere chance.

In this study, we approach the lottery system from a network analysis perspective. Each lottery number is considered a node in a network, and we use a network-based algorithm to infer the relationships among these numbers. This approach allows us to explore potential patterns and dependencies that are not immediately apparent from a purely statistical viewpoint.

To achieve this, we used the Expectation Reflection algorithm to fit a network model. This algorithm is designed to estimate the initial field values h_0 and interaction weights w that characterize the relationships between lottery numbers. The method operates on a feature matrix x, representing the features of previous lottery draws, and a target matrix y, containing the observed winning numbers. This approach is inspired by similar methodologies discussed in (Hoang et al., 22019).

If the lottery is not a completely random game, periodic patterns may exist among the winning numbers. Based on this assumption, we hypothesized that the winning numbers at time $t-1$ could be used to predict the winning numbers at time t. Therefore, we populated the feature matrix x with the lottery numbers from draws 1 to 1127, and the target matrix y with the lottery numbers from draws 2 to 1128. Subsequently, we modeled the influence of each number at time $t-1$ on each number at time t as links in the network.

[*] Equal contribution [1] Department of XXX, University of YYY, Location, Country. Correspondence to: Jeonghui Lee <jeonghui.lee@example.com>.

Accepted to the 2^{nd} *Workshop on Generative AI and Law*, colocated with the *International Conference on Machine Learning*, Vienna, Austria. 2024.

2. Network Inference of Time Series Data Following Probabilistic Binary States

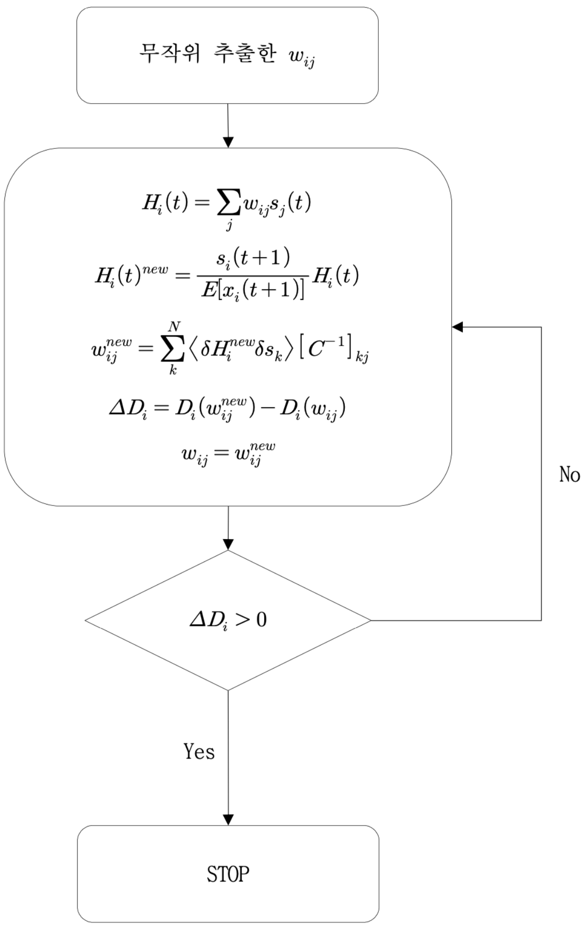

Figure 1. Flowchart of network inference for probabilistic binary time series data.

[Figure 1] shows the flowchart of network inference using the method for probabilistic binary time series data. First, randomly select w_{ij}. Then, using the randomly selected w_{ij}, calculate H_i. H_i is an $n \times t$ matrix. Update $H_i(t)$ once to obtain $H_i^{new}(t)$. Using the obtained $H_i^{new}(t)$, calculate w_{ij}^{new}. Using the randomly selected w_{ij} and the once-updated w_{ij}^{new}, calculate ΔD_i. Update w_{ij} with w_{ij}^{new} and check if ΔD_i is greater than 0. If ΔD_i is greater than 0, recalculate H_i using the updated w_{ij}, update to H_i^{new}, and repeat. If ΔD_i is less than or equal to 0, stop the iteration.

3. Results

The network inference among lottery numbers resulted in a heatmap as shown [figure 2]. The heatmap visualizes the correlation between each feature and target variable with color, where yellow indicates a strong positive correlation and purple indicates a strong negative correlation.

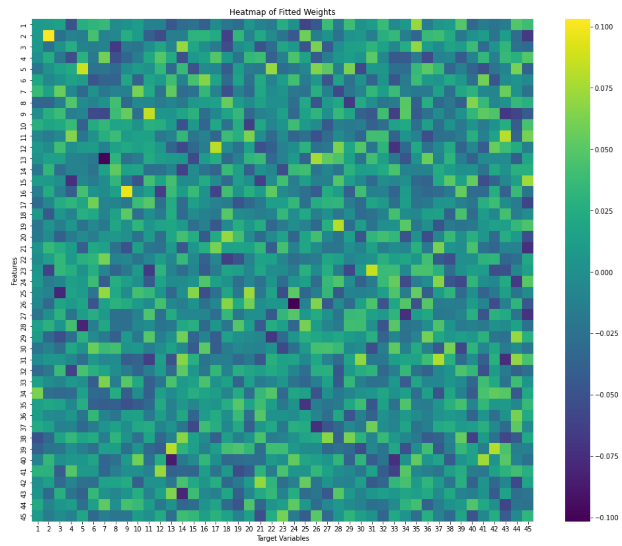

Figure 2. Heatmap illustrating the strength of connections between nodes in the network.

Using the winning numbers from the 1128th lottery draw to predict the 1129th lottery numbers resulted in the following:

- Predicted values: [44,22,27,7,5,14]
- Actual values: [5,10,11,17,28,34,22]

Among the predicted numbers, the number 22, 5 matches the actual value. The probability of randomly predicting at least 2 out of 7 numbers (including the bonus number) is approximately 30%. Considering this, it can be concluded that the model demonstrates a certain level of performance. Therefore, the model can be considered to have practical predictive capabilities for lottery number forecasting.

4. Conclusion

In this study, we approached the relationships among lottery numbers from a network analysis perspective and developed a predictive model. Utilizing the Expectation Reflection algorithm, we analyzed the interactions between each number and explored the potential for predicting lottery numbers.

The model's prediction, based on the winning numbers from the 1128th draw to forecast the numbers for the 1129th draw, resulted in matching 2 out of 7 numbers. Considering that the probability of randomly predicting at least 2 out of 7 numbers (including the bonus number) is approximately

30%, this indicates that the model possesses a certain level of predictive capability.

This study has several limitations. First, the data used in this study is limited to the lottery numbers from 1129 draws, which provides only 6-7 numbers per draw. In contrast, the network inference process requires the estimation of 45×45 links representing the interactions between numbers. Thus, the current data is insufficient to accurately model such a complex network.

Second, the algorithm can predict the influence of numbers from the previous draw on the next draw but needs to account for long-term influences and interactions among numbers within the same draw. Overcoming this limitation requires additional modeling that considers multiple lags and contemporaneous interactions.

Third, the study treated both the main winning numbers and the bonus numbers with equal weights when generating the network. However, since the values of winning first prize and second prize differ, it would be more realistic and accurate to analyze these numbers with different weights.

This study presented a novel approach to predicting lottery numbers. By using network analysis to understand the relationships among lottery numbers, we constructed a predictive model. Despite the current data limitations and model constraints, this approach has shown significant potential for predicting lottery numbers. Future research should leverage more extensive datasets and advanced modeling techniques to achieve more precise and accurate predictions. Additionally, different weights for the main and bonus numbers should be considered to enhance the analysis.

References

Danh-Tai Hoang, Juyong Song, Vipul Periwal, and Junghyo Jo. Network inference in stochastic systems from neurons to currencies: Improved performance at small sample size. *Physical Review E*, 22019.