A New Playing Method of the Guessing Football Lottery

Shunqi Liu, Yuxin Wang and Hong He*

School of Mechanical, Electronical and Information engineering, Shandong University, Shandong Province, China

*Corresponding author

Abstract. This paper analyzes the impact of sports lottery buyers’ opinions on the prediction of sports lottery and betting behavior, and designs a new “sky ladder” strategy. This strategy improves the existing playing methods in both amusement and flexibility. The author discusses the feasibility of the new strategy and its final influence over lottery companies and lottery buyers, simulates and predicts the effects of the brand-new strategy based on BP and LSTM neural network algorithms, and works out a series of simulation results that show the risks of the new playing method are well controlled.

Keywords: Guessing football lottery; Loss percent; Prediction; Dark horse strategy; Multiple strategy; Neural network.

1. Introduction

The author entered such key words as “football lottery”, “sports lottery”, “football betting”, “artificial intelligence risk control” and “guessing sports lottery” in Google Scholar, and obtained the following results by August 2019 which are shown in the Table 1.

Table 1. Classification of Current Sports Lottery and Quantity of Relevant Literature.

| Strategy | Football Lottery | Sports Lottery | Football Betting | Artificial Intelligence Risk Control | Guessing Sports Lottery |
|----------|------------------|----------------|------------------|-------------------------------------|------------------------|
| Quantity of Literature | 153 | 111 | 14 | 3 | 14 |

Constantinou[1][2] introduced TBNs and Dolores models and applied dynamic rating and mixed Bayesian network to predict the long-term performance of football teams. Boshnakov[3] designed a double variant Weibull count model to anticipate the score distribution in a football match. Berrar[4][5] discussed two feature engineering techniques which are used for prediction of match results, and integrated knowledge of different fields. Koopman[6] developed a multi-variate score-driven model to analyze the football match results of high-dimensional groups. Dubitzky[7] pointed out the finiteness of match result prediction and concluded that factors such as fewer goals being scored, lower winning percentage and hard-to-capture events usually determine the final outcome of a match. Bernardo[8] discovered the positive role of coach change in match results. Baker[9] proposed to estimate the strength of a football team with the passage of time in a certain rather than random manner based on a dynamic model. Hing[10] stated that factors such as the gender of lottery buyers would influence the choice over bet. This paper adopts the multiple strategy to design a new strategy based on four kinds of multiple strategy and five kinds of single strategy on the mainstream market. This new strategy incorporates an incentive...
system based on existing strategies, and increases the influence of lottery buyers’ intelligence strategy. Compared to current popular strategies, this strategy has been improved in both amusement and flexibility. We used BP and LSTM neural networks, and trained a stable prediction model after simulating the retrieval of real-time data on varied matches and passing combinations from the database. Then, we conducted risk control over the strategy through the return rate and profit sum of the strategy itself.

The remaining paper is organized as follows. In Section II, the design thoughts of the sky ladder climbing strategy is introduced. In Section III, the sky ladder model is illustrated by a logistic verification. In Section IV, we show a flexibility analysis based on examples. In Section V, two neural networks are introduced into the prediction of rate. Finally, conclusions are given in Section VI.

2. Design Thoughts of Sky Ladder Climbing Strategy

Basic design thoughts: Any nine matches + level passing + extra incentive mechanism. The entire game has three levels and the passing difficulty of each level increases based on the levels. Detailed level information is displayed in Table 2.

| Level | Reward | Betting category | right results/ levels |
|-------|--------|-------------------|----------------------|
| 123   | 1.5    | win, tie and loss | 2/3                  |
| 456   | 2      | goals             | 2/3                  |
| 789   | 3      | half-time match   | 3/3                  |

Table 2. Concrete Strategy Block.

The lottery buyers regard the given fourteen football matches as the guessing objects and respectively select nine from them. The system will automatically set the first three matches as “win, tie and loss prediction”. In other words, the first and second level are about prediction of goals scored while the third level is about the prediction of half-time and full-time match. After the player passes a level, extra reward will be given. The lottery buyer can gain the reward only after guessing the right result at the previous level.

The final incentive mechanism takes into account many factors such as the amount of the pool, the actual bet result of the lottery buyers and the deviation from the correct result and the ratio of bet. It is established under the prerequisite that the return rate is controlled between 75% and 80%. The bonus is calculated pursuant to the number of levels. If the player succeeds in passing the level, the basic bonus will be the amount of bet * the loss percent of the questions answered correctly. If the player answers the questions correctly in succession, the bonus will be the number of bet multiplying the basic bonus, which is the final bonus. The reward for the player passing three levels will be progressive times the basic bonus of each level, and the final reward increases progressively pursuant to the number of level being passed.

Stage incentive is the core that makes the strategy remain captivating. It avoids the situation of “one loses and all lose”, with certain fault tolerance for players. Moreover, different arrangement of matches make players enjoy freedom to some extent, thus bringing greater amusement.

The setting of ladder difficulty enhances the probability of reward for players and also controls the possibility of final reward. It not only maintains the enthusiasm of players in the game but also keeps the vitality of the strategy based on players’ psychology of pursuing higher reward.

3. Risk Control Model of the Sky Ladder Climbing Strategy

The combinations of lottery buyers vary according to individuals’ preferences and strategies. Therefore, for lottery buyers, they tend to put stable prediction of match results in the front so as to guarantee their minimum benefits. In this way, they are more confident in winning the corresponding reward. The conventional practice on the market is to arrange those with large loss percent difference in the front, that is, matches with the least possibility of the weak overcoming the strong.

This study defines it as follows:

\[ S_A = \{ S | S = z_1 z_2 \ldots z_9, z_{1,2,3} \in A \{3,1,0\}, z_{4,5,6} \in B \{0,1,2,3,4,5,6,7\}, z_{7,8,9} \in C \{uu\}, u \in \{3,1,0\} \} \]

is the total of all bet possibilities;
Spre = \{S|S = z_1 \ldots z_9, z_1, z_2, \ldots, z_9, z_1, z_2, \in A\{3,1,0\}, z_4, z_5, z_6, \in B\{0,1,2,3,4,5,6,7\}, z_7, z_8, \in C\{u\}, u \in \{3,1,0\}\} is a predicted bet.

Q = (1 - |Spre|/|SA|) \ast |Spre| is the risk ratio.

Level I: Win, loss and tie are respectively represented by 3, 0 and 1. Level II: 0~7+ of the total number of goals scored is indicated with 0~7 in order. Level III: Win, loss and tie at half-time and full-time matches are expressed in form of double combination being underlined, such as WW represented by 33.

Analysis of Level Passing Probability: Level I: Any lottery or actual match result is a triple array of the set \{3,1,0\}. Through calculation, the passing probability is 10/27. Level II: Players can pass the level by selecting two options correctly, which is much easier than guessing the score because the scoring results are relatively concentrated. It’s found that although there are eight options, lottery buyers’ choice takes on a trend of normal distribution based on the assumption that the number of goals scored is 2.

\[ Y = -\frac{1}{4} \times (x - 2)^2 + 0.12 \]  
\[ Y = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]  

the presentation of the form function image 1 is based on x=2, which means the total number of goals scored is 2. Through comprehensive data analysis, it’s noted that usually two teams are equal in strength without much difference, and the presentation of the form function image 2 is based on general axis of symmetry, that is, the value of \(\mu\). It’s usually between 2 and 4. Therefore, the cluster point of data on lottery buyers’ preferences is within the scope of 2 to 4, the total number of goals scored. When the priority is given to concentration points with a high ratio of bet, the win probability is larger and the probability of three concentration points is about 54%, which implies a great probability of winning. Regards to the probability of winning in the current round, according to rational choice, if what’s near the wave crest of the normal distribution function is considered first, the data collected in this study on 100 lottery buyers in recent two months basically complies with the data probability outcome tested hereunder. The final win probability is as follows:

\[ P = C^3_9 \times (0.125 \times 0.33 + 0.1 \times 0.08 + \ldots ) + 0.4^3 + \ldots \approx 12.1\% \]

Thereof, P means the win probability within a wave crest range of 1 to 2 of a high bet ratio point.

Level III: Compared with the previous levels, the difficulty of guessing the result of the half-time and full-time matches is strengthened in this level. Unlike the number of goals scored, no obvious pattern can be sought in both the distribution of match results and the bet ratio of lottery buyers. The correction rate is approximately 11.11%, which may fluctuate due to numerous factors. According to the statistics of the Major League Soccer in 2012, a balance of power was noted and the fluctuation was between ±3% and 4.2%. Hence, no distinct pattern such as concentration points of normal distribution and wave crest is shown, which makes the game more difficult. As the requirement for Level III is that players must answer all three questions correctly, the win probability is calculated based on three matches, with a mean value of 10.54%, and P (passing Level III) is shown below:

\[ P = \left( \frac{1}{9} \pm 4.2\% \right)^3 \approx 0.1\% \]

It’s concluded from the actual conditions of the three levels and the solving of theoretical win probabilities that the probability of eventual win in theory is around 0.0045% which is within the controllable scope.

4. Feasibility Analysis and Example

Two matches are set for passing in Level II designed in this study and usually players will only focus on four or five of the eight results regarding the total number of goals scored. Therefore, in light of competitive ratio, the possible circumstances are listed below:
The risk degree is measured pursuant to its definition:
\[ Q = (1 - |\text{Spre}|/|\text{SA}|) \ast |\text{Spre}| = (1 - 91125/3^3 \ast 8^3 \ast 9^3) \ast 91125 \approx 34468. \]

According to the aforementioned analysis and calculation, it can be seen that the risk degree is quite high because of the large amount of options in Level II and Level III. As a result, the win rate will decrease. However, as stated previously, on the basis of the number of goals scored in Level II, the number of options is eight in theory, that is, the number of goals scored is predicted as 0~7+. Yet, there are not so many possibilities of concentration according to actual mathematical analysis; instead, the fact is that the probability is within 2 to 3 points near the wave crest with a high bet ratio and more than 75% concentrates between 4 to 5 points, which increases the multiple of the win rate:
\[ \frac{Q_1}{Q_2} = \frac{|S_2(8)|}{|S_2(5)|} \approx 4.10 \]
\[ L_2 = \frac{|S_2(8)|}{|S_2(4)|} = 8 \]

Risk degree and competitive ratio analysis:
Level I: C1=|Sbest|=1; Betting on those with the smallest loss percent is the safest way. Accordingly, the risk degree is evident:
\[ Q_1 = (1 - |\text{Spre}|/|\text{SA}|) \ast |\text{Spre}| = 1 \]

Level II: Calculation of the competitive ratio:
\[ C_2 = |\text{Sbest}| = \sum_{i=4}^{6} P(z_i) = 125 \]
\[ Q_2 = (1 - |\text{Spre}|/|\text{SA}|) \ast |\text{Spre}| = (1 - \frac{125}{512}) \ast 125 \approx 94.48 \]

It can be known from the data analysis above that the risk degree or competitive ratio of the first two levels is quite high, that is, the probability of passing these two levels is greater. This study selected the loss percent data on 10, 50 and 100 matches from the Major League Soccer, calculated the actual funds and obtained the respective return rate of 77.42%, 76.51% and 81.06%, which is basically between 75% and 80%. Therefore, it can be seen that the new strategy is quite reasonable with the guarantee of the return rate.

5. Prediction Based on Neural Network
Prediction of initial loss percent based on BP neural network: The predicted initial loss percent was compared with the real loss percent of American football lottery. Through constant iteration training of historical data, a function that conforms to actual conditions was matched. Then, required influencing factors were input to directly obtain an approximately accurate loss percent. Based on multiple iterations, the degree of accuracy is finally controlled around 90%, shown in Figure 1.

![Figure 1. Times of Iteration and Precision Fitting Image.](image)
Figure 2. Loss Rate of Prediction Results
Dynamic prediction of loss percent based on bidirectional LSTM model: The loss percent prediction of many matches may be jointly determined by historical data and emergency factors. Hence, a bidirectional recurrent neural network was proposed to make the predicted value closer to the real reference value. First of all, training was conducted at the forward layer to save the output results of the forward hidden layer at every moment. Afterwards, a reverse training was carried out at the backward layer to save the results of the backward hidden layer at every moment. In the end, the final output was worked out based on the output results of the backward and forward layers at every moment. The result is shown in Figure 3. As displayed in Figure 4, as tested data is increased, the prediction results tend to be identical and the data similarity constantly fitted. Eventually, the tendency of varied aspects is extremely similar.

Figure 3. Training Simulation Diagram.  
Figure 4. Actual Data and Predicted Loss Percent Data Simulation Diagram.

6. Summary
The study on the new strategy in this paper is stable and reliable in such aspects as the analysis of theoretical win rate and theoretical loss percent as well as the analysis of return rate based on actual tested data. For lottery buyers, this strategy is strongly operable and intriguing. For sports lottery companies, the market is controlled and risks lowered. The subsequent studies may make self-defined adjustments and changes in ladder reward and level difficulty in that the basic idea is the same and it is easy to improve. Many sub-strategies can be derived from this strategy, which is of certain significance to the reform of playing strategy.

References
[1] Constantinou, A., & Fenton, N. (2017). Towards smart-data: Improving predictive accuracy in long-term football team performance. Knowledge-Based Systems, 124, 93–104.
[2] Constantinou, A. (2018). Dolores: a model that predicts football match outcomes from all over the world. Machine Learning. https://doi.org/10.1007/s10994-018-5703-7.
[3] Boshnakov G, Kharrat T, McHale IG (2017). “A Bivariate Weibull Count Model for Forecasting Association Football Scores.” International Journal of Forecasting, 33(2), 458–466. doi:10.1016/j.ijforecast.2016.11.006.
[4] Berrar, D. (2019). Guest editorial: special issue on machine learning for soccer, vol.108, pp.1-7.
[5] Berrar, D., Lopes, P., & Dubitzky, W. (2018). Incorporating domain knowledge in machine learning for soccer outcome prediction. Machine learning.
[6] Koopman, SJ (2019). Forecasting football match results in national league competitions using score-driven time series models, vol.35, pp.797-809.
[7] Dubitzky, W., Lopes, P., Davis, J., & Berrar, D. (2018). The open international soccer database. Machine Learning. https://doi.org/10.1007/s10994-018-5726-0.
[8] Bernardo G, Ruberti M, Verona R. Semi-strong inefficiency in the fixed odds betting market: Underestimating the positive impact of head coach replacement in the main European soccer leagues. The Quarterly Review of Economics and Finance 2019, 71: 239–246.

[9] Baker, R. D., & McHale, I. G. (2015). Time varying ratings in association football: The all-time greatest team is.... Journal of the Royal Statistical Society Series A, 178, 481–492

[10] Hing, N., Russell, A. M. T., Vitartas, P., & Lamont, M. (2016). Demographic, behavioral and normative risk factors for gambling problems amongst sports bettors. Journal of Gambling Studies, 32(2), 625–641. doi:10.1007/s10899-015-9571-9

[11] Data resources: http://football-data.co.uk/data.php