ISSN 2959-6157

A UEFA Champions League knockout draw and odds update algorithm based on Pagerank and Hidden Markov Chain

Yi Ling

Tianjin University

Abstract:

This study introduces an innovative UEFA Champions League knockout draw and odds update algorithm, which builds upon the PageRank algorithm and incorporates a hidden Markov chain. In contrast to traditional odds updates that solely rely on teams' historical performance and expert evaluation, this approach takes into account the actual competition between teams throughout the season. To capture the teams' strength and performance during the season more accurately, this study proposes a model grounded in the PageRank algorithm. By analyzing the results of matches between teams in the Champions League knockout stage, it calculates the rank score for each team. Compared to traditional odds updates, the method employed in this study dynamically reflects the actual competition between teams, enhancing the accuracy of the draw. Furthermore, this study incorporates a hidden Markov chain model, which integrates factors such as the team's historical record, draw results, and odds changes to further refine the odds update process. Hidden Markov chain models capture the temporal relationship between game results, enabling better predictions of teams' future performance. By merging the PageRank algorithm with the Hidden Markov Chain model, this study establishes a comprehensive UEFA Champions League knockout draw and odds update algorithm, serving as a more precise and dependable reference for football gambling. Experimental results demonstrate that the proposed algorithm achieves impressive results in predicting the odds of the Champions League knockout stage, exhibiting higher accuracy and stability compared to traditional methods. This study offers a novel approach and method for updating odds in football gambling, possessing significant theoretical and practical value.

Keywords: Pagerank, hidden Markov chain, elimination draw, odds

1 Introduction

Football, as one of the most popular sports in the world, has attracted billions of spectators and gamblers worldwide. In modern football, the UEFA Champions League, as one of the highest level club football competitions, has attracted the attention of top teams and numerous fans from around the world. In this intense game, the strength and performance of a team often determine their victory or defeat, and directly affect the odds of the bet.

Traditionally, football betting companies determine odds by analyzing factors such as a team's historical performance, player lineup, home and away advantages, etc. However, this method often overlooks the actual competition between teams during the season, resulting in insufficient accuracy and stability of the odds. Especially in knockout stages like the UEFA Champions League, the results and performances of teams are constantly changing, and traditional odds update methods often fail to reflect these changes in a timely and effective manner, affecting the decisions of gambling companies and gamblers.

To address this issue, this paper proposes a UEFA Champions League knockout draw and odds update algorithm based on the Pagerank algorithm and hidden Markov chain. The Pagerank algorithm, as a classic network analysis algorithm, can effectively evaluate the importance of web pages. By applying it to the match relationship between teams in the Champions League knockout stage, it can more accurately reflect the strength and competitive relationship of teams. Although not very common, research has focused on the impact of the draw process in football matches. These studies may involve the impact of drawing lots on the distribution of team strength, as well as the strategic impact on the knockout stage.

Foreign researchers have conducted extensive research in the field of odds prediction. They use statistical models, machine learning, and data mining methods to improve the accuracy of betting results. Some studies also focus on the relationship between market odds and actual match results. The Hidden Markov Chain model can take into account the temporal relationship between game results and further optimize the process of updating odds.

This article aims to explore how to use the Pagerank algorithm and hidden Markov chain model to improve the accuracy and stability of predicting odds in the knockout stage of the Champions League, providing more reliable references for football betting companies and gamblers. Through empirical analysis and comparative experiments, this study will verify the effectiveness of the proposed algorithm in odds prediction, and explore its potential value and application prospects in practical applications.

2 Use Pagerank to calculate the draw results and rank scores

2.1 Data Collection and Matrix Establishment

After the Champions League group stage, teams that advance to the round of 16 need to meet the condition that they do not meet in the same group or league, and that the first and second teams in the group do not meet. Based on the latter limitation, if the 16 teams use a 16 * 16 matrix, the matrix is essentially a block matrix equal to $\begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix}$, where A is the matrix of the match situation between the first and second teams in the group. To avoid calculation redundancy, it can be simplified to an 8 * 8 matrix, where the first team in the group is listed as the second team in the group (the rows and columns can be interchanged, and the team's score is calculated based on the columns). The core idea of PageRank is based on a random walk model on a directed graph, which is a first-order Markov chain. This model describes how a random walker randomly moves along the edges of a graph, accessing from one node to another. Under certain conditions, this random walk process will eventually converge to a stationary distribution. In this stationary distribution, the probability of each node being visited is its PageRank value, which can be interpreted as the probability of drawing lots to that opponent. Simultaneously calculating ranking scores requires iteration, which coincides with Pagerank's basic algorithm.

The most important factor of Pagerank is the damping factor, which is to prevent a parameter that only enters but does not exit a node [2]. This also applies to the draw for the Champions League. In the 2021-2022 UEFA Champions League knockout stage, there was an error in the draw by UEFA. Manchester United and Villarreal, who were originally in the same group, drew together in the knockout stage. This led to a rescheduled draw, which directly changed the original draw result. Therefore, damping factors need to be considered. The setting of damping coefficient is to reduce the impact of re drawing lots.

$$PR(A) = (1-d) + d \cdot \sum_{i=1}^{n} \frac{PR(Ti)}{C(Ti)} \tag{1}$$

This study will refine the image, with different colors representing different national leagues. The first row of the group is the first row, and the second row of the same group is the second row. The draw matrix for the top 16 teams in the 2022-2023 UEFA Champions League is as follows:

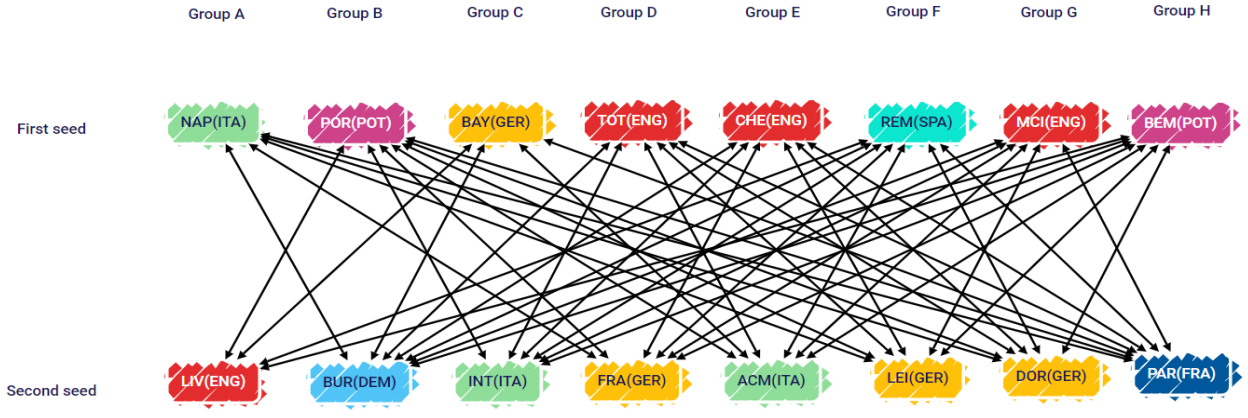

Figure 1 Directed graph of the draw for the 2022-2023 UEFA Champions League knockout stage

For the initial matrix, if there is a probability of encountering it, it is 1, and if there is no probability of encountering it, it is 0. The statistics are as follows:

$$\begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \tag{2}$$

2.2 Matrix processing and calculation results

After preliminary processing of the initial matrix using the Pagerank algorithm:

$$\begin{bmatrix} 0 & \frac{1}{5} & 0 & \frac{1}{5} & 0 & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{7} & 0 & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} \\ 0 & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & 0 & \frac{1}{7} & \frac{1}{7} \\ 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & 0 \end{bmatrix} \tag{3}$$

Due to the increase in iterative computation, the final result only needs to calculate the probability distribution of the lottery. Therefore, a normalization method was used instead of iteration, retaining the initial idea of pagerank, and ultimately obtaining the same result.

In the Pagerank definition, the more lines a point is connected to, the more important its properties are, while in drawing lots, this is exactly the opposite. Each team will exclude second placed teams from the same group, in addition to teams in the same level of league. This study found that Bayern Munich (German Bundesliga) has an extra low rank value. After inspection, the second placed teams in the same level of league to advance are Borussia Dortmund, Frankfurt, and Leipzig, resulting in fewer opponents that Bayern can match, which indirectly reflects the strong overall level of the Bundesliga. Of course, this is not rigorous. If the ranking of the second group is done by transposing the above matrix, it can be found that Liverpool (the English Premier League) has the lowest rank value. This is because there are three first groups in the same level of league (Manchester City, Chelsea, Tottenham Hotspur), which maps to this year's European standings and further illustrates the rationality of their Pagerank. For the points obtained from 2022-2023, England's first place is 22.428 points, and Germany's third place is 17.125 points. That year, Italian Serie A team Inter Milan won the runner up, so the second place is not within the scope of this study. Therefore, this is consistent with Pagerank's calculation results. Therefore, it can be concluded that the Pagerank value of a team is negatively correlated with the level of the league in which the team is located.

Compact | Detailed

#	Nation	19/20	20/21	21/22	22/23	23/24	Points
1	England	18.571	24.357	21.000	23.000	13.625	100.553
2	Spain	18.928	19.500	18.428	16.571	12.687	86.114
3	Italy	14.928	16.285	15.714	22.357	14.000	83.284
4	Germany	18.714	15.214	16.214	17.125	13.642	80.909

Figure 2 The latest European Championship standings [3]

3 Calculate the corresponding odds based on the Hidden Markov Chain to determine the winning or losing rate of the match

3.1 Calculate the probability of a team win-

ning or losing based on historical point data

This study provides an estimated analysis of Manchester City's knockout match against Copenhagen in February 2024:

Using a Markov chain model, consider the second home and then away matches in the group to predict the outcome of the two rounds.

The initial idea of this study was to use the lottery result as a bright chain, with a probability of a dark chain. The hidden Markov chain is essentially a finite state machine, and the dark chain can be converted to each other. However, since there is only one competition, this study used a Viterbi algorithm from a dynamic programming algorithm to find the Viterbi path that is most likely to generate an observation event sequence - the hidden state sequence, especially in the context of Markov information sources and hidden Markov models [4].

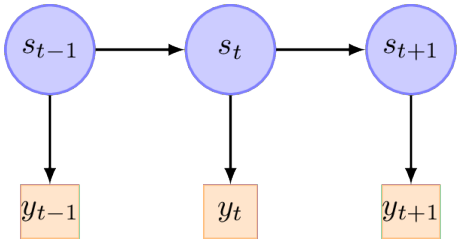

Figure 3 Hidden Markov Chain

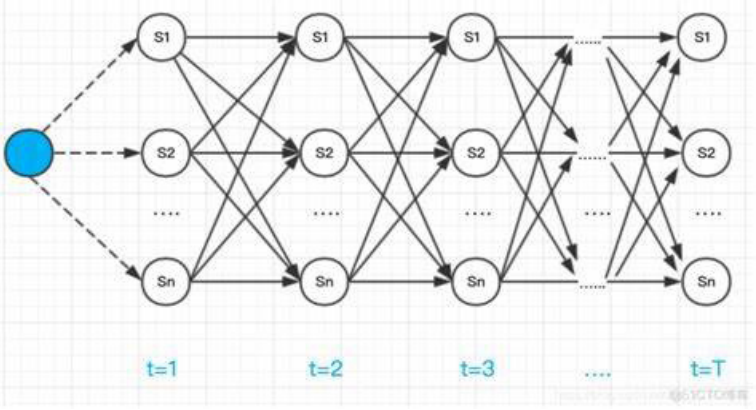

Figure 4 The basic idea of the Viterbi algorithm

This study converts the odds into whether to advance or not, and introduces two necessary factors: the victory or defeat of home and away games.

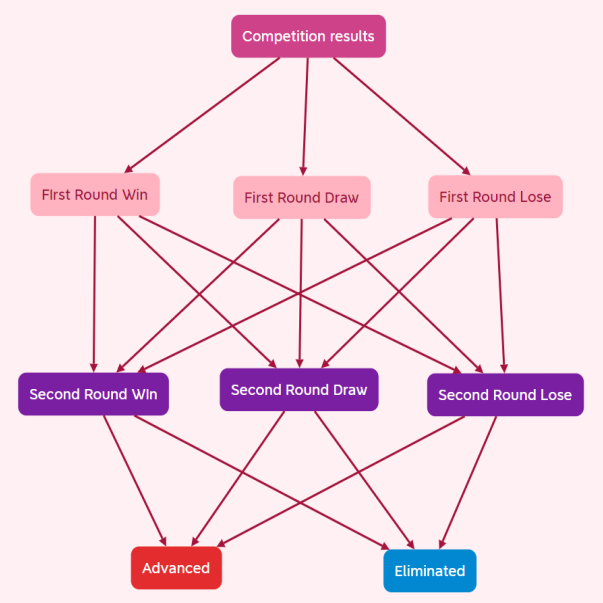

Figure 5 A Two Round Algorithm for UEFA Champions League knockout rounds based on Viterbi Algorithm

The status can include the match status of each team in each round:

Manchester City won the first leg (W1)
Manchester City drew in the first leg (D1)
Manchester City failed in the first round (L1)
Manchester City won in the second leg (W2)
Manchester City drew in the second leg (D2)
Manchester City failed in the second round (L2)

The result of the first game will directly affect the result of the second game.

$$Matrix = \begin{vmatrix} P(W1 \to W2) & P(W1 \to D2) & P(W1 \to L2) \\ P(D1 \to W2) & P(D1 \to D2) & P(D1 \to L2) \\ P(L1 \to W2) & P(L1 \to D2) & P(L1 \to L2) \end{vmatrix} \quad (4)$$

For the parameter setting, FIFA official releases the latest European game points for each team every year, using these points as the main criterion for team scoring. At the same time, rough parameters for home and away games are set based on various data provided by whoscore official (including strengths, weaknesses, home data, and away data). After setting the parameters, an algorithm is provided and the final calculation is carried out.

Among them, P (X1>Y2) represents the probability of transitioning to the second round state Y2 when the first round state is X1. This can be adjusted by analyzing the team's performance in past matches, home and away advantages, opponent strength, and other factors.

At first, we wanted to use the state transition matrix for victory or defeat prediction, but due to the excessive number of iterations that would dilute the matrix and increase the error rate, we improved the algorithm as follows:

Manchester City is away first and then home. In the first leg of the away game against Copenhagen, the probability of a basic win is 45%, the probability of a draw is 35%, and the probability of a loss is 20%. In the second leg of the home game against Copenhagen, the probability of a basic win is 75%, the probability of a draw is 15%, and the probability of a loss is 10% (all probabilities are rounded based on the European game score ratio of the two teams, as well as various data on home and away games and the probability of a win). If Manchester City wins the first leg, the probability of winning in the second leg will be multiplied by 120% on the original base winning probability, and the probability of drawing and losing will be reduced proportionally (code is needed to calculate, in short, 1=winning probability+drawing probability+losing probability). If the first leg is a draw, the probability of winning in the second leg will be multiplied by 105% on the original base winning probability. (According to traditional algorithms and games, a draw in the first leg away to the second leg has a certain psychological advantage, and there was an away goal advantage before 2022. Due to the cancellation of this mechanism, the probability of winning will be reduced proportionally.) The probability of drawing remains unchanged, but the probability of losing will be reduced proportionally. If The probability of losing in the first round and winning in the second round will be multiplied by 80% based on the original winning probability. The probability of drawing and losing will increase proportionally (code is required to calculate, in short, 1=winning probability+drawing probability+losing probability). Among them, W1L2, D1D2, and W1L2 may all result in the same overall score. The probability of W1L2 having the same overall score is 25%, the probability of Manchester City having a higher overall score is 50%, and the probability of Manchester City having a lower overall score is 25%. The probability of the total score of D1D2 being the same is 100%. The probability of the total score of L1W2 being the same is 30%, the probability of Manchester City having a higher total score is 60%, and the probability of Manchester City having a lower total score is 10% (the above data is determined by the team's home and away goals scored).

3.2 Calculate odds based on win rate

As is well known, the odds of each team before and after the knockout stage draw will be adjusted based on the draw results. For two teams with a large difference in European points, the odds of the stronger team will decrease, while the odds of the other team will increase. For teams with a very small difference in European points, the odds may also increase. For European strong teams with a low initial odds, the odds of encountering weaker opponents will not change much. In theory, it should be a three part function, but due to the complexity of the value calculation, this study based on the change table of odds before and after the draw released by Sina Sports in 2023-2024 and 2022-2023. Data was collected and calculated from five groups of matches, including Manchester City vs. Copenhagen (too strong vs. too weak), Arsenal vs. Porto (strong vs. weak), and Atletico Madrid vs. Inter Milan (at the same level), to obtain the relationship between odds ratio and win rate. The best fit line of GeoGebra software was used to find an approximate function, and the parameters were determined using the least squares method for function calculation.

Table 1 The relationship between the ratio of odds before and after and the outcome

X(win-pro)	15.6%	38.7%	54.7%	61.3%	73.5%
Y(odds-rate)	2	1.24	0.913	0.875	0.764

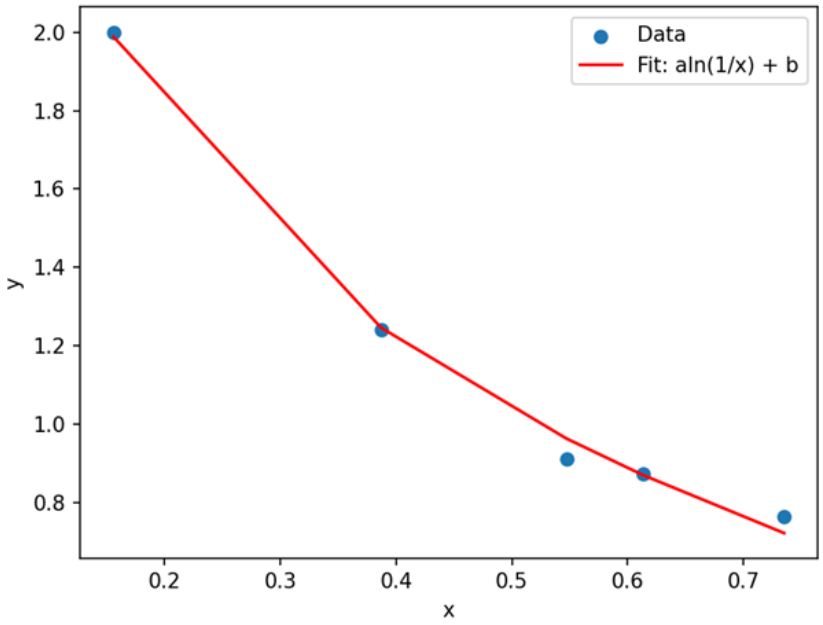

Figure 6 The relationship between the ratio of odds before and after and the outcome

4 Conclusion

This article constructs a comprehensive prediction model based on the Pagerank algorithm and hidden Markov chain based on the draw of the Champions League knockout stage, which can comprehensively consider the draw results and multiple factors. At the same time, a reasonable analysis and explanation of the Pagerank value were conducted to demonstrate its rationality in the calculation of draw probability, providing an understanding of the team's status. And completed model validation and performance evaluation, using historical data for validation to evaluate the performance and accuracy of the model, providing reliable support for practical applications.

However, there are still many areas for improvement: for situations where the result is a draw, we should discuss overtime and penalty kicks in the future. Overtime depends on the influence of the home team, which is the first in the group, and penalty kicks depend on the influence of the players and goalkeepers who take penalty kicks in the team. These data are available in Whascore, but they will become a fourth-order Markov chain, causing the computational workload to double by two indices. Therefore, this study did not collect these data. The data collection of odds is the final championship odds for each team, and the actual calculated results are more suitable for the probability of the team advancing to the quarterfinals. At the same time, the collected data is relatively small, so the calculation method of the obtained odds also has a large error. These shortcomings need to be further addressed in subsequent research.

Reference:

[1] Wayne Detailed explanation of PageRank algorithm https://link.zhihu.com/?target=https%3A//www.mlpod.com/36.html December 29, 2023

[2] Taozi Damping coefficient in PageRank https://blog.csdn.net/m0_72410588/article/details/131015965 June 3, 2023

[3] Live Bar. European Championship Points: Leading the gap in the Premier League, West, Italy, and Germany rank 2-4 respectively, while France and Netherlands compete for 5th place https://baijiahao.baidu.com/s?id=1791058945773842455&wfr=spider&for=pc February 16, 2024

[4] Leader Xiaobai. Detailed explanation of Viterbi algorithm in Hidden Markov Model (HMM) https://github.com/NLP-LOVE/Introduction-NLP February 8, 2020