

Design and Implementation of NBA Playoff Prediction Method Based on ELO Algorithm and Graph Database

Song Yan, Siyuan Meng, Qiwei Liu, Jing Li*

Department of Computer Science and Technology, Shandong University of Technology, Zibo, China

Email: yssdut@163.com, mengsiyuancm@gmail.com, *li_jing@sdut.edu.cn

How to cite this paper: Yan, S., Meng, S.Y., Liu, Q.W. and Li, J. (2019) Design and Implementation of NBA Playoff Prediction Method Based on ELO Algorithm and Graph Database. *Journal of Computer and Communications*, 7, 54-64.

<https://doi.org/10.4236/jcc.2019.711004>

Received: September 10, 2019

Accepted: November 2, 2019

Published: November 5, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

With the globalization of NBA, all eyes on the NBA playoffs are around the world. Ones celebrate the winning of their team which they like. Especially, NBA fans keep on predicting the playoffs game results. However, prediction of winning probability of teams in NBA playoffs is challenging. In order to meet the challenges, we proposed a method using ELO algorithm for prediction and leveraging Graph Database, Neo4j, for implementation. Experiment results show that, the design and implementation of the prediction system could work to some degree.

Keywords

The NBA Playoffs, Graph Database, Neo4j, ELO Algorithm

1. Introduction

Physical fitness has become a simple and effective way to keep fit in our daily life. It not only allows people to release and entertain themselves in today's fast-paced life but also makes their bodies stronger. As a sport, basketball is very popular among teenagers. As the highest level basketball game in the world, the NBA attracts billions of audiences every year in the playoffs, and the wins and losses of each game also create a very considerable operating profit for the gambling companies. The gambling companies give the winning odds of each team according to their unique prediction algorithm. Pan *et al.* put forward a method of NBA playoff prediction based on support vector machine, which has good prediction effect [1]. Qiu *et al.* put forward a new method for calculating the team's comprehensive strength, and established the Logistic model and Bayes

discriminant model [2]. The forecasting method we proposed is different from the above. We use graph database to implement ELO algorithm invented by Elo.

In this paper, our main contribution is that we proposed to use the improved ELO algorithm to predict the winning rate. ELO grading system is a method established by Elo, an American physicist of Hungarian origin, to measure the level of players in all kinds of games. It is an authoritative method to evaluate the level of games, and store all the data in graph database Neo4j. Experiment results show that, the design and implementation of the prediction system could work to some degree.

The rest of the paper is organized as follows: in Section 2, we introduce the preliminary. Section 3 introduces the architecture of this prediction system in detail, which consists of three parts: data preparation, data storage and query. Section 4 gives the algorithm of the system. In Section 5, we will discuss case testing. In Section 6, we review the relevant work and draw conclusions in Section 7.

2. Preliminary

2.1. Graph Database and Neo4j

A graph database is a database whose data model conforms to some forms of graph (or network or link) structure. The graph data model usually consists of nodes (or vertices) and (directed) edges (or arcs or links), where the nodes represent concepts (or objects) and the edges represent relationships (or connections) between these concepts (objects) [3]. Graph database management system is an online database management system, which also has the methods of adding, deleting, changing and searching graph data model. Graph database apply graph into the ability of storing data, which is a kind of high-performance data structure to store a large amount of data. It allows us to construct arbitrarily complex models freely by assembling nodes and connections with simple and abstract characteristics into relational structures, and to visually map the issues we want to describe. Graph databases show the advantages of its performance, flexibility, and agility. And now Neo4j has become one of the most commonly used graph databases.

Neo4j is one of the most prominent open source graph databases available. It allows developers to persist data more naturally from domains such as social networking and recommendation engines, where representing data as a graph of interconnected nodes is a natural choice. Neo4j significantly outperforms relational databases when querying graph data and it supports large data sets while preserving full transactional database attributes [4]. Neo4j is one of the NoSQL graph database management system. It stores data in a variety of graphs in the form of networks or trees. It can vividly and intuitively describe the real world. It is stable and efficient in the efficiency of the query and does not make the query performance to a lower level unlike the relational databases with the increase of the amount of data.

The main features of Neo4j: first, it consists of the nodes, relations, and attributes. Second, the attribute of a relation or a node is a Key-Value data set. Third, every relation has its own head node and tail node. Fourth, relationships can have no attribute.

The details are shown in **Figure 1**: the entities are represented as the four colored nodes in the diagram, where the red ones represent teams and the pink ones represent playoff rounds. The attributes in the figure are entities' names: "San Antonio", "Golden State", "First Round" and "Conference Finals". The relationship in the graph shows that WIN and RWIN represent the winning relationship of playoff and regular season respectively.

2.2. ELO Algorithm

With the development of the network and the improvement of people's living standards, many people will compete in all kinds of competitions on the network. At present, in all major competitive platforms, there is a lack of a ranking system to judge the competitive level of users in competitive competitions. International ranking is also called "FIBA ranking" or "ELO score". It was designed by Elo (1903-1992), an American Professor born in Hungary. It was drafted by the International Chess Federation Hierarchy Committee. It was adopted by the 1969 Plenary Session of the International Chess Federation and was formally implemented since 1970 [5].

ELO Rating Algorithm is widely used rating algorithm for ranking players in many competitive games. Players with higher ELO rating have a higher probability of winning a game than a player with lower ELO rating. ELO grading system is a method for calculating the overall level of both sides in a competition. It is an official method for evaluating the level of competition between two or groups at present. At present, it is mainly used in chess, football, basketball and electronic sports.

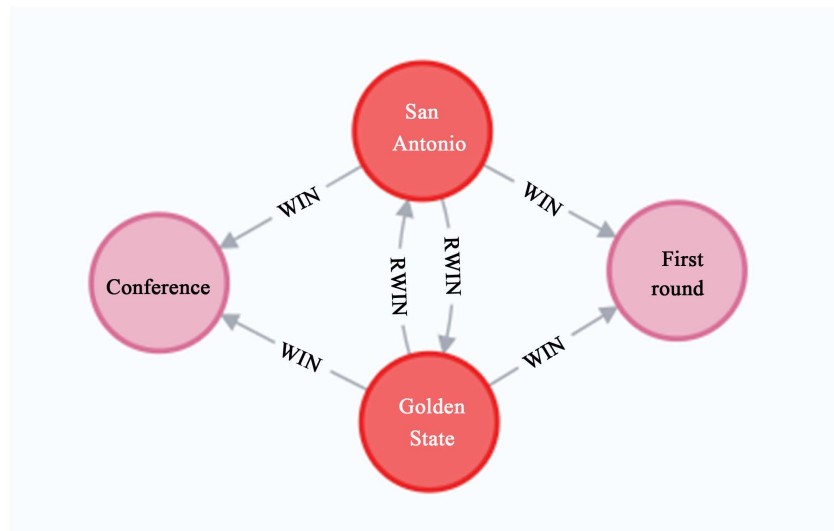


Figure 1. Neo4J diagram data example.

The computing method is listed as follows:

R_i : current score of player i ;

R'_i : score of player i after game;

E_{ij} : player i 's expectation of player j 's winning percentage.

The score difference between player i and player j : $D_{ij} = R_j - R_i$;

$$E_{ij} = \frac{1}{1 + 10^{\frac{D_{ij}}{400}}} \quad (1)$$

$$R'_i = R_i + K(S_{ij} - E_{ij}) \quad (2)$$

3. System Architecture

In this section, we mainly introduce the architecture of this prediction system, as shown in **Figure 2**. It consists of three parts: data preparation, data storage, and query.

Data preparation mainly includes data selection. We select the data of playoffs and regular season according to our forecast demand. Then, according to the team's fighting situation, the win-lose relationship between teams is determined.

The data storage part mainly constructs a graph to store the team's regular and playoff data and the relationship between teams in the database. In the Neo4j graph database, we can find the battle situation between a team and any team.

Preprocessing is mainly used for data prediction and preprocessing. For each team, the name of the team is created as the vertex, and the number of wins and losses between teams is created as the winning relationship of the team. If the team enters the playoffs, then on this basis, the relationship between the team and the new playoffs will be added.

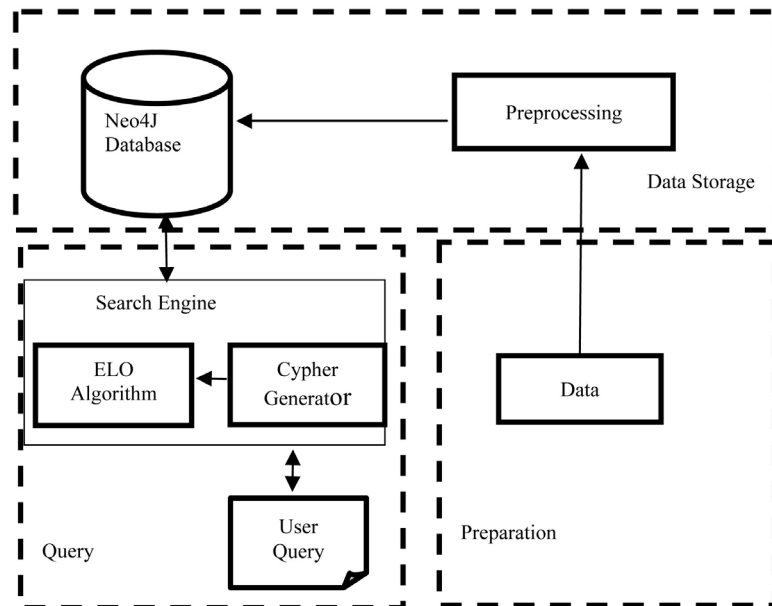


Figure 2. Framework of structure.

The query part mainly queries the data needed for team winning rate calculation, queries each part of the data through Cypher language, then calculates each part of the data through ELO algorithm, and finally obtains the team winning probability.

4. Modified ELO Algorithm

The ELO algorithm was originally used in chess to calculate and evaluate the rank of two players. So we need to modify it if we want to use it in basketball game prediction. The modified ELO algorithm is listed as follows:

t.name: the name of team;

R_i : The currently score of team i ;

R'_i : The new score of team i ;

E_{ij} : Regular-season team i 's expectation for team j 's winning percentage;

P_i : Whether team i join in the playoffs in current season;

P_{ij} : Playoff team i 's expectation of team j 's winning percentage;

Avg: Average winning rating of playoffs;

Reg: Average winning rating of regular-season.

The gap of score between player i and player j is $D_{ij} = R_j - R_i$;

$$E_{ij} = \frac{1}{1 + 10^{\frac{D_{ij}}{400}}} \quad (1)$$

$$R'_i = R_i + K(S_{ij} - E_{ij}) \quad (2)$$

Before calculating, we should consider the following question: when calculating the final winning probability, we need a playoff-regular ratio, and then what is the appropriate proportion? According to our predictive thinking, there are two kinds of teams that have entered the playoffs in the current season. One is to enter the playoffs in the past, and the other is to enter the playoffs for the first time in the current season. For the second case, we take DEN and SAS as examples.

The 2018-2019 season is DEN's first playoff season, and SAS has never missed the playoffs before. DEN ranked second in the West in the 2018-2019 season, and SAS ranked seventh in the West. If the playoffs: regular season = 4:6, the final probability of DEN winning is 40.52%, while the probability of SAS winning is as high as 49%. If the playoffs: regular season = 3:7, the probability of DEN winning is 43.10%, and the probability of SAS winning is 49.15%. If the playoffs: regular season = 2:8, the probability of DEN winning is 45.69%, and the probability of SAS winning is 49.15%. When the playoffs: regular season = 1:9, we consider the more extreme situation: in all the playoff data, select the team with the highest overall winning rate GSW, the winning rate is 63.19%. If we calculate the total probability of GSW according to the 1: 9 winning ratio, the result of regular season is too large to reflect the strong dominance of GSW in the playoffs. After the above calculation, we finally chose the playoffs: the regular season = 2:8. Among them, for teams like DEN who have not been promoted to the

playoffs, we calculate the winning rate of the regular season with the opponent: the winning rate of the regular season = 2:8. The verification method is the same as above.

Specific calculations algorithms are as follows: **Algorithm 1**, **Algorithm 2**.

5. Experiment

5.1. Experiment Environment

We run experiments with the following configurations, which are showed in **Table 1**.

5.2. Initial Score

The number of regular season wins in the 2018-2019 season is used as the initial score for each team (data from <https://china.nba.com/>), as shown in **Table 2**.

```

Input: t.name
Output: the word 'percent' as the final result
1  if t.name  $\neq$  NULL then
2      while t.name $\neq$ NULL do MATCH (t1:Team
3          {name:{t.name}})-[r:WIN]->()-[l:WIN]-() RETURN
4          sum(r.win)+sum(l.win) AS sum
5          calculate Eij
6          if sum>0 then
7              calculate Pij,Avg
8              percent = 0.2*Avg+0.8*Eij
9          end
10         if sum=0 then
11             calculate Reg
12             percent = 0.2*Reg+0.8*Eij
14         end
15     return percent;
        end
    end

```

Algorithm 1. ELO Algorithm for the calculation of the winning rate.

```

Input: Ri , Eij, Sij
Output: R'i
while Sij = 1||Sij = 0 do
1      if Ri<41 then
2          R'i = Ri + 4 * (Sij - Eij)
3      end
4
5      if Ri>=41 then
6          R'i = Ri + 2 * (Sij - Eij)
7      end
8      return R'i;
    end

```

Algorithm 2. ELO Algorithm for new scoring.

Table 1. Operating environment configuration.

Equipment	Configuration
CPU	Intel (R) Core (TM) i5-4200H CPU @2.80 Hz 2.79 GHz
Memory	8.00GB
Operating system	Windows 10
Database	Neo4J
Development tools	IntelliJ IDEA
Explorer	Chrome
Web service	Tomcat

Table 2. Initial score of playoff team in 2018-2019.

East	Initial score	West	Initial score
MIL	60	GSW	57
TOR	58	DEN	54
PHI	51	POR	53
BOS	49	HOU	53
IND	48	UTA	50
BKN	42	OKC	49
ORL	42	SAS	48
DET	41	LAC	48

In the formula $R'_i = R_i + K(S_{ij} - E_{ij})$, K is the limit value, which means that a player can win the most points or lose points. At first, we show the reference of K value and then prove it.

$$\begin{cases} K = 2 & WIN \geq 41 \\ K = 4 & WIN < 41 \end{cases}$$

We select the team with the biggest and smallest difference and the same winning game in the regular season of 2018-2019 to make explanation. The details are as follows:

The groups with the greatest difference in winning field are MIL and NYK.

We think of MIL as team A and NYK as team B. $R_a = 60$, $R_b = 17$. R'_a is the new score of team A and R'_b is the new score of team B. According to formula (1), $E_{ab} = 0.5615$; $E_{ba} = 0.4384$.

In the first case, MIL wins NYK:

Formula (2) gives $R'_a = 60 + 0.877 \approx 61$, that is, MIL wins only one point after winning NYK, while NYK loses only one point.

In the second case, NYK wins MIL:

Formula (2) gives $R'_b = 17 + 2.2464 \approx 19$, that is, NYK wins 2 points after winning MIL and MIL loses 2 points.

5.3. Case Study

All data in this paper are selected from the 2015-2018 playoffs and 2018-2019 regular season data (data resource from <https://china.nba.com/>). We chose two teams GSW and HOU as a simple example in this section. Cypher query statements for postseason winning rate:

```
MATCH (:Team{name: "Golden State"})-[w: WIN]->()<-[l: WIN]-()
RETURN SUM(w.win)*1.0/((SUM(w.win)+SUM(l.win))*1.0 AS Playoff
winning rate.
```

Cypher query statement on playoff match between two teams:

```
MATCH (:Team{name: "Golden State"})-[w: WIN]->()<-[l:
WIN]-(:Team{name: "Houston"}) RETURN SUM(w.win) AS GSW
playoff wins, SUM(l.win) AS HOU playoffs wins.
```

Specific query data are shown in **Table 3**.

For convenience, we define a presents GSW, and b represents HOU.

Winning gap in regular-season between GSW and HOU is:

$$D_{ab} = R_b - R_a = 53 - 57 = -4;$$

GSW's Winning Rate Expectation for HOU in Regular Season is:

$$E_{ab} = \frac{1}{1 + 10^{\frac{D_{ab}}{400}}} = 0.5058;$$

GSW's expectation of HOU's winning rate in the playoffs is:

$$P_{ab} = \frac{1}{1 + 10^{\frac{D_{ab}}{400}}} = 0.5058;$$

Average winning rate in the playoffs is:

$$Avg = \frac{E_{ab} + P_{ab}}{2} = 0.6319;$$

The final winning rate is:

$$\text{Percent} = 0.8 * E_{ab} + Avg * 0.2 = 0.5310;$$

The new score after GSW winning this round is:

$$R'_a = 47 + 2 * (1 - E_{ab}) \approx 58;$$

So $R'_b \approx 52$;

Table 3. Data of GSW and HOU.

Team	2018-2019 regular-season wins	2015-2018 playoffs wins	2015-2018 wins between two teams in playoffs	2015-2018 rating in playoffs
GSW	57	47	8	0.758
HOU	53	18	4	0.545

If GSW meets HOU again, then $R_a = 58$, $R_b = 52$. If GSW meets other teams, then $R_a = 57$. R_b is the initial score for team B.

5.4. Prediction Results

Table 4 is the comparison between the experimental results and the actual situation. According to the promotion situation in the table, there are three groups of prediction errors. The reason of DEN vs SAS prediction errors is that DEN participates in the playoffs for the first time. Compared with SAS, DEN has less experience in the playoffs, and its winning probability is lower than SAS in the calculation process. POR vs OKC's prediction error is due to POR: OKC = 0:4 in the regular season of 2018-2019, that is, POR is swept by OKC. In the calculation, the team's performance in the regular season of 2018-2019 accounts for a large proportion, so POR's prediction victory rate is lower than OKC.

6. Related Work

With the further globalization of the NBA, a playoff team brings more and more economic benefits, so it is very meaningful to predict whether a team can enter the playoffs. Gao *et al.* according to the data of 2011-2012 and 2012-2013 season, the Fisher discriminant model is established, and the cross-misjudgment rate is 20%. After converting the original index into the dominant score, the cross-misjudgment rate is reduced to 13.3%. Through the analysis of the misjudgment information, it is found that the western team has stronger strength [6].

Table 4. The actual promotion status and predicted promotion status of 2018-2019 season.

Team	Actual promotion	Forecast promotion
GSW vs LAC	GSW	GSW
HOU vs UTA	HOU	HOU
POR vs OKC	POR	OKC
DEN vs SAS	DEN	SAS
MIL vs DET	MIL	MIL
BOS vs IND	BOS	BOS
PHI vs BKN	PHI	PHI
TOR vs ORL	TOR	TOR
GSW vs HOU	GSW	GSW
DEN vs POR	POR	POR
MIL vs BOS	MIL	MIL
TOR vs PHI	TOR	TOR
GSW vs POR	GSW	GSW
MIL vs TOR	TOR	TOR
TOR vs GSW	TOR	GSW

Ji *et al.* examined the use of Neural Networks as a tool to predict the starting and reserve line up of All-Star game, in the National Basketball Association, from all the candidates [7]. Hu *et al.* think that predicting the outcome of a future game between two sports teams poses a challenging problem of interest to statistical scientists as well as the general public. To be effective such prediction must exploit special contextual features of the game [8]. Not only in the NBA, but also in other sports, there are many prediction methods. Stephanie Kovalchik proposed a Searching for the GOAT of tennis win prediction method. The evaluation models are divided into three categories: regression-based, point-based and pair-based comparison models. ELO algorithm is also used to judge, and the accuracy rate is 75% [9].

7. Conclusions

In this paper, we propose a method of using graph database to predict NBA playoffs, which uses graph database to store and ELO algorithm to predict NBA playoffs. This experiment uses graph database for data storage. Through the analysis of the real situation, the team is considered as a whole and the influence of players' ability and coaches' ability on the team is not considered. To achieve this goal, we have selected the most "new" data as far as possible, that is, the season data that represents the team's latest personnel allocation. In this way, we can ignore the influence of players and coaches in recent matches.

The limitation of this experiment is that it only considers the recent strength of the team, without paying attention to the impact of changes in players and coaches. For example, the current season, the 2018-2019 finals, TOR vs GSW, will advance based on the predicted results. However, the reality is that in this round of the series, GSW lost to TOR due to the absence of some star players. In the future, we plan to take such situations into consideration.

Acknowledgements

Jing Li is the corresponding author of this research work.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Zeng, P. and Zhu, A.M. (2016) NBA Playoff Prediction Method Based on Support Vector Machine. *Journal of Shenzhen University (Polytechnic Edition)*, **33**, 62-71. <https://doi.org/10.3724/SP.J.1249.2016.01062>
- [2] Sheng, Q., Chong, Y.D. and Zheng, C. (2010) NBA Playoff Performance Analysis and Prediction: Logistic and Bayes Model. *Statistical Education*, **10**, 46-51.
- [3] Wood, P.T. (2009) Graph Database. Birkbeck, University of London, London, UK.
- [4] Partner, J., Vukotic, A. and Watt, N. (2014) Neo4j in Action. Pearson Schweiz Ag, Zug.

- [5] Ouyang, M.S. (2013) Research on the Evaluation System of Competitive Events Based on ELO Algorithm. Doctoral Dissertation, Wuhan University of Technology, Wuhan.
- [6] Hong, X.G., Di, Y., et al. (2015) Qualification Prediction of NBA Playoffs Based on Fisher Model. *Journal of Chongqing University of Technology (Natural Science)*, **29**, 126-130.
- [7] Ji, B. and Li, J. (2014) NBA All-Star Lineup Prediction Based on Neural Networks. *International Conference on Information Science & Cloud Computing Companion*. Guangzhou, 7-8 December 2013, 864-869. <https://doi.org/10.1109/ISCC-C.2013.92>
- [8] Hu, F.-F. and Zidek, J.V. (2004) Forecasting NBA Basketball Playoff Outcomes Using the Weighted Likelihood. *Lecture Notes-Monograph Series*, **45**, 385-395. <https://doi.org/10.1214/lnms/1196285406>
- [9] Kovalchik, S.A. (2016) Searching for the GOAT of Tennis Win Prediction. *Journal of Quantitative Analysis in Sports*, **12**, 127-138. <https://doi.org/10.1515/jqas-2015-0059>