Ann Appl Sport Sci 10(1): e968, 2022.
http://www.aassjournal.com; e-ISSN: 2322–4479; p-ISSN: 2476–4981. 10.52547/aassjournal.968

ORIGINAL ARTICLE

The Evaluation of the Team Performance of MLB Applying PageRank Algorithm

[1]Yun Hyo-jun, [1]Park Jae-Hyeon, [1]Yoon Jiwun, [1]Jeon Minsoo*

[1]Department of Sports Science, Korea National Sport University, Seoul, Korea.

Submitted 19 January 2021; Accepted in final form 29 March 2021.

ABSTRACT

Background. There is a weakness that the win-loss ranking model in the MLB now is calculated based on the result of a win-loss game, so we assume that a ranking system considering the opponent's team performance is necessary. **Objectives.** This study aims to suggest the PageRank algorithm to complement the problem with ranking calculated with winning ratio in calculating team ranking of US MLB. **Methods.** PageRank figure is calculated by using the result of 4,861 matches in the 2017 season (2,430 matches) and 2018 season (2,431 matches) in the MLB. **Results.** There is a difference between ranking calculated in PageRank and ranking calculated with winning ratio both in the 2017 season and 2018 season, and there is a difference in performance per each district due to comparing performance per each league and district. In addition, as a result of calculating the predictive validity of PageRank and winning ratio ranking, it turns out that the ranking calculated with the PageRank algorithm has relatively high predictive validity. **Conclusion.** This study confirmed the possibility of predictive in the US MLB by applying the PageRank algorithm.

KEYWORDS: ***PageRank, MLB, Baseball, Ranking.***

INTRODUCTION

Evaluation of the athletic performance of teams and athletes in sporting events is one of significant interest to leaders who perform sports, athletes, and the public who watch sports (1). There are various ways of evaluating teams and athletes in sporting matches that are different in each sport. Especially, a ranking system can be introduced as a representative method to evaluate teams and players in a sports event. A ranking system is a method used to evaluate a team or players through the result of a match (2). In addition, the sports field has been used as information for determining the annual salary of professional athletes, selecting the national team, and deciding Olympic participation rights based on ranking, so it has a considerable interest (3, 4).

In sporting matches, a method to calculate ranking is applied differently according to sports. In the case of sports such as swimming, track and field, and weightlifting, ranking is calculated based on the game record. In the case of Taekwondo, badminton, and tennis, ranking is calculated using the accumulated points based on the results of a match. Moreover, in the case of baseball, soccer, and volleyball, ranking is calculated based on the difference between a win and a loss which is a traditional sports ranking method (5). It calculates the ranking by comparing the frequency of wins and losses in the total frequency of playing the game, and this is called a win·loss ranking model. The win and loss ranking models are widely used in league games where the number of matches is the same (1).

*. Corresponding Author:
Minsoo Jeon, Ph.D
E-mail: minsu1144@nate.com

In particular, as an example of representative sports applying the win-loss ranking model, U.S. major league baseball (MLB) can be introduced. The MLB provides the opportunity to compete in the division series to each first-ranked ranker of the 3 teams in national leagues (Western, Central and Eastern) and the 3 teams in American leagues (Western, Central, Eastern) on the win and loss ranking model. Specifically, major league teams play 76 games with teams in the same district and the same league, 66 games with teams in other districts and the same league, and 20 games with teams in other leagues. This is how to provide competing opportunities for division series to the first-ranked teams in each league district with a high winning ratio after performing 162 matches per team.

However, MLB applying the win and loss ranking model may have one question. In the case of MLB, the number of matches per team is the same, but the ranking is calculated without considering the opponent's performance level. For example, assuming that all the five teams of the western district in the league have good performance and five teams of the central district have relatively lower performance, the meaning of value for winning may be different in each district. In other words, winning the first and tenth rankings will have a relatively different value.

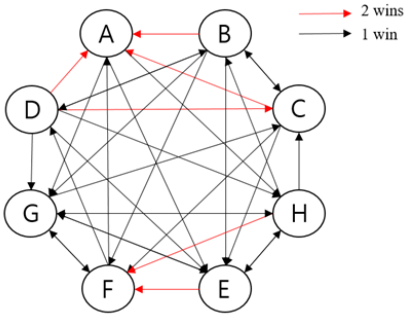
This kind of problem may occur in the performance of league in-out matches and each district's matches. However, there is a weakness that the win-loss ranking model in the major leagues is simply calculated based on the result of the win-loss game, so we assume that a ranking system considering the opponent's team performance is necessary.

Then, we can apply the PageRank algorithm based on the network theory to solve this problem. The PageRank algorithm is a method to present a priority page when Google calculates a search result, which is an algorithm that PageRank ranking varies according to the frequency of web page citations (6-9). PageRank algorithm is a calculating method of ranking by considering the opponent's performance level rather than calculating ranking based only on the result of the match, so it has the advantage of compensating for the problems of ranking currently being calculated in MLB (1, 10). This is a way of weighting relatively good teams according to their team performance level. In other words, even if Team A wins, the ranking may be variable depending on whether the opponent has a good performance (11-13).

Looking into Precedent Study on PageRank algorithm, it is used in methodology study to develop optimization model (14-18) application in the aspect of pedagogy (19-21) and sports field (22-26).

Therefore, this study aims to calculate MLB ranking using the PageRank algorithm. Precisely, to apply the PageRank algorithm, the team rankings are calculated based on the results of the regular season matches in 2017 and 2018, and the model's predictive validity is verified.

An Example of MLB Ranking Calculation Applying PageRank Algorithm. Looking at the match results of team A, they recorded seven wins and three losses in total; 2 wins against teams B, C, and D, one win against team F, and one win against teams E, G, and H. In the case of team E, they recorded six wins and four losses in total; 1 win against teams A, B, C, and D, two losses against team F, and one win and one loss against team H. These are illustrated in (Figure 1) below. If you look at (Figure 1), the object sending the arrow is interpreted as a loss, and the object receiving the arrow is interpreted as a win. If a matrix represents the game result data, it is as (Matrix 1) below. In (Matrix 1), the row is the frequency of victory, and the column is the frequency of defeat.



	A	B	C	D	E	F	G	H
A	0	0	0	0	1	0	1	1
B	2	0	1	1	1	1	1	0
C	2	1	0	0	1	1	0	0
D	2	1	2	0	1	0	1	1
E	0	0	0	0	0	2	1	1
F	1	0	0	1	0	0	1	0
G	0	0	1	0	1	1	0	1
H	0	1	1	0	1	2	1	0

Figure 1. The schematization of Match Results between Teams. Matrix 1. Initial Matrix Q

Q_{12} is the number of times Team A has won Team B, and Q_{51} is the number of times Team A has lost to Team E. In other words, in a matrix, the row is the team's victory, and the column is the defeat frequency. However, 1 of Q_{51} and 1 of Q_{32} is the same one loss, but the meaning can be interpreted differently. It is because 1 of Q_{51} is one among the three that team A sent, and 1 of Q_{32} is one among seven that team B sent. Therefore, it is necessary to convert to the matrix that considers the total weight sent by the team and can be expressed as (Figure 2) below. It is shown that Q_{51} in (Figure 2) is converted to 1/3=0.333, and Q_{32} is converted to 1/7=0.143.

	A	B	C	D	E	F	G	H
A	0.000	0.000	0.000	0.000	0.333	0.000	0.333	0.333
B	0.286	0.000	0.143	0.143	0.143	0.143	0.143	0.000
C	0.400	0.200	0.000	0.000	0.200	0.200	0.000	0.000
D	0.250	0.125	0.250	0.000	0.125	0.000	0.125	0.125
E	0.000	0.000	0.000	0.000	0.000	0.500	0.250	0.250
F	0.333	0.000	0.000	0.333	0.000	0.000	0.333	0.000
G	0.000	0.000	0.250	0.000	0.250	0.250	0.000	0.250
H	0.000	0.167	0.167	0.000	0.167	0.333	0.167	0.000

	A	B	C	D	E	F	G	H
A	0.019	0.019	0.019	0.019	0.302	0.019	0.302	0.302
B	0.262	0.019	0.140	0.140	0.140	0.140	0.140	0.019
C	0.359	0.189	0.019	0.019	0.189	0.189	0.019	0.019
D	0.231	0.125	0.231	0.019	0.125	0.019	0.125	0.125
E	0.019	0.019	0.019	0.019	0.019	0.444	0.231	0.231
F	0.302	0.019	0.019	0.302	0.019	0.019	0.302	0.019
G	0.019	0.019	0.231	0.019	0.231	0.231	0.019	0.231
H	0.019	0.160	0.160	0.019	0.160	0.302	0.160	0.019

Figure 2. Matrix 2, Conversion Matrix Q, Matrix 3, Conversion Matrix Q considering Damping Factor

Repetitive operation is performed as the PageRank algorithm is calculated by Markov Chain's radical root method. However, if the linked nodes do not send the link to other nodes, the sinking phenomenon occurs because PageRank continues to accumulate in the loop (6). To solve this problem, the PageRank algorithm applies the concept of the Damping Factor.

The Damping Factor is the random walker, which means the probability of moving to any other node, and it means randomly following a link on a web page or moving to a new random page by ending the current page due to various reasons. (6). Theoretically, it has a range of 0<d<1, and it is generally set to .15 to analyze, but it is also used to adjust the Damping factor depending on the situation. The process of generating a matrix considering the weight for the link and Damping Factor is shown in (Formula 1). In this example, Damping Factor is set to 0.15, and finally calculated matrix is as (Matrix 3).

$$(Formula\ 1)$$
$$Q_{i,j} = (1-d)\frac{A_{i,j}}{N} + \frac{d}{N}$$

Next, we can calculate the eigenvector matrix π that converges at any stage by calculating the radical root method of the Markov Chain, as shown in (Formula 2). Initial π^1 means the initial weight given to each team among the total weights. Specifically, team A wins seven times in 40 matches, so it is 0.175 (7/40), and team B wins three times, so it is .075(3/40). π^2 is produced by multiplying π^1 by (Matrix 2), and π^3 is produced by multiplying π^2 by (Matrix 2). When this process is repeatedly performed, a convergence step is formed in which the amount of change becomes insignificant, as shown in the following (Table 2), and the eigenvector matrix can be ranked in a large order.

$$(Formula\ 2)$$
$$\pi^T = \pi^T Q$$

In this example, team F is the 1st, and team G is 2nd, team E is 3rd, team A is 4th, team H is 5th, team C is 6th, team D is 7th, and team B is 8th. The rankings calculated using the match results and PageRank algorithm are shown in (Table 3). The calculation results show that team A tied for first place with Team F with a winning ratio of .700, but team A won 2 times against team B and team D with relatively low performance, so the PageRank value was shown to be relatively low. This is a feature of the PageRank algorithm that assigns weights based on their relative importance, and it is a way to reasonably evaluate team performance level in a system in which the number of matches of the opposing team varies according to the belonged leagues, such as MLB.

Table 1. Example of MLB Match Result Data

League	Match Result with the Opponent								Total
	A	B	C	D	E	F	G	H	
American League									
A	-	2 w	2 w	2 w	1 l	1 w	1 l	1 l	7 w, 3 l
B	2 l	-	1 w, 1 l	1 w, 1 l	1 l	1 l	1 l	1 l	3 w, 7 l
C	2 l	1 w, 1 l	-	2 w	1 l	1 l	1 w	1 l	5 w, 5 l
D	2 losses	1 w, 1 l	2 l	-	1 l	1 w	1 l	1 l	2 w, 8 l
National League									
E	1 w	1 w	1 w	1 w	-	2 l	1 w, 1 l	1 w, 1 l	6 w, 4 l
F	1 l	1 w	1 w	1 l	2 w	-	1 w, 1 l	2 w	7 w, 3 l
G	1 w	1 w	1 l	1 w	1 w, 1 l	1 w, 1 l	-	1 w, 1 l	6 w, 4 l
H	1 w	1 l	1 l	1 w	1 w, 1 l	2 l	1 w, 1 l	-	4 w, 6 l

w: wins**, l:** lost

Table 2. The Convergence Process of Eigenvector Matrix

π	A	B	C	D	E	F	G	H
π^1	0.175	0.075	0.125	0.050	0.150	0.175	0.150	0.100
π^2	0.140	0.059	0.085	0.077	0.150	0.173	0.178	0.137
π^3	0.127	0.061	0.100	0.075	0.146	0.181	0.174	0.136
π^20	0.134	0.062	0.098	0.077	0.145	0.179	0.173	0.132
π^21	0.134	0.062	0.098	0.077	0.145	0.179	0.173	0.132
π^22	0.134	0.062	0.098	0.077	0.145	0.179	0.173	0.132

Table 3. Match Result and PageRank Rangking

Ranking	Team	Match Results	WR	PR Value	Ranking	Team	Match Results	WR	PR Value
1	F	7 win 3 losses	0.700	0.179	5	H	4 win 6 losses	0.400	0.132
2	G	6 win 4 losses	0.600	0.173	6	C	5 win 5 losses	0.500	0.098
3	E	6 win 4 losses	0.600	0.145	7	D	2 win 8 losses	0.200	0.077
4	A	7 win 3 losses	0.700	0.134	8	B	3 win 7 losses	0.300	0.062

W.R.: Winning Ratio**, PR:** PageRank

MATERIALS AND METHODS

Study Data. This study used the match results of 2017 regular season and 2018 regular season to calculate PageRank rankings by the team. The number of game data is 4,861 games, with 2,430 games in 2017 and 2,431 games in 2018. To confirm the predictive validity of the PageRank algorithm, the 2017 and 2018 post-season match results were used. All the data was provided from MLB official website, and the data of this study are public data that can be used as secondary data published on the MLB website.

Data Processing Method. This study aims to propose an MLB ranking model using the PageRank algorithm. In order to achieve the purpose of this study, regular season and post-season results of 2017 and 2018 were collected, and the collected data were analyzed using MS-Excel and Python 3.7. Regular season data were classified into Source and Target with the standard of win and loss using MS-Excel, and Source was defined as the losing team and Target as the winning team. The separated data calculated PageRank value by converting to 1 mode matrix of Team × Team using Python. Damping factor d was set to 0.15. At this time, Damping Factor d is interpreted as the probability that the Google engine is not satisfied with the page while searching and clicks on another page link (1, 14), and in this study, it means the probability that a lower team wins a higher team or a higher team loses to a lower team. PageRank generally sets the Damping Factor d to 0.15 (27, 28). Our team produced the Python code for analysis. In addition, MS-Excel was used to compare relative performance levels by team, league, and district

and confirm the PageRank model's predictive validity.

RESULTS

The MLB plays 162 matches per team in one season; however, the opponents vary greatly depending on the region and league they belong to. Specifically, 76 matches are played against teams in the same league and the same region during the regular season, 66 matches against teams belonging to the same league, and 20 matches with other leagues. This structural problem has made it difficult to calculate the ranking of MLB teams. Therefore, this study calculated the ranking of MLB teams by applying the PageRank algorithm and suggested a measure to evaluate the performance level by team, league, and district. Also, we tried to evaluate the predictive validity of the PageRank algorithm based on the post-season match results. The results are as follows.

Ranking of MLB Teams Using PageRank. (Table 4) shows the results of calculating MLB team ranking in the 2017 and 2018 seasons using the PageRank algorithm. As a result, CLE (Cleveland Indians) (PR=0.04) was ranked first, and HOU (Houston Astros) (PR=0.0394) was ranked second in 2017. If looking at the PageRank value of the first ranking of each national league, the eastern district was ranked 8th with 0.0357 (WSH (Washington Nationals), the central district was ranked 9th with 0.0351 (CHC (Chicago Cubs)), and the western district was ranked 3rd with 0.0393 (LAD (Los Angeles Dodgers). In the American League, the eastern district ranked 5th with 0.0378 (BOS (Boston Red Sox)), the central district ranked 1st with 0.0400(CLE), and the western district ranked 2nd with 0.0394 (HOU).

In the 2018 season, BOS (PR=0.0408) ranked 1st, and HOU (PR=0.0407) ranked second. If looking at the PageRank value of the first ranking of each national league, the eastern district ranked 12th with 0.0361 (ATL (Atlanta Braves)), the central district ranked 5th with 0.0384 (MIL (Milwaukee Brewers), and the western district ranked 4th with 0.0386(LAD). In the American League, the eastern district ranked 5th with 0.0378(BOS), the central district ranked 1st with 0.0400(CLE), and the western district ranked 2nd with 0.0407 (HOU).

The following (Table 5) indicates teams that showed a difference of more than 5 in PageRank ranking and winning ratio ranking. In the 2017 season, MIA (Miami Marlins) and NYM (New York Mets) in the eastern district of the national league showed a difference of 5 between PageRank ranking and winning ratio ranking, and in the 2018 season, CLE and MIN (Minnesota Twins) in the central district of American League were 9 and 6, respectively. In particular, the reason for this result is that CLE is the top team in the central district of the American League in the 2018 season is because of the performance level of the teams who play the matches relatively more in the central strict of the American League. PageRank ranking of the teams in the central district of the American League is MIN (2nd in American League Central) ranked 25th, CWS (Chicago White Sox) (4th in American League Central) ranked 27th, DET (Detroit Tigers) (3rd in American League Central) ranked 28th and K.C. (Kansas City Royals) (5th in American League Central) ranked 30th. These teams have relatively lower performance, and it reflects the ranking calculation in consideration of the opposing team's performance, which is a characteristic of the PageRank algorithm. In other words, CLE won the American League Central with a winning ratio of .562; however, the best teams in the same league and the same district had low performance, and the ranking of the CLE team, who played many matches with that team was calculated low.

Comparison of Performance by League and District Using PageRank. The following (Table 6) compares the performance of each league and district using PageRank, and it uses an average PageRank value of each league and district. In the 2017 season, the Eastern League of the American League (PR=0.0349) was the highest, and the Eastern League of the National League (PR=0.0300) was the lowest. By league comparison, the American League (PR=0.0342) was higher than the National League (PR=0.0324). In the 2018 season, the Western League (PR=0.0358) was the highest in the American League, while the Central League (PR=0.0271) was

the lowest. By league comparison, the National League (PR=0.0343) was higher than the American League (PR=0.0324).

The PageRank algorithm is one in which the sum of node sizes (PageRank values) is one and has a proportional measure allocated by relative size from 1. Therefore, the PageRank value can calculate the relative performance between two teams. For example, if Team A's PageRank is 0.5 and Team B's PageRank is 0.3, Team A has about 1.67 times (0.5/0.3) better performance than Team B. From this point of view, the result of MLB (Table 4) shows that in the 2017 season, the Eastern District of the American League has about 1.16 times (0.0300/0.0349) better performance than the Eastern district of the National League, and the American League has about 1.06 times (0.0342/0.0324) better performance than the National League. In the 2018 season, the western district of the American League has about 1.32 times (0.0358/0.0271) better performance than the central district of the American League, and the National League has about 1.06 times (0.0343/0.0324) better performance than the American League.

Predictive Validity of Post Season Applying PageRank. The post-season match result was used to confirm the validity of the MLB ranking model by applying the PageRank algorithm. Specifically, it confirmed how well the PageRank rankings calculated as a match result of the regular season distinguish against the match result of the post-season. In other words, the PageRank ranking calculated how likely the high team won in the post-season.

Table 4. Ranking of MLB Tteams in 2017 Season and 2018 Season Using the PageRank Algorithm (Top 20)

Ranking	2017				2018			
	Team Name	PR	Winning Ratio	League Ranking	Team Name	PR	Winning Ratio	League Ranking
1	CLE	0.0400	0.630	AL_M_1st	BOS	0.0408	0.667	AL_E_1st
2	HOU	0.0394	0.623	AL_W_1st	HOU	0.0407	0.636	AL_W_1st
3	LAD	0.0393	0.642	NL_W_1st	NYY	0.0401	0.617	AL_E_2nd
4	ARI	0.0381	0.574	NL_W_2nd	LAD	0.0386	0.564	NL_W_1st
5	BOS	0.0378	0.574	AL_E_1st	MIL	0.0384	0.589	NL_M_1st
6	NYY	0.0372	0.562	AL_E_2nd	COL	0.0382	0.558	NL_W_2nd
7	COL	0.0364	0.537	NL_W_3rd	CHC	0.0381	0.583	NL_M_2nd
8	WSH	0.0357	0.599	NL_E_1st	OAK	0.0380	0.599	AL_W_2nd
9	CHC	0.0351	0.568	NL_M_1st	SEA	0.0376	0.549	AL_W_3rd
10	MIN	0.0350	0.525	AL_M_2nd	TB	0.0369	0.556	AL_E_3rd
11	MIL	0.0346	0.531	NL_M_2nd	STL	0.0366	0.543	NL_M_3rd
12	TB	0.0344	0.494	AL_E_3rd	ATL	0.0361	0.556	NL_E_1st
13	KC	0.0340	0.494	AL_M_3rd	ARI	0.0352	0.506	NL_W_3rd
14	LAA	0.0337	0.494	AL_W_2nd	PIT	0.0339	0.509	NL_M_4위
15	TOR	0.0330	0.469	AL_E_4th	PHI	0.0334	0.494	NL_E_3rd
16	STL	0.0328	0.512	NL_M_3rd	WSH	0.0329	0.506	NL_E_2nd
17	OAK	0.0327	0.463	AL_W_5th	CLE	0.0328	0.562	AL_M_1st
18	TEX	0.0327	0.481	AL_W_4th	LAA	0.0327	0.494	AL_W_4th
19	SEA	0.0325	0.481	AL_W_3rd	NYM	0.0323	0.475	NL_E_4th
20	BAL	0.0320	0.463	AL_E_5th	SF	0.0323	0.451	NL_W_4th

CLE: Cleveland Indians, **HOU:** Houston Astros, **LAD:** Los Angeles Dodgers, **ARI:** Arizona Diamondbacks, **BOS:** Boston Red Sox, **NYY:** New York Yankees, **COL:** Colorado Rockies, **WSH:** Washington Nationals, **CHC:** Chicago Cubs, **MIN:** Minnesota Twins, **MIL:** Minnesota Twins, **T.B.:** Tampa Bay Rays, **K.C.:** Kansas City Royals, **LAA:** Los Angeles Angels, **TOR:** Toronto Blue Jays, **STL:** St. Louis Cardinals, **OAK:** Oakland Athletics, **TEX:** Texas Rangers, **SEA:** Seattle Mariners, **BAL:** Baltimore Orioles, **ATL:** Atlanta Braves, **PIT:** Pittsburgh Pirates, **PHI:** Philadelphia Phillies, **WSH:** Washington Nationals, **NYM:** New York Mets, **S.F.:** San Francisco Giants, **NL:** National League, **AL:** American League, **E:** East, **W:** West, **M:** Middle

Table 5. Teams with the Difference of More Than 5 between PageRank Ranking and Winning Ratio Ranking

Year	League	District	PR (Ranking)	Winning Ratio (Ranking)	Ranking Difference	Remarks
2017						
MIA	National	Eastern	0.0298 (23th)	0.475 (18th)	5	NL_E_2nd
NYM	National	Eastern	0.027 (30th)	0.432 (25th)	5	NL_E_4th
2018						
CLE	American	Central	0.0328 (17th)	0.562 (8th)	9	AL_M_1st
MIN	American	Central	0.0292 (25th)	0.481 (19th)	6	AL_M_2nd

Table 6. Descriptive Statistics of PageRank Values and Winning Ratio by League and District

League	2017 Season				2018 Season			
	PageRank		Winning Ratio		Page Rank		Winning Ratio	
	M	SD	M	SD	M	SD	M	SD
NL								
Eastern	0.0300	0.0034	0.4714	0.0754	0.0326	0.0028	0.4844	0.0602
Central	0.0325	0.0027	0.4988	0.0581	0.0355	0.0033	0.5276	0.0713
Western	0.0348	0.0044	0.5172	0.1005	0.0348	0.0038	0.4972	0.0680
Total	0.0324	0.0039	0.4958	0.0765	0.0343	0.0033	0.5031	0.0645
AL								
Eastern	0.0349	0.0025	0.5124	0.0522	0.0342	0.0075	0.5162	0.1499
Central	0.0337	0.0043	0.4916	0.0944	0.0271	0.0039	0.4358	0.0844
Western	0.0342	0.0029	0.5084	0.0650	0.0358	0.0044	0.5384	0.0877
Total	0.0342	0.0031	0.5041	0.0680	0.0324	0.0064	0.4968	0.1128

NL: National League, **AL:** American League

The following (Figure 3) shows the match result and predictive validity of the 2017 post-season. In the 2017 season, the team with the highest PageRank rankings won 7 times in 9 situations (2 times of wide cards, four times of division series, two times of league championships, and one time of world series), corresponding to 77.8% predictive validity. On the other hand, if the ranking of the winning ratio calculates the predictive validity, it is about 66.7%. The following (Figure 4) shows the match result and predictive validity of the 2018 post-season. In (Figure 4), the team with the highest PageRank ranking won 7 times in 9 situations, which corresponds to 100.0% of predictive validity.

On the other hand, if the ranking of the winning ratio calculates the predicted validity, it is about 77.8%. Therefore, the PageRank ranking calculated during the regular season has higher predictive validity than the winning ratio calculated by the winning ratio. It is thought that objective comparison is possible if the PageRank algorithm is used to compare teams' performance in MLB.

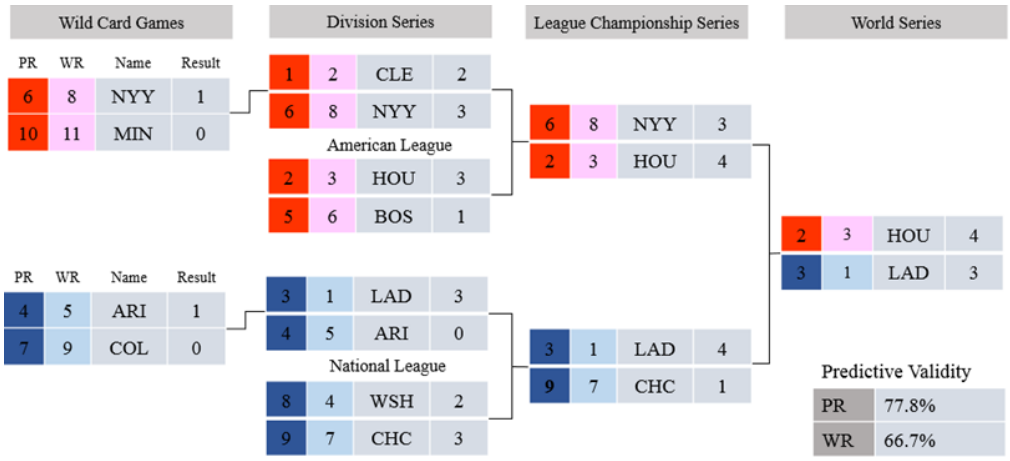


Figure 3. Match Result and Predictive Validity of the 2017 MLB Postseason.

***NYY:** New York Yankees, **MIN:** Minnesota Twins, **ARI:** Arizona Diamondbacks, **COL:** Colorado Rockies, **CLE:** Cleveland Indians, **HOU:** Houston Astros, **BOS:** Boston Red Sox, **LAD:** Los Angeles Dodgers, **WSH:** Washington Nationals, **CHC:** Chicago Cubs

Figure 4. Match Result and Predictive Validity of the 2018 MLB Postseason.
*** NYY:** New York Yankees**, OAK:** Oakland Athletics**, CHC:** Chicago Cubs**, COL:** Colorado Rockies**, BOS:** Boston Red Sox**, HOU:** Houston Astros**, CLE:** Cleveland Indians**, MIL:** Minnesota Twins**, COL:** Colorado Rockies**, LAD:** Los Angeles Dodgers**, ATL:** Atlanta Braves

DISCUSSION

In the MLB, calculating ranking is important because it is used to evaluate team performance and each athlete's performance (4). Especially, MLB is giving a chance to qualify for the Division series to the top-ranked team per each district, so it is a matter of interest to the athletes and team officials. However, the current calculating ranking method of the major league is that the number of matches per team is the same, but there is a limitation that they do not consider the opponent's performance level. Therefore, this study calculated major league ranking by applying the PageRank algorithm to make up for this limitation. Discussion according to result is as follows. First, looking into MLB team's ranking applying PageRank, there is a difference between the old ranking method applied in the league and PageRank ranking.

In particular, the reason for this result, despite that CLE is the top team in the central district of the American League in the 2018 season, is the performance level of the teams who play the matches relatively more in the central strict American League. PageRank ranking of the teams in the central district of the American League is MIN (2nd in American League Central) ranked 25th, CWS (Chicago White Sox) (4th in American League Central) ranked 27th, DET (Detroit Tigers) (3rd in American League Central) ranked 28th and K.C. (Kansas City Royals) (5th in American League Central) ranked 30th. These teams have relatively lower performance, and it reflects the ranking calculation in consideration of the opposing team's performance, which is a characteristic of the PageRank algorithm. In other words, CLE won the American League Central with a winning ratio of .562; however, the best teams in the same league and the same district had low performance, and the ranking of the CLE team, who played many matches with that team was calculated low. This is one of the advantages of PageRank, which is a method to consider the opponent's performance (1, 2, 10). This advantage is thought to evaluate more reasonably when we do a comparative evaluation of performance between regions or teams in MLB

Furthermore, prediction validity has a high index when PageRank is applied in the MLB league. This reports that it has high validity in the method considering the performance of the opponent's team and athlete in other sports such as Taekwondo, badminton, gymnastics, football, and baseball (1, 2, 11, 12). Therefore, if the PageRank algorithm is applied in many spots and MLB, we judge that it is available to evaluate athletes and teams more fairly and reasonably.

CONCLUSION

The conclusion is as follows. First, as a result of calculating PageRank value using win and loss in the major league of the 2017 season, CLE ranked 1st, HOU 2nd, LAD 3rd, ARI 4th, BOS 5th, and NYY sixth. Compared to the team at the

top of each league in the 2017 season, which means the team allowed to play in the Division Series, two teams may have a variable chance to play automatically. Second, as a result of calculating PageRank value using win and loss in the major league of the 2018 season, BOS ranked 1st, HOU 2nd, NYY 3rd, LAD 4th, MIL 5th, and COL 6th. Compared to the team allowed to play in the Division Series of the 2018 season, two teams may have a variable chance to play automatically. Third, as a result of comparing performance per each league and district using PageRank, the eastern district of the American League had the highest performance, and the eastern district of the National League had the lowest performance in the 2017 season. In the 2018 season, the western district of the American League had the highest performance, and the central district of the American League had the lowest. Fourth, comparing the predictive validity of PageRank ranking and winning ratio ranking, the ranking calculated by the PageRank algorithm has relatively higher predictive validity than the winning ratio ranking.

In this study, a new ranking method is presented, along with the problem of calculating major league ranking. However, indeed, the predictive validity presented to confirm the validity of the PageRank algorithm is insufficient to verify the validity of the model. Those various methods are suggested to evaluate the validity of ranking calculation; however, the only evidence of validity is presented in this study. Accordingly, various pieces of evidence to verify the validity of PageRank need to be supported in the future.

APPLICABLE REMARKS

- PageRank algorithm can evaluate team and player's performance more reasonably. MLB's data is used and applied in this study, but it is applicable in sports such as football, basketball, tennis, and others.
- To apply it to many kinds of sports, it must confirm the PageRank algorithm's validity considering each sport's features.

REFERENCES

1. Yoon JW, Park JH. Historical ranking of vault players in artistic gymnastics using PageRank algorithm. *Korean J Sport Sci*. 2017:506-516. **doi:** 10.24985/kjss.2017.28.2.506
2. Kim EH, Jeon MS. Proposal for implementation of a ranking model for Olympic Taekwondo competitions using PageRank. *Int J Performance Anal Sport*. 2019:227-235. **doi:** 10.1080/24748668.2019.1586506
3. Motegi S, Masuda N. A network-based dynamical ranking system for competitive sports. *Sci Rep*. 2012;**2**:904. **doi:** 10.1038/srep00904 **pmid:** 23226590
4. Soren PS. An overview of some methods for ranking sports. Knoxville: University of Tennessee.; 1999.
5. Park J, Newman ME. A network-based ranking system for U.S. college football. Journal of Statistical Mechanics. *Theory Experiment*. 2005:10014. **doi:** 10.1088/1742-5468/2005/10/P10014
6. Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: Bringing order to the web1999.
7. Shi J, Tian X.Y. Learning to Rank Sports Teams on a Graph. *Appl Sci*. 2020:5833. **doi:** 10.3390/app10175833
8. Beggs CB, Shepherd SJ, Emmonds S, Jones B. A novel application of PageRank and user preference algorithms for assessing the relative performance of track athletes in competition. *PLoS One*. 2017;**12**(6):e0178458. **doi:** 10.1371/journal.pone.0178458 **pmid:** 28575009
9. Brown S. A PageRank model for player performance assessment in basketball, soccer and hockey. arXiv preprint arXiv2017. 1704.00583 p.
10. Kim BS. Development of Taekwondo ranking model based on Google PageRank algorithm. *Int J Pure Appl Mathemat*. 2018:1267-1278.
11. Jo EH, Park JH, Choi CH. A PageRank Algorithm for the Rankings of Korea Badminton Players. *J Korean Data Anal Soc*. 2018:373-382. **doi:** 10.37727/jkdas.2018.20.1.373
12. Kim HS, Park JH, Jo EH, Choi CH. The Most Successful Team in AFC Asian Cup History : Country Rankings Using PageRank Algorithms. *Korean Soc Measure Evaluat Physic Educat Sport Sci*. 2019:89-101.
13. Brown S. A PageRank model for player performance assessment in basketball, soccer and hockey. *ArXiv preprint arXiv*. 2017:1704.00583.
14. Langville AN, Meyer CD. Deeper inside pagerank. *Internet Mathemat*. 2004:335-380. **doi:** 10.1080/15427951.2004.10129091

15. Li P, Chien E, Milenkovic O. Optimizing generalized pagerank methods for seed-expansion community detection. *ArXiv preprint arXiv*. 2019:1905-10881.
16. Kloumann IM, Ugander J, Kleinberg J. Block models and personalized PageRank. *Proc Natl Acad Sci U S A*. 2017;**114**(1):33-38. **doi:** 10.1073/pnas.1611275114 **pmid:** 27999183
17. Kohlschütter C, Chirita PA, Nejdl W. Efficient parallel computation of pagerank. *Europe Conference Inform Retrieval*. 2006:241-252. **doi:** 10.1007/11735106_22
18. Engström C, Silvestrov S. A componentwise pagerank algorithm. In 16th Applied Stochastic Models and Data Analysis International Conference (ASMDA2015) with Demographics 2015 Workshop, 30 June-4 July 2015. Greece: University of Piraeus; 2015.
19. Engström C. PageRank in evolving networks and applications of graphs in natural language processing and biology. Doctoral dissertation: Mälardalen University; 2016.
20. Ishii H, Tempo R, Bai EW. A web aggregation approach for distributed randomized PageRank algorithms. *IEEE Transaction Auto Control*. 2012:2703-2717. **doi:** 10.1109/TAC.2012.2190161
21. Wang Y, Kawai Y, Miyamoto S, Sumiya K. A students' mutual evaluation method for their reports using pagerank algorithm. In Proc. of the 22nd International Conference on Computers in Education2014.
22. Hochbaum DS. Ranking sports teams and the inverse equal paths problem. *Int Workshop Internet Network Economic*. 2006:307-318. **doi:** 10.1007/11944874_28
23. Lazova V, Basnarkov L. PageRank approach to ranking national football teams. *ArXiv preprint arXiv*. 2015:1503.01331.
24. Rojas-Mora J, Chávez-Bustamante F, del Río-Andrade J, Medina-Valdebenito N. A methodology for the analysis of soccer matches based on pagerank centrality. *Sport Manage Emerging Economic Activity*. 2017:257-272. **doi:** 10.1007/978-3-319-63907-9_16
25. Govan AY, Meyer CD, Albright R. Generalizing Google's PageRank to rank national football league teams. In Proceedings of the SAS Global Forum2008.
26. Júnior PSP, Gonçalves MA, Laender AH, Salles T, Figueiredo D. Time-aware ranking in sport social networks. *J Inform Data Manage*. 2012:195.
27. Boldi P, Santini M, Vigna S. PageRank as a function of the damping factor. In Proceedings of the 14th international conference on World Wide Web2005.
28. Hu ZH, Zhou JX, Zhang MJ, Zhao Y. Methods for ranking college sports coaches based on data envelopment analysis and PageRank. *Expert Sys*. 2015:652-673. **doi:** 10.1111/exsy.12108