KBO League Scheduling Using Optimization

Kim Taeyeoun
DGIST 201511058
Kim77ty @ dgist.ac.kr

***Abstract*—This project is aiming to optimizing the baseball league scheduling to minimize the total moving distance.**

***Keywords*—Optimization, scheduling**

I. Introduction

In KBO, Korea professional baseball league, there are 10 teams, and each team play 144 games in a year, and they have 53 times of moving day in a season[1]. Also, as the teams are spread all over the country, the distance to move is very important for each team. In this project, the purpose is to minimize the total distance of moving for all teams.

II. Problem Statement

A. Variables, constants, constraints definition

In real, one baseball team have 144 games in a year. However, the matches are always in 3 or 2 series of matches. So, it can be thought that one team has 54 big games in a year. Also, there are 10 teams in KBO league, and each of team is denoted as a number 1 to 10 (1: Doosan, 2: LG, 3: SK, 4: Nexen, 5: Hanhwa, 6: KIA, 7: Samsung, 8: Lotte, 9: NC, 10: KT). We define the binary variable x_{ijk} $(i = 1, 2, \ldots., 10 \;\; j = 1, 2, \ldots., 54 \;\; k = 1, 2, \ldots., 10)$ means team i have a match in jth match, k's home stadium, binary variable y_{ijkl} $(i = 1, 2, \ldots., 10 \;\; \mathrm{j} = 1, 2, \ldots., 54 \;\; \mathrm{k} = 1, 2, \ldots., 10 \;\; \mathrm{l} = 1, 2, \ldots, 10)$ means team i moves from stadium k to l after jth match, and constant d_{kl} $(k = 1, 2, \ldots., 10 \;\; l = 1, 2, \ldots., 10)$ means the distance between stadium k and l.

For season's games there are some constraints as followed.

1. Each team has 27 home games and 3 away games with each team assuming that each team has 54 big games.
2. In a specific day a team have a match only in one stadium.
3. There are 5 home matches in a day, and in a specific day and stadium, only less than 2 teams have game (0 or 2)
4. The first match-up is given.
5. As Doosan and LG have a same home stadium, they can't have home games in same day.

B. Mathematical Represntation

For first constraint, there are 27 home games, and 3 away games for each can be represented like below.

$$\sum_{j} x_{ijk} = 27 \quad , \forall i, k \quad (if\ i = k) \qquad (1)$$

$$\sum_{j} x_{ijk} = 3 \quad , \forall i, k \quad (if\ i \neq k) \qquad (2)$$

For second constraint, it can be represented like below.

$$\sum_{k} x_{ijk} = 1 \ , \forall i, j \qquad (3)$$

For third constraint, there are 5 home matches in a day can be represented like below.

$$\sum_{i} x_{iji} = 5 \quad , \forall \mathrm{j} \qquad (4)$$

$$\sum_{i} x_{ijk} \le 2 \ , \forall\, j, k \qquad (5)$$

For fourth constraint, x_{i1k} is given value.

For last constraint, it can be represented like below.

$$\sum_{i} (x_{ij1} + x_{ij2}) \le 2 \qquad (6)$$

To minimize the total distance, we define the variable y_{ijkl} with variable x_{ijk}.

$$(x_{ijk} + x_{i(j+1)l} - 1) \le y_{ijkl} \qquad (7)$$

With these constraints, if team i have a match in k in j day, and in l in j+1 day, y_{ijkl} becomes better than 1, otherwise becomes better than 0. As a result, as the objective function would be the sum of constant multiplication of the y_{ijkl}, and y_{ijkl} is the binary variable, it works as we defined.

C. Optimization Problem Definition

We defined the constraints, and for objective function, as y_{ijkl} = 1 means that in the day j to (j+1), team i moves to stadium k to l. So, we can model the optimization problem like this.

$$\min \ 2\sum_i \sum_j \sum_k \sum_l y_{ijkl} d_{kl}$$

$$\text{s.t.} \quad \sum_j x_{ijk} = 27 \ , \forall i,k \quad (if\ i = k) \qquad (1)$$

$$\sum_j x_{ijk} = 3 \ , \forall i,k \quad (if\ i \neq k) \qquad (2)$$

$$\sum_k x_{ijk} = 1 \ , \forall i,j \qquad (3)$$

$$\sum_i x_{iji} = 5 \ , \forall j \qquad (4)$$

$$\sum_i x_{ijk} \ \leq \ 2 \ , \forall j,k \qquad (5)$$

$$\sum_i (x_{ij1} + x_{ij2}) \leq 2 \qquad (6)$$

$$(x_{ijk} + x_{i(j+1)l} - 1) \leq y_{ijkl} \qquad (7)$$

$$x_{i1k} \text{ is given}$$

III. SIMULATION

A. Simplified scenario

Distance	1	2	3	4
	SK	Doosan	Lotte	KIA
SK	0.00	39.99	391.54	296.47
Doosan	39.99	0.00	367.28	289.47
Lotte	391.54	367.28	0.00	247.00
KIA	296.47	289.47	247.00	0.00

Table 1. distance for simplified scenario

1st Match-up	
SK-Lotte	Doosan-KIA
SK	Doosan

Table 2. first match-up for simplified scenario

First, we did simulation with a simplified scenario to check if the problem and its codes work well. We used Python library "pulp" to solve this problem [2]. In the simplified scenario, we assumed that there are 4 teams, and each team has 3 home, 3 away games. With this scenario, we can check that constraint 1 should change to equal 3, constraint 2 should change to equal 1, constraint 4 should change to equal 2, we don't need constraint 6, and the i, j, k, l should be to 4, 6, 4, 4. Table 1 and 2 shows the basic set up for this scenario.

To solve this problem easier, we mapped the variables. In the problem statement, the variable x is 3-dimensional, and y s 4-dimensional matrix. To calculate and write Python code easier, we mapped these variables to 1-dimensional variables as followed.

$$x_{ijk} = x[4(i-1) + k + 16(j-1)]$$

$$y_{ijkl} = y[4(i-1) + k + 16(l-1) + 64(j-1)]$$

With these mapping, we can think this problem as a big linear system problem.

day 1

	SK	Doosan	Lotte	KIA
SK	1	0	0	0
Doosan	0	1	0	0
Lotte	1	0	0	0
KIA	0	1	0	0

day 4

	SK	Doosan	Lotte	KIA
SK	1	0	0	0
Doosan	0	0	1	0
Lotte	0	1	0	0
KIA	0	0	0	1

day 2

	SK	Doosan	Lotte	KIA
SK	0	1	0	0
Doosan	0	1	0	0
Lotte	0	0	1	0
KIA	0	0	1	0

day5

	SK	Doosan	Lotte	KIA
SK	0	0	0	1
Doosan	1	0	0	0
Lotte	0	0	1	0
KIA	0	0	0	1

day 3

	SK	Doosan	Lotte	KIA
SK	1	0	0	0
Doosan	0	1	0	0
Lotte	0	0	0	1
KIA	1	0	0	0

day6

	SK	Doosan	Lotte	KIA
SK	0	0	1	0
Doosan	0	0	0	1
Lotte	0	0	1	0
KIA	0	0	0	1

Table 3. first result of simulation

Table 3 Shows the result of simulation, and the highlighted parts mean there were something wrong. The number '1' means that in that stadium, that team has a game, but, in the highlighted parts, there is only one team in a stadium, which is impossible as with only one team, the game cannot be played.

The reason for this wrong result was on the constraint 5. Constraint 5 means that in a stadium, less than 2 teams can play, so there were cases which only one team in a stadium. Thus, we revised the constraint 5 as followed

$$x_{iji} - \sum_k x_{kji} = 0 \ \ (k \neq i) \qquad (5)$$

If x_{iji} is 1, which means team i has a home match, there should be one more team which has a match in i. in other case, there should be no team in i.

	INCHEON	SEOUL	BUSAN	GWANGJU
day 1	SK - Lotte	Doosan - KIA		
day 2	SK - KIA	Doosan - Lotte		
day 3	SK - Doosan			KIA - Lotte
day 4		Doosan - SK	Lotte - KIA	
day 5			Lotte - SK	KIA - Doosan
day 6			Lotte - Doosan	KIA - SK

Table 4. Second result of simulation

	SK	Doosan	Lotte	KIA	
move1	0	0	39.99	39.99	
move2	0	39.99	289.47	296.47	
move3	39.99	39.99	247	247	
move4	367.28	289.47	0	247	
move5	247	247	0	0	
SUM	654.27	616.45	576.46	830.46	2677.64

Table 5. total moving distance

With this revised constraint, and same problem set up, table 4 is the optimal result's schedule. Based on this schedule, with re calculated the total moving distance and table 5, and it was same with the result of Python code's optimal value.

Considering that it is the linear programming problem, the fact that the problem has its feasible set and optimal value means that the problem works well, and as LP uses the simplex algorithm, it would be the real optimal value. Also, we checked that there is no calculation error in codes. Thus, we can say that the problem and its codes are works well.

B. *full scenario*

Distance		Doosan	LG	SK	Nexen	Hanhwa	KIA	Samsung	Lotte	NC	KT
		1	2	3	4	5	6	7	8	9	10
Doosan	1	0	0	39.99	21.48	155.93	289.47	271.89	367.28	333.26	31.26
LG	2	0	0	39.99	21.48	155.93	289.47	271.89	367.28	333.26	31.26
SK	3	39.99	39.99	0	19.78	163.77	296.47	297.81	391.54	356.81	39.00
Nexen	4	21.48	21.48	19.78	0	158.09	292.52	287.27	381.00	346.39	28.15
Hanhwa	5	155.93	155.93	163.77	158.09	0	164.76	145.39	239.12	199.41	130.66
KIA	6	289.47	289.47	296.47	292.52	164.76	0	205.20	247.00	198.56	264.42
Samsung	7	271.89	271.89	297.81	287.27	145.39	205.20	0	95.43	87.83	259.40
Lotte	8	367.28	367.28	391.54	381.00	239.12	247.00	95.43	0	50.91	354.80
NC	9	333.26	333.26	356.81	346.39	199.41	198.56	87.83	50.91	0	319.45
KT	10	31.26	31.26	39.00	28.15	130.66	264.22	259.40	354.8	319.45	0

Table 6. distance for full scenario

1st match-up				
Doosan-Samsung	Nexen-Hanhwa	SK-Lotte	KT-KIA	LG-NC
Doosan	Nexen	SK	KT	NC

Table 7. first match-up for full scenario

We checked that the problem and its codes work well. For next step, we did the simulation with the full scenario, which was our goal. There are 10 teams, and each team has 27 home and 27 away games in a season. The basic set up is on table 6 and 7. In table 7 we made the first match-up based on the 2018 season's schedule. Also, we mapped the variables as followed.

$$x_{ijk} = x[10(i-1) + k + 100(j-1)]$$

$$y_{ijkl} = y[10(i-1) + k + 100(l-1) + 1000(j-1)]$$

We didn't get the result of this simulation as its size was too big, and it takes too long time as the simplex algorithm itself checks all possible points. As an example, for constraint 7, we had to define all y variables, so, only with this one constraint, it needs 54000 number of constraints. Thus, it must take very long time. Although we cannot get the result of full scenario, we checked that it doesn't cause error, as there was no infeasible or other error while we are running this code for 3 days. Therefore, it could solve this optimization problem if there was enough time.

IV. CONCLUSION

With simulation, we revised the problem little bit, and the final optimization problem is as followed.

$$\min \ 2\textstyle\sum_i \sum_j \sum_k \sum_l y_{ijkl} d_{kl}$$

$$\text{s.t.} \quad \sum_j x_{ijk} = 27 \ ,\forall i,k \ \ (if \ i = k) \qquad (1)$$

$$\sum_j x_{ijk} = 3 \ ,\forall i,k \ \ (if \ i \neq k) \qquad (2)$$

$$\sum_k x_{ijk} = 1 \ ,\forall i,j \qquad (3)$$

$$\sum_i x_{iji} = 5 \ ,\forall \mathrm{j} \qquad (4)$$

$$x_{iji} - \sum_k x_{kji} = 0 \ (k \neq i) \qquad (5)$$

$$\sum_i (x_{ij1} + x_{ij2}) \leq 2 \qquad (6)$$

$$(x_{ijk} + x_{i(j+1)l} - 1) \leq y_{ijkl} \qquad (7)$$

$$x_{i1k} \ is \ given$$

With this project, we formulated mathematical problem based on the scheduling rules, and made codes to solve this problem. The codes are opened in github.

(https://github.com/KimTaeyeoun/Convex-Optimization-Project)

For advance, we thought about some points which would be helpful if someone do more research in this area.

First, in the simplified scenario, in table 5, considering the sum of distance, its standard deviation is 11.97, mean is 669.41 and its maximum distance difference for each team is 161.05. Even in this simplified scenario, we cannot say that the optimal result is fair for all of teams. Thus, it would be better to consider the fairness by adding the constraint of standard deviation.

Second, in the real league, there is a break time. That is, it will be better to consider this break time.

Lastly, in the full scenario, as the problem size was too big, it would be good to simplify the codes or equations to reduce the running time. It can be done by re formulating the problem, simplify the constraint, or make user defined function to solve the problem faster.

REFERENCES

[1] KBO League homepage

https://www.koreabaseball.com/Default.aspx

[2] Python library pulp in github

https://github.com/pulp/pulp