Proceedings of the International MultiConference of Engineers and Computer Scientists 2012 Vol I,
IMECS 2012, March 14 - 16, 2012, Hong Kong

A Reinforcement Learning for Train Marshaling Based on the Processing Time Considering Group Layout of Freight Cars

Yoichi Hirashima

Abstract—**In this paper a new method for generating marshaling plan of freight cars in a train is proposed. In the proposed method, marshaling plans based on the processing time can be obtained by a reinforcement learning system. In order to evaluate the processing time, the total transfer distance of a locomotive and the total movement counts of freight cars are simultaneously considered. Moreover, by grouping freight cars that have the same destination, candidates of the desired arrangement of the outbound train is extended. This feature is considered in the learning algorithm, so that the total processing time is reduced. Then, the order of movements of freight cars, the position for each removed car, the layout of groups in a train, the arrangement of cars in a group and the number of cars to be moved are simultaneously optimized to achieve minimization of the total processing time for obtaining the desired arrangement of freight cars for an outbound train. Initially, freight cars are located in a freight yard by the random layout, and they are moved and lined into a main track in a certain desired order in order to assemble an out bound train. Learning algorithm in the proposed method is based on the Q-Learning, and, after adequate autonomous learning, the optimum marshaling plan can be obtained by selecting a series of movements of freight cars that has the best evaluation.**

Index Terms—**Scheduling, Container Transfer Problem, Q-Learning, Freight train, Marshaling**

I. INTRODUCTION

IN recent years, logistics with freight train has important role in ecological aspects, because railway logistics is known to have smaller environmental load as compared to goods transportation with trucks ([1]). Train marshaling operation at freight yard is required to joint several rail transports, or different modes of transportation including rail. Transporting goods are carried in containers, each of which is loaded on a freight car. A freight train is consists of several freight cars, and each car has its own destination. Thus, the train driven by a locomotive travels several destinations disjointing corresponding freight cars at each freight station. In addition, since freight trains can transport goods only between railway stations, modal shifts are required for area that has no railway. In intermodal transports including rail, containers carried into the station are loaded on freight cars and located at the freight yard in the arriving order. The initial layout of freight cars is thus random. For efficient shift in assembling outbound train, freight cars must be rearranged before jointing to the freight train. In general, the rearrangement process is conducted in a freight yard that consists of a main-track and several sub-tracks. Freight cars are initially placed on sub tracks, rearranged, and lined into the main track. This series of operation is called marshaling, and several methods to solve the marshaling problem have been proposed [2], [3]. Also, many similar problems are treated by mathematical programming and genetic algorithm[4], [5], [6], [7], and some analyses are conducted for computational complexities [7], [8]. However, these methods do not consider the processing time for each transfer movement of freight car that is moved by a locomotive.

Faculty of Information Science and Technology, Osaka Institute of Technology, 1-79-1, Kita-yama, Hirakata City, Osaka, 573-0196, Japan. Tel/Fax: +86-72-866-5187 Email: hirash-y@is.oit.ac.jp

In this paper a new method for generating marshaling plan of freight cars in a train is proposed. In the proposed method, marshaling plans based on the processing time can be obtained by a reinforcement learning system. A movement of a freight car consists of 4 elements: 1. moving a locomotive to the car to be transferred, 2. jointing cars with the locomotive, 3. transferring cars to their new position by the locomotive, and 4. disjointing the cars from the locomotive. The processing times for elements 1. and 3. are determined by by the transfer distance of the locomotive, the weight of the train, and the performance of the locomotive. The total processing time for elements 1. and 3. is determined by the number of movements of freight cars. Thus, the transfer distance of the locomotive and the number of movements of freight cars are simultaneously considered, and used to evaluate and minimize the processing time of marshaling for obtaining the desired layout of freight cars for an outbound train. The total processing time of marshaling is considered by using a weighted cost of a transfer distance of the locomotive and the number of movements of freight cars. Then, the order of movements of freight cars, the position for each removed car, the arrangement of cars in a train and the number of cars to be moved are simultaneously optimized to achieve minimization of the total processing time. The *original* desired arrangement of freight cars in the main track is derived based on the destination of freight cars. In the proposed method, by grouping freight cars that have the same destination, several desirable positions for each freight car in a group are generated from the original one, and the optimal group-layout that can achieve the smallest processing time of marshaling is obtained by autonomous learning. Simultaneously, the optimal sequence of car-movements as well as the number of freight cars that can achieve the desired layout is obtained by autonomous learning. Also, the feature is considered in the learning algorithm, so that, at each arrangement on sub track, an evaluation value represents the smallest processing time of marshaling to achieve the best layout on the main track. The learning algorithm is derived based on the Q-Learning [9], which is known as one of the well established realization algorithm of the reinforcement learning.

ISBN: 978-988-19251-1-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

In the learning algorithm, the state is defined by using a layout of freight cars, the car to be moved, the number of cars to be moved, and the destination of the removed car. An evaluation value called Q-value is assigned to each state, and the evaluation value is calculated by several update rules based on the Q-Learning algorithm. In the learning process, a Q-value in a certain update rule is referred from another update rule, in accordance with the state transition. Then, the Q-value is discounted according to the transfer distance of the locomotive. Consequently, Q-values at each state represent the total processing time of marshaling to achieve the best layout from the state. Moreover, in the proposed method, only referred Q-values are stored by using table look-up technique, and the table is dynamically constructed by binary tree in order to obtain the best solution with feasible memory space. In order to show effectiveness of the proposed method, computer simulations are conducted for several methods.

II. PROBLEM DESCRIPTION

The yard consist of 1 main track and m sub tracks. Define k as the number of freight cars placed on the sub tracks, and they are carried to the main track by the desirable order based on their destination. In the yard, a locomotive moves freight cars from sub track to sub track or from sub track to main track. The movement of freight cars from sub track to sub track is called removal, and the car-movement from sub track to main track is called rearrangement. For simplicity, the maximum number of freight cars that each sub track can have is assumed to be n, the ith car is recognized by an unique symbol c_i $(i = 1, \cdots, k)$. Fig.1 shows the outline of freight yard in the case $k = 30, m = n = 6$. In the figure,

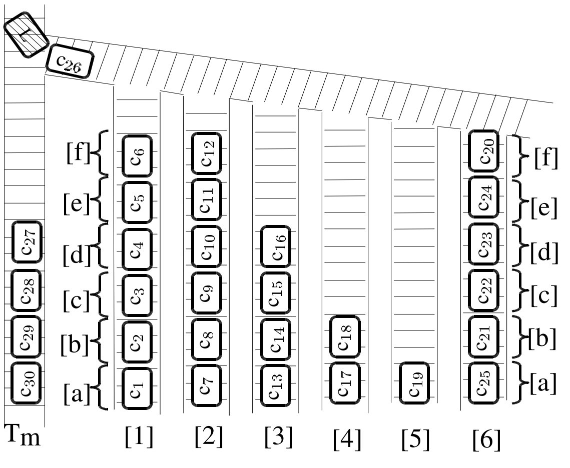

Fig. 1. Freight yard

track Tm denotes the main track, and other tracks [1], [2], [3], [4], [5], [6] are sub tracks. The main track is linked with sub tracks by a joint track, which is used for moving cars between sub tracks, or for moving them from a sub track to the main track. In the figure, freight cars are moved from sub tracks, and lined in the main track by the descending order, that is, rearrangement starts with c_{30} and finishes with c_1. When the locomotive L moves a certain car, other cars locating between the locomotive and the car to be moved must be removed to other sub tracks. This operation is called removal. Then, if $k \leq n \cdot m - (n-1)$ is satisfied for keeping adequate space to conduct removal process, every car can be rearranged to the main track.

In each sub track, positions of cars are defined by n rows. Every position has unique position number represented by $m \cdot n$ integers, and the position number for cars at main track is 0. Fig.2 shows an example of position index for $k = 30, m = n = 6$ and the layout of cars for fig.1.

Position index

	[1]	[2]	[3]	[4]	[5]	[6]
[f]	31	32	33	34	35	36
[e]	25	26	27	28	29	30
[d]	19	20	21	22	23	24
[c]	13	14	15	16	17	18
[b]	7	8	9	10	11	12
[a]	1	2	3	4	5	6

Yard layout

	[1]	[2]	[3]	[4]	[5]	[6]
[f]	c_6	c_{12}				c_{20}
[e]	c_5	c_{11}				c_{24}
[d]	c_4	c_{10}	c_{16}			c_{23}
[c]	c_3	c_9	c_{15}			c_{22}
[b]	c_2	c_8	c_{14}	c_{18}		c_{21}
[a]	c_1	c_7	c_{13}	c_{17}	c_{19}	c_{25}

Fig. 2. Example of position index and yard state

In Fig.2, the position "[a][1]" that is located at row "[a]" in the sub track "[1]" has the position number 1, and the position "[f][6]" has the position number 36. For unified representation of layout of car in sub tracks, cars are placed from the row "[a]" in every track, and newly placed car is jointed with the adjacent freight car. In the figure, in order to rearrange c_{25}, cars $c_{24}, c_{23}, c_{22}, c_{21}$ and c_{20} have to be removed to other sub tracks. Then, since $k \leq n \cdot m - (n-1)$ is satisfied, c_{25} can be moved even when all the other cars are placed in sub tracks.

In the freight yard, define $x_i (1 \leq x_i \leq n \cdot m, i = 1, \cdots, k)$ as the position number of the car c_i, and $\boldsymbol{s} = [x_1, \cdots, x_k]$ as the state vector of the sub tracks. For example, in Fig.2, the state is represented by $\boldsymbol{s} = [1, 7, 13, 19, 25, 31, 2, 8, 14, 20, 26, 32, 3, 9, 15, 21, 4, 10, 5, 36, 12, 18, 24, 30, 6, 0, 0, 0, 0, 0]$. A trial of the rearrange process starts with the initial layout, rearranging freight cars according to the desirable layout in the main track, and finishs when all the cars are rearranged to the main track.

III. DESIRED LAYOUT IN THE MAIN TRACK

In the main track, freight cars that have the same destination are placed at the neighboring positions. In this case, removal operations of these cars are not required at the destination regardless of layouts of these cars. In order to consider this feature in the desired layout in the main track, a group is organized by cars that have the same destination, and these cars can be placed at any positions in the group. Then, for each destination, make a corresponding group, and the order of groups lined in the main track is predetermined by destinations. This feature yields several desirable layouts in the main track.

Fig.3 depicts examples of desirable layouts of cars and the desired layout of groups in the main track. In the figure, freight cars c_1, $\cdots$, c_6 to the destination$_1$ make group$_1$, c_7, $\cdots$, c_{18} to the destination$_2$ make group$_2$, c_{19}, $\cdots$, c_{25} to the destination$_3$ make group$_3$, and c_{26}, $\cdots$, c_{30} to the destination$_4$ make group$_4$. Groups$_{1,2,3,4}$ are lined by ascending order in the main track, which make a desirable layout. In the figure, examples of layout in group$_1$ are in the dashed square.

Also, the layout of groups lined by the reverse order do not yield additional removal actions at the destination of each

ISBN: 978-988-19251-1-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

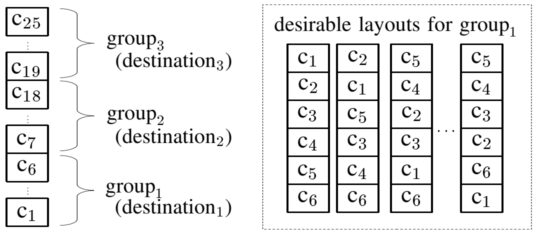

Fig. 3. Example of groups

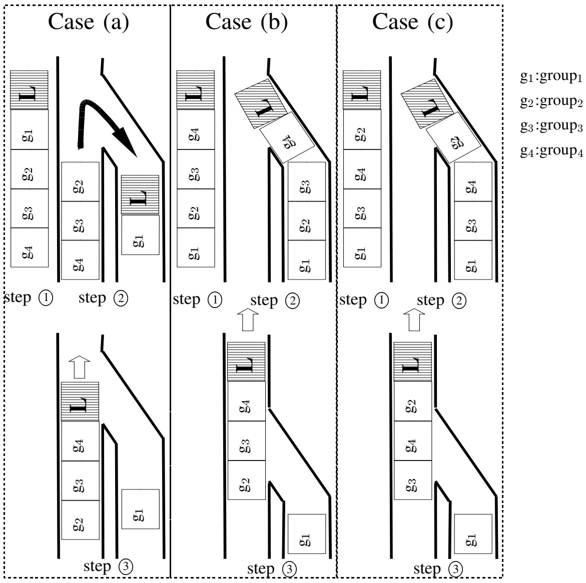

Fig. 4. Group layouts

group. Thus, in the proposed method, the layout lined groups by the reverse order and the layout lined by decoupling order from both ends of the train are regarded as desired layouts. Fig.4 depicts examples of material handling operation for extended layout of groups at the destination of group_1. In the figure, step ① shows the layout of the incoming train. In case (a), cars in group_1 are separated at the main track, and moved to a sub-track by the locamotive L at step ②. In cases (b),(c), cars in group_1 are carried in a sub-track, and group_1 is separated at the sub-track. In the cases, group_1 can be located without any removal actions for cars in each group. Thus, these layouts of groups are regarded as candidate for desired one in the learning process of the proposed method.

IV. Direct rearrangement

When rearranging car that has no car to be removed on it is exist, its rearrangement precede any removals. In the case that several cars can be rearranged without a removal, rearrangements are repeated until all the candidates for rearrangement requires at least one removal. If several candidates for rearrangement require no removal, the order of selection is random, because any orders satisfy the desirable layout of groups in the main track. In this case, the arrangement of cars in sub tracks obtained after rearrangements is unique, so that the movement counts of cars has no correlation with rearrangement orders of cars that require no removal. This operation is called direct rearrangement. When a car in a certain sub track can be rearrange directly to the main track and when several cars located adjacent positions in the same sub track satisfy the layout of group in main track, they are jointed and applied direct rearrangement.

Fig.5 shows an example of arrangement in sub tracks existing candidates for rearranging cars that require no removal. At the top of figure, from the left side, a desired layout of cars and groups, the initial layout of cars in sub tracks, and the position index in sub tracks are depicted for $m = n = 4, k = 9$. c_1, c_2, c_3, c_4 are in group_1, c_5, c_6, c_7, c_8 are in group_2, and group_1 must be rearranged first to the main track. In each group, any layouts of cars can be acceptable. In both cases, c_2 in step 1 and c_3 in step 3 are applied the direct rearrangement. Also, in step 4, 3 cars c_1, c_4, c_5 located ajacent positions are jointed and moved to the main track by a direct rearrangement operation. In addition, at step 5 in case 2, cars in group_2 and group_3 are moved by a direct rearrangement, since the positions of c_7, c_8, c_6, c_9 are satisfied the desired layout of groups in the main track.

In case 1 of the example, the rearrangement order of cars that require no removal is c_1, c_2, c_3, c_4, and in case 2, the order is c_3, c_2, c_1, c_4. Although 2 cases have different orders of rearrangement, the arrangements of cars in sub tracks and the numbers of movements of cars have no difference.

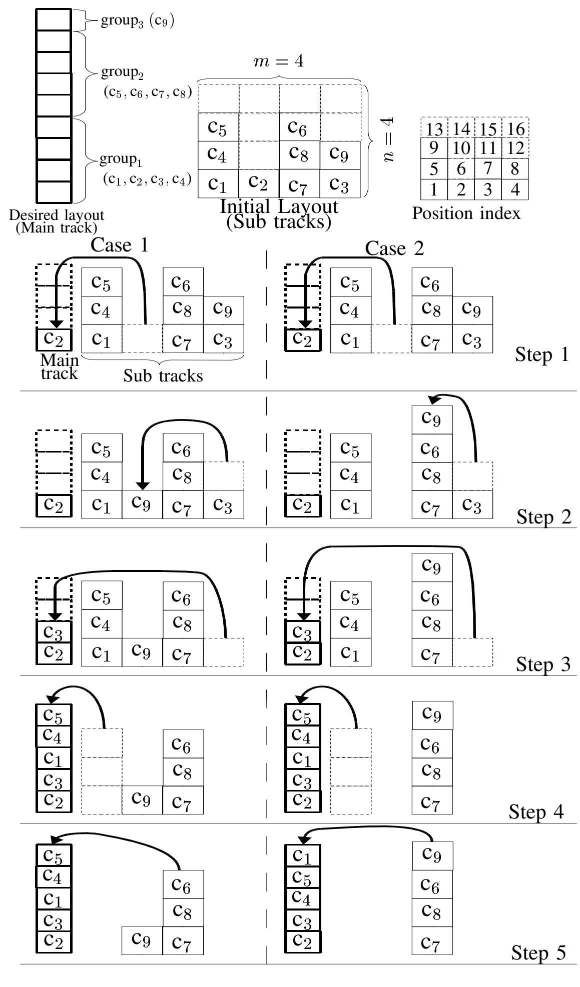

Fig. 5. Direct rearrangements

V. Rearrangement process

The rearrangement process for cars consists of following 6 operations :

ISBN: 978-988-19251-1-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

(I) selection of a layout of groups in the main track, and rearrangement for all the cars that can apply the direct rearrangement into the main track,
(II) selection of a freight car to be rearranged into the main track,
(III) selection of a removal destinations of the cars in front of the car selected in (I),
(IV) selection of the number of cars to be moved,
(V) removal of the cars to the selected sub-track,
(VI) rearrangement of the selected car.

These operations are repeated until one of desirable layouts is achieved in the main track, and a series of operations from the initial state to the desirable layout is define as a trial.

Now, define h as the number of candidates of the desired layout of groups. Each candidate in operation (I) is represented by u_{j_1} $(1 \le j_1 \le h)$.

In the operation (II), each group has the predetermined position in the main track. The car to be rearranged is defined as c_T, and candidates of c_T can be determined by excluding freight cars that have already rearranged to the main track. These candidates must belong to the same group.

Also, define r as the number of groups, g_l as the number of freight cars in group$_l (1 \le l \le r)$, and $u_{j_2}(h+1 \le j_2 \le h+g_l)$ as candidates of c_T.

In the operation (III), the removal destination of car located on the car to be rearranged is defined as c_M. Then, defining $u_{j_3}(h+g_l+1 \le j_3 \le h+g_l+m-1)$ as candidates of c_M, excluding the sub-track that has the car to be removed, and the number of candidates is $m-1$.

In the operation (IV), defining p as the number of removal cars required to rearrange c_T, and defining q as the number of removal cars that can be located the sub-track selected in the operation (III), the candidate numbers of cars to be moved are determined by $u_{j_4}, 2m \le j_4 \le 2m + \min\{p, q\} - 1$.

In both cases of Fig.5, the direct rearrangement is conducted for c_2 at step 1, and the selection of c_T conducted at step 2, candidates are $u_{h+1} = [1], u_{h+2} = [4]$, that is, sub-tracks where cars in group$_1$ are located at the top. u_{h+3}, u_{h+4} are excluded from candidates. Then, $u_{h+2} = [4]$ is selected as c_T. Candidates for the location of c_T are $u_{h+5} = [1], u_{h+6} = [2], u_{h+7} = [3]$, sub-tracks [1],[2], and [3]. In case 1, $u_6 = [2]$ is selected as c_M, and in case 2, $u_{h+7} = [3]$ is selected. After direct rearrangements of c_3 at step 3 and $\mathrm{c}_1, \mathrm{c}_4, \mathrm{c}_5$ at step 4, the marshaling process is finished at step 5 in case 2, whereas case 1 requires one more step in order to finish the process. Therefore, the layout of cars and groups in the main track, the number of cars to be moved, the location the car to be rearranged and the order of rearrangement affect the total movement counts of freight cars.

VI. PROCESSING TIME FOR A MOVEMENT OF LOCOMOTIVE

A. Transfer distance of locomotive

When a locomotive transfer freight cars, the process of the unit transition is as follows: (E1). starts without freight cars, and reaches to the joint track, (E2) restart in reverse direction to the target car to be moved, (E3). joints them, (E4) pull out them to the joint track, (E5) restart in reverse direction, and transfers them to the indicated location, and (E6) disjoints them from the locomotive. Then, the transfer distance of locomotive in (E1), (E2), (E4) and (E5) is defined as D_1, D_2, D_3 and D_4, respectively. Also, define the unit distance of a movement for cars in each sub track as $D_{\min_v}$, the length of joint track between adjacent sub tracks, or, sub track and main track as $D_{\min_h}$. The location of the locomotive at the end of above process is the start location of the next movement process of the selected car. Also, the initial position of the locomotive is located on the joint track nearest to the main track.

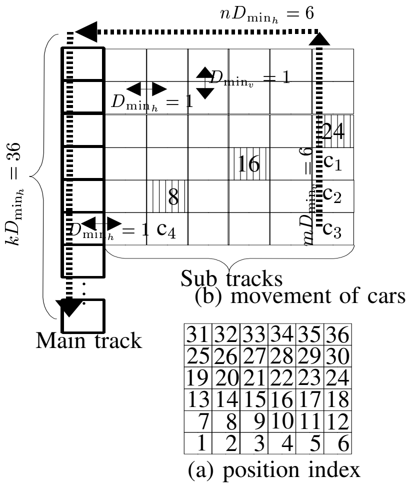

Fig. 6. Calculation of transfer distance

Fig.6 shows an example of transfer distance. In the figure, $m = n = 6$, $D_{\min_v} = D_{\min_h} = 1, k = 18$, (a) is position index, and (b) depicts movements of locomotive and freight car. Also, the locomotive starts from position 8, the target is located on the position 18, the destination of the target is 4, and the number of cars to be moved is 2. Since the locomotive moves without freight cars from 8 to 24, the transfer distance is $D_1 + D_2 = 12$ $(D_1 = 5, D_2 = 7)$, whereas it moves from 24 to 16 with 2 freight cars, and the transfer distance is $D_3 + D_4 = 13$ $(D_3 = 7, D_4 = 6)$.

B. Processing time for the unit transition

In the process of the unit transition, the each time for (E3) and (E6) is assumed to be the constant $\mathrm{t_E}$.

The processing times for elements (E1), (E2), (E4) and (E5) are determined by the transfer distance of the locomotive $D_i (i = 1, 2, 3, 4)$, the weight of the freight cars W moved in the process, and the performance of the locomotive. Then, the time each for (E1), (E2), (E4) and (E5) is assumed to be obtained by the function $f(D_i, W)$ derived considering dynamics of the locomotive, limitation of the velocity, and control rules. Thus, the processing time for the unit transition t_{U} is calculated by $t_{\mathrm{U}} = \mathrm{t_E} + \sum_{i=1}^{2} f(D_i, 0) + \sum_{j=4}^{5} f(D_j, W)$. The maximum value of t_{U} is define as $t_{\max}$ and is calculated by

$$\begin{aligned} t_{\max} &= \mathrm{t_E} + f(kD_{\min_v}, 0) + f(mD_{\min_h}, 0) \\ &\quad + f(mD_{\min_h} + n, W) + f(kD_{\min_v}, W) \quad (1) \end{aligned}$$

VII. LEARNING ALGORITHM

Define h as the number of candidates of the desired layout of groups. Each candidate is represented by u_{j_1} $(1 \le u_{j_1} \le$

h), and evaluated by $Q_1(u_{j_1})$. Then, $Q_1(u_{j_1})$ is updated by the following equation when one of desired layout is achieved in the main track:

$$Q_1(u_{j_1}) \leftarrow \max\{Q_1(u_{j_1}), (1-\alpha)Q_1(u_{j_1}) + \alpha\gamma^l R\}, \quad (2)$$

where l denotes the total movement counts required to achieve the desired layout, α is learning rate, γ is discount factor, R is reward that is given only when one of desired layout is achieved in the main track.

Define $\boldsymbol{s}(t)$ as the state at time t, r_{M} as the sub track selected as the destination for the removed car, p_{M} as the number of removed cars, q as the movement counts of freight cars by direct rearrangement, and $\boldsymbol{s}'$ as the state that follows $\boldsymbol{s}$. Also, Q_2, Q_3, Q_4 are defined as evaluation values for $(\boldsymbol{s}_1, u_{j_2})$, $(\boldsymbol{s}_2, u_{j_3})$, $(\boldsymbol{s}_3, u_{j_4})$, respectively, where $\boldsymbol{s}_1 = \boldsymbol{s}, \boldsymbol{s}_2 = [\boldsymbol{s}, \mathrm{c}_T]$, $\boldsymbol{s}_3 = [\boldsymbol{s}, \mathrm{c}_T, r_{\mathrm{M}}]$. $Q_2(\boldsymbol{s}_1, u_{j_2})$, $Q_3(\boldsymbol{s}_2, u_{j_3})$ and $Q_4(\boldsymbol{s}_3, u_{j_4})$ are updated by following rules:

$$Q_2 \quad (\boldsymbol{s}_1, \mathrm{c}_T) \leftarrow \max_{u_{j_3}} Q_3(\boldsymbol{s}_2, u_{j_3}), \quad (3)$$

$$Q_3 \quad (\boldsymbol{s}_2, r_{\mathrm{M}}) \leftarrow \max_{u_{j_4}} Q_4(\boldsymbol{s}_3, u_{j_4}), \quad (4)$$

$$Q_4 \quad (\boldsymbol{s}_3, p_{\mathrm{M}}) \leftarrow \quad (5)$$

$$\begin{cases} (1-\alpha)Q_4(\boldsymbol{s}_3, p_{\mathrm{M}}) + \alpha[R + \gamma^{q+1}V_1] \\ \qquad \text{(next action is rearrangement)} \\ (1-\alpha)Q_4(\boldsymbol{s}_3, p_{\mathrm{M}}) + \alpha[R + \gamma V_2] \\ \qquad \text{(next action is removal)} \end{cases}$$

$$V_1 = \max_{u_{j_1}} Q_2(\boldsymbol{s}'_1, u_{j_2}),$$

$$V_2 = \max_{u_{j_2}} Q_3(\boldsymbol{s}'_2, u_{j_3})$$

where α is the learning rate, R is the reward that is given when one of desirable layout is achieved, and γ is the discount factor that is used to reflect the processing time of the marshaling and calculated by the following equation.

$$\gamma = \delta \frac{t_{\max} - \beta t_{\mathrm{U}}}{t_{\max}}, \; 0 < \beta < 1, 0 < \delta < 1 \quad (6)$$

Propagating Q-values by using eqs.(3)-(6), Q-values are discounted according to the processing time of marshaling. In other words, by selecting the removal destination that has the largest Q-value, the processing time of the marshaling can be reduced.

In the learning stages, each $u_j \, (1 \le j \le h + 2m + \min\{p_{\mathrm{S}}, p_{\mathrm{d}}\} - 1)$ is selected by the soft-max action selection method[10]. Probability P for selection of each candidate is calculated by

$$\tilde{Q}_i(\boldsymbol{s}, u_{j_i}) = \frac{Q_i(\boldsymbol{s}, u_{j_i}) - \min\limits_u Q_i(\boldsymbol{s}, u_{j_i})}{\max\limits_u Q_i(\boldsymbol{s}, u_{j_i}) - \min\limits_u Q_i(\boldsymbol{s}, u_{j_i})} \quad (7)$$

$$P(\boldsymbol{s}_i, u_{j_i}) = \frac{\exp(\tilde{Q}_i(\boldsymbol{s}_{j_i}, u_{j_i})/\xi)}{\sum\limits_{u \in u_{j_i}} \exp(\tilde{Q}_i(\boldsymbol{s}_i, u)/\xi)}, \quad (8)$$

$$(i = 1, 2, 3, 4).$$

In the addressed problem, Q_1, Q_2, Q_3, Q_4 become smaller when the number of discounts becomes larger. Then, for complex problems, the difference between probabilities in candidate selection remain small at the initial state and large at final state before achieving desired layout, even after repetitive learning. In this case, obtained evaluation does not contribute to selections in initial stage of marshaling process, and search movements to reduce the transfer distance of locomotive is spoiled in final stage. To conquer this drawback, Q_1, Q_2, Q_3, Q_4 are normalized by eq.(7), and the thermo constant ξ is switched from ξ_1 to ξ_2 ($\xi_1 > \xi_2$) when the following condition is satisfied:

$$\begin{aligned} &[\text{The count of } Q_i(\boldsymbol{s}_{j_i}, u_{j_i})] > \eta, \\ &\text{s.t. } Q_i(\boldsymbol{s}_{j_i}, u_{j_i}) > 0, \\ &0 < \eta \le [\text{the number of candidates for } u_{j_i}] \end{aligned} \quad (9)$$

where η is the threshold to judge the progress of learning.

The proposed learning algorithm can be summarized as follows:

1) Initialize all the Q-values as 0
2) Determine the layout of the main track among u_{j_1}.
3) Conduct direct rearrangements.
4) If no cars are in sub tracks, go to 10, otherwise go to 5
5) ⓐ Determine c_T among the candidates by roulette selection (probabilities are calculated by eq. (8)),
   ⓑ Put reward as R = 0,
   ⓒ Update the corresponding $Q_4(\boldsymbol{s}_3, p_{\mathrm{M}})$ by eq.(5)
   ⓓ Store $\boldsymbol{s}_1, \mathrm{c}_T$
6) ⓐ Determine r_{M}(probability for the selection is calculated by eq.(8))
   ⓑ Update corresponding $Q_3(\boldsymbol{s}_2, r_{\mathrm{M}})$ by eq.(4),
   ⓒ store $\boldsymbol{s}_2, r_{\mathrm{M}}$
7) ⓐ Determine p_{M}(probability for the selection is calculated by eq.(8))
   ⓑ Update corresponding $Q_4(\boldsymbol{s}_3, p_{\mathrm{M}})$ by eq.(5)
   ⓒ Store $\boldsymbol{s}_3, p_{\mathrm{M}}$
8) Remove p_{M} cars and place at r_{M}
9) Go to 3
10) Receive the reward R, update $Q_1(\boldsymbol{s}_1, \mathrm{c}_T)$ by eq.(3)

VIII. COMPUTER SIMULATIONS

Computer simulations are conducted for $m = 12, n = 6, k = 36$ and learning performances of following 5 methods are compared:

(A) proposed method that evaluates the processing time of the marshaling operation, considering the layout of groups,
(B) a method that evaluates the transfer distance of the locamotive considering the layout of groups[11],
(C) a method that evaluates the number of movements of freight cars, considering the layout of gruops[11]
(D) a method that evaluates the processing time, with single layout of groups.

The initial arrangement of cars in sub tracks is described in Fig.7. In this case, the rearrantement order of groups is $\mathrm{group}_1, \mathrm{group}_2$, group_3, group_4. Cars $\mathrm{c}_1, \cdots, \mathrm{c}_9$ are in group_1, $\mathrm{c}_{10}, \cdots, \mathrm{c}_{18}$ are in group_2, $\mathrm{c}_{19}, \cdots, \mathrm{c}_{27}$ are in group_3, and $\mathrm{c}_{28}, \cdots, \mathrm{c}_{36}$ are in group_4. Other parameters are set as $\alpha = 0.9, \beta = 0.2, \delta = 0.9, R = 1.0, \eta = 0.95, \xi_1 = 0.1, \xi_2 = 0.05$. In method (C), the discount factor γ is assumed to be constant, and set as $\gamma = 0.9$ instead of calculationg by eq.(6).

The locomotive assumed to accerarate and deaccerarate the train with the constant force 100×10^3N, and to be $100 \times$

ISBN: 978-988-19251-1-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE I
TOTAL PROCESSING TIME

methos	processing time (sec.) best	average	worst
method (A)	5328.06	5376.33	5407.88
method (B)	5331.81	5390.69	5423.99
method (C)	5337.18	5366.46	5416.54
method (D)	5688.26	5763.90	5839.88

10^3kg in weight. Also, all the freight cars have the same weight, 10×10^3kg. The locomotive and freight cars assumed to have the same length, and $D_{\min_v} = D_{\min_h} = 20$m. The velocity of the locomotive is limited to no more than 10m/s. Then, the locomotive accerarates the train until the velocity arrives 10m/s, keeps the velocity, and deaccerarates until the train stops within the indicated distance. When the velocity does not arrive 10m/s at the half way point, the locomotive starts to deaccerarate immediately. Then, $t_{\max} = 462$.

Fig.8 show the results. In Fig.8, horizontal axis expresses the number of trials and the vertical axis expresses the minimum prcessing time to achieve a desirable layout found in the past trials. Each result is averaged over 20 independent simulations. In Fig.8, the learning performance of method (A) is better than that of method (D), because solutions derived by method (A) uses the extended layout of groups effectively for reducing the total processing time. In method (C), the learning algorithm evaluates the number of movements of freight cars, and is not effective to reduce the total processing time. In method (B), only the total transfer distance of locamotive is evaluated, so that the total processing time is not improved adequately even if many trials are repeated. Total transfer distances of the locomotive at 1.5×10^6th trial are described in table.I for each method. Fig.9 shows final arrangements of freight cars generated by the best solutions derived by methods (A) and (D). Since the layout of group is extended, method (A) learns the layout of groups in order to reduce the total processing time, whereas the layout is fixed to the ascending order in method (D).

				c_{19}	c_{32}	c_{31}					
c_{10}		c_{34}	c_{28}	c_{16}	c_4	c_5		c_{25}	c_{29}		c_{36}
c_2	c_{13}	c_8	c_{11}	c_{14}	c_{17}	c_{20}	c_{23}	c_{26}	c_1	c_{33}	c_{35}
c_3	c_6	c_9	c_{12}	c_{15}	c_{18}	c_{21}	c_{24}	c_{27}	c_{30}	c_{22}	c_7

Fig. 7. Initial layout

IX. CONCLUSIONS

A new scheduling method has been proposed in order to rearrange and line cars in the desirable order onto the main track. The learning algorithm of the proposed method is derived based on the reinforcement learning, considering the total processing time of marshaling. In order to reduce the total processing time of marshaling, the proposed method learns the layout of groups, as well as the arrangement of freight cars in each group, the rearrangement order of cars, the number of cars to be moved and the removal destination of cars, simultaneously. In computer simulations, learning performance of the proposed method has been improved by using normalized evaluation and switching thermo constants in accordance with the progress of learning.

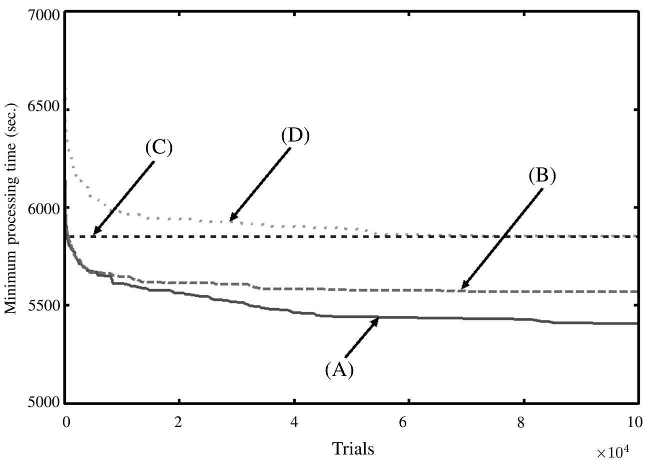

Fig. 8. Comparison of learning performances

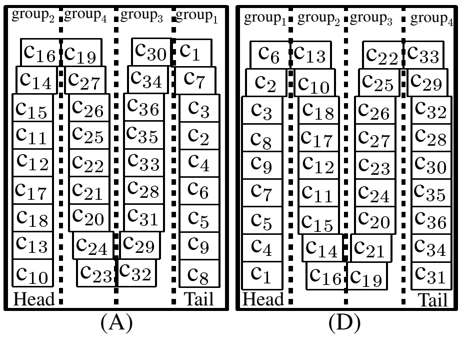

Fig. 9. Final layouts

REFERENCES

[1] G. LI, M. MUTO, N. AIHARA, and T. TSUJIMURA, “Environmental load reduction due to modal shift resulting from improvements to railway freight stations,” *Quarterly Report of RTRI*, vol. 48, no. 4, pp. 207–214, 2007.

[2] U. Blasum, M. R. Bussieck, W. Hochstättler, C. Moll, H.-H. Scheel, and T. Winter, “Scheduling trams in the morning,” *Mathematical Methods of Operations Research*, vol. 49, no. 1, pp. 137–148, 2000.

[3] R. Jacob, P. Marton, J. Maue, and M. Nunkesser, “Multistage methods for freight train classification,” in *Proceedings of 7th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems*, 2007, pp. 158–174.

[4] L. Kroon, R. Lentink, and A. Schrijver, “Shunting of passenger train units: an integrated approach,” *Transportation Science*, vol. 42, pp. 436–449, 2008.

[5] N. TOMII and Z. L. Jian, “Depot shunting scheduling with combining genetic algorithm and pert,” in *Proceedings of 7th International Conference on Computer Aided Design, Manufacture and Operation in the Railway and Other Advanced Mass Transit Systems*, 2000, pp. 437–446.

[6] S. He, R. Song, and S. Chaudhry, “Fuzzy dispatching model and genetic algorithms for railyards operations,” *European Journal of Operational Research*, vol. 124, no. 2, pp. 307–331, 2000.

[7] E. Dahlhaus, F. Manne, M. Miller, and J. Ryan, “Algorithms for combinatorial problems related to train marshalling,” in *Proceedings of the 11th Australasian Workshop on Combinatorial algorithms*, 2000, pp. 7–16.

[8] C. Eggermont, C. A. J. Hurkens, M. Modelski, and G. J. Woeginger, “The hardness of train rearrangements,” *Operations Research Letters*, vol. 37, pp. 80–82, 2009.

[9] C. J. C. H. Watkins and P. Dayan, “Q-learning,” *Machine Learning*, vol. 8, pp. 279–292, 1992.

[10] R. Sutton and A. Barto, *Reinforcement Learning*. MIT Press, 1999.

[11] Y. Hirashima, “A new reinforcement learning method for train marshaling based on the transfer distance of locomotive,” *Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2011, IMECS 2011*, pp. 80–85, 2011.

ISBN: 978-988-19251-1-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)