

On Reinforcement Learning Methods for Generating Train Marshaling Plan Considering Group Layout of Freight Cars

Yoichi Hirashima

Abstract—This paper proposes a new reinforcement learning method for train marshaling. In the proposed method, marshaling plans for freight cars in a train are generated based on the processing time. 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. In computer simulations, a performance of the proposed method is compared to that of the method based on the movement counts of freight cars and of the method based on the transfer distance of locomotive.

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

I. INTRODUCTION

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 delivering corresponding freight cars to each freight station. Also, logistics with freight train is paid attention as an effective transport in ecological aspects, because railway logistics is known to have smaller environmental load as compared to goods transportation with trucks ([1]).

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. Recently, methods that generates each transfer movement of freight car in a marshaling are designed based on the movement counts of freight cars, and based on the transfer distance of a locomotive[9]. By extending these methods, a reinforcement learning method that can obtain and improve marshaling plans based on the processing time has been proposed[10].

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. coupling cars with the locomotive, 3. transferring cars to their new position by the locomotive, and 4. decoupling the cars from the locomotive. The processing times for elements 1. and 3. are determined 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

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 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 computer simulations, a performance of the proposed method is compared to that of the method based on the movement counts of freight cars and of the method based on the transfer distance of locomotive.

The diagram shows six sequences, labeled [1] through [6], each consisting of a column of boxes representing elements. Brackets on the right side of each column group the elements into six categories, labeled [a] through [f] from bottom to top. Sequence [1] contains elements c₁, c₂, c₃, c₄, c₅, c₆. Sequence [2] contains c₇, c₈, c₉, c₁₀, c₁₁, c₁₂. Sequence [3] contains c₁₃, c₁₄, c₁₅, c₁₆. Sequence [4] contains c₁₇, c₁₈. Sequence [5] contains c₁₉. Sequence [6] contains c₂₀, c₂₁, c₂₂, c₂₃, c₂₄, c₂₅. On the far left, a vertical axis is labeled T_m at the bottom. Above the sequences, a diagonal line slopes downwards from left to right. To the left of this line, there is a box labeled c₃₀ and another box labeled c₃₇. To the right of the line, there is a shaded triangular region with a hatched pattern, and a box labeled c₃₅ is positioned near the top-left corner of this region.

—

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

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.

Figure 1 consists of two parts: Position index and Yard layout.

Position index: A 6x6 grid with rows labeled [f], [e], [d], [c], [b], [a] and columns labeled [1], [2], [3], [4], [5], [6]. The values in the grid are:

[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: A triangular grid with rows of containers C6, C5, C4, C3, C2, C1 and columns labeled [1], [2], [3], [4], [5], [6]. The values in the grid are:

C6	C12				C20
C5	C11				C24
C4	C10	C16			C23
C3	C9	C15			C22
C2	C8	C14	C18		C21
C1	C7	C13	C17	C19	C25

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

γ β α

freight cars c_1, \dots, c_6 to the destination₁ make group₁, c_7, \dots, c_{18} to the destination₂ make group₂, c_{19}, \dots, c_{25} to the destination₃ make group₃, and c_{26}, \dots, c_{30} to the destination₄ make group₄. 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 group₁ are in the dashed square.

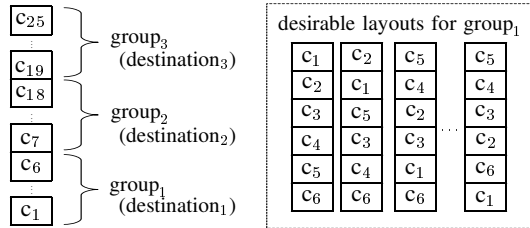


Fig. 3. Example of groups

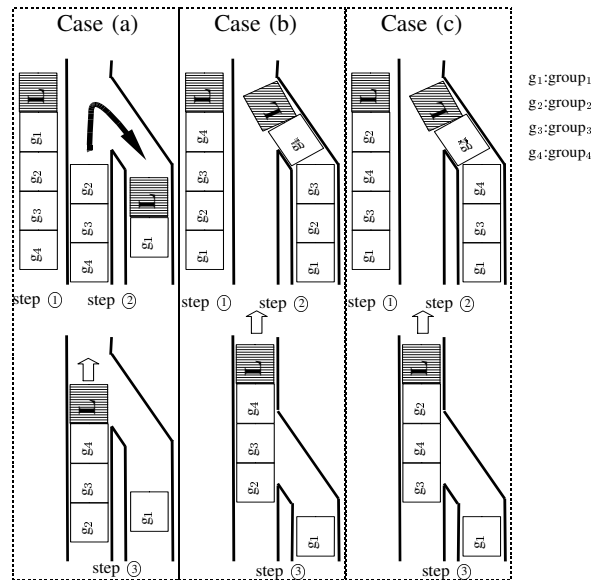


Fig. 4. Group layouts

Also, the layout of groups lined by the reverse order do not yield additional removal actions at the destination of each 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 group₁. In the figure, step ① shows the layout of the incoming train. In case (a), cars in group₁ are separated at the main track, and moved to a sub-track by the locomotive L at step ②. In cases (b),(c), cars in group₁ are carried in a sub-track, and group₁ is separated at the sub-track. In the cases, group₁ 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 coupled 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 group₁, c_5, c_6, c_7, c_8 are in group₂, and group₁ 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 adjacent positions are coupled and moved to the main track by a direct rearrangement operation. In addition, at step 5 in case 2, cars in group₂ and group₃ 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.

V. REARRANGEMENT PROCESS

The rearrangement process for cars consists of following 6 operations :

- (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} ($1 \leq j_1 \leq h$).

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} ($1 \leq l \leq r$), and u_{j_2} ($h + 1 \leq j_2 \leq h + g_l$) as candidates of c_T .

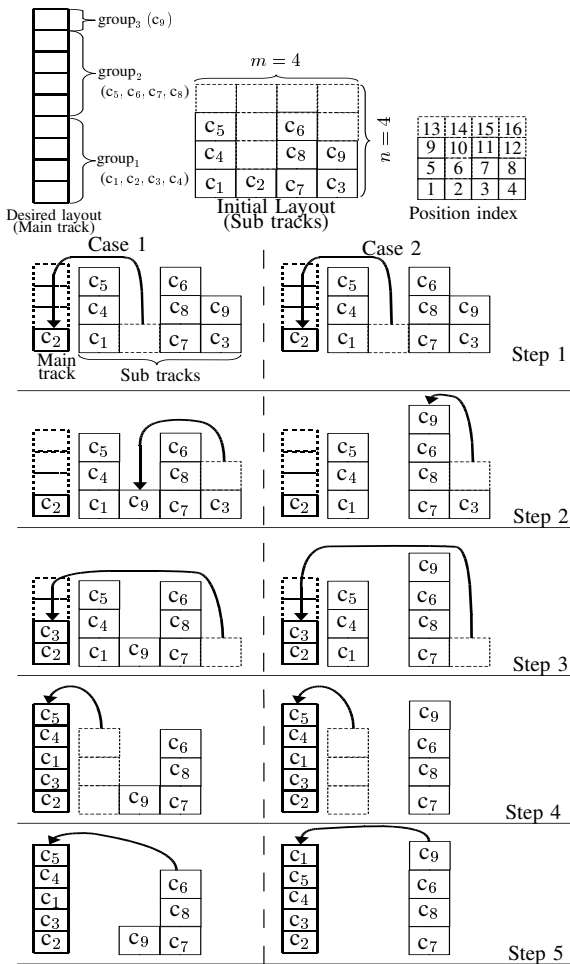


Fig. 5. Direct rearrangements

In the operation (III), the removal destination of car located on the car to be rearranged is defined as c_M . Then, defining $u_{j_3} (h+g_l+1 \leq j_3 \leq h+g_l+m-1)$ 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 $u_{j_4}, 2m \leq j_4 \leq 2m + \min\{p, q\} - 1$.

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 group₁ 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 c_1, c_4, 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). couples them, (E4) pull out them to the joint track, (E5) restart in reverse direction, and transfers them to the indicated location, and (E6) decouples 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.

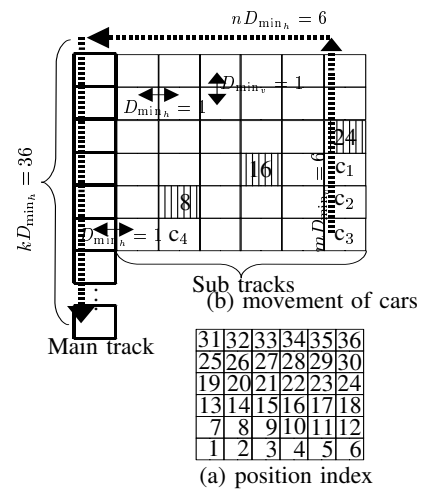


Fig. 6. Calculation of transfer distance

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 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 $t_U = t_E + \sum_{i=1}^2 f(D_i, 0) + \sum_{j=4}^5 f(D_j, W)$. The maximum value of t_U is define as t_{\max} and is calculated by

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

VII. LEARNING ALGORITHM

Define h as the number of candidates of the desired layout of groups. Each candidate is represented by u_{j_1} ($1 \leq u_{j_1} \leq h$), 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 $s(t)$ 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 s' as the state that follows s . Also, Q_2, Q_3, Q_4 are defined as evaluation values for (s_1, u_{j_2}) , (s_2, u_{j_3}) , (s_3, u_{j_4}) , respectively, where $s_1 = s$, $s_2 = [s, c_T]$, $s_3 = [s, c_T, r_M]$. $Q_2(s_1, u_{j_2})$, $Q_3(s_2, u_{j_3})$ and $Q_4(s_3, u_{j_4})$ are updated by following rules:

$$Q_2(s_1, c_T) \leftarrow \max_{u_{j_3}} Q_3(s_2, u_{j_3}), \quad (3)$$

$$Q_3(s_2, r_M) \leftarrow \max_{u_{j_4}} Q_4(s_3, u_{j_4}), \quad (4)$$

$$Q_4(s_3, p_M) \leftarrow \begin{cases} (1 - \alpha)Q_4(s_3, p_M) + \alpha[R + \gamma^{q+1}V_1] \\ \text{(next action is rearrangement)} \\ (1 - \alpha)Q_4(s_3, p_M) + \alpha[R + \gamma V_2] \\ \text{(next action is removal)} \end{cases} \quad (5)$$

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

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

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.

A. Calculation of γ

1) *discount for each movement of freight car:* When Q-values evaluate the total movement counts of freight cars, the discount factor γ is set as constant.

2) *evaluation of transfer distance of a locomotive:* When Q-values are used to evaluate the total transfer distance of locomotive in a marshaling process, the discount factor γ is calculated by the following equation:

$$\gamma = \delta \frac{D_{\max} - \beta D}{D_{\max}}, \quad 0 < \beta < 1, 0 < \delta < 1. \quad (6)$$

3) *evaluation of total processing time of a marshaling:* When Q-values are used to reflect the processing time of the marshaling and calculated by the following equation:

$$\gamma = \delta \frac{t_{\max} - \beta t_U}{t_{\max}}, \quad 0 < \beta < 1, 0 < \delta < 1. \quad (7)$$

In this case, propagating Q-values by using eqs.(3)-(7), 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 ($1 \leq j \leq h + 2m + \min\{p_s, p_d\} - 1$) is selected by the soft-max action selection method[12]. Probability P for selection of each candidate is calculated by

$$\tilde{Q}_i(s, u_{j_i}) = \frac{Q_i(s, u_{j_i}) - \min_u Q_i(s, u_{j_i})}{\max_u Q_i(s, u_{j_i}) - \min_u Q_i(s, u_{j_i})} \quad (8)$$

$$P(s_i, u_{j_i}) = \frac{\exp(\tilde{Q}_i(s_{j_i}, u_{j_i})/\xi)}{\sum_{u \in u_{j_i}} \exp(\tilde{Q}_i(s_i, u)/\xi)}, \quad (9)$$

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

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.(8), and the thermo constant ξ is switched from ξ_1 to ξ_2 ($\xi_1 > \xi_2$) when the following condition is satisfied:

$$[\text{The count of } Q_i(s_{j_i}, u_{j_i})] > \eta, \quad \text{s.t. } Q_i(s_{j_i}, u_{j_i}) > 0, \quad (10)$$

$$0 < \eta \leq [\text{the number of candidates for } u_{j_i}]$$

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) (a) Determine c_T among the candidates by roulette selection (probabilities are calculated by eq. (9)),
- (b) Put reward as $R = 0$,
- (c) Update the corresponding $Q_4(s_3, p_M)$ by eq.(5)
- (d) Store s_1, c_T
- 6) (a) Determine r_M (probability for the selection is calculated by eq.(9))
- (b) Update corresponding $Q_3(s_2, r_M)$ by eq.(4),
- (c) store s_2, r_M
- 7) (a) Determine p_M (probability for the selection is calculated by eq.(9))

- ## VIII. COMPUTER SIMULATIONS

methods	processing time (sec.)		
	best	average	worst
method (A)	7995.37	8210.09	8291.27
method (B)	7976.36	8241.77	8431.66
method (C)	8581.96	8894.30	9021.15
method (D)	9835.27	9835.27	9835.27
method (E)	9835.27	9835.50	9835.60
method (F)	9848.57	9887.30	9914.53

weight of a freight car	number of cars in a movement method (A)	method (B)
$10 \times 10^3 \text{ kg}$	1.543	1.447
$250 \times 10^3 \text{ kg}$	1.385	1.415

- Fig.11 shows final arrangements of freight cars generated by the best solutions derived by method (A) for cars $10 \times 10^3\text{kg}$ and for $250 \times 10^3\text{kg}$ in weight. The rearrangement orders of freight cars are different in order to reduce the total processing time in accordance with the weight of cars.

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

IX. CONCLUSIONS

(Advance online publication: 28 August 2012)

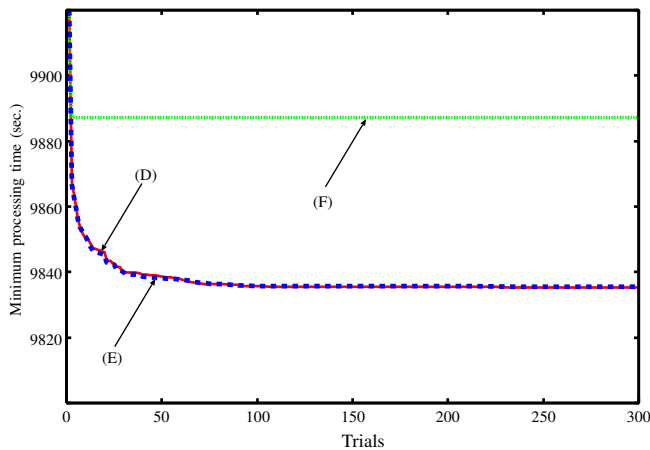


Fig. 8. Comparison of learning performances for methods (D),(E),(F)

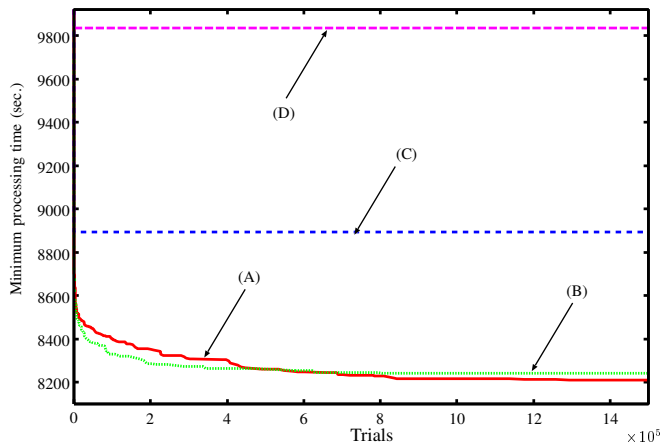


Fig. 9. Comparison of learning performances for methods (A),(B),(C),(D)

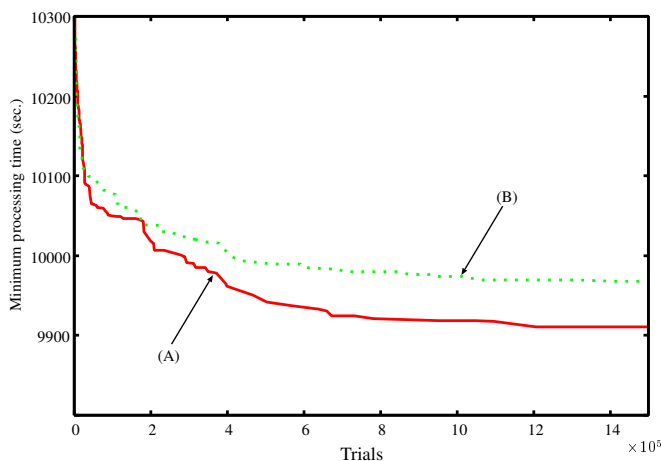


Fig. 10. Comparison of learning performances for heavy freight cars

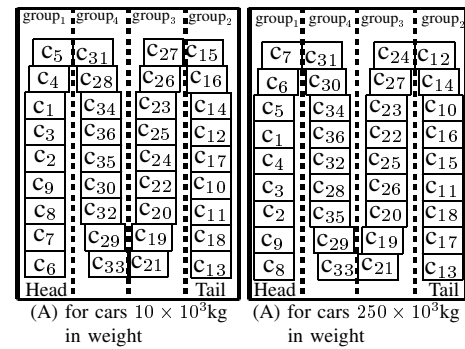


Fig. 11. Final layouts

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, the proposed method generates and improves marshaling plans according to the weight of a freight car.

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] Y. Hirashima, "A reinforcement learning method for train marshaling based on movements of locomotive," *IAENG International Journal of Computer Science*, vol. 38, no. 3, pp. 242–248, 2011.
- [10] —, "A reinforcement learning for train marshaling based on the processing time considering group layout of freight cars," *Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2012, IMECS 2012, 14-16 March, 2012, Hong Kong*, pp. 47–52, 2012.
- [11] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, pp. 279–292, 1992.
- [12] R. Sutton and A. Barto, *Reinforcement Learning*. MIT Press, 1999.