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 outbound 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

1. 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 disjoining 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.

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. disjoining 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 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.
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, ..., k). Fig.1 shows the outline of freight yard in the case k = 30, m = n = 6. In the figure, 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.

![Fig. 1. Freight yard](image)

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. 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, ..., k) as the position number of the car c_i, and s = [x_1, ..., 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.

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, ..., c_6 to the destination1 make group1, c_7, ..., c_18 to the destination2 make group2, c_19, ..., c_25 to the destination3 make group3, and c_26, ..., c_30 to the destination4 make group4. Groups 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.

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 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 locomotive 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 \(c_9\) 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 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 \(c_7, c_8, c_9\) 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:
The unit transition is as follows: (E1) starts without freight cars. In the operation (III), the candidate numbers of cars to be moved
are determined by excluding freight cars that have already been rearranged to the main track. These candidates must belong to the same group.
In the operation (III), the removal destination of the car located on the car to be rearranged is defined as $c_{LM}$. Then, defining $u_{j} = (h + g + 1) + j$, the target is located on the position $18$, the destination of the target is $4$, and the number of cars to be moved is $2$. The distance of 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 $t_U = t_E + \sum_{i=1}^{n} f(D_i, W) - f(mD_{min}, W)$, the maximum value of $t_U$ is defined as $t_{max}$ and is calculated by

$$t_{max} = t_E + f(kD_{min}, W) + f(mD_{min}, W)$$

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 $D_{min}$, the length of joint track between adjacent sub tracks, or, sub track and main track as $D_{min}$. 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 shows an example of transfer distance. In the figure, $m = n = 6$, $D_{min} = D_{min} = 1$, $k = 18$, (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 $D_3 + D_4 = 13$ ($D_3 = 7, D_4 = 6$).
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:

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

where $l$ denotes the total movement counts required to achieve the desired layout, $\alpha$ is learning rate, $\gamma$ 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 $(s_1, u_{j_1}), (s_2, u_{j_2}), (s_3, u_{j_3})$, respectively, where $s_1 = s, s_2 = \{s, c_T \}, s_3 = \{s, c_T, r_M \}$. $Q_2(s_1, u_{j_1}), Q_3(s_2, u_{j_2}), Q_4(s_3, u_{j_3})$ are updated by following rules:

$$Q_2(s_1, c_T) \leftarrow \max_{s_2} Q_3(s_2, u_{j_2}) \tag{3}$$

$$Q_3(s_2, r_M) \leftarrow \max_{s_3} Q_4(s_3, u_{j_3}) \tag{4}$$

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

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

$$\gamma = 0 < \beta < 1, 0 < \delta < 1 \tag{6}$$

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 $u_j (1 \leq j \leq h + 2m + \min \{p_M, p_d\} - 1)$ is selected by the soft-max action selection method[10]. Probability $P$ for selection of each candidate is calculated by

$$P(s_i, u_{j_i}) = \frac{\exp(\bar{Q}(s_i, u_{j_i})/\xi)}{\sum_{(s_i, u_{j_i})} \exp(\bar{Q}(s_i, u_{j_i})/\xi)} \tag{8}$$

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 stage 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 $\xi$ is switched from $\xi_1$ to $\xi_2$ ($\xi_1 > \xi_2$) when the following condition is satisfied:

$$[\text{The count of } Q_i(s_{j_i}, u_{j_i}) > \eta]$$

s.t. $Q_i(s_{j_i}, u_{j_i}) > 0 \tag{9}$

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

where $\eta$ 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 $u_{j_i}$.
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. (8)),
6. (b) Put reward as $R = 0$,
7. (c) Update the corresponding $Q_4(s_3, p_M)$ by eq.(5)
8. (d) Store $s_1, c_T$
9. (e) Determine $c_T = \text{probability for the selection is calculated by eq.(8)}$
10. (f) Update corresponding $Q_3(s_2, r_M)$ by eq.(4), store $s_2, r_M$
11. (g) Determine $p_M = \text{probability for the selection is calculated by eq.(8)}$
12. (h) Update corresponding $Q_4(s_3, p_M)$ by eq.(5)
13. (i) Store $s_3, p_M$
14. (j) Remove $p_M$ cars and place at $r_M$
15. (k) Go to 3
16. (l) Receive the reward $R$, update $Q_1(s_1, c_T)$ 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 locomotive considering the layout of groups[11],
(C) a method that evaluates the number of movements of freight cars, considering the layout of groups[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 rearrangement order of groups is group1, group2, group3, group4. Cars $c_1, \cdots, c_8$ are in group1, $c_{10}, \cdots, c_{18}$ are in group2, $c_{20}, \cdots, c_{28}$ are in group3, and $c_{30}, \cdots, c_{36}$ are in group4. Other parameters are set as $\alpha = 0.9, \beta = 0.2, \delta = 0.1, R = 1.0, \eta = 0.95, \xi_1 = 0.1, \xi_2 = 0.05$. In method (C), the discount factor $\gamma$ is assumed to be constant, and set as $\gamma = 0.9$ instead of calculation by eq.(6).

The locomotive assumed to accelerate and decelerate the train with the constant force $100 \times 10^2 N$, and to be $100 \times$
10^3kg in weight. Also, all the freight cars have the same weight, 10 \times 10^3kg. The locomotive and freight cars assumed to have the same length, and \( D_{\text{min}} = D_{\text{min}} = 20\text{m}. \) The velocity of the locomotive is limited to no more than 10m/s. Then, the locomotive accelerates the train until the velocity arrives 10m/s, keeps the velocity, and decelerates 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 decelerate immediately. Then, \( t_{\text{max}} = 462. \)

Fig.8 show the results. In Fig.8, horizontal axis expresses the number of trials and the vertical axis expresses the minimum processing 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 locomotive 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 \times 10^4th trial are described in Table.I for each method.

![Fig. 7. Initial layout](image)

**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](image)

**Fig. 9. Final layouts**

---

**REFERENCES**


