2020 2nd International Conference on Information Technology and Computer Application (ITCA) | 978-1-6654-0378-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/ITCA52113.2020.00053

Order Batch Optimization Strategy Based on Improved K-Means Clustering Algorithm

Wenwu Ping
Hangzhou Dianzi University
Hangzhou, China
Email: pingww@hdu.edu.cn

Houpan Zhou
Hangzhou Dianzi University
Hangzhou, China
Email: houpanzz@hdu.edu.cn

***Abstract*—Order picking is the most important part of warehousing logistics. In order to improve the efficiency of order picking, more and more automatic picking equipment is applied to logistics distribution center. The automatic picking equipment adopts the model described as "Out Stock -Picking- In Stock ". So the number of times of bin extractions becomes the key factor to restrict the picking efficiency. In order to reduce the number of bin extractions and improve the efficiency of picking equipment, this paper takes the automatic picking equipment introduced by a company as the research object, establishes the order batch model based on the information of merchandise storage bins, redefines the clustering center and the distance from order to class center, takes the distance from order to class center as the clustering index, and transforms the optimization goal from improving the order picking efficiency to reducing the number of times of bin extractions. On this basis, the K-Means algorithm is improved by using the idea of seed order to solve the model. And with the example data of a company to experiment, the results show that the improved algorithm proposed in this paper has significantly improved the picking efficiency compared to the typical order picking algorithm FCFS and a mainstream cross-border e-commerce warehouse picking algorithm.**

Keywords-component; automatic picking equipment; K-Means; Order in batches.

I. INTRODUCTION

With the rapid popularization of online shopping, e-commerce enterprises have to deal with more and more orders every day. Order picking is one of the essential part of warehousing and logistics. Its workload accounts for about 60% of the total workload of the distribution center, and its operation time accounts for more than 40% of the total operation time of the distribution center[1]. In order to improve the efficiency of picking, a company introduced automatic picking equipment, but the order batching strategy[2] is still one of the most important factors that directly affects the efficiency and the cost of order picking. Therefore, how to improve the work efficiency by optimizing the order picking batch has become a problem to be solved in the automatic picking system.

In recent years, scholars at home and abroad have carried out a preliminary study on the order batching problem of unmanned warehouse systems. Wu Yingying et al[3]. put forward an improved K-Means clustering algorithm to optimize the order sorting based on the order coupling factor, aiming at solving the picking optimization problem of the goods to manual picking system. Li et al[4]. established an integer programming model aiming at minimizing the total cost of order batch picking, and designed a K-Max clustering algorithm based on classified data. Shao Zeyi and Dong Baoli[5] used density and minimum distance to determine multiple initial clustering centers, and use improved genetic K-Means algorithm to determine the optimal batch number and optimize the order batch. Hsieh and Huang[6] constructed two new order merging rules: K-Means order merging (KMB) and self-organizing neural network order merging (SOMB), and verified the superiority of the algorithm through simulation experiments. Pan et al[7]. proposed an order batching method based on population genetic algorithm to balance the workload of each picking area and minimize the number of batches in the picking system, so as to improve the system performance. It can be seen from the above literature that most research is limited to allocating high similarity products to the same picking station or combining them into a batch. This goal emphasizes that the time of picking a single batch is the shortest, but it cannot guarantee the highest order picking efficiency for completing all orders.

In order to improve the efficiency of order picking, this paper sets up a batch model based on the information of merchandise storage bin. We redefined the clustering center and the distance between the order and the clustering center. Meanwhile the distance between the order and the cluster center is taken as the clustering index. We change the optimization target to improve the efficiency of order picking by reducing the number of times of bin extractions, use the idea of seed orders[8] to improve the K-Means algorithm and replace the model to achieve the overall optimal order picking.

II. PROBLEM DESCRIPTION AND MODEL CONSTRUCTION

A. *Automatic Picking Equipment*

The automatic sorting equipment is a new type of automatic three-dimensional warehouse, which is composed of two rows of high-density shelves and Automatic Guided Vehicles. The merchandise is placed in the material bins and placed on the shelf. The AGV(Automatic Guided Vehicles) group can travel flexibly and rapidly in the roadway of the equipment, enter each storage position in the aisle horizontally or vertically, and continuously extract the material bins from the shelves on both sides of the roadway for storage and retrieval. The goods are sent to the picking table for the manual picking process, as shown in Figure 1. The picked goods are transmitted to the automatic seeding equipment through the track, and the goods from single order are seeded into a material bin to complete the order picking. In order to save the cost of equipment operation and improve the efficiency of equipment picking, we need to develop a reasonable batch strategy.

978-1-6654-0378-8/20/$31.00 ©2020 IEEE
DOI 10.1109/ITCA52113.2020.00053

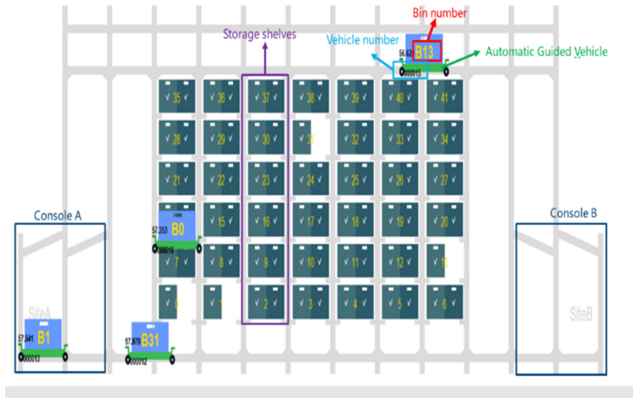

Figure 1. Simplified model of automated picking equipment

In order to ensure the feasibility and correctness of the model, we make the following assumptions:

- Each order contains at least one item.
- There can be multiple kinds of merchandise placed in the same bin, but no more than 8 kinds, each merchandise will correspond to a bin.
- The number of merchandise in every bin is sufficient to meet the task of each batch of picking, and there will be no shortage of goods and urgent insertion of orders.
- Assume it takes the same amount of time for every bin from their original position to get to the picking station.
- It is not allowed to split the same order into different batches.
- During the transportation process, the Automatic Guided Vehicle has sufficient power supply and the running speed is consistent.

B. *Description of Order-Batching Model*

The order batching problem in the automatic picking system can be described as: there are M kinds of goods in the picking equipment, which are stored in S bins, and the bins are placed on the shelves. The storage position of each merchandise in the material bin of the equipment is known, and the storage quantity of each merchandise in the corresponding storage position is sufficient. Supposing that there are N orders in the order pool that need to be picked in a certain period of time, how to batch these N orders can minimize the number of times of bin extractions. In order to establish the mathematical model of order batch problem, the following symbols are introduced:

i: the number of order, $i = 1,2,\dots,N$;

k: the number of order batch, $k = 1,2,\dots,K$;

s: the number of bin, $s = 1,2,\dots,S$;

t: product number, $t = 1,2,\dots,M$;

q: maximum number of orders allowed per batch;

Define decision variables:

$$x_{ik} = \begin{cases} 1, \textit{if order } i \textit{ in batch } k \\ 0, \textit{otherwise} \end{cases} \tag{1}$$

$$y_{ks} = \begin{cases} 1, \textit{if batch } k \textit{ requires extraction bin } s \\ 0, \textit{otherwise} \end{cases} \tag{2}$$

The distance between the order i and the batch class center based on the information of the goods storage bin is defined as:

After adding order i to batch K, the number of newly added picking bins when picking all order items in batch K. Let O_i be the vector of order i containing the bin where the merchandise is located[9], namely:

$$O_i = (o_{i1}, o_{i2}, \dots, o_{iS}) \tag{3}$$

Among them:

$$o_{is} = \begin{cases} 1, \textit{if the picking order } i \textit{ needs to extract the bin } s \\ 0, \textit{otherwise} \end{cases} \tag{4}$$

R_k refers to the clustering center of batch k based on bin storage information when order i is not added to batch k:

$$R_k = (r_{k1}, r_{k2}, \dots, r_{kS}) \tag{5}$$

among them:

$$r_{ks} = \begin{cases} 1, \textit{if batch } k \textit{ requires extraction bin } s \\ 0, \textit{otherwise} \end{cases} \tag{6}$$

$\hat{R}_k$ indicates that after order i is added to batch k, the class center of batch k based on the information of the merchandise storage bin.

$$\hat{R}_k = (\max(o_{i1}, r_{k1}), \max(o_{i2}, r_{k2}), \dots, \max(o_{iS}, r_{kS})) \tag{7}$$

Let L_{ik} indicates the class center increment of batch k after order I is added, that is:

$$\begin{aligned} L_{ik} &= (l_{ik}^1, l_{ik}^2, \dots, l_{ik}^S) = \hat{R}_k - R_k \\ &= ((\hat{r}_{k1} - r_{k1}), (\hat{r}_{k2} - r_{k2}), \dots, (\hat{r}_{kS} - r_{kS})) \end{aligned} \tag{8}$$

among them:

$$l_{ik}^s = \begin{cases} 1, \textit{if } o_{is} - r_{ks} = 1 \\ 0, \textit{otherwise} \end{cases} \tag{9}$$

Define the distance between order i and batch k based on the information of merchandise storage bins:

$$d_{ik} = \sum_s l_{ik}^s \tag{10}$$

Based on the above symbols and variables: the order batching problem of automatic picking equipment can be expressed as the following model:

Objective function：

$$D = \min \sum_{k=1}^{K} \sum_{s=1}^{S} \hat{r}_{ks} \quad (11)$$

Subject to:

$$\sum_{k=1}^{K} x_{ik} = 1,\ i \in (1, N) \quad (12)$$

$$\sum_{i=1}^{N} x_{ik} \le q,\ k \in (1, K) \quad (13)$$

The objective function (11) represents minimizing the number of bins extracted. Constraint (12) indicates that each order is allocated to exactly one batch; constraint (13) indicates that the number of orders allocated to each batch does not exceed the batch capacity limit.

III. USING IMPROVE K-MEANS ALGORITHM TO SOLVE THE MODEL

The traditional K-Means clustering algorithm is suitable for the sample clustering problem with numerical value[10], but cannot be used to solve the order batch problem directly. Therefore, the K-Means clustering algorithm is improved by combining the classification data features and mathematical model structure: redefining the distance between the class center and the order center. On this basis, a new clustering algorithm is designed to solve the order batching problem. The order picking process is shown in Figure 2. The overall algorithm flow of the experimental process is as follows:

Step1: Read the order data in the order pool, num is the number of orders in the statistical order pool, and determine the batch quantity $K = \lceil num/q \rceil$, $\lceil\rceil$ is the rounding up sign according to the maximum order quantity q accommodated in each batch.

Step2: Calculate the merchandise items contained in each order, and create the storage vector $O_i = (o_{i1}, o_{i2}, \dots, o_{iS})$ of the bin where the order i is based on the storage bin mapped by the merchandise item, and randomly select K orders Store the vectors as initial cluster centers and then delete them from orders that have not yet been batched. The class center of batch k based on commodity storage bin information is R_k, as in formula (5).

Step 3: Calculate the distance d_{ik} from each order to any class center k according to formula (7-10).

Step 4: Select the batch k corresponding to the smallest d_{ik}. If the total number of orders contained in batch k has not reached the maximum number of orders contained in the batch, allocate the order i to be allocated to batch k, and update the clustering center according to formula (7). If the number of orders in batch k reaches the maximum number of batch orders, batch k will be saved. Select the batch h with the minimum distance from the order to be allocated i, allocate the order i to the batch h, and update the clustering center of the batch h. Treat batch h as a new batch k.

Step5: Judge whether there is any unprocessed order, if not, it will end, if there is any, return to Step4.

Step6: After the order is completed in batches, for each batch k=1, 2, 3, ..., K, calculate the final cluster centers respectively, and calculate the number of bins to be extracted according to formula (11).

Step7: Repeat Step2 to Step5, select the batch result of the order with the least number of bins, until the number of bins corresponding to each batch result no longer changes or reaches the maximum value of iteration, output the batch results.

Algorithm program flow chart is shown in Figure 2.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In order to verify the effectiveness of the improved K-Means algorithm based on seed orders, this paper uses real data from a company's warehouse logistics distribution center for experiments. An intelligent picking device in the distribution center that can store 1288 bins and 10 robot carts, and each cart can only transport one bin at a time. During the experiment, 1200 kinds of merchandise were placed in the intelligent sorting equipment, and 1178 bins were stored. The seeding wall can accommodate up to 200 bins to distribute a single order, so the experimental order batches are up to 200 orders. Randomly sample the order data for the three days in the first quarter of 2019 as experimental data. The overall solution is based on python implementation, and the code compilation is implemented in the IDE (integrated development environment) pycharm of python in the win10 operating system.

Based on the first-come-first-served[11] (FCFS) batching strategy, a company's current order batching strategy and the improved K-Means algorithm based on seed orders batch processing calculates the number of times of bin extractions, compares the differences between the three, and verifies the use of the algorithm. Batching can effectively reduce the number of times of bin extractions. The solution is as follows:

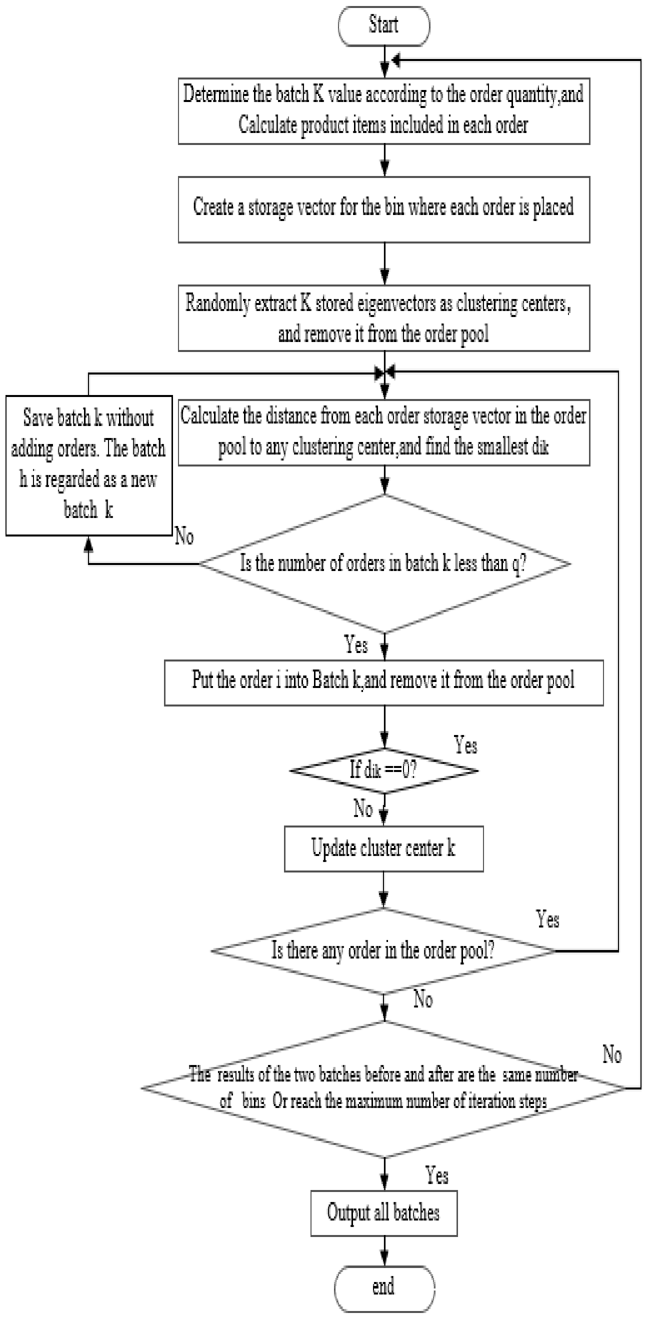

Figure 2. Flow chart of order batching strategy based on improved K-Means clustering algorithm

A. *First-Come First-Served (FCFS) Batching*

The first-come-first-served strategy means that when the order reaches the maximum batch processing capacity of 200, the 200 orders will be processed first, and then continue to wait for the order to arrive and be processed in the same way until the order is completed. Then calculate the bins extracted from each batch, as shown in Table 1.

TABLE 1. FCFS PROCESSING EXPERIMENTAL RESULTS

Order data	Order data I	Order data II	Order data III
Number of orders	3594	3416	2487
Number of batches	18	18	13
Number of extraction bins	1114	1074	946

B. *Company's Current Order Batching Algorithm（CCA）*

A company analyzes the characteristics of order data and finds that there will be single SKU(Stock Keeping Unit) single product, single SKU multiple products, multiple SKUs multiple products in the order, and some orders often appear repeatedly. Therefore, the company adopts the method of sequentially filtering orders in batches. First, the orders of single items are filtered, and then processed first, and then the bulk orders are filtered. Bulk orders refer to orders with the same SKU and the same quantity of goods. If it exceeds a certain amount, then concentrate on processing. Then, the remaining orders are scattered orders, and the degree of how the orders are scattered is high. The order is processed by the hierarchical clustering method until the orders in the order pool are all completed. Finally, the number of times each bin is transported is calculated, as shown in Table 2.

TABLE 2. CCA PROCESSING EXPERIMENTAL RESULTS

Order data	Order data I	Order data II	Order data III
Number of orders	3594	3416	2487
Number of batches	19	18	14
Number of extraction bins	1011	947	837

C. *Improved K-Means Algorithm Based on Seed Orders*

As described in Chapter III, the improved K-Means algorithm based on seed orders is used to process the orders, and the number of bins extracted after completing all orders is calculated according to the formula, as shown in Table 3.

TABLE 3. IMPROVED K-MEANS ALGORITHM PROCESSING EXPERIMENTAL RESULTS

Order data	Order data I	Order data II	Order data III
Number of orders	3594	3416	2487
Number of batches	18	18	13
Number of extraction bins	955	898	809

D. *Comparison of Experimental Results*

It can be seen from Figure 3. that the improved K-Means algorithm can effectively reduce the number of times of bin extractions. In order to better exhibit the comparative experimental results, this paper puts the experimental results of the three order batching strategies in Table 4. It can be seen that the improved K-Means algorithm is optimized by an average of about 15% compared with the First-Come-First-Served picking strategy. Compared with the company's current order batch strategy, the improved K-Means algorithm has an average increase of 4.7%.

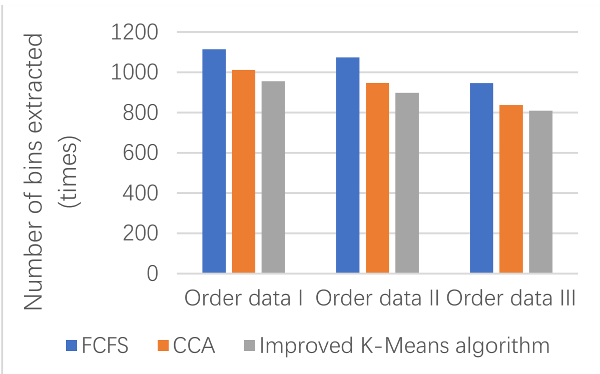

Figure 3. Comparison of the number of bins extracted by three order batching strategies

TABLE 4. COMPARISON OF THREE ORDER BATCHING STRATEGIES

Order data	Order data I	Order data II	Order data III
FCFS(Number of extraction bins)	1114	1074	946
CCA(Number of extraction bins)	1011	947	837
Improved K-Means Algorithm(Number of extraction bins)	**955**	**989**	**809**
Reduction Percentage Compare With FCFS (%)	**14.3%**	**16.4%**	**14.5%**
Reduction Percentage Compare With CCA (%)	**5.6%**	**5.2%**	**3.3%**

V. CONCLUSION

This article takes the automatic picking equipment introduced by a company as the background. Through the research and analysis of the automatic picking system, it is found that the frequency of the bin extraction is the key factor affecting the picking efficiency. Therefore, reducing the extraction frequency of the bin has become a necessary way to improve the system operating efficiency and reduces the equipment input cost. In response to this problem, this article proposes to establish an order batch model based on the information of merchandise storage bin, redefines the clustering center and the distance between the order and the clustering center, takes the distance between the order and the clustering center as the clustering index, and optimizes the goal by increasing order picking efficiency. This article transforms the optimization goal from improving the order picking efficiency to reducing the number of times of bin extractions, using the idea of seed orders to improve the K-Means algorithm and solve the model. Then randomly select the order data of a company in the first quarter of 2019 for experiment. By comparing the first-come-first-served service order batching strategy with a company's current order batching strategy, the K-Means algorithm is improved to deal with order batching. It has a better performance which is about 15% less than the First-Come-First-Served extraction bins, and 4.7% less than the company's current algorithm.

REFERENCES

[1] Y Y Wang, Y H Wu, G H Sun, et al. Research on Picking Order Batching Policy of A Distribution Center [J]. Journal of Shandong University (Engineering Science), 2010, 40(2): 43–46.

[2] De Koster M B M, Van Der Poort E S, Wolters M. Efficient Order Batching Methods in Warehouses[J]. International Journal of Production Research, 1999, 37(7): 1479–1504.

[3] Y Y Wang, X G Meng, Y Y Wang, et al. Order Sequence Optimization for "part-to-picker" Order Picking System [J]. Journal of Mechanical Engineering, 52(4): 212–218.

[4] Z P Li, Y X Tian, X Q Bu, L Y Wu. The Order Batching Problem of Unmanned Warehouse System and the K-max Clustering Algorithm [J/OL]. Computer Integrated Manufacturing Systems, 2020:1-26.

[5] Z Y Shao, B L Dong. Multi-variety Small Batch Order Batching Method Based on Improved Genetic K-Means Algorithm[J]. Journal of Zhejiang Sci-Tech University (Natural Sciences), 2018, 039(006): 732–738.

[6] Hsieh L F, Huang Y C. New Batch Construction Heuristics to Optimise the Performance of Order Picking Systems[J]. International Journal of Production Economics, 2011, 131(2): 618–630.

[7] Pan J C H, Shih P H, Wu M H. Order Batching in A Pick-and-Pass Warehousing System with Group Genetic Algorithm[J]. Omega, 2015, 57: 238–248.

[8] Y Zhang, Y Wu, W Ma. Seed Combine Accompanying Selection Rule of Order-Batching Methods in A Multi-Shuttle Warehouse System[C]//2017 Chinese Automation Congress (CAC). Jinan: IEEE, 2017: 6228–6232.

[9] Y Wang, Z Wang, S Mi. An Order Batching Clustering Algorithm of Fixed Maximum Order Number Based on Order Picking System[C]//2017 4th International Conference on Industrial Economics System and Industrial Security Engineering (IEIS). Kyoto, Japan: IEEE, 2017: 1–6.

[10] J G Sun. Clustering Algorithms Research: Clustering Algorithms

[11] Aboelfotoh A, Singh M, Suer G. Order Batching Optimization for Warehouses with Cluster-Picking[J]. Procedia Manufacturing, 2019, 39: 1464–1473.