

Study on a storage location strategy based on clustering and association algorithms

Li Zhou¹ · Lili Sun¹ · Zhaochan Li¹ · Weipeng Li¹ · Ning Cao²  · Russell Higgs³

Published online: 26 December 2018
© The Author(s) 2018

Abstract

In this paper, we study the improvement of a storage location strategy through the use of big data technology, including data collection, cluster analysis and association analysis, to improve order picking efficiency. A clustering algorithm is used to categorize the types of goods in orders. Classification is performed based on the turnover of goods, value, sales volume, favorable commodity ratings, whether free shipping is provided and whether cash on delivery is supported. An association algorithm is used to determine the relationships among goods by studying the habits of consumers who buy them. A method for improving the class-based storage strategy is proposed. The picking distance of the improved storage strategy is compared with that of the traditional strategy via simulation experiments. The picking efficiency is shown to be enhanced by the improved strategy.

Keywords Cluster · Association rules · Class-based storage · Order picking · Picking efficiency

1 Introduction

As an intermediary step between producers and consumers, warehousing is an important part of logistics systems. The operation of warehouses has long been a focus of industry research. Faced with rapidly growing business needs,

improving storage efficiency at low cost and reducing customer response times have become key issues for improving the operational efficiency of warehouse systems. While guaranteeing quality of service, enterprises must reduce equipment investment and control operational costs to realize expected benefits (Zhou et al. 2018). However, due to rising land prices, increasing rents and rapid industrial development, storage areas that were originally sufficient will gradually become insufficient, requiring either the expansion of one or more existing storage sites or the construction of another. Companies are very cautious about investing in fixed assets and are reluctant to invest heavily. Therefore, enterprises will tend to optimize and improve warehouse management instead of making new investments. In actual production operations, there are many possible methods of reducing expenditure on stocks and improving the operational efficiency of a warehouse. Improving the storage strategy is one such method. Given a fixed amount of warehouse space, optimizing or improving the storage strategy can reduce the cost of goods handling, improve the efficiency of storage and delivery, accelerate the overall operational efficiency of the warehouse, and reduce logistical costs, which can help a company to improve warehouse efficiency while lowering costs.

A storage strategy is a strategy for the placement of goods in a storage center and is one of the key factors influencing

Communicated by B. B. Gupta.

✉ Ning Cao
ning.cao2008@hotmail.com

Li Zhou
zhoulbit@126.com

Lili Sun
sunlibjwy@126.com

Zhaochan Li
zhaochanlibwz@126.com

Weipeng Li
walter1900@126.com

Russell Higgs
russell.higgs@ucd.ie

¹ School of Information, Beijing Wuzi University, Beijing, China

² College of Information Engineering, Qingdao Binhai University, Qingdao, China

³ School of Mathematics and Statistics, University College Dublin, Dublin, Ireland

picking efficiency. At present, three kinds of storage strategies are implemented: location-based storage, class-based storage and random storage. Due to the influence of product variety, the frequencies of loading and unloading, order batches, the number of orders and many other factors, goods allocation in e-commerce storage centers are generally based on the frequencies at which goods come into or are taken out of the warehouse or the values of the goods; that is, a class-based storage strategy is applied. In a class-based storage strategy, all goods are classified according to certain attributes (Liu et al. 2017); different types of goods are allocated to different locations, and goods of the same types are allocated according to certain principles.

Experts and scholars at home and abroad have conducted considerable research on warehouse management. For example, when analyzing warehouse management from the perspective of storage space allocation, it has been found that the ABC class-based strategy offers better efficiency of warehouse operations compared with two other allocation strategies (Hausman et al. 1976). An improved multicriterion ABC class-based strategy has been proposed to classify large inventory items based on both qualitative and quantitative criteria (Hatefi et al. 2014). In class-based storage, goods are usually distributed according to their relevance, liquidity, volume or other characteristics (Yang et al. 2014). The storage area is laid out in accordance with the characteristics of the goods, and the space provided for each type of product must be sufficient for its maximum inventory. Therefore, the average utilization of the storage space is low (Xiao and Zheng 2008). In actual operations, therefore, it is disadvantageous to classify products based on only a single standard. To overcome this disadvantage, many new classification approaches have been proposed. The implementation of ABC classification in inventory management has been improved by considering the contribution of each material to the profit of the company, whether the material is available on the market, and other factors (Hu 2010). The distances between warehousing locations and the distances from the entrance and exit have been considered when applying the ABC class-based method for inventory management (Zhao et al. 2015). A multicriterion ABC inventory management method based on a fuzzy clustering analysis of the materials to be stored has also been proposed (Long 2017).

The most widely used classification method in class-based storage strategies is ABC classification. This method classifies goods according to their values, but it is not perfect since it considers only the values of goods. Ideally, the relationships among different types of goods should also be considered during classification, and the number of attributes considered should also be increased to enable more accurate classification. The relationships among goods, namely, their relevance, should be assessed from three main perspectives. (1) Association, that is, similarity or complementarity among products. If

the combination of two or more products can produce greater value or better satisfy real needs than can any of the products alone, then a demand for one of those products will inevitably lead to a demand for the other(s). (2) Bill of materials (BoM). When a warehouse manager must pick materials in accordance with the BoM for a certain production plan, a stable relevance relationship can be found in accordance with the BoM structure (Teng and Ma 2016). (3) Common needs, that is, the tendency for certain groups of products to be ordered together, which can be identified through the analysis of a large quantity of order data. Many scholars assign corresponding positions based on certain correlations between goods. The distribution of goods based on product BOMs has been studied by using mathematical methods to cluster materials and assign them to their associated goods (Wang 2016). Based on material frequencies and material correlations, a mathematical model for the optimization of location allocation has been established (Jin 2014).

In the era of big data, various novel information technologies are continuously being introduced, and some of them, such as the Internet of Things, cloud computing, and big data, have been widely implemented. Some scholars have studied the application of big data technology for inventory management. A robust injection-point-based framework has been proposed for resolving cross-site scripting (XSS) vulnerabilities in online social networks (OSNs) to address security issues (Gupta and Gupta 2017). A new Arabic text indexing approach has been proposed to improve cloud e-learning systems for the Arabic language (Haffar et al. 2017). The use of XML to store and exchange data has been studied, and decision-making based on OLAP cubes in a cloud environment has been analyzed in pursuit of the Data Warehouse as a Service (DWaaS) concept (Dkaich et al. 2017).

Various big data algorithms have been used in many diverse fields. A new dynamic firefly algorithm has been proposed for estimating the demand for water resources (Wang et al. 2018). The random forest algorithm has been applied to analyze big data collected from insurance companies (Lin et al. 2017a, b, c). A graph mining algorithm has been used to extract process patterns to automate business process consolidation (Huang et al. 2016). For the analysis of resource scheduling in cloud computing, a virtual machine placement algorithm based on peak workload characteristics can be used (Lin et al. 2017a, b, c). Multiresource scheduling and power models have been used to extend CloudSim, which is one of the most popular and powerful simulation platforms in cloud computing (Lin et al. 2017a, b, c). Other scholars have studied and analyzed various methods used in big data technology. A distance-metric-optimization-driven learning approach has been proposed to solve the problem of facial recognition across ages by integrating the traditional steps of the process with a deep convolutional neural network (Li et al. 2018). The grammatical evolution of algorithms and

programs has been analyzed (He et al. 2016, 2017). The maximum duo-preservation string mapping problem and the minimum integral solution problem, which may be encountered in programming, have also been analyzed (Chen et al. 2013, 2014). Scholars have analyzed issues related to network security and capacity and have proposed a secure and flexible EHR sharing scheme suitable for mobile health clouds (Cai et al. 2017a, b). A community weakening control strategy (CWCS) has been proposed for the enhancement of network capacity (Cai et al. 2017a, b). Other scholars have used data mining methods, such as clustering and association algorithms, to analyze historical product transaction records to obtain optimized plans for warehouse allocation (Chen 2015).

Using big data technology, many scholars have solved various network and data problems. Several scholars have studied ABC class-based storage strategies and location optimization problems from different perspectives. Some researchers have studied the application of clustering and association algorithms to optimize warehousing configurations. However, the utilization of big data technology to improve storage strategies and optimize goods allocation in warehouses remains rare. Since it is difficult to directly obtain internal data from an enterprise, previous studies have tended to rely on random number generation to simulate the generation of orders, which makes it difficult to ensure the accuracy and authenticity of the order data. In this paper, big data crawling technology is used to obtain real orders directly to avoid this problem. Therefore, substantial room for further research remains. This paper focuses on using big data technology to directly obtain real order data from an enterprise, increase the number of dimensions considered during classification, and improve the ABC class-based method to more accurately classify goods. Clustering and association algorithms are used to optimize the traditional ABC class-based storage strategy. Compared with the traditional strategy, the improved method considers classification based on comprehensive features instead of only a single level of classification, and this improvement enables higher classification precision.

The web crawler technology, clustering and association algorithms and other big data technologies used in this paper have enabled remarkable achievements in retail, fast-moving e-commerce, finance, search engines, smart recommendations and other fields. These technologies can also be used to improve class-based storage in warehouses by optimizing the storage locations, thereby improving the overall storage efficiency. In view of these goals, big data technology is applied to improve the storage strategy applied in a warehouse. First, web crawler technology is used to acquire order data, and data processing, such as data cleaning and data conversion, is performed. Second, data mining technology is used to cluster the ordered goods and perform association rule analysis

to improve the ABC class-based storage strategy to optimize the storage locations. Finally, the results obtained with and without the proposed improvement to the storage strategy are compared through simulations. The big data approach is used to analyze historical orders to derive an empirical classification of the materials in the warehouse. This process enables more precise order classification and makes it possible for distribution to continue at normal efficiency when a network attack occurs. This paper highlights the unique advantages of big data technology in order analysis and introduces new avenues for research on warehouse picking efficiency.

2 Collection and preprocessing of order data

2.1 Collection of order data

To study the relevance of goods, we need to start from order data. Web crawling is adopted to obtain order data. A web crawler must be provided with two pieces of information: the URL to be crawled, that is, the URL of the data, and the target data to be crawled. The target URL is the URL of a large-scale e-commerce enterprise in China, and the data to be crawled are data on daily commodities and order data stored on the enterprise's website. The product information includes eight variables: the product ID, the price, the number of favorable comments about the product (positive feedback), the number of neutral comments about the product (neutral feedback), the number of negative comments about the product (negative feedback), whether free shipping is provided for the package, whether cash on delivery is supported for the package, and sales. The order data include order release times and consumer location information.

Since the objects to be crawled are the information and order data for daily commodities, we first enter the name of such a commodity (we consider 113 daily necessities, such as glasses, shampoo and paper towels) into the website to obtain the corresponding goods ID. Second, we use the acquired product ID to construct the URL to be crawled. During the construction of the target URL, a bug may arise that may lead to termination of the crawler or may cause the target data to not be accurately crawled.

Therefore, a fault-tolerance mechanism is implemented to ensure that if a program error occurs during the crawling process, the page that triggers the bug will be skipped to ensure that the program runs smoothly. The content of the webpage source code is loaded into memory, and the target data to be fetched are extracted. Finally, the fetched target data are stored in a database for subsequent analysis.

Following this data capturing process, we obtained a total of 1,313,025 pieces of data on 109 daily necessities, from which we extracted and sorted data on 75,000 single-product orders placed in North China in September. The pick list is

ordered over a certain period of time and covers all the sorted products.

2.2 Preprocessing of order data

More than 1.3 million pieces of data were collected for this study. Using the time and province of each order and excluding invalid data, only product data and order data from North China during September were retained, resulting in 75,000 eligible items. For storage, each piece of data is encapsulated in a JSON file. JSON is a lightweight data exchange format that is mainly expressed in the form of key-value pairs. An example of a JSON data entry is {"generalCount": 390, "orderTime": "2016-09-16 21:39:18", "goodCount": 759, "userAddress": "Beijing", "poorCount": 167, "productID": 767296, "productName": "electric-lunchbox", "price": 89, "pay": 1, "express": }, where the key appears to the left side of the colon in each pair and the corresponding value appears to the right side of the colon. Here, productID is the ID number of the product; productName is the name of the product; orderTime is the order generation time; goodCount is the number of high ratings for the product; generalCount is the number of moderate ratings for the product; poorCount is the number of poor ratings for the product; price is the price of the product; pay represents whether cash on delivery is supported, with a value of 1 indicating that it is supported and a value of 0 indicating that it is not; and express represents whether shipping is free, with a value of 1 indicating free shipping and a value of 0 indicating no free shipping.

Data expressed in the format of the above example can be transformed as follows. We replace the numbers of favorable and unfavorable comments on a product with a single favorable rate to reflect the quality of the product. The favorable rate is defined as the proportion of favorable comments relative to the total number of favorable, neutral and unfavorable comments. Thus, the example given above can be simplified to {"orderTime": "2016-09-16 21:39:18", "goodRate": 58%, "userAddress": "Beijing", "poorCount": 167, "productID": 767296, "productName": "electric-lunchbox", "price": 89, "pay": 1, "express": 0}.

3 Clustering and association analysis of the ordered goods

3.1 Clustering analysis of the ordered goods

Many brands of goods of the same type are available on the e-commerce website considered in this study. The characteristics of similar products of different brands are not the same. These characteristics include the price, sales volume, favorable rate, support for cash on delivery, and availability of free shipping. We collected data on 109 types of daily necessities,

each of which is available from up to 30 different brands. Therefore, if these products were to be treated separately, it is possible that different products of the same type could be clustered into different categories, resulting in goods of the same type being placed in different regions of the warehouse, which is not consistent with the desire to concentrate similar goods in the same area. Therefore, the data for each type of product must be homogenized (Rigutini and Maggini 2005).

Through the homogenization process, we compiled four representative numerical values, in the form of either means or probabilities of occurrence, to represent the overall characteristics of similar products of up to 30 different brands. These four numerical values represent the price of these products, the favorable rate, whether cash on delivery is supported, and whether free shipping is provided, and the total number of sales of all products of the same type was also recorded. An example of a piece of homogenized data is provided as follows: {"meangoodRate": 96%, "totalsales": 167, "productName": "Sports-Bottle", "meanprice": 59.8, "payprobability": 50%, "expressprobability": 14%}. Among 30 different brands of sports water bottles, a total of 167 were sold within the month considered in this study. The average price of these 30 brands of sports water bottles was 59.8 yuan. The average favorable rate was 96%, 50% of the 30 brands supported cash on delivery, and free shipping service was provided for 14% of the 30 brands.

In the traditional class-based storage strategy, especially in a production-based warehouse center, products are often classified based on value, volume, turnover, out-of-stock costs and other factors (Li 2016). In the e-commerce model, the most direct information about a product can be obtained from the e-commerce website itself, and the profit-making means provided by businesses to attract consumers to buy goods can also be determined. As the basis for the classification, in addition to considering the traditional attributes of goods, we consider three additional attributes: the favorable rate, support for cash on delivery, and availability of free shipping.

The goods studied in this article are necessities for daily life. The products are small in size, so ease of delivery is not a consideration. Moreover, many alternative brands of goods are available on the e-commerce site to satisfy consumers' demand. Thus, a shortage of goods will not lead to the loss of a large number of customers, meaning that out-of-stock costs can be ignored.

In the general case, a bowl-shaped relationship exists between the time required to access the goods in the storage system and the number of classification categories (Yu et al. 2015). It has been theoretically revealed that regardless of the quantity of goods, the optimal storage efficiency can always be achieved by classifying the goods into 3-5 categories. Therefore, k is set to three in the k -means classification algorithm, meaning that the commodities are divided into three categories.

Table 1 Clustering results based on the traditional classification

Clustering results	Category 1	Category2	Category 3
Category	B	A	C
Quantity	15	12	85
<i>Cluster center</i>			
Sales volume	-0.100210	2.483434	-0.332918
Price	2.075549	-0.496433	-0.296189

For the purposes of this research, the results of dividing the 109 types of commodities using two different methods will be compared. The first method is traditional classification based on clustering analysis using the attributes of price and sales volume to classify the commodities into three categories. In the second method, the clustering-based classification is improved by considering three additional attributes: the favorable rate, whether cash on delivery is supported, and whether free shipping is provided.

The goal of clustering analysis is to identify categories of the 109 types of products with similar attributes to determine which types of products are selling well and to attract more online consumer purchases. Moreover, products that are most frequently purchased should be placed at the shortest picking distance. Therefore, a combination of qualitative and quantitative methods is needed to analyze the clustering results. The clustering centers identified through *k*-means clustering can provide an understanding of the characteristics of each type of product. The Z-score standardization of the numerical values of each attribute allows the specific meaning of each class to be analyzed by observing the value of each variable associated with each cluster center (Scott and Longuet-Higgins 1990). We have observed that the variables associated with a cluster center may have both positive and negative values. If the value of a variable is less than 0 for a cluster center, then the representative value of that variable for that cluster is less than the average value for the population as a whole (Kannan et al. 2000).

Table 1 shows the goods clustering results according to the traditional classification based on sales volume and price. The product types are aggregated into three categories, with a total of 15 product types in the first category (category 1), 12 product types in the second (category 2) and 85 product types in the third (category 3). The sales volume values associated with the cluster centers of these three categories are -0.100210, 2.483434 and -0.332918. In terms of these values, the categories are ordered as follows: category 2 > 0 > category 1 > category 3. The values for categories 1 and 3 are both less than 0, indicating that the sales in categories 1 and 3 are lower than the overall average; in addition, the absolute value of the sales volume in category 3 is greater than the absolute value of the sales volume in category 1, indicating

that the sales in category 3 are the smallest. The price values associated with the cluster centers of the three categories are 2.075549, -0.496433, and -0.296189. In terms of these values, the categories are ordered as follows: category 1 > 0 > category 3 > category 2. The values for categories 2 and 3 are both less than 0, indicating that the price levels of categories 2 and 3 are lower than the overall average, and the absolute value of the price variable in category 2 is greater than the absolute value of the price variable in category 3, indicating that the products in category 2 are the cheapest. In summary, the 15 product types in category 1 account for 14% of the total goods. The prices of these goods are the highest, and the sales are moderate. There are 12 types of products in category 2, accounting for 11% of the total goods. These products are the cheapest and have the highest sales volume. There are 85 types of products in category 3, accounting for 75% of the total goods. These products constitute more than 50% of the total. These items are cheap but have the lowest sales volume.

According to the above analysis of the cluster centers, the goods in category 1 on the e-commerce website are the most popular consumer goods, and their prices are reasonable. Category 1 is followed by category 2 in terms of popularity, and the least popular goods are those in category 3. Therefore, in the terminology of ABC classification, category 1 corresponds to category B, category 2 corresponds to category A, and category 3 corresponds to category C. Table 2 shows the specific product classification results.

Table 3 shows the clustering results based on the new classification. These results are based on five attributes: sales volume, price, favorable rate, the probability of free shipping availability and the probability of support for cash on delivery.

The products are aggregated into three categories: categories 1, 2 and 3. There are 13 product types in category 1, 35 product types in category 2 and 61 product types in category 3. Only the values of the price and free shipping variables are less than 0 for the cluster center of category 1. In other words, the price and probability of free shipping availability in this category are lower than the overall averages. For the cluster center of category 2, among the five variables, only the sales volume is less than 0; thus, the sales volume of goods of this type is less than the overall average. By contrast, the variable values for the cluster center of category 3 are all negative, indicating that these commodities have lower values than the overall averages in terms of all considered attributes. In summary, the 13 types of products in category 1 account for 12% of the total goods. The sales volume, favorable rate, and probability of support for cash on delivery are the highest in this category. Since the e-commerce website requires a price greater than 99 yuan to provide free shipping and the value of the price variable associated with this cluster center is the smallest, goods in

Table 2 Product classification results 1

Goods	Type	Goods	Type	Goods	Type
Cooking appliances	B	Pillows	C	Teapots	C
Juicers	B	Sheets	C	Paper towels	A
Rice cookers	B	Blankets	C	Laundry products	A
Electric pressure cookers	B	Mattresses	C	Cleaning supplies	C
Soy milk	B	Mosquito netting	C	Deworming supplies	C
Coffee machines	B	Cushions	C	Family cleaning supplies	A
Microwave ovens	B	Bath towels	C	Leather care	C
Electric ovens	B	Electric blankets	C	Hand and foot skin care	C
Induction cookers	B	Curtains/screens	C	Weight loss products	C
Bread machines	B	Home decoration fabrics	C	Body care	C
Egg boilers	C	Summer sleeping mats	C	Nursing clothing	C
Yogurt machines	C	Heat protection	C	Massage oil	C
Electric cookers	C	Storage supplies	C	Hairdressing tools	C
Thermoses	C	Umbrellas and rain gear	C	Hair dye	C
Electric baking pans	C	Bathroom supplies	C	Aromatherapy oil	C
Multipurpose pots	C	Knitting supplies	C	Bath salts	C
Electric grills	C	Laundry/ironing supplies	C	Handmade soap	C
Fruit and vegetable processing machines	C	Decontamination products	C	Shampoo	A
Health pots	C	Kitchen knives	C	Conditioner	A
Electric lunchboxes	C	Scissors	C	Hair coloring	C
Shavers	C	Tool sets	C	Scrubs	A
Shavers/epilators	C	Cutting boards	C	Soap	A
Oral care	A	Fruit knives/planes	C	Toothpaste/dental powder	C
Hair dryers	C	Multifunction knives	C	Toothbrushes/floss	C
Beauty equipment	B	Plastic cups	C	Mouthwash	A
Barbering supplies	A	Sports bottles	C	Nursing supplies	A
Volumizer/hair straightener	C	Glasses	C	Dusters	C
Massagers	B	Porcelain cups	C	Ottomans	C
Foot tubs	C	Insulated cups	C	Seat cushions	B
Sphygmomanometers	B	Insulated pots	C	Seat covers	B
Electronic scales	C	Wine glasses/wine	C	Headrests and lumbar support	A
Blood glucose meters	A	Bowls/plates	C		
Thermometers	C	Chopsticks/spoons, knives and forks	C		
Tablecloths	C	Fruit baskets	C		
Carpets and floor mats	C	Cups	C		
Sofa cushion covers	C	Teapots	C		
Bedding kits	C	Tea trays	C		
Quilts	C	Earmuffs	C		

this category also have the lowest probability of free shipping availability. There are 35 types of products in category 2, accounting for 31% of the total goods. Only the sales volume value is less than 0, and this value lies between those of the other two categories, indicating that the overall sales of these products are less than those of category 1 but more than those of category 3. The values of the remaining four variables are all greater than 0. The absolute values of the

price and free shipping variables are the largest, indicating that the overall price of these products is the highest. Because of these high prices, the probability of free shipping availability is also the highest. There are 61 types of products in category 3, accounting for 54% of the total. The five variable values associated with the cluster center are all less than 0, indicating that although these goods have the lowest overall prices, their sales volume is also the lowest.

Table 3 Clustering results based on the new classification

Clustering results	Category 1	Category2	Category 3
Category	A	B	C
Quantity	13	37	59
Sales volume	2.228878	-0.233553	-0.341001
<i>Cluster center</i>			
Price	-0.514112	1.035768	-0.484728
Favorable rate	1.077490	0.214221	-0.352542
Free shipping	-0.717296	1.234126	-0.555239
Cash on delivery	1.262652	0.227063	-0.399372

According to the above analysis, the goods in category 1 on the e-commerce website are the most popular consumer goods. The price of these goods is the cheapest, their quality is good, and the probability of support for cash on delivery is the highest. This category is followed by the goods in category 2 in terms of popularity, and the least popular goods are those in category 3. Therefore, based on the analysis of the cluster centers, category 1 corresponds to category A, category 2 corresponds to category B, and category 3 corresponds to category C. Table 4 lists the specific product classification results.

3.2 Association analysis of the ordered goods

The relationships among the various products under study can be described in terms of association rules. We study the e-commerce order data to determine whether associations exist among the various products purchased by consumers. Due to e-commerce websites' strong awareness of the need to protect consumer information, we cannot obtain complete historical product purchasing data through web crawling. By assuming that all collected orders are fulfilled from the same storage center, we can replace the order purchasing data with order picking data.

Because the data were collected based on purchasing information at a single point in time, the data can be organized by date, and the daily order data can be sorted in their natural chronological order. This process makes it easy to determine the order in which each item was purchased on a given day. Then, the association rules among different purchased goods can be determined by means of an association algorithm. Figure 1 illustrates the data preparation process.

For simplicity in subsequent research, we replaced the product type names with digital codes, each consisting of the letter G plus a number. The codes are listed in Table 5.

Each ordered picking list can be regarded as a shopping basket. As shown in Table 6, under the assumption that the picking vehicles in the warehouse can accommodate 20 items

at a time on average, the order data for all single items ordered per day are divided into picking lists of 20 items.

In this paper, we use the Apriori algorithm to analyze the association rules relating the commodities contained in the picking lists for the daily commodity storage area in an e-commerce warehouse to determine the implicit links among various types of commodities. In the Apriori algorithm, the quantity of data is large, the number of iterations is large, and the operation time is long and inefficient. After many tests, the parameter settings of the algorithm were limited to achieve more efficient operation. The support for frequent 1-item sets was set to be greater than 50, and the support for other frequent item sets containing more than 2 items was set to be greater than or equal to 2.

Table 7 shows the partial results for frequent 2-item sets and larger frequent item sets that satisfy the set parameters. Frequent 2-item sets represent sets of 2 related products. Frequent 3-item sets represent sets of 3 related products, and so on. The support value represents the number of orders of goods that satisfy the association rule.

The calculation results show that there are sets of at most five frequently co-occurring items that satisfy the support condition, but the support for each rule is very low. Frequent 2-item sets are relevant to the purpose of this research, that is, finding a classification storage strategy that optimizes the placement of goods in an e-commerce storage center. Identifying strongly related items and placing them in neighboring locations will inevitably decrease the distance traveled when picking goods, thereby improving the picking efficiency and reducing the logistical costs.

The table also shows that frequent 3-item sets and larger frequent item sets relate many products, but their support is not high. Thus, these item sets are of little significance from the perspective of the specific placement of goods in a warehouse. Therefore, this paper focuses on frequent 2-item sets when investigating how to set appropriate support thresholds and confidence thresholds for identifying strongly associated products.

Table 8 presents examples of the results for the confidence levels of frequent 2-item sets. Two goods are considered to be strongly related when both the degree of support and the confidence level are greater than their associated thresholds.

In Table 8, the first column of data shows the identified frequent 2-item sets, where X is the first good in the set and Y is the second. The data in the second column are the support values of the frequent 2-item sets: a higher support value indicates a higher probability of two products appearing in adjacent positions in a picking list. The third column is the number of occurrences in the original data of the first commodity, X, in the frequent 2-item set. The fourth column is the conditional probability that commodity Y appears given that commodity X appears, that is, the confidence. The fifth column is the number of occurrences in the original data of

Table 4 Product classification results 2

Goods	Type	Goods	Type	Goods	Type
Cooking appliances	B	Pillows	B	Teapots	C
Juicers	B	Sheets	B	Paper towels	A
Rice cookers	B	Blankets	C	Laundry products	A
Electric pressure cookers	B	Mattresses	B	Cleaning supplies	C
Soy milk	B	Mosquito netting	C	Deworming supplies	C
Coffee machines	B	Cushions	C	Family cleaning supplies	A
Microwave ovens	B	Bath towels	C	Leather care	C
Electric ovens	B	Electric blankets	C	Hand and foot skin care	C
Induction cookers	B	Curtains/screens	C	Weight loss products	C
Bread machines	B	Home decoration fabrics	C	Body care	C
Egg boilers	C	Summer sleeping mats	C	Nursing clothing	C
Yogurt machines	B	Heat protection	C	Massage oil	C
Electric cookers	B	Storage supplies	C	Hairdressing tools	A
Thermoses	B	Umbrellas and rain gear	C	Hair dye	C
Electric baking pans	B	Bathroom supplies	C	Aromatherapy oil	C
Multipurpose pots	B	Knitting supplies	C	Bath salts	C
Electric grills	B	Laundry/ironing supplies	C	Handmade soap	C
Fruit and vegetable processing machines	B	Decontamination products	C	Shampoo	A
Health pots	B	Kitchen knives	C	Conditioner	A
Electric lunchboxes	B	Scissors	C	Hair coloring	C
Shavers	C	Tool sets	B	Scrubs	A
Shavers/epilators	B	Cutting boards	C	Soap	A
Oral care	A	Fruit knives	C	Toothpaste/dental powder	C
Hair dryers	A	Multifunction knives	C	Toothbrushes/floss	C
Beauty equipment	B	Plastic cups	C	Mouthwash	A
Barbering supplies	B	Sports bottles	C	Nursing supplies	A
Volumizer/hair straightener	B	Glasses	C	Dusters	C
Massagers	B	Porcelain cups	C	Ottomans	B
Foot tubs	B	Insulated cups	C	Seat cushions	B
Sphygmomanometers	B	Insulated pots	B	Seat covers	B
Electronic scales	C	Wine glasses/wine	B	Headrests and lumbar support	C
Blood glucose meters	A	Bowls/plates	B		
Thermometers	C	Chopsticks/spoons, knives and forks	C		
Tablecloths	C	Fruit baskets	C		
Carpets and floor mats	C	Cups	C		
Sofa cushion covers	C	Teapots	C		
Bedding kits	B	Tea trays	C		
Quilts	B	Earmuffs	C		

Fig. 1 The process of correlation analysis for data conversion

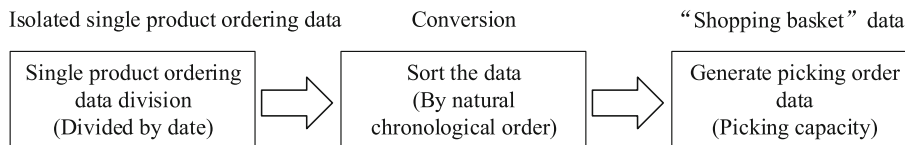


Table 5 Product code

Goods	Code	Goods	Code	Goods	Code
Cooking appliances	G1	Pillows	G39	Teapots	G40
Juicers	G2	Sheets	G40	Paper towels	G41
Rice cookers	G3	Blankets	G41	Laundry products	G42
Electric pressure cookers	G4	Mattresses	G42	Cleaning supplies	G43
Soy milk	G5	Mosquito netting	G43	Deworming supplies	G44
Coffee machines	G6	Cushions	G44	Family cleaning supplies	G40
Microwave ovens	G7	Bath towels	G45	Leather care	G41
Electric ovens	G8	Electric blankets	G46	Hand and foot skin care	G42
Induction cookers	G9	Curtains/screens	G47	Weight loss products	G43
Bread machines	G10	Home decoration fabrics	G48	Body care	G44
Egg boilers	G11	Summer sleeping mats	G49	Nursing clothing	G40
Yogurt machines	G12	Heat protection	G50	Massage oil	G41
Electric cookers	G13	Storage supplies	G51	Hairdressing tools	G42
Thermoses	G14	Umbrellas and rain gear	G52	Hair dye	G43
Electric baking pans	G15	Bathroom supplies	G53	Aromatherapy oil	G44
Multipurpose pots	G16	Knitting supplies	G54	Bath salts	G40
Electric grills	G17	Laundry/ironing supplies	G55	Handmade soap	G41
Fruit and vegetable processing machines	G18	Decontamination products	G56	Shampoo	G42
Health pots	G19	Kitchen knives	G57	Conditioner	G43
Electric lunchboxes	G20	Scissors	G58	Hair coloring	G44
Shavers	G21	Tool sets	G59	Scrubs	G40
Shavers/epilators	G22	Cutting boards	G60	Soap	G41
Oral care	G23	Fruit knives	G61	Toothpaste/dental powder	G42
Hair dryers	G24	Multifunction knives	G62	Toothbrushes/floss	G40
Beauty equipment	G25	Plastic cups	G63	Mouthwash	G41
Barbering supplies	G26	Sports bottles	G64	Nursing supplies	G42
Volumizer/hair straightener	G27	Glasses	G65	Dusters	G43
Massagers	G28	Porcelain cups	G66	Ottomans	G44
Foot tubs	G29	Insulated cups	G67	Seat cushions	G40
Sphygmomanometers	G30	Insulated pots	G68	Seat covers	G41
Electronic scales	G31	Wine glasses/wine	G69	Headrests and lumbar support	G42
Blood glucose meters	G32	Bowls/plates	G70		
Thermometers	G33	Chopsticks/spoons, knives and forks	G71		
Tablecloths	G34	Fruit baskets	G72		
Carpets and floor mats	G35	Cups	G73		
Sofa cushion covers	G36	Teapots	G74		
Bedding kits	G37	Tea trays	G75		
Quilts	G38	Earmuffs	G76		

the second commodity, Y. The sixth column is the conditional probability that commodity X appears given that commodity Y appears, which is called the reverse confidence. The seventh column is the greater of the two values in the fourth and sixth columns.

In this table, the support value of a frequent item set represents the strength of the correlation between the two commodities, while the confidence represents the degree of

confidence in the correlation. The analysis of these association rules can be used as the basis for the further optimization of product placement.

Goods are considered to be strongly correlated if the support for the corresponding frequent item set is 100 or more and the confidence is greater than 20% (Feng et al. 2015). Based on the above criteria, the most strongly correlated

Table 6 Example of a picking list

Product name	Product code
Thermos	G14
Mattress	G42
Cushion	G44
Bath towel	G45
Home decoration fabric	G48
Decontamination product	G56
Glass	G65
Chopsticks/spoon, knife and fork	G71
Teapot	G77
Cleaning supplies	G80
Electric cooker	G13
Knitting supplies	G54
Electric grill	G17
Bedding kit	G37
Cup	G73
Handmade soap	G93
Insulated cup	G67
Pillow	G39
Teapot	G101

product sets can be identified in order to develop a class-based storage strategy. The results are shown in Table 9.

4 Improved warehousing strategy

4.1 Goods placement principle in the traditional class-based storage strategy

Most class-based storage strategies are based on ABC classification. This classification method is based mainly on the turnover rate of goods, and the goods are divided into three categories: A, B and C. The cargo turnover is highest in category A, followed by category B and then C. Goods with high turnover are placed, to the greatest extent possible, near the warehouse I/O point. ABC classification also considers the ease of handling and the value of goods: goods that are difficult to handle or are high in value are generally classified as belonging to category A, followed by B and C (Zhu 2017).

The number of goods in category A generally constitutes 5 to 15% of the total number of goods, but these goods represent 60 to 80% of the total value. The goods in category B constitute 15 to 25% of the total number and represent 15 to 25% of the total value. Finally, the goods in category C constitute 60 to 80% of the total number but represent only 5 to 15% of the total value (Luo and Ye 2017).

Let D denote the distance from the warehouse I/O point to a single cargo space. The distance from the I/O point to

the picking location is calculated for each cargo space, and all obtained D values are sorted from smallest to largest. Let a denote the maximum D value among the top 15% of the values, and let b denote the maximum D value among the middle 25%.

In Fig. 2, a semicircle of radius a is drawn with its center at the I/O point; the cargo in category A is located within this semicircle. Similarly, the cargo in category B is located in the area between this semicircle and a concentric semicircle of radius b . Finally, cargo in category C is located in the remaining cargo space.

4.2 Improved class-based storage strategy

The traditional ABC class-based method is based on the out-of-stock frequencies, values and sales volumes of the goods to be classified. However, for goods offered on e-commerce websites, the characteristics of the goods themselves can be combined with the characteristics of the online shopping process; therefore, it is appropriate to consider additional attributes when dividing goods, such as product reviews, payment methods and whether the merchant provides free shipping.

This improved process can result in more accurate classification based on the popularity of the goods. The most popular goods with the highest sales volumes are assigned to category A, goods with moderate popularity and sales belong to category B, and the least popular goods with the lowest sales volumes are in category C. This improvement can be achieved using clustering analysis.

Although the goods are more precisely classified with this process, disadvantages can also arise since separating goods solely by popularity and sales volume disregards the possible internal relations among goods of different grades. E-commerce consumers typically do not purchase only one item at a time. To reduce delivery costs, consumers will purchase multiple items at once from their 'shopping carts'. Merchants on e-commerce websites often offer promotions to increase sales, for example, 'bundling' a product with another product at a special discount or accumulating a commodity with a price exceeding a certain amount. Merchants may also provide discounted or free shipping. Therefore, consumers do not expect the purchasing of different goods to be unrelated; there are certain inherent relationships. A correlation analysis of large amounts of data can reveal some of these inherent relationships, which can then be used to further optimize the classification used in storage strategies. A reasonable arrangement of the cargo space will increase the picking efficiency in an e-commerce storage center, accelerate the retrieval of goods from storage, reduce customer wait times and enhance customer satisfaction.

Any improvements to a class-based storage strategy need to be based on actual data to ensure that the optimized lay-

Table 7 Results of the association analysis

Frequent 2-item sets	Support	Frequent 3-item sets	Support	Frequent 4-item sets	Support	Frequent 5-item sets	Support
G32-G19	342	G65-G43-G15	231	G65-G43-G15-G1	89	G65-G43-G15-G1-G100	25
G78-G81	316	G78-G81-G80	209	G78-G81-G80-G79	74	G78-G81-G80-G79-G74	19
G51-G56	298	G51-G56-G53	187	G51-G56-G53-G15	63	G51-G56-G53-G15-G96	19
G98-G45	273	G98-G45-G61	176	G98-G45-G61-G34	54	G98-G45-G61-G34-G10	12
G23-G21	254	G23-G21-G57	164	G23-G21-G57-G22	46	G23-G21-G57-G22-G83	11
G82-G103	231	G82-G103-G53	151	G82-G103-G53-G43	41	G82-G103-G53-G43-G16	11
G24-G95	224	G24-G95-G35	143	G24-G95-G35-G51	32	G24-G95-G35-G51-G69	8
G24-G95	197	G24-G95-G16	124	G24-G95-G16-G43	21	G24-G95-G16-G43-G102	8
G101-G65	167	G101-G65-G80	112	G101-G65-G80-G15	16	G101-G65-G80-G15-G24	5
...

Table 8 Examples of the confidence of frequent 2-item sets

Frequent 2-item sets	Support	Frequency of X occurrence	Confidence, P(X Y) (%)	Frequency of Y occurrence	Frequency of Y occurrence (%)	Higher degree of confidence (%)
G32-G19	342	628	54.46	723	47.30	54.46
G78-G81	316	908	34.80	810	39.01	39.01
G51-G56	298	1123	26.54	794	37.53	37.53
G98-G45	273	987	27.66	881	30.99	30.99
G23-G21	254	782	32.48	890	28.54	32.48
G82-G103	231	617	37.44	574	40.24	40.24
G24-G95	224	792	28.28	681	32.89	32.89
G24-G95	197	510	38.63	656	30.03	38.63
G101-G65	182	623	29.21	1092	16.67	29.21
G14-G50	167	962	17.36	642	26.01	26.01
G57-G60	163	491	33.20	673	24.22	33.20
G68-G33	152	552	27.54	527	28.84	28.84
G49-G43	140	709	19.75	792	17.68	19.75
...

Table 9 Strongly correlated product types

Related product code-1	Related product code-2	Support	Confidence (%)
G32	G19	342	54.46
G78	G81	316	39.01
G51	G56	298	37.53
G98	G45	273	30.99
G23	G21	254	32.48
G82	G103	231	40.24
G24	G95	224	32.89
G101	G65	197	38.63
G14	G50	182	29.21
G57	G60	167	26.01
G68	G33	163	33.20
G73	G75	130	24.57
G79	G74	109	33.18

Fig. 2 Example of cargo location layout in ABC class-based storage

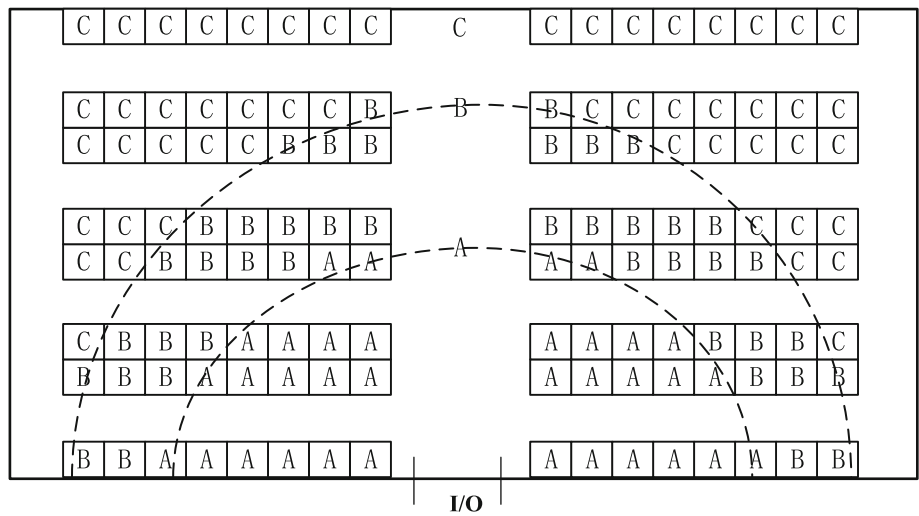
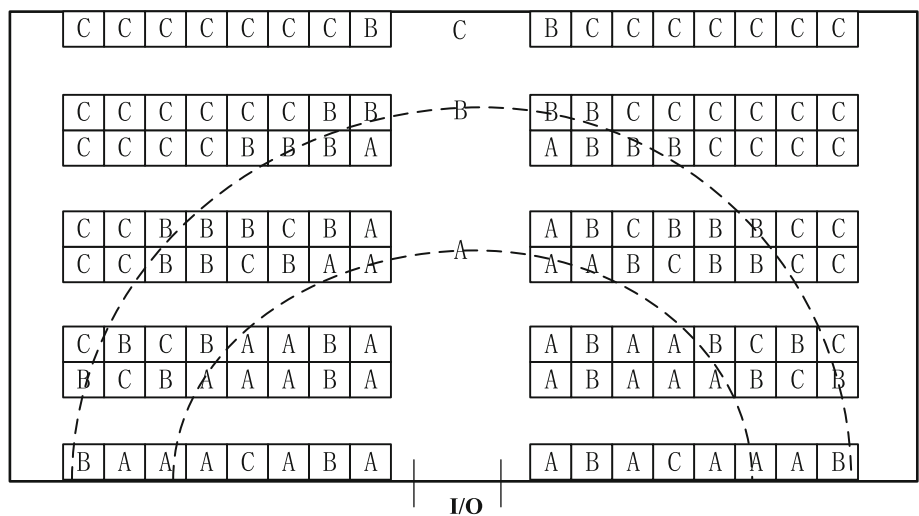


Fig. 3 Improved cargo layout example



out of the goods is realistic. As shown in Fig. 3, cargo in categories B and C may also be placed in the storage space belonging primarily to goods of category A when there is a strong correlation between these category B or C goods and category A goods. Similarly, the space belonging primarily to category B goods may also contain associated category C goods.

The specific rules adopted in the improved ABC class-based storage strategy to aid in cargo placement are as follows.

As the first step, cluster analysis is performed to divide the goods into three categories, namely, categories A, B and C, using the numerical proportions of the various types of goods. The radii a and b are then calculated to divide the storage space into areas associated with the three categories of goods, thereby obtaining the specific coordinates encoding the locations of each type of cargo.

The second step is to sort the category A goods in order of increasing sales.

In the third step, the category A goods are placed in accordance with the order of preparation of the corresponding cargo. Priority is given to goods placed to the right of the main aisle based on the ordering from the second step. There are ten product codes in category A: 21, 13, 16, 18, 54, 32, 22, 35, 67, and 89. The corresponding position coordinates are 111, 121, 131, 141, 151, 112, 122, 132, 142, and 152, respectively. The associated locations can be expressed as follows: A_{111}^{21} , A_{121}^{13} , A_{131}^{16} , A_{141}^{18} , A_{151}^{54} , A_{112}^{32} , A_{122}^{22} , A_{132}^{35} , A_{142}^{67} , and A_{152}^{89} .

In the fourth step, the cargo relationships are considered to adjust the assigned locations. If there is a strong correlation between two goods in the same category (A, B or C), then these goods are placed in adjacent locations (Wang and Lyu 2016).

Suppose that $A_M^\gamma \rightarrow A_N^\beta$, meaning that a purchase of the former product is accompanied by a purchase of the latter with high probability, and that A_S^α is adjacent to A_M^γ . Position adjustment should be performed by replacing adjacent goods

with related goods, that is, $A_S^\alpha \Rightarrow A_N^\alpha$ and $A_N^\beta \Rightarrow A_S^\beta$. If two strongly related goods are not of the same type, as in the case of $A_M^\gamma \rightarrow B_T^\theta$, the related goods (B_T^θ) will still replace the adjacent goods (A_S^α). In this case, the adjacent goods (A_S^α) will replace the category A goods that are farthest from the I/O point (A_R^t), and goods from A_R^t will be placed in the original location of B_T^θ : $B_T^\theta \Rightarrow B_S^\theta$, $A_S^\alpha \Rightarrow A_R^\beta$, and $A_R^t \Rightarrow A_R^t$. If there are no strong correlations among the goods, then the locations of the goods are not adjusted.

5 Order picking efficiency test

5.1 Picking environment

In this research article on an improved storage strategy for daily necessities, a limited range of commodities are considered, and the dimensions of the storage area considered for simulation and verification are relatively small. As shown in Fig. 4, the warehouse has the following features: a horizontal shelf layout, with two single rows of shelves next to the walls and double rows of back-to-back shelves in the remaining space, one I/O point for the entrance and exit of goods, one cross-aisle with access for eight sorting operations, and eight picking points on each side of each picking aisle, for a total of 128 picking points.

The specific parameters of the warehouse are listed in Table 10.

The settings and assumptions for the picking environment are as follows:

- (1) The three-dimensional shelf space is ignored, that is, the number of layers of each shelf is set to 1.
- (2) Only one type of commodity is placed in each cargo location.

- (3) The number of pickers is assumed to be one, regardless of the number of simultaneous picks. The picking distance is calculated as the cumulative distance traveled by an individual picker, not the sum of the pickers' walking distance.
- (4) The maximum number of items to be picked at one time is 20, that is, the number of items in a pick list is less than or equal to 20.
- (5) The picking path is an S-type picking path. The picker enters the warehouse on the left side of the I/O port and exits the warehouse on the right side of the I/O port, as shown in Fig. 4.

The simulation model was developed using the Python programming language. The purpose of the simulation is to assess whether the improved ABC class-based storage strategy is more efficient than the traditional ABC class-based storage strategy: a decrease in picking distance indicates an improvement in picking efficiency, meaning that the improved ABC class-based storage strategy is superior to the traditional strategy.

The purpose of the simulation model used in this paper is to simulate an S-type picking path for each input pick list to obtain the corresponding distance traveled and then calculate the average walking distance per pick list. For the same pick list data, the different storage strategies yield different picking distance results. These results show that the average picking distance of the improved storage strategy is shorter than that of the traditional storage strategy, demonstrating that the improved storage strategy improves the picking efficiency.

5.2 Picking distance calculation formulas

Let D represent the distance traveled for a single pick list, and let D_k denote the distance traveled for the k th pick list on a

Fig. 4 Warehouse schematic

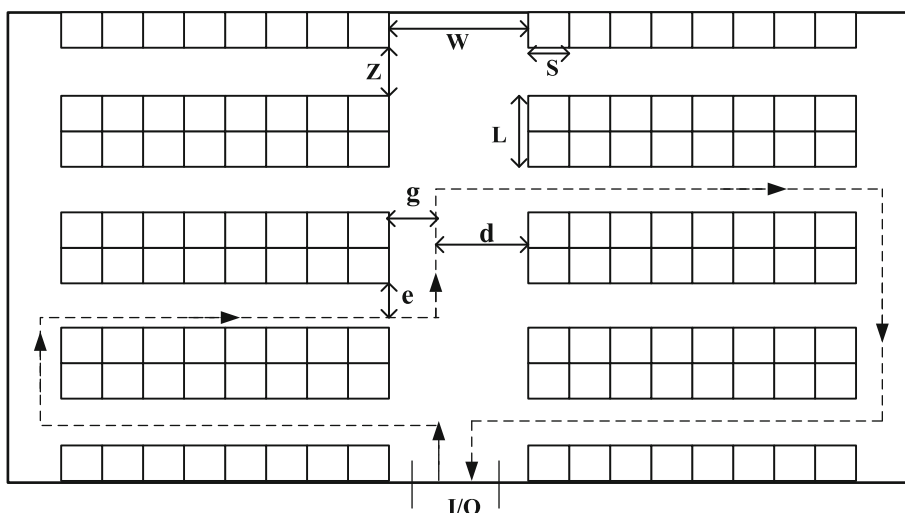


Table 10 Warehouse parameters

Parameter(unit: m)	Parameter description
$W = 3$	Cross-aisle width
$Z = 1.5$	Picking aisle width
$S = 1$	Width of the opening
$L = 3$	Width of each double row of back-to-back shelves
$d = 2$	Distance from the cross-aisle to a picking aisle
$e = 0.75$	Distance walked from the picking aisle for a picking operation
$g = 1$	Distance from a picking aisle to the cross-aisle

working day. The picking distance D is the walking distance that the order picker must travel through the picking aisles hosting the goods to be picked plus the distance traveled to bypass picking aisles without goods to be picked. The formula for D is as follows:

$$D = LD + MidD + RD \quad (1)$$

In Formula 1, the distance D for a single pick list is composed of three parts:

LD : The walking distance in the area to the left of the cross-aisle.

$MidD$: The walking distance in the cross-aisle.

RD : The walking distance in the area to the right of the cross-aisle.

The cargo coordinates are defined based on the I/O point as the origin. The y -axis of the coordinate system points into the warehouse from the I/O point along the longitudinal (vertical) direction, and the x -axis lies in the lateral (horizontal) direction.

Each position is indicated by three coordinate values, (x, y, z) , where x represents the horizontal coordinate, y represents the vertical coordinate, and z represents the quadrant.

As previously stated, the warehouse has eight picking aisles, with eight picking locations on each side of each picking aisle.

Therefore, the values of the x and y coordinates lie in the range $1 \leq x, y \leq 8$, and the value of z is either 1 or 2, where 1 indicates the first quadrant, that is, to the right of the main channel, and 2 indicates the second quadrant, that is, to the left of the main channel.

Table 10 defines the parameters that appear in the calculations.

- W : Cross-aisle width.
- Z : Picking aisle width.
- S : Width of the opening.
- L : Width of each double row of back-to-back shelves.
- d : Distance from the cross-aisle to a picking aisle.
- e : Distance walked from the picking aisle for a picking operation.

- g : Distance from a picking aisle to the cross-aisle.
- m : Number of picking aisles passed transversely during picking.
- n : Number of picking aisles passed longitudinally during picking.
- r : Number of double rows of back-to-back shelves passed longitudinally during picking.
- G : Number of picking aisles with no goods to be picked on either side.
- X : The transverse distance traveled for the picking task.
- Y : The longitudinal distance traveled for the picking task.
- LD : The total picking distance traveled in the area to the left of the cross-aisle, which is the distance traveled transversely plus the distance traveled longitudinally in this area.

Case 1 If no item to be picked is in the area to the left of the cross-aisle, then the corresponding walking distance is 0, that is, $LD = 0$.

Case 2 If any item to be picked is in the area to the left side of the cross-aisle, then the corresponding walking distance is not 0, that is, $LD \neq 0$.

The specific formula is as follows:

$$LD = X + Y \quad (2)$$

$$X = m \times (2g + 8S) \quad (3)$$

$$Y = n \times Z + r \times L \quad (4)$$

This formula depends on the specific values of the parameters m , n , and r and the specific placement of the goods, that is, the related coordinates.

Y_1 is the set of all y -axis values in the first quadrant, and y_{\max} is the maximum in this set. G represents the number of picking aisles with no goods to be picked on either side, that is, the number of picking aisles through which the picking operator need not pass. The range of G is $0 \leq G < 4$, $G \in N$.

Due to the S-type picking path, only one-way path segments pass through the picking aisles, so there is no turning back halfway. Therefore, the specific values of the parameters m , n and r are as follows.

When $1 \leq y_{max} \leq 4, m = 2, n = 1.5,$ and $r = 1.5.$ In this case, the goods are concentrated in the first two picking aisles nearest the I/O port, so the number of transverse passages through a picking aisle during picking is $m = 2,$ the number of picking aisles passed longitudinally during picking is $n = 1.5,$ and the number of double rows of back-to-back shelves passed longitudinally during picking is $r = 1.5.$

When $5 \leq y_{max} \leq 6,$ if $G = 0,$ then $m = 4, n = 3.5,$ and $r = 3.5.$ If $G \neq 0,$ then $m = 2, n = 2.5,$ and $r = 2.5.$ In this case, the cargo to be picked is located in the first three picking aisles nearest the I/O port. Moreover, when goods to be picked are located on one side or both sides of all three picking aisles, that is, $G = 0,$ the number of transverse passages through a picking aisle during picking is $m = 4,$ the number of picking aisles passed longitudinally during picking is $n = 3.5,$ and the number of double rows of back-to-back shelves passed longitudinally during picking is $r = 3.5.$ When there are no goods to be picked on either side of any of the three picking aisles, that is, $G \neq 0,$ the number of transverse passages through a picking aisle during picking is $m = 2,$ the number of picking aisles passed longitudinally during picking is $n = 2.5,$ and the number of double rows of back-to-back shelves passed longitudinally during picking is $r = 2.5.$

When $7 \leq y_{max} \leq 8,$ if $0 \leq G \leq 1, G \in N,$ then $m = 4, n = 3.5,$ and $r = 3.5;$ if $2 \leq G < 4, G \in N,$ then $m = 2, n = 3.5,$ and $r = 3.5.$ In this case, goods to be picked are stored on one or both sides of the innermost picking aisle on the left side of the warehouse. When there is no item to be picked on either side of at most one picking aisle in this area, that is, $0 \leq G \leq 1, G \in N,$ the number of transverse passages through a picking aisle during picking is $m = 4,$ the number of picking aisles passed longitudinally during picking is $n = 3.5,$ and the number of double rows of back-to-back shelves passed longitudinally during picking is $r = 3.5.$ When there are no goods to be selected on either side of two or three of the picking aisles in this area, that is, $2 \leq G < 4, G \in N,$ the number of transverse passages through a picking aisle during picking is $m = 2,$ the number of picking aisles passed longitudinally during picking is $n = 3.5,$ and the number of double rows of back-to-back shelves passed longitudinally during picking is $r = 3.5.$

RD is the picking distance traveled in the area to the right of the cross-aisle, which is the distance traveled laterally plus the distance traveled longitudinally in this area.

Case 1 If no item to be picked is in the area to the right of the cross-aisle, then the corresponding walking distance is 0, that is, $RD = 0.$

Case 2 If any item to be picked is in the area to the right of the cross-aisle, then the corresponding walking distance is not 0, that is, $RD \neq 0.$

Table 11 Simulation results (unit: m)

Traditional simulation results	1st improvement simulation results	2nd improvement simulation results
70.83	66.72	60.63
75.30	70.25	62.42
76.83	73.83	66.15
77.45	74.42	63.43
79.39	73.75	67.95
78.83	75.75	65.66
73.76	69.48	62.54
75.14	72.21	62.59
75.85	71.90	64.32
75.22	72.29	62.36
74.72	71.43	62.24
71.62	67.47	61.31
70.35	67.60	59.65
73.85	69.86	62.62
71.37	67.66	60.09
73.49	67.54	63.27
73.92	70.07	60.54
76.08	72.12	63.37
74.76	70.42	62.28
72.94	70.10	60.47
75.91	71.96	65.36
74.17	69.86	61.78
70.22	66.57	58.49
74.91	69.89	64.49
71.08	68.31	58.21
75.64	72.31	64.75
76.64	72.19	65.98
80.81	76.60	66.18
79.15	76.06	64.82

The specific formula is the same as for $LD.$

$$RD = X + Y \tag{5}$$

$$X = m \times (2g + 8S) \tag{6}$$

$$Y = n \times Z + r \times L \tag{7}$$

The specific values of $m, n,$ and r are the same as the values noted for the LD calculation formula.

$MidD$ is the distance traveled via the cross-aisle, including the distance traveled in the transverse direction and the distance traveled in the longitudinal direction.

The specific formula is as follows:

$$MidD = X + Y \tag{8}$$

$$X = W - 2d \tag{9}$$

$$Y = |Y_{LD} - Y_{RD}| \tag{10}$$

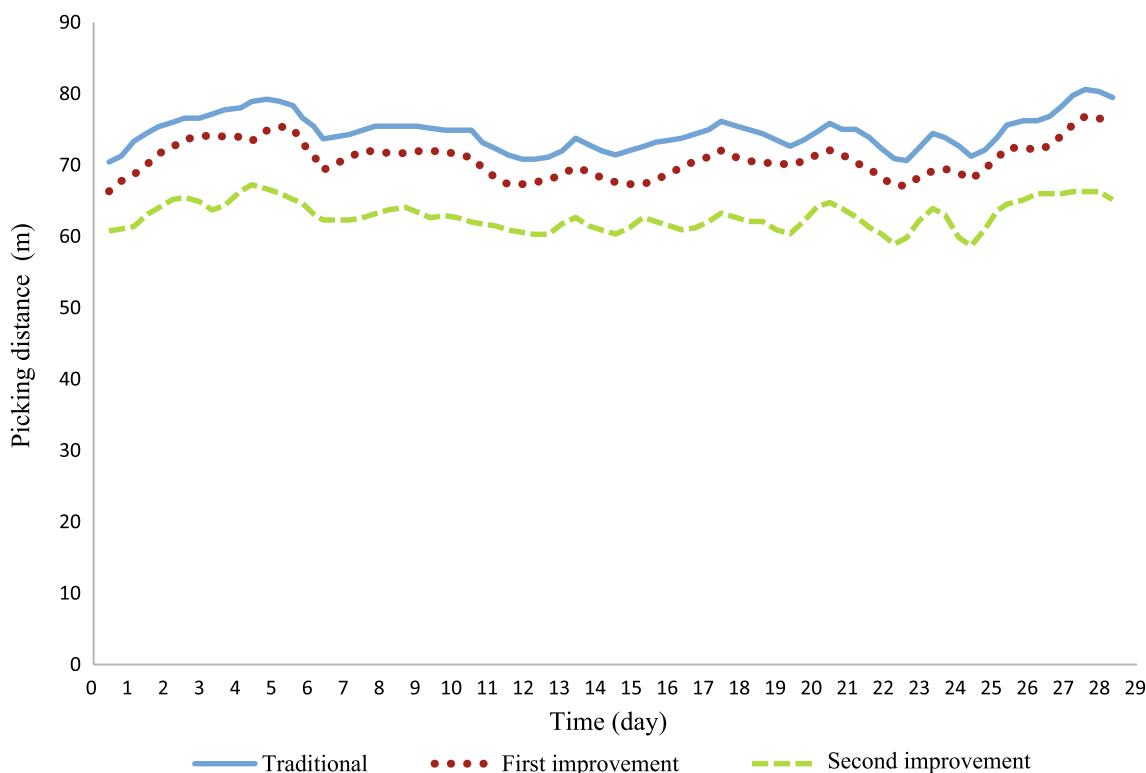


Fig. 5 Line chart of the simulation results

X is the transverse distance traveled in the cross-aisle, that is, from the left area to the right area; Y is the longitudinal distance traveled in the cross-aisle, which is also the absolute value of the difference between the longitudinal travel distances on the left and right sides of the cross-aisle.

5.3 Simulation model objective function

The objective function for the simulation model, D_{mean} , is the average daily distance traveled per pick list.

$$D_{\text{mean}} = \frac{\sum_1^{\text{order_quantity}} D_k}{\text{order_quantity}} \tag{11}$$

Constraint:

$$\sum_1^{\text{per_order_num}} X_i \leq 20, i \in \{1, 2, \dots, \text{order_num}\} \tag{12}$$

In the objective function, D_k represents the distance traveled for the k th pick list on a working day, and order_num represents the total number of pick lists fulfilled on a working day.

In the constraint, X_i represents a product on the i th pick list, and $\sum_1^{\text{per_order_num}} X_i$ represents the total number of products on the i th pick list. Equation (2) shows that the

number of products on any one pick list must be less than or equal to 20.

5.4 Comparison of the simulation results with the traditional and improved class-based storage strategies

The data used as the basis of the simulations were obtained from the orders placed on an e-commerce site by customers in North China during September, sorted by order date. Three simulations were conducted to determine the picking efficiencies under the traditional ABC class-based storage strategy, the ‘first improvement’ to the ABC class-based storage strategy, and the ‘second improvement’ to the ABC class-based storage strategy. The first improvement refers to the expansion of the basis for the ABC classification of the products from two attributes to five. The second improvement is the addition of the consideration of associations between goods in combination with the first improvement. The simulation results are presented in Table 11.

Figure 5 show a line chart of the simulation results. The picking efficiency is increased by both improvements, and the average picking distance is decreased. After the first improvement, the average picking distance per pick list is 5.1% lower than that of the traditional strategy. After the second improve-

ment, the average picking distance per pick list is reduced by 16% compared to that of the traditional strategy.

6 Conclusion

With the intensifying competition among e-commerce companies, the picking efficiency achieved in e-commerce warehouses is a very important concern in e-commerce logistics. The practical significance of this paper is that it presents an improved class-based storage strategy for e-commerce storage centers. This improvement is of great significance to e-commerce enterprises because it helps them to increase their picking efficiency and picking speed in their warehouses and thus to shorten the time required for order fulfillment logistics, reducing the time consumers must wait for their goods and enhancing the shopping experience.

This paper presents theoretical research on storage strategies and proposes an improved class-based storage strategy. After using web crawler technology to collect commodity information and product ordering data from an e-commerce website, the collected data were preprocessed and homogenized for data mining purposes. Clustering and association analysis were applied to the data to extract ordering patterns for daily commodities. By considering additional attributes of the goods, clustering results were obtained that could more precisely divide the goods into three categories (A, B, and C). The commodity ordering data were used to generate warehouse pick lists, and these pick lists were used to conduct an association analysis to find correlations between various types of goods in categories A, B, and C. The data mining results were used to refine the locations of goods in the storage center to improve the picking efficiency. Thus, in combination with expanding the attributes used as the basis for ABC class-based, the principle applied when placing classified goods in storage was reformulated. The improvements achieved with these two changes to the traditional class-based storage strategy were verified via simulations. The simulation results show that the average picking distance is shortened, indicating successful improvement.

Many potential directions for improvement remain to be explored in the future. Storage strategy research can still be improved in several respects. First, the product types considered in this paper were classified into only three categories, which is insufficient for subsequent in-depth research. Additionally, due to the limitations on the data collection process, the data span is not sufficiently large to truly reflect the actual situation. In future research, different types of goods should be more finely classified based on a larger amount of data, and a longer time span should be considered to identify patterns that more closely reflect reality. Second, several conditions were assumed in the simulations reported in this paper. In future research, a more realistic simulation model

should be used. Third, many factors affect the picking and sorting efficiencies in a storage center, including the picking path, order batching, the warehouse layout, and the storage strategy. However, the storage strategy was the only one of these factors considered in this paper. Other factors should be addressed in future research to determine how to further improve the efficiency.

Acknowledgements The study is supported by the National Natural Science Foundation of China “Research on the warehouse picking system blocking influence factors and combined control strategy” (No. 71501015), the Beijing Great Wall scholars program (No. CIT&TCD20170317), and the Beijing Collaborative Innovation Center, the Beijing Great Wall Fellowship program (No. CIT & TCD20170317), and the Beijing Collaborative Innovation Center.

Compliance with ethical standards

Conflict of interest This study was funded by the National Natural Science Foundation of China; no conflicts of interest exist.

Ethical approval This article does not report any studies on human participants or animals performed by any of the authors.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Cai Z, Yan H, Li P et al (2017a) Towards secure and flexible EHR sharing in mobile health cloud under static assumptions. *Clust Comput* 20(3):2415–2422
- Cai J, Wang Y, Liu Y et al (2017b) Enhancing network capacity by weakening community structure in scale-free network. *Future Gener Comput Syst*. <https://doi.org/10.1016/j.future.2017.08.014>
- Chen X (2015) Application of data mining technology for Apriori algorithm in warehousing configuration optimization. *J Tonghua Teach Coll* 36(06):9–11
- Chen W, Peng L, Wang J et al (2013) Inapproximability results for the minimum integral solution problem with preprocessing over infinity norm. *Theor Comput Sci* 478:127–131
- Chen W, Chen Z, Samatova NF et al (2014) Solving the maximum duo-preservation string mapping problem with linear programming. *Theor Comput Sci* 530:1–11
- Dkaich R, el Azami I, Mouloudi A (2017) XML OLAP Cube in the Cloud towards the DWaaS. *Int J Cloud Appl Comput: IJCAC* 7(1):47–56. <https://doi.org/10.4018/IJCAC.2017010103>
- Feng Q, Le M, Zhao Y (2015) Optimization of storage space assignment under material cluster analysis. *J Liaoning Tech Univ* 34(10):1207–1212
- Gupta S, Gupta BB (2018) Robust injection point-based framework for modern applications against XSS vulnerabilities in Online Social Networks (OSNs). *Int J Inf Comput Secur IJICS* 10(2/3):170–200. <https://doi.org/10.1504/IJICS.2018.091455>
- Haffar N, Maraoui M, Aljawarneh S et al (2017) Pedagogical indexed arabic text in Cloud E-learning system. *Int J Cloud Appl Comput: IJCAC* 7(1):32–46. <https://doi.org/10.4018/IJCAC.2017010102>

- Hatefi SM, Torabi SA, Bagheri P (2014) Multi-criteria abc inventory class-based with mixed quantitative and qualitative criteria. *Int J Prod Res* 52(3):776–786
- Hausman WH, Schwarz LB, Graves SC (1976) Optimal storage assignment in automatic warehousing systems. *Manag Sci* 22(6):629–638
- He P, Deng Z, Wang H et al (2016) Model approach to grammatical evolution: theory and case study. *Soft Comput* 20(9):3537–3548
- He P, Deng Z, Gao C et al (2017) Model approach to grammatical evolution: deep-structured analyzing of model and representation. *Soft Comput* 21(18):5413–5423
- Hu C (2010) Implementation and improvement of ABC class-based in inventory management. *Logist Eng Manag* 9:62–65
- Huang Y, Li W, Liang Z et al (2016) Efficient business process consolidation: combining topic features with structure matching. *Soft Comput* 22(2):645–657
- Jin X (2014) Research on B2C e-commerce enterprise location allocation considering demand correlation. Beijing Jiaotong University, Beijing, pp 35–39
- Kannan R, Vempala S, Vetta A (2000) On clustering's good, bad and spectral. In FOCS, pp 367–377
- Li R (2016) H company inventory control strategy. *Technol Enterp Technol* 35(1):90–91
- Li Y, Wang G, Nie L et al (2018) Distance metric optimization driven convolutional neural network for age invariant face recognition. *Pattern Recognit* 75:51–62
- Lin W, Wu Z, Lin L et al (2017a) An ensemble random forest algorithm for insurance big data analysis. *IEEE Access* 5:16568–16575
- Lin W, Xu S, Li J et al (2017b) Design and theoretical analysis of virtual machine placement algorithm based on peak workload characteristics. *Soft Comput* 21(5):1301–1314
- Lin W, Xu S, He L et al (2017c) Multi-resource scheduling and power simulation for cloud computing. *Inf Sci* 397:168–186
- Liu B, Sun L, Yu Y (2017) New progress in research on warehouse, logistics and supply chain management. *J Univ Sci Technol China* 47(02):176–187
- Long X (2017) Multi-criteria ABC inventory management based on cluster analysis method. *J Logist Sci* 40(2):62–65
- Luo C, Ye M (2017) ABC classification and ranking analysis method for the comparison of hospital drugstore management effect. *China Pharm* 26(04):92–94
- Rigutini L, Maggini M (2005) A semi-supervised document clustering algorithm based on EM. In: *Web intelligence*, pp 200–206
- Scott GL, Longuet-Higgins HC (1990) Feature grouping by relocalisation of eigenvectors of proximity matrix. In: *Proceedings British machine vision conference*, pp 103–108
- Teng X, Ma J (2016) Research on the optimization of logistics center warehouse location. *Mod Bus* 32:14–15
- Wang Y (2016) Research on the optimization of cargo storage center of H Machinery Company. Changchun University of Technology, Changchun, pp 21–33
- Wang Q, Lyu W (2016) An ABC class-based method based on the relevance of stock material. *Mark Mod* 30:107–108
- Wang H, Wang W, Cui Z et al (2018) A new dynamic firefly algorithm for demand estimation of water resources. *Inf Sci* 438:95–106
- Xiao J, Zheng L (2008) Optimization model of cargo space in maintenance spare parts warehouse. *J Tsinghua Univ Nat Sci* 48(11):1883–1886
- Yang W, Zhang W, Chang Y et al (2014) Optimization of location allocation in automated warehouses. *Mod Manuf Eng* 12:134–140
- Yu Y, De Koster RBM, Guo X (2015) Class-based storage with a finite number of items: using more classes is not always better. *Prod Oper Manag* 24(8):1235–1247
- Zhao L, Gong S, Liu Z (2015) Application of AHP improved ABC class-based in automobile spare parts inventory management. *Logist Technol* 9:247–250
- Zhou L, Liu H, Zhao X et al (2018) Study on the estimation of blocking rate in wide-aisle picking system. *Soft Comput*. <https://doi.org/10.1007/s00500-018-3148-3>
- Zhu Z (2017) ABC classification in supply chain management. *Mod Mark* 11(1):44–45

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.