Expert Systems

Article

DOI: 10.1111/exsy.12006

Data mining based storage assignment heuristics for travel distance reduction

David Ming-Huang Chiang,[1] Chia-Ping Lin[1] and Mu-Chen Chen[2]

(1) Department of Business Administration, National Taiwan University, College of Management 9F, No. 1, Section 4, Roosevelt Road, Taipei 106, Taiwan, R.O.C.
Email: cmh@ccms.ntu.edu.tw; d95741004@ntu.edu.tw
(2) Institute of Traffic and Transportation, National Chiao Tung University, 4F, No. 118, Section 1, Chung Hsiao W. Road, Taipei 100, Taiwan, R.O.C.
Email: ittchen@mail.nctu.edu.tw

Abstract: *Among the warehousing activities in distribution centres, order picking is the most time-consuming and labour-intensive. As a result, order picking may become a bottleneck preventing distribution centres from maximizing the effectiveness of their warehousing activities. Although storage location assignment (or product allocation) is a tactical decision, it is especially influential on the effectiveness of order picking. In previous studies, most storage assignment approaches considered the order frequency of individual products rather than that of product groups, which often are purchased together. This paper proposes a new association measure, weighted support count (WSC), based on association rule mining, to represent both the intensity and nature of the relationships between products in a distribution centre. This paper presents two storage assignment heuristics, the modified class-based heuristic (MCBH) and the association seed based heuristic (ASBH), designed to facilitate efficient order picking by applying WSC. The real-world data set of a grocery distribution centre is used to verify the effectiveness of the proposed approaches. From the computational results, MCBH cuts at most 4% from the travel distance for order picking per month, as compared with the traditional class-based approach. Meanwhile, ASBH achieves at most a 13% reduction in travel distance.*

Keywords: data mining, association rules, warehousing systems, storage assignment

1. Introduction

Nowadays, supply chain management has to focus on reducing both cost and time-to-market because a variety of products are continuously launched to fit better customers' diversified requirements. Because of limited shelf space, retail stores can only display product items in small quantities, thus requiring frequent replenishment from distribution centres. In supply chains, the distribution centre is an essential intermediary facility between suppliers and customers. The performance of distribution centres deeply affects the competency of supply chains. Today, the challenges that distribution centres face include working not only with a large set of product items, but also with a tremendous number of orders that include various combinations of those product items.

Generally speaking, warehousing activities in distribution centres include receiving, storage, order picking, consolidation, sorting and shipping. Among warehousing activities, order picking is the most time-consuming and labour-intensive (Koster *et al.*, 2007), and it generally accounts for 55% of warehouse operating expenses (Tompkins *et al.*, 1996). As a result, order picking becomes a warehouse bottleneck in distribution centres. In previous studies, four categories of approaches were shown to increase the efficiency of order-picking activity. They are (1) determining the appropriate order-picking route; (2) zoning the warehouse; (3) assigning products to the right storage locations and (4) assigning orders in batches (Roodbergen & Koster, 2001). Although the third approach, storage assignment (or product allocation) of product items, is a tactical decision, it is more influential on the effectiveness of order picking than any of the other three approaches.

One storage assignment method is to create a set of rules to assign products to storage locations (Koster *et al.*, 2007). A well-designed storage assignment approach could significantly reduce the travel distance or time of order picking. The storage assignment problem usually becomes more complicated as the numbers of locations and products involved increase. Thus, in most previous studies, heuristic approaches were developed to address the storage assignment problem. Common guidelines to design those heuristic approaches included maximizing utilization of the facility or allocating those products with a higher order frequency closer to the warehouse exit. However, some products are frequently ordered together and potentially can be allocated close to each other for the convenience of order picking. A few storage assignment approaches have developed specific measurements to consider the cross-product relationship, such as statistical measure (Frazelle & Sharp, 1989) and contact frequency (van Oudheusden *et al.*, 1988; van Oudheusden & Zhu, 1992). However, these studies did not clearly reveal the nature of product relationships. Basically, the nature of product relationships can be divided into three categories: substitutive, complementary and independent. Where relationships are of different natures, they should follow different assignment logics. For instance, if the relationship between two products is complementary, these two products should be allocated to locations close together, to reduce the travel distance of order picking.

Continuous technological innovation is important to enterprises' survival and development (Wu & Olson, 2009). The relatively new technique, data mining, can be applied to support the development of enterprises by extracting valuable information from enterprises' databases. One data mining technique, association rule mining, is able to discover both the intensity and the nature of relationships between certain products that often are purchased together by analysing the transaction database. Because of the enormous decrease in the cost of data collection and storage in the last few decades, abundant transaction data can now be stored and kept in information systems such as POS (Point of Sale) and WMS (Warehouse Management System). Therefore, it is practical to apply association rule mining to extract useful information for storage location assignment decision making.

This paper proposes an association measure, *weighted support count* (WSC), based on association rule mining, to represent both the intensity and nature of relationships between products (see Section 3.2 for details). Correspondingly, two storage assignment heuristics, modified class-based heuristic (MCBH) and association seed based heuristic (ASBH), are developed to improve order picking by allocating products frequently ordered together in the same aisle. In both MCBH and ASBH, WSC is applied to measure the intensity and nature of relationships between products. The real-world data set of a grocery distribution centre is then used to verify the performance of MCBH and ASBH in comparison to the traditional class-based storage assignment approach. The remainder of this paper is organized as follows. The literature related to storage assignment and association analysis is reviewed in the next section. The proposed heuristics, MCBH and ASBH, are presented in Section 3. The computational results are presented in Section 4. Finally, the conclusions are drawn in Section 5.

2. Literature review

In previous studies, assignment policies were mainly divided into three categories: rule-of-thumb policy, class-based policy, and family grouping policy (Koster *et al.*, 2007). In the past, the primary purpose of building a warehouse was storing inventory, so the storage cost was the main concern. In managing a warehouse, maximizing both the utilization of the storage facility and the convenience of putting products away were emphasized. In the past, only information pertaining to the utilization of storage locations could be obtained and used to resolve the storage location assignment problem (SLAP). The literature proposed some rule-of-thumb policies such as random location assignment, closest open location, farthest open location and longest open location (Gu *et al.*, 2007). The random location assignment policy allocates products to available locations in a random manner. The closest open location policy assigns products to the unallocated location nearest to the outbound exit, whereas the farthest open location policy assigns products to the location that is the farthest. The longest open location policy selects the available location with the longest unoccupied time. The storage assignment policy significantly directly affects the utilization of the storage facility. However, when considering the labour cost and service level, the performance of order picking has a greater impact than the utilization of the storage facility. As a result, later studies focused more on enhancing the performance of order picking through proper storage assignment than on utilization of the storage facility.

With the development of WMS, products' order frequency could be recorded and applied to the storage assignment process for better warehousing performance. The class-based policy stems from inventory control and divides various product classes according to the product turnover rate or order frequency. This kind of policy is designed to reduce the order-picking time while simultaneously maximizing the storage usage. A class-based policy enhances the efficiency of order picking by means of assigning products with a higher turnover rate to a zone near the outbound exit. Each product class is allocated to a dedicated zone according to the turnover rate of the class. The products in a product class are randomly allocated to locations within their zone to increase storage utilization and allocation convenience. Heskett (1963, 1964) proposed a cube-per-order index (COI) to categorize products and developed a storage assignment procedure based on this index in which products with a higher COI are placed nearer the outbound exit of a warehouse. Based on Heskett's works, Kallina and Lynn (1976) considered four practical issues – compatibility, complementarity, popularity and space – in their class-based assignment procedure. However, most previous studies used turnover rate as the basis to classify products for storage assignment. Hausman *et al.* (1976) compared the closest open location, turnover-based and class-based assignment approaches in terms of travel time and distribution of turnover of products by simulating an automated storage/retrieval system (AS/RS). From their comparisons, turnover-based and class-based approaches perform better than the closest open location approach in terms of reducing the travel time of order picking. Graves *et al.* (1977) extended Hausman's work by additionally considering the dual commands of depositing and picking in AS/RS. Jarvis and McDowell (1991) developed a heuristic storage assignment method based on the stochastic model, which minimizes the expected travel distance by allocating products to aisles according to their picking frequency. van de Berg (1996) also used dynamic programming to solve a class-based allocation problem in a single-command situation. Later on, Larson *et al.* (1997) took the storage space of different floors into account and proposed a class-based heuristic for storage assignment. Manzini *et al.* (2007) ran a set of designed experiments to explore the effect of factors presented in the class-based policy. Renaud and Ruiz (2008) considered discrete order picking and proposed a turnover-based heuristic to reduce the picking distance. Muppani and Adil (2008a) applied simulated annealing to solve a complex 0-1 integer model that simultaneously assigned products to a class and a storage location. Muppani and Adil (2008b) also addressed the class-based allocation problem with a non-linear integer-programming model and developed a branch-and-bound algorithm to solve the problem. Chan and Chan (2011) attempted to improve the performance of order picking by introducing class-based policies in a case study. The COI index and EIQ (entry-item-quantity) analysis are both methods to generate product classes in class-based policies. Class-based polices assume that all products are independent in ordering and picking.

Some researchers also worked on family grouping policy. The family grouping policy considers product relationships, which are described as the frequency that products are ordered together. The basic idea is that products frequently ordered together should be stored closer to each other for the efficiency of order picking (van Oudheusden & Zhu, 1992). Frazelle and Sharp (1989) proposed a correlation assignment policy by using the statistical correlation, which is the ratio of the number of orders in which two products appeared to the number of all orders. In their study, products are assigned pairwisely to storage locations, in descending order of statistical correlation. van Oudheusden *et al.* (1988) developed a pairwise interchange procedure, based on the distance and closeness between products, to allocate spare parts in the warehouse of a steel mill. In their work, closeness is defined as the frequency with which two parts were retrieved together. By interchanging parts iteratively, the parts with higher closeness were allocated nearer to each other. van Oudheusden and Zhu (1992) further considered recurrent orders in AS/RS and proposed a storage assignment approach that employed contact frequency, which counts the frequency with which two products are ordered together. They formulated SLAP as a set-partitioning problem by considering contact frequency, which could be resolved by a heuristic method. Lee (1992) used the items' propensity to classify similar products into groups and to assign storage locations by each group's COI. The items' propensity was based on the frequency a pair of items requested together in customer orders. Rosenwein (1994) grouped products according to order patterns and formulated the clustering problem as a *p*-median 0-1 integer programme. Liu (1999) considered item quantity in orders and measured the similarity coefficient with the probability that the pair of items appear on the same order. Liu's paper used both the correlation of products and the correlation of customers to formulate SLAP as a 0-1 integer programme. Jane and Laih (2005) applied family grouping policy in a synchronized zone order picking system. They developed a similarity measurement named order request and a corresponding heuristic to solve the *p*-median cluster problem. Manzini (2006) adopted statistical correlation, as proposed by Frazelle and Sharp (1989), and designed algorithms to generate product families. Xiao and Zheng (2010) proposed a bill-of-material (BOM) oriented, class-based storage assignment method. The BOM information was used to deal with SLAP. The frequency of a pair of parts presented together in all BOMs was the similarity measurement. Although both order pattern and BOM structure were used to find the relationship between products, the relationship between products could be complementary or substitutive. These two different types of relationships should follow different assignment logics. In the case of a complementary relationship, products should be allocated closer for order-picking convenience, whereas in the case of a substitutive relationship, products should not be allocated closer.

Data mining is a relatively new technique, which is used to extract the interesting patterns or rules from large amounts of data (Han & Kamber, 2012), and it has been applied in various areas such as banking (Emrouznejad and Anouze, 2010) and medicine (Delen, 2009; Yeh & Wu, 2010). In data mining, association rule mining can discover both the intensity and the types of relationship between products. Agrawal *et al.* (1993) and Srikant and Agrawal (1997) used association rule mining to discover the set of products recurrently purchased together by customers in a single store visit. Association rule mining is prevalently adopted in marketing (e.g. Liou & Liu, 2012; Chen *et al.*, 2012). The characteristics of association rule mining results are generally presented by three indexes: support, confidence and lift. Support evaluates the popularity of a rule. Family grouping policy proposes a similar concept of support value. Confidence measures the certainty of a rule. Lift illustrates the type of relationship between products. These three measurements are usually employed to select useful rules. To consider the intensity and types of relationships between products, the concepts of support and lift value are combined into an association measure, WSC, in this study. Also, two association-based heuristic approaches for storage assignment based on the WSC measure are developed. The new association measure and two association-based heuristic approaches will be described in the next section.

3. The association-based heuristic approaches

3.1. Overview

Two heuristics, MCBH and ASBH, are designed to solve SLAPs by considering relationships between products. The logic behind these two heuristics is to allocate the products that are frequently ordered together to the same aisle as much as possible by maximizing the aggregation of association measure, WSC, between products allocated to the same aisle. In Section 3.2, association rule mining is briefly introduced, followed by a description illustrating the new association measurement, WSC. The detailed procedures of MCBH and ASBH are interpreted in Sections 3.3 and 3.4, respectively.

The scenario we consider in this study is described as follows. In the distribution centre, each product can be only stored in one storage location. There are only one entrance and one outbound exit in the distribution centre. A picking tour follows an S-shape strategy (Hall, 1993). The picker will not enter an aisle if that aisle contains no product item that needs to be picked. Orders are picked on a first-come-first-served basis, which means that this distribution centre does not apply order batching. Besides, only retrieval of requested products is considered in the order-picking route, which is called a single command. Additionally, only relationships between pairs of products are considered. Before WSC and the two heuristics are presented, the notation related to the new association measure, WSC, and the two heuristics is defined in Table 1.

3.2. Weighted support count

The approach of association rule mining can be briefly described as follows. Let $K = \{1, 2, \ldots, p\}$ denote a set of literals, namely items. There are p items in the set K. Moreover, let D represent a set of transactions, where each transaction TS is a set of items such that $TS \subset K$. A unique identifier, namely TID, is associated with each transaction TS. Let Y and Z be the itemsets. A transaction TS is said to contain Y if $Y \subset TS$. The result of association rule mining takes the form of $Y \rightarrow Z$, which means that customers buying a set of items Y will also buy a set of items Z, where $Y \subset K$, $Z \subset K$ and $Y \cap Z = \emptyset$. Support, defined as $P(Y \cup Z)$, evaluates the popularity of a rule. There are $P(Y \cup Z)$ transactions in a data set that contains both items of Y and Z. Confidence, defined as $P(Z|Y)$, measures the certainty of a rule. $P(Z|Y)$

Table 1: *Notation*

	Indices
C	The set of class zones; $c = 1, 2 \ldots l$
I	The set of aisles; $i = 1, 2 \ldots n$
J	The set of locations; $j = 1, 2 \ldots o$
K	The set of unallocated products; $k = 1, 2 \ldots p$
Parameters	
WSC_{YZ}	The *weighted support count* between product Y and product Z
LW_{YZ}	The transformed weight of lift value which is positive if lift value between product Y and Z is greater than 1; negative if less than 1; 0 if equal to 1
$supc_{YZ}$	The support count between product Y and product Z
$Lift_{YZ}$	The lift value between product Y and product Z.
R	The proportion of the reserved locations in an aisle
UJ_c	The set of unreserved locations in zone c
RJ_c	The set of reserved locations in zone c
AU_c	The set of products are allocated in the unreserved locations of zone c
AR_c	The set of products are allocated in the reserved locations of zone c
AWS_{jk}	The aggregation of *weighted support counts* between products already allocated in the aisle of reserved location j and the unallocated product k
Z_j	The set of products are already allocated in the aisle of location j
A_i	The set of products are allocated in aisle i
L_i	The set of available locations in aisle i
Decision variables	
X_{jk}	$= \begin{cases} 1, & \text{if product } k \text{ is allocated in location } j \\ 0, & \text{otherwise} \end{cases}$

represents the probability that customers who buy items of Y will also buy items of Z. Lift, defined as $P(Z|Y)/P(Z)$ or $P(Y \cup Z)/P(Y)P(Z)$, illustrates the type of relationship between the products. Lift compares the observational probability, $P(Y \cup Z)$, with the theoretical probability, $P(Y)P(Z)$. If the lift value is greater than 1.0, which means the actual probability of Y and Z appearing in the same order is greater than the theoretical probability, the relationship of products is complementary; if the lift value is less than 1.0, the relationship is substitutive; if the lift value is equal to 1.0, the relationship is independent. In this paper, association rules are developed from the order data by using the *a priori* algorithm (Agrawal *et al.*, 1993; Srikant & Agrawal, 1997; Han & Kamber, 2012), which is an efficient algorithm for mining association rules. Because the proposed WSC measure represents the association between any pair of products, we set the rule length to be 2 in this paper.

The WSC measure involves the concepts of support and lift value. Support and lift value are used to represent the intensity and nature of relationships between products, respectively. However, the huge number of product pairs and the large set of orders may lead to relatively small support values. Support value is not suitable to apply to WSC. Therefore, *support count*, which means the count of any pair of products that appear in the same order, is adopted in our proposed heuristics to avoid the possible round-off error. The nature of relationships between products can be revealed by lift value. Different types of relationships should follow different assignment logics. A complementary relationship between two products indicates a positive fitness to allocate them in the same aisle. On the contrary, a substitutive relationship indicates a negative fitness. To take these logics into account, the support counts need to be transformed. WSC is positive if the lift value is greater than 1; it is negative if the lift value is less than 1; it is 0 if the lift value is equal to 1. Based on the above discussion, WSC is then expressed as

$$WSC_{YZ} = LW_{YZ} \times supc_{YZ} \begin{cases} LW_{YZ} = 1, & \text{if } Lift_{YZ} > 1 \\ LW_{YZ} = 0, & \text{if } Lift_{YZ} = 1 \\ LW_{YZ} = -1, & \text{if } Lift_{YZ} < 1 \end{cases} \tag{1}$$

WSC will be used in the two proposed storage assignment heuristics, so it should be calculated in advance.

3.3. *Modified class-based heuristic (MCBH)*

The first heuristic, MCBH, combines traditional class-based policy and family grouping policy to solve the SLAP. In the traditional class-based policy, products are sorted into classes by considering turnover rate, and locations are grouped into several zones according to the distance between each location and the outbound exit. The number of product classes and the number of location zones are decided correspondingly. Previous studies usually recommended two or three classes and the parallel layout of zones. The location zone near the outbound exit is first assigned to the product class with a higher turnover rate. After allocating all product classes to location zones, product items are allocated randomly within each location zone. In MCBH, the turnover rate of products, intensity and types of relationship between products are considered simultaneously. Therefore, two modifications are made accordingly. The first modification is that a certain proportion of locations in each aisle are reserved at the beginning. The reserved proportion, R, and the number of class zones, $|C|$, are predetermined. In addition to dividing all the storage locations into several class zones, the aisle locations in a class zone are further separated into reserved and unreserved locations based on a predetermined proportion, R. The second modification is that products are allocated zone by zone iteratively, instead of allocating at the same time. The procedure starts from the class zone nearest the outbound exit. The iteration proceeds one class zone at a time and can be broken down further into two stages. In the first stage, the unallocated products in set K are assigned to unreserved locations in the class zone based on their turnover rate. This stage is similar to traditional class-based approaches. After executing the first stage, the unallocated product set K needs to be updated. In the second stage, the reserved locations are allocated to the rest of the products in set K that have a higher association, in terms of WSC, with the products already placed in the same aisle. A 0-1 integer-programming model is applied to determine the product allocation to maximize the association between products allocated in the same aisle. The allocation of products in the second stage is similar to the idea of family grouping policy. Next, the unallocated product set K and class zones set C are updated. Then, the assignment procedure repeats until all class zones or all products have been allocated.

The *aggregated weighted support count* (AWS) is used to measure the association between products that need to be put away and products already allocated in the aisle of a reserved location. AWS_{jk} adds the WSCs between unallocated products k ($\forall k \in K$) and products already allocated in the aisle of location j. Let Z_j represent the set of products already allocated in the aisle of location j. All the AWS_{jk}s cross all combinations of reserved locations in the class zone c, and unallocated products should be calculated in advance and input to the 0-1 integer-programming model. AWS_{jk} takes the following form:

$$AWS_{jk} = \sum_{k' \in Z_j} WSC_{kk'} \forall j \in RJ_C, \forall k \in K \quad (2)$$

The 0-1 integer-programming model of location assignment for maximizing the association of products in the same aisle is formulated as follows:

$$\text{Max} \sum_{j \in RJ_c} \sum_{k \in K} X_{jk} \times AWS_{jk} \quad (3)$$

$$\text{Subject to} \sum_{k \in K} X_{jk} = 1 \quad \forall j \in RJ_c \quad (4)$$

$$X_{jk} \in \{0, 1\} \quad (5)$$

Objective function (3) maximizes the association between products allocated in the reserved locations and products already allocated in the same aisle. Constraint set (4) limits a reserved storage location to a particular product. Constraint set (5) guarantees the binary solution. However, when the number of products that must be put away is less than the number of reserved locations, Constraint (4) is replaced by Constraint (6), which ensures that each location allocates, at most, one unallocated product.

$$\sum_{j \in RJ_c} X_{jk} \leq 1 \quad \forall k \in K \quad (6)$$

The MCBH procedure is schematically illustrated in Figure 1, and the steps are described as follows:

Step 0: Set R and $|C|$.

Step 1: Perform *a priori* algorithm to generate association rules, and compute WSCs.

Step 2: Sort products in descending order by turnover rates, and sort storage zones in descending order by average distance between the locations in a zone and the outbound exit.

Step 3: Reserve R percentage of storage locations in each aisle randomly. In each zone, the locations are divided into two sets, reserved location set RJ_c and unreserved locations set UJ_c.

Step 4: Select the unallocated class zone c that is nearest to the outbound exit from the class zone set C.

Step 5: Assign the locations in UJ_c to unallocated products in set K according to the turnover rate in zone c.

(a) Select $|UJ_c|$ (the number of locations in UJ_c) products that have the first $|UJ_c|$ highest turnover rate in K into the AU_c.

Assign the locations in UJ_c to the products in AU_c randomly.

Exclude the products in AU_c from the unallocated product K., $K = K \backslash \{A_c\}$.

Step 6: Assign the locations in RJ_c to unallocated products in K according to the association between products.

(a) Compute the AWS_{jk} between all locations in RJ_c and all products in K by using equation (2).

(b) Assign the locations in RJ_c by solving the 0-1 integer-programming model (equations (3)–(6) to maximize the association between products within each aisle of the selected zone.

(c) Exclude the class zone c from the zone set C and the products allocated in zone c from unallocated product set K, $C = C \backslash \{c\}$, $K = K \backslash \{AR_c\}$.

Step 7: If the zone set C is empty or unallocated product set K is empty, proceed to Step 8. Otherwise, return to Step 4.

Step 8: Output the allocation of all products, and terminate MCBH.

3.4. *Association seed based heuristic (ASBH)*

ASBH is a heuristic approach, which performs assignment aisle by aisle and tries to maximize the aggregated association between products allocated in the same aisle. ASBH approximates the situation in which all locations are assigned based on the ordering relationship between products in MCBH. The assignment procedure starts to allocate products from the aisle nearest to the outbound exit in I. In the beginning of allocating an aisle, two unallocated products in K have the highest WSC between each other and so are chosen as the seed for the set of products allocated in the aisle i, A_i. Next, the unallocated product that has the highest WSC with one of the products in set A_i is added into set A_i iteratively until the number of products in set A_i equals the number of available locations in aisle i. The products in A_i are randomly allocated in aisle i. The process repeats until all aisles are assigned or all the products are allocated. The procedure of ASBH is schematically illustrated in Figure 2, and the steps are described as follows:

Step 1: Perform *a priori* algorithm to generate association rules, and compute WSCs.

Step 2: If the number of products ($|K|$) is less than the number of locations ($|J|$), proceed to Step 3. Otherwise, proceed to Step 4.

Step 3: Adjust the set of available locations in each aisle, L_i. To maximize space utilization and prevent congestion in picking, unassigned locations are appointed to aisles randomly. Then, $|L_i|$ (the number of available locations) is reduced in aisle i.

Step 4: Select aisle i, which is nearest to the outbound exit from the aisle set I, and initially set A_i to an empty set, $A_i = \emptyset$.

Step 5: Select 2 products, x and y, which have the highest WSC between each other from the unallocated products, and set K as the seed.

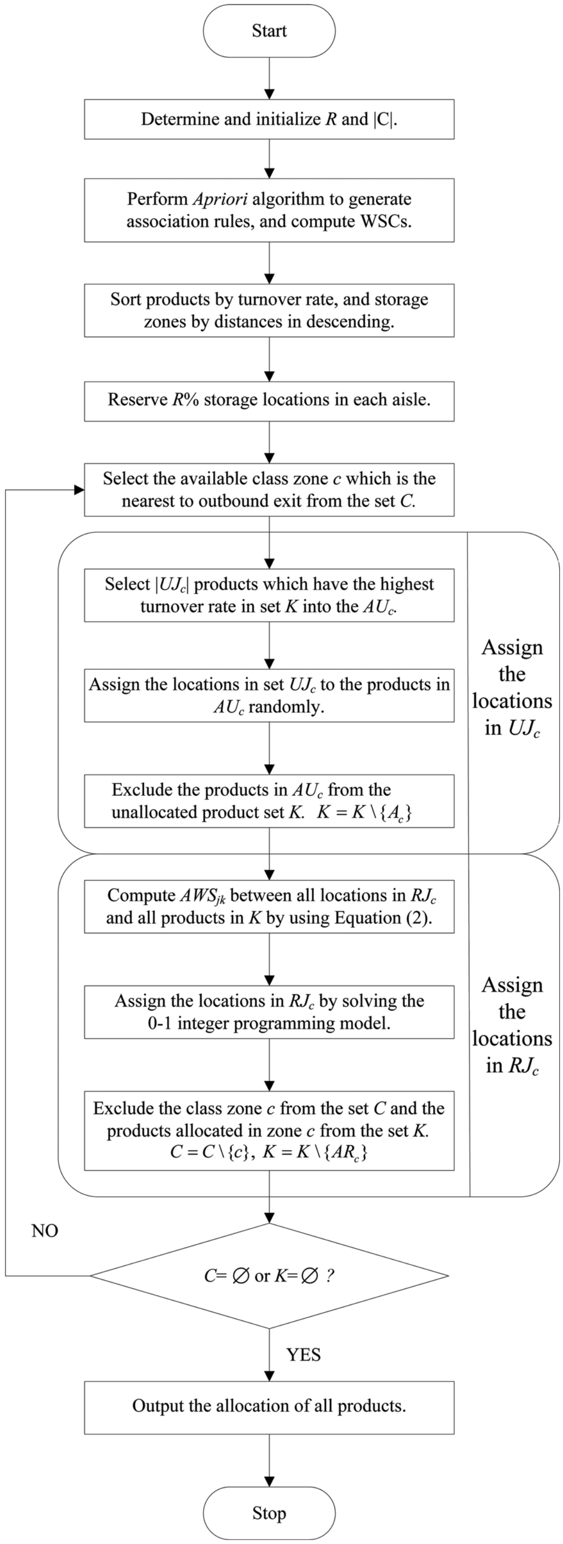

Figure 1: *Flowchart of MCBH.*

$\{x, y\} = \arg\max\{WSC_{xy}; x, y \in K\}$. If the WSCs between unallocated products are all equal, select two products, x and y, randomly as the seed.

Step 6: Add x and y to A_i, $A_i = A_i \cup \{x, y\}$. Then, exclude x and y from unallocated product set K, $K = K \backslash \{x, y\}$.

Step 7: Select product k, which has the highest WSC with one of the products in A_i. $\{k\} = \arg\max_k\{WSC_{zk}; z \in A_i, k \in K\}$. If the WSCs between unallocated products and the products in A_i are all equal, select a product, k, randomly.

Step 8: Add k to A_i, $A_i = A_i \cup \{k\}$. Then, exclude product k from unallocated product set K, $K = K\backslash\{k\}$.

Step 9: If all the locations in the selected aisle i have been assigned $(|A_i| = |L_i|)$ or the unallocated product set K is empty ($K = \emptyset$), proceed to Step 10. Otherwise, return to Step 7.

Step 10: Allocate the products in A_i in the locations of aisle i randomly. Exclude the selected aisle i from the aisle set I, $I = I\backslash\{i\}$.

Step 11: If aisle set I is empty ($I = \emptyset$) or the unallocated product set K is empty ($K = \emptyset$), proceed to Step 12. Otherwise, return to Step 4.

Step 12: Output the allocation of products and terminate ASBH.

Both the proposed heuristics solve SLAP based on associations between products. MCBH mixes together the features of the traditional class-based approach and family group policy. Products are allocated partially according to their turnover rate and partially according to their relationships with each other. Which approach plays the main role may depend on the setting of reserved percentage R. On the other hand, ASBH assigns storage locations entirely based on the relationship between products. It is a special case of MCBH in which R is set to 100%. ASBH purely maximizes the AWS between products allocated in the same aisle. The implementation of these two heuristics is presented in the next section through a case study of a grocery distribution centre.

4. Model verification

In this section, the proposed two heuristics, MCBH and ASBH, are implemented with a real data set of a grocery distribution centre in Taiwan. The characteristics of this data set are summarized in Section 4.1. Data preprocessing and the experimental setting are presented in Section 4.2. The experimental results and analysis are reported in Section 4.3.

4.1 Data set

The data set is extracted from the order database of a distribution centre that belongs to the logistics group of a company in Taiwan. The distribution centre we studied supplies the daily orders of grocery stores in the north of Taiwan. To balance the workload and avoid traffic congestion around the stores served by the distribution centre, the grocery stores place their orders and request the time at which they will receive goods, which usually is at night or in the early morning. As a result, the distribution centre is usually packed with orders in the evening. All order-picking activities need to be finished in a very short time frame. In addition, customer orders are often small in size and high in variety, such that the order picking in the picking area becomes very labour-intensive.

MCBH and ASBH are suitable to be applied in this kind of distribution centre for two reasons. First, some recurrent patterns may exist in the orders of the grocery stores. These recurrent patterns can represent the ordering relationship between products and can be used to assign storage allocations to improve order picking. Second, order picking creates a

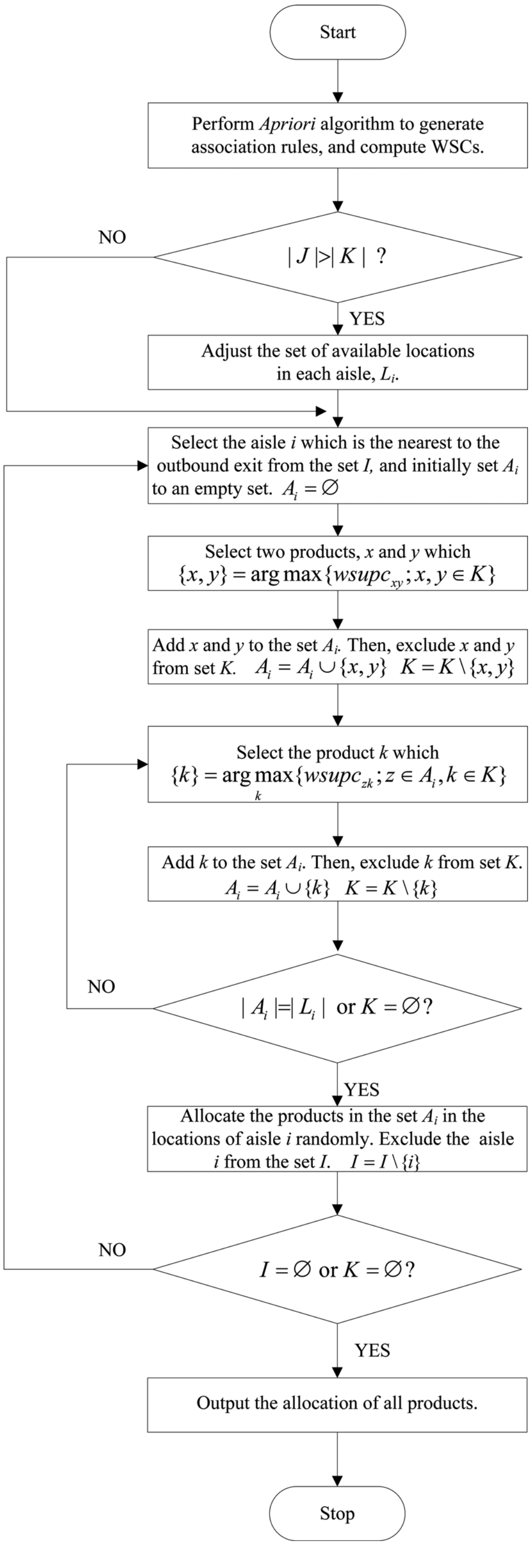

Figure 2: *Flow chart of ASBH.*

bottleneck in warehousing activities. Reducing the time of order picking is critical to the distribution centre's operation. It is worthwhile to allocate products appropriately in advance to enhance order picking.

4.2 Data preprocessing and experimental setting

The data set of order transactions is extracted from the WMS in the distribution centre. This data set includes 338113 daily orders from all served grocery stores in one year. There are 1308 distinct products recorded in the order transaction table, but only 787 products are stocked in the picking area. The original data set is divided into two parts for evaluating the performance of the proposed heuristics. A total of 312665 transaction records from January to November are used to generate WSCs and the remaining 25448 records for the month of December are taken as the new incoming orders for validation. The performance of different allocation approaches is measured by the travel distance of the December orders. The experimental setting is described as follows:

- The distribution centre has a rectangular shape with 800 storage locations. There are 21 aisles and 20 shelves. Each shelf has a two-sided space and 20 locations with one layer on each side. Figure 3 shows the layout of the distribution centre.
- There are 787 products that need to be allocated within the picking area.
- The width of the shelf is 1.5 m, the total width of two-sided shelf is 3 m and the length of an aisle is 32 m.

5. Results and analysis

Based on the experimental setting described in Section 4.2, a set of experiments is conducted to demonstrate the viability of the proposed heuristics. The traditional class based allocation approach (CBAA) is chosen as the benchmark to compare with the performance of MCBH and ASBH. Because these three approaches partially involve random assignment, each experimental setting is run for four trials to ensure robustness. After obtaining the storage assignment, the order data of December are used to represent the incoming orders for calculating the travel distance of order picking under different storage assignments.

In implementing CBAA and MCBH, products and locations are divided into two product classes and two zones for simplicity. The travel distance of four runs by CBAA for one month is summarized in Table 2. The results of average travel distance show that the travel distance of the four trials is roughly the same. It implies that CBAA is a robust approach for storage assignment.

To explore the effect of the relationship between products, four different levels of reserved proportion, 10%, 20%, 30%, and 40%, are studied in this paper. Products are allocated according to association in reserved locations. Table 3 shows the travel distance and computation time of order picking for one month under the allocation generated by MCBH with 10%, 20%, 30%, and 40% reserved proportions. Similar to CBAA, the results of MCBH are stable between trials. Table 4 summarizes the average travel distance generated from allocation by CBAA and MCBH. In comparison to CBAA, the travel distance generated by MCBH decreased from 1% to 4% and was dependent on the level of reserved proportion. MCBH, which combines the class-based allocation and family grouping approaches, outperforms the traditional class-based approach in terms of travel distance. The results indicate that allocating products that are frequently ordered together in the same aisle can enhance the efficiency of order picking.

From Table 3, it seems that travel distance decreases as the reserved proportion increases. Furthermore, from the

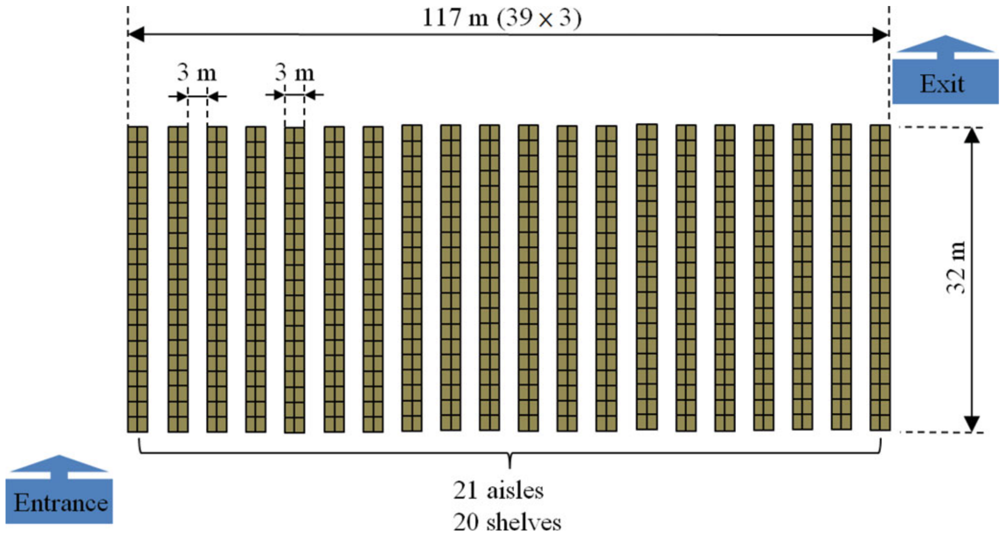

Figure 3: *The layout of the distribution centre.*

Table 2: *The travel distance of allocation by CBAA*

Trial	*Data set*	*distance (m) Travel*	*Average travel distance (m)*
Trial 1	Orders of December	6174879	6207567
Trial 2	Orders of December	6257439	
Trial 3	Orders of December	6201183	
Trial 4	Orders of December	6196767	

one-factor ANOVA shown in Table 5, the main effect of the reserved proportion is a significant effect on the improvement of travel distances at the significance level $\alpha = 0.01$. Therefore, increasing the reserved proportion of locations in MCBH can significantly achieve a greater reduction in the travel distance. More reserved proportions of locations allow for more pairs of products with a higher association to be allocated in the same aisle. Operators can pick more products in the entrance of an aisle such that the travel distance is shortened. From the experimental results, raising the reserved proportion can achieve better picking performance. The proposed heuristics are run on a laptop with 2.20 GHz CPU and 3.49 GB DRAM. Table 3 shows that the computational requirement also increases enormously as the reserved proportion increases because more reserved locations result in more computational effort spent on calculating AWS_{jk}. By using MCBH, increasing reserved locations leads to a better result but needs more computational requirements. ASBH is designed to approximate the situation of allocating all products based on their association between one another. The CPU time of ASBH is about 5 min, and it is much shorter than that of MCBH. Table 6 summarizes the travel distances of four runs generated by ASBH for different data sets. Table 7 summarizes the average travel distance of CBAA and ASBH.

Compared to CBAA, the travel distance generated by ASBH is decreased by approximately 13%, which is better than the distance generated by MCBH. ASBH attempts to store products frequently ordered together in the same aisle as much as possible so that the operator can pick more items in an entrance of an aisle. The results demonstrate sizable distance savings achieved solely by considering product relationships when allocating products. Intuitively, the decrease in travel time would be less than 10% according to the results from Table 4. However, the reduction appears to be larger than expected. The additional reduction may have been caused by eliminating the unreserved area. Without the

Table 3: *The travel distance and computation time of allocation by MCBH with December orders*

Reserved proportion (%)	*Trial*	*Travel distance (m)*	*Average travel distance (m)*	*Computation time*	*Average computation time (min)*
10	Trial 1	6143327	6132639	52	54.00
10	Trial 2	6173215		56	
10	Trial 3	6136863		55	
10	Trial 4	6077151		53	
20	Trial 1	6088479	6095215	136	138.25
20	Trial 2	6103135		139	
20	Trial 3	6105567		141	
20	Trial 4	6083679		137	
30	Trial 1	6030815	6009535	320	333.00
30	Trial 2	5996959		335	
30	Trial 3	5975711		343	
30	Trial 4	6034655		334	
40	Trial 1	5975071	5954207	436	441.25
40	Trial 2	5931103		444	
40	Trial 3	6036767		439	
40	Trial 4	5873887		446	

Table 4: *Comparisons of travel distance between CBAA and MCBH*

Approach	*Data set*	*Average travel distance for four trails*	*Improvement*
CBAA	December	6207567	Benchmark
MCBH/10%	December	6132639	1.21%
MCBH/20%	December	6095215	1.81%
MCBH/30%	December	6009535	3.19%
MCBH/40%	December	5954207	4.08%

Table 5: *The ANOVA table of the experiment with MCBH*

Source	*Sum of squares*	*df*	*Mean square*	*F value*	*p value*
Model	7.868E+010	3	2.623E+010	14.42	<0.0003
R.P.*	7.868E+010	3	2.623E+010	14.42	<0.0003
Pure error	2.183E+011	12	1.819E+009		
Total	1.005E+011	15			

*R.P. stands for reserved proportion.

Table 6: *The travel distance of allocation by ASBH*

Trial	*Data set*	*Travel distance (m)*	*Average travel distance (m)*
Trail 1	Orders of December	5418591	5399599
Trail 2	Orders of December	5381535	
Trail 3	Orders of December	5383967	
Trail 4	Orders of December	5414303	

Table 7: *Comparisons of travel distance between CBAA and ASBH*

Approach	*Data set*	*Average travel distance for four trails*	*Improvement*
CBAA	December	6207567	Benchmark
ASBH	December	5399599	13.02%

unreserved area, all locations in each aisle are available, which means that ASBH has more flexibility to assign products. On the other hand, the basic idea of CBAA has been embedded in ASBH. When a pair of products has a high WSC between each other, both of those products each have a high-order frequency individually. ASBH first selects and allocates the products with a higher association between each other to the locations near the outbound exit. From above-mentioned point of view, ASBH that keeps both the advantages of family grouping policy and CCBA is superior to MCBH in both effectiveness and efficiency. However, ASBH might not function well if the association between products is sparse. In such a situation, the maximum association mechanism of each aisle would be replaced by random selection. The effect of picking several products frequently ordered together in an aisle might be insignificant.

The experimental results demonstrate how the efficiency of order picking can be improved by using MCBH and ASBH to re-assign storage locations without requiring additional pickers or picking facilities. The proposed association-based heuristics can gather up more pairs of products frequently ordered together in the same aisle. Under this kind of assignment, the operators can pick several products with high association in an entrance of an aisle. The number of entering aisles and the number of stops for picking are reduced, and the travel distance is shortened. The newly developed association measure, WSC, applied in the two proposed heuristics, is generated from order data. When re-assigning the storage locations, changes in order patterns are considered in the new allocation. Enhancing order picking usually requires employing more pickers, purchasing superior equipment, or applying new operation processes, all of which result in more investment and cost. In this study, however, simply adjusting the allocations by association-based heuristics can reduce travel distance by up to 13%. With this significant savings, distribution centres can achieve a higher service level with a shorter picking distance and potentially reduce labour costs by needing fewer pickers.

6. Conclusion

In supply chains, the distribution centre is an essential intermediary facility between suppliers and customers. The performance of distribution centres affects the competency of supply chains. Order picking is the most time-consuming and labour-intensive activity in distribution centres. Today, order picking is more difficult because distribution centres have to store more types of products to fulfill more small orders within a shorter response time. In previous studies, the approaches used to solve SLAP seldom took product relationships into account for enhancing order picking. In this paper, a new association measurement, *WSC*, based on association rule mining, was developed. WSC represents both the intensity and nature of relationships between products. Based on WSC, two association-based heuristics, MCBH and ASBH, are proposed to solve the SLAP. MCBH combines the features of the traditional class based approach and family group policy. Locations are assigned to products partially by using the turnover rate and partially by using associations between products. Alternatively, ASBH conducts the allocation assignment to maximize the aggregated association between products allocated in the same aisle.

Both of these proposed heuristics solve the SLAP based on associations between products and contain the following features. First, the intensity and nature of relationships between products are considered in both heuristics. Second, because products frequently appearing in the same orders are picked in the same aisle, pickers can pick more products in a single aisle entrance. The number of aisle entrances and the number of stops are reduced as well. Third, changes in order patterns may be captured by using WSCs during every re-assigning. Fourth, re-assigning storage locations can enhance order-picking performance without incurring any additional overhead or capital investment.

Acknowledgement

This work was partially supported by National Science Council, Taiwan, R.O.C. under grant NSC 98-2410-H-009-009-MY2.

References

Agrawal, R., T. Imielinski and A. Swami (1993) Mining association rules between sets of items in large databases, in *Proceedings of the 1993 ACM SIGMOD Conference*, Washington, DC: ACM Press, 207–216.

Chan, F.T.S. and H.K. Chan (2011) Improving the productivity of order pocking of a manual-pick and multi-level rack distribution

warehouse through the implementation of class-based storage, *Expert System with Applications*, **38**, 2686–2700.
CHEN, M.C., C.M. CHAO and K.T. WU (2012) Pattern filtering and classification for market basket analysis with profit-base measures, *Expert Systems*, **29**(2), 170–182.
DELEN, D. (2009) Analysis of cancer data: a data mining approach, *Expert Systems*, **26**, 100–112.
EMROUZNEJAD, A. and A.L. ANOUZE (2010) Data envelopment analysis with classification and regression tree-a case of banking efficiency, *Expert Systems*, **27**, 231–246.
FRAZELLE, E.A. and G.P. SHARP (1989) Correlated assignment strategy can improve order-picking operation, *Industrial Engineering*, **4**, 33–37.
GRAVES, S.C., W.H. HAUSMAN and L.B. SCHWARZ (1977) Storage-retrieval interleaving in automatic warehousing systems, *Management Science*, **23**, 935–945.
GU, J., M. GOETSCHALCKX and L.F. MCGINNIS (2007) Research on warehouse operation: a comprehensive review, *European Journal of Operational Research*, **177**, 1–21.
HALL, R.W. (1993) Distance approximation for routing manual pickers in a warehouse, *IIE Transactions*, **25**, 76–87.
HAN, J. and M. KAMBER (2012) *Data Mining: Concepts and Techniques*, 2nd ed., San Francisco, CA: Morgan Kaufmann.
HAUSMAN, W.H., L.B. SCHWARZ and S.C. GRAVES (1976) Optimal storage assignment in automatic warehousing systems, *Management Science*, **22**, 629–638.
HESKETT, J.L. (1963) Cube-per-order index – a key to warehouse stock location, *Transport and Distribution Management*, **3**, 27–31.
HESKETT, J.L. (1964) Putting the cube-per-order index to work in warehouse layout, *Transport and Distribution Management*, **4**, 23–30.
JANE, C.C. and Y.W. LAIH (2005) A clustering algorithm for item assignment in a synchronized zone order picking system, *European Journal of Operational Research*, **166**, 489–496.
JARVIS, J.M. and E.D. MCDOWELL (1991) Optimal product layout in an order picking warehouse, *IIE Transactions*, **23**, 93–102.
KALLINA, C. and J. LYNN (1976) Application of the cube-per-order index rule for stock location in a distribution warehouse, *Interfaces*, **7**, 37–45.
KOSTER, R.D., T. LE-DUC and K.J. ROODBERGEN (2007) Design and control of warehouse order picking: a literature review, *European Journal of Operational Research*, **182**, 481–501.
LARSON, T.N., H. MARCH and A. KUSIAK (1997) A heuristic approach to warehouse layout class-based storage, *IIE Transactions*, **29**, 337–348.
LEE, M.K. (1992) A storage assignment policy in a man-on-board automated storage/retrieval system, *International Journal of Production Research*, **30**, 2281–2292.
LIOU, C.H. and D.R. LIU (2012) Hybrid recommendations for mobile commerce based on mobile phone features, *Expert Systems*, **29**(2), 108–123.
LIU, C.M. (1999) Clustering techniques for stock location and order-picking in a distribution center, *Computer and Operations Research*, **26**, 989–1002.
MANZINI, R. (2006) Correlated storage assignment in an order picking system, *International Journal of Industrial Engineering*, **13**, 384–394.
MANZINI, R., M. GAMBERI, A. PERSONA and A. REGATTIERI (2007) Design of a class based picker to product order picking system, *The International Journal of Advanced Manufacturing Technology*, **32**, 811–821.
MUPPANI, V.R. and G.K. ADIL (2008a) Efficient formation of storage classes for warehouse storage location assignment: a simulated annealing approach, *The International Journal of Management Science*, **36**, 609–618.
MUPPANI, V.R. and G.K. ADIL (2008b) A branch and bound algorithm for class based storage location assignment, *European Journal of Operational Research*, **189**, 492–507.
RENAUD, J. and A. RUIZ (2008) Improving product location and order picking activities in a distribution centre, *Journal of the Operational Research Society*, **59**, 1603–1613.
ROODBERGEN, K.J. and R.D. KOSTER (2001) Routing methods for warehouses with multiple cross aisles, *International Journal of Production Research*, **39**, 1865–1883.
ROSENWEIN, M.B. (1994) An application of cluster analysis to the problem of locating items within a warehouse, *IIE Transactions*, **26**, 101–103.
SRIKANT, R. and R. AGRAWAL (1997) Mining generalized association rules, *Future Generation Computer System*, **13**, 161–180.
TOMPKINS, J.A., J.A. WHITE, Y.A. BOZER, E.H. FRAZELLE, J.M.A. TANCHOCO and J. TREVINO (1996) *Facilities Planning*, 2nd ed., New York: Wiley & Sons Inc.
VAN DEN BERG, J.P. (1996) Class-based storage allocation in a single-command warehouse with space requirement constraints, *International Journal of Industrial Engineering*, **3**, 21–28.
VAN OUDHEUSDEN, D.L., Y.J.J. TZEN and H.T. KO (1988) Improving storage and order picking in a person-on-board AS/R system, *Engineering Costs and Production Economics*, **13**, 273–283.
VAN OUDHEUSDEN, D.L. and W. ZHU (1992) Storage layout of AS/RS racks based on recurrent orders, *European Journal of Operational Research*, **58**, 48–56.
WU, D. and D.L. OLSON (2009) Enterprise risk management: small business scorecard analysis, *Production Planning & Control*, **120**, 362–269.
XIAO, J. and L. ZHENG (2010) A correlated storage location assignment problem in a single-block-multi-aisles warehouse considering BOM information, *International Journal of Production Research*, **48**, 1321–1338.
YEH, J.Y. and T.H. WU (2010) Cascade of genetic algorithm and decision tree for cancer classification on gene expression data, *Expert Systems*, **27**, 201–218.

The authors

David Ming-Huang Chiang

David Ming-Huang Chiang is a Professor of Operations Management and Logistics in the Department of Business Administration at The National Taiwan University, Taipei, Taiwan. He earned his PhD in management science from The University of Iowa in 1992. Dr. Chiang's research interests include supply chain management, production scheduling, and inventory management. His work has been published in *Annals of Operations Research*, *Journal of Management and System*, *BJU International*, *Journal of Management*.

Chia-Ping Lin

Chia-Ping Lin is a PhD candidate of Operations Management and Logistics in the Department of Business Administration at The National Taiwan University, Taipei, Taiwan. He received his MSc degree in Institute of Commerce Automation and Management (ICAM) from National Taipei University of Technology and BS degree in Business Administration from Fu Jen Catholic University.

Mu-Chen Chen

Mu-Chen Chen is a Professor of Institute of Traffic and Transportation in National Chiao Tung University, Taipei, Taiwan. He received his PhD and MSc degrees both in Industrial Engineering and Management from National Chiao Tung University, and his BS degree in Industrial Engineering from Chung Yuan Christian University. His teaching and research interests include Data Mining, Logistics and Supply Chain Management and Meta-heuristics.