International Journal of Production Research
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tprs20

A data mining approach to optimise shelf space allocation in consideration of customer purchase and moving behaviours

Chieh-Yuan Tsai[ab] & Sheng-Hsiang Huang[a]

[a] Department of Industrial Engineering and Management, Yuan-Ze University, Chung-Li, Taiwan

[b] Innovation Center for Big Data and Digital Convergence, Yuan-Ze University, Chung-Li, Taiwan

Published online: 16 Jul 2014.

To cite this article: Chieh-Yuan Tsai & Sheng-Hsiang Huang (2015) A data mining approach to optimise shelf space allocation in consideration of customer purchase and moving behaviours, International Journal of Production Research, 53:3, 850-866, DOI: 10.1080/00207543.2014.937011

To link to this article: http://dx.doi.org/10.1080/00207543.2014.937011

International Journal of Production Research, 2015
Vol. 53, No. 3, 850–866, http://dx.doi.org/10.1080/00207543.2014.937011

A data mining approach to optimise shelf space allocation in consideration of customer purchase and moving behaviours

Chieh-Yuan Tsai[a,b]* and Sheng-Hsiang Huang[a]

[a]*Department of Industrial Engineering and Management, Yuan-Ze University, Chung-Li, Taiwan;* [b]*Innovation Center for Big Data and Digital Convergence, Yuan-Ze University, Chung-Li, Taiwan*

(*Received 10 November 2013; accepted 14 June 2014*)

A good shelf space allocation strategy can help customers easily find product items and dramatically increase store profit. Previous studies generally relied on the space elasticity formula to optimise space allocation models, but space elasticity requires estimates of many parameters, resulting in high costs and frequent errors in the mathematical models. In this study, a three-stage data mining method is proposed for solving the shelf space allocation problem with consideration of both customer purchase and moving behaviours. In the first stage, the customer's purchasing behaviour is derived from records of previous transactions, while moving behaviour is collected through radio frequency identification systems. In the second stage, the A priori algorithm is applied to obtain frequent product association rules from purchase transactions. In addition, the $UMSP_L$ algorithm is adopted to derive high-utility mobile sequential patterns from customer mobile transaction sequences. In the third stage, all product items are classified as either major, minor or trivial according to a set of criteria. A Location preference evaluation procedure is then developed to calculate location preference if a minor item is placed at a given section of the store. Based on the location preference matrix, minor items are reassigned to optimal shelves. The experimental results show the proposed method can reassign items to suitable shelves and dramatically increase cross-selling opportunities for major and minor items.

Keywords: shelf space allocation; association rules; high-utility mobile sequential patterns; customer purchase and moving behaviour

1. Introduction

To attract customers and survive in a competitive environment, retailers need to implement appropriate retail-mix strategy, the elements of which include store location, product assortment, pricing, advertising and promotion, store design and shelf display, services and personal selling (Levy and Weitz 1995). Among these, shelf space allocation is seen as one of the most important factors in customer purchasing decisions without incurring additional costs on the part of the retailer (Yang 1999). In literatures, space elasticity has been widely used to estimate the relationship between sales and allocated space (Hansen and Heinsbroek 1979; Corstjens and Doyle 1981; Yang and Chen 1999; Lim, Rodrigues, and Zhang 2004). These efforts used space elasticity to construct the relationship between shelf space and product demand, and solved it with different optimisation techniques. However, using space elasticity for shelf space allocation requires estimating a great number of parameters, resulting in high costs and high error rates in the mathematical models (Borin, Farris, and Freeland 1994).

Recently, advances in information technology have made it easier for retailers to collect various types of customer data (Song and Kusiak 2009). Mining this data for insight into customer behaviour can help retailers to solidify ephemeral relationships with customers into long-term loyalty. As part of this effort, retailers have modified their shelf space management using product association analysis (Agrawal and Srikant 1994; Chen, Chen, and Tung 2006; Chen and Lin 2007; Cil 2012) which is a powerful data mining tool that can detect significant co-purchase rules underlying a large amount of purchase transaction data. Although these studies demonstrated that product association analysis can improve the efficiency of shelf space usage, they simply assumed that customer purchase behaviour is consistent before and after shelf space rearrangement. However, this assumption may not be valid if the products the customers are looking for have been moved from their original locations or shelves. That is, the purchase patterns/rules may change if the locations of all products in a store are reorganised. Failure to consider this factor makes the shelf space allocation results impractical.

*Corresponding author. Email: cytsai@saturn.yzu.edu.tw

When shopping in a store, a customer moves around the aisles, stops at certain locations, deliberates about various purchase options and makes certain selections. This process is repeated until the shopping trip is complete. Larson, Bradlow, and Fader (2005) used a multivariate clustering algorithm to explore the paths taken by individual shoppers in an actual grocery store, with data collected through RFID (radio frequency identification) tags installed on their shopping carts. Liao and Lin (2007) proposed an RFID deployment model to track shopping paths and purchased items by individual customers. Hui, Bradlow, and Fader (2009) developed an individual-level integrated Bayesian model for store shopping trips to capture the relationship between a consumer's shopping path through the store and his/her purchasing behaviour. While these studies used RFID technology to efficiently track shopping paths, none discusses the application of this knowledge to solve the shelf space allocation problem (Irani, Gunasekaran, and Dwivedi 2010).

This paper proposes a solution to the shelf space allocation problem using a data mining approach which considers both customer purchase and moving behaviours. The proposed approach consists of three main stages: data collection and pre-processing, data mining and product-to-shelf assignment. To the best of our knowledge, this paper is the first work to integrate purchasing records and moving records to solve shelf space allocation problems. In addition, this study develops the concept that product items in the store are classified as major items, minor items and trivial items according to a set of regulations. Only minor items are relocated to ensure customers can follow their preferred shopping paths. Thus, this paper develops a Location Preference Evaluation (LPE) procedure to determine the location preference for minor items. The remainder of this paper is organised as follows. Related works are reviewed in Section 2. Section 3 formally defines the research problem and introduces important components of the proposed approach. Section 4 presents an empirical evaluation and describes a set of experiments. Finally, conclusions are drawn and future work directions are proposed in Section 5.

2. Literature review

Shelf space is an essential resource in logistics decisions and store management since well-designed space allocation can attract more consumers (Yang and Chen 1999). Space elasticity is the ratio of relative changes in unit sales to relative change in shelf space (Curhan 1972). The measurement of space elasticity can be divided into direct elasticity (main effect) and cross-elasticity (cross-effect). Hansen and Heinsbroek (1979) used space elasticity to estimate product demand and constructed optimisation models for product selection and allocation. Corstjens and Doyle (1981) extended this concept to develop a model which considers both space-elasticity and cross-elasticity. Yang and Chen (1999) disregarded cross-elasticities and allowed the profit of each product to vary when allocated to different shelves by formulating their shelf space allocation problem. This model was then optimised by Lim, Rodrigues, and Zhang (2004) through combining a local search technique with meta-heuristics. However, using space elasticity to determine product assortment and shelf space allocation requires the estimation of a great quantity of parameters, resulting in increased costs and errors in the mathematical model. In addition, some research used price policy to consider product items location (Murray, Talukdar, and Gosavi 2010; Nafari and Shahrabi 2010) and explored the product categories allocation problem (Chen 2012). However, these results of research do not consider the customer purchase and moving behaviour.

With the novel information technology, retailers can solidify ephemeral relationships with customers into long-term and fruitful relationships if they can discover customer behaviour from collected data. Data mining is one of the most popular technologies that discover potential customer knowledge from business databases to assist a policy decision. In fact, if we can know what product items are frequently purchased together, the cross-selling opportunity will dramatically increase (Agrawal and Srikant 1994). Chen, Chen, and Tung (2006) proposed a novel representation scheme and developed a robust algorithm based on association analysis to discover implicit, yet meaningful, relationships between the relative spatial distance of displayed products and unit sales. Chen and Lin (2007) applied multi-level association rule mining to explore relationships between products as well as between product categories to optimise product assortments and allocation in retail contexts. Cil (2012) used buying association measurements to create a category correlation matrix and applied the multi-dimensional scale technique to display a set of products in store spaces.

In recent years, mining user behaviours in mobile environments has emerged as an interesting topic in the field of pattern mining. Lee et al. (2007) proposed the T-Map algorithm to find temporal mobile access patterns that can efficiently discover temporal behaviour patterns of mobile users associated with location and requested services in different time intervals. Yun and Chen (2007) proposed the TJ_{PF} algorithm to mine mobile sequential patterns through considering both the movement and purchase patterns of customers. Tseng, Lu, and Huang (2007) proposed the TMSP-Mine algorithm to efficiently generate the most appropriate time segmentation intervals based on the fitness function of a genetic algorithm, while simultaneously mining mobile sequential patterns associated with moving paths and time intervals in location-based service (LBS) environments. On the other hand, different customer groups may produce mobile transaction sequences for different behaviours in mobile environments. Lu and Tseng (2009) proposed the Cluster-based

Mobile Sequential Pattern (CMSP)-Mine algorithm to efficiently discover CMSPs associated with moving paths and customer groups in LBS environments.

Recently, utility mining is one of the most important research issues in data mining fields due to its ability to consider the non-binary frequency values of items in transactions and different profit values for every item. Chan, Yang, and Shen (2003) first proposed the concept of utility to capture highly desirable statistical patterns and level-wise item set mining algorithm in association mining. Tseng et al. (2010) proposed a novel algorithm named UP-Growth to mine high-utility item sets with a set of techniques for pruning candidate item sets. Shie, Tseng, and Yu (2010) proposed GUIDE algorithm which can find temporal maximal utility item sets from data streams. They utilised TMUI-tree structure to capture the utility of each item set with one-time scanning. Shie, Hsiao, and Tseng (2013) first proposed level-wise and tree-based algorithms which integrate sequential purchasing patterns with the moving paths and utility mining to discover high-utility mobile sequential patterns in mobile environments.

3. Proposed method

As shown in Figure 1, the proposed shelf space allocation method consists of three main stages: data collection and pre-processing, data mining and product-to-shelf assignment. The detail description for each stage is introduced in the following sections.

3.1 *Data collection and pre-processing*

Figure 2 provides a sample store layout for consideration. In the figure, a cabinet contains six shelves for product placement. RFID readers are deployed in appropriate positions to partition the aisles into several sections. The grey oval in Figure 2 indicates the coverage area of an RFID reader. Let $I = \{i_1, i_2, \ldots, i_g\}$ be the set of all product items sold in the store, $Z = \{z_1, z_2, \ldots, z_y\}$ be the set of all shelves in the store, $S = \{S_1, S_2, \ldots, S_n\}$ be the set of sections in the store, and $R = \{R_1, R_2, \ldots, R_m\}$ be the set of RFID readers.

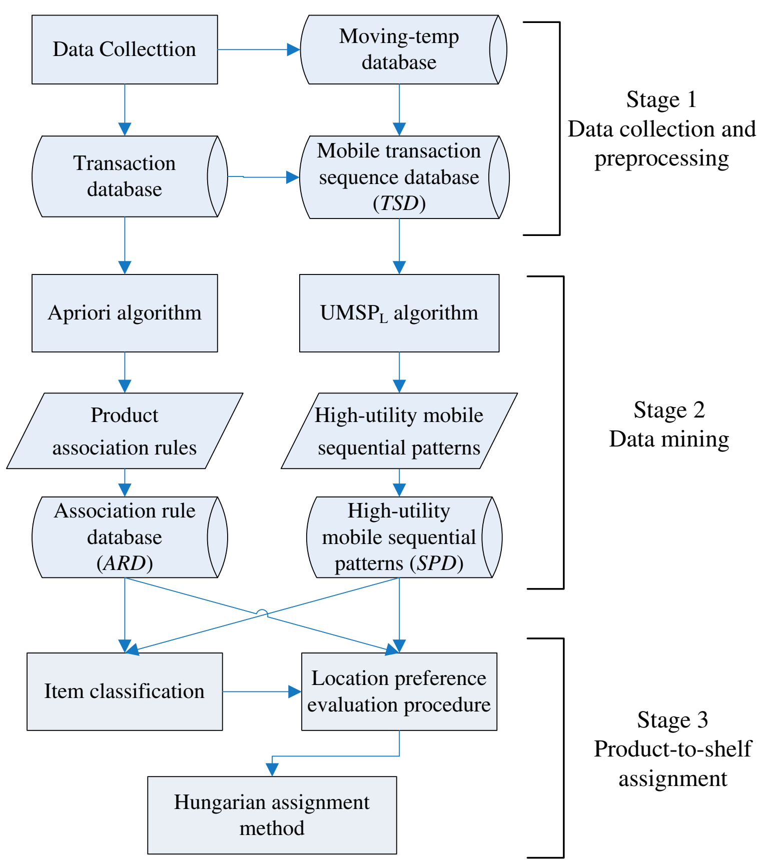

Figure 1. Proposed shelf space allocation method.

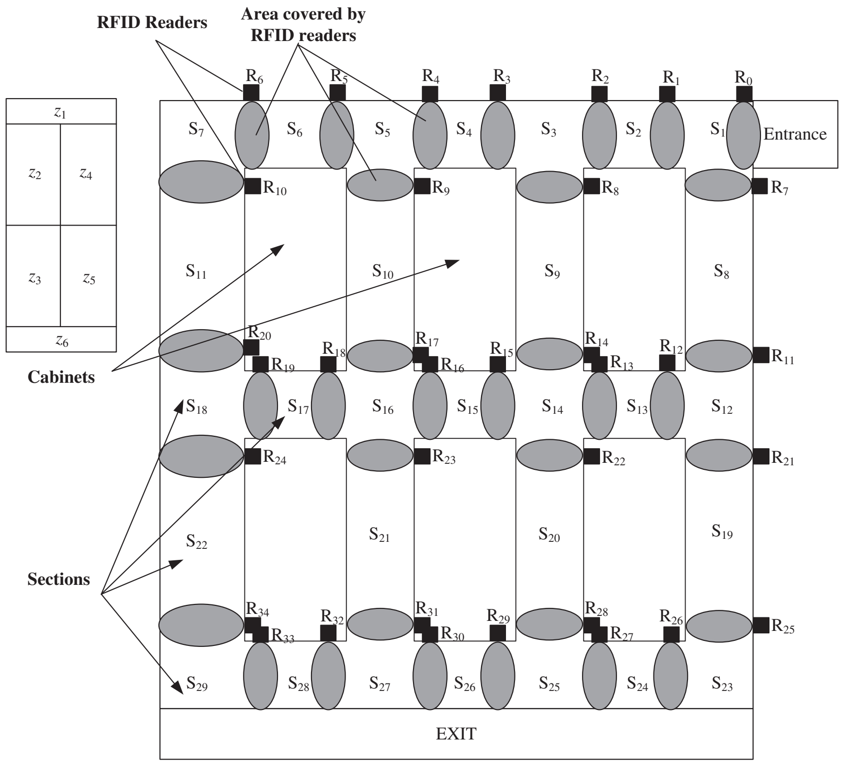

Figure 2. Layout notations.

To collect behaviour data for each customer, shopping carts are equipped with active RFID tags. When a cart passes through the coverage area of an RFID reader, the passing event with the form of (rid_i, t_i) will be recorded into the moving-temp database where $rid_i \in R$ and t_i is the timestamp. The system keeps track of the shopping carts as they pass in and out of the coverage areas of the various readers until the customer completes his/her shopping and cheques out at a counter. After the customer leaves the checkout counter, his/her purchase transactions are stored into the transaction database in the form of $\{[it_1, q_1], [it_2, q_2], \ldots, [it_h, q_h]\}$ where q_i is the purchased quantity of item $it_i \in I$. At the same time, the customer's moving path with the form of $<rid_1, rid_2, \ldots, rid_p>$ is retrieved from the moving-temp database where $t_1 < t_2 < \ldots < t_p$. Next, the moving path represented by readers $<rid_1, rid_2, \ldots, rid_p)>$ is transformed into a moving sequence represented by sections $<sid_1, sid_2, \ldots, sid_p>$ where $sid_p \in S$ according to the following transformation procedure.

With the purchase transaction and moving sequence, a customer's mobile transaction sequence can be derived and defined as $TS = <T_1, T_2, \ldots, T_n>$ where transaction $T_j = (Se_j; \{[it_{j1}, q_{j1}], [it_{j2}, q_{j2}], \ldots, [it_{jh}, q_{jh}]\})$ represents that a customer purchased $\{[it_{j1}, q_{j1}], [it_{j2}, q_{j2}], \ldots, [it_{jh}, q_{jh}]\}$ in section Se_j where q_{jp} is the purchased quantity of pth item $it_{jp} \in I$ in transaction T_j. A path is denoted as $Se_1Se_2 \ldots Se_r$, where $Se_j \in S$ and $1 \leq j \leq r$. All mobile transaction sequences (*TSs*) are stored in the mobile transaction sequence database (*TSD*).

Example I

Table 1 shows five mobile transaction sequences (*TSs*) in the mobile *TSD*. For example, customer CID 1 bought one item i_5 in section S_2, five items i_{11} in section S_4 and six items i_{58} in section S_{20}. Customer CID 2 bought five items i_{22} and three items i_{24} in section S_8, five items i_{38} in section S_{13} and so on.

3.2 *Product association rules and mining*

Association rule mining is used to discover the rules by which the presence of one set of items implies the presence of another set of items from a given transaction database. The form of a rule can be represented as $X \rightarrow Y$ where X and Y are, respectively, the antecedent and consequent parts of the rule. In association rule mining, a rule is called strong if it simultaneously satisfies a user-specified minimum support and a user-specified minimum confidence. The support of a

Table 1. Mobile transaction sequences in the *TSD*.

CID	Mobile transaction sequence
1	<(S_1; Ø), (S_2; {[i_5,1]}), (S_3; Ø), (S_4; {[i_{11},5]}), (S_3; Ø), (S_9; Ø), (S_{14}; Ø), (S_{20}; {[i_{58},6]}), (S_{25}; Ø)>
2	<(S_1; Ø), (S_8; {[i_{22},5],[i_{24},3]}), (S_{12}; Ø), (S_{13}; {[i_{38},5]}), (S_{14}; Ø), (S_{15}; {[i_{43},2]}), (S_{16}; Ø), (S_{21}; {[i_{62},3]}), (S_{27}; Ø)>
3	<(S_1; Ø), (S_8; {[i_{22},3]}), (S_{12}; Ø), (S_{13}; {[i_{38},8]}), (S_{14}; Ø), (S_{15}; {[i_{43},3]}), (S_{16}; Ø), (S_{17}; {[i_{50},3]}), (S_{16}; Ø), (S_{21}; {[i_{62},2]}), (S_{27}; Ø)>
4	<(S_1; Ø), (S_2; {[i_5,1]}), (S_3; Ø), (S_4; {[i_{11},3]}), (S_5; Ø), (S_{10}; Ø), (S_{16}; Ø), (S_{15}; {[i_{43},3]}), (S_{14}; Ø), (S_{20}; {[i_{58},3]}), (S_{25}; Ø)>
5	<(S_1; Ø), (S_8; {[i_{22},3],[i_{24},2]}), (S_{12}; Ø), (S_{13}; {[i_{38},6]}), (S_{14}; Ø), (S_{15}; {[i_{43},2]}), (S_{16}; Ø), (S_{21}; {[i_{62},3]}), (S_{27}; Ø)>

rule is defined as the number of transactions including all items in the antecedent and consequent parts of the rule, while the confidence of a rule is a percentage value that shows how frequently the antecedent part occurs among all transactions containing the rule.

This research uses the Apriori algorithm (Agrawal and Srikant 1994), one of the most popular frequent pattern mining methods. The input to the Apriori algorithm is the transaction database and minimum support, while the output is the set of all frequent item sets. In the algorithm, L_k denotes the set of all frequent k-item sets and C_k denotes the set of candidate k-item sets. In the first phase, L_1 is derived from C_1 where C_1 can be generated by listing all distinct items in the database. Next, the infrequent item set results in L_1 are removed. The kth phase produces L_k. Finally, if no further frequent item sets are generated, the algorithm will stop and return all values of L_k. In this research, only 'single item to singe item' rules are needed. Therefore, this research will stop at L_2 in the Apriori algorithm and generate rules based on the provided minimum confidence. All association rules then are stored in the product association rule database (*ARD*).

3.3 *High-utility mobile sequential patterns and mining*

3.3.1 *Definitions for high-utility mobile sequential patterns*

The high-utility mobile sequential pattern defined in Shie, Hsiao, and Tseng (2013) is adopted and modified in this paper. A section-item, denoted as $<Se_j; it_{je}>$, stands for item it_{je} purchased in the section s_j, where $Se_j \in S$ and $it_{je} \subseteq I$. The utility of a section-item $<Se_j; it_{je}>$ in a mobile transaction sequence TS_q is defined as

$$u\left(<Se_j; it_{je}>, TS_q\right) = q_{je} \times w(it_{je}) \tag{1}$$

where $w(it_{je})$ is the unit profit of item it_{je}, which is recorded in a utility table. Similarly, a section-item set, denoted as $<Se_j; \{it_1, it_2, \ldots, it_h\}>$, stands for the item set $\{it_1, it_2, \ldots, it_h\}$ that occurred in Se_j, where $Se_j \in S$ and $\{it_1, it_2, \ldots, it_h\} \subseteq I$. The utility of a section-item set $<Se_j; \{it_{j1}, it_{j2}, \ldots, it_{jh}\}>$ in TS_q is denoted as $u(<Se_j; \{it_{j1}, it_{j2}, \ldots, it_{jh}\}>, TS_q)$ and defined as $\sum_{k=1}^{h} u(<Se_j; it_{jk}>, TS_q)$. Moreover, the utility of section-item set $Y = <Se_j; \{it_{j1}, it_{j2}, \ldots, it_{jh}\}>$ in the mobile *TSD* is denoted as

$$u(Y) = u(<Se_j; it_{j1}, it_{j2}, \ldots, it_{jh}>) = \sum_{(Y \subseteq TS_q) \wedge (TS_q \in TSD)} u(Y, TS_q) \tag{2}$$

A section-pattern X is a list of section-item sets, denoted as $<Se_1;\{it_{11}, it_{12}, \ldots, it_{1h_1}\}><Se_2;\{it_{21}, it_{22}, \ldots, it_{2h_2}\}>\ldots<Se_m; \{it_{m1}, it_{m2}, \ldots, it_{mh_m}\}>$. The utility of a section-pattern X in TS_q is denoted as $u(X, TS_q)$ and defined as $\sum_{\forall Y \in X} u(Y, TS_q)$. The utility of a section-pattern X in *TSD* is denoted as

$$u(X) = \sum_{(X \subseteq TS_q) \wedge (TS_q \in TSD)} u(X, TS_q) \tag{3}$$

A moving pattern is composed of a section-pattern and a path. It is recorded in the form as $<\{<Se_1; \{it_{11}, it_{12}, \ldots, it_{1h_1}\}><Se_2; \{it_{21}, it_{22}, \ldots, it_{2h_2}\}> \ldots <Se_m; \{it_{m1}, it_{m2}, \ldots, it_{mh_m}\}>\}; Se_1Se_2 \ldots Se_m>$. The support of a moving pattern P, denoted as $sup(P)$, is defined as the number of mobile transaction sequences which contain P in *TSD*. On the other hand, the utility of a moving pattern P, denoted as $u(P)$, is defined as the summation of utilities of the section-patterns of P in the mobile transaction sequences which contain the path of P in *TSD*. Given a minimum support threshold δ and a minimum utility threshold ε, a moving pattern P is called a *mobile sequential pattern* if $\sup(P) \geq \delta$. P is called a *high-utility mobile sequential pattern*, abbreviated as *UMSP*, if $\sup(P) \geq \delta$ and $u(P) \geq \varepsilon$. The length of a pattern is the number of section-item sets it contains. A pattern with a length k is denoted as k-pattern.

The sequence utility of a mobile transaction sequence TS_q, which is the sum of the utilities of all items in TS_q, is defined as

$$SU(TS_q) = \sum_{<Se_n;it_{je}> \subseteq TS_q} u\left(<Se_j; it_{je}>, TS_q\right) \tag{4}$$

The *sequence-weighted utilisation*, abbreviated as *SWU*, of a section-item set Y is defined as

$$SWU(Y) = \sum_{(Y \subseteq TS_q) \wedge (TS_q \in TSD)} SU(TS_q). \tag{5}$$

Moreover, the sequence-weighted utilisation of a section-pattern X is defined as

$$SWU(X) = \sum_{(X \subseteq TS_q) \wedge (TS_q \in TSD)} SU(TS_q) \tag{6}$$

In addition, the sequence-weighted utilisation of a moving pattern P is defined as the summation of *SWUs* of the section-patterns of P in the mobile transaction sequences which contain the path of P in *TSD*. A pattern Z is called a *high sequence-weighted utilisation pattern*, abbreviated as *WUP*, if $\sup(Z) \geq \delta$ and $SWU(Z) \geq \varepsilon$. In addition, *high sequence-weighted utilisation section-item set* is abbreviated as *WULI*, *high sequence-weighted utilisation section-pattern* is abbreviated as *WULP*, and *high sequence-weighted utilisation mobile sequential pattern* is abbreviated as *WUMSP*.

3.3.2 $UMSP_L$ algorithm

This paper adopts the $UMSP_L$ (high-Utility Mobile Sequential Pattern by a Level-wised method) algorithm proposed by Shie, Hsiao, and Tseng (2013) to obtain high-utility mobile sequential patterns. The $UMSP_L$ algorithm consists of four steps. The input of the $UMSP_L$ algorithm includes a mobile *TSD*, a pre-defined utility table, a minimum support threshold (δ) and a minimum utility threshold (ε). The first three steps find *WUMSPs* based on the sequence-weighted downward closure property, while the fourth step finds high-utility mobile sequential patterns (*UMSPs*). In step 1, the mobile *TSD* is scanned several times to generate all *WULIs*, each of which is mapped to a specific identity in a mapping table. Note that the mapped *WULIs* are 1-*WULPs*. In step 2, the mobile *TSD* is transformed into a trimmed database (*TD*) by mapping the *WULIs* to their new identities. The section-items which cannot be elements of high-utility mobile sequential patterns are removed from the database. In step 3, the trimmed database (*TD*) is used to find the *WUMSPs* by the proposed level-wise method. In step 4, the *WUMSPs* are checked to find high-utility mobile sequential patterns (*UMSPs*) through an additional scan of the mobile *TSD*. The *WUMSPs* whose utilities are greater than or equal to ε are regarded as *UMSPs*. All *UMSPs* are then stored in the high-utility mobile sequential pattern database (*SPD*).

3.4 *Product item classification*

As mentioned in Section 1, not all items are suitable for rearrangement since previous popular shopping behaviours might not be consistent if the major attractive items are moved from their previous locations. Therefore, a product item in this study is classified as a *major item*, *minor item* or *trivial item* based on the high-utility mobile sequential patterns and product association rules derived from the store.

Definition 1: A *major item* is an item that has previously appeared in high-utility mobile sequential patterns in the high-utility mobile sequential pattern database (*SPD*). Major items are considered to be top-selling commodities. If major items are reassigned to other section(s)/shelves(s), the high-utility mobile sequential patterns might be invalid because shoppers may not be able to find items they planned to purchase. Therefore, the positioning of major items should not be changed. In the following discussion, the set of major items is denoted as *MA*.

Definition 2: A *minor item* is an item which can be relocated. Minor items are considered to be commodities affiliated with major items. Minor items can be found according to the following rules. First, all association rules in the product *ARD* are checked. If the item on the antecedent part of an association rule is a major item, the item on the consequent part of the association rule is a candidate minor item. Next, if the candidate minor item is not a major item, the item is a minor item. The set of minor items is denoted as *MI*.

Definition 3: A *trivial item* is an item which is neither a major item nor a minor item. Trivial items are items which do not typically bring in much revenue for the store.

Example II

To explain the generation of the set of major items *MA* and the set of minor items *MI*, the high-utility mobile sequential patterns in Table 2 and the product association rules in Table 3 are used as an example. Table 2 shows three high-utility mobile sequential patterns $<\{<S_8;\ i_{22}><S_{21};\ i_{62}>\};\ S_8S_{12}S_{13}S_{14}S_{15}S_{16}S_{21}>$, $<\{<S_8;\ i_{22}><S_{13};\ i_{38}><S_{21};\ i_{62}>\};\ S_8S_{12}S_{13}S_{14}S_{15}S_{16}S_{21}>$ and $<\{<S_8;\ i_{22}><S_{13};\ i_{38}><S_{15};\ i_{43}><S_{21};\ i_{62}>\};\ S_8S_{12}S_{13}S_{14}S_{15}S_{16}S_{21}>$. According to the definition of a major item, $MA = \{i_{22}, i_{38}, i_{43}, i_{62}\}$ can be identified. Next, based on the definition of a minor item, items i_8, i_{25}, i_{38}, i_{43}, i_{52}, i_{62} and i_{70} are candidate minor items since the items on the consequent part of all association rules belong to major items. However, items i_{38}, i_{43} and i_{62} are already known to be major items. Therefore, the set of minor items is $MI = \{i_8, i_{25}, i_{52}, i_{70}\}$.

3.5 *Product-to-shelf assignment*

According to definition 1, major items should not be relocated from their original positions because doing so would disrupt frequent customer-visiting behaviours (i.e. high-utility mobile sequential patterns). In addition, trivial items do not significantly contribute to store revenue. Therefore, only minor items will be rearranged.

3.5.1 *Location preference evaluation*

To increase their cross-sale potential, minor items should be moved to locations as close as possible to their corresponding major items according to previous customer purchase and moving behaviours. In this study, customer purchase and moving behaviours are, respectively, represented as the product association rules in the product *ARD* and high-utility mobile sequential patterns in the high-utility mobile sequential pattern database (*SPD*). From these two databases, this study develops a LPE procedure to calculate location preference if a minor item is placed in a certain section in the store.

First, the procedure scans the product *ARD* and retrieves all major items on the consequent part of a rule while the antecedent part of the rule is mi_j. The set of major items related to mi_j is denoted as GM_j. Then, for each major item ma_k in GM_j, the procedure scans the high-utility mobile sequential pattern database (*SPD*) to find the set of high-utility mobile sequential patterns containing ma_k, denoted as GP_{jk}. For each high-utility mobile sequential pattern $UMSP_m$ in GP_{jk}, the procedure will evaluate the *movement distance* for which minor item mi_j is assigned to section s_n. Let $D^{j,n}_{k,m}$ be the movement distance in $UMSP_m$ if mi_j is moved from the section in which major item ma_k is located to section s_n. If no relationship exists between minor item mi_j, major item ma_k and section s_n, provided that high-utility mobile sequential pattern $UMSP_m$ can be found, $D^{j,n}_{k,m}$ is set as β. Note that β is the threshold of maximum section movement and is provided by users.

Table 2. Sample high-utility mobile sequential patterns.

Pattern ID	High-utility mobile sequential pattern (*UMSP*)
1	$<\{<S_8;\ i_{22}><S_{21};\ i_{62}>\};\ S_8S_{12}S_{13}S_{14}S_{15}S_{16}S_{21}>$
2	$<\{<S_8;\ i_{22}><S_{13};\ i_{38}><S_{21};\ i_{62}>\};\ S_8S_{12}S_{13}S_{14}S_{15}S_{16}S_{21}>$
3	$<\{<S_8;\ i_{22}><S_{13};\ i_{38}><S_{15};\ i_{43}><S_{21};\ i_{62}>\};\ S_8S_{12}S_{13}S_{14}S_{15}S_{16}S_{21}>$

Table 3. Product association rules.

Rule ID	Association rule	Rule ID	Association rule
1	$i_{22} \rightarrow i_{25}$	6	$i_{43} \rightarrow i_{52}$
2	$i_{22} \rightarrow i_{38}$	7	$i_{43} \rightarrow i_{62}$
3	$i_{38} \rightarrow i_8$	8	$i_{62} \rightarrow i_{38}$
4	$i_{38} \rightarrow i_{25}$	9	$i_{62} \rightarrow i_{52}$
5	$i_{38} \rightarrow i_{43}$	10	$i_{62} \rightarrow i_{70}$

Example III

Take high-utility mobile sequential pattern $UMSP_2 = <\{<S_8;\ i_{22}><S_{13};\ i_{38}><S_{21};\ i_{62}>\};\ S_8S_{12}S_{13}S_{14}S_{15}S_{16}S_{21}>$ in Table 2 as an example. Assume there are eight sections of S_8, S_{12}, S_{13}, S_{14}, S_{15}, S_{16}, S_{21} and S_{25} in the store. According to association rule 3 in Table 3, we know that minor item i_8 is related to major item i_{38}. Since minor item i_8 is already located in section S_{13}, according to $UMSP_2$ it does not need to be relocated. That is, movement distance $D^{i_8,S_{13}}_{i_{38},UMSP_2} = 0$. If minor item i_8 is placed in section S_{12} or S_{14}, $D^{i_8,S_{12}}_{i_{38},UMSP_2} = D^{i_8,S_{14}}_{i_{38},UMSP_2} = 1$ since the movement distance from S_{13} to section S_{12} or S_{14} is 1 according to the path information ($S_8S_{12}S_{13}S_{14}S_{15}S_{16}S_{21}$) in $UMSP_2$. Similarly, $D^{i_8,S_8}_{i_{38},UMSP_2} = D^{i_8,S_{15}}_{i_{38},UMSP_2} = 2$, $D^{i_8,S_{16}}_{i_{38},UMSP_2} = 3$ and $D^{i_8,S_{21}}_{i_{38},UMSP_2} = 4$. Finally, $D^{i_8,S_{25}}_{i_{38},UMSP_2} = \beta$ since section S_{25} does not appear in $UMSP_2$.

If $D^{j,n}_{k,m}$ is close to 0, minor item mi_j should have a high likelihood of being relocated to section s_n. Therefore, the standardisation of assigning mi_j to s_n under the condition of high-utility mobile sequential pattern $UMSP_m$ and major item ma_k is defined as:

$$W^{j,n}_{k,m} = \frac{D^{j,n}_{k,m}}{\beta}, \quad 0 \le D^{j,n}_{k,m} \le \beta \tag{7}$$

where β is the threshold of maximum section movement and $0 \le W^{j,n}_{k,m} \le 1$.

Next, every minor item might have one or more related major item(s). Therefore, Equation (7) is applied for all major items related with minor item mi_j. Equation (8) defines the weight of assigning mi_j to s_n under the condition of $UMSP_m$.

$$A^{j,n}_m = \frac{\sum_{k \in GM_j} W^{j,n}_{k,m}}{|GM_j|} \tag{8}$$

where $GM_j$is the set of major items in MA related to mi_j. Next, Equation (8) is applied for all high-utility mobile sequential patterns, so that the weight of mi_j is placed on s_n, and $F_{j,n}$ can be obtained as:

$$F_{j,n} = \frac{\sum_{m=1}^{|SPD|} A^{j,n}_m}{|SPD|} \tag{9}$$

where $|SPD|$ is total number of $UMSP$s in the high-utility mobile sequential pattern database (SPD). Although $F_{j,n}$ can be derived for all minor items mi_j and section s_n, a section might contain many shelves. Therefore, based on $F_{j,n}$, the location preference weight for item j is placed on shelf z_y, and $f_{j,y}$ can be derived according to Equation (10).

$$f_{j,y} = \begin{cases} F_{j,n} & \text{if } z_y \in St_n \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where St_n is set of shelves on section s_n. That is, if shelf z_y belongs to St_n, $f_{j,y}$ will be $F_{j,n}$; otherwise, $f_{j,y}$ will be 0. Figure 3 shows the pseudo-code for the LPE procedure. The input of the procedure is from the high-utility mobile sequential patterns database (SPD), the product ARD, major item set (MA) and minor item set (MI), while the output is the location preference matrix $[f_{j,y}]$.

3.5.2 *Objective function*

This research assumes that every product item has the same size and quantity so that two minor items on different shelves can be easily exchanged. The reassignment considers the location preference matrix $[f_{j,y}]$ and reassigns products to the most suitable shelves. Therefore, the objective of product rearrangement is to rearrange minor items while keeping total shelf movement to a minimum:

$$\text{MIN} \sum_{j}^{\text{J}} \sum_{y}^{\text{Y}} (E_{jy} \cdot f_{jy}) \tag{11}$$

Subject to

$$\sum_{y}^{\text{Y}} E_{jy} = 1, \quad j = 1, 2, \ldots, J$$

```
Input: High-utility mobile sequential patterns database (SPD), Products association rule
database (ARD), Major item set (MA), Minor item set (MI)
Output: location preference matrix [f_{j,y}]
    For each minor item mi_j in MI
        For each major item ma_k in MA
            If major item ma_k is related to minor item mi_j according to ARD
                For each UMSP_m in SPD
                    If UMSP_m includes ma_k
                        For each section s_n in UMSP_m
                            Calculate D_{k,m}^{j,n};
                            Transform D_{k,m}^{j,n} into W_{k,m}^{j,n} using Equation (7);
                        Next
                    End If
                Next
                Derive A_m^{j,n} using Equation (8);
            End If
        Next
        Derive F_{j,n} using Equation (9);
        Derive f_{j,y} using Equation (10);
    Next
    Return [f_{j,y}];
```

Figure 3. Pseudo-code of the LPE procedure.

$$\sum_{j}^{J} E_{jy} = 1, \quad y = 1, 2, \ldots, Y$$

$$E_{jy} \geq 0 \quad \text{for all } j \text{ and } y$$

Let $E_{jy} = 1$ if minor item j is assigned to the shelf where minor item y is displayed, otherwise $E_{jy} = 0$. The item assignment problem shown in Equation (11) is solved using the Hungarian method (Kuhn 1955), a combinatorial optimisation algorithm that can solve the assignment problem in polynomial time.

4. Implementation and experimental results

4.1 *Environment and data description*

Figure 4 illustrates a simplified supermarket to demonstrate the feasibility of the proposed shelf space allocation method. The supermarket is divided into 37 sections (s_1 to s_{37}) and 52 shelves (z_1 to z_{52}) according the instructions mentioned in Section 3.1. Customers enter the supermarket from entrance s_1 and check out at section s_{32} or section s_{37}. There are 119 product items in this store where an item belongs to one of 16 product classes: Alcoholic beverage, Bread, Beverage, Candy, Canned food, Fresh food, Dairy, Cookies, Seasoning/Dressing, Jam, Powder/Cereal, Rice/Noodle, Tonic, Cooked food, Fruits and Vegetables. Table 4 summarises some of the product items and their classes. Product items utilities are listed in Table 5. Figure 5 illustrates the locations of products on the shelves prior to the proposed shelf space allocation method.

Because the RFID system is not currently deployed in this example, a mobile transaction sequence generator is developed to simulate customer shopping behaviour. In this generator, each product class is assigned to at least one of the following four categories: family, personal use, parties and picnics. For example, the family group covers product classes including canned food, fresh food, dairy, cookies, seasoning/dressing, jam, powder/cereal, tonic or fruit. Based on these four groups, five types of customer behaviours (i.e. mobile transaction sequences) are generated. Table 6 shows the percentages of items purchased for each group and type in mobile transaction sequences. For example, for type A behaviour, 70% of purchased products is from group 1, 15% is from group 2, 10% is from group 3 and 5% is from group 4. In addition, types A–E, respectively, account for 30, 30, 20, 10 and 10% of mobile transaction sequences.

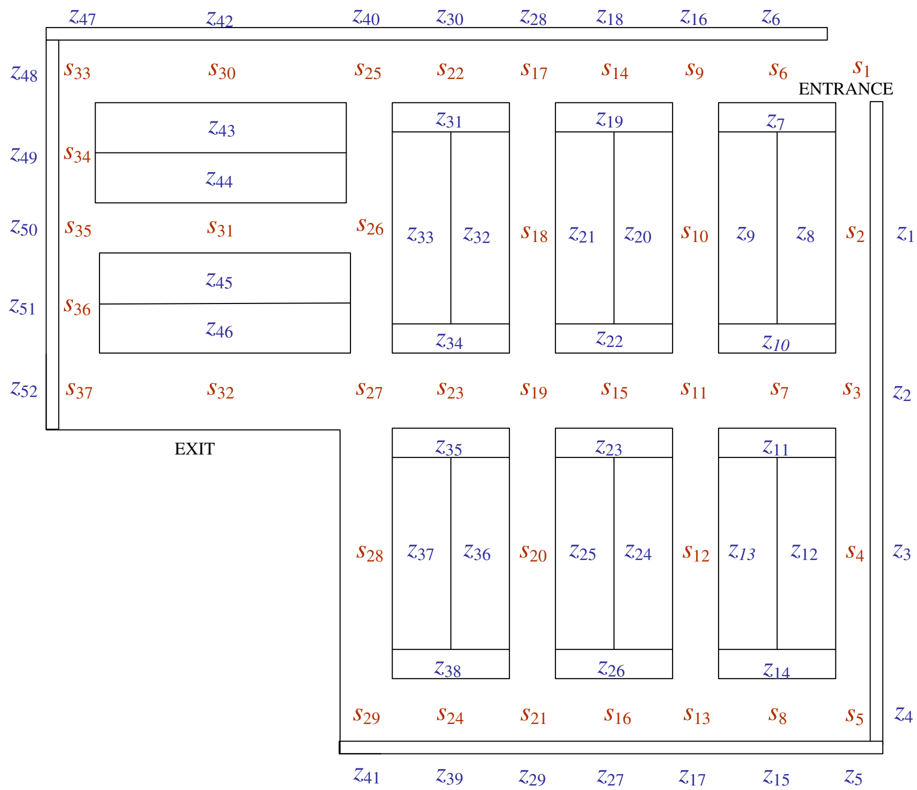

Figure 4. Layout of a simplified supermarket.

Table 4. Product classes and corresponding items in the store.

Product class	Product items
(1) Alcoholic drink	i_1(Apple wine), i_2(Japanese sake), i_3(Champagne), i_4(Beer), i_5(Raki) and i_6(Whisky)
(2) Bread	i_7(Dinner roll), i_8(Donut), i_9(Hearth Bread), i_{10}(Honey cake), i_{11}(Pineapple bread), i_{12}(Rise the crisp cake) and i_{13}(Toast)
(3) Beverage	i_{14}(Apple juice), i_{15}(Barley water), i_{16}(Black tea), i_{17}(Chrysanthemum tea), i_{18}(Coca cola), i_{19}(Cranberry Juice), i_{20}(Grape juice), i_{21}(Green tea, i_{22}(Orange juice), i_{23}(Lemon juice) and i_{24}(Mineral water)
(4) Candy	i_{25}(Chewing gum), i_{26}(Chocolate), i_{27}(Fruit candy) and i_{28}(Milk caramel)
…	…
(14) Cooked food	I_{101}(Beef steak), i_{102}(Chicken steak), i_{103}(Dried bean curd), i_{104}(Grilled chicken leg), i_{105}(Grilled chicken wing), i_{106}(Pork chop) and i_{107}(Shrimp)
(15) Fruit	i_{108}(Apple), i_{109}(Banana), i_{110}(Grapes), i_{111}(Hami melon), i_{112}(Pear) and i_{113}(Tomato)
(16) Vegetable	i_{114}(Bean sprouts), i_{115}(Black mushrooms), i_{116}(Cabbage), i_{117}(Carrot), i_{118}(Gold needle mushroom) and i_{119}(Herba corallodisci)

Purchase frequency is determined by a random uniform distribution in a range of 6–12 in the mobile transaction sequence. In this study, the total number of mobile transaction sequences in the generator is set at 1000. The transaction of each customer can be obtained from the mobile transaction sequences. Tables 7 and 8, respectively, show the mobile transaction sequences and transactions for some customers.

4.2 *Experimental results*

Based on the transactions in Table 8, 24 product association rules are generated using the Apriori algorithm when minimum support = 10% and minimum confidence = 60%. Part of the product association rules is illustrated in Table 9. Next,

Table 5. Utility table for product items.

Product item	Utility (US$ per unit)
i_1	3.5
i_2	6.5
i_3	5
i_4	4
i_5	6.5
…	…
i_{117}	0.65
i_{118}	0.65
i_{119}	0.65

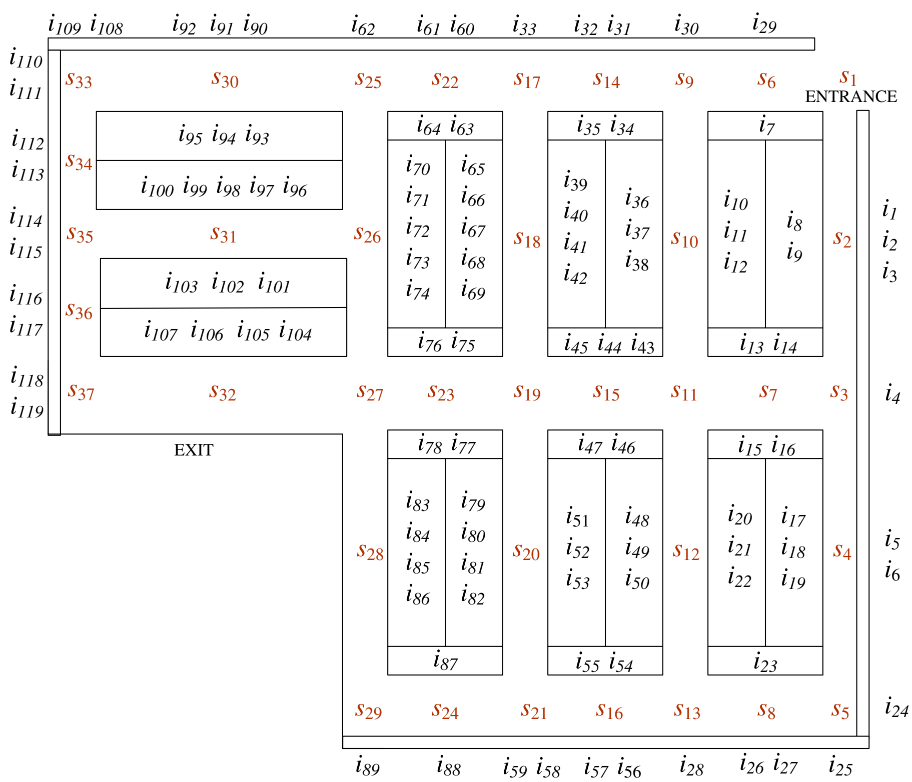

Figure 5. Physical locations of all products.

Table 6. Five types of mobile transaction sequences.

Customer type	Group 1 (%)	Group 2 (%)	Group 3 (%)	Group 4 (%)
A	70	15	10	5
B	40	30	15	15
C	30	30	20	20
D	20	20	30	30
E	25	25	25	25

Table 7. Part of mobile transaction sequences.

CID	Path
1	<(S_1; Ø), (S_2; Ø), (S_3; Ø), (S_4; Ø), (S_5; {[i_{24},4]}), (S_8; Ø), (S_{13}; Ø), (S_{12}; {[i_{50},1]}), (S_{13}; Ø), (S_{16}; Ø), (S_{21}; Ø), (S_{20}; {[i_{53},2]}), (S_{19}; Ø), (S_{18}; {[i_{67},4]}), (S_{17}; Ø), (S_{22}; Ø), (S_{25}; {[i_{62},1]}), (S_{26}; {[i_{74},4]}), (S_{31}; Ø), (S_{35}; {[i_{115},4]}), (S_{36}; Ø), (S_{37}; {[i_{118},1]})>
2	<(S_1; Ø), (S_2; {[i_9,1]}), (S_3; Ø), (S_7; Ø), (S_{11}; Ø), (S_{12}; {[i_{22},3], [i_{49},1]}), (S_{13}; Ø), (S_{16}; Ø), (S_{21}; {[i_{58},1]}), (S_{24}; Ø), (S_{29}; {[i_{89},4]}), (S_{28}; Ø), (S_{27}; Ø), (S_{26}; Ø), (S_{31}; {[i_{97},1]}), (S_{35}; Ø), (S_{36}; Ø), (S_{37}; Ø), (S_{32}; {[i_{106},1]})>
3	<(S_1; Ø), (S_2; {[i_1,1]}), (S_3; Ø), (S_7; Ø), (S_{11}; Ø), (S_{10}; {[i_{38},3]}), (S_{11}; Ø), (S_{12}; Ø), (S_{13}; Ø), (S_{16}; {[i_{56},3]}), (S_{21}; Ø), (S_{20}; {[i_{82},2]}), (S_{19}; Ø), (S_{18}; {[i_{42},4]}), (S_{17}; Ø), (S_{22}; Ø), (S_{25}; {[i_{62},3]}), (S_{30}; {[i_{93},1]}), (S_{33}; Ø), (S_{34}; Ø), (S_{35}; Ø), (S_{36}; {[i_{116},2]}), (S_{37}; Ø)>
4	<(S_1; Ø), (S_2; Ø), (S_3; {[i_4,1]}), (S_4; {[i_{18},1]}), (S_5; {[i_{24},2]}), (S_8; {[i_{26},1]}), (S_{13}; Ø), (S_{16}; Ø), (S_{21}; Ø), (S_{20}; {[i_{79},1]}),(S_{19}; Ø),(S_{18}; Ø),(S_{17}; Ø),(S_{22}; Ø), (S_{25}; {[i_{62},4]}), (S_{30}; {[i_{92},1],[i_{94},1]}), (S_{33}; {[i_{108},3]}), (S_{34}; Ø), (S_{35}; Ø), (S_{36}; Ø), (S_{37}; {[i_{119},1]})>
⋮	⋮
998	<(S_1; Ø), (S_2; Ø), (S_3; Ø), (S_4; Ø), (S_5; {[i_{25},2]}), (S_8; Ø), (S_{13}; {[i_{28},3]}), (S_{16}; Ø), (S_{21}; Ø), (S_{20}; {[i_{51},1]}), (S_{19}; Ø), (S_{18}; {[i_{65},1]}), (S_{17}; Ø), (S_{22}; {[i_{60},3]}), (S_{25}; Ø), (S_{26}; {[i_{72},4]}), (S_{31}; {[i_{102},1]}), (S_{35}; {[i_{115},1]}), (S_{36}; Ø), (S_{37}; {[i_{119},1]})>
999	<(S_1; Ø), (S_2; Ø), (S_3; Ø), (S_7; {[i_{14},4]}), (S_{11}; Ø), (S_{12}; {[i_{21},1]}) (S_{13}; Ø), (S_{16}; Ø), (S_{21}; {[i_{59},1]}), (S_{20}; {[i_{51},3]}), (S_{19}; Ø), (S_{23}; {[i_{78},1]}), (S_{27}; Ø), (S_{32}; {[i_{107},2]})>
1000	<(S_1; Ø), (S_6; {[i_7,4]}), (S_9; Ø), (S_{10}; Ø), (S_{11}; Ø), (S_{12}; {[i_{49},1]}), (S_{13}; Ø), (S_{16}; {[i_{56},3]}), (S_{21}; {[i_{58},2]}), (S_{20}; {[i_{81},2]}), (S_{19}; Ø), (S_{18}; {[i_{67},2]}), (S_{17}; Ø), (S_{22}; Ø), (S_{25}; Ø), (S_{30}; {[i_{92},3]}), (S_{33}; Ø), (S_{34}; Ø), (S_{35}; {[i_{114},1]}), (S_{36}; {[i_{116},2]}), (S_{37}; {[i_{119},1]})>

Table 8. Part of purchase transactions.

CID	Item(s)
1	i_{24}, i_{50}, i_{53}, i_{67}, i_{62}, i_{74}, i_{115}, i_{118}
2	i_9, i_{22}, i_{49}, i_{58}, i_{89}, i_{97}, i_{106}
3	i_1, i_{38}, i_{56}, i_{82}, i_{42}, i_{62}, i_{93}, i_{116}
4	i_4, i_{18}, i_{24}, i_{26}, i_{79}, i_{62}, i_{92}, i_{94}, i_{108}, i_{119}
⋮	⋮
998	i_{25}, i_{28}, i_{51}, i_{65}, i_{60}, i_{72}, i_{102}, i_{115}, i_{119}
999	i_{14}, i_{21}, i_{59}, i_{51}, i_{78}, i_{107}
1000	i_7, i_{49}, i_{56}, i_{58}, i_{81}, i_{67}, i_{92}, i_{114}, i_{116}, i_{119}

Table 9. Association rules (minimum support = 10% and minimum confidence = 60%).

Rule ID	Rule	Rule ID	Rule	Rule ID	Rule
1	$i_4 \rightarrow i_{43}$	9	$i_{22} \rightarrow i_{62}$	17	$i_{43} \rightarrow i_{107}$
2	$i_{43} \rightarrow i_4$	10	$i_{62} \rightarrow i_{22}$	18	$i_{107} \rightarrow i_{43}$
3	$i_4 \rightarrow i_{62}$	11	$i_{30} \rightarrow i_{42}$	19	$i_{58} \rightarrow i_{65}$
4	$i_{62} \rightarrow i_4$	12	$i_{42} \rightarrow i_{30}$	20	$i_{65} \rightarrow i_{58}$
5	$i_4 \rightarrow i_{76}$	13	$i_{30} \rightarrow i_{65}$	21	$i_{58} \rightarrow i_{80}$
6	$i_{76} \rightarrow i_4$	14	$i_{65} \rightarrow i_{30}$	22	$i_{80} \rightarrow i_{58}$
7	$i_8 \rightarrow i_{48}$	15	$i_{34} \rightarrow i_{58}$	23	$i_{62} \rightarrow i_{105}$
8	$i_{48} \rightarrow i_8$	16	$i_{58} \rightarrow i_{34}$	24	$i_{105} \rightarrow i_{62}$

the $UMSP_L$ algorithm is applied to generate high-utility mobile sequential patterns based on the utility data in Table 5 and mobile transaction sequences in Table 7. There are 16 high-utility mobile sequential patterns generated when the minimum support count is 6 and the minimum utility is 10. Part of the high-utility mobile sequential patterns is shown in Table 10.

After association rules and high-utility mobile sequential patterns are generated, major items and minor items can be found. In this simulation, nine major items are found including i_4 (Beer), i_{30} (Fish canned food), i_{43} (Miso instant noodle), i_{48} (Low-fat milk), i_{58} (Soda cracker), i_{62} (Chilli sauce), i_{65} (Ketchup), i_{83} (Chocolate powder) and i_{105} (Grilled

Table 10. High-utility mobile sequential patterns (minimum support count = 6).

Pattern ID	High-utility mobile sequential pattern	Path	Sup.
1	$<S_3; i_4><S_{15}; i_{43}><S_{25}; i_{62}><S_{32}; i_{105}>$	$S_1, S_2, S_3, S_{11}, S_{15}, S_{19}, S_{25}, S_{32}$	6
2	$<S_3; i_4><S_{15}; i_{43}><S_{25}; i_{62}>$	$S_1, S_2, S_3, S_{11}, S_{15}, S_{19}, S_{25}$	12
3	$<S_3; i_4><S_{15}; i_{43}><S_{32}; i_{105}>$	$S_1, S_2, S_3, S_{11}, S_{15}, S_{19}, S_{32}$	15
		$S_1, S_2, S_3, S_{11}, S_{15}, S_{19}, S_{25}, S_{32}$	6
4	$<S_3; i_4><S_{25}; i_{62}><S_{32}; i_{105}>$	$S_1, S_2, S_3, S_{25}, S_{32}$	26
		$S_1, S_2, S_3, S_{11}, S_{15}, S_{19}, S_{25}, S_{32}$	6
		$S_1, S_2, S_3, S_{11}, S_{15}, S_{19}, S_{25}$	12
		$S_1, S_{11}, S_{15}, S_{19}, S_{23}, S_{27}, S_{26}, S_{25}$	8
…	...	...	…
14	$<S_{15}; i_{43}><S_{32}; i_{105}>$	$S_1, S_{11}, S_{15}, S_{19}, S_{32}$	30
		$S_1, S_{11}, S_{15}, S_{19}, S_{25}, S_{32}$	20
		$S_1, S_2, S_3, S_{11}, S_{15}, S_{19}, S_{32}$	15
		$S_1, S_2, S_3, S_{11}, S_{15}, S_{19}, S_{25}, S_{32}$	6
15	$<S_{18}; i_{65}><S_{21}; i_{58}>$	$S_1, S_{18}, S_{19}, S_{20}, S_{21}$	12
		$S_1, S_6, S_9, S_{14}, S_{17}, S_{18}, S_{19}, S_{20}, S_{21}$	6
16	$<S_{25}; i_{62}><S_{32}; i_{105}>$	S_1, S_{25}, S_{32}	38
		$S_1, S_2, S_3, S_{25}, S_{32}$	26
		$S_1, S_{11}, S_{15}, S_{19}, S_{25}, S_{32}$	20
		$S_1, S_2, S_3, S_{11}, S_{15}, S_{19}, S_{25}, S_{32}$	6

chicken wing). In addition, seven minor items are identified including i_8 (Donut), i_{22} (Orange juice), i_{34} (Beef instant noodle), i_{42} (Kimchi stock), i_{76} (Vegetable oil), i_{80} (Strawberry jam) and i_{107} (Shrimp). Given the major and minor items, the threshold of maximum section movement is set at 3, and the location preference weight matrix $[f_{j,y}]$ can be obtained according to Equations (7)–(10). Figure 6 shows the values in the location preference weight matrix. For example, $f_{2,1} = 0.9$ is the location preference weight if minor item i_{22} is placed on the shelf of minor item z_8, and $f_{5,7} = 0.87$ is the location preference weight if minor item i_{76} is placed on the shelf of minor item z_{46}.

The final stage of the proposed shelf space allocation method reassigns minor items to their best shelf location using the Hungarian method. Figure 7 depicts the final physical location of minor items after reassignment. Checking Table 9 shows that the reassignment result is satisfactory. Table 9 indicates that i_8 (Donut) is strongly related to major item i_{48} (Low-fat milk), i_{42} (Kimchi stock) is strongly related to major item i_{30} (Fish canned food), i_{76} (Vegetable oil) is strongly related to major item i_4 (Beer) and i_{107} (Shrimp) is strongly related to major item i_{43} (Miso instant noodle). Thus, the four minor items (i_8, i_{42}, i_{76} and i_{107}) are reorganised to locations close to their associated major items (i_{48}, i_{30}, i_4 and i_{43}). The location of minor item i_{80} (Strawberry jam) does not change since i_{80} is already located very close to its major item i_{58} (Soda cracker) in the original layout. Minor item i_{34} is not assigned to the best shelf z_{36} since the location preference weights are calculated based on average concept.

4.3 *Sensitivity analysis*

This section analyses how parameters in our method affect the final reassignment result. These parameters include minimum support count in the $UMSP_L$ algorithm, as well as minimum support and minimum confidence in the Apriori algorithm.

$$
\boldsymbol{F} = \begin{array}{c|ccccccc}
 & z_8 & z_{13} & z_{19} & z_{21} & z_{34} & z_{36} & z_{46} \\
i_8 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\
i_{22} & 0.9 & 1 & 1 & 1 & 1 & 1 & 0.33 \\
i_{34} & 1 & 1 & 1 & 0.87 & 1 & 0.33 & 1 \\
i_{42} & 1 & 1 & 0.33 & 0.78 & 1 & 1 & 1 \\
i_{76} & 0.33 & 1 & 1 & 1 & 1 & 1 & 0.87 \\
i_{80} & 1 & 1 & 1 & 0.87 & 1 & 0.33 & 1 \\
i_{107} & 1 & 0.67 & 1 & 1 & 0.67 & 1 & 0.9
\end{array}
$$

Figure 6. Location preference weight matrix, $[f_{j,y}]$.

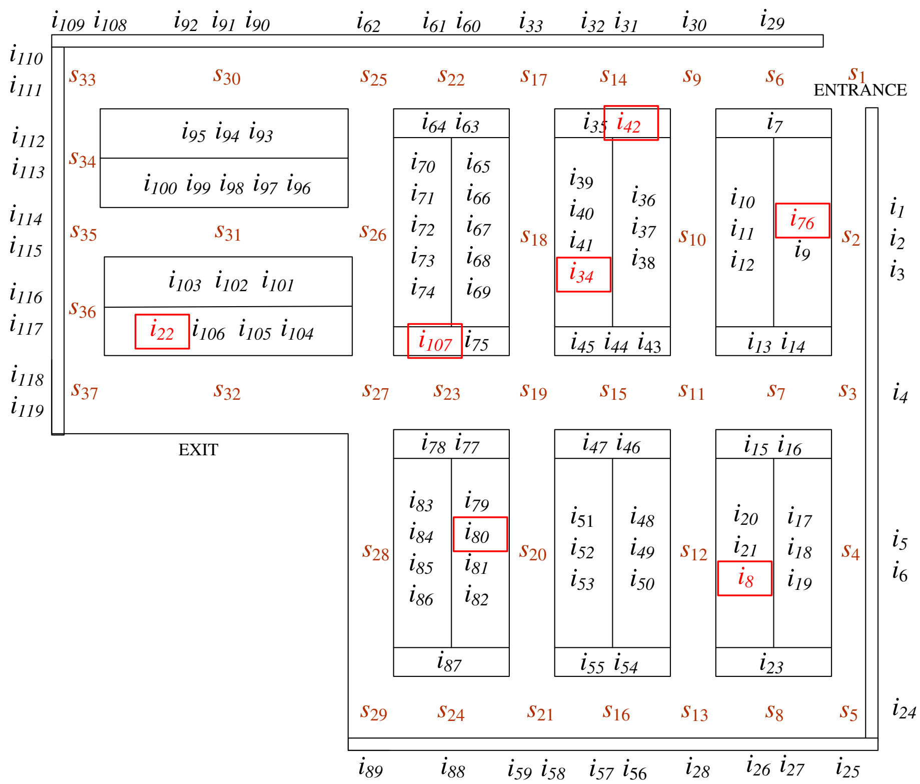

Figure 7. Physical locations of products after reassignment.

4.3.1 *Analysis of minimum support count in $UMSP_L$*

According to definition 1 in Section 3.4, the minimum support count in $UMSP_L$ has a direct effect on the number of major items. Therefore, the following analysis result focuses on how the number of major items affects the reassignment result. Table 11 shows the experimental results of applying five different minimum support counts in $UMSP_L$. The table shows that the reassignment result is feasible when the minimum support count is 4 or 6. Two types of infeasible situations are identified. For infeasible situation – (a), the number of major items is 18 when the minimum support count is 2. The large number of major items corresponds with a very small number of minor items (i.e. 1). Since no minor items can be exchanged, it becomes an infeasible solution. For infeasible situation – (b), the number of major items falls when the minimum support count is higher (8 or 10). According to the definition in Section 3.4, a minor item is one which is related to major items. In other words, a small number of major items will correspond with a small number of minor items. Figure 8 illustrates the feasible and infeasible zones given different minimum support counts in $UMSP_L$. It is clear that feasible results require a minimum support count in $UMSP_L$ which is neither too high nor too low.

4.3.2 *Analysis of minimum support and minimum confidence in association rules*

The setting of different minimum supports and minimum confidences in the association analysis also causes different reassignment results because the number of minor items is significantly affected by the number of association rules.

Table 11. Reassignment results for five minimum support counts in the $UMSP_L$ algorithm.

Minimum support count	Number of major items	Number of minor items	Reassignment result
2	18	1	Infeasible[(a)]
4	14	5	Feasible
6	9	7	Feasible
8	5	2	Infeasible[(b)]
10	2	1	Infeasible[(b)]

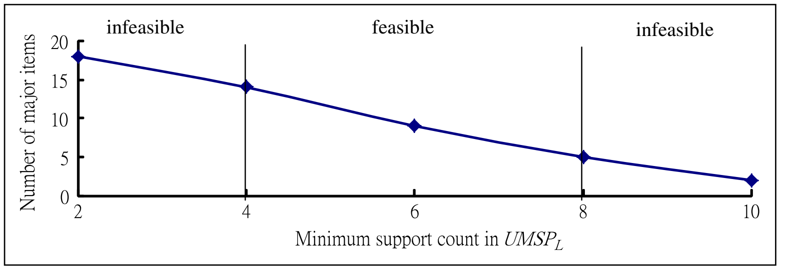

Figure 8. Reassignment results for different minimum support counts in $UMSP_L$.

Table 12. Reassignment results for different minimum supports and minimum confidences in association rule analysis.

Minimum support (%)	Minimum confidence (%)	Number of minor items	Reassignment result
5	40	8	Feasible
5	50	8	Feasible
5	60	8	Feasible
5	70	0	Infeasible
10	40	7	Feasible
10	50	7	Feasible
10	60	7	Feasible
10	70	0	Infeasible
15	40	6	Feasible
15	50	5	Feasible
15	60	5	Feasible
15	70	0	Infeasible
20	40	0	Infeasible
20	50	0	Infeasible
20	60	0	Infeasible
20	70	0	Infeasible

Therefore, the following discusses the reassignment result among four different minimum supports and four different minimum confidences, while the minimum support count in $UMSP_L$ algorithm is kept constant (i.e. 6). The 16 experimental results are summarised in Table 12 which shows that if the minimum support is set as a value greater than or equal to 20%, no association rule is generated. Thus, there is no minor item to reassign. Similarly, if the minimum confidence is set as a value greater than or equal to 70%, no association rule is generated.

Based on the above observations, Figure 9 summarises our suggestions for setting the minimum support and minimum confidence. The figure is divided into three sections. Section B shows the feasible result when minimum support and confidence are neither too high nor too low. Section C shows the infeasible result if both minimum support and confidence are set to too high. Section A shows the feasible but non-preferred result in which too many minor items are generated in Section A, thus increasing the chance that customers will have trouble following their previously preferred moving patterns since too many item locations have been changed.

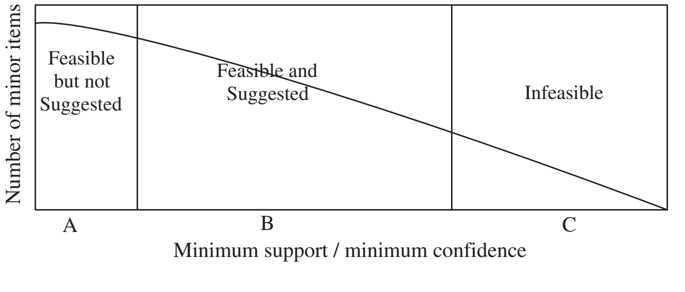

Figure 9. Trend analysis for setting minimum support/minimum confidence.

5. Conclusions

In retail shops, different shelf space allocation strategies can significantly impact customer purchasing decisions. Thus, this research proposes a novel data mining approach to solve shelf space allocation problems through considering both customer product association relationships and shopping path knowledge. The proposed method allows retail store managers to optimise product allocation and positioning. The main contributions of the proposed method are summarised as follows. First, the proposed method considers customer purchasing behaviour derived from data on previous transactions, along with customer browsing behaviour collected through the use of RFID technology. This paper is the first work to integrate purchasing records and moving records to solve shelf space allocation problems. Second, this study develops the concept that moving all items to other section(s)/shelves(s) may render customers' previous shopping patterns invalid since they will not be able to locate items they originally intended to purchase. Therefore, instead of relocating all items, product items in the store are classified as major items, minor items and trivial items according to a set of regulations. Only minor items are relocated to ensure customers can follow their preferred shopping paths. Third, this paper develops a LPE procedure to calculate location preference if a minor item is placed in a certain section in the store. The minor items are relocated to be in closer proximity to their corresponding major items, thus increasing the likelihood of successful cross-selling. Fourth, the proposed method is elastic in that store managers can set different minimum support and confidence levels in the association rule analysis, or a minimum support count in $UMSP_L$ algorithm to control the number of major and minor items to generate the optimal product layout.

Unfortunately, no similar methods can be found and compared. To validate the proposed method, a mobile transaction sequence generator is developed to simulate customer shopping behaviour. A series of analyses and comparisons on the performance of different types of shopping behaviour are conducted through experimental evaluations. The implementation and experimental results show the proposed method is feasible and helpful for store managers.

Some potential extensions for this research are as follows. First, in the current study, all products on a given shelf are limited to the same major category (e.g. 'Food'). Future work could consider grouping items from multiple product classes. Second, this study assumes product volumes are identical and can be smoothly be exchanged. However, in practice, different products might be displayed in different volumes, and future work can consider this issue. Third, this study solves the product-to-shelf assignment using the Hungarian method, which would run too slowly given large quantities of data, and future work can test different assignment algorithms. Fourth, the proposed method is now validated and evaluated through the simulated behaviour data by the mobile transaction sequence generator. In the future, a practical data collected from the RFID system in a store will be tested and implemented.

References

Agrawal, R., and R. Srikant. 1994. "Fast Algorithms for Mining Association Rules." In *Proceedings of the 20th International Conference on Very Large Data Bases*, 487–499, Santiago, Chile.

Borin, N., P. W. Farris, and J. R. Freeland. 1994. "A Model for Determining Retail Product Category Assortment and Shelf Space Allocation." *Decision Sciences* 25 (3): 359–384.

Chan, R., Q. Yang, and Y. D. Shen. 2003. "Mining High Utility Itemsets." In *Proceedings of Third IEEE International Conference on Data Mining*, 19–26, Melbourne, FL.

Chen, M. 2012. "Constructing a Model of Shelf Space Allocation in Product Categories." In *Proceedings of 2012 IEEE Symposium on Robotics and Applications*, 701–704, Kuala Lumpur.

Chen, M. C., and C. P. Lin. 2007. "A Data Mining Approach to Product Assortment and Shelf Space Allocation." *Expert Systems with Applications* 32 (4): 976–986.

Chen, Y. L., J. M. Chen, and C. W. Tung. 2006. "A Data Mining Approach for Retail Knowledge Discovery with Consideration of the Effect of Shelf-space Adjacency on Sales." *Decision Support Systems* 42 (3): 1503–1520.

Cil, I. 2012. "Consumption Universes Based Supermarket Layout through Association Rule Mining and Multidimensional Scaling." *Expert Systems with Applications* 39 (10): 8611–8625.

Corstjens, M., and P. Doyle. 1981. "A Model for Optimizing Retail Space Allocations." *Management Science* 27 (7): 822–833.

Curhan, R. C. 1972. "The Relationship between Shelf Space and Unit Sales in Supermarkets." *Journal of Marketing Research* IX: 406–412.

Hansen, P., and H. Heinsbroek. 1979. "Product Selection and Space Allocation in Supermarkets." *European Journal of Operational Research* 3 (6): 474–484.

Hui, S. K., E. T. Bradlow, and P. S. Fader. 2009. "Testing Behavioral Hypotheses Using an Integrated Model of Grocery Store Shopping Path and Purchase Behavior." *Journal of Consumer Research* 36 (3): 478–493.

Irani, Z., A. Gunasekaran, and Y. K. Dwivedi. 2010. "Radio Frequency Identification (RFID): Research Trends and Framework." *International Journal of Production Research* 48 (9): 2485–2511.

Kuhn, H. W. 1955. "The Hungarian Method for the Assignment Problem." *Naval Research Logistics Quarterly* 2: 83–97.

Larson, J. S., E. T. Bradlow, and P. S. Fader. 2005. "An Exploratory Look at Supermarket Shopping Paths." *International Journal of Research in Marketing* 22 (4): 395–414.

Lee, S. C., J. Paik, J. Ok, I. Song, and U. M. Kim. 2007. "Efficient Mining of User Behaviors by Temporal Mobile Access Patterns." *International Journal of Computer Science and Network Security* 7 (2): 285–291.

Levy, M., and B. A. Weitz. 1995. *Retailing Management*. Chicago, IL: Irwin.

Liao, I. E., and W. C. Lin. 2007. "Shopping Path Analysis and Transaction Mining Based on RFID Technology." 2007 1st Annual RFID Eurasia, Istanbul, Turkey, 292–296.

Lim, A., B. Rodrigues, and X. Zhang. 2004. "Metaheuristics with Local Search Techniques for Retail Shelf-space Optimization." *Management Science* 50 (1): 117–131.

Lu, E. H. C., and V. S. Tseng. 2009. "Mining Cluster-based Mobile Sequential Patterns in Location-based Service Environments." In *Proceedings of the IEEE International Conference on Mobile Data Management*, 273–278, Taipei.

Murray, C. C., D. Talukdar, and A. Gosavi. 2010. "Joint Optimization of Product Price, Display Orientation and Shelf-Space Allocation in Retail Category Management." *Journal of Retailing* 86 (2): 125–136.

Nafari, M., and J. Shahrabi. 2010. "A Temporal Data Mining Approach for Shelf-space Allocation with Consideration of Product Price." *Expert Systems with Applications* 37 (6): 4066–4072.

Shie, B. E., H. F. Hsiao, and V. S. Tseng. 2013. "Efficient Algorithms for Discovering High Utility User Behavior Patterns in Mobile Commerce Environments." *Knowledge and Information Systems* 37 (2): 363–387.

Shie, B. E., V. S. Tseng, and P. S. Yu. 2010. "Online Mining of Temporal Maximal Utility Itemsets from Data Streams." In *Proceedings of the 25th Annual ACM Symposium on Applied Computing*, 1622–1626, Sierre.

Song, Z., and A. Kusiak. 2009. "Optimising Product Configurations with a Data-mining Approach." *International Journal of Production Research* 47 (7): 1733–1751.

Tseng, V. S., E. H. C. Lu, and C. H. Huang. 2007. "Mining Temporal Mobile Sequential Patterns in Location-based Service Environments." In *Proceedings of the International Conference on Parallel and Distributed Systems*, 1–8, Hsinchu.

Tseng, V. S., C. W. Wu, B. E. Shie, and P. S. Yu. 2010. "UP-Growth: An Efficient Algorithm for High Utility Itemsets Mining." In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 253–262, Washington, DC.

Yang, M. H. 1999. "An Efficient Algorithm to Allocate Shelf Space." *European Journal of Operational Research* 131 (1): 107–118.

Yang, M. H., and W. C. Chen. 1999. "A Study on Shelf Space Allocation and Management." *International Journal of Production Economics* 60–61: 309–317.

Yun, C. H., and M. S. Chen. 2007. "Mining Mobile Sequential Patterns in a Mobile Commerce Environment." *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)* 37 (2): 278–295.