Big data analysis on the business process and management for the store layout and bundling sales

Shu-hsien Liao and Yi-Shan Tasi
Tamkang University, New Taipei City, Republic of China

Received 28 January 2018
Revised 27 December 2018
10 February 2019
Accepted 17 February 2019

Abstract

Purpose – In the retailing industry, database is the time and place where a retail transaction is completed. E-business processes are increasingly adopting databases that can obtain in-depth customers and sales knowledge with the big data analysis. The specific big data analysis on a database system allows a retailer designing and implementing business process management (BPM) to maximize profits, minimize costs and satisfy customers on a business model. Thus, the research of big data analysis on the BPM in the retailing is a critical issue. The paper aims to discuss this issue.

Design/methodology/approach – This paper develops a database, ER model, and uses cluster analysis, C&R tree and the *a priori* algorithm as approaches to illustrate big data analysis/data mining results for generating business intelligence and process management, which then obtain customer knowledge from the case firm's database system.

Findings – Big data analysis/data mining results such as customer profiles, product/brand display classifications and product/brand sales associations can be used to propose alternatives to the case firm for store layout and bundling sales business process and management development.

Originality/value – This research paper is an example to develop the BPM of database model and big data/data mining based on insights from big data analysis applications for store layout and bundling sales in the retailing industry.

Keywords Retailing, Business process management, Big data, Database management, Bundling sales, Data mining

Paper type Research paper

1. Introduction

A business process is a collection of related, structured activities or tasks that produce a specific service or product (serve a particular goal) for a particular customer or customers. It may often be visualized as a flowchart of a sequence of activities with interleaving decision points or as a process matrix of a sequence of activities with relevance rules based on data in the process (Trkman *et al.*, 2015). In addition, business process management (BPM) can be defined as assigning each manager in-charge of a whole set of activities that produce a valuable product/service for current or future internal/external customers. It is also considered as the horizontal method of management based on end-to-end processes, and contrasts with the traditional hierarchical approach (Hammer, 2007; Paim *et al.*, 2008; Palmberg, 2010; Bernardo *et al.*, 2017). BPM aims to assign responsibilities based on expected products/services for internal/external customers rather than on duties, and "to organize around outcomes not tasks" (Hammer, 2007). Mounting empirical evidence suggest that having a process orientation, results in enhancement of products/services, decrease in costs and faster functions (Psomas *et al.*, 2011; Jayaraman, 2016).

Organizations are increasingly seeking to improve the implementation of strategy, given the need to align organizational actions and goals to increase effectiveness (Niehaves *et al.*, 2014). BPM is a management approach that contributes to this purpose by aligning the organization with customer demands (ABPMP, 2013). Business processes largely determine

This research was funded by the Ministry of Science and Technology, Taiwan, Republic of China (NSC 101-2622-H-032-001-CC3).

1463-7154
DOI 10.1108/BPMJ-01-2018-0027

the quality, degree of innovation and productivity of organizations (Lindman *et al.*, 2016). When effective, business processes are considered unique and critical corporate assets that account for a significant portion of the organization's costs, offering significant opportunities to improve market share, decision-making capabilities and performance management (Ohlsson *et al.*, 2017).

In the retailing industry, store layout allows a retailer to maximize sales volume per square foot of the allocated selling space within the store. Store layouts generally show the size and location of each department, any permanent structure, fixture locations and customer traffic patterns. A large portion of the sales and marketing effort for any store is related to store layout and merchandise displays process design. Therefore, it is critical that the floor plan not only maximizes available space, but presents merchandise in the most effective layout possible. This is as true for online stores as for traditional retailing outlets, and there are several previous studies focused on online retailing store layout. For example, Cordier *et al.* (2003) present a web application that provides more powerful access to and manipulation of clothing to facilitate clothing design, pattern derivation and sizing. Vrechopoulos *et al.* (2004) propose an experimental investigation into the use of three different layouts for online grocery retailing: freeform, grid and racetrack. These three conventional retailing layout types were transformed into virtual layouts for computer-mediated interfaces. In another direction, Griffith's (2005) study uses information processing theory and aspects of the technology acceptance model to theorize how two types of online store layout (tree and tunnel website structures) influence consumer elaboration and response. Hu and Jasper (2007) presented a cross-cultural examination of the effects of social perception styles on consumers' store image formation. Their results showed that, as hypothesized, Chinese students were more significantly affected by the social cues embedded within the online store environment than American students were. In addition, unlike American women, Chinese women formed a favorable impression of a store with low social orientation.

On the other hand, data can be an incredibly powerful tool for BPM. Companies using it to their advantage – from recruiting to customer retention to selling process – will be the ones that thrive. Big data is a relatively new term coined to label the exponential growth and availability of data, both structured and unstructured. The availability of massive amounts of data provides unprecedented opportunities for organizations. The research of big data on the BPM is flourishing. Guo *et al.* (2017) combined the techniques of web crawler, natural language processing and machine learning algorithms with data visualization to develop a big data competitor analysis system that informs a business process that operations managers about competitors and meaningful relationships among them. The authors illustrate the approach using the fitness mobile app business. In addition, among the new techniques developed for big data, data mining is the process of discovering significant knowledge, such as patterns, associations, changes, anomalies and significant structures from large amounts of data stored in databases In the literature, there are also many big data application methods for the research issue of BPM (Maita *et al.*, 2015). Thus, in terms of BPM, how to effectively process and use big data analysis on various business process are becoming increasingly important. The big data analysis approach, such as data mining, is to help analyze, understand or even visualize the huge amounts of the big data gathered from business and research applications for application by retailing firms (Liao *et al.*, 2014). In the literature, there are many data mining models such as classification, estimation, predictive modeling, clustering/segmentation, affinity grouping or association rules, description and visualization, as well as sequential modeling. Similarly, there are also many application methods, including association rules, sequential patterns, grouping analysis, classification analysis and probability heuristic analysis (Maita *et al.*, 2015). Customer knowledge extracted through data mining can be integrated with product/brand purchased knowledge from research and then provided to retailers, thereby serving as a valuable reference for business process mining in the retailing industry.

Accordingly, this study investigates store layout and bundling sales process management issue in a Taiwan outdoor articles chain store company, GoHiking. There are three data mining stages implemented in this study. The cluster analysis is focused on investigating customer segmentation and C&R tree is a methodology for implementing classification analysis to explore the customer profile and product/brand mix on the case firm database system. Following that, the *a priori* algorithm is a methodology that consists of the association rules for data mining, and which is implemented to extract knowledge from data mining results. This is illustrated as knowledge patterns and rules integrating with a business process map in order to propose alternatives to the case firm for its store layout and possible brand/product cross-selling bundling recommendations. The rest of this study is organized as follows. In Section 2, we present the background of the case firm. Section 3 introduces the database design and structure, the relational database. Section 4 presents the data mining process, including the Classification and regression trees (C&RT), *a priori* algorithm, and knowledge extraction. Section 5 analyzes data mining results. Managerial implications are presented in Section 6, and Section 7 presents a brief conclusion.

2. The case firm: an outdoor articles chain store in Taiwan – GoHiking

Beginning in 2010, Taiwan textile fiber manufacturers looked downstream to purchase conservative fight gross margin improvement and revenue increase, textile fibers plant is self-seeking vertical integration to enhance product value. Thus, manufacturers must create their own brands. For example, Li Li/Li Peng, a listed chemical engineering company in Taiwan is the second largest nylon 6 manufacturing firm in the world invested in the retail channel brand GoHiking that had opened 14 stores on Taiwan. Clothing and equipment lines are tailored to climbers, mountaineers, snowboarders, snowboarders, hikers, water/aquatic sports and endurance athletes.

In 2012, March 13, GoHiking, was one of the three Taiwan-based brands to be voted on to the list of the world's most effective rebrands in the eighth annual REBRAND 100® International Awards. The awards are the highest recognition for excellence in brand repositioning, and are part of the first and most-respected international program of its kind. World-leading brands such as Cisco, Pfizer and UBS were also among this year's winners. GoHiking, received this award due to its new identity, including color, store layout and design, packaging, and bag design, that together transformed GoHiking into a stylish, contemporary and fashionable brand. The store layout and design process, when used in combination with consumer profile/brand/product/selling space/purchase behavior segmentation considerations, created an in-store design that was unprecedented in the local market, achieving high levels of acceptance among a wide spectrum of consumers and creating significant differentiation from its competitors.

3. Research design

3.1 Research framework

The study framework is shown in Figure 1. Since there is a large database for the case firm, we developed a relational database to obtain information linking consumers purchasing preferences and behaviors. We then constructed a system architecture and database design to describe consumer behaviors and preferences. C&RT analysis is the first stage of data mining in order to investigate a possible store layout. By doing so, some specific product/brand sales segments can be described as different possible store layout design patterns. Developing association rules is the second stage for mining specific patterns of time-based product and brand bundling and customer profiles sales promotion for sales activities.

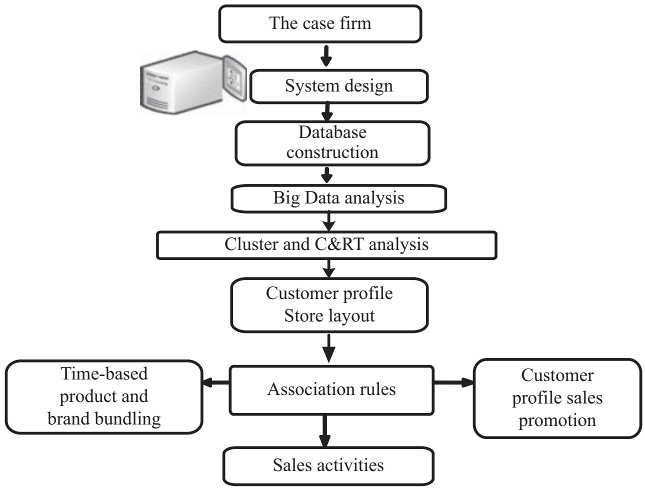


Figure 1.
Research framework

3.2 Database structure design – a relational database (E-R model) development

In terms of support the store layout and bundling sales analysis, this study develops a relational database, ER model, for a database structure design. The concept of the relational databases was first developed in the Codd (1970) to represent interrelated data in the form of a table, and the representation of data in an interrelated table has since become the main characteristic of relational databases. Relational databases organize data as a collection of tables in which all data relationships are represented by common values in related tables. These databases can relate data stored in one table to data in another, as long as the two tables share a common data element. Many organizations maintain relational databases, and since relational patterns reliably portray patterns embedded within databases, relational patterns can be utilized by these organizations to support a variety of efforts for building their database management systems and E-business. By organizing detail protocols from the information on the overall structure, a database design and structure are usually presented in the form of graphics, and described by the use of the ER model or object modeling symbols. In this study, the relational database contains 24 entities, 5 relationships and 191 attributes with data normalization, a data warehousing schema. From May 2015 to June 2016, the case firm database system and relational database then provided a normalization database system, including a total of 1,576,423 valid and selected transaction data for big data and data mining analysis.

4. Big data analysis – data mining approach

4.1 Cluster analysis

The process of partitioning a large set of patterns into disjoint and homogeneous clusters is fundamental in knowledge acquisition. It is called clustering in most studies and it has been applied in various fields, including data mining, statistical data analysis, compression and vector quantization. The k-means is a very popular algorithm and is one of the best for implementing the clustering process. K-means clustering proceeds in the following order. First, the K numbers of observations are randomly selected from all N number of observations according to the number of clusters, and these become centers of the initial clusters. Second, for each of the remaining N–K observations, the nearest cluster is found in terms of the Euclidean distance with respect to $xi = (x_{i1}, x_{i2}, \ldots, x_{ip}, \ldots, x_{iP})$. After each observation is assigned to the nearest cluster, the center of the cluster is re-computed.

Finally, after the allocation of all observations, the Euclidean distance between each observation and the cluster's center point is calculated to confirm whether or not they have been allocated to the nearest cluster. In addition, several studies have discussed implementation of the *k*-means algorithm for cluster analysis as a data mining approach (Khlif, 2017).

4.2 Classification and regression trees (C&RT)

C&RT, a recursive partitioning method, builds classification and regression trees for predicting continuous dependent variables (regression) and categorical predictor variables (classification). The classic C&RT algorithm was popularized by Breiman *et al.* (1984). There are numerous algorithms for predicting continuous variables or categorical variables from a set of continuous predictors and/or categorical factor effects. For example, in general linear models and general regression models, we can specify a linear combination (design) of continuous predictors and categorical factor effects (e.g. with two-way and three-way interaction effects) to predict a continuous dependent variable. In general discriminant function analysis (GDA), we can specify such designs for predicting categorical variables, i.e., to solve classification problems (Ripley, 1996).

For example, classification-type problems are generally those where analysis attempts to predict values of a categorical dependent variable (class, group membership, etc.) from one or more continuous and/or categorical predictor variables. For example, we might be interested in predicting which one of multiple different alternative consumer products (e.g. makes of cars) a person decides to purchase, or which type of failure occurs with different types of engines. In those cases, there are multiple categories or classes for each categorical dependent variable. There are a number of data mining methods for analyzing classification-type problems and to compute predicted classifications, either from simple continuous predictors (e.g. binomial or multinomial logit regression in GLZ), from categorical predictors (e.g. log-linear analysis of multi-way frequency tables), or both (e.g. via ANCOVA-like designs in GLZ or GDA). The CHAID also analyzes classification-type problems and produces results that are similar in nature to those computed by C&RT. Note that various neural network architectures are also applicable to solve classification-type problems.

On the other hand, C&RT classification techniques can produce accurate predictions or predicted classifications based on few logical if-then conditions, and they have a number of advantages over many alternative techniques. In most cases, the interpretation of results can be summarized in a tree very simply. This simplicity is useful for rapid classification of new observations since it is much easier to evaluate just one or two logical conditions, than to compute classification scores for each possible group, or predicted values, based on all predictors and using some possibly complex nonlinear model equations. For example, when analyzing business process problems, it is much easier to present management with if-then statements to management than elaborate equations (Witten and Frank, 2000).

4.3 Association rules

Discovering association rules is an important data mining problem (Agrawal *et al.*, 1993), and there has been considerable research on using association rules for data mining problems. The association rules algorithm is used mainly to determine the relationships between items or features that occur synchronously in databases. For instance, during a trip to the shopping center, if the people who buy item X also buy item Y as well, there exists a relationship between items X and Y. Such information is useful for decision makers. Therefore, the main purpose of implementing the association rules algorithm is to find synchronous relationships by analyzing random data and to use these relationships as a reference for decision-making. The association rules are defined as follows (Wang *et al.*, 2004).

Make $I = \{i_1, i_2, ..., i_m\}$ the item set, in which each item represents a specific literal. D stands for a set of transactions in a database in which each transaction T represents an item set such that T ⊆ I. That is, each item set T is a non-empty sub-item set of I. The association rules are an implication of the form X→Y, where X ⊂ I, Y ⊂ I and X∩Y = Φ. The rule X→Y holds in the transaction set D according to two measurement standards – support and confidence. Support (denoted as_Sup(X, D)) represents the rate of transactions in D containing the item set X. Support is used to evaluate the statistical importance of D, and the higher its value, the more important the transaction set D is. Therefore, the rule $X \rightarrow Y$ which has support Sup(X∪Y, D) represents the rate of transactions in D containing X∪Y. Each rule X→Y also has another measuring standard called confidence (denoted as Conf(X→Y)), representing the rate of transactions in D that contain both X and Y. That is, Conf (X→Y) = Sup(X∩Y)/Sup(X, D).

In this case, Conf(X→Y) denotes that if a transaction includes X, the chance that this transaction also contains Y is relatively high. The measure of confidence is then used to evaluate the level of confidence about the association rules X→Y. Given a set of transactions, D, the problem of mining association rules is used to generate all transaction rules that have certain levels of user-specified minimum support (called Min sup) and confidence (called Minconf). According to Agrawal and Shafer (1996), the problem of mining association rules can be broken down into two steps. The first step is to detect a large item set whose support is greater than Min sup, and the second step is to generate association rules, using the large item set. Such rules must satisfy the following two conditions:

$$\mathrm{Sup}(\mathrm{X} \cup \mathrm{Y}, \mathrm{D}) \geqslant \text{ Min sup}, \tag{1}$$

$$\mathrm{Conf}(\mathrm{X} \rightarrow \mathrm{Y}) \geqslant \mathrm{Minconf}. \tag{2}$$

To explore association rules, many researchers use the *a priori* algorithm (Agrawal *et al.*, 1993). In order to reduce the possible biases incurred when using these measurement standards, the simplest way to judge the standard is to use the lift judgment. Lift is defined as: Lift = Confidence(X→Y)/Sup(Y) (Wang *et al.*, 2004).

4.4 Big data/data mining tool – SPSS Modeler

In this research, SPSS Modeler is employed as data mining tool for analysis. The difference between Modeler and other software is that its data processing works through the use of nodes, which are then connected to form a stream frame. In addition, data visualization can be presented to users after the mining process has been completed. The data processing in Modeler is performed through the use of nodes, which are then connected together to form a stream frame. In addition, data visualization can be presented to users after the mining process has completed. The nodes can be divided into six categories: source node, record options node, field options node, graphs node, modeling node and output node. SPSS Modeler provides a different classification of clustering in that the modeling node; the data analysis process and the main set of nodes are linked together to complete the analysis of the data stream processing. Therefore, this study implements ODBC bridge into the Modeler data in order to establish the analysis process, and employs the SPSS Modeler 14 to analyze data using C&RT, followed by application of the *a priori* algorithm to analyze association rules.

5. Big data/data mining analysis results

5.1 Cluster analysis – customer segmentation

In the example of Spring season data in 2016, using selected seasonal data as an example, there two clusters found by clustering analysis. Cluster 1 is called the easy tasting new delicacy type male consumers group. Cluster 2 is called household sport consumers group. The purchasing motivation of each cluster is behavior and preference of outdoor activities. The purchasing habit is to choose brands/products with preference ranking and the source of product information comes from television advertising and purchase on-site from store shelves. The similarities and differences of the customers are as follows (Figure 2).

According to Table I, in Cluster 1 the easy tasting new delicacy type young consumers group is a specific customer segment. Their favorite outdoor activities are mountain climbing, go boating, climbing, camping, indoor aerobic exercise, etc. In Cluster 1, the purchase preference of outdoor products, including backpack and pant and usually purchase frequency is 6~12 months once in the average. Their purchasing brand preference includes brand image and imported brand about promotion and service frequency or the quick response of service personnel at the purchasing time. In contrast with Cluster 1, mature adult consumers group is more participated in golf, mountaineering, sailing and skiing, etc. In addition to, in Cluster 2, shoes, functional clothing and outdoor equipment are usually purchased in terms of brand and service preference including amiable, leisure and health, expertise service personnel and brand trust, etc. Their primary purchasing motivation is the desire to trust their acquainted product. On the other hand, marriage status and age are also segmentations on outdoor activities and purchase patterns. Young segment group is inclined to accept new outdoor activities and imported brand with brand image. Mature adult group has a nature of inertia, such as regular activities and brand trust on purchase behaviors. Catalog purchase is a difference on channel preference comparing with young segment.

Due to limited space, this study uses a seasonal data (Spring season) and selected two clusters for describing how to implement data mining results to design store layout and possible brand/product bundling sales alternatives on the case firm.

5.2 C&RT classification analysis – store layout process

The purpose of classification is to establish a model containing good information, which can be classified, analyzed and retrieved from the decision tree using the selected rules. Each decision point in the internal nodes and external nodes uses if-then logical rules to describe the data classification. This approach's main advantage is that it is very fast and it was soon learned that C&RT classification and regression trees could be used to classify data to the department as classification points, through a single input variable function. Separate data on each node were used to construct a dichotomous decision tree.

In order to get more information from the database about customers' preferences and their background, the C&RT classification analysis was performed using the variables of customer profile. A group of customer's attributes were obtained and then analyzed by association rules as possible consequences, along with classification variables that are composed of the following attributes: who (customer profile)/where (purchase way)/when (purchase time)/why (sales promotion way)/how (pay way)/ and what (purchase items). The results of classification analysis are shown in Figure 3.

According to C&RT analysis, the case firm then has a classification base for its product categories, brand attributes, fixture locations and customer traffic patterns for store layout and design. Initially, forming customer profiles is the first step in customer segmentations based on membership data, such as gender, age, membership status, education, living area, purchase behaviors, purchase preferences, leisure/sport characteristics, job category, etc. The first level of classification is for product categories. Either A or B area is a complementary product category classification, and belongs to complementary products that might have some cross-selling

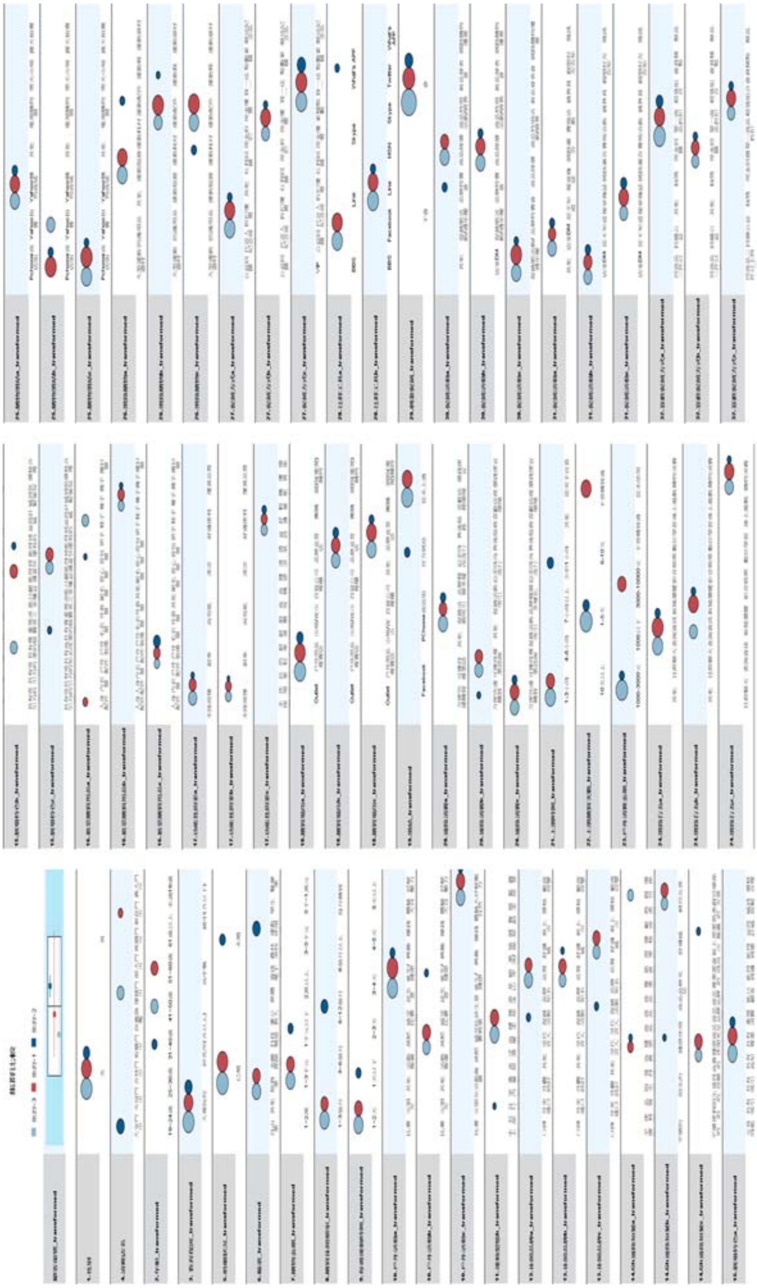


Figure 2.
K-means
cluster analysis

Cluster category		Cluster 1	Cluster 2
Name/description		(Easy tasting new delicacy type young consumers group) Single, young adult, high involvement in outdoor brands and products, individual participant, limit purchase capability, lower purchase frequency and higher promotion preference	(Mature adult consumers group) Married, mature adult, good purchase capability, higher service and quality preference, frequent outdoor activities, higher purchase frequency, higher involvement in outdoor information
Gender		Female (64.3%) Male (35.7%)	Female (54.8%) Male (45.2%)
Age		31~40	41~55
Marriage status		Single (52.1%)	Married (61.3%)
Job category		1. Service industry (29.3%) 2. Catering industry (25.5%) 3. Leisure sports (23.3%) 4. Financial and banking (15.3%) 5. Expatriate manpower (6.6%)	1. Service industry (22.6%) 2. Financial and banking (20.3%) 3. Engineering (19.05) 4. Public employees (15.05%) 5. Educational personnel (13.03%) 6. Freelancer (9.97%)
Purchase amount of money in average		Below 1,000 NT dollars	Between 1,000 to 3,000 NT dollars
Channel		Internet Physical store	Physical store Catalog Internet
Online purchase		Line Facebook	Yahoo Facebook
Outdoor activities	Purchase frequency	6~12 months once in average	1~3 months once in average
	Participate frequency	1. 1 day in average/week 2. Accept new things	1. Over 2 days in average/week 2. Regular activity
	Participate activities	1. Mountain climbing 2. Go boating 3. Climbing 4. Camping 5. Walking race 6. Indoor aerobic exercise 7. Water sports 8. River trekking	1. Golf 2. Mountain climbing 3. Indoor aerobic exercise 4. Walking race 5. Camping 6. Mountain climbing 7. Mountaineering 8. Sailing and skiing 9. Take a walk
Outdoor product preference	Similarity and difference	1. Backpack 2. Pant 3. Outdoor equipment	1. Shoes 2. Functional clothing 3. Outdoor equipment
Brand and service preference	Similarity and difference	1. Leisure and health 2. Promotion 3. Service frequency 4. Brand image 5. Imported brand 6. Quick response	1. Amiable 2. Leisure and health 3. Expertise service personnel 4. Quick response 5. Brand trust

Table I.
K-means clusters and categories

purchases patterns on these two categories, respectively. On the other hand, in contrast with A, B categories are incompatible with the A category product in that customers seldom purchase A and B products at the same time. Thus, product complementary and incompatibility is a classification tree result generated by C&RT analysis. In addition, C area is an open space for seasonal flagship and monthly sales products on display. The middle area is the location for hot sale products of A and B areas. The second level of C&RT analysis is brand classification, including private brand and store brand products. Private brand products were developed by

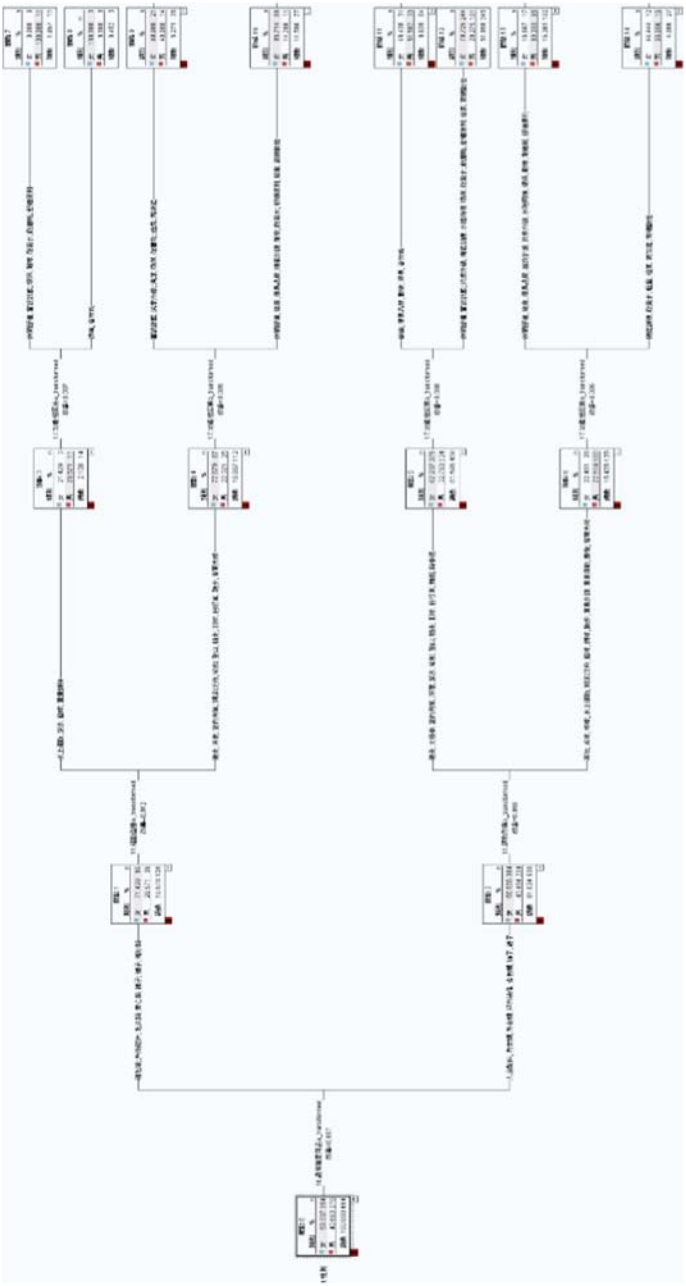


Figure 3.
C&RT classification analysis results

GoHiking itself and non-private brands include slotting brand, sale on consignment brand and royalty-free brand products. The third level of C&RT analysis generates a possible cabinet and frame-level layout by classifying product/brand purchase behaviors. By doing so, using space (product category/areas) and merchandise (brand/product items) to display (cabinet and frame levels), C&RT analysis provides a store layout design process blueprint for a 7,000 square feet selling space store (Figure 4).

Usually, in retailing, the straight floor plan process is an excellent store layout for most types of retail store. It makes use of the walls and fixtures to create small spaces within the retail store. Thus, the straight floor plan process is one of the most economical store designs. In Figure 5, by incorporating customer purchase knowledge, the GoHiking proposes a mixed floor plan. This mixed floor plan process incorporates the straight, diagonal and angular floor plans to create the most functional store layout and design. For example, this is a self-service type of store and so it offers excellent visibility for cashiers and customers. This store layout process also invites smooth movement and traffic flow through the store. The curves and angles of fixtures and walls make for a more spatially efficient store design, and the soft angles create better traffic flow throughout the store.

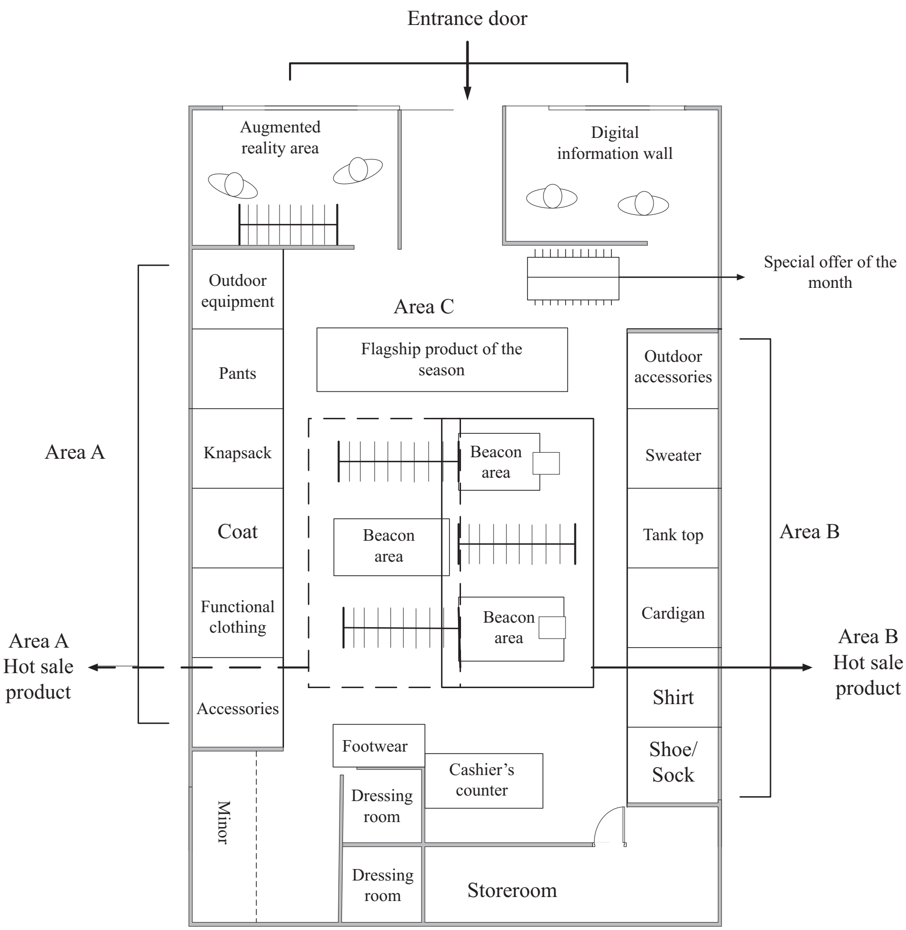


Figure 4.
Store layout and design process

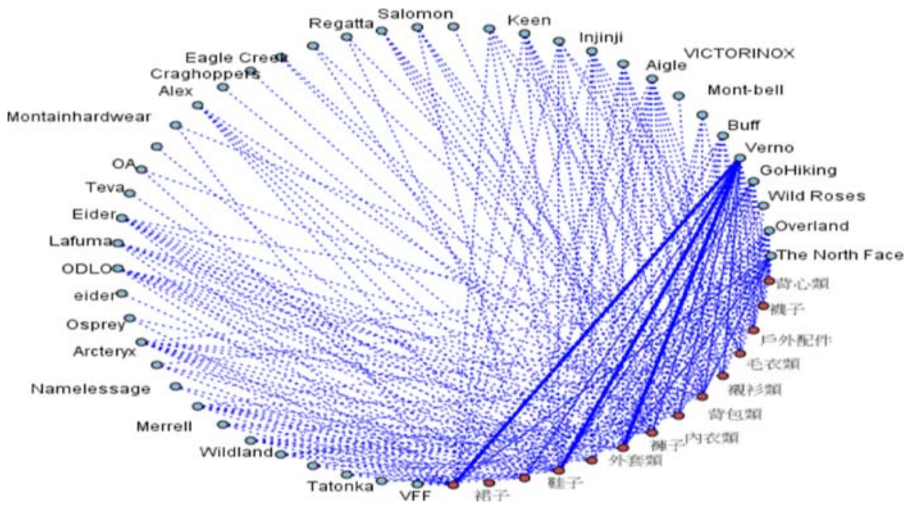


Figure 5.
The association diagram of brands/products crossing-selling bundling

5.3 Association rules–building brand/product bundling sales process management

By considering the customers' purchase behavior, we determine the segments of the brand/product mix for the case firm. Based on association rules and diagrams, we deploy useful information to determine the most effective associations for product mix for single brands.

Based on brand/product sales associations, we find certain product bundling for sales promotion for customer purchase behaviors. In Table II, this study established the minimum support range as from 1.4 to 6.58 percent, the minimum confidence as from 5.0 to 62.5 percent and selects 23 rules which have a lift value larger than 1. Thus, a concept of basket analysis, the

Table II.
The association rules of product mix for single brands

Rule	Sup	Con	Lift	Consequent	Antecedent
RA1	3.45	9.38	12.42	Sprayway	Sweater/tank top
RA2	3.45	12.5	9.66	F.N.ICE	Sweater/sock
RA3	4.10	7.90	9.15	Snow Travel	Coat/pant/outdoor equipment
RA4	4.53	7.14	8.28	Namelessage	Knapsack/outdoor equipment
RA5	6.58	9.83	7.01	VFF	Outdoor accessory/cardigan
RA6	3.88	13.89	6.78	Arcteryx	Shirt/sock
RA7	3.13	6.90	6.39	Wildland	Pant/outdoor equipment
RA8	4.10	5.26	6.10	Namelessage	Coat/pant
RA9	5.61	7.70	5.49	Lafuma	Footwear/functional clothing
RA10	3.02	7.14	5.09	VFF	Footwear/outdoor equipment
RA11	3.13	6.90	4.92	Lafuma	Knapsack/accessory
RA12	4.32	5.0	4.64	Wildland	Coat/pant/functional clothing
RA13	4.53	11.91	4.60	ODLO	Sock/shirt
RA14	3.13	10.35	4.57	Keen	Knapsack/outdoor equipment
RA15	3.67	5.88	4.54	F.N.ICE	Shoe/tank top
RA16	1.40	23.08	4.37	Rewoolution	Sock/cardigan
RA17	5.29	6.12	3.55	Injinji	Cardigan/sweater
RA18	4.64	9.30	3.45	Buff	Accessory/knapsack
RA19	4.10	5.26	3.05	Injinji	Coat/outdoor equipment
RA20	4.32	5.0	2.73	Eider	Accessory/pant/coat
RA21	3.13	20.69	2.09	GoHiking	Pant/outdoor equipment
RA22	4.32	62.5	1.87	Verno	Shoe/sock/shirt
RA23	4.32	62.5	1.87	The North Face	Tank top/cardigan

possible product bundling for a specific brand, is then found for sales promotion. On the other hand, in case of a specific brand, Sprayway, customers who are accustomed to purchasing sweaters can also buy tank tops, using certain price discounts for a second item as a rule of recommendation. In addition, customers who buy F.N.ICE sweater can also buy F.N.ICE socks through a promotion with coupons. Similarly, customers who purchase coats can be encouraged to buy pants and outdoor equipment using an extra dividend allowance that is paid by some specific credit cards with a Snow Travel product promotion consideration (Table II).

In addition, using terms of brand/product bundling in Figure 5 and Table III, and based on association rules analysis and 16 rules, some patterns of brand/product cross-selling, as well as possible integrated brands/products bundling rules, is then found for sales promotions process management. On the other hand, some customers (profile is as males/19–35 years old/single/online-to-offline (O2O) purchase mode/pay by credit card/junior membership) who are accustomed to purchasing Verno/outdoor equipment can also buy Buff/knapsack and Eider/accessory with a half-price discount for a cross-selling as a rule for sales recommendation. In addition, some customers (profile is as female/36–55 years old/married/physical store purchase mode/pay by credit card/senior membership) who purchase GoHiking/outdoor equipment are also encouraged to buy Wildland/pants and F.N.ICE/shoes/tank tops using membership discounts as a kind of complementally brand/products bundling.

In terms of bundling sales, the associations of product category on Cluster 1 including shirt, pant, outdoor equipment, knapsack and functional clothing, etc., are the possible product bundling. The easy tasting new delicacy type young consumers group accepts most of promotion ways on selling and purchases products both on physical and internet channels. Imported brand is a brand preference on this consumer segment. Cluster 2, the mature adult consumers group, has a product bundling pattern, such as coat, outdoor accessory, accessory and tank top, etc. They prefer discount for special item and extra dividend for sales promotions. Most channels of selling are accepted. Verno and GoHiking, two self-brands of case firm, shows that Cluster 2 has a brand loyalty on local brand. Accordingly, integrating the above knowledge from data mining results for the two cluster customers, this study concludes in Figure 6 with a BPM map for bundling sales

Rule	Brands	Product crossing-selling bundling
RB1	Verno/Buff/Eider	Pant/shoe/outdoor equipment/shirt/knapsack/accessory
RB2	GoHiking/Wildland/F.N.ICE	Coat/shoe/pant/accessory/outdoor equipment/outdoor accessory/tank top
RB2	Wildland/Sprayway	Knapsack/shoe/sweater/tank top/pant
RB4	Namelessage/Keen/Namelessage	Outdoor accessory/coat/outdoor equipment/shoe/sock/accessory/knapsack
RB5	Snow travel/Injinji/VFF	Coat/pant/outdoor equipment/outdoor accessory/accessory/shirt/shoe
RB6	Eider/Keen/Snow Travel	Outdoor equipment/coat/knapsack/functional clothing/accessory/outdoor accessory
RB7	Keen/Lafuma/Buff	Pant/accessory/shoe/functional clothing/knapsack/sweater/outdoor equipment
RB8	Arcteryx/Verno	Accessory/shoe/shirt
RB9	Sprayway/Namelessage/VFF	Outdoor accessory/outdoor equipment/shoe/cardigan/knapsack
RB10	VFF/Lafuma/Snow Travel	Footwear/coat/accessory/outdoor accessory/outdoor equipment/shoe
RB11	Buff/F.N.ICE/Eider	Knapsack/shoe/outdoor accessory/accessory/pant
RB12	Injinji/Namelessage/Eider	Tank top/coat/shirt/knapsack/outdoor accessory
RB13	F.N.ICE/Sprayway	Coat/outdoor equipment/shoe/knapsack/sweater
RB14	Rewoolution/Rewoolution	Sock cardigan/outdoor equipment/coat
RB15	The North Face/Keen/Namelessage	Accessory/knapsack/coat/shoe/outdoor equipment
RB16	ODLO/Injinji/Rewoolution	Sweater/sock/pant/knapsack/accessory/shoe

Table III. Brands/products crossing-selling bundling recommendations allied with product category, channel, promotions and product/brand bundling sales associations.

On the other hand, big data analysis is not only used to recommend brand/product cross-selling bundling but also to investigate customers' leisure/sport preferences in order to develop further new product/service bundling according to customers' needs, wants and demands for their outdoor/indoor lifestyle. For example, salespeople can recommend functional clothing to customers who purchases shirts along with products for sport walking and indoor aerobic exercise. We also found that functional clothing users also participated in other outdoor activities such as golf, mountaineering, water sports, river trekking, sailing and skiing. Some indoor and outdoor sport courses are also introduced to customers on the internet shop, apps, augmented reality and the digital information wall (DIW), together with product/service recommendations process management.

6. Managerial implications

(1) In regard to a whole picture of annual data analysis on a product line analysis, this study presents a multiple seasons comparison on case firms' product line development. In terms of Spring season, the products purchased by male customers

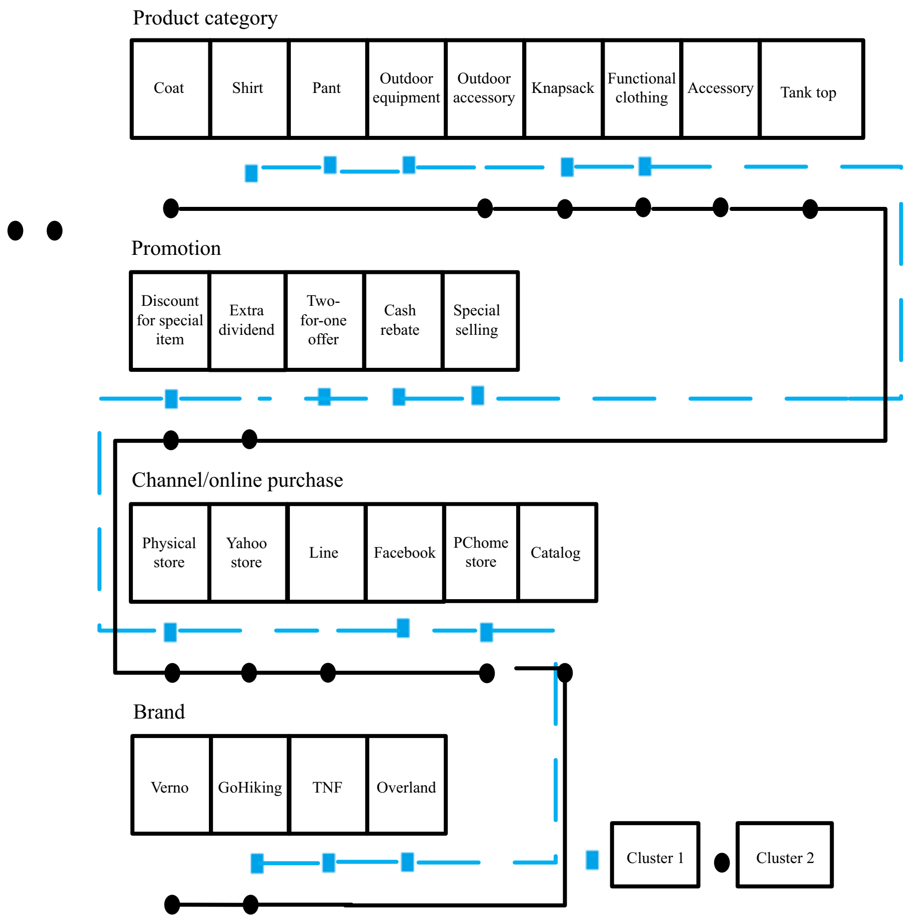


Figure 6.
Business process management map for bundling sales

in the Spring are still mainly based on clothing, pants, jackets and other products that keep warm, but the products that are matched are more diverse. The weather in Spring is changing and cold, so the warm-length long-sleeved pants are the preferred products for female customers. In terms of Summer season, Summer is a popular season for outdoor activities, and the characteristics of this season are also reflected in the hot-selling products. For example, various types of hats, long-sleeved clothes, gloves, sleeves, etc., that can be shaded, water bottles that add moisture, or materials that provide cool and ventilated cooling, are popular among female customers. The hot-selling products of male customers pay more attention to the functionality of the products and the comfort of sweating and quick-drying. Sunscreen is not the focus of their attention though. Thus, in the face of male customers, the staff of the on-site store should strengthen the professional knowledge of product features, materials, functions and matching sales products and increase the sales of on-site products to customers. In terms of Autumn season, the hot-selling products in the Autumn are similar to those in Spring, and most of them are mainly warm-keeping products, such as brush coats and thermal underwear. For female customers, case firm can match their own brands and their own products to increase consumer familiarity and product quality in this season. For the product mix of male customers, attention should be paid to the variety of product line content. For example, the case firm can emphasize the purchase of agent products for male customers, and donate its own products (long-sleeved underwear, long-sleeved thermal underwear and warm pants) and other products that are high-quality and attractive to male customers. In terms of Winter season, male customers have more diversified preferences for their products. In comparing with the Spring, Summer and Autumn seasons, they have high-level preferences and different combinations for other brands in addition to their own products. On the other hand, female customers are more loyal to their own products on this season. Products such as bristle jackets and thermal underwear are warm-keeping products. Compared with the Autumn product structure, more brand-name products will be purchased by both male and female consumers in this season. The data mining of annual data analysis on product line analysis is illustrated on Figure 7.

(2) In regard to a whole picture of annual data analysis on a brand development analysis, spectrum, originally used in physics, is currently used in brand development (Liao *et al.*, 2009). It allows firms to clearly understand where they are in the market position and where their competitors are. Thus, this study examines the case firm in the form of brand spectrum, and how to plan the brand development of the store layout in terms of strength and weak brand preference on an annual data mining analysis basis. With annual data mining results, we can see that due to the home advantage of the channel, the number of transactions of its own brand Verno accounted for more than 25 percent of the total number of transactions. In a further analysis, we selected case firm's own brand Gohiking and Verno, and selected an average single spending amount higher than Verno's spending, and whether Gohiking and Verno have cut-in possible for other brands. The selected brands include Wildland (WL), Namelessage (NL), TNF, FUSALP (FU), Overland (OL), Aigle (AI), Merrell, Cloudveil (CL), Keen, Eider (EI) and other ten brands for brand spectrum analysis. This study found that under the advantages of its own channel and price, Verno's sales volume is at its highest level; and following brands are Wildland, Namelessage, The North Face, FUSALP, Overland, Aigle, Cloudveil, Eider and Gohiking. In the various seasons analyzed, Wildland's consumption is second only to Verno's brand, but its average consumer price is about 25 percent

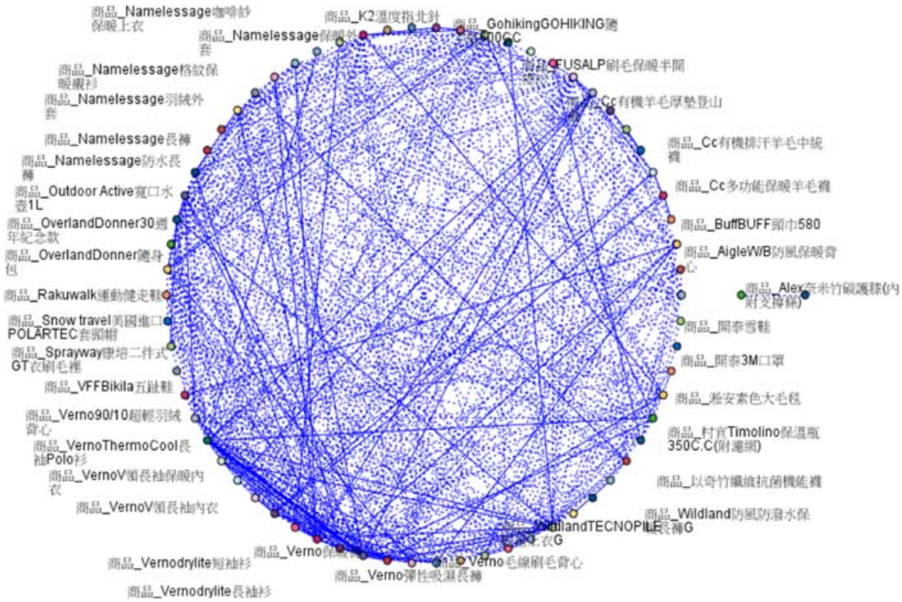
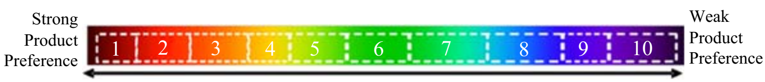


Figure 7.
The association diagram of annual data analysis on product line analysis

higher than Verno. The average consumer price of Namelessage is three times that of Verno, and the average consumer price of The North Face is 3.5 times that of Verno. Thus, for Gohiking and Verno, imitating the products of these brands will help directly increase the brand's profit margin. The annual brand development analysis results (brand spectrum) of case firm are illustrated on Figure 8.

(3) In regards to qualitative process management analysis, in addition to the quantitative analysis method using data mining, GoHiking also considered some qualitative indexes for store layout and service using store managers' experience and knowledge. These processes included: store appearance – interior: the store's atmosphere and décor is appealing to the consumer; the floor free of debris and appears clean, all light fixtures are working properly, the store is well-lit and the store appears to be full of inventory; customer Service: customers are greeted by a member of the staff upon entering the store; employees are easily identifiable with a uniform or name tags; the staff seems knowledgeable about the products they sell, the employees of this store are qualified to handle customer complaints, returns and other customer service issues directly and promptly and the staff consistently treats customers with respect; product offerings: products are well signed, labeled with



Spectrum 1: Verno
Spectrum 2: Wildland
Spectrum 3: Namelessage
Spectrum 4: The North Face
Spectrum 5: FUSALP
Spectrum 6: Overland
Spectrum 7: Aigle
Spectrum 8: Cloudveil
Spectrum 9: Eider
Spectrum 10: Gohiking

Figure 8.
A brand spectrum – annual data analysis on brand development analysis

prices and neatly displayed, there is a good variety and selection of products, the quality of merchandise is the highest available for the price; traffic flow: the store layout makes it easy to navigate around the store, each department within the store is clearly defined, the aisles are wide enough and free of boxes. By integrating quantitative and qualitative analysis, the case firm developed a store layout, sales and service mechanism for its branch stores.

(4) In the regard of sales power and profitability, cross-selling and one-stop shopping are businesses promotion goals in the retailing (Liao *et al.*, 2014). Such as inventory turnover, cash flow, return on space investment, associated purchase rate, sell-through rate, product gross profit, financial current ratio and slotting allowance, etc., are some indexes for measuring store sales power and profitability. Mostly, these indexes are embedded in a database system in order to help retailers to manage business operation online. Therefore, not only sales power but also profitability capability is another source of information for store layout and sales process management.

(5) In terms of electronic store displays process, both an augmented reality (AR) area and a DIW area technologies for delivering product information. AR is a live direct or indirect view of a physical, real-world environment whose elements are augmented (or supplemented) by computer-generated sensory input such as sound, video, graphics or GPS data. It is related to a more general concept called mediated reality, in which a view of reality is modified by computer. As a result, the technology functions by enhancing customer's current perception of reality. Artificial sales information about the store and its objects can be overlaid on the real world. DIW, the interactive window display has data on more than a thousand products, both outdoors and indoors, available in the store. Customers may select an item and the DIW then presents details on the selected item, such as price, color, location, measurements, etc. GoHiking wanted to emulate an online shopping experience while also making an impression in the store for O2O development. Another type of electronic display is Beacon, which is an intentionally conspicuous device designed to attract attention to a specific location. In the middle area of store, using Beacon combined with flashing lights or other indicators important information can be provided, such as the status of promotions, according to the price pattern of customer membership, or of cross-selling discounts as indicated on the case firm App. When used in such fashion, the Beacons can be considered a form of optical telegraphy. Simultaneously, all real-time data are uploaded to the cloud for further computing with data mining analysis.

(6) For store space process management, in terms of product display and sales, store space process management is a critical factor for retailers. For example, return on space investment is an index used to measure the efficiency of space operation on profitability. The case firm considered that space process management should determine how to lay out brands/products in store areas (by product category), in cabinet and frame level (by brand/product items) for maximized return on space investment. Thus, each area, cabinet and each frame-level layout becomes part of an overall picture for exploring how to manage each square foot of the allocated selling space.

(7) In terms of customer satisfaction, customers' needs, wants and demands are a sensitive and complicated aspect of market intelligence, if a retailer can understand customers and attempt to fulfill their wants and demands by providing exactly products/services suited to them, then customers could be more supportive

of the enterprise. During the process of development, from the store layout to the actual product sales, customers can only passively receive partial information and can only select from the products that are currently on sale in a store. No matter which type of brand/product, the customer cannot individually come up with a brand/product concept and then develop it for purchase satisfaction. Furthermore, buying what is available in a store does not mean that customers are satisfied with the current brand/product, because the customers' preferences and experiences were not considered in developing the brand/product bundling with customer segmentations, and as a result, they can only choose from the brands/products that are offered. As a result, retail firms have the responsibility to develop brand/product sales through an appropriate store layout and effective service mechanisms that fulfill their customers' needs, wants and demands, since this will increase the retail firm's competitiveness and is an essential criterion for earning higher profits.

7. Conclusion

Though there are many businesses in the retailing market, only a few really succeed with their business models. Only by creating unique sales strategies and by having a competitive advantage in regard to brand/product sales, can businesses effectively cultivate customer satisfaction and succeed in getting each customer to purchase and engage in repeat buying/purchase. In this way, retailers can have a greater chance of success by implementing an intelligent BPM approach. Thus, this study considers an outdoor articles chain store's buying behavior patterns, including customer purchase preferences and customer purchase behaviors, in order to generate a store layout and sales process management alternatives for retailers. These research results provide retailers with useful references to find potential store layouts and sales process management methods, to develop possible brand/product cross-selling bundling, to propose effective promotion activities and to earn higher profits through knowledge extraction using data mining in the big data and BPM development era.

References

ABPMP (2013), *Guide to the Business Process Management Common Body of Knowledge* (BPM CBOK), V.3.0, Association of Business Process Management Professionals, Springfield, IL.

Agrawal, R. and Shafer, J. (1996), "Parallel mining of association rules", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8 No. 6, pp. 962-969.

Agrawal, R., Imilienski, T. and Swami, A. (1993), "Mining association rules between sets of items in large databases", *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 207-216.

Bernardo, R., Galina, S.V. and Dallavalle de Pádua, S.I. (2017), "The BPM lifecycle: how to incorporate a view external to the organization through dynamic capability", *Business Process Management Journal*, Vol. 23 No. 1, pp. 155-175.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984), *Classification and Regression Trees*, Wadsworth International Group.

Codd, E.F. (1970), "A relational model of data for large shared data banks", *Communications of the ACM*, Vol. 13 No. 6, pp. 377-387.

Cordier, F., Seo, H. and Magnenat-Thalmann, N. (2003), "Made-to-measure technologies for an online clothing store", *IEEE Computer Graphics and Applications*, Vol. 23 No. 1, pp. 38-48.

Griffith, D.A. (2005), "An examination of the influences of store layout in online retailing", *Journal of Business Research*, Vol. 58 No. 10, pp. 1391-1396.

Guo, L., Sharma, R., Yin, L., Lu, R. and Rong, K. (2017), "Automated competitor analysis using big data analytics: evidence from the fitness mobile app business", *Business Process Management Journal*, Vol. 23 No. 3, pp. 735-762.

Hammer, M. (2007), "The process audit", *Harvard Business Review*, Vol. 85 No. 4, pp. 111-123.

Hu, H. and Jasper, C.R. (2007), "A cross-cultural examination of the effects of social perception styles on store image formation", *Journal of Business Research*, Vol. 60 No. 3, pp. 222-230.

Jayaraman, R. (2016), "Project cost control: a new method to plan and control costs in large projects", *Business Process Management Journal*, Vol. 22 No. 6, pp. 1247-1268.

Khlif, W. (2017), "A methodology for the semantic and structural restructuring of BPMN models", *Business Process Management Journal*, Vol. 23 No. 1, pp. 16-46.

Liao, H.S., Ho, H.H., Hsian and Yang, F.C. (2009), "Ontology-based data mining approach implemented on exploring product and brand spectrum", *Expert Systems with Applications*, Vol. 36 No. 8, pp. 11730-11744.

Liao, H.S., Wen, C.H., Hsian, P.Y., Li, C.W. and Hsu, C.W. (2014), "Mining customer knowledge for a recommendation system in convenience stores", *International Journal of Data Warehousing and Mining*, Vol. 10 No. 2, pp. 55-86.

Lindman, M., Pennanen, K., Rothenstein, J., Scozzi, B. and Vincze, Z. (2016), "The value space: how firms facilitate value creation", *Business Process Management Journal*, Vol. 22 No. 4, pp. 736-762.

Maita, R.C., Martins, L.C., López Paz, C.R., Marques Peres, S. and Fantinato, M. (2015), "Process mining through artificial neural networks and support vector machines: a systematic literature review", *Business Process Management Journal*, Vol. 21 No. 6, pp. 1391-1415.

Niehaves, B., Poeppelbuss, J., Plattfaut, R. and Becker, J. (2014), "BPM capability development – a matter of contingencies", *Business Process Management Journal*, Vol. 20 No. 1, pp. 90-106.

Ohlsson, J., Han, S. and Bouwman, H. (2017), "The prioritization and categorization method (PCM) process evaluation at Ericsson: a case study", *Business Process Management Journal*, Vol. 23 No. 2, pp. 377-398.

Paim, R., Caulliraux, H.M. and Cardoso, R. (2008), "Process management tasks: a conceptual and practical view", *Business Process Management Journal*, Vol. 14 No. 5, pp. 694-723.

Palmberg, K. (2010), "Experiences of implementing process management: a multiple-case study", *Business Process Management Journal*, Vol. 16 No. 1, pp. 93-113.

Psomas, E.L., Fotopoulos, C.V. and Kafetzopoulos, D.P. (2011), "Core process management practices, quality tools and quality improvement in ISO 9001 certified manufacturing companies", *Business Process Management Journal*, Vol. 17 No. 3, pp. 437-460.

Ripley, B.D. (1996), *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge.

Trkman, P., Mertens, W., Viaene, S. and Gemmel, P. (2015), "From business process management to customer process management", *Business Process Management Journal*, Vol. 21 No. 2, pp. 250-266.

Vrechopoulos, A.P., O'Keefe, R.M., Doukidis, G.I. and Siomkos, G.J. (2004), "Virtual store layout: an experiment comparison in the context of grocery retail", *Journal of Retailing*, Vol. 80 No. 1, pp. 13-22.

Wang, Y.F., Chuang, Y.L., Hsu, M.H. and Keh, H.C. (2004), "A personalized recommender system for the cosmetic business", *Expert Systems with Applications*, Vol. 26 No. 1, pp. 42-52.

Witten, I.H. and Frank, E. (2000), *Data Mining. Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, San Francisco, CA.

Corresponding author
Shu-hsien Liao can be contacted at: michael@mail.tku.edu.tw