IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 4 February 2023, accepted 25 March 2023, date of publication 31 March 2023, date of current version 5 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3263729

RESEARCH ARTICLE

Probabilistic Customer Purchase Evolution Graph

YUNG-TZU J. LIN[1], CHUAN-YI CHANG[2], SHEIN-YUNG CHENG[2], AND MENG-YUN T. LIN[1]
[1]Department of Business Management, National Taiwan University of Science and Technology, Taipei City 106335, Taiwan
[2]Department of Information and Computer Engineering, Chung Yuan Christian University, Taoyuan City 320314, Taiwan

Corresponding author: Chuan-Yi Chang (chang.maverick@gmail.com)

ABSTRACT Following evolutionary theory, this study defines the commodity consumption gene of the customer journey in retail transactions to discuss the distribution and evolution process of the consumer group. First, through relevant retail research, each transaction is defined as a data point in the SPC space (sales-product-customer). The customer journey is a pivot transformation of the transaction data points in the SPC space. Customer purchase products are defined as consumption genes in the customer journey, forming customer consumption species. Furthermore, evolutionary operations with probabilities between consumer species (genes) can be used to analyze the evolution of each customer's purchase consumption over time in the retail database. An algorithm for the Customer Purchase Evolution Graph (CPEG) is proposed. To prove practicability, nearly 300,000 actual transactions from over 27,000 consumers are used to establish the CPEGs with evolutionary probabilities of the overall customers and the CPEG of the SVIP. The CPEG of all customers can be used to determine the main consumption distribution, main consumption starting point (first purchase) behavior, and repurchase behavior of all customers and new customers. The CPEG of SVIP customers can further reveal the main consumption genes of mature customers (species) and the evolution process of their consumption journey. These findings can be of specific help in a company's commodity strategy and operational marketing.

INDEX TERMS Retail transaction, repurchase, evolution theory, customer journey.

I. INTRODUCTION

A. PURCHASE AND REPURCHASE

Retailing involves the sale of certain products or services to customers. In retail, customers buying a certain product have a behavioral tendency toward purchase intention [1]. Repurchase intention or repeat patronage indicates the possibility that customers will repurchase products or avail services of the same brand [2]. Therefore, investigating customer repurchase behavior and constructing a model of multiple customer transactions is critical. The repurchase intentions of customers can be measured from three aspects: (a) intentions to repurchase, which measures customers' willingness to repurchase the company's products or services in the future, and (b) an indicator of customers' future behavioral intentions. (b) Primary behavior: This behavior includes the number of purchases, frequency, amount, and quantity of customers. (c) Secondary behaviors: This behavior refers to behaviors in which customers help the company to introduce, recommend, and build reputation [3]. Among the three behaviors, only the primary behaviors can be measured using actual transaction data. Each customer's primary behavior changes over time. Thus, each customer has a distinct journey. This was the basic starting point of this study.

B. CUSTOMER JOURNEY

A customer's process of purchasing goods and services is a well-defined service process called a customer journey [4]. The customer journey gathers information about the process and experience from the customer's perspective through the product or service offered by the brand [5]. Formally, this customer journey can be described as a repeated interaction between a customer and brand (service provider) [6]. The aforementioned interactions or communications between customers and brands can be defined as "touchpoints" [7], [8], [9]. A touchpoint represents an abstract form of the customer experience (the customer's product purchase in this study) [10]. Thus, the customer journey can

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Tan.

be visualized as a well-defined service process with clear beginning and end [4]. In the customer journey of purchase analysis, the initial touchpoint can be regarded as the first purchase behavior of a new customer and the final touchpoint can be regarded as the purchasing characteristics of a loyal customer.

C. EVOLUTION THEORY

During the customer journey, the goods or services purchased by customers change over time and are affected by the external environment (including shopping mall environment and product mix); therefore, the customer journey can be regarded as an evolutionary process of customer purchase. Formally, evolution refers to changes in heritable traits between generations [11], [12]. The evolution of organisms occurs through changes in their heritable traits. In biology, hereditary traits are controlled by an organism's genome (genetic material), known as the genotype. On the other hand, the completely observable behavior or traits that result from the influence of these genes are called phenotypes [13]. A species is a group of individuals carrying the same genes. All the genotypic variations observed in the world today form a variety of evolutionary operations, such as mutation, genetic drift, gene flow and natural selection, called ''force of evolution'' [14]. Each evolutionary operation can be represented as a transition that switches genome x to genome y with a transition probability in the form of a Markov chain [15].

Scholars believe that human culture can also be explained by evolution, which is called cultural genes or memes [16]. Therefore, some scholars have studied the evolution of an enterprise, which will also undergo internal evolution due to the influence of the external environment [17], [18], [19]. In retail applications, there is an opportunity to define the customer's consumption gene as the combination of products (categories) consumed by customers, so as to follow the evolution theory to discuss the evolution process of customer purchase.

D. RESEARCH OBJECTIVES

This study applies evolution theory to the investigation of customer journeys in retail purchase transactions and proposes a customer purchase evolution graph (CPEG). For the study of retailing, the retailing elements are formulated in Section II, and an SPC data space for retailing is constructed for customer journeys in Section III. Section IV defines the consumption genes of customer purchases and the transition probabilities between consumption species to represent the customer purchase evolution graph (CPEG). CPEG can be used to investigate customer purchase insights for real retail transactions. Finally, Section VI concludes the study.

II. FORMULATION OF RETAILING ELEMENTS

A. PRODUCT SET AND CUSTOMER SET

Retailing refers to the sale of products or services to customers. A brand not only launches a series of products or services but also delivers six meanings to consumers in a series of features: (a) attributes, (b) value, (c) user, (d) benefits, (e) personality, and (f) users [20]. Product (expandable to brand) knowledge can be constructed from concrete (products) to abstract (brands) with these six levels. Moreover, the formal definition is obtained as follows:

Definition 1: Product $p \in P$ can be defined as a vector of the specific product features p_j: product $p = [p_1, p_2, \ldots, p_j, \ldots]$.

For example, the product feature set of clothes $P_a = \{\text{coat, pants, } \ldots\}$, and the product attribute of clothes gender $P_b = \{\text{male, female, child}\}$. Similarly, customer features can come from the scope (levels, from abstract to concrete) of the consumption market: (a) classified by geographic variables, such as area, city size, population density, and climate; (b) classified by demographic variables, such as gender, age (corresponding to the above P_a), family size, income, occupation, education, and nationality; (c) classified by social class, lifestyle, personality, and other psychological variables; (d) classified by behavioral variables, such as use timing, benefits, user status, and purchase preparation [21].

Definition 2: A customer is a vector of its features: customer $c = [c_1, c_2, \ldots, c_i, \ldots] \in C$, where the specific customer feature c_i is the feature of certain customers.

For example, c_i can be a customer feature set of customer features in a certain area, such as customer age group $C_a = \{\text{children, teenagers, adults, senior citizens}\ldots\}$, or customer features of sales channels, such as customer value $C_b = \{\text{one-time customers, repeat customers, frequent visitors, VIP,}\ldots\}$. The cardinality of set A is denoted as n(A). Thus, the cardinality of product set P and customer set C are n(P) and n(C), respectively.

B. CP PLANE

Product set P and consumer behavior set C can be used to construct a Cartesian product CP plane.

Definition 3: A CP plane from product set P and customer set C can be formally defined as a Cartesian product of C and P, that is, $\{(c_i, p_j): c_i \in C \text{ and } p_j \in P\}$.

With the partial ordering of the customer and product sets, C and P form two independent coordinate axes: the consumer (C) axis from customer to community and the product (P) axis from product to brand. In our previous research, we proposed the use of these two coordinate axes to weave a CP plane, as shown in ***Figure 1.*** [22], [23]. Each $CP = C \times P = \{(c_i, p_j)\}$ cell located in the CP plane represents a coordinate sequence pair (c_i, p_j) with different degrees of detail. There were $n(P) \times n(C)$ cells in total.

C. SALES BEHAVIORS

In the retail process, sales staff interact with customers to persuade them to purchase. The research divided the sales process into seven stages: (a) Prospecting: identifying potential buyers of products or services. (b) Pre-approach: Collects relevant information from potential buyers to prepare

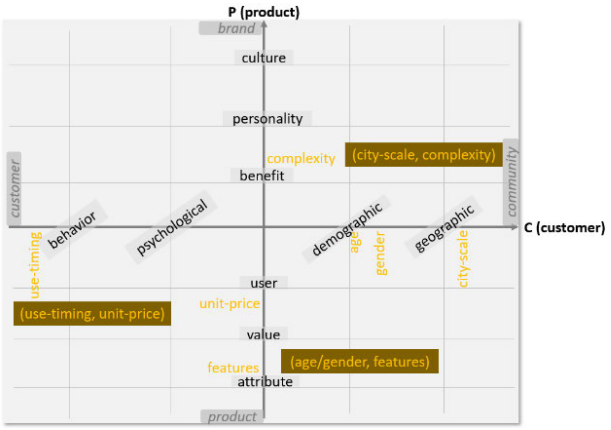

FIGURE 1. The weaved CP plane.

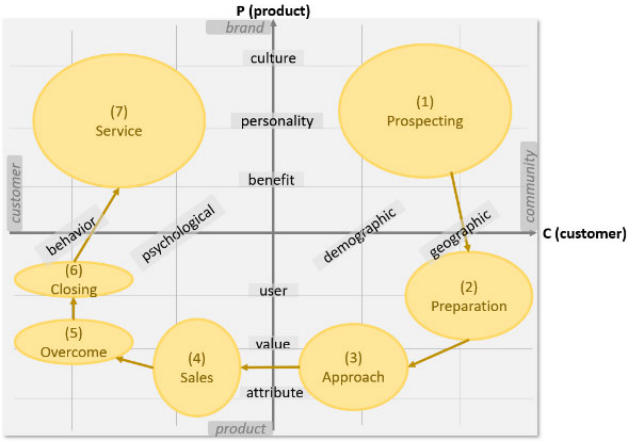

FIGURE 2. Seven stages of selling over CP plane.

sales visits. (c) Approach: Starting selling to specific buyers. (d) Sales Presentation: Provide the characteristics and advantages of products or services to awaken purchasing desires. (e) Handling Objections and Overcoming Resistance: Strive to overcome consumers' refusal and rejection. (f) Closing: Make customers complete purchases using suitable and effective methods. (g) Post-Sale Follow-Up: Continue to emphasize after-sale customer satisfaction [24], [25], [26]. Following the above derivation, these customer consumption behaviors in the sales process can also be formally defined.

Definition 4: Given a product set $P = \{ p_j \}$ and a customer set $C = \{c_i\}$, the customer consumption behavior $S = [\ s_k(c_i, p_j)\ ]$ in the sales process can be expressed as a series of purchase actions s_k, where $s_k(c_i, p_j)$ represents the customer c_i performing some actions s_k on product p_j.

Through the CP plane, this decision-making process can be easily expressed as the seven steps of the sales process (S) activity above the CP plane, as shown in ***Figure 2.*** In ***Figure 2***, the seven-stage sales process is clearly illustrated in the clockwise direction on the CP plane.

III. RETAIL SPC DATA SPACE AD ITS TRANSFORMATION

A. SPC DATA SPACE

The CP plane encircles all elements of products (P) and customers (C). Moreover, the entire sales process (S) of retailing can be carried out on this CP plane. Therefore, a complete SPC model was proposed as the research foundation for retail data analytics.

Definition 5: Given a customer set $C = \{ c_i \}$ and a product set $P = \{p_j\}$ of retailing, the corresponding SPC model can be constructed through the formulation of SPC data space (S, P, C), where the sales set $S = \{s_k(c_i, p_j)\}$ contains all possible customer consumption behaviors $s_k(.,.)$'s.

As ***Figure 2*** shows, the coordinatized CP plane can be divided into four quadrants: Brand, Service, Sales, and Marketing, representing quadrants I, II, III, and IV respectively in ***Figure 2.*** In the 1st Brand quadrant, corporate prospects are potential products (abstract brand) and potential (general) customer communities. (Step 1 of the sales process) After that, sales prepare the product and try to approach customers (attracted from possible community) through some channels in the 4th Marketing quadrant (such as physical stores or e-commerce sites). (Steps 2 and 3) When the real sales begin (in the 3rd Sales quadrant), the salesperson will perform a series of procedures (Steps 4-6) to prompt (specific) products to (specific) customers. After completing the purchase, the brand (abstract product) needs to take follow-up actions to the customer in the 2nd Service quadrant. (Step 7) Obviously, different quadrants focus on different research targets, and the coordinates of the CP plane may vary depending on the application.

B. RETAIL TRANSACTION DATABASE IN SPC SPACE

As previously indicated, retail is a sales activity in which retailers sell products or abstract services (P) directly to consumers or end users (C) at certain prices in certain channels (e.g., stores). During period ($T = \{t_i\}$), transactions record these actual consumer behaviors within a database in the form of SPC space.

Definition 6: Given an SPC data space (S, P, C) with customer set $C = \{c_i\}$, product set $P = \{p_i\}$ with some product prices ($M = \{ m_i \}$), and sales set $S = \{ s_i(c_i, p_i)\}$, a retail transaction database D within the period $T = \{t_i\}$ is defined as a collection of retail transaction d_i's: $D = \{ d_i = [t_i, s_i, c_i, p_i, m_i]\}$.

To study retail transactions in the (S, P, C) data space, one small real transaction database is considered as an example.

Example 1. Table 1 shows transaction D_1 for 20 transactions in a real retail business in 2020. The SPC data space is $S_1 = \{s_i\} = \{$shop-1, …, shop-5$\}$, $P_1 = \{$product-1, …, product-19$\}$, and $C_1 = \{$customer-1, …, customer-5$\}$ from the channels, product, and customer, respectively. Note that the field category still represents more general features (tops, skirts, pants, dresses, coats, shoes) of products along the P-axis. In addition, the fields datetime and invoiceNo represent the time tag t_i's in time period T_1. The field quantity and amount represent the consumption measures of

TABLE 1. Retail transaction database D1 in *Example 1*.

Datetime	invoiceNo	channel	product	category	quantity	amount	customer
20200403	C004020299	shop-7	product-6	category-3	1	3509	customer-1
20200403	C004020299	shop-7	product-16	category-1	1	2448	customer-1
20201101	C011023902	shop-8	product-2	category-1	1	4311	customer-2
20201101	C011023902	shop-8	product-3	category-2	1	2869	customer-2
20201108	C011027813	shop-8	product-9	category-3	1	3590	customer-2
20201108	C011027813	shop-8	product-18	category-1	1	2792	customer-2
20200902	C009017164	shop-5	product-7	category-3	1	2657	customer-3
20200913	C009022817	shop-5	product-19	category-1	1	1791	customer-3
20201024	C010030564	shop-5	product-10	category-4	1	6464	customer-3
20200609	C006027867	shop-1	product-4	category-3	1	3224	customer-4
20200609	C006027867	shop-1	product-5	category-3	1	3224	customer-4
20200609	C006027867	shop-1	product-8	category-3	1	3224	customer-4
20200928	C009031421	shop-1	product-11	category-5	1	7632	customer-4
20200928	C009031318	shop-3	product-17	category-1	1	3141	customer-4
20200117	C001039291	shop-2	product-14	category-1	1	1328	customer-5
20200117	C001039291	shop-2	product-15	category-1	1	1612	customer-5
20200413	C004023508	shop-4	product-12	category-1	1	1328	customer-5
20200525	C005035872	shop-4	product-13	category-1	1	1612	customer-5
20200904	C009018332	shop-6	product-1	category-1	1	3058	customer-5
20200904	C009018332	shop-6	product-8	category-3	1	2902	customer-5

the transaction (s_i, p_i, c_i) points of the SPC space. These 20 transactions in D_1 with time tags can be described as 20 time-tagged points scattered in a three-dimensional (S, P, C), as ***Figure 3*** shown.

Since a transaction is the result of one purchase, regardless of whether the purchase object is a product or service, this record is a variable of retail purchase behavior and has nothing to do with the other three consumer behavior (psychological, demographic, and geographical) factors. Only when the customer is a member of the brand and their basic data, such as gender or age has been collected, can the other three factors be processed. The following will begin the investigation of some in-depth customer behaviors from the viewpoint of the customer journey.

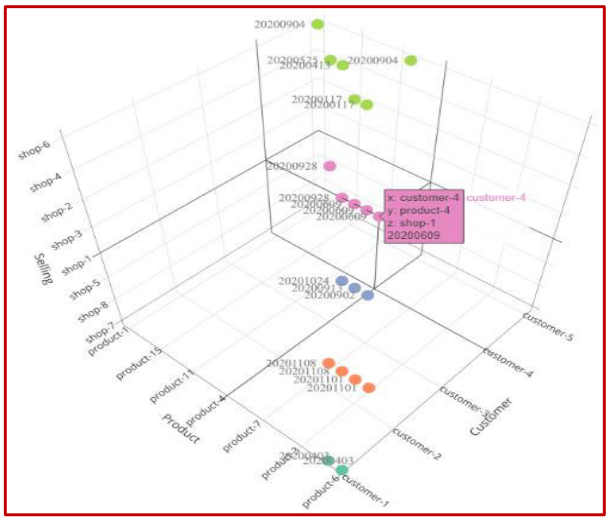

FIGURE 3. PM knowledge space of Example 1.

C. PIVOT TRANSFORMATION AND CUSTOMER JOURNEY

In a data space, it is possible to perform data flow transformation, which is a mapping to move data from a data source to a data destination [27]. For the SPC data space, one type of useful data flow transformation can be defined as follows:

Definition 7: For a retail transaction database $D = \{d_i = [t_i, s_i, c_i, p_i, m_i]\}$ in an SPC data space, a type of synchronous transformation with pivot operation on one specific axis (axis A, which may be S, P, or C here), called the axis-oriented pivot transformation $P_A(D) = [\ a_i(D)\ ]$, turns (aggregate) values from one dimension into a (new) dataset.

Under this definition, the SPC data space may possess three types of pivot transformations: sales(operation)-oriented pivot transformation $P_s(D)$, product(brand)-oriented pivot transformation (projection) $P_p(D)$, and customer-oriented pivot transformation (projection) $P_c(D)$. The first two pivot transformations $P_s(D)$ and $P_p(D)$ can be used to examine retail from the perspective of sales/operations and brand/products. This study focuses on the customer's viewpoint, $P_c(D) = [\ c_i(D)\ ]$.

The customer-oriented pivot transformation (COPT) $P_c(D)$ projects all transactions onto the customer axis, which can then be used to build all customer models. As ***Figure 4*** shows, $P_c(D_1) = [\ c_i(D_1)\ ]$ aggregates all transaction data in D_1 related to customer-i in $P_c(D_1)$.

Proposition 1: (a) $P_c(D)$ is not unique. (b) Each transaction (point) in $P_c(D)$ has the form $c_i(s_k, p_j)$.

Proof. (a) Different operations of $d_i = [t_i, s_i, c_i, p_i, m_i]\}$ will produce different types of aggregation results. (b) Each transaction has the form $d_i = [t_i, s_i, c_i, p_i, m_i\ ]$, which can be easily rewritten as $c_i(s_i, p_i)$. □

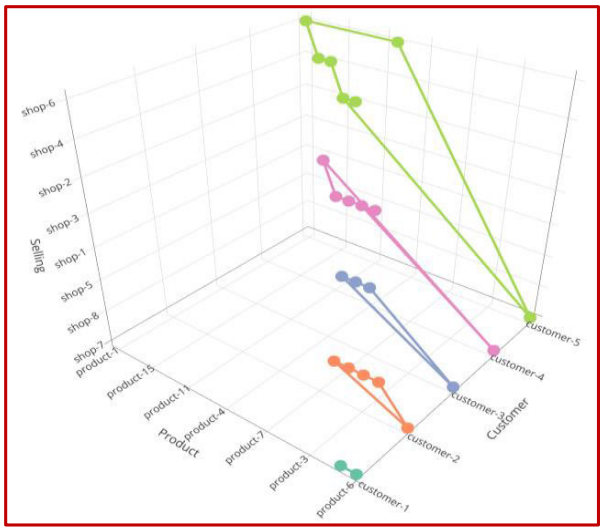

FIGURE 4. Customer-Oriented Pivot Transformation of D1 to Customer Axis.

In the SPC space, the transaction format $c_i(s_k, p_j)$ creates the aforementioned interaction/communication touchpoints between the customer (c_i) and the product/service provider (p_j) [7], [8], [9]. A formal definition of the customer journey can then be obtained.

Definition 8: For a retail transaction database $D = \{d_i = [t_i, s_i, c_i, p_i, m_i]\}$ in an SPC data space, the customer journey of customer c_i is defined as a sequence of touchpoints: $C_J[c_i] = [c_i(s_k,p_j)] = P_c(D)[c_i]$, which implies that the customer journey can be obtained through pivot transformation.

In practice, time tags are used in $C_J[.]$. For *Example 1*, $P_c(D_1)$[customer-1] = [(2020/11/01, shop-8, product-2/3), (2020/11/08, shop-8, product-9/18)], which will make the ''engaging story'' of the customer journey more alive [28].

IV. CUSTOMER PURCHASE EVOLUTION GRAPH

A. CUSTOMER CONSUMPTION GENE

Among the many COPTs mentioned above, the most iconic transformation worthy of in-depth discussion is how to extract a customer's purchasing characteristics from the customer journey [$c_i(s_k,p_j)$]. This is an attempt to treat customer-purchasing characteristics as a biological heritable trait, thereby introducing the evolutionary concept of organisms. [29].

The genotype (gene set within an organism's genome) biologically controls inherited traits. The interaction of genotypes with the environment forms observable traits of organisms, which is also called the phenotype [13], [30]. Similarly, from the perspective of human culture, many behaviors can be explained by evolutionary memes, whereas a set of cultural genetic factors evolve through the process of duplication (imitation), mutation, and selection [31]. Under the concept of social evolution, the customer consumption gene $P_{cg}(D)$ in retail is a kind of cultural gene in which genotype (purchasing motivation, etc.) cannot be seen explicitly, but the phenotype (purchased product/category) is clearly recorded in a transaction.

Definition 9: For a retail transaction database $D = \{d_i = [t_i, s_i, c_i, p_i, m_i]\}$ in an SPC data space, the consumption gene of customer c_i is defined as a type of COPT, P_{cg}:$C_G[c_i] = P_{cg}(D)[c_i] = \{p_j: c_i(s_k,p_j) \in C_J[c_i]\} \equiv \gamma_{ci}$, $\forall c_i \in C$.

Here, the products p_j are used to define the customer gene $C_G[c_i]$, which is more often defined in category g_j practically, to describe customer traits more generally. As shown in ***Table 1*** of *Example 1*, the customer gene of customer customer-1 is $P_{cg}(D_1)$[customer-1] = {category-1, category-3} = [1, 0, 1, 0, 0, 0] = 101000, with each bit coded as the category [tops, skirts, pants, dresses, coats, shoes]. To pack all the consumption genes of customers in a vector, $P_{cg}(D_1)$ = [101000, 111000, 101100, 101010, 101000] was obtained. Notably, $C_G[c_1] = 101000 = C_G[c_5]$, which defines the species.

Definition 10: For a retail transaction database, $D = \{d_i = [t_i, s_i, c_i, p_i, m_i]\}$, the consumption species of the customer gene γ_k $C_S(D, \gamma_k) = \{c_i: C_G[c_i] = \gamma_k, \forall c_i \in C\}$.

In biology, a species is a group of individuals with the same genes that can reproduce. For D_1 in *Example 1*, this definition divides customer set C into four species: $C_S(D_1, 101000) = \{c_1, c_5\}$, $C_S(D_1, 111000) = \{c_2\}$, $C_S(D_1, 101100) = \{c_3\}$, and $C_S(D_1, 101010) = \{c_4\}$, within which customers in the same species have the same consumption behavior.

B. EVOLUTIONARY OPERATIONS OF CUSTOMER CONSUMPTION GENE

Similar to biological evolution, evolutionary operations such as duplication, mutation, and recombination of customer consumption genes result from the mutual influence of customer consumption habits and changes in customer product consumption habits. Evolution occurs from one generation to another. First, this type of generation should be defined.

Definition 11: For a retail transaction database $D = \{d_i = [t_i, s_i, c_i, p_i, m_i]\}$, consumption generation is defined in a specific time period; that is, $D[T] = \{d_i = [t_i, s_i, c_i, p_i, m_i], \forall t_i \in T\}$.

With the definition of generation T, the customer (consumption) genes aggregate all the product (category) traits for all customers in the transaction database D[T]. For two consecutive time periods (generations), T_1 and T_2, the in-between evolution can be defined.

Definition 12: For two consecutive generations T_1, T_2, and $T_1 \subseteq T_2$, the three kinds of evolutionary operations (φ) of transaction database $D = \{d_i = [t_i, s_i, c_i, p_i, m_i]\}$ can be defined as $\varphi(D, T_1, T_2, \gamma_k, \gamma_k')$, often symbolized as $\varphi(D, T_1 \rightarrow T_2, \gamma_k \rightarrow \gamma_k')$, which means (a) duplication (φ_d) if $\gamma_k' = \gamma_k$, (b) mutation (φ_m): $\gamma_k \rightarrow \gamma_k'$, $\gamma_k \subseteq \gamma_k'$ or $\gamma_k' \subseteq \gamma_k$, (c) recombination (or crossover, φ_c): $\gamma_k \rightarrow \gamma_k'$, $\gamma_k' = (\gamma_k' \setminus \gamma_k) \cup (\gamma_k' \cap \gamma_k)$. The evolution probability of the evolutionary operation φ is defined as $P_r(\varphi(D, T_1 \rightarrow T_2, \gamma_k \rightarrow \gamma_k')) = nC(\gamma_k') / nC(\gamma_k)$.

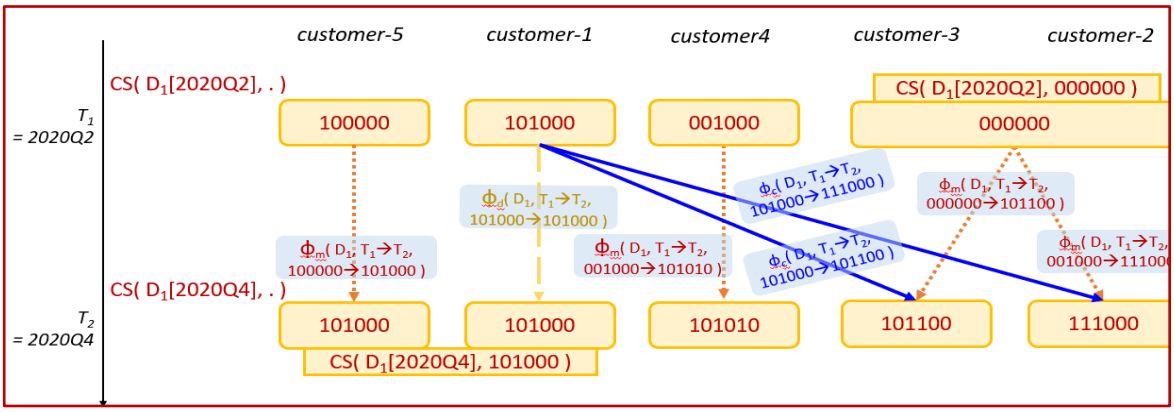

FIGURE 5. Evolution of customer-consumption genes.

When illustrated by *Example 1*, ***Figure 5*** shows the evolution from generation T_1 = 2020Q1 to T_2 = 2020Q4. Four different customer species are obtained in two generations, and three kinds of evolution can be interpreted as follows: (a) Duplication 101000→101000 means reservation with the same clothing; (b) Mutations 100000→101000, 001000→101010, 000000→ 101100, and 000000→111000 indicate the additive dressing of the same customers; and (c) Crossovers 101000→111000 and 101000→101100 demonstrate the transfer of dressing traits from one customer (e.g., customer-1) to another customer (e.g., customer-3). Because the number of customers is too small in *Example 1*, the calculation of the evolution probability is illustrated using practical cases in the next section.

C. EVOLUTION OF CUSTOMER CONSUMPTION GENES

The evolution direction of ***Figure 5*** is from parental node(s) to child node(s), which are the elements of a graph. In biological evolution, an evolutionary tree or phylogenetic tree is usually used to represent the entire evolutionary process of the genetic relationship between individuals of different species or different ethnic groups of the same species [32], [33]. Practically, this evolution tree is generalized into an evolution graph, which can be constructed using the following approach:

As T_1, T_2, T_3, and T_4 are chosen as 2020Q1, 2020Q2, 2020Q3, and 2020Q4, respectively, from *Example 1*, the corresponding evolution graph can be constructed as ***Figure 6***. Regarding the time study of species evolution, vertical development means that members of the same species have vertical development at different ages, and lateral development is the process of co-evolution of members of different species simultaneously [34]. In terms of retail consumption behavior, the vertical development of customers of a certain consumer species, similar to the evolution path 000000→001000→101010 of customer-2 in ***Figure 6***, and the co-evolution of customers of different consumer species simultaneously will be seen in a larger database in later discussion.

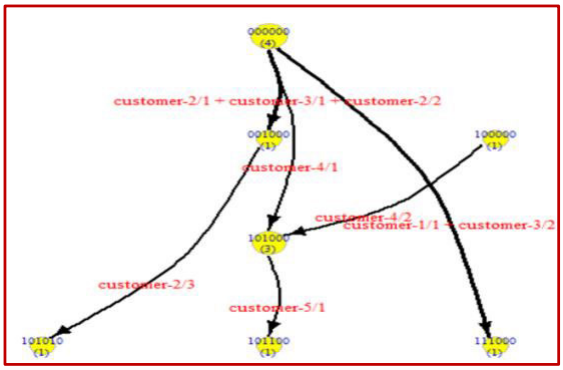

FIGURE 6. Evolution tree of Example 1 with four quarters as generations.

V. PRACTICAL EXPERIMENT OF REAL RETAIL TRANSACTIONS

For retail transaction data, this study analyzes the customer's journey, customer value, and customer's consumption genes to form an evolution tree of customer repurchase behavior and proposes the process of evolution analysis. This section will use actual retail transaction data to conduct a real case analysis and propose practical analysis and some application considerations.

A. THE PRACTICAL EXAMPLE

This experiment was based on actual retail data for big data analysis. The data come from a 40-year retail brand of a real company in the Far East. This brand has nearly 300 e-commerce and physical stores. The company is mainly engaged in the apparel industry, and the main product categories are tops, jackets, shoes, bags, accessories, etc.. The brand has had more than 40,000 active customers over the past three years.

The data of this experiment come from N=292197 transactions in 2020 and $D_2 = \{d_i, i=1, \ldots, N\}$. These data only take six major categories: tops, skirts, pants, dresses, coats, and shoes, namely P = {category-1, category-2, …, category-6}. These transactions are real buying behaviors generated by N_c = 27019 active customers $C = \{c_i, i = 1, \ldots, N_c\}$.

TABLE 2. Customer value model of retail data D2.

F \ M	(-1e+05,0]	(0,999]	(999,1e+04]	(1e+04,1e+05]	(1e+05,1e+06]	Sum
(0,1]	65	336	7417	1412	0	9230
(1,9]	272	54	6309	8332	39	15006
(9,49]	13	2	61	2056	534	2666
(49,999]	0	0	1	15	101	117
Sum	350	392	13788	11815	674	27019

Algorithm 1 Algorithm for CPEG (Customer Purchase Evolution Graph)

Given a transaction database $D = \{d_i = [t_i, s_i, c_i, p_i, m_i], c_i \in C, p_i \in P, t_i \in T\}$, the Customer Purchase Evolution Graph CPEG(γ, φ) can be constructed through the following steps:

(a) Split the time interval T into a sequence of consumption generations $T_1, \ldots, T_M$ with nesting $T_1 \subseteq \ldots \subseteq T_k \subseteq T_{k+1} \subseteq \ldots \subseteq T_M$. (*Definition 11*).

(b) Find all customer genes γ_{ci} of all customer $c_i \in C$ for all consumption generations T_m by their consumption behaviors $c_i(s_k, p_j) \in C_J[c_i]$. (*Definition 9*).

(c) The customer set is classified into a collection of customer species $C_S(D, \gamma_k) = \{c_i: C_G[c_i] = \gamma_k, \forall c_i \in C\}$. (*Definition 10*).

(d) Find any two consecutive generations T_1, T_2, and $T_1 \subseteq T_2$, find all the evolutions $\varphi(D, T_1, T_2, \gamma_k, \gamma_k')$, concisely symbolized as $\varphi(\gamma_k \rightarrow \gamma_k')$, and the evolution probabilities $P_r(\varphi(\gamma_k \rightarrow \gamma_k'))$. (*Definition 12*).

Thus, the CPEG (consumer purchase evolution graph) can be obtained as a weighted directed graph CPEG($\gamma = \{\gamma_k\}$, $\varphi = \{\varphi(\gamma_k \rightarrow \gamma_k')\}$, weight $= \{P_r(\varphi(\gamma_k \rightarrow \gamma_k'))\}$).

As mentioned in Section III, the standard transaction data database uses each item of each transaction as a record d_i as ***Table 1***.

B. CUSTOMER SEGMENTATION BY RFM MODELS

Customer segmentation is the first step in developing retail transaction data. Among the various methods for segmenting customers, the most popular is the RFM data-analysis technique proposed by Hughes in 1994 [35]. The RFM model includes three terms: (a) recency of customer consumption/purchase (R), (b) purchase frequency over a period of time (F), and (c) purchase amount (monetary) during this period (M) [36].

Customer value is the perception attitude of customers towards giving and getting. This subjective attitude affects consumers' overall evaluations of products [37]. In this experiment, through the segmentation of frequency F and monetary M, a more concrete "customer value table" was used, as shown in ***Table 2***. The rows in the table represent the frequency of customer visits (F) and the column represents the consumption money (M). ***Table 2*** contains 27,019 customers, of whom the first row (10,294 customers) is a one-time customer. In ***Table 2***, (a, b] = {x: a<x≤b }, which represents all the values from a to b (not including a), and the number 1e+04 (1.0·e+04) represents a scientific notation. This table also shows that in 2020, there will be 55 frequent customers who have visited more than 50 times and 674 VIP customers who have spent more than NT$100,000.

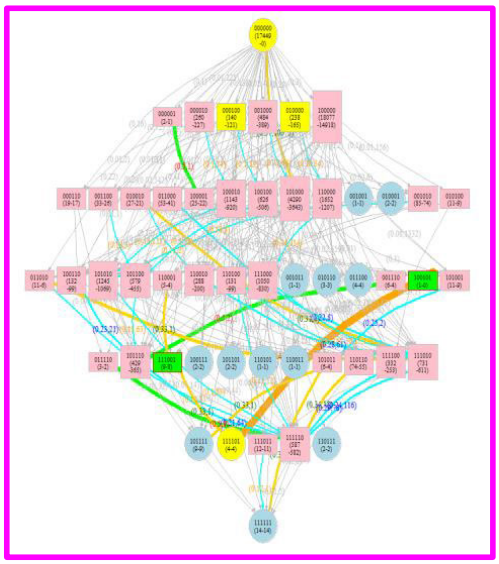

FIGURE 7. The CPEG graph of all the 27019 customers.

C. THE CPEG GRAPH OF ALL THE CUSTOMERS

For each edge (species evolution) in ***Figure 7***, there are two numbers in parentheses, as (pC, nC), where (a) nC is the number of customers participating in the evolution of this group and (b) pC is the evolution probability of this species evolution. To make the evolution more evident, the evolution edges with the highest 30 weights are listed in ***Table 3***. Using ***Algorithm 1***, the CPEG can be derived. First, the four quarters $[T_k] = [Q_1, Q_2, Q_3, Q_4]$ are used to generate the time period in step (a) of ***Algorithm 1*** and then step (b) generates the consumption genes of all customers with a total of 51 consumer gene species, such as 000000 (nC=17449), 001110 (nC=6), and 100000 (nC=18077). Moreover, 248 species evolution possibilities are formed among these 51 species, for example, 000000→100000 (pC=0.37, nC=10675), 111100→111110 (pC=0.29, nC=78), 000000→001000 (nC=451), 000000→111110 (nC=46) and 001110→101110 (pC=0.5, nC=2) 100000→101001 (nC=4). Thus, step (d) brings out a CPEG (Customer Purchase Evolution Graph), as shown in ***Figure 7***.

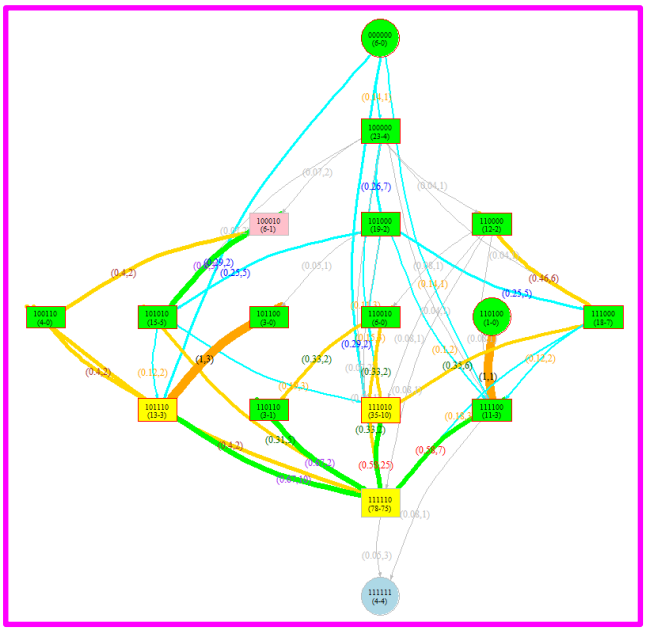

FIGURE 8. The CPEG graph of the 117 SVIP customers.

Some insights can be observed from ***Figure 7*** and ***Table 3***. (a) The most (18077 customers) consumption gene is "100000." Even in Q4, there are still 14918 customers who only have this gene; that is, they only buy tops. (b) The consumption genes with the second largest number of customers were 4290 purchasers of clothes and pants ("101000") and 1652 purchasers of clothes and skirts ("110000"). (c) From the repurchase analysis of first-purchase customers, the most consumption evolution is 332 customers with "100000" → "100010" (those who buy tops and then buy coats) and then, the other 332 customers with "100000" → "110000" (buy tops and then buy skirts). From the repurchase analysis of mature customers, the most consumption evolution is 78 customers with "111100" → "111110" (those who buy tops/skirts/pants/dresses will then buy coats with 0.29 probability) and 61 customers with "110010" → "111010" (buy tops/skirts/coats and then buy pants in 0.28 probability). (d) For new customers (the edges whose starting point is "000000"), it can be seen that the customer is most likely to buy pants ("100000," 10675 people), and what is more interesting is that a considerable proportion of clothes and pants are purchased together ("110000," 559 customers) and tops and pants ("101000," 2571 customers). (e) In addition, ***Table 3*** shows the evolution of all purchases with probability > 0.18, such as 100101→111101 with probability $= 1$, which means that those who bought X will definitely buy Y, but there is only one customer. These high-probability repurchase evolutions are represented in ***Figure 7*** as thicker lines with more prominent colors. These observations are helpful for product development, channel operations, and marketing strategies. Next, we will further discuss the evolution process of customer purchases for SVIP (Super Very Important Persons) customers.

TABLE 3. Evolution edges with the highest 10 weights.

Source genes	Destination genes	Edge weights (nC)	Evolution probability
100101	111101	1	1
000001	100001	1	0.5
001110	101110	2	0.5
011110	111110	1	0.5
011010	111010	4	0.44
000000	100000	10675	0.37
110110	111110	18	0.34
101011	101111	1	0.33
101011	111111	1	0.33
110001	111001	1	0.33
111001	111101	1	0.33
010010	111010	4	0.31
111100	111110	78	0.29
110010	111010	61	0.28
101001	101011	2	0.25
111010	111110	116	0.24
100110	101110	21	0.23
011000	111000	8	0.22
101110	111110	64	0.21
001000	101000	64	0.18

D. THE CPEG GRAPH OF SVIP CUSTOMERS

As shown in ***Table 2***, 117 SVIPs visited more than 50 times and mostly consumed more than 10,000 in a year. In 2020, purchases of these SVIPs generate 20409 transactions, about 6.98% of overall 292197 transactions. Using ***Algorithm 1***, 17 customer purchase genes and 44 purchase evolutions can be found, which form a CPEG graph, as ***Figure 8***.

Similarly, several insights into the evolution of customer purchases can be observed as follows: (a) The most (78 customers) mature SVIP consumption gene is "111110." Even in Q4, 75 customers still have this gene; that is, they buy all categories except shoes. Among them, only three customers buy all categories, and only four customers buy all categories. (b) The customer with the most evolution starting point is "100000" (23 people), which is the entry category of the SVIP. In addition, the starting point of SVIP evolution is tops and pants ("101000," 19 people), and its derivative gene tops/pants and skirts ("111000," 18 people) and tops/pants and coats ("101010," 15 people). The evolution of these three consumer groups is diverse, and the corresponding evolution probabilities are mostly less than 0.3. (c) From the perspective of the mature evolution of consumption genes, the mature consumption genes of "111110" can be evolved from "111010" (25 people), "101110" (10 people), and "111100" (7 people), and the evolution probabilities are more than half (0.58-0.67). In other words, as long as SVIPs purchase the four categories, they can easily evolve into mature consumption genes. (d) What's more, there are two consumption species that will definitely evolve further (the evolution probability is 1), namely 101100–>101110 and 110100–>111100, but the number of customers is not

much (3 and 1). (e) Because they are SVIPs, many customers start to purchase tops/pants (“101000,” 19 people), and this consumption species can evolve into tops/pants and coats (“101010,” 5 people) or skirts (“111000,” 5 people), so that the union gene (“111010,” 3 people), and even the mature gene (“111110,” five people).

VI. CONCLUSION

Using actual retail transactions, this study utilizes evolution theory to investigate customer journeys. To pave the research foundation, Section II formulates the analysis elements, customer (C) and product (P), to weave a coordinate CP plane, and the selling procedure (S) is a series of activities over the CP plane. Accordingly, a (S, P, C) data space is constructed in Section III to locate retail transaction data, and the customer journey is then formulated as a pivot transformation in the (S,P,C) space. With the definition of consumption genes of customer purchases, Section IV introduces their evolution operations and then builds up a customer purchase evolution graph (CPEG) by ***Algorithm 1***. Finally, Section V applies the CPEG to a real transaction set of an actual retail company for the customer journeys of overall 27019 customers (27019 customers) and 117 SVIPs to prove its practicability by exploring several major customer purchase species and their significant purchase evolution (repurchase) patterns with evolutionary operations, which can be used as an important basis for planning retail merchandise operation strategies.

This study examines the structure of a customer journey using evolutionary theory, which is just a starting point, and the actual case adopted is only a preliminary experiment. In the future, CPEG can be applied to more practical cases to explore the deeper evolutionary nature of the customer purchase journey. However, CPEG will conduct a more in-depth analysis of evolutionary characteristics in the field of customer purchase and consumption, including the schemata of evolutionary species, calculation of evolutionary operators, and analysis of various evolutionary patterns. Finally, it is possible to conduct a broader discussion on the analysis of customer purchases in the retail data space, in combination with other analysis theories. These are the goals that we will work on in the future.

REFERENCES

[1] W. B. Dodds, K. B. Monroe, and D. Grewal, “Effects of price, brand, and store information on buyers’ product evaluations,” *J. Marketing Res.*, vol. 28, no. 3, pp. 307–319, 1991, doi: 10.2307/3172866.

[2] M. Tsiros and V. Mittal, “Regret: A model of its antecedents and consequences in consumer decision making,” *J. Consum. Res.*, vol. 26, no. 4, pp. 401–417, Mar. 2000, doi: 10.1086/209571.

[3] L. Gronholdt, A. Martensen, and K. Kristensen, “The relationship between customer satisfaction and loyalty: Cross-industry differences,” *Total Quality Manage.*, vol. 11, nos. 4–6, pp. 509–514, Jul. 2000, doi: 10.1080/09544120050007823.

[4] S. R. Whittle and M. Foster, “Customer profiling: Getting into your customer’s shoes,” *Manag. Decis.*, vol. 27, no. 6, pp. 27–31, 1989, doi: 10.1108/00251748910132575.

[5] A. Følstad and K. Kvale, “Customer journeys: A systematic literature review,” *J. Service Theory Pract.*, vol. 28, no. 2, pp. 196–227, Mar. 2018, doi: 10.1108/JSTP-11-2014-0261.

[6] A. Meroni and D. Sangiorgi, *Design for Services. Gower, Surrey*. London, U.K.: Routledge, 2011, doi: 10.4324/9781315576657.

[7] L. G. Zomerdijk and C. A. Voss, “Service design for experience-centric services,” *J. Service Res.*, vol. 13, no. 1, pp. 67–82, Feb. 2010, doi: 10.1177/1094670509351960.

[8] L. G. Zomerdijk and C. A. Voss, “NSD processes and practices in experiential services*,” *J. Product Innov. Manage.*, vol. 28, no. 1, pp. 63–80, Jan. 2011, doi: 10.1111/j.1540-5885.2010.00781.x.

[9] L. Patrício, R. P. Fisk, J. Falcão e Cunha, and L. Constantine, “Multilevel service design: From customer value constellation to service experience blueprinting,” *J. Service Res.*, vol. 14, no. 2, pp. 180–200, May 2011, doi: 10.1177/1094670511401901.

[10] C. Diana, E. Pacenti, and R. Tassi, “Visualtiles: Communication tools for (service) design,” in *Proc. ServDes*, Linköping, Sweden, 2009, pp. 65–76.

[11] D. E. McCoy, “Evolutionary change,” in *Encyclopedia of Evolutionary Psychological Science*, T. Shackelford and V. Weekes-Shackelford, Eds. Cham, Switzerland: Springer, 2018, pp. 1–16, doi: 10.1007/978-3-319-16999-6_2094-1.

[12] A. Judy Stamps and M. Alison Bell, “Combining information from parental and personal experiences: Simple processes generate diverse outcomes,” *PLoS ONE*, vol. 16, no. 7, 2021, Art. no. e0250540, doi: 10.1371/journal.pone.0250540.

[13] H. Pearson, “What is a gene?” *Nature*, vol. 441, no. 7092, pp. 398–401, 25 May 2006, doi: 10.1038/441398a.

[14] B. Shook, K. Nelson, and K. Aguilera, *Explorations: An Open Invitation To Biological Anthropology*. Hopewell, VA, USA: American Anthropological Association, 2019.

[15] A. McAvoy and B. Allen, “Fixation probabilities in evolutionary dynamics under weak selection,” *J. Math. Biol.*, vol. 82, no. 3, p. 14, Feb. 2021, doi: 10.1007/s00285-021-01568-4.

[16] R. Dawkins, *The Selfish Gene*, Oxford, U.K.: Oxford Univ. Press, 1976.

[17] R. L. Daft, *Organization Theory and Design*, 8th ed. Columbus, OH, USA: South-Western, 2004.

[18] W. J. Bock, “Principles of biological comparison,” *Acta Morphol. Neerlando-Scandinavica*, vol. 27, nos. 1–2, pp. 17–32, 1989.

[19] W. J. Bock, “Explanations in evolutionary theory,” *J. Zool. Systematics Evol. Res.*, vol. 45, no. 2, pp. 89–103, May 2007.

[20] Q. Deng and R. Paul Messinger, “Dimensions of brand-extension fit,” *Int. J. Res. Marketing*, vol. 39, no. 3, pp. 764–787, 2022, doi: 10.1016/j.ijresmar.2021.09.013.

[21] M. Cleveland, N. Papadopoulos, and M. Laroche, “Identity, demographics, and consumer behaviors: International market segmentation across product categories,” *Int. Marketing Rev.*, vol. 28, no. 3, pp. 244–266, May 2011, doi: 10.1108/02651331111132848.

[22] Y. Lin, C. Liang, and J. Heh, “From e-learning to e-shopping,” in *Taiwan E-Learning Forum (TWELF)*. Taichung: National Taichung Univ., 2013.

[23] Y. Lin, C. Liang, K. Li, C. Cheng, H. Liang, and J. S. Heh, “Brand angel: An on-job-training eBook for fashion brands,” in *Proc. Nat. Comput. Symp. (NCS)*, 2013, pp. WS17–WS25.

[24] J. Alan Dubinsky, “A factor analytic study of the personal selling process,” *J. Pers. Selling Sales Manag.*, vol. 1, no. 1, pp. 26–33, 2013, doi: 10.1080/08853134.1981.10754192.

[25] E. R. Hite and A. J. Bellizzi, “Differences in the importance of selling techniques between consumer and industrial salespeople,” *J. Pers. Selling Sales Manag.*, vol. 5, no. 2, pp. 19–30, 2013, doi: 10.1080/08853134.1985.10754398.

[26] W. C. Moncrief and G. W. Marshall, “The evolution of the seven steps of selling,” *Ind. Marketing Manage.*, vol. 34, no. 1, pp. 13–22, Jan. 2005, doi: 10.1016/j.indmarman.2004.06.001.

[27] M. L. Rodrigues, T. S. Körting, G. R. de Queiroz, C. P. Sales, and L. A. R. da Silva, “Detecting center pivots in Matopiba using Hough transform and web time series service,” in *Proc. IEEE Latin Amer. GRSS ISPRS Remote Sens. Conf. (LAGIRS)*, Santiago, Chile, Mar. 2020, pp. 189–194, doi: 10.1109/LAGIRS48042.2020.9165648.

[28] M. Stickdorn and J. E. Schneider, *This is Service Design Thinking: Basics, Tools, Cases*. Amsterdam, The Netherlands: BIS Publishers, 2010.

[29] R. Sturm, “Eye colour: Portals into pigmentation genes and ancestry,” *Trends Genet.*, vol. 20, no. 8, pp. 327–332, Aug. 2004, doi: 10.1016/j.tig.2004.06.010.

[30] P. M. Visscher, W. G. Hill, and N. R. Wray, “Heritability in the genomics era—Concepts and misconceptions,” *Nature Rev. Genet.*, vol. 9, no. 4, pp. 255–266, Apr. 2008, doi: 10.1038/nrg2322.

[31] S. Blackmore, *The Meme Machine*. Oxford, U.K.: Oxford Univ. Press, 2000, pp. 75–76.
[32] J. Felsenstein, *Inferring Phylogenies*. Sunderland, MA, USA: Sinauer Associates, 2004.
[33] I. Letunic and P. Bork, "20 years of the SMART protein domain annotation resource," *Nucleic Acids Res.*, vol. 46, no. D1, pp. D493–D496, Jan. 2018, doi: 10.1093/nar/gkx922.
[34] K.-C. Hung, "An evolution framework cross-border mergers and acquisitions by living systems approach," Ph.D. dissertation, Ph.D. Program Bus., Feng-Chia Univ., Taiwan, 2014.
[35] A. H. L. Chen, Y.-C. Liang, W.-J. Chang, H.-Y. Siauw, and V. Minanda, "RFM model and K-means clustering analysis of transit traveller profiles: A case study," *J. Adv. Transp.*, vol. 2022, Aug. 2022, Art. no. 1108105, doi: 10.1155/2022/1108105.
[36] A. Handojo, N. Pujawan, B. Santosa, and M. L. Singgih, "A multi layer recency frequency monetary method for customer priority segmentation in online transaction," *Cogent Eng.*, vol. 10, no. 1, Dec. 2023, doi: 10.1080/23311916.2022.2162679.
[37] W. Zang, Y. Qian, and H. Song, "The effect of perceived value on consumers' repurchase intention of commercial ice stadium: The mediating role of community interactions," *Int. J. Environ. Res. Public Health*, vol. 19, no. 5, p. 3043, 2022, doi: 10.3390/ijerph19053043.

YUNG-TZU J. LIN received the M.S.B.A. degree from the Department of Business Management, National Taiwan University of Science and Technology, Taiwan, in 2014, where she is currently pursuing the Ph.D. degree in business administration. In Taiwan, she was the New Business Development Manager of Levi's, from 2005 to 2007, the License Manager of Disney, from 2007 to 2009, and the Assistant General Manager of Tommy Hilfiger, from 2010 to 2012. Her research interests include brand management, retail marketing segmentation, and its analytics.

CHUAN-YI CHANG received the B.S.C.E. degree from the Department of Information and Computer Engineering, Chung Yuan Christian University, in 1994, and the M.S.C.E. degree from the Department of Computer Science and Engineering, Yuan Ze University, in 2001. He is currently pursuing the Ph.D. degree in computer science with Chung Yuan Christian University. He is the Local GM of Global ODM Company, Nanchang, China. His research interests include data mining, e-commerce data technology, and e-learning technology.

SHEIN-YUNG CHENG received the B.S.C.E., M.S.C.E., and Ph.D. degrees from the Department of Information and Computer Engineering, Chung Yuan Christian University, in 1987, 1991, and 2005, respectively. He is currently with the Computer Center, Chung Yuan Christian University, where he is an Assistant Professor with the Department of Information and Computer Engineering. His research interests include database technology, data mining, and e-learning technology.

MENG-YUN T. LIN received the Ph.D. degree from the Warwick Business School, The University of Warwick, U.K., in 1996. He is currently a Professor with the Department of Business Management, National Taiwan University of Science and Technology. His courses include marketing management, marketing theory, management, and research methodology. His research interests include word-of-mouth marketing and market segmentation.