RESEARCH ARTICLE

Probabilistic customer purchase evolution graph

Liu Xiong

X Liu. Probabilistic customer purchase evolution graph. J Pure Appl Math. 2023; 7(4):293-303.

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.

Key words: *Retail transaction; Repurchase; Evolution theory; Customer journey*

INTRODUCTION

Retailing is selling certain products or services to customers. In retail, customers buying a certain product is a behavioral tendency of their purchase intention [1].

Purchase and Repurchase

Repurchase intention or repeat patronage indicates the possibility that customers will repurchase products or avail services of the same brand again [2]. Therefore, investigating the customer's repurchase behavior and constructing a model of multiple transactions of customers is critical. The repurchase intentions of customers can be measured from three aspects, namely (a) Intents to repurchase: This behavior measures customers' willingness to repurchase the company's products or services in the future, which is an indicator of customers' future behavioral intentions. (b) Primary behaviors: This behavior includes the number of purchases, frequency, amount, and quantity of customers. (c) Secondary behaviors: This behavior refers to behaviors that customers help the company introduce, recommend, and build reputation [3]. Among the three behaviors, only primary behaviors can be measured by actual transaction data. For each customer, his/her primary behaviors change over time. Thus, each customer has a distinct customer journey. This is the basic starting point of this research.

Customer journey

A customer's process of purchasing goods and services is a well-defined service process, called the 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 a brand (service provider) [6]. The aforementioned interactions or communications between customers and brands can be defined as "touchpoints" [7-9]. A touchpoint represents an abstract form of the customer experience (the customer's product purchase in this paper) [10]. The customer journey can thus be visualized as a well-defined service process with a clear beginning and end [4]. In the customer journey of purchase analysis, the initial touchpoint can be regarded as the first purchase

Department of Business Management, National Taiwan University of Science and Technology, Taipei City 106335, Taiwan;

Correspondence: Liu Xiong, Department of Information and Computer Engineering, Chung Yuan Christian University, Taoyuan City 320314, Taiwan (R.O.C.), e-mail chang.maverick@gmail.com

Received: 8 August, 2023, Manuscript No. puljpam-23-6643, Editor Assigned: 9 August, 2023, PreQC No puljpam-23-6643(PQ), Reviewed: 19 August, 2023, QC No. puljpam-23-6643(Q), Revised: 20 August, 2023, Manuscript No puljpam-23-6643(R), Published: 31 August, 2023, DOI:-10.37532/2752-8081.23.7(4).293-303

behavior of a new customer, and the final touchpoint can be regarded as the purchasing characteristics of a loyal customer.

Evolution theory

During the customer journey, the goods or services purchased by customers will change over time and be affected by the external environment (including shopping mall environment and product mix, etc.), so 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 heritable traits. In biology, hereditary traits are controlled by the organism's genome (genetic material), known as the genotype. On the other hand, the complete observable behavior or traits that result from the influence of these genes are called phenotypes [13]. A species is a group of individuals with 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 also study the evolution of an enterprise, which will also undergo internal evolution due to the influence of the external environment [17- 19], In the application of retail, 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.

Research objectives

This paper applies evolution theory to the investigation of customer journey in retail purchase transactions and then propose a Customer Purchase Evolution Graph (CPEG). For the study of retailing, the retailing elements are formulated in Section II and a SPC data space of retailing is constructed for customer journey in Section III. Accordingly, 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.

FORMULATION OF RETAILING ELEMENTS

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:

<u>Definition 1</u>

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

For example, the product feature set of clothes P_a = {coat, pants, ...}, the product attribute of clothes gender P_b = {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, nationality; (c) Classified by social class, lifestyle, personality and other psychological variables; (d) Classified by behavioral variables, such as use timing, benefits, user status, purchase preparation [21].

<u>Definition 2</u>

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 a feature of some certain customers.

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

CP Plane

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

<u>Definition 3</u>

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

With the partial orderings of customer set and product set, C and P form two independent coordinate axes: consumer (C) axis from customer to community and product (P) axis from product to brand. Our previous research proposed to use 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 on the CP plane represents a coordinate sequence pair (c_i, p_j) with different degrees of detail. There are $n(P) \times n(C)$ cells in totals.

Sales Behaviors

In the retail process, sales staff interact with customers to persuade customers to purchase. Research divided such sales process into seven stages: (a) Prospecting: Identify potential buyers of products or services. (b) Pre-approach: Collect relevant information of potential buyers to prepare sales visits. (c) Approach: Start selling to specific buyers. (d) Sales Presentation: Provide the characteristics and advantages of products or services to awaken the purchasing desires. (e) Handling Objections and Overcoming Resistance: Strive to overcome the refusal and rejection of consumers. (f) Closing: Make customers complete purchasing by some suitable and effective methods. (g) Post-Sale Follow-Up: Continue to emphasize after-sale customer satisfaction [24-26]. Following the above derivation, these customer consumption behaviors in the sales process can also have a formal definition.

Definition 4

Given a product set P = { p_j } and a customer set C = { c_i }, the customer consumption behaviors S = [$s_k(c_i, p_j)$] in the sales process can be expressed as a series of purchase actions s_k's, where $s_k(c_i, p_j)$ represents the customer c_i performs some actions s_k on the product p_j.

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

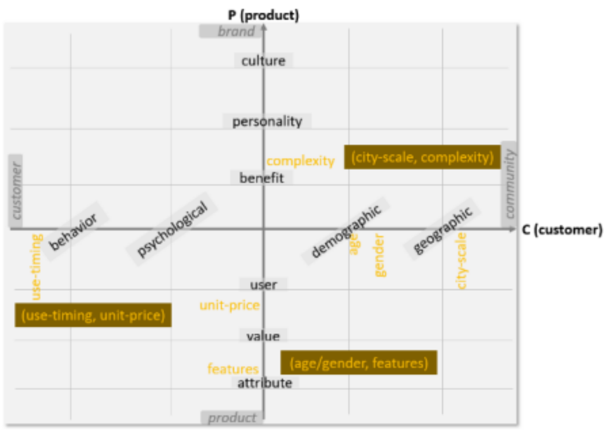

Figure 1) *The weaved CP plane*

RETAIL SPC DATA SPACE AD ITS TRANSFORMATION

SPC Data space

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 is proposed as the research foundation of 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 the possible customer consumption behaviors $s_k(.,.)$'s.

As Figure 2 shown, the coordinatized CP plane can be divided into four quadrants: Brand, Service, Sales, and Marketing represent quadrants from I, II, III, to IV respectively in Figure 2. In the 1st Brand quadrant, corporate prospects potential products (abstract brand) and potential (general) customer community. (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 some 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 CP plane may vary by applications.

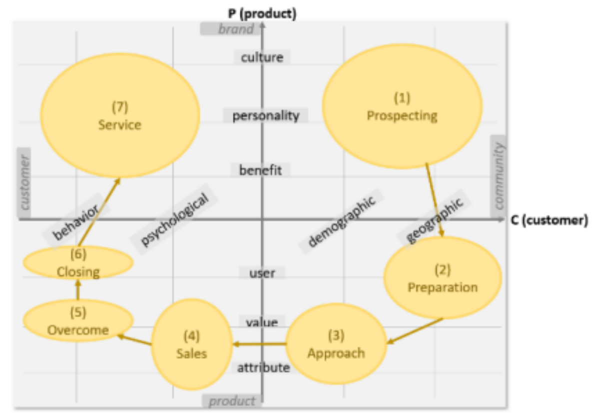

Figure 2) *Seven stages of selling over CP plane*

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 some prices in certain channels (S, e.g., stores). During the period (T = { t_i }), transactions record these actual consumer behaviors within some database in the form of SPC space.

Definition 6

Given a SPC data space (S, P, C) with customer set C = {c_i}, a product set P = p_j} with some product prices (M = { m_i }), and the 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 transaction in (S, P, C) data space deeply, one small real transaction database is taken as an example.

Example 1

Table 1 is a transaction D_1 of 20 transactions in a real retail business of 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 fields of channel, product and customer, respectively. Note that the field category represents still more general features (tops, skirts, pants, dresses, coats, shoes, here) of products along the P-axis. Besides, the fields datetime and invoice no represent the time tag t_i's in the time period T_1. The fields quantity and amount represent the consumption measures of the transactions (s_i, p_i, c_i)-point of 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)-space, as Figure 3 shown.

TABLE 1
Retail transaction database D1 in Example 1

Datetime	Invoice No	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

Since a transaction is the result of one purchase, no matter the purchase object is a product or a 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 his/her basic data such as gender or age has been collected, can the other three factors be processed. The following will start the investigation of some in-depth customer behaviors from the viewpoint of customer journey.

Pivot transformation and customer journey

In a data space, it is possible to carry out data flow transformation which is a mapping to moves data from a data source to a data destination [27]. For SPC data space, one kind 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 a SPC data space, a kind 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 $P_A(D)$ = [$a_i(D)$], to turn (aggregate) values from one dimension into a (new) data set. Under this definition, SPC data space may possess three kinds 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 look at retail from the perspective of sales/operation and brand/ products. The focus of this paper will be on the customer's viewpoint, $P_c(D)$ = [$c_i(D)$].

The Customer-Oriented Pivot Transformation (abbreviated as COPT) $P_c(D)$ projects all the transactions into the customer- axis, which can then be used to build all customer models. As Figure 4 shown, $P_c(D_1)$ = [$c_i(D_1)$] aggregates all the transaction data in D_1 related to customer-i in $P_c(D_1)$.

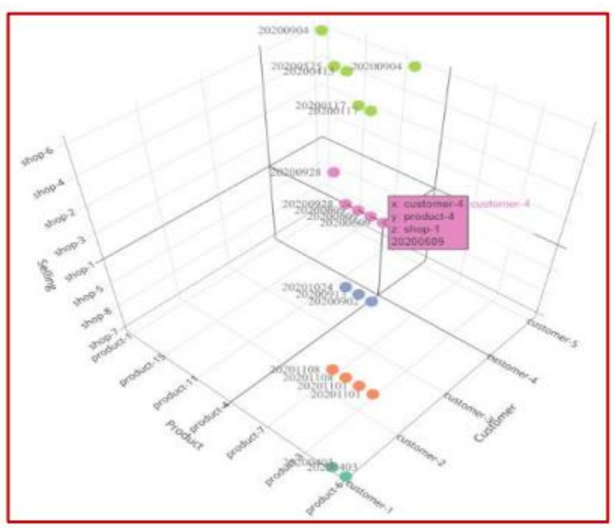

Figure 3) *PM knowledge space of Example 1*

Proposition 1

(a) $P_c(D)$ is not unique.

(b) Each transaction (point) in $P_c(D)$ will own the form $c_i(s_k, p_j)$.

Proof

(a) Obvious since 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 is easily rewritten as $c_i(s_i, p_i)$.

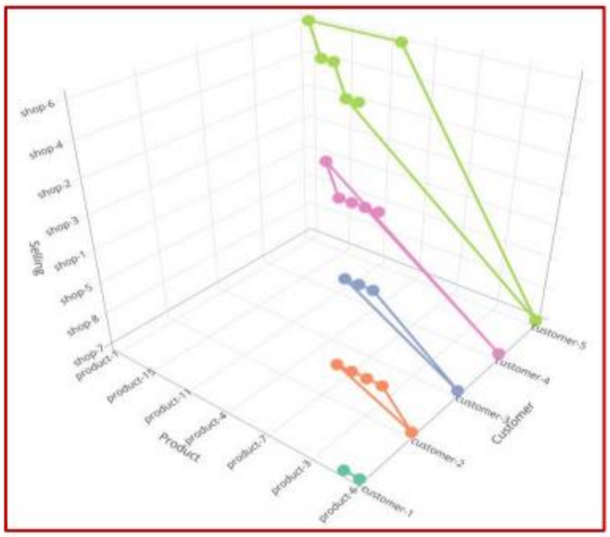

Figure 4) *Customer-Oriented Pivot Transformation of D1 to Customer Axis.*

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

Definition 8

For a retail transaction database $D = \{d_i = [t_i, s_i, c_i, p_i, m_i]\}$ in a 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 means customer journey can be obtained through pivot transformation.

Practically, time tags will be 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].

CUSTOMER PURCHASE EVOLUTION GRAPH

Customer consumption gene

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

The genotype (gene set within an organism's genome), biologically, controls inherited traits. And, the interaction of genotypes with the environment forms the observable traits of organisms, which is also called the phenotype [13, 30]. Similarly, from the perspectives of human culture, a lot of behaviors can be explained by evolutionary meme whereas a set of culture genetic factors evolve through the process of duplication (imitation), mutation, and selection [31]. Under such concept of social evolution, 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 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 a SPC data space, the consumption gene of the customer c_i is defined as one kind of COPT $P_{cg}(.)$:$C_G[c_i] = P_{cg}(D)[c_i] = \{ p_j: c_i(s_k,p_j) \in C_J[c_i] \} \equiv \gamma_{ci}$, $\forall c_i \in C$. Although here the products p_j's is used to define the customer gene $C_G[c_i]$, which is more often defined on category g_j's 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] is obtained. It is notable that $C_G[c_1] = 101000 = C_G[c_5]$, which brings out the definition of 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 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 with each other. For D_1 in Example 1, this definition divided the 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 the customers in the same species have the same consumption behaviour.

Evolutionary operations of customer consumption gene

Just like biological evolution, the evolutionary operations such as the duplication, mutation, and recombination of customer consumption genes come from the mutual influence of customer consumption habits and changes in customer product consumption habits. Since evolution happens from generation to generation. This kind of generation should be defined first.

Definition 11

For a retail transaction database $D = \{d_i = [t_i, s_i, c_i, p_i, m_i]\}$, the consumption generation is defined in a specific time period, i.e. $D[T] = \{d_i = [t_i, s_i, c_i, p_i, m_i], \forall t_i \in T\}$.

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, it is possible to define 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): $\gamma_k \rightarrow \gamma_k'$, $\gamma_k' = (\gamma_k' \backslash \gamma_k) \cup (\gamma_k' \cap \gamma_k)$. Moreover, the evolution probability of evolutionary operation φ is defined as $P_r(\varphi(D, T_1 \rightarrow T_2, \gamma_k \rightarrow \gamma_k')) = nC(\gamma_k') / nC(\gamma_k)$.

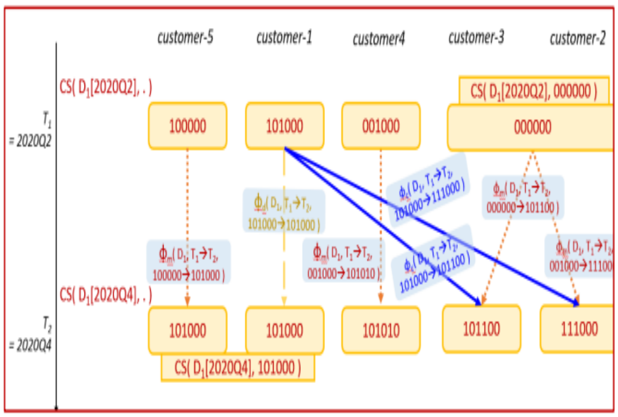

Figure 5) *Evolution of customer-consumption genes.*

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

Evolution of customer consumption genes

The evolution direction of Figure 5 is evolved from parental node(s) to child node(s), which are the element of a graph. In biological evolution, an evolutionary tree, or phylogenetic tree, is usually used to represent the whole evolution 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 by 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

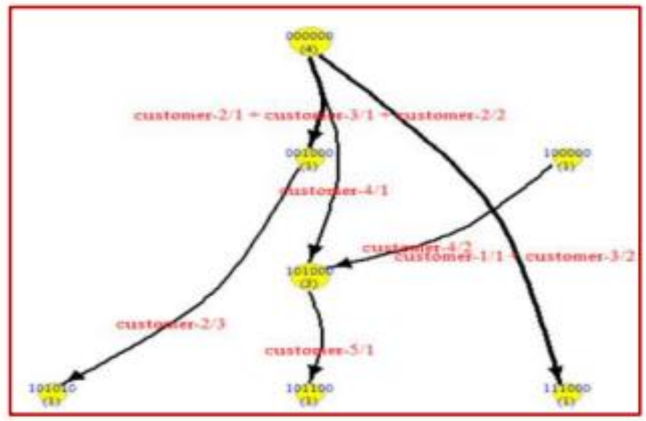

Figure 6) *Evolution tree of Example 1 with four quarters as generations*

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.

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, . . . , 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, . . . , 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 construction of Customer Purchase Evolution Graph CPEG(γ, φ) can be built up 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 the customer genes γ_{ci} of all the customer $c_i \in C$ for all consumption generations T_m's by their consumption behaviors $c_i(s_k, p_j) \in C_J[c_i]$. (Definition 9)

(c) Classify the customer set into a collection of the 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 $\varphi(\gamma_k \gamma_k')$ and the evolution probabilities $P_r(\varphi(\gamma_k \gamma_k'))$. (Definition 12)

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

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

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

PRACTICAL EXPERIMENT OF REAL RETAIL TRANSACTIONS

For retail transaction data, this paper 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 real case analysis, and put forward practical analysis and some application considerations.

The practical example

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

The data of this experiment comes from N=292197 transactions in 2020, set $D_2 = \{d_i, i=1, \ldots, N\}$. This data only takes six major categories such as 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 $C = \{c_i, i = 1, \ldots, N_c\}$. As mentioned in Section III, the standard transaction data database uses each item of each transaction as the record d_i as Table 1.

Customer segmentation by RFM models

In actual processing, customer segmentation is always 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 Arthur 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 get-ting. This subjective attitude affects consumers' overall evaluation of products [37]. In this experiment, through the segmentation of frequency F and monetary M, a more concrete "customer value table" is used in Table 2. The rows in the table represent the frequency of customer visits (F), and the column is consumption money (M). Table 2 contains 27,019 customers, of which 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 as 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.

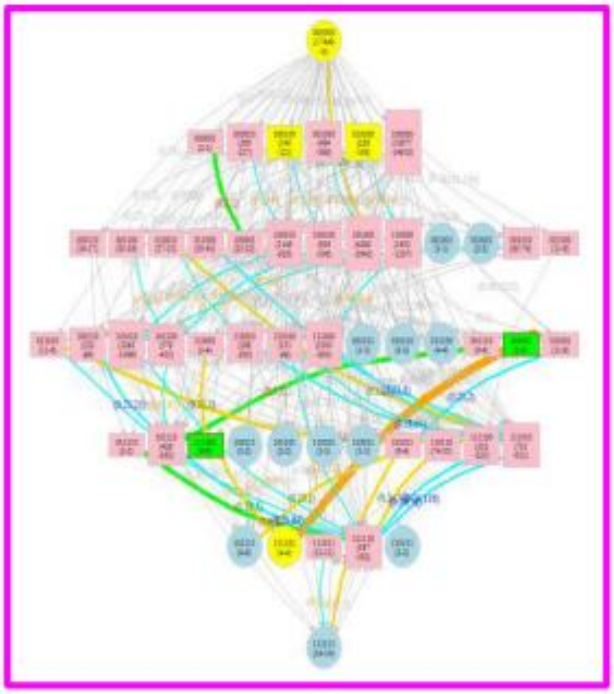

Figure 7) *The CPEG graph of all the 27019 customers*

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 obvious, the evolution edges with the highest 30 weights are listed in Table 3. By 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.

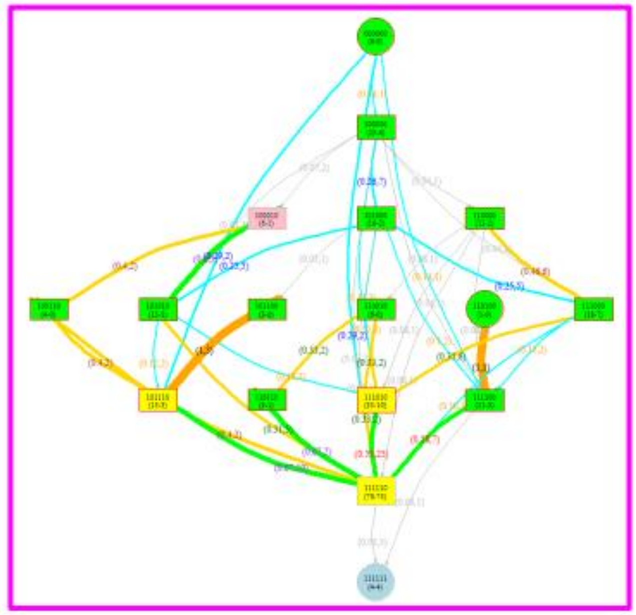

Figure 8) *The CPEG graph of the 117 SVIP customers*

From Figure 7 and Table 3, some insights can be observed as follows. (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 are 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 very helpful for product development, channel operation, 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
1	100001	1	0.5
1110	101110	2	0.5
11110	111110	1	0.5
11010	111010	4	0.44
0	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
10010	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
11000	111000	8	0.22
101110	111110	64	0.21
1000	101000	64	0.18

The CPEG graph of SVIP Customers

From Table 2, there are 117 SVIPs who visit more than 50 times and mostly consume more than 10,000 in a year. In 2020, the purchases of these SVIPs generate 20409 transactions, about 6.98% of overall 292197 transactions. By Algorithm 1, 17 customer purchase genes can be found and 44 purchase evolutions, which form a CPEG graph, as shown in Figure 8.

Likewise, several insights into the evolution of customer purchases can be observed as follows. (a) The most (78 customers) SVIP mature consumption gene is "111110". Even in Q4, there are still 75 customers who have this gene, that is, they buy all categories except shoes. Among them, there will be only 3 customers who buy all categories and there are only 4 customers who buy all categories. (b) The customer with the most evolution starting point is "100000" (23 people), which is the entry category of SVIP. In addition, the more SVIP evolution starting point 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 outgoing evolution of these three consumer groups is very 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). That is to say, as long as SVIPs purchase 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) Since 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",5 people).

CONCLUSION

With actual transaction of retailing, this paper utilizes evolution theory in the investigation of customer journey. To pave the research foundation, Section II formulates the analysis elements, customer (C) and product (P), to weave a coordinate CP plane, and then the selling procedure (S) is a series of activities over CP plane. Accordingly, a (S, P, C) data space is constructed in Section III for locating retail transaction data and customer journey is then formulated as a pivot transformation in (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 into a real transaction set of actual retailing 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 paper studies the structure of customer journey with evolutionary theory, which is just a starting point, and the actual case adopted is only a preliminary experiment. In the future, CPEG will be applied to more practical cases to explore the deeper evolutionary nature of the customer purchase journey. On the other hand, 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, etc. 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 we will work on in the future.

REFERENCES

1. Dodds WB, Monroe KB, Grewal D. Effects of price, brand, and store information on buyers' product evaluations. J Marketing Res. 1991;28(3):307-319.
2. Tsiros M, Mittal V. Regret: A model of its antecedents and consequences in consumer decision making. J Consum Res. 2000;26(4):401-417.
3. Gronholdt L, Martensen A, Kristensen K. The relationship between customer satisfaction and loyalty: Cross-industry differences. Total Qual Manag. 2000;11(4-6):509-514.
4. Whittle SR, Foster M. Customer profiling: Getting into your customer's shoes. Manag Decis. 1991;27(6).
5. Folstad A, Kvale K. Customer journeys: A systematic literature review. J Serv Theory Pract. 2018;28(2):196-227.
6. Meroni A, Sangiorgi D. Design for Services. Surrey: Gower; 2011.
7. Zomerdijk LG, Voss CA. Service design for experience-centric services. J Serv Res. 2010;13(1):67-82.
8. Zomerdijk LG, Voss CA. NSD processes and practices in experiential services. J Prod Innov Manag. 2011;28(1):63-80.
9. Patrício L, Fisk RP, Cunha JF e, Constantine L. Multilevel service design: From customer value constellation to service experience blueprinting. J Serv Res. 2011;14(2):180-200.
10. Diana C, Pacenti E, Tassi R. Visualtiles: Communication tools for (service) design. In: Proceedings of ServDes 2009. Linköping, Sweden; 2009:65-76.
11. McCoy DE. Evolutionary Change. In: Shackelford T, Weekes-Shackelford V, eds. Encyclopedia of Evolutionary Psychological Science. Cham: Springer; 2018.
12. Stamps JA, Bell AM. Combining information from parental and personal experiences: Simple processes generate diverse outcomes. PLOS ONE. 2021;16(7):e0250540.
13. Pearson H. Genetics: What is a gene?. Nature. 2006;441(7092):398-401.
14. Shook B, Nelson K, Aguilera K. Explorations: An Open Invitation To Biological Anthropology. American Anthropological Association; 2019.
15. McAvoy A, Allen B. Fixation probabilities in evolutionary dynamics under weak selection. J Math Biol. 2021;82(14).
16. Dawkins R. The Selfish Gene. Oxford University Press; 1976.
17. Daft RL. Organization theory and design (8th edition). Ohio: South-Western; 2004.
18. Bock WJ. Principles of biological comparisons. Acta Morphol Neerl Scand. 1989;27(1-2):17-32.
19. Bock WJ. Explanations in evolutionary theory. J Zool Syst Evol Res. 2007;45(2):89-103.
20. Deng Q, Messinger PR. Dimensions of brand-extension fit. Int J Res Mark. 2022;39(3):764-787.
21. Cleveland M, Papadopoulos N, Laroche M. Identity, demographics, and consumer behaviors: International market segmentation across product categories. Int Mark Rev. 2011;28:244-266.
22. Lin C, Liang C, Heh J. From eLearning to eShopping. Taiwan E-Learning Forum (TWELF); 2013.
23. Lin C, Liang C, Li K, Cheng C, Liang H, Heh J. Brand angel: An on-job-training eBook for fashion brands. National Computer Symposium (NCS); 2013.
24. Dubinsky AJ. A Factor Analytic Study of the Personal Selling Process. J Pers Sell Sales Manag. 1981;1(1):26-33.
25. Hite RE, Bellizzi JA. Differences In the Importance of Selling Techniques Between Consumer and Industrial Salespeople. J Pers Sell Sales Manag. 1985;5(2):19-30.
26. Moncrief WC, Marshall GW. The evolution of the Seven Steps of Selling. Ind Mark Manag. 2005;34(1):13-22.
27. Rodrigues ML, Körting TS, de Queiroz GR, Sales CP, Silva LARd. Detecting Center Pivots In Matopiba Using Hough Transform And Web Time Series Service. IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS); 2020.

28. Stickdorn M, Schneider J. This Is Service Design Thinking: Basics, Tools, Cases. BIS Publishers; 2010.

29. Sturm RA, Frudakis TN. Eye colour: Portals into pigmentation genes and ancestry. Trends Genet. 2004;20(8):327-332.

30. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era: Concepts and misconceptions. Nat Rev Genet. 2008;9(4):255-266.

31. Blackmore S. The Meme Machine. Oxford University Press; 2000.

32. Felsenstein J. Inferring Phylogenies. Sunderland, MA: Sinauer Associates; 2004.

33. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 2018;46(D493-D496).

34. Hung K-C. An Evolution Framework Cross-Border Mergers and Acquisitions by Living Systems Approach. Ph.D. dissertation, Feng-Chia Univ. Taiwan; 2014.

35. Chen AHL, Liang Y-C, Chang W-J, Siauw H-Y, Minanda V. RFM Model and K-Means Clustering Analysis of Transit Traveller Profiles: A Case Study. J Adv Transport. 2022;2022:1108105.

36. Handojo A, Pujawan N, Santosa B, Singgih ML. A multi-layer recency frequency monetary method for customer priority segmentation in online transaction. Cogent Eng. 2022;10(1).

37. Zang W, Qian Y, Song H. 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. 2022;19(5):3043.