Digital Intelligence Banking of Adaptive Digital Marketing with Life Needs Control

Ryosuke KONISHI [a] Fumito NAKAMURA [a] Yasushi KIYOKI [b]

[a] *Generic Solution Corporation, Nampeidai-chou, Shibuya-ku, Tokyo, Japan*
[b] *Faculty of Environmental Information, KEIO University, Fujisawa, Kanagawa 252, Japan*

Abstract. While individuals benefit from the goods and services provided by companies that enrich their lives and that have adapted to a dynamic environment that is always changing, these companies pay a high communication cost to access opportunities to provide these goods and services and to seek a better understanding of individual customers' changing needs. Although vast amounts of information can be obtained, databases and machine learning are playing an increasingly important role in extracting meaning from this information, turning it into meaningful information assets that consider circumstances and contexts, and individualizing the economy of information. I propose an implementation method for providing information to enrich the profiles of individual customers by consolidating different data, calculating the individual customers' needs through the relationships between customers and products, evaluating the change in relationships between individual customers and products over time, and providing goods and services to suit different intervals of change to factors such as lifestyle and living environment. As there are different factors involved in estimating the incidence of needs, and different frequencies and rates at which they occur, based on the special characteristics of products, different data are required to estimate such needs. By profiling individuals over the long term, it is possible to build an information provision environment that is conducive to companies' customer acquisition.

Keywords. Recommendation system, Mathematical Model of Meaning, Logistic Regression Mixture Model, Local Variational Inference, Retail Application, Hawkes process

1. Introduction

We live in an era in which digitalization technology, along with high-speed communication, large volumes of data, and advanced artificial intelligence technologies such as machine learning, can effectively support the use of information. Today, companies can not only send various types of messages to a massive number of customers and obtain their responses but also interact with them and get feedback on products and services through messaging. There is a growing expectation that further development of information digitalization will make it possible to store a huge amount of information in the form of databases that can be applied to digital marketing. In addition, the popularization of smartphones has increased the number of digital points of contact with customers, enabling the consideration of reachability, segmentation, perception, and recordability

when evaluating such messaging media and recording their position, time, and response in more detail. Given the need to find meaning from information, convert it to a meaningful property that considers the situation and context, and personalize its economic efficiency, information technologies involving databases and machine learning are expected to be advanced further.

The era of digital marketing is forcing companies to change their business formats and break away from single channels with only brick-and-mortar stores. The same can be said about marketing channels, which require a paradigm shift to digital omni-channels and need to stop relying only on traditional distribution channels and sales methods. Thus, in this new era, adopting an omni-channel strategy is essential for survival. The essence of marketing is to generate and control customer needs for products and direct the flow of products and services from manufacturers to consumers and users by combining products, distribution, promotion, and prices. To this end, it is necessary to capture the behavior of each customer, record it as data, and monitor its trend for long periods.

The advertisement industry uses a technique that considers the processes of attention, interest, desire, memory, and action (AIDMA). For each of these processes, it identifies elements of perception, knowledge, feelings and emotions, intentions, and actions that form a consumer's attitude and measures advertising effectiveness by applying an attitude scale used in psychology. Companies are required to turn these processes into data and understand each process from the perspective of customers for all products and services. To ensure communication that encompasses the entire living environment of the dweller or consumer and stimulate these processes, as each person assigns a different value to information and requires it at different times, it is important for companies to specialize for each customer, design the value of information, record it, and use it to predict demand.

Response: Customer responses to messages Recommend: Recommendations made by companies based on analysis of customer responses Reach: Messages reach both companies and customers Relation: Sell products and services through relationship building Repeat: Multiple uses of products and services Retention: Maintain a customer base by transforming them into repeat purchasers Royalty: Win customer loyalty by gaining their trust

Several advertisement techniques are based on passive information, but users normally provide poor responses to information they did not request. Merely recorded information increases asymmetry, driving companies to implement massive promotion measures for marketing activities that explore the emergence of customer needs. Therefore, for companies to detect customer interest in products that emerges with life events, obtain information close to the timing of their purchase intention, and understand the customer's needs, they must evaluate the economic efficiency of the information through the use of a database.

When companies regard each customer as a human being, they can record marketing activities from the customer's human side, taking into consideration what kind of products and services would make their lives more comfortable, from birth till death. It can be said that understanding customers through communication is becoming an increasingly important element. In this context, companies are increasingly expected to implement digital marketing strategies that predict the timing of customer needs according to their life stages and events.

The expansion of contact points with customers faced limits. In face-to-face channels —because customer information, service history, and past suggestions are not shared

across channels in real time—when collecting information regarding a first-time customer, it is necessary to interview them from the ground up; thus, the process from suggestion to purchase takes a considerable amount of time.

It is possible to identify the relationships among products from different backgrounds such as durable goods (e.g., cars, houses, furniture, household appliances), consumer goods (e.g., food, daily-use items), non-durable goods (e.g., eating out, cultural entertainment), and luxury consumer goods (e.g., travel, jewelry) as well as those among the methods of payment used to purchase such goods and financial products such as mortgages, fund management, and asset formation. Customers' opportunity and frequency of purchase of these products and services varies in their life. Companies are also expected to ensure consistency between customers' life events, evaluate their economic power including consumption capability (e.g., work and family history, asset status), analyze how it changes over time, and suggest products and services with adequate pricing. Financial institutions, particularly banks, possess a large number of information assets obtained through payment and deposit data for each customer. These information assets create opportunities for them and other companies to suggest financial products and offer services that suit customers' life events and style. Such use of information is expected to play an important role in the stimulation of processes through corporate marketing activities in the next years. While companies are working to stimulate marketing processes individually by offering products and services with different frequencies and cycles, banks and other financial institutions are expected to understand customers' global needs from their perspective and effectively support the stimulation of marketing processes. Banks record not only customers' attributes but also their purchase history and behavior through various points of contact, and, based on the potential needs identified, make suggestions from various channels through establishing contact in real time and immediately evaluate the causal effect and prediction accuracy of those suggestions. Based on the assessment results, they can quickly make further suggestions according to the customer needs identified. In such processes, the advancement of information technologies related to databases, machine learning, and big data analysis plays an essential role.

The advancement of such digital banking services will allow companies to build full profiles of their customers, enabling them to offer products and services according to their life events and styles and suggest related financial products. The effect of this process can be maximized if companies reduce the asymmetry of the information volume designed to learn about customers and make the information acquisition process more efficient. Targeting can play the most important role in maximizing the return on marketing investments.

2. Related Works

We have previously proposed a recommendation system that utilizes action and demand vector simultaneously and updates a recommendation list in a same cycle [1]. After this work, we have proposed another system that utilizes the action vector more frequently than the demand vector [2]. In this paper, we introduce a new concept life time that is longer interval to update than the action and the demand.

3. Basic Theories

Here, we introduce the methodologies used in this study. Specifically, we use the mathematical model of meaning (MMM) and machine learning; Bayesian inference of a logistic regression mixture model (LRMM) and Hawkes process.

3.1. Mathematical Model of Meaning

The MMM was first used to extract semantic information behind data deterministically [3]. Let $X \in \mathbb{R}^{N \times M}$ be a data matrix, where N is the number of data and M is the number of features. We normalize each column of the matrix by the 2-norm, and denote the resulting normalized matrix by $\tilde{X}$, i.e. the (i, j)th element of the matrix is

$$\tilde{X}_{i,j} = \frac{X_{i,j}}{\sqrt{\sum_{k=1}^{N} X_{k,j}^2}}. \tag{1}$$

This is referred to as the fundamental data matrix. For the product of (1) $\tilde{X}^T \tilde{X}$, we calculate the eigenvalues and eigenvectors. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M \geq 0$ be the eigenvalues and $v_1, v_2, \ldots, v_M$ the corresponding eigenvectors , $Y = [Y_1^T, Y_2^T, \ldots, Y_t^T] \in \mathbb{R}^{t \times M}$ $t \times M$ context matrix, and $Q = [v_1, v_2, \ldots, v_M] \in \mathbb{R}^{M \times M}$. Then, for a threshold ϵ, MMM determines the index set of the chosen eigenvectors Λ_ϵ by

$$\Lambda_\epsilon = \left\{ j \in \{1, 2, \ldots, M\} | 0 < \epsilon < 1, \frac{(Q_Y)_j}{\|Q_Y\|_\infty} > \epsilon \right\}. \tag{2}$$

Here, $Q_Y = \sum_{i=1}^{t} Y_i Q$, $(Q_Y)_j$ is the j-th element of Q_Y, and $\|Q_Y\|_\infty = \max_{1 \leq j \leq M} |(Q_Y)_j|$. Arranging the chosen eigenvectors, we obtain the semantic projection

$$P(Y) = [v_j^T | j \in \Lambda_\epsilon] \in \mathbb{R}^{|\Lambda_\epsilon| \times M}, \tag{3}$$

where $|\Lambda_\epsilon|$ represents the cardinality of Λ_ϵ. When we apply the projection to each datum $g \in \mathbb{R}^M$, the datum is regarded as an element of a semantic space, denoted by $P(Y)g$. In this semantic space, we can measure the distance between the datum and the semantic centroid $\bar{D}$:

$$\bar{D} = \frac{1}{\|P_Y\|_\infty} P_Y, \tag{4}$$

where $P_Y = \sum_{i=1}^{t} P(Y)Y_i$. Let $\bar{D}_j$ and $(P(Y)g)_j$ be the j-th elements of the vectors; the distance is calculated as the weighted Euclidean distance $dist(\bar{D}, g)$,

$$dist(\bar{D}, g) = \sqrt{\sum_{j=1}^{|\Lambda_\epsilon|} c_j \left(\bar{D}_j - (P(Y)g)_j \right)^2}, \tag{5}$$

where $c_j = \frac{(P_Y)_j}{\|P_Y\|_\infty}$.

3.2. Logistic Regression Mixture Model and Local Variational Inference

LRMM is a mixture of a logistic regression model, and used to solve classification problems like those found in marketing, clinical trials, and psychology [4,5,6]. This model represents the probability of a binary output y given an input data x as follows:

$$p(y|x,w) = \sum_{k=1}^{K} a_k r(x,b_k)^y (1 - r(x,b_k))^{1-y}, \tag{6}$$

where $y \in \{0,1\}$, $x \in \mathbb{R}^M$, $b_k \in \mathbb{R}^M$, $0 \le a_k \le 1$ for $1 \le k \le K$, $\sum_{k=1}^{K} a_k = 1$, and $r(\cdot)$ is a sigmoid function $r(x,b_k) = 1/(1 + \exp(-x^{\mathrm{T}} b_k))$. $w = (a_1, a_2, \ldots, a_K, b_1, b_2, \ldots, b_K)$ is a parameter, and we must estimate it or the probability distribution itself from n pairs of input and output $(x^n, y^n) = (x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n)$.

Several methods have been proposed to estimate the parameters [7,8,9]. In this study, we use the method proposed by [1] to estimate the probability distribution because it is more accurate than the other methods. Moreover, the method leverages mini-batch learning; thus it learns efficiently. The calculation is as follows. In the first step, we assume a prior and a posterior distribution of the parameter given the training data. Let $\varphi(w)$ be the prior distribution, and $p(y^n|x^n, w)$ the likelihood of (6); then, the posterior distribution of the parameter $p(w|x^n, y^n)$ is obtained as follows:

$$p(w|x^n, y^n) = \frac{1}{Z} p(y^n|x^n, w) \varphi(w), \tag{7}$$

where $Z = \int p(y^n|x^n, w) \varphi(w) dw$. By using the posterior distribution, we obtain the output distribution $p(y|x, x^n, y^n)$ that we want to predict, which we call the predictive distribution in this paper:

$$p(y|x, x^n, y^n) = \int p(y|x,w) p(w|x^n, y^n) dw. \tag{8}$$

However, the normalization constant Z of the posterior distribution is intractable. To avoid the problem, we leverage an approximated posterior distribution replacing (7). Local variational inference (LVI) with mini-batch learning can construct the approximated distribution efficiently and accurately. [1]. In mini-batch learning, we first divide the whole data (x^n, y^n) into T mini-batches of data:

$$\{(x^n, y^n)\} = \{(x^{n_1}, y^{n_1}), (x^{n_2}, y^{n_2}), \ldots, (x^{n_T}, y^{n_T})\}.$$

Let the prior distribution $\varphi(w)$ be

$$\varphi(w) \propto \prod_{k=1}^{K} a_k^{\alpha-1} \exp\left(-\frac{\beta}{2} b_k^{\mathrm{T}} b_k\right), \tag{9}$$

where $\alpha, \beta \in \mathbb{R}_+$ are hyperparameters that are tuning parameters. Then, the approximated posterior given the t-th partition $(x^{n_t}, y^{n_t}) = ((x_{1,t}, y_{1,t}), \ldots, (x_{n_t,t}, y_{n_t,t}))$ is written as follows [1]:

$$\hat{p}(w|x^{n_t}, y^{n_t}) \propto \prod_{k=1}^{K} a_k^{\hat{\alpha}_k^t - 1} \exp\left(-\frac{1}{2}(b_k - \hat{\mu}_k^t)^{\mathrm{T}} \hat{\beta}_k^t (b_k - \hat{\mu}_k^t)\right) \tag{10}$$

$$\hat{\alpha}_k^t = \hat{\alpha}_k^{t-1} + \sum_{i=1}^{n_t} u_{ik,t}(\xi), \tag{11}$$

$$\hat{\beta}_k^t = \hat{\beta}_k^{t-1} - \sum_{i=1}^{n_t} 2 u_{ik,t}(\xi) v_{ik,t}(\eta) x_{i,t} x_{i,t}^{\mathrm{T}}, \tag{12}$$

$$\hat{\mu}_k^t = (\hat{\beta}_k^t)^{-1} \hat{\kappa}_k^t, \tag{13}$$

$$\hat{\kappa}_k^t = \hat{\kappa}_k^{t-1} + \sum_{i=1}^{n_t} u_{ik,t}(\xi)(y_{i,t} - 0.5) x_i. \tag{14}$$

For simplicity, we set the initial conditions as $\hat{\alpha}_k^0 = \alpha$, $\hat{\beta}_k^0 = \beta I_M$, and $\hat{\kappa}_k^0 = 0$. In (10)-(14), it is necessary to determine the local parameters $u_{ik}(\xi)$ and $v_{ik}(\eta)$. They are represented as follows:

$$u_{ik,t}(\xi) = \frac{\exp(h_{ik,t}(\xi))}{\sum_{l=1}^{K} \exp(h_{il,t}(\xi))}, \tag{15}$$

$$v_{ik,t}(\eta) = -\frac{\tanh(\frac{\sqrt{g_{ik,t}(\eta)}}{2})}{4\sqrt{g_{ik,t}(\eta)}}, \tag{16}$$

$$h_{ik,t}(\xi) = \Psi(\hat{\alpha}_k^t) - \Psi\left(\sum_{k=1}^{K} \hat{\alpha}_k^t\right) + (y_{i,t} - \frac{1}{2}) x_{i,t}^{\mathrm{T}} \hat{\mu}_k^t - \log 2 \cosh\left(\frac{\sqrt{g_{ik,t}(\eta)}}{2}\right), \tag{17}$$

$$g_{ik,t}(\eta) = x_{i,t}^{\mathrm{T}}(\hat{\mu}_k^t (\hat{\mu}_k^t)^{\mathrm{T}} + (\hat{\beta}_k^t)^{-1}) x_{i,t}, \tag{18}$$

where $\Psi(\cdot)$ is the digamma function, i.e. $\Psi(x) = \frac{d}{dx} \log \Gamma(x)$, where $\Gamma(\cdot)$ is the gamma function. Note that (10) - (14) and (15) - (18) include their opponent each other, thus by iterative update the method can attain a local minimum. Algorithm 1 outlines this approach.

The predictive distribution for the t-th partition is approximated by replacing the obtained approximated posterior distribution with the posterior distribution used in (8), as follows:

$$p(y|x, x^{n_t}, y^{n_t}) \approx \int p(y|x, w) \hat{p}(w|x^{n_t}, y^{n_t}) dw \approx \begin{cases} \sum_{k=1}^{K} \hat{a}_k \tilde{r}(x) & (y = 1) \\ \sum_{k=1}^{K} \hat{a}_k (1 - \tilde{r}(x)) & (y = 0) \end{cases}. \tag{19}$$

Here, $\hat{a}_k^t = \hat{\alpha}_k^t / \sum_{l=1}^{K} \hat{\alpha}_l^t$ and $\tilde{r}(x)$ is

$$\tilde{r}(x) = r\left(\frac{x^{\mathrm{T}} \hat{\mu}_k^t}{\sqrt{(1 + \pi x^{\mathrm{T}} (\hat{\beta}_k^t)^{-1} x)/8}}\right).$$

Algorithm 1 Mini-batch LVI for LRMM

1: **procedure** MINIBATCH($\alpha, \beta, iteration, tol, partition$)
2:     $F \leftarrow \infty$
3:     Initialize $\hat{\alpha}_k, \hat{\beta}_k, \hat{\mu}_k$
4:     **for** $t \leftarrow 1, \infty$ **do**
5:         pick up next (x^{n_t}, y^{n_t})
6:         **while** $ite \leq iteration$ **do**
7:             $\hat{g}_{ik}(\eta) \leftarrow$ (18), $\hat{v}_{ik}(\eta) \leftarrow$ (16)
8:             $\hat{h}_{ik}(\xi) \leftarrow$ (17), $\hat{u}_{ik}(\xi) \leftarrow$ (15)
9:             $\hat{\alpha}_k^t \leftarrow$ (11)
10:            $\hat{\beta}_k^t \leftarrow$ (12)
11:            $\hat{\mu}_k^t \leftarrow$ (13)
12:            $ite \leftarrow ite + 1$
13:        **end while**
14:    **end for**
15: **end procedure**

3.3. Hawkes Process

Hawkes process is one of a point process which is modeled to describe the timing of events like the timing of purchase, e-mail message, tweet, earthquake, and so on [10,11]. Especially, Hawkes process leverages past events, that is the more events happen in the past, the more likely they are to happen at the current time. We utilize the process to classify a current status of life event.

Let $\lambda(t)$ be an intensity function at time t, then the probability distribution of the number of event occurrence $N(t, t+\delta)$ between t and $t+\delta$, where $\delta > 0$ is a sufficiently small real number, is described as,

$$p(N(t, t+\delta) > 1) = o(\delta),$$
$$p(N(t, t+\delta) = 1) = \lambda(t)\delta + o(\delta),$$
$$p(N(t, t+\delta) = 0) = 1 - \lambda(t)\delta + o(\delta).$$

Here, $o(\delta)$ represents a term that converges to 0 faster than δ. In particular, the Hawkes process consists of the following intensity function:

$$\lambda(t) = \mu + \sum_{j|t_j < t} g(t - t_j), \tag{20}$$

where $\mu \in \mathbb{R}_+$ and t_j is the timing of the event. $g(t - t_j)$ is non-negative real valued function and usually used the following exponential kernel:

$$g(t - t_j) = \alpha\beta \exp(-\beta(t - t_j)),$$

where $\alpha, \beta \in \mathbb{R}_+$. Figure 1 shows a intensity function of the Hawkes process and timing of events. The solid line shows the intensity function $\lambda(t) = 10 + \sum_{j|t_j < t} \exp(-(t - t_j))$, and the cycle shows the timing.
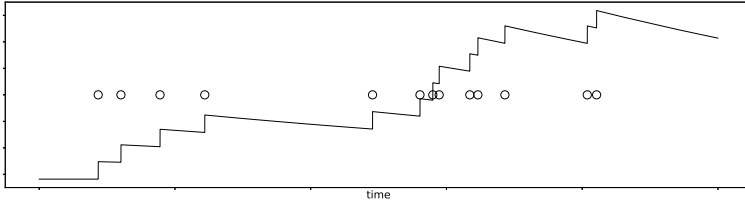
Figure 1. Example of Hawkes process with intensity function $\lambda(t) = 10 + \sum_{j|t_j<t} \exp(-(t-t_j))$.

In practice, we should decide the parameters μ, α, and β from the timing data $t^n = (t_1, t_2, ..., t_n)$. Several methods to infer the parameters have been proposed [12,11,13,14]. In this paper, we optimize the following objective function: the function $L(\mu, \alpha, \beta | t^n)$ is formalized by negative log likelihood with a regular regularization term

$$L(\mu, \alpha, \beta | t^n) = -\sum_{i=1} \log\left\{\mu + \sum_{j|j<i} \alpha\beta \exp(-\beta(t_i - t_j))\right\}$$

$$+ \left\{\mu T + \sum_{i=1}^{n} \alpha(1 - \exp(-\beta(T - t_i)))\right\} + \frac{c_1}{2}\mu^2 + \frac{c_2}{2}\alpha^2 + \frac{c_3}{2}\beta^2, \quad (21)$$

where T is total time that observes the events and $c_1, c_2, c_3 \in \mathbb{R}_+$ are hyperparameters. To minimize the objective function, we use quasi-newton method.

4. Implementation for Architecture

This section specifically shows the method to implement our recommendation system. Previous work [2] assumed that personalized recommendation system is necessary because the needs highly depend on each customer, and utilizes demand and action vectors to personalize the system. Additionally, we assume that the customer's behavior highly depends on their stage of life event, therefore we detect the status of the life event in this paper. Figure 2 shows a calculation flow that takes the above concept into account. In the term and cycle phase, the demand and action vector are updated respectively, and the stage of the life event is classified in the life time phase. The difference among the phases is their interval; the cycle phase is the shortest interval among the phases such as 1 hour, 30 minutes, 1 minutes, the term phase is the intermediate such as 1 week, 1 month, and the life time phase is the longest such as 1 year, 10 years. Note that we assume that the products information like assortment, sale products is updated in the term phase. We will explain each phase one by one.

4.1. Initialization phase

The system initializes the setting here. Since the system has no information about action vector at this phase, it utilizes the product information and the purchased information for each customer at the first cycle. The process comprises the following steps.
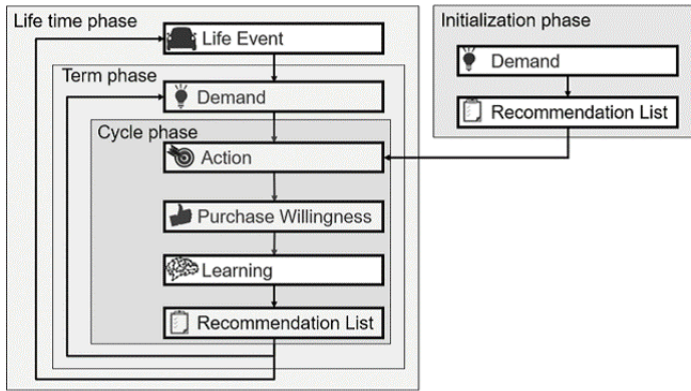
Figure 2. Calculation flow considered the above concept.

Step 1-1.

Determination for Recommendable products:

The system picks up products that can be recommended in the current term, and regard the normalized ones as a fundamental data matrix $\tilde{X} \in \mathbb{R}^{N \times M}$ as we explained at section 3.1. Here, N is the number of possible products to recommend, and M the number of their features.

Step 1-2.

Generation of a Demand Vector:

Let Y_j be a context matrix that represents feature matrix of products that was purchased previously by the j-th customer, then the system can calculate the semantic projection $P(Y_j)$ in (3). Using the projection, the system projects a feature vector of a product onto the meaning space which we call the demand vector in this paper. For feature vector $g_s \in \mathbb{R}^M$ of an product s, the system projects it onto the demand vector $P(Y_j)g_s$.

Step 1-3.

Determination of Recommended Items:

The system calculates a weighted Euclidean distance $dist(\bar{D}, g_s)$ in (5) between the s-th demand vector $P(Y_j)g_s$ and the centroid of the meaning space $\bar{D}$ in (4), and recommends the products in ascending order of the distance by the number of possible indicating products L.

4.2. Cycle phase

After initialization phase, a customer's phase moves to the cycle phase. In the cycle phase, the system updates the recommendation list of customers that purchased something in the previous cycle. The process is the following steps.

Step 2-1.

Aggregating Customer's Action:

In this step, the system aggregates the customer's action log in the previous cycle.

Step 2-2.

Establishing Purchase Willingness:

After aggregating the customer's action, the system calculates a purchase willingness for each product. The value of the purchase willingness W_s for action vector r_s of product s is set as follows:

$$W_s = \begin{cases} 1 & (dist(\bar{a}, r_s) \leq \epsilon_r) \\ 0 & \text{(otherwise).} \end{cases} \tag{22}$$

Here, $\bar{a}$ represents a centroid of action vectors that was purchased previously, and $dist(\bar{a}, r_s)$ represents the Euclidean distance between $\bar{a}$ and r_s. The purchase willingness represents whether the action vector is close to the action vectors previously purchased or not.

Step 2-3.

Updating Learning Model:

In this step, the system adopts the mini-batch LVI for LRMM in section 3.2 to score the products. The demand and the action vectors of the products are the inputs x^{n_t}, and the purchase willingness of the products are the outputs y^{n_t}. Note that the learning model assumes that a dimension of the input does not change dynamically, thus the system initializes the model if the context is updated because the new semantic projection may have different dimensions or represent other demand even same dimensionality.

Step 2-4.

Updating Recommendation Items:

In this step, the system updates the recommendation products on the basis of predicted strength of the purchase willingness, which is obtained by $\hat{p}(y = 1|x, x^{n_t}, y^{n_t})$ in (19). Since there is no information on which products will be purchased or reacted in the next cycle, the system substitutes the input of the next cycle for the current ones that was utilized for learning. The system recommends the products by ascending order of the strength $\hat{p}(y = 1|x, x^{n_t}, y^{n_t})$ by the recommendation size L.

4.3. Term phase

After several cycle updates, the customer's phase moves to the term phase. In the term phase, products information is updated, therefore the system updates the demand vectors. The process is the following steps.

Step 3-1.

Updating Recommendable Products:

Like step 1-1, the system picks up the products matrix that that can be recommended in the current term, and normalizes it.

Step 3-2.

Updating Demand Vector:

In this step, the system updates the context for the customers that purchase something in the previous term. After that, the semantic projection is recalculated, and the demand vector for the current term is updated.

4.4. Life time phase

After several term and cycle updates, the customer's phase moves to the life time phase. In the life time phase, the system classifies the stage of the life time, and the context is specialized on the basis of the life time stage. The process is the following steps.

Step 4-1.

Defining Life Time Products:
In this step, the system sets the life time products for each life event such as marriage, child care, education, and so on. For example, a marriage ring is in the marriage event, a house is in the child care event.

Step 4-2.

Predicting Status of Life Event:
In this step, the system detects the status of the life event. Let $\{1, 2, ..., L\}$ be a category set of each life event $t_l^n = (t_l^{(1)}, ..., t_l^{(n_l)})$ the timing of the life event l, where $l = 1, ..., L$, then the estimated intensity function $\lambda_l(t)$ for the life event l is obtained by minimizing (21). Since the intensity function is proportional to the event occurrence probability,

$$\hat{l} = \arg \max_{1 \leq l \leq L} \lambda_l(t)$$

is regarded as the label of the current life stage.

Step 4-3.

Updating Context Based on Life Stage:
In this step, when the estimated life stage $\hat{l}$ is different from the previous one, the system detects a change of life event. In that situation, the context is initialized on the basis of products that was previously purchased by other customers in the same life event.

4.5. Algorithm

Summarizing sections 4.1, 4.2, 4.3, 4.4, we can describe Algorithm 2 to implement them. Note that the sections 4.1, 4.2, 4.3, and 4.4 are based on one customer, but in practice, the entire system needs to consider all customers, therefore a step of registration for customers to be calculated is appended in the algorithm.

Algorithm 2 Proposed Recommendation System

Require: $C = \emptyset$: A set of customers
 1: **procedure** RECOMMENDATION(L)
 2:     **for** $i \leftarrow 1$, sufficiently long time **do**
 3:         Wait for a cycle.
 4:         **if** i = 1 **then**
 5:             Check a set of customers C' who purchased items in the past time.
 6:         **else**
 7:             Check a set of customers C' who purchased items among the cycle.
 8:         **end if**
 9:         $C = C \bigcup C'$
10:         **for each** $c \in C$ **do**
11:             **if** c is appended at time i **then**
12:                 Procedure for initialization phase, i.e. section 4.1.
13:             **else**
14:                 **if** Life time is changed **then**
15:                     Procedure for life time phase, i.e. section 4.4.
16:                 **end if**
17:                 **if** Term is changed **then**
18:                     Procedure for term phase, i.e. section 4.3.
19:                 **end if**
20:                 **if** Cycle is changed **then**
21:                     Procedure for cycle phase, i.e. section 4.2.
22:                 **end if**
23:             **end if**
24:             Display the recommendation list for customer c with the size L.
25:         **end for**
26:     **end for**
27: **end procedure**

---

5. Conclusion

In this study, we proposed a personalized recommendation system that not only consider about simple use of purchased information and action log, but also the purchased timing to detect specific life events. Compared to previous works, it is considered that our recommendation system may be proper for the area where products highly depend on the customer's life event such as finance. To validate our method and improve our model in tuning level, conducting numerical experiments based on synthetic and real data is our future work.

References

[1] R. Konishi, F. Nakamura, and Y. Kiyoki, "Estimating Adaptive Individual Interests and Needs Based on Online Local Variational Inference for a Logistic Regression Mixture Model," in *Knowledge Creation and Intelligent Computing (IES-KCIC)*, 2018.

[2] R. Konishi, F. Nakamura, and Y. Kiyoki, "Responsive Calibrated Web Personalization System with Online Local Variational Inference for Logistic Regression Mixture Model.," in *International Conference on Information Modeling and Knowledge Bases*, pp. 94–110, 2019.

[3] T. Kitagawa and Y. Kiyoki, "A mathematical model of meaning and its application to multidatabase systems," in *Research Issues in Data Engineering, 1993: Interoperability in Multidatabase Systems, 1993. Proceedings RIDE-IMS'93., Third International Workshop on*, pp. 130–135, 1993.

[4] J. Rost, "Logistic Mixture Models," in *Handbook of Modern Item Response Theory*, pp. 449–463, New York, NY: Springer New York, 1997.

[5] J. Rounds, "Logistic Models with Missing Categorical Covariates," *All Graduate Plan B and other Reports*, 2009.

[6] M. Kaptein and P. Ketelaar, "Maximum likelihood estimation of a finite mixture of logistic regression models in a continuous data stream," *arXiv preprint arXiv:1802.10529*, 2018.

[7] P. Wang and M. L. Puterman, "Mixed Logistic Regression Models," *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 3, no. 2, p. 175, 1998.

[8] D. Flake, "Separation of Points and Interval Estimation in Mixed Dose-Response Curves with Selective Component Labeling," *All Graduate Theses and Dissertations*, 2016.

[9] F. Nakamura, R. Konishi, and Y. Kiyoki, "Inference for Logitistic Regression Mixture Model with Local Variational Approximation and Study for Variational Free Energy," in *Information-Based Induction Science Workshop*, pp. pp.29–36, 2018. (in Japanese).

[10] M. Miyake, T. Ikegami, M. Oka, and Y. Hashimoto, "Web data analysis with hawkes process and histogram bin-width optimization," *The 31st Annual Conference of the Japanese Society for Artificial Intelligence*, 2017.

[11] M. Rizoiu, Y. Lee, S. Mishra, and L. Xie, "A tutorial on hawkes processes for events in social media," *arXiv preprint arXiv:1708.06401*, 2017.

[12] T. Ozaki, "Maximum likelihood estimation of Hawkes' self-exciting point processes," *Annals of the Institute of Statistical Mathematics*, 1979.

[13] S. Linderman, Y. Wang, and D. Blei, "Bayesian inference for latent Hawkes processes," *Advances in Neural Information Processing Systems*, 2017.

[14] R. Zhang, C. Walder, and M. Rizoiu, "Sparse Gaussian Process Modulated Hawkes Process," *arXiv preprint arXiv:1905.10496*, 2019.