Life-stage Prediction for Product Recommendation in E-commerce

Peng Jiang[†] Yadong Zhu[†] Yi Zhang[‡] Quan Yuan[†]
[†] Alibaba Inc, Beijing, China
[‡] University of California, Santa Cruz, USA
{jiangpeng.jp, edgewind.zyd, yuanquan.yq}@alibaba-inc.com
{yiz}@soe.ucsc.edu

ABSTRACT

Although marketing researchers and sociologists have recognized the large impact of life stage on consumer's purchasing behaviors, existing recommender systems have not taken this impact into consideration. In this paper, we found obvious correlation between life stage and purchasing behavior in many E-commerce categories. For example, a mum may look for different suitable products when her baby is at different ages. Motivated by this, we introduce the conception of life stage into recommender systems and propose to predict a user's current life-stage and recommend products correspondingly. We propose a new Maximum Entropy Semi Markov Model to segment and label consumer life stage based on the observed purchasing data over time. In the mom-baby product category where the life stage transition is deterministic, we develop an efficient approximate solution using large scale logistic regression and a Viterbi-like algorithm. We also propose a Gaussian mixture model to efficiently handle multi-kids life stage prediction problem. We integrate the life stage information predicted into the recommender system behind the largest online shopping website taobao.com. Both offline and online experiments demonstrate the effectiveness of the proposed life-stage based recommendation approach.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database applications–Data mining

Keywords

E-commerce, Recommender System, Life Stage

1. INTRODUCTION

Recommender systems have achieved much commercial success and is one of the most important data mining applications. For example, online stores such as Taobao, Amazon, and Walmart.com provide customized recommendations for additional products or services based on a user's historical purchasing records. It's well known that these recommendations are extremely important and have great impact on consumers' online shopping experience. Thus scholars and industry researchers in the data mining community have spent much effort on further improving recommendation performance, exploring all possible approaches. Most of the existing recommendation approaches can be mainly divided into two categories: content-based approaches [1, 2, 3] and collaborative filtering approaches [4, 5]. The content-based filtering recommend items similar to what a user likes before by representing the user's preferences using item content features, whereas the collaborative filtering approaches assume users with similar preferences on some items may also have similar preferences on other items.

On the other hand, marketing researchers and sociologists have recognized the importance of life stages on consumer's purchasing behaviors for many years[6, 7, 8]. For example, consumers go through various life stages, such as bachelor stage (i.e. young and single), newly married couples (young, no children), full nest (married couple with dependent children), empty nest (i.e. elder married couples with no children living together). Full nest stage can be further divided into a few substage, largely depending on the age of the dependent children; while empty nest stage can be further divided into sub stages such as head in labor force, retired, solitary survivors (i.e. old and single) etc. It's well recognized that both durable products and consumable products are sensitive to changes in life cycle stage.

In our E-commerce system, we also found strong correlation between life cycle stages and consumer purchasing behavior. For example, a woman will buy maternal vitamin during the pregnancy stage, then buy a baby car when her baby is born. A mum will have different purchasing needs during different life stages, which can be illustrated as figure 1. The figure shows that a user's needs and interests are dynamic and determined by her current life stage. Therefore, explicitly modeling and predicting life stages might be very useful for predicting the purchasing timed of products. We expect making recommendations based on this additional information could further improve the recommendation effectiveness. Unfortunately, most of the existing methods do not consider the concept of life stages, which exists in many verticals in E-commerce.

In this paper, we introduce the concept of life-stage into E-commerce recommendation systems. The system first predicts the current life stage of a user, then recommends products to the user based on the prediction and our proposed

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
KDD'15, August 10-13, 2015, Sydney, NSW, Australia.
© 2015 ACM. ISBN 978-1-4503-3664-2/15/08 ...$15.00.
DOI: http://dx.doi.org/10.1145/2783258.2788562.

probabilistic recommendation model, which can explicitly incorporate the life-stage information into an E-commerce recommender system.

To predict the life stage of a user, we proposed a new Maximum Entropy Semi Markov model for stochastic life stage segmentation and prediction over the user's past purchasing sequence. Besides, we developed a practical efficient large scale industry solution for life stage segmentation and labeling in mum-baby domain, where the life stage transition is deterministic. More specifically, we utilize the logistic regression classifier to predict the label and associated probability of each behavior sequence, and segment the whole sequence by Viterbi-like approach efficiently. We also handle the multi-kids problem via Gaussian mixture model.

To evaluate the effectiveness of the proposed life-stage based approach, we conduct extensive experiments with both offline data and online user studies for the mum-baby domain. The offline experimental results on a large dataset demonstrate that the proposed approach can significantly improve the performance in terms of prediction accuracy. Online experimental results also demonstrate the effectiveness of introducing life stage concept into a recommendation system. This work is integrated into the largest online shopping web site Taobao.com. Figure 2 is a snapshot of mum-baby product recommender system[1]. On Figure 2, the left side vertical navigation menu contains several life stages: pregnant, new born to 6 months, 6-12 months, 1-3 years, 3-6 years and 6-14 years, which describes the natural growth process of children. The system automatically locates a customer into a predicted (i.e. system selected) life stage. The right side recommends various products corresponding to the selected life stage.

The main contributions of this paper are:

1. We introduced the conception of life stages into E-commerce recommendation systems;
2. We proposed a new Maximum Entropy Semi Markov model for stochastic life stage segmentation and prediction;
3. We developed a practical efficient large scale industry solution for life stage segmentation and labeling in mum-baby domain, where the life stage transition is deterministic;
4. We proposed a probability model which can explicitly incorporate the life-stage information into an E-commerce recommender system.
5. We proposed a solution for modeling multi-kids scenario via Gaussian mixture models.
6. We verified of the effectiveness of the life-stage based approach in both offline and online scenarios.

The rest of the paper is organized as follows. The paper first provides some example based analysis in Section 2. Secondly, the life-stage based recommendation approach proposed is described in Section 3, followed by the analysis and modeling of multi-kids scenario in Section 4. Section 5 presents the experimental results, Section 6 describes the related work, and Section 7 concludes.

[1]http://www.taobao.com/market/baobao/2014/

Figure 1: An illustration of the changing needs during different life stages of a baby.

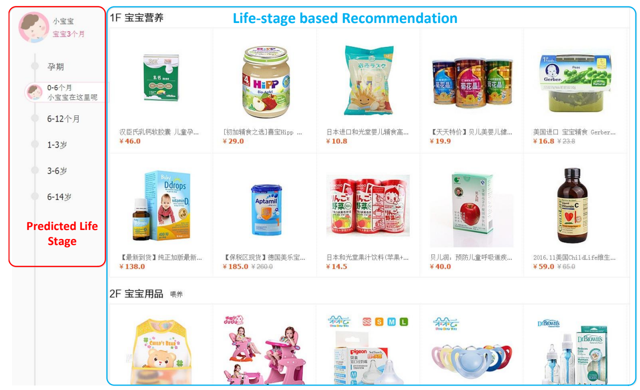

Figure 2: A screenshot of life-stage based recommender system deployed in Taobao.com.

2. LIFE STAGE AND CONSUMER BEHAVIORS

In this paper, we define life stage broadly and assume life stages are linked to certain tasks, which are the hidden reasons behind various purchasing choices. In addition, new purchasing tasks may emerge as a consumer transits from one life stage to the next one. Based on our broad definition, the conception of life stage exists in many vertical domains (i.e. categories) in E-commerce, such as home remodeling, car purchasing and wedding planning.

2.1 Data Analysis in Multiple Verticals

We carried out detailed data analysis in several verticals at taobao.com. We find that the conception of life-stage can be recovered from the purchasing behavior of users. The results on house remodeling vertical are shown in Figure 3. It shows that when a person enters into the house remodeling process, he or she may first buy some tiles for floor and emulsion paints for wall at the early stage. Next, the person may buy ceiling lights and shower faucets, followed by furniture such as dining table and beds. When all the above remodeling are completed, the person may buy formaldehyde clear spray for cleaning and bedding for the final stage decoration. We also find similar temporal purchasing patterns for other complex events that involve multiple stages, such as wedding planning and car purchasing.

2.2 Mum-Baby Domain

Although the proposed idea and major models can be applied to various events, the experiments and large scale solution described in this paper are focusing on the vertical recommendation of mum-baby products. In the mum-baby

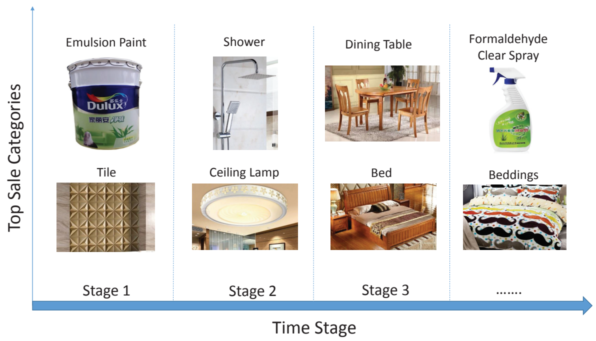

Figure 3: A data analysis of home remodeling in Taobao.com

vertical domain, the effect of life-stage on consumer behavior is the most obvious and very important. At different life stages of a baby, a woman needs to buy different matching products. For example, a woman is likely to buy maternity clothes at the prenatal stage, then she may look for milk powder after the baby is born. For a vertical recommender system, how to predict a user's life stage precisely and recommend appropriate products for the current life stage is a very important and challenging problem.

3. LIFE-STAGE BASED RECOMMENDATION

Life-stage of a user is a predefined set of phases in life, and each life stage spans a period of time. Changes in life stages may lead to changes in a person's purchasing needs. Examples of life-stage in maternal and child domain include pregnancy, new-born-baby, 1 - 3 months old, etc.

In our system, there are two major sub tasks. First, we label each user's behavior sequences so that we can tell a user's current life stage. Second, we generate recommendations based on the predicted life-stage. Thus the key components of life-stage based recommendations include how to model and predict life stages and how to give appropriate product recommendation correspondingly. The details of each component is described in this sections. Our experiments focus on the mum-baby domain, which is a special domain we will describe in more details.

3.1 Notations

The notations to be used in the rest of this paper are:

- X: the behavior sequence of a user. Behavior sequence X of a user is a sequence of user actions ordered by time. User actions in our system include clicking/collecting/buying an item, issuing a query, clicking navigating tips, etc.
- T: the length of X.
- X_t: the observed behavior sequence at time t.
- d_t: the duration of a life stage at time t.
- y_t: the life stage label at time t.
- l_{min},l_{max}: the minimum and maximum lengths of life stage respectively which are defined by domain experts.

3.2 MESMM for Life-stage Prediction and Segmentation

We considered several structured models (e.g. Hidden Semi-Markov Model[9], Maximum Entropy Markov Model[10], Conditional Random Fields[11], etc.) that may be used to model the joint distribution of life stages and user actions. Based on the characteristics of our problem, we propose a Maximum Entropy Semi Markov Model (MESMM) to segment and label consumer life stage based on the observed purchasing data over time. This model is motivated by Maximum Entropy Markov Models and Hidden Semi Markov Models. It is similar to Maximum Entropy Markov model except that the probability of changing the hidden state depends on the amount of time that has elapsed since entry into the current state. This model is general enough to model various scenarios.

In this model(see Figure 4), the probability of life stage y_t at time t depends on the previous life stage y_{t-1} at time $t-1$, how long the user has been in the previous life stage (i.e. d_{t-1}), and the observed user behavior sequence. The variable d changes deterministically. When life stage changes, d is reset to 0. Otherwise d decreases as time goes on. Therefore, d can capture the duration of user being in a life stage.

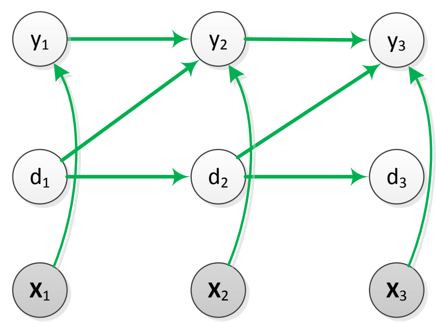

Figure 4: The Maximum Entropy Semi Markov Model for segmenting and labeling consumer life stages based on the observed purchasing data over time

Given an observed behavior sequence X, our goal is to find the best underlying life-stage sequence $y_1, ..., y_k$ and the corresponding duration $d_i, ..., d_k$:

$$\begin{aligned} &\{y_1, ..., y_k, d_i, ..., d_k\} \qquad (1)\\ &= \operatorname*{argmax}_{k,d_t,y_t} \prod_{t=1}^{k} P(y_t|y_{t-1}, d_{t-1}, X_t) P(d_t|d_{t-1}) \\ &\text{s.t. } l_{min} \le d_t \le l_{max} \end{aligned}$$

where X_t denotes the observed behavior sequence at time t. T denotes the length of X. l_{min} and l_{max} are the minimum and maximum lengths of a stage, which are defined by domain experts. $p(y_t|y_{t-1}, d_{t-1}, X_t)$ is the probability of being in y_t at time t given the previous stage y_{t-1}, the previous stage duration d_{t-1} and the observed behavior sequence X_t. The optimal segmentations and labels can be found using dynamic programming.

3.3 Mum-Baby Domain: An Efficient Solution

In our experiments, we mainly focus on the life-stage based recommendation in the mum-baby domain. The candidate life-stage labels/states are shown on Table 1. They are provided by marketing experts based on baby product standards and their domain knowledge.

Although MESMM is very general, learning and inference are computationally expensive for large data set and complex features. Fortunately, in the mom-baby product category, the transition and duration of life stages for a child are deterministic. We only need to find the birthday date, then the whole sequence can be labeled by a baby life stage template with a fixed sequence of life stages and fixed length in each stage.

Since most of our mum-baby consumers only have one child per family (due to the one child policy in our country), we first assume all our training data are from one child families. Multi-kids family scenarios are referred to the next Section.

In this scenario, we only need to focus on the problem of estimating the probability $P(y_t|y_{t-1}, d_{t-1}, X_t)$ in E.q 1. In our system, we train a large scale multinomial logistic regression (LR) model with $l1$ regularization. y_{t-1}, d_{t-1} and X_t are all input features for the LR model. To get training data for the multinomial logistic regression model, we first label users' behavior sequences in our training dataset based on the birth dates provided by users. The prediction and segmentation of life-stage for all users are implemented offline on parallel machines.

Table 1: Life-stage labels for Mom-baby domain

Label	Baby Age Category
0	Pregnant
1	New born - 6 months
2	6 - months
3	1 - 3 years
4	3 - 6 years

3.4 Features for the Classifier

To train $P(y|y', d', X_*)$, we convert each behavior subsequence X_* into a feature vector for training the multinomial logistic regression model. In general, the features fall in the following groups:

Category features: Many user actions are related to some items. Given a behavior sequence, we obtain a set of product IDs and could use a bag of IDs as features. However, this will lead to poor performance because of the sparsity issue. Besides, the set of items available in our system are changing frequently over time. To alleviate this problem, we represent each item using its category.

All items in Taobao are mapped to a category hierarchy. We expect consumer purchasing behavior patterns at different stages are more obvious on the categorical level. For example, a pregnant women is more likely to browse or buy products in 'Maternity Cloths' and 'Pregnant Bras' categories. A new parent is likely to browse 'Baby stroller' category. Figure 5 shows how purchasing ratio of top level categories relates to baby age.

We utilize all parent categories and first level categories an item belongs to as features. In order to reduce the influence of popular categories, the feature weight is set based on the TFIDF equation commonly used in the Information Retrieval community. To do so, each user is treated as a document and each category is treated as a term.

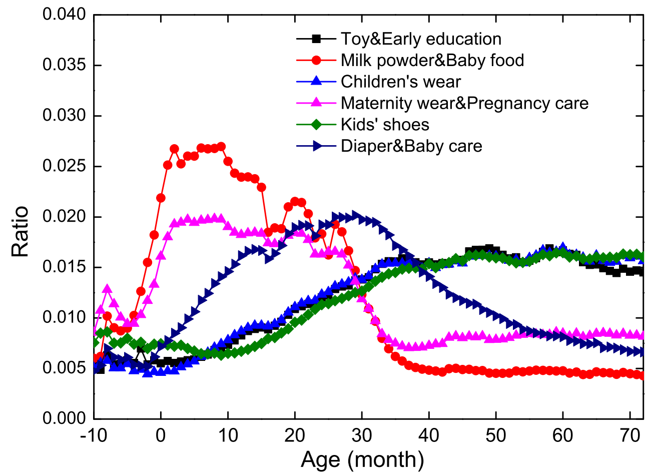

Figure 5: The purchasing ratio of 6 top level categories changes with a baby's age (unit: month). We normalize the value of a category at each time stage using a category specific normalization factor (i.e. the total number of products in the category bought over all those years)

Queries: A user behavior sequence may contain user product search activities. User search queries can directly reflect users' requirements, which can indirectly reflect a baby's age. For example, a user may search "large-size diaper" or "3 years old children's garments". This information is very important for age prediction. Therefore, search queries in E-commerce are also utilized as features. In our system, search queries are pre-processed using Chinese word segmentation and stop words removing techniques, and represented as word vectors.

Product property features: Taobao is a distributed market space where sellers sell products. Many sellers provide meta data about the products. For example, a seller will label 'Size' as "M" or "L" on children clothes, or label 'Age' as "newborn" or "1-3 years" etc. Each product facet-value pair (i.e. a product property) is associated with an unique feature.

Product title features: Product titles are created by sellers who are not affiliated with Taobao. Sellers are very creative about product titles. As a result, many product titles can be very informative about the life stage of the consumer. For example, *'Diaper for new born'* on a product title suggests the product is targeting for consumers with new babies. We preprocess these product titles via Chinese segmentation and phrase extraction techniques. Most of the extracted words are used as features, except stop words and words with very high inverse-document frequency.

Temporal Effect of Features: Whether a user has purchased a diaper 1 month ago or 2 years ago has deferent meanings in terms of predicting the current life stage. To capture the temporal patterns, we divide each consumer behavior time sequence X_* into multiple subsequences with fixed size time windows. A feature vector for each time window is generated. Then we concatenate all the feature

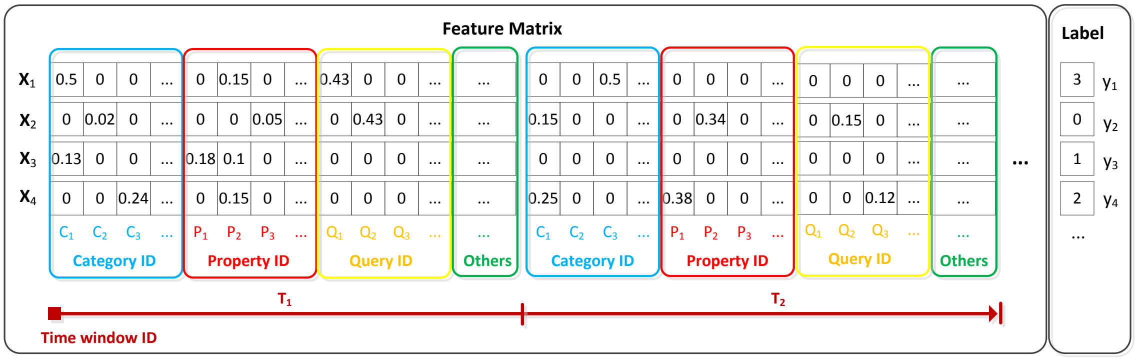

Figure 6: An example of feature matrix in our approach.

vectors into a big feature vector to represent X_*. Figure 6 illustrates an example of feature matrix, where each row corresponds to a training data point.

3.5 Recommendation Based on Life-stage

To make product recommendations, we propose a model to estimate the probability of a user purchasing a product at a specific age a. That is the joint probability $P(p\_product_j, a)$, where a is the baby's age. Let $P(p\_product_j)$ be the probability of the user purchasing the product j, and we calculate it for each user-product pair based on the existed recommendation algorithms. Let $P(a|p\_product_j)$ be the conditional probability of making the purchase at age a given that the user will purchase product j. Based on the chain rule, we have:

$$\begin{aligned} &P(p\_product_j, a) \\ &= P(a|p\_product_j)P(p\_product_j) \end{aligned} \tag{2}$$

Given a baby's age a, we can use the above value to rank all candidate products. For a user u without the age information, we first estimate his/her age distribution $p_u(a)$ based on the Gaussian mixture model, which will be introduced in Section 4. Then we rank the products based on:

$$P(p\_product_j) \int p(a|p\_product_j)p_u(a)da \tag{3}$$

where $p(a|p\_product_j)$ is the probability of the age of a random user purchasing product j (i.e. purchasing age), which we assume follows a Gaussian distribution $\mathcal{N}(\mu_j, \sigma_j^2)$. This distribution can be estimated from the age information of all users who have purchased this product. For a user without age information, we estimate the baby's birth date using the life stage prediction and segmentation method described previously, then infer the age of the user when purchasing each product in the past. Figure 7(a) shows an example of fitting a Gaussian distribution ($\mu = 43$) based on the histogram of purchasing age of users who purchased a product. It suggests that the product is most suitable for children at around 43-month old, where μ_j is the expected baby age for product j.

To estimate $P(p\_product_j)$ in E.q 3, we use a large scale logistic regression model:

$$P(p\_product_j) = \frac{1}{1 + e^{-\mathbf{w}^T\mathbf{x}}} \tag{4}$$

where x is a feature vector that represents a purchasing action, which includes scores generated by basic recommendation algorithms such as item-item based collaborative filtering, as well as other features that might be useful. $\mathbf{w}$ is learnt by maximizing the likelihood of the training data.

4. MIXTURE MODEL FOR MULTI-KIDS

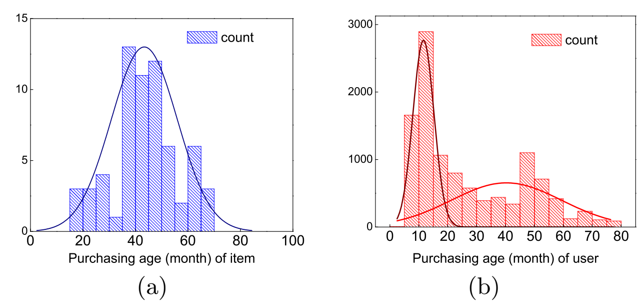

Figure 7: The distributions of item and user purchasing age (month). (a) Gaussian distribution of item purchasing age. This item is suitable for 41-month old children. (b) Gaussian mixture model with two components for user purchasing age. One component represents 11-month old, and the other one represents 40-month old.

Our previous analysis is based on the assumption that each user account corresponds to a family with only one child, which is very common in our market due to the one-child policy in China. However, we also find about 10% of our users bought maternity dress and baby products together with products for older kids at the same time. Many of those cases do not look like buying gifts for others. Recently, the one child policy is adjusted in China, which leads to an increasing number of families having two or more children. This motivates us to work on multi-kids life stage modeling and prediction.

One of the major challenges is that most of our users who registered their babies' age seldom register for multiple babies, thus we do not have much labeled training data for multi-kids modeling. Therefore we approach this problem based on unsupervised learning techniques.

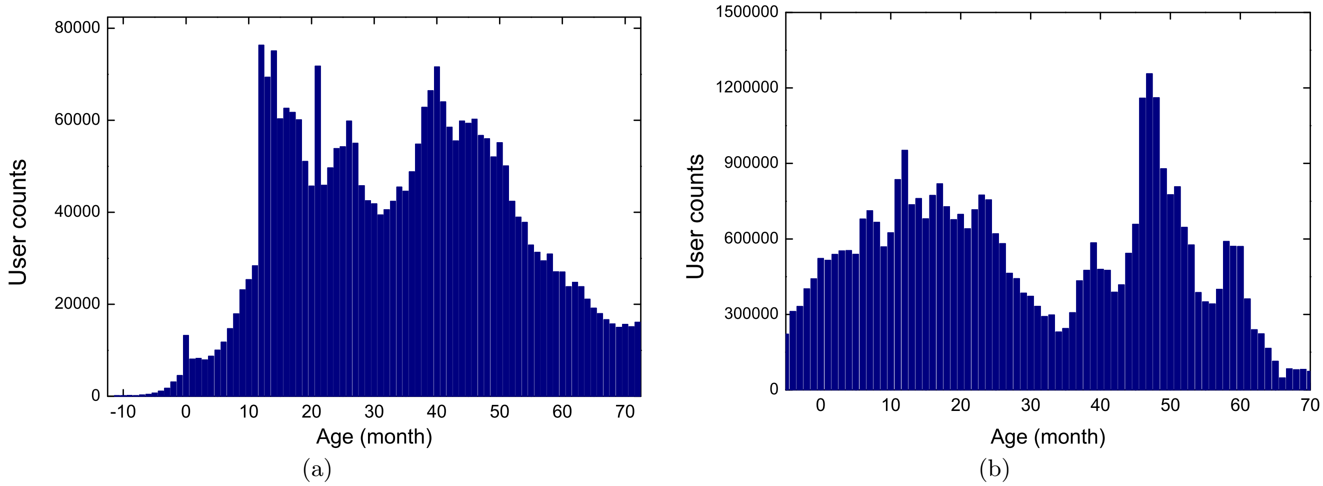

Figure 8: The distribution of user children's age (unit: month): (a) training dataset and (b) predicted dataset.

For a given user account, we observe a sequence of product purchasing behaviors. For each product j purchased by the user at time t_j, we estimate the expected purchasing age of the consumer $a_j = E(a|P(age|p\_product_j))$ at time t_j based on the age information of all consumers who have purchased the product, as described in the previous section. Then we can map the expected purchasing age to the projected age of the user at time t by adding the time difference between t and t_j: $a_{j,t} = a_j + (t - t_j)$.

We model the distribution of $a_{j,t}$ with a Gaussian Mixture Model (GMM):

$$p(a_{j,t}) = \sum_{c=1}^{K} w_c \mathcal{N}(a_{j,t} : \mu_c, \sigma_c^2) \quad (5)$$

where $c \in [1, K]$ is the index for each child, K is the total number of children, μ_c and σ_c are the mean and variance of c-th Gaussian component respectively, and w_c is the mixture weight for the component. Given a fixed K and the set of projected age of the user, the model parameters w_c, μ_c, σ_c can be found by the maximum likelihood estimation. To find the optimal K, we use the Akaike information criterion (AIC) and Bayesian information criterion (BIC) [12]. Based on this model, the age of the cth child at time t follows a Gaussian distribute centered on μ_c with variance σ_c as illustrated by the example in Figure 7(b).

5. EXPERIMENTS

In this section, we will evaluate the effectiveness of our approach empirically. We first introduce the experimental setup, then analyze the classification accuracy on offline dataset. We also give a data analysis of temporal effect. Additionally, we make an analysis of multi-kids situation. Finally, a real online evaluation is launched on Taobao.com.

5.1 Experimental Setup

We collected a dataset[2] with more than 8 millions children birth date information provided by consumers who share the information in order to receive better recommendations or search results. The age distribution till Feb 2015 is shown in Figure 8(a). The distribution is not uniform, because users usually do not disclose the age information when they are in the pregnancy stage. For each user with known birth date, we divided his/her activity sequence over time into different life stages, such as *'pregnant'*, *'new born-6 month'*, *'6-12 months'*, etc, using the fixed baby life stage template. This gives us labeled sequences for further analysis.

[2]The dataset sample can be downloaded from http://tianchi.aliyun.com/datalab/dataSet.htm?id=3

Table 2: Effects of features for maternal and child life-stage prediction.

Features&Approach	Accuracy
Basic	73.86%
Basic+Prop	76.14%
Basic+Prop+Query	76.31%
Basic+Prop+Query+Title	76.83%
Basic+Prop+Query+Title+Temp	81.99%
Basic+Prop+Query+Title+Temp+Seg	83.14%

5.2 Life-stage Prediction and Segmentation

For offline analysis, classification accuracy is used to measure the performance of mum-baby life-stage prediction and template based segmentation. We firstly study the effect of different feature sets and segmentation. For the multinomial logistic regression model, 5-fold cross validation is used. Table 2 shows the performance as more features are added. The logistic regression model with only product category features serves as the basic approach. It shows that the product meta data properties (Prop), title features (Title) and query features (Query) are effective. The biggest contribution is capturing the temporal effect of features (temp) with feature vectors at different time windows.

"+Seg" means the fixed baby life stage template is introduced to segment each user activity sequence, and the best starting point of the template is found to maximize the joint likelihood of the label sequence given the classifiers at each segment. The results show that introducing the template to force consistency improves the performance, as it achieves the goal of smoothing segments with less activities (i.e. poor features) by soft constraints from other segments.

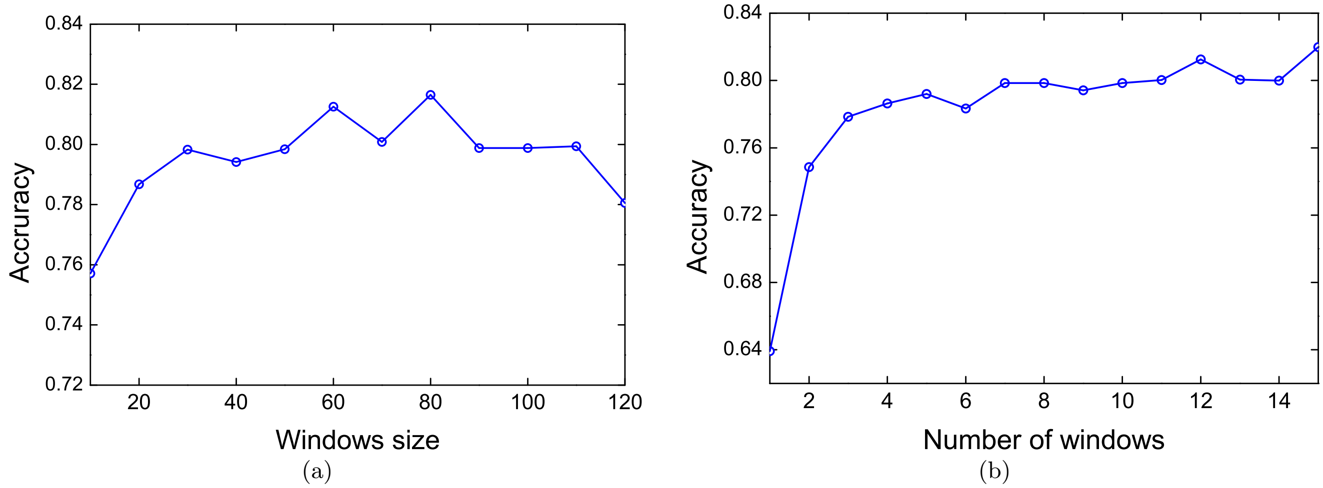

Figure 9: The analysis of window size and number.

The prediction and segmentation module is deployed to taobao.com. The predicted age distribution of all users till Feb. 2015 is shown in Figure 8(b). The KL-divergence is 0.2086 between predicted distribution and the distribution of the training dataset in Figure 8(a). There are relatively more users at the pregnancy stage than that of the training dataset.

5.3 Analysis of Temporal Effect

Temporal patterns play very important roles in baby age stage prediction. We divide each consumer behavior time sequence X_* into multiple time windows. The detailed analysis results are shown as figure 9. As we increase the window size, the prediction accuracy first increases, then does not change much, finally decreases. The best accuracy is achieved when the window size is around 40 to 90 as shown in Figure 9(a). On one hand, when the windows size is too small, the activities in a window are too sparse and limited. On the other hand, the discriminative power of temporal features is too weak when the window size is set too big. In the rest of this paper, the window size for each time span is set to 60, which is within the optimal range.

Given a fixed window size 60, we also change the number of windows from 1 to 15 to see how it affects the performance. As the window number increases, the performance starts to increase significantly before hitting a plateau (Figure 9(b)). This suggests a longer consumer behavior history gives more accurate prediction of the user's current life-stage.

5.4 Analysis of Multi-kids

As discussed before, an increasing number of families in China are having more than one child over the last several years, thus we propose the Gaussian Mixture Modeling approach (GMM) to predict the baby age(s) for each account. Although very few users provide multiple children's birth date, the learned GMMs suggest around 16.03% $(1,133,678)$ users in our training dataset have more than one child.

Figure 10 shows an example of choosing the number of children for an account using AIC and BIC as model selection criteria. And the lower values of y-axis means a better choice for the number of children.

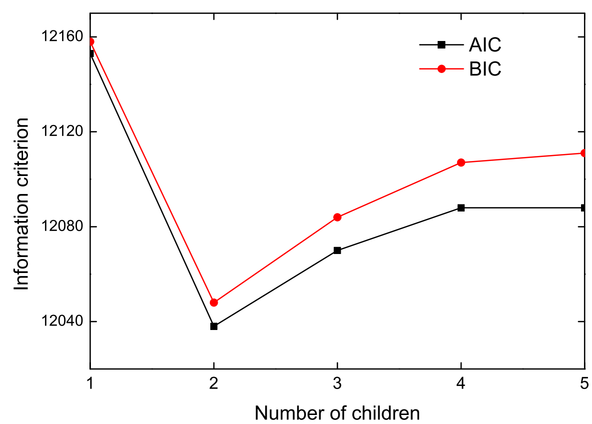

Figure 10: A case of AIC and BIC analysis to determine the number of children.

5.5 Online Experiments

Online experiments with real world E-commerce users are carried out to study the effect of life-stage based product recommendation. We conduct A/B testing based on a recommendation scenario in Taobao.com($100M$+clicks/day). We randomly sample some users with predicted mum-baby life stages and divide them into 2 buckets.

Our system has an independent module to generate candidate mum-baby products, which are the same for both buckets. Then for users in the A bucket (i.e., the baseline group), the system ranks the candidate products based on our existing recommendation solution that predicts $P(p\_product_j = yes)$ based on various factors, such as CF predictions, consumer brand preferences, consumer shop preferences, quantity of sale, consumer purchasing power, etc. For users in the B bucket (i.e., the experimental group), the system integrates life stage prediction and ranks candidate products based on Equation 3, where $P(p\_product_j = yes)$ is estimated in the same way as the A bucket. Therefore, even if a product has high $P(p\_product_j = yes)$ (E.g,. it is more relevant to user past behaviors or is popular), it may not be recommended if it does not match the life stage of the user.

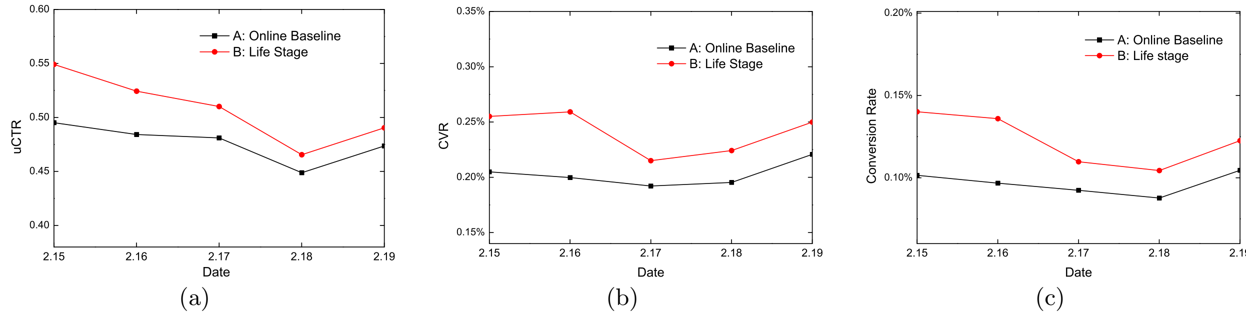

Figure 11: uCTR, CVR and Conversion Rate for two buckets.

We compared the two buckets using user Click Through Rate (uCTR), Click Conversion Rate (CVR) and Conversion Rate:

$$uCTR = \frac{\#user\_click}{\#user\_pv}$$
$$CVR = \frac{\#user\_trade}{\#user\_click}$$
$$Conversion\_Rate = uCTR \times CVR$$

where #user_click is the number of user clicking, #user_pv is the number of user browsing, #user_trade is the number of user trading in the scenario.

The results are shown in Figure 11[3]. We found that the performance on the experimental group (bucket B) is significantly better ($p<0.05$) than that on the baseline group (bucket B) in all three measures. More specifically, the improvement of uCTR means that users are more likely to slide and click. The improvement of CVR implies that after a user's clicks the proposed approach has higher probability to convert the clicking action into a purchasing action. Conversion rate combines both uCTR and CVR, and reflects the probability of a browsing action leading to a purchasing action. The improvement on it is very important for an E-commerce recommendation system and has significant business value.

6. RELATED WORK

Recommender systems are the most important and popular data mining tools that have been utilized by many E-commerce web sites to help consumers find products. There are two major recommendation approaches: content-based filtering and collaborative filtering.

The content-based approach is based on the assumption that explicit features of an item (category, words in title, price, style, size, etc.) a user liked before tell much about a user's preference. Thus the system recommends items that are similar to what the user like before. Various approaches [1, 2, 3] have been proposed, such as vector space models, language models, Bayesian classifiers, clustering models, etc. Although content based filtering has gained popularity in news recommender systems, it is success in E-commerce products recommendation is limited, because content features such as product title words, product style and brands, can hardly describe all the implicit characters of products and the fine-grained preference of users. The content-based approach is useful to increase recommendation coverage. Therefore, it is usually used as a part of solution in a hybrid recommender system [13].

The collaborative filtering approach[4, 5] assumes that users with similar tastes on some items may also have similar preferences on other items. Thus the system recommends items to a user based on other users who share similar behaves similarity preferences on other items. This approach usually operates on a user-item matrix, in which each row is a user vector and each column is an item vector. User-based methods [14, 15] find similar users to the current user. Item-based methods [16, 17] directly find items that are similar to the item that a user has clicked or purchased. Various similarity measures, such as cosine similarity, jaccard similarity, Pearson correlation coefficient, conditional probability-based similarity have been proposed to find the top-n similar items. Compared to user-based method, item-based method is more commonly used in large scale E-commerce systems because of the effectiveness and efficiency. Recently, factorization models, such as such as Singular Value Decomposition (SVD) [18], Probabilistic Matrix Factorization (PMF) [19], have gained much attention from both the research community and industry due to the their good performjance on benchmark datasets. The main benefit of this approach is introducing latent factors of user and item, which better solves the data sparsity issue.

Recently, the effect of time in recommender system has received some research attention [20, 5, 21]. One aspect studied is the drift of user preferences over time [22, 23]. Koren et al.[20] improved movie rating prediction accuracy on Netflix dataset by modeling the temporal dynamics via factorization model. Xiang et al.[22] used graph to explicitly model users' long-term and short-term preferences for top-N recommendation. Rendel et al.[23] proposed a tensor factorization modeling approach that combines Markov chain and matrix factorization for personalized next-basket recommendation. Wang et al.[24] adapted the proportional hazards modeling approach in survival analysis and explicitly incorporate time information to recommend right product at right time. However, we are not aware of any work that tries to explicitly model and predict life stages for product recommendation, especially in the context of a large scale E-commerce system.

[3]Normalized related value instead of the original absolute value of the online measures are shown for business reason. The relative improvement is consistent with real experimental results. The fluctuating performance over time is due to the effect of Chinese New Year.

7. CONCLUSION

We introduce the concept of life stage into recommendation systems research. In particular, we propose a new Maximum Entropy Semi Markov Model to segment and label consumer life stages based on observed purchasing data over time. In the mum-baby product category where the life stage transition is deterministic, we develop an efficient approximate solution using large scale logistic regression and a template based approach. We also propose a Gaussian mixture model to efficiently handle multi-kids cases. We integrate our solution into the largest online shopping website taobao.com. Both the offline and online experiments demonstrate the effectiveness of the proposed approach.

This is the first step towards life-stage based recommendations. In the future, we will study scalable solutions for the more general Maximum Entropy Semi Markov Model and apply it to other categories such as wedding and home remodeling. Besides, we plan to apply life stage labeling to other E-commerce applications such as personalized search and advertising.

8. REFERENCES

[1] Raymond J. Mooney and Loriene Roy. Content-based book recommending using learning for text categorization. In *Proceedings of the Fifth ACM Conference on Digital Libraries*, DL '00, pages 195–204, New York, NY, USA, 2000. ACM.

[2] Michael Pazzani and Daniel Billsus. Learning and revising user profiles: The identification of interesting web sites. *Machine Learning*, 27(3):313–331, 1997.

[3] Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. Content-based recommender systems: State of the art and trends. In Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, editors, *Recommender Systems Handbook*, pages 73–105. Springer US, 2011.

[4] G. Linden, B. Smith, and J. York. Amazon.com recommendations: item-to-item collaborative filtering. *Internet Computing, IEEE*, 7(1):76–80, Jan 2003.

[5] Jian Wang, Badrul Sarwar, and Neel Sundaresan. Utilizing related products for post-purchase recommendation in e-commerce. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys '11, pages 329–332, New York, NY, USA, 2011. ACM.

[6] WILLIAM D. WELLS and GEORGE GUBA. Life cycle concept in marketing research. *Journal of Marketing Research*, 1966.

[7] S. G. Barton. The life cycle and buying patterns. *Consumer Behavior*, 1955.

[8] David C. Bojanic. The impact of age and family life experiences on mexican visitor shopping expenditures. *Tourism Management*, 32(2):406 – 414, 2011.

[9] Shun-Zheng Yu. Hidden semi-markov models. *Artificial Intelligence*, 174(2):215 – 243, 2010. Special Review Issue.

[10] Andrew McCallum, Dayne Freitag, and Fernando C. N. Pereira. Maximum entropy markov models for information extraction and segmentation. In *Proceedings of the Seventeenth International Conference on Machine Learning*, ICML '00, pages 591–598, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.

[11] Charles Sutton, Andrew McCallum, and Khashayar Rohanimanesh. Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data. *J. Mach. Learn. Res.*, 8:693–723, May 2007.

[12] H. Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723, Dec 1974.

[13] Robin Burke. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4):331–370, 2002.

[14] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Analysis of recommendation algorithms for e-commerce. In *Proceedings of the 2Nd ACM Conference on Electronic Commerce*, EC '00, pages 158–167, New York, NY, USA, 2000. ACM.

[15] John S. Breese, David Heckerman, and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, UAI'98, pages 43–52, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc.

[16] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web*, WWW '01, pages 285–295, New York, NY, USA, 2001. ACM.

[17] Mukund Deshpande and George Karypis. Item-based top-n recommendation algorithms. *ACM Trans. Inf. Syst.*, 22(1):143–177, January 2004.

[18] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, Aug 2009.

[19] Ruslan Salakhutdinov and Andriy Mnih. Bayesian probabilistic matrix factorization using markov chain monte carlo. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, pages 880–887, New York, NY, USA, 2008. ACM.

[20] Yehuda Koren. Collaborative filtering with temporal dynamics. *Commun. ACM*, 53(4):89–97, April 2010.

[21] Gang Zhao, Mong Li Lee, Wynne Hsu, and Wei Chen. Increasing temporal diversity with purchase intervals. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 165–174, New York, NY, USA, 2012. ACM.

[22] Liang Xiang, Quan Yuan, Shiwan Zhao, Li Chen, Xiatian Zhang, Qing Yang, and Jimeng Sun. Temporal recommendation on graphs via long- and short-term preference fusion. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, pages 723–732, New York, NY, USA, 2010. ACM.

[23] Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. Factorizing personalized markov chains for next-basket recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 811–820, New York, NY, USA, 2010. ACM.

[24] Jian Wang and Yi Zhang. Opportunity model for e-commerce recommendation: Right product; right time. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '13, pages 303–312, New York, NY, USA, 2013. ACM.

[25] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, pages 452–461, Arlington, Virginia, United States, 2009. AUAI Press.

[26] Steffen Rendle, Zeno Gantner, Christoph Freudenthaler, and Lars Schmidt-Thieme. Fast context-aware recommendations with factorization machines. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pages 635–644, New York, NY, USA, 2011. ACM.