8-2017

# Time-aware conversion prediction

Wendi JI
*East China Normal University*

Xiaoling WANG
*East China Normal University*

Feida ZHU
*Singapore Management University*, fdzhu@smu.edu.sg



## Citation

JI, Wendi; WANG, Xiaoling; and ZHU, Feida. Time-aware conversion prediction. (2017). *Frontiers of Computer Science*. 11, (4), 702-716.
**Available at:** https://ink.library.smu.edu.sg/sis_research/3701

**Published in Frontiers of Computer Science, 2017 August, Volume 11, Issue 4, Pages 702-716**
**http://doi.org/10.1007/s11704-016-5546-y**

# Time-aware conversion prediction

**Wendi JI[1], Xiaoling WANG (✉)[1], Feida ZHU[2]**

1 Shanghai Key Laboratory of Trustworthy Computing, Institute for Data Science and Engineering, East China Normal University, Shanghai 200062, China

2 School of Information Systems, Singapore Management University, Singapore 188065, Singapore

**Abstract** The importance of product recommendation has been well recognized as a central task in business intelligence for e-commerce websites. Interestingly, what has been less aware of is the fact that different products take different time periods for conversion. The "conversion" here refers to actually a more general set of pre-defined actions, including for example purchases or registrations in recommendation and advertising systems. The mismatch between the product's actual conversion period and the application's target conversion period has been the subtle culprit compromising many existing recommendation algorithms.

The challenging question: what products should be recommended for a given time period to maximize conversion—is what has motivated us in this paper to propose a rank-based time-aware conversion prediction model (rTCP), which considers both recommendation relevance and conversion time. We adopt lifetime models in survival analysis to model the conversion time and personalize the temporal prediction by incorporating context information such as user preference. A novel mixture lifetime model is proposed to further accommodate the complexity of conversion intervals. Experimental results on two real-world data sets illustrate the high goodness of fit of our proposed model rTCP and demonstrate its effectiveness in time-aware conversion rate prediction for advertising and product recommendation.

**Keywords** conversion time, survival analysis, product recommendation, advertising

## 1 Introduction

The last decade has witnessed the generation of massive amounts of behavioral data on a daily basis, thanks to the rapid development of online shopping, social media, and location-based services. These behavioral data reveals user preferences, needs and consumption habits through their feedbacks from online recommendation systems and advertising platforms. Based on the feedback information, targeted recommendations or advertisements could be delivered to the users who are more likely to respond. The recommendation performance is measured by conversion rate. In a commercial recommendation system, conversion is the consumption or rating of a recommended product. In advertising system, conversion refers to pre-defined behaviors such as purchases, registrations and email subscriptions. A core mission of product recommendation and advertising is to optimize conversion rate prediction.

Traditional recommendation systems and advertising systems focus on finding the right product to recommend or the right user to deliver a piece of advertisement. However, in real applications it is typical that the actual conversion rate also depends on the timing of the recommendation or advertising. For instance, recommending TVs to a user who has just recently bought a new one is likely to result in negative impressions as most consumers would not purchase another big-ticket item of the same kind, such as a car or a house, within a short period of time. They might consider replacing it or buying an extra one month or year later. A right product, recommended at a wrong time, is an opportunity wasted huge. The pain is felt most acutely in product recommendation where users are disappointed by a recommendation list full of products unwanted at the time, which drives home the importance of time-awareness in the performance of conversion rate prediction.

On the other hand, the task conversion rate prediction in real life often comes together with a given time period constraint. Online shopping websites care most about products that will be bought in a particular upcoming period, such as one week or half a month. Advertising systems, especially those specializing in behavior re-targeting, are more interested in users who are most likely to convert in a near future. What is essential in all these scenarios is the ability to predict the conversion rate for a given period of time. The question we should answer is not just whether it is the right product to recommend but also whether it is the right time to recommend.

Some heuristic approaches [1–4] have been proposed to improve recommendation performance with temporal information. Previous works focus on selecting the right items in which timing is treated just as temporal characteristics to improve the prediction accuracy, e.g., holiday seasons are found to be the favorite time for shopping [5]. In this paper, we propose a data-driven model to predict the probability of a user to convert in a specified upcoming time period based on historical behavior.

We adopt the lifetime models in survival analysis, where survival time originally represents the lifetime of a patient in treatment experiments. In this work, the survival time is the conversion interval between a piece of recommendation/advertisement and the eventual conversion. In order to achieve personalized recommendation, the model is complemented by a linear regression of contextual covariates, including among others user consumption preferences and product quality. We also note that conversions tend to have different types, such as long term conversions and short term ones. For this reason, a novel mixture model is proposed to classify and identify the conversion types. Furthermore, the model has been extended with a hierarchical Bayesian framework for regularization. The proposed model is called the context-aware Weibull mixture model (cWMM for short). As there is no closed-form solution for parameter learning, the generalized expectation-maximization (EM) algorithm has been used, with a gradient strategy in the maximization step.

However, the conversion interval does not exist if the conversion never occurs, which is different from survival analysis where patients will die eventually. This is the reason why we need two models: one to predict the conversion rate and the other to predict the conversion interval. The probability that a user converts within a particular time period is given by a rank-based time-aware conversion prediction model (rTCP), which is proposed to integrate the two submodels.

Finally, the proposed model is evaluated on two real-world datasets: the conversion data from a display advertising platform (Criteo), and the online consumption data from an e-commerce website (T-mall). The experimental results show that the proposed approach effectively models conversion interval and improves performance in temporal advertisement conversion prediction and product recommendation.

The major contributions of this paper are summarized as follows:

1) A new research problem has been proposed to identify what products should be recommended for a given upcoming time period, such as three days or one week.
2) A lifetime model based on Weibull distribution has been proposed to model the conversion intervals in the consideration of personalization. A more comprehensive model has been proposed using a Bayesian framework to regularize and utilize the generalized EM for parameter estimation.
3) A rank-based time-aware conversion prediction model (rTCP) with cWMM has been proposed to estimate behaviors over a specific prediction period.
4) The proposed cWMM has been evaluated using two real-world datasets. Experimental results confirmed the goodness of fit of cWMM. Furthermore, the effectiveness of rTCP in temporal advertising and recommendation systems has been demonstrated, especially for short term prediction.

The remainder of the paper is organized as follows. Section 2 discusses the related work. Section 3 gives the preliminaries and the problem definition. Section 4 introduces a temporal behavior prediction model. Section 5 provides details of some lifetime models, including a proposed context-aware Weibull mixture model i.e., cWMM. The experimental evaluations are detailed in Section 6, and the paper is concluded in Section 7.

## 2 Related work

We summarize research work in product recommendation and computational advertising which are most related to conversion rate prediction.

The central task of commercial recommendation systems is to recommend products catering to users' consumption interest. Some context-based methods have been proposed in Refs. [6–9] which make recommendations by analyzing textual information and finding regularities in the content. Another popular technique, collaborative filtering, assumes that users who rate items similarly have similar consumption behaviors [10]. To alleviate the data sparsity problem, which is a crucial challenge in collaborative filtering, many dimensionality reduction approaches have been proposed, including matrix factorization methods [12, 13] and latent semantic models [14, 15]. Recent works also take the sequential behaviors of users into account. The research in Ref. [16] focused on the sequential order of purchases, and combined Markov chains and matrix factorization to predict the next action. A matrix factorization method combining a personalized Markov chain and region localization has been proposed in Ref. [17] to recommend successive POI.

Many recent research papers have explored the problem of conversion rate prediction in the context of computational advertising. An analysis work [18] posted a detailed analysis of conversion rates in the setting of non-guaranteed delivery targeted advertising and illustrated that the click-to-conversion delay is a challenge to conversion prediction. Logistic regression is the most popular model used in real-life advertising platforms because it can be parallelized easily to handle large-scale problems [19]. Lee et al. [20] combined some individual estimators, such as user and publisher, and used logistic regression to tackle extreme data sparsity of conversions.

However, users' purchase decisions tend to change with time. Lathia et al. [21] showed that temporal diversity was an important factor in recommendation systems, which had been used as a new measure in many subsequent research of temporal recommendation. Xiang et al. [22] proposed a method that the score of a given product was determined by both long-term preference (the relevance between a produce and a user) and short-term bias due to special events (such as seasonal sales). Koren [23] modeled the temporal dynamics along the whole time period to separate transient factors from lasting ones. A purchase interval cube was proposed in [24, 25] to measure the temporal similarity of users. A fourier-assisted auto-regressive integrated moving average (FARIMA) process was proposed in Ref. [26] to tackle with the year-long seasonal period of purchasing data to achieve daily-aware preference predictions. An intelligence recommender system was proposed in Ref. [27] to detect users' purchase intents from their microblogs in near real-time and make product recommendation based on matching the users' demographic information. In all these works, temporal factors are used as temporal characteristics and cannot provide explicit recommendation for a specified upcoming period.

In this work, survival analysis is used to have a fine-grained modeling of conversion intervals. The term "survival analysis" refers to the study and modeling of observed product lifetime with various fields of application, including actuarial science, economics, engineering, and social and behavioral sciences [28]. It focuses on the occurrence time of a particular operation, such as death in biological survival studies and failure in mechanical reliability. In this study, lifetime is the interval of the transition from the preceding behavior to the eventual conversion. Some research have used lifetime models in recommendation and advertising systems. Exponential model was used in Ref. [29] to capture the conversion delay in display advertising. Wang et al. proposed an opportunity method to model the interval of purchases [30] and career switches [31], but used the joint probability of the time probability and occurrence probability to predict the conversion rate of a given time, which was different from the rank-based method of this work. We find that the joint probability is not suitable for the motivation of this work, which will be detailed in Section 4.

## 3 Preliminaries and problem definition

In this section, we first introduce some notations about time-aware conversion prediction, followed by the problem formulation.

### 3.1 Preliminaries

In this paper, we aim to find how to recommend the right products or advertisements for a specified upcoming period based on the behavior sequences of users. Figure 1 shows an ad browsing sequence and an online shopping sequence. A behavior is defined as a quaternion tuple {action, object, timestamp, contextual variables}, which means that a user takes an action with respect to an object at some timestamp with corresponding contextual variables. Except for the timestamp, the other three tuples can vary in different scenarios. For instance, in online shopping, the user is a consumer, the object is a product (identified by name, brand, or category), the activity is click or purchase, and contextual variables includes user profile variables (user age, gender, income, etc.), product profile variables (product category, price, discount, etc.), and time-dependent variables (recent user clicks and purchases, etc.).

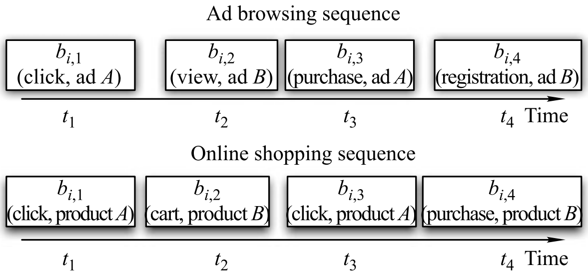



**Fig. 1** Behavioral sequences on an advertising system and an online shopping website. Each sequence consists of a chronological sequence of actions by user $i$ and user $j$, where $i, j \in U$

The following notations are used in this paper:

- $U$: a set of users, each of which has a behavior sequence.
- $X'$: a set of features determining whether a conversion will be performed.
- $C \in \{0, 1\}$: indicating whether a conversion will be performed.
- $\Delta t$: an upcoming period specified by applications.
- $T$: the current time.
- $Cat$: a set of product/ad categories. In product recommendation, it can be the nature category of the product, such as books & audible and home, garden & tools or the category learned by clustering models, like laten Dirichlet allocation (LDA) and $k$-means. In advertising, the category are the brand of advertisements or the advertising campaign of an advertiser.
- $M$: a set of transitions, each of which refers to the type of transition from the current behavior to the conversion. A transition is related to the category of the object and the action, but is not specific to a particular user. In Fig. 2, the transition $m \in M$ is $\{$(action $A$, object $B$) $\rightarrow$ (action $C$, object $D$)$\}$. If object $B$ is in category $p$ and object $D$ is in category $q$, the transition is formed by category $\{p \rightarrow q\}$. In the ad browsing sequence shown in Fig. 1, the transition from behavior $b_{i,2}$ to behavior $b_{i,4}$ is $\{$(view, ad $B$) $\rightarrow$ (registration, ad $B$)$\}$, whereas in the online shopping sequence, it is $\{$(cart, product $A$) $\rightarrow$ (purchase, product $B$)$\}$. Only conversion and purchase are defined as the latter action because in product recommendation, advertising platforms and e-commerce web sites pay more attention to final consequences.
- $N_m$: a set of observations, each of which is an observed transition between two behaviors in a behavior sequence of transition $m$, where $m \in M$. The former behavior of an observation is the triggering behavior and the latter one the pre-defined conversion. Each observation consists of two parts: the time interval $y_{m,n}$ and the covariates $x_{m,n}$, as shown in Fig. 2.
- $Y$: the conversion interval if $C = 1$. As shown in Fig. 2, $y_{m,n} \in Y$ is the time difference of the observation $n$ in transition $m$: $y_{m,n} = t_p - t_q$, where $n \in N_m$, $m \in M$ and observation $n$ consists of behaviors $b_{i,p}$ and $b_{i,q}$.
- $X$: a set of features determining when a behavior will be performed, of which each element $x_{m,n}$ is a vector that presents the contextual information of an observation pair $n$ in transition $m$.

Figure 2 shows an example of the relationship among transition $m \in M$, observation $n \in N_m$, time interval $y_{m,n}$, and covariates $x_{m,n}$. It is assumed here that observations in the same transition share similar properties, and therefore that observations in the same transition should be considered as having the same scope, like a document in the LDA model [15]. Furthermore, covariates should contain the respective properties of each behavior, the characteristics of the corresponding transition, and the preferences of the user. For example, the covariates $x_{m,n}$ incorporate the inherent behavioral variables $\vec{v}_{i,p}$ and $\vec{v}_{i,q}$ (such as the ratings of objects $o_{i,q}$ and $o_{i,p}$), the features extracted from transition $m$ (such as the distance between two locations and the similarity of two products), and the preferences of user $u_i$ (such as gender, age, and favorite categories).

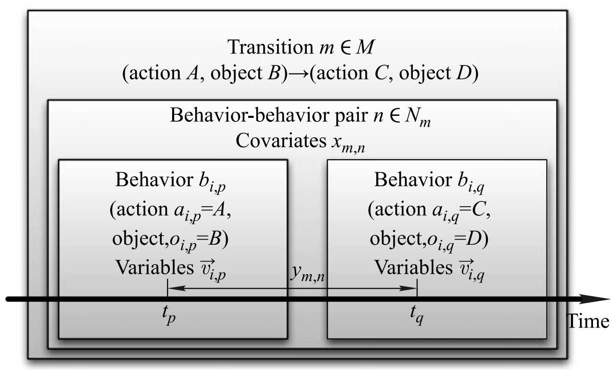



**Fig. 2** Illustration of the relationship among behaviors, transitions, and observations with corresponding components

## 3.2 Problem definition

To recommend the right products or advertisements for a specified upcoming period $\Delta t$, we should first find some candidate products with high conversion rate for each user, then exam which candidate has higher probability to convert. Regardless of time, deciding what products or advertisements should be recommended needs to predict the conversion rate $Pr(C = 1|X')$. However, the conversion rate only gives the probability of the occurrence in the future, but not the guidance of the conversion time. This problem results in the waste of recommending opportunity and poor user satisfaction when the conversion has low possibility to occur within $\Delta t$. We calculate the percentage of ad conversions (all ads are of the same ad campaign) within different time intervals. As shown in Fig. 3, about 70% of ad conversions occur within ten hours of the clicks, about 20% of the conversions occur in the interval between12 hours and seven days and the rest 10% happen much latter.

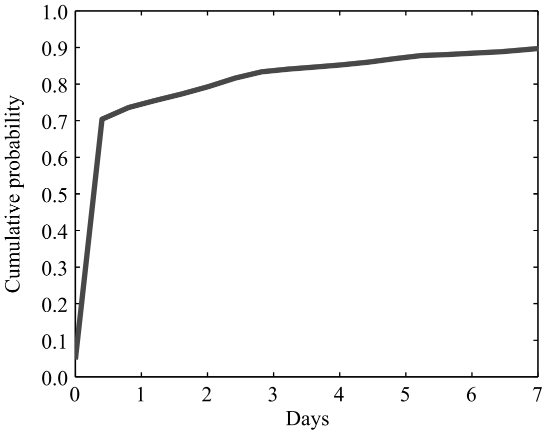



**Fig. 3** Cumulative distribution of the interval between ad click and ad conversion. All intervals are of the same ad campaign

The probability of conversion occurrence within an upcoming period $\Delta t$ is determined by the specified period $\Delta t$ and the distribution of the conversion interval $p(Y)$. Suppose there are two candidate ads with the same conversion rate. Some advertiser pursuits short-term return; for instance, they want more conversions within one day or three days. We should deliver the ad with higher probability to convert within short-term.

If we have a proper lifetime distribution, the first task is how to use the lifetime model to improve the performance of time-aware conversion prediction. It is difficult to combine these two predictions: what product or advertisement should be recommended, and when they should be recommended. The problem will be discussed in Section 4. However, traditional statistical lifetime distributions are inadequate to model the intervals between behaviors. Indeed, building a framework for analyzing the time factor of a user behavior sequence effectively remains an open problem. So the second task is find a proper lifetime distribution to model the personality and complexity of interval distributions, which is more difficult. Section 5 will introduce the proposed lifetime models.

## 4 Time-aware conversion prediction

This section discusses how to use a proper lifetime distribution to improve the performance of time-aware conversion prediction. The lower-case letters are the observed values of the random variables in Section 3.1. In time-aware conversion prediction, the data set is made up of tetrads $(x'_{u,q}, c_{u,q}, t_{u,p}, x_{u,p,q})$ of user $u$. The triggering behavior is $b_p$, the conversion is $b_q$, $c_{u,q}$ indicates whether the conversion happens, $t_{u,p}$ is the timestamp of $b_p$, $x'_{u,q}$ determines whether conversion will happen, and $x_{i,p,q}$ influences when the conversion will happen. If $c_{u,q} = 1$, the interval $y_{u,p,q}$ is also given.

Given the behavioral sequence of a user, two sub-models are needed to predict whether he or she will perform a certain conversion within a specific period $\Delta t$ (e.g., three days or half a month): a model of the conversion probability $Pr(C|X')$, and a model of the conversion time $Pr(Y - T < \Delta t|X, C = 1, Y, T, \Delta t)$. To estimate $Pr(C = 1|X')$, i.e., the probability that a user will convert or the probability of purchasing a product, we can use the existing classification models in recommendation systems. In this study, the most widely used model, the logistic regression model, has been applied to conversion prediction (CP) as the baseline model:

$$\Pr(C = 1|X' = x') = \frac{1}{1 + \exp(-w^{\mathrm{T}}x'))}, \tag{1}$$

where w is a vector of weights that can be learned by the maximum likelihood estimation.

If a conversion occurs ($C = 1$), the time interval $Y$ will be modeled by a lifetime distribution $p(y)$. When predicting the occurrence time of a conversion, the model should focus on a time period, not a single point in time. For example, in advertising platforms, advertisers pay more attention to the probability that a user will convert within one day or one week; in product recommendation systems, marketers would prefer to know the products that a user is most likely to buy in the next week or the next month. Therefore, the time prediction model (TP) is designed to provide the probability that a conversion will be performed within $\Delta t$:

$$\Pr(Y - T < \Delta t|X, C = 1, Y, T, \Delta t) = \int_0^{\Delta t} p(y)\mathrm{d}y. \tag{2}$$

The cumulative function of cWMM and $p(y)$ is the probability density function of a lifetime model. This represents the probability that a conversion happens within the period of $[T, T + \Delta t]$.

Under these models, a straightforward approach is using the joint probability of CP (Eq. (1)) and TP (Eq. (2)) to calculate the probability that a user converts within $\Delta t$. This is called the naive TCP model, which is the opportunity model in Refs. [30, 31]. However, in naive TCP, the value of TP suppresses that of CP. If $C = 1$, TP provides the probability that a user will perform the conversion within $[T, T + \Delta t]$. Regardless of whether a behavior will be performed, $Pr(Y - T < \Delta t | X, C = 1, Y, T, \Delta t)$ is close to one when $\Delta t$ is large enough. In this situation, a negative conversion will be misestimated as positive when $Pr(C = 1|X')$ is small and $Pr(Y - T < \Delta t | X, C = 1, Y, T, \Delta t)$ is large.

Therefore, a rank-based time-aware conversion prediction method, named rTCP, is proposed here to integrate CP and TP, which is shown in Algorithm 1. We use a threshold $\delta$ to determine the relationship between CP and TP. When CP is high, TP dominates the output of rTCP, and when CP is low, TP dominates. That means TP has stronger effect on the behaviors when the CP predicted value is lower than that when the CP predicted value is higher. Therefore, in rTCP, threshold $\delta$ changes for different predicted values of CP: if the CP predicted value is high (e.g., $[0.8, 1]$), $\delta$ is small; whereas if the CP predicted value is low (e.g., $[0, 0.4]$), $\delta$ is large. On the other hand, rTCP also reflects that if the cumulative probability is too small, it is too early to say that a behavior will not be performed. Here, the heuristic ranking strategy given in Algorithm 1 is proposed, but there are other approaches to set $\delta$, which can be explored in the future.

**Algorithm 1** Ranking strategy in rTCP

```
Input: A, B
Output: RANK of A and B
  δ = 1 − max(CP(A), CP(B));
  If |CP(A) − CP(B) ⩽ δ|
  return rank(TP(A),TP(B)); //ordering A and B with TP.
  else
  return rank(CP(A),CP(B)); //ordering A and B with CP.
```

## 5 Mixture temporal model

This section presents a novel lifetime model for exploring the time factor in sequential behaviors. First, the characteristics of conversion intervals and the Weibull distribution are introduced. Then an extension of the basic distribution to capture personalization and complexity is illustrated, and the parameter estimation using the EM algorithm is detailed.

### 5.1 Conversion intervals

Figures 4(a) and 4(b) show the distributions of the intervals and log intervals within the same ad campaign over 14 days. The oscillating shape is due to daily cyclic trends, where the peak hour can be considered as the preferred hour of a user. Figures 4(c) and 4(d) show the purchase intervals from an e-commerce website. As shown in Fig. 4, the distributions of the intervals in are much more complex than a basic lifetime distribution such as the exponential or Weibull distribution. A related study [29] assumed that the ad conversion intervals followed an exponential distribution. However, an exponential distribution gives a straight-line regression for log intervals, which fails to match the distribution shown in Fig. 4(b). The Weibull distribution with a shape parameter less than one may provide a better fit for short intervals. However, even the Weibull distribution tends either to under-predict short intervals (< 2 days) or to over-predict long intervals (> 2 days), which means that the interval distribution appears to be a mixture of short and long intervals. Furthermore, unlike the common lifetime distributions, the shapes of the purchase interval distributions fluctuate widely. This phenomenon indicates that the distributions tend to change over time, which reflects the complexity and personality of behavior patterns and user preferences. It also implies that a mixture lifetime model is required to model the conversion intervals.

### 5.2 Weibull distribution

The Weibull distribution is one of the most widely used lifetime distributions in survival analysis. Here, the lifetime is conversion interval $Y$. In statistics, the Weibull distribution, which is characterized by a shape parameter $\alpha$ and a scale parameter $\theta$, is a versatile distribution that can take on the characteristics of various other distributions. The exponential distribution is an exceptional type of Weibull distribution in which $\alpha = 1$.

Let $p(y_{m,n})$ be the probability density function of the Weibull distribution:

$$p(y_{m,n}) = \frac{\alpha}{\theta}\left(\frac{y_{m,n}}{\theta}\right)^{\alpha-1} \exp\left[-\left(\frac{y_{m,n}}{\theta}\right)^{\alpha}\right], \tag{3}$$

where $y_{m,n}$ is the time interval between behavior $b_i$ and behavior $b_j$ of observation $n$ in transition $m$, and $p(y_{m,n})$ is the probability density function of the continuous variable $y_{m,n}$.

The transition modes between behaviors are denoted by $\alpha$ and $\theta$. As shown in Fig. 5, the shape parameter $\alpha$ is concerned with the tail behavior, and the scale parameter $\theta$ is related to the distribution peak.

- A value of $\alpha < 1$ indicates that the event occurrence rate decreases over time. The shape parameter $\alpha$ for most intervals between ad impressions (or clicks) and conversions is less than one, which indicates that the influence of an ad fades quickly with time, as can be observed in Figs. 4(a) and 4(b).
- A value of $\alpha = 1$ indicates that the event occurrence rate does not change with time. This indicates that the willingness of users to go to some location or buy some product is independent of time.
- A value of $\alpha > 1$ indicates that the event occurrence rate increases over time. This happens in some re-consumption cases in which the purchase willingness of a product can at first be low when the user has just bought it. Unlike distributions with $\alpha \leqslant 1$, $p(y_{m,n} = 0)$ is not the distribution peak when $\alpha > 1$.
- The scale parameter $\theta$ reflects the values around which the $y$s mainly cluster (the distribution peak). When $\alpha = 1$ or $\alpha \gg 1$, the mean value, i.e., $\bar{y}_{m,n} = \theta\Gamma\left(\dfrac{\alpha+1}{\alpha}\right)$, equals to $\theta$.

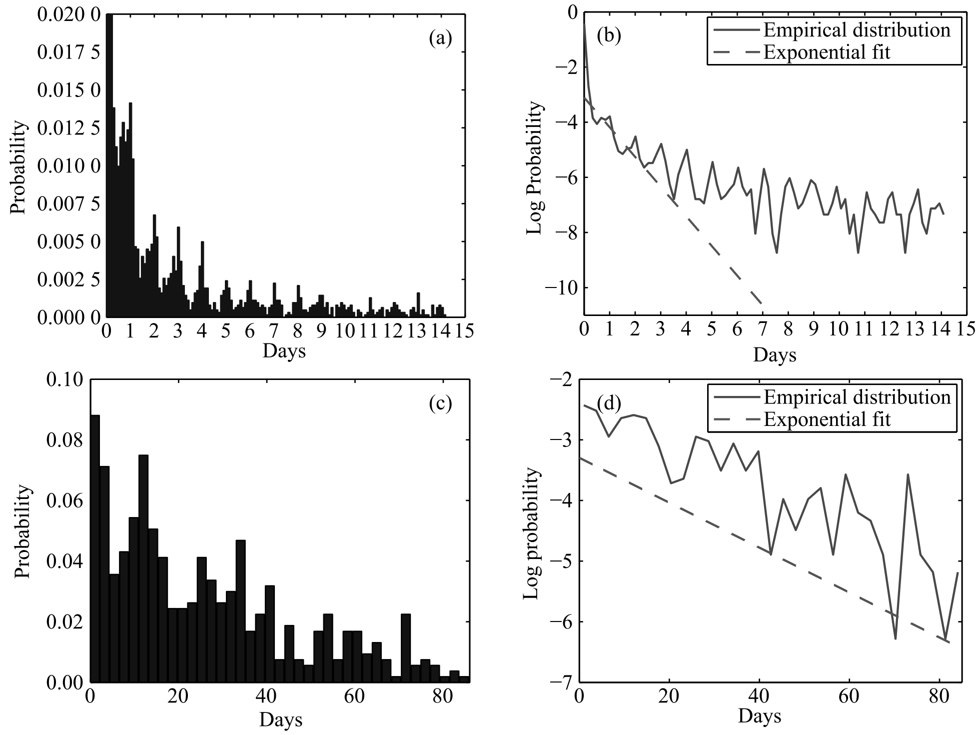



**Fig. 4** Probability density function (pdf) of intervals and log intervals between behaviors. (a) pdf of intervals between clicks and conversions (mean: 0.9 day); (b) log pdf of intervals between clicks and conversions (mean: 0.9 day); (c) pdf of intervals between purchases and purchases (mean: 27.0 days); (d) log pdf of intervals between purchases and purchases (mean: 27.0 days)

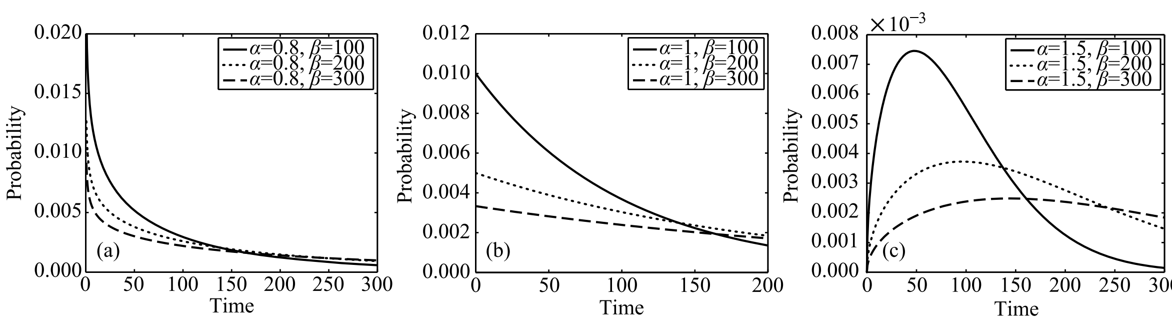



**Fig. 5** Probability density functions for various values of the shape parameter $\alpha$s and the scale parameter $\theta$s (time is in hours). (a) $\alpha < 1$; (b) $\alpha = 1$; (c) $\alpha > 1$

The Weibull distribution has been proved to perform better as a lifetime distribution than other continuous distributions [32]. However, both exponential distribution and Weibull distribution are adequate to model conversion intervals, since each of them has both advantage and disadvantage. Weibull distribution fits the interval data better than exponential distribution with adaptable shape parameter $\alpha$, while exponential distribution is much simpler than Weibull distribution. Therefore, applications have the right to analyze the trade-off and choose which distribution is more suitable for individual requirements. If exponential distribution is chosen, we can simply set $\alpha$ to 1.

Considering personalization, the interval distribution is also strongly related to contextual information. Therefore, the scale parameter $\theta$ is re-characterized using the covariate vector $x_{m,n}$ and the weights $\beta$. The parameter $\theta$, which is regarded as the pseudo-mean of the time interval $y_{m,n}$, can be represented by the exponential linear combination of $x_{m,n}$ and $\beta$ as $\exp\left[\beta^{\mathrm{T}} x_{m,n}\right]$. Then the probability density function is changed to:

$$p(y_{m,n}) = \frac{\alpha}{\exp\left[\beta^{\mathrm{T}} x_{m,n}\right]} \left(\frac{y_{m,n}}{\exp\left[\beta^{\mathrm{T}} x_{m,n}\right]}\right)^{\alpha-1} \exp\left[-\left(\frac{y_{m,n}}{\exp\left[\beta^{\mathrm{T}} x_{m,n}\right]}\right)^{\alpha}\right]. \tag{4}$$

### 5.3 Context-aware Weibull mixture model

As discussed in Section 5.1, it is difficult to model intervals using a single Weibull distribution. In this section, we propose a novel context-aware Weibull mixture model (cWMM) which incorporates covariates to introduce contextual information and leverages a mixture of models to cover the variance of behavioral patterns and motivations. This model is only an optional extension to fit the conversion intervals better, but it is not necessary when applications attach more importance to efficiency than accuracy.

• **Model description** The proposed cWMM is a probabilistic mixture generative model of time interval $y_{m,n}$ in a certain transition $m$ with contextual covariates $x_{m,n}$, which can be represented by the graphical model in Fig. 6. The greatest strength of the cWMM is that it can take into account the complexity of multiple behavioral transition patterns in different time periods. The model considers the inherent differences among transitions, and hence the observations in each transition $m$ share the same parameters.

Because mixture models without priors are sensitive to singularities, the mixture Weibull model was extended to the Bayesian version. Parameters in different transitions are independent, but share the same global priors. To represent the mixture of Weibull models, a hidden variable $z_{m,n}$ was introduced to indicate which basic distribution generated observation $y_{m,n}$. Given the Weibull parameter $\phi_m$ and the mixture parameter $\omega_m$, the joint distribution of the time interval $y_{m,n}$ and its mixture variable $z_{m,n}$ is:

$$\begin{aligned} & p\left(y_{m,n}, x_{m,n}, z_{m,n} \mid \phi_m, \omega_m\right) \\ & = \sum_{k}^{K} p\left(z_{m,n} = k \mid \omega_m\right) p\left(y_{m,n}, x_{m,n} \mid \phi_{m,k}\right) \\ & = \sum_{k}^{K} \omega_{m,k} \mathrm{Weibull}\left(y_{m,n}; \alpha_{m,k}, \exp\left[\beta_{m,k}{}^{\mathrm{T}}, x_{m,n}\right]\right), \end{aligned} \tag{5}$$

where $\omega_m$ is a $K$-dimensional vector containing the weight of each Weibull distribution. Algorithm 2 illustrates the process of generating the cWMM.

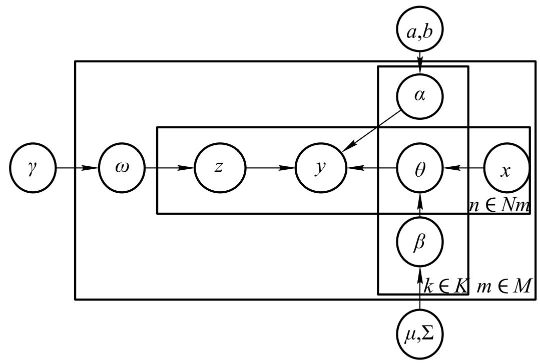



**Fig. 6** Illustration of cWMM as a generated probability graph, where observation $n$ is of transition $m$, $M$ is the number of transitions and $N_m$ the number of observations of transition $m$. $y_{m,n}$ is the time interval of observation $n$ and $x_{m,n}$ is the corresponding set of covariates. $z_{m,n}$ is the mixture indicator of the $K$ Weibull distributions, $\omega$ is the parameter of $z_{m,n}$ with prior $\gamma$, and $\alpha_{m,k}$ and $\beta_{m,k}$ are the parameters of the $k$th Weibull distribution with prior $a, b, \mu, \omega$

**Algorithm 2** Probabilistic generative process in cWMM

```
for each transition m, m ∈ M do
    Draw ω_m ~ Dirichlet(ω_m; γ/K, ..., γ/K);
    for each Weibull distribution k, k ∈ K do
        Draw α_{m,k} ~ Gamma(α_{m,k}; a, b);
        Draw β_{m,k} ~ Gaussian(β_{m,k}; μ, Σ);
    end for
    for each observation n, n ∈ N_m do
        Draw z_{m,n} ~ Discrete(z_{m,n}; ω_m);
        for each single Weibull distribution k, k ∈ K do
            let θ_{m,n,k} = exp[β_{m,k}^T x_{m,n}];
        end for
        Draw y_{m,n}|z_{m,n} = k ~ Weibull(y_{m,n}; α_{m,k}, θ_{m,n,k});
    end for
end for
```

• **Model training** This model is an extension of the mixture model, in which each basic distribution is a Weibull distribution. The EM algorithm has been used to perform parameter estimation. Because the scale parameter of each Weibull distribution is assigned to the linear regression of the contextual covariates, there is no closed-form solution in step M. The generalized EM method is used to optimize the Weibull parameters $\{\alpha, \beta\}$ using a gradient-based method. As for the global priors and hyper-parameters, they are treated as fixed values (i.e., $a = 2$, $b = 2$, $\mu = \{0.1, \ldots, 0.1\}, \sigma = 0.5\mathrm{I}, \gamma = \{0.1, \ldots, 0.1\}$, $k = 3$). For the current iteration $t$:

In the expectation step, $\alpha, \beta, \omega$, and the sample $z_{m,n}$ are fixed for all the observations as follows:

$$\begin{aligned} {r_{m,n,k}}^t &= p\left(z_{m,n} = k \mid y_{m,n}, x_{m,n}, {\phi_{m,n}}^{t-1}, {\omega_m}^{t-1}\right) \\ &= \frac{{\omega_{m,k}}^{t-1} p\left(y_{m,n}|x_{m,n}, {\phi_{m,n,k}}^{t-1}\right)}{\sum_j^K {\omega_{m,j}}^{t-1} p\left(y_{m,n}|x_{m,n}, {\phi_{m,n,j}}^{t-1}\right)}, \end{aligned} \tag{6}$$

where $r_{m,n,k}$ is the probability that this observation is generated by the $k$th Weibull distribution.

In the maximization step, the first task is to update the parameters $\{\alpha_{m,k}, \beta_{m,k}\}$ for each Weibull distribution in transition $m$. Because there is no closed-form solution to optimize the log likelihood of the complete data, a gradient-based method has been leveraged to perform parameter estimation with maximum a posteriori probability (also known as MAP).

The log likelihood of the observed data of transition $m$ is:

$$\begin{aligned} L(\phi) \propto & \sum_k^K \sum_{n^{(k)}}^{{N_m}^{(k)}} r_{m,n,k} \ \log[p\left({y_{m,n}}^{(k)}|\alpha_{m,k}, \exp\left[{\beta_{m,k}}^{\mathrm{T}} {x_{m,n}}^{(k)}\right]\right) \\ & p\left(\alpha_{m,k}|a, b\right) p\left(\beta_{m,k}|\mu, \Sigma\right)]. \end{aligned} \tag{7}$$

The first-order derivatives of the shape parameter $\alpha_{m,k}$ and the covariate weights $\beta_{m,k}$ can be directly calculated from the log likelihood as follows:

$$\begin{aligned} \frac{\partial L(\phi)}{\partial \alpha_{m,k}} = & \sum_{n^{(k)}}^{{N_m}^{(k)}} r_{m,n,k}\Big(\frac{a}{\alpha_{m,k}} + \left[1 - \left(\frac{{y_{m,n}}^{(k)}}{\exp\left[{\beta_{m,k}}^{\mathrm{T}} {x_{m,n}}^{(k)}\right]}\right)^{\alpha_{m,k}}\right] \\ & \left(\ln {y_{m,n}}^{(k)} - \left[{\beta_{m,k}}^{\mathrm{T}} {x_{m,n}}^{(k)}\right]\right) - \frac{1}{b}\Big), \end{aligned} \tag{8}$$

$$\begin{aligned} \frac{\partial L(\phi)}{\partial \beta_{m,k}} = & \sum_{n^{(k)}}^{{N_m}^{(k)}} r_{m,n,k}\Big(-\alpha_{m,k} \left[1 - \left(\frac{{y_{m,n}}^{(k)}}{\exp\left[{\beta_{m,k}}^{\mathrm{T}} {x_{m,n}}^{(k)}\right]}\right)^{\alpha_{m,k}}\right] \\ & \left({x_{m,n}}^{(k)}\right)^{\mathrm{T}} - \frac{1}{2}\left[\Sigma^{-1} \left(\beta_{m,k} - \mu\right)\right]^{\mathrm{T}}\Big). \end{aligned} \tag{9}$$

In the update procedure, given the current hidden variables $z_{m,n}$, $\alpha_{m,k}$ and $\beta_{m,k}$ are estimated using an optimization strategy based on the L-BFGS optimizer in the current implementation. Then ${\theta_{m,k}}^t = ({\alpha_{m,k}}^t, {\beta_{m,k}}^t)$ are set using the new estimates for $k = 1 : K$.

Finally, the mixture weights $\omega_m$ are updated using the tractable posterior probabilities of the pseudo-counts, denoted as $r_{m,k} = \sum_n^{N_m} r_{m,n,k}$. The mixture parameters are updated by the expectation of the Dirichlet posterior:

$$p\left(\omega_m|y;\gamma\right) = Dirichlet\left(r_{m,1} + \frac{\gamma}{K}, \ldots, r_{m,K} + \frac{\gamma}{K}\right), \tag{10}$$

and $\omega_m$ is updated using:

$${w_{m,k}}^t = \frac{{r_{m,k}}^t + \dfrac{\gamma}{K}}{\sum_j^K \left({r_{m,j}}^t + \dfrac{\gamma}{K}\right)}. \tag{11}$$

## 6 Experimental evaluation

### 6.1 Datasets of sequential behaviors

To analyze the distribution of conversion intervals, two real-world datasets are used.

• **Criteo** The conversion behavior dataset is obtained from Criteo[1)], a display advertising platform specializing in retargeting [29]. In various campaigns, display ads are posted in different forms (e.g., iFocus ads and floating ads) through various channels (e.g., portal web sites and social web sites). An ad campaign aims to deliver a message about a specific brand to the target audiences. This dataset contains the click and conversion records for one month, including 15 998 883 click-conversion observations (25% being positive samples and 75% negative samples)[2)] in 13 073 ad campaigns. Behavioral timestamps are in seconds.

This scenario focuses on the transitions between clicks and conversions. Post-click attribution means that a conversion assigns the whole credit to a particular click, which means in turn that the conversion behavior is caused by the click. This is a commonly accepted attribution model in the online advertising industry. The time intervals between clicks and conversions are valuable in temporal conversion-rate prediction and in campaign performance analysis. In statistical terms, approximately 46% of conversions occur within one day of the click and 13% after two weeks. The statistics of the whole intervals is quite different from interval distribution of a champaign, which is shown in Fig. 3. This fact indicates that interval distributions of different transitions should be modeled separately. However, the number of observations for each campaign tends to follow the power law distribution, which means that the observations in most campaigns are too rare to build a model. Therefore, 1 603 025 click-conversion

1) http://labs.criteo.com/downloads/2014-conversion-logs-dataset/

2) In the real case, the conversion rate $Pr(conversion|click)$ can be as low as 1%, and the dataset is sampled as one positive: three negatives for model training

pairs from the largest 100 campaigns are used in the experiments.

• **Tmall** The online consumption behavior dataset is obtained from the competition for the Tmall Recommendation Prize 2014[3)] and is provided by Ali. The Tmall dataset contains a user history log of browsing and consumption for four months, including clicking, adding an item to the wish list, adding an item to the shopping cart, and purchasing. Behavioral timestamps are in days. After filtering out records where the user made no purchase or where the brand was clicked fewer than 50 times or purchased fewer than ten times, the dataset contains 182 879 records (including 9 684 conversions), 2 149 brands, and 748 users.

Online stores always want to know what products a user will purchase in the next week or next month. However, a crucial challenge of product recommendation is that users tend to buy "new products". User behavior log reveals that most products (as much as 75%) which are bought in the last month are never clicked or bought in the former three months. Therefore, this research considers two kinds of intervals, including click-to-purchase transitions (the term "click" here includes actions except purchases) and purchase-to-purchase transitions. Purchase to purchase transitions are important in this scenario because the products users have purchased have strong intention of user preference and can be used to predict the potential upcoming products.

The products users have purchased also reveal the latent relationship between products. Therefore, the brands are clustered into ten categories using LDA [15] based on the consuming records of users. The conversion intervals between two brands are extremely sparse. Under this condition, the interval distributions between two brands tend to be stochastic and can not be modeled by any well-known distribution. Even if the dataset contains more users and brands, it would still be impossible to model intervals between brands because most users usually purchase at most dozens of products per month, and thus the conversion intervals between brands are still sparse.

## 6.2 Effectiveness of lifetime models

Next, the goodness of fit of the proposed lifetime models will be demonstrated. First, the experimental setup, including data preparation and the diverse approaches used, will be described. Then the main experimental results will be demonstrated on each dataset. Finally, the question about how to analyze different transition modes will be explored.

[3)] http://102.alibaba.com/competition/addDiscovery/index.htm

### 6.2.1 Data preparation

The experimental setup is constructed as follows. For the Criteo dataset, 100 campaigns are used in the experiments. Each campaign is treated as a transition, and the click-conversion pairs in each campaign are divided into four groups. The average number of observations in each transition is 13 356. The contextual covariates include eight counting features and nine categorical features, which are directly provided by the publisher and anonymized. As for the Tmall dataset, all users are divided into four groups, and click-purchase and purchase-purchase transitions are examined among categories (all brands are clustered into ten categories) as separate transitions. There are on average 1 782 click-purchase observations and 613 purchase-purchase observations in each group. The covariates include various features, including how many times the category transition has appeared before and how many times the product has been bought before. For both datasets, fourfold cross validation is applied to the cWMM and to all competing models.

### 6.2.2 Contrasting approaches

To evaluate the rationality and necessity of the proposed cWMM, the following three components are examined: 1) the effect of the Weibull distribution; 2) the effect of the contextual covariates; and 3) the effect of the mixture extension. The following contrasting models are evaluated:

- EMM: exponential mixture model, $k = 1, 2, 3, \ldots$ (when $k = 1$, this is a basic exponential model (EM).)
- WMM: Weibull mixture model, $k = 1, 2, 3, \ldots$ (when $k = 1$, this is a basic Weibull model (WM). )
- cEMM: context-aware exponential mixture model, $k = 1, 2, 3, \ldots$ (when $k = 1$, this is a context-aware exponential model (cEM).)
- cWMM: context-aware Weibull mixture model, $k = 1, 2, 3, \ldots$ (when $k = 1$, this is a context-aware Weibull model (cWM). )

The log likelihood and perplexity have been used to evaluate how well the model fits the test data. The log likelihood formula has been defined as:

$$\log p(y) = \sum_{m}^{M} \log p\left(y_m\right). \tag{12}$$

The perplexity is used by convention in language models [11] and topic models [15]. It decreases monotonically as the likelihood of the test data increases. For a test set of M transitions, the perplexity can be defined according to [33]:

$$\text{perplexity}(y) = \exp\left[-\frac{\sum_m^M \log p(y_m)}{\sum_m^M N_m}\right]. \qquad (13)$$

In general, a better fit means a larger log likelihood and less perplexity. For both log likelihood and perplexity, the likelihood of a transition in the purposed cWMM is:

$$p(y_m) = \prod_n^{N_m} \sum_k^K \omega_k p\left(y_{m,n}, x_{m,n} \mid \phi_{m,k}\right). \qquad (14)$$

### 6.2.3 Experimental results

The experimental results obtained are shown in Fig. 7. In general, the proposed cWMM has the best performance among all the competing models. First, it can be seen that the Weibull distribution always provides a better fit than the exponential distribution, regardless of whether it is used with contextual covariates or mixture extensions. Second, contextual information helps both the Weibull and exponential distributions fit real-world behavioral data more closely. This can be explained by the statistics that models with contextual covariates have a self-adapting scale parameter, which means that the mean value of the model for each interval $y_{m,n}$ is determined by the corresponding contextual covariates $x_{m,n}$. Finally, mixtures further improve the fitting ability of models because the mixed models all performed better than the non-mixed models. Once $K$ is large enough to cover the divergence of the time factor, performance will not be further improved by increasing $K$. In the Criteo dataset, the cWMM fits the test set well when $k = 2$; whereas in the Tmall dataset, the cWMM fits the test set well when $k = 3$. The models of click-conversion intervals (in Criteo) all combine a small shape parameter $\alpha$ (approximately 0.3) and a bigger one (approximately 0.9). The models of purchase-purchase intervals (in Tmall) have a shape parameter $\alpha$ greater than one (approximately 1.6). It is also apparent that contextual regression and mixture extension improve fitting performance more in Tmall than that in Criteo, which means that in Tmall the individual context and the transition mode of difference observations have a stronger effect on interval probabilities. This result explains the irregular shapes of the interval distributions in Figs. 4(c) and 4(d).

### 6.2.4 Transition mode analysis

The proposed cWMM provides a non-monotonic hazard function to identify the various transition modes in different periods, with the character of the shape parameter. In survival analysis, the hazard rate reflects the occurrence rate of a behavior. In this research, it is used to reflect the transition mode from one triggering behavior $b_{i,p}$ to the conversion $b_{i,q}$. One great advantage of the Weibull distribution is that the shape parameter $\alpha$ directly reflects the tendency of the hazard rate. If the shape parameter $\alpha_{m,k} = 1$, the hazard rate is a constant, which means that the behavior occurrence rate is independent of time. If $0 < \alpha_{m,k} < 1$, the hazard rate decreases with time, which means that the positive influence of the triggering behavior is fading. If $\alpha_{m,k} > 1$, the hazard rate increases with time, which means that the negative influence or inhibiting effect of the triggering behavior is fading.

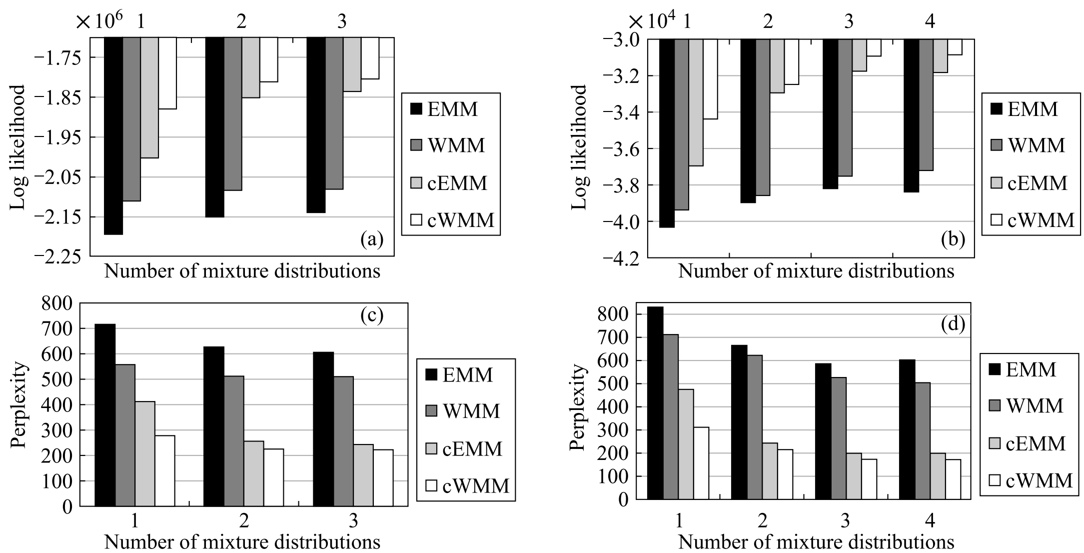



**Fig. 7** Overall comparison between EMM, WMM, cEMM, and the proposed cWMM. (a) Log likelihood for Cretio; (b) log likelihood for Tmall; (c) perplexity for Cretio; (d) perplexity for Tmall

**Table 1** Transition modes

| Ordered Hazard rates | Click→Purchase | | | Purchase→Purchase | | |
|---|---|---|---|---|---|---|
| | Brand 1 | Brand 2 | Brand 3 | Brand 1 | Brand 2 | Brand 3 |
| $\alpha(p_1)<\alpha(p_2)<\alpha(p_3)$ | 0.011 | 0.019 | 0.009 | 0.019 | 0.015 | 0.019 |
| $\alpha(p_1)<\alpha(p_3)<\alpha(p_2)$ | 0.899 | 0.781 | 0.863 | 0.187 | 0.203 | 0.327 |
| $\alpha(p_2)<\alpha(p_1)<\alpha(p_3)$ | 0.006 | 0.006 | 0.018 | 0.007 | 0.041 | 0 |
| Click and buy | **0.916** | **0.806** | **0.890** | 0.213 | 0.259 | 0.346 |
| $\alpha(p_2)<\alpha(p_3)<\alpha(p_1)$ | 0.031 | 0.094 | 0.048 | 0.206 | 0.239 | 0.106 |
| $\alpha(p_3)<\alpha(p_1)<\alpha(p_2)$ | 0.029 | 0.063 | 0.027 | 0.510 | 0.404 | 0.308 |
| $\alpha(p_3)<\alpha(p_2)<\alpha(p_1)$ | 0.023 | 0.037 | 0.035 | 0.080 | 0.098 | 0.240 |
| Re-consumption | 0.083 | 0.194 | 0.110 | **0.786** | **0.741** | **0.654** |

This section introduces how to use the order of $\alpha$s in different periods to reflect the transition modes of observations. In the cWMM ($k = 3$), if the $\alpha$s in three periods are less than one, equal (or approximately equal) to one, and greater than one, the model is competent to model the transition modes between behaviors. There are six orders of $\alpha$, and $\alpha(p_i)$ is used here to represent the shape parameter of period $p_i$. $p_1$, $p_2$, and $p_3$ represent three periods in which $\alpha$ is increasing. The value of $p_i$ can be set to the mean value of the Weibull distribution: $\overline{y}^{(k)} = \theta_k \Gamma\left(\frac{\alpha_k + 1}{\alpha_k}\right)$ $(\theta_k = \exp\left[\beta_k{}^{\mathrm{T}} x\right])$. Table 1 shows the percentages of the six transitions in the Tmall dataset. The six types can be summarized in two categories: click and buy and re-consumption.

The first three orders belong to the first category: click and buy. In this category, the hazard function first decreases and finally converges to a constant. This indicates that each triggering purchase promotes the following one and that the promotional effect decreases with time to a constant hazard rate (0.02 or so). If $\alpha(p_2) > 1$, the hazard rate might increase over the short term. If $\alpha(p_3) > 1$, the probability of the latter behavior is relatively high in period $p_3$, but the hazard rate will still converge to a constant value.

The second three orders belong to the second category: re-consumption. In this category, the hazard function first increases and then decreases, finally converging to some constant. When $\alpha(p_1) > 1$, $p_1$ can be regarded as the fatigue period, which means that the negative effect of the triggering consumption suppresses later consumption. If $\alpha(p_2) > 1$, this situation is much more complex, and the negative effect of the triggering consumption is closely related to other factors, such as the length of the periods and the value of $\alpha$.

The statistical results in Table 1 demonstrate that most click → purchase observations belong to click and buy and most purchase → purchase observations belong to re-consumption. However, only about 75% of purchase → purchase observations are found in re-consumption, which indicates that consumption fatigue does not happen every time in purchase → purchase transitions.

### 6.3 Just-in-time conversion rate prediction

In advertising systems, the probability that an ad will be converted within a specific time period is useful for optimizing advertising strategies. Because the ranked conversion rates of clicked ads are helpful for real-time bidding, the area under the ROC curve (AUC) is chosen as the metric in this case. The competing models are CP and naive TCP. The two feature vectors $X'$ and $X$ are the same in this experiment, which is introduced in Section 3.1. In the mixture lifetime models, we set $K$ to 2, according to the results in Section 6.2.3. The prediction periods in this experiment are one, three, and seven days.

The results are presented in Table 2. It is clear that it is more difficult to predict conversion rates over a short period than over a long period. The proposed rTCP greatly improves the baseline method CP when the prediction period is one day. As the period increases, the AUCs of CP and rTCP become closer. This result can be explained by observing that approximately 50% of conversions occur within one day of the triggering click (see Section 3.1). If the elapsed time is too short, it is too early to tell that the click does not eventually lead to conversion. As for the CP predictions of similar values, the proposed rTCP gives a higher rank to observations with higher time-aware conversion prediction. This is why the predictions of CP may suffer from false negatives over a short period. It is also apparent that naive TCP fails in this task. The joint probabilities of CP and TP make more erroneous predictions over long prediction periods because the predictions of TP increase with the prediction period, which impairs the effect of CP.

Comparing the lifetime models in Table 2, we can see that even the most simple exponential distribution improves the performance of time-aware conversion prediction a lot, especially when the prediction period is short. Context information obviously improves the accuracy, which implies that the history behaviors and preferences of a user have significant influence on the conversion time. And context-aware exponential model performs better than Weibull models without context information. It also indicates that personalization is an important factor of conversion interval prediction. Furthermore, mixture models perform better than the others without mixture, and mixture exponential models even perform a little better than Weibull models without mixture. The comparison of these lifetime models shows that all the models improve the accuracy of conversion prediction for a specified upcoming period, and more comprehensive models have better performance. Each specific application could choose a proper model by considering the tradeoff between accuracy and computing efficiency.

**Table 2** AUC and AUC improvement

| Method | Method | 1 day | | 3 days | | 7 days | |
|---|---|---|---|---|---|---|---|
| | | AUC | Imp. | AUC | Imp. | AUC | Imp. |
| CP | | 0.734 | – | 0.819 | – | 0.847 | – |
| naive TCP | cWMM | 0.653 | –0.081 | 0.648 | –0.171 | 0.613 | –0.234 |
| rTCP | EM | 0.778 | 0.044 | 0.827 | 0.008 | 0.851 | 0.004 |
| | WM | 0.785 | 0.051 | 0.830 | 0.011 | 0.853 | 0.006 |
| | cEM | 0.801 | 0.067 | 0.834 | 0.015 | 0.856 | 0.009 |
| | cWM | 0.813 | 0.079 | 0.836 | 0.017 | 0.859 | 0.012 |
| | cEMM | 0.817 | 0.083 | 0.839 | 0.020 | 0.864 | 0.017 |
| | cWMM | **0.824** | 0.090 | **0.841** | 0.022 | **0.865** | 0.018 |

### 6.4 Just-in-time product recommendations

In the context of product recommendation over time, an online shopping web site wants to know what products will be bought in a specific upcoming period. This is a typical item selection and recommendation problem[4] and can be measured by the F1-score. The CP, naive TCP, and rTCP models can predict only products that the user has clicked or purchased before. A high percentage of new purchases requires a candidate product set based on recommendation algorithms. Therefore, in addition to the three methods used in the last experiment, the following recommendation models are chosen:

- TopPop: a non-personalized baseline to recommend the most popular products to the user.
- LDA: a widely used probabilistic model to estimate the preference value $P(\text{product}|\text{user})$. It has better performance in item selection applications than recommendation models based on matrix factorization, such as SVD and PMF [34].

In this research, TopPop or LDA was used to generate a candidate product set for each user. Then CP, naive TCP, and rTCP re-estimated the product purchase rate for the candidate sets. In the three proposed models, the feature vector $X'$ contains a number of features, including the preference value of LDA and how many times the product has been purchased. In practice, 20 latent factors are used in LDA to achieve better recommendation results. Theoretically, these can be the same as in the LDA used earlier for product clustering (Section 6.1), but the motivation is different. The prediction periods in this experiment are 7, 15, and 30 days.

The specified period $\Delta t$ and the transition $m$ are determined by the triggering product $b_{i,p}$ and the candidate product $b_{i,q}$ of $u_i$. The triggering product can be any historical behavior of $u_i$. In this research, the behavior with the highest transition probability is chosen[5] as the triggering product. In this recommendation system, the top $N$ products are recommended to a user, where $N$ is the average number of products that the user buys in the same period. If a user normally purchases two products within seven days, then $N = 2$ for 7 days; if a user normally purchased five products within 30 days, then $N = 5$ for 30 days.

The results of these experiments are shown in Fig. 8. CP, naive TCP, and rTCP can recommend only products that users clicked or bought in the training set. In collaboration with TopPop or LDA, the three models gain the ability to handle new purchases. The recall of TopPop and LDA is higher than their precision, especially over short prediction periods, which reveals that some potential products, which will be bought eventually, may not be bought in the short term. Therefore, the temporal factor should be taken into consideration. However, like the results of conversion rate prediction, the naive joint probabilities CP and TP introduce extra erroneous estimates, and the results are worse for long prediction periods. In conclusion, the proposed rTCP has achieved the best performance. It markedly improves the recommendations, especially for short prediction periods.

## 7 Conclusions

This research has focused on recommending the right products for a specified period. Considering both relevance and

[4] Item selection is in comparison to item rating. Item selection focuses on which items will be selected by users; whereas item rating focuses on the rating that a user will give to an item

[5] The transition probability is: $Pr(b_{i,p}, b_{i,q}) = \dfrac{\#(b_{i,p}, b_{i,q})}{\#(b_{i,*}, b_{i,q}) + 1}$

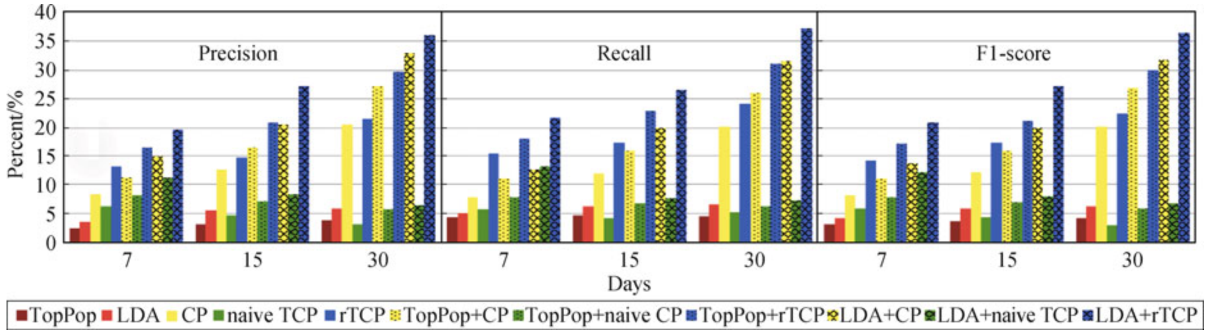



**Fig. 8** Precision, recall, and F1-score of competing product recommendation models

conversion time, a rank-based time-aware conversion prediction model (rTCP) is proposed. We make an analogy between conversion intervals and lifetimes in survival analysis and propose a novel lifetime model (cWMM) based on the Weibull distribution to embody the influence of personalized preference and the complexity of behavioral motivation. And experimental results have shown that the cWMM provides a good fit to the conversion intervals. Moreover, rTCP is helpful in improving the performance of time-aware ad conversion rate prediction and product recommendation.

**Acknowledgements** This work was supported by the National Natural Science Foundation of China (NSFC) (Grant Nos. 61472141, 61532021 and 61021004), Shanghai Knowledge Service Platform Project (ZF1213) and Shanghai Leading Academic Discipline Project (B412).

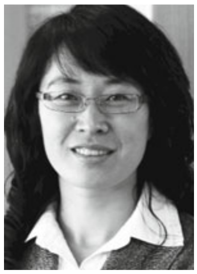
## References

1. Yu J J, Zhu T Y. Combining long-term and short-term user interest for personalized hashtag recommendation. Frontiers of Computer Science, 2015, 9(4): 608–622
2. Ding Y, Li X. Time weight collaborative filtering. In: Proceedings of the 14th ACM International Conference on Information and Knowledge Management. 2005, 485–492
3. Campos P G, Diez F, Cantador I. Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols. User Modeling and User-Adapted Interaction, 2014, 24(1): 67–119
4. Yuan Q, Cong G, Ma Z Y, Sun A, Thalmann N M. Time-aware point-of-interest recommendation. In: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2013, 363–372
5. Bannur S, Alonso O. Analyzing temporal characteristics of check-in data. In: Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion. 2014, 827–832
6. Pazzani M, Billsus D. Learning and revising user profiles: the identification of interesting web sites. Machine Learning, 1997, 27(3): 313–331
7. Zhu T S, Greiner R, Haubl G. Learning a model of a web user's interests. In: Proceedings of the 2003 International Conference on User Modeling. 2003, 65–75
8. Pazzani M J. A framework for collaborative, content-based and demographic filtering. Artificial Intelligence Review, 1999, 13(5): 393–408
9. Guan Y, Cai S M, Shang M S. Recommendation algorithm based on item quality and user rating preferences. Frontiers of Computer Science, 2014, 8(2): 289–297
10. Goldberg K, Roeder T, Gupta D, Perkins C. Eigentaste: a constant time collaborative filtering algorithm. Information Retrieval, 2001, 4(2): 133–151
11. Brown P F, Pietra V J D, Mercer R L, Pietra S A D, Lai J C. An estimate of an upper bound for the entropy of English. Computational Linguistics, 1992, 18(1): 31–40
12. Billsus D, Pazzani M J. Learning Collaborative Information Filters. In: Proceedings of the 15th International Conference on Machine Learning. 1998, 46–54
13. Ma H, Liu C, King I, Lyu M R. Probabilistic factor models for Web site recommendation. In: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2011, 265–274
14. Breese J S, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence. 1998, 43–52
15. Blei D M, Ng A Y, Jordan M I. Latent Dirichlet allocation. Journal of Machine Learning Research, 2003, 3: 993–1022
16. Rendle S, Freudenthaler C, Schmidt-Thieme L. Factorizing personalized markov chains for next-basket recommendation. In: Proceedings of the 19th International Conference on World Wide Web. 2010, 811–820
17. Cheng C, Yang H Q, Lyu M R, King I. Where you like to go next: successive point-of-interest recommendation. In: Proceedings of the 23rd International Joint Conference on Artificial Intelligence. 2013, 2605–2611
18. Rosales R, Cheng H, Manavoglu E. Post-click conversion modeling and analysis for non-guaranteed delivery display advertising. In: Proceedings of the 5th ACM International Conference on Web Search and Data Mining. 2012, 293–302
19. Chapelle O, Manavoglu E, Rosales R. Simple and scalable response prediction for display advertising. ACM Transactions on Intelligent Systems and Technology, 2015, 5(4)
20. Lee K C, Orten B, Dasdan A, Li W J. Estimating conversion rate in display advertising from past erformance data. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discov-

ery and Data Mining. 2012, 768–776

21. Lathia N, Hailes S, Capra L, Amatriain X. Temporal diversity in recommender systems. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2010, 210–217
22. Xiang L, Yuan Q, Zhao S W, Chen L, Zhang X T, Yang Q, Sun J M. 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. 2010, 723–732
23. Koren Y. Collaborative filtering with temporal dynamics. Communications of the ACM, 2010, 53(4): 89–97
24. Zhao G, Lee M L, Hsu W, Chen W. Increasing temporal diversity with purchase intervals. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2012, 165–174
25. Gu W R, Dong S B, Zeng Z Z. Increasing recommended effectiveness with markov chains and purchase intervals. Neural Computing and Applications, 2014, 25(5): 1153–1162
26. Zhang Y F, Zhang M, Zhang Y, Lai G K, Liu Y Q, Zhang H H, Ma S P. Daily-aware personalized recommendation based on feature-level time series analysis. In: Proceedings of the 24th International Conference on World Wide Web. 2015, 1373–1383
27. Zhao X W, Guo Y W, He Y L, Jiang H, Wu Y X, Li X M. We know what you want to buy: a demographic-based system for product recommendation on microblogs. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014, 1935–1944
28. Nelson W B. Applied Life Data Analysis. New York: John Wiley & Sons, 2005
29. Chapelle O. Modeling delayed feedback in display advertising. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Aining. 2014, 1097–1105
30. Wang J, Zhang Y. 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. 2013, 303–312
31. Wang J, Zhang Y, Posse C, Bhasin A. Is it time for a career switch? In: Proceedings of the 22nd International Conference on World Wide Web. 2013, 1377–1388
32. Richards S J. A handbook of parametric survival models for actuarial use. Scandinavian Actuarial Journal, 2012, 2012(4): 233–257
33. Murphy K P. Machine Learning: A Probabilistic Perspective. Cambridge, Massathusetts: The MIT press, 2012
34. Barbieri N, Manco G. An analysis of probabilistic methods for top-*n* recommendation in collaborative filtering. In: Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases. 2011, 172–187

Wendi Ji is currently working toward the doctoral degree in the Software Engineering Institute at East China Normal University, China. Her research interests mainly include user behavior analysis and data mining.



Xiaoling Wang received the bachelor master and doctoral degrees from Southeastern University, China in 1997, 2000, and 2003, respectively. She is currently a professor, and the vice dean in Software Engineering Institute, East China Normal University, China. Her research interests mainly include Web data management, data service technology and applications.

Feida Zhu is an assistant professor at the School of Information Systems of Singapore Management University (SMU), Singapore. He obtained his PhD in computer science from the University of Illinois at Urbana-Champaign (UIUC), USA in 2009 and his BS in computer science from Fudan University, China in 2001. His current research interests include large scale data mining, graph/network mining, and social network analysis.