The figure 1 visualizes the generation of the browsing sequence.

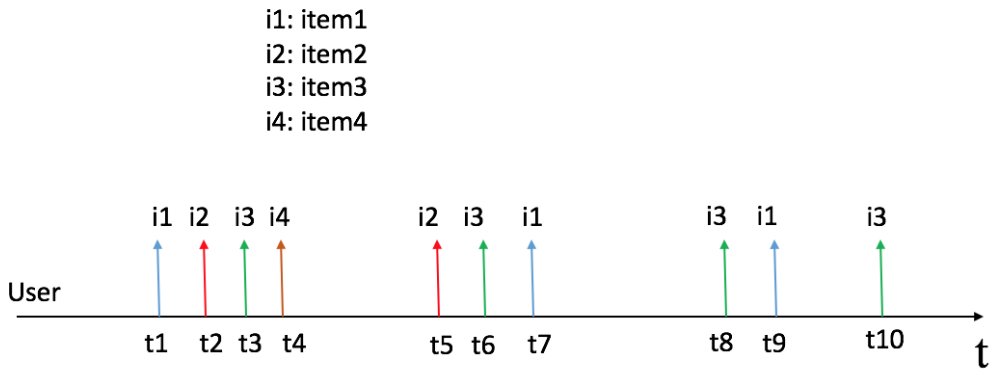


Figure 1: browsing sequence of a purchase of a user

We use $u_i$ to represent the user's browsing preference vector and $v$ to represent the item's attribute vector. $V$ represents the whole set of items.

We believe attributes of browsed items can shape both user's browsing preferences and user's purchasing preferences over time. $\alpha$ represents the shaping pattern adopted by user $u$. Shaping patterns $\{\alpha_1, ..., \alpha_k, ..., \alpha_K\}$ are shared among users. In addition, we believe that not all browsed items will influence the users' preferences, so we use a latent variable $z$ to distinguish whether the browsed item affects the user's preference. When $z = 1$, the browsed item affects the user's preference. When $z = 0$, the browsed item does not affect the user's preference.

So the Eq.4 transforms into the following format,

$$\vec{u}_i(t) = (1 - z_t)\vec{u}_i(t-1) + z_t\Big\{\vec{u}_i(0) + \sum_{t_j<t}(\vec{\alpha}_{t_j} \odot \vec{v}_{t_j})exp\big(-\gamma_1(t-t_j)\big)\Big\} \tag{1}$$

where the $Kernel(t, t_j)$ is replaced with $exp\big(-\gamma_1(t-t_j)\big)$

Since latent variable $z$ is a binary variable, we assume it is generated through a Bernoulli distribution. In addition, we impose Beta distribution as a prior to this Bernoulli distribution.

The generation of $t$ is modeled via a Hawkes Process:

$$\lambda(t) = \lambda_0 + \sum_{t_j<t} z_{t_j}\langle\vec{\beta}^1_{u_i,t_j}, \vec{v}_{t_j}\rangle \exp\big(-\gamma_2(t-t_j)\big) + \sum_{t_j<t}(1-z_{t_j})\langle\vec{\beta}^0_{u_i,t_j}, \vec{v}_{t_j}\rangle \exp\big(-\gamma_2(t-t_j)\big) \tag{2}$$

Given the current user's browsing preferences $u_i(t)$ and items' attributes, the probability of browsing the item $v$ at time $t$ is

$$p(v_t) = \frac{exp(\vec{u}_i(t)^T\vec{v}_t)}{\sum_v^V exp(\vec{u}_i(t)^T\vec{v})} \tag{3}$$

where $v_t$ means the item attributes vector browsed at time $t$.

The whole generative process of the sequence data is:
for each sequence of a user

1. sample $\theta \sim Beta(a, b)$
2. for each action in the sequence
   (a) sample the on/off variable $z \sim \theta$
   (b) sample the timestamp $t \sim \lambda(t)$
   (c) draw the user preference vector $\vec{u}_i(t)$ according to Eq.1
   (d) browse the item $v$ with probability $p(v_t)$ according to Eq.5

If we think of whether browsing an item or not at each timestamp as a binary classification problem, our setting is that we use one set of parameters of logistic regression to predict each browsing behavior and the parameters are linked via the Hawkes Process as the equation 1 shows. Our current setting of logistic regression, which we call ?Hawkes Process Logistic Regression?, lies between two extreme settings of logistic regression. One extreme is using one set of parameters of logistic regression to predict the browsing behaviors of the whole sequence. The other extreme is for each browsing behavior using one set of parameters of logistic regression to predict. To verify whether our intermediate setting can improve the accuracy, we can first verify another simpler intermediate setting that we divide a sequence into multiple groups of browsing behaviors and for each group we use a logistic regression to predict, which we call ?Group Logistic Regression?. If ?Group Logistic Regression? can work better than two extremes, we may expect our current setting can perform best among these models. Since how to divide the group matters a lot in ?Group Logistic Regression? and it is difficult to achieve an ideal division, ?Hawkes Process Logistic Regression? which do not need to divide the sequence into groups may outperform ?Group Logistic Regression?. As a result, ?Hawkes Process Logistic Regression? may outperform other models.

# 1 Baselines

### SVD

update equations:

$$\begin{aligned} b_u &= b_u + \alpha(2 * e_{uv} - \beta b_u) \\ b_v &= b_v + \alpha(2 * e_{uv} - \beta b_v) \\ u &= u + \alpha(2 * e_{ui} * v - \beta u) \\ v &= v + \alpha(2 * e_{ui} * u - \beta v) \end{aligned}$$

### timeSVD++

The predicted score

$$\begin{aligned} \hat{r}(t) &= avg + b_u(t) + b_v(t) + v^T(u + |R(u)|^{-\frac{1}{2}} \sum_{j \in R(u)} y_j) \\ b_u(t) &= b_u + \lambda_u dev(u) + b_{ut} \\ b_v(t) &= b_v + b_{vt} \\ e_{uv} &= r - \hat{r}(t) \end{aligned}$$

update equations

$$
\begin{aligned}
b_u &= b_u + \alpha(2 * e_{uv} - \beta b_u) \\
b_v &= b_v + \alpha(2 * e_{uv} - \beta b_v) \\
b_{ut} &= b_{ut} + \alpha(2 * e_{uv} - \beta b_{ut}) \\
b_{vt} &= b_{vt} + \alpha(2 * e_{uv} - \beta_v b_{vt}) \\
\lambda_u &= \lambda_u + \alpha(2 * e_{uv} * dev(u) - \beta_\lambda \lambda_u) \\
u &= u + \alpha(2 * e_{ui} * v - \beta_u u) \\
v &= v + \alpha(2 * e_{ui} * (u + |R(u)|^{-\frac{1}{2}} \sum_{j \in R(u)} y_j) - \beta_u v) \\
y_j &= y_j + \alpha(2 * e_{ui} * v * |R(u)|^{-\frac{1}{2}} - \beta_y y_j)
\end{aligned}
$$

**Dynamic Logistic Regression**

This model captures the dynamic changes of users' preferences vectors, but does not consider the time information as a variable. So the user's preference vector is defined as followsing,

$$
\vec{u}_i(t) = \vec{u}_i(0) + \sum_{t_j < t} (\vec{\alpha}_{t_j} \odot \vec{v}_{t_j}) Kernel(t, t_j) \tag{4}
$$

Given the current user's browsing preferences $u_i(t)$ and items' attributes, the probability of browsing the item $v$ at time $t$ is

$$
p(v_t) = \frac{exp(\vec{u}_i(t)^T \vec{v}_t)}{\sum_v^V exp(\vec{u}_i(t)^T \vec{v})} \tag{5}
$$

where $v_t$ means the item attributes vector browsed at time $t$.

The parameters of this model is $\alpha$. We assume $\alpha$ is $M$ dimensions. The cluster memberships $z$ of $\alpha$ are latent variables. To infer the cluster membership $z$ and estimate the parameters $\alpha$, we utilize EM algorithm.

In E-step, we estimate the probability of drawing $kth$ $\alpha$ for user $u$ is:

$$
\begin{aligned}
p(z_u = k | S_{u_1}, ..., S_{u_n}, T_{u_1}, ..., T_{u_n}) &= \frac{p(z_u = k, S_{u_1}, ..., S_{u_n}, T_{u_1}, ..., T_{u_n})}{p(S_{u_1}, ..., S_{u_n}, T_{u_1}, ..., T_{u_n})} && (6) \\
&= \frac{L(u)}{\sum_k L(u)} && (7) \\
&= \frac{\prod_{i=1}^{u_n} \prod_{v_t}^{S_{u_i}} \left\{ \frac{\exp(u_b(t)^T v_t)}{\sum_v^V \exp(u_b(t)^T v)} \right\}}{\sum_k^K \prod_{i=1}^{u_n} \prod_{v_t}^{S_{u_i}} \left\{ \frac{\exp(u_b(t)^T v_t)}{\sum_v^V \exp(u_b(t)^T v)} \right\}} && (8)
\end{aligned}
$$

In M-step, To estimate $\alpha$s, we first compute the log complete data likelihood.

$$
\begin{aligned}
L = \sum_u \log p(z_u = k, S_{u_1}, ..., S_{u_n}) &= \sum_u \sum_i^{u_n} \sum_{v_t}^{S_{u_i}} \Big\{ \log \exp(u_b(t)^T v_t) - \log \sum_v^V \exp(u_b(t)^T v) \Big\} \\
&= \sum_u \sum_i^{u_n} \sum_{v_t}^{S_{u_i}} \Big\{ u_b(t)^T v_t - \log \sum_v^V \exp(u_b(t)^T v) \Big\} \\
&= \sum_u \sum_i^{u_n} \sum_{v_t}^{S_{u_i}} \Big\{ u_b(0)^T v_t + \sum_{t_j<t} \sum_k^K 1(z^u = k)(\alpha_{t_j} \odot v_{t_j})^T v_t Kernel(t, t_j) - \\
&\log \sum_v^V \exp\Big( u_b(0)^T v + \sum_{t_j<t} \sum_k^K 1(z^u = k)(\alpha_{t_j} \odot v_{t_j})^T v Kernel(t, t_j)\Big)\Big\}
\end{aligned}
$$

the update equation of $\alpha$ is

$$
\begin{aligned}
\frac{\partial L}{\partial \alpha_{km}} = \sum_u \sum_i^{u_n} \sum_{v_t}^{S_{u_i}} \Big\{ \sum_{t_j<t} p(z^u = k | S_{u_1}, ..., S_{u_n}) v_{t_j m} v_{tm} Kernel(t, t_j) - \qquad (9) \\
\sum_v^V \Big( \frac{\exp\big(\boldsymbol{u_b(0)^T v + \sum_{t_j<t} \sum_k^K \alpha_{km} p(z^u = k | S_{u_1}, ..., S_{u_n}) v_{t_j m} v_m Kernel(t, t_j)}\big)}{\sum_v^V \exp\big(\boldsymbol{u_b(0)^T v + \sum_{t_j<t} \sum_k^K \alpha_{km} p(z^u = k | S_{u_1}, ..., S_{u_n}) v_{t_j m} v_m Kernel(t, t_j)}\big)} \qquad (10) \\
\times \sum_{t_j<t} p(z^u = k | S_{u_1}, ..., S_{u_n}) v_{t_j m} v_m Kernel(t, t_j)\Big)\Big\} \qquad (11)
\end{aligned}
$$

where $Kernel(t, t_j) = exp\big( - \gamma_1(t - t_j)\big)$

$$
\begin{aligned}
L &= \sum_u \log p(z_u = k, S_{u_1}, ..., S_{u_n}) \\
&= \sum_u \sum_i^{u_n} \sum_{v_t}^{S_{u_i}} \Big\{ \log \frac{\exp(u_b(t)^T v_t)}{\sum_v^V \exp(u_b(t)^T v)} \Big\} \\
&= \sum_u \sum_i^{u_n} \sum_{v_t}^{S_{u_i}} \Big\{ \log \frac{1}{1+\exp(-u_b(t)^T v_t)} + \sum_{v, v \neq v_t}^{V} \log \frac{1}{1+\exp(u_b(t)^T v)} \Big\}
\end{aligned}
$$

$$
\Big\{ \log \frac{\exp(u_b(t)^T v_t)}{\sum_v^V \exp(u_b(t)^T v)} \Big\}
$$

$$
\Big\{ \log \frac{1}{1+\exp(-u_b(t)^T v_t)} + \sum_{v, v \neq v_t}^{V} \log \frac{1}{1+\exp(u_b(t)^T v)} \Big\}
$$

$$
\Big\{ \log \frac{1}{1+\exp(-u_b(t)^T v_t)} + \sum_{m, m \neq v_t}^{M} \log \frac{1}{1+\exp(u_b(t)^T m)} \Big\}
$$