The figure 1 visualizes the generation of the browsing sequence.

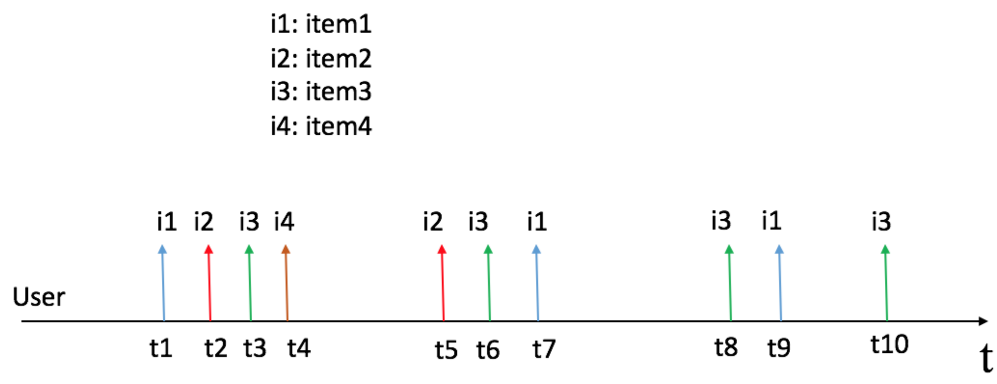

Figure 1: browsing sequence of a purchase of a user

We use a latent vector u to represent a user and a latent vector i to represent an item. The vector $u(t)$ means the user's preferences (aspect weights) at time t. On the other hand, the vector i_t means the item $i's$ attributes (aspect qualities), which is browsed by the user at time t. We believe the attributes of browsed items can shape the user's preferences over time and how the attributes of browsed items shape the users preferences can be categorized into groups shared by users. We use α_t to represent the shaping pattern at time t. Thus, at time t, the preference of a user u can be represented as,

$$u(t) = u_0 + \sum_{t_j < t} \alpha_u^{t_j} i_{t_j} \tag{1}$$

$\sum_{t_j<t}$ captures the sequential influence of previous events on the current browsing behavior. And for time t_j, the corresponding shaping pattern is $\alpha^{t_j} \sim Gaussian(\mu, \sigma)$. Given the current user's preferences $u(t)$ and items' attributes, the probability of browsing the item k at time t is

$$p(k) = \frac{exp(u(t)^T k)}{\sum_i exp(u(t)^T i)}$$

where i is the item $i's$ latent vector.

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.