

Development and Derivation of The Psychological Model

June 29, 2018

This document provides development logs of the psychological model, including intuitions, derivations, and result summaries.

To begin with, we must outline several key concepts in building the blocks of our models.

- K aspects. We assume there are K aspects based on which a user measure the items. The number K is predefined. Consequently, we define the variables, U, V where $v_{i,k}$ denotes item i 's feature on aspect k and $u_{p,k}$ denotes user p 's preference on aspect k . The range of U, V is specified by the model, thus a probabilistic generative model and a numerical model might provide apply different range constraints.
- Prominent aspect(s). We assume that the K aspects are not treated equally by the users. One or some (depending on the model) of the aspects are most important.
- Distinguish prominent and non-prominent aspects. We adopt different scoring or ranking methods for prominent and non-prominent aspects. This part essentially makes our model non-compensate.

Part I

Ranking Based Models

This part devotes to models that observe rankings. In each session, we assume that the user is given a set of items, and the user generates a set of pairwise rankings. For example, if a user has access to items i, j , and he buys i , clicks on j . Then we can make a reasonable conclusion that to the user $i \succ j$. The ranking based models mimic the generation of pairwise rankings. The details of observing rankings, i.e. which actions to be taken into accounts and how to grade these actions, are out of the scope of this document (but should be considered in the experiments).

0.1 The BTL Model

Part II

Rating Based Models

In this section, we develop a series of ranking based models. These models have the following properties. (1) The models are based on rating observations, i.e. we assign a rating to every action. The actions include the old-fashioned preference ratings (in such case, each rating is a standalone session), actions in a session, positive and negative samples, etc. (2) The models are Bayesian generative models that have good explainability. (3) We control the assumptions by adjusting the hyper-parameters in these models.

0.2 The Sigmoidal model

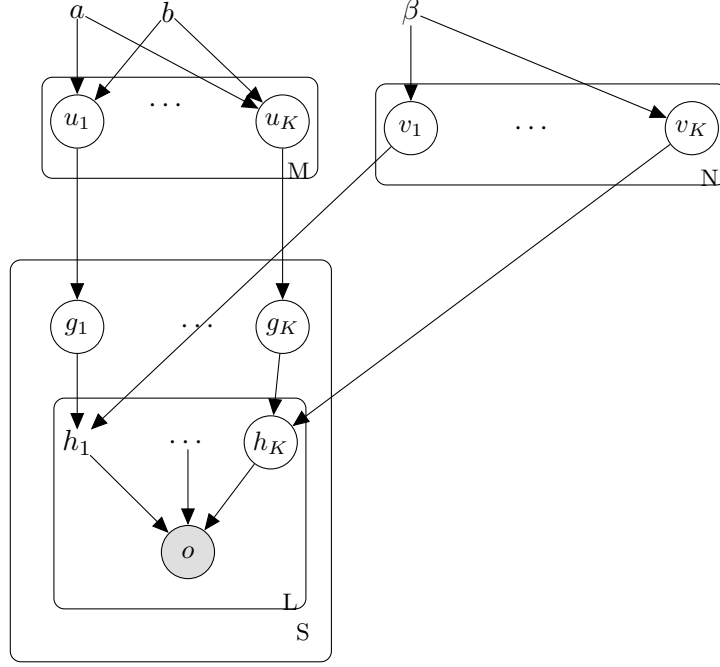
We are motivated to design this model by the following assumptions: (1) we allow multiple ($0 < K$) aspects in a session to be prominent. (2) We use positive and negative observations, e.g. a click is negative and a buy is positive. (3) For prominent aspects, we use the sigmoidal function as the likelihood $p(o_v = 1|g_k = 1) = \sigma(v_k)$. (4) For non-prominent aspects, we assume that they do not contribute to the likelihood of being positive or negative, ie. $p(o = 1|g_k = 0) = 0.5$.

As such, in experiments, we should monitor model performances on the following issues. (1) Shall we induce sparse constraints on the number of prominent aspects? This could be done by choosing proper values for a, b . (2) Shall we use confidence levels for the negative samples, especially for implicit feedback? (3) Shall we consider ratings, instead of positive and negative samples? A fair comparative study with MF is needed to answer this question.

This model might also suffer from the following weaknesses. (1) The likelihood is obtained by a multiplication of all aspects. However, this is proved by preceding pilot studies, that a complete BTL models multiplying all aspects is better than a simplified BTL model on just one aspect. (2) The non-prominent aspects are discarded. This could be fixed by introducing a step function, i.e. $p(o_v = 1|g_k = 0) = f(v - \theta)$.

First we generate the user preferences and item features.

- For aspect $k = 1 : K$, for user $u = 1 : M$, sample user preference on aspect $k : u_k \sim \text{Beta}(\alpha)$. Unlike most of previous research, the user preference vector u in our work is not the normalized weight over all aspects.
- For aspect $v = 1 : N$, sample item features for the item universe $v \sim \mathcal{N}(0, \beta)$. The item features resemble the factors in previous research.



Next, given S sessions of user u with L ranking observations with known item ids.

- For each session s , for aspect $k = 1 : K$, generate aspect indicator $g_{s,k} \sim \text{Bern}(u_k)$
- For each aspect, generate the indicator $p(h_k = 1|v_k, g_k) = \frac{1}{1+\exp v_k} \frac{g_k}{2} \frac{1}{2}^{1-g_k}$
- For each observed rating o , generate ranking observation $p(o) = \text{XOR}(h_1, \dots, h_K)$ where $o = 1$ if and only if $\forall k, h_k = 1$

The joint probability is given by

$$p(U, V, G, H, D|a, b, \beta) = \prod_v \{p(v|0, \beta) \prod_u \prod_k [p(u_k|a, b) \prod_s p(o|H)p(g_{s,k}|u_k)p(h_{s,k}|g_{s,k}, v_k)]\}$$

Where:

The upperbound for $p(h_k|g_{s,k}, v_k)$ is

$$\begin{aligned} p(h_{s,k}|g_{s,k}, v_k) &= \frac{1}{2}^{1-g_{s,k}} [\sigma(v_k)^{h_{s,k}} (1 - \sigma(v_k))^{1-h_{s,k}}]^{g_k} \\ &\leq \frac{1}{2}^{1-g_{s,k}} \exp(v_k h_{s,k}) \sigma(\xi) \exp[-\frac{v_k + \xi}{2} - \lambda(\xi)(v_k^2 - \xi^2)] \end{aligned}$$

By variational inference, we have factorize the posterior probability $p(U, V, G, H|D, a, b, \Lambda) = q(U)q(V)q(G)q(H)$, where $\Lambda = \beta^{-1}$

$$\begin{aligned}\ln q(u_k) &= \mathbb{E}_{G,V,H}[\ln p(u_k|a, b) + \ln p(g_k|u_k)] + \text{const} \\ &= (a-1) \ln u_k + (b-1) \ln(1-u_k) + \sum_s \mathbb{E}[g_{s,k}] \ln u_k + (1 - \mathbb{E}[g_{s,k}] \ln(1-u_k) + \text{const} \\ \mathbb{E}[u_k] &= \frac{a-1 + \sum_s \mathbb{E}[g_{s,k}]}{a-1 + \sum_s \mathbb{E}[g_{s,k}] + b-1 - \mathbb{E}[g_{s,k}]}\end{aligned}$$

$$\begin{aligned}\ln q(v) &= \mathbb{E}_{G,U,H}[\ln p(H|v, G) + \ln(v|\Lambda)] + \text{const} \\ &= \sum_u \sum_s \sum_k \left\{ \mathbb{E}[g_{s,k}] [v_k h_{s,k} - \frac{v_k + \xi}{2} - \lambda(\xi)(v_k^2 - \xi^2)] - \frac{1}{2} v^T \Lambda v \right\} + \text{const} \\ \Sigma[v]^{-1} &= \mathbf{I}(\Lambda - 2\lambda(\xi) \sum_s \sum_u \mathbb{E}[g_{s,k}]) \\ \mathbb{E}[v] &= \Sigma[v] \left(\sum_u \sum_s \mathbb{E}[g_{s,k}] (h_{s,k} - 1/2) \right)\end{aligned}$$

$$\begin{aligned}\ln q(g_{s,k}) &= \mathbb{E}_{U,V,H}(\ln(g_{s,k}|u_k) + \ln(H|G, U)) + \text{const} \\ &\geq \sum_s g_{s,k} \mathbb{E}[\ln u_k] + (1 - g_{s,k}) \mathbb{E}[1 - \ln u_k] + \ln \frac{1}{2} (1 - g_{s,k}) + g_{s,k} \{ \mathbb{E}[h_k] \mathbb{E}[\sigma(v_k)] + \mathbb{E}[1 - h_k] \mathbb{E}[1 - \sigma(v_k)] \} \\ \mathbb{E}[g_{s,k}] &= \frac{\mathbb{E}[\ln u_k + \{ \mathbb{E}[h_k] \mathbb{E}[\sigma(v_k)] + \mathbb{E}[1 - h_k] \mathbb{E}[1 - \sigma(v_k)] \}]}{\mathbb{E}[\ln u_k + \{ \mathbb{E}[h_k] \mathbb{E}[\sigma(v_k)] + \mathbb{E}[1 - h_k] \mathbb{E}[1 - \sigma(v_k)] \} + \mathbb{E}[1 - \ln u_k] + \ln \frac{1}{2}}\end{aligned}$$

$$\begin{aligned}\ln q(H_s) &= \mathbb{E}_{G,U,V}(\ln p(H_s|V_s, G_s) + \ln p(o|H)) \\ &\geq \sum_k \mathbb{E}[g_{s,k}] \{ h_{s,k} \ln \sigma(v_k) + (1 - h_{s,k}) \mathbb{E}[\ln(1 - \sigma(v_k))] \} + \ln p(o|H) + \text{const} \\ \mathbb{E}[H_s] &=?\end{aligned}$$