Development and Derivation of The Psychological Model

July 2, 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 Beta 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 directly model the likelihood by the Beta function. (4) For non-prominent aspects, we assume that they do not contribute to the likelihood of being positive or negative, ie. $p(o = 1|g_k = 0) = 0.5$. (5) We assume that the final label is assigned by a very strict measurement, i.e. the item must be "good" on all aspects.

As such, in experiments, we should monitor model performances on the following issues. (1) A proper baseline would be to also treat the recommendation system as binary classification problems. For simplicity, I would recommend starting with those traditional MF or classifiers without any number, or time related features. (2) The binary classification framework is easily extend to one class classification scheme for implicit feedback. However, I would recommend to first testify the performance for two classes. When we involve implicit feedback, a fair comparison is important. It would be best if we could verify the capacity of our model on two-class problems, before we use confidence levels or sampling techniques for the negative implicit feedback. (3) Shall we induce sparse constraints on the number of prominent aspects? This could be done by choosing proper values for a, b. (4) Tuning the parameter K is the last step in experimental studies.

For future extensions, we should keep in mind that this model might suffer from the following weaknesses. (1) The likelihood is obtained by a multiplication of all aspects. Although 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, I am not very confident about the conclusion. We can easily modify the model to account for an opposite assumption, that the "good" performance on one aspect can override other aspects. (2) The non-prominent aspects are discarded. This could be fixed

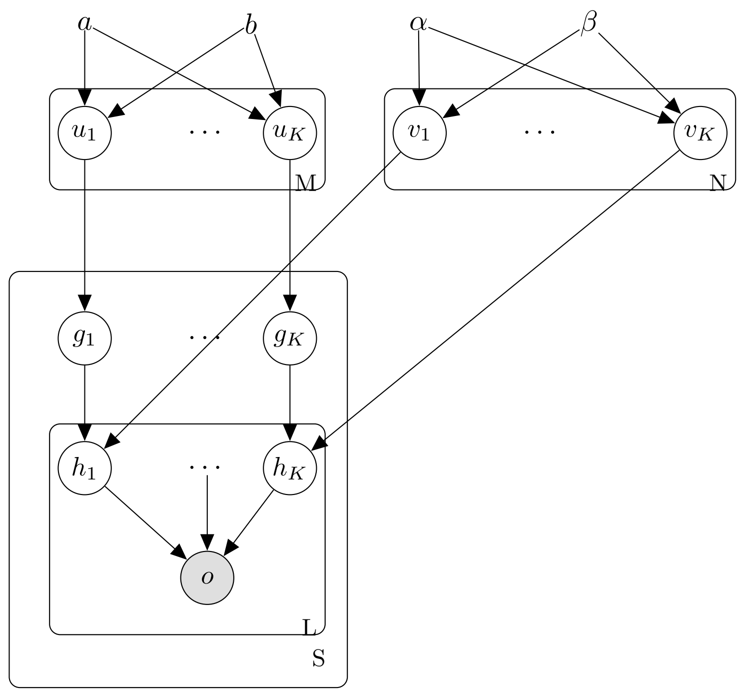

Figure 1: Plate notation of the proposed sigmoidal model

by introducing a step function, i.e. $p(o_v = 1|g_k = 0) = f(v-\theta)$. (3) Shall we consider ratings (normalized to $(0, 1)$), instead of positive and negative samples? A fair comparative study with MF is needed to answer this question. If the answer is, we can modify the generation of observations. All the above directions are possible and can be implemented without much difficulty.

First we generate the user preferences and item features.

- For aspect $k = 1 : K$, for user $u = 1 : M$, sample user preference on aspect $k : u_k \sim Beta(a, b)$. Unlike most of previous research, the user preference vector u in our work is not the normalized weight over all aspects.

- For items $v = 1 : N$, sample item features for the item universe $v_k \sim Beta(\alpha, \beta)$. Again, unlike most of previous research, the user preference vector u in our work is not the normalized weight over all aspects.

Next, given S sessions of user u

- For each session s, for aspect $k = 1 : K$, generate aspect indicator $g_{s,k} \sim Bern(u_k)$. If $g_{s,k} = 1$, then k is a prominent aspect for user u in the session s.

- For each item v within the session s of length L

  - For each aspect k, generate the indicator $p(h_{s,v,k}|g_{s,k}, v_k) = [v_k^{h_{s,v,k}}(1-v_k)^{1-h_{s,v,k}}]^{g_{s,k}}[\frac{1}{2}^{h_{s,v,k}}\frac{1}{2}^{1-h_{s,v,k}}]^{1-g_{s,k}}$
  - For each observed rating o_v, generate label observation $o_{s,v} = 1 - \prod_k(1 - h_{s,v,k})$ where $o_v = 1$ if and only if $\forall k, h_{s,v,k} = 1$

The joint probability is given by

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

We should notice that $o_{v,k} = 1$ indicates $\forall k, h_{v,k} = 1$. Thus the likelihood must be decomposed to a term with observations and a term with hidden variables. We have:

$$\begin{aligned} \ln p(U, V, G, H, D|a, b, \Lambda, \xi) &= \sum_s \sum_{o_{s,v}=1} \sum_k p(h_{s,v,k} = 1|g_{s,k}, v_k) \\ &+ \sum_s \sum_{o_{s,v}=0} p(o_{s,v} = 0|H) \sum_k p(h_{s,v,k}|g_{s,k}, v_k) \\ &+ \sum_u \sum_k p(u_k|a, b) + \sum_v \sum_k p(v_k|\alpha, \beta) + \sum_u \sum_s \sum_k p(g_{s,k}|u_k) \end{aligned} \quad (1)$$

By variational inference, we have factorize the posterior probability $p(U, V, G, H|D, a, b, \alpha, \beta) = q(U)q(V)q(G)q(H)$. We can see from the following derivations that $q(u_k) \sim Bern(a', b')$

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

Because $p(h_{s,v,k}|g_{s,k}, v_k) = [v_k^{h_{s,v,k}}(1-v_k)^{1-h_{s,v,k}}]^{g_{s,k}}[\frac{1}{2}^{h_{s,v,k}}\frac{1}{2}^{1-h_{s,v,k}}]^{1-g_{s,k}}$, we have $\ln p(h_{s,v,k}|g_{s,k}, v_k) = g_{s,k}[h_{s,v,k}\ln v_k + (1 - h_{s,v,k}\ln(1 - v_k)] + (1 - g_{s,k}\ln\frac{1}{2}$

$$
\begin{aligned}
\ln q(v) &= \mathbb{E}_{G,U,H}[\ln p(H|v,G) + \ln(v|\alpha,\beta)] + const \\
&= \sum_s \sum_{o_{s,v}=1} \sum_k p(h_{s,v,k}=1|g_{s,k},v_k) + \sum_s \sum_{o_{s,v}=0} \sum_k p(h_{s,v,k}|g_{s,k},v_k) + \sum_v \sum_k p(v_k|\alpha,\beta) \\
\ln q(v_k) &= \{\sum_s \sum_{o_{s,v}=1} \mathbb{E}[g_{s,k}] + \sum_s \sum_{o_{s,v}=0} \mathbb{E}[g_{s,k}]\mathbb{E}[h_{s,v,k}]\} \ln v_k \\
&\quad + \sum_s \sum_{o_{s,v}=0} \mathbb{E}[g_{s,k}]\mathbb{E}[1-h_{s,v,k}] \ln(1-v_k) + (\alpha-1)\ln v_k + (\beta-1)\ln(1-v_k) + const \\
\mathbb{E}[v_k] &= \frac{\sum_s \sum_{o_{s,v}=1} \mathbb{E}[g_{s,k}] + \sum_s \sum_{o_{s,v}=0} \mathbb{E}[g_{s,k}]\mathbb{E}[h_{s,v,k}] + \alpha}{\sum_s \sum_{o_{s,v}=1} \mathbb{E}[g_{s,k}] + \sum_s \sum_{o_{s,v}=0} \mathbb{E}[g_{s,k}]\mathbb{E}[h_{s,v,k}] + \alpha + \beta + \sum_s \sum_{o_{s,v}=0} \mathbb{E}[g_{s,k}]\mathbb{E}[1-h_{s,v,k}]}
\end{aligned}
$$

$$
\begin{aligned}
\ln q(G) &= \mathbb{E}_{U,V,H}(\ln p(G|U) + \ln(H|G,U)) + const \\
\ln q(g_{s,k}) &= \ln p(g_{s,k}|u_k) + \sum_{o_{s,v}=1} p(h_{s,v,k}=1|g_{s,k},v_k) + \sum_{o_{s,v}=0} p(h_{s,v,k}|g_{s,k},v_k) \\
&= g_{s,k}\{\mathbb{E}[\ln u_k] + \mathbb{E}[\ln v_k] + \sum_{o_{s,v}=1} \mathbb{E}[\ln v_k] + \sum_{o_{s,v}=0} [\mathbb{E}[h_{s,v,k}]\mathbb{E}[\ln v_k] + \mathbb{E}[1-h_{s,v,k}]\mathbb{E}[1-\ln v_k] \\
&\quad +(1-g_{s,k})\{\mathbb{E}[\ln(1-u_k)] + \sum_{o_{s,v}} \ln \frac{1}{2}\} \\
\mathbb{E}[g_{s,k}] &= \frac{\mathbb{E}[\ln u_k] + \mathbb{E}[\ln v_k] + \sum_{o_{s,v}=1} \mathbb{E}[\ln v_k] + \sum_{o_{s,v}=0}[\mathbb{E}[h_{s,v,k}]\mathbb{E}[\ln v_k] + \mathbb{E}[}{\mathbb{E}[\ln u_k] + \mathbb{E}[\ln v_k] + \sum_{o_{s,v}=1} \mathbb{E}[\ln v_k] + \sum_{o_{s,v}=0}[\mathbb{E}[h_{s,v,k}]\mathbb{E}[\ln v_k] + \mathbb{E}[1-h_{s,v,k}]\mathbb{E}[1-\ln v_k}
\end{aligned}
$$

where $\mathbb{E}[\ln(u_k) = \phi(a')\phi(a'+b'), \mathbb{E}[\ln(v_k) = \phi(\alpha')\phi(\beta')$.

$$
\begin{aligned}
\ln q(H) &= \mathbb{E}_{G,U,V}(\ln p(H|V,G) + \ln p(O|H)) \\
\ln q(h_{s,v,k}=1) &= \sum_{o_{s,v}=1} \sum_k p(h_{s,v,k}=1|g_{s,k},v_k) + \sum_{o_{s,v}=0} \sum_{h_{s,v,k'}} p(o_{s,v}=0|H) p(h_{s,v,k'}|g_{s,k'},v_k) + const \\
\mathbb{E}[h_{s,v,k}] &= q(h_{s,v,k}=1) + q(h_{s,v,k}=0)
\end{aligned}
$$