

A Psychological Model for Consumption Prediction

Paper 277

ABSTRACT

The study of consumer psychology reveals two categories of procedures used by consumers to make consumption related choices: compensatory rules and non-compensatory rules. Existing recommendation models such as matrix factorization assume the consumers follow the compensatory rules, which is to make decisions based on a weighted or summated score over different attributes. In this paper, we present a novel model which adopts non-compensatory decision rules. The selected item is superior on the most important attribute and beyond the minimally acceptable level on other attributes. Furthermore, we incorporate ordinal utility in the psychological model for consumption prediction in click sessions. We experimentally demonstrate that this model outperforms state-of-the-art methods.

KEYWORDS

consumer psychology, non-compensatory decision rules, ordinal utility, early prediction

1 INTRODUCTION

Ever since the dawn of psychological science, many attempts have been made to explain human choice behavior. It is well regarded that there are two categories of procedures to make consumption related choices: compensatory rules and non-compensatory rules [?]. Consumers who adopt compensatory rules evaluate every product over multiple relevant aspects and compute a weighted or summated score for each product. Then they will select the products with highest scores. The shortcomings of a product are balanced out by its attractive features. On the contrary, non-compensatory rules do not allow a good performance on one aspect of a product to compensate for poor performances on other aspects. Non-compensatory rules include lexicographic, conjunction and disjunction rules. Under lexicographic rules, products are compared on the most important aspect. Under conjunction and disjunction rules, the consumer imposes requirements for minimally acceptable values on each aspect separately. The conjunction and disjunction rules are often used in conjunction with other decision rules.

One essential goal of recommender systems (RS) is to understand consumption patterns. In the fruitful literature of RS, almost all models are implementations of the compensatory decision rules, i.e. they assume that the probability of a consumption is calculated by a weighted combination of item performances over several hidden aspects [?, ?]. To the best of our knowledge, there is no model that is based on non-compensatory decision making rules.

However, many researches in the field of consumer psychology revealed that users use both compensatory and non-compensatory rules [?]. The selection of decision rules is affected by many factors. For example, a survey [?] pointed out that when the consumer is familiar with the product and the product is low in complexity,

he is most likely to establish a set of evaluative criteria and adopt conjunction rules.

Our first contribution in this work is a novel probabilistic model that represents the cognitive process of consumers following non-compensatory rules. The problem being tackled here is to predict what item the consumer will buy in an activity session (i.e. a set of clicks on a e-commerce website). We assume that an item is purchased because (1) it is superior on the aspect which is currently of the most pertinence to the consumer, and (2) it surpasses a minimum level of acceptable performance on other aspects.

When modeling consumers' decision making process, another important issue is that consumers' preferences are described by ordinal utility rather than cardinal utility. Ordinal utility is the idea that user's satisfaction (utility) cannot be measured in absolute quantity, while cardinal utility theory presumes utility is measurable and additive. Suppose there are two items to evaluate, it is possible to ask a consumer which item is better, but it is meaningless to ask him how much better it is. In modern economics, ordinal utility theory is applied to study consumer behavior.

RS models are usually expressed in terms of cardinal utility, largely because RS models are dealing with choices under uncertainty. In the few recent works [?] which employ ordinal utility, only the final decisions (i.e. differences among explicit ratings) are considered to be in the domain of ordinal variables.

According to study of consumer behavior [?], after a desire of purchase is triggered, the consumer will proceed with a series of alternative evaluations using limited memory and attention. Our hypothesis in this paper is that ordinal utility is applied by the consumer throughout the decision process. We give an example below to show that, when coupling with non-compensatory rules, if the model does not reflect the complete decision making process, the prediction might be misleading.

Suppose there are 5 options $< A, B, C >$, exposed to the customer sequentially. The consumer evaluates each option on two attributes.

Figure 1: default

Our second contribution is to incorporate ordinal utility. We assume that a sequence of evaluations. ordinal utility is kept in mind. present several heuristics .

This paper is structured as follows.

2 NON-COMPENSATORY MODEL

Suppose we have the item universe $V = \{v_1, \dots, v_M\}$ of M items, the user universe $U = \{u_1, \dots, u_N\}$ of N users, input data $D = d$ consists of buying sessions, each of which is a pair of a winning item and a set of losing items, $d = \langle W^d, L^d \rangle = \langle v_1, v_2, \dots, v_l \rangle$. We use subscripts to denote the elements in each vector, i.e. the preference value of user u in the k -th aspect is u_k .

Imagine that there are K underlying aspects, the user preferences and item features are both quantified in the aspects. Therefore we

assign a k -dimensional vector for each user as the user preference vector. Without ambiguity, we use the same notion u_i . We do the same to each item. The consumption session d for user u_d is a generative process:

- Choose a favorite aspect a according to u_d
 - Generate d according to Equ. 1

As mentioned in Sec. 1, the winning item $w \in W^d$ is a skyline object. To define the probability of skyline events, we turn to the BTL model [?] in social science. In the BTL model, a ranking $i > j$ is likely to happen if the latent utility score s_i is relevantly larger than s_j . To be specific, we have define $p(i > j), p(i < j)$ as follows to guarantee $p(i > j) + p(i < j) = 1$

$$p(i > j) = \frac{s_i}{s_i + s_j}$$

$$p(i < j) = \frac{s_j}{s_j + s_i}$$

If the favorite aspect in current session is k , then following the definition of skyline object, $w \in W^d$ is superior in the favorite aspect, and is not worse than other items in other aspects. $w \in W^d$ ranks higher than item $v \in L^d$ in the aspect k is $p^k(w > v) = \frac{w_k}{w_k + \theta v_k}$; the probability of item $w \in W^d$ ranks not lower than item $v \in L^d$ in the aspect $k' = k$. Under the extended BTL model, we have $p^{k'}(w = v) = (\theta^2 - 1) \frac{w_{k'} v_{k'}}{[w_{k'} + \theta v_{k'}][\theta w_{k'} + v_{k'}]}$. Therefore, $p^{k'}(w \geq v) = 1 - p^{k'}(v > w) = \frac{\theta w_{k'}}{v_{k'} + \theta w_{k'}}$. Suppose g is a K -dim vector, with one and only one component to be equal to 1, for each pair, $p(< w, v > |g_k = 1) = p^k(w > v) \times \prod_{k' \neq k} p^{k'}(w \geq v)$. then the probability of generating a session observation d given the hidden aspect a is defined as:

$$p(d|g, \theta, V, U) = \prod_{w \in W^d, v \in L^d} \prod_{k=1}^K \left[\frac{w_k}{w_k + \theta v_k} g_k \frac{\theta w_k}{v_k + \theta w_k}^{1-g_k} \right] \quad (1)$$

The parameter $\theta > 1$ plays as a tolerance threshold. If the absolute difference between the winning item and the losing item in aspect a is not significant to the user, $|w_a - v_a| \leq \theta$, then the consumer will tolerate such a difference and calls a tie. Again, the property $p(w_a > v_a) + p(v_a > w_a) + p(w_a = v_a) = 1$ is guaranteed for any aspect and item.

The model parameters are denoted as $\Theta = \{\theta, v \in V, u \in U\}$, which include user preferences u , item features v and tolerance threshold θ . To inference the model parameters, we present the following EM algorithm to maximize the log-likelihood $\ln p(D|u, v, \theta)$.

First, let's use $\gamma(d, k, \Theta^t)$ to denote the conditional probability $p(g_k = 1|d, \Theta^t)$ given parameters in the t -th round, when the current session specific favorite aspect is $g_k = 1$, defined as follows

$$\gamma(d, k, \Theta^t) = \frac{p(d, g|\Theta^t)}{\sum_g p(d, g|\Theta^t)} = \frac{p(g|\Theta^t)p(d|\Theta^t, g)}{\sum_g p(g|\Theta^t)p(d|\Theta^t, g)} \quad (2)$$

$$= \frac{u_k \prod_{w \in W^d, v \in L^d} \frac{w_k}{w_k + \theta^t v_k} \prod_{k' \neq k} \frac{\theta^t w_{k'}}{v_{k'} + \theta^t w_{k'}}}{\sum_{k=1}^K u_k \prod_{w \in W^d, v \in L^d} \frac{w_k}{w_k + \theta^t v_k} \prod_{k' \neq k} \frac{\theta^t w_{k'}}{v_{k'} + \theta^t w_{k'}}}$$

Note that $\forall d, \sum_k \gamma(d, k, \Theta^t) = 1$.

In the E-step of t -th EM round, compute the expectation $Q(\Theta^t) = E_G \ln p(D, G|\Theta)$

$$E_G \ln p(D, G|\Theta) = \sum_d \sum_{k=1}^K \gamma(d, k, \Theta^t) \ln p(d, g|\Theta) \quad (3)$$

$$= \sum_d \sum_{k=1}^K \gamma(d, k, \Theta^t) \left\{ \ln u_k + \sum_{w \in W^d, v \in V_d} \left[\ln \frac{w_k}{w_k + \theta v_k} + \sum_{k' \neq k} \ln \frac{\theta w_{k'}}{v_{k'} + \theta w_{k'}} \right] \right\}$$

In the M-step, we first maximize $Q(\Theta^t)$ with respect to U . For each $u \in U$, eliminating constant terms, we have:

$$\min -\sum_{u(d)=u} \sum_{k=1}^K \gamma(d, k, \Theta^t) \ln u_k \quad (4)$$

w.r.t $\sum_k u_k = 1$

Solving the above Lagrange function Equ. 4, we get

$$u_k = \frac{\sum_{u(d)=u} \gamma(d, k, \Theta^t)}{\sum_{s=1}^K \sum_{u(d)=u} \gamma(d, s, \Theta^t)} \quad (5)$$

By applying $\ln v_k - \ln(v_k + \theta^t v_k') \geq \ln v_k + 1 - \ln(v_k^t + \theta^t v_k'^t) - \frac{v_k + \theta^t v_k'}{v_k^t + \theta^t v_k'^t}$, we obtain a minorization function of $\tilde{Q}(\Theta^t)$.

$$\tilde{Q}(\Theta^t) = \sum_d \sum_{k=1}^K \gamma(d, k, \Theta^t) \sum_{w \in W^d, v \in L_d} \left\{ [\ln w_k + 1 - \ln(w_k^t + \theta^t v_k^t) - \frac{w_k + \theta v_k}{w_k^t + \theta^t v_k^t}] + \sum_{k' \neq k} [\ln(\theta w_{k'}) + 1 - \ln(v_{k'}^t + \theta^t w_{k'}^t) - \frac{v_{k'} + \theta w_{k'}}{v_{k'}^t + \theta^t w_{k'}^t}] \right\} \quad (6)$$

One advantage of $\tilde{Q}(\Theta^t)$ is that it can be separated for each item v . Considering only the k -th component $v_k, \tilde{Q}(v_k, \Theta^t)$ involves two terms, one of which is relevant to observations $d \in W(v)$ where v acts as skyline object, the other is relevant to observations $d \in L(v)$ where v acts as comparisons, $\tilde{Q}(v_k, \Theta^t) = \tilde{Q}^1(v_k, \Theta^t) + \tilde{Q}^2(v_k, \Theta^t)$. Removing all constants and irrelevant terms for v_k , we have the following minorizing function:

$$\tilde{Q}^1(v_k, \Theta^t) = \sum_{d \in W(v)} |L_d| \ln v_k - v_k \sum_{d \in W(v)} \sum_{v' \in L_d} \left[\frac{\gamma(d, k, \Theta^t)}{\alpha(v, v', k, \Theta^t)} + \sum_{k' \neq k} \frac{\theta^t \gamma(d, k', \Theta^t)}{\alpha(v', v, k, \Theta^t)} \right]$$

$$\tilde{Q}^2(v_k, \Theta^t) = -v_k \sum_{d \in L(v)} \sum_{v' \in W_d} \left[\frac{\theta^t \gamma(d, k, \Theta^t)}{\alpha(v', v, k, \Theta^t)} + \sum_{k' \neq k} \frac{\gamma(d, k', \Theta^t)}{\alpha(v, v', k, \Theta^t)} \right]$$

where $|L_d|$ is the number of objects being dominated in d , $\alpha(v, v', k, \Theta^t) = v_k^t + \theta^t v_k'^t$. By setting the partial derivative of $\frac{\partial \tilde{Q}(v_k, \Theta^t)}{\partial v_k} = 0$, we have:

$$\frac{1}{v_k} = \frac{\sum_{d \in W(v)} \sum_{v' \in L_d} \left[\frac{\gamma(d, k, \Theta^t)}{\alpha(v, v', k, \Theta^t)} + \sum_{k' \neq k} \frac{\theta^t \gamma(d, k', \Theta^t)}{\alpha(v', v, k, \Theta^t)} \right]}{\sum_{d \in W(v)} |L_d|} \quad (7)$$

$$+ \frac{\sum_{d \in L(v)} \sum_{v' \in W_d} \left[\frac{\theta^t \gamma(d, k, \Theta^t)}{\alpha(v', v, k, \Theta^t)} + \sum_{k' \neq k} \frac{\gamma(d, k', \Theta^t)}{\alpha(v, v', k, \Theta^t)} \right]}{\sum_{d \in W(v)} |L_d|}$$

Fix $v \in V$ and $u \in U$, rearranging Equ.6, we have the solution for $\frac{\partial \tilde{Q}(\Theta^t)}{\partial \theta} = 0$ as:

$$\theta = \frac{(K-1)\Sigma_d |W_d| |L_d|}{\Sigma_d \Sigma_k \gamma(d, k, \Theta^t) \Sigma_{w,v} [\frac{v_k}{\alpha(w,v,k,\Theta^t)} + \Sigma_{k' \neq k} \frac{w_{k'}}{\alpha(w,w,k',\Theta^t)}]} \quad (8)$$

Finally, the EM algorithm is summarized below:

Input: A set $D = \{d\}$ of sessions, where $u(d) = u$ and $d = \{< w, v >, w \in W_d, v \in L_d\}$ is a set of skyline objects w and normal objects v
Output: Model parameters $u \in U, v \in V, \theta$

```

1 Randomly initialize  $u \in U, v \in V, \theta > 1$ ;
2 for  $t \leftarrow 1$  and not converge do
3    $\gamma(d, k, \Theta^t) = \frac{u_k \Pi_{w \in W_d, v \in L_d} \frac{w_k}{w_k + \theta^t v_k} \Pi_{k' \neq k} \frac{\theta^t w_{k'}}{v_{k'} + \theta^t w_{k'}}}{\Sigma_{k=1}^K u_k \Pi_{w \in W_d, v \in L_d} \frac{w_k}{w_k + \theta^t v_k} \Pi_{k' \neq k} \frac{\theta^t w_{k'}}{v_{k'} + \theta^t w_{k'}}}$ ;
4   for  $u \in U$  do
5     for  $k \leftarrow 1$  to  $K$  do
6       | Compute  $u_k$  by Equ. 5;
7     end
8   end
9   for  $v \in V$  do
10    for  $k \leftarrow 1$  to  $K$  do
11      | Compute  $v_k$  by Equ. 7;
12    end
13  end
14  Normalize  $V$ ;
15   $\theta^{t+1} = \frac{\Sigma_d |W_d| [\Sigma_{v \in L_d} |v^{t+1}|]}{\Sigma_d (|W_d| |L_d|)}$ ;
16 end
17 return  $\Theta = \{\theta, U, V\}$ ;
```

Algorithm 1: EM algorithm for Model

3 DISCUSSION

In this section, we prove that under the skyline assumption, the purchases over the item space will have a long tail distribution. To simplify the argument, let $\theta = 0$ and each session to be of length $|V|$ where only one skyline object is selected. Then the probability of item v being selected in an arbitrary session would be

$$p(v) = \Pi_k v_k$$

Suppose for each item, the feature specific value v_k is drawn independently and identically from an arbitrary probability distribution. Applying the central limit theorem in the log domain, we know that for large K , the above probability will be a log-normal distribution, say $p(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp -\frac{(\ln x - \mu)^2}{2\sigma^2}$. The probability of item v being purchased for k times, is defined as

$$p(v) = \Pi_k v_k$$

3.1 Recommendation

The first step is to estimate the most possible aspect that the user is interested on, given the current browsing history \tilde{d} . We have

$$p(a = k | \tilde{d}, V) = \frac{\Sigma_{v \in \tilde{d}} v_k}{\Sigma_k \Sigma_{v \in \tilde{d}} v_k} \quad (9)$$

And the possibility of generating a skyline object $w \in \tilde{d}$ is

$$p(w | \tilde{d}, \theta, V, U, g) = \Pi_{k=1}^K [\frac{w_k}{w_k + \theta v_k} \frac{g_k}{v_k + \theta w_k}^{1-g_k}] \quad (10)$$

4 ORDINAL UTILITY

5 EXPERIMENT

5.1 Experimental setup

5.2 Comparative Performance of Decision Rules

5.3 Comparative Performance of Utility

6 CONCLUSION

7 RELATED WORK

7.1 Rank Aggregation

Due to its wide applications in economics, biomedicines and social science, rank aggregation has been studied extensively in those communities. Researchers within these communities are often asked to find a consensus ranking, given a set of individual rankings over different alternatives. Consider a voting process in social science, each voter demonstrates a preference list over political candidates, the goal is to achieve an optimal ranking to which each voter agrees in a sense. In completing such a goal, there are usually two types of methods in literature. The first type is directly based on the ordinal data. Most commonly, the ordinal data is employed to construct a graph over alternatives, and a permutation of alternative positions can be achieved explicitly by finding the minimum feedback arc set [?], in which case the resulting ranking has minimal Kendall Tau distance to all inputs; or by finding the stationary access distribution in a random walk framework [?], in which case the result approximates the MLE estimator of the BTL model. The second type is driven by random utility theory, as it assumes that a real-valued utility score is associated with each alternate, and the individual ranking is regarded as noisy observation of the ground truth ranking, which is the ordering of utility scores. The objective is thus transformed to inferencing the utility scores.

In social choice and biomedical communities, popular random utility models include the Bradley-Terry-Luce model (BTL for short) [?], the Plackett-Luce model (PL) [?], Mallows models [?], and so on. In the BTL model, suppose each individual is assigned a score s_i , the probability of pair-wise comparison $i > j$ is $p(i > j) = \frac{s_i}{s_j}$. In the PL model, the probability for a ranking $r = s_1 > s_2 \dots > s_M$ is $p(r|s) = \frac{s_1}{\Sigma_{l=1}^M s_l} \times \frac{s_2}{\Sigma_{l=2}^M s_l} \dots \times \frac{s_{M-1}}{s_{M-1} + s_M}$. The pair-wise comparison probability is further studied in [?]. BTL can be regarded as a special case of PL via marginalization. In the mallows model, the probability of ranking $p(r|\sigma, \phi) = \frac{1}{Z} e^{-\lambda d(r, \sigma)}$, where σ is the hidden true ranking, λ is the negative lognormal of a dispersion parameter, and Z is the normalizer. Other extensions are also

proposed under the random utility assumption, e.g. exponential distribution family is introduced in [?].

We should notice that usually there is only one single ranking as output. It is believed that any rank aggregation requires some degree of compromise. The group nature of individual rankings is brought to attention in a recent study [?]. In this study, a two stage strategy is adopted, where the inputs are clustered in a projected space, and then the true state is inferred in each cluster.