Quantifying Aspect Bias in Ordinal Ratings using a Bayesian Approach

Lahari Poddar        Wynne Hsu        Mong Li Lee
School of Computing  National University of Singapore
{lahari, whsu, leeml}@comp.nus.edu.sg

Abstract
User opinions expressed in the form of ratings can influence an individual’s view of an item. However, the true quality of an item is often obfuscated by user biases, and it is not obvious from the observed ratings the importance different users place on different aspects of an item. We propose a probabilistic modeling of the observed aspect ratings to infer (i) each user’s aspect bias and (ii) latent intrinsic quality of an item. We model multi-aspect ratings as ordered discrete data and encode the dependency between different aspects by using a latent Gaussian structure. We handle the Gaussian-Categorical non-conjugacy using a stick-breaking formulation coupled with Polya-Gamma auxiliary variable augmentation for a simple, fully Bayesian inference. On two real world datasets, we demonstrate the predictive ability of our model and its effectiveness in learning explainable user biases to provide insights towards a more reliable product quality estimation.

1 Introduction
With easy availability of information on the web, user ratings have become increasingly important in molding people’s perception of an item. However, an item typically has many aspects and not all aspects are equally important to all users. To some user, the cleanliness of a hotel is most important and he/she tends to rate this aspect stringently, but is lenient when rating food or amenities. Other users may have a different set of preferences and their aspect ratings for the same item could be vastly different. Hence, it is difficult to interpret conflicting ratings without knowing the underlying user biases. For an item with only few ratings this is aggravated, since even its average ratings are highly susceptible to the users’ biases.

To enable proper interpretation of ratings, we propose a unified probabilistic model for quantifying the underlying user biases for different aspects that lead to the observed ratings. We model the correlation between aspects by allowing a covariance structure among them. This is realistic since a user’s bias, and in turn his rating, of one aspect may be correlated with another aspect.

We detect the underlying aspect preferences of individual users that are consistent across their ratings on different items. People with a negative bias tend to be more critical about the aspect and generally underrate the aspect than other users, whereas people with a positive bias for an aspect tend to overrate it. Knowing the aspect biases of individuals, users can better interpret their ratings. Furthermore this is beneficial for service providers to focus on improving the aspects of an item that consumers truly care about.

While existing works assume ratings to be continuous, in reality most observed ratings in e-commerce websites are ordinal in nature. Our model incorporates the ordinal nature of observed ratings through proper statistical formulation. However, modeling the ordinal nature of observed ratings as well the correlation between aspects introduce non-conjugacy into our model, making Bayesian inference very challenging.

To eliminate the non-conjugacy of Gaussian-Categorical likelihood, we utilize stick-breaking formulation with Polya-Gamma auxiliary variable augmentation. The construction proposed in the paper is efficient and generic. It will help developing inference mechanisms for various applications that need to model ordinal data in terms of continuous latent variables with a correlation structure.

Figure 1: A sample restaurant’s ratings with color coded user aspect bias. This is a real output produced by the proposed model.

We can learn the aspect bias of users even with few ratings, by introducing latent user groups, based on the similarity of users’ rating behavior on various aspects. For example, one user group might generally give low ratings for ambiance while another user group gives high ratings for food.

Figure 1 shows an example application of the model where the learned user aspect bias is displayed beside the ratings. We can learn the aspect bias of users even with few ratings, by introducing latent user groups, based on the similarity of users’ rating behavior on various aspects. For example, one user group might generally give low ratings for ambiance while another user group gives high ratings for food.

Figure 1 shows an example application of the model where the learned user aspect bias is displayed beside the ratings. We can learn the aspect bias of users even with few ratings, by introducing latent user groups, based on the similarity of users’ rating behavior on various aspects. For example, one user group might generally give low ratings for ambiance while another user group gives high ratings for food.

Figure 1 shows an example application of the model where the learned user aspect bias is displayed beside the ratings. We can learn the aspect bias of users even with few ratings, by introducing latent user groups, based on the similarity of users’ rating behavior on various aspects. For example, one user group might generally give low ratings for ambiance while another user group gives high ratings for food.
Experiments on two real world datasets from TripAdvisor and OpenTable demonstrate that the proposed model provides new insights in users’ rating patterns, and outperforms state-of-the-art methods for aspect rating prediction.

To the best of our knowledge, this is the first work to model ordinal aspect ratings parameterized by latent multivariate continuous responses, with a simple, scalable and fully Bayesian inference.

2 Ordinal Aspect Bias Model

In this section, we describe the design of our Ordinal Aspect Bias model and present a Bayesian approach for inference.

Suppose we have $J$ users and $I$ items. Let $R$ be the set of observed ratings where $r_{ij}$ is an $A$ dimensional vector denoting the rating of user $j$ for item $i$ on each of its aspects. Each $r_{ij}$ is a discrete value between 1 and $K$ corresponding to a $K$-level scale (poor to excellent). We assume that $r_{ij}$ arises from a latent multivariate continuous response $v_{ij}$ which is dependent on (i) the intrinsic quality of the item on the aspect and (ii) the bias of the user for the aspect.

The intrinsic quality of an item $z_i$ is an $A$ dimensional vector, drawn from a multivariate normal distribution, with mean $\mu$ and covariance matrix $\Sigma$. We use multivariate normal distribution to account for the correlation among the subsets of aspects of an item. For example, it is highly unlikely for a hotel to have excellent room quality but very poor cleanliness, but it is possible to have a good location and average food choices. Such correlations among subset of aspects are captured by the covariance matrix. The parameters $(\mu, \Sigma)$ are given a conjugate normal-inverse Wishart (NIW) prior.

The preference of a user for an aspect is captured by a bias vector $m_g$ of dimension $A$. If a user places great importance on a particular aspect (e.g. cleanliness), this will be reflected in his ratings across all hotels. In other words, his rating on the cleanliness aspect will tend to be lower than the majority’s rating for cleanliness on the same hotel. We cluster users with similar preferences into different user groups and associate a bias vector $m_g$ with each group. The membership of a user $j$ in a user group is denoted as $s_j$ where $s_j$ is drawn from a categorical distribution $\theta$ with Dirichlet prior parameter $\alpha$.

Given the intrinsic quality $z_i$ and bias $m_g$, the latent response $v_{ij}$ is drawn from a multivariate Gaussian distribution with $z_i + m_g$ as mean and a hyper-parameter $B$ as covariance. This is intuitive as a user’s response depends on the item’s intrinsic quality for an aspect as well as his own bias.

With the latent response $v_{ij}$, we sample the observed rating vector $r_{ij}$. Note that since the observed ratings are ordered and discrete, they should be drawn from a categorical distribution. However, the latent response $v_{ij}$ is given a multivariate Gaussian prior. In order to have a fully Bayesian inference, we need to transform this categorical distribution to a Gaussian form to exploit conjugacy. This is the central technical challenge for our proposed model.

We develop a stick-breaking mechanism with logit function to map the categorical likelihood to a binomial form. Thereafter, leveraging the recently developed Pólya-Gamma auxiliary variable augmentation scheme [13], the binomial likelihood is transformed to Gaussian, thus establishing conjugacy and enabling us to achieve an effective posterior inference. The generative process of the model is as follows:

1. Draw a multinomial group distribution $\theta$ from Dirichlet $\alpha$.
2. For each group $g \in 1, \cdots, G$ draw a bias offset $m_g$ from $N_A(0, \Lambda)$.
3. For each user $j \in 1, \cdots, J$ sample a group $s_j$ from Cat $\theta$.
4. For each item $i \in 1, \cdots, I$ sample an intrinsic rating $z_i$ from $N_A(\mu, \Sigma)$.
5. For each rating $r_{ij} \in R$
   (a) draw latent continuous rating $v_{ij}$ from $N_A(z_i + m_g, B)$
   (b) draw observed ordinal rating $r_{ij}$ from Cat $(SB(v_{ij}, c))$

where $SB(v_{ij}, c)$ refers to the stick-breaking parameterization of the continuous response $v_{ij}$ using cut-points $c$. Figure 2 shows the proposed graphical model using plate notation.

2.1 Stick-Breaking Likelihood

We first discuss how to map the categorical likelihood of $v_{ij}$, denoted as $Lik(v_{ij})$, to a binomial form.

Let $r_{ija}$ denote the observed ordinal rating of item $i$, by user $j$ on aspect $a$, and is drawn from a categorical distribution over $K$ categories. Since the categories are ordered, we utilize a stick-breaking parameterization for the probabilities $P(r_{ija} = k)$ where $k \in \{1, \cdots, K\}$. Suppose we have a unit length stick where the continuum of points on this stick represents the probability of an event occurring. If we break this stick at some random point $p$, then we have a probability mass function over two outcomes (with probabilities $p$ and $1 - p$). By breaking the stick multiple times, we obtain a probability mass function over multiple categories.

Let $c = [c_1, \cdots, c_{K-1}]$ be a cut-point vector where $c_1 < c_2 < \cdots < c_{K-1}$ represent the boundaries between the ordered categories. The probability of each ordinal rating $r_{ija}$ being assigned the categorical value $k$, is parametrized using a function of the covariate $n_{ija}^k = c_k - v_{ija}$. Then the probability of observing the vector of ratings $r_{ij}$ is a product of probabilities of observing each of the aspect ratings $r_{ija}$ given the values of $\eta_{ija}$. Hence the likelihood of $v_{ij}$ is:

$$Lik(v_{ij}) = P(r_{ij}|v_{ij}, c) = P(r_{ij}|\eta_{ija}) = \prod_{a=1}^{A} P(r_{ija}|\eta_{ija})$$

To squash $\eta_{ija}$ within $[0, 1]$ we use a sigmoid function on it denoted by $f(x) = \frac{1}{1+e^{-x}}$. Sigmoid function enables us to use Pólya-Gamma augmentation scheme [13] to handle the non-conjugacy subsequently. For identifiability, we set $f(\eta_{ija}^C) = 1$. The stick-breaking likelihood can be written as:

$$P(r_{ija} = k) = \prod_{k' < k} (1 - f(\eta_{ija}^k)) f(\eta_{ija})$$

![Figure 2: Ordinal Aspect Bias Model](image-url)
By encoding \( r_{ij} \) with a 1-of-\( K \) vector \( x_{ij}^k \) where
\[
x_{ij}^k = \begin{cases} 1 & \text{if } r_{ij} = k \\ 0 & \text{otherwise} \end{cases}
\]
we now rewrite the likelihood of \( v_{ij} \) in binomial terms:
\[
P(r_{ij} | \eta_{ij}) = P(x_{ij}^k | \eta_{ij}) = \prod_{k=1}^{K} \text{Binom}(x_{ij}^k | N_{ij}, f(\eta_{ij}^k))
\]
where
\[
N_{ij} = 1 - \sum_{k < k} x_{ij}^k
\]

### 2.2 Polya-Gamma Variable Augmentation

Next, we explain how to transform the binomial likelihood to a Gaussian form via Polya-Gamma (PG) auxiliary variable augmentation scheme. The integral identity at the heart of the PG augmentation is:
\[
\frac{e^{\psi_a}}{(1 + e^{\psi})^2} = 2^{-b} e^{\psi} \int_0^\infty e^{-\omega^{2/2}} p(\omega) d\omega
\]
where \( \kappa = a - b/2, b > 0 \) and \( \omega \sim PG(b, 0) \).

By expanding the binomial likelihood in Eqn. 4, we get
\[
P(x_{ij}^k | \eta_{ij}) = \prod_{k=1}^{K} \sum_{a=1}^{K} \left( \frac{N_{ij}^k e^{\psi_a}}{e^{\psi_a} + \omega} \right) \prod_{k=1}^{K} \left( \frac{e^{\psi_a} + \omega}{e^{\psi_a} + \omega} \right)^{N_{ij}^k - \eta_{ij}^k}
\]
Using the integral identity of PG augmentation, we can now rewrite the categorical likelihood of \( v_{ij} \) as:
\[
L_i(k_j) = \prod_{a=1}^{A} P(x_{ij}^k | \eta_{ij})
\approx \prod_{a=1}^{A} \prod_{k=1}^{K-1} e^{\kappa_{ij}^k - \kappa_{ij}^a} \int_0^\infty e^{-\omega_{ij}^k (\omega_{ij}^k)^2/2} p(\omega_{ij}^k) d\omega_{ij}^k
\]
where \( \kappa_{ij}^k = x_{ij}^k - N_{ij}^k/2, \omega_{ij}^k = \eta_{ij}^k \) and \( p(\omega_{ij}^k) \) is PG\( (N_{ij}^k/2, 0) \) independent of \( \omega_{ij}^a \).

By property of PG distribution [13], we can draw the auxiliary variable \( \omega_{ij}^k \) from \( PG(N_{ij}^k, \eta_{ij}^k) \). Conditioning on \( \omega_{ij}^k \), \( L_i(k_j) \) can be transformed to a Gaussian form:
\[
L_i(k_j) \propto \prod_{a=1}^{A} \prod_{k=1}^{K} \left( \frac{2^{K-1} \pi^{K-1}}{\Gamma(K-1)} \right)^{1/2} \left( 1 + \frac{\kappa_{ij}^k - \kappa_{ij}^a}{\omega_{ij}^k} \right)^{K-1} \prod_{k=1}^{K} \left( \frac{1}{\omega_{ij}^k} \right)^{(K-1)/2}
\]

### 2.3 Bayesian Inference

Finally, we describe the sampling of user groups \( s \), bias offset of user groups \( m \), intrinsic ratings \( z \), cut-points \( c \) and latent continuous ratings \( v \) using fully Bayesian MCMC inference. We factor the joint probability of these variables as:
\[
P(r, v, m, z, s, c) = P(r | v, c) P(v | m, z, s) P(c) P(z) P(s) P(m)
\]

#### Sampling Bias Offset of User Groups

For each user group \( g \), we sample its bias offset \( m_g \) from the Gaussian posterior:
\[
P(m_g | A, v, z) \propto P(m_g | A) \prod_{j \in J[g]} \prod_{i \in I[j]} P(v_{ij} | m_g, z_i, B)
\]
where \( J[g] \) is the set of users belonging to group \( g \) and \( I[j] \) is the subset of items rated by user \( j \).

Since the prior is a multivariate Gaussian \( N_A(0, \Lambda) \) and the observations \( v_{ij} \) are also drawn from a multivariate Gaussian \( N_A(z_i + m_g, B) \), the posterior of \( m_g \) is given by a Gaussian \( N_A(m_g, \Sigma_g) \) with
\[
m_g = \Sigma_g (B^{-1} \sum_{j \in J[g]} \sum_{i \in I[j]} (v_{ij} - z_i))
\]

#### Sampling Intrinsic Ratings

Similar to the bias offsets of user groups, we sample intrinsic rating \( z_i \) of each item \( i \) from a Gaussian distribution \( N_A(\mu_i, \Sigma_i) \) where
\[
\mu_i = \Sigma_i (B^{-1} \sum_{j \in I[i]} (v_{ij} - m_{ij}) + \Sigma^{-1} \mu)
\]
\[
\Sigma_i = (n_i B^{-1} + \Sigma)^{-1}
\]
where \( n_i \) is the total number of ratings observed for users belonging to group \( g \).

#### Sampling Latent Continuous Ratings

The latent continuous ratings \( v_{ij} \) have a Gaussian prior \( N_A((z_i + m_{ij}), B) \) and a categorical likelihood \( P(r_{ij} | v_{ij}) \). We have transformed the categorical likelihood to the conditional Gaussian form (recall Eqn. 3). The posterior can be formulated as:
\[
P(v_{ij} | m_{ij}, z_i, B) \ast L_i(k_j) \omega_{ij}^k \]
\[
\propto \prod_{k=1}^{K} \left( \frac{1}{\omega_{ij}^k}(\kappa_{ij}^k - \kappa_{ij}^a)^2 \right) \prod_{k=1}^{K} \left( \frac{1}{\omega_{ij}^k}\right)^{(K-1)/2}
\]
\[
\times e^{\psi_a} \int_0^\infty e^{-\omega_{ij}^k (\omega_{ij}^k)^2/2} p(\omega_{ij}^k) d\omega_{ij}^k
\]
where \( \kappa_{ij}^k, \omega_{ij}^k \) are vectors of dimension \( A \), \( \Omega_{ij}^k \) is a diagonal matrix of \( \omega_{ij,1}^k, \omega_{ij,2}^k, \ldots, \omega_{ij,J}^k \).

Here, we assume the values in the \( A \)-dimensional cut-point vector \( c_k \) are all equal to \( c_k \). In practice, if we need different cut-points for different aspects, \( c_k \) can be set accordingly.
Since both the prior and likelihood are now Gaussian, we have the following Gibbs sampler:

\[ \mathbf{v}_{ij} \sim \mathcal{N}_d(\mu_{ij\omega}, \Sigma_{ij\omega}) \]

\[ \omega_{ij} \sim PG(\mathbf{N}_{ij\omega}, \mathbf{v}_{ij} - c) \]

where

\[ \mu_{ij\omega} = \mathbf{B}^{-1}(\mathbf{z}_i + \mathbf{m}_j) + \sum_{k=1}^{K-1} \Omega_{ij}^k (c_k - \omega_{ij}^k) / \omega_{ij}^k \]

\[ \Sigma_{ij\omega} = \mathbf{B}^{-1} + \sum_{k=1}^{K-1} \Omega_{ij}^k \]

**Sampling Cut-Points.** Sigmoid function in the stick-breaking formulation allows us to sample cut-points while ensuring their relative order without additional constraints. Figure 3 shows probability distributions for simulated cut-points.

![Probability Distributions for Simulated Cut-Points](image)

**Figure 3:** Category probabilities for cut-points (-5,-1,2,7)

The following lemma gives the relationship between cut-points, latent continuous ratings, and the observed ratings.

**Lemma 2.1.** If \( v_{ij} > c_k - \ln((1 - e^{-(c_{k+1} - c_k)}) \), then \( P(r_{ij} = k + 1) > P(r_{ij} = k) \).

**Proof.** Let \( \delta_k \geq -\ln((1 - e^{-(c_{k+1} - c_k)}) \). By replacing \( v_{ij} \) with \( (c_k + \delta_k) \) in Eqn. 2, we have

\[ P(r_{ij} = k) = \prod_{q<k} (1 - f(q - c_k - \delta_k)) (f(c_k - c_k - \delta_k)) \]

\[ P(r_{ij} = k + 1) = \prod_{q<k} (1 - f(q - c_k - \delta_k)) (1 - f(\delta_k)) (f(c_{k+1} - c_k - \delta_k)) \]

Taking the ratio, we have

\[ \frac{P(r_{ij} = k + 1)}{P(r_{ij} = k)} = \frac{(1 - f(\delta_k)) (f(c_{k+1} - c_k - \delta_k))}{f(\delta_k)} = \frac{(1 + e^{\delta_k} + \frac{1}{1 + e^{\delta_k} + \frac{1}{1 + e^{\delta_k} + \ldots}})}{\frac{1}{1 + e^{\delta_k} + \frac{1}{1 + e^{\delta_k} + \frac{1}{1 + e^{\delta_k} + \ldots}}}} \]

Since \( \delta_k \geq -\ln((1 - e^{-(c_{k+1} - c_k)}) \), we see that \( 1 + e^{\delta_k} + \frac{1}{1 + e^{\delta_k} + \frac{1}{1 + e^{\delta_k} + \ldots}} > 1 \). Hence, \( P(r_{ij} = k + 1) > P(r_{ij} = k) \).

Hence, given the sampled values of \( v_{ij} \) we can constrain the possible set of values for the cut-points. We sample cut-point \( c_k \) from a uniform distribution within the range:

\[ c_k \sim U[\max(v_{ij}, \arg \max P(r_{ij} | v_{ij}, k^*) = k) - \ln(1 - e^{-c_k - c_k - 1}), \min(v_{ij}, \arg \max P(r_{ij} | v_{ij}, k^*) = k + 1) - \ln(1 - e^{-c_k - c_k - 1})] \]

## 3 Experiments

For evaluation we use hotel ratings from TripAdvisor and restaurant ratings from Opentable.com. We crawled OpenTable.com for all the restaurant ratings in New York Tri-State area. Table 1. shows the details of the datasets.

| Dataset       | # Items | # Users | # Ratings | Aspects rated |
|---------------|---------|--------|-----------|---------------|
| TripAdvisor   | 12,773  | 781,203| 1,621,956 | Service, Value, Room, Location |
| OpenTable     | 2805    | 1997   | 73,469    | Ambience, Food, Service, Value |

**Table 1:** Statistics of experimental datasets.

### 3.1 Rating Prediction

One application of Ordinal Aspect Bias model is predicting observed aspect ratings. We perform five-fold cross validation on user-item pairs, and take expected value of an aspect rating as the predicted rating. Note that all the aspect ratings for the same user-item pair will be in the same training or test set. By default, the number of user groups are set to 10. For comparison, we also implemented the following models:

- **Continuous Aspect Bias model** is the continuous variant of our model where observed ratings are assumed to be continuous. Observed ratings are drawn from a (conjugate) multivariate Gaussian distribution, with mean as the true rating of the item offset with the bias of the user’s group.

- **Ordinal and Continuous No Bias model** assume users are not biased. The observed ratings for an item are drawn from only the true rating of the item.

- **Ordinal and Continuous Global Bias model** assume all users have the same bias. All ratings for an item are drawn from the true rating of the item offset with a global bias.

| Model                      | TripAdvisor Data | OpenTable Data |
|----------------------------|------------------|----------------|
| log LL                     | 1.20             | 1.03           |
| RMSE                       | 556.76           | 549.25         |
| Continuous Aspect Bias     | -1.00            | -1.03          |
| Continuous No Bias         | -1.00            | -1.03          |
| Continuous Global Bias     | -1.00            | -1.03          |
| Ordinal Aspect Bias        | -1.00            | -1.03          |
| Ordinal No Bias            | -1.00            | -1.03          |
| Ordinal Global Bias        | -1.00            | -1.03          |

**Table 2:** Test set log likelihood (the higher, the better) and RMSE (the lower, the better). All comparisons are statistically significant (paired t-test with \( p < 0.0001 \)).

**Table 2** shows mean log likelihood and RMSE (root mean square error) on test data. For both datasets Ordinal Aspect Bias model performs the best, demonstrating the need to consider both user bias and the proper ordinal nature of ratings.

Next, we compare the performance of our model with state-of-the-art rating prediction models, namely, PMF, BPMF, URP, SVD++, and BHFree. For each of these, we used the best parameter settings published.
### 3.2 Evaluation of User Groups

A significant advantage of our model is that it can infer latent user groups depending on their rating behaviors across multiple items. In this set of experiments, we show that if users are assigned to the same group, then their ratings on the same items for the same aspects are similar.

We look at the standard deviation of the set of users belonging to the same group who have rated the same entity. For each aspect of each item, we compare the standard deviation of the ratings of each user group with that of a control group comprising of all the users who have rated the item.

Figure 4 shows the scatter plots of the standard deviations for both datasets. We observe that most of the points lie above the line $y = x$, indicating that users who belong to the same group have smaller standard deviation compared to the control group. This implies that the latent user groups obtained by the proposed model can effectively cluster users who give similar aspect ratings to the same item.

| Model       | TripAdvisor Data | OpenTable Data |
|-------------|------------------|----------------|
|             | RMSE | FCP  | RMSE | FCP  | RMSE | FCP  | RMSE | FCP  |
| PMF         | 2.006 | 0.501 | 1.915 | 0.526 | 1.886 | 0.392 | 2.127 | 0.603 |
| BPMF        | 1.414 | 0.586 | 1.373 | 0.571 | 1.314 | 0.614 | 1.209 | 0.651 |
| URP         | 1.179 | 0.489 | 1.156 | 0.515 | 1.194 | 0.513 | 1.001 | 0.492 |
| SVDD++      | 1.064 | 0.578 | 1.079 | 0.562 | 1.093 | 0.639 | 0.894 | 0.665 |
| BHFree      | 1.143 | 0.553 | 1.199 | 0.582 | 1.124 | 0.624 | 1.007 | 0.671 |
| LARA        | 1.193 | 0.576 | 1.221 | 0.531 | 1.087 | 0.558 | 1.170 | 0.672 |
| OrdRec + SVDD++ | 1.348 | 0.619 | 1.344 | 0.613 | 1.359 | 0.654 | 1.173 | 0.702 |
| AspectBias  | 1.067 | 0.646* | 1.063* | 0.645* | 1.045 | 0.678* | 0.854* | 0.717* |

| Method       | TripAdvisor Data | OpenTable Data |
|-------------|------------------|----------------|
| PMF         | 0.016 | 0.142 |
| BPMF        | 0.233 | 0.177 |
| URP         | 0.364 | 0.201 |
| SVDD++      | 0.359 | 0.205 |
| BHFree      | 0.289 | 0.152 |
| LARA        | 0.148 | 0.262 |
| OrdRec + SVDD++ | 0.404 | 0.298 |

**Table 4:** Pearson’s correlation of aspect ranking.

The relative ranking of aspects for a user-item pair is also important to understand which aspects of an item the user liked better. For different methods, Table 4 shows the Pearson correlation coefficient of aspect ranking for a user-item pair, compared to its ground truth ranking. Clearly, Ordinal Aspect Bias model outperforms all other methods for the task of relative ranking of aspects. This validates that our model is able to learn aspect rating behavior of users accurately.

### 3.3 Intrinsic Quality of Items

Often one forms a judgment about the quality of an item by the average rating it has received. However, if an item has
received only a few ratings, it is difficult to form an accurate opinion concerning its quality. In this set of experiments, we show that the intrinsic quality, learned by the proposed model, is correlated with users’ perception of the item’s true quality, even for items with few ratings.

We focus on items with less than 30 ratings and whose intrinsic quality and average rating for an aspect differ by at least 0.5. Since an item’s true quality is unknown, we estimate it by the relative difference in the observed ratings of the same user on a pair of items. This is because if the qualities of two items are similar, a user will rate them similarly.

For each pair of items rated by the same user on the same aspect, let their difference in observed ratings be $\Delta \text{obs}$, difference between their average ratings be $\Delta \text{avg}$ and difference between the learned intrinsic ratings be $\Delta \text{int}$. Figure 6 shows the correlation between $\Delta \text{obs}$ and $\Delta \text{int}$, as well as the correlation between $\Delta \text{obs}$ and $\Delta \text{avg}$ aggregated over all aspects. We observe that for both datasets, as $\Delta \text{int}$ increases, $\Delta \text{obs}$ also increases. However, $\Delta \text{avg}$ remains almost constant. This indicates that $\Delta \text{obs}$ is closely correlated with $\Delta \text{int}$, whereas $\Delta \text{avg}$ appears to be independent of $\Delta \text{obs}$. This confirms that the learned intrinsic rating is better able to reflect users’ perception of the true quality of an item compared to using average ratings of the items.

![Figure 6: Correlation with $\Delta \text{obs}$](image)

### 3.4 Case Study

Finally, we present the reviews of a user from OpenTable to demonstrate that the aspect bias learned by our model correlates with their review texts (see Figure 7). The user is from group G2 in Figure 6, that is particularly critical about Value.

From the reviews of this user, as well as the reviews of randomly selected users from other groups for the same item, we see that the user from group 2 is indeed critical. We further confirm this observation by manually going through 100 randomly sampled reviews and tabulate the sentiment distribution of each item. We observe that the user is consistently critical even though the majority opinion is positive. This strengthens the fact that the group bias captured by our model is accurate and can help us better interpret a users’ rating.

### 4 Related Work

Existing works on aspect rating prediction use reviews to analyze latent aspect ratings [19, 20] and ignore the explicit aspect ratings provided by users. While the widely used CF approaches for rating prediction view ratings as continuous values and do not encode aspect dependencies [8, 6, 4, 9, 12, 16, 15, 11, 21].

There have been very few attempts to address the ordinal nature of ratings. The authors in [17] develop a model combining CF and content-based filtering using regression to handle ordinal ratings as a special case. The work in [7] proposes a wrapper around a CF method for ordinal data. Both of these works use a logit model for ordinal regression.

In contrast, most statistical approaches handle ordinal data using an ordinal probit model [11, 14, 10]. Although they allow a Bayesian inference but it necessitates using truncated Gaussian distributions and forced ordering of cut-off points. This leads to complicated and even sub-optimal inference.

The authors of [18] used stick-breaking formulation to parameterize the underlying continuous rating. However, since the non-conjugacy made an MCMC sampling non-trivial, they performed an approximate variational Bayesian inference. For correlated topic models [3], Pólya-Gamma auxiliary variable augmentation is used with logistic-normal transformation, whereas the work in [5] used stick-breaking likelihood for categorical data. However, none of these works use stick-breaking likelihood with a Pólya-Gamma variable augmentation to exploit conjugacy to facilitate Gibbs sampling.

### 5 Conclusion

We have presented a novel approach to understand users’ aspect bias, while capturing aspect dependencies as well as the proper ordinal nature of user responses. Our construction of the stick-breaking likelihood coupled with Pólya-Gamma auxiliary variable augmentation has resulted in an elegant Bayesian inference of the model.

Empirical evaluation on two real world datasets demonstrates that through proper statistical modeling of data we are able to capture users’ rating behavior and outperform state-of-the-art approaches. Furthermore, our model is effective in user modeling, analyzing users’ aspect preferences and provides a better product quality estimation even when the product has received few ratings. Most importantly, the construction of the model described here is generic and presents new possibilities for modeling such data in a wide-range of domains. Our work is orthogonal to works involving texts and
social graph of rating domains and it will be interesting to know the connection between bias groups and social groups.

References

[1] J. H. Albert and S. Chib. Bayesian analysis of binary and polychotomous response data. Journal of the American statistical Association, 1993.

[2] N. Barbieri. Regularized gibbs sampling for user profiling with soft constraints. In Advances in Social Networks Analysis and Mining (ASONAM), 2011.

[3] J. Chen, J. Zhu, Z. Wang, X. Zheng, and B. Zhang. Scalable inference for logistic-normal topic models. In Advances in Neural Information Processing Systems, 2013.

[4] T. Hofmann. Collaborative filtering via gaussian probabilistic latent semantic analysis. In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, 2003.

[5] M. E. Khan, S. Mohamed, B. M. Marlin, and K. P. Murphy. A stick-breaking likelihood for categorical data analysis with latent gaussian models. In AISTATS, 2012.

[6] Y. Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In ACM SIGKDD, 2008.

[7] Y. Koren and J. Sill. Collaborative filtering on ordinal user feedback. In IJCAI, 2013.

[8] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In Advances in neural information processing systems, 2001.

[9] B. M. Marlin. Modeling user rating profiles for collaborative filtering. In Advances in neural information processing systems, 2003.

[10] S. Muthukumarana and T. B. Swartz. Bayesian analysis of ordinal survey data using the dirichlet process to account for respondent personality traits. Communications in Statistics-Simulation and Computation, 2014.

[11] D. M. Pennock, E. Horvitz, S. Lawrence, and C. L. Giles. Collaborative filtering by personality diagnosis: A hybrid memory-and model-based approach. In Uncertainty in artificial intelligence, 2000.

[12] C. Pierre. Balancing prediction and recommendation accuracy: hierarchical latent factors for preference data. 2012.

[13] N. G. Polson, J. G. Scott, and J. Windle. Bayesian inference for logistic models using pòlya–gamma latent variables. Journal of the American statistical Association, 2013.

[14] P. E. Rossi, Z. Gilula, and G. M. Allenby. Overcoming scale usage heterogeneity: A bayesian hierarchical approach. Journal of the American Statistical Association, 2001.

[15] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In NIPS, 2007.

[16] R. Salakhutdinov and A. Mnih. Bayesian probabilistic matrix factorization using markov chain monte carlo. In International conference on Machine learning, 2008.
[17] D. H. Stern, R. Herbrich, and T. Graepel. Matchbox: large scale online bayesian recommendations. In Proceedings of the 18th international conference on World wide web, 2009.

[18] S. Virtanen and M. Girolami. Ordinal mixed membership models. In International Conference on Machine Learning, 2015.

[19] H. Wang and M. Ester. A sentiment-aligned topic model for product aspect rating prediction. In EMNLP, 2014.

[20] H. Wang, Y. Lu, and C. Zhai. Latent aspect rating analysis without aspect keyword supervision. In Proceedings of the 17th ACM SIGKDD, 2011.

[21] P. Wang, C. Domeniconi, and K. B. Laskey. Latent dirichlet bayesian co-clustering. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2009.