A Bayesian Latent Variable Model of User Preferences with Item Context

Aghiles Salah and Hady W. Lauw
School of Information Systems, Singapore Management University, Singapore
{asalah, hadywla} @ smu.edu.sg

Abstract

Personalized recommendation has proven to be very promising in modeling the preference of users over items. However, most existing work in this context focuses primarily on modeling user-item interactions, which tend to be very sparse. We propose to further leverage the item-item relationships that may reflect various aspects of items that guide users’ choices. Intuitively, items that occur within the same “context” (e.g., browsed in the same session, purchased in the same basket) are likely related in some latent aspect. Therefore, accounting for the item’s context would complement the sparse user-item interactions by extending a user’s preference to other items of similar aspects. To realize this intuition, we develop Collaborative Context Poisson Factorization (C2PF), a new Bayesian latent variable model that seamlessly integrates contextual relationships among items into a personalized recommendation approach. We further derive a scalable variational inference algorithm to fit C2PF to preference data. Empirical results on real-world datasets show evident performance improvements over strong factorization models.

1 Introduction

Recommender systems are essential in guiding users as they navigate the myriads of options offered by modern applications. They rely chiefly on information about which items users have consumed—rated, purchased, etc.—in the past, which can be represented as user-item preference matrix. The most prominent framework for recommendation is Matrix Factorization (MF) [Mnih and Salakhutdinov, 2008; Hu et al., 2008; Koren et al., 2009]. The principle is to decompose the preference matrix into low-dimensional user and item latent factor matrices. The bilinear combination of user and item’s latent factors can be used to predict unknown preferences.

Classical probabilistic MF models [Mnih and Salakhutdinov, 2008] typically assume that a user’s preference for an item is drawn from a Gaussian distribution centered at the inner product of their latent factor vectors. A distinct form of probabilistic MF, referred to as Poisson Factorization (PF) [Canny, 2004; Cemgil, 2009]—where the Poisson distribution is substituted for the usual Gaussian—recently demonstrates natural aptness for modeling discrete data such as ratings or purchases commonly found in recommendation scenarios. As documented in [Gopalan et al., 2015], thanks to the properties of the Poisson distribution, PF realistically models user preferences, fits well to sparse data, enjoys scalable variational inference with closed-form updates, and substantially outperforms previous state-of-the-art MF models based on Gaussian likelihoods [Mnih and Salakhutdinov, 2008; Shan and Banerjee, 2010; Koren et al., 2009].

Nevertheless, existing PF models for recommendation are primarily focused on user-item interactions, which are very sparse. A PF model that relies on user-item interactions alone may not necessarily associate similar items with similar representations in the latent space. This is due to the fact that such items are not necessarily rated by exactly the same users. Furthermore, on average, any given user may have had the opportunity to rate or purchase relatively few items. Thus, modeling and generalizing her preference across the large vocabulary of items based on the few user-item interactions alone is an onerous task. Fortunately, there are auxiliary information that could augment user-item interactions. One that we focus on in this paper is the contextual relationships among items.

Real-world behavior data often hold clues on how items may be related to one another. For instance, items found in the same shopping cart may work well together, e.g., shirt and matching pair of jeans. Items clicked or viewed on an e-commerce site in the same session may be alternatives for a particular need, e.g., shopping for a phone. Songs found in the same playlist probably share a coherent theme, e.g., country music of the 90s. As an abstraction of such scenarios, we introduce the notion of “context”, which may refer to a shopping cart, session, playlist, etc., depending on the specific problem instance. Intuitively, items that share similar contexts are implicitly related to one another in terms of some aspect that guides the choices one makes, such as specification, functionality, visual appearance, compatibility, etc. Note that contextual relatedness is not necessarily synonymous with feature-based similarity, e.g., shirt and jeans may share similar contexts, though they have different features.

The question is how to exploit and incorporate such contextual relationships among items within the PF framework. In this work, we posit that there could be two reasons that might explain the preference of a user for an item. The first
reason is that the user’s latent preference matches the latent attributes of the item of interest. The second reason is that the user’s latent preference matches those of other related items, i.e., those sharing similar contexts with the item of interest.

Based on the above assumption, we propose Collaborative Context Poisson Factorization (C²PF), a new Bayesian latent variable model of user preferences which takes into account contextual relationships between items; this is our first contribution. Under C²PF, the preference of a user for an item is driven by two components. One component is the interaction between the user’s and item’s latent factors, as in traditional PF. The other component consists of interactions between the user’s latent factor and item’s context latent factors. In this paper, “the context set of an item i” refers to the set of items sharing the same contexts (e.g., browsing sessions) with i. As the second contribution, we derive a scalable variational algorithm for approximate posterior inference, to fit our model to preference data. As the third contribution, through extensive experiments on six real-world datasets, we demonstrate the benefits of leveraging item context; C²PF noticeably improves upon the performance of Poisson factorization models, especially in the sparse scenario in which users express few ratings only.

2 Related Work

Given the breadth of scope of recommender systems in the literature [Bobadilla et al., 2013], we focus on those closely related to ours, to sharpen and clarify our contributions.

Approaches based on matrix factorization rely primarily on user-item interactions [Mnih and Salakhutdinov, 2008; Hu et al., 2008; Koren et al., 2009]. The sparsity of such information motivates the exploration of side information in several directions. On the users’ side, these include leveraging social networks [Ma et al., 2008; Zhou et al., 2012] or common features [Rao et al., 2015] to bring related users’ latent factors closer. On the items’ side, these include exploiting item content [Wang and Blei, 2011] or product taxonomy [Koenigstein et al., 2011] to pull together item latent factors.

In this work, we focus on item-item relationships. The closest such work to ours is [Park et al., 2017], which proposes Matrix Co-Factorization (MCF) model. The latter falls into the large class of collective matrix factorization [Singh and Gordon, 2008], which consists in jointly decomposing multiple data matrices, user-item and item-item matrices in MCF, with shared latent factors. This is a widely used approach in the recommendation literature to exploit different sources of data. The model we propose is radically different from MCF. First, here we investigate another architecture for leveraging item relationships with new modeling perspectives. More precisely, as opposed to collective MF-based models like MCF, in our approach, the user-item preferences are the only observations being factorized, and the auxiliary information (item-item relationships) is embedded into the model’s architecture. Second, MCF relies on the Gaussian distribution and uses stochastic gradient descent for learning, whereas our model builds on the Poisson distribution and enjoys scalable variational inference with closed-form updates. The benefits of our model are reflected in experiments.

In contrast, the CoFactor model [Liang et al., 2016] induces item relationships from the same user-item matrix, instead of a separate item-item matrix. It is also an instance of collective MF, relies on a Gaussian likelihood, and designed specifically for implicit feedback data [Hu et al., 2008].

Our model is also a novel contribution to the body of work on recommendation models based on Poisson factorization [Canny, 2004; Cemgil, 2009]. To our best knowledge, item context has not been explored within the PF framework.

Various other extensions of PF have been proposed. Gopalan et al. [2014a] develop Bayesian non-parametric PF, which does not require dimension of the latent factors to be specified in advance. Gopalan et al. [2014b] propose Collaborative Topic Poisson Factorization (CTPF) to model both article contents and reader preferences. CTPF is also an instance of collective MF and could be viewed as a “Poisson” alternative to MCF. Chaney et al. [2015] extend PF to incorporate social interactions. Charlin et al. [2015] propose a model which accounts for user and item evolution over time.

3 Collaborative Context Poisson Factorization

This section describes Collaborative Context Poisson Factorization (C²PF), a Bayesian latent variable model of user preferences that accounts for the item’s context.

Let \( \mathbf{X} = (x_{ui}) \) denote the user-item preference matrix of size \( U \times I \), where \( x_{ui} \) is the integer rating\(^1\) that user \( u \) gave to item \( i \), or zero if no preference was expressed. Let \( \mathbf{C} = (c_{ij}) \), of size \( I \times J \), be the item-context matrix, where \( c_{ij} = 1 \) if item \( j \) belongs to the context of item \( i \), and \( c_{ij} = 0 \) otherwise. Subsequently, we refer to \( j \) as the context item, and to the set of items \( j \) for which \( c_{ij} = 1 \) as the context of item \( i \).

C²PF builds on Poisson factorization [Friedman et al., 2001] to jointly model user preferences and leverage item’s context. Formally, C²PF represents each user \( u \) with a vector of latent preferences \( \theta_u \in \mathbb{R}_+^K \), each item \( i \) with a vector of latent attributes \( \beta_i \in \mathbb{R}_+^K \), and each context item \( j \) with a vector of latent attributes \( \xi_j \in \mathbb{R}_+^K \). C²PF also assumes additional latent variables \( \kappa_{ij} \in \mathbb{R}_+ \) for each observed item-context pair, that we shall discuss shortly. Conditional on these latent variables, the user preferences \( x_{ui} \) are assumed to come from a Poisson distribution as follows:

\[
x_{ui} \sim \text{Poisson}(\theta_u^\top \beta_i + \sum_j c_{ij} \kappa_{ij} \theta_u^\top \xi_j).
\]

The preference \( x_{ui} \) is affected by both how well the user \( u \)’s latent factors \( \theta_u \) matches the target item \( i \)’s latent factors \( \beta_i \), and how well \( \theta_u \) matches the context latent factors \( \xi_j \) of other items in \( i \)’s context. Given that \( i \) may have multiple context items, it is natural to expect that different context items may affect \( i \) to different degrees. This is the intuition behind each variable \( \kappa_{ij} \), which represents the effect a context item \( j \) has on item \( i \); we refer to these variables as the context effects.

The latent user preferences \( \theta_{uk} \), item attributes \( \beta_{ik} \), context item attributes \( \xi_{ik} \) and context effects \( \kappa_{ij} \) are all drawn from Gamma distributions. The Gamma is an exponential family distribution over positive random variables, governed

\(^1\)Other user-item interactions indicative of preferences are also possible, e.g., number of clicks.
by a shape and rate parameters [Bishop, 2006], which is a
conjugate prior to the Poisson distribution.
Moreover, in real-world data the items have very unbal-
anced context sizes, i.e., some have many more items in their
context set than others. To account for this diversity in context
size, C²PF assumes additional priors on the rate parameter of
the Gamma distribution over the context effects κ\(ij\), which
govern the average magnitude of the latter variables. This
induces a hierarchical structure over the κ\(ij\)’s that makes it possible to model item context more realistically.

The graphical model of C²PF is depicted in Figure 1, and
its generative process is as follows:

1. Draw user preferences: \(θ_{uk} \sim \text{Gamma}(α_θ^u, α_θ^r)\).
2. Draw context item attributes: \(ξ_{jk} \sim \text{Gamma}(α_ξ^s, α_ξ^r)\).
3. For each item \(i:\)
   a. Draw attributes: \(β_{ik} \sim \text{Gamma}(α_β^s, α_β^r)\).
   b. Draw the average magnitude of the context effects:
      \(γ_i \sim \text{Inverse-Gamma}(δ^s, δ^r)\).
   c. For each context item \(j\) of \(i\) draw a context effect: \(κ_{ij} \sim \text{Gamma}(α_{κ}^s, α_{κ}^r)\).
4. For each user-item pair \((u,i)\) sample a preference:
\(x_{ui} \sim \text{Poisson}(θ_{ui} β_{i} + \sum_{j} c_{ij} κ_{ij} ξ_{jk})\).

Note that C²PF includes as special cases other simpler
models, such as the original Bayesian PF, which can be de-
riveried by modifying C²PF’s specific components. In experi-
ments, we consider some of such simpler variants of C²PF.

In practice, we are given \(X\) and \(C\), and we are interested in
reversing the above generative process so as to infer the poste-
rrior distribution of the latent user preferences, context effects,
item and context item attributes, i.e., \(p(θ, β, ξ, γ, Z, Z^c|ν)\), with a factorized form, i.e., the lat-
tent variables are assumed to be independent and each gov-
erned by its own variational parameters, as follows:
\[
q(ν|μ) = \prod_{u,k} q(θ_{uk}|X^θ_u) \prod_{i,k} q(β_{ik}|X^β_i) \prod_{j,k} q(ξ_{jk}|X^ξ_{jk}) \prod_{i,j} q(κ_{ij}|X^κ_{ij}) \prod_{i,j} q(γ_i|η_i) \prod_{i,j} q(z_{ui}^c, z_{ui}^c|φ_{ui}),
\]
where \(ν = \{λ, η, φ\}\). The form of each factor in the above
equation is specified by the corresponding complete condi-
tional: the conditional distribution of each variable given the
other variables and observations. That is, the factors over
the Gamma variables are also Gamma distributions with vari-
ational parameters \(λ\), e.g., \(X^θ_u = (λ^θ_{uk}^1, λ^θ_{uk}^2)\), the superscripts \(s\) and \(r\) refer to the shape and rate parameters.
The factors over the Inverse-Gamma variables \(γ_i\) are also
Inverse-Gamma distribution with shape \(s\) and scale \(sc\) variational parameters,
e.g., \(η_i = (η^s_i, η^{sc}_i)\). Finally, the factors over \(z_{ui} = (z_{ui}^s, z_{ui}^c)\)
are Multinomial distributions with free parameters \(φ_{ui}\). The
latter result follows from the additive property of Poisson random
variables [Kingman, 1993], i.e., if \(x_1 \sim \text{Poisson}(λ_1)\), \(x_2 \sim \text{Poisson}(λ_2)\) and \(x = x_1 + x_2\), then \(x \sim \text{Poisson}(λ_1 + λ_2)\).

Given the variational family \(q\), VI is to fit its parameters by
solving the following optimization problem:
\[
ν^* = \arg\min_{ν} KL(q(ν|μ)||p(ν|X, C))
\]
This equation makes it clear how the observed data, $X$ and $C$, enter the variational distribution. Once $\nu^*$ is found, we use $q(\cdot | \nu^*)$ as a surrogate to the true posterior to compute the prediction in (2) and subsequently make recommendations.

Coordinate ascent learning. We derive an efficient coordinate ascent mean-field algorithm to solve the optimization problem (4). The principle is to alternate the update of each variational parameter while holding the others fixed. Iterating on such updates is guaranteed to monotonically decrease the KL in (4), and to converge into a locally optimal solution.

Thanks to the auxiliary variables, our model is conditionally conjugate. That is, each complete conditional is in the exponential family [Ghahramani and Beal, 2001; Blei et al., 2017]. Thereby, each coordinate update can be performed in closed form, by setting the expected parameter equal to the expected natural parameter (w.r.t. $q$) of the corresponding complete conditional. This is indeed the optimal update for the variational parameter.

The complete conditional for the user preference, $p(\theta_{uk} | \cdot)$, is a Gamma with shape and rate parameters given by:

\[
(\alpha_\beta + \sum_i z_{ui}^x + \sum_{j} z_{uijk}, \beta_\beta + \sum_{i} c_{ij} \nu_{ij} \xi_{jk}).
\]

The complete conditionals for the other Gamma variables are:

\[
p(\beta_{ik} | \cdot) = \text{Gamma}(\alpha_{\beta} + \sum_i z_{uijk}, \alpha_{\beta} + \sum_{u} \theta_{uk}). \tag{5}
\]

\[
p(\xi_{jk} | \cdot) = \text{Gamma}(\alpha_{\xi} + \sum_{u,i} z_{uijk}, \alpha_{\xi} + \sum_{u,i} c_{ij} \nu_{ij} \theta_{uk}). \tag{6}
\]

\[
p(\nu_{ij} | \cdot) = \text{Gamma}(\alpha_{\nu} + \sum_{u,k} z_{uijk}, \alpha_{\nu} + \sum_{u,k} \theta_{uk} \xi_{jk}). \tag{7}
\]

The complete conditional for the average intensity of the context effect is as follows:

\[
p(\gamma_{i} | \cdot) = \text{Inv-Gamma}(\delta^\theta + \alpha_{\theta} \sum_{j} c_{ij}, \delta^\nu + \alpha_{\nu} \sum_{j} \nu_{ij}). \tag{8}
\]

The complete conditional for the auxiliary variables is:

\[
p(z_{ui} | \theta, \beta, \xi, \kappa, C, X) = \text{Multinomial}(x_{ui}, \log p_{ui}), \tag{9}
\]

where $z_{ui} = (z_{ui}^x, z_{ui}^c)$, $p_{ui} = (p_{ui}^x, p_{ui}^c)$ is a point on the $(K + K x c)$-simplex, and for all $k,j$: $p_{uijk}^x \propto \theta_{uk} \beta_{ik}$ and $p_{uijk}^c \propto c_{ij} \nu_{ij} \theta_{uk} \xi_{jk}$.

The expected natural parameters (w.r.t. $q$) of these conditionals give the optimal updates for the variational parameters, e.g., the update for Gamma variational parameter $\lambda_{uk}^\theta$ is obtained by taking expectation of (5), which yields:

\[
\lambda_{uk}^\theta = \alpha_{\beta} + \sum_i x_{ui}(\delta_{ui}^x + \sum_j \phi_{uijk}), \quad \lambda_{uk}^\nu = \alpha_{\nu} + \beta_{\nu} \sum_{i,j} c_{ij} \nu_{ij} \phi_{uijk}. \tag{10}
\]

where we have used the standard results about the expectation of Gamma and Multinomial random variables. That is, if $\theta \sim \text{Gamma}(\lambda^\theta, \lambda^\nu)$, then $\mathbb{E}(\theta) = \frac{\lambda^\theta}{\lambda^\nu}$, and if $z_{ui} \sim \text{Multinomial}(x_{ui}, \phi_{ui})$, then the expectation of the $k$th component of $z_{ui}$ is $\mathbb{E}(z_{ui}) = x_{ui} \phi_{ui}$. Using the standard results of the expectation of the log of a Gamma variable, i.e., $\mathbb{E}(\log \theta) = \psi(\lambda^\theta) - \log \lambda^\nu$ with $\psi(\cdot)$ denoting the digamma function, the updates for the components of the variational parameter $\phi_{ui} = (\phi_{ui}^x, \phi_{ui}^c)$ are:

\[
\phi_{uijk}^x \propto \exp \left( \psi(\lambda_{uijk}^\theta) - \log \lambda_{uijk}^\nu + \psi(\lambda_{uijk}^\nu - \log \lambda_{uijk}^\theta) \right). \tag{11}
\]

\[
\phi_{uijk}^c \propto \exp \left( \mathbb{E}(\log \theta_{uk}) + \mathbb{E}(\log \xi_{jk}) + \mathbb{E}(\log \kappa_{ij}) \right), \tag{12}
\]

for brevity we did not develop the expectations in (13).

The updates for the remaining variational parameters can be derived in the same way. The full variational inference for C$^2$PF is depicted in Algorithm 1.

---

**Algorithm 1** Variational inference for C$^2$PF.

**Input:** X, C, K, $\delta$, $\alpha_{\theta}$, $\alpha_{\nu}$, $\kappa$, $\nu$.

**Output:** The set of variational parameters $\nu^*$.

**Steps:**

1. Initialization: $\eta_{i}^\theta = \delta^\theta + c_{ij} \times \alpha_{\nu}$, randomly initialize the remaining Gamma variational parameters $\lambda^\theta, \lambda^\nu$.

2. For each observed preference $x_{ui}$, update the variational Multinomial parameter $\phi_{ui}$ using equations (12) and (13).

3. Update the user related parameters, $\forall u, k$: $\lambda_{uk} = \alpha_{\beta} + \sum_i x_{ui} \phi_{ui}^c + \sum_{i,j} x_{ui} \phi_{uijk}^c$.

4. Update the item related parameters, $\forall v, k$: $\lambda_{vk} = \alpha_{\beta} + \sum_i \xi_{jk} \phi_{uijk}^c$.

5. Update the context item related parameters, $\forall v, s$: $\lambda_{vs} = \alpha_{\beta} + \sum_{i,j} \nu_{ij} \phi_{uijk}^c$.

6. Update the context parameters, $\forall u, j$, such that $c_{ij} > 0$: $\eta_{ij}^\nu = \delta^\nu + \sum_{k,j} \lambda_{i,j}^\nu x_{uk} \phi_{uijk}^c$.

7. Until convergence.

---

**Efficient implementation.** A key property of the variational C$^2$PF algorithm is efficiency. The operations involving users and items need to be carried out for only the non-zero elements in $X$ and $C$. Furthermore, we can avoid explicitly computing and storing the Multinomial parameters $\phi$. We need to store only the following matrices, $L_\theta = \{\exp(\mathbb{E}_q(\log \theta_{uk}))\}$, $L_\beta = \{\exp(\mathbb{E}_q(\log \beta_{ik}))\}$, $L_\xi = \{\exp(\mathbb{E}_q(\log \xi_{jk}))\}$ and $L_\kappa = \{\exp(\mathbb{E}_q(\log \kappa_{ij}))\}$. We can then use these quantities directly in the updates of the variational shape parameters.

**Computational time complexity.** The Proposition below shows that the computational complexity of Algorithm 1 scales linearly with the number of non-zero entries in $X$ and $C$. In practice $X$ and $C$ are extremely sparse, and Algorithm 1 converges within 100 iterations. Furthermore, the updates of the variational parameters are trivially parallelizable across users and items, hence our variational inference for C$^2$PF can easily scale to large datasets.

**Proposition 1.** Let $nz_x$ and $nz_c$ denote respectively the number of non-zero in $X$ and $C$. The computational complexity per iteration of Algorithm 1 is $O(K \cdot (nz_x + nz_c + U + I))$.

**Proof.** The computation bottleneck of Algorithm 1 is with the update blocks 3 to 6. The computational complexity of
updating $\lambda_{uk}^{\theta,s}$ is $O(nz_w^{uk})$, such that $nz_w^{uk}$ is the number of ratings expressed by user $u$. This complexity holds since the sum over $j$ can be precomputed once for each $i, k$ and stored in a $I \times K$ matrix, the total cost of this operation is $O(K \cdot nz_c)$. The complexity of updating all $\lambda_{uk}^{\theta,r}$ parameters is $O(K \cdot (I + U + nz_c))$. Therefore, the computational complexity of block 3 is $O(K \cdot (nz_x + nz_c + U + I))$.

Similarly, we can show that the complexity of block 4 is $O(K \cdot (nz_x + U + I))$ and that of blocks 5 and 6 is $O(K \cdot (nz_x + nz_c + U + J))$. Putting it all together, the complexity per iteration of Algorithm 1 is $O(K \cdot (nz_x + nz_c + U + I))$, where we have assumed that $I$ is of the same order as $J$.\]

5 Experimental Study

Our objective is to study the impact of item context, and our modeling assumptions, on personalized recommendation.

5.1 Datasets

We use six datasets from Amazon.com, provided by McAuley et al.: McAuley et al. [2015b; 2015a]. These datasets include both the user-item preferences and the “Also Viewed” lists that we treat as the item contexts. We preprocess all datasets so that each user (resp. item) has at least ten (resp. two) ratings, and the sets of items and context items are identical. Table 1 describes the resulting datasets.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>#Users</th>
<th>#Items</th>
<th>#Ratings</th>
<th>nz_X (%)</th>
<th>nz_C (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office</td>
<td>3,703</td>
<td>6,523</td>
<td>53,282</td>
<td>0.22</td>
<td>0.25</td>
</tr>
<tr>
<td>Grocery</td>
<td>8,938</td>
<td>22,890</td>
<td>148,735</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>Automotive</td>
<td>7,280</td>
<td>15,635</td>
<td>63,477</td>
<td>0.05</td>
<td>0.15</td>
</tr>
<tr>
<td>Sports</td>
<td>19,049</td>
<td>24,095</td>
<td>211,582</td>
<td>0.04</td>
<td>0.09</td>
</tr>
<tr>
<td>Pet Supplies</td>
<td>16,462</td>
<td>20,049</td>
<td>164,017</td>
<td>0.05</td>
<td>0.16</td>
</tr>
<tr>
<td>Clothing</td>
<td>41,809</td>
<td>97,619</td>
<td>420,377</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 1: Statistics of the Datasets.

5.2 Comparative Models

We benchmark our model, C²PF, against strong comparable generative factorization models.

- **MCF**: Matrix Co-Factorization [Park et al., 2017], which incorporates item-to-item relationships into Gaussian MF.
- **PF**: Bayesian Poisson Factorization [Gopalani et al., 2015] which arises as a special case from our model without the item context. Therefore, we can effectively assess the impact of the item context by comparing C²PF to PF.
- **CTPF**: Collaborative Topic Poisson Factorization [Gopalani et al., 2014b] is a co-factorization approach that jointly models user preferences and item topics. It can also be used to leverage the item context by substituting the item-context matrix $C$ for the item-word matrix.
- **CoCTPF**: Content-only CTPF [Gopalani et al., 2014b] is a variant of CTPF without the document topic offsets; please refer to [Gopalani et al., 2014b] for details.

Note that the above baselines have been found to perform better than several other models on the task of item recommendation. To examine the contributions of our modeling choices, we also include the results for two simplified variants of C²PF:

- **rC²PF**: reduced C²PF that drops the item factors $\beta_i$, resulting in a simpler model where only the context part in (1) is responsible for explaining the user preferences $x_{ui}$.
- **tC²PF**: tied C²PF that constrains the context factors $\xi$ to be the same as the item factors $\beta_i$, that is $\xi_i = \beta_i$ for all $i$.

5.3 Experimental Settings

For each dataset, we randomly select 80% of the ratings as training data and the remaining 20% as test data. Random selection is carried out five times independently on each dataset. The average performance over the five samples is reported as the final result.

For most experiments, we set the number of latent components $K$ to 100. Later, we will also vary $K$ and indeed find 100 to be a good trade-off between accuracy and model complexity. To encourage sparse latent representations, we set $\alpha_k = \alpha_\beta = \alpha_\xi = (0.3, 0.3)$—resulting in exponentially shaped Gamma distributions with mean equal to $1$. We further set $\delta = (2, 5)$ and $\alpha_\zeta = 2$, fixing the prior mean over the context effects to 0.5. Note that we set $\alpha_\zeta > 1$ to avoid sparse distributions over the $\kappa_{ij}$ variables and thereby encourage C²PF to rely on item’s context to explain user preferences. For an illustration, please refer to Figure 2 in [Bogadilla et al., 2013].

To set the different hyperparameters of MCF, we follow the same strategy, grid search, as in [Park et al., 2017].

5.4 Evaluation Metrics

We assess the recommendation accuracy on a set of held-out items—the test set. We retain four widely used measures for top-$M$ recommendation, namely the Normalized Discount Cumulative Gain (nDCG), Mean Reciprocal Rank (MRR), Precision@$M$ (P@$M$) and Recall@$M$ (R@$M$), where $M$ is the number of items in the recommendation list [Bogadilla et al., 2013]. Intuitively, nDCG and MRR measures the ranking quality of a model, while Precision@$M$ and Recall@$M$ assess the quality of a user’s top-$M$ recommendation list. These measures vary from 0.0 to 1.0 (higher is better).

5.5 Empirical Results and Discussion

Table 2 depicts the average performances of the various competing models in terms of different metrics, over all datasets. In order to ease interpretation, we provide another presentation of Recall@20 in Figure 2—the results are consistent across all metrics.

We note that C²PF, and its variants, substantially outperforms the other competing models on all datasets and across

---

1http://jmcauley.ucsd.edu/data/amazon/
all measures. Recall that without the item context information $C^{2\text{PF}}$ degenerates to the basic $\text{PF}$. We can therefore attribute the performance improvements reached by $C^{2\text{PF}}$, relative to $\text{PF}$, to the modeling of the item context. The importance of the item context is also strongly supported by the high performance of $rC^{2\text{PF}}$ relative to $\text{PF}$, though $rC^{2\text{PF}}$ relies solely on item's context to make recommendations.

Overall, the results from Table 2 suggest that the item context underlies different aspects of items that explain the user behaviour. To gain further insights into the performance of the proposed model and the impact of our modeling choices, we now delve into specific research questions.

- **Q1. How important is the Poisson distribution?**
  We observe that even though $\text{PF}$ does not leverage the relationships among items, it still outperforms $\text{MCF}$ in most cases. This provides empirical evidence that the Poisson distribution is a better alternative to Gaussian in modeling user preferences.

- **Q2. How important are the $C^{2\text{PF}}$'s modeling assumptions?**
  $\text{CTPF}$ and $\text{CoCTPF}$ offer alternative $\text{PF}$-based architectures to $C^{2\text{PF}}$ for leveraging item's context, with different modeling assumptions. More precisely, $\text{CTPF}$ and $\text{CoCTPF}$ fall into the class of collective matrix factorization, and consist in jointly factorizing the user-item $X$ and item-context $C$ matrices, with shared item factors. This is a popular strategy in the recommendation literature to model different sources of data. The proposed models, $C^{2\text{PF}}$ and its variants, substantially outperform $\text{CTPF}$ and $\text{CoCTPF}$ in all cases, demonstrating the benefits of the assumptions behind $C^{2\text{PF}}$.

- **Q3. Why does $\text{CoCTPF}$ performs better than $\text{CTPF}$?**
  $\text{CoCTPF}$ arises as a special case from $\text{CTPF}$ without the item offset. Surprisingly, the former performs better than the latter. A careful investigation reveals that the magnitudes of the item offsets (noted $\epsilon$ in the original paper) tend to be bigger than those of the shared item attributes $\theta$. This means that, in $\text{CTPF}$, the item offsets, which are specific to

![Figure 2: Comparison of average Recall@20 over different datasets.](image-url)

<table>
<thead>
<tr>
<th>Data</th>
<th>Metric</th>
<th>MCF</th>
<th>PF</th>
<th>CTPF</th>
<th>CoCTPF</th>
<th>$C^{2\text{PF}}$</th>
<th>$L^{2\text{PF}}$</th>
<th>$C^{2\text{PF}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office</td>
<td>nDCG</td>
<td>0.1286</td>
<td>0.1568</td>
<td>0.1553</td>
<td>0.1717</td>
<td>0.1722</td>
<td>0.1776</td>
<td>0.1763</td>
</tr>
<tr>
<td></td>
<td>MRR</td>
<td>0.0145</td>
<td>0.0452</td>
<td>0.0429</td>
<td>0.0529</td>
<td>0.0590</td>
<td>0.0595</td>
<td>0.0585</td>
</tr>
<tr>
<td></td>
<td>R@20</td>
<td>0.0122</td>
<td>0.0343</td>
<td>0.0358</td>
<td>0.0440</td>
<td>0.0478</td>
<td>0.0519</td>
<td>0.0500</td>
</tr>
<tr>
<td></td>
<td>R@20</td>
<td>0.0106</td>
<td>0.0329</td>
<td>0.0341</td>
<td>0.0461</td>
<td>0.0480</td>
<td>0.0509</td>
<td>0.0500</td>
</tr>
<tr>
<td>Groceries</td>
<td>nDCG</td>
<td>0.1286</td>
<td>0.1568</td>
<td>0.1553</td>
<td>0.1717</td>
<td>0.1722</td>
<td>0.1776</td>
<td>0.1763</td>
</tr>
<tr>
<td></td>
<td>MRR</td>
<td>0.0120</td>
<td>0.0100</td>
<td>0.0103</td>
<td>0.0337</td>
<td>0.0394</td>
<td>0.0387</td>
<td>0.0392</td>
</tr>
<tr>
<td></td>
<td>R@20</td>
<td>0.0122</td>
<td>0.0343</td>
<td>0.0358</td>
<td>0.0440</td>
<td>0.0478</td>
<td>0.0519</td>
<td>0.0500</td>
</tr>
<tr>
<td></td>
<td>R@20</td>
<td>0.0122</td>
<td>0.0343</td>
<td>0.0358</td>
<td>0.0440</td>
<td>0.0478</td>
<td>0.0519</td>
<td>0.0500</td>
</tr>
<tr>
<td>Automotive</td>
<td>nDCG</td>
<td>0.1286</td>
<td>0.1568</td>
<td>0.1553</td>
<td>0.1717</td>
<td>0.1722</td>
<td>0.1776</td>
<td>0.1763</td>
</tr>
<tr>
<td></td>
<td>MRR</td>
<td>0.0120</td>
<td>0.0100</td>
<td>0.0103</td>
<td>0.0337</td>
<td>0.0394</td>
<td>0.0387</td>
<td>0.0392</td>
</tr>
<tr>
<td></td>
<td>R@20</td>
<td>0.0122</td>
<td>0.0343</td>
<td>0.0358</td>
<td>0.0440</td>
<td>0.0478</td>
<td>0.0519</td>
<td>0.0500</td>
</tr>
<tr>
<td></td>
<td>R@20</td>
<td>0.0122</td>
<td>0.0343</td>
<td>0.0358</td>
<td>0.0440</td>
<td>0.0478</td>
<td>0.0519</td>
<td>0.0500</td>
</tr>
<tr>
<td>Sports</td>
<td>nDCG</td>
<td>0.1286</td>
<td>0.1568</td>
<td>0.1553</td>
<td>0.1717</td>
<td>0.1722</td>
<td>0.1776</td>
<td>0.1763</td>
</tr>
<tr>
<td></td>
<td>MRR</td>
<td>0.0120</td>
<td>0.0100</td>
<td>0.0103</td>
<td>0.0337</td>
<td>0.0394</td>
<td>0.0387</td>
<td>0.0392</td>
</tr>
<tr>
<td></td>
<td>R@20</td>
<td>0.0122</td>
<td>0.0343</td>
<td>0.0358</td>
<td>0.0440</td>
<td>0.0478</td>
<td>0.0519</td>
<td>0.0500</td>
</tr>
<tr>
<td></td>
<td>R@20</td>
<td>0.0122</td>
<td>0.0343</td>
<td>0.0358</td>
<td>0.0440</td>
<td>0.0478</td>
<td>0.0519</td>
<td>0.0500</td>
</tr>
<tr>
<td>Pet Supplies</td>
<td>nDCG</td>
<td>0.1286</td>
<td>0.1568</td>
<td>0.1553</td>
<td>0.1717</td>
<td>0.1722</td>
<td>0.1776</td>
<td>0.1763</td>
</tr>
<tr>
<td></td>
<td>MRR</td>
<td>0.0120</td>
<td>0.0100</td>
<td>0.0103</td>
<td>0.0337</td>
<td>0.0394</td>
<td>0.0387</td>
<td>0.0392</td>
</tr>
<tr>
<td></td>
<td>R@20</td>
<td>0.0122</td>
<td>0.0343</td>
<td>0.0358</td>
<td>0.0440</td>
<td>0.0478</td>
<td>0.0519</td>
<td>0.0500</td>
</tr>
<tr>
<td></td>
<td>R@20</td>
<td>0.0122</td>
<td>0.0343</td>
<td>0.0358</td>
<td>0.0440</td>
<td>0.0478</td>
<td>0.0519</td>
<td>0.0500</td>
</tr>
<tr>
<td>Clothing</td>
<td>nDCG</td>
<td>0.1286</td>
<td>0.1568</td>
<td>0.1553</td>
<td>0.1717</td>
<td>0.1722</td>
<td>0.1776</td>
<td>0.1763</td>
</tr>
<tr>
<td></td>
<td>MRR</td>
<td>0.0120</td>
<td>0.0100</td>
<td>0.0103</td>
<td>0.0337</td>
<td>0.0394</td>
<td>0.0387</td>
<td>0.0392</td>
</tr>
<tr>
<td></td>
<td>R@20</td>
<td>0.0122</td>
<td>0.0343</td>
<td>0.0358</td>
<td>0.0440</td>
<td>0.0478</td>
<td>0.0519</td>
<td>0.0500</td>
</tr>
<tr>
<td></td>
<td>R@20</td>
<td>0.0122</td>
<td>0.0343</td>
<td>0.0358</td>
<td>0.0440</td>
<td>0.0478</td>
<td>0.0519</td>
<td>0.0500</td>
</tr>
</tbody>
</table>
the user-item interaction component, dominate the prediction of unknown preferences (please refer to equation 1 in [Gopalan et al., 2014b]).

- **Q4. When does C²PF offer the most improvements?**
In Figure 3, we report the performances, in terms of Recall@20, of C²PF and PF, on users with different number of ratings. C²PF consistently achieves the best performance over different scenarios. Though this may be data-dependent, C²PF seems to provide the most improvement on users with few ratings. The relative difference between C²PF and PF tends to decrease with more ratings. It is challenging to infer good user representations when there is a lack of information in the preference matrix. By leveraging additional signals from items’ contexts, C²PF mitigates this lack of information.

- **Q5. What is the impact of the number of factors on the performance of C²PF?**
In Figure 4, we report the performance of the different models, on Office, over different K. C²PF consistently outperforms the competing methods. It is not very sensitive to the value of K and seems to provide better performances when K ≥ 100. Because the complexity of the models increases with K, we recommend to set the number of factors to 100, which is a good tradeoff between recommendation quality and model complexity.

### 6 Conclusion & Perspectives

Based on the assumption that items sharing similar contexts are related in some latent aspect that guides one’s choices, we develop Collaborative Context Poisson Factorization (C²PF), a Bayesian latent factor model of user preferences which takes into account the contextual relationships among items. Under C²PF, not only do items (through latent attributes) contribute to explain user behaviour, but so do their contexts. Empirical results on real-world datasets show that C²PF noticeably improves the performance of Poisson factorization models, especially in the sparse scenario in which users express few ratings, suggesting that the item context underlies aspects of items that can explain the user preferences.

A flexible model with strong theoretical foundations, C²PF can be extended in several directions. For instance, it would be interesting to extend C²PF to account for user-user social relationships to further alleviate the sparsity issue. Another possible line of future work is to compose C²PF with other graphical models. For instance, one could combine C²PF and CTPF to jointly model item’s context and textual content.

### Acknowledgments

This research is supported by the National Research Foundation, Prime Minister’s Office, Singapore under its NRF Fellowship Programme (Award No. NRF-NRFF2016-07).
References


