A. Supplement to "A Listwise Approach to Collaborative Ranking"

A.1. Proofs in Theory section

Proof of Theorem 1. Notice that Π_{i1} is the argument, k, that minimizes Y_{ik}, and $\mathbb{P}\{\Pi_{i1} = k\} = \mathbb{P}\{Y_{ik} \leq \min\{Y_{ij}\}_{j\neq k}\}$. Furthermore, $\min\{Y_{ij}\}_{j\neq k}$ is exponential with rate parameter $\sum_{j\neq k}\phi(X_{ij})$ and is independent of Y_{ik}. Hence,

$$\begin{aligned}&\mathbb{P}\left\{Y_{ik} \leq \min\{Y_{ij}\}_{j\neq k}\right\}\\&= \int_0^\infty \phi(X_{ik})e^{-u\phi(X_{ik})}e^{-\sum_{j\neq k}u\phi(X_{ij})}\mathrm{d}u\\&= \frac{\phi(X_{ik})}{\sum_j \phi(X_{ij})}.\end{aligned}$$

Furthermore,

$$\begin{aligned}&\mathbb{P}\{\Pi_i|\Pi_{i1}\} = \mathbb{P}\{Y_{\Pi_{i2}} \leq \ldots \leq Y_{\Pi_{im}}|Y_{\Pi_{ij}} \geq Y_{\Pi_{i1}}, \forall j\}\\&= \mathbb{P}\{Y_{\Pi_{i2}} - Y_{\Pi_{i1}} \leq \ldots \leq Y_{\Pi_{im}} - Y_{\Pi_{i1}}|Y_{\Pi_{ij}} \geq Y_{\Pi_{i1}}, \forall j\}.\end{aligned}$$

By the memorylessness property, we have that the joint distribution of $Y_{\Pi_{i2}} - Y_{\Pi_{i1}}, \ldots, Y_{\Pi_{im}} - Y_{\Pi_{i1}}|Y_{\Pi_{ij}} \geq Y_{\Pi_{i1}}$, $\forall j > 1$ is equivalent to the distribution of $Y_{\Pi_{i2}}, \ldots, Y_{\Pi_{im}}$. Hence, we can apply induction with the previous argument, and the tower property of conditional probability. □

Proof of Lemma 1. By optimality,

$$\frac{1}{n}\sum_{i=1}^n -\log P_{\hat{X}_i}(\Pi_i) \leq \frac{1}{n}\sum_{i=1}^n -\log P_{X_i^\star}(\Pi_i).$$

Which is equivalent to

$$\frac{1}{n}\sum_{i=1}^n -\log\frac{P_{X_i^\star}(\Pi_i)}{P_{\hat{X}_i}(\Pi_i)} \geq 0.$$

Thus, we can subtract the expectation,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n \mathbb{E}\log\frac{P_{X_i^\star}(\Pi_i)}{P_{\hat{X}_i}(\Pi_i)}\\&\leq -\frac{1}{n}\sum_{i=1}^n\left(\log\frac{P_{X_i^\star}(\Pi_i)}{P_{\hat{X}_i}(\Pi_i)} - \mathbb{E}\log\frac{P_{X_i^\star}(\Pi_i)}{P_{\hat{X}_i}(\Pi_i)}\right)\end{aligned}$$

where the expectation $\mathbb{E}$ is with respect to the draw of Π_i conditional on X. □

Lemma 2. *Let π be a permutation vector and x be a score vector each of length m. Suppose that $|\log\phi(x_j)| \leq C$ for all $j = 1,\ldots,m$. Define the relative loss function,*

$$L_{x,x'}(\pi) := \log\frac{P_x(\pi)}{P_{x'}(\pi)}.$$

Consider translating an item in position ℓ to position ℓ' in the permutation π, thus forming π' where $\pi'_{\ell'} = \pi_\ell$. Specifically, $\pi'_k = \pi_k$ if $k < \min\{\ell,\ell'\}$ or $k > \max\{\ell,\ell'\}$; if $\ell < \ell'$ then $\pi'_k = \pi_{k+1}$ for $k = \ell,\ldots,\ell'-1$; if $\ell' < \ell$ then $\pi'_k = \pi_{k-1}$ for $k = \ell'+1,\ldots,\ell$. The relative loss function has bounded differences in the sense that

$$|L_{x,x'}(\pi) - L_{x,x'}(\pi')| \leq C_0\|\log\phi(x) - \log\phi(x')\|_\infty,$$

where $\phi(x)$ is applied elementwise and $C_0 = 2+e^{2C}\ln(m+1)$.

Proof. Suppose that $\ell < \ell'$, and define the following shorthand,

$$\lambda_j = \phi(x_{\pi_j}), \quad \Lambda_j = \sum_{k=j}^m \lambda_k,$$

and let λ'_j, Λ'_j be similarly defined with x'. Then by replacing the permutation π with π' causes the Λ_j to be replaced with $\Lambda_j - \lambda_j + \lambda_\ell$ for $j = \ell+1,\ldots,\ell'$. Hence,

$$\begin{aligned}&\log Q_x(\pi) - \log Q_x(\pi')\\&= \sum_{j=\ell+1}^{\ell'}\log(\Lambda_j - \lambda_j + \lambda_\ell) - \log(\Lambda_j)\\&= \sum_{j=\ell+1}^{\ell'}\log(\Lambda_{j-1} + \lambda_\ell) - \log(\Lambda_j)\\&= \sum_{j=\ell}^{\ell'-1}\log\left(1 + \frac{\lambda_\ell}{\Lambda_j}\right) + \log\Lambda_\ell - \log\Lambda_{\ell'}.\end{aligned}$$

So we can bound the difference,

$$\begin{aligned}|L_{x,x'}(\pi) - L_{x,x'}(\pi')| &\leq \left|\log\frac{\Lambda_\ell}{\Lambda'_\ell} - \log\frac{\Lambda_{\ell'}}{\Lambda'_{\ell'}}\right|\\&+ \sum_{j=\ell}^{\ell'-1}\left|\log\left(1+\frac{\lambda_\ell}{\Lambda_j}\right) - \log\left(1+\frac{\lambda'_\ell}{\Lambda'_j}\right)\right|.\end{aligned}$$

Suppose that for each j, $|\log\lambda_j - \log\lambda'_j| \leq \delta$ and that $|\log\lambda_j| \leq C$. Then we have that

$$\left|\log\frac{\Lambda_\ell}{\Lambda'_\ell}\right| \leq \max_{j\geq\ell}\left|\log\frac{\lambda_j}{\lambda'_j}\right| \leq \delta.$$

The same equation can be made for this term with ℓ'. Let

$$\alpha_j = \max\left\{\frac{\lambda_\ell}{\Lambda_j}, \frac{\lambda'_\ell}{\Lambda'_j}\right\}.$$

Then we have

$$\begin{aligned}&\left|\log\left(1+\frac{\lambda_\ell}{\Lambda_j}\right) - \log\left(1+\frac{\lambda'_\ell}{\Lambda'_j}\right)\right|\\&\leq \left|\log(1+\alpha_j) - \log(1+e^{-\delta}\alpha_j)\right| \leq \left|1-e^{-\delta}\right||\alpha_j|.\end{aligned}$$

Furthermore, because $\Lambda_j \geq (m-j+1)e^{-C}$ then $|\alpha_j| \leq (m-j+1)^{-1}e^{2C}$ and

$$\begin{aligned}|L_{x,x'}(\pi) - L_{x,x'}(\pi')| &\leq 2\delta \\ &+ \sum_{j=\ell}^{\ell'-1} |1-e^{-\delta}| \frac{e^{2C}}{m-j+1} \\ &\leq 2\delta + |1-e^{-\delta}| e^{2C} H_m \\ &\leq \delta(2 + e^{2C}\ln(m+1)).\end{aligned}$$

In the above equation, H_m is the mth harmonic number. A similar calculation can be made when $\ell' > \ell$. Setting $\delta = \|\log\phi(x) - \log\phi(x')\|_\infty$ concludes the proof. □

Proof of Theorem 2. Define the empirical process function to be

$$\rho_n(x) := \frac{1}{n}\sum_{i=1}^n \left(\log\frac{P_{X_i^\star}(\Pi_i)}{P_{X_i}(\Pi_i)} - \mathbb{E}\log\frac{P_{X_i^\star}(\Pi_i)}{P_{X_i}(\Pi_i)}\right).$$

By the listwise representation theorem, $\rho_n(x)$ is a function of $n\times m$ independent exponential random variables. Moreover, if we were to change the value of a single element y_{ik} then this would result in a change of permutation of the type described in Lemma 2. Notice that the bound on the change in the relative loss is $C_0\|\log\phi(X_i) - \log\phi(X_i')\|_\infty$, where $C_0 = 2 + e^{2C}\ln(m+1)$, and notice that the sum of squares of these bounds are,

$$\begin{aligned}&\sum_{i,k} C_0^2 \|\log\phi(X_i) - \log\phi(X_i')\|_\infty^2 \\ &= mC_0^2 \sum_{i=1}^n \|\log\phi(X_i) - \log\phi(X_i')\|_\infty^2 \\ &= mC_0^2\|Z - Z'\|_{\infty,2}^2,\end{aligned}$$

where Z, Z' are $\log\phi$ applied elementwise to X, X' respectively. By Lemma 2 and McDiarmid's inequality,

$$\mathbb{P}\{n(\rho_n(x) - \rho_n(x')) > \epsilon\} \leq \exp\left(-\frac{2\epsilon^2}{mC_0^2\|Z-Z'\|_{\infty,2}^2}\right).$$

Hence, the stochastic process $\{n\rho_n(X) : X \in \mathcal{X}\}$ is a subGaussian field with canonical distance,

$$d(X, X') := \sqrt{m}C_0\|Z - Z'\|_{\infty,2}.$$

The result follows by Dudley's chaining (Talagrand, 2006). □

Lemma 3. *If $\log\phi$ is 1-Lipschitz then we have that $g(\mathcal{Z}) \leq g(\mathcal{X})$.*

Proof. Let $X, X' \in \mathcal{X}$, and $Z = \log\phi(X), Z' = \log\phi(X')$. Then

$$|z_{ij} - z'_{ij}| \leq |x_{ij} - x'_{ij}|,$$

by the Lipschitz property. Hence, $\|Z - Z'\|_{\infty,2} \leq \|X - X'\|_{\infty,2}$, and so

$$\mathcal{N}(u, \mathcal{Z}, \|.\|_{\infty,2}) \leq \mathcal{N}(u, \mathcal{X}, \|.\|_{\infty,2}).$$

□

Proof of Corollary 1. Consider two matrices X, X' in the model (7) then

$$\|x_i - x_i'\|_\infty = \max_j |\beta^\top z_{ij} - \beta'^\top z_{ij}| \leq \|\beta - \beta'\|_2 \|Z_i\|_{2,\infty}.$$

Let $\zeta = \max_i \|Z_i\|_{2,\infty}$ then

$$\|X - X'\|_{\infty,2} \leq \zeta\|\beta - \beta'\|_2.$$

The covering number of $\mathcal{X}$ is therefore bounded by

$$\mathcal{N}(\mathcal{X}, u, \|.\|_{\infty,2}) \leq \mathcal{N}(B_{c_b}, u/\zeta, \|.\|_2) \leq \left(\frac{\zeta C_0 c_b}{u}\right)^s,$$

for an absolute constant C_0, where B_{c_b} is the ℓ_2 ball of radius c_b. The result follows by Lemma 3 and Theorem 2. □

Proof of Corollary 2. Let $X = UV^\top$ and $X' = U'V'^\top$ such that U, V, U', V' bounded by 1 in Frobenius norm (modifying the Frobenius norm bound does not change the substance of the proof). Consider

$$\begin{aligned}&|u_i^\top v_j - u_i'^\top v_j'| \leq |u_i^\top v_j - u_i'^\top v_j| + |u_i'^\top v_j - u_i'^\top v_j'| \\ &\leq \|u_i - u_i'\|\|v_j\| + \|u_i'\|\|v_j - v_j'\|.\end{aligned}$$

Maximizing this over the selection of j,

$$\begin{aligned}&\max_j |u_i^\top v_j - u_i'^\top v_j'| \\ &\leq \|u_i - u_i'\|_2\|V\|_{2,\infty} + \|u_i\|_2\|V - V'\|_{2,\infty}.\end{aligned}$$

Hence,

$$\begin{aligned}&\|X - X'\|_{\infty,2} \\ &\leq \|U - U'\|_F\|V\|_{2,\infty} + \|U'\|_F\|V - V'\|_{2,\infty} \\ &\leq \|U - U'\|_F + \|V - V'\|_{2,\infty}.\end{aligned}$$

Consider the vectorization mapping from the $m\times r$ matrix to the mr dimensional vectors. The Frobenius norm is mapped to the ℓ_2 norm, and we can consider the $2,\infty$ norm to be, the norm $\|x\|_\rho = \max_j \|(x_{jr+1}, \ldots, x_{j(r+1)})\|_2$. The ρ-norm unit ball (B_ρ) is just the Cartesian product of the ℓ_2 norm ball in K dimensions. The volume of a d-dimensional ball, V_d, is bounded by

$$C_l \leq \frac{V_d}{(e\pi)^{d/2}} d^{\frac{d}{2}} \leq C_u,$$

where $C_l < C_u$ are universal constants. So the volume ratio between the ℓ_2 norm ball and the ρ norm ball is bounded by

$$\frac{V(B_2)}{V(B_\rho)} \leq C\left(\frac{r^{r/2}}{(e\pi)^{r/2}}\right)^m / \left(\frac{(rm)^{rm/2}}{(e\pi)^{rm/2}}\right)$$
$$\leq Cm^{-rm/2},$$

where $C = C_u/C_l$.

$$\mathcal{N}(\epsilon, B_2, \|.\|_\rho) \leq C_r\left(\frac{2}{\epsilon}+1\right)^{rm} m^{-rm/2}$$
$$\leq C\left(\frac{3}{\epsilon\sqrt{m}}\right)^{rm},$$

for $\epsilon \leq 1$. This is also the covering number of the Frobenius norm ball in the $2,\infty$ norm. Moreover, we know that the covering number of the unit Frobenius norm ball in $n \times K$ matrices (B_F) in the Frobenius norm is

$$\mathcal{N}(\epsilon, B_F, \|.\|_F) \leq \left(\frac{c}{\epsilon}\right)^{nr},$$

for some constant c. Consider covering the space $\mathcal{X}$, by selecting centers U, V from the $\epsilon/2$-coverings of B_F in the F-norm and $2,\infty$ norm respectively. By the above norm bound, this produces an ϵ-covering in the $\infty, 2$ norm. Dudley's entropy bound is thus

$$\int_0^\infty \sqrt{\log\mathcal{N}(\epsilon, B_F, \|.\|_F) + \log\mathcal{N}(\epsilon, B_F, \|.\|_{2,\infty})}\mathrm{d}\epsilon$$
$$\leq \int_0^c \sqrt{nr\log(c/\epsilon)}\mathrm{d}\epsilon + \int_0^{3/\sqrt{m}} \sqrt{-mr\log(\epsilon\sqrt{m}/3)}\mathrm{d}\epsilon.$$

So that

$$\int_0^c \sqrt{nr\log(c/\epsilon)}\mathrm{d}\epsilon$$
$$\leq c'\sqrt{nr}$$

for some absolute constant c' and

$$\int_0^{3/\sqrt{m}} \sqrt{-mr\log(\epsilon\sqrt{m}/3)}\mathrm{d}\epsilon \leq \int_0^3 \sqrt{r\log(u/3)}\mathrm{d}u$$
$$\leq c'\sqrt{r}.$$

Hence, for $g(\mathcal{X}) \leq c'\sqrt{nr}$ and we have the result. □

A.2. Algorithms

Algorithm 3 Compute gradient for V when U fixed

```
Input: Π, U, V, λ, ρ
Output: g {g ∈ ℝ^{r×m} is the gradient for f(V)}
g = λ · V
for i = 1 to n do
  Precompute h_t = u_i^T v_{Π_it} for 1 ≤ t ≤ m̄ {For implicit
  feedback, it should be (1 + ρ) · m̃ instead of m̃, since
  ρ · m̃ 0's are appended to the back}
  Initialize total = 0, tt = 0
  for t = m̄ to 1 do
    total += exp(h_t)
    tt += 1/total
  end for
  Initialize c[t] = 0 for 1 ≤ t ≤ m̄
  for t = m̄ to 1 do
    c[t] += h_t · (1 − h_t)
    c[t] += exp(h_t) · h_t · (1 − h_t) · tt
    total += exp(h_t)
    tt −= 1/total
  end for
  for t = 1 to m̄ do
    g[:, Π_it] += c[t] · u_i
  end for
end for
Return g
```

Algorithm 4 Gradient update for V (Same procedure for updating U)

```
Input: V, ss, rate {rate refers to the decaying rate of
the step size ss}
Output: V
Compute gradient g for V {see alg 3}
V −= ss · g
ss *= rate
Return V
```

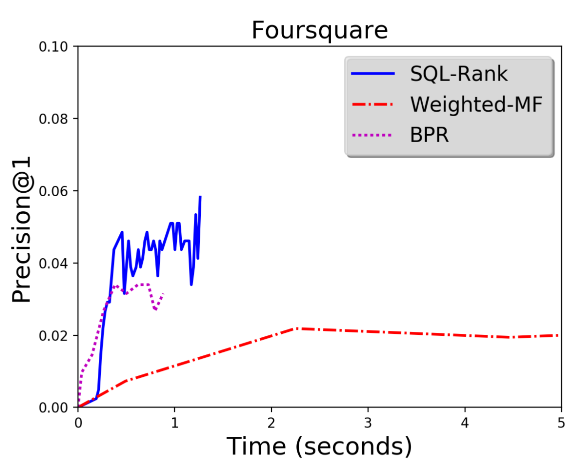

Figure 3. Comparing implicit feedback methods.

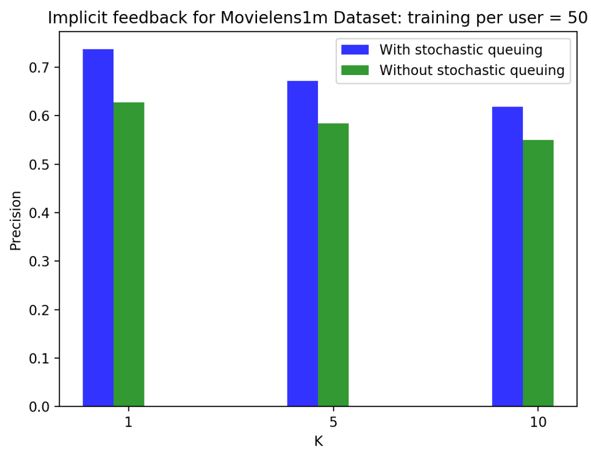

Figure 4. Effectiveness of Stochastic Queuing Process.

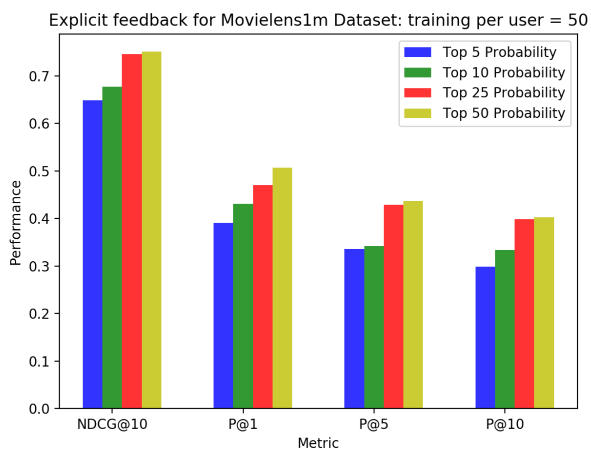

Figure 5. Effectiveness of using full lists.