Robust Asymmetric Recommendation via Min-Max Optimization

Peng Yang[†], Peilin Zhao[‡], Vincent W. Zheng[#], Lizhong Ding[†], Xin Gao[†]

[†]KAUST, Saudi Arabia; [‡]SCUT, China; [#]Advanced Digital Sciences Center, Singapore
{peng.yang.2, lizhong.ding, xin.gao}@kaust.edu.sa; peilinzhao@hotmail.com; vincent.zheng@adsc.com.sg

ABSTRACT

Recommender systems with implicit feedback (e.g. clicks and purchases) suffer from two critical limitations: 1) imbalanced labels may mislead the learning process of the conventional models that assign balanced weights to the classes; and 2) outliers with large reconstruction errors may dominate the objective function by the conventional L_2 -norm loss. To address these issues, we propose a robust asymmetric recommendation model. It integrates cost-sensitive learning with capped unilateral loss into a joint objective function, which can be optimized by an iteratively weighted approach. To reduce the computational cost of low-rank approximation, we exploit the dual characterization of the nuclear norm to derive a min-max optimization problem and design a subgradient algorithm without performing full SVD. Finally, promising empirical results demonstrate the effectiveness of our algorithm on benchmark recommendation datasets.

CCS CONCEPTS

• Information systems → Collaborative filtering; • Theory of computation → Semi-supervised learning;

KEYWORDS

Robust Asymmetric Learning; Concave-Convex Optimization

ACM Reference Format:

Peng Yang[†], Peilin Zhao[‡], Vincent W. Zheng[#], Lizhong Ding[†], Xin Gao[†]. 2018. Robust Asymmetric Recommendation via Min-Max Optimization. In *SIGIR '18: The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, July 8–12, 2018, Ann Arbor, MI, USA*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3209978.3210074>

1 INTRODUCTION

Recommender systems are attractive for content providers, through which they can increase sales or views. The user-item matrix, i.e., $Y \in \mathbb{R}^{n \times m}$, encodes the individual preference of users for items. In this work, we assume $m \leq n$. Most often user preference is not explicitly but implicitly expressed with binary (0 or 1) values. In the implicit feedback setting, the observed feedbacks marked as 1's

are positive instances, indicating that a user viewed or purchased an item, while the others annotated as 0's are a mixture of negatives and missing values. However, imbalanced labels and outliers often occur in the implicit feedback datasets. To tackle these problems, a large number of studies have been conducted, among which one-class collaborative filtering (CF) model [5, 8] treated all unlabeled values as negatives, which is biased in estimation since some missing values are positives. Rendle et al. [14] sampled a subset of negative instances from unlabeled values, which often decreases the predictive power of the model due to insufficient data coverage. To address the outlier issue, robust matrix decomposition [9][17] employed a sparse matrix to capture the noise so that the low-rank structure can be recovered. However, the proposed L_2 -norm loss function may be easily misled by the outliers and noisy data.

Although these methods are effective, they mainly rely on nuclear norm for the low-rank approximation. The optimization for nuclear norm minimization can be solved by first-order optimization methods such as subgradient descent [10] and proximal gradient descent [11]. Although these methods are guaranteed to converge, they are inefficient because a singular value decomposition (SVD) step, which takes $O(m^2n)$ time, is required at each iteration. To reduce the computational complexity, one efficient way is to replace the full SVD with an approximate SVD [3]. However, those approaches require the loss function to be smooth [7]. The alternative way relies on the bilinear factorization, which utilizes alternating direction methods to iteratively optimize the decomposed variables. However, these approaches break the convexity of the original problem, which results in no global convergence [8].

In this paper, we propose a robust asymmetric recommendation method via min-max optimization. To reduce the negative impact of imbalanced and noisy labels, we propose a capped L_p -norm asymmetric learning model. It derives an iteratively weighted algorithm to minimize the objective function. By leveraging the capped-norm and cost-sensitive learning, our model seamlessly integrates them into a joint framework, where the result of one task reinforces the other one. To reduce the computational cost of low-rank approximation, we develop an SVD-free optimization algorithm. Based on the dual characterization of the nuclear norm, we reformulate the problem with a min-max optimization problem and solve it by a subgradient method. In each iteration, we only need to compute the largest singular vector instead of a full SVD, thus reducing the time complexity from $O(m^2n)$ to $O(mn)$. Finally, the promising experimental results demonstrate the effectiveness of our algorithm on several benchmark recommendation datasets.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
SIGIR '18, July 8–12, 2018, Ann Arbor, MI, USA

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-5657-2/18/07...\$15.00
<https://doi.org/10.1145/3209978.3210074>

2 ALGORITHM

We assume that a 0-1 rating matrix $Y \in \{0, 1\}^{n \times m}$ with a bounded nuclear norm $\|Y\|_* \leq \epsilon$, where n and m are the numbers of items and users, respectively. Generally, only a subset of 1's of Y are observed. Let the observations Ω be a subset of positive entries sampled from $\{(i, j) | Y_{ij} = 1\}$. We define the observation matrix as A where $A_{ij} = 1$ if $(i, j) \in \Omega$, and $A_{ij} = 0$ otherwise. Our objective is that, given the observation matrix A , we aim to find a low-rank matrix X to minimize the loss between X and A :

$$\min_{X \in \mathbb{R}^{n \times m}} \sum_{i=1}^n \sum_{j=1}^m \ell(X_{ij}, A_{ij}) + \lambda \text{rank}(X), \quad (1)$$

where $\lambda \geq 0$ is a trade-off parameter, $\ell(X, A)$ is a loss function between the matrices X and A , and $\text{rank}(X)$ constrains X into a low-rank matrix.

2.1 Framework for Class Imbalance and Outlier

We introduce a framework to tackle class imbalance and outliers. **Class-Imbalance Learning:** In the 0-1 rating matrix, a small portion of positives is observed, which may result in biased estimation by the conventional models that treat each class equally. To address this issue, we assume that the cost of missing a positive target is much higher than that of having a false-positive, and study an appropriate performance metric: *weighted cost*,

$$\text{cost} = c_p \times M_p + c_n \times M_n, \quad (2)$$

where M_p and M_n are the numbers of false negatives and false positives, c_p and $c_n \in [0, 1]$ ($c_p + c_n = 1$) denote the predefined costs of false negatives and false positives, respectively. The lower the cost value, the better the performance.

LEMMA 1. *Minimizing the weighted cost in Eq. (2) is equivalent to minimizing the following objective:*

$$\sum_{A_{ij}=+1} \alpha \mathbb{I}(X_{ij} < q) + \sum_{A_{ij}=0} \mathbb{I}(X_{ij} \geq q), \quad (3)$$

where $\alpha = \frac{c_p}{c_n}$ is a biased coefficient, $q \in (0, 1)$ is a threshold and $\mathbb{I}_\pi$ is the indicator function that outputs 1 if π holds and 0 otherwise.

Lemma 1 gives the explicit objective for optimization, but the indicator function in Eq. (3) is not convex. Thus, we replace the indicator function by its convex surrogate,

$$\ell_\alpha(X_{ij}) = \left(\alpha \mathbb{I}_{(A_{ij}=1)} + \mathbb{I}_{(A_{ij}=0)} \right) \hat{\ell}(X_{ij}), \quad (4)$$

and $\hat{\ell}(\cdot)$ is a convex unilateral loss for the 0-1 classes,

$$\hat{\ell}(X_{ij}) = \begin{cases} [1 - X_{ij}]_+^2, & A_{ij} = 1 \\ [X_{ij}]_+^2, & A_{ij} = 0 \end{cases}$$

where $[\cdot]_+ = \max(\cdot, 0)$ is the hinge loss. We assume that $0 \leq c_n \leq c_p$, since we would prefer to improve the accuracy of the positive class. Fig. 1(a) illustrates the loss $\ell_\alpha(X_{ij})$ for $A_{ij} = 1$ with $\alpha = 2$. We observe that for $\ell_\alpha(X_{ij})$, the slope of the loss function changes for the positive class, leading to more "aggressive" updating. The cost-sensitive learning has been widely studied in [15]. Different from previous techniques, we derive a cost-sensitive unilateral loss for binary classification with 0-1 label feedback.

Capped L_p -norm Loss: A large squared loss occurs in Eq. (4) if the data point is not correctly classified. Outliers may result in such

Figure 1: Illustration of robust cost-sensitive loss

issues and mislead the learning process. To be resistant to outliers, we propose a robust cost-sensitive loss for $0 < p < 2$:

$$h(\ell_\alpha(X_{ij})) = \left(\alpha \mathbb{I}_{(A_{ij}=1)} + \mathbb{I}_{(A_{ij}=0)} \right) \min(\hat{\ell}(X_{ij})^{\frac{p}{2}}, \xi_\alpha), \quad (5)$$

where $h(\ell_\alpha(x))$ enforces a capped L_p -norm over the loss $\hat{\ell}(x)$ with an upper bound $\xi_\alpha > 0$. As shown in Fig. 1(b) ($\alpha = 2$, $p = 1$ and $\xi_\alpha = 0.5$), no matter how misclassified the data point is, the loss $h(\ell_\alpha(x))$ is capped by ξ_α . This makes the loss function robust to outliers since their effect to the model is fixed. However, optimizing Eq. (5) is difficult since $h(\ell_\alpha(x))$ is a concave non-smooth function in the domain of $\ell_\alpha(x)$. Motivated by concave duality [19], Lemma 2 provides an iteratively weighted function to solve it.

LEMMA 2. *Eq. (5) can be relaxed to iteratively minimizing a weighted L_2 -norm formulation:*

$$X^{t+1} = \underset{X}{\operatorname{argmin}} \sum_{i,j} \tilde{\beta}_{ij}^t \hat{\ell}(X_{ij}) (\alpha \mathbb{I}_{(A_{ij}=1)} + \mathbb{I}_{(A_{ij}=0)}), \quad (6)$$

where $\tilde{\beta} = \nabla_u h(u)|_{u=\hat{\ell}(x)}$ denotes the supergradient of the concave function $h(u)$ at point $u = \hat{\ell}(x)$ for $p \in (0, 2)$:

$$\tilde{\beta}_{ij}^t = \begin{cases} \frac{p}{2} \hat{\ell}(X_{ij}^t)^{\frac{p-2}{2}}, & \hat{\ell}(X_{ij}^t)^{\frac{p}{2}} \leq \xi_\alpha; \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

PROOF. Let $u = \hat{\ell}(x)$ and $h(u) = \min(u^{\frac{p}{2}}, \xi_\alpha)$. As $\hat{\ell}(x)$ is convex in the domain of x , $h(u)$ can be rewritten:

$$\min(u^{\frac{p}{2}}, \xi_\alpha) = \inf_{\tilde{\beta} \geq 0} [\tilde{\beta}u - h^*(\tilde{\beta})], \quad (8)$$

where $h^*(\tilde{\beta})$ is the concave dual of $h(u)$ given below:

$$h^*(\tilde{\beta}) = \inf_u [\tilde{\beta}u - h(u)] \stackrel{(8)}{=} \inf_u [\tilde{\beta}u - \min(u^{\frac{p}{2}}, \xi_\alpha)] \\ = \begin{cases} \frac{p-2}{p} \left(\frac{p}{2} \right)^{\frac{2}{2-p}} \left(\frac{1}{\tilde{\beta}} \right)^{\frac{p}{2-p}}, & \text{if } u^{\frac{p}{2}} < \xi_\alpha \\ \tilde{\beta} \xi_\alpha^{\frac{2}{p}} - \xi_\alpha, & \text{if } u^{\frac{p}{2}} \geq \xi_\alpha. \end{cases}$$

Equipped $h^*(\tilde{\beta})$ into Eq. (8), $\tilde{\beta}$ can be solved as in Eq. (7). $\square$

From the updating rule (7), we observe that the misclassified points $\hat{\ell}(x)^{\frac{p}{2}} > \xi_\alpha$ are considered as outliers, and ignored, which is expected for robustly learning a model.

Remark: The sequence $\{X^t\}$ generated by Eq. (6) iteratively minimizes the upper bound of the concave problem $\sum_{i,j} h(\ell_\alpha(X_{ij})) + \lambda \text{rank}(X)$. Since $h(\ell_\alpha(x))$ is a concave function, for any x ,

$$h(\ell_\alpha(x)) \leq h(\ell_\alpha(x^t)) + \langle \tilde{\beta}^t, \ell_\alpha(x) - \ell_\alpha(x^t) \rangle_F,$$

Algorithm 1: ROMA: RObust asyMmetric recommendA-tion

```

1 Input:  $\mathbf{A} \in \mathbb{R}^{n \times m}$ ;
2 Initialize:  $\mathbf{X}_0 = \mathbf{U}_0 = \mathbf{0} \in \mathbb{R}^{n \times m}$ ,  $\eta_0$  and  $\tau_0$ ;
3 for  $t = 1, \dots, T-1$  do
4    $\tilde{\beta}_{ij}^t = \begin{cases} \frac{\rho}{2} \hat{\ell}(X_{ij}^t)^{\frac{p-2}{2}}, & \hat{\ell}(X_{ij}^t)^{\frac{p}{2}} \leq \xi_\alpha \\ 0, & \text{otherwise} \end{cases}$  as Eq. (7);
5    $\mathbf{X}_{t+1} = \mathbf{X}_t - \eta_t (\tilde{\beta}^t \circ \partial \ell_\alpha(\mathbf{X}_t) + \lambda \mathbf{U}_t)$ ;
6    $\mathbf{U}_{t+1} = \mathbf{U}_t + \tau_t (\lambda \mathbf{X}_t - \rho \partial[\|\mathbf{U}\|_2 - 1]_+)$ ;
7 end
8 Output:  $\hat{\mathbf{X}}_T = \sum_{t=1}^T \mathbf{X}_t / T$ ;

```

where $\tilde{\beta}^t = \nabla_u h(u)|_{u=\hat{\ell}(x^t)}$. Thus we obtain an upper bound of $\sum_{ij} h(\ell_\alpha(X_{ij})) + \lambda \text{rank}(\mathbf{X})$ via a linear approximation at $\ell_\alpha(\mathbf{X}^t)$:

$$\begin{aligned} \forall \mathbf{X} \in \mathbb{R}^{n \times m} : \lambda \text{rank}(\mathbf{X}) + h(\ell_\alpha(\mathbf{X})) \\ \leq \lambda \text{rank}(\mathbf{X}) + h(\ell_\alpha(\mathbf{X}^t)) + \langle \tilde{\beta}^t, \ell_\alpha(\mathbf{X}) - \ell_\alpha(\mathbf{X}^t) \rangle_F. \end{aligned} \quad (9)$$

Since $\ell_\alpha(\mathbf{X}^t)$ is constant w.r.t. $\mathbf{X}$, we have

$$\begin{aligned} \mathbf{X}^{t+1} &= \arg \min_{\mathbf{X}} \lambda \text{rank}(\mathbf{X}) + h(\ell_\alpha(\mathbf{X}^t)) + \langle \tilde{\beta}^t, \ell_\alpha(\mathbf{X}) - \ell_\alpha(\mathbf{X}^t) \rangle_F \\ &= \arg \min_{\mathbf{X}} \lambda \text{rank}(\mathbf{X}) + \langle \tilde{\beta}^t, \ell_\alpha(\mathbf{X}) \rangle_F, \end{aligned}$$

which obtains an iterative solution to minimize the upper bound of the concave objective function.

2.2 SVD-free Min-Max Optimization

The objective, $\lambda \text{rank}(\mathbf{X}) + \langle \tilde{\beta}^t, \ell_\alpha(\mathbf{X}) \rangle_F$, is a nonconvex function and computationally intractable [1]. To address the issue, we relax $\text{rank}(\mathbf{X})$ to its convex norm,

$$\mathbf{X}^{t+1} = \arg \min_{\mathbf{X}} \langle \tilde{\beta}^t, \ell_\alpha(\mathbf{X}) \rangle_F + \lambda \|\mathbf{X}\|_*, \quad (10)$$

where $\|\mathbf{X}\|_* = \sum_i \sigma_i(\mathbf{X})$ is the nuclear norm and $\sigma_i(\mathbf{X})$ is a singular value. The nuclear norm minimization can be solved by subgradient descent (GD) [12] or proximal gradient descent (PGD) [11]. Although GD and PGD are guaranteed to converge, they have to perform SVD of $\mathbf{X}_t$ or $\mathbf{X}_t - \eta_t \nabla \ell_\alpha(\mathbf{X}_t)$ in each iteration [2], which takes a high computational cost of $O(m^2 n)$. To reduce the computational complexity, we propose an SVD-free min-max optimization algorithm. According to the dual form of the nuclear norm [16],

$$\|\mathbf{X}\|_* = \max_{\mathbf{U} \in \mathbb{R}^{n \times m}} \text{tr}(\mathbf{U}^\top \mathbf{X}) \quad \text{s.t. } \|\mathbf{U}\|_2 \leq 1,$$

where $\|\cdot\|_2$ represents the spectral norm, we can cast Eq. (10) to a concave-convex optimization problem,

$$\min_{\mathbf{X}} \max_{\mathbf{U}} \langle \tilde{\beta}^t, \ell_\alpha(\mathbf{X}) \rangle_F + \lambda \text{tr}(\mathbf{U}^\top \mathbf{X}) \quad \text{s.t. } \|\mathbf{U}\|_2 \leq 1.$$

Due to the spectral norm constraint, we have to project the intermediate solution onto the unit spectral norm ball, which again requires a full SVD operation. To address this issue, we replace the constraint $\|\mathbf{U}\|_2 \leq 1$ with a regularization term in the objective function to control the spectral norm of $\mathbf{U}$:

$$\min_{\mathbf{X}} \max_{\mathbf{U}} \langle \tilde{\beta}^t, \ell_\alpha(\mathbf{X}) \rangle_F + \lambda \text{tr}(\mathbf{U}^\top \mathbf{X}) - \rho[\|\mathbf{U}\|_2 - 1]_+,$$

where $\rho > 0$ is a parameter and $[\cdot]_+ = \max(\cdot, 0)$. Since the above optimization problem is convex-concave, we can apply the above standard subgradient method to solve it, which iterates as follows:

$$\begin{aligned} \mathbf{X}_{t+1} &= \mathbf{X}_t - \eta_t (\tilde{\beta}^t \circ \partial \ell_\alpha(\mathbf{X}_t) + \lambda \mathbf{U}_t), \\ \mathbf{U}_{t+1} &= \mathbf{U}_t + \tau_t (\lambda \mathbf{X}_t - \rho \partial[\|\mathbf{U}\|_2 - 1]_+). \end{aligned}$$

Note that the subgradient $\partial[\|\mathbf{U}\|_2 - 1]_+$ can be computed efficiently. We denote σ_1 as the leading singular value of $\mathbf{U}$, $\mathbf{p}_1$ and $\mathbf{q}_1$ as the corresponding left and right singular vectors. Then we have

$$\mathbf{p}_1 \mathbf{q}_1^\top \mathbb{I}_{(\sigma_1 > 1)} \in \partial[\|\mathbf{U}\|_2 - 1]_+.$$

In each iteration, we only need to compute the leading singular vector of $\mathbf{U}_t$ with $O(mn)$ time. In contrast, a full SVD takes $O(m^2 n)$ time. The entire procedure is summarized in Algorithm 1.

3 EXPERIMENTAL RESULTS

We start with experimental data and benchmark setup, and then present the comparison results.

3.1 Experimental Settings

Datasets: The experiments were conducted on three real datasets: MovieLens-100K [4], MovieLens-1M, and EachMovie¹. Table 1 summarizes the details of the three datasets. These datasets are class-imbalanced due to a small portion of observations. To simulate the implicit setting on the datasets, we followed previous studies [13] by treating the ratings larger than 3 as observed positive feedbacks.

Evaluation Metrics: Each dataset was randomly partitioned into two non-overlapping sets for training and testing: for each user, 80% of its observed feedbacks were randomly chosen as the training data and the remaining 20% were used for testing. For the sake of fair comparison, we repeated the random partition for five times. At each time, we computed the results for each user, then reported the averaged results over all users. We measured the performance with two widely-used evaluation metrics: 1) NDCG@N, the normalized discounted cumulative gain in the ranking; 2) F1@N, the weighted harmonic mean of precision and recall. We chose the top-N ranked items for evaluation with $N = \{5, 15\}$. Specifically, the higher these measures, the better the performance is.

Baselines: We compared our algorithm with four state-of-the-art baselines: 1) BPRMF is a one-class CF approach, which assigns each example with a heuristic prior [14]. 2) RMF-MM solves the L_1 -norm based low-rank matrix factorization problem via majorization minimization [9]. 3) SPMF employs a self-paced learning to resist outliers and noise [20]. 4) MCSHIFT is a cost-sensitive method for matrix completion [6]. All parameters of the baselines were tuned according to their recommended instructions. For bilinear factorization methods, the number of latent factors was tuned from $\{10, 20, \dots, 50\}$. For ROMA, we fixed $p = 1$, $\lambda = 10$ and $\xi_\alpha = 1$. We let $\alpha = 10$ for ML-100K and $\alpha = 100$ for ML-1M and EachMovie. We tuned the parameters η_0 and τ_0 from $\{10^{-3}, \dots, 10^2\}$, and set $\eta_t = \eta_0 / \sqrt{t}$, $\tau_t = \tau_0 / \sqrt{t}$ according to [16].

3.2 Comparison Results

The comparison results, in terms of NDCG and F1, are summarized in Table 2. It shows that ROMA always outperforms state-of-the-art robust MF methods, SPMF and RMF-MM, over all metrics.

¹<http://www.grouplens.org/>

Table 1: Description of Datasets

Datasets	#Users	#Items	#Ratings	Data Density
ML-100K	943	1,682	100,000	6.30%
ML-1M	6,039	3,628	1,000,209	4.47%
Each-MV	72,916	1,628	2,811,983	1.91%

Figure 2: Objective value versus the CPU time **Table 2: Mean and Standard Deviation of Evaluation Metrics**

Algorithm	F1@5	F1@15	NDCG@5	NDCG@15
ML-100K				
BPRMF	19.53±0.46	27.93±0.34	69.37±0.88	68.73±0.62
SPMF	18.74±0.53	27.72±0.26	68.02±0.51	68.17±0.46
RMF-MM	18.80±0.77	27.88±0.42	68.13±0.73	69.39±0.56
MCShift	19.12±0.54	29.00±0.29	70.16±0.54	71.33±0.49
ROMA	21.32±0.53	31.15±0.31	72.45±0.57	72.93±0.41
ML-1M				
BPRMF	15.81±0.48	23.64±0.35	62.44±1.01	63.47±0.62
SPMF	16.63±0.46	23.42±0.44	62.03±0.62	64.05±0.46
RMF-MM	16.71±0.51	23.81±0.41	62.19±0.78	64.09±0.72
MCShift	16.52±0.42	24.25±0.31	63.07±0.60	64.03±0.47
ROMA	17.13±0.42	25.92±0.33	64.67±0.46	65.32±0.39
Each-MV				
BPRMF	31.91±0.46	28.30±0.20	60.36±1.02	60.92±0.66
SPMF	31.10±0.53	27.77±0.34	59.65±0.55	60.58±0.47
RMF-MM	31.14±0.46	27.59±0.36	59.78±0.86	60.73±0.75
MCShift	31.31±0.38	27.32±0.30	59.65±0.51	60.67±0.45
ROMA	32.22±0.38	29.07±0.30	61.68±0.54	62.11±0.48

These MF methods treat imbalanced labels equally, which result in biased estimation of classification. In addition, ROMA outperforms the cost-sensitive methods, MCShift and BPRMF, over all the datasets. Although these methods exploit prior weights on imbalanced classes, their performance is degraded due to outliers [18].

To show the efficiency of the proposed min-max (MM) optimization, we compared our method with two classical methods: subgradient descent (SD) and proximal subgradient descent (PSD). The same step $\eta_t = \eta_0 / \sqrt{t}$ is used for SD and PSD. We plotted the objective value along the running time in Fig. 2. As can be seen, MM decreases much faster than SD and PSD. This is as expected, because MM is SVD-free and time-efficient, taking much less time in each iteration than the other two methods.

We also studied the impact of the parameter α to show how cost-sensitive weight improves the results. The results in Fig. 3 show that a small value of α would degrade the performance due to the equal weights on the imbalanced classes. However, increasing the value of α can reduce the negative impact of imbalanced classes via more “aggressive” update for the positive class.

Figure 3: Parameter sensitivity analysis of α

4 CONCLUSION

We proposed a robust asymmetric recommendation via efficient low-rank approximation. Our proposed model ROMA can well leverage both cost sensitive learning and capped L_p -norm to solve the problem. It novelly formulates the objective as an iteratively weighted convex function, which can be efficiently solved with a tight bound. To reduce the computational cost of low-rank approximation, we proposed a min-max optimization without performing full SVD. Finally, we validated the efficiency and effectiveness of our algorithm in the promising empirical results.

REFERENCES

- [1] Edoardo Amaldi and Viggo Kann. 1998. On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems. *Theoretical Computer Science* 209, 1-2 (1998), 237–260.
- [2] JianFeng Cai, Emmanuel J Candes, and Zuowei Shen. 2010. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization* 20, 4 (2010), 1956–1982.
- [3] Lihong Ding and Shizhong Liao. 2017. An approximate approach to automatic kernel selection. *IEEE Transactions on Cybernetics* 47, 3 (2017), 554–565.
- [4] F Maxwell Harper and Joseph A Konstan. 2016. The movielens datasets: History and context. *TIIS* 5, 4 (2016), 19.
- [5] Xiangnan He and Hanwang Zhang. 2016. Fast matrix factorization for online recommendation with implicit feedback. In *SIGIR*. ACM, 549–558.
- [6] Cho-Jui Hsieh, Nagarajan Natarajan, and Inderjit S Dhillon. 2015. PU learning for matrix completion. In *ICML*. 2445–2453.
- [7] Cho-Jui Hsieh and Peder Olsen. 2014. Nuclear norm minimization via active subspace selection. In *ICML*. 575–583.
- [8] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *ICDM*. Ieee, 263–272.
- [9] Zhouchen Lin, Chen Xu, and Hongbin Zha. 2017. Robust Matrix Factorization by Majorization Minimization. *PAMI* (2017).
- [10] Yurii Nesterov. 2004. *Introductory lectures on convex optimization*. Vol. 87. Springer Science & Business Media.
- [11] Yu Nesterov. 2013. Gradient methods for minimizing composite functions. *Mathematical Programming* 140, 1 (2013), 125–161.
- [12] Yurii Nesterov. 2013. *Introductory lectures on convex optimization: A basic course*. Vol. 87. Springer Science & Business Media.
- [13] Weike Pan and Li Chen. 2013. GBPR: group preference based Bayesian personalized ranking for one-class collaborative filtering. In *IJCAI*. 2691–2697.
- [14] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *UAI*. AUAI, 452–461.
- [15] Clayton Scott. 2011. Surrogate losses and regret bounds for cost-sensitive classification with example-dependent costs. In *ICML-11*. 153–160.
- [16] Yichi Xiao, Zhe Li, Tianbao Yang, and Lijun Zhang. 2017. SVD-free convex-concave approaches for nuclear norm regularization. In *IJCAI*. 3126–3132.
- [17] Peng Yang, Peilin Zhao, and Xin Gao. 2017. Robust online multi-task learning with correlative and personalized structures. *IEEE Transactions on Knowledge and Data Engineering* 29, 11 (2017), 2510–2521.
- [18] Peng Yang, Peilin Zhao, Xin Gao, and Yong Liu. 2017. Robust Cost-Sensitive Learning for Recommendation with Implicit Feedback. *arXiv preprint arXiv:1707.00536* (2017).
- [19] Tong Zhang. 2010. Analysis of multi-stage convex relaxation for sparse regularization. *JMLR* 11 (2010), 1081–1107.
- [20] Qian Zhao, Deyu Meng, Lu Jiang, Qi Xie, Zongben Xu, and Alexander G Hauptmann. 2015. Self-Paced Learning for Matrix Factorization.. In *AAAI*. 3196–3202.