Diversely Regularized Matrix Factorization for Accurate and Aggregately Diversified Recommendation

Jongjin Kim, Hyunsik Jeon, Jaeri Lee, and U Kang

Seoul National University, Seoul, South Korea
{j2kim99, jeon185, jlunits2, ukang}@snu.ac.kr

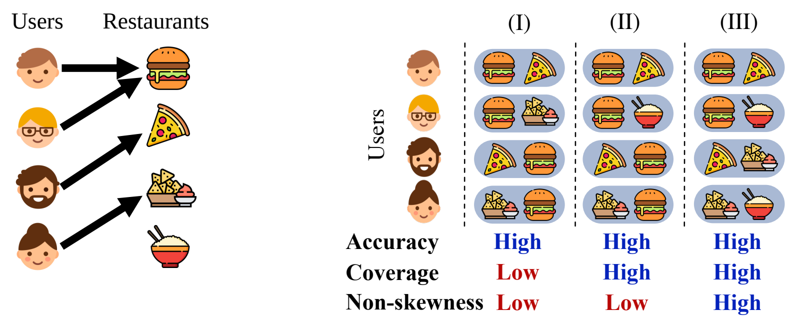
Abstract. When recommending personalized top-k items to users, how can we recommend them diversely while satisfying users' needs? Aggregately diversified recommender systems aim to recommend a variety of items across whole users without sacrificing the recommendation accuracy. They increase the exposure opportunities of various items, which in turn increase the potential revenue of sellers as well as user satisfaction. However, it is challenging to tackle aggregate-level diversity with matrix factorization (MF), one of the most common recommendation models, since skewed real-world data lead to the skewed recommendation results of MF.

In this work, we propose DivMF (Diversely Regularized Matrix Factorization), a novel matrix factorization method for aggregately diversified recommendation. DivMF exploits novel coverage regularizer and skewness regularizer which consider the top-k recommendation results of an MF model to aggregately diversify the recommendation results. We also propose a carefully designed training algorithm for effective training. Extensive experiments on real-world datasets show that DivMF gives the state-of-the-art performance, improving up to 34.7% aggregate-level diversity in the similar level of accuracy, and up to 27.6% accuracy in the similar level of aggregate-level diversity compared to the best competitors.

Keywords: Diversified Recommendation · Aggregate-level Diversity · Matrix Factorization

1 Introduction

When recommending personalized top-k items to users, how can we recommend them diversely while satisfying users' needs? Customers heavily rely on recommender systems [10,12,15] to choose items due to the flood of information nowadays. Thus, it is desired to expose as many items as possible to users to maximize the potential revenue of sales platforms [2] while improving users' experience [3]. Achieving aggregate-level diversity means fairly distributing items for the overall recommendation results. It requires that the results are of high



(a) Users' ground-truth preferences (b) Three different recommendation results

Fig. 1. Comparison of three different recommendation results. Note that all three results achieve high accuracy by recommending the ground-truth item to each user. However, the aggregate-level diversities (i.e., coverage and non-skewness) of the results (I), (II), and (III) are significantly different. Aggregately diversified recommendation aims to achieve high coverage and non-skewness while maintaining high accuracy as in the result (III).

coverage and low skewness; coverage indicates the proportion of recommended items among all items, and skewness indicates the degree of unfair frequencies of recommended items. Fig. 1 demonstrates the coverage and non-skewness of three different recommendation results. Note that all three results achieve high accuracy but only the result (III) obtains high aggregate-level diversity by recommending every item twice. In other words, only the result (III) achieves a high aggregate-level diversity, recommending each item by the same amount, maximizing the potential revenue of sales platforms.

Matrix factorization (MF) [16] is the most widely used collaborative filtering method due to its powerful scalability and flexibility [13, 19]. However, the traditional MF has a limitation in achieving high aggregate-level diversity on real-world data because it is vulnerable to the skewness of data [23]. To overcome this problem, previous works on aggregately diversified recommendation rerank the recommendation lists or recommendation scores of a given MF model [1,5,14,17]. However, these approaches do not give the best diversity since they focus only on post-processing the results of MF, which is already trained with skewed data. Thus, it is desired to deal with aggregate-level diversity in the training process of MF to achieve both high accuracy and diversity.

In this work, we propose Diversely Regularized Matrix Factorization (DivMF), a novel approach for aggregately diversified recommendation. DivMF regularizes a recommendation model in its training process so that more diverse items appear uniformly on top-k recommendations. DivMF effectively maximizes the coverage and non-skewness of the recommendation by utilizing two regularizers: coverage and skewness regularizers both of which consider the item occurrences in top-k recommendation list. This allows the model to achieve optimal aggregate-level diversity in the training process. We also propose a carefully designed training algorithm that first focuses on accuracy and then on diversity, and an unmasking mechanism for accurate and effective learning of DivMF.

Our contributions are summarized as follows:

- **Method.** We propose DivMF, a method for aggregately diversified recommendation. DivMF provides a new way to accurately and efficiently optimize an MF model to achieve both high accuracy and aggregate-level diversity for top-k recommendation.
- **Theory.** We theoretically prove that DivMF provides an optimal solution to maximize the aggregate-level diversity in top-k recommendation.
- **Experiments.** Extensive experiments show that DivMF achieves up to 34.7% higher aggregate-level diversity in the similar level of accuracy, and up to 27.6% higher accuracy in the similar level of aggregate-level diversity in personalized top-k recommendation compared to the best competitors, resulting in the state-of-the-art performance (see Fig. 2). The code and datasets are available at `https://github.com/snudatalab/DivMF`.

2 Aggregately Diversified Recommendation

In recent years, diversification has attracted increasing attention in recommendation research [11, 25]. We focus on increasing diversity at the aggregate-level. Aggregate-level diversity considers the diversity in the overall recommendation results of all users to improve the potential profit of service platforms [2].

Aggregately diversified recommendation aims to improve two aspects of recommendation: coverage and non-skewness. Coverage is the total number of unique items recommended at least once. Non-skewness is the balance between frequencies of recommended items. The details of their evaluation are as follows.

- **Coverage.** Coverage measures how many different items a recommendation result contains from the whole items. It is defined as follows:

$$\mathit{Coverage} = \left|\bigcup_{u\in\mathbb{U}}\mathbb{L}(u)\right| / \left|\mathbb{I}\right|, \tag{1}$$

where k is the number of items recommended, and $\mathbb{L}(u)$ is the set of recommended items for user u. $\mathbb{U}$ and $\mathbb{I}$ are sets of users and items, respectively. Coverage ranges from 0 to 1, and a higher value represents better coverage.

- **Gini index.** Gini index measures the inequality between item frequencies in recommendation results. It is defined as follows:

$$\mathit{Gini} = \frac{1}{|\mathbb{I}|-1}\sum_{j=1}^{|\mathbb{I}|}(2j-|\mathbb{I}|-1)p_j, \tag{2}$$

where p_j is the j-th least value in $\{\frac{f(i)}{\sum_{j\in\mathbb{I}} f(j)} | i \in \mathbb{I}\}$ and $f(i)$ indicates the frequency of item i in the recommendation results for whole users. Gini index ranges from 0 to 1, and a lower value represents better non-skewness.

3 Proposed Method

In this section, we propose DivMF (Diversely Regularized Matrix Factorization), a matrix factorization method for accurate and aggregately diversified recommendation.

3.1 Overview

We address the following challenges to achieve high performance of aggregately diversified recommendation:

– **Coverage maximization.** Matrix factorization (MF) is prone to obtaining top-k recommendations with low coverage where only a few items are recommended. How can we train MF to recommend every item at least once?
– **Non-skewed frequency.** MF is liable to achieving skewed top-k recommendations. How can we train MF to recommend all items with similar frequencies?
– **Non-trivial optimization.** It is difficult to simultaneously handle both accuracy and diversity which are disparate criteria. How can we train MF to optimize both the accuracy and diversity?

The main ideas to address the challenges are as follows:

– **Coverage regularizer.** The coverage regularizer evenly balances the recommendation scores at the item-level, enabling us to recommend each item to at least one user.
– **Skewness regularizer.** The skewness regularizer equalizes all the recommendation scores to assist the coverage regularizer to make the model recommends all items by the same numbers of times.
– **Careful training.** We carefully design a training algorithm which first focuses on accuracy and then on diversity. This allows a model to be trained stably and efficiently, despite the conflict between accuracy and diversity. We also propose an unmasking mechanism for effective training.

3.2 Definition of Diversity Regularizer

Coverage Regularizer We design a *coverage regularizer* to maximize the coverage. Focusing on the recommended items in the score matrix, we mask the scores of non-recommended items for each user to zero. After masking, a column filled with zeros corresponds to an item that is not recommended to any user. Hence, the coverage regularizer is required to distribute the remaining values in the masked matrix among all columns. In the following, we show how we construct the coverage regularizer from the fact that the equality condition of the arithmetic-geometric mean inequality states the equal distribution of values.

Assume that $\hat{\mathbf{R}} = [\hat{r}_{ui}] \in \mathbb{R}^{|\mathbb{U}| \times |\mathbb{I}|}$ is the recommendation score matrix where $\hat{r}_{ui}$ is a dot product between user u's embedding and item i's embedding. For $u \in \mathbb{U}$, consider $\mathbf{S} = [s_{ui}]$ where $\mathbf{S}_u = softmax(\hat{\mathbf{R}}_{\mathbf{u}})$, which means $(s_{u1}, ..., s_{u|\mathbb{I}|}) = softmax(\hat{r}_{u1}, ..., \hat{r}_{u|\mathbb{I}|})$. Then, we keep top-$k$ elements of each row in $\mathbf{S}$ while masking others to zero to construct a matrix $\mathbf{T} = [t_{ui}]$. Note that the nonzero t_{ui} implies that the top-k recommendation list of user u includes item i. Then, the coverage regularizer Reg_{cov} is defined as follows:

$$Reg_{cov} = -\log\left(\prod_{i \in \mathbb{I}} \sum_{u \in \mathbb{U}} t_{ui}\right) = -\sum_{i \in \mathbb{I}} \log\left(\sum_{u \in \mathbb{U}} t_{ui}\right).$$

This regularizer is useful to maximize coverage, as shown in Theorem 1.

Theorem 1. *If Reg_{cov} is minimized, then coverage is maximized.*

Proof. $\sum_{u\in\mathbb{U}} t_{ui} \leq \sum_{u\in\mathbb{U}} s_{ui}$ for all $i \in \mathbb{I}$ since $0 \leq t_{ui} \leq s_{ui}$ for all $u \in \mathbb{U}$ and $i \in \mathbb{I}$. Thus, using the fact that $\sum_{i\in\mathbb{I}} s_{ui} = 1$ for all $u \in \mathbb{U}$,

$$\sum_{i\in\mathbb{I}}\sum_{u\in\mathbb{U}} t_{ui} \leq \sum_{i\in\mathbb{I}}\sum_{u\in\mathbb{U}} s_{ui} = \sum_{u\in\mathbb{U}}\sum_{i\in\mathbb{I}} s_{ui} = |\mathbb{U}|.$$

We thus obtain

$$\exp(-Reg_{cov}) = \prod_{i\in\mathbb{I}}\sum_{u\in\mathbb{U}} t_{ui} \leq (\frac{|\mathbb{U}|}{|\mathbb{I}|})^{|\mathbb{I}|}, \tag{3}$$

from the arithmetic geometric mean inequality. Equality holds if and only if for all i, $\sum_{u\in\mathbb{U}} t_{ui} = |\mathbb{U}|/|\mathbb{I}|$. In this case, every column of $\mathbf{T}$ has at least one nonzero element. Thus, every item is included in at least one user's top-k recommendation list, so the coverage is 1. □

Skewness Regularizer Although the condition to minimize the coverage regularizer guarantees the coverage of the model to be 1, this does not guarantee the non-skewness to be maximized. For example, assume that $(t_{11}, t_{21}, ..., t_{|\mathbb{U}|1}) = (\frac{1}{2}, \frac{1}{2}, 0, 0, ..., 0)$ and $(t_{12}, t_{22}, ..., t_{|\mathbb{U}|2}) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, 0, ..., 0)$. In this case, $\sum_{u\in\mathbb{U}} t_{u1} = \sum_{u\in\mathbb{U}} t_{u2}$ but the item 1 is recommended twice while the item 2 is recommended three times. In other words, it is possible to meet the equality condition of Equation (3) even if the non-skewness of the model is not maximized, since the value of each nonzero element could vary.

To address this problem, we propose a *skewness regularizer*. Since the problem occurs because of the variance of nonzero elements, we design the skewness regularizer to equalize values of nonzero t_{ui}. After equalization, $\sum_{u\in\mathbb{U}} t_{ui}$ and $\sum_{u\in\mathbb{U}} t_{uj}$ would be equal if and only if items i and j are recommended for the same number of times, so the coverage regularizer would also optimize the non-skewness in recommendation lists.

Let $\mathbf{T}' = [t'_{ui}]$ be a row-normalized $\mathbf{T}$ which means $t'_{ui} = t_{ui}/\sum_{j\in\mathbb{I}} t_{uj}$. The skewness regularizer Reg_{skew} is defined as follows:

$$Reg_{skew} = \sum_{u\in\mathbb{U}}\sum_{i\in\mathbb{I}} t'_{ui}\log t'_{ui} = -\sum_{u\in\mathbb{U}} entropy(\mathbf{T}_u).$$

Since each entropy function is maximized if and only if nonzero elements of each $\mathbf{T}_u$ are equal, Reg_{skew} is minimized if and only if all nonzero elements of each row of $\mathbf{T}$ are equal.

Diversity Loss Function Finally, we define the loss function for aggregate-level diversity in DivMF as $\mathcal{L}_{div}(\hat{\mathbf{R}}) = Reg_{cov} + Reg_{skew}$. This loss function satisfies the Theorem 2.

Theorem 2. *If $\mathcal{L}_{div}(\hat{\mathbf{R}})$ is minimized, then coverage and non-skewness are both maximized.*

Proof. The condition to minimize Reg_{cov} is $\sum_{u \in \mathbb{U}} t_{ui} = |\mathbb{U}|/|\mathbb{I}|$ for every item i, and the condition to minimize Reg_{skew} is that nonzero elements of $\mathbf{T}_u$ are equal for every user u. Thus, the condition to minimize $Reg_{cov} + Reg_{skew}$ is that each row of $\mathbf{T}$ contains k nonzero elements with value of $\frac{1}{k}$, and each column of $\mathbf{T}$ contains $\frac{|\mathbb{U}|k}{|\mathbb{I}|}$ nonzero elements. In this case, every item appears in the recommendation results with equal frequency. Therefore, both coverage and non-skewness are maximized if $\mathcal{L}_{div}(\hat{\mathbf{R}})$ is minimized. □

3.3 Model Training

Objective Function and Training Algorithm In order to maximize accuracy and aggregate-level diversity of recommendation results simultaneously, we propose the following objective function.

$$\mathcal{L}_{total}(\theta; \mathbf{R}) = \mathcal{L}_{acc}(\hat{\mathbf{R}}) + \mathcal{L}_{div}(\hat{\mathbf{R}}),$$

where $\mathcal{L}_{total}(\cdot)$ is the total loss to be minimized, $\mathcal{L}_{acc}(\cdot)$ and $\mathcal{L}_{div}(\cdot)$ are losses for accuracy and aggregate-level diversity, respectively, $\mathbf{R}$ is the observed interaction matrix, $\hat{\mathbf{R}}$ is the recommendation score matrix, and θ is the parameter to be optimized. We use BPR loss function as an accuracy loss since it is known to show the best performance in top-k recommendation [21]. Thus,

$$\mathcal{L}_{acc}(\hat{\mathbf{R}}) = \sum_{u \in \mathbb{U}, (i,j) \in \mathbb{Z}(u)} \log\Big(1 + \exp\Big(\hat{\mathbf{R}}_{uj} - \hat{\mathbf{R}}_{ui}\Big)\Big),$$

where $\mathbb{Z}(u) = \{(i, j) | \mathbf{R}_{ui} = 1, \mathbf{R}_{uj} = 0\}$.

A challenge in minimizing the loss $\mathcal{L}_{total}$ is that directly minimizing $\mathcal{L}_{total}$ or optimizing $\mathcal{L}_{acc}$ and $\mathcal{L}_{div}$ in an iterative, alternating fashion leads to poor performance (see Section 4.4). We presume that this problem happens because the gradients of accuracy loss and diversity regularizer cancel each other out. The accuracy loss tries to increase the gap between recommendation scores of high scored items and low scored items, while the diversity regularizer tries to decrease the gap. Thus, the net gradient is not large enough to prevent the model from being trapped in bad local optima.

Our idea to avoid this issue is to train DivMF model with only accuracy loss $\mathcal{L}_{acc}$ until the accuracy converges, and then train the model with the diversity regularizer $\mathcal{L}_{div}$. In this way, the gradients of accuracy loss and diversity regularizer do not cancel each other out since the optimizer minimizes only one loss at a time. To adjust the trade-off between accuracy and diversity, we control the number n_{ep} of epochs to optimize $\mathcal{L}_{div}$, since the model achieves higher diversity and lower accuracy as we increase n_{ep}.

Unmasking Mechanism Gradients from $\mathcal{L}_{div}(\hat{\mathbf{R}})$ do not flow directly into unrecommended items since $\mathbf{T}$ masks $|\mathbb{I}| - k$ items with the lowest scores in $\mathbf{S}$ of each user. Thus, a straightforward gradient descent with $\mathcal{L}_{div}(\hat{\mathbf{R}})$ has limitation to find new items for diversity, optimizing only k scores of initially selected items.

We propose an unmasking mechanism to overcome this problem. The idea is to keep additional unmasked elements in each row of $\mathbf{S}$ when building $\mathbf{T}$. In this way, rarely recommended items have an opportunity to be unmasked. DivMF finds new rarely recommended items by a gradient descent with this unmasking mechanism. DivMF unmasks a fixed number of the highest-scored items other than already recommended items during each iteration of training, which is the best unmasking scheme as experimentally shown in Section 4.5.

4 Experiments

We perform experiments to answer the following questions:

Q1. **Diversity and accuracy (Section 4.2).** Does DivMF show high aggregate-level diversity without sacrificing the accuracy of recommendation?

Q2. **Regularizer (Section 4.3).** How do the diversity regularization terms Reg_{cov} and Reg_{skew} of DivMF help improve the diversity of DivMF?

Q3. **Training algorithm (Section 4.4).** Does the training algorithm of DivMF prevent the training from being trapped in bad local optima?

Q4. **Unmasking mechanism (Section 4.5).** How does the unmasking mechanism of DivMF affect the performance?

4.1 Experimental Setup

We introduce our experimental setup including datasets, evaluation protocol, baseline approaches, evaluation metrics, and the training process.

Datasets. We use five real-world rating datasets as summarized in Table 1. We preprocess extremely sparse datasets (Yelp, Gowalla, and Epinions) as core-15 following a previous work [17]. In other words, we make the datasets include only users and items that have at least 15 interactions. MovieLens-10M and MovieLens-1M datasets [9] contain movie ratings constructed by the GroupLens research group. Yelp-15 contains 15-core restaurant rating data collected from a restaurant review site with the same name. Epinions-15 [18] contains 15-core rating data of products constructed from a general consumer review site. Gowalla-15 [4] contains 15-core data of a friendship network of users constructed from a location-based social networking website. We remove the rating scores of datasets and obtain user-item interaction data which indicate whether the user has rated the item or not.

Evaluation Protocol. We employ *leave-one-out* protocol where one of each user's interaction instances is removed for testing. If the dataset includes timestamp, the latest instance of each user is removed, and if not, randomly sampled instances are removed.

Baselines. We compare DivMF with existing methods for aggregately diversified recommendation.

- **Kwon.** Kwon et al. [1] adjust recommendation scores of items based on their frequencies to achieve aggregate level diversity.
- **Karakaya.** Karakaya et al. [14] replace items on recommendation lists with infrequently recommended similar items.

Table 1. Summary of datasets.

Dataset	Users	Items	Interactions	Density(%)
Yelp-15 [1]	69,853	43,671	2,807,606	0.0920
Gowalla-15 [2]	34,688	63,729	2,438,708	0.1111
Epinions-15 [3]	5,531	4,286	186,995	0.7888
MovieLens-10M [4]	69,878	10,677	10,000,054	1.3403
MovieLens-1M [5]	6,040	3,706	1,000,209	4.4684

[1] `https://www.yelp.com/dataset`
[2] `https://snap.stanford.edu/data/loc-gowalla.html`
[3] `http://www.trustlet.org/downloaded_epinions.html`
[4] `https://grouplens.org/datasets/movielens/10m/`
[5] `https://grouplens.org/datasets/movielens/1m/`

- **Fairmatch.** Fairmatch [17] utilizes a maximum flow problem to find important items.
- **UImatch.** UImatch [5] assigns recommendation capacity to each item and greedily constructs recommendation lists.

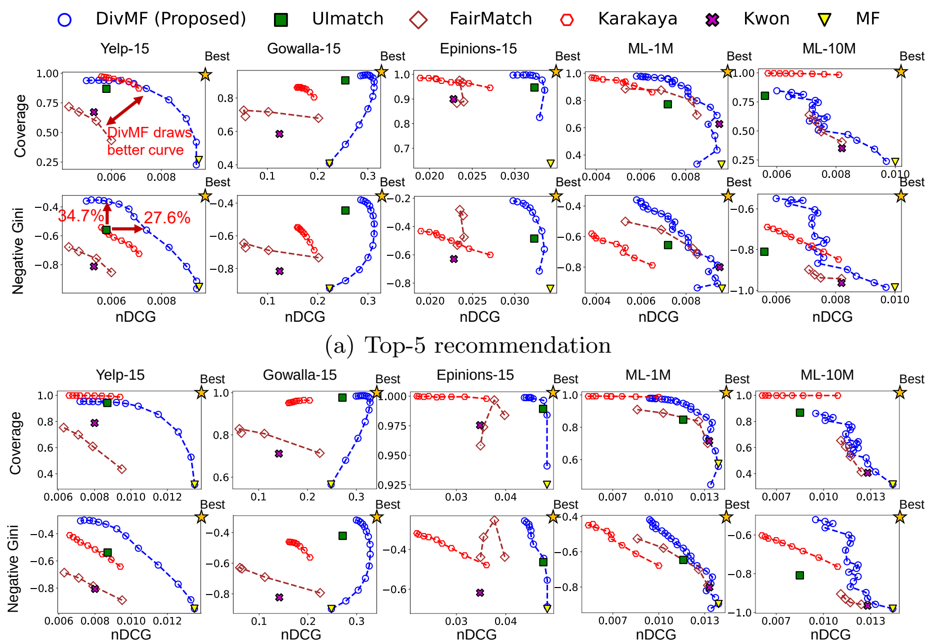
Evaluation metrics. We evaluate the performance of the methods in two categories: accuracy and diversity. Accuracy metric checks whether a model recommends correct items or not, and diversity metrics evaluate aggregated diversity of the recommendation. For each experiment, a list of recommendation to each user is created and evaluated by the following metrics.

- **Accuracy.**
  - **nDCG@k.** nDCG@k measures the overall accuracy of the top-k recommendation. It ranges from 0 to 1, where the value 0 indicates the lowest accuracy and the value 1 represents the highest accuracy.
- **Diversity.**
  - **Coverage@k.** The coverage of the top-k recommendation.
  - **Negative Gini index@k.** The negative value of the Gini index of the top-k recommendation.

Training Details. We first train the MF model until convergence.Then, we apply each baseline and DivMF on the trained MF model. We min-max normalize the recommendation scores for Kwon and Karakaya since they need prediction ratings on a finite scale. We use reverse prediction scheme [1] and set $T_H = 0.8, T_R = 0.9$ for Kwon. We vary t in $\{30, 50, 75, 100\}$ and set $\alpha = 0.5$ for FairMatch. We unmask 50 items in Epinions-15 dataset, 100 items in ML-1M/ML-10M datasets, and 500 items in Gowalla-15/Yelp-15 datasets to apply DivMF. All the models are trained with Adam optimizer with learning rate 0.001, l_2 regularization coefficient 0.0001, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. We vary k in $\{5, 10\}$ for all datasets.

4.2 Diversity and Accuracy (Q1)

We show the change of accuracies and diversities of DivMF and the competitors on five real-world datasets in Fig. 2. For each method, we adjust hyperparameters to mark points on the plot and connect them to obtain the trade-off curve. We mark the point with the highest accuracy and the highest diversity



(a) Top-5 recommendation

(b) Top-10 recommendation

Fig. 2. Accuracy-diversity trade-off curves of top-5 and top-10 recommendations on five real-world datasets. DivMF achieves the highest aggregate-level diversity while sacrificing minimal accuracy.

in each plot as the 'best' point of the plot. Note that DivMF achieves the highest diversity while sacrificing the least accuracy compared to other baselines considering the balance of coverage and non-skewness.

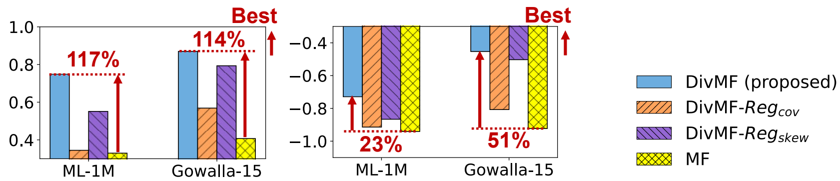
4.3 Regularizer (Q2)

To verify the impact of coverage regularizer and skewness regularizer, we examine how much the diversity of top-5 recommendation results improves. We compare DivMF, DivMF-Reg_{skew}, and DivMF-Reg_{cov} on ML-1M and Gowalla-15 datasets; DivMF-Reg_{skew} and DivMF-Reg_{cov} are DivMF without the skewness regularizer and the coverage regularizer, respectively. For the fair comparison, we train each model until the nDCG is dropped by 5% compared to MF.

Fig. 3 shows that DivMF increases both the coverage and the non-skewness the most, compared to other models. This verifies that both regularizers contribute to improving the aggregate-level diversity.

4.4 Training Algorithm (Q3)

To prove the effectiveness of our training algorithm, we compare top-5 recommendation performances of DivMF and DivMF $_{alter}$ on ML-1M dataset during training. Instead of sequentially optimizing accuracy loss and diversity loss as in DivMF, DivMF $_{alter}$ alternately optimizes two losses.



(a) Coverage (b) Negative Gini index

Fig. 3. Diversities of DivMF and its variants on ML-1M dataset compared to MF when nDCG is decreased by 5%. DivMF improves diversity the most.

Fig. 4 shows that DivMF significantly increases the diversity compared to DivMF $_{alter}$ while sacrificing a similar amount of accuracy. This proves that our training algorithm prevents the model from being trapped in bad local optima.

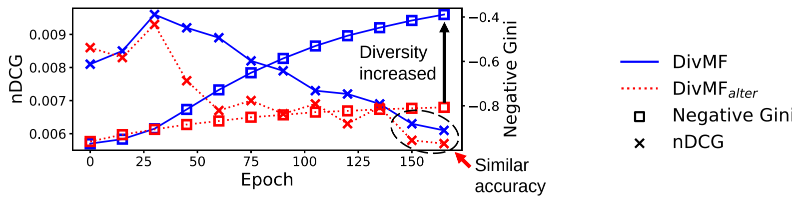


Fig. 4. Change of nDCG and Gini index of DivMF and DivMF $_{alter}$ during training on ML-1M dataset. DivMF improves diversity better by avoiding bad local optima.

4.5 Unmasking Mechanism (Q4)

To find the best unmasking policy for DivMF, we compare three policies: *No unmasking*, *Top*, and *Random* on ML-1M dataset. In addition to top-k items, *Top* unmasks n items with the highest prediction scores while *Random* unmasks random n items. We set $n = 100$ since it shows the best performance in both schemes. *No unmasking* does not unmask any item other than top-k items.

Fig. 5 shows performances of the three policies in top-5 recommendation. We have two observations. First, *No unmasking* fails to increase aggregate-level diversity, while *Top* and *Random* further improve both coverage and non-skewness. Second, *Top* performs better than *Random* since it achieves higher coverage while non-skewnesses of the two schemes are comparable in the case. In summary, *Top* is the best unmasking scheme to achieve high aggregate-level diversity.

5 Related Works

Individually diversified recommendation. Individually diversified recommendation recommends diversified items to each user [25]. Maximizing individual diversity can maximize item novelty in each user's view, but it may recommend already known items in overall recommendation list for all users. Thus, maximizing individual-level diversity does not guarantee the improvement in aggregate-level diversity [1].

Fair recommendation. Fair recommendation aims to design an algorithm that makes fair predictions devoid of discrimination [8]. Fairness in recommendation could be observed between different item groups [6] or between distinct

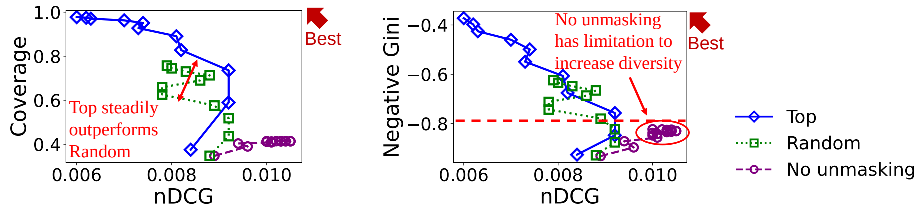


Fig. 5. Accuracy-coverage (left) and accuracy-negative Gini index (right) trade-off curves of different unmasking policies. '*Top*' shows the best overall performance.

items with similar attributes [20]. Aggregately diversified recommendation does not require any group or attribute of items, which is the main difference compared to the fair recommendation.

Popularity debiased recommendation. Popularity debiased recommendation aims to improve the quality of recommendation for long-tail items. Traditional recommender systems tend to show poor accuracy for infrequently appearing items because of the skewness in dataset [22]. There are researches to eliminate the popularity bias to achieve high accuracy in recommending long-tail items as well as popular items [7, 24]. Aggregately diversified recommendation focuses on increasing the frequencies of long tail items instead of their accuracies, which is the main difference from popularity debiased recommendation.

6 Conclusion

We propose DivMF, a matrix factorization method which maximizes aggregate-level diversity while sacrificing minimal accuracy in top-k recommendation. DivMF exploits coverage regularizer and skewness regularizer for MF via a carefully designed training algorithm. Experiments on five real-world datasets show that DivMF achieves the state-of-the-art performance in aggregately diversified recommendation, outperforming the best competitor with up to 34.7% reduced Gini index in the similar level of accuracy and up to 27.6% higher nDCG in the similar level of diversity. Future works include extending DivMF for other recommendation models beyond the matrix factorization.

Acknowledgments

This work was supported by Jung-Hun Foundation. The Institute of Engineering Research and ICT at Seoul National University provided research facilities for this work. U Kang is the corresponding author.

References

1. Adomavicius, G., Kwon, Y.: Improving aggregate recommendation diversity using ranking-based techniques. IEEE TKDE (2012)
2. Brynjolfsson, E., Hu, Y., Simester, D.: Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. Management Science (2011)

3. Brynjolfsson, E., Hu, Y., Smith, M.D.: Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. Management science (2003)
4. Cho, E., Myers, S.A., Leskovec, J.: Friendship and mobility: User movement in location-based social networks. In: KDD. ACM (2011)
5. Dong, Q., Xie, S.S., Li, W.J.: User-item matching for recommendation fairness. IEEE Access (2021)
6. Ekstrand, M.D., Kluver, D.: Exploring author gender in book rating and recommendation. UMUAI (2021)
7. Ferraro, A.: Music cold-start and long-tail recommendation: bias in deep representations. In: RecSys (2019)
8. Gajane, P.: On formalizing fairness in prediction with machine learning. ArXiv (2017)
9. Harper, F.M., Konstan, J.A.: The movielens datasets: History and context. TiiS (2015)
10. Jeon, H., Jang, J.G., Kim, T., Kang, U.: Accurate bundle matching and generation via multitask learning with partially shared parameters. Plos one (2023)
11. Jeon, H., Kim, J., Lee, J., Lee, J., Kang, U.: Aggregately diversified bundle recommendation via popularity debiasing and configuration-aware reranking. In: PAKDD (2023)
12. Jeon, H., Kim, J., Yoon, H., Lee, J., Kang, U.: Accurate action recommendation for smart home via two-level encoders and commonsense knowledge. In: CIKM. ACM (2022)
13. Jeon, H., Koo, B., Kang, U.: Data context adaptation for accurate recommendation with additional information. In: BigData (2019)
14. Karakaya, M., Aytekin, T.: Effective methods for increasing aggregate diversity in recommender systems. KAIS (2018)
15. Koo, B., Jeon, H., Kang, U.: Accurate news recommendation coalescing personal and global temporal preferences. In: PAKDD (2020)
16. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer (2009)
17. Mansoury, M., Abdollahpouri, H., Pechenizkiy, M., Mobasher, B., Burke, R.: Fairmatch: A graph-based approach for improving aggregate diversity in recommender systems. In: UMAP. ACM (2020)
18. Massa, P., Souren, K., Salvetti, M., Tomasoni, D.: Trustlet, open research on trust metrics. SCPE (2008)
19. Park, H., Jung, J., Kang, U.: A comparative study of matrix factorization and random walk with restart in recommender systems. In: BigData (2017)
20. Patro, G.K., Biswas, A., Ganguly, N., Gummadi, K.P., Chakraborty, A.: Fairrec: Two-sided fairness for personalized recommendations in two-sided platforms. In: WebConf (2020)
21. Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L.: BPR: bayesian personalized ranking from implicit feedback. In: UAI. AUAI Press (2009)
22. Steck, H.: Item popularity and recommendation accuracy. In: RecSys (2011)
23. Wang, H., Ruan, B.: Matrec: Matrix factorization for highly skewed dataset. In: BigData (2020)
24. Wei, T., Feng, F., Chen, J., Wu, Z., Yi, J., He, X.: Model-agnostic counterfactual reasoning for eliminating popularity bias in recommender system. In: KDD (2021)
25. Wu, Q., Liu, Y., Miao, C., Zhao, Y., Guan, L., Tang, H.: Recent advances in diversified recommendation (2019)