Comparative Study of Adversarial Training Methods for Cold-Start Recommendation

Haokai Ma
Shandong University
mahaokai@mail.sdu.edu.cn

Xiangxian Li
Shandong University
xiangxian_lee@mail.sdu.edu.cn

Lei Meng*
Shandong University
lmeng@sdu.edu.cn

Xiangxu Meng
Shandong University
mxx@sdu.edu.cn

ABSTRACT

Adversarial training in recommendation is originated to improve the robustness of recommenders to attack signals and has recently shown promising results to alleviate cold-start recommendation. However, existing methods usually should make a trade-off between model robustness and performance, and the underlying reasons why using adversarial samples for training works has not been sufficiently verified. To address this issue, this paper identifies the key components of existing adversarial training methods and presents a taxonomy that defines these methods using three levels of components for perturbation generation, perturbation incorporation, and model optimization. Based on this taxonomy, different variants of existing methods are created, and a comparative study is conducted to verify the influence of each component in cold-start recommendation. Experimental results on two benchmarking datasets show that existing state-of-the-art algorithms can be further improved by a proper pairing of the key components as listed in the taxonomy. Moreover, using case studies and visualization, the influence of the content information of items on cold-start recommendation has been analyzed, and the explanations for the working mechanism of different components as proposed in the taxonomy have been offered. These verify the effectiveness of the proposed taxonomy as a design paradigm for adversarial training.

CCS CONCEPTS

• **Information systems** → **Recommender systems**.

KEYWORDS

Cold-start recommendation, Adversarial training, Taxonomy

ACM Reference Format:
Haokai Ma, Xiangxian Li, Lei Meng, and Xiangxu Meng. 2021. Comparative Study of Adversarial Training Methods for Cold-Start Recommendation. In *Proceedings of the 1st International Workshop on Adversarial Learning for Multimedia (ADVM '21), October 20, 2021, Virtual Event, China.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3475724.3483600

*Corresponding author

ADVM '21, October 20, 2021, Virtual Event, China

ACM ISBN 978-1-4503-8672-2/21/10…$15.00
https://doi.org/10.1145/3475724.3483600

1 INTRODUCTION

Recommendation for cold-start users and items is a well-recognized issue. It leads to the ill-posed learning of conventional collaborative filtering recommenders, such as Matrix Factorization (MF) [8, 15], such that active users enjoy higher recommendation accuracy than the cold-start users and popular items have a much higher chance to be recommended than the cold-start items. Existing studies typically alleviate it by introducing content [7, 11, 26] and attributes [9, 16] information of users and items. However, this does not well-address the "biased learning" problems. Notably, recent studies attempt to incorporate adversarial training as a new direction to alleviate the cold-start problems in recommendation, which performs data augmentation by imposing adversarial perturbations [1, 2, 20].

Existing adversarial training methods can be categorized into different classes, based on their methods to generate the perturbations and the way to add them to the training process. Specifically, commonly-used perturbation generation methods include stochastic perturbations [8, 22], statistical perturbations, gradient-based perturbations [8, 22, 25] and GAN-based perturbations [1, 2, 20]. Such perturbations are added to either the input data [1, 2, 20, 22] or the pre-extracted intermediate features [8, 25]. It is worth mentioning that most of the existing adversarial training methods focus on the trade-off between model robustness [13, 17] and recommendation performance [8, 22, 25]. Therefore, their effectiveness and the underlying principles on cold-start recommendation have not been fully-verified. This leads to the need of a comparative study on these algorithms and an experimental verification on the key components that improve the cold-start recommendation.

To this end, this paper presents an investigation on existing adversarial training algorithms for recommenders and proposes a taxonomy that redefines them with three levels of key components: 1) information level defines their ways to introduce the perturbation into the training process, including data perturbation [1, 2, 20, 22] and feature perturbation [8, 25]; 2) methodology level classifies the approaches for perturbation generation into stochastic perturbations [8, 22], statistical perturbations, gradient-based perturbations [8, 22, 25], and GAN-based perturbations [1, 2, 20]; and 3) optimization level indicates the loss signals that are used for model optimization, termed adversarial optimization and hybrid optimization [8, 22, 25]. This taxonomy covers all of the existing adversarial training algorithms and may serve as a design paradigm, where

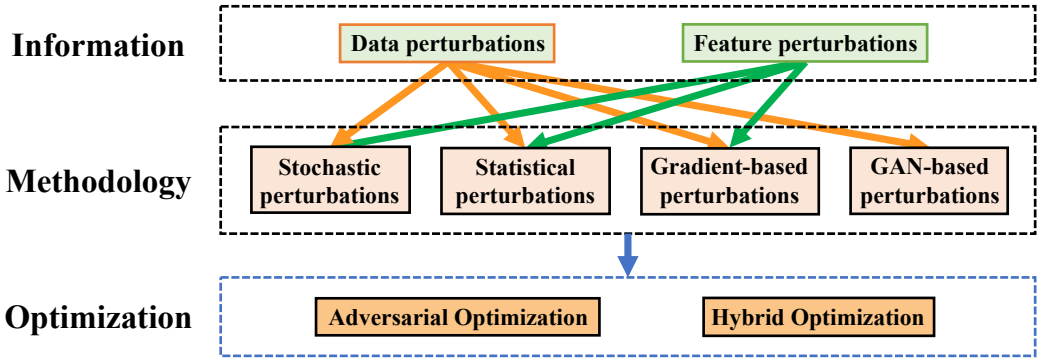


Figure 1: Illustration on the proposed taxonomy of adversarial training methods for recommendation.

an algorithm can be produced by a combination of the key components at different levels. Based on this taxonomy, we evaluated the effectiveness of existing adversarial training algorithms and their variants for cold-start recommendation on two benchmarking datasets. Experimental results with respect to performance comparison, cold-start performance, and embedding visualization, suggest that adversarial training can both effectively alleviate the cold-start problems in recommendation and improve the overall performance and robustness of the recommendation model. The reason may lie in its ability to facilitate the learning of representations in the embedding space compared to BPR-MF, allowing the embeddings of users and the items they interacted with tend to be closer. To summarize, the main contributions of this paper include:

(1) A taxonomy is proposed as a design paradigm of adversarial training methods for recommendation. It enables the personalized development of adversarial training algorithms by combining the key components of different levels in the taxonomy.
(2) A comparative study is conducted to verify the performance of existing adversarial training algorithms for cold-start recommendation on two real-world datasets and analyze the underlying principles through embedding visualization.

2 RELATED WORK

2.1 Cold-Start Recommenndation

The cold-start problem is common in existing recommendation datasets, where a few users or items dominate the interactions in the dataset, due to the imbalance of interactions between users and items [1, 2, 11, 12]. It may lead to the recommendation 'bias' that popular items have a much higher chance to be recommended than the cold-start ones. To alleviate this issue, most of the existing methods attempts to introduce auxiliary data to improve the learning of cold-start user (item) representations, such as the users' personal [11] and social network information [16], and the items' multi-modal [7, 10, 24, 26] and affinity information [9]. However, such additional information may not be available in many real-world cases. To address this problem, a line of studies [12, 14, 19] attempt to use the rich semantic information from the higher-order graph structures, targeting at the augmented user-item interactions. Another line of research [1, 2, 8, 20, 22, 25] uses adversarial training to augment the cold-start data. Interestingly, a recent study [23] adopts the dropout mechanism for the input interactions during training to improve the model generalization capability.

2.2 Adversarial Training in Recommendation

Christian Szegedy discovered that neural networks are vulnerable to small but intentional adversarial perturbations [21]. This leads to the research of adversarial training, which aims to improve the robustness of recommenders by introducing the adversarial samples in training process [3, 6, 8, 13, 17, 18]. The main procedures of adversarial training include perturbation generation and addition. Despite different ways to generate the perturbation, existing algorithms introduce the perturbation in four ways. First, adding perturbations to the intermediate features is the commonly-used method, for example, Adversarial Matrix Factorization (AMF) [8] adds stochastic and gradient-based perturbations to the low-dimensional item features, and Fine-grained Adversarial Collaborative Auto Encoder (FG-ACAE) [25] imposes gradient-based perturbations to the output features from the encoder. Besides, Adversarial Multimedia Recommendation (AMR) [22] imposes stochastic perturbations and gradient-based perturbations on the pre-extracted content features, Rating Augmentation Generative Adversarial Networks with bias treatment (RAGAN^{BT}) [1] apply GAN-based perturbations to generate new interactions. Interestingly, Adversarial Pairwise Learning (APL) [20] and Augmented Reality Collaborative Filtering (AR-CF) [2] generate new users and items with interactions.

3 TAXONOMY OF ADVERSARIAL TRAINING METHODS FOR RECOMMENDATION

As illustrated in Figure 1, the proposed taxonomy describes the existing adversarial training methods with three levels of components, including information level, methodology level, and optimization level, which are detailed in the following sections.

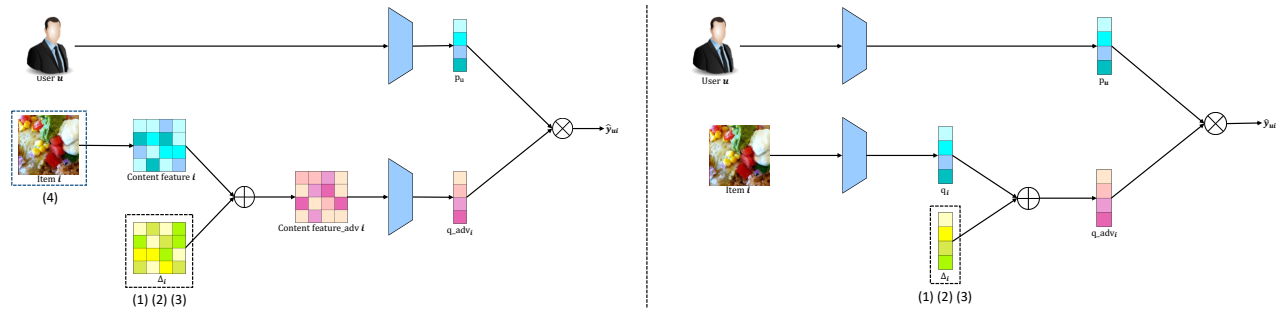
3.1 Information Level

As shown in Figure 2, information level categorizes existing adversarial training methods into two classes, i.e. data perturbations and feature perturbations, based on their ways to introduce the perturbations, as illustrated below.

3.1.1 Data perturbations. Data perturbations refer to adding adversarial perturbations to the model inputs, defined by

$$\mathbf{M_{adv}} = \mathbf{M} + \lambda\Delta \tag{1}$$

where $\mathbf{M_{adv}}$ denotes the perturbed data, $\mathbf{M}$ denotes the original data, Δ denotes the adversarial perturbations, and λ controls the magnitude of the adversarial perturbations. Existing studies add the adversarial perturbations to either the interaction data [1, 2, 20] or the pre-extracted content features of items [22].



(a) Data perturbations (b) Feature perturbations

Figure 2: Illustration on the adversarial training flowchart using BPR-MF for (a) data perturbations and (b) feature perturbations. (1) stochastic perturbations, (2) statistical perturbations, (3) gradient-based perturbations, and (4) GAN-based perturbations are introduced in different stages in the forward pass of the model.

3.1.2 Feature perturbations. Feature perturbations add adversarial perturbations to the intermediate features produced by the recommendation model, as defined by

$$\mathbf{v_{adv}} = \mathbf{v} + \lambda\Delta \tag{2}$$

where $\mathbf{v_{adv}}$ denotes the perturbed features, $\mathbf{v}$ denotes the original features, Δ denotes the adversarial perturbations on features, and λ controls the magnitude of the adversarial perturbations. As illustrated in (1) – (3) in Figure 2(b), existing studies mainly add the adversarial perturbations to either the item features [8] or to the features extracted by a CNN encoder [25].

3.2 Methodology Level

Methodology level defines four ways to generate perturbations including stochastic perturbations, statistic-based perturbations, gradient-based perturbations, and GAN-based perturbations. The details of these methods are described as follows.

3.2.1 Stochastic perturbations. Stochastic perturbations are produced by the generative functions with certain probability density functions, such as the normal distribution, as defined by

$$\Delta_{sto} \approx \mathcal{N}\left(u, \sigma^2\right) \tag{3}$$

where Δ_{sto} denotes stochastic perturbation, $\mathcal{N}(.)$ denotes a Gaussian distribution function, u and σ^2 denotes its expectation and variance, respectively. Existing studies [8, 22] figure out that using a simple Gaussian perturbation following the distribution N(0, 0.01) exhibits a better performance than using a fully stochastic perturbation, and this may improve the robustness of the commonly-used collaborative filtering algorithms, such as MF.

3.2.2 Statistical perturbations. Statistical perturbations usually follow a pre-defined distribution and therefore do not consider the statistics of the input data. This may harm the effectiveness of adversarial training when the generated perturbations are much smaller or larger than the input values. To address this issue, we propose the statistical perturbations, which generates the probabilistic distribution for stochastic perturbations based on the statistical information of the input data, as defined by

$$\Delta_{sta} \approx \mathcal{G}\left(u, \sigma^2, \mathbf{v}\right) \tag{4}$$

where Δ_{sta} denotes statistic-based perturbations, $\mathbf{v}$ denotes the original data/features, u, σ^2 denotes the pre-determined expectation and variance based on $\mathbf{v}$, and $\mathcal{G}(.)$ denotes a function which could generate statistic-based perturbations based on u, σ^2, and $\mathbf{v}$. To avoid the impact of bias in data or features in each epoch due to the training, we calculate the mean and distribution of v from the 10th to the 100th batch in each epoch and use it as parameters to generate statistical perturbations based on the normal distribution. As shown in Figure 2(a) and 2(b), statistical perturbations can be added both on data and on features.

3.2.3 Gradient-based perturbations. Gradient-based perturbations are generated based on the gradients back-propagated to optimize the model, which is usually obtained by

$$\Delta_{gra} \approx \epsilon\mathcal{H}(g) \tag{5}$$

where Δ_{gra} denotes gradient-based perturbations, ϵ controls the magnitude of gradient-based perturbations, g denotes the gradient information, and $\mathcal{H}(.)$ denotes a transformation function. Existing studies usually use the fast gradient method (FGM) [6, 8, 25] and the fast gradient sign method (FGSM) [6, 22] to generate the gradient-based perturbations, which are approximated by a linear function.

3.2.4 GAN-based perturbations. GAN-based perturbations are generated based on Generative Adversarial Networks (GAN) [5], which can capture the data distribution through an adversarial process between generators and discriminators and thus generate synthetic but realistic data. GAN-based perturbations can be defined as

$$\mathbf{v_{GAN}} = \mathbf{v} + \Delta_{GAN} \tag{6}$$

where $\mathbf{v_{GAN}}$ denotes virtual data generated by GAN, $\mathbf{v}$ denotes the original data, and Δ_{GAN} denotes GAN-based perturbations which are generated from $\mathbb{E}_{\mathbf{v}\sim p_{\text{data}}(\mathbf{v})}[\ln \mathcal{D}(\mathbf{v})] - \mathbb{E}_{\mathbf{z}\sim p_z(\mathbf{z})}[\ln \mathcal{D}(\mathcal{G}(\mathbf{z}))]$. As shown in (4) in Figure 2(a), a common approach for existing GAN-based perturbation research in recommendation is to generate virtual interactions based on GAN to perform data augmentation on the original interactions [1, 2, 20]. Notably, applying GAN-based perturbations to features has not been investigated so far.

3.3 Optimization Level

The optimization methods of existing adversarial training algorithms mainly include adversarial optimization and hybrid optimization, both of which can be treated as playing a minimax game, as defined by

$$\Theta^*, \Delta^* = \arg\min_{\Theta} \max_{\Delta, \|\Delta\| \leq \epsilon} \text{Loss} \tag{7}$$

where *Loss* indicates the loss terms used in a specific algorithm, the learning algorithm for the model parameters Θ is the minimizing player, while the procedure of generating perturbations Δ acts

as the maximizing player, whose aim is to generate worst-case perturbations for the current model.

3.3.1 Adversarial optimization. Adversarial optimization refers to using solely the adversarial samples to optimize the model. It holds by the hypothesis that iteratively training the model using the adversarial samples is sufficient. It defines the loss function as

$$\begin{aligned} \text{Loss} &= L(\mathbf{I}; \Theta, \Delta) \\ \text{where } \Delta &= \arg \max_{\Delta, \|\Delta\| \leq \epsilon} L(\mathbf{I}; \Theta, \Delta) \end{aligned} \tag{8}$$

where $\mathbf{I}$ denotes the input on the model, Θ denotes the current model parameters, Δ denotes the adversarial perturbations, $L(.)$ denotes the loss function of adversarial samples (the most commonly used in recommendation are Bayesian Personalized Ranking (BPR) [7] and cross-entropy (CE) [25] loss function), and $\epsilon \geq 0$ controls the magnitude of the perturbations. The data used for testing is the clean one without perturbations. Therefore, if adversarial optimization can still perform better in the testing stage, it could not only illustrate that adversarial training solely based on adversarial samples can also bring performance improvement, but also prove that this method effectively improves the model's robustness.

3.3.2 Hybrid optimization. Hybrid optimization means that the recommendation model is optimized by the combination of original samples and adversarial samples. To address the two requirements of being suitable for personalized ranking and remaining robustness against adversarial perturbations, in addition to minimizing the loss $L(\mathbf{I}; \Theta)$ of the original samples, hybrid optimization also regularizes the model by minimizing the loss $L(\mathbf{I}; \Theta, \Delta)$ of the adversarial samples. The loss function is defined as

$$\begin{aligned} \text{Loss } &= L(\mathbf{I}; \Theta) + \lambda L(\mathbf{I}; \Theta, \Delta), \\ \text{where } \Delta &= \arg \max_{\Delta, ||\Delta|| \leq \epsilon} L(\mathbf{I}; \Theta, \Delta) \end{aligned} \tag{9}$$

where $\mathbf{I}$ denotes the input on the model, Θ denotes the current model parameters, Δ denotes the adversarial perturbations, λ controls the magnitude of the adversarial term, $L(.)$ denotes the loss function of original and adversarial samples and $\epsilon \geq 0$ controls the magnitude of the perturbations. The optimization objective of the recommendation models in existing studies [8, 22, 25] is usually the original samples and the adversarial samples, i.e., minimizing the loss $L(\mathbf{I}; \Theta)$ of the original samples and the loss $L(\mathbf{I}; \Theta, \Delta)$ of the adversarial samples at the same time, and train it based on the minimax game mentioned above until convergence.

4 EXPERIMENTS

4.1 Experimental Setup

4.1.1 Datasets. We conduct experiments on the reorganized adversarial training algorithms based on the above taxonomy with two publicly available datasets. Table 1 summarizes the statistics of the datasets. One, called Allrecipes, was crawled from Allrecipes.com by Gao et al. in [4]. Each of the interactions in Allrecipes represents that the user has tried this recipe. The other is MeishiChina, which was crawled from a Chinese food sharing platform, MeishiChina, by Meng et al. in [15]. It is the first published Chinese food dataset so far, and each interaction in MeishiChina represents the user who has collected or commented on the recipe. We used the interaction information and image information from the above two datasets to verify the reorganized algorithms in the taxonomy we proposed.

Table 1: Statistics of the experimented datasets.

Datasets	#Interactions	#Users	#Items	#Sparsity
Allrecipes	1,093,845	68,768	45,630	99.97%
MehishiChina	515,082	64,029	72,796	99.99%

4.1.2 Evaluation measures. Five popular evaluation measures were employed to evaluate the performance of recommendation, including Precision (P), Recall (R), F1-Score (F), Normalized discounted cumulative gain (NDCG), and Area under ROC curve (AUC) [15]. Five hundred items are randomly selected from the dataset along with all positive items to form a ranking list for each user. P@K, R@K, F@K, and NDCG@K compute their performance for the Top-k ranked items in all sampled items. AUC measures the probability that a recommender will rank a positive sample higher than a randomly chosen negative one. To alleviate the problem of randomness, we repeated the evaluation process five times, and the average value is taken as the final performance.

4.1.3 Implementation details. Note that the purpose of this work is to propose an adversarial training-based taxonomy to alleviate the cold-start problems in recommendation, rather than developing new recommendation models. Therefore, we selected MF as the backbone of this study and optimized it with BPR. MF has been recognized as the basic yet most effective model in recommendation which represents each user and item in the form of an embedding vector, and its core idea is to estimate the user's preference score on an item as the inner product between their embedding vectors. BPR-MF (ResNet18) replaces the latent item embedding with the pre-extracted visual features extracted with ResNet18. To fairly evaluate all algorithms above, we fix the embedding size to 64, tune the learning rate in [0.001, 0.005, ..., 0.5] and optimize them based on Adagrad with the batch size of 64 for baseline models BPR-MF [8] and BPR-MF (ResNet18) [15]. For adversarial training algorithms, we tune ϵ in [0.0001, 0.001, ..., 1000] and λ in [0, 0.05, ..., 1].

4.2 Performance Comparison

This section reports the experimental performance of two baseline algorithms (BPR-MF and BPR-MF (ResNet18)) and eighteen reorganized adversarial training algorithms based on the taxonomy we proposed on two datasets, Allrecipes and MeishiChina. The hyperparameters of all algorithms were tuned based on the implementation details proposed in Section 4.1.3 to obtain the best performance. From Table 2, we can observe the following points:

- BPR-MF with pre-extracted virtual features of items achieves an increase on all performance measures than with items' latent embeddings on Allrecipes. The opposite effect was achieved on MeishiChina, which is due to its sparsity.
- Most of the reorganized adversarial training algorithms from the taxonomy obtain better results than baseline algorithms on all performance indicators of these datasets, which proves the necessity of applying adversarial training in recommendation.
- Adding perturbations to data achieves more competitive performance than adding perturbations to features among all the reorganized adversarial training algorithms with pre-extracted

Table 2: Performance comparison between two baseline algorithms and eighteen reorganized adversarial training algorithms based on our proposed taxonomy. (Base1: BPR-MF; Base2: BPR-MF (ResNet18) with visual features extracted by ResNet18; DP: Data perturbations; FP: Feature perturbations; STO: Stochastic perturbations; STA: Statistical perturbations; FGM: Fast gradient method; FGSM: Fast gradient sign method; AO: Adversarial optimization; HO: Hybrid optimization)

Item Representation	Algorithms	Allrecipe Dataset					MeishiChina Dataset				
		P@10	R@10	F@10	NDCG@10	AUC	P@10	R@10	F@10	NDCG@10	AUC
Latent Embedding	BPR-MF [15] (Base1)	0.0783	0.2635	0.0965	0.3139	0.8254	0.0554	0.2130	0.0584	0.2016	0.7675
	Base1+FP+STO+AO	0.0785	0.2645	0.0968	0.3146	0.8258	0.0555	0.2122	0.0583	0.2018	0.7672
	Base1+FP+STA+AO	0.0785	0.2646	0.0969	0.3149	0.8258	0.0560	0.2155	0.0591	0.2029	0.7672
	Base1+FP+FGM+AO	0.0792	0.2640	0.0971	0.3143	0.8255	0.0556	0.2131	0.0585	0.2019	0.7666
	Base1+FP+FGSM+AO	0.0792	0.2646	0.0973	0.3153	0.8272	0.0559	0.2149	0.0590	0.2023	0.7676
	Base1+FP+FGM+HO [8] (AMF)	0.0797	0.2659	0.0978	0.3166	0.8285	0.0572	**0.2183**	**0.0603**	**0.2071**	0.7703
	Base1+FP+FGSM+HO	**0.0800**	**0.2669**	**0.0982**	**0.3170**	**0.8295**	**0.0574**	0.2177	0.0602	0.2061	**0.7709**
Pre-Extracted Features	BPR-MF$_{ResNet18}$ [15] (Base2)	0.0799	0.2674	0.0981	0.3185	0.8036	0.0507	0.1974	0.0525	0.1972	0.7633
	Base2+DP+STO+AO	0.0821	0.2728	0.1005	**0.3232**	0.8226	0.0522	0.2076	0.0554	0.2026	0.7680
	Base2+DP+STA+AO	0.0820	0.2726	0.1005	0.3225	0.8298	0.0518	0.2013	0.0537	0.2016	0.7656
	Base2+DP+FGM+AO	0.0820	0.2726	0.1005	0.3222	0.8297	0.0520	0.2050	0.0547	0.2016	0.7685
	Base2+DP+FGSM+AO	0.0820	0.2726	0.1005	0.3220	0.8288	0.0523	0.2081	0.0557	0.2046	0.7701
	Base2+DP+FGM+HO [22] (AMR)	**0.0822**	**0.2733**	**0.1008**	0.3222	0.8287	0.0534	**0.2136**	**0.0573**	**0.2055**	**0.7753**
	Base2+DP+FGSM+HO	0.0821	0.2726	0.1006	0.3223	0.8293	**0.0537**	0.2098	0.0564	0.2048	0.7701
	Base2+FP+STO+AO	0.0819	0.2725	0.1005	0.3219	0.8339	0.0515	0.2002	0.0536	0.1999	0.7653
	Base2+FP+STA+AO	0.0804	0.2674	0.0985	0.3190	0.8054	0.0515	0.2018	0.0535	0.2015	0.7660
	Base2+FP+FGM+AO	0.0754	0.2441	0.0912	0.3000	0.7720	0.0510	0.1983	0.0528	0.1977	0.7633
	Base2+FP+FGSM+AO	0.0820	0.2723	0.1005	0.3222	**0.8369**	0.0500	0.1952	0.0516	0.1936	0.7626
	Base2+FP+FGM+HO	0.0818	0.2721	0.1003	0.3214	0.8331	0.0519	0.2010	0.0538	0.2005	0.7633
	Base2+FP+FGSM+HO	0.0820	0.2727	0.1005	0.3221	0.8329	0.0516	0.1989	0.0532	0.1988	0.7630

visual features. This is mainly because in this way recommendation models can mine adversarial samples directly and deeply so as to improve their performance on original samples.

- Adversarial training algorithms with gradient-based perturbations outperform those with stochastic or statistical perturbations in most performance metrics. It is due to the fact that gradient-based perturbations can maximize the objective function in the above minimax game, and then bring the best optimization effect.
- Adversarial training algorithms based on hybrid optimization outperform those based on adversarial optimization, and they both achieve better performance than the baseline algorithms. This not only verifies that both of the above-mentioned optimization methods can effectively improve the precision of recommendation models, but also demonstrate that these optimization methods promote the robustness of recommendation models.

4.3 Ablation Study for Performance of Cold-Start and Warm-Start Data

The imbalance of interactions in recommendation datasets is significant. To verify the effectiveness of adversarial training in alleviating the cold-start problems in recommendation, we divided Allrecipes and MeishiChina into two parts, i.e., the cold-start set with a few interactions and the warm-start set with a relatively large number of interactions. Finally, we conducted experiments to test the performance of the reorganized adversarial training algorithms on these sets. The observations form Table 3 can be drawn as following:

- **Adversarial training algorithms with pre-extracted visual features could alleviate cold-start problems:** "Base1" represents items with latent embeddings, and does not yield significant precision improvement; while "Base2" uses pre-extracted visual features, and achieves significant improvement on both two datasets and their divided datasets. This verifies that the importance of visual information in adversarial training in alleviating cold-start problems in recommendation.

Table 3: The recommendation performance of different baseline algorithms and reorganized adversarial training algorithms on the cold-start and warm-start sets.

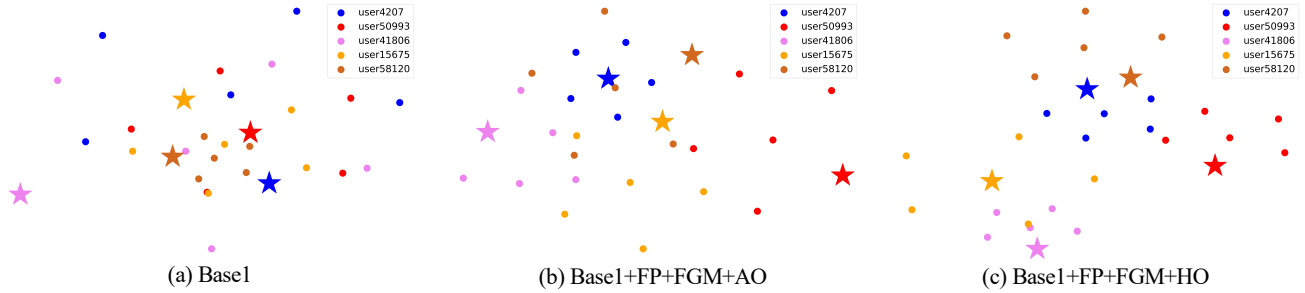
Datasets / Algorithms	Allrecipes			MeishiChina		
	Cold	Warm	All	Cold	Warm	All
BPR-MF [15] (Base1)	0.0348	0.1682	0.0783	0.0249	0.1210	0.0554
Base1+FP+STO+AO	0.0349	0.1679	0.0785	0.0246	0.1215	0.0555
Base1+FP+STA+AO	0.0349	0.1682	0.0785	0.0250	0.1214	0.0560
Base1+FP+FGM+AO	0.0346	0.1705	0.0792	0.0249	0.1209	0.0556
Base1+FP+FGSM+AO	0.0348	0.1712	0.0792	0.0248	0.1214	0.0559
Base1+FP+FGM+HO [8]	0.0349	0.1718	0.0797	**0.0253**	0.1247	0.0572
Base1+FP+FGSM+HO	**0.0351**	**0.1723**	**0.0800**	0.0250	**0.1252**	**0.0574**
Improvement of the best model over baseline	0.78%	2.42%	2.12%	1.61%	3.53%	3.53%

(a) Performance in Recommendation based on Latent Embedding Measured by Precision@10

Datasets / Algorithms	Allrecipes			MeishiChina		
	Cold	Warm	All	Cold	Warm	All
BPR-MF$_{ResNet18}$ [15] (Base2)	0.0352	0.1712	0.0799	0.0235	0.1080	0.0507
Base2+DP+STO+AO	0.0358	**0.1772**	0.0821	0.0246	0.1110	0.0522
Base2+DP+STA+AO	0.0358	0.1770	0.0820	0.0240	0.1100	0.0518
Base2+DP+FGM+AO	0.0358	0.1771	0.0820	0.0243	0.1105	0.0520
Base2+DP+FGSM+AO	0.0358	0.1765	0.0820	0.0246	0.1114	0.0523
Base2+DP+FGM+HO [22]	0.0358	0.1771	**0.0822**	**0.0252**	0.1138	0.0534
Base2+DP+FGSM+HO	**0.0359**	0.1769	0.0821	0.0246	**0.1152**	**0.0537**
Base2+FP+STO+AO	0.0358	0.1765	0.0819	0.0236	0.1095	0.0515
Base2+FP+STA+AO	0.0352	0.1733	0.0804	0.0242	0.1093	0.0515
Base2+FP+FGM+AO	0.0319	0.1649	0.0754	0.0236	0.1089	0.0510
Base2+FP+FGSM+AO	0.0358	0.1771	0.0820	0.0232	0.1066	0.0500
Base2+FP+FGM+HO	0.0357	0.1767	0.0818	0.0239	0.1110	0.0519
Base2+FP+FGSM+HO	0.0358	0.1769	0.0820	0.0237	0.1107	0.0516
Improvement of the best model over baseline	2.00%	3.46%	2.97%	7.24%	6.60%	6.03%

(b) Performance in Recommendation based on Pre-Extracted Features Measured by Precision@10

- **Adversarial training algorithms provide better alleviation of cold-start problems on "colder" datasets:** The average number of users' interactions for Allrecipes is 15.91, compared to 8.05 for MeishiChina. However, adversarial training algorithms achieve better precision enhancements on MeishiChina. This verifies that adversarial training could alleviate cold-start problems on "colder" datasets better.



(a) Base1 (b) Base1+FP+FGM+AO (c) Base1+FP+FGM+HO

Figure 3: Visualization of the learned t-SNE transformed representations derived from Base1, Base1+FP+FGM+AO, and Base1+FP+FGM+HO. Wherein, each star is a user, and the points with the same color denote the relevant items.

- **Adversarial training algorithms with gradient-based perturbations and hybrid optimization achieve the maximum performance improvement:** Adversarial training algorithm with gradient-based perturbations and hybrid optimization achieves the highest improvement, regardless of the baseline algorithms or the datasets. This indicates that gradient-based perturbations can maximize the objective function in the above minimax game, and the hybrid optimization method performs adversarial regularization on the basis of optimizing original samples, which jointly bring the highest improvement in cold-start recommendation.

4.4 Visualization of Embedding vectors transformed by t-SNE

In this section, we attempt to investigate how the adversarial training algorithms facilitate representation learning in the embedding space. For this purpose, we randomly selected five users and items they had interacted with, which have been fully trained in the training phase, to explore how their embedding representations change in different algorithms. Figures 3(a), 3(b), and 3(c) show the visualization of the representations derived from Base1, Base1+FP+FGM+AO, and Base1+FP+FGM+HO, respectively.

We notice that the connectivity of users and items is well reflected in the embedding space, that is, they are embedded in a proximal part. In particular, the representations of Base1+FP+FGM+AO, and Base1+FP+FGM+HO, based on adversarial training, exhibit clear clustering compared to Base1, i.e., points with the same color (items interacted with the same user) tend to form clusters. These observations verify that adversarial training algorithms can effectively facilitate the learning of representations in the embedding space, allowing the embeddings of users and items they interacted with tend to be closer. This may be a reason for the improved performance of adversarial training algorithms in recommendation.

4.5 Discussion

From the performance comparison between the existing methods and the reorganized algorithms based on the taxonomy in Section 4.2, it can be concluded that the state-of-the-art algorithms can be further improved by rational replacement of components in the taxonomy. We also found that the recommendation based on pre-extracted features is worse than that based on latent embedding in MeishiChina, which may due to the combination of the more 'colder' dataset and the pre-trained model based on ImageNet cannot effectively extract food visual features. As shown in Table 3, by comparing the precision between the reorganized algorithms with the existing studies [8, 22] on the cold-start and warm-start sets, we found that the reorganized algorithms achieve better performance in most of the divided datasets. At the same time, we also found that adding visual information to the adversarial training can improve its ability in alleviating the cold-start problems in recommendation by contrasting the improvements of the best adversarial training algorithms over the baseline algorithm. From Figure 3, we discover that adversarial training can promote the learning of the representations of users and items in the embedding space, so that the embedding of items tends to be closer to the users they have interacted with and further away from the users they have never interacted, and thus improve the precision of recommendation model. This is possibly the reason why adversarial training algorithms can effectively alleviate the cold-start problems in recommendations.

5 CONCLUSION

This paper presents a taxonomy that may serve as a design paradigm of adversarial training algorithms for cold-start recommendation. Existing adversarial training methods for recommendation typically consider one aspect of perturbation generation or incorporation process, and the underlying principles on how these methods work have not been fully-verified. The taxonomy addresses this issue by dismantling and reorganizing existing adversarial training algorithms at three levels of key components, including information, methodology, and optimization, and then exploring their effectiveness and possible mechanism for cold-start recommendations. Experimental results show that the reorganized adversarial training algorithms based on this taxonomy outperform existing studies in some cases, which reflects the effectiveness of the taxonomy. Finally, this paper demonstrates through visualization that adversarial training alleviates the cold-start problem by further facilitating the learning of user and item representations in the embedding space.

Future work of this research may focus on two directions. First, we plan to conduct full coverage experiments for each reorganized algorithm of the taxonomy to analyze the applicability of each method through it. Second, we desire to propose a new algorithm based on the taxonomy that could improves its ability to represent users and items through adversarial training, and thus achieves the best performance on all evaluation measures for each datasets.

ACKNOWLEDGMENTS

This work is supported in part by the National Natural Science Foundation of China (Grant no. 62006141) and the Special Project of Science and Technology Innovation Base of Key Laboratory of Shandong Province for Software Engineering (Project ID:11480004042015).

REFERENCES

[1] Dong-Kyu Chae, Jin-Soo Kang, Sang-Wook Kim, and Jaeho Choi. 2019. Rating augmentation with generative adversarial networks towards accurate collaborative filtering. In *The World Wide Web Conference*. 2616–2622.

[2] Dong-Kyu Chae, Jihoo Kim, Duen Horng Chau, and Sang-Wook Kim. 2020. AR-CF: Augmenting Virtual Users and Items in Collaborative Filtering for Addressing Cold-Start Problems. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1251–1260.

[3] Fuli Feng, Xiangnan He, Jie Tang, and Tat-Seng Chua. 2019. Graph adversarial training: Dynamically regularizing based on graph structure. *IEEE Transactions on Knowledge and Data Engineering* (2019).

[4] Xiaoyan Gao, Fuli Feng, Xiangnan He, Heyan Huang, Xinyu Guan, Chong Feng, Zhaoyan Ming, and Tat-Seng Chua. 2019. Hierarchical attention network for visually-aware food recommendation. *IEEE Transactions on Multimedia* 22, 6 (2019), 1647–1659.

[5] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).

[6] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. 2014. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572* (2014).

[7] Ruining He and Julian McAuley. 2016. VBPR: visual bayesian personalized ranking from implicit feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 30.

[8] Xiangnan He, Zhankui He, Xiaoyu Du, and Tat-Seng Chua. 2018. Adversarial personalized ranking for recommendation. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 355–364.

[9] Dong-Gyun Hong, Yeon-Chang Lee, Jongwuk Lee, and Sang-Wook Kim. 2019. CrowdStart: Warming up cold-start items using crowdsourcing. *Expert Systems with Applications* 138 (2019), 112813.

[10] Xiaopeng Li and James She. 2017. Collaborative variational autoencoder for recommender systems. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 305–314.

[11] Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. 2014. Facing the cold start problem in recommender systems. *Expert Systems with Applications* 41, 4 (2014), 2065–2073.

[12] Siwei Liu, Iadh Ounis, Craig Macdonald, and Zaiqiao Meng. 2020. A heterogeneous graph neural model for cold-start recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2029–2032.

[13] Yang Liu, Xianzhuo Xia, Liang Chen, Xiangnan He, Carl Yang, and Zibin Zheng. 2020. Certifiable robustness to discrete adversarial perturbations for factorization machines. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 419–428.

[14] Yuanfu Lu, Yuan Fang, and Chuan Shi. 2020. Meta-learning on heterogeneous information networks for cold-start recommendation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1563–1573.

[15] Lei Meng, Fuli Feng, Xiangnan He, Xiaoyan Gao, and Tat-Seng Chua. 2020. Heterogeneous Fusion of Semantic and Collaborative Information for Visually-Aware Food Recommendation. In *Proceedings of the 28th ACM International Conference on Multimedia*. 3460–3468.

[16] Chanyoung Park, Donghyun Kim, Jinoh Oh, and Hwanjo Yu. 2016. Improving top-K recommendation with truster and trustee relationship in user trust network. *Information Sciences* 374 (2016), 100–114.

[17] Surabhi Punjabi and Priyanka Bhatt. 2018. Robust factorization machines for user response prediction. In *Proceedings of the 2018 World Wide Web Conference*. 669–678.

[18] Ali Shafahi, Mahyar Najibi, Amin Ghiasi, Zheng Xu, John Dickerson, Christoph Studer, Larry S Davis, Gavin Taylor, and Tom Goldstein. 2019. Adversarial training for free! *arXiv preprint arXiv:1904.12843* (2019).

[19] Chuan Shi, Yitong Li, Jiawei Zhang, Yizhou Sun, and S Yu Philip. 2016. A survey of heterogeneous information network analysis. *IEEE Transactions on Knowledge and Data Engineering* 29, 1 (2016), 17–37.

[20] Zhongchuan Sun, Bin Wu, Yunpeng Wu, and Yangdong Ye. 2019. Apl: Adversarial pairwise learning for recommender systems. *Expert Systems with Applications* 118 (2019), 573–584.

[21] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. 2013. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199* (2013).

[22] Jinhui Tang, Xiaoyu Du, Xiangnan He, Fajie Yuan, Qi Tian, and Tat-Seng Chua. 2019. Adversarial training towards robust multimedia recommender system. *IEEE Transactions on Knowledge and Data Engineering* 32, 5 (2019), 855–867.

[23] Maksims Volkovs, Guang Wei Yu, and Tomi Poutanen. 2017. DropoutNet: Addressing Cold Start in Recommender Systems.. In *NIPS*. 4957–4966.

[24] Hao Wang, Naiyan Wang, and Dit-Yan Yeung. 2015. Collaborative deep learning for recommender systems. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 1235–1244.

[25] Feng Yuan, Lina Yao, and Boualem Benatallah. 2019. Adversarial collaborative neural network for robust recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1065–1068.

[26] Yu Zhu, Jinghao Lin, Shibi He, Beidou Wang, Ziyu Guan, Haifeng Liu, and Deng Cai. 2019. Addressing the item cold-start problem by attribute-driven active learning. *IEEE Transactions on Knowledge and Data Engineering* 32, 4 (2019), 631–644.