Non-Compensatory Psychological Models for Recommender Systems

ID: 335

Abstract

The study of consumer psychology reveals two categories of consumption decision procedures: compensatory rules and non-compensatory rules. Existing recommendation models which are based on latent factor models assume the consumers follow the compensatory rules, i.e. the consumer evaluate an item over multiple aspects and compute a weighted or/and summated score which is further used to derive the rating or the rankings among items. However, it has been shown in the literature of consumer psychology that, consumers adopt non-compensatory rules more than compensatory rules. Our main contribution in this paper is to study the unexplored utilization of non-compensatory rules in recommendation models.

Our general assumptions are (1) there are K universal hidden aspects. In each evaluation session, only one aspect is chosen as the prominent aspect according to user preference. (2) Evaluations over prominent and non-prominent aspects are non-compensatory. Evaluation is manly based on item performance on the prominent aspect. For non-prominent aspects the user sets a minimal acceptable value. We give a conceptual model for these general assumptions and show how this model can be applied to a wide range of existing recommender systems, including point-wise rating prediction models and pair-wise ranking prediction models. We experimentally show that adopting non-compensatory rules constantly improve ranking performance of existing models on a variety of real-world recommendation data sets.

Introduction

The majority of state-of-the-art recommendation models are based on latent factor models. Generally, latent factor models transform both user preferences and item features into the same hidden feature spaces with K aspects. To recover the observations (i.e. ratings or rankings) in any recommender system, they adopt the inner product of the user preferences and the item features. There are fruitful successful applications of latent factor models in rating predictions (Koren, Bell, and Volinsky 2009; Koren 2010; Lee et al. 2014) and ranking reconstructions (Rendle et al. 2009; Steck 2015; Zhao et al. 2018; Shi, Larson, and Hanjalic 2010).

From the perspective of consumer decision making, all existing latent factor models fall into the category of *compensatory rules*. Consumers who adopt compensatory rules evaluate every item over multiple aspects and compute a weighted or/and summated score for each item. Then they will rate or rank items based on the score. The key property of compensatory rules is that a good performance on one aspect of an item compensates for poor performances on other aspects.

However, in the study of human choice behavior, it is well regarded that there are two categories of decision making procedures, namely *compensatory rules* and *non-compensatory rules*. Furthermore, it is found in many surveys (Engel, Blackwell, and Miniard 1986) that in most cases consumers make consumption related choices based on non-compensatory rules. Non-compensatory rules do not allow the shortcomings of a product to be balanced out by its attractive features. We next illustrate two typical non-compensatory rules.

Example. Alice wants to buy a smart phone and she ranks her alternatives over three relevant aspects: battery life, price and storage space. If Alice's priority is long-lasting battery, than she will adopt *lexicographic rules*, a type of non-compensatory rules, to rank phones first based on the battery life. Clearly Honor and iPhone will be ranked higher than Galaxy, the other benefits offered by Galaxy do not outweigh her desire for a long-life battery . If Alice has a second layer of demands that she wants the phone to be cheap and with plenty of storage space, then she will adopt *conjunctive rules*, another type of non-compensatory rules, to set cut-off points (e.g. 600\$ and $64GB$ on the corresponding aspects). iPhone fails to meet the cut-off point, it will not outrank Honor which satisfies the minimal acceptable value on each aspect. In either case, adopting a compensatory rule based recommendation model (e.g. MF) is problematic.

Table 1: Illustrative example of non-compensatory rules

Item	Prominent aspect	Not prominent aspects	
	Battery life	Price	Memory
iPhone SE	13	700\$	64GB
Galaxy S8	9	500\$	128GB
Honor 10	24	589\$	128GB

Non-compensatory rules are widely used in many

Decision **S**upport **S**ystems (DSS) (Lee 2009). Despite of the commercial success, current DSS are labor costly, i.e. they ask consumers to control or manipulate the rules, e.g. specify the value of an aspect. To the best of our knowledge, no previous work has been devoted to modeling and learning non-compensatory rules in recommender systems. Our goal in this paper is to study this unexplored area.

Our primary contribution is to give a conceptual model of how users adopt non-compensatory rules in recommender systems. We assume that, (1) there are $K-$ hidden aspects which user preferences and item features are transformed into, (2) in each evaluation session, the user picks a prominent aspect according to his/her preference, (3) the user adopts different evaluation strategies on prominent and non-prominent aspects. The evaluation is mainly based on item performance on the prominent aspect. The evaluation is less influenced by item performance with respect to a user-defined minimal acceptance value on non-prominent aspects.

Our second contribution is to realize the conceptual model in a wide range of recommendation frameworks, including point-wise rating prediction models such as the conventional Matrix Factorization (MF (Koren, Bell, and Volinsky 2009)), Matrix Factorization with neighborhood collaborative filtering (AMF (Koren 2008)), and locally low-rank matrix approximation (LLORMA (Lee et al. 2013)) and pairwise ranking reconstruction models such as BTL model (Hu and Li 2016) and BPR style Thurstonian model (Rendle et al. 2009).

We conduct comprehensive experiments on a variety of real world data sets. We experimentally show that the non-compensatory versions of these models significantly improve ranking performances of the original models.

The paper is organized as follows. In Sec. , we start with surveying the most commonly adopted latent factor models in the community of recommendation research. We categorize previous research work on the basis of combinations of different rating approximation formulas and loss functions. In Sec. , we describe our non-compensatory assumptions and develop non-compensatory versions of existing models. In Sec. , we experimentally show that the non-compensatory versions significantly outperform the original versions of existing models on a variety of real-world data sets. Finally, in Sec. we conclude our work and future directions.

Compensatory Recommendation Models

In this section, we summarize and categorize recommendation models based on the rating prediction formulas and loss functions. We restrict our discussions to latent factor models, i.e. models where a universe of K factors is used to project user preferences and item features. Hereafter, unless stated otherwise, we use lower-case letters for indices, upper-case letters for universal constants, lower-case bold-face letters for vectors and upper-case bold-face letters for matrices. Specifically, $\mathbf{X} \in \mathcal{R}^{M\times N}$ denotes the rating matrix, $\hat{\mathbf{X}} \in \mathcal{R}^{M\times N}$ denotes the predicted rating matrix, $\mathbf{p}, \mathbf{q} \in \mathcal{R}^K$ denotes the item features, $\mathbf{u} \in \mathcal{R}^K$ denotes the user preferences.

Rating Prediction Formulas

One goal of recommendation research is to recover the rating matrix $\mathbf{X}$, by minimizing a loss function $\mathcal{L}(\mathbf{X}, \hat{\mathbf{X}})$, which is usually defined as the regularized square loss between the predicted rating $\hat{\mathbf{X}}_{u,q}$ and the observed rating $\hat{\mathbf{X}}_{u,q}$ for each user u who has rated item q. We list some of the most successful rating prediction formulas for $\hat{\mathbf{X}}$.

Matrix Factorization. In conventional matrix factorization (Koren, Bell, and Volinsky 2009), the predicted rating can be computed as an inner product of user preferences and item features as follows.

$$\hat{\mathbf{X}}_{u,q} = \sum_{k=1}^{K} \mathbf{q}_k \mathbf{u}_k \tag{1}$$

For simplicity we ignore the user specific or item specific bias (Koren, Bell, and Volinsky 2009). A massive amount of techniques have been proposed based on Equ. 1. Most of them modified the loss function ,e.g. by incorporating prior distributions over $\mathbf{p}, \mathbf{u}$ (Salakhutdinov and Mnih 2008), adding priors over unknown values (Devooght, Kourtellis, and Mantrach 2015), weighing different samples (Pilászy, Zibriczky, and Tikk 2010) and so on.

Neighborhood Factorization. in traditional memory based collaborative filtering strategies, neighborhood information has been proved to be useful. It is possible to embed such neighborhood information in latent factor models. Instead of directly modeling user preferences $\mathbf{u}$, each user is represented by items that he/she gives explicit or implicit feedback. For example, if we consider explicit feedback only, then each item is associated with two types of vectors $\mathbf{p}, \mathbf{q}$, the rating prediction formula of AMF in (Koren 2008) is stated as follows.

$$\hat{\mathbf{X}}_{u,q} = \sum_{k=1}^{K} \mathbf{q}_k \Big(\sum_{p\in R(u)} \mathbf{p}_k / \sqrt{|R(u)|} \Big), \tag{2}$$

where $R(u)$ is the set of rated items for u. AMF has been extended to SVD++ (Koren 2008) with implicit feedback.

Local Low-Rank Matrix Approximation. The third type of rating prediction formula is LLORMA (Lee et al. 2013). The intuition of LLORMA is that the entire rating matrix $\mathbf{X}$ is not low-rank but a sub-matrix restricted to a neighborhood of similar users and items is low-rank. Therefore, the predicted rating is aggregated over S sub-matrices as follows:

$$\hat{\mathbf{X}}_{u,q} = \sum_{t=1}^{S} \sum_{k} \mathbf{u}_{t,k} \frac{K((\mathbf{u}_t, \mathbf{i}_t), (\mathbf{u}, \mathbf{q}))}{\sum_{s=1}^{S} K((\mathbf{u}_s, \mathbf{i}_s), (\mathbf{u}, \mathbf{q}))} \mathbf{q}_{t,k} \tag{3}$$

$\mathbf{u}_t, \mathbf{q}_t$ are the factorized user preferences and item features in the $t-$th sub-matrix, $\mathbf{i}_s, \mathbf{i}_t$ are anchor points in the corresponding matrix to locate a neighborhood for low-rank decomposition, $K(\cdot)$ is a smoothing kernel.

Ranking Models

Another goal of recommendation research is to reveal the observed rankings. We here consider pair-wise rankings $p \succ_u q$, where user u prefers item p over q. The pair-wise rankings can be generated from pre-processing the ratings, i.e. $\mathbf{X}_{u,p} > \mu, \mathbf{X}_{u,q} < \mu$ (Hu and Li 2017), or from explicit and implicit feedback, i.e. $\mathbf{X}_{u,p} \neq 0$ and $\mathbf{X}_{u,q}$ doesn't exist (Rendle et al. 2009).

A large body of previous research has been presented by employing a ranking aware loss function $\mathcal{L}(p(p \succ_u q), o(p \succ_u q))$, where $p(p \succ_u q)$ is the predicted possibility and $o(p \succ_u q)$ is an indicator function of whether or not the ranking is observed. To generate the probability of pair-wise rankings $p(p \succ_u q)$, each user-item combination is associated with a score, i.e. $\hat{\mathbf{X}}_{u,p}, \hat{\mathbf{X}}_{u,q}$. We list two most commonly adopted ranking models .

Thurstone Model The most frequently adopted ranking model in recommendation systems is the Thurstone model (Thurstone 1927) which uses a non-linear transformation of the predicted ratings.

$$p(p \succ_u q) = \frac{1}{1 + \exp[-(\hat{\mathbf{X}}_{u,p} - \hat{\mathbf{X}}_{u,q})]} \tag{4}$$

Bradley-Terry Model. The famous BTL model (Hunter 2004) is extensively studied in learning to rank scenarios. BTL models the generation of ranking pairs by a division.

$$p(p \succ_u q) = \frac{\exp \hat{\mathbf{X}}_{u,p}}{\exp \hat{\mathbf{X}}_{u,p} + \exp \hat{\mathbf{X}}_{u,q}} \tag{5}$$

In either ranking model, the score $\hat{\mathbf{X}}$ can be approximated by different ranking prediction formulas. We categorize existing ranking aware methods based on the combination of rating prediction formulas and the ranking models.

- BTL model has been leveraged with MF prediction formula in (Hu and Li 2016);
- Thurstone model with standard matrix factorization prediction formula is first presented as BPR (Rendle et al. 2009), which maximizes the Bayesian posterior with respect to Thurstonian modeling of standard matrix factorization predictions. Abundant research has been carried out to improve BPR-style systems by modifying the sampling methods in optimization, including BTR++ (Lerche and Jannach 2014), WARP (Weston, Bengio, and Usunier 2011), DNS (Zhang et al. 2013), RankMBPR (Yu et al. 2016) and so on.
- Thurstone model with neighborhood factorized prediction formula AMF is first incorporated in a point-wise ranking framework In (Steck 2015), FSBPR (Zhao et al. 2018) implants AMF in a Thurstone model and maximizes its likelihood.
- Thurstone model with local low-rank factorization prediction formula is utilized in LCR (Lee et al. 2014).

The list is by no means exclusive. However, we believe that most of existing recommender systems are covered. It is worthy to point out that (1) we do not restrict the form of loss functions. For example, many ranking approaches consider Bayesian maximum posterior, cross entropy and other forms of loss functions. Nevertheless, the core ranking model is either BTL or Thurstone. (2) Although we only study pair-wise ranking , the conclusion is insightful for other ranking-aware systems, i.e. point-wise and list-wise approaches. The reason is that, as shown in (Steck 2015), point-wise and list-wise loss functions can be decomposed to components which are directly based on each score $\hat{\mathbf{X}}_{u,p}$ and components that are not related to $\hat{\mathbf{X}}$. Thus our proposed strategy in Sec. is also applicable to point-wise and list-wise ranking models.

Non-Compensatory Recommendation Models

We begin this section by reviewing the findings in consumer psychology study. Ever since the dawn of consumption psychology study, psychologists have been studying how consumers adopt different heuristics to facilitate brand (or other consumption related) choices. Two distinct categories of decision rules are found (Engel, Blackwell, and Miniard 1986): compensatory rules and non compensatory rules. The decision rules can be naturally explained in the latent factor models. For example, compensatory rules are adopted if a consumer determines options in terms of each factor and computes a weighted or summated score for each item, then selects the item that scores the highest among the alternatives evaluated. It is clear that all related work that has been described in previous section is the application of compensatory rules.

Non-compensatory rules include *lexicographic, conjunction* and *disjunction* rules. The conjunctive and disjunctive rules are often used in conjunction with lexicographic rules.

We illustrate the non-compensatory rules using three toy item vectors $\mathbf{p}, \mathbf{q}, \mathbf{i}$ in Table. . Under lexicographic rules, the user will first rank the factors, e.g. pick the prominent aspect according to user preference. Then items are evaluated on the prominent aspect, i.e. $p \succ_u q \succ_u i.$. Under conjunctive and disjunctive rules, the consumer imposes requirements for minimally acceptable values on each aspect separately. Suppose the user's minimal acceptable value is 0.3. If the user adopts conjunctive rules only, then the chosen items must be better than the minimal acceptable value on all aspects, i.e. $p.i$ will be picked. If the user adopts disjunctive rules only, then the chosen items must satisfy the minimal acceptable value on at least one aspect, i.e. p, q, i are all chosen. Finally, these rules an be combined. If the user adopts both lexicographic rules and conjunctive rules, then the result is $p \succ_u i$ and i will not be chosen.

We can see that non-compensatory rules differ from compensatory rules in two key points. (1) *Distinguished factors*. In compensatory rules, different factors are essentially equivalent (i.e. all factors contribute to the final score), while in non-compensatory rules factors are not interchangeable (i.e. only the prominent factor is considered if there are no ties). (2) *Distinguished evaluation metrics on each factor*. In compensatory rules, the evaluations on each factor follow the same framework (i.e. a product of user preference and item feature on the specific factor), while in non-compensatory rules, the evaluations on each factor are dissimilar (i.e. numerical comparisons on the prominent factor and acceptance/rejection on other factors).

For computational convenience, inspired by the psychological findings, we present the following two assumptions based on lexicographic and conjunction rules. (1) We assume that in each evaluation session[1], there is a prominent aspect. The choice of the prominent aspect is dependent on the user preferences. (2) We assume two types of evaluation strategies are adopted, one for the prominent aspect and the other for other non-prominent aspects. According to the above assumptions, we provide the non-compensatory versions of rating prediction formulas and ranking models.

Non-Compensatory Rating Prediction Formulas

Our goal here is to modify the rating prediction formulas as little as possible, while still preserving the most important properties of non-compensatory rules. Therefore, we follow the same notations for user preferences and item features. In each evaluation session, the hidden prominent aspect is sampled by $\frac{\exp \mathbf{u}_k}{\sum_{k'} \mathbf{u}_{k'}}$. We use a parameter θ to control the strength of prominent aspect, i.e. the evaluation on the prominent aspect is magnified by $\exp\theta$. The prediction is generated across all possible hidden prominent aspects. This gives us the following non-compensatory versions of rating prediction formulas.

Matrix Factorization: MF-NCR

$$\hat{\mathbf{X}}_{u,q} = \sum_{k=1}^{K} \frac{\exp \mathbf{u}_k}{\sum_{k'} \exp \mathbf{u}_{k'}} [\exp\theta \mathbf{q}_k + \sum_{k' \neq k} \mathbf{q}_{k'}]. \quad (6)$$

Neighborhood Factorization: AMF-NCR implements a similar scheme by setting $u_k = \sum_{p \in R(u)} \mathbf{p}_k / \sqrt{|R(u)|}$,

$$\hat{\mathbf{X}}_{u,q} = \sum_{k=1}^{K} \frac{\exp(\sum_{p \in R(u)} \mathbf{p}_k)}{\sum_{k'} \exp(\sum_{p \in R(u)} \mathbf{p}_{k'})} [\exp\theta \mathbf{q}_k + \sum_{k' \neq k} \mathbf{q}_{k'}]. \quad (7)$$

LLORMA-NCR uses the same decomposition for each sub-matrix.

$$\hat{\mathbf{X}}_{u,q} = \sum_{t=1}^{S} \sum_k \frac{\exp \mathbf{u}_k}{\sum_{k'} \exp \mathbf{u}_{k'}} \frac{K((\mathbf{u}_t,\mathbf{i}_t),(\mathbf{u},\mathbf{q}))}{\sum_{s=1}^{S} K((\mathbf{u}_s,\mathbf{i}_s),(\mathbf{u},\mathbf{q}))} [\exp\theta \mathbf{q}_{t,k} + \sum_{k' \neq k} \mathbf{q}_{t,k'}] \quad (8)$$

We can see that all these NCR versions are combinations of lexicographic and conjunction rules, where $\exp\theta \to \infty$ indicates that the user adopts lexicographical rules only.

Non-Compensatory Ranking Models

Thurston-NCR. The modification of Thurston model is straightforward, as the ranking probability involves a subtraction component of $\hat{\mathcal{R}}_{u,q}$ which can be rreplaced by any NCR-version of rating prediction formulas.

[1]The evaluation session could be either a true user interaction session with multiple actions, or a pseudo session which contains one rating action. The impact of availability of session information is discussed in experiments.

Inference of Thurston models is easily extensible. For example, if we use the Bayesian maximum posterior estimator as in BPR (Rendle et al. 2009), the loss function is defined as:

$$\mathcal{L} = -\sum_u \sum_{p \succ_u q} \ln \frac{1}{1 + \exp -[\hat{\mathbf{X}}_{u,p} - \hat{\mathbf{X}}_{u,q}]} - \lambda \|\Theta\|, \quad (9)$$

where Θ is the set of all parameters. Thus the inference procedure is accomplished by stochastic gradient descent (SGD) with $\frac{\partial \mathcal{L}}{\partial \Theta} = \sum_u \sum_{p \succ_u q} \frac{\partial \mathcal{L}}{\partial \Delta \hat{\mathbf{X}}_{u,p,q}} \frac{\partial \Delta \hat{\mathbf{X}}_{u,p,q}}{\partial \Theta}$, where $\Delta \hat{\mathbf{X}}_{u,p,q} = \hat{\mathbf{X}}_{u,p} - \hat{\mathbf{X}}_{u,q}$.

BTL-NCR. Finally we propose the non-compensatory version of BTL ranking model. In order to treat prominent and non-prominent aspects differently, we define the probability of any ranking pair $p \succ_u q$ as the product of results by factor-wise comparisons, based on a variant of BTL model with ties (Hunter 2004). Again, in each evaluation session, a hidden prominent aspect k is sampled by user preference $\mathbf{u}$. The overall prediction is aggregated over all possible hidden prominent aspect k.

$$p(p \succ_u q) = \prod_{k=1}^{K} \mathbf{u}_k [\frac{\mathbf{p}_k}{\mathbf{p}_k + \theta \mathbf{q}_k} \prod_{k' \neq k} \frac{\theta \mathbf{p}_{k'}}{\mathbf{q}_{k'} + \theta \mathbf{p}_{k'}}]. \quad (10)$$

where $\mathbf{u}_k > 0, \sum_k \mathbf{u}_k = 1$ and $\theta > 1$. BTL-NCR models the non-compensatory rules in a manner that (1) the evaluation is mainly based on the prominent aspect. The item p is more likely to be preferred than q by user u if p is significantly better than q on the prominent aspect, i.e. $p_k > \theta q_k, \theta > 1$. (2) The performance on other aspects are less important. Because p is considered to be as good as q, as long $\forall k' \neq k, \theta p_{k'} > q_{k'}, \theta > 1$. BTL-NCR is also a combination of lexicographic rules and conjunction rules. An interpretation is that we dynamically set a minimal acceptance value for $p_{k'}$ on factor $k' \neq k$ based on the compared alternative $q_{k'}$, where the minimal acceptance value is $q_{k'}/\theta$. The parameter θ controls the tolerance range. When $\theta \to \infty$, the users adopt lexicographic rules only.

To infer the parameters of BTL-NCR, we implement a stochastic expectation maximization (SEM) algorithm. In each E-step, we first draw the value of prominent aspect k for each evaluation session by

$$k \sim u_k^t \frac{\mathbf{p}_k^t}{\mathbf{p}_k^t + \theta^t \mathbf{q}_k^t} \prod_{k' \neq k} [\frac{\theta^t \mathbf{p}_{k'}^t}{\mathbf{q}_{k'}^t + \theta^t \mathbf{p}_{k'}^t}]. \quad (11)$$

where t indicates the value obtained from the $t-$th round of SEM algorithm. In each M-step, we incorporate the MM bound in (Hunter 2004) and maximize the log-likelihood of complete data.

Experiments

We conduct experiments to evaluate the performance of non-compensatory rules in recommendation models. We conduct three sets of experiments on real world datasets. The first set of experiments is conducted to examine whether the NCR versions of rating prediction models outperform the original versions on rating data sets. The second set of experiments is conducted to examine whether NCR versions of ranking aware models outperform the original versions on data sets with implicit feedback. The third set of experiments is conducted to examine whether NCR versions of ranking aware models outperform the original versions on data sets with user interaction session information.

Rating Prediction Performance

Data Sets We use the standard benchmarking datasets with user-item ratings. (1) Movielens (2) FilmTrust (3) CiaoDVD. The ratings are in the range of $1-5$ stars. Statistics of the datasets are described in Table. 5. For each dataset, we reserve users with at least 5 ratings and randomly split training and test set by avoiding cold-start users and items. We consider each rating as an evaluation session. The reported results are averaged using 5-fold cross validation,

Table 2: Statistics of Datasets with ratings and implicit feedback

Dataset	#users	#items	#ratings	#sessions
Movielens	942	1650	80000	4641262
FilmTrust	1235	2062	35497	623516
CiaoDVD	2665	14280	72665	2478836

Comparative Methods. We compare the NCR improved versions with the original versions on three widely adopted rating prediction methods (1) MF (Koren, Bell, and Volinsky 2009): standard matrix factorization (2) AMF (Koren 2008): neighborhood factorization (3) LLORMA (Lee et al. 2013): local low-rank matrix factorization.

Evaluation Metrics. The goal is to reconstruct the observed user-item ratings as accurate as possible. Hence we evaluate different approaches based on the following metrics. (1) AUC: (2) NDCG (3) RMSE (4) MAE (5) MRR

Table 3: Comparative rating prediction performance

Dataset	Method	AUC	NDCG	RMSE	MAE	MRR
Movielens	MF					
	MF-NCR					
	AMF	0.6043	0.5003	1.1682	0.9924	0.7506
	AMF-NCR	0.6129	0.5027	1.4392	1.2560	0.7559
	LLORMA		0.8990	0.9291	0.7322	0.5761
	LLORMA-NCR		0.8994	0.9217	0.7288	0.5761
Filmtrust	MF					
	MF-NCR					
	AMF	0.6244	0.5055	0.8601	0.7448	0.7622
	AMF-NCR	0.6436	0.5098	1.1535	0.9330	0.7717
	LLORMA		0.8672	0.8341	0.6437	0.6481
	LLORMA-NCR		0.8684	0.8310	0.6417	0.6533
CiaoDVD	MF					
	MF-NCR					
	AMF	0.6211	0.5048	1.2025	1.0261	0.7607
	AMF-NCR	0.7993	0.5657	1.1475	0.9723	0.8950
	LLORMA		0.7827	1.0436	0.8188	0.4883
	LLORMA-NCR		0.7838	1.0608	0.8271	0.4904

Ranking Performance for Implicit Feedback

Comparative Methods. We compare the NCR improved versions with the original versions on three widely adopted rating prediction methods (1) MF (Koren, Bell, and Volinsky 2009): standard matrix factorization (2) AMF (Koren 2008): neighborhood factorization (3) LLORMA (Lee et al. 2013): local low-rank matrix factorization.

Evaluation Metrics. The goal is to reconstruct the observed user-item ratings as accurate as possible. Hence we evaluate different approaches based on the following metrics. (1) AUC: (2) NDCG (3) RMSE (4) MAE (5) MRR

Table 4: Comparative ranking prediction performance with implicit feedback

Dataset	Method	MAP	NDCG	Prec	Recall	MRR
Movielens	BTL	0.7654	0.5070	0.5307	0.7553	0.7654
	BTL-NCR	0.8440	0.5425	0.6879	0.8340	0.8440
	BPR	0.8478	0.5443	0.6956	0.8478	0.8478
	BPR-NCR	0.8623	0.5508	0.7246	0.8623	0.8623
	FSBPR	0.7474	0.4993	0.4968	0.7583	0.7484
	FSBPR-NCR	0.7964	0.5205	0.5908	0.7856	0.7954
	LCR					
	LCR-NCR					
Filmtrust	BTL	0.7674	0.5070	0.5307	0.7551	0.7654
	BTL-NCR	0.8182	0.5312	0.6379	0.8089	0.8190
	BPR	0.7825	0.5147	0.5649	0.7825	0.7825
	BPR-NCR	0.8365	0.5392	0.6730	0.8365	0.8365
	FSBPR	0.7484	0.4996	0.4980	0.7387	0.7490
	FSBPR-NCR	0.7956	0.5205	0.5908	0.7953	0.7954
	LCR					
	LCR-NCR					
CiaoDVD	BTL	0.8009	0.5230	0.6016	0.7903	0.8008
	BTL-NCR	0.9394	0.5857	0.8787	0.9291	0.9393
	BPR	0.7241	0.4883	0.4481	0.7240	0.7240
	BPR-NCR	0.9537	0.5922	0.9074	0.9537	0.9537
	FSBPR	0.7501	0.5001	0.5004	0.7401	0.7502
	FSBPR-NCR	0.8906	0.5637	0.7815	0.8806	0.8908
	LCR					
	LCR-NCR					

Ranking Performance for Graded Sessional Feedback

Data Sets In our model the prominent aspect is sampled for each evaluation session. In the previous experiments, the evaluation sessions are considered to be associated with each rating/click action. However, when the user interaction session information is available, the definition of evaluation session is different. We use two real world datasets with user-item interaction sessions. (1) Tmall (2) FilmTrust (3) .

Table 5: Statistics of Datasets with graded sessional feedback

Dataset	#users	#items	#ratings	#sessions
Tmall-single	33815	176231		364844
Tmall-hybrid	62101	198344		475503
Yoochose	1	30852		341396

Comparative Methods. We compare the NCR improved versions with the original versions on three widely adopted rating prediction methods (1) MF (Koren, Bell, and Volinsky 2009): standard matrix factorization (2) AMF (Koren 2008): neighborhood factorization (3) LLORMA (Lee et al. 2013): local low-rank matrix factorization.

Evaluation Metrics. The goal is to reconstruct the observed user-item ratings as accurate as possible. Hence we evaluate different approaches based on the following metrics. (1) AUC: (2) NDCG (3) RMSE (4) MAE (5) MRR

Table 6: Comparative ranking prediction performance with graded sessional feedback

Dataset	Method	MAP	NDCG	Prec	Recall	MRR
Tmall-single	BTL	0.4348	0.2814	0.2787	0.7253	0.4890
	BTL-NCR	0.4408	0.2853	0.2810	0.7308	0.4973
	BPR	0.4359	0.2826	0.2789	0.7252	0.4932
	BPR-NCR	0.4410	0.2854	0.2810	0.7305	0.4977
	FSBPR	0.4163	0.2732	0.2734	0.7092	0.4717
	FSBPR-NCR	0.4193	0.2747	0.2749	0.7130	0.4740
	LCR					
	LCR-NCR					
Tmall-hybrid	BTL	0.5015	0.3056	0.2934	0.7931	0.5458
	BTL-NCR	0.5592	0.3305	0.3044	0.8249	0.6063
	BPR	0.5463	0.3248	0.3006	0.8132	0.5950
	BPR-NCR	0.5635	0.3324	0.3050	0.8267	0.6112
	FSBPR	0.4398	0.2770	0.2768	0.7431	0.4817
	FSBPR-NCR	0.4597	0.2865	0.2831	0.7624	0.5007
	LCR					
	LCR-NCR					
Yoochoose	BTL	0.6368	0.4742	0.4569	0.8732	0.7156
	BTL-NCR	0.7112	0.5166	0.4786	0.8966	0.7882
	BPR	0.6821	0.5019	0.4711	0.8934	0.7639
	BPR-NCR	0.7049	0.5144	0.4775	0.9030	0.7844
	FSBPR	0.5685	0.4379	0.4374	0.8405	0.6541
	FSBPR-NCR	0.6825	0.5445	0.5362	0.8839	0.7987
	LCR					
	LCR-NCR					

Effect of Non-compensatory Rules

Related Work

Conclusion

References

Devooght, R.; Kourtellis, N.; and Mantrach, A. 2015. Dynamic matrix factorization with priors on unknown values. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, 189–198. New York, NY, USA: ACM.

Engel, J. F.; Blackwell, R. D.; and Miniard, P. W. 1986. *Consumer Behavior*. The Dryden Press.

Hu, J., and Li, P. 2016. Improved and scalable bradley-terry model for collaborative ranking. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, 949–954.

Hu, J., and Li, P. 2017. Decoupled collaborative ranking. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, 1321–1329. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee.

Hunter, D. R. 2004. Mm algorithms for generalized bradley-terry models. *The Annals of Statistics* 32(1):384–406.

Koren, Y.; Bell, R.; and Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *Computer* 42(8):30–37.

Koren, Y. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 426–434. ACM.

Koren, Y. 2010. Factor in the neighbors: Scalable and accurate collaborative filtering. *ACM Transactions on Knowledge Discovery from Data.* 4:1:1–1:24.

Lee, J.; Kim, S.; Lebanon, G.; and Singer, Y. 2013. Local low-rank matrix approximation. In *International Conference on Machine Learning*, 82–90.

Lee, J.; Bengio, S.; Kim, S.; Lebanon, G.; and Singer, Y. 2014. Local collaborative ranking. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, 85–96. New York, NY, USA: ACM.

Lee, I. 2009. *Transforming E-Business Practices and Applications: Emerging Technologies and Concepts: Emerging Technologies and Concepts*. IGI Global.

Lerche, L., and Jannach, D. 2014. Using graded implicit feedback for bayesian personalized ranking. In *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys '14, 353–356. New York, NY, USA: ACM.

Pilászy, I.; Zibriczky, D.; and Tikk, D. 2010. Fast als-based matrix factorization for explicit and implicit feedback datasets. In *Proceedings of the fourth ACM conference on Recommender systems*, RecSys '10, 71–78. New York, NY, USA: ACM.

Rendle, S.; Freudenthaler, C.; Gantner, Z.; and Schmidt-Thieme, L. 2009. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, 452–461. Arlington, Virginia, United States: AUAI Press.

Salakhutdinov, R., and Mnih, A. 2008. Probabilistic matrix factorization. *Advances in neural information processing systems* 20:1257–1264.

Shi, Y.; Larson, M.; and Hanjalic, A. 2010. List-wise learning to rank with matrix factorization for collaborative filtering. In *Proceedings of the Fourth ACM Conference on Recommender Systems*, RecSys '10, 269–272. New York, NY, USA: ACM.

Steck, H. 2015. Gaussian ranking by matrix factorization. In *Proceedings of the 9th ACM Conference on Recommender Systems*, RecSys '15, 115–122. New York, NY, USA: ACM.

Thurstone, L. L. 1927. A law of comparative judgment. *Psychological review* 34(4):273.

Weston, J.; Bengio, S.; and Usunier, N. 2011. Wsabie: Scaling up to large vocabulary image annotation. In *IJCAI*, volume 11, 2764–2770.

Yu, L.; Zhou, G.; Zhang, C.; Huang, J.; Liu, C.; and Zhang, Z.-K. 2016. Rankmbpr: Rank-aware mutual bayesian personalized ranking for item recommendation. In Cui, B.; Zhang, N.; Xu, J.; Lian, X.; and Liu, D., eds., *Web-Age Information Management*, 244–256. Cham: Springer International Publishing.

Zhang, W.; Chen, T.; Wang, J.; and Yu, Y. 2013. Optimizing top-n collaborative filtering via dynamic negative item sampling. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '13, 785–788. New York, NY, USA: ACM.

Zhao, Q.; Zhang, Y.; Ma, J.; and Duan, Q. 2018. Factored item similarity and bayesian personalized ranking for recommendation with implicit feedback. *Arabian Journal for Science and Engineering*.